

Certificate

It is certified that the work contained in the thesis titled “DESIGN AND DEVELOPMENT OF SOME METHODS AND MODELS FOR CROWD ANALYSIS USING COMPUTER VISION AND DEEP LEARNING APPROACHES” 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 the requirements of Comprehensive, Candidacy and State-of-the-art seminar.



Signature of Supervisor

Prof. Rajeev Srivastava

Professor, Department of Computer Science and Engineering

Indian Institute of Technology (BHU), Varanasi

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियंत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी Varanasi-221005

Declaration by the Candidate

I, *Santosh Kumar Tripathy* certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of **Prof. Rajeev Srivastava** from *July 2018 to July 2022*, at the *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 wilfully lifted up any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and included them in this thesis and cited as my own work.

Date: 4.7.22

Place: IIT (BHU), Varanasi

Santosh Kumar Tripathy
Signature of the Student

(Santosh Kumar Tripathy)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Rajeev
4.7.22

Signature of Supervisor

(Prof. Rajeev Srivastava)

Professor, Department of Computer Science and Engineering

Indian Institute of Technology (BHU), Varanasi

पर्यवेक्षक/Supervisor

संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी Varanasi-221005

Sanjay Kumar Singh
Signature of Head of Department

(Prof. Sanjay Kumar Singh)

आचार्य व. विभागाध्यक्ष

Professor & Head

संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg

भारतीय प्रौद्योगिकी संस्थान

Indian Institute of Technology

(बनारस हिन्दू यूनिवर्सिटी)

(Banaras Hindu University)

Copyright Transfer Certificate

Title of the Thesis: *Design and Development of Some Methods and Models for Crowd Analysis using Computer Vision and Deep Learning Approaches.*

Name of the Student: Santosh Kumar Tripathy

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: 04-07-2022

Place: IIT (BHU), Varanasi



Signature of the Student

(Santosh Kumar Tripathy)

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.

Acknowledgement

I would like to express my deepest gratitude to all those people who helped me directly or indirectly to complete this thesis work. Firstly, I would like to express my heartfelt gratitude to my supervisor, **Prof. Rajeev Srivastava**, for being a great mentor and the best adviser I could ever have. His advice, encouragement and critics are source of innovative ideas, inspiration are causes behind the successful completion of this Thesis work. The confidence shown on me by him was the biggest source of inspiration for me. It has been a privilege working with him from several years.

I would like to express my deepest appreciation to my research progress evaluation committee members Prof. K. K. Shukla of the Department of Computer Science and Engineering and Prof. Subir Das, Department of Mathematical Sciences IIT (BHU), for providing continuous support, encouragement, and advice.

I express my sincere thanks to all the Professors, Deans, office staff, supporting staff and PhD Research Scholars of Indian Institute of Technology (BHU) Varanasi India. I express my gratitude to Director, Registrars, Deans, Heads, and Student Alumni of the Indian Institute of Technology (BHU) Varanasi.

Special thanks to Dr. Subodh Srivastava (Assistant Professor, NIT, Patna) for his continuous support during my study. I'd like to extend my gratitude to Dr. Ankit Jaiswal (Assistant Professor, Bennet University), Dr. Roshan Singh (System Analyst IIT BHU Varanasi), Dr. Vibhav Prakash Singh (Assistant Professor MNNIT Allahabad), Dr. Jani Kuntesh Ketan (GEC, Gujrat), Dr. Gargi Srivastava (Assistant Professor, RGPIT Jais Amethi). I also wish to thank my lab members Saurav Arora, Divya Singh and Suryakant Singh for their consistent support and help during my research work.

I extend special thanks to the non-teaching staff in the Department, particularly, Mr. Manoj Singh, Mr. Ravi Bharti, Mr. Subham Pandey, Mr. Prakhar for their consistent support.

My parents, Smt. Rama Tripathy and Shri Trilochan Tripathy, who gave me the power and brain to work out on this research and their help at every level made me to see this success.

I owe thanks to my wife, Mrs. Sagarika Mishra and my son Sriyansh Aditya Tripathy for their continued and unfailing love, support and understanding during my pursuit of Ph.D. degree that made the completion of thesis possible. Words are insufficient to express my profound sense of gratitude to my loving brother, sister, sister-in-law, nephews and neice. Their encouragement and support gave me physical and moral strength throughout my career as well in the present research. Finally, I would like to wind up by paying my heartfelt thanks and prayers to the Almighty Lord Shiva, Mata Parvati, Sri Laxmi Narayan, Mata Saraswati, Sri Ganesh and Mata Santoshi for their unbound love and grace.

