MACHINE LEARNING BASED CROWD BEHAVIOUR ANALYSIS AND PREDICTION

A Dissertation

Submitted

In Partial Fulfillment of the Requirements for
The Degree of

Master of Technology Advanced Computing and Data Science

Submitted by:

Neha Srivastava

Enroll No. 1900100274

Roll No. 1901209004

Under the Supervision of:

Dr. Mohammad Zunnun Khan

(Assistant Professor)

Department of Computer Science & Engineering
Faculty of Engineering

INTEGRAL UNIVERSITY, LUCKNOW, INDIA
August, 2021

CERTIFICATE

This is to certify that **Mrs. Neha Srivastava** (Enroll. No. 1900100274) ha**s** carried out the research work presented in the dissertation titled **"Machine Learning Based Crowd Behaviour Analysis and Prediction"** 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 original work of the candidate 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 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 Zunnun Khan
Dissertation Guide
(Assistant Professor)
Department of CSE,
Integral University, Lucknow

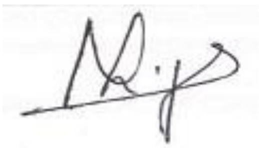
Date:
Place: Lucknow

DECLARATION

I hereby declare that the dissertation titled "**Machine Learning Based Crowd Behaviour Analysis and Prediction**" is an authentic record of the research work carried out by me under the supervision of **Dr. Mohammad Zunnun Khan**, Department of Computer Science & Engineering, for the period from August, 2020 to August, 2021 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, theses, etc., or available at web-sites without their permission, and have not included those in this M.Tech dissertation citing as my own work.

Date:



Signature
Neha Srivastava
Enroll. No.: 1900100274

RECOMMENDATION

On the basis of the declaration submitted by "**Neha Srivastava**", a student of M.Tech CSE (Advanced Computing and Data Science), successful completion of Pre presentation on 20/7/2021 and the certificate issued by the supervisor, Dr. Mohammad Zunnun Khan, Assistant Professor, Computer Science and Engineering Department, Integral University, the work entitled "Machine Learning Based Crowd Behaviour Analysis and Prediction", submitted to department of CSE, in partial fulfillment of the requirement for award of the degree of Master of Technology Advanced Computing and Data Science, is recommended for examination.

________________________	________________________
Program Coordinator Signature	HOD Signature
Dr. Faiyaz Ahamad	Dr. M Akheela Khanum
Dept. of Computer Science &Engineering	Dept. of Computer Science & Engineering
Date:________________	Date:__________________

COPYRIGHT TRANSFER CERTIFICATE

Title of the Dissertation: **Machine Learning Based Crowd Behaviour Analysis and Prediction**

Candidate Name: **Neha Sriavastava**

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.

NEHA SRIVASTAVA

ACKNOWLEDGEMENT

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

It gives me a great pleasure to express my deep sense of gratitude and indebtedness to my guide **Dr. Mohammad Zunnun Khan, Assistant 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, **Dr. Mohammadi Akheela Khanum (Associate Professor, Department of Computer Science and Engineering)** and PG Program Coordinator **Dr. Faiyaz Ahamad, 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 faculties for encouraging me constantly to work hard in 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

INDEX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

SVM	Support Vector Machine
CNN	Convolutional neural Network
RNN	Recurrent Neural Network
MMH	Maximum Marginal Hyperplane
DDoS	Distributed Denial of Service Attack
SaaS	Software-as-a-Service
PaaS	Platform-as-a-Service
IaaS	Infrastructure-as-a-Service
HRM	Hybrid Random Matrix
RIP	Restricted Isometry Property
GP	Genetic Programming
YOLO	You Only look Once
LSTM	Long Short-term Memory
GTD	Ground Truth Density
ED	Estimated Density
ReLU	Rectified Linear Unit
MAE	Mean Absolute Error
MSE	Mean Squared Error

ABSTRACT

Crowd behaviour has been widely known to have the ability to forecast the events a crowd could create. Crowd management can become extremely efficient if situations such as riots, mob lynching, traffic jams, accidents, stampede, etc. could be predicted beforehand. In this paper we have introduced a new approach to predict crowd behaviour using Multicolumn Convolutional Neural Network (MCNN). First of all, we process the input image and extract its features. Then we calculate the head count of the crowd in the image and crop patches from the image. For each patch of the image, we extract low level features. Next, we generate density maps of the objects in the image. Then our algorithm learns a linear mapping between the extracted features and their object density maps. Finally, we apply MCNN algorithm to count the crowd and predict. On performing have tested our algorithm onUCF-QNRFdataset.

CHAPTER - 1

INTRODUCTION

1.1 CROWD ANALYSIS:

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 crowd at some point of 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 crowd brings us to the most important task i.e. crowd analysis.

Crowd analysis is the process of understanding the overall behaviour 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 crowd can help us to forecast multiple real-world situations such as mob lynching, traffic jams, riots, stampede, violence etc. Using these important predictions, we can notify corresponding authorities beforehand to take preventive actions. For example, if we are able to 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.

To analyze a crowd the input used is an image or a video of that crowd.

The behaviour of the crowd is made of the aggregate behaviour of each individual in that crowd. We need to perform an analysis of this collective emotion to understand the crowd behaviour. We need to remember a few things:

- The crowd emotion or behaviour 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 behaviour of a crowd and then

comparing it with the current behaviour.

1.2 NEED FOR CROWD ANALYSIS

Crowd management is of extreme importance. Unmanaged crowd has always been the cause of dangerous situations such as stampede 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 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 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 the field of anomalous crowd detection and crowd behaviour 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 behaviour analysis. Some of these factors are listed below.

- Visual Occlusions
- Severe Clutters
- Scale
- Computational complexity
- Size
- Perspective
- Boundary restrictions

- Determination of abnormal and normal behaviour
- 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 basically two approaches that are followed in Crowd Behaviour Analysis. These are:

2.1.1 OBJECT BASED APPROACH

In this approach the crowd is taken to be a group of individual persons and to estimate the behaviour 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.

2.1.2 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 analyze the crowd as a single entity.

1.5 APPLICATIONS OF CROWD BEHAVIOUR ANALYSIS

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

Figure 1.1: Applications of crowd Analysis

- Population Counting
- Public events management
- Disaster management
- Safety Monitoring
- Military management
- Suspicious activity detection

1.6 MACHINE LEARNING

It is a branch 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 own 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

Machine Learning is traditionally divided into three approaches i.e., Supervised learning, Unsupervised learning and Reinforcement learning.

2.1.1 SUPERVISED LEARNING

Supervised learning is a branch of computing where we provide multiple examples to a machine to learn and then let the machine to 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 to learn on those rules on its own by learning from the example instances. Thus, machine learning has two phases i.e., training phase where machine learns a model and prediction phase where the machine uses that implements that model to make predictions. Examples of supervised learning algorithms are Support Vector Machine (SVM), Decision Tree, Random Forest, Bayesian classifier, etc.

2.1.2 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 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.

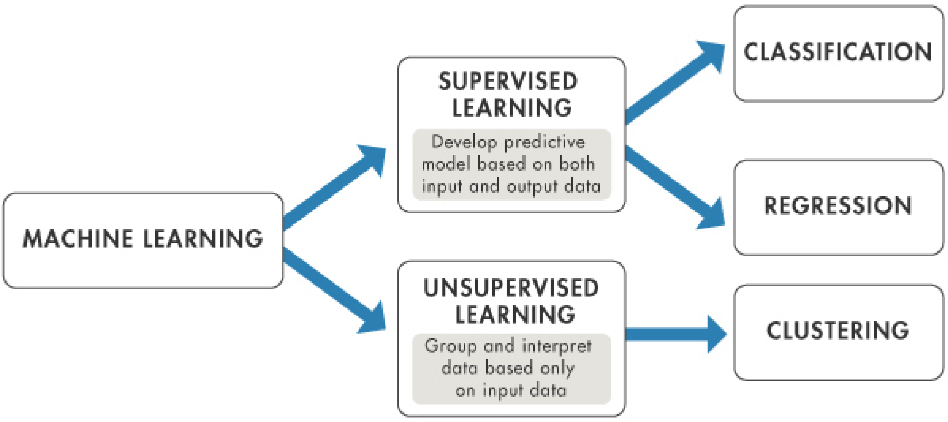


Figure 1.3: Supervised and Unsupervised learning

2.1.3 REINFORCEMENT LEARNING

This approach of 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 to learn the right steps to perform in a given environment. This learning is called reinforcement.

1.7 DEEP LEARNING

Deep learning is a branch of machine learning that has got inspiration from our biological brains. Just like 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).

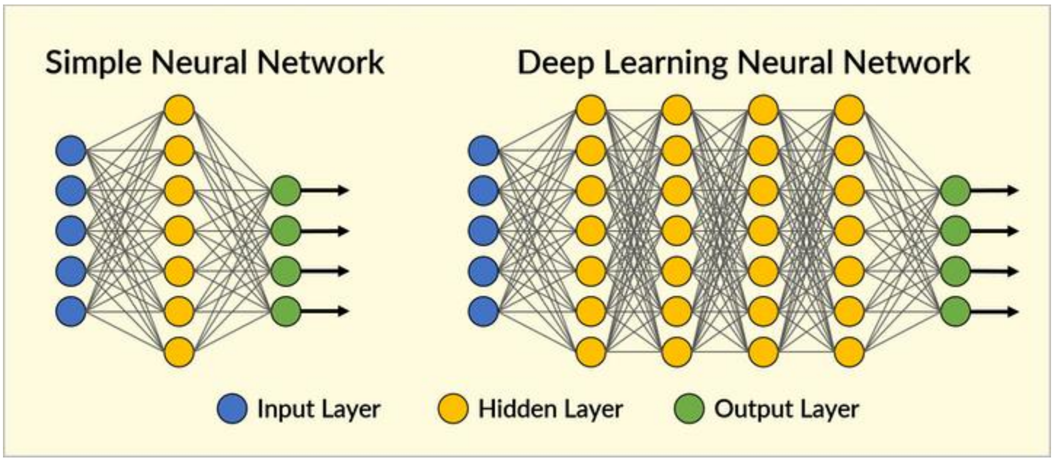


Figure 1.4: Deep learning neural networks

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 researches are still being carried out in the field of crowd behaviour 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 the past researchers.

1.9 REVIEW OF ALGORITHMS USED IN CROWD BEHAVIOUR ANALYSIS

2.1.1 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 researches

[5],[10],[31]. 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 on to 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 but 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:

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

Where W is a weight vector, such as, W = {w1, w2, …, wn};

N is the number of attributes;

b is a scalar (also called as bias)

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.

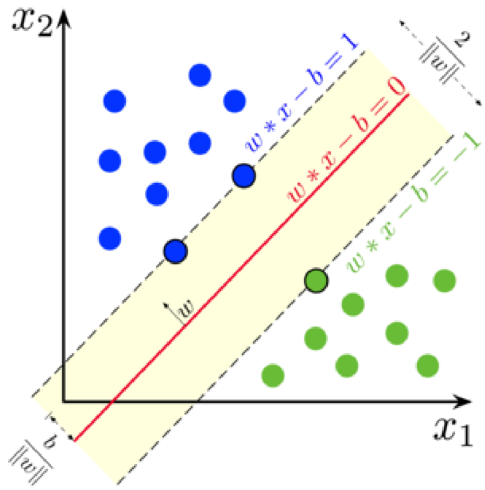


Figure 1.5: Support Vector Machine

The merits and demerits of SVM are discussed as follows [31]:

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

Advantage of SVM:

- ❖ High Accuracy: 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.

2.1.2 CONVOLUTIONAL NEURAL NETWORK

CNN are a kind Artificial neural network that have been specifically designed for analyzing images encouraged by the working of human vision [14]. This makes CNN the most apt machine learning tool to be used for the analysis of crowd images or videos. This is proven by the vast application of CNN [1],[6],[14] and its variations [2],[5],[12],[13] in crowd counting and crowd event prediction. The layers in the CNNs are 3 dimensional and unlike the regular neural networks the neurons in one layer are not necessarily connected to all the neurons of the next layer. CNNs comprise of two components: (i) the feature extraction part (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 of the learning in the model. The classification part is where we calculate the likeliness of the image being a part of a particular class. This is achieved by assigning probabilities for the same.

Convolution means combining two functions to give a new third function. It is performed in CNN by applying kernels or filters repeatedly upon 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 [13]. The entire operation is in 3 dimensions due to colours in an image. Each filter produces a separate feature map. All the features maps constitute the output of the convolution layer of CNN. This output is then fed into an activation function such as ReLU activation function. Padding is done to stop the feature map from shrinking.

A CNN has four important hyperparameters:

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

After convolution layer a pooling layer is added to reduce the dimensionality. At last, the classification layer is included that is set of fully connected layers. But these layers can accept input only as 1 dimensional data.

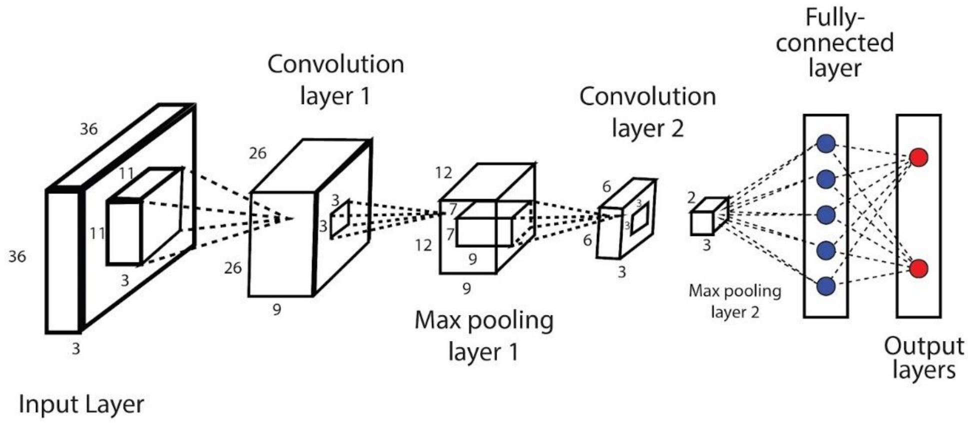


Figure 1.6: Detailed architecture of Convolutional Neural Network

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

Chapter 4

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 5

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

Chapter 6

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

CHAPTER -2
DATA SECURITY

2.1 VIRUS AND RELATED THREATS

Data stored in computing systems is always threatened to be compromised by malicious programs and software. These could be application software as well as utility programs like compilers and editors. In this section we will study some such programs called virus and worms posing danger to the data.

2.1.1 WHAT IS A VIRUS?

Virus is a program or code that attaches itself to other programs. These programs onto which the virus attaches itself are called hosts. Once the virus infects the host program, the host program gets modified maliciously. On modification a copy of the virus is also generated that can infect many other software.

2.1.1.1 Phases Of Virus

Usually there are 4 phases of a virus.

- **Dormant Phase**: In this phase the virus sits idle waiting for an appropriate event to infect the host program. This event could be time, availability of a program, etc. This phase is not mandatory to all viruses.
- **Propagation Phase**: This is the phase where the virus starts replicating. That means it starts attaching itself to a program and creating copies of itself.
- **Triggering Phase**: This phase activates the virus i.e., prepares itself to infect and perform the malicious function it was created to perform.
- **Execution Phase**: This is the phase of actual functioning of the virus. The result could as harmful as destroying files from a system or as harmless as

displaying an error message on the screen.

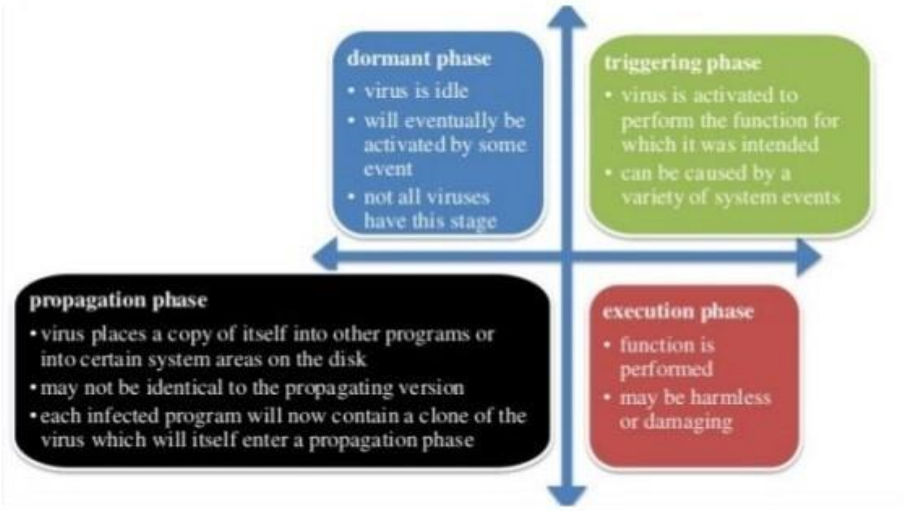


Figure 2.1: Phases of a computer virus

2.1.1.2 Types Of Viruses

Computer viruses are of multiple types. Some of the most important types of viruses are:

- **Parasitic virus**: It is the most common type of virus. It replicates itself once it has infected on program, it starts infecting others as well.
- **Memory-resident virus**: This type of virus lives in main memory and infects every program that is executed by the system.
- **Boot Sector virus**: This virus infects the boot record and spreads when a system is booted from that corrupted boot disk.
- **Stealth virus**: Such viruses are designed in such a way that they stay hidden from antivirus software.
- **Polymorphic virus**: This is a virus that modifies itself with each infected

program. Hence, it is very difficult to recognize and track this virus.

