

**MACHINE LEARNING-BASED CROWD DETECTION AND
CLASSIFICATION FOR SAFETY CONTROL SYSTEM**

A Thesis

Submitted

In Partial Fulfilment of the Requirements

for the Degree of

MASTER OF TECHNOLOGY

In

ADVANCED COMPUTING AND DATA SCIENCE

Submitted by:

Nida Khan

(2001209001)

Under the Supervision of:

Dr. Mohd Haroon

(Associate Professor)



Department of Computer Science & Engineering

Faculty of Engineering

INTEGRAL UNIVERSITY, LUCKNOW, INDIA

JULY, 2022



INTEGRAL UNIVERSITY

इंटीग्रल विश्वविद्यालय

Accredited by NAAC. Approved by the University Grants Commission under Sections 2(f) and 12B of the UGC Act, 1956, MCI, PCI, IAP, BCI, INC, CoA, NCTE, DEB & UPSMF. Member of AIU. Recognized as a Scientific & Industrial Research Organization (SIRO) by the Dept. of Scientific and Industrial Research, Ministry of Science & Technology, Government of India.

CERTIFICATE

This is to certify that **Ms. Nida Khan** (Roll No. 2001209001) has carried out the research work presented in the dissertation titled “**Machine Learning-based crowd detection and classification for safety control system**” submitted for partial fulfillment for the award of the **Master of Technology Advanced Computing and Data Science** from **Integral University, Lucknow** under my supervision.

It is also certified that:

- i. This dissertation embodies the candidate’s original work and has not been earlier submitted elsewhere for the award of any degree/diploma/certificate.
- ii. The candidate has worked under my supervision for the prescribed period.
- iii. The dissertation fulfills the requirements of the norms and standards prescribed by the University Grants Commission and Integral University, Lucknow, India.
- iv. No published work (figure, data, table, etc.) has been reproduced in the dissertation without the express permission of the copyright owner(s).

Therefore, I deem this work fit and recommend for submission for the award of the aforesaid degree.

Dr. Mohammad Haroon
(Associate Professor)
Department of CSE,
Integral University, Lucknow

Date:

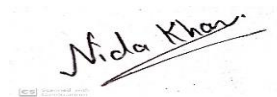
Place: Lucknow

DECLARATION

I hereby declare that the dissertation titled “**Machine Learning-Based Crowd detection and classification for safety control system**” is an authentic record of the research work carried out by me under the supervision of **Dr. Mohd Haroon**, Department of Computer Science & Engineering, for the period from September 2021 to July 2022 at Integral University, Lucknow. No part of this dissertation has been presented elsewhere for any other degree or diploma earlier.

I declare that I have faithfully acknowledged and referred to the works of other researchers wherever their published works have been cited in the dissertation. I further certify that I have not willfully taken other’s work, para, text, data, results, tables, figures, etc. reported in the journals, books, magazines, reports, dissertations, thesis, etc., or available at websites without their permission, and have not included those in this M.Tech dissertation citing as my own work.

Date:



Signature

Nida Khan

Enroll. No. **1300101376**

RECOMMENDATION

On the basis of the declaration submitted by “**Nida Khan**”, a student of M.Tech CSE (Advanced Computing and Data Science), successful completion of Pre presentation on 24/06/2022 and the certificate issued by the supervisor. **Dr. Mohd Haroon**, Associate Professor, Computer Science, and Engineering Department, Integral University, the work entitled “**Machine Learning-based Crowd detection and classification for safety control system**”, submitted to the department of CSE, in partial fulfillment of the requirement for the award of the degree of Master of Technology Advanced Computing and Data Science, is recommended for examination.

Program Coordinator Signature

Dr. Faiyaz Ahmad

Dept. of CSE

Date:

HOD Signature

Mrs. Kavita Agrawal

Dept. of CSE

Date:

COPYRIGHT TRANSFER CERTIFICATE

Title of the Dissertation: **Machine Learning-based crowd detection and classification for safety control system**

Candidate Name: **Nida Khan**

The undersigned hereby assigns to Integral University all rights under copyright that may exist in and for the above dissertation, authored by the undersigned and submitted to the University for the Award of the Master of Technology Advanced Computing and Data Science degree.

The Candidate may reproduce or authorize others to reproduce material extracted verbatim from the dissertation or derivative of the dissertation for personal and/or publication purpose(s) provided that the source and the University's copyright notices are indicated.

NIDA KHAN

ACKNOWLEDGEMENT

I am highly to the Head of the Department of Computer Science and Engineering for giving me proper guidance and facility for the successful completion of my dissertation.

It gives me great pleasure to express my deep sense of gratitude and indebtedness to my guide **Dr. Mohd Haroon, Associate Professor, Department of Computer Science and Engineering**, for his valuable support and encouraging mentality throughout the project. I am highly obliged to him for providing me this opportunity to carry out the ideas and work during my project period and helping me to gain the successful completion of my Project.

I am also highly obliged to the Head of Department, **Mrs. Kavita Agrawal (Head of Department of Computer Science and Engineering)** and PG Program Coordinator **Dr. Faiyaz Ahmad, Assistant Professor, Department of Computer Science and Engineering**, for providing me all the facilities in all activities and for his support and valuable encouragement throughout my project.

My special thanks are going to all of the faculty for encouraging me constantly to work hard on this project. I pay my respect and love to my parents and all other family members and friends for their help and encouragement throughout this course of project work.

Date:

Place: Lucknow

TABLE OF CONTENT

CONTENT	PAGE NO.
Title Page	i
Certificate/s (Supervisor)	ii
Declaration	Iii
Recommendation	iv
Copyright Transfer Certificate	v
Acknowledgment	vi
List of Tables	x
List of Figures	xi-xii
List of Abbreviations and Symbols	xiii
Abstract	xv
Chapter 1: Introduction	1
1.1 Crowd Analysis	3
1.2 Need for Crowd Analysis	4
1.3 Factors Affecting Crowd Analysis	4
1.4 Approaches to Crowd Behavior Analysis	5
1.4.1 Object-Based Approach	5
1.4.2 Holistic Approach	5
1.5 Applications of Crowd Behavior Analysis	6
1.6 Machine Learning	6
1.6.1 Supervised Learning	7
1.6.2 Unsupervised Learning	9
1.6.3 Reinforcement Learning	10
1.7 Deep Learning	11
1.8 Related work	11

1.9	Review of Algorithms Used in Crowd Behavior Analysis	12
1.9.1	Support Vector Machine	12-13
1.9.2	Convolutional Neural Network	13-15
1.10	Dissertation Outline	15-16
Chapter 2: Literature Review		17
2.1	Literature Summary	18 – 25
2.2	Conclusion	26
Chapter 3: System Model & Architecture		27
3.1	Model Architecture	28
3.1.1	Convolution and Deconvolution	28
3.1.2	Long Short-term memory cells (LSTM)	29
3.1.3	Convolutional LSTM	30
3.1.4	Final Model Architecture	31
3.2	Regularity Score	32
3.3	Proposed Methodology	33
3.3.1	The Approach	33
3.3.2	Preprocessing	34
3.4	Training and Testing	35 – 36
3.5	Model for crowd detection behavior	38
3.5.1	Crowd dataset	38
3.5.2	Feature Selection	40
3.5.3	Data Flow	41
3.5.4	Classification	41
3.5.5	Data Structure	42
3.5.6	Feature Statistics	44 – 46

Chapter 4: Result Analysis and Comparative Study	47
4.1 Result Analysis	48
4.1.1 Evaluation Metrics	48
4.1.1.1 Mean Absolute Error	48
4.1.1.2 Mean Squared Error	49
4.1.1.3 Area under ROC curve (AUC) and EER	50
4.1.2 Dataset	51
4.1.3 Quantitative Results (Area under ROC curve (AUC) and EER)	52
4.1.3.1 Qualitative Analysis (Area under ROC curve (AUC) and EER)	53
Chapter 5: Conclusion and Future Work	58
5.1 Conclusion	59
5.2 Future Work	60
References	
Appendix	
Plagiarism check report	
Publications from this work	
Publications	

LIST OF TABLES

Table 2.1: The Summary of Algorithms used in some past researches.	24 - 25
Table 4.1: Comparative study of different methods and our approach for crowd abnormal behavior	53
Table 4.2: Comparative study of different methods for crowd abnormal behavior	54
Table 4.3: Model comparison by MSE	55
Table 4.4: Model comparison by RMSE	55
Table 4.5: Model comparison by MAE	55

LIST OF FIGURES

Figure 1.1: Applications of crowd Analysis	6
Figure 1.2: Phases in Machine Learning	7
Figure 1.3: Supervised and Unsupervised learning	9
Figure 1.4: Deep learning neural networks	11
Figure 1.5: Support Vector Machine	13
Figure 1.6: Detailed architecture of Convolutional Neural Network	15
Figure 2.1: General Structure of Crowd detection system	19
Figure 3.1: Convolution and Deconvolution operation using a 3×3 kernel	29
Figure 3.2: A Standard LSTM cell	30
Figure 3.3: Final model architecture	31
Figure 3.4: Algorithmic flow of our crowd analysis framework. Each input is reconstructed using our neural network and a reconstruction cost is calculated. Furthermore, this cost is used to predict if the input contains normal events or abnormal events.	34
Figure 3.5: Training workflow of our abnormal event detection framework. A sequence of 10 images is fed as an input to our model. The model reconstructs the given input and per pixel Euclidean loss is calculated from reconstructed sequence and raw frames. This loss is back propagated for the network to learn better reconstruction.	36
Figure 3.6: Testing and deployment workflow of our abnormal event	36

detection framework	
Figure 3.7: Data Flow Diagram	37
Figure 3.8: Dataset for detection	40
Figure 3.9: Data Flow	41
Figure 3.10: Data View	42
Figure 3.11: Image Detection for model development	43
Figure 3.12: Feature Statistics	44
Figure 3.13: Line Plot	45
Figure 4.1: Formula for Mean Absolute Error	48
Figure 4.2: Diagram representing Mean Absolute Error	49
Figure 4.3: Formula for Mean Squared Error	49
Figure 4.4: Diagram representing ROC(AUC) and EER	50
Figure 4.5: Normal (top) and abnormal (bottom) samples of UCSD ped1 dataset	51
Figure 4.6: Normal (top) and abnormal (bottom) samples of UCSD ped2 dataset	52
Figure 4.7: Visualization of various frames from UCSD dataset	52
Figure 4.8: Visualizing the regularity score on Ped1 Test video 1, video 8, video 24 respectively. Abnormal events are marked within red bounding boxes. And abnormal events have a red background in the graph.	54

LIST OF ABBREVIATIONS AND SYMBOLS

ML	Machine Learning
AI	Artificial Intelligence
SVM	Support Vector Machine
CNN	Convolutional neural Network
kNN	K nearest neighbors
DBSCAN	Density-based spatial clustering of applications with noise
RNN	Recurrent Neural Network
MMH	Maximum marginal hyperplane
ReLU	Rectified Linear Unit
LBP	Local Binary Pattern
GLCM	Gray-level Co-occurrence Matrix
MCNN	Multi-Column Convolutional Neural Network
GAN	Generative adversarial network
2D CNN	2-Dimensional Convolutional Neural Network
3D CNN	3-Dimensional Convolutional Neural Network
CD	Crowd divergence
CC	Crowd Counting
CMS	Crowd Monitoring System
LDA	Linear Discriminant Analysis
SFM	Social Force Model
MII	Motion Information Images
SD-CNN	Scale Driven Convolutional Neural Network
DISAM	Density Independent and Scale Aware Model

LSTM	Long Short-term memory
ConvLSTM	Convolutional Long Short-term memory
FC-LSTM	Fully connected Long Short-term memory
UCSD	UCSD Anomaly Detection Dataset
AUC	Area Under ROC Curve
ROC	Receiver Operator Characteristic
EER	Equal Error Rate
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error

ABSTRACT

A crowd is a recognizable group of people or anything included in a community or society. Many academic disciplines, including sociology, civil engineering, and physics, among others, are very familiar with the crowd phenomena. It has currently evolved into the area of computer vision research that is the most active and popular. The three processing stages that generally makeup crowd analysis are pre-processing, object detection, and event or behavior identification. These phases are pre-processing, object detection, and event recognition. The three processing stages that generally makeup up crowd analysis are pre-processing, object detection, and event or behavior identification. These phases are pre-processing, object detection, and event recognition. This study provides a crowd analysis paradigm and a taxonomy of the most common crowd analysis techniques. It could be beneficial to researchers and act as a solid introduction to the field of the work that has been done. We also provide a scalable and effective way for identifying unusual events in videos. We employ unsupervised learning to develop a measure of the regularity of the input video because real-world scenarios lack labeled data. The regularity score produced by our neural network is utilized to decide if the input sequence's events are normal or aberrant. In movies of crowded settings, we suggest a spatiotemporal architecture for anomalous event identification. Our architecture develops efficient and later methods for encoding both spatial and temporal data. The detection accuracy of our system is comparable to state-of-the-art methods, according to experimental results on the UCSD (ped1 and ped2) benchmarks.

CHAPTER – 1

INTRODUCTION

One of the wonderful human mental capacities that develop in early development is the ability to perceive objects, activities, and events visually. Despite obstructions and confusion, the human visual system can complete computationally challenging tasks like counting or finding similar items in a scene with apparent ease. Researchers in the field of computer vision has been creating mathematical tools that are comparable to human abilities for object detection, object recognition, and behavior discovery in visual settings. Understanding human behavior is particularly important in all of these endeavors for both practical use and academic inquiry. It opens the door to understanding how human visual cognition and social abilities evolve. Additionally, several applications, including surveillance and human-computer interaction, benefit from scientific knowledge.

People perform actions in groups because human activities are frequently social. Activities associated with complexity, such as self-organizing or emergent behaviours, which come up as a result of interactions between people, are found in crowds of many people. The visual comprehension of crowd scenes is a difficult problem because of these complex issues. Our research and observations show that even humans would have a difficult time executing basic visual identification tasks in busy settings, including counting or tracking items. This is due, in part, to the fact that scenes with more things present have more opportunities for visual inspection and recognition. Additionally, additional people would increase the chance of occlusion and increase the clutter in a scene. This prevents the human brain's parallel perceptual and integration pathways from using the pop-up effect to draw focus on the target object. Because it is difficult to visually recognize congested situations, the traditional pipeline of object detection, tracking, and behavior analysis in computer vision research would perform poorly. Despite all the achievements in computer vision research, crowd analysis and human activity detection have remained unresolved problems because of the intrinsic complexity

and broad variability seen in crowded contexts. Crowd behavior analysis in computer vision research opens up new application domains, such as the automatic detection of riots or other chaotic crowd behavior, the localization of abnormal regions in scenes for high-resolution analysis, the understanding of group behavior, and performance evaluation. This thesis develops algorithms and representations that efficiently describe crowd detection and classification.

1.1 CROWD ANALYSIS:

A crowd as we know is a term that is used to refer to a collection or group of individuals. All of us have witnessed or been a part of the crowd at some point in time. Hence, we can safely say that crowd is a very important part of our lives. Most of the places that we visit such as markets, streets, parks, stadiums, malls, etc. are flooded with people all the time. This immense relationship with the crowd brings us to the most important task i.e., crowd analysis.

Crowd analysis is the process of understanding the overall behavior of the crowd and using that understanding to make important inferences such as estimation of the count of people in the crowd or the nature of the crowd. The analysis of the crowd can help us to forecast multiple real-world situations such as mob lynching, traffic jams, riots, stampedes, violence, etc. Using these important predictions, we can notify corresponding authorities beforehand to take preventive actions. For example, if we can predict that the crowd near a crossroad is not normal the traffic authorities can plan accordingly to prevent any traffic jams at that location and hence avoid any probable accidents.

Images or videos of the audience are used as the crowd's input when analysing it. The behavior of the crowd is made of the aggregate behavior of each individual in that crowd. We need to perform an analysis of this collective emotion to understand the

crowd behavior. We need to remember a few things:

- The crowd emotion or behavior is not just the summation of the individual emotions
- People in a crowd have different positions and are usually moving in different directions.
- It is very difficult to determine the normal behavior of a crowd and then compare it with the current behavior.

1.2 NEED FOR CROWD ANALYSIS

Crowd management is of extreme importance. The unmanaged crowd has always been the cause of dangerous situations such as stampedes and accidents. With the growing crowd each day, we need some concrete steps and techniques to bring the crowd in control efficiently and effectively. Today we have cameras installed in public places giving us enormous data regarding the crowd in the form of images and videos. For example, video surveillance cameras are installed at airports, stadiums, and train stations. This gives an added impetus to crowd analysis by making the input available in such a huge amount. Not only do we have data in quantity but also of varying quality that is collected over a wide range of time.

But even though we have input at our hands we have not been able to use the resources optimally. There is still a huge scope in anomalous crowd detection and crowd behavior analysis with better techniques to achieve higher efficiency and accuracy of results.

1.3 FACTORS AFFECTING CROWD ANALYSIS:

Numerous factors come into play while performing crowd behavior analysis. Below is a list of a few of these elements.

- Visual Occlusions

- Severe Clutters
- Scale
- Computational complexity
- Size
- Perspective
- Boundary restrictions
- Determination of abnormal and normal behavior
- Irregular illumination conditions
- Image resolutions

Other challenges that we face during crowd analysis are that getting good quality images round the clock from cameras is difficult Also, the processing time is not real-time

1.4 APPROACHES TO CROWD BEHAVIOUR ANALYSIS

There are two approaches that are followed in Crowd Behavior Analysis.

These are:

1.4.1 OBJECT-BASED APPROACH

In this approach, the crowd is taken to be a group of individual persons, and to estimate the behavior of the crowd we need to track the motion and context of each of these individuals separately. Then information from each individual is aggregated in some logical fashion to arrive at the overall crowd estimation.

1.4.2 HOLISTIC APPROACH

This approach looks at the crowd as one whole entity instead of the combination of multiple individual entities which allows us to measure the emotion and behavior of the entire crowd as a whole. Following this

approach, we evaluate the group of people as a whole.

1.5 APPLICATIONS OF CROWD BEHAVIOUR ANALYSIS

The field of crowd analysis has immense scope and a wide range of real-world applications. These applications are:



Figure 1.1: Applications of crowd Analysis

- Population Counting
- Management of public events
- Disaster management
- Safety Monitoring
- Military management
- Suspicious activity detection

1.6 MACHINE LEARNING

It is a subset of computing where the machines are not pre-programmed to solve any kind of problem. Instead, the machine is subjected to multiple instances that are used by the machine to develop its model or function based

on its learning and then uses that model to make predictions on new or unprecedented instances. Thus, machine learning has two phases.



Figure 1.2: Phases in Machine Learning

Software systems can predict outcomes more correctly with the use of machine learning (ML), a type of artificial intelligence (AI), without needing to be explicitly told to do so. Machine learning algorithms use historical data as input to forecast new output values. Supervised learning, Unsupervised learning, and Reinforcement learning are the three traditional methodologies in machine learning.

1.6.1 SUPERVISED LEARNING

Supervised learning is a branch of computing where we provide multiple examples to a machine to learn and then let the machine build a model based on that learning. Then this model can be used to make predictions about new instances which were not provided to it as examples while training. Hence, instead of programming the machine or providing the rules to the machine for making predictions, we let the machine learn those rules on its own by learning from the example instances. Thus, machine learning has two phases i.e., the training phase where the machine learns an m, model, and the prediction phase where the machine uses and implements that model to make predictions. Methods like Support Vector Machine (SVM), Decision Tree, and Random Forest, Bayesian classifier, etc. are examples of supervised learning algorithms.

There are two categories of supervised learning algorithms:

Classification: There is a classification challenge when the output variable is a category, such as "Red" or "Blue," "Illness" or "No Disease."

Regression: When the output variable, such as "dollars" or "weight," has a real value, a regression issue arises.

Data that has been "labeled" is used in supervised learning. This suggests that some data has already been assigned the right response.

Types of supervised learning:

- Regression
- Logistic Regression
- Classification
- Naive Bayes Classifiers
- K-NN (k nearest neighbors)
- Decision Trees
- Support Vector Machine

Advantages:

- Learning under supervision enables data collection and generates data output based on prior experiences.
- uses experience to assist optimize performance criteria.
- A variety of real-world computation issues are solved with the use of supervised machine learning.

Disadvantages:

- Big data classification can be difficult.
- A lot of calculation time is required for supervised learning training. As a result, it takes some time.



Steps

1.6.2 UNSUPERVISED LEARNING

When we supply instances to the machine without labels for building the model, the machine performs unsupervised learning. It uses various types of algorithms that can find relations between instances supplied and create a model based on that learning. Again, the model built will be used to make predictions on new unknown instances. This is the reason it is called unsupervised learning as there are no labels in the training phase to guide the model building. It also has the same two phases. Examples of unsupervised learning algorithms are Hierarchical clustering, K means, and DBSCAN.

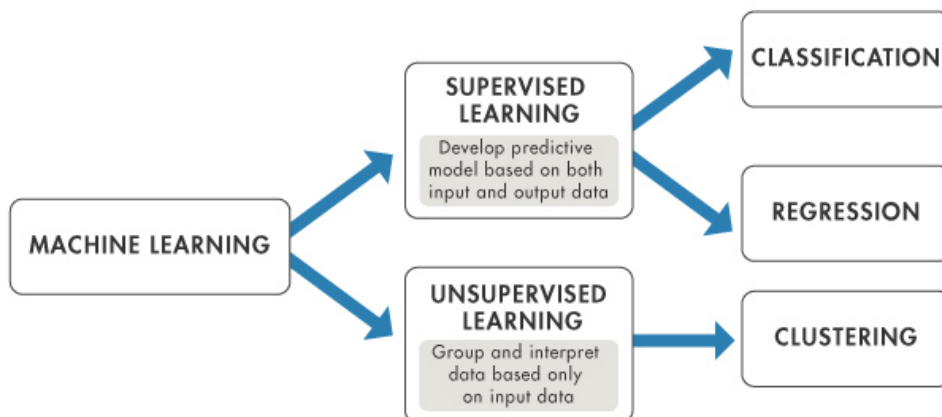


Figure 1.3: Supervised and Unsupervised learning

Unsupervised learning is divided into two groups by the algorithms they fall

under:

Clustering: Identifying the natural groupings in the data, such as classifying clients based on their purchasing patterns, is a clustering problem.

Association: When you wish to find rules that broadly characterize your data, such as "those who buy X also tend to buy Y," you have an association rule learning problem.

1.6.3 REINFORCEMENT LEARNING

This approach to machine learning is different from the others. In this approach, the machine learns by interacting with the environment. The machine performs one step at a time and records the responses of the environment. In case the response of the environment is favorable the machine is rewarded else not. This helps the machine learn the right steps to perform in a given environment. This learning is called reinforcement.

There are two distinct categories of reinforcement:

Positive - Positive When an event happens as a result of a certain behavior, this is known as reinforcement, and it enhances the frequency and strength of the behavior. In other words, it influences behavior favorably.

Reinforcement learning has the following benefits:

- Performance is maximized when change is sustained over a lengthy period.
- A surplus of states brought on by excessive reinforcement may have a negative impact on the outcomes.

Negative – Negative reinforcement is the strengthening of behavior as a result of stopping or avoiding a negative condition.

Reinforcement learning benefits include:

- Enhances Behavior
- Show disobedience to the required minimal level of performance
- It only offers what is necessary to meet the minimum standard of behavior.

