

Attention-Driven Body Pose Encoding for Human Activity Recognition

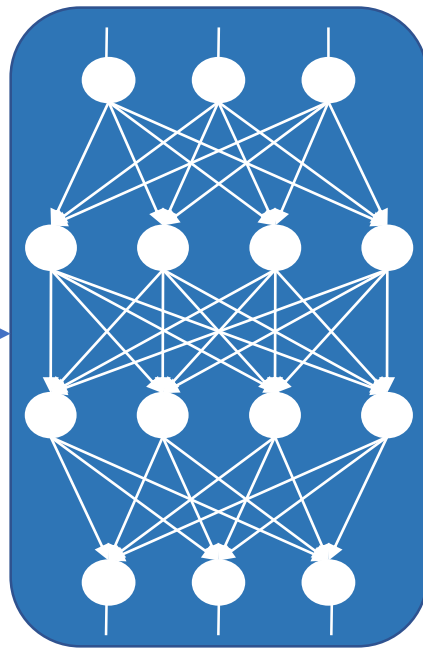
Bappaditya Debnath*, Mary O'Brien†, Swagat Kumar* and Ardhendu Behera**

Department of Computer Science, †Faculty of Health, Social Care & Medicine

Edge Hill University, Ormskirk, UK, L394QP

Email:{debnathb, obrienm, kumars, beheraa}@edgehill.ac.uk





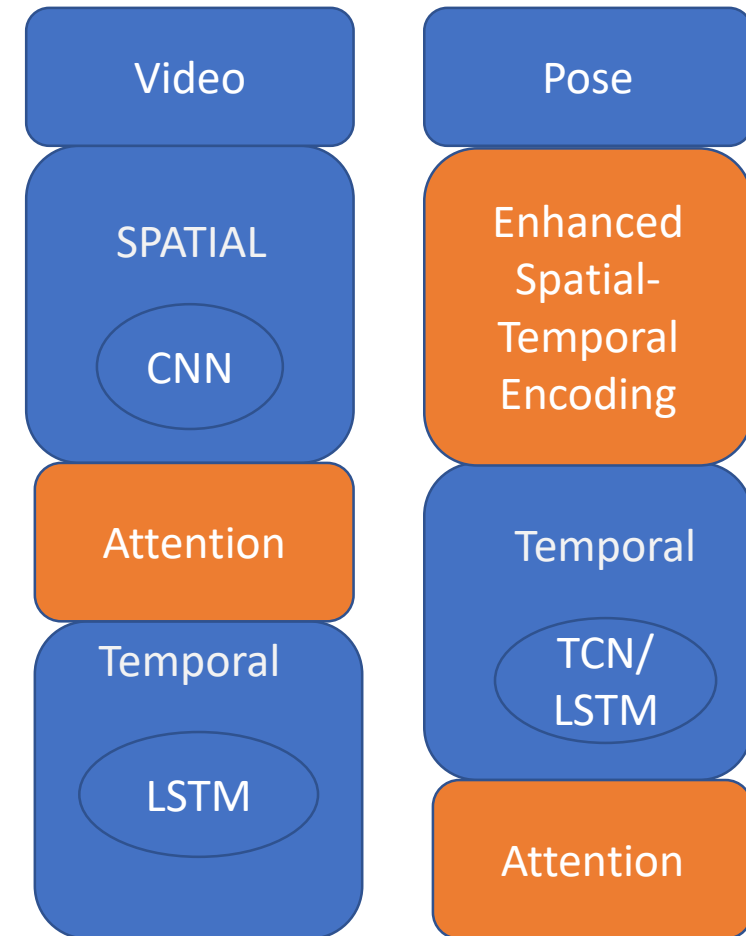
Drinking
Sitting
Standing
Brushing Hair
Bending
Answering Phone
Wearing Glasses
Brushing Floor
Reaching Above
Clapping

- Aim: Detect human activity from videos
- Potential Applications:
 - Home-Based Rehabilitation,
 - Surveillance,
 - Human-Robot Interaction