Santosh Kumar Tripathy
Santosh Kumar Tripathy

Abstract

In today's world, the exponential growth of the worldwide population has caused a considerable increase in crowd densities in places like rallies, public speeches, stadiums, mass transit, and tourist or pilgrim sites. Such places are vulnerable to crowd disasters. Crowd disasters are controlled by efficiently analyzing crowd scenes. The crowd analysis (CA) is beneficial for drawing better crowd management strategies, and public space design. It helps in developing an AI-based visual surveillance system to provide security and safety to the crowd. However, the CA is a tedious task that comprises several correlated tasks, out of which crowd counting and density estimation (CCDE), crowd congestion-level analysis (CCA), and crowd behavior analysis (CBA) are the minimum tasks required to control crowd disasters. Most existing solutions for crowd analysis are done manually in real world scenario, which are very complex and prone to error. Recently, deep learning techniques have significantly improved the performance of several computer vision tasks and contributed considerably to the development of AI-based solutions. Thus, the drawback of the manual process for the CA can be overcome by developing automated AI-based solutions using computer vision and deep learning solutions.

Further, CA using several single-task AI models will incur computational complexity overheads, which can be minimized by drawing better multitasking CA models. So, to control crowd disasters and provide security and safety to the crowd, efficient solutions using computer vision and deep learning approaches for significant tasks of CA such as CCDE, CCA, CBA, and multitasking CA, are required. However, as per the literature review reported later, the performance of several computer vision-based CA models is mainly affected by cluttered background, varying crowd densities, crowd

shape changes due to perspective distortion, illumination changes, and lack of availability of largescale CA datasets. Therefore, to address the issues mentioned above, the problem statement of the thesis is defined as the design and development of some methods and models for crowd analysis using computer vision and deep learning techniques.

This thesis mainly focuses on studying and analyzing the state-of-the-art CA techniques, finding their advantages and limitations, and proposing new methods and models to accomplish the objectives. This thesis aims to conduct a comprehensive literature review on four tasks of CA, i.e., CCDE, CCA, CBA, and multitasking CA. For each task, various methods and models concerning current research trends have been analyzed by mentioning their pros and cons and identifying possible research scopes. Various models using computer vision and deep learning approaches have been proposed in this thesis to fulfill the research scopes of each of the four tasks of CA.

The first contribution in the thesis is related to the task of Crowd Counting and Density Estimation (CCDE), where two models using deep learning techniques have been proposed. The first proposed model for CCDE is an **Attentive Multi-Stream CNN (AMS-CNN)** for video-based crowd counting. The main objective behind the AMS-CNN is to enhance the feature representation for crowd videos, minimize the effect of the cluttered background, and design attention mechanism for each stream to improve the counting performance. The second proposed model for CCDE is a novel cascaded deep architecture with weak supervision for video crowd counting. This model comprises two deep models named **Local Density-map Regressor (LDR)** and **Global Crowd Counting Regression (GCCR)** modules. The LDR focuses on extracting multiscale spatial-temporal features using a multicolumn 3D Atrous (Dilated) CNN to tackle crowd shape changes due to perspective distortion. It also minimizes the effect of the cluttered background using a **Head Attention Module (HAM)**. The LDR considers the local distribution of

crowds and generated crowd density maps. On the other hand, the GCCR model is trained in a weakly supervised manner to exploit global crowd properties from the predicted density maps to obtain final crowd counting using a multi-layer perceptron neural network.

The second contribution to the thesis is related to the task of Crowd Congestion Level Analysis (CCA), where deep learning-based real-time **Two Input Stream Multi-Column Multi-Stage CNN (TIS-MCMS-CNN)** is proposed. In the proposed method, each of the two streams of TIS-MCMS-CNN has been built with three columns of multi-layers of CNN with different receptive fields to extract multiscale spatial and temporal features from two cues of video frames, i.e., frame and the flow magnitude of the frame. The extracted multiscale features are also known as scale-invariant features, which can handle crowd shape change due to perspective distortion. For experimental analysis, a dataset for the CCA is prepared using three publicly available benchmark crowd datasets. It is observed that the TIS-MCMS-CNN can process the frames in real-time.