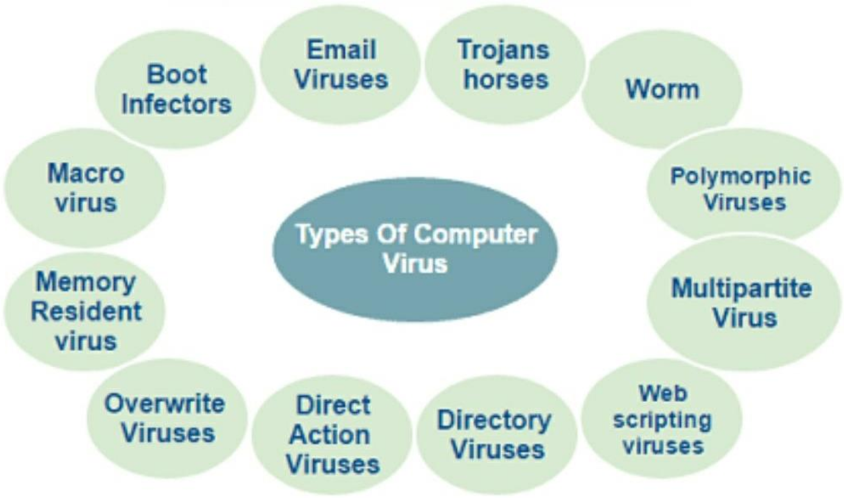


Figure 2.2: Types of a computer virus

2.1.2 WHAT IS A WORM?

Worms are programs that can replicate themselves and move from one computer to another through network connections. When a worm arrives at a computer, the worm may be activated to replicate and propagate again. In addition to propagation, the worm usually performs some unwanted function. An e-mail virus has some of the characteristics of a worm

2.2 VIRUS COUNTERMEASURES

We need to save our system from virus and for that purpose we have to follow some approaches.

2.2.1. ANTIVIRUS APPROACHES

The solution to the virus problem can be designed using any of the approaches given below.

- **Prevention**: Though this is the best approach but in general it is impossible to achieve. In this approach we try to prevent the virus from entering the computer system in the first place itself.

- **Detection**: But usually the virus is able to enter our system and we need to determine that that virus has got into the system and locate it.
- **Identification**: After the detection we have to recognize the virus that has infected the programs in the system.
- **Removal**: At last, we need to remove the virus from all infected programs so that it can not spread and infect other programs. Sometimes it is also required to remove the entire infected program and restore a fresh copy.

2.2.2. ADVANCED ANTIVIRUS TECHNIQUES

In this section we are going to discuss some important antivirus techniques introduced in the recent times.

- **Generic decryption**: This technique helps an antivirus program to detect even the most difficult virus i.e., polymorphic virus. As we know that polymorphic virus need to decrypt themselves to get activated, this antivirus technique leverages that fact to identify it.
- **Digital Immune System**: This technique was developed by IBM. It addresses the virus that propagate through the internet. The use of Integrated mail Systems and Mobile Program Systems have led to rising cases of internet-based viruses. The Digital Immune System has a very short response time and can detect the virus almost as soon as it enters the system. This system captures the virus, analyzes it and applies protection from it. Along with this the Digital Immune System sends the information of the virus to other IBM antivirus so that they can be detected before they are able infect any other

system.

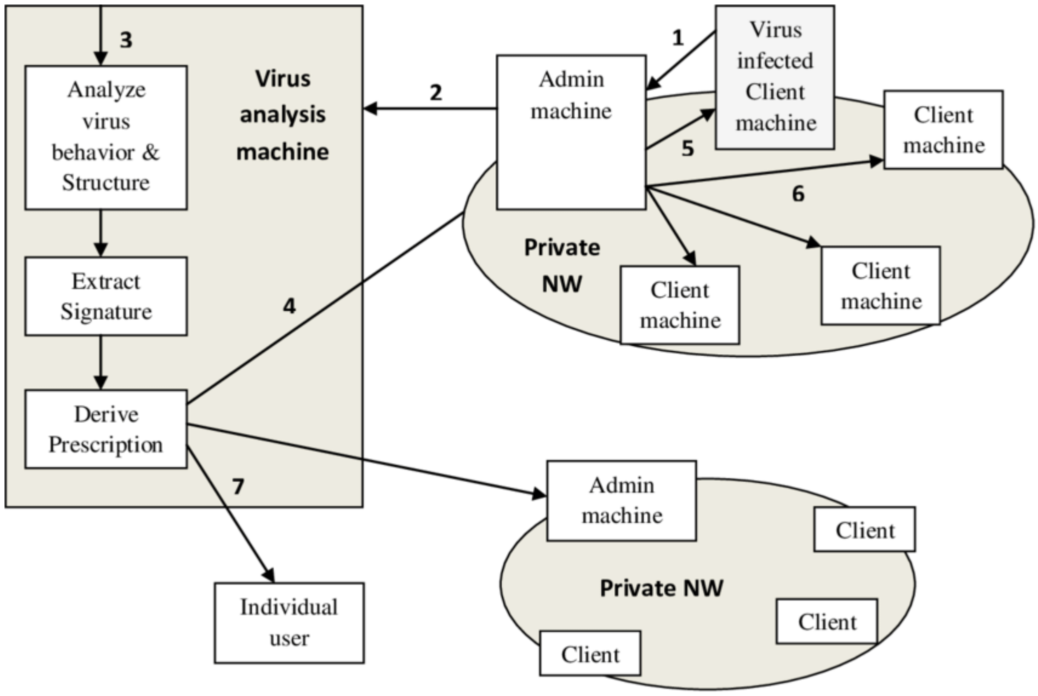


Figure 2.3: Digital Immune System

2.3 DISTRIBUTED DENIAL OF SERVICE ATTACKS

Distributed denial of service attack (Figure 2.5) is the most serious, large-scale DoS attack in which the attacker uses multiple sources i.e., the different IP addresses or machine to attack the target at the same time. This is done from thousands of malware infected hosts [40]. What makes this attack a very serious one that the redundant requests or packets comes from various sources at the same time, making it hard for the victim to identify the host. Due to the distributed nature of the attack, it is neither possible to stop this attack as it cannot be resolved by any simple filtering techniques, nor the authorized users can be distinguished from the non-authorized ones, because there is no sole point of origin.

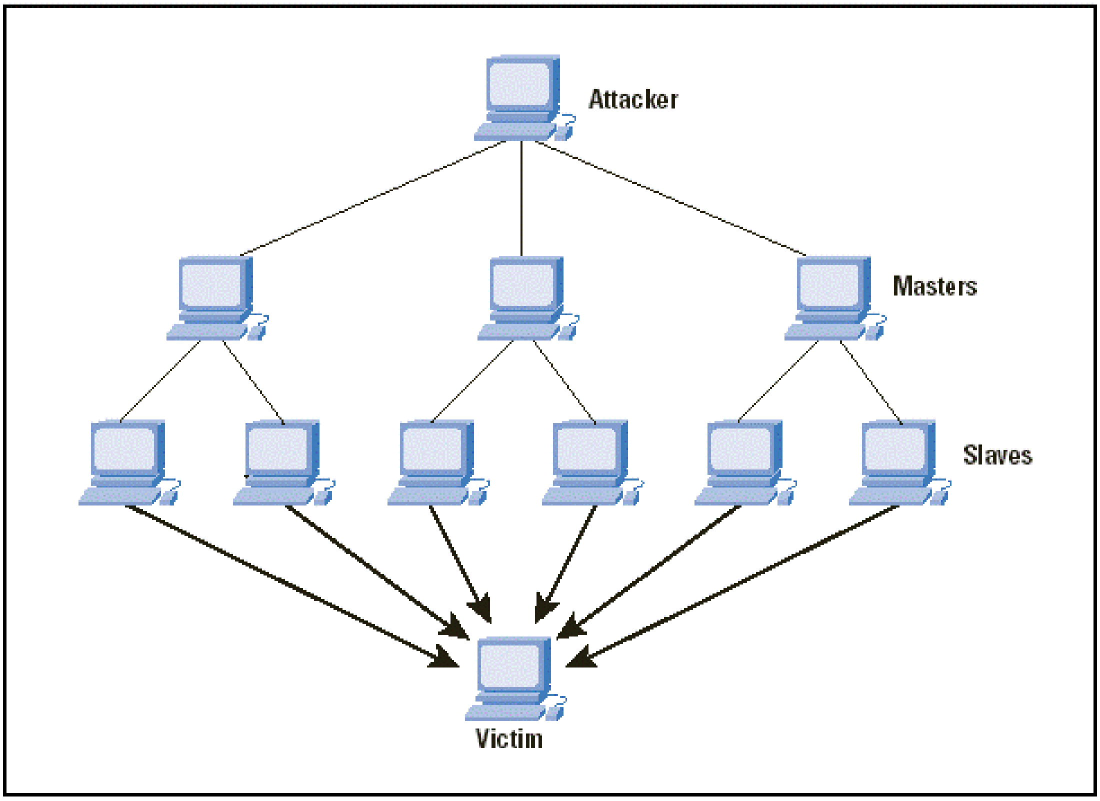


Figure 2.4: Distributed Denial of Service Attack

2.3.1. DETECTION OF DDoS ATTACK

There are various indications that can tell whether a particular system is under denial-of-service attack or not. They are:

- The files will take longer time to load, due to the slow network performance.
- Sudden loss of connection to a network.
- Unresponsiveness of a particular website, when loaded on web page.
- Increase in processing time.
- Unnatural behavior of the system

2.4 DATA SECURITY IN CLOUDS

Cloud security is a type of Cyber Security, also called as cloud computing security. It has various set of policies, controls, procedures and technology that are used to secure cloud-based systems, data and environment from both external and internal cyber security threats such as theft, leakage and destruction.

The cloud security can be optimized according to the business need, from authentication to filtering of traffic. Since the management and optimization of these rules can be can be done from a single place, within the organization or even outside, overheads created by its administration can be reduced drastically.

2.4.1. TYPES OF CLOUD COMPUTING

On the basis of category of cloud computing used, the cloud security also differs. Following are the four chief categories of cloud computing, according to which cloud security also can be categorized:

2.4.1.1 Public Cloud

We refer to cloud as a public cloud when a third party provides computing services and infrastructure on demand and other organizations can also share the services through the internet. It provides all three kinds of infrastructures such as software-as-a-service (SaaS), infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS).

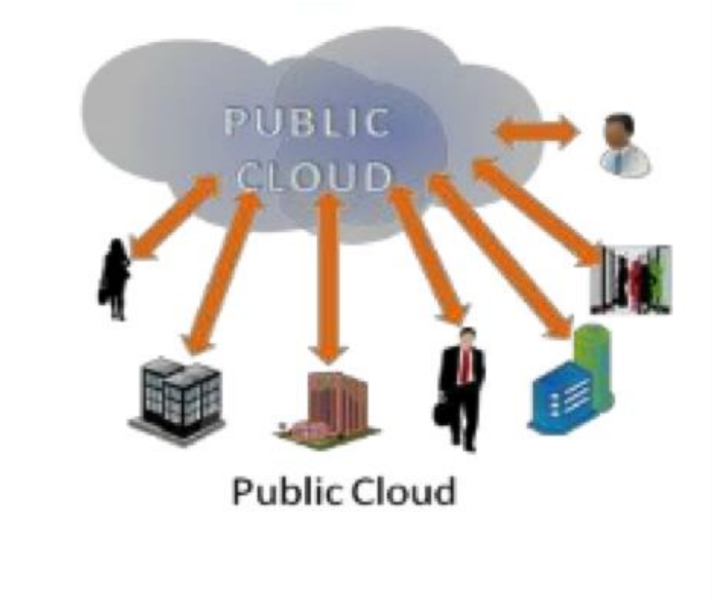


Figure 2.5: Public Cloud

2.4.1.2 Private Cloud

A cloud is called a private cloud if all the computing services and infrastructure provided by the cloud are for the specific use of one organization only.

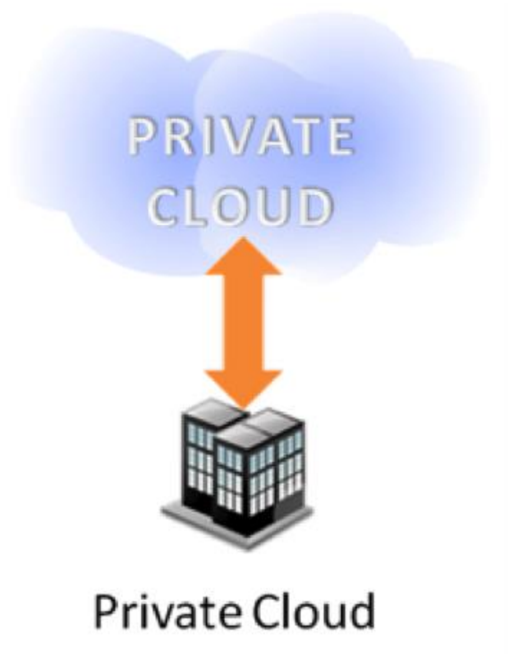


Figure 2.6: Private Cloud

2.4.1.3 Hybrid Cloud

In a hybrid cloud an organization can combine its private cloud with the public cloud. The key is to balance this combination in such a way that the organization is able to use the optimal cloud solution for each of its application.

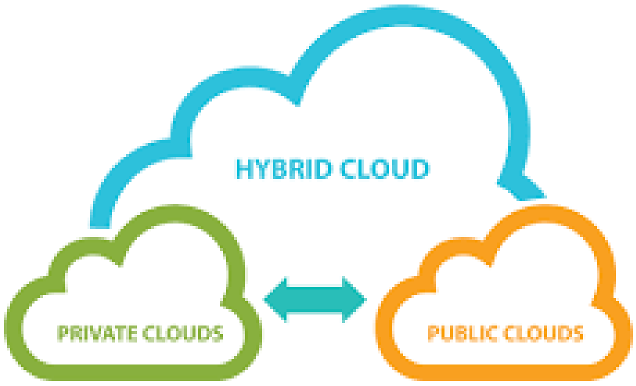


Figure 2.7: Hybrid Cloud

2.4.2. CLOUD COMPUTING SERVICES

There are basically three types of computing services that are provided by the cloud providers.

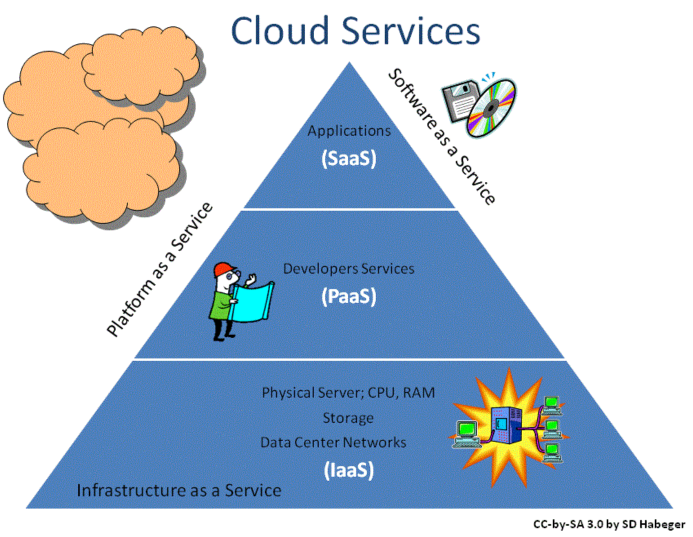


Figure 2.8: Various cloud services offered

- **SaaS (Software-as-a-Service)**

In this service security of data and user access are the sole responsibilities of the user.

- **PaaS (Platform-as-a-Service)**

In this service the responsibility of the customer is also to secure application loaded on cloud along with the data and user access

- **IaaS (Infrastructure-as-a-Service)**

In this service since the entire infrastructure in involved thus only securing the applications, users' access and data is not enough, the protection of operating systems and network traffic is also necessary by the user.

2.4.3. CLOUD SECURITY

Usually, concerns regarding security have been the primary challenge for organizations that need cloud services, especially public cloud services. But, in response to demand, the security provided by cloud service providers is growing rapidly and is outperforming on-premises security.

- Some of the best practices followed in cloud security include the following
- Shared responsibility for security
- Data encryption
- User identity and access management
- Collaborative management
- Security and compliance monitoring

CHAPTER – 3

LITERATURE REVIEW

3.1 LITERATURE SUMMARY

[1] The researchers in this paper have proposed a new strategy of performing abnormal crowd event detection by using Motion Information Images (MII) and with Convolutional Neural Networks (CNN).

[2] This paper aims to come up with a novel methodology of detecting crowd emotions and behaviour by making use of 2D Convolutional Neural Networks (ConvNets). The goal of this network is to perform a classification of the general behaviour of the crowd. In order to perform this experiment, the researchers built up a dataset of images that focused on six kinds of emotions such as anger, sadness, excitement, happiness, fright and neutral.

[3] The goal of this paper is to detect of using a combined approach of deep learning networks and compressed sensing. With this aim, the researchers have created a new Hybrid random Matrix (HRM). This matrix is helped to fulfil the Restricted Isometry property (RIP).

[4] The researchers in this paper learn various methods of data mining and improve the accuracy of detection of abnormal behaviour in crowd videos by machine. This paper brings forward a new IDS video crowd. It conducts experiments by using Unix user’s shell command data. The experiment has three parts.

[5] This paper introduces a lightweight and low complexity end-to-end network to analyze crowd with congested scenes. This is called as a lightweight network since it uses only 0.86 M parameters. This is a very small number as compared to other neural

networks.

[6] This purpose of this paper is to find the crowd density map and the number of people in a group by implementing Convolutional Neural Network (CNN) and Deep learning models. The researcher has laid down the lack of training samples, drastic blockages, and changes of perspective to be the reason for the exploration of deep learning for crowd analysis.

[7] In this paper the researcher has built an automated anomalous behaviour model for the detection of crowd video sequences. The methodology introduced in the paper has made use of the gradient based approach for modelling motion of the objects by implementing the concept of cuboids (spatio-temporal volumes) to find the difference between various activities within a time varying video sequence.

[8] The aim of this paper is to propose a new real time architecture with the aim of crowd event detection. The fundamentals of this architecture lie in object detection that is done with fixed-width clustering. To achieve this, the crowd events in this research have been divided into two main classes. One is either walking or running and the other is either splitting or merging. The second class is later classified as local or global. The author has used YOLOv2 (You Only Look Once) for detecting objects in the video frames.

