CERTIFICATE

It is certified that the work contained in the thesis titled ***"Video-based Human Action Recognition in the Wild using Deep Learning"*** by ***Nitika Nigam*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor

Dr. Tanima Dutta
Assistant Professor,
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Uttar Pradesh, INDIA 221005.

DECLARATION BY THE CANDIDATE

I, ***Nitika Nigam***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Dr. Tanima Dutta*** from ***July 2018*** to ***August 2023*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, thesis, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 10/2/23
Place: Varanasi

Nitika
(Nitika Nigam)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Dr. Tanima Dutta)

Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

10/02/23

Signature of Head of Department (Actg)
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Video-based Human Action Recognition in the Wild using Deep Learning

Name of the Student: Nitika Nigam

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date: 10/2/23
Place: Varanasi

Nitika.
(Nitika Nigam)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

Dedicated to my parents,

Mrs. Varsha Nigam

and

Mr. Raj Narain Nigam

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor, Dr. Tanima Dutta, for her valuable asset and assistance. I regard it as tremendous happiness to express my profound sense of gratitude and sincere regard for her constant feedback and expertise during all these years. I am eternally grateful to have had the opportunity to work on my thesis under her supervision. She supported my research directions and allowed me to explore new ideas in the domain of computer vision and deep learning. My cordial thanks to all the members of the Department of Computer Science and Engineering for creating an excellent working atmosphere. I would also like to thank the other members of my Doctoral committee, Dr. Bhaskar Biswas, Department of Computer Science and Engineering, and Dr. Rajeev, Department of Mathematical Sciences, for their help and support throughout the tenure of my studies. Special thanks to Dr. Hari Prabhat Gupta for his consistent assistance in both work and life aspects. I would also like to convey my sincere gratitude to Dr. S.K Singh, Head of Department of Computer Science and Engineering and all the RPEC and DPGC members for their suggestions and endorsement to this work.

I am grateful to Deepali Verma, Rahul Mishra, Ashish Gupta, Randheer Bagi, and all my dear lab members for the long discussions and their brilliant insights that have helped me to overcome the challenges I have faced in the development of this work. I also want to thanks my dear friends Suneha Verma, Pratima Yadav, Paras, Nitish, Lokendra, Nistha Dubey, Kritika Singh, Nidhi Chaubey, and Darshna for supporting me in my hard times.

Finally, I express my heartfelt gratitude to my parents Mrs. Varsha Nigam and Mr. Raj Narain Nigam, and my sister Shipra Nigam for their constant support, love, encouragement, and sacrifices. Their affectionate love and care cannot be expressed in words.

(Nitika Nigam)

Contents

List of Figures

List of Tables

List of Symbols

Symbol	Description
H or $\mathcal{H}$	Height of feature map
W or $\mathcal{W}$	Width of feature map
C or $\mathcal{C}$	Channels of feature map
t	Timestep
T or $\mathcal{T}$	Temporal length
i, j, k	Temporary Variables
epr	Facial expression
gst	Limb Gesture
max	Maximum value
min	Minimum value
tanh	tanh function
att	attention
ReLU	ReLU function
$\boldsymbol{W}$	Dilated kernel
$\mathfrak{W}$	Kernel
${}^{n}C_{r}$	Combination
wce	Weighted Cross-entropy Loss
wfl	weighted focal-loss
λ	hyperparameters
η	learning rate
softmax or $\delta(\cdot)$	Softmax function
sigmoid or $\sigma(\cdot)$	Sigmoid function
m	Input modulation gate
i	Input gate
o	Output gate
h	Hidden state
$\circ$	Hadamard product

Symbol	**Description**
$*$	Convolution operation
$\circledast$	Kronecker product

Abbreviations

Abbreviation	Description
CNN	Convolutional Neural Network
DNN	Deep Neural Network
DL	Deep Learning
LSTM	Long short-term memory
SVM	Support Vector Machine
AR	Action Recognition
RPN	Region Proposal Network
ReLU	Rectified Linear Unit
GRU	Gated recurrent unit
NMS	Non-maximal suppression
IoU	Intersection-over-union
maxpool	Maxpool Operation
FCN	Fully Convolutional Network
GT	Ground Truth
AP	average pooling
MDB	Multi-scale Deformable Backbone
AOSA	Actor-Object-Scene Attention
LRC	Long-Range Context
CTAP	Class-aware Temporal Attention Pooling
SBN	Shallow Backbone Network
VAN	Visual Attention Network
LAN	long-term attention network
TAP	temporal attention pooling
HAANet	Human Action Attention Network
GMP	Global Max-Pool
TDC	Temporal Deformable Convolution Layer
Spatial AP	Spatial Average Pooling

Abbreviation	**Description**
US	Uncertain-Action Supervision
TCDC	Temporal cascaded deformable convolution
SN	Supervision network
SSV2	Something-Something V2
SOTA	State-Of-The-Art
BN	Batch Normalization
FCL	Federated contrastive learning
HRCA	Human-robot collaboration assembly
MHA	Multi-head attention
FFN	Feed-forward network
FDL	Federated deep learning
GNN	Graph neural network
IoT	Internet of Things
Acc.	Mean accuracy
mAP	Mean average precision
OFB	Object filtering branch
FC	Fully-connected