November 09, 2024 Thesis Proposal Defense

Skeleton-based Human Action Recognition Based on modified Graph Convolution Network

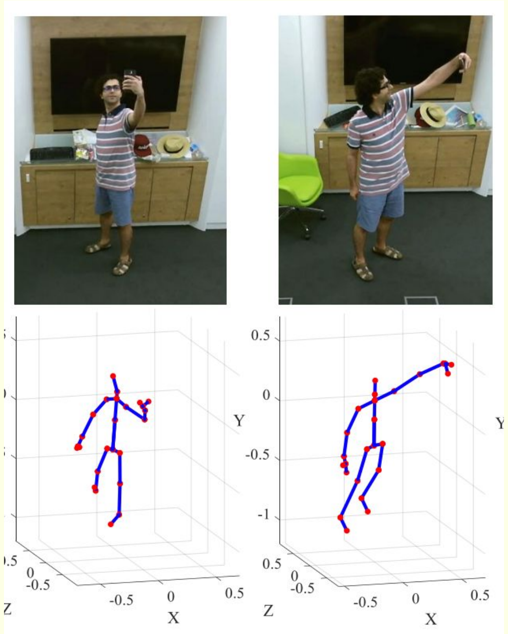
Samihah Binte Sharif [C211250]
Takia Sultana Nova [C211219]
Kazi Taspia Tahlin [C211246]

Department of Computer Science and Engineering
International Islamic University Chittagong

Supervisor: Md. Khaliluzzaman

Why Skeleton-based Action Recognition?

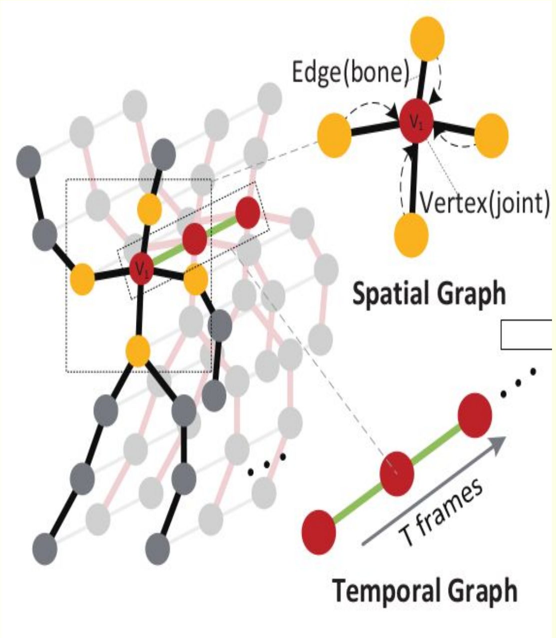
- Robust Against:
    - ➔ Complex Backgrounds
    - ➔ Camera View Changes
- Represents natural human body topology
- Helps models focus on semantically relevant features
- Lower computational cost



Graph Convolutional Networks(GCN)

- Skeletons can be represented as graphs where:
  - ❑ Joints -> vertices
  - ❑ natural topology (bones) between adjacent joints -> spatial edges
  - ❑ temporal correlations of adjacent frames -> temporal edges

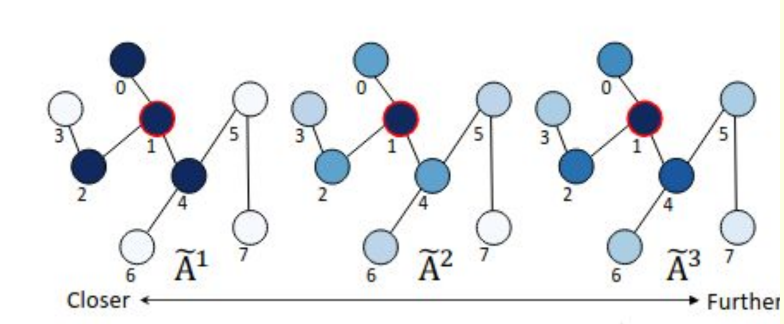
- GCNs can effectively model these spatial relationships.