[9] The purpose of this paper is to detect unusual crowd behaviour in a video sequence by using probability models of speeds. Firstly, two kinds of probability densities are assigned to each video frame. One represents the speed of the crowd and the other probability density represents the direction of the crowd. The author has

used Expectation -Maximization algorithm (EM) for converting the optical flow vectors into models of probability for speed and direction for each video frame

[10] This aim of this paper is to find a new technique for the classification of large-scale crowd density by using dynamic texture analysis. The paper also estimates the crowd flow direction. The analysis consists of three parts. The first part deals with the locations of points of interest using Hessian matrix. The second part is about the extraction of local spatio-temporal features of each interest point detected before.

[11] this paper introduces a Multiview-based Parameter Free framework (MPF) and Structural context descriptor for the identification of coherent groups in crowd videos. This paper the brings out three aspects of the research. First, the individuals in a crowd are tied closely with their neighbours which represents their structural property.

[12] This purpose of this paper is to introduce a strategy to detect and classify abnormal crowd events. This is especially for the events of panic and fight. This new strategy states that to correctly understand the motion in the crowd videos (both spatial and temporal context) it is important to analyze the optical flow combined with density heat maps since motion in the crowd might be triggered by things apart from people.

[13] The researchers in this paper have introduced a novel neural network called DA-Net. This network is more robust and much deeper. The network contains two parts. One contains eight blocks and the other is a multi-layer aggregation of deform blocks that makes the network capable to learn extra offsets to start the spatial sampling locations.

[14] The aim of this research is to calculate the measure of how many people are present in the images of dense crowd with the help of deep convolutional neural networks. The researcher has developed a special convolutional neural network that is containing two parallel modules finding different features from the crowd images.

[15] This paper aims to perform crowd counting on still images of crowds that are taken from various sources with the help of a drone camera. Multiple sources of information such as HOG, LBP, CANNY are used for this purpose. A method of feature fusion is introduced in which different types of features of crowd images are extracted and studied such as texture features.

[16] This aim of this paper is to create the combined distribution of the pixels in an image while considering the temporal information as well as the spatial information of the input and study the properties of the pixels of the image for multiple adjacent video frames. This leads to anomaly detection in crowd images or video frames.

TABLE 3.1

THE SUMMARY OF ALGORITHMS USED IN SOME PAST RESEARCHES

S.NO	TITLE	TECHNIQUE USED	FINDINGS
1	Abnormal Crowd Behaviour Detection Using Motion Information Images and Convolutional Neural Networks	Motion Information Images and Convolutional Neural Networks	Results of the evaluation of UMN and PETS2009 datasets prove that the suggested approach was effective and accurate.

2	Crowd Emotion Analysis Using 2D ConvNets	2D convolutional neural network with residual connections	Proposed architecture is suitable for use on single images, bulk images as well as on videos. It can be implemented on the local and web servers for real time emotion classification.
3	Violent crowd behaviour detection using deep learning and compressive sensing	deep learning network, compressive sensing, Hybrid Random Matrix (HRM), Violent Flows (ViF) algorithm	As opposed to other approaches the proposed strategy called Vif-HRM-NET has performed more effectively.
4	Video crowd detection and abnormal behaviour model detection based on machine learning method	IDS anomaly detection model	Experimental results show that the detection performance of the new model is much better than that of Lane et al.
5	A LIGHTWEIGHT NEURAL NETWORK FOR CROWD ANALYSIS OF IMAGES WITH CONGESTED SCENES	CNN, Cascaded convolutions, Max pooling	It was found that the density map generated are very accurate and can also perform crowd counting accurately as well.
6	Crowd Scene Analysis Using Deep Learning Network	CNN, Python OpenCV	This paper performs crowd scene analysis using a convolutional crowd dataset comprising of 100 videos collected from 800 crowd scenes and extracting a set of 94 attributes.
7	Anomalous Crowd Behaviour Detection in Time Varying Motion Sequences	Statistical modelling, Genetic Programming	Results profess that the suggested strategy outperforms the conventional and existing techniques in accuracy of classification as well as time involved.

8	Cluster-based Crowd Movement Behaviour Detection	YOLOv2, fixed-width clustering algorithm	For evaluation this paper uses six video sequences form PETS2009 and the results have shown the accuracy of the proposed methodology to be between 80%-95%.
9	Abnormal Crowd Behaviour Detection Using Speed and Direction Models	Expectation maximization algorithm, Optical flow vectors, Probability models	Abnormal behaviour can be detected by calculating the distance between the probability models of current model with those of the previous ones. If the distance is more than a threshold value it indicates an anomalous event in the crowd
10	A Large-Scale Crowd Density Classification using Spatio-Temporal Local Binary Pattern	dynamic texture analysis, multi-class support vector regression, spatio-temporal local binary (RIST-LBP) pattern	It was observed that the proposed strategy has lower computational complexities and high efficiencies as compared to the conventional pixel-based techniques.
11	Detecting Coherent Groups in Crowd Scenes by Multiview Clustering	Multiview-based Parameter Free framework (MPF), Structural Context descriptor, Self-weighted Multiview Clustering method	Results of the analysis on various datasets have proven the effectiveness of the proposed approach.
12	Abnormal Behaviour Detection in Crowded Scenes Using Density Heatmaps and Optical Flow	convolutional LSTMs, density heat maps, Optical Flow	Once trained on activities that are either normal or abnormal the network is ready to classify the abnormal events in crowd videos. The paper also curates a new dataset using GTA V engine that contain videos with abnormal behaviour.

13	DA-Net: Learning the Fine-Grained Density Distribution With Deformation Aggregation Network	DA-Net, multi-layer aggregation, deform blocks	Multi-layer extraction allows the network to immensely improve the upkeeping of the spatial locations. Also, the paper introduces adaptive weights to replace direct aggregation to evaluate the reliability of multilevel outputs.
14	Deep and Wide Convolutional Neural Network Model for Highly Dense Crowd	CNN	This approach was tested on ShanghaiTech Part-A datasets with performance metrics as mean square error and mean absolute error. As a result, the approach tested successfully on comparing these metrics with other approaches.
15	A Method Based on Multi-source Feature Detection for Counting People in Crowded Areas	HOG, LBP, CANNY, SVM, regression analysis	This approach was found to be easy and fast to implement. This experiment gives comparative results as compared to the joint HOG.
16	Machine and Deep Learning for Crowd Analytics	Bayesian model, MATLAB	As a solution this paper the Bayesian model as a significant feature. The proposed method is implemented using MATLAB. And on average achieves 88.83% accuracy when applied on the videos from the same dataset.

3.2 CONCLUSION

The purpose of performing the literature review is to highlight the researches done in the past by other researchers in the field of crowd behaviour analysis, abnormal crowd behaviour 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 approach such as Motion Information Images (MII) and optical flow vectors [1]. [7] had used gradient to create 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 as a separate entity. On the other hand, the holistic approach considers 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. This in turn, is solved by [5] making use of cascaded convolution filters.

Crowd analysis is still a work in progress striving for more accuracy and efficiency than the past approaches.

CHAPTER – 4

PROPOSED WORK

4.1 PROPOSED WORK

This section describes our proposed methodology in detail. Figure 1 shows the overview of the methodology. We have presented a novel approach to predict crowd behaviour that uses Multicolumn Convolutional Neural Network (MCNN). The pre-processing of the input image is the first step. We use OpenCV for the purpose of pre-processing of the image. The aim of this pre-processing is the extraction of low-level features. For that purpose, we first get the image spectrum and then filter it to crop it into patches depending on the frame set. For each such patch we extract low level features. From this we create density maps of the objects and linearly map them with the extracted features to obtain the Ground Truth Density of the image. At this step we introduce our MCNN algorithm to predict the crowd density or the Estimated Density.

4.2 AIM OF THE RESEARCH

Today one of the biggest challenges that we face is crowd management. Be it any public places such as streets, markets, train stations, airports, etc., crowd management is the need of the hour. A well-managed crowd can avoid problematic situations such as traffic jams, riots, etc. Also crowd management might also help avoid some extremely dangerous scenarios like stampedes, mob, lynching and accidents.

The aim of this research is to find out what kind of crowd is present in an image. We try to find whether a crowd is normal or anomalous by analysis the crowd behaviour. This is done so that on find any abnormal crowd behaviour we can inform the concerned authorities that can take the right steps at the right time to avoid critical situations. If we detect the congestion happening before the right time, then according to that, we can make the right arrangement so that this congestion does not turn into a problem.

Through this approach, we can find out that this crowd is normal traffic, any violent mob or any type of ordinary mob.

4.3 OBJECTIVE

The objectives of this research have been stated below.

1. Preprocessing the image to extract low level features using OpenCV. This includes following tasks:
    - Getting the image spectrum
    - Based on the frameset, cropping it to create patches
    - Finally extracting the low-level features for each of these patches
2. Calculating the ground truth density of the image
    - Using the density maps
    - Mapping the density maps linearly with the extracted features.
3. Estimating the crowd count and crowd density of the image.
    - Using the Multi Column Convolutional Neural Network.

4.4 MACHINE LEARNING MODEL USED

4.4.1 MULTI COLUMN CONVOLUTIONAL NEURAL NETWORK (MCNN)

Multi Column Neural Networks are special CNNs that use more than one CNN in the same architecture. Each CNN is used to take input images of different resolutions. Final output is calculated as a linear combination of the outputs of the multiple CNNs used. The best part of this CNN is that it does an automatic feature extraction.

Figure 4.1, shows an example architecture of MCNN.

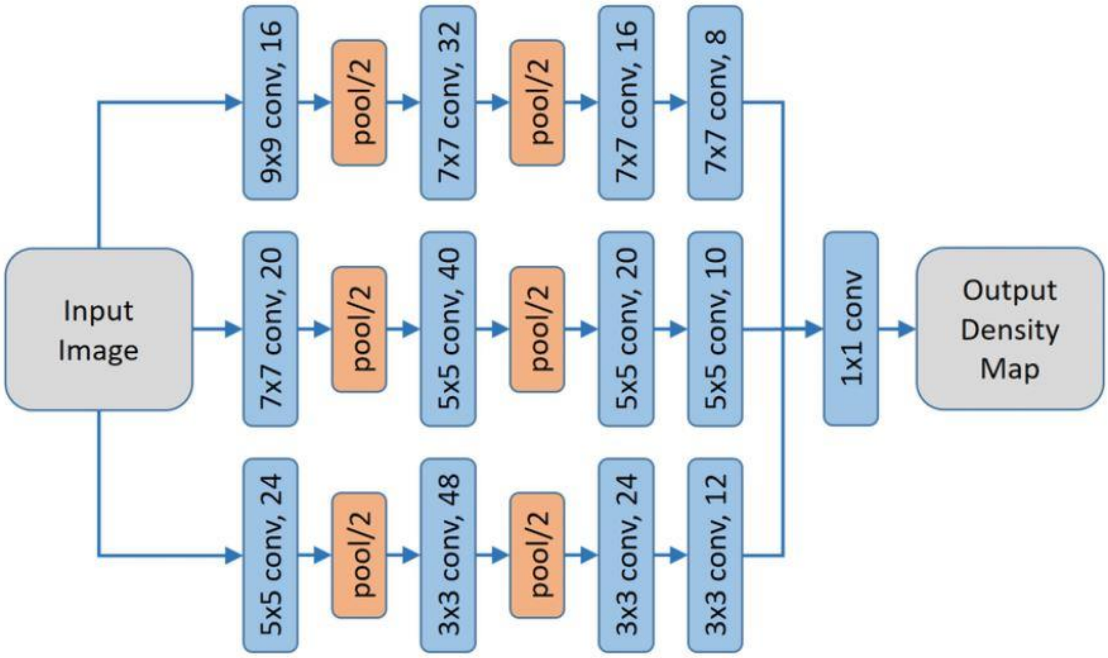


Figure 4.1: Sample Architecture of a MCNN

4.4.2 WORKING OF MCNN

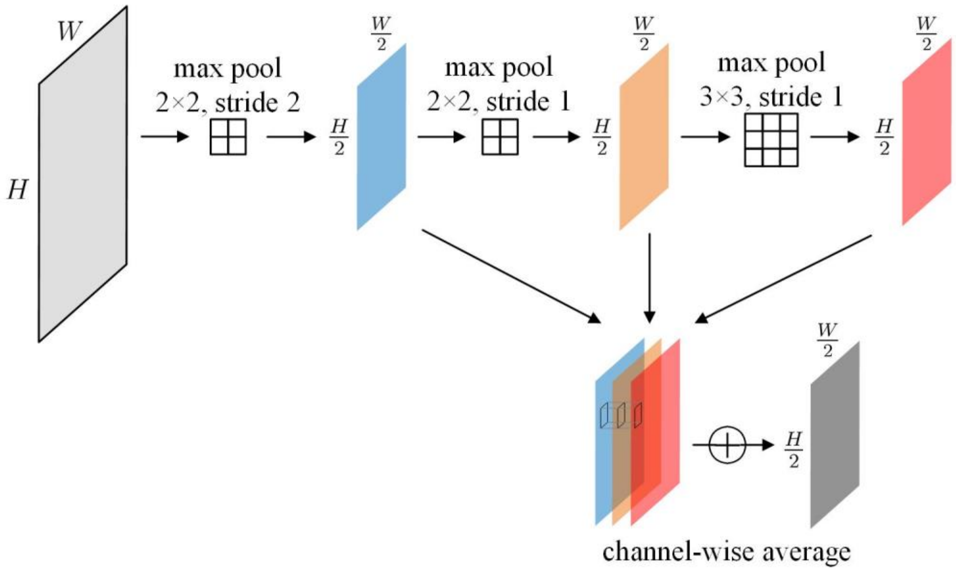


Figure 4.2: Working of a single column of MCNN

Figure 4.2, shows the working of a single column of MCNN.

Each Column is actually a CNN that has a combination of two layers.

❖ **Convolution Layer**: Convolution layer is basically a filter that is applied to an input image to get a filtered image. If we apply the same filter again and again, we get a mapping of various filtered images of the same input that indicate the strength of the image features.

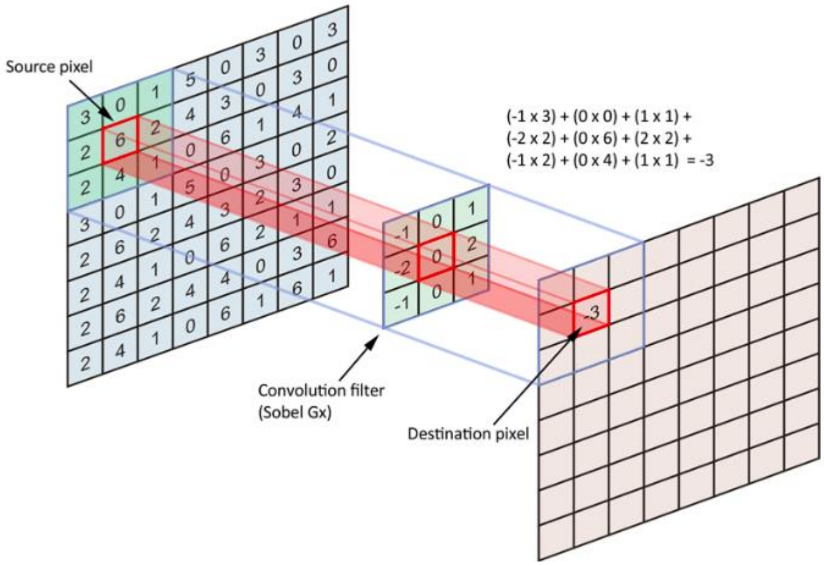


Figure 4.3: Working of a Convolution layer

These filters are learned by the model during the training

❖ **Max Pooling Layer**: Max pooling layer performs max pooling i.e., from each region on the feature map covered by the filter the maximum element is extracted. Hence, at the end of the max pooling layer we get a reduced feature map that contains the most prominent features of the input image. The reason that we use max pooling in CNNs is that it down samples the feature detection in feature maps. This helps in two ways. One, that we get the most important features of the image and second, we reduce the processing complexity.

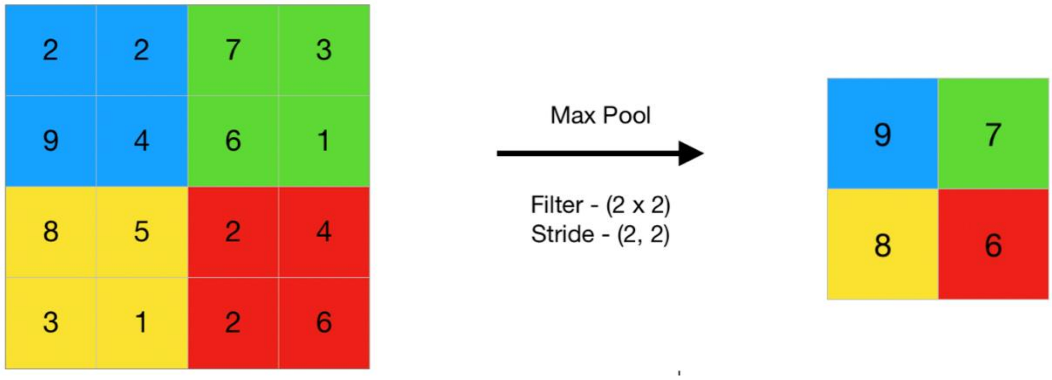


Figure 4.4: Working of a Max pooling layer

In this example max pooling we have taken a filter of size 2×2 and the stride taken the filter is 2,2.

4.5 METHODOLOGY

1. Our proposed methodology uses Multicolumn Convolutional Neural Network (MCNN).
2. According to this approach, first we will create a data set of traffic related, mob lynching related, market congestion images.
3. After this we will filter the data set, then prepare a training and testing data set.
4. Then in the next step we, we will take images of the crowd and process them to extract their features, write their dimensions, shapes, gray scaling to calculate the ground truth density.
5. After this, we will input the images in the MCNN's Algorithm and find out the estimated density of the images.
6. We compare the ground truth count and density with the estimated account and density to finalize the weights of the model.
7. Once the model is trained, we can use it calculate the density ratio of the

images based on which we predict whether the crowd in the image is normal or abnormal.