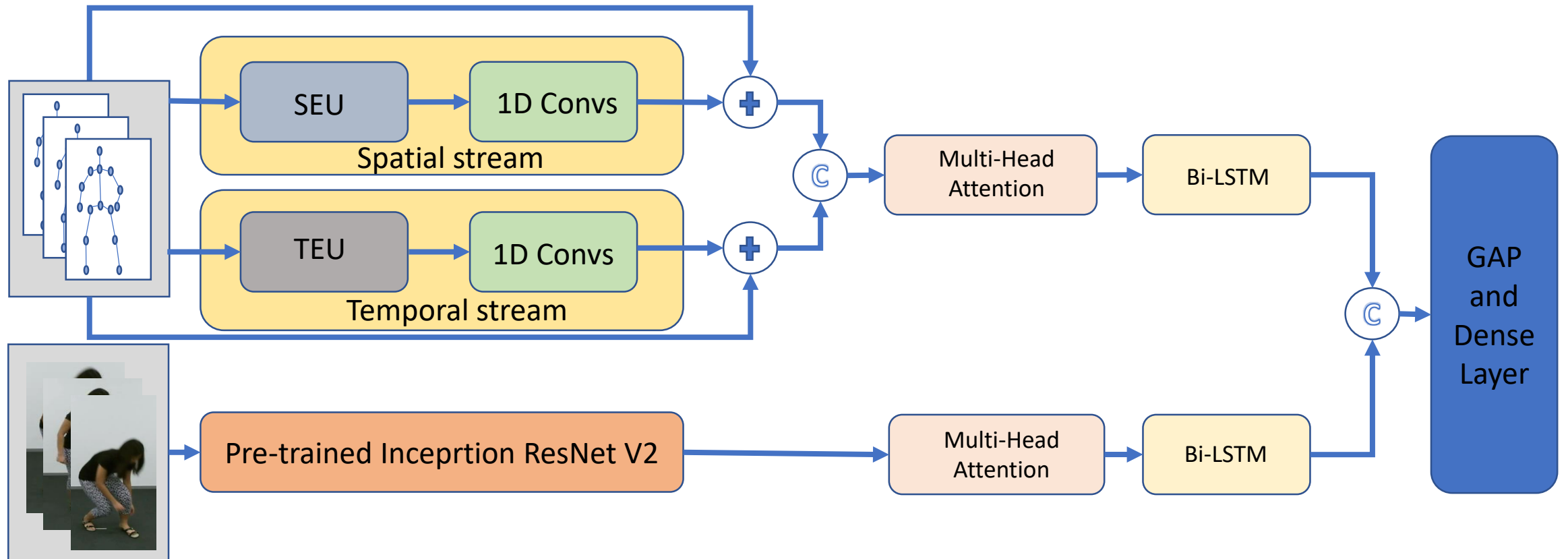


Proposed Approach

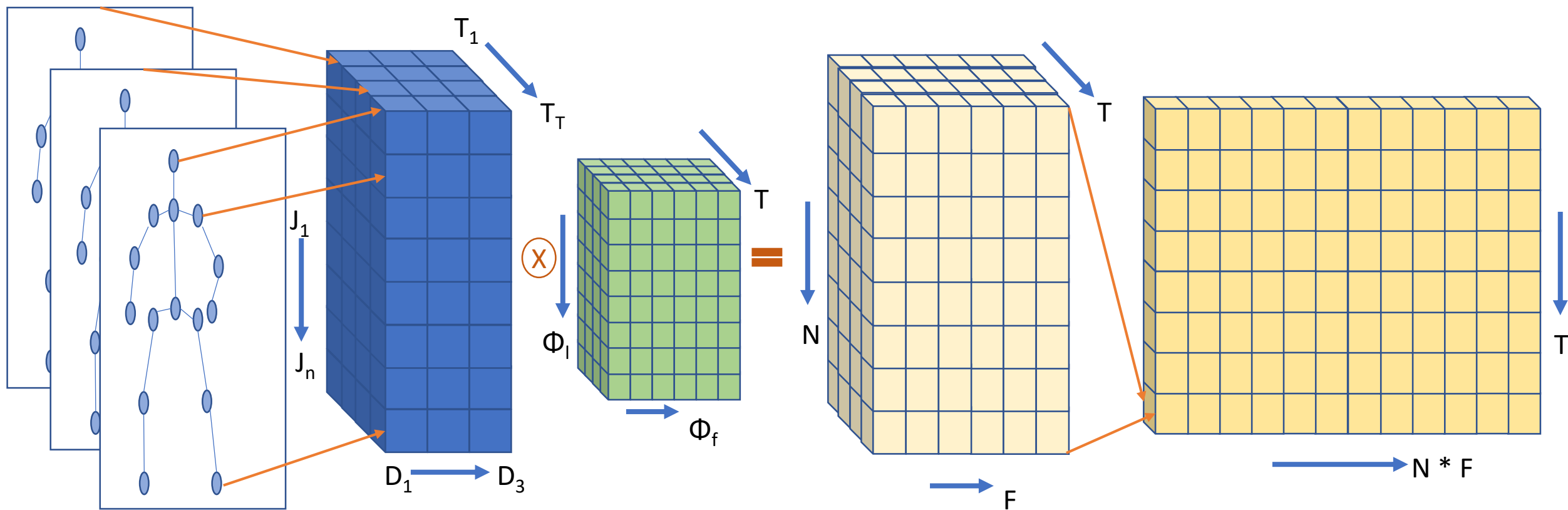
- Multimodal:
 - Monocular video data
 - Human body-pose data
- Philosophy:
 - Learn the spatial and temporal relationships between various body-joints for enhanced body-pose representation



