

3D POINT CLOUD DATA ANALYSIS FOR
ENHANCED HUMAN ACTIVITY RECOGNITION

WANG RUYA

UNIVERSITI KEBANGSAAN MALAYSIA

3D POINT CLOUD DATA ANALYSIS FOR ENHANCED HUMAN ACTIVITY
RECOGNITION

WANG RUYA

PROJECT SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF
MASTER OF DATA SCIENCE

FACULTY OF INFORMATION SCIENCE AND TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2025

3D POINT CLOUD DATA ANALYSIS FOR ENHANCED HUMAN ACTIVITY
RECOGNITION

WANG RUYA

PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT UNTUK MEMPEROLEH IJAZAH SARJANA SAINS
DATA

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI
2025

DECLARATION

I hereby declare that the work in this project is my own except for quotations and summaries which have been duly acknowledged.

I have used a AI for 10% of the work related to LR in preparing this report.

24 January 2025

WANGRUYA
P139574

ACKNOWLEDGEMENT

Time flies, my one-year graduate student so quickly passed. I still remember the discomfort I felt when I first came to a foreign country and the stress of facing a new environment, but now I have completely adapted to the life here and even want to live in Malaysia. I am very happy that I have met many interesting people and learned a lot of knowledge through this one-year master's degree.

I would like to thank my tutor for her careful guidance in my study, as well as her concern and care for me in daily life, so that I can adapt to my study and life in Malaysia faster and better.

I would like to thank my friend Shen Ruihan. If it were not for her, I would not have had the opportunity to learn about studying in Malaysia. She helped me adapt to the life in Malaysia quickly while taking care of my own study, so that I could spend my master's life more happily.

I would like to feel grateful to my parents for providing me with such an opportunity to study in Malaysia for my master's degree. I wish them a smooth and healthy New Year.

Finally, I would like to thank myself for shortening the learning time of one and a half years to one year and successfully writing my graduation thesis. May my life go smoothly in the future.

ABSTRAK

Pengecaman pergerakan manusia (HAR) menggunakan data awan titik 3D telah menjadi bidang penyelidikan yang penting, dengan keperluan mendesak untuk analisis pergerakan manusia yang tepat dalam banyak bidang seperti penjagaan kesihatan, pemantauan pintar, dan interaksi manusia-komputer. Kerja ini bertujuan untuk meningkatkan keberkesanan HAR dengan menggabungkan PointMapNet, yang cemerlang dalam pengekstrakan ciri tempatan, dengan GCN, yang memodelkan kebergantungan global. Tiga set data penanda aras — NTU RGB+D, NTU RGB+D 120, dan Kinect — telah dipilih, merangkumi data rangka 3D yang kaya dan pelbagai sebagai asas untuk menilai prestasi sistem pengecaman pergerakan manusia. Semasa prapemprosesan, kualiti data input dipastikan melalui penyahnotaan, pembersihan, dan penormalan. Model yang dicadangkan menyepadukan keupayaan perwakilan ciri spatial tempatan PointMapNet dengan pemodelan perhubungan patio-temporal global GCN untuk mencipta rangka kerja HAR hibrid yang cekap. Keputusan menunjukkan bahawa model PointMapNet+GCN bersepadu mengatasi GCN tradisional dalam semua set data, mencapai kadar ketepatan klasifikasi sebanyak 90% pada Kinect, 88% pada NTU RGB+D, dan 90% pada NTU RGB+D 120. Ini menunjukkan ketepatan dan ingatan semula yang tinggi, terutamanya dalam membezakan perbezaan halus dalam postur dan pergerakan manusia. Selain itu, model ini menangani cabaran seperti oklusi data, maklumat yang hilang, dan pertindihan kategori dengan berkesan, menunjukkan kebolehpercayaan dalam aplikasi praktikal. Kajian ini memperkenalkan pendekatan hibrid yang baru, menonjolkan faedah menggabungkan pemodelan ciri tempatan dan global untuk HAR. Penemuan ini mencadangkan bahawa rangka kerja tersebut mempunyai potensi untuk aplikasi masa nyata dan kebolehskalaan kepada set data yang pelbagai. Arah masa hadapan termasuk memajukan gabungan data berbilang mod, meningkatkan generalisasi silang set data, dan mengoptimumkan prestasi pengkomputeran kelebihan. Hasil ini meletakkan asas yang kukuh untuk inovasi dalam pemantauan perubatan, realiti maya, dan penyelesaian bandar pintar, serta menyediakan peta jalan untuk sistem pengecaman pergerakan manusia generasi akan datang.