4.5.1 APPROACH FLOW DIAGRAM

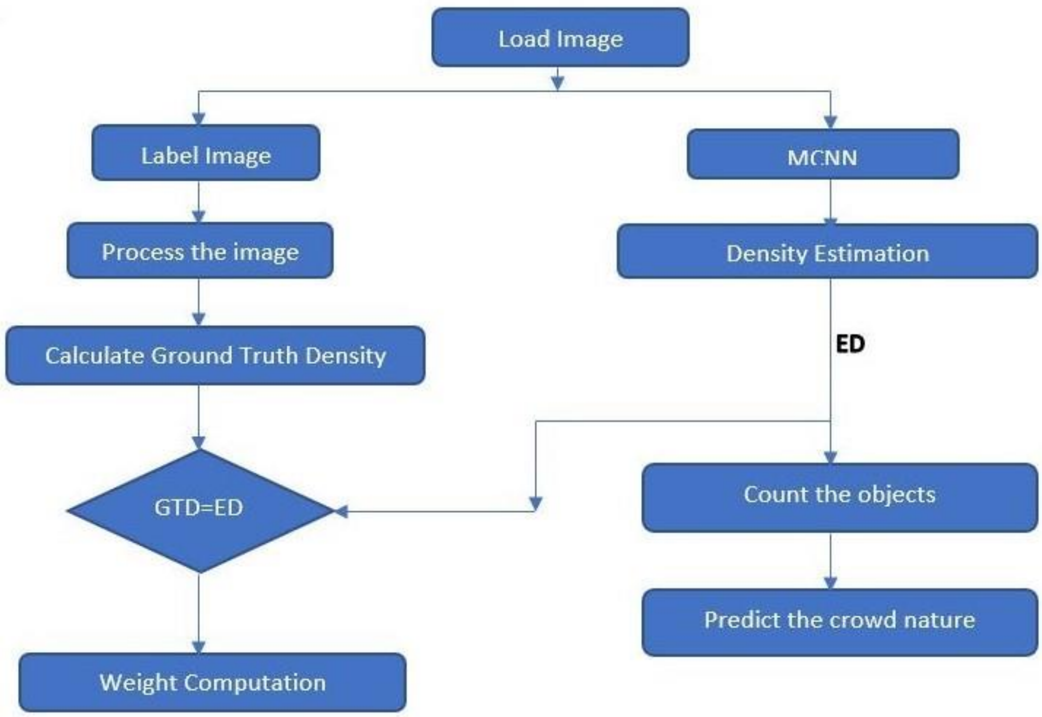


Figure 4.5: Flow Diagram representing the model approach

The process starts with the input image. We perform the pre-processing of the input image using OpenCV technology developed for this purpose. This very first step will lead to the extraction of low-level features from the input image. To get the low-level features, an image spectrum is obtained. Then based on the frame set we apply filter on the image spectrum to crop it into patches. Now, we extract the low-level features of each such patch obtained. This ends the process of preprocessing of the image.

Density maps of the objects are formed from these low-level features and as a final step the density maps are linearly mapped with the extracted features to obtain the Ground Truth Density of the image.

Once we have arrived at the ground truth density, we now input the image into our MCNN algorithm. The MCNN algorithm is now used to predict the crowd density i.e., provide the Estimated Crowd Density.

The MCNN outputs a density maps of the crowd. The aggregation of these density maps provide us with the total crowd count. Now is the turn to train our MCNN model. For this we check the Estimated Density (ED) against the Ground Truth Density (GTD) calculated earlier.

If the difference between the Ground Truth Density (GTD) and the Estimated Density (ED) is within the agreed threshold, then we can save the model by finalizing the weights. But in case the difference is beyond the threshold the model is trained again on new weights.

At last, when our model is trained, we check the density ratio to segregate normal from abnormal crowds. Table 1 is used to predict the normal and abnormal behaviour of the crowd.

TABLE 4.1

CROWD AREA AND THEIR ASSUMED NORMAL CROWD DENSITY

Crowd Area	Normal Crowd Density
100 m^2	70
200 m^2	120
300 m^2	320
400 m^2	380
500 m^2	400

4.5.2 ALGORITHM USED:

Step 1: Input the image for preprocessing

Step 2: Get the image spectrum

Step 3: Crop patches of the image and for each obtained patch extract low level features.

Step 4: Generate the density maps and map them linearly with the features extracted in the previous step. This results in the ground truth crowd density (GTD).

Step 5: Use MCNN to calculate the estimated crowd density (ED)

Step 6: Train the model by comparing the estimated density with the ground truth density to check the model accuracy.

Step 7: Predict the density of an input image using the trained MCNN model.

Step 8: label the crowd as normal or abnormal based on the estimated density ratio

4.5.3 MODEL ARCHITECTURE

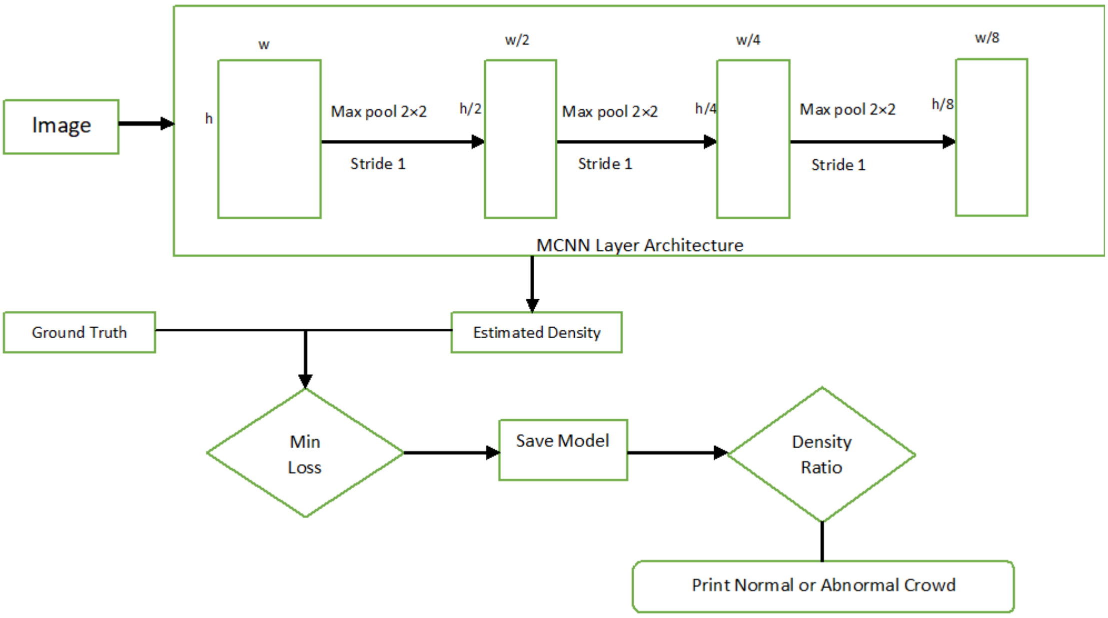


Figure 4.6: Detailed architecture of the MCNN model used

Figure 4.6 shows the complete architecture of the MCNN model used in our approach. There are three parallel CNNs, also called columns. For each of these columns the size of the receptive fields are different. The layers used in these columns are same i.e., the convolution layer and the max pooling layer. Each column has 3 sets of these layers with varying filter sizes and numbers. Hence, there are a total of three max pooling layers that are responsible for creating an output feature map that is down sampled from the original input feature map to 1/8th.

Size of max pooling filter is 2×2 and we have used the Rectified linear unit (ReLU) as the activation function. To map the output feature maps obtained from all CNNs to a final density map we use a filter of size 1 ×1. Euclidean distance method is adopted to calculate the difference between the estimated density map and ground truth.

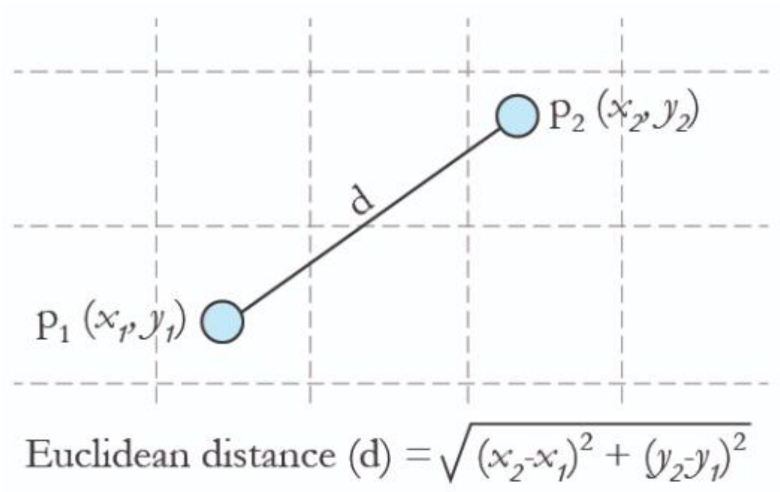


Figure 4.7: Euclidean Distance Calculation

4.6 PROGRAMMING TECHNOLOGY

The Programming language used to implement the proposed approach is Python. Python is the most used language for implementing machine learning models. For developing the code, we have used a Notebook environment called Jupyter Notebook.

4.6.1 LIBRARIES USED

Python uses multiple libraries and packages for various functionalities. We have used Keras for building and training the MCNN model. Apart from this, other libraries used are numpy and matplotlib.

<u>CHAPTER – 5</u>

RESULT ANALYSIS

AND

DISCUSSION

5.1 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 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 MCNN 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 CNN and LSTM.

5.1.1 EVALUATION METRICS

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

5.1.1.1 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.

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

Figure 5.1: Formula of Mean Absolute Error

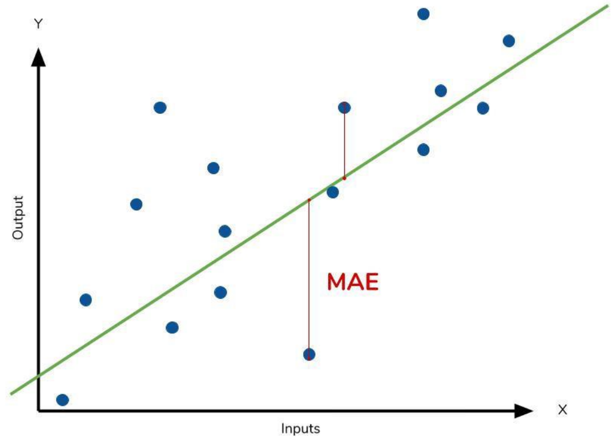


Figure 5.2: Diagram representing Mean Absolute Error

5.1.1.2 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.

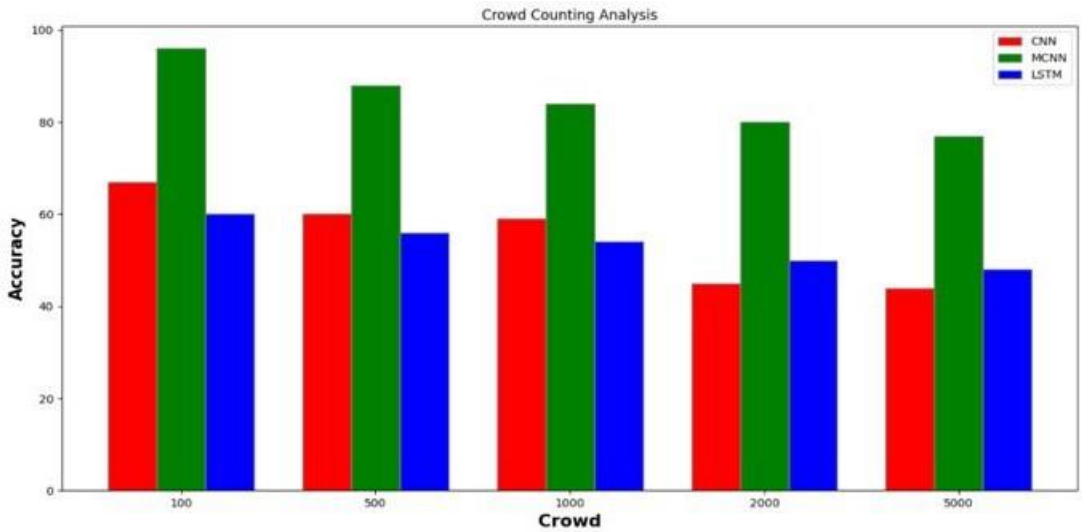
$$MSE = \frac{1}{N}\sum_{1}^{N}(z_i - \hat{z}_i)^2$$

Figure 5.3: Formula of Mean Absolute Error

N = Number of images

Zi = Number of people in ith image

ẑI = Estimated number of people in the ith image



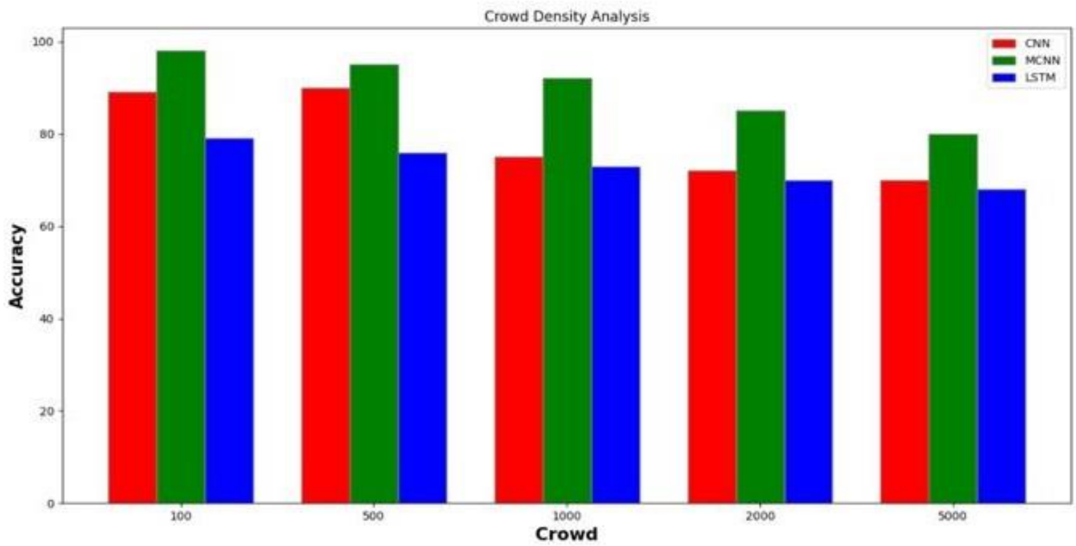


Figure 5.4: Graphs depicting the comparison of other methods and MCNN method

5.1.2 DATASET

In our analysis for training and testing the proposed model of Multi Column Convolutional Neural Network. We are using the UCF-QNRF dataset for the training and testing of our proposed methodology. This dataset has a total of 1535 annotated images of densely crowded scenarios. Out of these images 1201 images are for the purpose of training and 334 images are used for testing.

5.1.3 RESULTS

On applying our proposed methodology on the UCF-QNRF dataset we have observed that our method outperforms the other methodologies used for comparison in our experiment i.e., CNN and LSTM. Following figure shows the graph representing the accuracy of these methods as compared to our proposed method (MCNN). This gives the evidence that MCNN predicts with a better accuracy in crowd counting and crowd density analysis as compared to CNN and LSTM.

TABLE 5.1

COMPARATIVE STUDY OF DIFFERENT METHODS AND OUR APPROACH FOR CROWD COUNT ACCURACY

Crowd	100	500	1000	2000	5000
CNN	67	60	59	45	44
MCNN	**96**	**88**	**84**	**80**	**77**
LSTM	60	56	54	50	48

TABLE 5.2

COMPARATIVE STUDY OF DIFFERENT METHODS AND OUR APPROACH FOR CROWD DENSITY ACCURACY

Crowd	100	500	1000	2000	5000
CNN	89	90	75	72	70
MCNN	**98**	**95**	**92**	**85**	**80**
LSTM	79	76	73	70	68

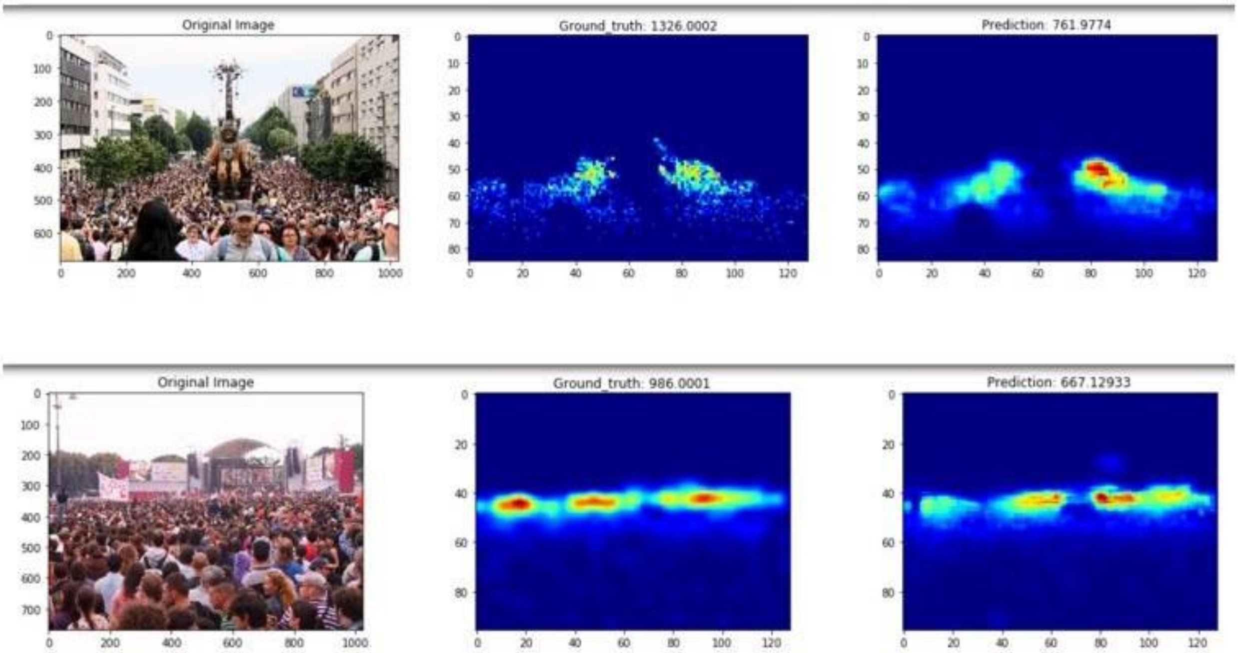


Figure 5.5: Examples of test images and their ground truth and predicted estimated density

CHAPTER – 6

CONCLUSION

AND

FUTURE SCOPE

6.1 CONCLUSION

This dissertation aims to provide a completely new methodology for performing image analysis that has a crowd scene and predict the behaviour of the crowd in an image by generating a Multi Column Convolutional Neural Network (MCNN) consisting of three parallel CNNs. Each column is a complete CNN that is using three sets of convolution layer and max pooling layers. Our model is able to down sample the output image by $1/8^{th}$ of the original input image. This gives us more accuracy. We have evaluated the performance of our Multi Column Convolutional Neural Network (MCNN) model along with two more well-known state of the art methods i.e., Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) on the UCF-QNRF dataset.