The third contribution to the thesis is related to the task of Crowd Behavior Analysis (CBA). Under this contribution, two models have been proposed out of which the first model i.e., a **Multiscale Spatial-Temporal 3D Atrous-Net with PCA-guided OC-SVM (MuST-POS)** is designed using conventional machine learning and deep learning approaches to classify normal and panic crowd behaviors. On the other hand, the second model is a **Two-Stream Multiscale Deep Architecture (TS-MDA)** is developed using deep learning techniques for multiclass crowd behavior prediction. Both the models resolve the issue of human shape variation in the crowd videos, while the second model takes measures to minimize the effect of cluttered backgrounds in the crowd video.

The fourth contribution to the thesis is related to multitasking Crowd Analysis (CA), where an efficient multitasking deep model is proposed using a backbone structure

of multi-layer **Depth-wise Separable CNN (DSCNN)** to predict crowd behaviors with crowd counting. The proposed model exploits spatial-temporal features with **Flow Attention Blocks (FABs)** to provide optical flow attention to different scales features of the backbone network. The FABs focus on the moving pixels, thereby minimizing the effect of cluttered background. The multiscale flow attentive features are fused to handle crowd shape variation and perform multitask CA using a feed-forward network. In addition, a largescale multitasking CA dataset is also developed from the available benchmark crowd behavior datasets. Around 1,20,000 frames have been annotated to obtain ground truth crowd counting information.

The proposed models have been implemented, experimented and evaluated on several publicly available datasets. Experimental results and analysis show that the proposed models perform better than state-of-the-art methods reported in the literature. Extensive ablation studies for each of the proposed models have also been conducted to show the influence of several components of the proposed models.

The thesis concludes with an overall conclusion of the proposed research work, followed by a discussion of possible future research scopes in the areas of CA.

Table of content

Certificate	ii
Declaration by the Candidate	iii
Copyright Transfer Certificate	iv
Acknowledgement	v
Abstract	vii
Table of content	xi
List of Figures	xvi
List of Tables	xx
List of Symbols	xxii
List of Abbreviations	xxiii
Chapter 1 Introduction	1
1.1 Background	1
1.2 Vision-based Crowd Analysis.....	4
1.2.1 Crowd Analysis Tasks.....	6
1.2.1.1 Crowd Counting and Density Estimation	6
1.2.1.2 Crowd Congestion-Level Analysis	6
1.2.1.3 Crowd Behaviour Analysis.....	7
1.2.1.4 Multitasking Crowd Analysis	7
1.2.2 Need for Vision-based Crowd Analysis	7
1.3 Motivation.....	7
1.4 Problem Statement	8
1.5 Thesis Objectives	9
1.6 Contributions to the Thesis	10
1.7 Thesis Organization	11
Chapter 2 Literature Review	13
2.1 Literature Review on Crowd Counting and Density Estimation Approaches	13
2.1.1 Taxonomy of CCDE approaches.....	14
2.1.1.1 Taxonomy of CCDE-based on Mode of Implementation.....	14
2.1.1.2 Taxonomy of CCDE-based on dealing with labelled data.....	15
2.1.1.3 Taxonomy of CCDE-based on Learning Mechanism.....	17
2.1.1.4 Taxonomy of CCDE-based on Dataset Modality	18
2.1.2 Review on Image-based CCDE.....	20
2.1.2.1 Conventional Machine Learning Approaches for Image-based CCDE	20

2.1.2.2 Deep Learning Approaches for Image-based CCDE	24
2.1.3 Review on Video-based CCDE approaches	32
2.1.3.1 Conventional Machine Learning.....	32
2.1.3.2 Deep-Learning-based Techniques.....	37
2.1.4 Summary of Vision-based CCDE.....	38
2.2 Literature review on crowd congestion-level analysis	39
2.2.1 Conventional Machine Learning-based Approaches for CCA	40
2.2.2 Deep Learning-based Approaches for CCA	44
2.2.3 Summary of Vision-based CCA Approaches	44
2.3 Literature Review on Crowd Behavior Analysis.....	45
2.3.1 OCC-based Crowd Behavior Prediction.....	46
2.3.1.1 Traditional Approaches for OCC-based CBP	46
2.3.1.2 Deep-Learning Approaches for OCC-based CBP.....	48
2.3.2 MCC-based Crowd Behavior Prediction	50
2.3.3 Summary of CBA Approaches	51
2.4 Literature review on Multitasking Crowd Analysis	52
2.4.1 Summary of Multitasking Crowd Analysis	52
2.5 Research Gaps Identified for Vision-based Crowd Analysis	53
2.5.1 Research Gaps Identified for Video-based CCDE.....	53
2.5.2 Research Gaps Identified for Vision-based CCA	54
2.5.3 Research Gaps Identified for Vision-CBP	55
2.5.3.1 Research Gaps Identified for OCC-based Crowd Panic Detection (CPD).....	55
2.5.3.2 Research Gaps Identified for MCC-based CBP.....	56
2.5.4 Research Gaps Identified for Multitasking CA	57
2.6 Datasets used for Experimental Analysis	57
2.6.1 Datasets used for Video-based CCDE.....	57
2.6.1.1 Generating Ground Truth Crowd Density Maps.....	59
2.6.2 Datasets used for Crowd Panic Detection.....	60
2.6.3 Datasets used for Crowd Behavior Prediction	62
2.7 Performance Metrics	63
2.8 Conclusion.....	66