Problem Statement

Challenge of Modeling Long-Range Dependencies:

- Over-smoothing issue
- Indirect dependencies

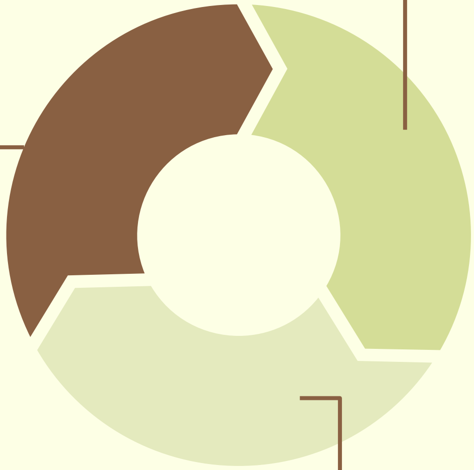


Temporal Redundancy in Skeleton Frames:

- Keyframe Selection
- Loss of Semantic Focus

Research Objectives

To model long-range dependencies more accurately between joints



To improve temporal robustness by focusing on key, semantics-related poses

To design an efficient, scalable model that balances accuracy with computational cost

Literature Review

Ref.	Findings	Dataset	Methods used	Accuracy				Limitations
				Top-1	**Top-5**	**X-Sub**	**X-View**	
[1]	Introduces **ST-GCN**, a model that applies graph convolutional networks (GCNs) to human skeleton data for action recognition	Kinetics	Graph convolution, Temporal convolution, Specialized convolution kernels	30.7%	52.8%			Fixed Skeleton Graph, Lack of Flexibility Across Layers, Inflexibility for Different Actions
		NTU-RGB+D				81.5 %	88.3%	

Literature Review

Ref.	Findings	Dataset	Methods used	Accuracy				Limitations
				Top-1	**Top-5**	**X-Sub**	**X-View**	
[2]	**2s-AGCN** allows the model to adaptively learn graph topologies rather than relying on fixed, manually set ones.	Kinetics - Skeleton	Adaptive Graph Convolution, Two-Stream Framework	36.1%	58.7%			Increased Computational Cost, Potential Overfitting, Limited Testing on Unseen Actions
		NTU-RGB+D				88.5 %	95.1%	

Literature Review

Ref.	Findings	Dataset	Methods used	Accuracy				Limitations
				Top-1	**Top-5**	**X-Sub**	**X-View**	
[3]	**Disentangles and unifies graph convolutions** to more effectively capture diverse spatial and temporal dependencies	Kinetics Skeleton 400	Disentangled Graph Convolutions, Unified Graph Convolution, Hierarchical Aggregation	38.0%	60.9%			Complexity in Model Design, Higher Computational Demand, Limited Adaptability to Unseen Skeleton Variants
		NTU-RGB+D 120				86.9 %	88.4%	
		NTU-RGB+D 60				91.5 %	96.2%	

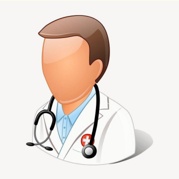
Literature Review

Ref.	Findings	Dataset	Methods used	Accuracy				Limitations
				Top-1	**Top-5**	**X-Sub**	**X-View**	
[4]	**AAM-GCN** solves the problem of over-smoothing, automatically increasing the number of convolution kernels with the increase of the complexity of the topological structure of graphs	Kinetics - Skeleton	Attention Adjacency Matrix , Dimension Attention Mechanism, Spatial-Temporal Convolution Blocks	37.5%	60.5%			Limited generalizability due to primary testing on large-scale datasets, Increased computational cost
		NTU-RGB+D				90.4 %	96.2%	

Methodology

1. Problem Formulation
2. Data Preprocessing
3. Graph Construction
4. Model Architecture
5. Receptive Field Expansion
6. Implementation
7. Training Strategy
8. Evaluation

Applications



Healthcare & Rehabilitation

Virtual Reality (VR)

Surveillance Systems

Human-Robot Interaction

Sports Analysis

Entertainment & Gaming

Expected Contribution and Outcome

Improved Accuracy

Enhance skeleton-based action recognition by effectively capturing long-range dependencies

Benchmark Performance

Expected to outperform existing models on standard datasets like NTU-RGBD and Kinetics-Skeleton.

Practical Applications

The model's efficiency and robustness will enable its use in real-world applications like virtual reality, robotics.

Expected Contribution and Outcome

This work will provide a foundation for future developments in human action recognition.

Thesis Significance

Enhanced Action Recognition

This thesis advances skeleton-based action recognition by addressing key gaps, improving accuracy, and robustness in challenging conditions.