Our dataset is made of 1535 annotated images of that contain highly congested scenes. We found that our model (MCNN) performs better than the other two methods when evaluated on the same dataset and demonstrates higher accuracy.

6.2 FUTURE WORK

There is still a lot of room available for the improvement in the strategies and algorithms implemented for the purpose of crowd behaviour 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 behaviour and the problem is only made worse by the absence of data regarding normal and abnormal crowd behaviour. This leads the path to continued research and strive to find the most optimal algorithms possible to detect and predict crowd behaviour.

REFERENCES

[1] C. Direkoglu, "Abnormal Crowd Behaviour Detection Using Motion Information Images and Convolutional Neural Networks," in IEEE Access, vol. 8, pp. 80408-80416, 2020, doi: 10.1109/ACCESS.2020.2990355.

[2] G. Tripathi, K. Singh and D. K. Vishwakarma, "Crowd Emotion Analysis Using 2D ConvNets," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 969-974, doi: 10.1109/ICSSIT48917.2020.9214208.

[3] M. Gao et al., "Violent crowd behavior detection using deep learning and compressive sensing," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 5329-5333, doi: 10.1109/CCDC.2019.8832598.

[4] Xie, S., Zhang, X. & Cai, J. Video crowd detection and abnormal behavior model detection based on machine learning method. Neural Comput & Applic 31, 175–184 (2019).

[5] X. Ma, S. Du and Y. Liu, "A Lightweight Neural Network For Crowd Analysis Of Images With Congested Scenes," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 979-983, doi: 10.1109/ICIP.2019.8803062.

[6] C. Santhini and V. Gomathi, "Crowd Scene Analysis Using Deep Learning Network," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8550851.

[7] I. Usman, "Anomalous Crowd Behavior Detection in Time Varying Motion Sequences," 2019 4th World Conference on Complex Systems (WCCS), Ouarzazate, Morocco, 2019, pp. 1-5, doi: 10.1109/ICoCS.2019.8930795.

[8] M. Yang et al., "Cluster-Based Crowd Movement Behavior Detection," 2018

Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 2018, pp. 1-8, doi: 10.1109/DICTA.2018.8615809

[9] A. CHIBLOUN, S. EL FKIHI, H. MLIKI, M. HAMMAMI and R. OULAD HAJ THAMI, "Abnormal Crowd Behavior Detection Using Speed and Direction Models," 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC), Rabat, Morocco, 2018, pp. 197-202, doi: 10.1109/ISIVC.2018.8709192.

[10] S. Lamba and N. Nain, "A Large Scale Crowd Density Classification Using Spatio-Temporal Local Binary Pattern," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 296-302, doi: 10.1109/SITIS.2017.57.

[11] Q. Wang, M. Chen, F. Nie and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 1, pp. 46-58, 1 Jan. 2020, doi: 10.1109/TPAMI.2018.2875002.

[12] L. Lazaridis, A. Dimou and P. Daras, "Abnormal Behavior Detection in Crowded Scenes Using Density Heatmaps and Optical Flow," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 2060-2064, doi: 10.23919/EUSIPCO.2018.8553620.

[13] Z. Zou, X. Su, X. Qu and P. Zhou, "DA-Net: Learning the Fine-Grained Density Distribution With Deformation Aggregation Network," in IEEE Access, vol. 6, pp. 60745-60756, 2018, doi: 10.1109/ACCESS.2018.2875495.

[14] M. A. Kizrak and B. Bolat, "Deep and Wide Convolutional Neural Network Model for Highly Dense Crowd," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 2019, pp. 1-6, doi: 10.1109/ASYU48272.2019.8946395.

[15] G. songchenchen and E. Bourennane, "A Method Based on Multi-source Feature Detection for Counting People in Crowded Areas," 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 2019, pp. 1021-1025, doi: 10.1109/SIPROCESS.2019.8868691.

[16] Siraj, Muhammad. “Machine and Deep Learning for Crowd Analytics.” ArXiv abs/1909.04150 (2019): n. pag.

[17] X. Li, M. Chen and Q. Wang, "Quantifying and Detecting Collective Motion in Crowd Scenes," in IEEE Transactions on Image Processing, vol. 29, pp. 5571-5583, 2020, doi: 10.1109/TIP.2020.2985284.

[18] S. S. A. Abbas, P. O. Jayaprakash, M. Anitha and X. V. Jaini, "Crowd detection and management using cascade classifier on ARMv8 and OpenCV-Python," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-6, doi: 10.1109/ICIIECS.2017.8275988.

[19] Li, Yan & Liu, Hong & Zheng, Xiangwei & Han, Yanbin & Li, Liang. (2019). A Top-Bottom Clustering Algorithm Based on Crowd Trajectories for Small Group Classification. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2902310.

[20] M. Abdou and A. Erradi, "Crowd Counting: A Survey of Machine Learning Approaches," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 48-54, doi: 10.1109/ICIoT48696.2020.9089594.

[21] A. N. Moustafa, M. E. Hussein and W. Gomaa, "Gate and Common Pathway Detection in Crowd Scenes Using Motion Units and Meta-Tracking," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, NSW, 2017, pp. 1-8, doi: 10.1109/DICTA.2017.8227438.

[22] N. Kumaran and U. S. Reddy, "Object detection and tracking in crowd environment — A review," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 777-782, doi: 10.1109/ICICI.2017.8365242.

[23] A. K. Jhapate, S. Malviya and M. Jhapate, "Unusual Crowd Activity Detection using OpenCV and Motion Influence Map," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170704.

[24] M. Mancas et al., "People Groups Analysis for AR Applications," 2018 International Conference on 3D Immersion (IC3D), Brussels, Belgium, 2018, pp. 1-8, doi: 10.1109/IC3D.2018.8657892.

[25] N. Nain and S. Lamba, "Oriented Tracklets Approach for Anomalous Scene Detection in High Density Crowd," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 435-441, doi: 10.1109/SITIS.2018.00073.

[26] A. Bera, S. Kim and D. Manocha, "Interactive Crowd-Behavior Learning for Surveillance and Training," in IEEE Computer Graphics and Applications, vol. 36, no. 6, pp. 37-45, Nov.-Dec. 2016, doi: 10.1109/MCG.2016.113.

[27] P. V. V. Kishore, R. Rahul, K. Sravya and A. S. C. S. Sastry, "Crowd Density Analysis and tracking," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, 2015, pp. 1209-1213, doi: 10.1109/ICACCI.2015.7275777.

[28] T. Senst, V. Eiselein, I. Keller and T. Sikora, "Crowd analysis in non-static cameras using feature tracking and multi-person density," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 6041-6045, doi: 10.1109/ICIP.2014.7026219.

[29] M. R. Rejitha and S. N. George, "An Unsupervised Abnormal Crowd Behavior Detection Technique using Farneback Algorithm," 2019 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2019, pp. 1-5, doi: 10.1109/CONECCT47791.2019.9012845.

[30] H. Huang, H. Guo, Z. Ding, Y. Chen and X. Wu, "3D virtual modeling for crowd analysis," 2015 IEEE International Conference on Information and Automation, Lijiang, 2015, pp. 2949-2954, doi: 10.1109/ICInfA.2015.7279793.

[31] Singh, J., & Khan, M. Z. (2019). Detection of Fake Profile in Social Media. Journal of Emerging Technologies and Innovative Research, 6(6), 38–42. Retrieved from http://www.jetir.org/papers/JETIR1906M04.pdf

[32] Khan, M. Z., & Shoaib, M. (2019). Healthcare Analytics in the Modern Era: A Survey. International Journal of Research in Advent Technology,7(S), March 2019, 132-134. Retrieved from http://www.ijrat.org/downloads/Conference_Proceedings/Special_Issue_March-19/Paper ID-156.pdf

[33] M. Z. Khan, M. S. Kidwai, F. Ahamad and M. U. Khan, "Hadoop based EMH framework: A Big Data approach," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 1068-1070, doi: 10.1109/ICACITE51222.2021.9404710

[34] M. Suhaib Kidwai and M. Zunnun Khan, "A new perspective of detecting and classifying neurological disorders through recurrence and machine learning classifiers," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 200-206, doi: 10.1109/ICACITE51222.2021.9404645.

PLAGIARISM CHECK REPORT

Plagiarism Checker X Originality Report

Similarity Found: 8%

Date: Saturday, August 07, 2021
Statistics: 651 words Plagiarized / 8120 Total words
Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

PUBLICATIONS FROM THIS WORK

1) **"Machine Learning Based Crowd Behaviour Analysis and Prediction"** has been published in the Journal of Network Security Computer Networks Volume 7 Issue 1 Year 2021. The journal is indexed in Google Scholar and Index Copernicus.

2) **"Machine Learning Based Crowd Behaviour Analysis and Prediction"** has been published in the UGC Care group I journal 'Dogo Rangsang Research Journal' with ISSN 2347-7180 at Volume-11 Issue-8, August 2021.

PUBLICATIONS

Literature Review Paper Publication Certificate


MAT JOURNALS

PUBLICATION CERTIFICATE

This is to certify that the manuscript entitled Machine Learning Based Crowd Behaviour Analysis and Prediction Submitted by Neha Srivastava has been published in Journal of Network Security Computer Networks (e-ISSN: 2581-639X) Volume 7 Issue 1 Year 2021

MAT JOURNALS GHAZIABAD U.P. INDIRAPURAM

Arvind Sharma

Arvind Sharma
Managing Director, MAT Journals
(A Unit of ARV InfomediaPvt. Ltd.)

INDEXING: DOI||Cosmos Impact Factor||
J-Gate||Figshare||Google Scholar||Cite
Factor||Zenodo||Scientific Indexing Services||Root Indexing

Machine Learning Based Crowd Behaviour Analysis and Prediction

Neha Srivastava[1]*, Mohammad Zunnun Khan[2]
[1]PG Student, Department of Computer Science and Engineering, Integral University, Lucknow, Uttar Pradesh, India
[2]Assistant Professor, Department of Computer Science and Engineering, Integral University, Lucknow, Uttar Pradesh, India

*Corresponding Author: nigamn.nigam@gmail.com

ABSTRACT
Crowd behaviour has been widely known to have the ability to forecast the events a crowd could create. Crowd management can become extremely efficient if situations such as riots, mob lynching, traffic jams, accidents, stampede, etc. could be predicted beforehand. To this end many researchers have made their contributions in the past and there is still immense work being carried out currently. All the researches worked with different algorithms and techniques to analyze images or videos of crowd scenes for counting the number of people in the crowd, predicting the behaviour of the crowd and classifying an image or video as normal or abnormal crowd event. This paper, hence, is directed towards underlining the some of the major researches in this field, the approaches and algorithms adopted by them and their comparisons. Overall, this paper reviews the past researches and presents a summary of the techniques and strategies employed. At the end of this paper is the future scope of work possible in the field of crowd behaviour analysis, prediction and crowd counting.

Keywords-- Anomalous crowd prediction, Crowd behaviour analysis, Deep learning

INTRODUCTION

Crowd or mob means a group or collection of large number of people. Living in a society makes us also an integral part of crowd and it affects our daily activities such as going to markets, banks, multiplexes, offices, parks, etc. so much so that it has necessitated the understanding and analysis of crowds [1]. This is important as we can find immense usage of the knowledge and insights obtained from this analysis in real world applications. Crowd analysis can be useful in predicting situations such as riots and mob lynching which in turn could help the corresponding authorities to take preventive actions. Another scenario where crowds pose a problem is traffic as it could lead to traffic jams and even accidents.

Many recent studies in this field have brought forward some very interesting methods involving CNNs and optical flow vectors [1-3]. The input to the analysis of crowd could be either an image or a video. Videos are sliced into a sequence of consecutive frames and then processed [4]. Many factors affect the analysis such as occlusion, scale, size, perspective, boundary restrictions, etc. [5, 6]. With the advancement of machine learning and deep neural networks, it is natural to be inclined towards this approach to seek a solution for the problem of crowd behaviour analysis and prediction.
It is not an easy feat to accomplish as it is complicated to understand and prepare a model for crowd behaviour. Crowd analysis can be performed using either of the two strategies [4].

i. Object based approach: Here the crowd is considered to be a collection of individual persons and the motion and context of each of these individuals is tracked to estimate the behaviour of the crowd.
ii. Holistic approach: Here the crowd is used as a single entity and we track the motion and behaviour of the entire crowd as a whole.

Object based approach requires tracking of individuals in a crowd which is extremely difficult when the scene is densely crowded such as at sports stadiums [6]. However, it is a feasible approach for lightly crowded scenes and many techniques such as optical flow vectors and speed and direction models have been devised [7-9]. The biggest issue in crowd behaviour analysis is ascertaining the normal behaviour patterns in crowd and the way a comparative is drawn between them and current behaviour patterns [10]. This brings the analysis of crowd behaviour under the category of unsupervised learning, a subfield of machine learning.

MACHINE LEARNING

It is a branch 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 own 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 (Fig. 1).

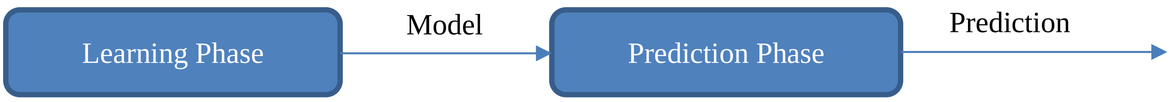


Figure 1: *Phases of machine learning.*

Conventionally there are three approaches of machine learning. These are:

Supervised Learning

This approach requires those sample data instances for training that are already labelled with the appropriate classes. Examples of supervised learning models are Support Vector Machine (SVM), Decision Tree, Random Forest, Bayesian classifier, etc.

Unsupervised Learning

This approach uses those sample data for training that are not pre-labelled with classes. Examples of unsupervised learning models are Hierarchical clustering, K means, DBSCAN.

Reinforcement Learning

This is a special approach where the machine has to interact with its environment while performing a task one step at a time. The machine gets rewarded if the step is correct else no reward.

DEEP LEARNING

It is a branch of machine learning that is inspired by our biological brains. It uses artificial neural networks. There are multiple layers in these networks due to which these are called deep neural network. Lower layers are responsible of extracting low level features while higher layers discover high level and detailed features. Most recent of them are the Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

REVIEW OF ALGORITHM USED IN CROWD BEHAVIOUR ANALYSIS:

Support Vector Machine (SVM)

SVM has been widely used in many of the researches in crowd emotion analysis as the main classifier with a few methods implementing variations of SVM too [5, 11, 12]. 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 on to 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 but 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:

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

Where **W** is a weight vector, such as,

W = {*w*1, *w*2, …, *wn*};

N is the number of attributes;

b is a scalar (also called as bias)

X is the data point

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

***w*0 + *w*1*x*1 + *w*2*x*2 = 0**

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 (Fig. 2).

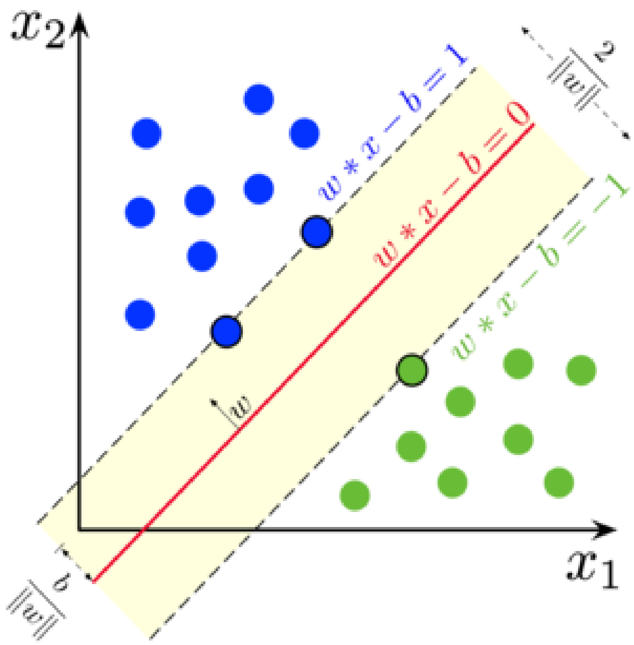


Figure 2: *Classification through SVM in 2-dimensional data space.*
(Source: https://en.wikipedia.org/wiki/Support-vector_machine)

The merits and demerits of SVM are discussed as follows [12]:
Disadvantage of SVM: Sometimes training the fastest SVMs can take huge lengths of time.
Advantage of SVM:

- Accuracy of SVM is very high, mainly because they can create complex nonlinear decision hyperplanes.
- Apart from this, in comparison to other methods SVMs are very less likely to suffer from overfitting.