1.7 DEEP LEARNING

Deep learning is a branch of machine learning that has gotten inspiration from our biological brains. Just like a neural network of a human brain, the models of deep learning use artificial neural networks. This branch is called deep learning as its networks contain multiple layers. Each layer is assigned a specific task. Lower layers are used to extract the low-level features of the input whereas the higher layers determine the high-level and detailed features. Most recent examples of these models are the Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

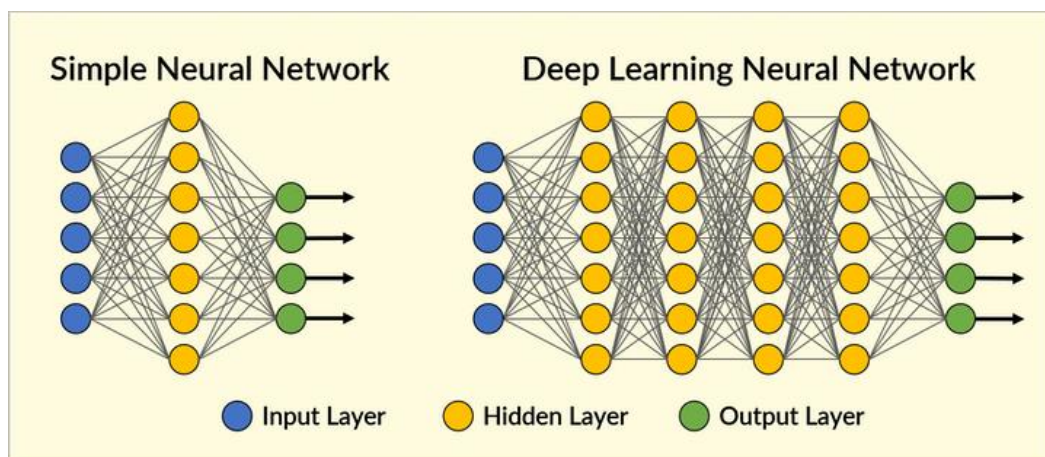


Figure 1.4: Deep learning neural networks

1.8 RELATED WORK

A review of the literature on crowd analysis proves that a lot of work has been done in the past and many types of research are still being carried out in the field of crowd behavior analysis for various purposes like public crowd management, disaster

management, military management, suspicious activity detection, etc. Different kinds of methodologies and algorithms have been used by past researchers.

1.9 REVIEW OF ALGORITHMS USED IN CROWD BEHAVIOUR ANALYSIS

1.9.1 SUPPORT VECTOR MACHINES (SVM)

SVM has been widely used as the main classifier in many of the researches in crowd emotion analysis. Other variations of SVM are also implemented in some research. SVM is a type classification model whose highlight is that it can classify not just linear data but also nonlinear data. First, the data points are plotted onto a multi-dimensional space. Then a decision boundary is found that can optimally segregate the data points between two different classes. This decision boundary is a linear hyperplane. Although there could be many such planes this hyperplane is a plane in that multi-dimensional space that is at maximum distance (margin) from the support vectors. Due to this reason, it is aptly called maximum marginal hyperplane (MMH). This is done to minimize classification errors for new data points. Now the support vectors are those data points that are nearest to the hyperplane. Once this hyperplane is found any new data point can be correctly assigned to its class using the decision boundary.

A separating hyperplane has the following equation:

$$\mathbf{W} \cdot \mathbf{X} + \mathbf{b} = 0$$

Where \mathbf{W} is a weight vector, such as, $\mathbf{W} = \{w_1, w_2, \dots, w_n\}$;

\mathbf{N} is the number of attributes;

As a scalar, \mathbf{b} (also called bias)

\mathbf{X} is the data point

If $n=2$ i.e., 2-dimensional data points are considered, then the MMH can be written

as:

The value of this equation is 0 for the points lying on the line and the value is +1 or -1 for points that are on either side of the line. Hence, we can classify new data points based on the value obtained from this equation.

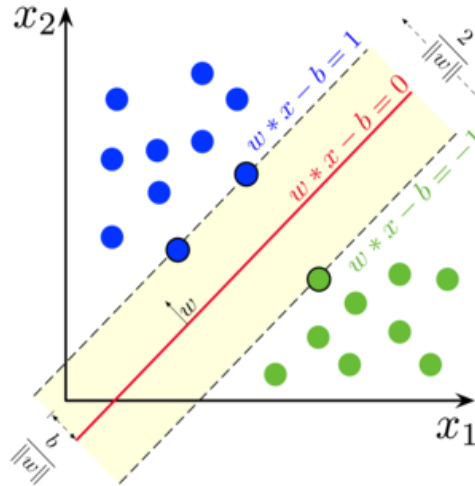


Figure 1.5: Support Vector Machine

The following is a discussion of the benefits and drawbacks of SVM [31]:

The disadvantage of SVM: Sometimes training the fastest SVMs can take huge lengths of time.

Advantages of SVM:

- High Accuracy: The accuracy of SVM is extremely high which is attributed to their capability to create complex nonlinear decision hyperplanes.
- Minimal Overfitting: Also, as compared to other methods it is very less likely for SVMs to suffer from overfitting.

1.9.2 CONVOLUTIONAL NEURAL NETWORK

CNN is a kind of artificial neural network that has been specifically designed for analyzing images encouraged by the working of human vision. This makes CNN the aptest machine learning tool to be used for the analysis of crowd images or videos. This is proven by the vast application of CNN and its variations in crowd counting

and crowd event prediction. The layers in the CNNs are 3 3-dimensional unlike the regular neural networks, the neurons in one layer are not necessarily connected to all the neurons of the next layer. CNN comprises two components: (i) the feature extraction part and (ii) the classification part. The former is also called the hidden layer part. It is where the features are drawn out by employing convolutions and pooling. In other terms, this part is responsible for the learning in the model. The classification part is where we calculate the likeliness of the image as a part of a particular class. This is achieved by assigning probabilities for the same.

Convolution is the process of fusing two functions to create a brand-new third function. It is performed in CNN by applying kernels or filters repeatedly to the image that has to be classified. The kernel is a matrix that is moved over the input image. We obtain a feature map by performing a matrix multiplication at each location. The receptive field refers to the area of the filter used. Dynamic receptive fields have been predominantly used in crowd counting and density estimation. The entire operation is in 3 dimensions due to the colors in an image. Each filter produces a separate feature map. All the feature maps constitute the output of the convolution layer of CNN. This output is then fed into an activation function such as the ReLU activation function. To prevent the feature map from contracting, padding is used.

A CNN has four important hyperparameters:

- the kernel size
- the filter count
- stride (how big are the steps of the filter)
- padding

After her convolution 1, Ayer a pooling layer is added to reduce the dimensionality. At last, the classification layer is included which is a set of fully connected layers. But these

layers can accept input only as 1-dimensional data.

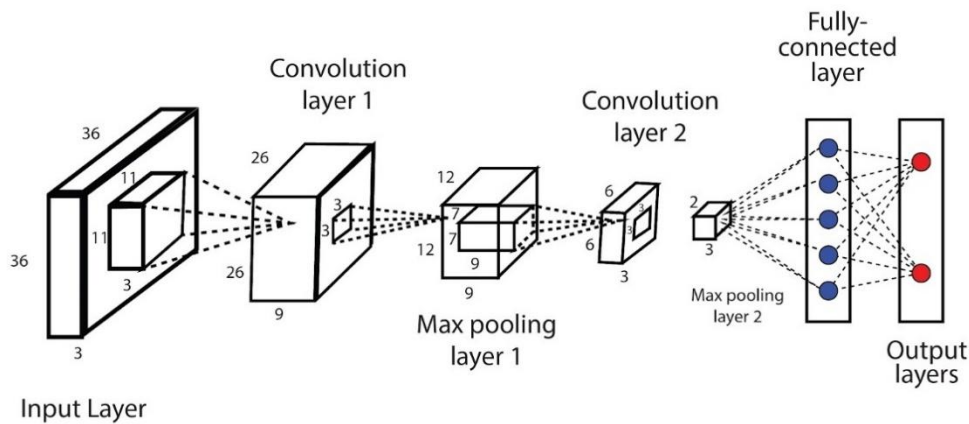


Figure 1.6: Detailed architecture of Convolutional Neural Network

1.10 DISSERTATION OUTLINE

The remaining part of the dissertation is organized as follows:

Chapter 2

This chapter gives a summary of the literature in the field of crowd analysis. A literature review was done in order to have a clear understanding of the topic, the problem statement, and the progress done so far.

Chapter 3

This chapter introduces the methodology used in our research. It also talks about the purpose and the objectives of the research. We present the model approach along with the flow diagram and the step-wise algorithm.

Chapter 4

In this chapter, we discuss the experimental results and analysis. It also presents the evaluation metrics used in our experiment.

Chapter 5

In this chapter conclusion and some of the future scopes are discussed of this work.

CHAPTER – 2

LITERATURE REVIEW

2.1 LITERATURE SUMMARY

Wafee Mohib Shalash et al. created a mobile-based crowd anomalous behavior identification and management system. IP surveillance cameras in the vicinity of the entrance gate(s) are connected to a server-side application, which is then used by a mobile application with various user permissions to get an alarm from the server-side application if crowd levels or unusual movement grow. All system users might be connected and warned quickly using the proposed architecture in the event of abnormal crowd behavior. The system has been tested using interface, unit, and usability tests to ensure that it will work and that users will be able to reply appropriately. The condition is satisfactory, according to the testing findings (**Wafaa Mohib, 2019**).

Atika Burney and Tahir Q. Syed wrote: "Crowd Video Classification Using Convolutional Neural Networks." A 2D CNN can categorize films using 3-channel image maps generated using spatial and temporal information, which reduces the amount of space and time required for video analysis when compared to a conventional 3D CNN. We were able to improve the model's accuracy by testing it against the dataset provided by them using current methods. The mid-level descriptors of the groups described in this study can be used to simplify crowd recording classification, but the classification accuracy can also be increased by treating the crowd as a whole. The extra processes of recognizing groups, computing their characteristics, and then combining them to define the crowd can be reduced to a single step of calculating the features of the crowd (**Atika Burney, 2016**).

"Single-Image Crowd Counting Via Multi-Column Convolutional Neural Network" by Yingying Zhang et al. aims to create a system that accurately estimates the crowd count from a single image, regardless of crowd density or viewpoint. This study includes 1198 photographs with over 330,000 annotations on the heads. The proposed MCNN

model, in particular, outperforms all previous methods. Furthermore, testing shows that once the model has been trained on one dataset, it may readily be transferred to another (Yingying Zhang,2016).

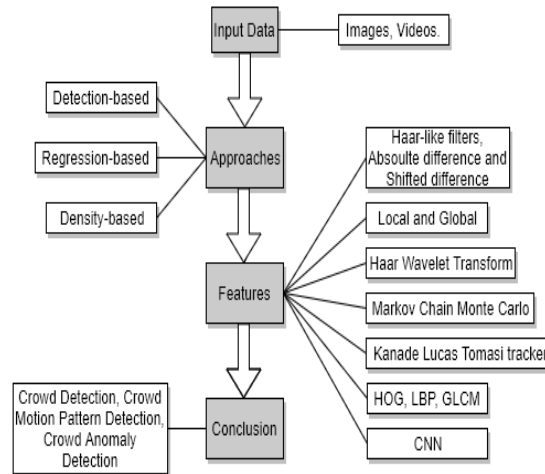


Figure 2.1: General Structure of Crowd Detection System from [5]

Mayur D. Chaudhari and Archana S. Ghotkar conducted "A Study on Crowd Detection" for safety concerns. The purpose of this research is to determine the crowd density in witness footage. It is feasible to estimate crowd size and density using face and detecting pattern recognition. Human faces have so many factors, such as colour, location, and orientation, that identifying a face in a crowd can be challenging. The counting performance has steadily improved thanks to the Deep Convolutional Neural Network. The components of a crowd detection system are depicted in Figure 2.1, which include input data, techniques, features, and a conclusion. Detection-based methods, regression methods, and density-based methods are three types of approaches to tackling the problem (Mayur D.,2018).

Akbar Khan et. al, are the members of the group. Deep Convolutional Neural Networks for Crowd Monitoring and Localization Research The real-time monitoring of huge groups of individuals may be accomplished through the application of machine learning

methods and approaches. Methods of crowd-monitoring have been thoroughly examined. These models employ a convolutional neural network driven by size as the basis for their crowd counting and localization capabilities, and we conclude that they are the only ones of their kind for dense crowd pictures. They can be used to identify the heads in a photograph based on the density and scale of the image (**Akbar Khan,2020**).

Beibei Song and Rui Sheng, "Multiscale GAN Network Combined with Deep Optical Flow," "Crowd Counting and Abnormal Behavior Detection," Multiscale feature extraction is confined to crowd counting, and a crowd counting model based on the multiscale network is provided in this research. An embedded GAN module is created by combining the multibranch generation network with the regional discrimination network, and then connecting it to the multiscale module using a pyramid pooling structure. As a total, the model is trained using three different loss functions so that the model may increase its capacity to extract multiscale features from the expected picture and its ability to count accurately and robustly. The usefulness of the model has been demonstrated by a large number of qualitative and quantitative studies using a public dataset of three crowd counts (**Beibei Song,2020**).

According to Dongyao Jia, who presented a crowd density classification approach based on pixels and texture data, crowd density classification has been a difficult problem in the field of computer vision, with various applications in the public and commercial sectors. Although the crowd density classification and recognition technique has been thoroughly investigated in the past, there are still difficulties of inaccuracy, low resilience, and inefficiency that must be addressed. This study proposes adaptive crowd density categorization based on pixels and texture data. In order to extract the texture features of crowd images, the WorldExpo'10 dataset is utilized to

integrate many texture characteristics, such as the local binary pattern (LBP), and the Gray-level co-occurrence matrix (GLCM), the Gabor, Haar-like, and Wavelet groups. Experiments have shown that the suggested technique has a classification rate of 98.2 percent (**Dongyao jia, 2021**).

Sherif Elbishlawi et. al, "Deep Learning-Based Crowd Scene Analysis Survey," (Deep Learning-Based Crowd Scene Analysis Survey) A overview of deep learning-based algorithms for assessing congested situations is presented in this work. The approaches under consideration are divided into two categories: (1) crowd counting and (2) crowd activity recognition. Furthermore, datasets from crowd scenes are examined. This work also provides assessment criteria for crowd scene analysis tools, which are discussed in further detail below. With the help of this statistic, you can determine how much of a difference there is between the computed and real crowd counts in crowd scene recordings. Crowd divergence (CD) is offered as a new performance indicator for the crowd scene analysis approach to give an accurate and robust evaluation of its performance. This may be done by comparing the actual trajectory/count to the projected trajectory/count and calculating the difference. The GAN framework and context-aware are promising prospects in crowd scene analysis, according to the results of this survey (**Sherif El bishlawi, 2020**).

Individuals are forced to contend with crowds or mass gatherings on an almost daily basis at a variety of places, including airports, hospitals, sports stadiums, and theme parks, to name just a few examples. The activities include a wide variety of realms, ranging from social and cultural to religious. In contrast to social gatherings and sporting events, it may not be feasible to escape the crowds that form during significant religious gatherings like

the Hajj and the Umrah. **(Zhan, B, 2008)**.

In order to keep the public safe, maintain a high pedestrian flow to avoid stampedes, provide better emergency services in the event of crowd-related emergencies, and optimize resources in order to provide good accessibility by avoiding congestion, an intelligent Crowd Monitoring System (CMS) is required. From a broader perspective, crowd management, monitoring, and analytics may be used in a number of scenarios **(Wang, Xiaofei, 2014)**. These include, but are not limited to, applications in the safety sector, emergency services, traffic flow and management in private and public areas, people counting and analyzing group behaviors, and other applications that are similarly swarm-based.

Because these applications are so important, there is a natural need for study and development in the field of crowd management and analysis, as well as the behavior of people inside crowds. This desire is due to the fact that there is a natural demand for these topics. This comprises the study of groups, counting and summarizing data, determining the density and making predictions, analyzing flow, predicting particular behavior, and monitoring mass. The recognition of the group applies generally and density estimation has been shown to be beneficial for the corresponding processes of intelligent analytics as well as various applications **(Cong, 2013)**.

A crowd is made up of a large number of people who have gathered in a disorganized or unmanageable way. It refers to a collection of people who have converged in a particular location. The crowd at a place of worship, for instance, will be distinct in comparison to the crowd in a commercial district. In addition, crowds change depending on the setting.

The context in which the word "crowd" is used reveals the nature of the gathering in terms of size, length, composition, inspiration, coherence, and proximity of the

individuals who make up the gathering. In today's world, video surveillance systems, the cost of whose equipment is relatively low, are being widely implemented in public places such as airports, subway stations, traffic intersections, schools, and many other types of establishments for the purpose of ensuring the safety of the general populace. However, the vast majority of today's video surveillance systems are only used as a recording system. This means that they are unable to recognize and analyze an odd event on their own, and it is also impossible for a person to monitor the screen continuously.

Crowd counting (CC) aims to count the number of objects, such as people, cars, cells, and drones in photographs or videos taken by them. Processing digital images, using machine learning, or even engaging in deep learning are only some of the available options for carrying it out. To be more precise, there are a variety of cutting-edge methods that may be used in the process of counting crowds. Some examples of these methods are clustering, density estimation, and counting by detection regression (**Zhan, Beibei, 2008**). Crowd counting is a difficult scientific topic that has to be addressed since it has such a broad range of applications, ranging from commercial to military goals, and because it is of great relevance in computer vision. A lot of scholars attempted to give extensive surveys and evaluations of earlier methodologies by taking into consideration a wide variety of crowd characteristics. These time-honored approaches to headcount estimation concentrate mostly on manually generated low-level crowd characteristics. These low-level characteristics are chosen, retrieved, and grouped before being fed into the regression model, which is then used to evaluate and minimize the loss function. Regarding this particular aspect, (**Zhan et al., 2008**) presented a complete study for the purpose of general crowd counting.

They focused mostly on reviewing eyesight and other types of impairments. Crowd modeling in vision-based issues is performed using information that has been retrieved

from visual data and is used for the purpose of crowd-event inference. On the other hand, non-vision techniques seek to explain and anticipate the accumulated consequences of crowd behavior by correcting the link between characteristics. Later on, experts concentrated their attention on crowd-counting models, paying particular attention to the shortcomings of these methods. The primary contribution that they made was the classification of crowd-modeling methods into three distinct categories: learned-appearance-based models, motion-flow-based models, and hybrid approaches. Further subdivision of the motion-flow-based models included optical-flow-based models, Lagrange-based approaches, and background-subtraction-based models.

TABLE 2.1

THE SUMMARY OF ALGORITHMS USED IN SOME PAST RESEARCHES

S.NO	TITLE	TECHNIQUE USED	FINDINGS
1	Crowd density classification method based on pixels and texture features	SVM classifier with K-means clustering iterative method	The proposed technique performs well with four datasets: World Expo'10, UCSD, Shanghai Tech A, and UCF CC 50 while dealing with different densities.
2	Crowd Detection Management System	Social Force Model (SFM) algorithm with LDA classifier	The findings of the tests indicate a satisfactory outcome.
3	Crowd Video Classification using Convolutional Neural Networks	Deep learning method CNN is used	Classify crowd videos using mid-level descriptors and also obtained accuracy if the descriptors are computed considering the whole crowd as a single entity.
4	Single-Image Crowd Counting via Multi-Column	Multi-column Convolutional Neural Network	Calculate the number of people in a crowd from a single photograph using

	Convolutional Neural Network	(MCNN)	arbitrary crowd density and arbitrary perspective.
5	A Study on Crowd Detection and Density Analysis for Safety Control	CNN	Calculate the concentration of the crowd in the videos of observers.
6	Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review	Scale Driven Convolutional Neural Network (SD-CNN) and DISAM model	Counting and localizing crowds in dense crowd images with the maximum degree of accuracy across several datasets.
7	Crowd Counting and Abnormal Behavior Detection via Multiscale GAN Network Combined with Deep Optical Flow	Multiscale GAN network	Crowd counting on a single image and the detection of anomalous behavior.
8	Deep Learning-Based Crowd Scene Analysis Survey	Survey of crowd counting and crowd activity detection techniques based on deep learning.	
9	Hajj Crowd Management Using CNN-Based Approach	CNN	Managing the large crowds and ensuring that stampedes and other overcrowding accidents are avoided.
10	Crowd Detection and Tracking in Surveillance Video Sequences	Kalman filter	Track location for each frame.
11	Crowd Behavior Classification based on Generic Descriptors	Graph-based descriptors	Crowd behavior analysis.

2.2 CONCLUSION

The purpose of performing the literature review is to highlight the research done in the past by other researchers in the field of crowd behavior analysis, abnormal crowd behavior detection, and crowd counting. Here we discuss all kinds of approaches employed by the researchers for this purpose. We have already stated that two approaches have been adopted, namely, object-based and holistic approaches. A lot of algorithms have been developed on the basis of object-based approaches such as Motion Information Images (MII) and optical flow vectors. In some researches gradient is used to create a model for individual motion in lines of Genetic Programming (GP) based classifier for easy hardware implementation. Object-based methodologies resulted in deep feature extraction keeping in mind each individual in the crowd is a separate entity. On the other hand, the holistic approach considers the entire crowd as a single unit. Most of the algorithms developed on this approach have implemented CNN. Even though CNN it is a powerful tool for analyzing crowd images and videos, many challenges such as occlusion, scale, size, perspective, etc., still exist. For resolving them huge spatial filters are generally used but at the cost of computational complexity. Crowd analysis is still a work in progress striving for more accuracy and efficiency than the past approaches.

CHAPTER – 3

SYSTEM MODEL & ARCHITECTURE

3.1 MODEL ARCHITECTURE:

In this section, we discuss the basic building blocks of our model and present how these blocks interact with each other to create our neural network architecture. Since our data is in the form of frames(images), using convolutional neural networks seems to be the logical choice as they are the standard choice when it comes to the image domain. Also, we have a series of frames which we give our model as input, recurrent models like RNN, and LSTMs are state of art in sequence modeling problems. Given this, we used a combination of both convolution and LSTMs as a building block for our Autoencoder. These types of blocks are known as Convolutional LSTMs or ConvLSTMs.

Let us briefly define what auto encoders are before continuing further.

Autoencoders: Neural networks that have been trained to recreate the input are called autoencoders. There are two components to the autoencoder:

1. **The Encoder:** which may learn effective representations of the input data (x) and is also known as the encoding $f(x)$. The input representation f is located in the bottleneck layer, which is the final layer of the encoder (x). Many times, embedding is another name for it.
2. **The Decoder:** Using the encoding in the bottleneck, the decoder reconstructs the input data using the formula $r = g(f(x))$.

Building blocks of our architecture:

3.1.1 Convolution and Deconvolution:

The primary goal of using convolution is to extract meaningful features from the images. Each convolutional layers contains a set of filters which are convolved with input feature maps to produce an output feature map which is propagated to the next layers. A convolutional network learns the values of these filters on its own during the training process, parameters such as the number of filters, filter size, and the number of layers before training

are user-defined and are mostly adjusted empirically. The goal of deconvolutional layers is to upscale the input feature maps into higher spatial dimensions. Here the layer learns the filters which will lead to minimum upscaling error. Figure 4, depicts the convolution and deconvolution operations along with their specific inputs.

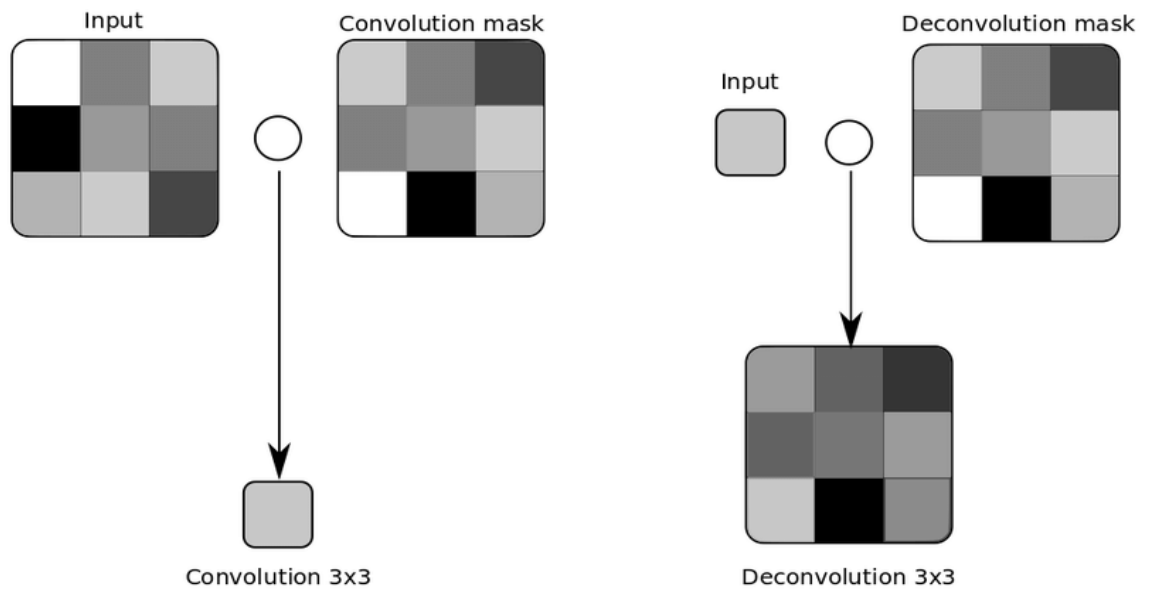


Figure 3.1: Convolution and deconvolution operation using a 3 x 3 kernel.