ABSTRACT

Human movement recognition (HAR) using 3D point cloud data has become an important research area, and there is an urgent need for accurate human movement analysis in many fields such as healthcare, intelligent monitoring, and human-computer interaction. This work is dedicated to improving HAR effectiveness by combining PointMapNet, which excels at local feature extraction, with GCN, which models global dependencies. Three benchmark datasets—NTU RGB+D, NTU RGB+D 120, and Kinect—were selected, encompassing rich and diverse 3D skeleton data to form the foundation for evaluating human motion recognition system performance. During preprocessing, input data quality was ensured through denoising, cleaning, and normalization. The proposed model integrates PointMapNet's local spatial feature representation capability with GCN's global spatio-temporal relationship modelling to create an efficient hybrid HAR framework. Results demonstrated that the integrated PointMapNet+GCN model outperformed traditional GCN across all datasets, achieving classification accuracy rates of 90% on Kinect, 88% on NTU RGB+D, and 90% on NTU RGB+D 120. It showed high accuracy and recall rates, particularly in distinguishing subtle differences in human posture and movement. Moreover, the model effectively addressed challenges like data occlusion, missing information, and category overlap, demonstrating reliability in practical applications. This study introduces a novel hybrid approach, highlighting the benefits of combining local and global feature modelling for HAR. Findings suggest the framework has real-time application potential and scalability to diverse datasets. Future directions include advancing multimodal data fusion, enhancing cross-dataset generalization, and optimizing edge computing performance. These outcomes lay a solid foundation for innovations in medical monitoring, virtual reality, and smart city solutions, providing a roadmap for next-generation human movement recognition systems.

TABLE OF CONTENTS

	Page
DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRAK	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF ILLUSTRATIONS	x
LIST OF ABBREVIATIONS	xii
CHAPTER I INTRODUCTION	
1.1 Research Background	1
1.2 Problem Statement	2
1.3 Research Objectives	3
1.4 Research Scopes	3
1.5 Research Contribution	4
1.6 Summary	4
CHAPTER II LITERATURE REVIEW	
2.1 Introduction	6
2.2 3D Point Cloud Data Analysis on Human Activity Recognition	6
2.3 3D Point Cloud Data	13
2.3.1 NTU RGB+D and NTU RGB+D 120	15
2.3.2 Kinect DATA	18
2.4 Graph Convolution Network	20
2.5 Summary	21
CHAPTER III METHODOLOGY	
3.1 Introduction	22
3.2 Datasets	24
3.2.1 NTU RGB+D dataset	24

	3.2.2	NTU RGB+D 120 dataset	24
	3.2.3	The Kinect dataset	25
3.3		Data Preparation	26
	3.3.1	Pointmapnet	26
	3.3.2	Data Processing	28
3.1		Classification	31
3.4		Preformance Evaluation	33
3.5		Summary	36
CHAPTER IV RESULTS AND DISCUSSION			
4.1		Introduction	37
4.2		Experimental Process Results	37
	4.2.1	Data Preprocessing	38
	4.2.2	Pointmapnet	39
	4.2.3	Experimental Resluts	41
4.3		Comparison Results	52
	4.3.1	Epoch Comparison for NTU RGB+D	52
	4.3.2	Method Conparison	56
4.4		Summary	57
CHAPTER V RESULTS AND DISCUSSION			
5.1		Conclusion	59
5.2		Contributions	60
5.3		Future Works	61
5.4		Final Remarks	62
	5.4.1	Challenges and Future Directions	62
	5.4.2	Technological Evolution and Trend	63
5.5		Summary	63
REFERENCES			64

LIST OF TABLES

Table No.		Page
Table 2.1	3D point cloud data analysis on human activity recognition	7
Table 2.2	First table in specific objective for chapter II	14
Table 3.1	Size of “NTU RGB+D” and “NTU RGB+D 120” dataset	24
Table 3.2	Size of the Kinect dataset	25
Table 4.1	Compare he performance of the two methods	56

LIST OF ILLUSTRATIONS

Figure No.		Page
Figure 3.1	General framework of proposed study	23
Figure 3.2	Sample frames of “NTU RGB+D” (Shahroudy et al.,2016) and “NTU RGB+D 120” (Liu et al., 2019) dataset	25
Figure 3.3	Sample frames of Kinect (Xia et al., 2012) dataset	26
Figure 3.4	Data processing	28
Figure 3.5	Preprocessing of NTU RGB+D and NTU RGB+D 120 datasets	29
Figure 3.6	Preprocessing of Kinect datasets	30
Figure 3.7	Classification	31
Figure 3.8	GCN STRUCTURE (Kipf & Welling, 2016)	32
Figure 4.1	Preprocessing of NTU RGB+D datasets	38
Figure 4.2	Preprocessing of NTU RGB+D 120 datasets	39
Figure 4.3	NTU RGB+D image	40
Figure 4.4	NTU RGB+D 120 image	41
Figure 4.5	Kinect image	41
Figure 4.6	NTURGB +D-Training Loss Curve	43
Figure 4.7	NTU RGB +D- Model Evaluation Metrics	44
Figure 4.8	NTU RGB +D-Confusion Matrix	45
Figure 4.9	NTURGB +D 120-Training Loss Curve	46
Figure 4.10	NTU RGB +D 120-Model Evaluation Metrics	47
Figure 4.11	NTU RGB +D120 -Confusion Matrix	48
Figure 4.12	Kinect Data-Training Loss Curve	49
Figure 4.13	Kinect Data- Model Evaluation Metrics	50
Figure 4.14	Kinect Data- Confusion Matrix	51