Convolution Neural Network (CNN)

CNN are a kind Artificial neural network that have been specifically designed for analysing images encouraged by the working of human vision [13]. This makes CNN the most apt 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 [1-3, 5, 9, 13, 14]. The layers in the CNNs are 3 dimensional and unlike the regular neural networks the neurons in one layer are not necessarily connected to all the neurons of the next layer. CNNs comprise of two components: (i) the feature extraction part (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 of the learning in the model. The classification part is where we calculate the likeliness of the image being a part of a particular class. This is achieved by assigning probabilities for the same.

Convolution means combining two functions to give a new third function. It is performed in CNN by applying kernels or filters repeatedly upon 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 [14]. The

entire operation is in 3 dimensions due to colours in an image. Each filter produces a separate feature map. All the features maps constitute the output of the convolution layer of CNN. This output is then fed into an activation function such as ReLU activation function. Padding is done to stop the feature map from shrinking.
A CNN has four important hyperparameters:

- The Kernel Size.
- The Filter Count.
- Stride (how big are the steps of the filter).
- Padding.

After convolution layer a pooling layer is added to reduce the dimensionality. At last, the classification layer is included that is set of fully connected layers. But these layers can accept input only as 1 dimensional data.

LITERATURE SUMMARY

In this paper the researchers have proposed a novel strategy of using Motion Information Images (MII) along with Convolution Neural Networks (CNN) to perform abnormal crowd event detection [1]. The paper emphasizes on detection of behaviours such as panic and escape that may be a result of natural disasters and violent events. The MIIs proposed here represent the motion in the crowd visually. First of all, magnitudes of optical flow vectors and angle difference between them in consecutive video frames are used to formulate these MIIs for each video frame. These MIIs are responsible for distinguishing normal crowd behaviour from abnormal one. Once the MIIs are ready they are fed to the CNN to train them to classify a crowd video as normal or abnormal crowd. Results of the evaluation of UMN and PETS2009 datasets prove that the proposed approach was effective and accurate.

The researchers in this paper have come up with a novel strategy of detecting crowd emotions and behaviour by implementing 2D Convolution Neural Networks (ConvNets) [2]. This network's goal is to classify the general behaviour of the crowd. For this experiment the researchers have curated a dataset of images that mainly displays six kinds of emotions such as anger, sadness, excitement, happiness, scaredness and neutral. The network used in this paper is a variation of Xception network and classifies the crowd emotions into six categories as givens. The model is built by stacking depth wise convolutional to reduce computational cost. It also involves residual connections to discover a shortcut in the sequential network and solve the issue of vanishing gradients. Results of the experiment prove that the proposed architecture is suitable for use on single images, bulk images and videos. It can be deployed on the local and web servers for real time emotion classification.

This paper is focussed towards the detection of violent crowd behaviour using a combined approach of deep learning networks and compressed sensing [8]. For this purpose, the researchers have designed a new Hybrid random Matrix (HRM). The HRM is proved to fulfil the Restricted Isometry property (RIP). The research proceeds in four stages. First step is the extraction of high dimensional violent flows features through Violent Flows Algorithm (ViF). The second step performs the dimensionality reduction by transforming the extracted high dimensional features into low dimensional features using HRM and compressed sensing. In the third step deep, neural networks are trained to learn more meaningful features. Lastly, Support Vector Machine are employed to perform the classification. For the testing and evaluation part a five-fold cross validation approach is used. When compared with other approaches the proposed strategy called Vif-HRM-NET has performed more effectively.

The purpose of this paper is to learn various methods of data mining and improve the accuracy of detection of abnormal behaviour in crowd videos by machine [10]. This paper proposes a new IDS video crowd and its behaviour anomaly detection model which is based on machine learning. It conducts experiments by using Unix user's shell command data. The experiment involves three main parts. These are optical flow extraction, optical flow to geospatial mapping, and spatial analysis of optical flow. The developed prototype of the system displays the optical flow field at a certain time. In this paper the author has performed scattered point interpolation and contour generation experiments. For the purpose of scatter point interpolation IDW method is used.

As a result, this model works much better than the fixed length command sequence model.

The aim of this paper is to introduce a lightweight and low complexity end-to-end network for analysis of crowd with congested scenes [5]. This is referred to as a lightweight network as it uses only 0.86 M parameters which is very less as compared to other neural networks. The author has divided the network into three parts. First is the multi scale feature extraction where the researcher has used four scale aware modules using cascaded convolutions with smaller spatial filters to extract the required features. This brings down the size of the input image to one fourth due to the use Max Pooling at two layers. Second is the density map estimation which is done by the regression of the extracted features. But it is a rough estimation as it is affected by occlusion. Hence, the third part is the density map correction using an autoencoder. As a result of the experiment, it was found that the density map generated are very accurate and can also perform crowd counting accurately as well.

This paper aims to find the crowd density map and the number of people arriving in a group through the use of Convolution Neural Network (CNN) and Deep learning models [3]. The author has stated the lack of training samples, drastic blockages, disorder in crowd scenes, and changes of perspective as the reason to explore deep learning for crowd analysis. In this paper author has proposed the use of VGG16 convolution neural network architecture. This strategy requires the videos to be sliced into frames and arranged into sequential manner. This paper performs crowd scene analysis using a convolutional crowd dataset comprising of 100 videos collected from 800 crowd scenes and extracting a set of 94 attributes.

In this paper the author has introduced an automated anomalous behaviour detection model for crowd video sequences [4]. The technique presented in the paper has used the gradient based approach for modelling motion of each object by adopting the concept of cuboids (spatio-temporal volumes) to differentiate between various activities within a time varying video sequence. Hence, it generates a 3D space that represents time dimension as well. First the author uses statistical models to extract the feature set. Then a classifier based on GP (Genetic Programming) is developed to perform a binary classification between normal and anomalous behaviour in crowd video. The classifier is generated by calculating the mean and variance of cuboids. This classifier has better performance when it comes to finding the hidden dependencies in the solution space. Results show that the method defined in this paper outperforms the conventional and existing techniques in terms of accuracy of classification as well as time involved.

This paper proposes a new real time architecture for crowd event detection [6]. The basis of this architecture is found in object detection that is in association with fixed-width clustering. For this purpose, the crowd events in this paper have been divided into two main categories. One is either walking or running and the other is either splitting or merging. The latter is further classified as local or global. The author has first used YOLOv2 (You Only Look Once) for detecting objects in each video frame. The detected objects are then plotted in a two-dimensional space as feature vectors. Then a novel fixed width clustering algorithm is applied to find clusters or groups of objects that are close to each other. Following this the behaviour of the movements of these clusters as depicted by their centroids is used to perform event classification or crowd change detection. For evaluation this paper uses six video sequences form PETS2009 and the results have shown the accuracy of the advocated strategy to be between 80%-95%.

In this paper the researcher aims to detect unusual crowd behaviour in a video sequence by making use of probability models of speeds and directions [7]. First of all, each video frame is assigned two kinds of probability densities. One represents the speed of the crowd and the other represents the direction. The author has stated the use of Expectation - Maximization algorithm (EM) to convert the optical flow vectors into probability models of speed along with that of direction for each video frame. In situations where there is any change in the speed or the direction between any two consecutive frames then it will get reflected in these densities. This research paper defines an anomaly as a change that will in turn change the speed or the direction. Hence, abnormal behaviour can be detected by determining the distance between the probability models of current model with those of the previous ones. If the distance exceeds a threshold value it indicates an anomalous event in the crowd.

This paper aims to establish a new strategy for performing the classification of large-scale crowd density by applying dynamic texture analysis and also estimating the crowd flow direction [11]. The research is comprised of three parts. In the first part, the locations of interest points are observed using Hessian matrix. The second part involves extracting local spatio-temporal features of each interest point detected before. These features are exhibited by RIST-LBP code. This generates the dynamic texture of moving crowd. The last part is all about training a multiclass SVM on these features for classifying the crowd densities at various levels. The research is continued by adding a tracking step that tracks the detected interest points across multiple video frames using a KLT tracker algorithm. This gives an estimation of the crowd flow. This research is performed on three datasets namely PETS, UCF, CUHK. As a result, it was observed that the proposed strategy has lower computational complexities and high efficiencies as opposed to the conventional pixel-based techniques.

A Multiview-based Parameter Free framework (MPF) and Structural context descriptor are introduced in this research paper for the purpose of the identification of coherent groups in crowd videos [15]. In this paper the authors bring out three new aspects of the research. Firstly, this paper states that the individuals in a crowd are closely tied with their neighbours that represent their structural property. This structural property is expressed by the suggested Structural Context (SC) descriptor for each point. Secondly, to incorporate the correlations of the feature points from the context view along with that of orientation, two versions of self-weighted multi view clustering method are formulated. Then lastly, to bring together the coherent local groups a tightness-based merging approach is developed. Results of the experiment on various datasets have proven the effectiveness of the proposed approach.

This paper aims to present an imaginative strategy to detect and classify abnormal crowd events specially those of panic and fight [9]. This new strategy emphasizes that to correctly analyse the motion in the crowd videos in both spatial and temporal context it is paramount that the analysis of the optical flow is combined with density heat maps as motion in the crowd may be initiated by things apart from people. This gives only the relevant regions in the crowd videos for motion analysis. Thus, it is called the two-stream architecture. Batches of ten frames are input into Long Short-Term Memory (LSTM) neural networks to learn the spatio-temporal patterns. Once trained on activities that are either normal or abnormal, the network is ready to classify the abnormal events in crowd videos.

At last, the paper also curates a new dataset using GTA V engine that contain videos with abnormal behaviour.

The researchers in this paper have designed a new neural network call DA-Net which is more robust and is much deeper [14]. The architecture of this network consists of two parts. One is the backbone that contains eight blocks and the other is multi-layer aggregation that contains deform blocks that gives the network capability of learning extra offsets to augment the spatial sampling locations. This makes it more robust to scale variations. Next the paper brings in a novel diamond architecture that pulls out high-level abstract information. Multi-layer extraction allows the network to immensely improve the maintenance of the spatial locations. Also, the paper introduces adaptive weights to replace direct aggregation to assess the reliability of multilevel outputs. As a result of the experiment, it was found that the approach advocated in the paper delivers high performance on various datasets.

The objective of this research is to get the measure of the number of people in images of dense crowd with the assistance of deep convolution neural networks [13]. The author has designed a new convolution neural networks that is made of two parallel modules that extract different features from the crowd images. From the first module lower-level features are pulled out that are used to formulate the density maps. Second module obtains higher level features that can identify parts of body such as head and upper body. Finally, these two modules are attached to a fully connected neural network to perform the crowd count estimation. This approach was tested on ShanghaiTech Part-A datasets with performance metrics as mean square error and mean absolute error. As a result, the approach tested successfully on comparing these metrics with other approaches.

This paper focuses to do crowd counting on still images of crowds taken from multiple sources using a drone camera [16]. This uses sources of information such as HOG, LBP, CANNY. For this purpose, a feature fusion method is proposed wherein multiple kinds of features of crowd images are drawn out and analysed such as that of texture features. This also detects the edges of crowd images. For each source separate count estimates and statistical measurements are calculated. After the estimates are collected, they are used to train a Support Vector Machine (SVM) classifier which learns the features and classifies highly dense crowd images using regression analysis. This approach was found to be easy and fast to implement. This experiment gives comparative results that are at par to the results of joint HOG.

This paper's goal is to discover the combined distribution of the pixels in an image and consider the temporal information and spatial information of the input and investigate the properties of the image pixels across multiple adjacent video frames to perform anomaly detection in crowd images or video frames [17]. To ensure this, it makes a general assumption that the distribution of the image pixels do not matter, i.e., it could be stationary or mobile in a given learning interval. Traditional machine learning methods assume that the anomalous situation is unknown, so the probability ratios cannot be properly computed. As a solution this paper the Bayesian model as a significant feature. The proposed method is implemented using MATLAB. And on average achieves 88.83% accuracy when applied on the videos from the same dataset (Table 1).

Table 1: *The summary of algorithms used in some past researches.*

S. No.	Title	Technique Used	Findings
1	Abnormal Crowd Behaviour Detection Using Motion Information Images and Convolutional Neural Networks	Motion Information Images and Convolutional Neural Networks	Results of the evaluation of UMN and PETS2009 datasets prove that the suggested approach was effective and accurate.
2	Crowd Emotion Analysis Using 2D ConvNets	2D convolutional neural network with residual connections	Proposed architecture is suitable for use on single images, bulk images as well as on videos. It can be implemented on the local and web servers for real time emotion classification.
3	Violent crowd behaviour detection using deep learning and compressive sensing	deep learning network, compressive sensing, Hybrid Random Matrix (HRM), Violent Flows (ViF) algorithm	As opposed to other approaches the proposed strategy called Vif-HRM-NET has performed more effectively.

4	Video Crowd Detection and Abnormal Behaviour Model Detection Based on Machine Learning Method	IDS anomaly detection model	Experimental results show that the detection performance of the new model is much better than that of Lane et al.
5	A Lightweight Neural Network for Crowd Analysis of Images with Congested Scenes	CNN, Cascaded convolutions, Max pooling	It was found that the density map generated are very accurate and can also perform crowd counting accurately as well.
6	Crowd Scene Analysis Using Deep Learning Network	CNN, Python OpenCV	This paper performs crowd scene analysis using a convolutional crowd dataset comprising of 100 videos collected from 800 crowd scenes and extracting a set of 94 attributes.
7	Anomalous Crowd Behaviour Detection in Time Varying Motion Sequences	Statistical modelling, Genetic Programming	Results profess that the suggested strategy outperforms the conventional and existing techniques in accuracy of classification as well as time involved.
8	Cluster-based Crowd Movement Behaviour Detection	YOLOv2, fixed-width clustering algorithm	For evaluation this paper uses six video sequences form PETS2009 and the results have shown the accuracy of the proposed methodology to be between 80%-95%.
9	Abnormal Crowd Behaviour Detection Using Speed and Direction Models	Expectation maximization algorithm, Optical flow vectors, Probability models	Abnormal behaviour can be detected by calculating the distance between the probability models of current model with those of the previous ones. If the distance is more than a threshold value it indicates an anomalous event in the crowd
10	A Large-Scale Crowd Density Classification using Spatio-Temporal Local Binary Pattern	dynamic texture analysis, multi-class support vector regression, spatio-temporal local binary (RIST-LBP) pattern	It was observed that the proposed strategy has lower computational complexities and high efficiencies as compared to the conventional pixel-based techniques.
11	Detecting Coherent Groups in Crowd Scenes by Multiview Clustering	Multiview-based Parameter Free framework (MPF), Structural Context descriptor, Self-weighted Multiview Clustering method	Results of the analysis on various datasets have proven the effectiveness of the proposed approach.
12	Abnormal Behaviour Detection in Crowded Scenes	convolutional LSTMs, density heat maps,	Once trained on activities that are either normal or abnormal the network is ready to classify the abnormal events

	Using Density Heatmaps and Optical Flow	Optical Flow	in crowd videos. The paper also curates a new dataset using GTA V engine that contain videos with abnormal behaviour.
13	DA-Net: Learning the Fine-Grained Density Distribution with Deformation Aggregation Network	DA-Net, multi-layer aggregation, deform blocks	Multi-layer extraction allows the network to immensely improve the upkeeping of the spatial locations. Also, the paper introduces adaptive weights to replace direct aggregation to evaluate the reliability of multilevel outputs.
14	Deep and Wide Convolutional Neural Network Model for Highly Dense Crowd	CNN	This approach was tested on ShanghaiTech Part-A datasets with performance metrics as mean square error and mean absolute error. As a result, the approach tested successfully on comparing these metrics with other approaches.
15	A Method Based on Multi-source Feature Detection for Counting People in Crowded Areas	HOG, LBP, CANNY, SVM, regression analysis	This approach was found to be easy and fast to implement. This experiment gives comparative results as compared to the joint HOG.
16	Machine and Deep Learning for Crowd Analytics	Bayesian model, MATLAB	As a solution this paper the Bayesian model as a significant feature. The proposed method is implemented using MATLAB. And on average achieves 88.83% accuracy when applied on the videos from the same dataset.

CONCLUSION

This paper was written with the purpose of highlighting the researches done in the past in the field of crowd behaviour analysis, abnormal crowd behaviour detection and crowd counting. We have discussed the various methodologies employed for this purpose by the researchers. We have discussed that two approaches were adopted, namely, object-based and holistic. On one hand, object-based strategies made in depth feature extraction that was done considering each individual in the crowd as a separate entity. Many algorithms based on this object-based approach were studied such as Motion Information Images (MII) and optical flow vectors. Uses gradient to model individual motion followed by Genetic Programming (GP) based classifier for easy hardware implementation. The other approach takes entire crowd as a single unit. Most algorithms based on this approach use CNN. Though it is a powerful tool to analyse crowd images and videos, many issues such as occlusion, scale, size, perspective exist. For handling them large spatial filters are used but that increases the complexity. This is solved by using cascaded convolution filters.

The work is still going on in this field for getting more accuracy and efficiency contrary to the past approaches.

FUTURE SCOPE

There is an immense possibility of scope for improvement and optimization in the strategies and algorithms implemented for the purpose of crowd behaviour analysis. A need is still felt to bring a more practical and feasible approach as the applications of this research are immense and wide. Crowd behaviour is hard to model and the problem is only aggravated by the lack of data regarding normal and abnormal crowd behaviour. This gives the momentum to continue the research and strive to find the most optimal algorithms possible to detect and predict crowd behaviour.