Chapter 3 Video-based Crowd Counting and Density Estimation using Deep Learning Techniques **67**

3.1 Introduction	67
3.2 AMS-CNN: Attentive Multi-Stream CNN for Video-based Crowd Counting.....	68
3.2.1 Proposed Method and Model.....	68
3.2.1.1 Detail Architecture of AMS-CNN	71
3.2.1.2 Pre-processing.....	73
3.2.1.3 Loss Function for MS-CNN.....	74
3.2.1.4 Loss Functions for SADM, FADM and TADM	75

3.2.1.5 Final Loss Function and Optimization.....	76
3.2.2 Experimental Setup	77
3.2.3 Results Analysis and Discussion	78
3.2.3.1 The Mall Dataset	78
3.2.3.2 The Venice Dataset.....	79
3.2.3.3 The UCSD Dataset	81
3.2.3.4 Ablation Study.....	83
3.3 A Novel Cascaded Deep Architecture with Weak-Supervision for Video Crowd Counting	85
3.3.1 Proposed Method and Model.....	85
3.3.1.1 Pre-processing.	87
3.3.1.2 Working of the LDR Module	87
3.3.1.3 Working of the Weakly-Supervised GCCR module.....	94
3.3.2 Experimental Setup	96
3.3.3 Results Analysis and Discussion	97
3.3.3.1 The Venice Dataset.....	97
3.3.3.2 The Mall Dataset	99
3.3.3.3 The UCSD Dataset	101
3.3.3.4 Ablation study.....	103
3.4 Conclusion	106
Chapter 4 A Real time Two Input Stream Multi Column Multiscale CNN for Efficient Crowd Congestion-level Analysis.....	107
4.1 Introduction.....	107
4.2 The Proposed Model: A Real time Two Input Stream Multi Column Multiscale CNN for Efficient Crowd Congestion-level Analysis.....	108
4.2.1 Network Architecture	110
4.2.2 Pre-processing and Motion Magnitude Map Extraction.....	111
4.2.3 Problem Formulation & Learning Mechanism.....	112
4.2.4 Precautions to handle Overfitting	115
4.3 Dataset Preparation	117
4.4 Experimental Setup	119
4.5 Result Analysis and Discussion	120
4.5.1 Pets-2009.....	122
4.5.2 UCSD-Ped1	122
4.5.3 UCSD-Ped2.....	123
4.5.4 UMN Plaza1 and Plaza2.....	124
4.5.5 Ablation Study.....	124
4.5.6 Time Analysis	126
4.6 Conclusion	126
Chapter 5 Crowd Behaviour Analysis using Machine Learning and deep learning approaches ...	128