Figure 4.15	400 epochs	52
Figure 4.16	100 epochs	52

LIBRARY FTSM

LIST OF ABBREVIATIONS

GCN	Graph Convolutional Network
HAR	Human Activity Recognition
UKM	Universiti Kebangsaan Malaysia

LIBRARY FTSM

CHAPTER I

INTRODUCTION

1.1 RESEARCH BACKGROUND

With the development and progress of science and technology, human motion recognition has also developed rapidly and has provided convenience for our daily lives, such as healthcare and rehabilitation, sports and fitness, safety and monitoring, autonomous driving, robotics and automation. But Vernikos et al. (2023) proposed that traditional human motion recognition methods often rely on 2D data sources, such as 2D video and image sequences, which can be limited by occlusion, light changes, and viewpoint changes. How to improve human motion recognition ability has become a difficult problem to solve for many researchers. In order to solve this problem, in recent years, scientists have developed more flexible and convenient sensors and imaging equipment, and three-dimensional point cloud data has gradually entered our field of vision.

Laser radar and depth cameras are the most representative sensors and imaging devices. Although the current movement has mostly captured a minimal distance, LiDAR transmits a laser beam to measure the reflection time; in this manner, it can acquire the three-dimensional coordinate information of an object, which is advantageous owing to its high precision and long-distance measurement. High-throughput motion features with high computational efficiency can be derived from the depth image provided by the depth camera and the accelerometer signal supplied by the inertial body sensor, which has the merits of large quantities, fast acquisition speed, and low cost. High-quality 3D point cloud data can be easily acquired through these advancements in sensor technology.

Y. Ben-Shabat et al. (2023) compared to traditional 2D video data, which loses the detail of the spatial layout of objects, the 3D point cloud is a data set with an enormous quantity of 3D coordinate points for the most accurate description of the shape in three dimensions and spatial position of object structures. For these purposes, there is evidence to suggest that through point cloud data, human movement recognition has significantly increased both in detailed data or deep information and in spatial relationships of some body parts. Thus, it has successfully enhanced the field of activity recognition and human identification, which further resulted in improved accuracy and robust Human Activity Recognition, HAR system.

1.2 PROBLEM STATEMENT

This limitation is being overcome in the current study by applying PointMapNet to 3D point cloud data to improve the accuracy of human activities. These are the problems of this research:

1. How to enhance the ability to extract data from 3D point cloud data:

GCN can help solve this problem. WenjieYang et al. (2021) proposed the GCN to strengthen the model's ability to capture more abundance. By modeling point cloud data as a graph structure, GCN can relate the relationship to the context, taking into account the connectivity and relative position between the relevant nodes. Can better represent spatial relationships in the data, to enhance the ability to extract data from 3D point cloud data.

2. How to simplify 3D point cloud data processing:

To date, most of these current point cloud sequence coding methods entail high processing and extra pose estimation algorithms to avert the repeatability of the computation. Li, Xing et al. (2023) proposed PointMapNet is going to come up with a framework that will make the encoding simpler compared to traditional depth and skeleton sequences based on point cloud sequences. This would make implementing the data processing pipeline much simpler and more effective without losing the richness of the data collected.

3. How to improve accuracy using 3D point cloud data:

The PointMapNet system will answer. Li, Xing, et al. (2023) proposed that traditional HAR systems rely heavily on 2D data, which is inaccurate because there is a lack of depth or spatial information. Using the point cloud feature maps for extracting appearance and motion clues that are intrinsic to 3D data, the problem can be solved with PointMapNet. A large number of studies predict that the human activities' identification accuracy with spatiotemporal details perfectly captured will be extremely helpful in improving the performance of the system for reliability and efficiency.

1.3 RESEARCH OBJECTIVES

The main purpose of this study is to use a new point cloud sequence network, PointMapNet, to advance the development of 3D human activity recognition (HAR). Hence, the main objectives of this thesis are:

1. To identify the data analysis techniques for 3d point cloud data for HAR.
2. To improve HAR accuracy by implementing PointMapNet architecture on three selected 3D point cloud datasets.
3. To evaluate the effectiveness of PointMapNet by classifying using Graph Convolution Network measuring performance using performance metrics etc.

1.4 RESEARCH SCOPES

The scope of this study includes the following aspects:

1. Select and process three different 3D point cloud datasets that cover multiple human activity types and scenarios.
2. Design and implement PointMapNet for feature extraction and human activity recognition of 3D point cloud data.
3. The GCN was used to classify the model and evaluate its performance in different scenarios.
4. The performance of PointMapNet and the traditional HAR method was compared and analyzed to verify its improvement effect.

1.5 RESEARCH CONTRIBUTION

This study contributes to the field of 3D HAR. The main contributions of this study are summarized below:

1. Introduce and use a new model PointMapNet to enhance 3D HAR:

This study introduces PointMapNet, a novel point cloud sequence network specifically designed for 3D HAR. PointMapNet utilizes the rich spatial and motion information inherent in 3D point cloud data to significantly improve the accuracy and robustness of the HAR system.