Network Diagram



Spatial Encoding Unit

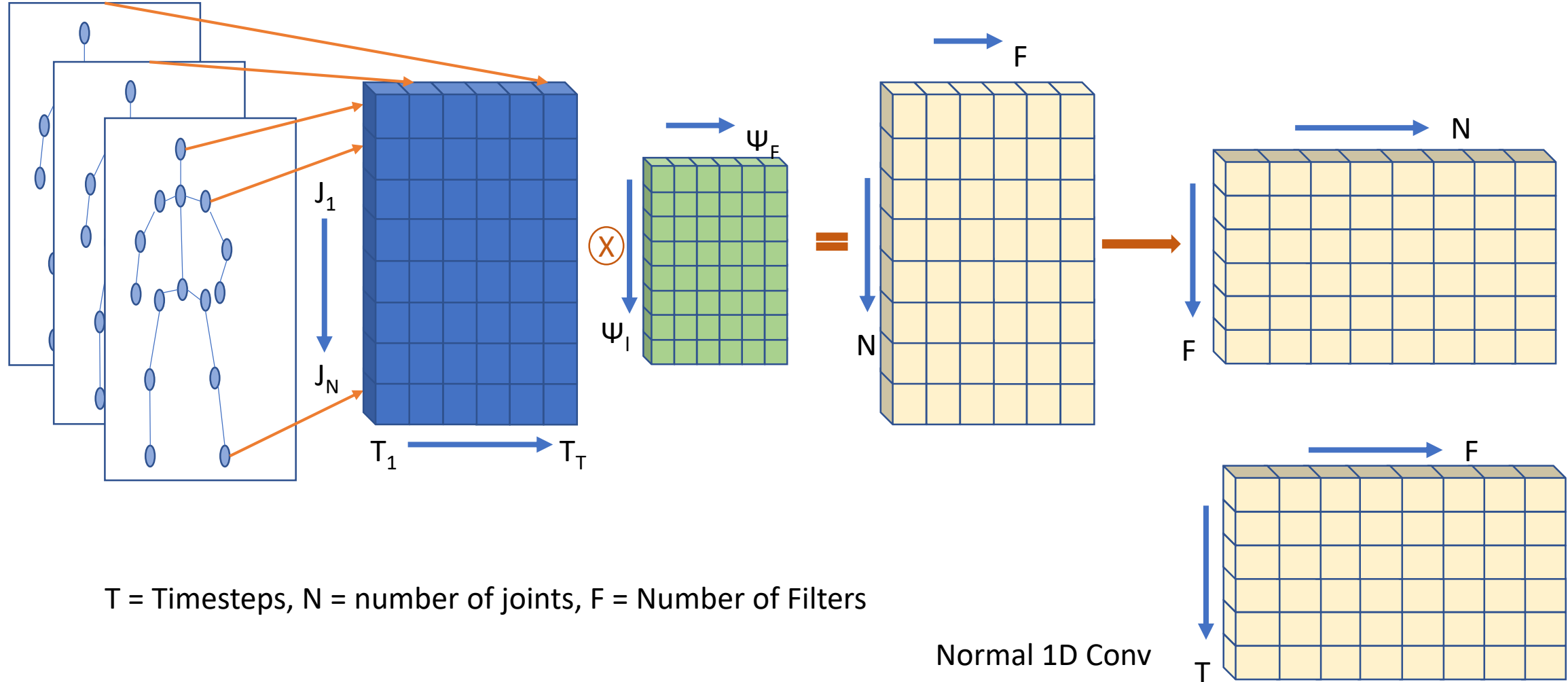


T = Timesteps, N = number of joints, F = Number of Filters

Normal 1D Conv = $T * F$



Temporal Encoding Unit