REFERENCES

1. C. Direkoglu. (2020). Abnormal crowd behaviour detection using motion information images and convolutional neural networks. *IEEE Access*, 8, pp. 80408-80416, Available at: https://doi.org/10.1109/ACCESS.2020.2990355.
2. G. Tripathi, K. Singh & D. K. Vishwakarma. (2020). Crowd emotion analysis using 2D convnets. *Third International Conference on Smart Systems and Inventive Technology*. IEEE, pp. 969-974, Available at: https://doi.org/10.1109/ICSSIT48917.2020.9214208.
3. C. Santhini & V. Gomathi. (2018). Crowd scene analysis using deep learning network. *International Conference on Current Trends towards Converging Technologies*. IEEE, pp. 1-5, Available at: https://doi.org/10.1109/ICCTCT.2018.8550851.
4. I. Usman. (2019). Anomalous crowd behavior detection in time varying motion sequences. *4th World Conference on Complex Systems*. IEEE, pp. 1-5, Available at: https://doi.org/10.1109/ICoCS.2019.8930795.
5. X. Ma, S. Du & Y. Liu. (2019). A lightweight neural network for crowd analysis of images with congested scenes. *IEEE International Conference on Image Processing*. IEEE, pp. 979-983, Available at: https://doi.org/10.1109/ICIP.2019.8803062.
6. M. Yang et al. (2018). Cluster-based crowd movement behavior detection. *Digital Image Computing: Techniques and Applications*. IEEE, pp. 1-8, Available at: https://doi.org/10.1109/DICTA.2018.8615809.
7. A. Chibloun, S. El Fkihi, H. Mliki, M. Hammami & R. Oulad Haj Thami. (2018). Abnormal crowd behavior detection using speed and direction models. *9th International Symposium on Signal, Image, Video and Communications*. IEEE, pp. 197-202, Available at: https://doi.org/10.1109/ISIVC.2018.8709192.
8. M. Gao et al. (2019). Violent crowd behavior detection using deep learning and compressive sensing. *Chinese Control and Decision Conference*. IEEE, pp. 5329-5333, Available at: https://doi.org/10.1109/CCDC.2019.8832598.
9. L. Lazaridis, A. Dimou & P. Daras. (2018). Abnormal behavior detection in crowded scenes using density heatmaps and optical flow. *26th European Signal Processing Conference*. IEEE, pp. 2060-2064, Available at: https://doi.org/10.23919/EUSIPCO.2018.8553620.
10. Xie S., Zhang X. & Cai J. (2019). Video crowd detection and abnormal behavior model detection based on machine learning method. *Neural Computing and Applications*, 31, pp. 175-184, Available at: https://link.springer.com/article/10.1007/s00521-018-3692-x.
11. S. Lamba & N. Nain. (2017). A large-scale crowd density classification using spatio-temporal local binary pattern. *13th International Conference on Signal-Image Technology & Internet-Based Systems*. IEEE, pp. 296-302, Available at: https://doi.org/10.1109/SITIS.2017.57.
12. Singh J. & Khan, M. Z. (2019). Detection of fake profile in social media. *Journal of Emerging Technologies and Innovative Research*, 6(6), PP. 38-42. Available at: http://www.jetir.org/papers/JETIR1906M04.pdf.
13. M. A. Kizrak & B. Bolat. (2019). Deep and wide convolutional neural network model for highly dense crowd. *Innovations in Intelligent Systems and Applications Conference* (ASYU). IEEE, pp. 1-6, Available at: https://doi.org/10.1109/ASYU48272.2019.8946395.

14. Z. Zou, X. Su, X. Qu & P. Zhou. (2018). DA-Net: Learning the fine-grained density distribution with deformation aggregation network. *IEEE Access*, 6, pp. 60745-60756, Available at: https://doi.org/10.1109/ACCESS.2018.2875495.
15. Q. Wang, M. Chen, F. Nie & X. Li. (2020). Detecting coherent groups in crowd scenes by multiview clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42(1), pp. 46-58, Available at: https://doi.org/10.1109/TPAMI.2018.2875002.
16. G. Songchenchen & E. Bourennane. (2019). A method based on multi-source feature detection for counting people in crowded areas. *IEEE 4th International Conference on Signal and Image Processing*. IEEE, pp. 1021-1025, Available at: https://doi.org/10.1109/SIPROCESS.2019.8868691.
17. Siraj Muhammad. (2019). Machine and deep learning for crowd analytics. *Computer Vision and Pattern Recognition*, Available at: https://arxiv.org/abs/1909.04150?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%253A+arxiv%252FQSXk+%2528ExcitingAds%2521+cs+updates+on+arXiv.org%2529.

Research
Paper Publication Certificate

DOGO RANGSANG
Research Journal
দগো ৰাংছাং
গৱেষণা পত্ৰিকা

ISSN : 2347-7180

CERTIFICATE OF PUBLICATION

This is to certify that the article entitled

MACHINE LEARNING BASED CROWD BEHAVIOUR ANALYSIS AND PREDICTION

Authored By

Neha Srivastava
M.Tech Student, Department of Computer Science and Engineering, Integral University, Lucknow, India

Published in
Dogo Rangsang Research Journal : ISSN 2347-7180
Vol. 11, Issue. 08, No. 01 : 2021
UGC Care Approved, Group I, Peer Reviewed, Bilingual and Referred Journal

ज्ञान-विज्ञान विमुक्तये
UGC
University Grants Commission

Chief Editor
(Hon.) – Dr. Upen Rabha Hakacham

MACHINE LEARNING BASED CROWD BEHAVIOUR ANALYSIS AND PREDICTION

Neha Srivastava M.Tech Student, Department of Computer Science and Engineering, Integral University, Lucknow, India
Dr. Mohammad Zunnun Khan Department of Computer Science and Engineering, Integral University, Lucknow, India

Abstract
Crowd behaviour has been widely known to have the ability to forecast the events a crowd could create. Crowd management can become extremely efficient if situations such as riots, mob lynching, traffic jams, accidents, stampede, etc. could be predicted beforehand. In this paper we have introduced a new approach to predict crowd behaviour using Multicolumn Convolution Neural Network (MCNN). First of all, we process the input image and extract its features. Then we calculate the head count of the crowd in the image and crop patches from the image. For each patch of the image, we extract low level features. Next, we generate density maps of the objects in the image. Then our algorithm learns a linear mapping between the extracted features and their object density maps. Finally, we apply MCNN algorithm to count the crowd and predict. On performing have tested our algorithm on UCF-QNRF dataset.

Keywords: Crowd behaviour analysis, anomalous crowd prediction, deep learning

Introduction

Crowd denotes a group of large number of people. Management of crowd hence, is of utmost importance primarily due to the reason that if not managed properly crowd might lead to many problematic situations such as riots, stampede, violence, etc. For performing crowd management, we must analyze the behavior of the crowd that depends on the collective emotions of the people in that crowd. People in crowd are present at different positions and move in different directions which makes it even more difficult to find the effective features [1]. Understanding the crowd emotions is complicated for it is just not the sum total of individual emotions [2]. The biggest challenge in anomalous crowd detection is determining normal crowd behavior and comparing it to the current crowd behavior [4]. Spatial distribution information of crowd has immense use crowd analysis. A lot of research has been based on spatial and temporal feature-based models and can primarily be categorized as learning models with classical classifiers such as SVM and deep learning model [6].

A lot of recent studies for crowd analysis have introduced many methods involving CNNs and density maps [5]. Machine learning and deep neural networks have introduced a new approach for seeking a solution to the problem of crowd behavior analysis and prediction. Two strategies are generally used to perform crowd analysis [7].

a) Object based approach: This approach considers the crowd as a collection of individual persons and tracks the motion and context of each of these persons to calculate the behavior of the crowd.
b) Holistic approach: This approach uses the crowd as a single entity and tracks the motion and behavior of the entire crowd as a single unit.

Object based approach requires tracking of individuals in a crowd which is extremely difficult when the scene is densely crowded such as at sports stadiums. Many aspects such as size, scale, perspective, occlusion, boundary restrictions, etc. affect the analysis of the crowd behavior [8]. But object-based approach is a feasible approach for scenes that are lightly crowded. For this reason, crowd behaviour analysis is a problem under unsupervised learning.

Statement of the Problem

The problem that we are trying to address in this paper is the prediction of the nature or the type of crowd by analyzing the images of crowds using machine learning algorithms in order to prevent the crowd becoming risky and problematic.

Objectives of the study

- The crowd itself is a big problem in today's time and we detect the same from this approach.
- With this approach, we can find out what kind of crowd it is so that we can take the right steps at the right time.
- If we detect the congestion happening before the right time, then according to that, we can make the right arrangement so that this congestion does not turn into a problem.
- Through this approach, we can find out that this crowd is normal traffic, any violent mob or any type of ordinary mob.

Review of Literature

Literature provides a lot of work that has been done in the area of crowd behavior analysis and crowd counting. Motion Information Images (MII) representing the crowd motion, based on optical flow vectors' magnitude and angular difference have been used along with Convolution Neural Networks (CNN) to perform abnormal crowd event detection [1]. Tripathi et al. [2] implemented 2D Convolution Neural Networks (ConvNets), a variation of Xception Network performs the classification of crowd emotion into six categories such as anger, sadness, excitement, happiness, scaredness and neutral. It solves the issue of vanishing gradients by creating a shortcut in sequential network through residual connections. Gao et al. [3] combined deep learning networks and compressed sensing to detect violent crowd behaviour using Violent Flows Algorithm (Vif) and Hybrid Random Matrix (HRM) approach. Ma et al. [5] developed a light weight neural network that makes use of only 0.86 M parameters for crowd analysis of congested scenes which involves multi scale feature extraction, density map generation and an autoencoder. In [7] Umran presented an automated model for anomalous crowd detection that differentiates between activities in time varying video sequence by employing spatio-temporal cuboids.

Yang et al. [8] proposed a real time model for detecting event crowd based on fixed width clustering object detection. The objects detected by You Only Look Once (YOLOv2) are clustered whose centroids depicting their movement behaviour are used for classification. Lamba and Nain [10] extract local spatio-temporal features of each interest point detected by Hessian Matrix represent them by Rotation Invariant Spatio-Temporal Local Binary Pattern (RIST-LBP) code. The generated dynamic textures are used to classify crowd densities. Wang et al [11] introduced a Multiview-based Parameter Free framework (MPF) and Structural context descriptor for identifying coherent groups in crowd videos. Structural Context (SC) descriptor for each feature point. The correlations of the feature points from the context view and orientation is done by multi view clustering method and coherent local groups are detected using a tightness based merging approach. Lazaridis et al. [12] emphasizes that optical flow must be combined with density heat maps to analyse the motion in the crowd videos as motion in the crowd may be initiated by things apart from people. For this they use Long Short-Term Memory (LSTM) neural networks to learn the spatio-temporal patterns to classify normal and abnormal events in crowd.

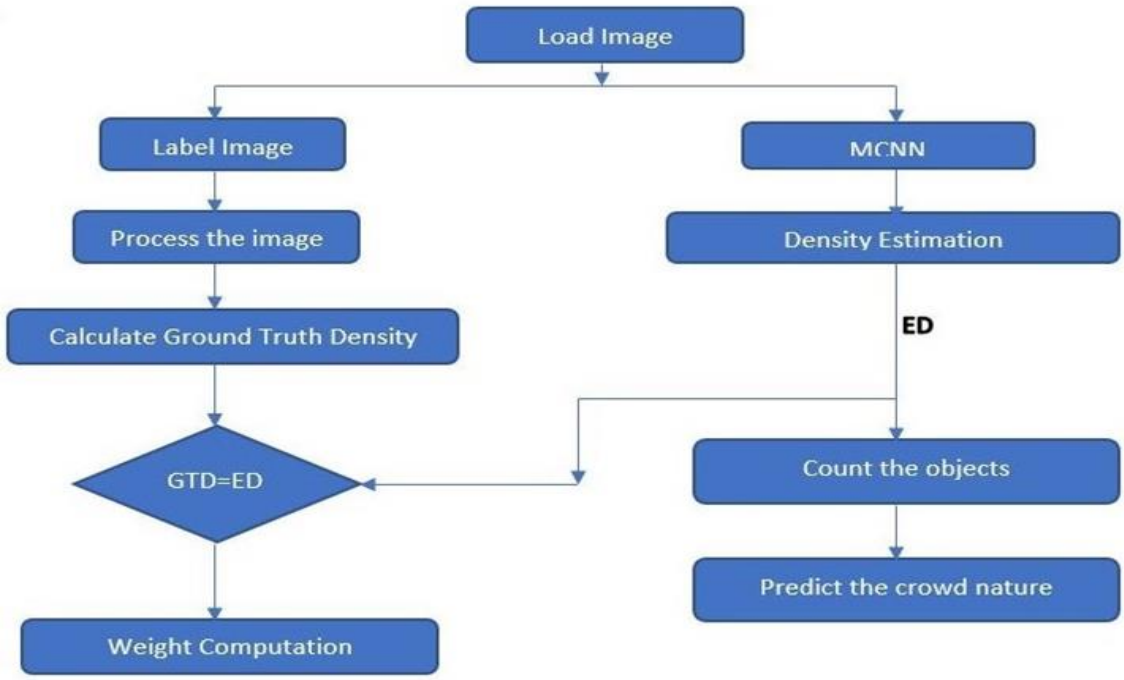
Zou et al. [13] introduce a Deformation Aggregation Network (DA-Net) with the capability of learning extra offsets to augment the spatial sampling locations making it more robust to scale variations. Kizrak and Bolat [14] designed a new convolution neural networks made of two parallel modules to extract different features from the crowd images. The first module pulls out lower-level features to generate density maps and second module obtains higher level features to identify body parts. Finally, these two modules are attached to a fully connected neural network to perform the crowd count estimation. Songchenchen and Bourennane [15] present a feature fusion method to draw and analyse multiple kinds of features of crowd images such as texture features and edges of crowd images. A Support Vector Machine (SVM) that is trained on the count estimates collected for each source, learns the features and classifies highly dense crowd images using regression analysis. In [16] Siraj proposes to implement the Bayesian model using MATLAB to find the properties of the image pixels across multiple adjacent video frames and perform anomaly detection in crowd images or video frames.

Remaining of the paper is organized is as follows: Research methodology, Results and discussions and Conclusion.

Research Methodology

This section describes our proposed methodology in detail. **Figure 1** shows the overview of the methodology. We have presented a novel approach to predict crowd behaviour that uses Multicolumn Convolution Neural Network (MCNN). The pre-processing of the input image is the first step. We use OpenCV for the purpose of pre-processing of the image. The aim of this pre-processing is the extraction of low-level features. For that purpose, we first get the image spectrum and then filter it to crop it into patches depending on the frame set. For each such patch we extract low level features. From this we create density maps of the objects and linearly map them with the extracted features to obtain the Ground Truth Density of the image. At this step we introduce our MCNN algorithm to predict the crowd density or the Estimated Density.

Figure 1: **Overview of the proposed methodology**



An image is input to the MCNN and the output is a density map of the crowd whose summation gives the total crowd count. To train the model we compare the Estimated Density (ED) with the Ground Truth Density (GTD) calculated earlier.

If the difference is within the threshold, then the model is saved by finalizing the weights, else the model is trained again on new weights. Once our model is ready the density ratio is checked to differentiate between normal and abnormal crowd. **Table 1** is used to predict the normal and abnormal behaviour of the crowd.

Table No.1: Crowd area and their assumed normal crowd density

Crowd Area	Normal Crowd Density
100 m^2	70
200 m^2	120
300 m^2	320

400 m^2	380
500 m^2	400

This entire process is summed up in **Algorithm 1**.

Algorithm 1: Step wise method to crowd analysis

Step 1: Input the image for crowd counting

Step 2: Pre-process the image to get the image spectrum

Step 3: Crop the image into patches and extract low level features in each patch.

Step 4: Generate the density maps and create a linear mapping between the density map and the features extracted in the previous step. This gives the ground truth density.

Step 5: Now the MCNN is used to calculate the estimated density

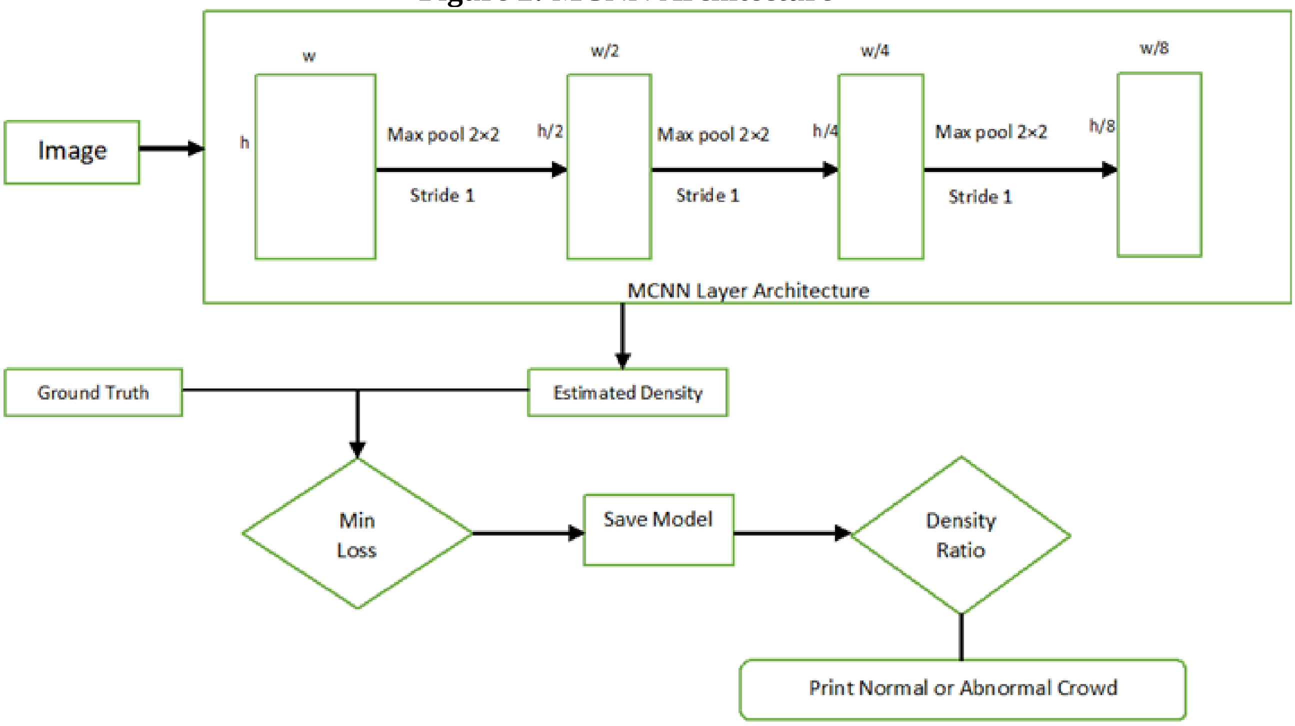
Step 6: The estimated density is compared with the ground truth density to check the accuracy of the model

Step 7: Once model is fit and the final weights(filters) are learned, the model can be used to calculate the density of an input image

Step 8: Based on the estimated density ratio we can label the crowd in the image as normal or abnormal.