2. The combination of graph convolutional network (GCN) and PointMapNet is proposed:

PointMapNet is combined with GCN to enhance the model's ability to capture complex spatial relationships in 3D point cloud data. By modelling point cloud data as a graph structure, connectivity and relative location between points can be effectively linked, thus improving data extraction and representation of spatial relationships.

3. Comprehensive evaluation using three classical data sets:

The effectiveness of PointMapNet was evaluated using three well-known 3D point cloud datasets: NTU RGB+D, NTU RGB+D 120, and Kinect datasets. The classification accuracy is measured by experiments and performance indexes.

1.6 SUMMARY

Chapter1 outlines the background, motivation, objectives and scope of the research. This study proposes the main research objective of improving 3D human motion recognition (HAR) by using PointMapNet model, and describes the required data preprocessing steps and model training methods in detail. Based on NTU RGB+D, NTU RGB+ D120 and Kinect datasets, this study will provide a high-quality data basis for subsequent model training and evaluation through data cleaning, normalization and enhancement techniques. Chapter2 summarizes the existing 3D human motion

recognition techniques and methods. This study reviews the advantages and disadvantages of traditional methods and deep learning methods, especially the multimodal data fusion technology. This chapter also introduces the main datasets used, including the NTU RGB+D, NTU RGB+D 120 and Kinect datasets, emphasizing the importance of these datasets in motion recognition research. By summarizing the existing research, this study provides a theoretical basis for the design and implementation of PointMapNet model. Chapter3 introduces the design and implementation details of PointMapNet model. This study describes the architecture of the model in detail, including the design of the input layer, convolutional layer, graph Convolutional Network (GCN) layer, fusion layer and output layer. In addition, this chapter also discusses the specific steps of model training, including data preprocessing, hyperparameter setting, loss function selection and optimization algorithm. Through the detailed description of the PointMapNet model, this study laid the foundation for subsequent experiments and evaluations aimed at verifying its effectiveness and superiority in 3D human motion recognition.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

In order to deeply understand the latest progress and research trend of 3D point cloud data analysis in the field of human activity recognition and select 3 relevant important data sets, 30 high-impact articles in 3D point cloud data analysis and related extension fields were read, and their data sets, methods, algorithms, results, advantages, and limitations were summarized into a Table2. 1.

2.2 3D POINT CLOUD DATA ANALYSIS ON HUMAN ACTIVITY RECOGNITION

According to the table, it can be found that many literatures have successfully enhanced the research accuracy of human motion recognition by using 3D point cloud data. Among them, the integration of deep learning technology and three-dimensional point cloud data is also promoting significant progress in human body recognition systems. Especially in the case of changes in occlusion and lighting conditions, it is difficult to draw useful conclusions if using 2D data, but using 3D point cloud data for HAR is particularly beneficial. For instance, Xing Li et al. (2023), have proposed PointMapNet in the 3D human motion recognition task. Consequently, such a network explicitly shuns the typical computational complexity with voxelization characteristics for easy point-cloud sequence modeling while still holding the structural details. The two-flow network architecture in the PointMapNet architecture both breaks spatiotemporal information encoding and, at the same time, prevents mutual interferences between appearance and motion features. Table2.1 showing the summary of 30 literatures is as follows:

Table 1.1 3D point cloud data analysis on human activity recognition

No	Reference	Datasets	Method	Algorithm	Results	Advantages/Contribution	Limitations
1	Li Xing, Huang Qian, Zhang Yunfei, Yang Tianjin, Wang Zhijian (2023)	NTU RGB+D 60 MSRActio n3D 3D Action	DMM MHI and MEI PCAM PCMM	Deep Learning for static point clouds	Using the PointMapNet for improving 3D human action recognition.	Avoids voxelization, simplifying point cloud modelling. Easier to encode than depth sequences. Richer details than skeleton sequences;	PointMapNet has limited capabilities in capturing point cloud sequences and is not suitable for semantic segmentation.
2	WenjieYang , JianlinZhang , Jingju Cai, Zhiyong Xu(2021)	NTU- RGB+D, NTU- RGB+D 120,	CWG MTC STAM GCNs	PyTorch deep learning framework	A trainable relation selection mechanism solve the problem about how to spare the adjacency matrices.	Symmetrical spatial-temporal attention module enhances sensitivity to complex contexts.	The CWG and MTC have obvious noise in the skeleton sequence, which increases the calculation cost.
3	Wang, Y, Xiong, F., Jiang, Zhou, J. T., Yuan, J. (2020)	NTU RGB+D 60 and120 N-UCLA and UWA3DII	3DV Deep learning network	A multi- stream deep learning model	PointNet++ is connected with 3DV to conduct end-to-end feature learning.	PointNet++ is smallest model size and running faster. It can be easier to train.	The discernment of 3DV model needs to be further enhanced.
4	A. Kamel, B. Sheng, P. Yang, P.Li, R.Shen(2019)	MSRActio n3D 3D Action	CNN	Depth Motion Image	A method for identifying human movements from depth maps and pose data by deep CNNs .	Three-channel deep CNN method uses depth maps and posture data.	Moving cameras which captures spontaneous actions from various views and distances.
5	Wu, Y. Li X. Ma (2022)	NTU RGB+D 60 & 120	DRDIS PEMS	RGB-Based Deep Learning Methods	Enhance the spatiotemporal multimodal learning ability of 3D models on depth and attitude data.	Propose DRDIS and PEMS for depth and pose video representations.	3Dnetwork architecture needs to enhance lightweight to reduce computing costs.