3.1.2 Long short-term memory cells (LSTM):

A time series is a collection of data gathered across a number of time periods. A model based on LSTM (Long Short-Term Memory), a sort of recurrent neural network architecture, is an effective strategy in such circumstances. The model transmits the prior hidden state to the following step in the sequence in this form of design. As a result, the network stores information about earlier data and uses it to make decisions. In other words, the sequence of the data is crucial.

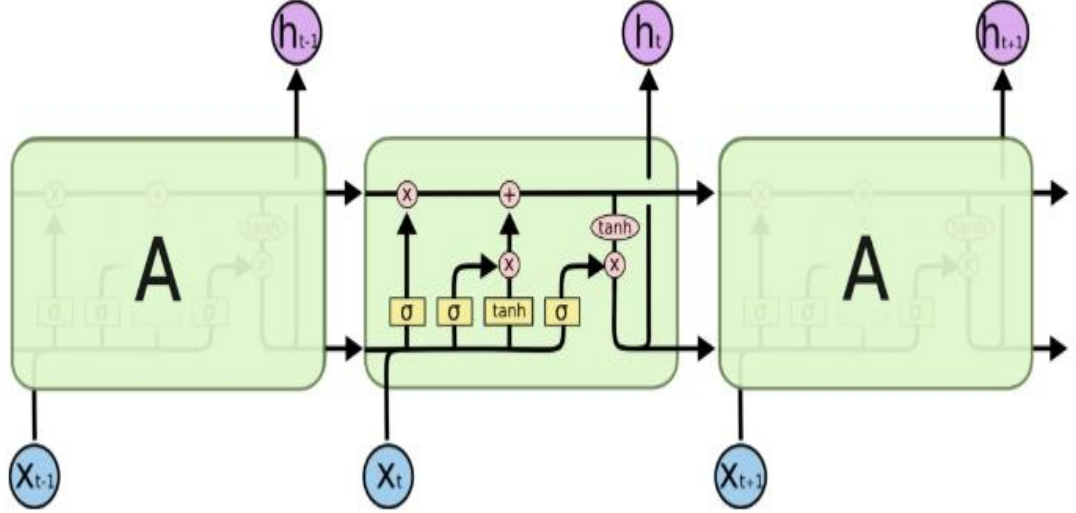


Figure 3.2: A standard LSTM cell

3.1.3 Convolutional LSTM:

For video frame prediction, the Convolutional Long Short-Term Memory (ConvLSTM) model, a version of the LSTM architecture, is utilized. Convolutions are used in place of matrix operations in ConvLSTM as opposed to the conventional fully connected LSTM (FC-LSTM). ConvLSTM uses fewer weights and produces superior spatial feature maps by employing convolution for input-to-hidden and hidden-to-hidden connections. The ConvLSTM unit's formulation can be perfectly described as (1) through (6).

$$f_t = \sigma(W_f * [h_{t-1}, x_t, C_{t-1}] + b_f) \quad (1)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t, C_{t-1}] + b_i) \quad (2)$$

$$\hat{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t \quad (4)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t, C_{t-1}] + b_o) \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

3.1.4 Final model architecture:

Using all the building blocks explained above, we create a spatio-temporal autoencoder which takes a sequence of 10 images as an input and tries to reconstruct these sequences of images as output. It has a spatial encoder which encodes the 2-dimensional information and a temporal decoder to learn patterns across the axis of time. The aim of the bottleneck layer is to force the encoder to extract and encode only meaningful information which can be decoded by the subsequent decoders. Detailed architecture can be seen in Figure 3.3.

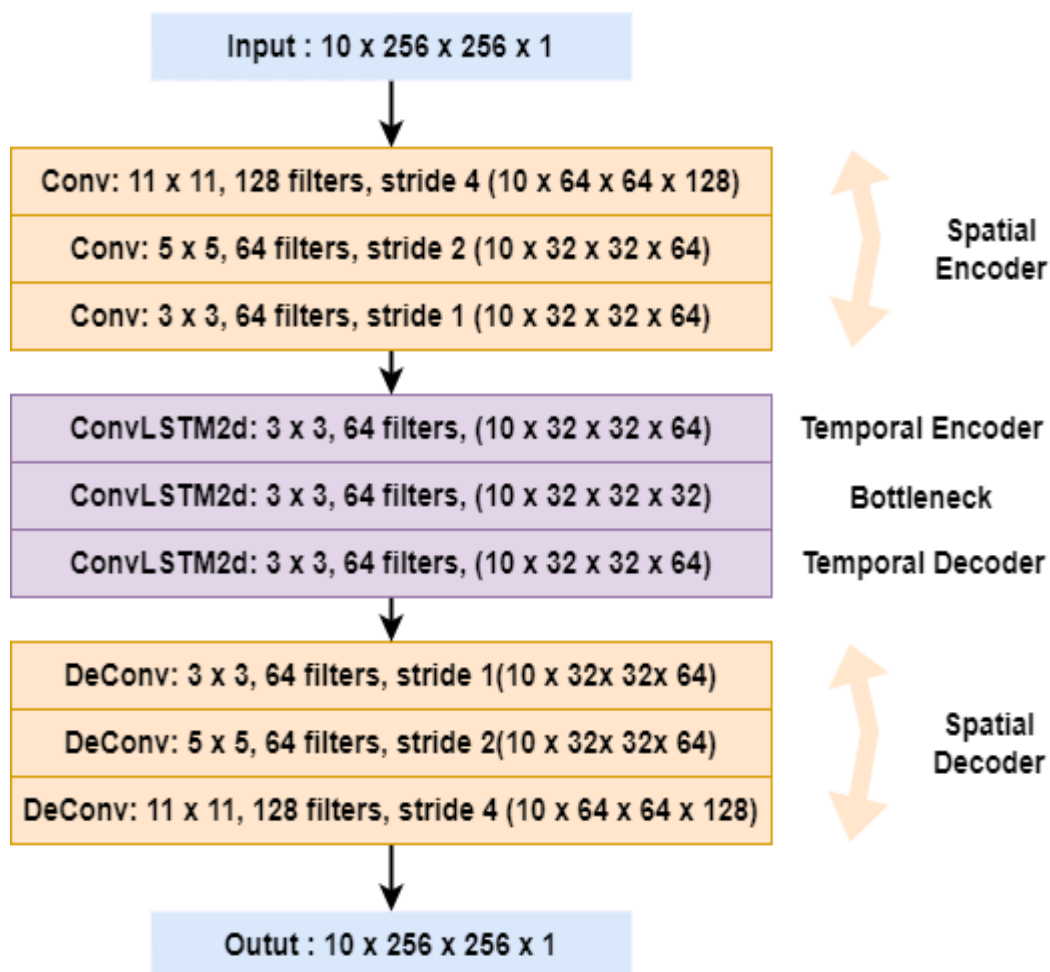


Figure 3.3: Final model architecture

3.2 REGULARITY SCORE:

Using the L2 norm, we calculate the reconstruction error of the intensity value I of a pixel at the location (x,y) in frame t of the clip:

$$e(x, y, t) = \left\| I(x, y, t) - fw(I(x, y, t)) \right\|_2 \quad (7)$$

fw is the model that the LSTM convolutional autoencoder learned, in this case. The reconstruction error of a frame t is then calculated by adding together all pixel-wise errors:

$$e(t) = \sum_{(x,y)} e(x, y, t) \quad (8)$$

The following formula can be used to determine the reconstruction cost of a 10-frame sequence that begins at time t :

$$\text{sequence reconstruction cost}(t) = \sum_{t'=t}^{t+10} e(t') \quad (9)$$

The abnormality score $S_a(t)$ is then calculated by scaling between 0 and 1.

$$S_a(t) = \frac{\text{sequence reconstruction cost}(t) - \text{sequence reconstruction cost}(t)_{min}}{\text{sequence reconstruction cost}_{max}} \quad (10)$$

By deducting abnormality scores from 1, we may obtain the regularity score $S_r(t)$.

$$S_r(t) = 1 - S_a(t) \quad (11)$$

For each t in the range $[0,190]$, we first compute the regularity score $S_r(t)$, and then we draw $S_r(t)$ and based on the thresholding we predict the abnormality of the input sequence.

3.3 PROPOSED METHODOLOGY

Ideally, we would prefer to approach the issue as a binary classification problem, but doing so requires a significant amount of labelled data, and doing so is challenging for the reasons listed below:

1. Occurrence of abnormal events is relatively rare as compared to the normal events.
2. Abnormal events present massive variations, categorizing them manually and manually labeling such events would come with huge manpower costs.

Also, Unusual things happen for one of two reasons:

1. non-pedestrian objects, such as skateboarders, cyclists, and small carts, in the walkway.
2. Unusual pedestrian motion patterns, such as people crossing a path or gazing at the grass in its vicinity.

Now, to solve this problem of lack of labeled data across various categories, as well as define new categories. We decided to tackle the problem using an unsupervised approach. We have the option to use unlabeled data or data with very few labels when employing unsupervised or semi-supervised approaches like dictionary learning, Spatio-temporal features, and autoencoders. In contrast to supervised algorithms, these methods just need unlabeled video footage that is simple to collect in practical applications and contains few to no anomalous events. Unsupervised approaches involving images mostly include autoencoder-based architectures.

3.3.1 The Approach

The reconstruction error is the key. To discover regularity in video sequences, we employ an autoencoder. It makes sense that the trained autoencoder will accurately reconstruct

motions in regular video sequences with little error, but not in irregular video sequences. Below is the algorithmic flow of our crowd analysis framework. We compare the clips generated by our neural network architecture and generate a regularity score or reconstruction cost. After that we compare this cost with a threshold which is selected empirically based on our observation. If the cost is greater than the threshold then we classify the input as abnormal.

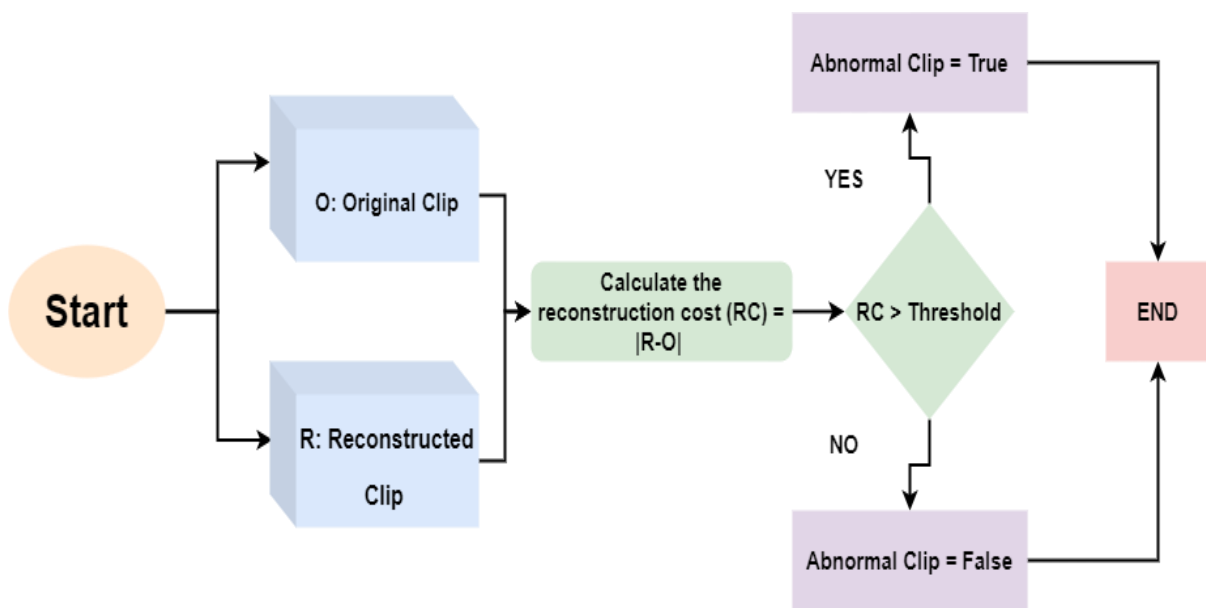


Figure 3.4: Algorithmic flow of our crowd analysis framework. Each input is reconstructed using our neural network and a reconstruction cost is calculated. Furthermore, this cost is used to predict if the input contains normal events or abnormal events.

3.3.2 Preprocessing:

Regular video frame sequences make up the training set; the model will be trained to recreate these sequences. To prepare the dataset that will be ingested by our model, We follow the following steps:

1. Using the sliding window technique, divide the training video frames into temporal

sequences of 10 frames each.

2. To guarantee that the input photos have the same resolution, resize each frame to 256×256 .

3. Divide each pixel by 256 to scale its value from 0 to 1.

One more thing: because there are so many parameters in this model, we need a lot more training data, so we add more data in the temporal dimension. We combine frames with different skipping strides to create more training sequences. As an illustration, the first stride-1 sequence consists of the frames (1, 2, 3, 4, 5, 6, 7, 8, 9, 10), while the first stride-2 sequence has the frames (1, 3, 5, 7, 9, 11, 13, 15, 17, 19).

3.4 TRAINING AND TESTING:

We use an open-source framework Pytorch for training and testing of our model. We train our model for 50 epochs for both UCSD ped1 and ped2 dataset. Using Batch size of 64 and initial learning rate of 0.0001. We use Adam optimizer for optimizing the training and use a cosine annealing learning rate scheduler (**Ilya Loshchilov,2016**) to change our learning rate. Figure 3.5 and 3.6 describes our testing and training workflows. During inference we take input of 10 image sequences and generate a regularity score. Based on this score our system determines if the input contains abnormal events or not.

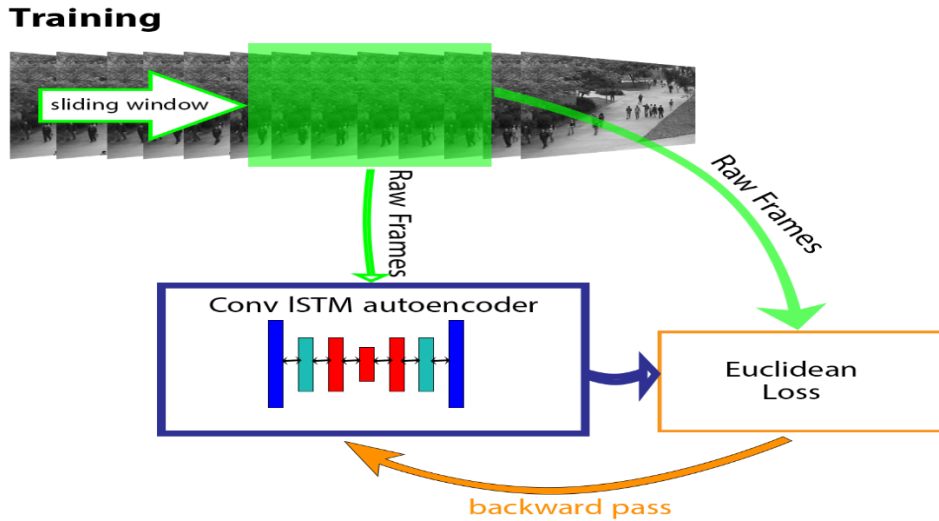


Figure 3.5: Training workflow of our abnormal event detection framework. A sequence of 10 images is fed as an input to our model. The model reconstructs the given input and per pixel Euclidean loss is calculated from reconstructed sequence and raw frames. This loss is back propagated for the network to learn better reconstruction.

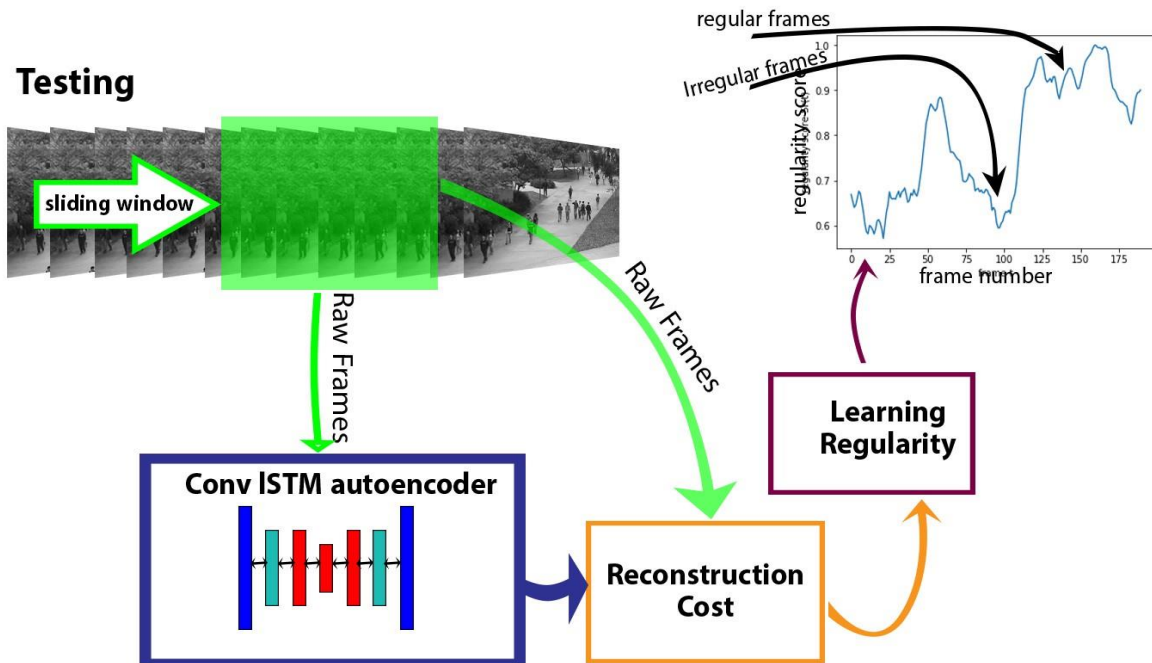


Figure 3.6: Testing and deployment workflow of our abnormal event detection framework

3.5 Model for crowd detection behavior

The proposed system entails developing and putting into action the system that provided the solutions the existing system needed. The task includes developing an Unsupervised Abnormal Crowd Behavior Detection system as well as thoroughly examining the system. By finishing the system's creation, the mission aims to give Urban professional security. We built a method that can discriminate between normal and pathological crowd behavior by developing a continuous picture observation system and using SVM, kNN, Neural Networks, and linear regression method to evaluate monitoring of congested metropolitan regions. (see figure 3.7).

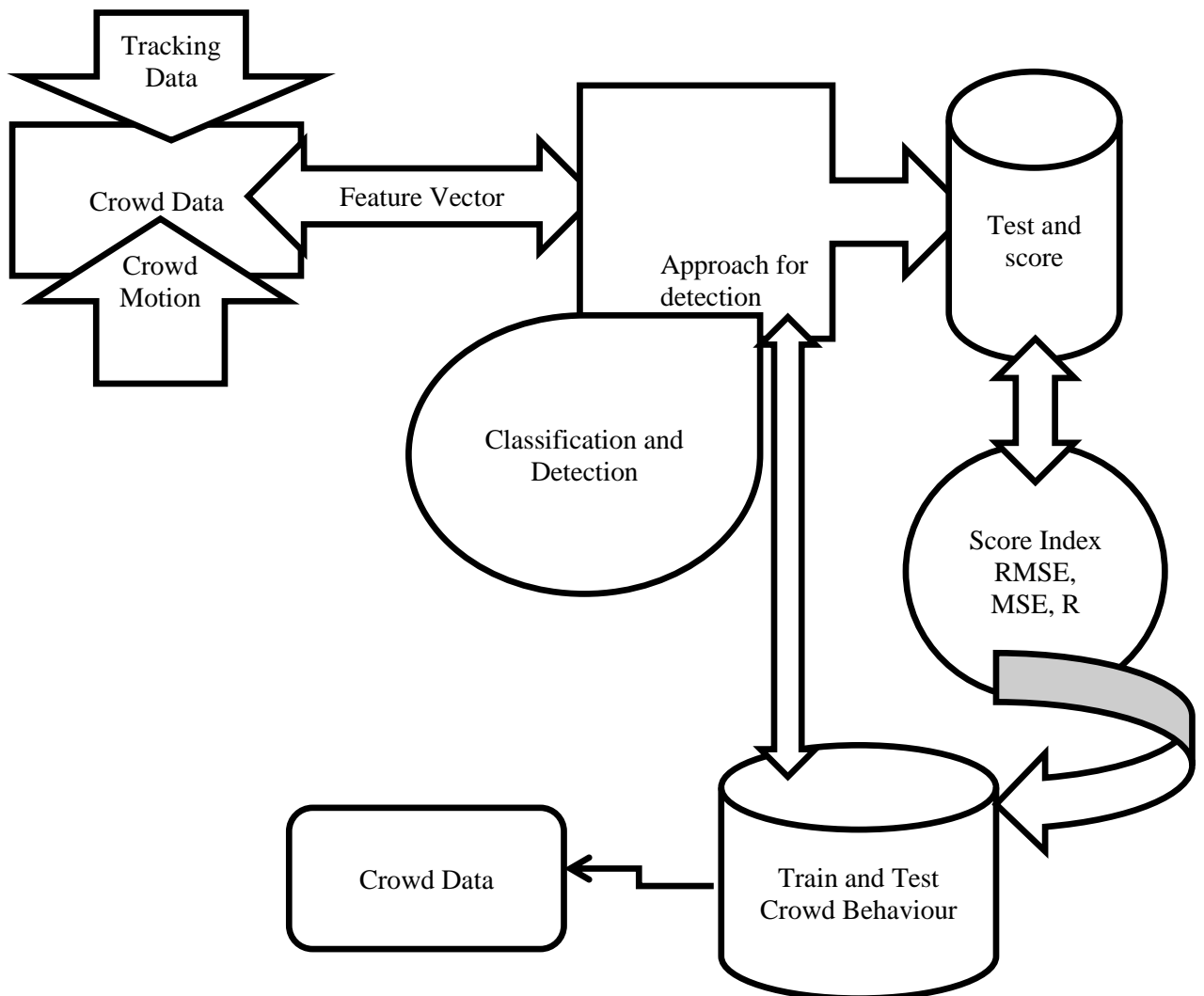


Figure 3.7: Data Flow Diagram

This is the analysis module which validates the system efficiency between the actual class label and the predicted class label. In various modules, we generate the actual count and density map of each image while in testing methods predicts the possible counts of the respective input images.

3.5.1 Crowd data set

Due to variations in perspective and scene, the distribution of crowd density in crowded crowd images is seldom consistent. Figure 3.8 displays a variety of photos for your perusal. Because of this, it is illogical to attempt to count the people in the crowd by taking in the full scene at once. As a direct consequence of this, our system was modified to use the divide-count-sum methodology. A regression model is used to translate each image segment to the relevant local count once the images have been divided into patches. The total number of these patches is then multiplied by a cumulative formula to arrive at the global image count.

Image segmentation offers two distinct advantages to its users: In the smaller image patches, the crowd density is initially dispersed in a way that is largely consistent. Second, by performing picture segmentation, the amount of training data that the regression model may access is boosted. Because of the advantages described above, we can now create a regression model with greater resilience. In spite of the fact that there isn't a uniformity to the crowd density, there is a continuous pattern in the overall crowd density distribution. This indicates that adjacent image patches need to have densities that are equal to one another.

When we want to split the picture, we often employ overlaps, which helps better the link between image patches. The estimated count between overlapping picture patches is smoothed down with the use of a Markov random field so that the overall result is closer to the genuine density distribution. This helps to correct any possible estimation mistakes that may have occurred while counting image patches. In order to learn a map from the aforementioned features to the local count, we make use of a fully connected neural network.

Additionally, we make use of a pre-trained deep residual network in order to extract features from picture patches.

Picture identification, object detection, and image segmentation are just some of the many computer vision applications that have benefited from the use of deep convolutional network features. This would indicate that the learnt characteristics of the deep convolutional network may be used to a broad variety of different computer vision applications. When there are more layers in the network, the learnt features have a better chance of accurately representing the data.

On the other hand, in order to properly prepare for a more in-depth model, you will need more data. For crowd counting, the datasets currently available are insufficient to completely train an extremely deep convolutional neural network.

