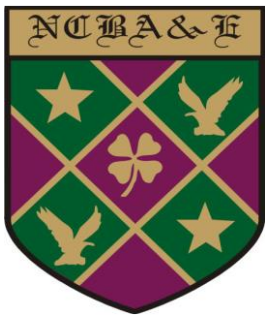
National College of Business Administration & Economics Lahore



REVOLUTIONIZING CROWD SURVEILLANCE: EMPOWERING SURVEILLANCE SYSTEM USING TRANSFER LEARNING

BY

ZOHAIB AHMAD CHUGHTAI

MASTER OF PHILOSOPHY
IN
COMPUTER SCIENCE

SEPTEMBER, 2023

NATIONAL COLLEGE OF BUSINESS ADMINISTRATION & ECONOMICS

REVOLUTIONIZING CROWD SURVEILLANCE: EMPOWERING SURVEILLANCE SYSTEM USING TRANSFER LEARNING

BY

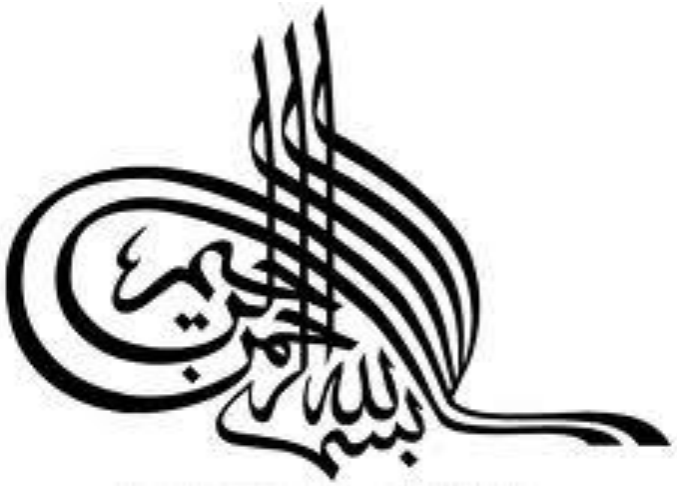
ZOHAIB AHMAD CHUGHTAI

A dissertation submitted to
Faculty of Computer Sciences

In Partial Fulfillment oft he
Requirements for the Degree of

MASTER OF PHILOSOPHY
IN
COMPUTER SCIENCE

SEPTEMBER, 2023



In the name of ALLAH,
The Most Beneficial,
The Most Merciful,

NATIONAL COLLEGE OF BUSINESS ADMINISTRATION & ECONOMICS LAHORE

REVOLUTIONIZING CROWD SURVEILLANCE: EMPOWERING SURVEILLANCE SYSTEM USING TRANSFER LEARNING

BY
ZOHAIB AHMAD CHUGHTAI

A dissertation submitted to Faculty of Computer Sciences, in partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY IN COMPUTER SCIENCE

Dissertation Committee:

_____
Chairman

_____
Member

_____
Member

DECLARATION

It is to declare that this research work has not been submitted for obtaining similar degree from any other university/college.

Zohaib Ahmad Chughtai
September, 2023

Dedicated
To

My Parents

&

My Teachers

ACKNOWLEDGEMENT

All glory, honour, and praise are due to Allah, the Most Gracious, Merciful, and Magnificent, who has enabled me to finish my dissertation. I ask Him for direction and protection throughout my entire life.

We would like to express our gratitude to Dr. Muhammad Saleem, our supervisor, for all of his technical support and advice. Without you, it would not have been feasible to finish this project.

RESEARCH COMPLETION CERTIFICATE

Certified that the research work contained in this thesis entitled **"Revolutionizing Crowd Surveillance: Empowering Surveillance System using Transfer Learning"** has been carried out and completed by **Mr. Zohaib Ahmad Chughtai** under my supervision during his **M.Phil. Computer Science** Programme.

(Dr. Muhammad Saleem)
Supervisor

SUMMARY

Crowd surveillance experiences critical difficulties in actually observing and breaking down complex group ways of behaving, including impediments, swarm thickness varieties, and unexpected changes in conduct. Addressing these difficulties is significant to upgrade the proficiency and unwavering quality of reconnaissance frameworks. Move learning is an arising procedure in PC vision, which arises as a promising arrangement by utilizing pre-prepared models to extricate significant elements from visual information, empowering exact recognition, following, and examination of people in jam-packed scenes.

In this research work, a transfer learning-based model is proposed for improving crowd surveillance detection accuracy, tracking precision, and system adaptability in order to enhance the security, reliability, integrity in smart cities. The effectiveness of the proposed model in revolutionizing crowd surveillance is offering benefits such as enhanced situational awareness, early threat detection, and optimized resource allocation. This research presents a valuable contribution to the field by harnessing transfer learning to address the evolving challenges of crowd monitoring and analysis while predicting the crowd surveillance, the proposed model for 2D image classification, the proposed system have achieved 95.34% validation accuracy on self-collected dataset.

LIST OF ABBREVIATION

Abbreviation	Description
TLBCSS	Transfer Learning Based Crowd Surveillance System
RNNs	Recurrent Neural Networks
CNNs	Convolutional Neural Networks
SVMs	Support Vector Machines
LSTM	Long Short-Term Memory
DNNs	Deep Neural Networks
COCO	Common Objects in Context
GPS	Global positioning system
RFID	Radio-frequency identification
UCSD	University of California, San Diego
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformers
CCTV	Closed-circuit television
ANPR	automatic number plate recognition
GPU	Graphics processing unit
TPU	Tensor Processing Unit
SIFT	scale-invariant feature transformation
SURF	Speeded Up Robust Features
HOG	Histogram of oriented gradients
LBP	local binary pattern
VGG	Visual Geometry Group
ResNet	Residual Network

Abbreviation	Description
GRU	Gated Recurrent Units
GANs	Generative adversarial networks
DBN	Deep Belief Network
CAD	Computer-aided Design
AUC	Area Under the ROC Curve
HTM	hierarchical temporal memory
KNN	k-nearest neighbors algorithm
SSD	single shot detector
SORT	Simple Online and Real-time Tracking
DCT Nets	Deep crowd transfer network
ICCE	International Conference on Consumer Electronics

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Crowd surveillance assumes a fundamental part in guaranteeing public wellbeing and security across different spaces, including public occasions, transportation center points, and metropolitan regions. The capacity to really screen and examine swarm conduct might assume a crucial part in early danger recognition, proficient asset designation, and opportune reactions. The group reconnaissance frameworks can be introduced in different settings, including public vehicle centers, air terminals, arenas, and downtown areas. These frameworks are fit for following people and checking their developments progressively, as well as investigating information from numerous sources to recognize examples and likely dangers. A portion of the vital highlights of a group observation framework incorporate facial acknowledgment, social examination, and item following. Facial acknowledgment innovation can recognize people in a group in view of their facial highlights, while conduct examination can distinguish dubious way of behaving, for example, dallying, sporadic development, or forceful way of behaving. Object following can be utilized to follow explicit items or individuals as they travel through a group. The objective of a group reconnaissance framework is to work on open wellbeing by recognizing and forestalling potential security dangers in packed regions. Nonetheless, these frameworks additionally raise worries about security and common freedoms, as they can be utilized to screen the developments and exercises of enormous gatherings.

A surveillance system is an innovation based framework intended to screen and catch data about an area or explicit movement. The motivation behind reconnaissance frameworks is to further develop wellbeing and security, forestall wrongdoing, and upgrade situational mindfulness. There are various kinds of reconnaissance frameworks, including video observation, sound reconnaissance, and electronic reconnaissance that might be utilized in different settings, including public places, for example, air terminals, train stations, and shopping centers, as well as in confidential homes and organizations. These can be utilized to screen worker conduct, forestall burglary, identify security dangers, traffic the board, untamed life observing, and ecological checking.

Lately, progresses in computerized reasoning (computer based intelligence) and AI (ML) have prompted the improvement of shrewd observation frameworks that can naturally recognize and break down exercises progressively. In any case, the utilization of reconnaissance frameworks raises

worries about security and common freedoms. It is fundamental with comply to appropriate moral and legitimate rules to guarantee the capable and straightforward utilization of group reconnaissance situation while regarding individual security freedoms. Swarm reconnaissance presents a provoking errand because of the unpredictable and consistently changing nature of group conduct.

Move learning is a ML method that can be utilized to address this test by utilizing pre-prepared models on related errands to work on the exhibition of group observation strategies. Move gaining is a strong strategy where information acquired from preparing a model on one errand can be used to work on the exhibition of an alternate however related task, like in the improvement of group reconnaissance frameworks. By utilizing pre-prepared models, move learning upgrades the exactness and productivity of group reconnaissance by adjusting existing information to identify explicit articles or exercises of interest inside observation recordings.

1.1 OVERVIEW

Crowd surveillance system frameworks are utilized to screen and track enormous gatherings in broad daylight spaces like roads, arenas, air terminals, and malls with the blend of advancements like cameras, sensors, and examination programming to distinguish and follow people or gatherings. Sensors can recognize different sorts of information like sound, temperature, and movement. Information capacity is important to store a lot of information gathered by the framework, and examination programming is utilized to process and investigate the information to distinguish examples and oddities.

Transfer learning, a powerful technique in computer vision, allows for the transfer of knowledge gained from pre-trained models to improve the performance and offers significant potential to address the challenges posed by complex and dynamic crowd behaviors. The revolution in crowd surveillance may entail by leveraging transfer learning in order to empower surveillance systems. By utilizing this model on large-scale image datasets, valuable features and representations can be extracted from visual data, enabling more accurate detection, tracking, and analysis of individuals within crowded scenes. The transfer learning allows the system to adapt to different surveillance scenarios and effectively handle challenges such as occlusions, crowd density variations, and sudden behavior changes.

1.2 BACKGROUND

According to a 2014 report by the U.N 54% of the all out people as of now lives in metropolitan districts with this dare to augmentation to 66% by 2050. This looks at to an improvement in an overall metropolitan people of 2.5 billion people all through the accompanying 32 years. With this fast extension in the metropolitan people, significantly stopped up gatherings will transform into a basic piece of everyday presence, acquainting immense troubles with the help of public security and the useful improvement of people in current metropolitan networks. Reliably numerous people are hurt or killed in thickly populated metropolitan locales in light of charges and squashes. Another outline of this was the 2014 New Year's Eve charge in Shanghai, China where 36 people, tragically, passed on. This loss of life could really be prevented with better assessment and understanding of gathering behavior and stop up levels across colossal metropolitan locales. Our country is the sixth most-transcontinental country on earth with a general population outperforming 230,639,535 people [65]. The yearly improvement speed of Pakistan is 1.91% [66]. On account of this tremendous number of people abiding in Pakistan, there are by and large immense gatherings at public spots as well as open, political, and definitive parties. Such places are presumably going to make lamentable results assuming there ought to emerge an event of sudden episodes like disasters and dread based oppressor attacks.

In recent times, the field of crowd surveillance has been undergoing significant advancements in enhancing surveillance systems through the application of transfer learning techniques. Traditional surveillance systems face challenges in effectively monitoring and analyzing crowded scenes, limiting their accuracy and efficiency. However, with the emergence of transfer learning, which leverages pre-trained models and knowledge from one domain to improve performance in related domains, crowd surveillance systems can benefit from the wealth of knowledge gained from large-scale datasets and complex visual features. By empowering surveillance systems with transfer learning, the aim is to revolutionize crowd surveillance, enabling improved detection and tracking of individuals, identification of abnormal behaviors, prediction of potential threats or emergencies, and real-time situational awareness in crowded environments.

1.3 AN OVERVIEW OF VISUAL SURVEILLANCE

Traditional visual surveillance systems are comprised of a number of modules, each performing a unique function. A highly generalized visual surveillance system is depicted in Figure 1.1. This surveillance network consists

of multiple cameras and a sequence of processing steps which are classified broadly into the following categories:

- Environment modelling.
- Feature extraction and/or Object detection.
- Higher-level event and behavior understanding.
- Fusion of information from multiple cameras.

The information from each camera is fused to acquire operationally meaningful data. The terminology used here is very general: a surveillance system may be designed to track the individual body parts of a single person; or to track the position of multiple individuals in a scene; or to measure the holistic properties of a crowd. Each of the stages in this framework are described briefly in the following sections.