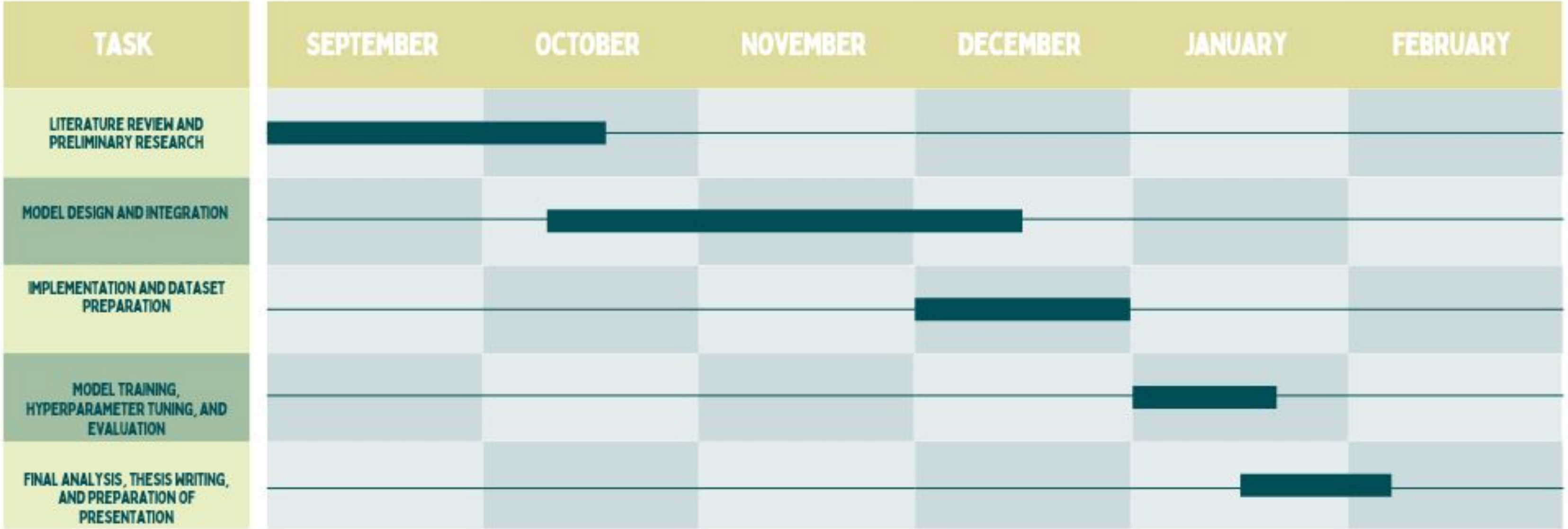
Broad Applicability

The model can benefit fields like VR, healthcare, robotics, and surveillance, where precise movement detection is essential.

Efficiency and Scalability

The balanced design ensures the model is both accurate and resource-efficient, suitable for various environments.

GANTT CHART



TASK
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
JANUARY
FEBRUARY
LITERATURE REVIEW AND PRELIMINARY RESEARCH
MODEL DESIGN AND INTEGRATION
IMPLEMENTATION AND DATASET PREPARATION
MODEL TRAINING, HYPERPARAMETER TUNING, AND EVALUATION
FINAL ANALYSIS, THESIS WRITING, AND PREPARATION OF PRESENTATION

Conclusion

Research Focus
Developed adaptive strategies for enhanced skeleton-based action recognition.

Expected Outcomes

- Improved recognition accuracy and robustness across datasets.
- Reduced computational costs while maintaining model adaptability.

Potential Impact
Applications in fields like virtual reality, robotics, and abnormal behavior detection.

References

[1] S. Yan, Y. Xiong, and D. Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition, 2018. 1, 3, 7, 8

[2] L. Shi, Y. Zhang, J. Cheng, and H. Lu. Adaptive spectral graph convolutional networks for skeleton-based action recognition. CoRR, abs/1805.07694, 2018. 1, 3, 5, 6, 7, 8

[3] Z. Liu, H. Zhang, Z. Chen, Z. Wang, and W. Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 140–149, 2020. doi: 10.1109/CVPR42600.2020.00022 1, 2, 3, 7

[4] B. Jxa, B. Qma, B. Rla, B. Wxa, T. C. Lei, Z. C. Sheng, and D. Xg. Attention adjacency matrix based graph convolutional networks for skeleton-based action recognition - sciencedirect. Neurocomputing, 440:230–239, 2021. 3, 6, 7

Thank you