No	Reference	Datasets	Method	Algorithm	Results	Advantages/Contribution	Limitations
6	Y. Wang, S. He, X. Wei, S. A. (2022)	Kinetics-700 dataset	Kinetics-700 dataset +	Deep convolutional network	Optimizes frame image processing, reducing time and harder to enhance performance accuracy.	Kinetics dataset yields higher accuracy and avoids overfitting, training deeper 3D ResNet models.	The human action dataset and recognition speed will be enhanced model accuracy and application experience.
7	J. Liu, A. Shahroudy, M. Perez, G. Wang, L. -Y. Duan, A. C. Kot (2019)	NTU RGB+D 120 dataset	3D Action Recognition Methods	Part-Aware LSTM and FSNet	A large scale RGB+D motion recognition dataset in cameras with different heights is presented.	Based on cross-topic and cross-setting criteria, this paper compares different data patterns to enhance action recognition performance.	No limitation
8	J. Park, J. Kim, Y. Gil and D. Kim(2024)	DGU-HAO data.	The Skeleton Bone method	MMNet model	A novel motion-capture dataset that can be tailored to human motion analysis	DGU-HAO features multi-modality with five data types and ample samples.	No limitation
9	X. Gao et al(2019)	NTU RGB-D dataset	GCN	Graph-Based Representation Learning	The motion recognition problem of 3D skeleton data is successfully solved.	Propose skeleton feature expression to enhance limb and center of gravity relationships.	The NTU-RGB+D has limitations in terms of size, diversity, and representation.
10	P. Wei, H. Sun, N. Zheng(2019)	UTKinect-Action 3D Dataset. MSRAction3D	CLS	CAIP:	They succeeded in solving a model for recognizing three-dimensional human behavior called the CLS model.	Their method learns applicable to multimedia applications like content retrieval and educational entertainment.	CLS models need to be combined with deep learning to better enhance HAR.
11	Y. Ji, Y. Yang, F. Shen, H. T. Shen W(2021)	A new RGB-D action Skeleton data	VS-CNN C3D LRCN TCN	C3D	A broad RGB-D action dataset was successfully designed for any perspective analysis	VS-CNN excels in wide view ranges, outperforming related approaches. Extensive evaluations show superiority over eight related methods.	Unable to handle large view changes.
12	D. C. Luvizon, D. Picard, H. Tabia(2021)	NTU RGB+D	CNN	A multi-task deep learning	Achieves precise 3D pose estimation without volumetric heat maps, and trains with mixed 2D and 3D enhancing accuracy.	Uses only RGB images for 3D poses and visual information. Scalable network architecture without extra training.	No limitation

No	Reference	Datasets	Method	Algorithm	Results	Advantages/Contribution	Limitations
13	M. Karim, S. Khalid, A. Algerian, N. Tauran, Z. Ali and F. Ali(2024)	the novel HADE dataset	the HADE I and HADE II CNN models	SOA machine learning	HADE approach achieves 83.57% accuracy, surpassing existing benchmarks, confirming its ability to enhance recognition accuracy and overall HAR system performance.	Introduce HADE dataset for precise action categorization using HADE I and II CNN models. Employ SOA feature extraction and advanced ML to handle occlusion, lighting	Expanding the range of actions within the HADE dataset to cover more complex and diverse human activities.
14	Y. Han, S. -L. Chung, Q. Xiao, W. Y. Lin S. -F. Su(2020)	NTU RGB+D	GSA and ALC GL-LSTM LSTM Diff	RNN	This paper integrates GSA and ALC models into LSTM framework, and successfully proposes GL-LSTM+Diff model	Global spatial attention evaluates skeleton joint significance, enhancing accuracy.	Outperforms STA-LSTM lacks exploration in generalization, scalability, robustness, and efficiency for real-world applications.
15	L. Wang, D. Q. Huynh P. Koniusz(2024)	MSRAAction3D Action NTU RGB+D Dataset	RNN LSTM	HOPC HPM+TM P-LSTM RNN	Multiple state-of-the-art 3D motion recognition algorithms are used on benchmark datasets to present, analyse, and compare manual and deep learning features.	Assessed cross-view vs. cross-subject performance, and impact of camera view on recognition using depth, skeleton, or combined features.	With handcrafted features showing superiority over deep learning ones in smaller datasets and deep learning methods excelling with larger datasets.
16	J. Zhu, W. Zou, Z. Zhu, L. Xu, G. Huang(2019)	NTU RGB+D, Northwestern UCLA MSR Daily Activity3D	I3D Align	Roi Action Machine	Motion recognition machine has achieved competitive performance on video motion data set by using character bounding box and human posture.	Enhance performance by modelling body movements, multi-task training, and fusing RGB and pose predictions. Demonstrate framework's generality through extensive experiments.	Motion recognition machines are also an effective framework in other application scenarios where further enhancements are needed.
17	T. Murakami, T.Nakamura(2020)	3D pose	deep neural network (DNN)	R-CNN.	This study achieves enhanced accuracy of 3D pose estimation during high-speed motion DNNS using generated training datasets.	Eliminates retraining by aligning 2D joint locations with training data, regardless of camera parameters.	In the face of high-speed motion, 3D pose using the correction procedure proposed in this paper can become unstable.
18	M. Li, S.Chen, X. ChenY. Zhang, Y. (2022)	NTU-RGB+D, Kinetics,	Sybio-GNN	JGC PGC	Sybio-GNN is proposed to capture motion patterns through graph structure manipulation.	Multitasking framework enhances action recognition and motion prediction via mutual promotion.	The model is limited, the evaluation scope is narrow, and the applicability is not widely discussed.