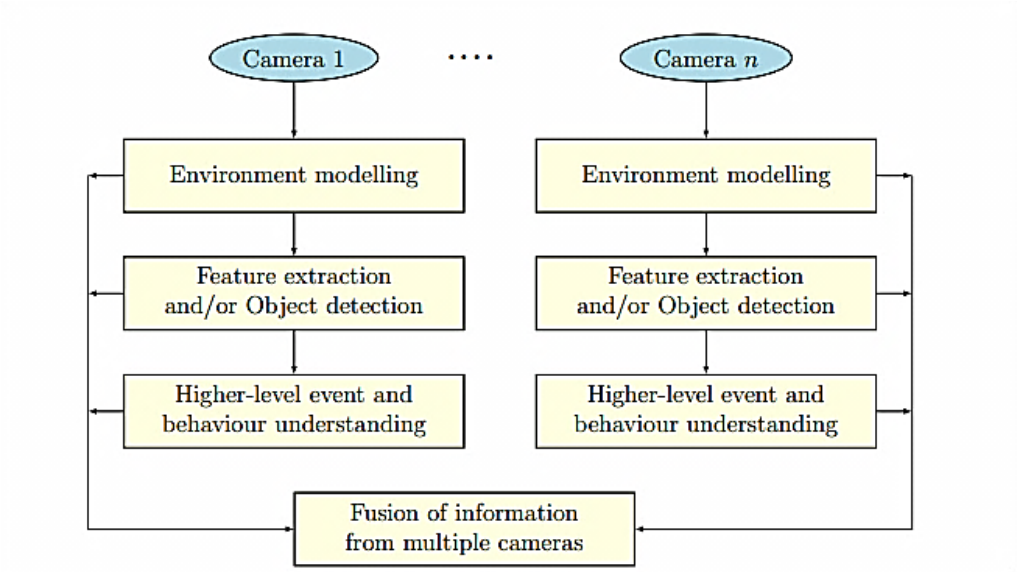


Figure 1.1: General Visual Surveillance Framework

1.3.1 Environment Modelling

Environment modeling is a crucial aspect of various fields, including computer graphics, virtual reality, simulation, and robotics. It involves creating digital representations of physical or virtual environments to simulate their behavior and interactions. By accurately capturing the geometry, appearance, and dynamics of the environment, modeling allows for realistic rendering, analysis, and simulation, enabling applications ranging from immersive experiences to architectural design and urban planning. The advancements in

environment modeling techniques have greatly contributed to creating more immersive and interactive virtual environments and improving our understanding of real-world phenomena.

1.3.2 Feature Extraction and/or Object Detection

Feature extraction refers to the process of transforming pixels in an image into meaningful descriptors through low-level image processing. These extracted low-level features from surveillance videos are then used for higher-level processing and comprehension. The specific components of this module may vary depending on the intended application, and they may encompass a range of techniques such as:

1.3.2.1 Motion Segmentation

Motion segmentation serves as the initial processing stage in numerous surveillance applications, aiming to separate foreground objects from the background. The widely employed approach for this task is the Stauffer Grimson background subtraction method [42].

1.3.2.2 Feature Extraction

Image features play a crucial role in extracting meaningful descriptors from an image, which can then be processed at a higher level. For instance, holistic image features like texture have been applied in crowd density analysis [43, 44], enabling the understanding of crowd patterns and behavior. On the other hand, local image features extracted from 'spatio-temporal patches' have been utilized in anomaly detection [45, 46, 47], allowing the identification of abnormal events or behaviors within a surveillance scene. These approaches demonstrate the importance of feature extraction in leveraging specific characteristics of an image for different analysis tasks.

1.3.2.3 Object Classification

When the underlying object in a foreground segment is unknown, these segments are often referred to as 'blobs'. Object classification aims to assign each segment to a specific object class, such as humans or cars. Shape features, including the blob area or aspect ratio of the bounding box [48], as well as

motion features like periodicity [49] and rigidity [50], are commonly used for this purpose.

1.3.2.4 Person Detection

Popular approaches for detecting humans in images include Dalal's Histogram of Oriented Gradients (HOG) [51] and Felzenszwalb's part-based models [52, 53]. Additionally, Viola's cascade of simple image features [54] is commonly used for face detection.

1.3.2.5 Trajectory Extraction

To enable subsequent analysis, objects are tracked over time, and their trajectories are captured. Existing tracking research is extensively reviewed in [55]. Moreover, the trajectories of image key points can be extracted using techniques like the Kanade-Lucas-Tomasi (KLT) feature tracker [56, 57].

1.3.3 Higher-Level Event and Behavior Understanding

Higher level understanding refers to providing information that is meaningful and useful to the end user, typically a human operator. To accomplish this, the system leverages ML tools to perform its designated task by analyzing the low-level features extracted in earlier modules. The target application can encompass a wide range of tasks, including but not limited to:

1.3.3.1 Behavior Understanding

The application of transfer learning in Crowd Surveillance Systems enables the system to leverage pre-existing knowledge from related tasks, enhancing its ability to understand and interpret complex behaviors exhibited by individuals within a crowd, leading to improved threat detection and crowd management. Tracking results are used to analyses trajectories and recognize behavior of individuals [58, 59, 60].

1.3.3.2 Personal Identification

By employing transfer learning techniques, Crowd Surveillance Systems can leverage pre-trained models and knowledge from related domains to

enhance personal identification capabilities, enabling accurate and efficient recognition of individuals within crowded environments, thereby improving security and public safety. Height, facial appearance and walking gait are the main biometric features used for personal identification [61].

1.3.3.3 Crowd Monitoring

The focus of this thesis is on monitoring crowded environments, which includes estimation of crowd size [62,63], crowd flow analysis [66, 67] and anomaly detection [64,65].

ML techniques such as support vector machines (SVM), neural networks (NN), hidden Markov models (HMM) and Gaussian mixture models (GMM) are commonly used for classification and regression.

1.3.4 Fusion of Information from Multiple Cameras

Fusion of information from multiple cameras is aided by the use of camera calibration as described in Section 1.3.1 which enables real-world positions to be mapped across viewpoints. Correspondence between objects from multiple viewpoints can improve the system's understanding of the scene. It may be necessary to account for blind spots or overlaps between the viewpoints.

1.4 MOTION DETECTION

This section provides an overview of the fundamental visual surveillance tasks that form the basis of existing research before delving into the review of crowd monitoring literature. One essential component in most visual surveillance systems is motion detection [90]. This initial step enables the execution of subsequent tasks, including object classification and crowd counting.

1.5 CROWD COUNTING

Crowd counting is a prominent research area that focuses on accurately estimating the number of individuals present in a crowded scene. With the rapid growth of urban areas and the increasing need for public safety and resource allocation, crowd counting plays a crucial role in various applications, including

crowd management, event planning, and urban infrastructure design. Traditional crowd counting methods relied on manual counting or simple computer vision techniques, but recent advancements in deep learning and computer vision have revolutionized this field. Deep learning-based crowd counting algorithms leverage Convolutional Neural Networks (CNNs) to extract intricate features and learn complex patterns from crowd images or videos. By training on large-scale annotated datasets, these models can effectively estimate crowd densities and counts with improved accuracy.

1.6 ANOMALY DETECTION

Anomaly detection in Crowd Surveillance Systems is enhanced through the application of transfer learning techniques. By leveraging knowledge from related tasks or domains, transfer learning enables the system to learn and recognize normal patterns of behavior within a crowd. This learned knowledge is then used to detect and identify anomalies or suspicious activities that deviate from the expected behavior. By incorporating transfer learning, the Crowd Surveillance System becomes more adept at detecting and alerting security personnel to potential threats, enhancing overall surveillance effectiveness and public safety.

1.7 ML INTRODUCTION

ML is a piece of PC put together insight that bright lights with respect to the improvement of estimations and models prepared for acquiring from data and making assumptions or decisions without unequivocal programming. It incorporates the use of quantifiable methodology and computational capacity to engage laptops to normally recognize models and associations inside colossal datasets. ML estimations can be broadly characterized into three sorts: directed learning, independent learning, and backing learning. Coordinated learning incorporates planning models on checked data to make definite assumptions on new, covered models. Solo learning computations, on the other hand, plan to reveal hidden away models and plans in unlabeled data. Support learning utilizes an expert helping out an environment, learning ideal exercises through a game plan of compensations and disciplines. ML has changed various endeavors, including clinical consideration, cash, and advancement, by engaging advanced data assessment, robotization, and perceptive capacities.

1.7.1 Types of ML

1.7.1.1 Supervised Learning

In this type, models are trained on labeled data, where the input data is accompanied by the correct output. The goal is to learn a mapping function that can accurately predict outputs for new, unseen inputs.

1.7.1.2 Unsupervised Learning

Here, models gain from unlabeled information, with no predefined yield names. The goal is to find examples, designs, and connections inside the information without explicit direction. Grouping and dimensionality decrease are normal assignments in solo learning.

1.7.1.3 Reinforcement Learning

This type involves an agent that interacts with an environment and learns through a system of rewards and penalties. The agent aims to maximize cumulative rewards by taking optimal actions in different situations. Reinforcement learning is commonly used in robotics, game playing, and autonomous systems.

1.7.2 Transfer Learning

Transfer learning is a strong strategy in PC vision that permits a pre-prepared model to be utilized as a beginning stage for another PC vision task. Rather than beginning without any preparation, move learning includes utilizing the elements advanced by a pre-prepared model as a beginning stage for another model, which can then be tweaked for the particular job needing to be done. There are multiple ways of performing move learning with PC vision calculations. One famous methodology is to utilize a pre-prepared (CNN), like VGG, ResNet, or Beginning, as an element extractor. In this methodology, the pre-prepared CNN is utilized to separate elements from the information pictures, which are then taken care of into another classifier or regressor. Another methodology is to calibrate the pre-prepared CNN for the particular main job. This includes freezing a portion of the layers of the pre-prepared CNN and preparing just the excess layers on the new dataset. Thusly, the pre-prepared CNN can adjust its learned highlights to the new dataset, while as yet holding the important information it has acquired from the first dataset. Move learning

can be especially valuable in circumstances where there is restricted preparation information accessible for another undertaking. By beginning with a pre-prepared model, the requirement for a lot of marked preparing information can be decreased, which can save time and assets. Moreover, move learning can assist with working on the exhibition of the new model, particularly in the event that the pre-prepared model was prepared on a huge and different dataset. Move learning is a strong method for PC vision that can assist with working on the exhibition of new models while decreasing the requirement for a lot of preparing information.1.7.3 Advantages of Transfer Learning in Crowd Surveillance System

Faster Training: Transfer learning allows for faster training of deep learning models for crowd surveillance tasks, as pre-trained models can be used as a starting point for training on a smaller dataset. This reduces the amount of time and resources required for training, allowing for faster deployment of the surveillance system.

1.7.3.1 Improved Accuracy

Pre-trained models that have been trained on large datasets such as ImageNet or COCO have already learned to recognize general visual patterns and features that are useful for many computer vision tasks, including crowd surveillance. By fine-tuning these pre-trained models on a smaller dataset of surveillance videos, the accuracy of the surveillance system can be significantly improved.

1.7.3.2 Robustness

Pre-trained models are generally more robust to variations in lighting, background, and other factors that can affect the performance of deep learning models. By using a pre-trained model as a starting point for training on surveillance videos, the resulting model is likely to be more robust and able to handle the challenging conditions that are common in real-world surveillance scenarios.

1.7.3.3 Adaptability

Transfer learning allows for the adaptation of pre-trained models to specific surveillance scenarios and tasks. For example, a pre-trained model can be fine-tuned to recognize specific behaviors, such as loitering or fighting, that

are of interest to the user. This makes the surveillance system more adaptable and flexible, allowing it to be customized to specific needs and requirements.

1.8 TRANSFER LEARNING BASED DEEP LEARNING ALGORITHM

1.8.1 Convolutional Neural Networks

CNNs are a kind of brain network that are broadly utilized in picture and video acknowledgment undertakings. They were first presented by Yann LeCun in 1998, and have since turned into the standard methodology for picture arrangement and other PC vision undertakings. CNNs depend on the idea of convolution, which is a numerical activity that consolidates two capabilities to deliver a third capability. On account of CNNs, convolution is utilized to extricate highlights from pictures. The convolution activity is applied to the picture utilizing a bunch of learnable channels, which are otherwise called portions or loads. Each channel is utilized to extricate a particular element from the picture, like edges, corners, or different examples.

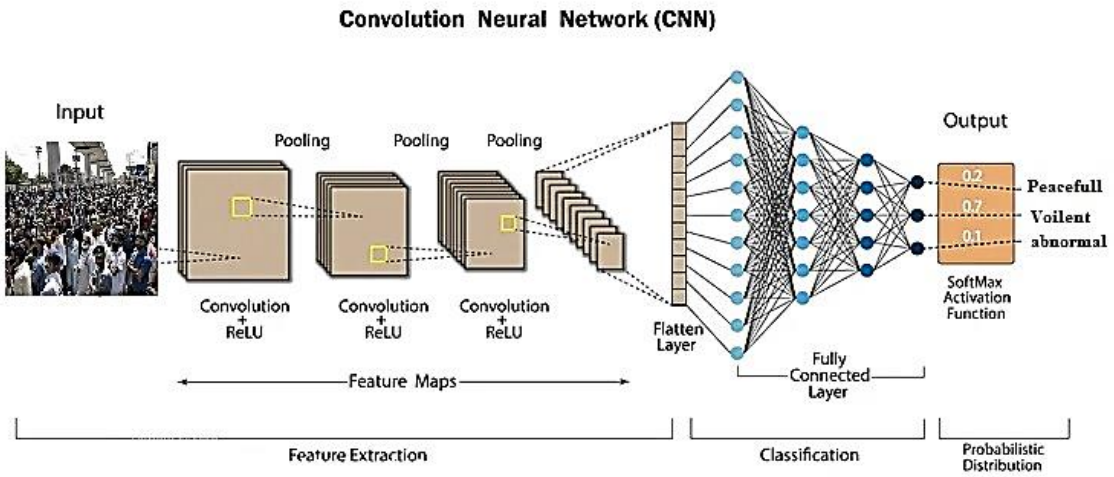


Figure 1.2: Convolutions Neural Network

In a Figure 1.2 the convolutional layers are followed by one or more fully connected layers, which are used to classify the image. During training, the CNN learns the optimal filters and weights by adjusting them based on the difference between the predicted output and the actual output. CNNs have been used in a wide range of applications, including image classification, object detection, facial recognition, and natural language processing. They have been shown to outperform traditional ML algorithms in many of these tasks, and are often used

in conjunction with other deep learning techniques such as RNNs and attention mechanisms.