We make advantage of a deep residual network that has already been pre-trained in order to extract features from picture patches. The team came up with the idea to rephrase the layers as learning residual functions with reference to the layer inputs as a solution to the deterioration issue. Because of this, they were able to avoid studying unreferenced functions, which was the method that had been used before. In order to extract the deep features that characterize the density of the crowd, we make use of the residual network, which was trained on the Picture Net dataset for the purpose of image classification. The network was trained using the data from the Image Net dataset.

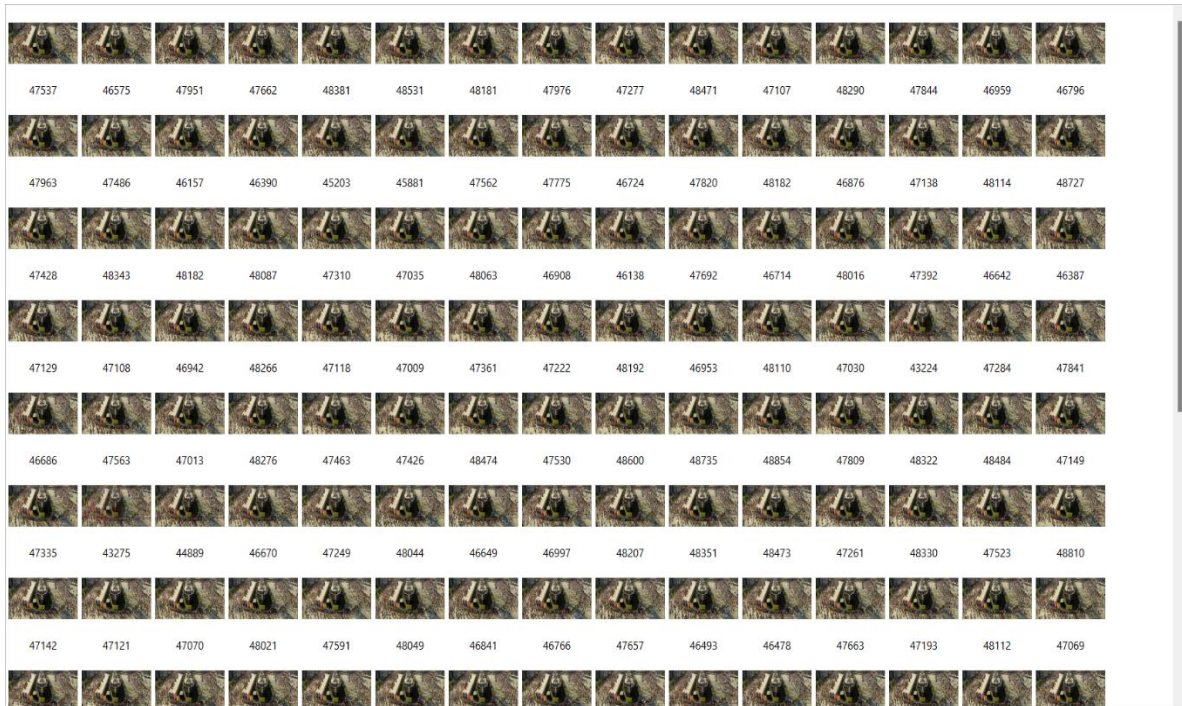


Figure 3.8: Data set for detection

3.5.2 Feature Selection

In this section, using feature selection the embedded methods category includes Random Forest. Filter and wrapper techniques' advantages are combined in embedded methods. They are put into practice by algorithms with built-in feature selection techniques. The following are some advantages of embedded methods:

- They are extremely accurate;
- They generalize better.
- They are comprehensible

3.5.3 Data Flow

The fundamental idea of Orange, as seen in figure 3.9, is known as visual programming. This implies that each analytical step is encapsulated inside a widget. The canvas is used to insert widgets, which are then coupled into an analytical workflow that runs in a clockwise direction from left to right. Orange never passes data backwards. Let us start with a simple workflow.

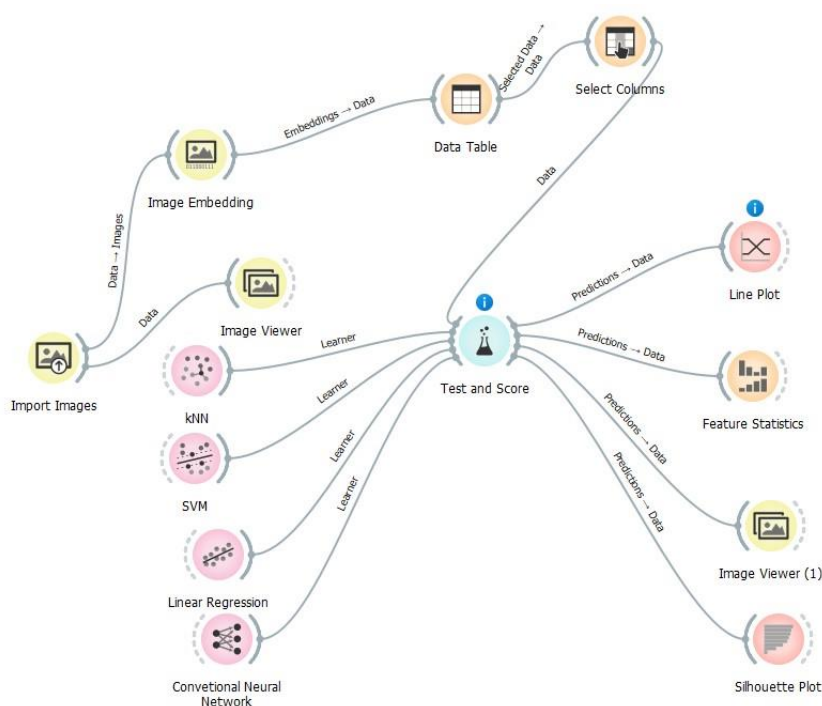


Figure 3.9: Data Flow

3.5.4 Classification

Despite the many improvements that have been made in the study of human behavior, it is still difficult to comprehend and control the behavior of large groups of people. Because of the individualized nature of abnormal conduct, It can be explained in a variety of ways. It has been the source of a great deal of uncertainty in the published work. Some of the

researchers talk about the anomaly in terms of how often it occurs. When something happens very seldom, it is referred to as an anomalous occurrence or something that takes place very infrequently. A crowd may be either structured or unstructured, depending on how it is organized. A well-organized crowd is simple to study, but an unorganized crowd is fraught with danger because of its tendency toward unpredictable movements. Understanding behavior requires a primary focus on velocities, the direction of flow, and anomalous occurrences like fighting, running, and other such activities.

3.5.5 Data Structure

The screenshot shows the Orange Data Mining interface. On the left, the 'Data Table - Orange' widget is active, displaying a table with 78 instances. The table columns are: hidden origin, image name, image, size, width, height, n0 True, n1 True, n2 True, n3 True, n4 True, and n5 True. The data rows show various image files with their respective sizes and dimensions, and binary values for the features n0 through n5. The interface also includes a sidebar with options like 'Show variable labels', 'Visualize numeric values', and 'Color by instance classes'.

hidden origin	image name	image	size	width	height	n0 True	n1 True	n2 True	n3 True	n4 True	n5 True
75	00075	00075.jpg	27702	640	480	0.11677	0.0633531	0.0824171	0.0442776	0.505155	0.8487
66	00066	00066.jpg	27827	640	480	0.11912	0.0663145	0.0940408	0.0540403	0.554097	0.9560
74	00074	00074.jpg	27579	640	480	0.121281	0.0619007	0.106465	0.0530738	0.53454	0.8370
73	00073	00073.jpg	27685	640	480	0.122367	0.0534121	0.0975186	0.0483799	0.555306	0.9125
76	00076	00076.jpg	27714	640	480	0.123849	0.0761848	0.076012	0.0419973	0.547559	0.8232
71	00071	00071.jpg	28040	640	480	0.12468	0.110817	0.073427	0.102164	0.622834	0.8927
77	00077	00077.jpg	28219	640	480	0.135614	0.087125	0.100264	0.0298855	0.627265	0.8044
69	00069	00069.jpg	27929	640	480	0.137623	0.0406709	0.0698968	0.0515337	0.684678	0.817
55	00055	00055.jpg	27493	640	480	0.142516	0.0914542	0.0784721	0.0641601	0.495647	1.016
78	00078	00078.jpg	27844	640	480	0.143221	0.0487325	0.111665	0.0520588	0.561391	0.8618
67	00067	00067.jpg	27709	640	480	0.143487	0.0785118	0.0860405	0.0452026	0.581561	0.9471
49	00049	00049.jpg	27448	640	480	0.153167	0.0663215	0.0790054	0.0906816	0.589906	0.860
63	00063	00063.jpg	27580	640	480	0.155021	0.0753621	0.0894536	0.0461607	0.529387	0.8792
72	00072	00072.jpg	27883	640	480	0.157732	0.0939744	0.114159	0.0613818	0.502192	0.9252
61	00061	00061.jpg	27540	640	480	0.160485	0.159711	0.0723886	0.0489143	0.589957	0.85
48	00048	00048.jpg	27468	640	480	0.16615	0.0643957	0.0704378	0.0569278	0.679889	0.8469
70	00070	00070.jpg	27991	640	480	0.168476	0.0608446	0.0730728	0.0626604	0.660253	0.7877
68	00068	00068.jpg	27924	640	480	0.170862	0.0623655	0.0613113	0.0535254	0.615068	0.8227
64	00064	00064.jpg	27611	640	480	0.172079	0.0730734	0.0826114	0.0400889	0.535329	0.8406
40	00040	00040.jpg	30163	640	480	0.174973	0.0962677	0.0790753	0.0931672	0.564783	0.7928
54	00054	00054.jpg	27516	640	480	0.175322	0.0681841	0.0985872	0.0620054	0.520449	0.9569
65	00065	00065.jpg	27623	640	480	0.175383	0.109305	0.0723302	0.0504172	0.589597	0.8047
59	00059	00059.jpg	27765	640	480	0.177975	0.120194	0.113573	0.0428624	0.5214	0.9554
41	00041	00041.jpg	30158	640	480	0.179608	0.0932907	0.0850379	0.0936521	0.53243	0.8060
56	00056	00056.jpg	27366	640	480	0.179666	0.100015	0.103131	0.0627708	0.531058	0.8831
47	00047	00047.jpg	27228	640	480	0.184645	0.0803136	0.0730087	0.0690057	0.54662	0.9620
60	00060	00060.jpg	27777	640	480	0.190118	0.150812	0.0870472	0.0621384	0.57217	0.9091
62	00062	00062.jpg	27630	640	480	0.192883	0.117454	0.0701055	0.051937	0.513479	1.008
51	00051	00051.jpg	27355	640	480	0.198926	0.117007	0.102982	0.0876909	0.596578	0.9854
53	00053	00053.jpg	27489	640	480	0.202017	0.126223	0.104548	0.0544163	0.526112	0.9542
58	00058	00058.jpg	27292	640	480	0.207123	0.109914	0.114647	0.0654637	0.4518	1.044

Figure 3.10: Data View

Using the File widget, we will load the data; let's use the well-known Iris data set as an example. To right-click on the canvas, go to the menu. There will be a menu presented. You should begin typing "File" before pressing the Enter key to confirm your selection. On the canvas, there will be a file widget installed. The output of the File widget may be seen in the "ear" that is located on the right side of the widget. Simply clicking on the "ear" and dragging a connection away from it will do this. After disengaging from the

connection, a menu will become visible. You should begin with the name of the widget you want to link with the File widget, for example, "Data Table." Choose the widget, then hit the enter button.

The widget is included in the canvas's composition. This is a straightforward process flow. The data is loaded via the File widget, and it is then sent to the output. The data is received by Data Tabular, which then presents it in a table format. Please be aware that Data Table is merely a viewer and only passes on the selection to further processes. The information may always be accessed directly from the source using the File widget.

The Test and Score widget is used to evaluate predictive models, while the Predictions widget is used to make predictions based on recently gathered data. Learners (algorithms to use for training the model), data (a data set for evaluating models), and a preprocessor (which is optional) are some of the different inputs that Test and Score accepts (for normalization or feature selection).

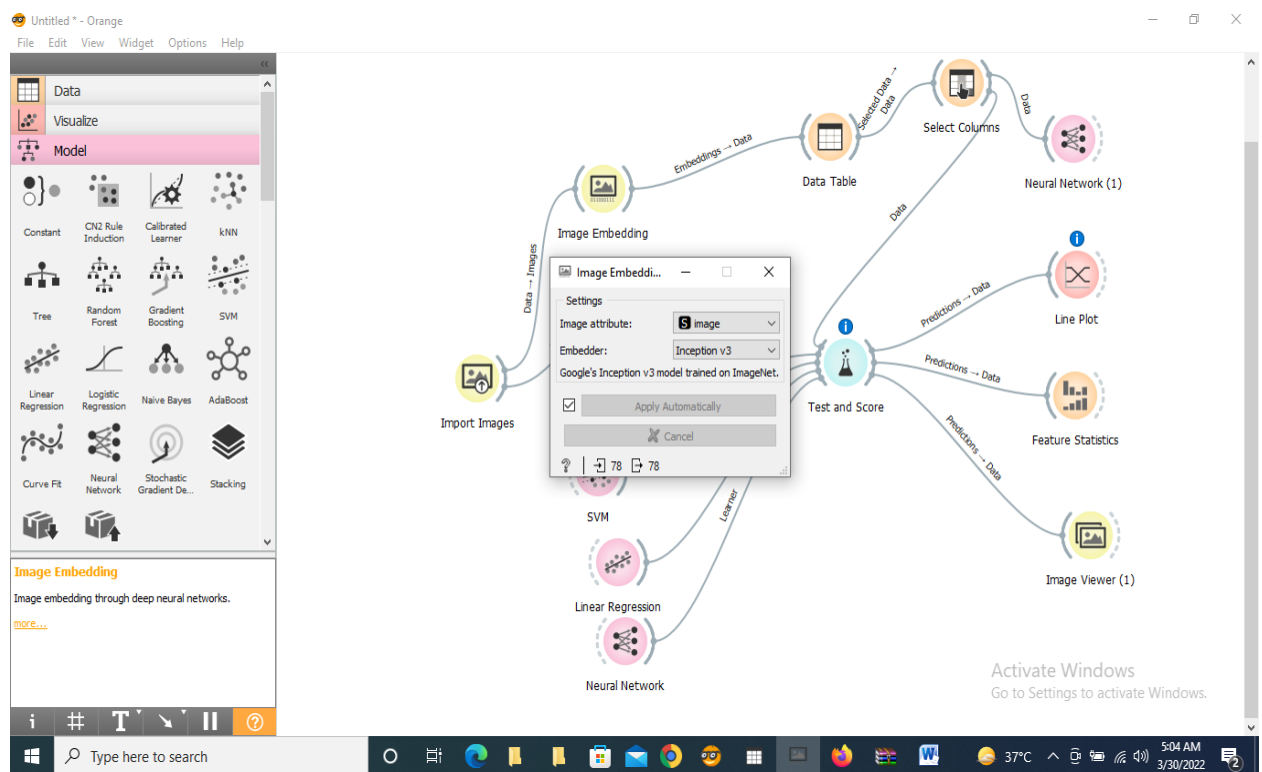


Figure 3.11: Image Detection for model development

3.5.6 Feature Statistics

The Feature Statistics (see figure 3.12) widget on the *crowd* data set. The feature was manually changed to a metavariable for illustration purposes.



Figure 3.12: Feature Statistics

- Information on the amount, quantity, and categories of characteristics of the currently available data set
- Any characteristic may be used to provide a color to the histograms on the right. If the characteristic that was chosen is categorical, then a distinct color palette will be used (as shown in the example). When a numerical characteristic is chosen to be analyzed, a color palette with no breaks is used. The statistics for each aspect of the data set are included in the table that can be seen on the right. The characteristics may be arranged in a hierarchy according to each statistic, which will now be discussed.
- There are many possible types of features, including category, numeric, temporal, and string.

- The name of the component being discussed.
- A histogram displaying the values of the features. If the feature is a number, we discretize the values into suitable bins according to the proper method. If the feature is categorical, then the histogram will have separate bars for each of the possible values.
- The predominant pattern shown by the feature values. This is the most common pattern for categorical characteristics. This is the mean value for the numerical characteristics.
- The degree to which the feature values are spread out. The entropy of the value distribution is what we mean here when talking about categorical characteristics. This is the coefficient of variation for a characteristic that has a numeric value.
- The minimal value. The computation for this takes into account both numerical and ordinal categorical characteristics.
- The highest possible value. The computation for this takes into account both numerical and ordinal categorical characteristics.
- The number of values that are absent from the dataset.

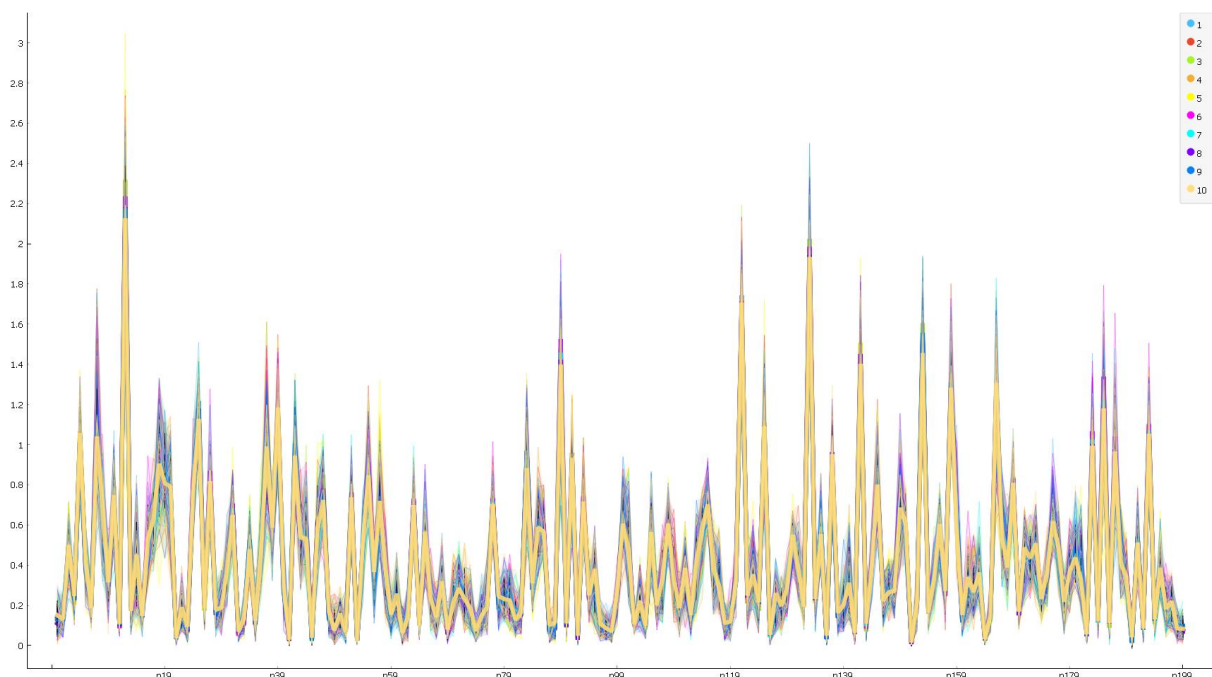


Figure 3.13: Line Plot

A line plot is a common kind of visualization widget that presents data profiles, often in the form of sorted numerical data. In this straightforward example, the crowd data will be shown as a line plot, with individual points being sorted according to the iris characteristic.

CHAPTER – 4
RESULT ANALYSIS
AND
COMPARATIVE STUDY

4.1 RESULT ANALYSIS:

This section is divided into three sections. In the first section we will discuss the various metrics that were used for the evaluation purpose and Area under ROC curve (AUC) and EER of our proposed methodology. In the second section we have described the dataset that has been used for the purpose of training, testing and analyzing our proposed ConvLSTM autoencoder model. And finally in the last section we have displayed the results of our experiments and presented a comparative study of the accuracy of our proposed methodology with respect to other models such as SF, ConvAE and Spatio-temporal AE.

4.1.1 EVALUATION METRICS

For the purpose of evaluating the performance of our proposed methodology we have employed two important evaluation metrics in our experiment. These metrics are Mean Absolute Error (MAE) and Mean Squared Error (MSE).

4.1.1.1 Mean Absolute Error (MAE)

The Mean Absolute Error also called MAE is the average of the total errors that we get in any set of predictions. The errors here refer to the difference between the actual observations and the predicted values. The average of this difference is called the Mean Absolute Error. In our experiment we have calculated the Ground Truth Density (GTD) which is the actual observation in our experiment. The Estimated Density (ED) of the crowd is the predicted value. Hence, the difference between the ED and the GTD gives the absolute error. Finding its average gives us the MAE of the prediction.

$$MAE = \frac{1}{N} \sum_{i=1}^N |z_i - \hat{z}_i|$$

Figure 4.1: Formula of Mean Absolute Error

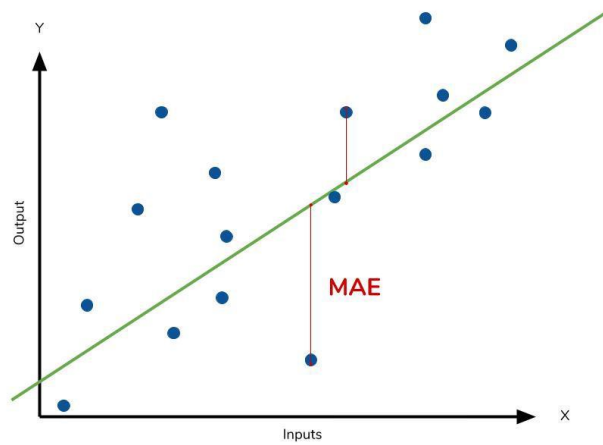


Figure 4.2: Diagram representing Mean Absolute Error

4.1.1.2 Mean Squared Error (MSE)

The Mean Squared Error also called MSE is the square of the mean of the square of the errors. This metric keeps in mind the direction of the difference. Hence, the result of MSE is always positive. We calculate the difference between the actual observation and the predicted values and find the square of this difference. The average of this difference gives us the Mean Squared Error.

$$MSE = \frac{1}{N} \sum_{1}^N (z_i - \hat{z}_i)^2$$

Figure 4.3: Formula of Mean Squared Error

N = Number of images

Z_i = Number of people in ith image

\hat{z}_i = Estimated number of people in the ith image

4.1.1.3 Area under ROC curve (AUC) and Equal Error Rate

A measurement tool for binary classification issues is the Receiver Operator Characteristic (ROC) curve. Essentially, it does this by graphing the TPR in comparison to the FPR at different threshold levels in order to distinguish the "signal" from the "noise." The capacity of a classifier to discriminate between classes is measured using the Area Under the Curve (AUC), which is used as a summary of the ROC curve.

The model performs better at differentiating between positive and negative classifications the higher the AUC. The identical error rate loses its mystique if ROC curves are understood. It's simply one method of attempting to balance the risk of false positives and false negatives to the best possible degree. A ROC or DET curve's EER is the point when the rates of erroneous acceptance and false rejection are equal.

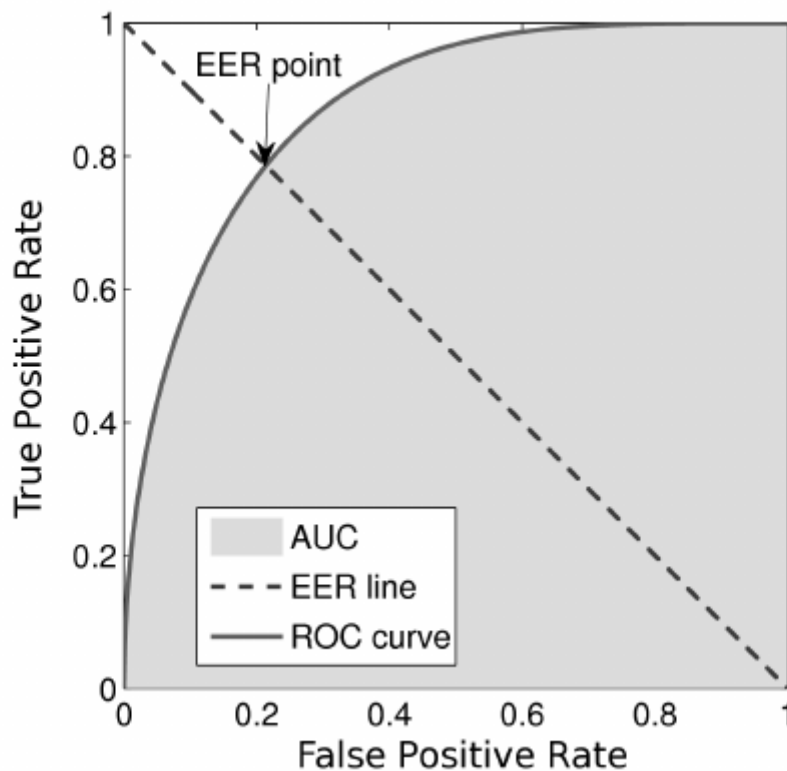


Figure 4.4: Diagram representing ROC(AUC) and EER

4.1.2 DATASET

We use the UCSD dataset for congested settings in our research. [22] A stationary camera installed at a height and looking down on pedestrian pathways was used to collect the UCSD Anomaly Detection Dataset. The walkways had varying densities of people, from very few to many. Only pedestrians are shown in the video at its regular setting. Odd things happen for one of two reasons: abnormal pedestrian motion patterns the movement of non-pedestrians in the walkways Bikers, skateboarders, tiny carts, and pedestrians crossing a walkway or in its surrounding grass are examples of often occurring anomalies. There were a few cases of persons using wheelchairs as well. All abnormalities are real; they weren't produced to create the dataset. They all occur spontaneously. Two separate subsets of the data were created, one for each scene. Each scene's video recording was divided into a number of clips, each with about 200 frames.