5.1 Introduction	128
5.2 MuST-POS: Multiscale Spatial-Temporal 3D Atrous-Net and PCA guided OC-SVM for Crowd Panic Detection.....	129
5.2.1 Proposed Method and Model.....	129
5.2.1.1 Architecture Details of MuST-3AN.....	129
5.2.1.2 Pre-Processing	132
5.2.1.3 Multiscale Spatial-Temporal feature extraction.....	132
5.2.1.4 Dimension Reduction.....	134
5.2.1.5 Crowd Panic Detection using OC-SVM	135
5.2.2 Experimental Setup.....	135
5.2.3 Result Analysis and Discussion.....	136
5.2.3.1 The UMN dataset.....	136
5.2.3.2 The MED Dataset	138
5.2.3.3 The Pets-2009 Dataset	139
5.2.3.4 Ablation Study	140
5.3 TS-MDA: Two-Stream Multiscale Deep Architecture for Crowd Behaviour Prediction	143
5.3.1 Proposed Method and Model.....	143
5.3.1.1 Pre-processing.....	145
5.3.1.2 Candidates for TS-MDA.....	147
5.3.1.3 Architecture Details	147
5.3.1.4 Multiscale Spatial-Temporal Feature Extraction and Prediction	149
5.3.1.5 Crowd Behaviour Prediction.....	151
5.3.1.6 Loss Function and Optimization	151
5.3.2 Experimental Setup.....	154
5.3.3 Results Analysis and Discussion	154
5.3.3.1 The MED dataset	154
5.3.3.2 The GTA dataset.....	156
5.3.3.3 Ablation Study	157
5.4 Conclusion.....	161
Chapter 6 Multiscale Flow Attentive Depth Separable CNN for Multitasking Crowd Analysis	163
6.1 Introduction	163
6.2 The Proposed Method and Model	164
6.2.1 Overview	164
6.2.2 Pre-processing	165
6.2.3 Network Overview	165
6.2.4 Spatial-Temporal Feature Modelling using Depth Separable CNN.	167
6.2.5 Working of Flow Attention Block.	169
6.2.6 Multiscale De-background Feature Modelling.	170
6.2.7 Multitasking Crowd Analysis and Optimization	170
6.3 Multitasking Crowd Analysis Dataset and Performance Metrics.....	171

6.3.1 Multitasking Crowd Analysis Dataset	171
6.4 Experimental Setup	173
6.5 Results Analysis	173
6.5.1 Results analysis for Crowd Behavior Prediction	175
6.5.1.1 The MED Dataset	175
6.5.1.2 The GTA Dataset	177
6.5.2 Comparative Results Analysis with Crowd Counting Models	178
6.5.2.1 The MED Dataset	178
6.5.2.2 The GTA Dataset	179
6.5.3 Ablation Study	179
6.6 Conclusion	181
Chapter 7 Conclusion and Future Directions.....	182
7.1 Conclusions	182
7.2 Suggestions for Future Research Work.....	184
List of Publications	186
References.....	188

List of Figures

Figure 1.1: Sample of different types of crowd scenes of the MED dataset [2]	2
Figure 1.2: Samples of crowd scenes. (a) and (b) are the samples of the Pets 2009 datasets [3] representing structured crowd scenes. (c) and (d) are the samples of the Venice datasets [4] representing unstructured crowd scenes.	2
Figure 1.3: Human deaths due to stampede [5].....	3
Figure 1.4: Overall structure of vision-based crowd analysis system and its applications	4
Figure 2.1: Categorisation of vision-based crowd counting approaches based on mode of implementation.....	14
Figure 2.2: Categorisation of vision-based crowd counting approaches based on dealing with labelled data	16
Figure 2.3: Categorisation of vision-based crowd counting approaches based on learning mechanism.....	17
Figure 2.4: Categorisation of vision-based crowd counting approaches based on dataset modality.....	19
Figure 2.5: Some results of detection-based approaches obtained from literature.	21
<i>Figure 2.6: Sample of crowd counting results based on people head detection [77].....</i>	<i>25</i>
Figure 2.7: Some samples of the results of detection-based approaches using conventional machine learning techniques	33
Figure 2.8: Example of crowd shape change due to perspective distortion in the Pets-2009 crowd panic dataset	56
Figure 2.9: Examples of human shape change due to perspective distortion in crowd scenes	56
Figure 2.10: Example of a Frame of Mall Dataset [57]	58
Figure 2.11: Example of a Frame of UCSD Dataset [59]	58
Figure 2.12: Example of a frame of Venice dataset [4]	58
Figure 2.13: A frame of the mall dataset [57]	60
Figure 2.14: density map.....	60
Figure 2.15:A frame of the Venice dataset [4].....	60
Figure 2.16: Density map.....	60

