Dynamic Sign Language Recognition using Deep Learning

SAMEENA JAVAID

02-284172-002

A thesis submitted in fulfillment of the requirement for the degree of Doctor of Philosophy (Computer Sciences)

Department of Computer Sciences

BAHRIA UNIVERSITY KARACHI

AUGUST 2023

APPROVAL FOR EXAMINATION

Scholar Name: Sameena Javaid Registration No. 36798 Program of study: Ph.D. (Computer Science) Thesis Title: Dynamic Sign Language Recognition using Deep Learning it is to certify that above scholars' thesis has been completed to my satisfaction and, to my belief, its' standard is appropriate for submission for examination. I have also conducted a plagiarism test of this thesis using HEC prescribed software and found a similarity index of 14 % that is within the permissible limit set by the HEC for the Ph.D. degree thesis. I have also found the thesis in a format recognized by Bahria University for the Ph.D. thesis.

Principal Supervisor's Signature: ____________________

Date: 08th August 2023

Name: Dr. Syed Safdar Ali Rizvi

AUTHOR'S DECLARATION

I, ______Sameena Javaid______hereby state that my Ph.D. thesis titled "Dynamic Sign Language Recognition using Deep Learning" is my own work and has not been submitted previously by me for taking any degree from this university ______Bahria University____ or anywhere else in the country/world. At any time if my statement is found to be incorrect even after my completion, the University has the right to withdraw/cancel my Ph.D. degree.

Name of scholar: ____Sameena Javaid______

Date: __08th August 2023____

PLAGIARISM UNDERTAKING

I, solemnly declare that research work presented in the thesis titled "Dynamic Sign Language Recognition using Deep Learning" is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that I have written the complete thesis. I understand the zero-tolerance policy of the HEC and Bahria University towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited. I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D. degree, the university reserves the right to withdraw / revoke my Ph.D. degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Scholar / Author's Sign: ____________________

Name of the Scholar: Sameena Javaid

DEDICATION

After almighty Allah who has blessed me always, I dedicate this thesis to my supervisor Dr. Syed Safdar Ali Rizvi.

ACKNOWLEDGMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed to my understanding and thoughts. In particular, I wish to express my sincere appreciation to my thesis supervisor, Senior Associate Professor Dr. Syed Safdar Ali Rizvi, for his encouragement, guidance, critics, and motivation. Without his continued support and interest, this thesis would not have been the same as presented here.

The administration at Bahria University also deserves special thanks for their assistance in providing the relevant resources, environment, and sustenance. My kindred postgraduate understudies and partners ought to likewise be perceived for their help. My true appreciation additionally reaches out to every one of my reviewer, examiner, and other people who have helped on different events. Their perspectives and tips are valuable for sure. Last but not least I am grateful to all my family members who are a great source of encouragement and motivation, without their support I cannot be here where I am today.

ABSTRACT

Sign Language-based communication is a major source of communication between people with hearing impairments and the general public. Hand, head, body, and gesture developments are immediate aspects of communicating specific feelings or techniques of correspondence in any language, whether marked or spoken. Hand sign recognition is known as manual sign language recognition in sign language, whereas facial expressions and body gesture understanding are known as non-manual sign language recognition. Several researchers are focusing either on manual gestures or non-manual gestures separately; a rare focus is on manual and non-manual gestures concurrently, making the loss of content or complete meaning of the sentence. Nonetheless, one of the great challenges of Sign Language Recognition is dealing with formal and non-formal parameters simultaneously within a single framework or methodology whereas using dynamic sign language recognition. However, dynamic sign language recognition has several difficulties recognizing complex features, accurate classifications, and extensive video sequence data training and testing in the Spatio-temporal domain. To issues, the current research study presents a Multimodal Dynamic Sign Language Recognition in the Spatio-temporal domain based on vision-based deep learning. This research consists of two-fold contributions. First, for the training and testing, there is a lack of such a dataset, where manual and non-manual modalities combine with affective facial expressions. So our first contribution is compiling a Pakistan Sign Language (PSL) dataset with Manual and Non-Manual modalities named PkSLMNM. Further, we proposed a system called Sign Language Action Transformer Network (SLATN) is restricts hand, body, and facial signals in video arrangements. Here we are using a Transformer-style structural design as a "base network" to remove highlights from a spatiotemporal space. The model hastily figures out how to follow individual people and their setting of activity in different edges. Further, a "head network" at the same time classify and region out hand movement and facial expression, which is frequently critical to figuring out communication through signing. It uses its attention mechanism for creating tight bounding boxes around classified gestures. Later, the model's performance is evaluated against state-of-the-art datasets and conventional identification techniques. It not only completes tasks more efficiently but also with good accuracy. Our suggested network achieves 82.66% testing accuracy and 94.13 Giga FLOPs of a notable processing performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D-CNN	3 Dimension – Convolutional Neural Network
3D-GSNET	3 Dimension – Gesture Segmentation Network
3DMobileNet	3-Dimensional Mobile Network
3DRCNN	3-Dimensional Recurrent Convolutional Neural Network
AAM	Active Appearance Model
AFE	Affective Facial Expression
ANN	Artificial Neural Network
API	Application Program Interface
ArSL	Arabic Sign Language
ASL	American Sign Language
BOF	Bag of Features
BSL	Brazilian Sign Language
CMOS	Complementary Metal-Oxide-Semiconductor
CPU	Central Processing Unit
CRF	Conditional Random Forest
CV	Computer Vision
CSL	Chinese Sign Language
D3D	Dissected 3 Dimension
DCT	Discrete Cosine Transform
DL	Deep Learning
DLSTM	Deep Long Short Term Memory
DoF	Degree of Freedom
DSLR	Dynamic Sign Language Recognition
DTW	Dynamic Time Wrapping

DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
EOH	Edge Oriented Histogram
ESL	Egyptian Sign Language
FACS	Facial Action Coding System
FC-RNN	Fully Connected - Recurrent Neural Network
FE	Facial Expression / Feature Extraction
FESF	Family Education Services Foundation
FN	False Negative
FP	False Positive
FPS	Frames per Second
FRNN	Fuzzy Rough Nearest Neighbors
FRCNN	Faster Region based Convolutional Neural Network
FU	Facial Units
G-FLOPS	Giga-Floating Point Operations per Second
GB	Giga Bytes
GFE	Grammatical Facial Expression
GMM	Gaussian Mix Models
GPU	Graphical Processing Unit
GSL	German Sign Language
HAR	Human Action Recognition
HCI	Human Computer Interaction
HCRF	Hidden Conditional Random Field
HD	High Definition
HMM	Hidden Markov Model
HoG	Histogram of Oriented Gradients
HTF	Heterogeneous Texture Feature
HU	Histogram Units
I3D	Inflated 3 Dimension
IBCC	Independent Bayesian Classification Combination
ICT	Integrated Computing Technologies

IMU	Inertial Measurement Unit
IoT	Internet of Things
ISL	Indian Sign Language
JSL	Japanese Sign Language
KSL	Korean Sign Language
L-RELU	Leaky – Rectified Linear Unit
LBP	Local Binary Pattern
LED	Light Emitting Diodes
LIBRAS	Língua Brasileira de Sinais
LMC	Leap Motion Controller
LSTM	Long Short Term Memory
mAP	Mean Average Precision
ML	Machine Learning
MNIST	Modified National Institute of Standards and Technology
MoSL	Moroccan Sign Language
MTT	Multi Type Tree
NN-DTW	Neural Network - Dynamic Time Wrapping
PC	Personal Computer
PCA	Principle Component Analysis
PkSLMNM	Pakistan Sign Language Manual Non-Manual
PSL	Persian / Polish / Pakistan Sign Language
QP	Quantization Parameter
QPr	Query Pre-Processor
RAM	Random Access Memory
RCU	Recurrent Convolutional Unit
ReLU	Rectified Linear Unit
RFC	Random Forest Classifier
RGB	Red Green Blue
RGB-D	Red Green Blue – Depth
RNN	Recurrent Neural Network
RO	Research Objective

RoI	Region of Interest
RP	Research Problem / Region Proposal
RPN	Risk priority number
RQ	Research Question
RSL	Russian Sign Language
SIFT	Scale Invariant Feature Transform
SL	Sign Language
SLATN	Sign Language Action Transformer Network
SLR	Sign Language Recognition
SSLR	Static Sign Language Recognition
SVD	Single Value Decomposition
SVM	Support Vector Machine
TN	True Negative
ToF	Time-of-Flight
TP	True Positive
TSL	Turkish Sign Language
UN	United Nations
VCEG	Video Coding Expert Group
VGA	Video Graphic Array
WHO	World Health Organization
X3D	Extended 3 Dimensional