Figure 4.5: Normal (top) and abnormal(bottom) samples of UCSD ped1 dataset

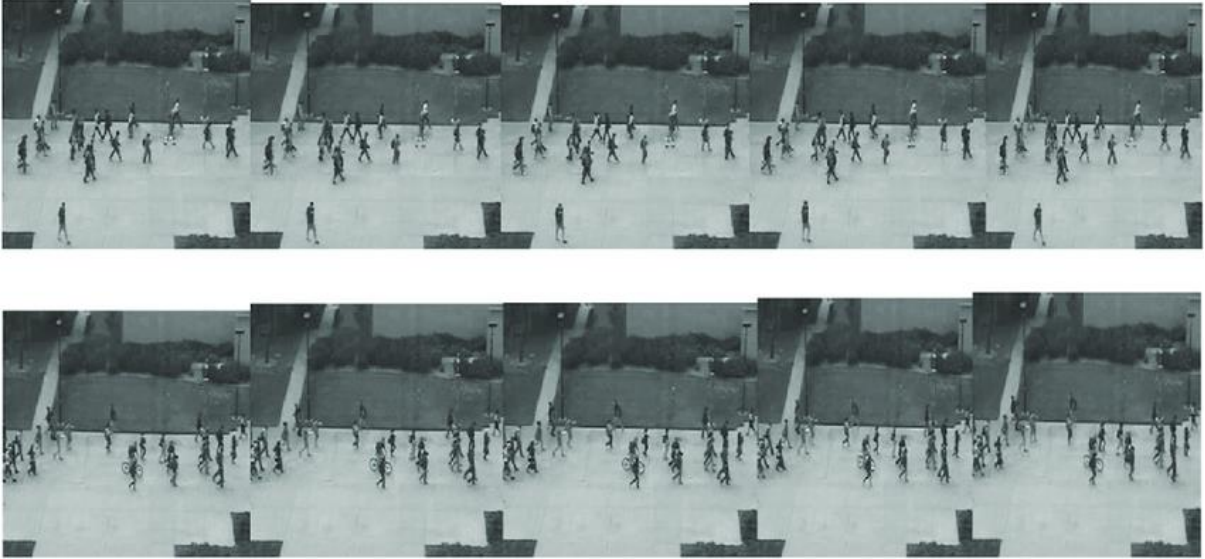
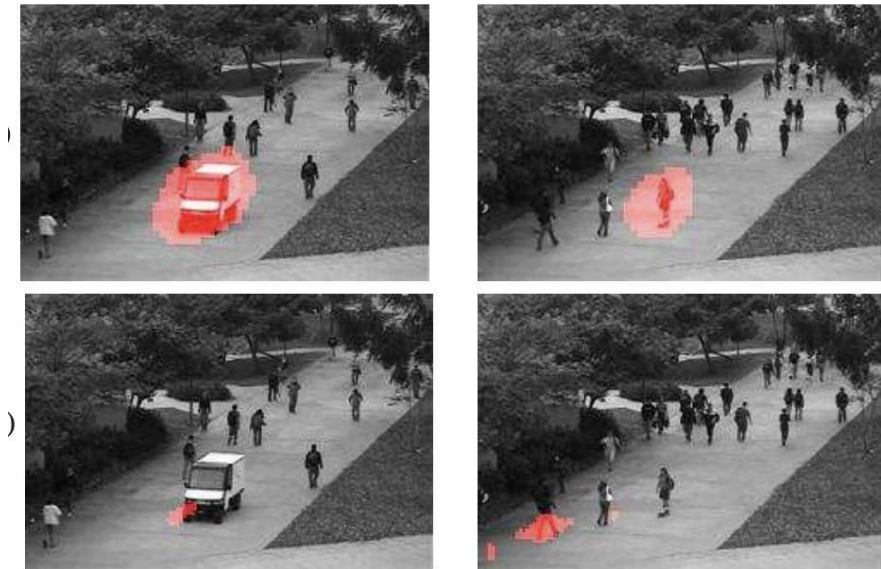


Figure 4.6: Normal (top) and abnormal(bottom) samples of UCSD ped2 dataset



a. Abnormal elements (highlighted red) in the dataset

Figure 4.7: Visualization of various frames from UCSD dataset.

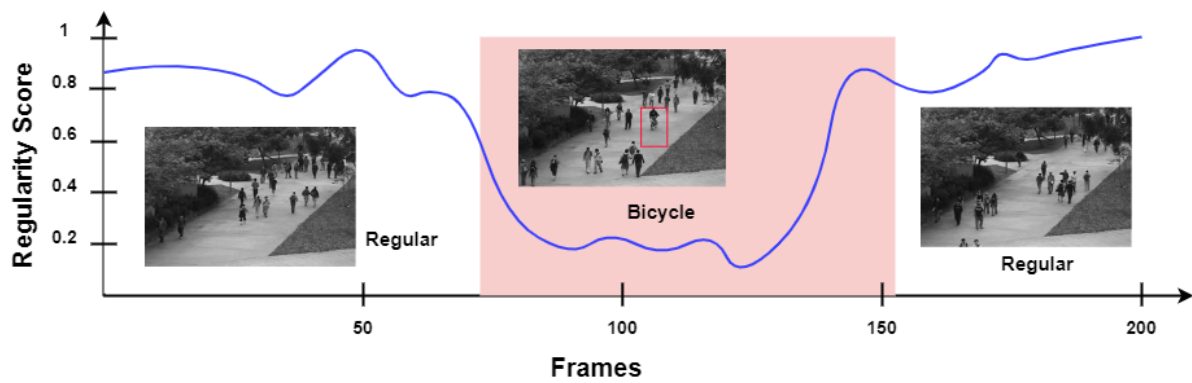
4.1.3 Quantitative results (Equal Error Rate and Area Under the ROC Curve (AUC))

We compare our model with other methods and observe that our model performs better in Ped1 dataset as compared to Ped2 dataset. We also observe that our results are competitive with respect to other methods. Results can be viewed in table below:

TABLE 4.1**COMPARATIVE STUDY OF DIFFERENT METHODS AND OUR APPROACH FOR CROWD ABNORMAL BEHAVIOUR**

Methods	Ped1 (AUC/EER)	Ped2 (AUC/EER)
SF [22]	67.5/31.0	55.6/42.0
MPPCA [23]	66.8/40.0	69.3/30.0
MPPCA+SF [22]	74.2/32.0	61.3/36.0
HOFME [24]	72.7/33.1	87.5/20.0
ConvAE [25]	81.0/27.9	90.0/21.7
Spatio-temporal AE[26]	89.9/12.5	87.4/12.0
Ours	90.1/11.5	86.2/14.8

4.1.3.1 Qualitative Analysis: Visualizing regularity scores with respect to frames



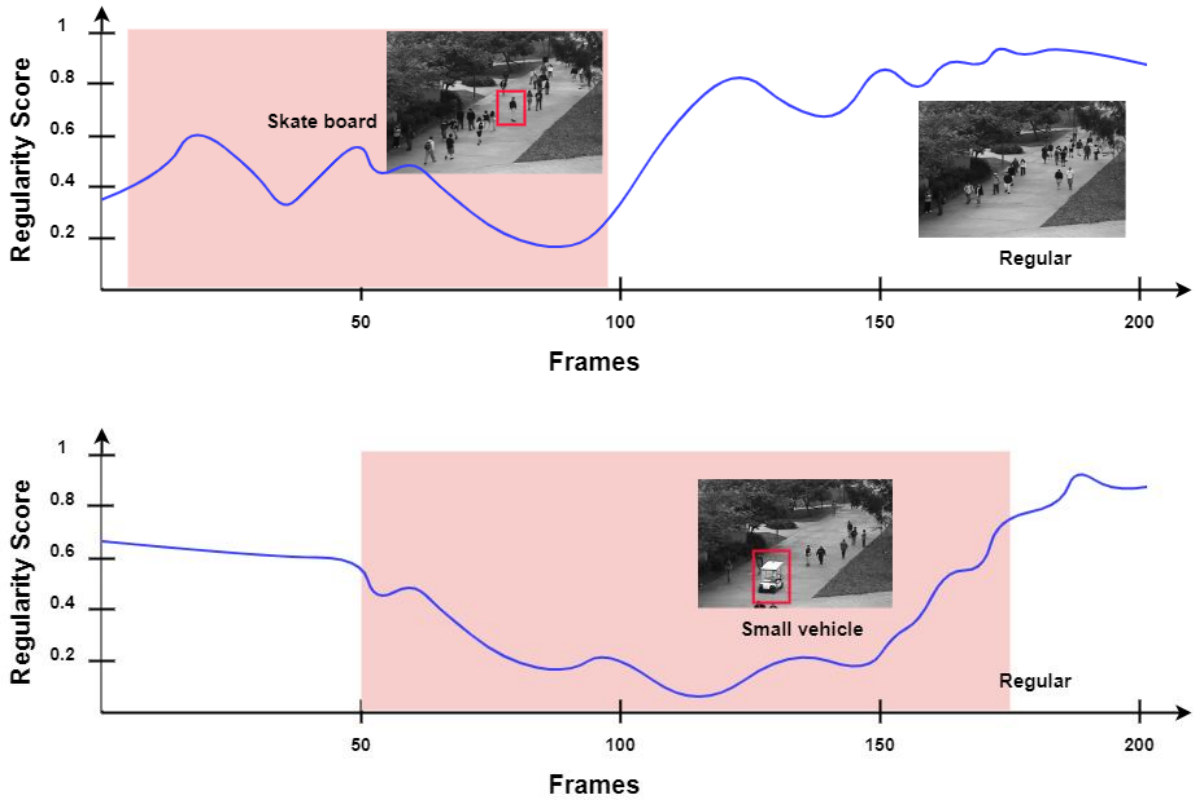


Figure 4.8: Visualizing the regularity score on Ped1 Test video 1, video 8, video 24 respectively. Abnormal events are marked within red bounding boxes. And abnormal events have a red background in the graph.

We have also evaluated numerous algorithms utilizing various datasets and have used pictures to fully capture changes in the crowd.

TABLE 4.2

COMPARATIVE STUDY OF DIFFERENT METHODS FOR CROWD DETECTION

Model	MSE	RMSE	MAE	R2
kNN	495658.962	704.031	655.890	0.419
SVM	863320.526	929.150	690.460	-0.013
Neural Network	2119205776.612	46034.832	46022.115	-2484.642
Linear Regression	523921.069	723.824	584.909	0.385

TABLE 4.3**MODEL COMPARISON BY MSE**

	kNN	SVM	Neural Network	Linear Regression
kNN		0.057	0.000	0.348
SVM	0.943		0.000	0.916
Neural Network	1.000	1.000		1.000
Linear Regression	0.652	0.084	0.000	

TABLE 4.4**MODEL COMPARISON BY RMSE**

	kNN	SVM	Neural Network	Linear Regression
kNN		0.051	0.000	0.388
SVM	0.949		0.000	0.916
Neural Network	1.000	1.000		1.000
Linear Regression	0.612	0.084	0.000	

TABLE 4.5**MODEL COMPARISON BY MAE**

	kNN	SVM	Neural Network	Linear Regression
kNN		0.255	0.000	0.890
SVM	0.745		0.000	0.910
Neural Network	1.000	1.000		1.000
Linear Regression	0.110	0.090	0.000	

In Table 4.2 and 4.3, We are characterized Figure shows that, in comparison to kNN and linear regression, the MSE was significantly high. This is as a result of taking a density-level repressor approach into account while classifying picture patches. Additionally, when

compared to the total dataset, the MSE was relatively low. This is due to the skip link being considered when using scale-oriented training to address issues of various magnitudes. The MSE showed a very low standard error value compared to SVM and kNN. This happened as a result of the network's convolution and deconvolution layers taking into account a controlled information flow. We so conclude that for datasets with a dense and heterogeneous range of densities, the employment of a task-oriented repressor and convolution improves the accuracy when estimating the level of a better density map.

In order to improve information flow, it is possible to deal with the low density of datasets by using patch-based augmentation procedures (variable-scale) and an emphasis on solving the scale-varying problem in the convolution and deconvolution layers. It is possible to address scale-varying problems caused by the perspective view. The sentences before this one may have information on each of these strategies.

Critical Observations

Validation of crowd behavior is another problematic difficulty since ground truth video material revealing particular anomalous actions in the average crowd is not widely accessible and available. This makes it difficult to validate crowd behavior.

They distinguished between object-based and holistic approaches to crowd behaviour analysis. The holistic approach places more focus on the group as a whole than on individual differences, whereas the object-based approach defines a crowd as a collection of distinct individuals. With this approach, it is assumed that every individual in a crowd moves uniformly throughout the analysis.

This research examines the numerous dimensions of crowd modelling and crowd analysis, two disciplines that employ a wide range of techniques for a wide range of real-world applications. The most recent advances in methods for counting the population have been thoroughly analysed, along with discussions of the systems' advantages and disadvantages.

The results of this poll provide a glimpse into the foreseeable future of crowd monitoring and categorization. For effective control of crowds, an integrated system that can handle any kind of crowd analysis is necessary. This system must be able to handle everything from small-scale disorderly crowds to large-scale panic situations.

This is done to set the framework for upcoming research in this particular topic and to identify the common gaps that are present in the approaches already in use. Even if experts have come to a conclusion, there are still some unanswered questions that necessitate additional research.

CHAPTER- 5

CONCLUSION

AND

FUTURE SCOPE

5.1 Conclusion

In this study we presented an efficient and scalable method to determine abnormal events in videos involving crowded scenes. We conducted various experiments and proposed an architecture with competitive results with respect to other methods. Furthermore, our unsupervised approach makes our method more application intensive as labelling of high amounts of data is both time and resource consuming task.

Crowd analysis encompasses a wide range of research fields, including visual surveillance, machine learning, computer vision, and pattern recognition. The identification, categorization, and recognition of individuals and groups of individuals have all been made possible via the application of crowd analysis. We believe our work will further enable many downstream applications mentioned above. This paper presents a novel method to identify anomalous behavior in crowd scenarios using a competitive machine learning model. A method of There have been suggestions for extracting features and then characterizing them in order to learn more about particle movement and crowd motion. The space-time feature cube method is used. Then, a detection method for identifying anomalous events in large crowds is provided that integrates a competitive neural network model with space-time feature cubes. The goal of this algorithm is to locate crowds across the globe. The experimental results obtained by using our test video sequences have demonstrated that our system is capable of identifying and localizing abnormal crowd behavior. They divided the approaches to studying crowd movement into two groups: object-based and holistic-based approaches. In contrast to the individual variances that exist within an organization, the holistic technique lays more attention on the organization as a whole, whereas According to the object-based paradigm, a crowd is made up of numerous distinct people. With this process, it is predicted that every single person in a crowd would move consistently throughout the duration of the study. This study explores the wide range of elements

connected to crowd modelling and crowd analysis. These two domains employ a wide range of techniques for a wide range of real-world applications. This study explores the wide range of elements connected to crowd analysis and crowd modelling. It has been described how a complete examination of the most recent advancements in population counting techniques has been provided, in addition to a discussion of both advantages and disadvantages of the different strategies. The survey's findings shed some light on the most likely path for crowd surveillance and classification in the near future. To manage crowds effectively, it is essential to have a unified system that can do any type of crowd analysis. This system must be capable of handling various chaotic circumstances, including those in which both people in small groupings and extremely large numbers of people. In order to build the groundwork for future study on this particular topic, in order to find the common problems in the practices now in use, this is done. This study aims to increase knowledge of this particular topic's research methodology. Despite the fact that researchers have come to a conclusion, a number of issues have not yet been completely resolved, which suggests that more research is necessary.

5.2 FUTURE WORK

The potential for improvement is still quite great in the strategies and algorithms implemented for the purpose of crowd behavior analysis. A good amount of optimization is still required, pressing more explicitly on finding a more practical and feasible approach. Given that there are numerous and wide applications of this research, methodologies with higher accuracy and efficiency are required. It is indeed difficult to model crowd behavior and the problem is only made worse by the absence of data regarding normal and abnormal crowd behavior. This leads the path to continued research and strive to find the most optimal algorithms possible to detect and predict crowd behavior.

REFERENCES

1. Akbar Khan, Jawad Ali Shah, Kushsairy Kadir, Waleed Albattah and Faizullah Khan. “Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review”, in Appl. Sci. 2020, 10, 4781; doi:10.3390/app10144781. Online - www.mdpi.com/journal/applsci.
2. Atika Burney, Tahir Q. Syed. “Crowd Video Classification using Convolutional Neural Networks”: 2016 International Conference on Frontiers of Information Technology.
3. Beibei Song and Rui Sheng. “Crowd Counting and Abnormal Behavior Detection via Multiscale GAN Network Combined with Deep Optical Flow”, in Hindawi Mathematical Problems in Engineering Volume 2020, Article ID 6692257, 11 pages <https://doi.org/10.1155/2020/6692257>.
4. BILAL TAHA AND ABDULHADI SHOUFAN. “Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research”, in IEEE Digital Object Identifier 10.1109/ACCESS.2019.2942944.
5. Camille Dupont, Luis Tobias, Bertrand Luvison. “Crowd-11: A Dataset for Fine-Grained Crowd Behaviour Analysis”, in IEEE Xplore provided by the Computer Vision Foundation.
6. Cong, Yang, Junsong Yuan, and Ji Liu. “Abnormal event detection in crowded scenes using sparse representation.” Pattern Recognition 46.7 (2013): 1851–1864.
7. Dongyao Jia, Chuanwang Zhang, Bing Zhang. “Crowd density classification method based on pixels and texture features”, Machine Vision and Applications (2021) 32:43. Springer.
8. “Floor Fields for Tracking in High-Density Crowd Scenes” by Saad Ali and Mubarak Shah.

9. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 733–742 (June 2016).
10. Ilya Loshchilov, & Frank Hutter (2016). SGDR: Stochastic Gradient Descent with Restarts. *CoRR*, *abs/1608.03983*.
11. Jiangtao Wang, Yasha Wang, Qin Lv. “Crowd-Assisted Machine Learning: Current Issues and Future Directions”, in PUBLISHED BY THE IEEE COMPUTER SOCIETY 0018-9162/19©2019IEEE.
12. Lempitsky V., Zisserman A. Learning to count objects in images; Proceedings of the Advances in Neural Information Processing Systems; Vancouver, BC, Canada. 6–11 December 2010; pp. 1324–1332.
13. Li, Teng, et al. “Crowded scene analysis: A survey.” *Circuits and Systems for Video Technology*, *IEEE Transactions on* 25.3 (2015): 367–386.
14. L. Wang, F. Zhou, Z. Li, W. Zuo and H. Tan, "Abnormal Event Detection in Videos Using Hybrid Spatio-Temporal Autoencoder," *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2276-2280, doi: 10.1109/ICIP.2018.8451070.
15. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1975–1981 (2010).
16. Mayur D. Chaudhari, Archana S. Ghotkar. “A Study on Crowd Detection and Density Analysis for Safety Control”, in *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING* · April 2018, E-ISSN: 2347-2693.
17. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: 2009 IEEE Computer Society Conference on Computer Vision and

- Pattern Recognition Workshops, CVPR Workshops 2009. pp. 935–942 (2009).
18. Ming Li, Jian Weng, Anjia Yang, Wei Lu , Yue Zhang. “CrowdBC: A Blockchain-Based Decentralized Framework for Crowdsourcing”, in IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 30, NO. 6, JUNE 2019
 19. Mounir Bendali-Braham, Jonathan Weber, Germain Forestier, LhassaneIdoumghar, Pierre-Alain Muller. “Recent trends in crowd analysis: A review”, in Machine Learning with Applications 4 (2021) 100023.
 20. Pei Voon Wong, Norwati Mustapha, Lilly Suriani Affendey, Fatimah Khalid, Yen-Lin Chen. “Crowd Behavior Classification based on Generic Descriptors”, in 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 978-1-7281-3038-5/19/\$31.00 ©2019 IEEE.
 21. SherifElbishlawi, Mohamed H. Abdelpakey, AgwadEltantawy, Mohamed S. Shehata, and Mostafa M. Mohamed. “Deep Learning-Based Crowd Scene Analysis Survey”, in: J. Imaging 2020, 6, 95; doi:10.3390/jimaging6090095. Online – www.mdpi.com/journal/jimaging
 22. Sohail Salim, Othman O Khalifa, Farah Abdul Rahman, AdidahLajis. “Crowd Detection and Tracking in Surveillance Video Sequences”, in Proc. of the 2019 IEEE 6th International Conference on Smart Instrumentation, Measurement, and Applications (ICSIMA 2019) 27-29 August 2019, Kuala Lumpur, Malaysia.
 23. Sonu Lamba and Neeta Nain. “A Literature Review on Crowd Scene Analysis and Monitoring”, in Article · September 2016 DOI: 10.21742/ijuduc.2016.4.2.02
 24. Sonu Lamba and Neeta Nain. “Crowd Monitoring and Classification: A Survey”, in Springer Nature Singapore Pte Ltd. 2017 S.K. Bhatia et al. (eds.), Advances in Computer and Computational Sciences, Advances in Intelligent Systems and Computing 553, DOI 10.1007/978-981-10-3770-2_3.

25. Sugam Dembla, Niyati Dolas, Ashish Karigar, Dr. Santosh Sonavane. "Machine Learning based Object Detection and Classification using Drone", in International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 08 Issue: 06 | June 2021 www.irjet.net.
26. V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly detection in crowded scenes," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 1975-1981, doi: 10.1109/CVPR.2010.5539872.
27. Wafaa Mohib Shalash, Azzah Abdullah AlZahrani, Seham Hamad Al-Nufaii. "Crowd Detection Management System": 2019 IEEE.
28. Waleed Albattah, Muhammad Haris Kaka Khel, Shabana Habib, Muhammad Islam, Sheroz Khan and Kushsairy Abdul Kadir. "Hajj Crowd Management Using CNN-Based Approach", in Computers, Materials & Continua DOI:10.32604/cmc.2020.014227.
29. Wang, Xiaofei, et al. "A high accuracy flow segmentation method in crowded scenes based onstreakline." Optik-International Journal for Light and Electron Optics 125.3 (2014): 924–929.
30. Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, Yi Ma. "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network", in 2016 IEEE Conference on Computer Vision and Pattern Recognition.
31. Zhan, B.; Monekosso, D.N.; Remagnino, P.; Velastin, S.A.; Xu, L.Q. Crowd analysis: A survey. Mach. Vis. Appl. 2008, 19, 345–357.
32. Zhan, Beibei, et al. "Crowd analysis: a survey." Machine Vision and Applications 19.5–6 (2008): 345–357.