Figure 2.17: Examples of samples of the datasets. Figures (a), (b) and (c) are the examples of normal scenes of UMN S1, S2 and S3 respectively. Figures (d) and (e) are the normal scenes of Pets-2009 dataset. Figure (f) is the example of normal scene of MED dataset	62
Figure 2.18: Examples of different samples of the MED dataset [2]	63
Figure 2.19: Examples of different samples of the GTA dataset [146].....	63
Figure 3.1: Architecture of the proposed AMS-CNN model.....	69
Figure 3.2: Blocks of proposed AMS-CNN	70
Figure 3.3: Predicted crowd counts on the Mall dataset [57].	80
<i>Figure 3.4: Predicted crowd counts on the Venice dataset [4]</i>	<i>80</i>
Figure 3.5: Predicted crowd counts on the UCSD dataset [59]	82
Figure 3.6: Predicted Crowd Counts of different models on the Mall dataset [57].....	84
Figure 3.7: Predicted Crowd Counts of different models on the Venice dataset [4].....	84
Figure 3.8: Predicted Crowd Counts of different models on the UCSD dataset [59].....	84
Figure 3.9: Block diagram of the proposed cascaded deep model	86
Figure 3.10: Details of layers used in the proposed architecture.....	88
Figure 3.11: Details of DMRM.....	89
<i>Figure 3.12: Predicted versus Ground-truth Crowd Counts of LDR-Module on the Venice Dataset [4]</i>	<i>98</i>
<i>Figure 3.13: Predicted versus Ground-truth Crowd Counts of LDR+GCR-Module on the Venice Dataset [4]</i>	<i>99</i>
Figure 3.14: Predicted versus Ground-truth Crowd Counts of LDR-Module on the Mall Dataset [57].....	101
<i>Figure 3.15: Predicted versus Ground-truth Crowd Counts of LDR+GCCR Module on the Mall Dataset [57]</i>	<i>101</i>
Figure 3.16: Predicted versus Ground-truth Crowd Counts of LDR-Module on the UCSD Dataset [59].....	103
<i>Figure 3.17: Predicted versus Ground-truth Crowd Counts of LDR+GCCR Module on the UCSD Dataset [59].....</i>	<i>103</i>
Figure 4.1: Overall architecture of the proposed model	108
<i>Figure 4.2: Detail architecture of the model TIS-MCMS-CNN</i>	<i>109</i>
Figure 4.3: Examples of crowd scenes of different crowd congestion-levels	118

Figure 4.4: Confusion Matrix-Heatmap of TIS-MCMS-CNN for Pets-2009 of Dataset-1.	120
Figure 4.5: Confusion Matrix-Heatmap of TIS-MCMS-CNN for Pets-2009 of Dataset-2.	120
Figure 4.6: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UCSD-Ped1 of Dataset-1.	120
Figure 4.7: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UCSD-Ped1 of Dataset-2.	120
Figure 4.8: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UCSD-Ped2 of Dataset-1.	121
Figure 4.9: Confusion Matrix-Heatmap of TIS-MCMS-CNN of UCSD-Ped2 of Dataset-2.	121
Figure 4.10: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UMN-Plaza1 of Dataset-1.	121
Figure 4.11: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UMN-Plaza2 of Dataset-2.	121
Figure 4.12: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UMN-Plaza2 of Dataset-1.	121
Figure 4.13: Confusion Matrix-Heatmap of TIS-MCMS-CNN for UMN-Plaza2 of Dataset-2.	121
Figure 5.1: Overall block diagram of the proposed MuST-POS.....	130
Figure 5.2: The architecture of the proposed MuST-POS	130
Figure 5.3: Examples of output of the proposed model on the UMN dataset. Figures (a), (b), (c) are normal sequences, and the model predicted as normal, figures(d), (e), (f) belong to starting of panic behavior, and the model predicted as panic and figures (g), (h).	137
Figure 5.4: Comparison of average accuracy and average error rate between several approaches on three datasets.	137
Figure 5.5: Examples of output of the proposed model on the MED dataset. Figure (a) is the normal sequence, and the model is predicting as normal, figure (b) shows to starting of panic behaviour, and the model is predicting as panic, and figure (c) shows the panic situations.....	139