1.8.2 Recurrent Neural Networks

RNNs are a class of brain networks that are intended to deal with consecutive information, for example, time series or regular language sentences, by utilizing input associations with permit data to persevere across various time steps. The essential design of a RNN comprises of a secret state vector that gets refreshed at each time step, as well as an info and a result vector. The info vector is taken care of into the organization at each time step, and the result vector is produced by applying a non-direct capability to the ongoing secret state vector. One of the critical benefits of RNNs is their capacity to display transient conditions in information. This makes them appropriate for various assignments, including discourse acknowledgment, machine interpretation, and text age. Notwithstanding, one test with standard RNNs is that they can experience the ill effects of the disappearing angle issue, where slopes can turn out to be tiny as they are back propagated through time, prompting slow union and trouble learning long haul conditions. To resolve this issue, different kinds of RNNs have been proposed, like Long Momentary Memory (LSTM) organizations and Gated Intermittent Units (GRUs), which integrate specific gating instruments to more readily control the progression of data through the organization.

1.8.3 Resnet

ResNet, short for Remaining Brain Organization, is a profound learning design that was presented by Kaiming He et al. in 2015. It is one of the most powerful and broadly involved CNN models for PC vision undertakings like picture arrangement, object discovery, and picture division. The vital thought behind ResNet is the utilization of leftover associations, otherwise called skip associations, to resolve the issue of evaporating slopes in profound brain organizations. Profound organizations frequently experience the ill effects of debasement, where the exactness of the organization diminishes as the organization profundity increments. ResNet resolves this issue by presenting alternate way associations that permit data to sidestep a few layers, empowering the organization to learn remaining mappings.

1.8.4 VGG-16

VGG16 is a CNN design that acquired critical prevalence because of its straightforwardness and solid execution in picture grouping undertakings. Created by the Visual Calculation Gathering (VGG) at the College of Oxford, VGG16 has a profound construction with a sum of 16 layers, including 13 convolutional layers and 3 completely associated layers.

The convolutional layers in VGG16 consist of multiple 3x3 filters, which are followed by rectified linear unit (ReLU) activation functions. These layers are organized into five blocks, with each block containing a varying number of convolutional layers. The number of filters in the convolutional layers increases deeper into the network, starting from 64 filters and doubling after each block (64, 128, 256, 512, 512).

After each block of convolutional layers, VGG16 applies max pooling with a 2x2 filter and stride 2, effectively reducing the spatial dimensions of the feature maps while retaining the most important features. This down sampling process helps in capturing hierarchical and translation-invariant representations.

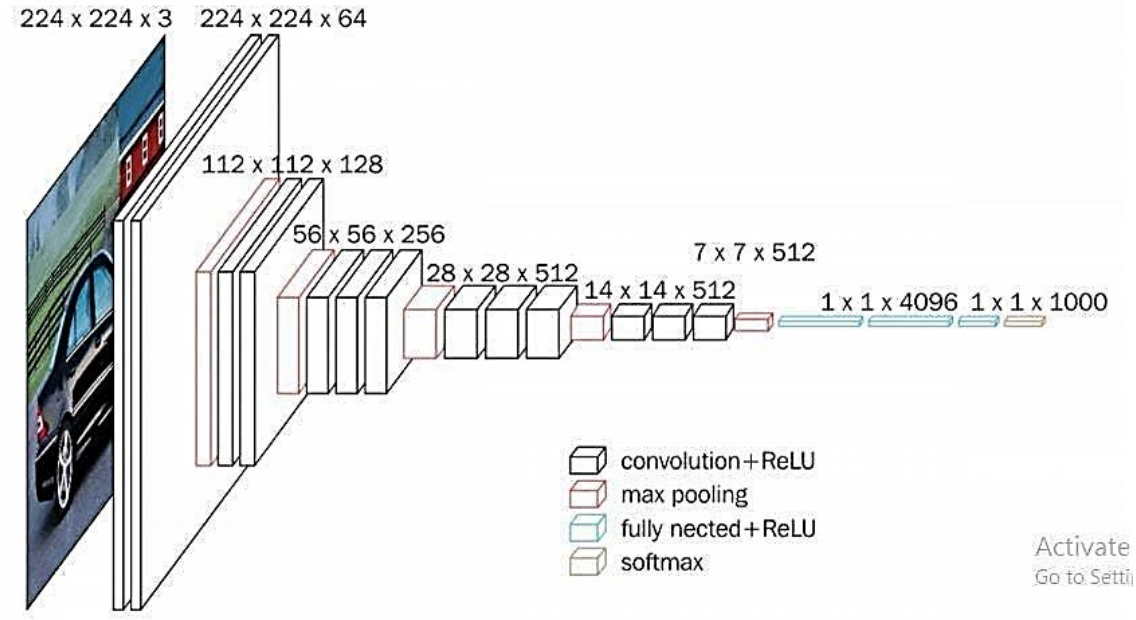


Figure 1.5: VGG 16

In the following Fig 1.5 convolutional layers, the output is flattened into a vector and passed through three fully connected layers. The first two fully connected layers have 4,096 neurons each, while the final fully connected layer serves as the output layer with 1,000 neurons representing the predicted probabilities for 1,000 different classes in the ImageNet dataset.

1.9 CROWD DENSITY ESTIMATION

Crowd density estimation is an important task in crowd surveillance that involves the estimation of the number of people present in a given area. Transfer learning is a technique that allows us to leverage knowledge learned from one domain to another domain. In the context of crowd density estimation, transfer learning can be used to improve the performance of crowd density estimation models by leveraging pre-trained models or data from related tasks. One approach to crowd density estimation using transfer learning is to use pre-trained CNNs for image classification. The pre-trained CNNs can be fine-tuned on the crowd density estimation task by training the last few layers of the network on the crowd density estimation dataset. This approach has been shown to be effective in improving the accuracy of crowd density estimation models. Another approach is to use transfer learning to learn features from related tasks that can be used for crowd density estimation. For example, features learned from object detection or semantic segmentation tasks can be used as input to crowd density estimation models. This approach has been shown to be effective in improving the accuracy of crowd density estimation models. Transfer learning can be used to improve the performance of crowd density estimation models by leveraging pre-trained models or data from related tasks. This approach can help overcome the challenges of limited training data and improve the accuracy of crowd density estimation models.

1.10 CROWD BEHAVIOR ANALYSIS

Crowd behavior analysis is a fundamental part of move learning-based swarm reconnaissance frameworks. Move gaining includes utilizing information gained starting with one space then onto the next area, and with regards to swarm reconnaissance, this includes utilizing pre-prepared models and datasets to distinguish and examine swarm conduct continuously. To perform swarm conduct investigation, the framework regularly utilizes a mix of PC vision and ML procedures. The framework initially identifies and tracks people in the group utilizing object identification calculations. Then, the framework applies ML calculations to break down the development examples and cooperation between people in the group. This investigation can give significant experiences into the way of behaving of the group, for example, recognizing expected episodes, assessing swarm thickness, and anticipating swarm developments. One of the essential difficulties in move learning-based swarm reconnaissance frameworks is adjusting the pre-prepared models to the particular setting of the observation climate. This includes adjusting the pre-prepared models on information from the reconnaissance climate to work on their exactness and execution. Swarm conduct examination is a urgent part of

move learning-based swarm reconnaissance frameworks. By utilizing pre-prepared models and datasets, these frameworks can recognize and examine swarm conduct progressively, giving significant experiences into the way of behaving of the group and working on generally security and wellbeing out in the open spaces.

1.11 CROWD FLOW ANALYSIS

Crowd flow analysis in transfer learning-based crowd surveillance technique that includes utilizing profound learning calculations to break down the development examples of groups. This examination assists with identifying peculiarities and possibly risky circumstances progressively. Move learning is utilized to work on the exactness of the investigation by utilizing the pre-prepared models on an enormous dataset and retraining them on the objective dataset with restricted marked information. In swarm stream examination, the development examples of people are dissected utilizing PC vision procedures, like article identification and following. This includes identifying and following people in a group as they move starting with one area then onto the next. The information gathered can then be investigated to decide the progression of the group, including the heading and speed of development. By breaking down the stream designs, the framework can distinguish areas of clog or possible perils. Move learning is utilized to work on the exactness of the examination by utilizing the pre-prepared models on an enormous dataset and retraining them on the objective dataset with restricted named information. This strategy is especially valuable in circumstances where there is restricted marked information accessible for preparing the model. By utilizing pre-prepared models, the framework can exploit the immense measures of marked information that have proactively been gathered and use it to work on the exactness of the investigation on a new, more modest dataset.

1.12 LIMITATION OF CURRENT CROWD SURVEILLANCE SYSTEM

Crowd surveillance systems typically involve using cameras and computer vision algorithms to analyze the behavior and movements of large groups of people. Transfer learning is a popular approach in crowd surveillance, where a pre-trained model is fine-tuned on a new dataset to improve its accuracy on a specific task. However, there are several limitations of current crowd surveillance systems in using transfer learning. Lack of diversity in training data: Transfer learning relies on having a large and diverse dataset for pre-training. However, many crowd surveillance datasets are limited in terms of the

variety of scenes, lighting conditions, and crowd sizes. This can lead to overfitting on specific scenarios and poor generalization to new ones.

1.12.1 Difficulty in Domain Adaptation

Transfer learning also requires the source and target domains to have some level of similarity. However, crowd surveillance systems may need to operate in different environments, such as indoor and outdoor settings, which can differ significantly in terms of lighting, weather, and background.

1.12.1.1 Privacy Concerns

Crowd surveillance systems can raise privacy concerns as they involve collecting and analyzing sensitive information about individuals. Fine-tuning pre-trained models on new datasets may exacerbate these concerns, especially if the pre-training dataset includes sensitive information.

1.12.1.2 Limited Interpretability

Transfer learning models can be difficult to interpret, especially when fine-tuned on new datasets. This can make it challenging to identify and troubleshoot errors or biases in the system.

1.12.1.3 Ethical Considerations

Crowd surveillance systems may be used for purposes that raise ethical concerns, such as monitoring political protests or targeting specific groups based on race, ethnicity, or religion. Transfer learning models may amplify these biases if the pre-training data is not diverse and inclusive. While transfer learning has shown promise in improving the accuracy of crowd surveillance systems, these limitations need to be carefully considered and addressed to ensure their safe and responsible use.

1.12.1.4 Importance of Crowd Surveillance System in Public Safety

Crowd surveillance systems are turning out to be progressively significant in open security, particularly in regions where enormous groups accumulate like public occasions, transportation center points, and downtown areas. The

utilization of group reconnaissance frameworks can assist with recognizing potential security dangers and forestall crime before it happens. Move realizing, which includes utilizing pre-prepared models to work on the exactness of another model, can be especially helpful in creating successful group reconnaissance frameworks. This is on the grounds that move learning takes into account the exchange of information and mastery from prior models to new models, accordingly working on their precision and diminishing how much preparation information required. With regards to swarm observation frameworks, move learning can be utilized to prepare models on previous informational indexes of group conduct and afterward calibrate these models on new informational indexes intended for the area and occasion being checked. This can altogether work on the exactness of the models and empower them to more readily distinguish and recognize potential security dangers. In addition, swarm observation frameworks can give ongoing data to policing and security faculty, permitting them to answer rapidly to any security dangers that might emerge. This can assist with forestalling crime and work on open security.

1.13 MOTIVATION

The inspiration driving creating is to work on open wellbeing and security in jam-packed public spaces. Conventional observation frameworks frequently depend on manual checking, which can be work escalated and inclined to mistakes. Profound learning methods, for example, move learning, have shown huge headway in object discovery and acknowledgment, making them reasonable for swarm observation applications. By utilizing pre-prepared models, the proposed framework can rapidly and precisely distinguish and follow people in a group, recognize possible dangers, and ready security work force progressively. Additionally, the proposed framework can address the constraints of conventional observation frameworks, for example, low exactness, high misleading problem rates, and restricted adaptability. The framework can likewise adjust to new and changing conditions by calibrating the pre-prepared models on new information, making it an adaptable and savvy answer for public wellbeing and security. The proposed Move Learning Based Group Observation Framework (TLBCSS) can give a significant device to working on open wellbeing and security in packed public spaces, by giving continuous checking, examination, and cautions to expected dangers.