PLAGIARISM CHECK REPORT

neeeeeda khan

by Prakriti Mishra

Submission date: 20-Jul-2022 01:13AM (UTC-0500)

Submission ID: 1866839644

File name: Final_Thesis_Nida_Khan.docx (4.02M)

Word count: 10463

Character count: 56272

neeeeeda khan

ORIGINALITY REPORT

12%	8%	8%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	towardsdatascience.com Internet Source	1%
2	www.mdpi.com Internet Source	1%
3	link.springer.com Internet Source	1%
4	Fuqiang Zhou, Lin Wang, Zuoxin Li, Wangxia Zuo, Haishu Tan. "Unsupervised Learning Approach for Abnormal Event Detection in Surveillance Video by Hybrid Autoencoder", Neural Processing Letters, 2019 Publication	<1%
5	www.hindawi.com Internet Source	<1%
6	Submitted to University of Central Florida Student Paper	<1%
7	www.groundai.com Internet Source	<1%

PUBLICATIONS

PUBLICATIONS FROM THIS WORK

- 1) **“Comparative Study of Various Crowd Detection and Classification Methods for Safety Control System”** has been published in International Journal of Engineering and Management Research Volume-12, Issue-3 of June 2022 (<https://www.ijemr.net/ojs/index.php/ijemr/issue/view/37>).

PUBLICATION CERTIFICATE



CERTIFICATE

This is to certify that
International Journal of Engineering and Management Research
has scored a Publication Impact Factor (PIF) of
5.965 for the year 2021

Powered by
International Institute of Organized Research(I2OR)
India | Australia

A handwritten signature in blue ink, appearing to be 'Smy'.

Chief Editor

A handwritten signature in blue ink, appearing to be 'Smy'.

Managing Editor



editor.i2or@gmail.com

i2or.com

Comparative Study of Various Crowd Detection and Classification Methods for Safety Control System

Nida Khan¹ and Dr. Mohd Haroon²

¹PG Student, Department of CSE, Integral University, Lucknow, Uttar Pradesh, INDIA

²Associate Professor, Department of CSE, Integral University, Lucknow, Uttar Pradesh, INDIA

¹Corresponding Author: nidaintegral01@gmail.com

ABSTRACT

A crowd is a distinct collection of people or anything that is involved in a community or society. The phenomenon of a crowd is fairly well known in a wide range of academic fields, including sociology, civil engineering, and physics, amongst others. At this point in time, it has developed into the most active-oriented research and fashionable issue in the field of computer vision. Pre-processing, object detection, and event or behavior identification are the three stages of processing that are traditionally included in crowd analysis. These stages are pre-processing, object detection, and event recognition. Pre-processing, object detection, and event or behaviour identification are the three stages of processing that are traditionally included in crowd analysis. These stages are pre-processing, object detection, and event recognition. This study gives a model of crowd analysis as well as a taxonomy of the most prevalent method to crowd analysis. It may be helpful to researchers and would serve as a good introduction connected to the area of work that has been conducted.

Keywords-- Crowd Analysis, Crowd Detection, Crowd Counting

keep track of the actions of a large number of people in an extremely crowded setting at the same time. In our day-to-day lives, virtually everyone may encounter a crowd at some point. These crowds can be assembled for a variety of reasons, such as a cultural event or a sporting event. Additionally, one may encounter a throng at an airport or train station when travelling. In some circumstances, such as a sporting event, it is possible to escape the crowd; yet, in other circumstances, such as cultural events, it is quite difficult to avoid obtaining the sense of being in a crowd. Therefore, in an area with a high population density, there is always the potential for unwelcome behaviour, like as stampeding, to endanger the lives of humans. For this reason, it is essential to have a system in place for monitoring the crowd and analysing the behaviour of those individuals throughout the course of the allotted amount of time to ensure the safety of those who are present in the crowd. Manually analysing a large group of people is a highly challenging and time-consuming operation. So that we may create a system based on computer vision that is capable of carrying out these duties. There has been a significant amount of study carried out in the subject of crowd behaviour analysis over the last ten to fifteen years with only a moderate degree of success [4], indicating that there is a significant amount of room for expansion in the research field in this particular area. Because of the significance of the issue, the study subfield of computer vision that focuses on monitoring and analyzing the behavior of crowds is now accepting new topics. In the last ten years, several approaches of accomplishing these tasks have been suggested. These methodologies are intended to carry out a variety of tasks for the crowd, some of which include determining the size of the crowd in terms of its numerical strength for the purpose of effective crowd management in real time or for security reasons, forecasting the behavior of the crowd in the future, and other similar activities. Even though quite a few complicated methods have been put into practise for evaluating crowds, there is always room for improvement in terms of methods that can evaluate crowds in real time, particularly for unorganized crowds.

I. INTRODUCTION

A crowd is a distinct collection of people or anything that is involved in community or society. Tracking the movement of a crowd is quite different from following the movements of individuals inside a crowd. When tracking people, the information is computed at the level of each person being monitored. Crowded situations have become more commonplace in the actual world than they ever have been before [1], due to the growth in population as well as the variety of human activities. It poses significant difficulties in terms of public administration, safety, and security. Humans have the capacity to gather relevant information about the behaviour patterns in the surveillance area, watch the scene in real time for odd events, and give the opportunity for fast intervention. Research in psychophysics demonstrates, however, that humans have significant shortcomings in their capacity to monitor many signals at the same time [2]. Even for a human observer, it is a substantial effort to

1.1 Detection and Machine Learning Concept

The field of artificial intelligence includes the subfield of machine learning (AI). In spite of the fact that machine learning is a subfield of computer science, it is distinct from more conventional methods to computational problem solving. Object detection plays an important part in both the theoretical foundations of computer vision and its implementation in the real world. Object detection includes the tracking, detection, and counting of individual objects. The use of deep learning strategies may be helpful in a number of computer vision applications, including object identification. The world as we know it is now going toward automation, and one of the cutting-edge technologies that is being employed in automation is machine learning. Machine learning is a subfield of artificial intelligence that emulates human behaviour by training itself using data and mathematical formulas in order to perform more accurately. The field of machine learning encompasses the processes of deep learning. Deep learning is a subfield of machine learning that attempts to simulate the functioning of the human brain by using neural networks that have several layers. In the actual world, deep learning may be used in a variety of contexts. Object detection is one of its many applications, and the emphasis of this study is on that particular one. Object detection is a subfield of image processing and computer vision that identifies and locates things inside still photos, films, or real time movies. This may be done using the data included within the image or video. The techniques for detecting objects may be broken down into two distinct categories: the neural approach and the non-neural approach. Item identification using non-neural methods entails first extracting the features from an image and then feeding those features to a regression model in order to predict the position and label on the object in an image. This process is known as feature extraction. The selection of input features, the design of the model structure, and the selection of the learning algorithm are the three most important aspects of the process of developing a neural network system. Each space-time feature cube is labelled by its position, direction, and speed to build our fundamental feature set. This is accomplished by computing the change that occurs in the video of particle motion as it is occurring in the motion region.

II. CROWD ANALYSIS

A crowd's density, position, pace, and color, among other characteristics, are among its most crucial component traits. The information may be gleaned from the computer visions in either an automated or a manual fashion, depending on personal preference. The topology sensor and the typology sensor are the two kinds of sensors that are used throughout the scene capture process. The

process of extracting information from a crowd scene should be dependent on the conditions of the environment, such as changes in illumination (the transition from day to night, shadows of background images, and non-static backgrounds such as leaves blown by the wind could be detected as moving object), handling the occlusion, multiple input channels and the amount number of cameras, the changes in motion, and detecting different characteristics, such as hu or lu. This will allow for more accurate information to be obtained from the crowd scene. In most cases, the crowd model is constructed based on the extracted information that either implicitly or explicitly represents the state, while the event detection is carried out with the help of the computational model.

Both of the models have been modified to reflect the newly extracted information.

2.1 Crowd Density Estimation

Estimating the density of the crowd is one of the most difficult aspects of visual surveillance. The safety of public events with huge audiences has always been a primary concern, particularly given the existence of a significant danger of deterioration. Because of their high level of effectiveness in information collection and relatively cheap costs in terms of human resources, approaches involving video analysis are gaining a lot of traction in the field of visual monitoring of public spaces. The flow density of large malls, supermarkets, and places such as subway stations is becoming more and more serious, which brings security risks due to the crowd congestion. This is due to the rapid development of the economy as well as the increasing number of people participating in social activities.

2.2 Crowd Motion Detection

The Background Subtraction Method may be used to determine whether or not there is motion in the crowd. Background Subtraction is a technique that may be used to separate the foreground object from the background object in a series of video frames. This is indicated by the name of the method. It is possible to define a foreground object as an object of attention that not only contributes to the reduction of the quantity of data that has to be processed but also supplies essential information on the activity that is being considered. In many cases, the item that is considered to be in the front of a picture is one that is continuously moving [10]. Objects of interest may be extracted from a scene using a family of methods known as background subtraction, which can be useful for surveillance and other types of applications. The process of building a decent algorithm for background subtraction is fraught with several obstacles. To begin, it has to be resilient against shifting levels of light. Second, it should prevent itself from recognizing any backdrop items that aren't stationary, as well as any shadows thrown by moving objects. A good backdrop model should also be

able to respond fast to changes in the background and alter itself to accommodate changes that are happening in the background, such as the movement of a stationary chair from one spot to another. In addition to this, it need to have a high rate of foreground detection, and its processing time for the removal of the backdrop ought to be in real time.

III. PROPOSED METHOD

The proposed system consists of planning and executing the system which secured the answers for the

existing system required. Building an Unsupervised Abnormal Crowd Behavior Detection system is part of the assignment, which also involves doing a comprehensive analysis of the system. The goal of the task is to provide expert security for Urban by completing the construction of the system. We built a system that is capable of characterizing normal and abnormal behaviour in crowds by applying a continual image observation system and utilizing SVM, kNN, Neural Network, and linear regression algorithm to assess monitoring of crowded urban areas (see figure 1).

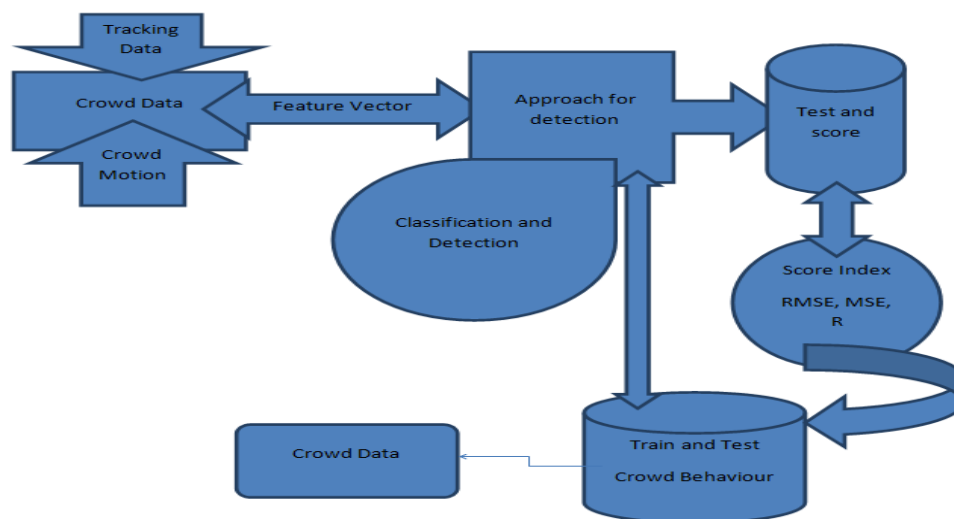


Figure 1: Proposed Model

3.1 Crowd Dataset

The distribution of crowd density in photographs of packed crowds is seldom uniform because of the differences in viewpoint and scene that exist in these photos. Figure 2 provides a selection of photographs for your viewing pleasure. Due of this, it is nonsensical to try to count the individuals in the crowd while simultaneously taking in the whole sight. The divide-count-sum mechanism was implemented into our system after it was updated as a direct result of this issue. After the images have been segmented into patches, a regression model is used in order to map each image segment to the associated local count. This step follows the segmentation of the photographs into patches. In conclusion, the number of global images may be obtained by applying a cumulative computation to the number of these patches. This yields the global image count. Users of image segmentation software may take use of not one, but two unique benefits: To begin, the density of the crowd is spread among the smaller picture patches in a way that is reasonably consistent with itself. Second, the amount of training data that may be made available to the regression model is

enhanced when picture segmentation is conducted because of the rise in the amount of data that is segmented from the image. We are now able to construct a regression model that is more robust than it was before because of the benefits that were discussed earlier. Even though there isn't any consistency in the distribution of crowd density, the overall distribution of crowd density has a continuous pattern [4]. This is despite the fact that there isn't any uniformity in the distribution of crowd density. This suggests that picture patches that are contiguous to one another should have densities that are comparable to one another. When we wish to cut the picture up into smaller pieces, we often use overlaps, which strengthen the connection between the various image patches. The introduction of a Markov random field helps to smooth out the rough edges of the estimated count across overlapping image patches, which brings the final result closer to the true density distribution [7]. This helps to rectify any potential estimate errors that could have taken place when counting the picture patches. We make use of a neural network that has all of its connections intact in order to learn a map that goes from the aforementioned attributes to

the local count. In addition, in order to extract features from different image patches, we make use of a pre-trained deep residual network. The usage of deep convolutional network features has been beneficial to a wide variety of computer vision applications, including but not limited to image segmentation, object recognition, and picture identification, to name just a few of those applications. This would imply that the learned properties of the deep convolutional network have the potential to be used to a

wide range of diverse computer vision applications. The learned features have a greater probability of properly capturing the data [9], when there are more layers in the network. On the other side, you will want more data in order to adequately prepare for a model that is more in-depth. It is not possible to train an exceptionally deep convolutional neural network from the ground up using the datasets that are presently available for crowd counting since they are insufficient.

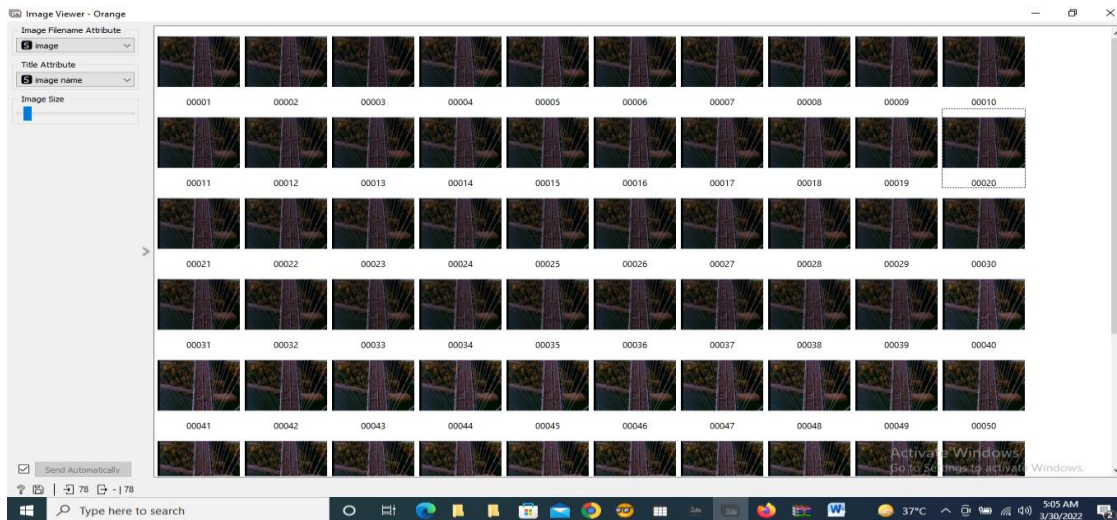


Figure 2: Images

3.2 Data Flow Architecture

Predictive models (figure 4) are assessed using the Test and Score widget, and forecasting based on newly collected information is carried out using the Predictions widget. Test and Score takes a number of different things

as input, including data (a data set for assessing models), learners (algorithms to use for training the model), and a preprocessor (which is optional) (for normalization or feature selection).

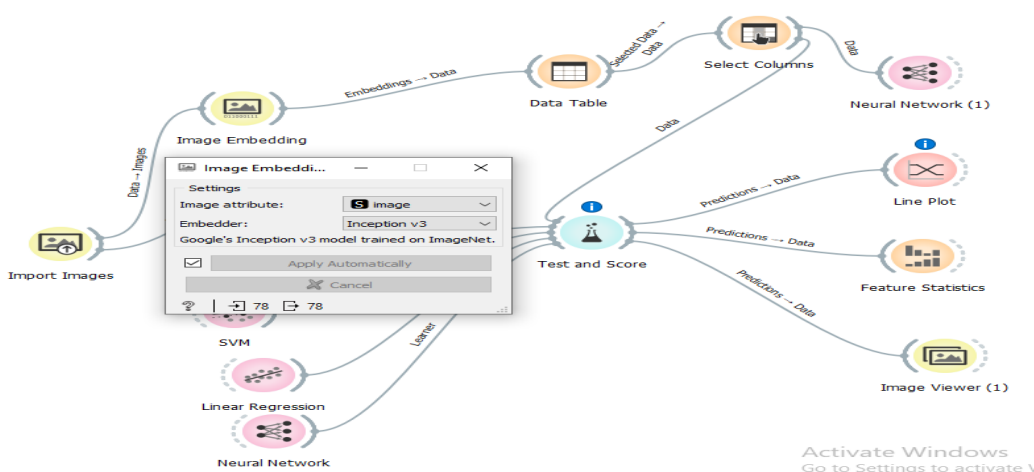


Figure 3: Flow structure

In figure 1.8, we have done the evaluation using the various methods and find the complete changes of crowd.

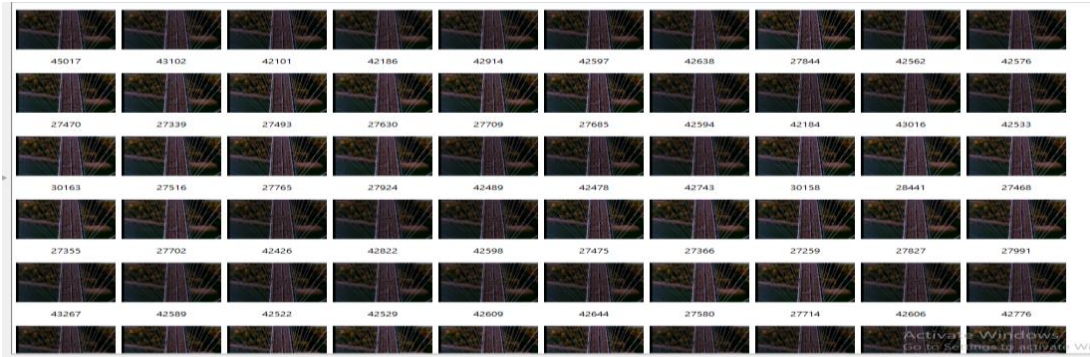


Figure 4: After Evaluation

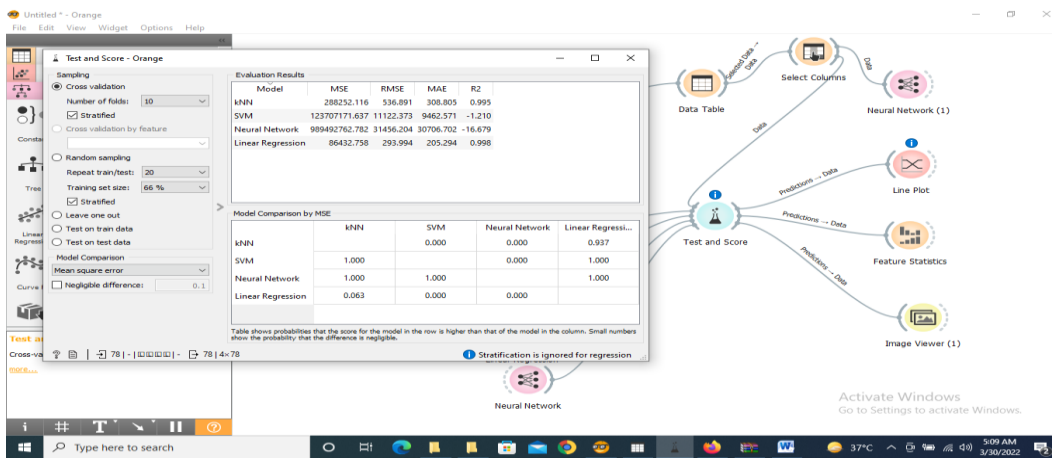


Figure 5: MSE evaluation

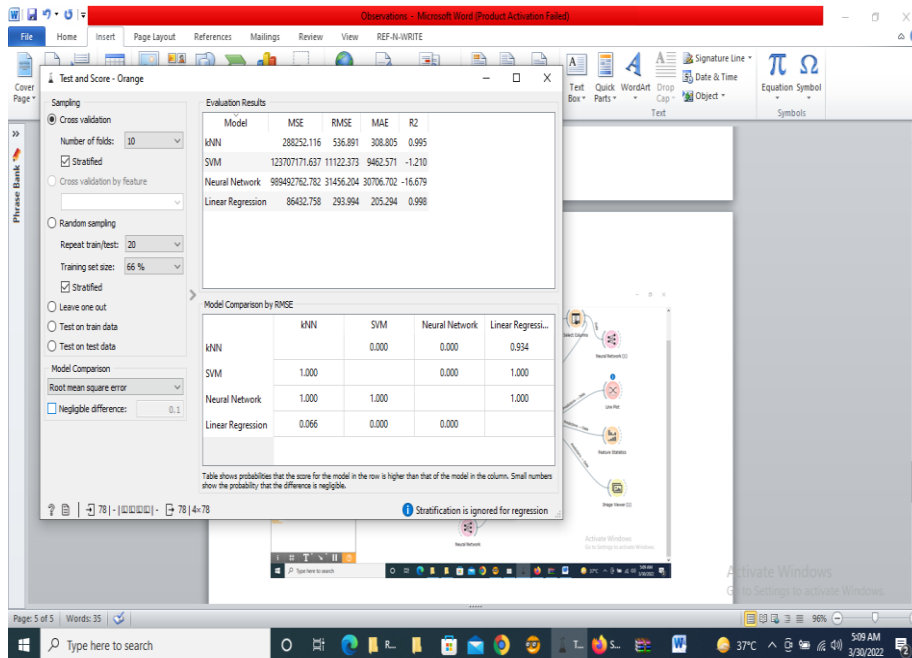


Figure 6: RMSE evaluation

In figure 5 and 6, we are illustrated Figure depicts that the MSE (0.96 to 0.99) of was relatively high when compared to kNN and linear regression. This is due to the consideration of density-level classification of image patches with a density-oriented-based repressor approach. In addition, the MSE was rather low when evaluated against the dataset as a whole. This is as a result of taking into account a skip link with scale-oriented training in order to deal with problems of varied magnitude. In comparison to SVM and kNN, the MSE had a very low standard error value. This occurred because a regulated flow of information was taken into account when passing through the convolution and de-convolution layers of the network. As a result, we have reached the conclusion that the use of a task-oriented repressor and convolution results in an improvement in accuracy when predicting the quality of a high-quality density map for datasets that include a dense and varied range of densities. As a result, the low density of datasets may be handled through the use of patch-based augmentation strategies (variable-scale) and a focus on tackling the scale-varying problem in the convolution and de-convolution layers to optimize information flow. Scale-varying issues brought on by the viewpoint view may be addressed. You may find information on each of these tactics in the paragraphs that came before this one.

IV. CONCLUSION

Using competitive machine learning model, this study introduce a novel approach to detect abnormal behaviors in crowd scenes. For the purpose of obtaining information on particle movement and crowd motion, a technique for first extracting features and then describing them has been proposed. This approach applies space-time feature cubes. Then, a detection approach that combines space-time feature cubes and a competitive neural network model is presented as a means of detecting anomalous occurrences in worldwide crowds. This algorithm is intended to identify crowds in global regions. Our system has been proved to be capable of recognizing and localizing anomalous crowd behaviors, as shown by the experimental findings produced via the use of our test video sequences. They classified the methods for evaluating crowd behavior into two distinct categories: object-based and holistic-based methods. In the object-based methodology, a crowd is seen to be a collection of distinct individuals, but in the holistic methodology, the focus is on the organization as a whole rather than on the individual distinctions that exist within it. It is anticipated that each and every individual in a crowd would move with the same characteristics throughout the whole of the investigation while using this methodology. This study investigates the many facets that are associated with crowd

modelling and crowd analysis. These are two fields that make use of a wide variety of methods for a wide variety of applications in the real world. This study investigates the many facets that are associated with crowd modelling and crowd analysis. It has been detailed how a complete review of the most recent advancements in methods for counting the number of people has been presented, together with a discussion of both the merits and shortcomings of the various approaches. The findings of this survey provide some insight into the likely trajectory of the monitoring and classification of crowds in the near future. It is vital to have a unified system that is capable of doing any kind of crowd analysis in order to achieve efficient crowd management. This system has to be able to manage a wide range of chaotic circumstances, from those involving small groups of people to those involving a huge number of people. This is done in order to discover the common gaps that are present in the existing procedures and to lay the foundation for future research in this specific subject. The goal of this study is to better understand how to do research in this particular topic. In spite of the fact that researchers have arrived at a conclusion, there are still a number of problems that have not been satisfactorily addressed, which indicates that further study is required.