Figure 5.6: Examples of output of the proposed model on the Pets-2009 dataset. Figure (a) is the normal sequence, and the model is predicting as normal, figure (b) shows to starting of panic behavior, and the model is predicting as panic, and figure (c) shows the panic frame and the model is predicting as panic.	140
Figure 5.7: Samples of panic situations which are detected as Normal by Single-Scale POS but are detected as Panic by the proposed MuST-POS	143
Figure 5.8: Architecture of proposed TS-MDA.....	144
Figure 5.9: Accuracies obtained by the proposed model using leave-one-sequence-out on the MED dataset.....	154
Figure 5.10: Confusion matrix of the proposed model on the MED dataset [2]	155
Figure 5.11: Confusion matrix of the proposed model on the GTA dataset [146].....	155
Figure 5.12: Confusion metrics of different modules during ablation study on the MED dataset [2]. The subfigures (a), (b), (c), (d), and (e) are the confusion metrics of MSS, MTS, WF-TS-MDA, WoF-TS-MDA, and WoMS-TS-MDA modules, respectively..	158
Figure 5.13: Comparison of accuracies of different models during ablation study using leave-one-sequence-out cross-validation on the MED dataset.	159
Figure 5.14: Confusion metrics of different modules during ablation study on the GTA dataset [146]. The subfigures (a), (b), (c), (d), and (e) are the confusion metrics of MSS, MTS, WF-TS-MDA, WoF-TS-MDA, and WoMS-TS-MDA modules, respectively..	159
Figure 6.1: Overall architecture of the proposed model	166
Figure 6.2: Details of FAB.	167
Figure 6.3: Details of Feed Foreword Network	167
Figure 6.4: Details of Spatial-Temporal Feature Modelling using a DSC-1.....	168
Figure 6.5: Examples of different samples of the MED dataset.	172
Figure 6.6: Examples of different samples of the GTA dataset.....	172
Figure 6.7 Confusion Matrix of the proposed model on the MED dataset.....	176
Figure 6.8: Confusion matrix of the proposed model on GTA dataset.....	178

List of Tables

Table 2.1: Comparative analysis of Image-based CCDE approaches.....	27
Table 2.2: Comparative analysis of Video-based CCDE approaches.....	35
Table 2.3: Comparative analysis of Vision-based CCA approaches.....	42
Table 2.4: Details of datasets used for the CCDE.....	59
Table 2.5: shows the properties of these three datasets.....	61
Table 2.6: Confusion matrix for binary classification.....	64
Table 3.1: Layer Details of AMS-CNN.....	72
Table 3.2: Comparative Analysis of Results on the Mall Dataset [57].....	78
Table 3.3: Comparative Analysis of Results on the Venice Dataset [4].....	80
Table 3.4: Comparison of results of several models on the UCSD Dataset [59].....	82
Table 3.5: Comparison of results of different models during ablation study.....	83
Table 3.6: Details of layers used in the proposed architecture.....	89
Table 3.7: Comparisons of Results of several approaches on Venice.....	98
Table 3.8: Comparisons of Results of several approaches on the Mall dataset [57].....	100
Table 3.9: Comparisons of Results of several approaches on UCSD.....	102
Table 3.10: Comparisons of results for ablation study on the different datasets.....	104
Table 4.1: TIS-MCMS-CNN layers information.....	110
Table 4.2: Details of five congestion levels.....	118
Table 4.3: Details of "Dataset-1".....	119
Table 4.4: Details of "Dataset-2".....	119
Table 4.5: Performance Analysis of several approaches using Dataset Pets-2009.....	122
Table 4.6: Performance analysis of several approaches on UCSD-Ped1.....	123
Table 4.7: Performance analysis of several approaches on UCSD-Ped2.....	123
Table 4.8: Performance analysis of several approaches on UMN-Plaza1.....	124
Table 4.9: Performance analysis of several approaches on UMN-Plaza2.....	124
Table 4.10: Performance analysis of Ablation Study on Different Datasets.....	125
Table 4.11: Test frames processing time of several approaches.....	126
Table 5.1: Block details of the MuST-POS.....	131
Table 5.2: Comparison of results with state-of-the-art methods on the UMN dataset..	136
Table 5.3: Comparison of results with state-of-the-arts on the MED dataset.....	138