No	Reference	Datasets	Method	Algorithm	Results	Advantages/Contribution	Limitations
19	C. Wang J. Yan(2023)	NTU RGB+D Kinetics- Skeleton	RGB- based, Skeleton- based	2D and 3D pose estimation algorithms	This paper gives a comprehensive overview of human action recognition methods through data type system reviews.	This paper successfully reviewed the application of 2D and 3D depth attitude estimation models in HAR,	The RGB method is very limited by background noise and shooting Angle.
20	C. Yang, X. Wang S. Mao(2023)	RF data	TARF	DANN CNN	In this paper, TARF is combined with various RF devices to successfully mitigate the impact of technology-independent data acquisition	TARF prototype: Demonstrates robust human activity recognition across FMCW radar, WIFI, and RFID, CNN	Real-world testing is needed to scale and compare performance under different conditions.
21	H. A. Ullah, S. Letchmunan, M. S. Zia, U. M. Butt aF. H. Hassan(2021)	DMLSmar tAcions. Keleton dataset RQ3	C3D	CNN RNN 3D-DNN	This paper selects a suitable deep neural network architecture by systematically reading literature review and using the latest deep learning paradigms.	HAR assists researchers and enhances real-world modelling and assesses the need for enhancement by highlighting video HAR research gaps.	HAR research can only be used for video activity recognition
22	A. Abdelgawwad , A. C. Mallofré, M. Pätzold(2021)	IMU data CSI data	The Geometric al Model	TV-MDS	This paper demonstrates the possibility of designing an IMU based human activity recognition model for non-static channels.	Expressions for TV speed, azimuth angle, and elevation angle. Consideration of TV path gains for realism. Micro-Doppler signature extraction using spectrogram.	No limitation
23	G. Zheng(2021)	WISDM and UCI- HAR	LGSTNet Model Architectur e	3D-CNN Module With 3D-DF	LGSTNet is proposed for addressing human activity recognition problem.	LGSTNet, a novel deep learning model, combines 2D and 3D CNNs for HAR.	Future work aims to enhance fine-grained activity recognition, such as eating, drinking, and brushing teeth
24	H. He, G. Liu, X. Zhu, L. He G. Tian(2019)	RGB data The depth data Dataset- SDK	IMM Kinect v2 camera IWCF	IMM Algorithm	In this paper, IMM method and IWCF are successfully used to integrate distributed multi-view information to solve the occlusion problem.	Evaluation conducted on Kinect SDK and OpenPose datasets. IWCF+IMM method achieves higher accuracy, especially with OpenPose dataset.	The limitation of this work is the lack of comprehensive theoretical analysis and comparison

No	Reference	Datasets	Method	Algorithm	Results	Advantages/Contribution	Limitations
25	S. Kumawat, M. Verma, Y. Nakashima S. Raman (2022)	Kinetics-400,	SGD	Deep CNNs X-STFT networks 3D CNNs	This paper successfully proposes a new layer class called STFT which can replace 3D convolutional layer.	Better feature learning; state-of-the-art accuracy demonstrated. Future extension: Tasks like classification, segmentation in 3D representations.	No limitation
26	P. V. V. Kishore, D. A. Kumar, R. C. Tanguturi, K.(2024)	3D skeletal datasets: NTU RGBD,	JMAM jg2gMDM	MRCNNSA CNN JMAM	The development of JMAMs for action recognition, showing enhanced accuracy rates using the MRCNNSA network and jg2gMDMs.	By providing grouped motion information, lightweight multi-resolution CNNs with spatial attention mechanisms are developed for classification.	There are too few datasets, and larger datasets are needed to further train the model
27	K. Bantupalli Y. Xie(2019)	The American Sign Language data	LSTM RNN	CNN SLR	Non-signers have difficulties in communication and promote sign language translation through deep learning	Promote sign language interpretation and using a ROI to isolate hand gestures from the images help accuracy.	No different RNN architectures are used for detection, and a capsule network can be used instead of Inception.
28	C. Kang, M. Kim, K. Kim and S. Lee (2024)	Body Motion Data contact state data	Bi-LSTM structure	LLF	The accuracy, recall and precision of this method are 0.99, 0.97 and 0.95, respectively.	Bi-LSTM was used to process sequence features and DNN was used solve the detection of contact parts.	Tagging frames with contact labels is time-consuming, prompting exploration of efficient data collection methods.
29	W. Ding, C. Ding, G. Li K. Liu(2021)	NTU RGB+D Kinect datasets	Graph LSTM	3D CNN 3D Convolutional Neural Networks	Dependence is extracted by JSG and RSG while 3D convolution and 3D pooling are performed.	Proposed SSG for adaptive skeleton structure learning and spatial relationship determination among body parts.	Focus on various methods for encoding skeleton data to adapt the CNN more effectively.
30	Amir Shahroudy, Jun Liu, Tian-Tsong Ng, Gang Wang (2016)	NTU RGB+D	LSTM	Utilizes Microsoft Kinect v2 to capture RGB videos	The method shows superior performance compared to state-of-the-art hand-crafted features on the NTU RGB+D dataset.	Large-scale, diverse dataset with multiple modalities and extensive variation.	Limited to indoor scenes due to the sensor's operational constraints.Generalization to diverse real-world scenarios and environments needs further validation.