REFERENCES

- [1] Zhan, B, Monekosso, D.N., Remagnino, P., Velastin, S.A. & Xu, L.Q. (2008). Crowd analysis: A survey. *Mach. Vis. Appl.*, 19, 345–357.
- [2] Andrade, E.L., S. Blunsden & R.B. Fisher. (2006). Modelling crowd scenes for event detection. *Proceedings of the 18th International Conference on Pattern Recognition*, 1, pp. 175-178.
- [3] Kim, C. & J.N. Hwang. (2002). Object-based video abstraction for video surveillance systems. *IEEE Trans. Circuits Syst. Video Technol.*, 12, 1128-1138.
- [4] Hazel, G.G. (2000). Multivariate gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, 38, 1199-1211.
- [5] Mathur, Garima, Devendra Somwanshi & Mahesh M. Bundele. (2018). Intelligent video surveillance based on object tracking. *3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*. IEEE.
- [6] Kumar, Chethan & R. Punitha. (2020). Yolov3 and yolov4: Multiple object detection for surveillance applications. *Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE.
- [7] Wang, Xiaofei, et al. (2014). A high accuracy flow segmentation method in crowded scenes based onstreakline. *Optik-International Journal for Light and Electron Optics* 125(3), 924–929.

[8] Cong, Yang, Junsong Yuan & Ji Liu. (2013). Abnormal event detection in crowded scenes usingsparse representation. *Pattern Recognition* 46(7), 1851–1864.

[9] Zhan, Beibei, et al. (2008). Crowd analysis: a survey. *Machine Vision and Applications* 19(5–6), 345–357.

[10] Li, Teng, et al. (2015). Crowded scene analysis: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on* 25(3), pp. 367–386.

[11] Lempitsky V. (2010). Zisserman a. learning to count objects in images. *Proceedings of the Advances in Neural Information Processing Systems; Vancouver, BC, Canada*, pp. 1324–1332.

[12] Camille Dupont, Luis Tobias & Bertrand Luvison. (2011). Crowd-11: A dataset for fine-grained crowd behaviour analysis. In: *IEEE Xplore provided by the Computer Vision Foundation*.

Machine Learning-based abnormal event detection in crowded scenes: An unsupervised approach

Nida Khan ¹, Dr Mohd Haroon ², Dr. Faiyaz Ahmad³

PG Student Department of CSE Integral University, Lucknow, Uttar Pradesh, India

Associate Professor Department of CSE Integral University, Lucknow, Uttar Pradesh, India

Assistant Professor Department of CSE Integral University, Lucknow, Uttar Pradesh, India

nidaintegral01@gmail.com

ABSTRACT-

We present an efficient and scalable method for detecting abnormal events in videos. Given the lack of labeled data in real-world scenarios, we use unsupervised learning to come up with a measure of regularity of input video. Our neural network generates a regularity score which is used to determine if the events in the input sequence are normal or abnormal. In videos of crowded situations, we suggest a spatiotemporal architecture for anomalous event identification. Our design develops effective encoding strategies for both spatial and temporal data throughout time. Experimental findings on the UCSD (ped1 and ped2) benchmarks show that our method's detection accuracy is comparable to state-of-the-art techniques.

KEYWORDS—Crowd Detection, CNN, Autocoders, ConvLSTM.

1. INTRODUCTION

Large numbers of people collected together or a large group of individuals clustered around a shared topic of interest constitute a crowd. It is believed that a crowd is a collection of people who are paying attention to some common object, their reaction being basic and prepotent, and it is accompanied by intense emotional responses. In a crowd, there are many people gathered together in a public area, such as a market, a train station, a magic show, a movie theatre, or a street. There are different kinds of crowds today, like protesting crowds, expressive crowds, traditional crowds, casual crowds, and demonstration crowds. The term "casual crowd" refers to any gathering of individuals who are just in the same vicinity at the same time and are not planning anything special. It does not have a clear objective or a distinct identity. A casual throng, like the one that was gathered here to cross the street, is an excellent illustration of this concept. Second, a conventional crowd is a group of individuals that congregate to accomplish a specific goal. Depending on the event, they could be attending a concert, a play, a movie, or a class lecture. It is possible to categorize a crowd as expressive or non-expressive based on the primary reason for their congregation. Political rallies for any candidate, as well as religious gatherings, are two examples of this. Fourth in terms of crowd behaviour, and acting mob engages in aggressive or other destructive conduct, such as looting, as the name indicates. The first example of an acting crowd is a mob, which is a group of people that are either engaged in or prepared to engage in violence. Another example of an acting crowd is a panic, which is a quick reaction by a crowd that results in self-destructive conduct. The fifth kind of crowd is a protest crowd, which is made up of individuals who have gathered to voice their displeasure with some aspect of society, culture, politics, or the economy.



Figure 1: Example of a crowded scene

Crowds participating in a sit-in, a march, or a rally are instances of protests, as are demonstrations. The method of evaluating data on the organic movement of things is known as "crowd analysis". In these crowd monitoring analysis, a particular crowd's movement patterns are examined, as well as when a movement pattern shifts. For example, when a certain number of people congregate in a specific location or a specific proportion of people are in a particular region, Crowd Detection will sound an alert. A picture can be counted or estimated by using Crowd Counting. There are several uses for accurately calculating the number of people in one picture, including town planning and public safety. Classification and detection of crowds utilizing surveillance cameras are made easier with this overview of crowd analysis and classification techniques.

2. Crowd Scene Detection Applications:

In real-world circumstances, crowd surveillance offers a wide range of uses.

- 2.1 Crowd Management:** The examination of crowd scenes aids in the development of crowd control methods. Crowd control is important in ensuring the safety of everyone who is present at an event, from the attendees to the staff to the artists and other participants [14, 15].
- 2.2 Virtual Environment:** Establishing a mathematical model of crowd investigations is crucial for improving the simulation of crowds and human life experiences. Virtual environments are necessary for developing a mathematical model of crowd investigations [14,15].
- 2.3 Intelligent Surveillance:** Humans' safety is the primary goal of intelligent monitoring. Intelligent surveillance systems should take the role of conventional ones, capable of crowd analysis and crowd management via alarms [14,15].
- 2.4 Public Space Design:** In the architecture of public spaces such as railroad lines, retail malls, stadiums, and airports, the detection, and categorization of crowds and their related dynamics can give prior instruction in order to ensure safety and comfort levels [14,15].
- 2.5 Visual Surveillance:** This kind of surveillance technique can aid in the automatic detection of suspicious activity and the overloading of alarms. Police can identify and apprehend offenders with the aid of visual tracking of identities [14,15].

3. Related works

According to Dongyao Jia, Crowd density classification has been a challenging subject in the field of computer vision, with several applications in the public and commercial sectors. He introduced a crowd density classification approach based on pixels and texture data. Although the crowd density

classification and recognition technique has been thoroughly investigated in the past, there are still difficulties of inaccuracy, low resilience, and inefficiency that must be addressed. This study suggests categorising adaptable crowd densities using texture and pixel data. The texture features of crowd photos can be extracted, the WorldExpo'10 dataset is utilized to integrate many texture characteristics, such as the LBP (local binary pattern), and the GLCM (Gray-level co-occurrence matrix), the Gabor, Haar-like, and Wavelet groups. Experiments have shown that the suggested technique has a classification rate of 98.2 percent [1].

A mobile-based crowd anomalous behaviour identification and management system was developed by Wafee Mohib Shalash and colleagues. IP surveillance cameras in the vicinity of the entrance gate(s) are connected to a server-side application, This is then utilised by an alarm from the server-side application by a mobile application with different user rights if crowd levels or unusual movement grow. All system users might be connected and warned quickly using the proposed architecture in the event of abnormal crowd behavior. To assure that the system will function, interface, unit, and usability tests have been performed and that users will be able to reply appropriately. The condition is satisfactory, according to the testing findings [2].

Atika Burney and Tahir Q. Syed wrote: "Crowd Video Classification Using Convolutional Neural Networks." A 2D CNN can categorize films using 3-channel image maps generated using spatial and temporal information, which reduces the amount of space and time required for video analysis when compared to a conventional 3D CNN. We were able to improve the model's accuracy by testing it against the dataset provided by them using current methods. The mid-level descriptors of the groups described in this study can be used to simplify crowd recording classification, but the classification accuracy can also be increased by treating the crowd as a whole. The extra processes of recognizing groups, computing their characteristics, and then combining them to define the crowd can be reduced to a single step of calculating the features of the crowd. [3].

The research by Yingying Zhang et al., "Single-Image Crowd Counting Via Multi-Column Convolutional Neural Network," attempts to develop a system that precisely calculates the crowd count from a single image, regardless of crowd density or viewpoint. This study includes 1198 photographs with over 330,000 annotations on the heads. The proposed MCNN model, in particular, outperforms all previous methods. Furthermore, testing shows that once the model has been trained on one dataset, it may readily be transferred to another. [4].

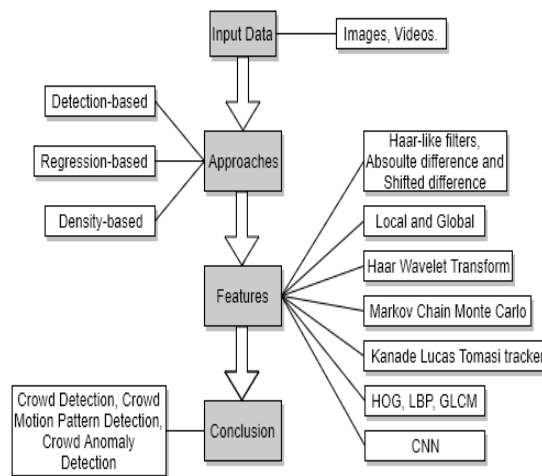


Figure2: Basic Architecture of Crowd Detection System from [5]

Mayur D. Chaudhari and Archana S. Ghotkar conducted "A Study on Crowd Detection" for safety concerns. The purpose of this research is to determine the crowd density in witness footage. It is feasible to estimate crowd size and density using face and detecting pattern recognition. Human faces have so many factors, such as color, location, and orientation, that identifying a face in a crowd can be challenging. The Deep Convolutional Neural Network has contributed to a steady improvement in the counting performance. The components of a crowd detection system are depicted in Figure 2, which include input data, techniques, features, and a conclusion. Regression techniques, density-based techniques, and detection techniques are three types of approaches to tackling the problem [5].

Akbar Khan et. al, are the members of the group. Deep Convolutional Neural Networks for Crowd Monitoring and Localization Research The real-time monitoring of huge groups of individuals may be accomplished through the application of machine learning methods and approaches. Methods of crowd-monitoring have been thoroughly examined. These models employ a convolutional neural network driven by size as the basis for their crowd counting and localization capabilities, and we conclude that they are the only ones of their kind for dense crowd pictures. They can be used to identify the heads in a photograph based on the density and scale of the image [6].

Beibei Song and Rui Sheng, "Multiscale GAN Network Combined with Deep Optical Flow," "Crowd Counting and Abnormal Behavior Detection," In this study, a crowd counting model based on the multiscale network is presented. Crowd counting is the only application of multiscale feature extraction. The regional discrimination network and the multibranch generation network are combined to form an embedded GAN module, which is then connected to the multiscale module through the use of a pyramid pooling structure. As a total, the model is trained using three different loss functions so that the model may increase its capacity to extract multiscale features from the expected picture and its ability to count accurately and robustly. The usefulness of the model has been demonstrated by a significant amount of qualitative and quantitative studies using a public dataset of three crowd counts [7].

Sherif Elbishlawi et. al, "Deep Learning-Based Crowd Scene Analysis Survey," (Deep Learning-Based Crowd Scene Analysis Survey) A overview of deep learning-based algorithms for assessing congested situations is presented in this work. The methods up for discussion can be broken down into two categories: crowd counting and crowd activity recognition. Furthermore, datasets from crowd scenes are examined. This work also provides assessment criteria for crowd scene analysis tools, which are discussed in further detail below. With the help of this statistic, you can determine how much of a difference there is between the computed and accurate crowd counts in recordings of crowd scenes. Crowd divergence (CD) is offered as a new performance indicator for the crowd scene analysis approach to give an accurate and robust evaluation of its performance. This may be done by comparing the actual trajectory/count to the projected trajectory/count and calculating the difference. The GAN framework and context-aware are promising prospects in crowd scene analysis, according to the results of this survey [8].

Crowd detection and density analysis methods are made up of Convolutional Neural Network-based methods, which include methods for analysing crowd density and identifying them using deep learning. CNN employs nonlinear functions to learn nonlinear functions ranging from crowd photos to counts. The following are some of the approaches that have been proposed in the literature. In terms of approach, CNN uses two methods of training: patch-based training, which uses small regions of images, and whole-image training, which uses the complete image [5].

Among the first to use CNNs for estimating crowd density were Wang et al. and Fu et al. In order to count people in images with a high density of people, Wang et al. developed an end-to-end deep CNN regression model. For crowd prediction, he created the AlexNet network, which uses a single neuron

instead of the 4096 neurons in a fully linked layer. He uses a patch-based inference procedure to arrive at his conclusions. Fu et al. categorized the image into five density levels instead of generating density maps: very high, high, medium, low, and very low. [5,8,9].

4. Methodology

Ideally, we would prefer to approach the issue as a binary classification problem, but doing so requires a significant amount of labelled data, and doing so is challenging for the reasons listed below:

1. Occurrence of unusual incident is relatively rare as compared to the normal incident.
2. Abnormal events present massive variations, categorizing them manually and manually labeling such events would come with huge manpower costs.

Also, Unusual things happen for one of two reasons:

1. Non-pedestrian objects, such as skateboarders, cyclists, and small carts, in the walkway.
2. Strange pedestrian movement patterns, such as people crossing a path or gazing at the grass in its vicinity.

Now, to solve this problem of lack of labeled data across various categories, as well as define new categories. We decided to tackle the problem using an unsupervised approach. We have the option to use unlabeled data or data with very few labels when employing unsupervised or semi-supervised approaches like dictionary learning, Spatio-temporal features, and autoencoders. In contrast to supervised algorithms, These techniques only require unlabeled video material that is simple to collect in practical applications and contains few to no anomalous events. Unsupervised approaches involving images mostly include autoencoder-based architectures.

The Approach

The reconstruction error is the key. To discover regularity in video sequences, we employ an autoencoder. The trained autoencoder should be able to recreate motions reliably and with little error in normal video sequences but not in irregular video sequences. Below is the algorithmic flow of our crowd analysis framework. We compare the clips generated by our neural network architecture and generate a regularity score or reconstruction cost. After that we compare this cost with a threshold which is selected empirically based on our observation. If the cost is greater than the threshold then we classify the input as abnormal.

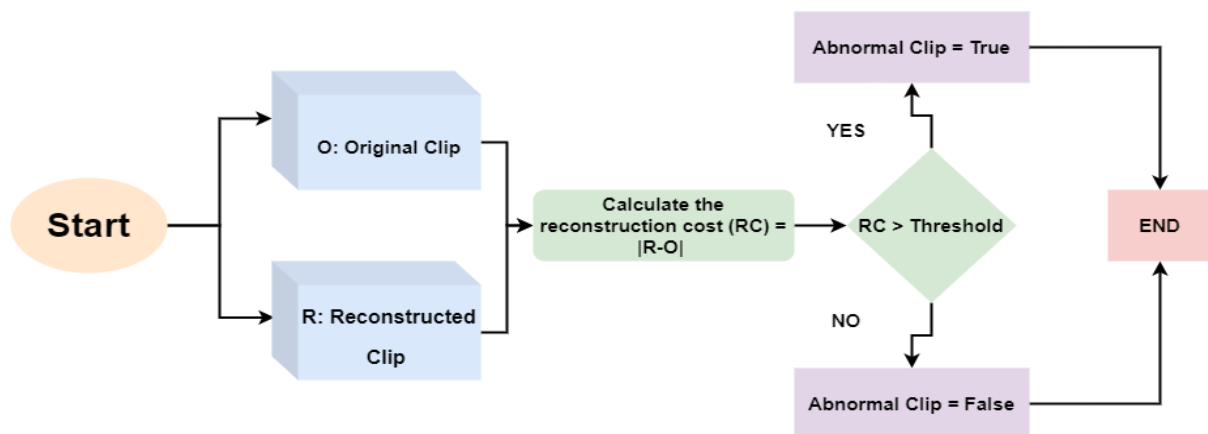


Figure 3: Algorithmic flow of our crowd analysis framework. Each input is reconstructed using our neural network and a reconstruction cost is calculated. Furthermore, this cost is used to predict if the input contains normal events or abnormal events.

4.1 Preprocessing:

Regular video frame sequences make up the training set; In order to replicate these sequences, the model will be trained. To prepare the dataset that will be ingested by our model, We follow the following steps:

1. Using the sliding window technique, create temporal sequences of 10 frames each from the training video frames.
2. To guarantee that the input photos have the same resolution, resize each frame to 256×256 .
3. Divide each pixel by 256 to scale its value from 0 to 1.

One more thing: because there are so many factors in this model, we add additional data in the temporal dimension since we require a lot more training data. We combine frames with different skipping strides to create more training sequences. As an illustration, The frames (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) make up the first stride-1 sequence, whereas the frames make up the first stride-2 sequence (1, 3, 5, 7, 9, 11, 13, 15, 17, 19).

4.2 Model architecture:

The fundamental components of our model are covered in this section. and present how these blocks interact with each other to create our neural network architecture. Since our data is in the form of frames(images), using convolutional neural networks seems to be the logical choice as they are the standard choice when it comes to the image domain. Also, we have a series of frames which we give our model as input, recurrent models like RNN, and LSTMs are state of art in sequence modeling problems. Given this, we used a combination of both convolution and LSTMs as a building block for our Autoencoder. These types of blocks are known as Convolutional LSTMs or ConvLSTMs.

Let us briefly define what auto encoders are before continuing further.

Autoencoders: Neural networks that have been trained to recreate the input are called autoencoders. There are two components to the autoencoder:

1. **The Encoder:** which may learn effective representations of the input data (x) and is also known as the encoding $f(x)$. The input representation f is located in the bottleneck layer, which is the final layer of the encoder (x). Many times, embedding is another name for it.
2. **The Decoder:** Using the encoding in the bottleneck, the decoder reconstructs the input data using the formula $r = g(f(x))$.

Building blocks of our architecture:

4.2.1 Convolution and Deconvolution:

The primary goal of using convolution is to extract meaningful features from the images. Each convolutional layers contains a set of filters which are convolved with input feature maps to produce an output feature map which is propagated to the next layers. A convolutional network learns the values of these filters on its own during the training process, parameters such as the number of filters, filter size, and the number of layers before training are user-defined and are mostly adjusted empirically. The goal of deconvolutional layers is to upscale the input feature maps into higher spatial dimensions. Here the layer learns the filters which will lead to minimum upscaling error. Figure 4, depicts the convolution and deconvolution operations along with their specific inputs.

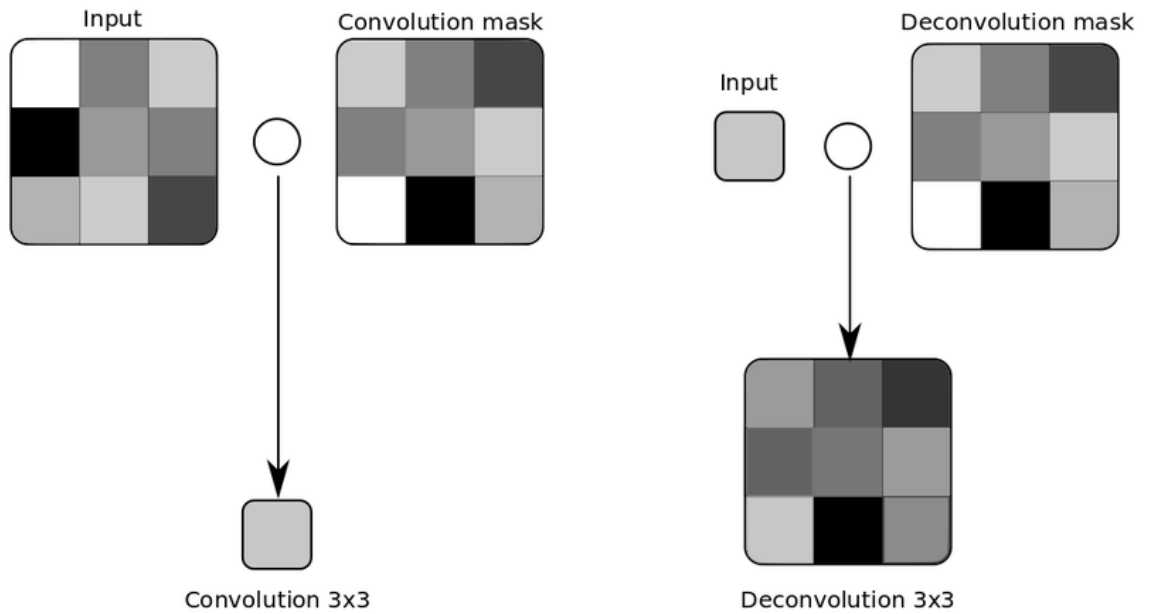


Figure 4: Convolution and deconvolution operation using a 3 x 3 kernel.

4.2.2 Long short-term memory cells (LSTM):

A time series is a collection of data gathered across a number of time periods. In such cases, an efficient approach is to use a model based on **LSTM** (Long Short-Term Memory), a type of Recurrent Neural Network architecture. In this type of architecture, the model passes the previous hidden state to the next step of the sequence. As a result, the network stores information about earlier data and uses it to make decisions. In other words, the data order is extremely important.

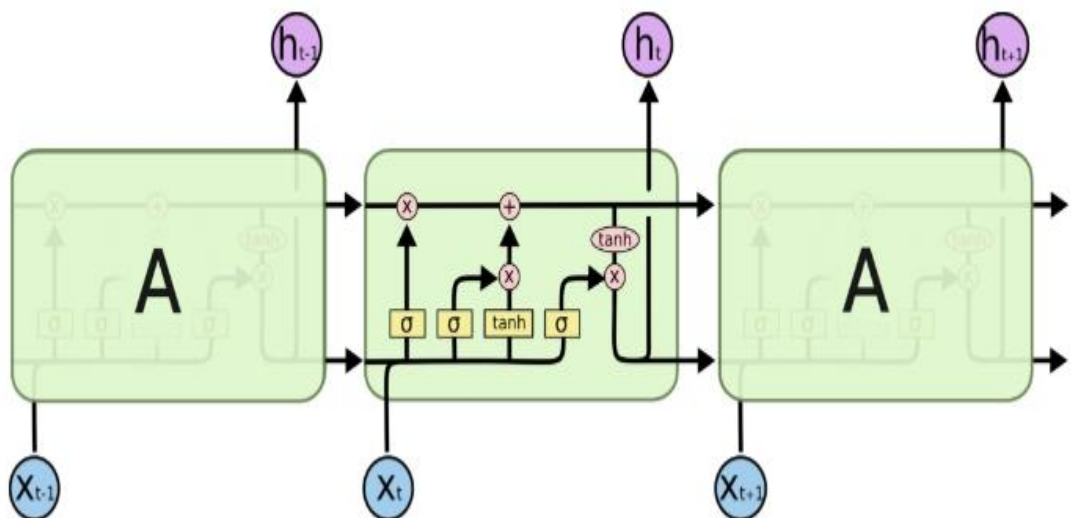


Figure 5: A standard LSTM cell

4.2.3 Convolutional LSTM:

For video frame prediction, the Convolutional Long Short-Term Memory (ConvLSTM) model, a version of the LSTM architecture, is utilized. Convolutions are used in place of matrix operations in ConvLSTM as opposed to the conventional fully connected LSTM (FC-LSTM). Convolution is used for input-to-hidden and hidden-to-hidden connections in ConvLSTM, which employs less weights and generates superior spatial feature maps. The ConvLSTM unit's formulation can be perfectly described as (1) through (6).

$$f_t = \sigma(W_f * [h_{t-1}, x_t, C_{t-1}] + b_f) \quad (1)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t, C_{t-1}] + b_i) \quad (2)$$

$$\hat{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t \quad (4)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t, C_{t-1}] + b_o) \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

4.2.4 Final model architecture:

Using all the building blocks explained above, we create a spatio-temporal autoencoder which takes a sequence of 10 images as an input and tries to reconstruct these sequences of images as output. It has a spatial encoder which encodes the 2-dimensional information and a temporal decoder to learn patterns across the axis of time. The aim of the bottleneck layer is to force the encoder to extract and encode only meaningful information which can be decoded by the subsequent decoders. Detailed architecture can be seen in Figure 6.