The complete architecture of the MCNN is illustrated in **Figure2**. It consists of three parallel CNNs that uses filters connected to receptive fields that are of different sizes. These parallel CNNs are called the columns each of which has the same three layers (Convolution and Max Pooling) with varying filter sizes and numbers. This means that there are a total of three max pooling layers that create an output feature map that is down sampled to 1/8th of the input feature map. Size of max pooling filter is 2×2 and the activation function used is Rectified linear unit (ReLU). We use a filter of size 1 ×1 to map the output feature maps of all CNNs to a final density map. We adopt the method of Euclidean distance to calculate the difference between the estimated density map and ground truth.

Figure 2: MCNN Architecture



Results and Discussion

We have divided this section into three sections. The first section discusses the metrics used for the evaluation of our proposed methodology. The second section describes the dataset that is

used in the analysis for training and testing our model. The last section shows the results of our experiments and
presents a comparative study with respect to other techniques such as CNN and LSTM.

a. Evaluation Metrics

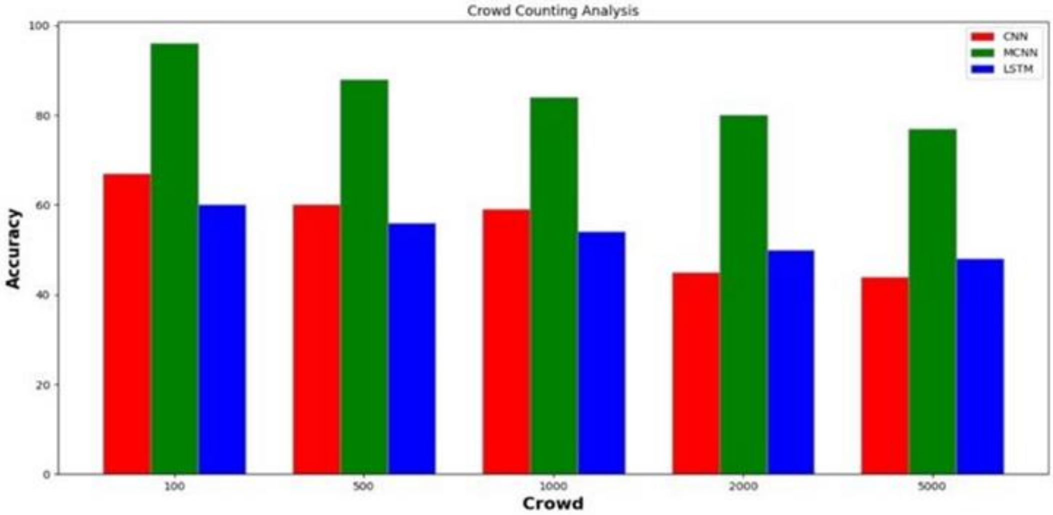
To evaluate the performance of our proposed model (MCNN) we have used two evaluation metrics in our experiment. These metrics are Mean Absolute Error (MAE) and Mean Squared Error (MSE). The MSE is just the square root of the MAE but it gives a standardized evaluation. Hence, we evaluate our proposed method against other methods using these metrics.

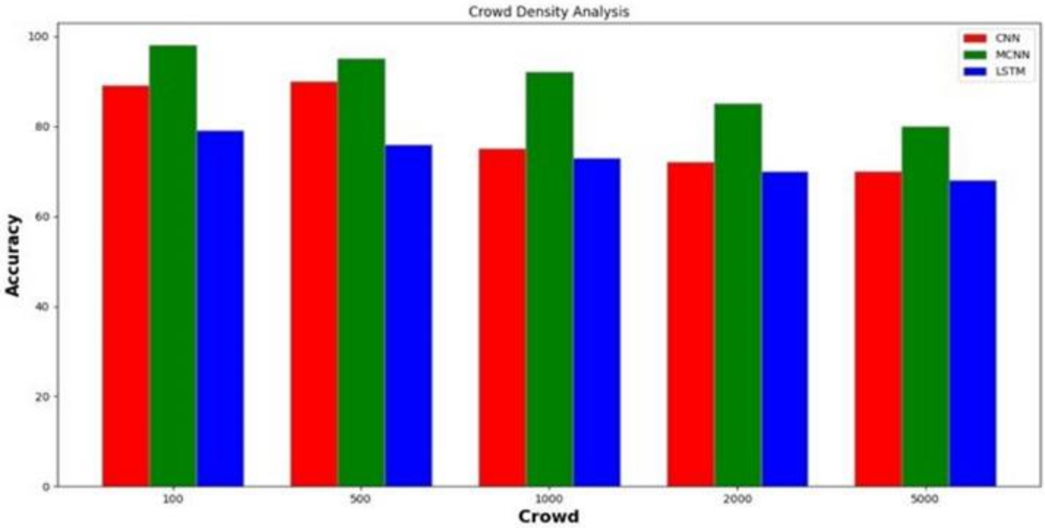
$$MAE = \frac{1}{N}\sum_{1}^{N}|z_i - \hat{z}_i|$$

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

N = Number of images
Z_i = Number of people in ith image
$\hat{z}_I$ = Estimated number of people in the ith image

Figure 3: Graphs depicting the comparison of other methods and MCNN method.





b. Dataset

We are using the UCF-QNRF dataset for the training and testing of our proposed methodology.

This dataset has a total of 1535 annotated images of densely crowded scenarios. Out of these images 1201 images are for the purpose of training and 334 images are used for testing.

c. Results

On applying our proposed methodology on the UCF-QNRF dataset it is observed that our method outperforms the other techniques such as CNN and LSTM. Figure 3 shows the graphical comparison of the accuracy of these methods to our proposed method (MCNN). This provides the evidence that MCNN shows better accuracy in crowd counting and crowd density analysis as compared to CNN and LSTM. Table 2 and Table 3 compares the performance of our model and that of CNN and LSTM for crowd counting and crowd density analysis respectively.

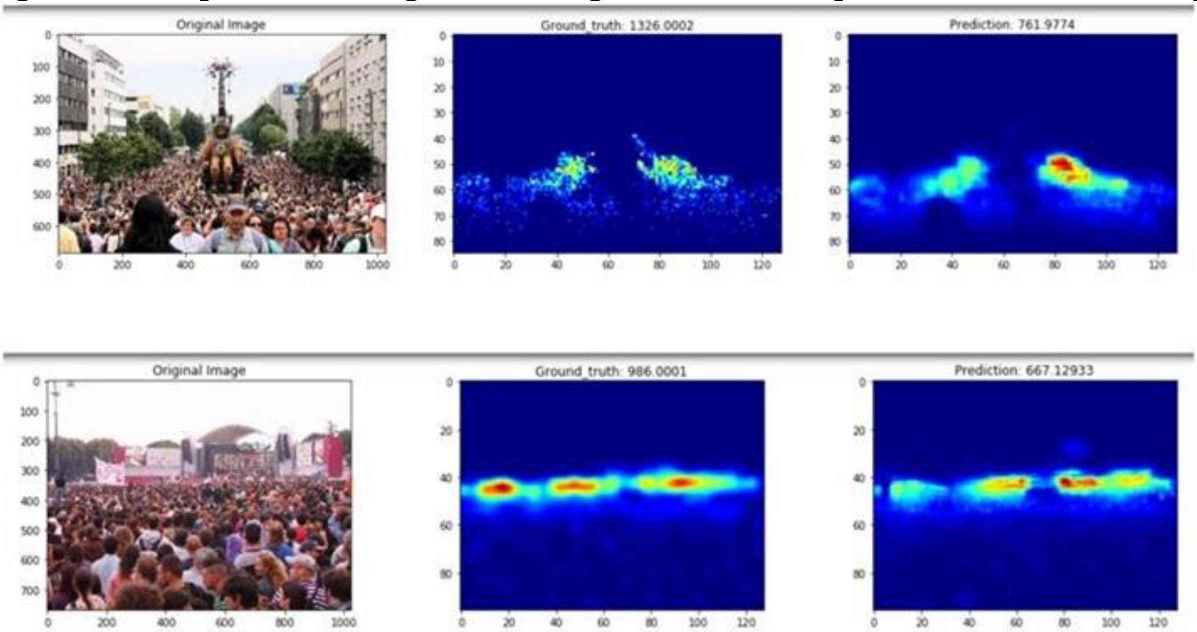
Table No.2: Comparative study of different methods and our approach for Crowd Count Accuracy

Crowd	100	500	1000	2000	5000
CNN	67	60	59	45	44
MCNN	**96**	**88**	**84**	**80**	**77**
LSTM	60	56	54	50	48

Table No.3: Comparative study of different methods and our approach for Crowd Density Accuracy

Crowd	100	500	1000	2000	5000
CNN	89	90	75	72	70
MCNN	**98**	**95**	**92**	**85**	**80**
LSTM	79	76	73	70	68

Figure 4: Examples of test images and their ground truth and predicted estimated density



Conclusion

This paper presents a novel methodology to perform analysis of an image of a crowd scene and predict the crowd behavior in an image by proposing a Multi Column Convolution Neural Network (MCNN) with three parallel CNNs each using three max pooling layers. The performance of our model along with two other state of the art methods is evaluated on the UCF-QNRF dataset. This dataset is comprising of 1535 annotated images of highly congested scenes.

It is seen that our model performs better than the other two methods and demonstrates higher accuracy when evaluated on this dataset.

References

1. C. Direkoglu, "Abnormal Crowd Behaviour Detection Using Motion Information Images and Convolutional Neural Networks," in IEEE Access, vol. 8, pp. 80408-80416, 2020, doi: 10.1109/ACCESS.2020.2990355.

2. G. Tripathi, K. Singh and D. K. Vishwakarma, "Crowd Emotion Analysis Using 2D ConvNets," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 969-974, doi: 10.1109/ICSSIT48917.2020.9214208.
3. M. Gao et al., "Violent crowd behavior detection using deep learning and compressive sensing," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 5329-5333, doi: 10.1109/CCDC.2019.8832598.
4. Xie, S., Zhang, X. & Cai, J. Video crowd detection and abnormal behavior model detection based on machine learning method. *Neural Comput & Applic* **31,** 175–184 (2019).
5. X. Ma, S. Du and Y. Liu, "A Lightweight Neural Network For Crowd Analysis Of Images With Congested Scenes," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 979-983, doi: 10.1109/ICIP.2019.8803062.
6. C. Santhini and V. Gomathi, "Crowd Scene Analysis Using Deep Learning Network," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8550851.
7. I. Usman, "Anomalous Crowd Behavior Detection in Time Varying Motion Sequences," 2019 4th World Conference on Complex Systems (WCCS), Ouarzazate, Morocco, 2019, pp. 1-5, doi: 10.1109/ICoCS.2019.8930795.
8. M. Yang et al., "Cluster-Based Crowd Movement Behavior Detection," 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 2018, pp. 1-8, doi: 10.1109/DICTA.2018.8615809.
9. A. CHIBLOUN, S. EL FKIHI, H. MLIKI, M. HAMMAMI and R. OULAD HAJ THAMI, "Abnormal Crowd Behavior Detection Using Speed and Direction Models," 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC), Rabat, Morocco, 2018, pp. 197-202, doi: 10.1109/ISIVC.2018.8709192.
10. S. Lamba and N. Nain, "A Large Scale Crowd Density Classification Using Spatio-Temporal Local Binary Pattern," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 296-302, doi: 10.1109/SITIS.2017.57.
11. Q. Wang, M. Chen, F. Nie and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 1, pp. 46-58, 1 Jan. 2020, doi: 10.1109/TPAMI.2018.2875002.
12. L. Lazaridis, A. Dimou and P. Daras, "Abnormal Behavior Detection in Crowded Scenes Using Density Heatmaps and Optical Flow," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 2060-2064, doi: 10.23919/EUSIPCO.2018.8553620.
13. Z. Zou, X. Su, X. Qu and P. Zhou, "DA-Net: Learning the Fine-Grained Density Distribution With Deformation Aggregation Network," in IEEE Access, vol. 6, pp. 60745-60756, 2018, doi: 10.1109/ACCESS.2018.2875495.
14. M. A. Kizrak and B. Bolat, "Deep and Wide Convolutional Neural Network Model for Highly Dense Crowd," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 2019, pp. 1-6, doi: 10.1109/ASYU48272.2019.8946395.
15. G. songchenchen and E. Bourennane, "A Method Based on Multi-source Feature Detection for Counting People in Crowded Areas," 2019 IEEE 4th International

Conference on Signal and Image Processing (ICSIP), Wuxi, China, 2019, pp. 1021-1025, doi: 10.1109/SIPROCESS.2019.8868691.
16. Siraj, Muhammad. “Machine and Deep Learning for Crowd Analytics.” *ArXiv* abs/1909.04150 (2019): n. pag.
17. X. Li, M. Chen and Q. Wang, "Quantifying and Detecting Collective Motion in Crowd Scenes," in IEEE Transactions on Image Processing, vol. 29, pp. 5571-5583, 2020, doi: 10.1109/TIP.2020.2985284.
18. S. S. A. Abbas, P. O. Jayaprakash, M. Anitha and X. V. Jaini, "Crowd detection and management using cascade classifier on ARMv8 and OpenCV-Python," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-6, doi: 10.1109/ICIIECS.2017.8275988.
19. Li, Yan & Liu, Hong & Zheng, Xiangwei & Han, Yanbin & Li, Liang. (2019). A Top-Bottom Clustering Algorithm Based on Crowd Trajectories for Small Group Classification. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2902310.
20. M. Abdou and A. Erradi, "Crowd Counting: A Survey of Machine Learning Approaches," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 48-54, doi: 10.1109/ICIoT48696.2020.9089594.
21. A. N. Moustafa, M. E. Hussein and W. Gomaa, "Gate and Common Pathway Detection in Crowd Scenes Using Motion Units and Meta-Tracking," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, NSW, 2017, pp. 1-8, doi: 10.1109/DICTA.2017.8227438.
22. N. Kumaran and U. S. Reddy, "Object detection and tracking in crowd environment — A review," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 777-782, doi: 10.1109/ICICI.2017.8365242.
23. A. K. Jhapate, S. Malviya and M. Jhapate, "Unusual Crowd Activity Detection using OpenCV and Motion Influence Map," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170704.
24. M. Mancas et al., "People Groups Analysis for AR Applications," 2018 International Conference on 3D Immersion (IC3D), Brussels, Belgium, 2018, pp. 1-8, doi: 10.1109/IC3D.2018.8657892.
25. N. Nain and S. Lamba, "Oriented Tracklets Approach for Anomalous Scene Detection in High Density Crowd," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 435-441, doi: 10.1109/SITIS.2018.00073.
26. A. Bera, S. Kim and D. Manocha, "Interactive Crowd-Behavior Learning for Surveillance and Training," in IEEE Computer Graphics and Applications, vol. 36, no. 6, pp. 37-45, Nov.-Dec. 2016, doi: 10.1109/MCG.2016.113.
27. P. V. V. Kishore, R. Rahul, K. Sravya and A. S. C. S. Sastry, "Crowd Density Analysis and tracking," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, 2015, pp. 1209-1213, doi: 10.1109/ICACCI.2015.7275777.
28. T. Senst, V. Eiselein, I. Keller and T. Sikora, "Crowd analysis in non-static cameras using feature tracking and multi-person density," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 6041-6045, doi: 10.1109/ICIP.2014.7026219.
29. M. R. Rejitha and S. N. George, "An Unsupervised Abnormal Crowd Behavior Detection Technique using Farneback Algorithm," 2019 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2019, pp. 1-5, doi: 10.1109/CONECCT47791.2019.9012845.
30. H. Huang, H. Guo, Z. Ding, Y. Chen and X. Wu, "3D virtual modeling for crowd analysis," 2015 IEEE International Conference on Information and Automation, Lijiang, 2015, pp. 2949-2954, doi: 10.1109/ICInfA.2015.7279793.

31. Singh, J., & Khan, M. Z. (2019). Detection of Fake Profile in Social Media. Journal of Emerging Technologies and Innovative Research, 6(6), 38–42. Retrieved from http://www.jetir.org/papers/JETIR1906M04.pdf
32. Khan, M. Z., & Shoaib, M. (2019). Healthcare Analytics in the Modern Era: A Survey. International Journal of Research in Advent Technology,7(S), March 2019, 132-134. Retrieved from http://www.ijrat.org/downloads/Conference_Proceedings/Special_Issue_March-19/Paper ID-156.pdf
33. M. Z. Khan, M. S. Kidwai, F. Ahamad and M. U. Khan, "Hadoop based EMH framework: A Big Data approach," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 1068-1070, doi: 10.1109/ICACITE51222.2021.9404710
34. M. Suhaib Kidwai and M. Zunnun Khan, "A new perspective of detecting and classifying neurological disorders through recurrence and machine learning classifiers," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 200-206, doi: 10.1109/ICACITE51222.2021.9404645.