# Hybrid Sparse GCN + Attention Routing Transformer for Real-Time 3D Human Pose Estimation

## Introduction

This project proposes a real-time 3D human pose detection system using a hybrid architecture combining Sparse Graph Convolutional Networks (GCNs) with Attention Routing Transformers. The model is optimized for accuracy, real-time inference, and research extensibility.

## Model Architecture

The architecture processes 2D keypoints (from MediaPipe) as input and outputs 3D joint positions:  
Input → SparseGCN → Attention Routing Transformer → MLP → Output 3D joints (28).

## Components Explained

Sparse GCN Layer:  
- Operates on a sparse, learnable graph of joints using PyTorch Geometric's GCNConv.  
  
Attention Routing Transformer:  
- Multi-head attention over joints enables selective attention to relevant joints only.  
  
MLP Head:  
- Regresses the final 3D joint positions from transformer output.

## Why This Model is Significant

- Sparse graphs reduce memory and computation.  
- Joint-level attention dynamically adapts per frame.  
- Works per-frame: ideal for real-time.  
- Easy to extend or analyze for research.

## Novelty

Compared to works like PoseFormer and SemGCN, this model:  
- Does not rely on fixed joint connectivity.  
- Uses routing-based attention instead of full self-attention.  
- Requires only single-frame input, making it suitable for edge devices.

## Training

Dataset: mpi\_inf\_combined.npz (converted from MPI-INF-3DHP)  
Batch size: 64  
Optimizer: Adam (lr=1e-3)  
Loss: MSE  
Epochs: 10  
Run: python train\_pose\_model.py

## Real-Time Inference

The pipeline uses MediaPipe to extract 2D keypoints and overlays the predicted 3D joints live on the webcam feed.  
Run: python predict\_live.py

## System Requirements

Python 3.12.9  
CUDA 12.8  
PyTorch Geometric, MediaPipe, OpenCV, Matplotlib

## Key Files

- model.py: Core model architecture  
- train\_pose\_model.py: Training pipeline  
- predict\_live.py: Live webcam demo  
- visualize.py: Drawing the 3D skeleton  
- extract\_mpi\_inf\_3dhp.py: Dataset conversion utility

## Planned Extensions

- Add MPJPE/P-MPJPE metrics  
- Learn top-K adjacency per joint  
- Export to ONNX/TensorRT  
- Expand to gesture classification

## Citation

Vedansh Tembhre, 2025. "Hybrid Sparse Graph and Transformer Architecture for Real-Time 3D Human Pose Estimation."