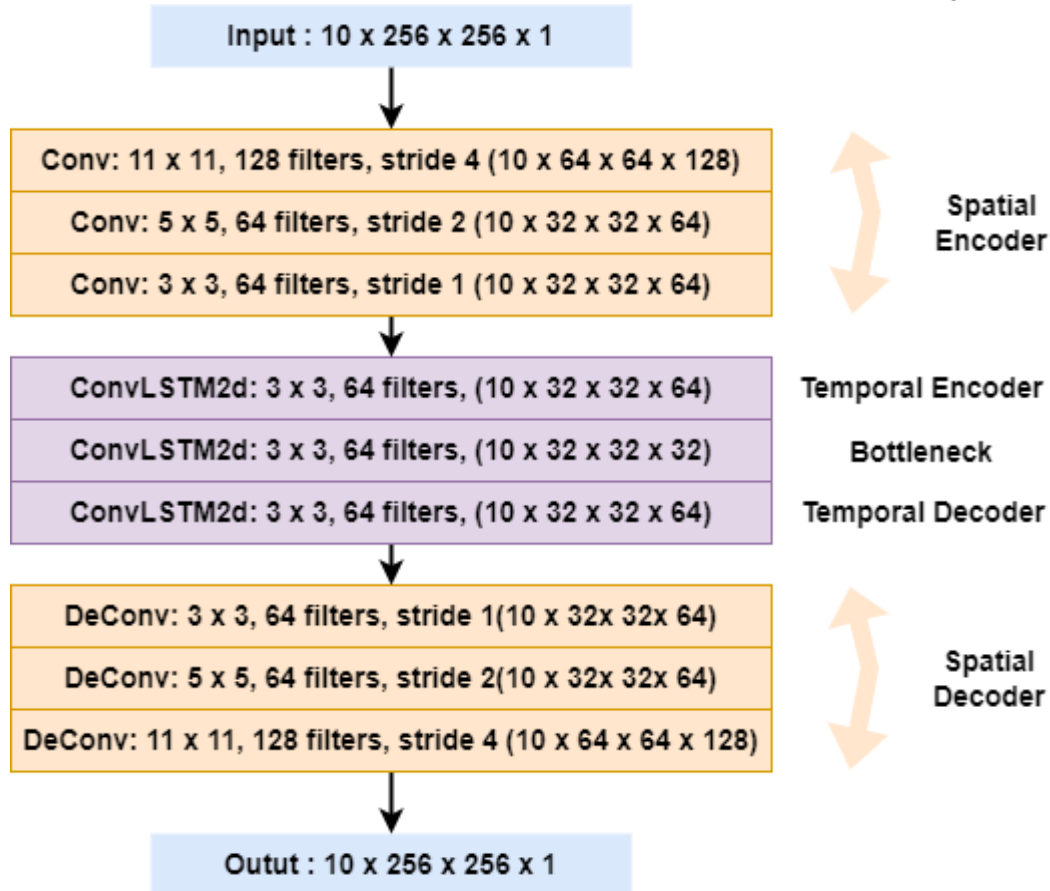


Figure 6: Final model architecture

4.3 Regularity score:

Using the L2 norm, we calculate the reconstruction error of the intensity value I of a pixel at the location (x,y) in frame t of the clip:

$$e(x, y, t) = \left\| I(x, y, t) - fw(I(x, y, t)) \right\|_2 \quad (7)$$

fw is the model that the LSTM convolutional autoencoder learned, in this case. The reconstruction error of a frame t is then calculated by adding together all pixel-wise errors:

$$e(t) = \sum_{(x,y)} e(x, y, t) \quad (8)$$

The following formula can be used to determine the reconstruction cost of a 10-frame sequence that begins at time t :

$$\text{sequence reconstruction cost}(t) = \sum_{t'=t}^{t+10} e(t') \quad (9)$$

The abnormality score $S_a(t)$ is then calculated by scaling between 0 and 1.

$$S_a(t) = \frac{\text{sequence reconstruction cost}(t) - \text{sequence reconstruction cost}(t)_{\min}}{\text{sequence reconstruction cost}_{\max}} \quad (10)$$

By deducting abnormality scores from 1, we may obtain the regularity score $S_r(t)$.

$$S_r(t) = 1 - S_a(t) \quad (11)$$

For each t in the range $[0,190]$, we first compute the regularity score $S_r(t)$, and then we draw $S_r(t)$ and based on the thresholding we predict the abnormality of the input sequence.

5. Experiments:

5.1 Dataset Description:

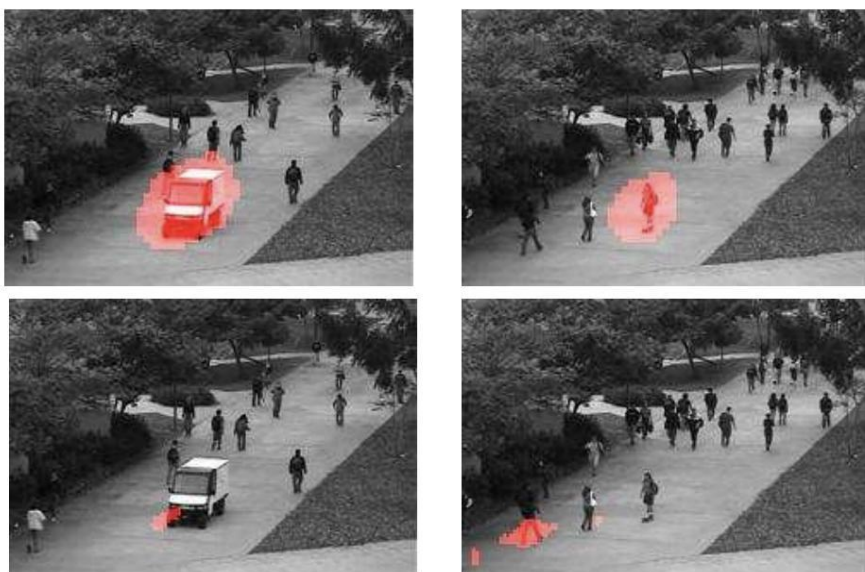
We use the UCSD dataset for congested settings in our research. [22] A stationary camera installed at a height and looking down on pedestrian pathways was used to collect the UCSD Anomaly Detection Dataset. The walkways had varying densities of people, from very few to many. Only pedestrians are shown in the video at its regular setting. Odd things happen for one of two reasons: abnormal pedestrian motion patterns the movement of non-pedestrians in the walkways Bikers, skateboarders, tiny carts, and pedestrians crossing a walkway or in its surrounding grass are examples of often occurring anomalies. There were a few cases of persons using wheelchairs as well. All abnormalities are real; they weren't produced to create the dataset. They all occur spontaneously. Two separate subsets of the data were created, one for each scene. Each scene's video recording was divided into a number of clips, each with about 200 frames.



Figure: Normal (top) and abnormal(bottom) samples of UCSD ped1 dataset



Figure 7: Normal (top) and abnormal(bottom) samples of UCSD ped2 dataset



a. Abnormal elements (highlighted red) in the dataset
Figure 8: Visualization of various frames from UCSD dataset.

5.2 Training and testing:

We use an open-source framework Pytorch for training and testing of our model. We train our model for 50 epochs for both UCSD ped1 and ped2 dataset. Using Batch size of 64 and initial learning rate of 0.0001. We use Adam optimizer for optimizing the training and use a cosine annealing learning rate scheduler [27] to change our learning rate. Figure 9 and 10 describes our testing and training workflows. During inference we take input of 10 image sequences and generate a regularity score. Based on this score our system determines if the input contains abnormal events or not.

Training

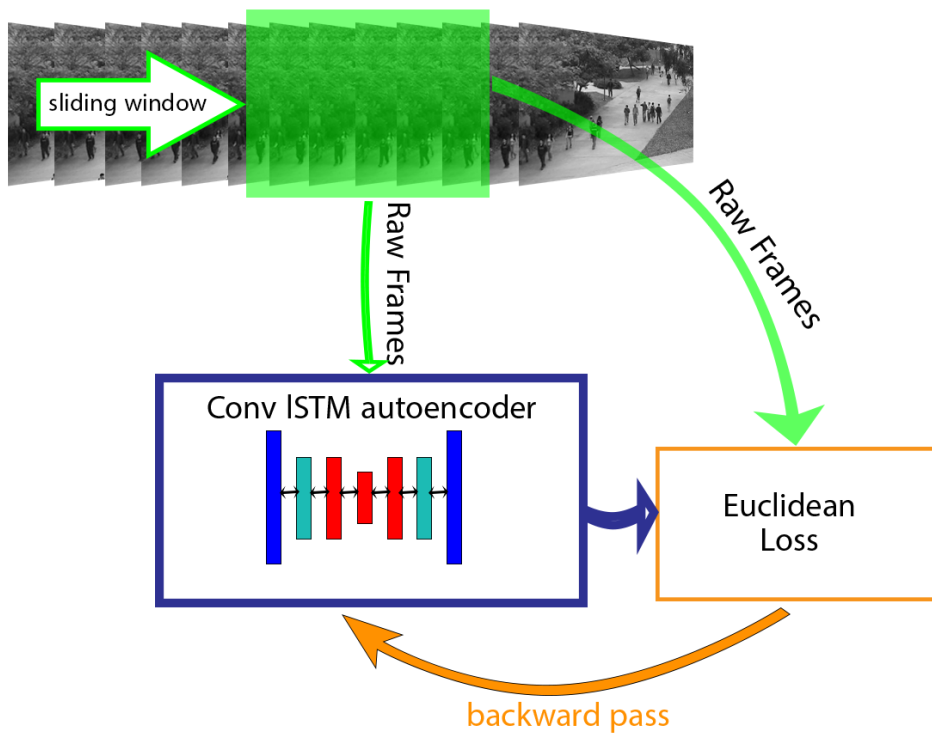


Figure 9: Training workflow of our abnormal event detection framework. A sequence of 10 images is fed as an input to our model. The model reconstructs the given input and per pixel Euclidean loss is calculated from reconstructed sequence and raw frames. This loss is back propagated for the network to learn better reconstruction.

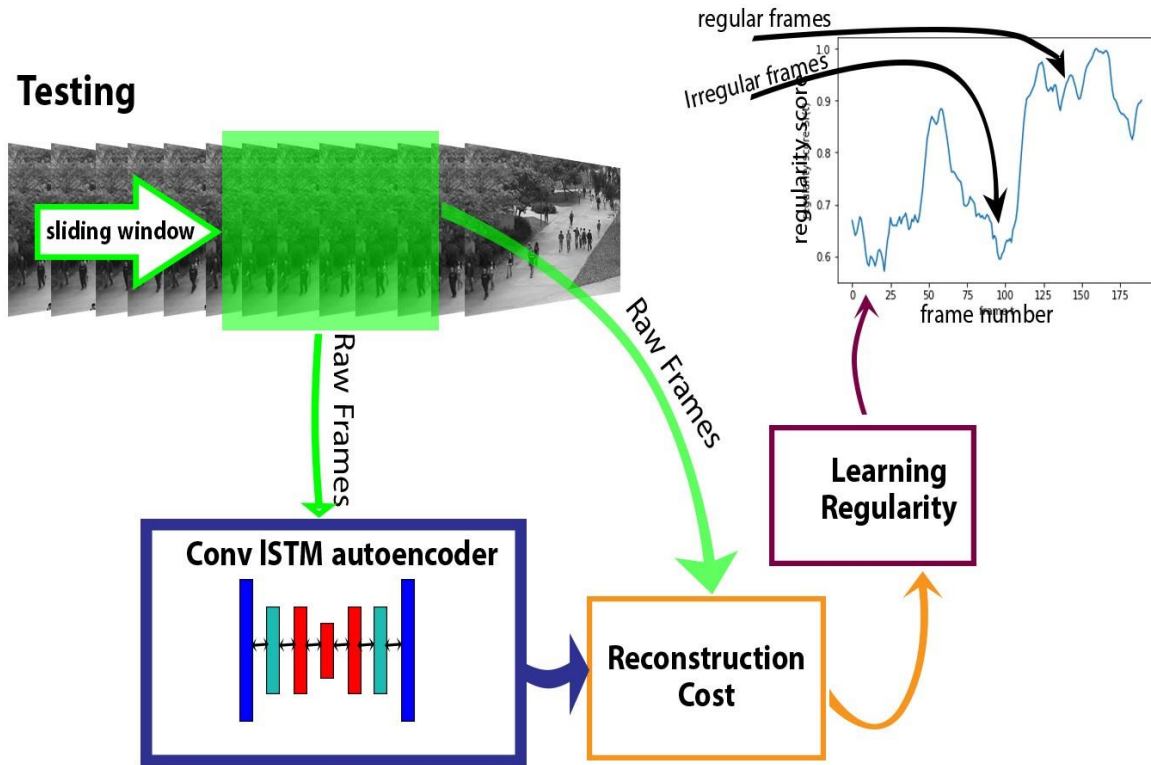


Figure 10: Testing and deployment workflow of our abnormal event detection framework

5.3 Results and Analysis

5.3.1 Quantitative results (Area under ROC curve (AUC) and Equal Error Rate)

We compare our model with other methods and observe that our model performs better in Ped1 dataset as compared to Ped2 dataset. We also observe that our results are competitive with respect to other methods. Results are shown in the table below.

Methods	Ped1 (AUC/EER)	Ped2 (AUC/EER)
SF [22]	67.5/31.0	55.6/42.0
MPPCA [23]	66.8/40.0	69.3/30.0
MPPCA+SF [22]	74.2/32.0	61.3/36.0
HOFME [24]	72.7/33.1	87.5/20.0
ConvAE [25]	81.0/27.9	90.0/21.7
Spatio-temporal AE[26]	89.9/12.5	87.4/12.0
Ours	90.1/11.5	86.2/14.8

5.3.2 Qualitative Analysis: Visualising regularity scores with respect to frames

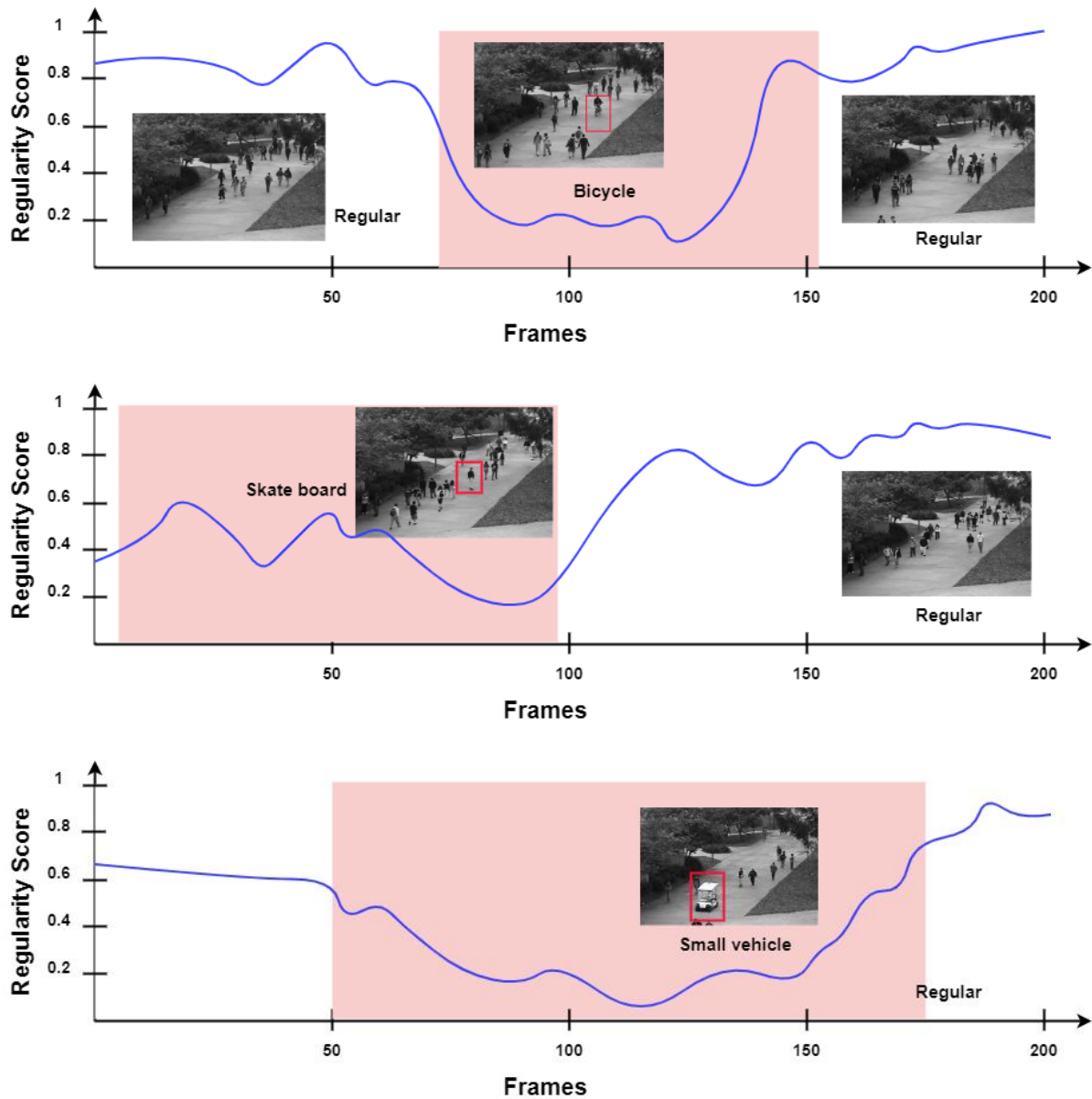


Figure 11: Visualizing the regularity score on Ped1 Test video 1, video 8, video 24 respectively. Abnormal events are marked within red bounding boxes. And abnormal events have a red background in the graph.

6. Conclusion and future work

In this paper we presented an efficient and scalable method to determine abnormal events in videos involving crowded scenes. We conducted various experiments and proposed an architecture with competitive results with respect to other methods. Furthermore, our unsupervised approach makes our method more application intensive as labelling of high amounts of data is both time and resource consuming task.

Crowd analysis encompasses a wide range of research fields, including visual surveillance, machine learning, computer vision, and pattern recognition. The identification, categorization, and recognition of individuals and groups of individuals have all been made possible via the application of crowd analysis. We believe our work will further enable many downstream applications mentioned above.

References

- [1] Dongyao Jia, Chuanwang Zhang, Bing Zhang. “Crowd density classification method based on pixels and texture features”, *Machine Vision and Applications* (2021) 32:43. Springer
- [2] Wafaa Mohib Shalash, Azzah Abdullah AlZahrani, Seham Hamad Al-Nufaii. “Crowd Detection Management System”: 2019 IEEE.
- [3] Atika Burney, Tahir Q. Syed. “Crowd Video Classification using Convolutional Neural Networks”: 2016 International Conference on Frontiers of Information Technology.
- [4] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, Yi Ma. “Single-Image Crowd Counting via Multi-Column Convolutional Neural Network”, in 2016 IEEE Conference on Computer Vision and Pattern Recognition
- [5] Mayur D. Chaudhari, Archana S. Ghotkar. “A Study on Crowd Detection and Density Analysis for Safety Control”, in *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING* · April 2018, E-ISSN: 2347-2693.
- [6] Akbar Khan, Jawad Ali Shah, Kushsairy Kadir, Waleed Albattah and Faizullah Khan. “Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review”, in *Appl. Sci.* 2020, 10, 4781; doi:10.3390/app10144781. Online - www.mdpi.com/journal/applsci.
- [7] Beibei Song and Rui Sheng. “Crowd Counting and Abnormal Behavior Detection via Multiscale GAN Network Combined with Deep Optical Flow”, in *Hindawi Mathematical Problems in Engineering* Volume 2020, Article ID 6692257, 11 pages <https://doi.org/10.1155/2020/6692257>.
- [8] SherifElbishlawi, Mohamed H. Abdelpakey, AgwadEltantawy, Mohamed S. Shehata, and Mostafa M. Mohamed. “Deep Learning-Based Crowd Scene Analysis Survey”, in: *J. Imaging* 2020, 6, 95; doi:10.3390/jimaging6090095. Online – www.mdpi.com/journal/jimaging
- [9] Waleed Albattah, Muhammad Haris Kaka Khel, Shabana Habib, Muhammad Islam, Sheroz Khan and Kushsairy Abdul Kadir. “Hajj Crowd Management Using CNN-Based Approach”, in *Computers, Materials & Continua* DOI:10.32604/cmc.2020.014227
- [10] Sohail Salim, Othman O Khalifa, Farah Abdul Rahman, AdidahLajis. “Crowd Detection and Tracking in Surveillance Video Sequences”, in *Proc. of the 2019 IEEE 6th International Conference on Smart Instrumentation, Measurement, and Applications (ICSIMA 2019)* 27-29 August 2019, Kuala Lumpur, Malaysia.
- [11] Pei Voon Wong, Norwati Mustapha, Lilly Suriani Affendey, Fatimah Khalid, Yen-Lin Chen. “Crowd Behavior Classification based on Generic Descriptors”, in 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 978-1-7281-3038-5/19/\$31.00 ©2019 IEEE
- [12] Sugam Dembla, Niyati Dolas, Ashish Karigar, Dr. Santosh Sonavane. “Machine Learning based Object Detection and Classification using Drone”, in *International Research Journal of Engineering*

and Technology (IRJET), e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 08 Issue: 06 | June 2021
www.irjet.net.

[13] Mounir Bendali-Braham, Jonathan Weber, Germain Forestier, LhassaneIdoumghar, Pierre-Alain Muller. "Recent trends in crowd analysis: A review", in *Machine Learning with Applications* 4 (2021) 100023.

[14] Sonu Lamba and Neeta Nain. "A Literature Review on Crowd Scene Analysis and Monitoring", in *Article* · September 2016 DOI: 10.21742/ijeduc.2016.4.2.02

[15] Sonu Lamba and Neeta Nain. "Crowd Monitoring and Classification: A Survey", in Springer Nature Singapore Pte Ltd. 2017 S.K. Bhatia et al. (eds.), *Advances in Computer and Computational Sciences, Advances in Intelligent Systems and Computing* 553, DOI 10.1007/978-981-10-3770-2_3

[16] BILAL TAHA AND ABDULHADI SHOUFAN. "Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research", in *IEEE Digital Object Identifier* 10.1109/ACCESS.2019.2942944

[17] Jiangtao Wang, Yasha Wang, Qin Lv. "Crowd-Assisted Machine Learning: Current Issues and Future Directions", in *PUBLISHED BY THE IEEE COMPUTER SOCIETY* 0018-9162/19©2019IEEE

[18] Ming Li , Jian Weng, Anjia Yang, Wei Lu , Yue Zhang. "CrowdBC: A Blockchain-Based Decentralized Framework for Crowdsourcing", in *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 30, NO. 6, JUNE 2019*

[19] Camille Dupont, Luis Tobias, Bertrand Luvison. "Crowd-11: A Dataset for Fine-Grained Crowd Behaviour Analysis", in *IEEE Xplore provided by the Computer Vision Foundation*.

[20] "Floor Fields for Tracking in High-Density Crowd Scenes" by Saad Ali and Mubarak Shah.

[21] Lempitsky V., Zisserman A. Learning to count objects in images; *Proceedings of the Advances in Neural Information Processing Systems*; Vancouver, BC, Canada. 6–11 December 2010; pp. 1324–1332.

[22] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly detection in crowded scenes," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 1975-1981, doi: 10.1109/CVPR.2010.5539872.

[23] Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009. pp. 935–942 (2009)

[24] Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1975–1981 (2010)

[25] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 733–742 (June 2016)

[26] L. Wang, F. Zhou, Z. Li, W. Zuo and H. Tan, "Abnormal Event Detection in Videos Using Hybrid Spatio-Temporal Autoencoder," *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2276-2280, doi: 10.1109/ICIP.2018.8451070.

[27] Ilya Loshchilov, & Frank Hutter (2016). SGDR: Stochastic Gradient Descent with Restarts. *CoRR*, *abs/1608.03983*.