Table 5.4: Comparison of results with state-of-the-art methods on the Pets-2009 dataset.	140
Table 5.5: Comparison of results of different modules during ablation study	141
Table 5.6: Details of the layers of the proposed model	148
Table 5.7: Performance comparison with other state-of-the-art approaches on the MED dataset	156
Table 5.8: Performance comparison with state-of-the-art approach on the GTA dataset. Values in bold letters represent best in the table.	157
Table 5.9: Comparative analysis of results of different modules of the proposed TS-MDA during ablation study on the MED dataset [2]. Values in bold letters represent best in the table.....	160
Table 5.10: Comparative analysis of results of different modules of the proposed TS- MDA during ablation study on the GTA dataset [146]. Values in bold letters represent best in the table	160
Table 6.1: Stats of multitasking crowd analysis dataset focusing on crowd behaviours and crowd counting	171
Table 6.2: Experimental analysis of performance of the proposed model on various values of weighted loss parameters on the MED dataset.....	174
Table 6.3: Experimental analysis of performance of the proposed model on various values of weighted loss parameters on the GTA dataset	174
Table 6.4: Comparative result analysis of proposed model for the CBP with state-of-the- art approaches for the MED dataset.....	175
Table 6.5: Comparison of results against Novel_Descriptor [187] for the CBP on the MED dataset [2].....	177
Table 6.6: Comparison of results for CBP on the GTA dataset	177
Table 6.7: Comparison of results for crowd counting on the MED dataset [2].....	179
Table 6.8: Comparison of results for crowd counting on the GTA dataset [146]	179
Table 6.9: Comparison of models during ablation study for the MED dataset [2].....	180
Table 6.10: Comparison of models during ablation study for the GTA dataset.....	181

List of Symbols

β_1	Decay rates of first moment
β_2	Decay rates of second moment
ξ	relaxation variable
ω	normal to the hyperplane of the feature space
$\phi()$	maps the primitive feature space to a higher dimension
#	Cardinality
$v^t(x)$	the pixel corresponding to location x in the t^{th} resized frame
η	the learning rate
\times	Not Applicable
\checkmark	Applicable
$.*$	Elementwise multiplication
\mathbb{R}^d	Set of Real number of d dimension
\emptyset and θ	Set of learnable parameters of deep models
λ	Lagrangian multiplier
\cup	Union
\cap	Intersection
Tanh	The activation function Tanh
L	Loss function
$Sum(.)$	The sum of all elements of the feature map(s)
\odot	Represents elementwise multiplication
nu	Hyperparameter of OC-SVM which defines the percentage of training samples will be treated as outliers.
$Concatenate()$	Concatenate the feature maps.
$CONCATE()$	Concatenate the feature maps.
\oplus	Concatenate the feature maps.
	Max Pooling
Σ	Sum of all the elements of the given feature map(s)
\otimes	Convolution Operation
$argmin$	Returns the input for minimum output.

List of Abbreviations

CA	Crowd Analysis
CCDE	Crowd Counting and Density Estimation
CCA	Crowd Congestion-level Analysis
CBA	Crowd Behaviour Analysis
CBP	Crowd Behaviour Prediction
CNN	Convolutional Neural Network
Conv2D/2DCNN/Conv_2D	2D Convolutional Neural Network
Conv3D/3D CNN/Conv_3D	3D Convolutional Neural Network
FCN	Fully Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
ConvLSTM	Convolutional Long Short-Term Memory
GAN	Generative Adversarial Neural Network
AI	Artificial Intelligent
HOG	Histogram of Oriented Gradients
HOF	Histogram of Optical Flow
LBP	Local Binary Patterns
AMS-CNN	Attentive Multi-Stream CNN
DMR	Density Map-based Regression
SCR	Single Count-based Regression
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
OC-SVM	One Class Support Vector Machine
SVR	Support Vector Regression
GLCM	Gray Level Co-occurrence Matrix
SIFT	Scale-invariant Feature Transform
GPR	Gaussian Process Regression
GMM	Gaussian Mixture Model
MuST-POS	Multiscale Spatial -Temporal 3D Atrous Net with PCA guided OC-SVM
MuST-3AN	Multiscale Spatial -Temporal 3D Atrous Net

TS-MDA	Two-Stream Multiscale Deep Architecture
AMS-CNN	Attentive Multi-Stream CNN
ML	Machine Learning
OCC	One Class Classification
MCC	Multi Class Classification
HOT	Histogram of Oriented Tracklets
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
FPR	False Positive Rate
ER	Error Rate
NA	Not Applicable
Mean-ACC	Mean Accuracy
Acc	Accuracy
MOG	Mixture of Gaussian
FV	Frame Volume
RGB	Red Green Blue
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
ReLU	Rectified Linear Unit
AP	Average Pooling
3DAP	3D Average Pooling
MP	Max Pooling
3DMP	3D Max Pooling
FC	Fully Connected
DC	Densely Connected
GAP	Global Average Pooling
Dil_ConvLSTM2D	Dilated Convolutional LSTM 2D
Dil_Conv3D	Dilated 3D Convolution
TADM	Temporal Attentive Density Map
FADM	Foreground Attentive Density Map
SADM	Spatial Attentive Density Map