1.14 PROBLEM STATEMENT

The ongoing issue is the absence of an exact, proficient, and protection consistent group reconnaissance framework that can distinguish and follow

people in a jam-packed climate, recognize expected dangers, and ready security faculty progressively. Conventional observation frameworks frequently depend on manual checking, which can be work serious, tedious, and inclined to mistakes. Moreover, existing profound learning models require a lot of named information to accomplish high exactness, making it trying to foster models for swarm observation applications. In addition, security concerns and possible infringement of individual privileges should be tended to while growing such frameworks. In this way, the issue proclamation is to create and carry out an exchange learning-based swarm observation framework that can use pre-prepared models, calibrate them on new information, and adjust to evolving conditions. The framework ought to be exact, proficient, protection agreeable, and keep away from victimization any gatherings in view of race, orientation, or religion. Also, the framework ought to be coordinated into existing reconnaissance frameworks to give continuous checking, examination, and alarms to expected dangers. At last, the proposed framework ought to be assessed and approved utilizing public datasets and genuine situations to guarantee its adequacy in working on open wellbeing and security in packed public spaces.

1.15 RESEARCH QUESTIONS

Based on the research gap and the objectives of the proposed TLBCSS, the following research questions can be formulated:

1. What are the main challenges and limitations of using transfer learning for crowd surveillance systems?

2. How to design a system while providing effective performance in crowd surveillance using transfer learning?

3. How transfer learning can be helpful in improving the scalability and real-time performance of crowd surveillance systems?

4. How can transfer learning be used to address the challenges of cross-dataset generalization in peaceful and unpeaceful crowd scene classification?

These research questions will help to address the research gap and achieve the objectives of the proposed TLBCSS.

1.16 RESEARCH OBJECTIVES

The point of the TLBCSS is to foster a continuous observing and examination device that can distinguish and follow people in a jam-packed climate, recognize likely dangers, and ready security work force progressively.

The objectives of the proposed system include:

1. To identify and analyze the main challenges and limitations associated with using transfer learning in crowd surveillance systems and provide insights into potential solutions.

2. To design a system for crowd surveillance that utilizes transfer learning and achieves effective performance, and develop a framework for integrating these techniques into the system.

3. To investigate the potential of transfer learning in improving the scalability and real-time performance of crowd surveillance systems, and explore ways to optimize the use of transfer learning in this context.

4. To address the challenges of cross-dataset generalization in peaceful and unpeaceful crowd scene classification, and evaluate their effectiveness in addressing this challenge.

1.17 THESIS ORGANIZATION

The proposal has been partitioned into five sections: part 1 presents an outline of the Engaging Observation Framework utilizing Move Learning issue and gives a prologue to the critical advancements in this space, section 2 presents different fundamental terms and hypothetical ideas that are pertinent to this expert postulation, section 3 portrays the strategy used to foresee swarm conduct analysiss, part 4 examinations the outcomes acquired from the execution, talks about the task result in light of the exploration inquiries in segment 1.4 and looks at the meaning of this expert proposition, and section 5 presents the end for the expert proposal and looks at what's in store work.

CHAPTER 2

LITERATURE REVIEW

Most researchers have extensively worked on Crowd Surveillance Systems using Transfer Learning, which contributes to a comprehensive understanding of the research and advancements in this field. The study extensively investigates the implementation of transfer learning techniques for augmenting the capabilities of surveillance systems in comprehending crowd behavior, identifying individuals, and detecting anomalies. Researchers have explored diverse transfer learning approaches, encompassing fine-tuning pre-trained models, feature extraction, and domain adaptation, with the aim of enhancing the performance of Crowd Surveillance Systems. These techniques harness knowledge acquired from related domains like object recognition, human pose estimation, and action recognition, to facilitate the comprehension of intricate crowd behavior. The literature review scrutinizes the efficacy of different transfer learning architectures, including CNNs, RNNs, and hybrid models, within the context of crowd surveillance. It also investigates the influence of various pre-training datasets, transfer learning strategies, and fine-tuning techniques on the system's overall performance.

Sharma et al. [1] proposed the Scale-mindful CNN for swarm thickness assessment and group conduct examination is an exploration paper that proposes another CNN design for precisely assessing swarm thickness and breaking down swarm conduct in pictures and recordings. The proposed model is fit for taking care of various sizes of groups and can assess the thickness of groups in complex scenes with high precision. The model includes three principal parts: a multi-scale highlight extractor, a thickness map assessor, and a conduct classifier. The multi-scale highlight extractor separates include from the information picture or video at various scales, which assists the model with recognizing hordes of various sizes. The thickness map assessor then produces a thickness map that shows the group thickness in every locale of the information picture. At last, the conduct classifier breaks down the thickness map and predicts the way of behaving of the group, like typical, blocked, or scattered. The proposed model is assessed on four benchmark datasets, and the outcomes show that it beats cutting edge models regarding swarm thickness assessment and group conduct examination. The creators recommend that the proposed model can be utilized in different applications, like group the board, metropolitan preparation, and public wellbeing. The Scale-mindful CNN for swarm thickness assessment and group conduct examination is a critical commitment to the field of PC vision,

giving a vigorous answer for swarm investigation in complex scenes with shifting group sizes.

Alharthi et al. [2] have as of late proposed the usage of C3D (Convolutional 3D) brain organizations to perceive unusual ways of behaving in enormous groups. C3D is a sort of profound gaining design that can separate spatio-fleeting elements from recordings. The proposed approach includes first gathering video film of an enormous group, then, at that point, portioning and naming people in the video outlines utilizing an item identification calculation. Then, the C3D model is prepared on the fragmented and marked recordings to perceive ordinary and strange ways of behaving. The unusual way of behaving can incorporate activities like battling, pushing, or rushing. The exploration shows promising outcomes in precisely perceiving strange ways of behaving in enormous groups, which can have huge applications in open wellbeing and security. The methodology can likewise be stretched out to different regions, like games examination or modern wellbeing, where it is significant to perceive unusual way of behaving.

In their examination, Danista et al. [3] presented a gadget free group counting approach called Cross Count. This approach use move learning strategies to accomplish uncommon exactness in counting the quantity of people inside an assigned region, without the requirement for any gadgets or sensors. The framework use existing pre-prepared brain network models, like VGG16 and Commencement V3, and adjusts them for the undertaking of group counting. This approach decreases the requirement for huge datasets and broad preparation time, making it a proficient and versatile arrangement. To accomplish exact group counting, Cross Count utilizes a method called cross-scene counting. This includes preparing the model on a source scene with named information, and afterward moving the learned information to an objective scene without marks. The model figures out how to distinguish visual elements of individuals and include them precisely in the objective scene. The creators assessed Cross Depend on two public datasets, and the outcomes show that it beats existing cutting edge techniques. The framework accomplished an exactness of 94.4% on the UCF-QNRF dataset and a precision of 90.2% on the NWPU-Group dataset. Also, Cross Include showed power to changes in lighting conditions and perspective changes. Cross Count is an effective and exact gadget free group counting approach that can be utilized in different applications, like group the executives, reconnaissance, and metropolitan preparation.

In their work, Khosro et al. [4] set forward a strategy that consolidates profound exchange learning with Web of Robots (IoD) advancements to perceive strange ways of behaving in packed scenes. This inventive

methodology plans to upgrade the exactness and productivity of unusual conduct acknowledgment by utilizing the force of both profound exchange learning procedures and IoD advances. The creators contend that conventional strategies for strange conduct acknowledgment in jam-packed scenes, like those in view of carefully assembled highlights or shallow learning calculations, are restricted in their capacity to deal with the intricacy and fluctuation of certifiable situations. To address these limits, they propose a technique that use the force of profound exchange learning and IoD innovations. The proposed strategy comprises of three phases: information assortment, move learning, and unusual conduct acknowledgment. In the information assortment stage, the creators gather a huge dataset of recordings of swarmed scenes utilizing an IoD framework. In the exchange learning stage, they utilize a pre-prepared profound brain organization (DNN) to remove significant elements from the video information and move the learned information to another DNN for strange conduct acknowledgment. In the unusual conduct acknowledgment stage, the creators utilize the new DNN to arrange the ways of behaving in the recordings as one or the other typical or strange. The creators assess the presentation of their technique utilizing a few measurements, including exactness, accuracy, review, and F1-score, on an openly accessible dataset. The outcomes show that their strategy outflanks a few cutting edge techniques for strange conduct acknowledgment in packed scenes. The paper shows the capability of profound exchange learning and IoD advances for tending to the difficulties of unusual conduct acknowledgment in jam-packed scenes and has suggestions for the improvement of shrewd observation frameworks.

In their distribution [5], Ali et al. presented a strategy for swarm counting that uses profound exchange learning. The creators contend that traditional ways to deal with swarm counting, which depend on carefully assembled highlights or shallow learning calculations, are compelled in their ability to deal with the unpredictability and variety present in genuine situations. In this manner, the proposed strategy consolidates profound exchange learning methods to beat these restrictions and work on the exactness and strength of group including in testing conditions. To address these limits, they propose a technique that use the force of profound exchange learning. The proposed strategy comprises of two phases: highlight extraction and move learning. In the component extraction stage, the creators utilize a pre-prepared profound CNN to separate important highlights from the info picture. In the exchange learning stage, they utilize another CNN to gain proficiency with the planning between the extricated highlights and the group count. The creators assess the presentation of their strategy utilizing a few measurements, including mean outright mistake (MAE) and mean squared blunder (MSE), on a few freely accessible datasets. The outcomes show that their strategy beats a few cutting edge techniques for swarm counting. The paper shows the capability of profound exchange learning for

tending to the difficulties of group counting and has suggestions for the advancement of canny observation frameworks and different applications that require exact group counting.

In this exploration Alrowais et al. [6] proposed a strategy for swarm thickness investigation utilizing profound exchange learning and item recognition. The creators contend that conventional techniques for swarm thickness examination, like those in light of hand tailored highlights or shallow learning calculations, are restricted in their capacity to deal with the intricacy and fluctuation of certifiable situations. To address these constraints, they propose a technique that use the force of profound exchange learning and item recognition. The proposed technique comprises of three phases: information assortment, move learning, and item discovery. In the information assortment stage, the creators gather an enormous dataset of recordings of swarmed scenes utilizing a video reconnaissance framework. In the exchange learning stage, they utilize a pre-prepared profound CNN to separate significant elements from the video information and move the learned information to another CNN for object location. In the item discovery stage, the creators utilize the new CNN to recognize and include the quantity of individuals in the scene. The creators assess the exhibition of their technique utilizing a few measurements, including accuracy, review, and F1-score, on an openly accessible dataset. The outcomes show that their technique beats a few cutting edge strategies for swarm thickness investigation. The paper exhibits the capability of profound exchange learning and article location for tending to the difficulties of group thickness examination in video reconnaissance frameworks and has suggestions for the advancement of shrewd observation frameworks.

Alafif et al. [7] proposed a strategy for unusual conduct location in packed situations utilizing generative ill-disposed networks (GANs). The creators contend that conventional strategies for unusual conduct recognition, like those in light of handmade highlights or shallow learning calculations, are restricted in their capacity to deal with the intricacy and changeability of certifiable situations. To address these impediments, they propose a strategy that use the force of GANs. The proposed strategy comprises of two phases: preparing and testing. In the preparation stage, the creators utilize a GAN to gain proficiency with the conveyance of typical conduct in jam-packed situations utilizing an enormous dataset of recordings caught during the Hajj journey. In the testing stage, they utilize the GAN to recognize unusual conduct in new recordings of swarmed situations. The creators assess the presentation of their strategy utilizing a few measurements, including accuracy, review, and F1-score, on an openly accessible dataset of recordings from the Hajj journey. The outcomes show that their strategy beats a few cutting edge techniques for strange conduct location in jam-packed situations. In the paper shows the capability of GANs

for tending to the difficulties of unusual conduct location in jam-packed situations and has suggestions for the advancement of canny observation frameworks for public wellbeing and security.

As indicated by Ahmed et al. [8] proposed a strategy for swarm observing utilizing IoT sensors and profound learning methods. The creators contend that customary strategies for swarm observing, like those in light of manual counting or video examination, are restricted in their capacity to deal with enormous and complex groups. To address these impediments, they propose a strategy that use the force of IoT sensors and profound learning procedures. The proposed strategy comprises of three phases: information assortment, move learning, and article recognition. In the information assortment stage, the creators use IoT sensors to gather information on swarm thickness and development. In the exchange learning stage, they utilize a pre-prepared profound CNN to extricate significant highlights from the information and move the learned information to another CNN for object location. In the item identification stage, the creators utilize the new CNN to distinguish and include the quantity of individuals in the group. The creators assess the exhibition of their technique utilizing a few measurements, including mean normal accuracy (Guide), on a freely accessible dataset of swarmed scenes. The outcomes show that their strategy beats a few cutting edge techniques for swarm checking. In the paper shows the capability of IoT sensors and profound learning procedures for tending to the difficulties of group checking and has suggestions for the advancement of canny observation frameworks and shrewd urban communities.