In related work, Yang et al. (2021) designed a mechanism for 3D point cloud data in which the connections exist, such that the model automatically finds the most valuable connections in the graph. This enables models to generate sparse adjacency matrices and, by all means, prevent the transmission of redundant information between nodes. Relation selection in Graph Convolutional Networks for action recognition shows that graph-based approaches might be strengthened for enhanced recognition by capturing richer and more complex spatiotemporal relationships in skeleton data extracted from point clouds.

Gao et al. (2019) also suggested a new GCN framework for the recognition of video motion based on 3D bones, and this utilizes natural connection inside bone data, which will allow the enhancement of the relation between limbs and their center of gravity. Such a methodology tends to bring out the capability of graphical representation in capturing complex details of human movements. In order to further enhance accuracy by a significant value, Kumawat et al. (2022) designed an H. action recognition technique by incorporating spatiotemporal Deep Short-Time Fourier Transform Convolutional Neural Networks (STFT-CNN) and deep learning for spatiotemporal feature extraction.

Kang et al. (2023) used manually labelled contact data and deep learning technologies to detect contact points from 3D human motion data. This study has greatly enhanced the accuracy of detection for contact parts by the combination of deep learning with the manually labelled data. For example, Ding et al. (2021) have presented a kind of method for human action recognition, where, in this case, the square grid skeletons are used; such a technique requires 3D CNN in the processing of the skeleton data, thereby considerably enhanced in its action recognition performance.

Specific works, geometric feature extraction, and statistical feature analysis have developed considerably. The work of Wang et al. (2020) introduced a multi-stream deep learning model to learn 3D motion and appearance features of a 3D dynamic voxel (3DV) method for deep video motion recognition; end-to-end features are learned from point cloud representations using Pointnet++. The training of this is compared to that

of the conventional 3D convolutional networks. This is more efficient and easily trained.

On the other hand, Wu et al. (2022) put forward a novel method by combining spatio-temporal multimodal learning with 3D Convolutional networks in video action recognition where depth and pose data are integrated to achieve better recognition performance. They derive these motion patterns by suggesting 3D CNN extension and proposing depth residual dynamic image sequences besides pose estimation mapping sequences to cover full features in the body motion.

Such motion iteration and the importance of multiple modalities were iterated by Kamel et al. (2019), where deep, convolutional neural networks were employed to recognize the action of human based on depth map and pose; the applied emphasis was on the fusion of depth information concerning the pose analysis. The precision of motion recognition in this method is increased by fusing the data from depth maps with post-data. It is this type of complex scene where added dimensional information by depth maps facilitates the capturing of slight variations in human movements.

Ji et al. (2021) developed an action recognition method for human actions that includes viewing from any angle using arbitrary angles through action sets of RGB-D data, which contain any-angle data on actions for training. To better the rate and robustness of recognizing information on action through multi-angle details, this paper proposed an arbitrary-viewing-angle human action recognition approach based on RGB-D data about the action sets of arbitrary-viewing-angle action information. Recognition of moving objects from one viewpoint using geometric features is an essential but complex problem.

In other words, the result of a systematic analysis of 30 kinds of literature is that, currently, deep learning and the extraction of geometric characteristics are two main research directions of 3D point cloud data analysis. This study chooses to combine PointMapNet with GCN to enhance the model's ability to capture complex spatial relationships in 3D point cloud data. And summarize three main frequently used 3D point cloud data.