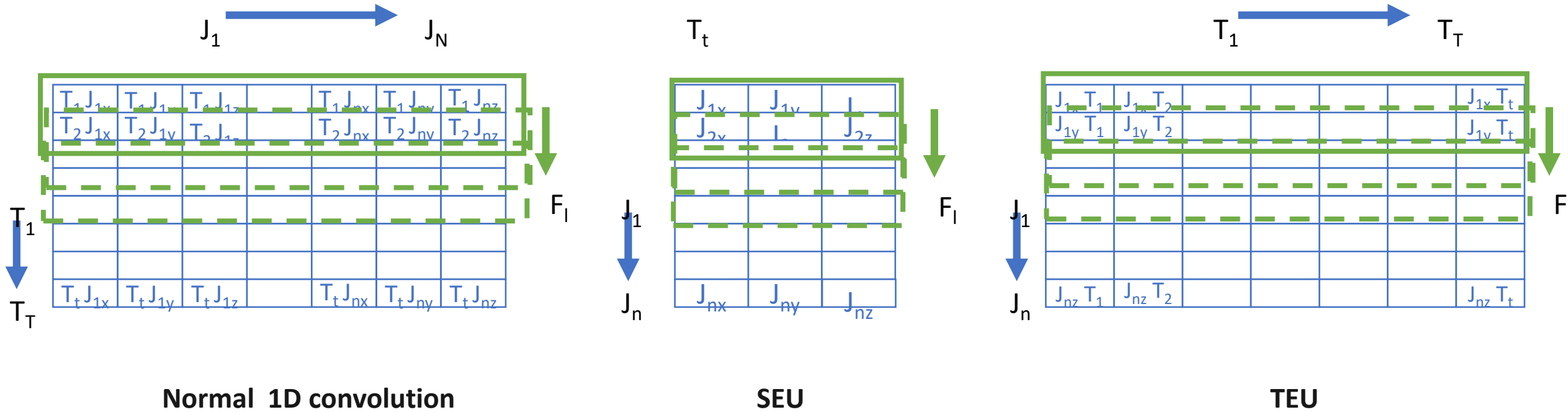


T = Timesteps, N = number of joints, F = Number of Filters

Normal 1D Conv



SEU and TEU Analysis



- Further SEU and TEU analysis
- Extensive ablation studies



QUESTIONS?

THANK YOU