Kun Liu et al. [9] proposed a method for overhauling abnormality distinguishing proof in observation accounts using move acquiring from action affirmation. The makers battle that traditional procedures for peculiarity acknowledgment in perception accounts, similar to those considering handcrafted features or shallow learning computations, are limited in their ability to manage the multifaceted design and alterability of genuine circumstances. To address these cutoff points, they propose a procedure that utilization the power of move acquiring from action affirmation. The proposed method contains two stages: planning and testing. In the readiness stage, the makers use a pre-arranged significant CNN for action affirmation to remove huge features from the video data and move the learned data to one more CNN for inconsistency ID. In the testing stage, they use the new CNN to recognize and bunch abnormalities in new surveillance accounts. The makers evaluate the display of their procedure using a couple of estimations, including exactness, survey, and F1-score, on a uninhibitedly open dataset of surveillance accounts. The results show that their strategy beats a couple of state of the art systems for abnormality disclosure in observation accounts. The assessment paper shows the ability of move acquiring from action affirmation for further developing

peculiarity area in perception accounts and has ideas for the progression of vigilant surveillance systems for public prosperity and security.

Ali Atghaei et al. [10] proposes a technique for strange occasion location in metropolitan reconnaissance recordings utilizing GANs and move learning. The creators contend that customary techniques for strange occasion discovery in observation recordings, like those in view of handmade highlights or shallow learning calculations, are restricted in their capacity to deal with the intricacy and fluctuation of certifiable situations. To address these constraints, they propose a strategy that use the force of GANs and move learning. The proposed strategy comprises of two phases: preparing and testing. In the preparation stage, the creators utilize a GAN to gain proficiency with the conveyance of ordinary occasions in metropolitan reconnaissance recordings utilizing an enormous dataset of recordings. In the testing stage, they use move figuring out how to adjust the learned GAN to identify unusual occasions in new observation recordings. The creators assess the exhibition of their technique utilizing a few measurements, including accuracy, review, and F1-score, on an openly accessible dataset of metropolitan observation recordings. The outcomes show that their technique outflanks a few best in class strategies for unusual occasion recognition in reconnaissance recordings. In the paper exhibits the capability of GANs and move learning for tending to the difficulties of strange occasion discovery in metropolitan observation recordings and has suggestions for the advancement of canny reconnaissance frameworks for public wellbeing and security in metropolitan conditions.

Al abdul karim L et al. [11] discussed the use of urban analytics for crowd management during the Hajj pilgrimage. The authors highlight the challenges faced by organizers in managing large crowds during the Hajj, including the risk of stampedes, overcrowding, and the spread of infectious diseases. They suggest that urban analytics, which involves the collection and analysis of data related to urban environments and human behavior, can be useful in addressing these challenges. The paper discusses various data sources that can be used for urban analytics, including CCTV cameras, mobile phone data, and social media. The authors also describe how data visualization and simulation techniques can be used to model crowd behavior and identify potential problem areas. The authors conclude that urban analytics has the potential to improve crowd management during the Hajj, but also note that further research is needed to fully explore its capabilities and limitations. They also suggest that urban analytics could be useful in other contexts where large crowds are present, such as music festivals and sporting events.

Maya acquire sasquatch Nakashima et al. [12] examines the difficulties of video rundown, including the need to distinguish key edges and minutes in a

video that precisely catch its substance. The creators propose an answer in light of profound learning and semantic element extraction. The technique includes first removing low-level visual elements from the video outlines utilizing a CNN. These highlights are then used to prepare a profound conviction organization (DBN) that can learn more elevated level semantic elements. The DBN is prepared on a huge dataset of video cuts with related marks that demonstrate their substance. When the DBN is prepared, it very well may be utilized to separate semantic highlights from new video cuts. The creators utilize these elements to distinguish key edges and minutes in the video that are generally illustrative of its substance. They likewise propose a technique for creating an outline of the video by choosing a subset of these vital casings and minutes. The creators assess their strategy on a few benchmark datasets and contrast it with other video synopsis procedures. They report that their strategy beats existing methods with regards to both precision and computational effectiveness. The paper presents an original way to deal with video rundown in view of profound semantic elements. The creators propose that their technique could be helpful in applications, for example, reconnaissance, video search, and content-based video recovery.

Hoatching shin et al. [13] presents an investigation of profound CNNs for PC supported recognition (computer aided design) in clinical imaging. Computer aided design is a procedure used to help radiologists in identifying irregularities in clinical pictures, for example, X-beams, CT sweeps, and X-ray examines. The creators explore the viability of various CNN models, dataset qualities, and move learning methods for computer aided design. They assess their techniques on a few benchmark datasets and report the outcomes with regards to responsiveness, particularity, and region under the beneficiary working trademark bend (AUC). The creators propose a technique for calibrating pre-prepared CNNs for explicit computer aided design errands. This includes preparing the CNN on an enormous dataset of general pictures, and afterward calibrating the organization on a more modest dataset of clinical pictures for a particular computer aided design task. The creators report that this technique can fundamentally work on the presentation of the CNN for computer aided design. The paper additionally talks about the difficulties of involving profound CNNs for computer aided design, including the requirement for a lot of marked information and the potential for overfitting. The creators propose a few systems to address these difficulties, including information expansion and regularization procedures. In the paper presents a thorough investigation of profound CNNs for computer aided design in clinical imaging. The creators recommend that their techniques could be helpful in clinical practice for working on the exactness and productivity of radiological determination.

As indicated by Lamba S, Nain N et al. [14] covers different parts of group observing and characterization, including information securing, information preprocessing, highlight extraction, and order. The creators examine various sorts of sensors that can be utilized for information obtaining, like cameras, mouthpieces, and ecological sensors. They additionally survey the various information preprocessing strategies used to clean and channel the information. The creators then, at that point, present a definite examination of element extraction methods utilized for swarm investigation, for example, thickness based highlights, movement based elements, and shape-based highlights. They likewise talk about various characterization strategies, for example, rule-based techniques, bunching based strategies, and ML-based techniques. The paper closes by featuring a portion of the difficulties and future bearings in the field of group checking and order, like the requirement for continuous handling and the improvement of additional exact and dependable strategies. The creators recommend that the future exploration in this field ought to zero in on growing more powerful and versatile methods that can deal with the intricacy and changeability of group information in different true situations.

As per Ravanbakhsh M et al. [15] proposed a fitting and-play structure for swarm movement examination, explicitly for distinguishing unusual occasions in swarms. The structure comprises of two primary parts: a pre-prepared CNN and an inconsistency discovery module. The pre-prepared CNN is utilized for highlight extraction from swarm movement information, like directions and optical stream. The peculiarity identification module then utilizes these highlights to identify unusual occasions in the group. The creators exhibit the adequacy of their methodology on a few benchmark datasets and report critical enhancements in strange occasion recognition contrasted with cutting edge techniques. One of the critical commitments of this paper is the utilization of pre-prepared CNNs for swarm movement investigation. This approach empowers the structure to be effortlessly adjusted to new datasets and assignments without the requirement for broad preparation. The creators likewise propose an original misfortune capability for preparing the irregularity discovery module that considers the sparsity of strange occasions in the information. The paper presents a promising methodology for swarm movement investigation and unusual occasion identification that is both successful and productive. The creators propose that their structure could be valuable in different certifiable applications, like reconnaissance, security, and group the executives.

In this Exploration Gong Cheng et al. [16] proposed an original methodology for remote detecting picture scene characterization utilizing profound learning and metric learning procedures. The proposed technique comprises of two fundamental parts a CNN for highlight extraction and a

measurement learning module for learning discriminative element portrayals. The CNN is utilized to remove highlights from remote detecting pictures, which are then taken care of into the measurement learning module. The measurement learning module learns a distance metric that expands the between class distance and limits the intra-class distance, in this manner learning discriminative component portrayals for various scene classes. The creators likewise propose an original misfortune capability for preparing the measurement learning module that consolidates both the order misfortune and the measurement learning misfortune. The adequacy of the proposed approach is exhibited on a few benchmark remote detecting datasets, and the outcomes show critical upgrades over cutting edge strategies. The creators likewise lead a removal study to research the commitment of every part of their way to deal with the general presentation. In the paper presents a promising methodology for remote detecting picture scene order that use the force of profound learning and metric learning strategies. The creators recommend that their methodology could be valuable in different applications, like ecological checking, debacle the executives, and metropolitan preparation.

Menschen prior et al. [17] proposed a profound learning-based approach for the programmed characterization of CT lessening designs for interstitial lung sicknesses. The proposed approach utilizes profound CNNs to remove highlights from CT pictures, which are then utilized for arrangement. The creators prepared their CNN on an enormous dataset of CT pictures with various weakening examples, including typical, ground-glass haziness, reticular, and honeycombing designs. They utilized an exchange learning approach, where a pre-prepared CNN model was tweaked on their dataset. The creators likewise proposed an original misfortune capability for preparing the CNN that integrates both grouping and regularization targets. The viability of the proposed approach is assessed on a different dataset of CT pictures with various weakening examples. The outcomes show that the proposed approach accomplishes cutting edge execution for the grouping of CT constriction designs for interstitial lung infections. In the exploration presents a promising methodology for the programmed order of CT lessening designs for interstitial lung sicknesses utilizing profound learning procedures. The creators recommend that their methodology could be valuable in clinical practice for the early recognition and determination of interstitial lung illnesses.

In this paper showing dance et al. [18] proposed a profound learning-based approach for human activity acknowledgment utilizing 3D CNNs. The creators contend that customary strategies for human activity acknowledgment, which depend close by made elements and shallow models, have restricted precision and speculation abilities. To resolve this issue, the creators propose a 3D CNN engineering that can gain spatiotemporal elements straightforwardly from crude video outlines. The proposed design comprises of different 3D convolutional

layers followed by max-pooling and completely associated layers. The creators likewise present another information expansion procedure that haphazardly tests and links different continuous edges to expand the size of the preparation set. The viability of the proposed approach is assessed on a few benchmark datasets for human activity acknowledgment, including UCF101 and HMDB51. The outcomes show that the proposed 3D CNN engineering outflanks past cutting edge techniques on these datasets.

As indicated by Muhammad zees et al. [19] proposed a profound learning-based approach for semantic investigation of news stories utilizing CNNs. The creators contend that customary strategies for news investigation, which depend on physically made highlights and complex rule-based frameworks, have restricted precision and versatility. To resolve this issue, the writers propose a CNN-based engineering that can gain significant elements straightforwardly from crude news stories. The proposed engineering comprises of different convolutional and pooling layers, trailed by completely associated layers and a softmax classifier. The writers likewise utilize pre-prepared word embeddings to address words in the info news stories. The viability of the proposed approach is assessed on a few benchmark datasets for news order and feeling examination. The outcomes show that the proposed CNN design beats past cutting edge techniques on these datasets. In the paper presents a promising methodology for semantic examination of news stories utilizing CNNs, which has possible applications in fields like media investigation, data recovery, and content proposal.

Kwang-eun koi et al. [20] proposed a profound learning-based structure for distinguishing unusual conduct in a brilliant observation framework. The proposed system comprises of three principal stages: (1) pre-handling and component extraction, (2) profound CNN displaying, and (3) unusual conduct identification. In the pre-handling stage, the creators use picture and video handling strategies to remove helpful elements from the information video information. The pre-handled information is then taken care of into a CNN model, which is prepared utilizing a huge dataset of typical and strange conduct recordings. The creators use move figuring out how to introduce the CNN model with pre-prepared loads, which assists with further developing the preparation productivity and accuracy.In the strange conduct discovery stage, the creators utilize a limit based way to deal with group the info video successions as typical or unusual. The edge is gained from the preparation information and depends on the conveyance of the CNN yield scores for ordinary and strange recordings. The proposed structure is assessed on a few benchmark datasets for strange conduct discovery, and the outcomes show that it outflanks a few cutting edge techniques. The creators likewise exhibit the viability of their structure in a genuine brilliant observation framework. In the paper presents a promising

methodology for distinguishing strange conduct in a savvy reconnaissance framework utilizing profound CNNs, which has expected applications in fields like security and public wellbeing.

Demetrio's purpose's et al. [21] presents a start to finish video foundation deduction approach utilizing 3D CNNs. The proposed technique figures out how to separate spatiotemporal highlights straightforwardly from crude video successions with next to no manual element designing. The 3D CNN engineering is intended to catch both spatial and worldly data of the video outlines, and a foundation model is advanced by the organization during preparing. The proposed technique is assessed on standard datasets and contrasted and cutting edge strategies, showing further developed execution. The outcomes exhibit the adequacy of the proposed approach in taking care of mind boggling video scenes with dynamic foundations and brightening changes.

