ARIA
APPLIED RESEARCH IN ACTION

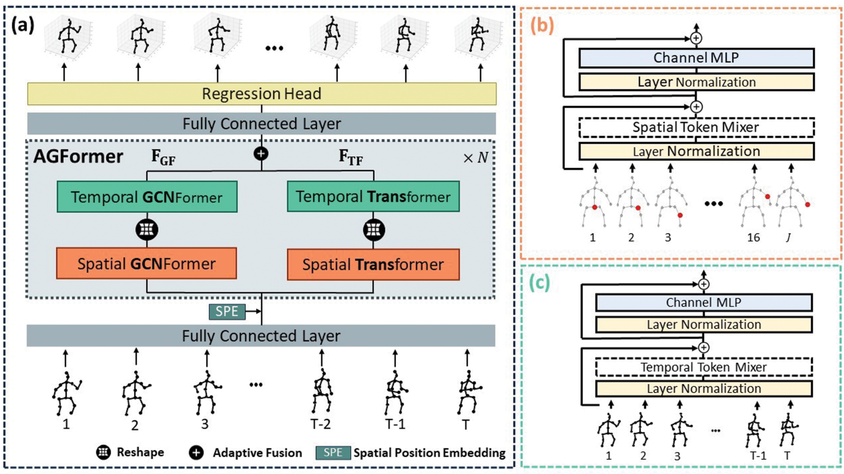
MotionAGFormer: Enhancing 3D Human Pose Estimation with a Transformer-GCNFormer Network

A model with less parameters and computation compared to SOTA models while being more accurate in 3D Human pose estimation

Soroush Mehraban, Vida Adeli, Babak Taati

Prof. Lueder Kahrs
ACADEMIC SUPERVISOR

Vicky Yu
INDUSTRY SUPERVISOR



PROJECT SUMMARY

Recent transformer-based approaches have demonstrated excellent performance in 3D human pose estimation. However, they have a holistic view and by encoding global relationships between all the joints, they do not capture the local dependencies precisely. In this project, we present a novel Attention-GCNFormer (AGFormer) block that divides the number of channels by using two parallel transformer and GCNFormer streams. Our proposed GCNFormer module exploits the local relationship between adjacent joints, outputting a new representation that is complementary to the transformer output. By fusing these two representation in an adaptive way, AGFormer exhibits the ability to better learn the underlying 3D structure. By stacking multiple AGFormer blocks, we propose MotionAGFormer in four different variants, which can be chosen based on the speed-accuracy trade-off. We evaluate our model on two popular benchmark datasets: Human3.6M and MPI-INF-3DHP. MotionAGFormer-B achieves state-of-the-art results, with P1 errors of 38.4mm and 16.2mm, respectively. Remarkably, it uses a quarter of the parameters and is three times more computationally efficient than the previous leading model on Human3.6M dataset.

IATSL: INTELLIGENT ASSISTIVE TECHNOLOGY AND SYSTEMS LAB

Computer Science
UNIVERSITY OF TORONTO