2.3 3D POINT CLOUD DATA

From the above Table 2.2, a derived table based on the data set is as follows:

Table 1.2 3D point cloud data

Datasets	Method
NTU RGB+D	1. DMM
NTU RGB+D 60	2. MHI and MEI
NTU RGB+D 120	3. AH-DMMs
	4. DMI
	5. PCAM and PCMM
	1. CWG
	2. MTC
	3. STAM
	1. Temporal rank pooling
	2. Temporal split
	3. A multi-stream deep learning model
	1. 3D Activity Analysis Datasets
	2. 3D Action Recognition Methods
	3. Word2Vec module
	1. Convolutional neural networks (CNNs)
	2. The powerful Word2Vec model
	3. APSR Framework
	A novel end-to-end model based on Graph Convolutional Network (GCN)
	CNN architecture
	3D ConvNet (I3D)
	RGB Action random crop
	RGB Action person crop
	KPS RGB Action
	5.KPS Pose Action RGB Action
	symbiotic graph neural network (Sybio-GNN)
	both RGB-based and skeleton-based approaches
	Joint Motion Affinity Maps
	The Joint Group to Group Motion Directed JMAMs
	Graph Representation of Skeleton
	Grid Representation of Skeleton Structure
	LSTM
MSR-Action3D , MSR-DailyActivity	1. Convolutional neural networks (CNNs)
	2. The powerful Word2Vec model
	3. APSR Framework
	1. 3D ConvNet (I3D)
	2. RGB Action random crop
	3. RGB Action person crop
	4. KPS RGB Action
	5. KPS Pose Action RGB Action
SBU Kinect interaction	1. The depth residual dynamic image sequence (DRDIS)
	2. The pose estimation map sequence (PEMS)
	Graph Representation of Skeleton
	Grid Representation of Skeleton Structure
	LSTM
Kinetics-400	Stochastic gradient descent (SGD)
	The STFT blocks-based networks
Kinetics-700 dataset	3D ResNet framework + Kinetics-700 dataset + OpenCV + Python.

2.3.1 NTU RGB+D and NTU RGB+D 120

Li et al. (2023) used the NTU RGB+ D60, UTD-MHAD, and MSR Action3D datasets in their study to propose a point cloud feature map network called PointMapNet for 3D human action recognition. Using the NTU RGB+ D60 dataset, they achieved 89.4% cross-subject accuracy and 96.7% cross-view accuracy.

The study by Wenjie Yang et al. (2021) utilized NTU-RGB+D and NTU-RGB+D 120 datasets. They run skeleton-based action recognition with heightened relevance to complicated contexts when the RS-GCN, which refers to Relation selection graph convolutional networks, is added. These two datasets provide many training samples for their models; therefore, upon graph generation, it can select valuable connections while creating a sparse adjacency matrix at the same time to keep down redundancy in information.

Wang et al. used the NTU RGB+D 120 and 60 datasets to suggest a new method, that is, a 3D dynamic voxel, for the detection of motion. They apply PointNet++ in 3DV for end-to-end feature learning. Those data sets offered large quantities of training data to support the training and testing of their models from many angles, which further advanced their generalization capacities.

For example, Wu et al. (2022) introduced the NTU RGB+D 60 and 120 datasets in their research to create a multimodal dual-flow 3D network framework aimed at understanding spatiotemporal multimodal learning capabilities for 3D models on depth and attitude data. This makes up the diversity and richness for them to perform many experiments and validate the performance of their methods over challenging RGB-D datasets.

In the research work, Jun Liu (2019) used the dataset of NTU RGB+D120 from one of the most extensive 3D human activity datasets, which comprises 114,480 video samples for 120 action categories and camera settings. They have proposed and validated a simple APSR framework for single-shot 3D motion recognition. They tested most of the state-of-the-art approaches in 3D motion recognition using this dataset to compare recognition performance on different modes of data and examined criteria on

cross-subject recognition. These data sets provide a rich training sample set for models, significantly improving the ability to generalize over different environments and perspectives.

The new 3D skeleton action recognition model by Liu et al. and Gou et al (2019) is built on a graph neural network, using the NTU RGB-D dataset. For example, the use of multi-tasking to learn both color and depth data streams simultaneously to estimate more accurate poses 17, or learning from synthetic data in the absence of large-scale annotated real data 18, have become standard methods. The proposal utilizes a skeleton feature representation for better model the relation between limbs and as well as limb with the center of mass. Also, a method is suggested to calculate the interaction features in one frame and then to convolve two skeletons at the same time. This rich sample training made possible for their model to be trained with the help of those datasets, which later yield some good results in the accuracy in estimating 3D poses and motion recognition. However, both the scale and diversity of the NTU RGB-D dataset have limitations in expression that might limit its applicability and generalization in specific contexts.

D. Luvizon et al. (2021) designed a multitask deep learning architecture—3D human pose estimation and action recognition—considering the addition of elastic network losses within the NTU RGB+D dataset. With the help of multi-task environment and a large number of training samples, they successfully realized and obtained high-precision attitude estimation and motion recognition.

In this recent review, Han et al. (2020) showed that using NTU RGB+D and SBU datasets is suitable to propose a global spatiotemporal attended action recognition framework through 3D human skeleton data. They have achieved action recognition based on skeleton data by combining GSA (Global Spatiotemporal Attention) and ALC models (Accumulative Learning Curve). This type of GSA model can enhance the accuracy of articulatory importance expression since it can evaluate the importance of every joint of the skeleton within the whole sequence of motion. To calculate the importance of each frame in this kind of ALC model, weight parameters are introduced into the LSTM architecture for adequate capacity reflection. Large numbers of these