Key hang et al. [22] proposes a review approach for video outline, where the synopsis is created in the wake of noticing the whole video succession. The proposed approach first builds a chart portrayal of the video, where every hub addresses a casing and edges between hubs catch the worldly relations between outlines. Then, at that point, a chart rundown calculation is applied to the diagram to recognize a bunch of keyframes that really sum up the video. At last, a video outline is created by choosing a subset of edges from the keyframes that are illustrative of the whole video succession. The proposed approach is assessed on standard datasets and contrasted and best in class strategies, showing cutthroat execution. The outcomes exhibit the adequacy of the proposed review approach for video rundown.

M. Bendali-Braham et al. [23] presents a technique for the grouping of group developments in video accounts utilizing move learning. The proposed approach utilizes pre-prepared profound learning models for huge scope picture datasets and tweaks them on more modest datasets of group recordings to accomplish exact arrangement. The creators assess the proposed technique on two group datasets and show that it beats a few benchmark strategies concerning order exactness. The outcomes propose that move learning can successfully use enormous scope picture datasets to work on the characterization of group developments in video accounts.

Ibrion M et al. [24] discusses the impact of cultural beliefs on disaster risk governance and resilience in Iran. It examines how cultural factors such as fatalism, religiosity, and trust in authorities can affect the preparedness and response to disasters in the country. The author argues that understanding and addressing the beliefscape, or the cultural beliefs and attitudes surrounding disaster risk, is crucial for improving disaster risk governance and enhancing

community resilience. The article also emphasizes the importance of engaging with local communities and their beliefs in disaster risk management efforts.

Ilyas, Z et al. [25] presents an original half and half profound organization based approach for swarm peculiarity location in observation recordings. The proposed approach joins the qualities of profound CNNs and LSTM organizations. The CNNs are utilized for highlight extraction and the LSTM network is utilized for arrangement displaying. The methodology is assessed on two benchmark datasets, to be specific, UCF-Wrongdoing and ShanghaiTech Group. The exploratory outcomes show that the proposed approach beats cutting edge strategies as far as discovery exactness and misleading problem rate. The article presumes that the proposed approach can actually distinguish swarm oddities in observation recordings and can be utilized in true situations for improved public security.

Wei Lin, Junyu Gao et al. [26] presents a clever methodology for identifying peculiarity occasions in swarm scenes utilizing manufactured information. The proposed strategy initially creates an enormous number of engineered preparing tests with different peculiarity examples and afterward prepares a profound learning model utilizing these examples. The model comprises of an element extractor and a classifier, which are together prepared utilizing a double cross-entropy misfortune capability. The exploratory outcomes show that the proposed strategy outflanks cutting edge techniques on two benchmark datasets, exhibiting the adequacy of involving manufactured information for swarm abnormality recognition.

Suresha, M et al. [27] proposes a profound learning approach for situation based irregularity identification in video observation frameworks. The proposed strategy uses a 3D CNN to catch spatio-fleeting elements from video groupings and characterize them into typical or unusual occasions. The strategy is assessed on a freely accessible dataset and accomplishes promising outcomes contrasted with best in class techniques. The paper features the capability of profound learning-based strategies for situation based irregularity discovery and their significance in upgrading video observation frameworks' security.

Alafif, T et al. [28] presents a strategy for unusual conduct recognition in gigantic group recordings utilizing GANs. The proposed technique is applied to the Hajj journey, where enormous groups assemble and unusual occasions can happen. The GAN-put together model is prepared with respect to norml swarm conduct to produce sensible video outlines, and any deviations from the created outlines are considered as unusual way of behaving. The strategy is assessed on genuine group recordings from the Hajj journey, and the outcomes show the

adequacy of the proposed approach for strange conduct identification in enormous scope swarm scenes.

Bamaqa, Amna et al. [29] proposes a various leveled fleeting memory (HTM) approach for both receptive and proactive anmaly recognition in swarm the executives. The proposed framework utilizes a blend of video and sensor information to identify irregularities in swarm conduct. The framework is intended to be versatile and ready to learn and refresh its abnormality discovery models over the long haul. The methodology is assessed on genuine information from a shopping center and the outcomes show that the proposed framework beats customary peculiarity discovery strategies.

Table 2.1
Literature Survey

Authors / Year	**Sharma et al. 2023**
Title	Scale-aware CNN for crowd density estimation and crowd behavior analysis [1]
Dataset	ShanghaiTech
Methodology	motion map generation, cascaded CNN architecture for crowd density estimation, K-means clustering
Results	crowd counting accuracy = 97.60%, Average MAE= 58.60% and MSE = 98.55%
Surveillance	No
Authors / Year	**Alharthi et al. 2023**
Title	Massive Crowd Abnormal Behaviors Recognition Using C3D [2]
Dataset	Hajj
Methodology	C3D Model, VGG-16
Results	C3D: Accuracy= 0.89, Precision= 0.9, Recall=0.89, F1-score=0.88
Surveillance	No
Authors / Year	**Danista et al. 2022**
Title	Cross Count: Efficient Device-Free Crowd Counting by Leveraging Transfer Learning [3]
Dataset	channel state information (CSI) Dataset Local
Methodology	CSI-based supervised crowd counting systems, Transfer learning and CNNs
Results	accuracy with transfer learning = 81.47%, CNN model accuracy =76.764%
Surveillance	No
Authors / Year	**Khosro et al. 2022**
Title	Deep-Transfer-Learning-Based Abnormal Behavior Recognition Using Internet of Drones for Crowded Scenes [4]
Dataset	UMN dataset, UCSD datase
Methodology	deep transfer learning, Internet of Drones, ModIFIEd ResNet-18
Results	UCSD-ped2: accuracy = 82%, Sensitivity= 81%, Specifi city= 82%
Surveillance	NO

Authors / Year	Ali et al. 2022
Title	Deep crowd transfer networks for an approximate crowd counting [5]
Dataset	ShanghaiTech, JHU-CROWD++, and UCF-QNRF. R-CNN algorithms
Methodology	Efficient crowd counting: combination of counting and estimate techniques
Results	DCTNet on ShanghaiTech Dataset: DCTNet = MAE= 60.7, MSE=99.1, DCTNet on JHUCROWD++ Dataset: DCTNet=MAE =63.2, MSE= 249.3, DCTNet on UCF-QNRF
Surveillance	No
Authors / Year	Alrowais et al. 2022
Title	Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems [6]
Dataset	CIFAR-10 dataset
Methodology	Image processing: Namely texture feature and edge feature, Regression function: Estimating aggregate person numbers , Density map regression: Estimating crowd density
Results	Fscore: 81% and 84.95%, Fscore of 92.87%
Surveillance	No
Authors / Year	Alafif et al. 2022
Title	Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study [7]
Dataset	Hajj
Methodology	optical fow framework based on GAN, U-Net, Flownet , transfer learning
Results	Transfer learning (%): HAJJ= 79.63%, Non-transfer: HAJJ=76.89%
Surveillance	NO
Authors / Year	Liang et al. 2022
Title	Liang, Dingkang, et al. "Transcrowd: weakly-supervised crowd counting with transformers." Science China Information Sciences 65.6 [8]
Dataset	NWPU-Crowd JHU-CROWD++ ShanghaiTech
Methodology	Transformer
Results	92.2%
Surveillance	No
Authors / Year	Rezaei et al 2021
Title	Real-time crowd behavior recognition in surveillance videos based on deep learning [9]
Dataset	PETS 2009
Methodology	Yolo for detection 3DCNN for behavior Crowd Counting Behaviour Recognition
Results	93.07%
Surveillance	Yes
Authors / Year	Ahmed et al. 2021
Title	IoT-based crowd monitoring system: Using SSD with transfer learning [10]
Dataset	Self Recorded Videos
Methodology	Mixture of Gaussian (MoG) Person Counting ,SSD-MobilenetV2,
Results	SSD-Mobilenetv2 without transfer learning: Precision = 89%, Recall = 91%, mAP = 92%, Counting accuracy = 92%, SSD-Mobilenetv2 with transfer learning: Precision = 91%, Recall = 93%, mAP = 95%, Counting accuracy = 95%
Surveillance	NO

Authors / Year	Kun Liu et al 2020
Title	Enhancing Anomaly Detection in Surveillance Videos with Transfer Learning from Action Recognition [11]
Dataset	CitySCENE
Methodology	2D-CNN based methods and 3D-CNN based approaches
Results	Accuracy 90.43%
Surveillance	Yes
Authors / Year	Ali Atghaei et al. 2020
Title	Abnormal Event Detection in Urban Surveillance Videos Using GAN and Transfer Learning [12]
Dataset	(UCSD Peds1 and UCSD Peds2
Methodology	Gunner-Farneback optical flow algorithm , GAN , VGG-16
Results	Results on UCSD PEDS1: Frame-level EER= 14% , Frame-level AUC =93% , Pixel-level EER= 36% , Pixel-level AUC= 73% , Results on UCSD PEDS2: Frame-level EER= 15% , Pixel-level EER= 17%
Surveillance	Yes
Authors / Year	Shehzed et al.2019
Title	Shehzed, Ahsan, Ahmad Jalal, and Kibum Kim. "Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection." 2019 international conference on applied and engineering mathematics (ICAEM). IEEE, 2019[13].
Dataset	PETS 2009
Methodology	Yolo for detection, 3DCNN for behavior
Results	93.07%
Surveillance	Yes
Authors / Year	Alanazi et al.2019
Title	Alanazi, Adwan Alownie, and Muhammad Bilal. "Crowd density estimation using novel feature descriptor."[14].
Dataset	PETS 2009
Methodology	Support Vector Machine (SVM)
Results	88.7%
Surveillance	No

The Table 2.1 is showing that the research works [1, 2, 3, 4, 5, 6,7, 8, 10, 14] did not focus on crowd surveillance system as the crowd surveillance is an important which can be helpful to identify potential targets for criminal activities, including terrorism, theft, or violence. The surveillance system helps to detect suspicious behavior, abnormal activities, or unattended objects, enabling early intervention and preventive measures to mitigate risks. By monitoring crowd behavior, crowd surveillance can help ensure public safety and maintain order during large-scale events, protests, or demonstrations.

Most of the research is based on object detection [6] and Crowd Counting [13, 8] in crowd surveillance system but the authors did not focus on behavior of the crowd and their category. The behavior of crowd is an important factor to understanding how crowds behave can help in managing and ensuring the safety of people in various settings such as stadiums, concerts, festivals, protests, and transportation hubs. By analyzing crowd behavior, authorities can identify

potential risks, anticipate crowd movements, and implement appropriate measures to prevent accidents, stampedes, or overcrowding.

2.1 Limitation of Research Work

Table 2.2
Limitation of Literature survey

Authors / Year	Dataset	Prediction Model	Decision Model	Use CNN
Sharma et al. 2023 [1]	ShanghaiTech	Scale-aware CNN for crowd analysis.	CNN for crowd analysis decisions.	CNN for crowd analysis modeling.
Alharthi et al. 2023 [2]	Hajj	C3D for abnormal behavior recognition.	Abnormal behavior recognition with C3D.	CNN-based abnormal behavior recognition.
Danista et al. 2022 [3]	channel state information (CSI) Dataset Local	Transfer learning for crowd counting.	Efficient crowd counting using transfer learning.	CNN-based crowd counting efficiency.
Khosro et al. 2022 [4]	UMN dataset , UCSD datase	IoT drones for abnormal behavior recognition.	Transfer learning for abnormal behavior recognition.	CNN-based abnormal behavior recognition.
Ali et al. 2022 [5]	ShanghaiTech, JHU-CROWD++, and UCF-QNRF. R-CNN algorithms	Transfer learning-based crowd counting	Transfer learning for crowd counting	CNNs utilized for approximate crowd counting task.
Alrowais et al. 2022 [6]	CIFAR-10 dataset	Transfer learning for crowd density analysis.	Transfer learning for crowd density analysis.	CNNs utilized for intelligent crowd analysis.
Alafif et al. 2022 [7]	Hajj	--	GAN-based abnormal behavior detection.	GAN with CNN for crowd behavior detection.
Liang et al. 2022 [8]	NWPU-Crowd JHU-CROWD ShanghaiTEc	Transformer-based crowd counting prediction model.	--	Weakly-supervised crowd counting with transformers.
Rezaei et al 2021 [9]	PETS 2009	Deep learning-based crowd counting model.	--	Deep learning with CNN implementation.
Ahmed et al. 2021 [10]	Self Recorded Videos	SSD with transfer learning for crowd monitoring.	--	Transfer learning for IoT crowd monitoring.

Kun Liu et al 2020 [11]	CitySCENE	Transfer learning for video anomaly detection.	Transfer learning for video anomaly detection.	CNN for video anomaly detection.
Ali Atghaei et al. 2020 [12]	(UCSD Peds1 and UCSD Peds2	GAN-based transfer learning for abnormal event detection.	GAN-based transfer learning for event detection.	GAN with CNN for event detection.
Shehzed et al.2019 [13]	PETS 2009	Multi-person tracking for crowd analysis.	Smart surveillance for crowd analysis.	CNN-based crowd tracking system.
Alanazi et al.2019 [14]	PETS 2009	Novel feature descriptor for density estimation.	Feature descriptor for density estimation.	CNN-based density estimation model.

3. Chapter

Proposed Methodology

The increasing demand for public safety and security necessitates the development of advanced surveillance systems capable of efficiently analyzing crowded scenes. Traditional approaches for crowd surveillance often suffer from limitations such as high computational costs, inadequate real-time performance, and a lack of robustness to variations in crowd dynamics. This thesis proposed the integration of transfer learning techniques into crowd surveillance systems, aiming to overcome these challenges and enhance the system's capabilities. the widespread availability of surveillance cameras and

the need for efficient monitoring of crowded environments have led to a surge in research and development of crowd surveillance systems. These systems aim to detect and track individuals, analyze crowd behavior, and identify potential threats or abnormal activities. However, the success of such systems heavily relies on the availability of large-scale labeled datasets, which are often scarce and time-consuming to create. To address this challenge, this research proposes a crowd surveillance system leveraging transfer learning techniques to achieve effective and scalable crowd monitoring.

3.1 Research Contribution

The contribution of a thesis on a crowd-based surveillance system using transfer learning can encompass several aspects. Here are some potential contributions that can be made in such a thesis:

- The developed Proposed System detects the event categories like Sports, Protest, Music Concert, Religious crowd and classify the crowd behavior such as peaceful and unpeaceful using transfer learning techniques.
- It gives an extensive assessment and benchmarking of move learning strategy with regards to swarm based reconnaissance. This elaborate contrasting the presentation of our model and cutting edge proposed models on standard assessment measurements, like exactness, accuracy, review, or F1 score.
- The Proposed Work looked at the exhibition of the proposed move learning approach with conventional strategies regularly utilized in swarm reconnaissance, like high quality elements or traditional ML methods. The correlation featured the benefits of move learning regarding precision, productivity, flexibility to assorted swarm situations, or speculation ability.

3.2 Methodology

3.2.1 Overall Methodology

In the extent of our task, the proposed research have recognized two modules inside our framework. The principal module centers on imaging-based order, while the subsequent module rotates around video-based peculiarity discovery. Since it can catch the two pictures and recordings of groups, there is no equivocalness with respect to the necessary information. Our essential goal is to order pictures in view of specific highlights to acquire the ideal result. AI (ML) calculations have arisen as well-known strategies for tending to characterization and relapse issues. A portion of the fundamental calculations, as portrayed in Figure 3.1, incorporate SVM, direct relapse, strategic relapse, grouping tree, credulous Bayes, and KNN.

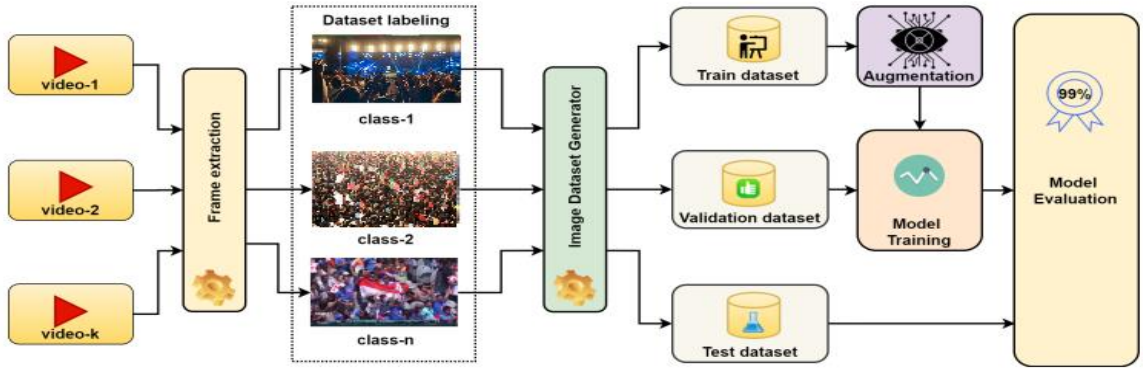


Figure 3.1: Proposed Methodology

Video Processing: The first step in crowd analysis involves extracting frames from videos. Videos containing crowd scenes are divided into individual frames to enable further analysis.

- **Data Labelling:** Once the frames are extracted, each frame needs to be labeled to identify specific elements of interest within the crowd. This process involves manually annotating objects, such as people or objects of interest, to create a labeled dataset for training the crowd analysis model.
- **Model Training:** The labeled dataset is utilized to train a crowd analysis model. This model learns to recognize and understand various aspects of crowd behavior, such as crowd density, flow patterns, or anomalous events. Training involves iteratively fine-tuning the model's parameters until it achieves satisfactory performance.
- **Data Augmentation:** To improve the presentation and vigor of the prepared model, information increase procedures are utilized. These strategies include applying changes or alterations to the named dataset, like turn, scaling, or presenting counterfeit clamor. By increasing the dataset, the model increases openness to a more extensive scope of group situations, further developing its speculation capacities.
- **Model Validation:** Subsequent to preparing the model, surveying its exhibition and speculation ability is fundamental. This is finished by assessing the model on a different approval dataset that comprises of named swarm scenes not seen during preparing. The model's exhibition measurements, like exactness, accuracy, review, or F1 score, are processed to quantify its viability in dissecting swarm conduct.
- **Model Evaluation:** Once the model has been validated, it can be deployed for real-world crowd analysis tasks. Its performance is assessed on unseen, real-time crowd footage or datasets. This evaluation helps determine its practical utility and identifies areas for potential improvement.

By following this flow, crowd analysis systems can effectively process videos, extract frames, label data, train models, augment data, validate model performance, and ultimately evaluate the model's effectiveness in analyzing crowd behavior.

3.3 Explanation of Work Flow

- Surveillance Cameras: The flow begins with the utilization of surveillance cameras strategically placed in areas where crowd behavior needs to be monitored. These cameras capture real-time video footage of the crowd scenes.
- Frame Extraction: From the video footage obtained from surveillance cameras, individual frames are extracted. Each frame represents a snapshot of the crowd at a specific moment in time.
- Deep Learning Model: The removed edges are then gone through a profound learning model. Profound learning models are a kind of man-made intelligence calculation fit for learning complex examples and connections inside information. For this situation, the model is explicitly prepared to investigate swarm conduct.
- Crowd Analysis: The deep learning model analyzes the frames to understand the crowd behavior. It examines various visual cues, such as crowd density, movement patterns, group formations, and individual actions. Based on these cues, the model can classify the crowd behavior into different categories, such as peaceful or violent.
- Behavior Classification: The result of the profound learning model is utilized to order the noticed group conduct. The model relegates a name to each examined outline, showing whether the group conduct in that edge is quiet or brutal. This arrangement gives bits of knowledge into the general idea of the group conduct over the long run.

In the Fig3.2 the surveillance cameras capture real-time footage, which is then converted into frames. These frames are fed into a deep learning model, which analyzes the crowd behavior and classifies it as either peaceful or violent. This flow enables the automated monitoring and analysis of crowd behavior, assisting in situations where timely identification of potentially disruptive or dangerous behavior is crucial.

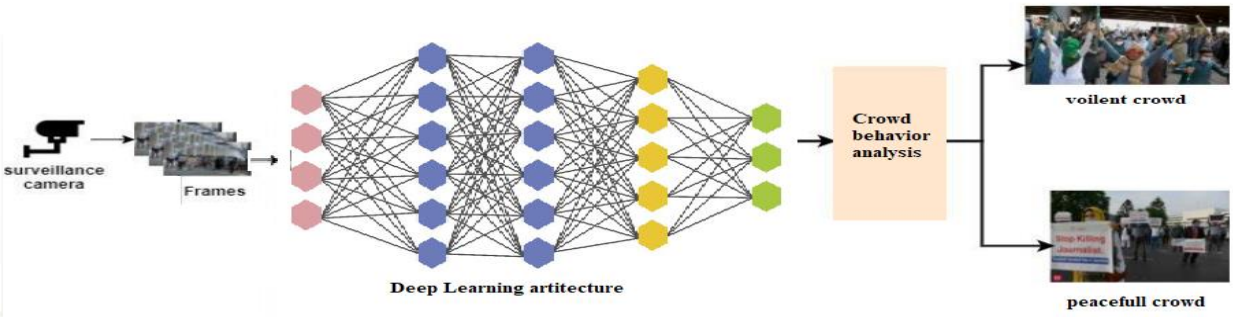


Fig 3.2 Work Flow Diagram

3.3.1 Detecting the Objects

Object detection is a strong procedure in PC vision that distinguishes the class of an item as well as gives its exact area inside a picture or video. It has a large

number of utilizations, for example, object counting and following, as well as precise naming. To recognize individuals on roads, a pre-prepared Tensor Flow model called SSD ResNet50 V1 FPN 640 × 640 is used. SSD, or Single Shot Multi Box Finder, is a calculation explicitly intended for continuous video examination, as it can distinguish different articles all the while in a solitary shot. This calculation accomplishes its fast by taking out the requirement for producing jumping box proposition. All things being equal, it utilizes different bouncing boxes and changes them as a feature of the forecast cycle. The organization's last couple of layers are liable for logically more modest bouncing box expectations, and the last expectation is a mix of these singular expectations. The SSD ResNet50 V1 FPN model fills in as a component extractor, shared box indicator, and utilizations central misfortune. It has been prepared on the COCO 2017 dataset, which comprises of pictures scaled to a goal of 640x640. The COCO dataset contains 80 classes, including the classes "weapon, stick, individual, place cards, tennis racket," which are specifically noteworthy in this unique situation.

3.3.2 Tracking Crowd

In object following, a remarkable ID is relegated to each recognized item, which stays unaltered as long as the item is inside the view or video feed. Following is a critical test in AI and has various commonsense applications. Since object location treats each edge freely and needs connection between items across outlines, an alternate calculation is expected to lay out connections between these items. Profound SORT (Straightforward On the web and Continuous Following) is one such calculation that resolves this issue and offers effective following capacities. Profound SORT is a broadly utilized following calculation known for its precision and speed. It works quickly and can accomplish countless casings each second. It integrates a special methodology by thinking about distance and speed as well as the presence of articles. This is achieved by processing profound highlights for each bouncing box and using the comparability between these elements to illuminate the following rationale. Profound SORT extricates a 128-layered vector for each bouncing box, catching the vital highlights of the item. By utilizing these profound highlights, Profound SORT empowers upgraded following execution, even in genuine situations including covering or firmly found objects. The utilization of Profound SORT for object following outcomes in more strong and solid following, especially while managing testing situations. It handles the following IDs of covering articles or items in closeness really. The result of the following system can be envisioned in Figure 3.3.2.

Fig 3.3.2 Tracking object

As well as relegating novel IDs to followed objects, the proposed model presents a variety coding framework in light of these following IDs. Each following ID is related with a particular tone, considering simple visual recognizable proof of individual items. Moreover, the model creates trails that follow the development of these items in the video film. These paths are likewise doled out an exceptional variety, particular from the item's bouncing box tone. The paths are set close to the foot of the jumping boxes to upgrade the precision of following the articles' genuine area. By catching the development design through these paths, the model can later picture the directions of the items on a guide or in a graphical portrayal. This gives significant bits of knowledge into the ways taken by the items and helps in understanding their development examples and conduct. Figure 3.3.3 delineates an illustration of such perception, exhibiting the paths and their related varieties for the followed objects. By consolidating variety based special IDs and trails, the proposed model upgrades the following system and empowers a more far reaching examination of item development in the video film.

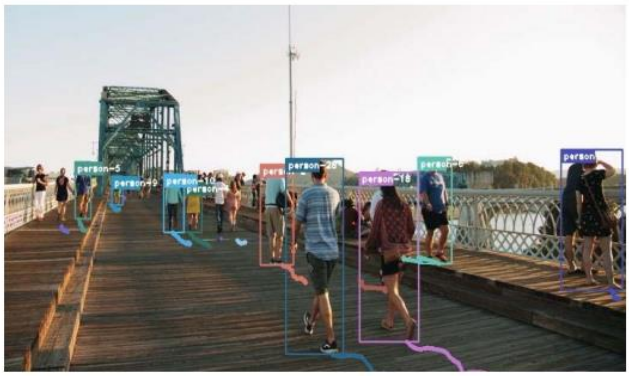


Fig. 3.3.3. Tracking with trail

3.3.4 Mapping the Trail

For each recognized item, a bunch of x and y organizes is gotten, addressing its area in the pixel space of the 2D edge caught by the camera [36]. To relate these pixel directions to this present reality spatial area, a planning is expected between the pixel space and another 2D plane, like a guide. This planning can

be accomplished utilizing a viewpoint change [37], which is a lattice activity that tasks focuses starting with one 2D plane then onto the next. The point of view change is a usefulness given by the Open CV library, explicitly the "get Point of view Change" capability, which plays out the important framework estimations. To plan the pixel directions to this present reality arranges, it is fundamental to decide the longitude and scope of explicit pixel areas in our casing [Figure 3.3.4]. By laying out a correspondence between pixel organizes and their relating certifiable directions, the viewpoint change can precisely extend the item's area from the pixel space to the guide. The point of view change activity assumes a pivotal part in adjusting the distinguished item's situation in the pixel space with its genuine area in reality, considering significant perceptions and examination of article developments on a guide.

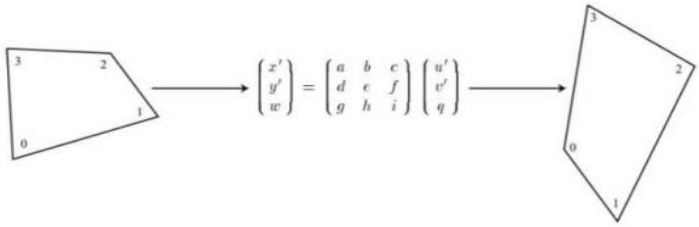


Figure 3.3.4 Image Warping and Texture Mapping for Crowds [42]

To guarantee an exact viewpoint change, it is important to characterize quadrilaterals with four focuses for both the pixel space (camera film) and this present reality space (scaffold's satellite picture). These quadrilaterals ought to be painstakingly situated, spread out precisely, and cover however much region as could be expected to improve the ongoing area exactness. In your particular situation, the camera film catches a well-known span, and the planning between the camera film and the scaffold's satellite picture is shown in Figures 3.3.5 and 3.3.6. These figures portray the quadrilaterals that have been laid out to lay out the correspondence between the camera's pixel space and this present reality space addressed by the scaffold's satellite picture. By exactly choosing and orchestrating the four focuses in both the camera film and the satellite picture, the viewpoint change can successfully project the article areas from the camera film to their relating positions on the extension's satellite picture. This planning empowers precise arrangement between the noticed items in the camera film and their certifiable situations on the scaffold. If it's not too much trouble, let me know as to whether there's anything explicit you would like me to improve or some other changes you might want to find in the substance.


Figure 3.3.5 Area of Effect in footage Figure 3.3.6 Area of Effect on map

When the planning between the pixel space and this present reality space is laid out, it is important to make a class that can decipher pixel areas (x, y) to true areas (longitude, scope). This class plays out a direct change in view of the planned places. In any case, it's critical to take note of that this class incorporates no rationale to confirm whether a foundation pixel falls past the disappearing point of the 2D surface or on the other hand on the off chance that a longitude-scope coordinate falls behind the camera's field of view. Wariness ought to be practiced while applying these changes to guarantee their precision and validity.The aftereffects of applying this change can be seen in Figures 3.3.7 and 3.3.8. These figures probably outline the result of making an interpretation of pixel areas to their comparing certifiable longitude and scope organizes utilizing the made class. It merits underscoring that while the straight change gives a way to switch pixel areas over completely to true facilitates, it doesn't represent likely restrictions or limitations of the camera's viewpoint or the idea of the scene. Extra contemplations and approval might be important to guarantee the changed areas precisely address this present reality positions.

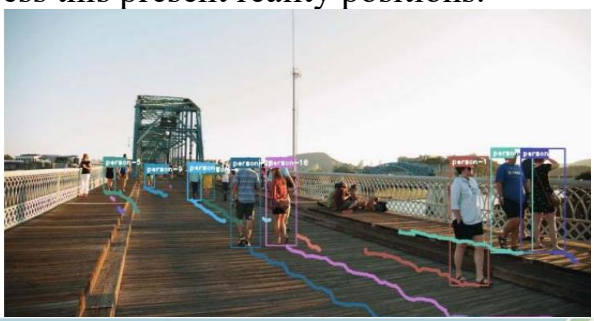



Fig. 3.3.7. Tracked Crowd Fig. 3.3.8. Crowd Movement Tracks

The unique color assigned to each tracking ID allows for effective visualization of movement patterns on the map. By associating a specific color with each

tracking ID, it becomes easier to track and differentiate the movement of individual objects or persons. This color-coded tracking enables a clear representation of the paths followed by each object or person, enhancing the analysis of their movement patterns. In Figure 3.3.9, you can observe the visualization of tracked movement patterns on the map, showcasing the distinct color trails associated with each tracking ID. This representation provides valuable insights into the trajectories and paths taken by the tracked objects or persons. By utilizing unique colors for each tracking ID, the model enables efficient visualization of individual movement patterns and facilitates detailed analysis of object or person tracking on the map.

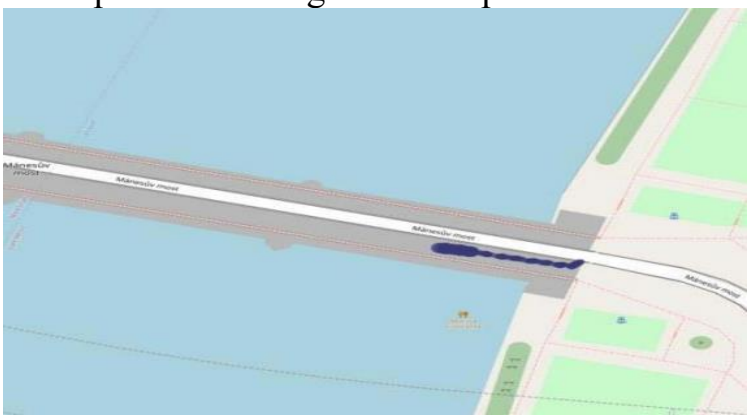


Fig.3.3.9. Tracking individual

The expanded circle, which features the individual's ongoing position, fills in as a visual mark of their nearby area on the guide. By growing the circle, it becomes more straightforward to recognize and pinpoint the specific place of the individual at a given second. This can be especially valuable for observation purposes, where constant following and it are fundamental for screen of people. Furthermore, the track, addressed by the variety coded trail, gives an obvious sign of the individual's heading of development. By following the track, it becomes clear to follow the way taken by the individual over the long haul. This component considers simple distinguishing proof and examination of the individual's development designs, helping with observation activities. The blend of the augmented circle and the track gives a proficient method for imagining and following need targets. It works with continuous observing and upgrades situational mindfulness by obviously showing the individual's ongoing position and development bearing on the guide. This usefulness is significant in observation situations where following explicit people is of most extreme significance.

3.5 Dataset Selection

For training and testing Deep learning models, a large amount of dataset is require because a deep learning model learns from data. After defining the scope first of all, the dataset was collected. According to the project scope, it is decided to do crowd type detection part using image classification that requires image dataset. parallel with dataset gathering, firstly the image dataset was tested using

a simple 2D CNN model. For more accuracy, the 2D CNN model was changed to inception V1 model and this model was trained on our own dataset. The activity recognition is a timely process and needs to be gathered. The activity recognition is implemented using a 2D NCC using DSL studio. The detail of each is given below:

3.6 Dataset Acquisition

Data acquisition is a fundamental beginning move toward tending to any profound learning issue. The quality and amount of the information straightforwardly impact the precision and execution of the subsequent model. Subsequently, it is urgent to gather information with a predefined heading and reason. On account of the proposed framework, the initial step is to decide the information sources. For web based information assortment, the framework essentially accumulates 2D picture information from Google and 2D video information from YouTube. These stages offer broad vaults of visual substance that can be used for preparing and testing profound learning models. Here are the particular subtleties for every information class:

3.6.1 2D Image Data from Google:

The system collects 2D image data primarily from Google. This may involve employing techniques such as web scraping or utilizing specific search queries to retrieve relevant images. The goal is to curate a diverse and representative dataset that aligns with the specific deep learning problem at hand. The collected images may encompass various object categories, scenes, or concepts.

3.6.2 2D Video Data from YouTube:

The system also gathers 2D video data from YouTube. This could involve extracting video clips or utilizing publicly available video datasets from the platform. The inclusion of video data adds a temporal dimension to the learning process, making it suitable for tasks like object detection, tracking, or action recognition. By obtaining data from both Google and YouTube, the proposed system ensures a comprehensive and varied dataset for training and evaluating deep learning models. The diverse sources of visual data contribute to more accurate and robust results in subsequent stages of the deep learning pipeline.

3.7 Dataset preprocessing

Dataset preprocessing is important for better results. These steps are important for getting a clean data. This preprocessed image and videos will help the model to better understand the dataset. The preprocessing steps are as follows:

4. Chapter
Implementation and Results

4.1 Implementation and results

For the recognition module, the organization was prepared for 1,000 ages, with bunch size of 16, over dataset. To limit the blunder, SGD-force was utilized with learning rate 0.02, energy 0.8. Xavier initializer was utilized to instate the boundaries. Adam analyzer is utilized having learning rate and weight rot equivalent to 0.0001. The organization was executed utilizing Tensor stream and Keras on a machine having 16 GB Smash, and a Nvidia 1080Ti 11GB GPU.

4.2 Data Set:

Dataset	Collected Data					Source
Images Dataset	Total	Religious	Political	Sports	Concert	Google
	4000	1000	1000	1000	1000	
Videos Dataset	Total	Violence	Sudden Panic	Property Damage	Peaceful	Youtube
	361	135	70	46	110	

4.1 Table Statistics of the dataset gathered

4.3 Confusion Matrix

Actual	Prediction					
	Input Sample	400	Sports	Protest	Religious	Public
	Sports	100	92	2	3	3
	Protest	100	3	91	2	4
	Religious	100	2	3	93	2
	Public	100	2	2	3	92

4.2 Table Confusion Matrix

4.4 Comparative Results

SR	Author	Methdolgy	Accuracy	Miss Rate
1	Sharma et al. 2023 [1]	motion map generation, cascaded CNN architecture for crowd density	90.60%	9.4

		estimation, K-means clustering		
2	Alharthi et al. 2023 [2]	C3D Model, VGG-16	89.01%	10.01
3	Danista et al. 2022 [3]	CSI-based supervised crowd counting systems, Transfer learning and CNNs	76.76%	23.24
4	Khosro et al. 2022 [4]	deep transfer learning, Internet of Drones, ModIFIEd ResNet-18	82%	18
5	Alrowais et al. 2022 [6]	Image processing: Namely texture feature and edge feature, Regression function: Estimating aggregate person numbers, Density map regression: Estimating crowd density	85%	15
6	Alanazi et al.2019 [14]	Support Vector Machine (SVM)	88.07%	11.3
7	Kun Liu et al 2020 [11]	2D-CNN based methods and 3D-CNN based approaches	90.43%	9.57
8	Our	CustomizedCNN Model with Modified Residual Block	95.34%	4.66

4.3 Table Comparative Table

4.5 Frameworks and Libraries

4.5.1 Framework

4.5.1.1. pycharm:

PyCharm is an incorporated advancement climate (IDE) explicitly intended for Python programming. It is created by JetBrains, a product improvement organization known for making efficiency instruments for designers. PyCharm gives a large number of elements and instruments to work with Python improvement, including code finishing, sentence structure featuring, code route, investigating capacities, and variant control mix.

4.5.1.2 Juypter note book:

Jupyter (previously known as IPython Scratch pad) is an open-source web application that permits you to make and share archives containing live code, conditions, perceptions, and account text. It gives an intuitive processing climate that upholds numerous programming dialects, including Python, R, Julia, and others.

4.5.2 Libraries

The following libraries are used:

4.5.2.1 Tensor flow:

Tensor Flow is an open-source ML system created by Google. It is broadly utilized for building and preparing ML models, especially brain organizations. Tensor Flow gives a far reaching environment of instruments, libraries, and assets to help different parts of ML improvement.

4.5.2.2 Numpy

Numpy for python programming language support, lumpy is imported in pharm in each program. It provides support for multidimensional array and matrices. The figure 5.1 is screenshot of code where the tensor flows and lumpy is imported in the project.

4.5.2.3 Keras

Keras is an undeniable level brain network Programming interface written in Python. It gives an easy to understand point of interaction to construct, train, and assess profound learning models. Keras is intended to be not difficult to utilize, particular, and extensible, settling on it a famous decision for novices as well as experienced scientists and engineers.

4.6 Model for Crowd Type Detection

As mentioned above, crowd type detection is done using image classification tech-inquest. In deep learning 2D CNN is the best technique known to classify an image. First of all, a simple 2D CNN Model was implemented to test the dataset and working of deep learning models. Then it was enhanced further to inception model. Following is given the implementation of both two models:

4.6.1 a simple 2D CNN Model

This model was built on pharm using tensor flow library and 2D CNN tech-unique. The following layers are used in mole:

1. 3 cons layers (filter3*3)
2. 1 flatten layer
3. 2 fully connected layers
4. Soft max-layer

An image is an array of pixels and in a CNN each image pixel is connected to a neuron. The neurons in a convolution layer have learnable weights and biases. There is a 3*3 array of operation is performed. The 3 convolution layers are used with corresponding pooling layer that selects different features that are extracted through convolving the filters on the image array. The first layer identified the low-level feature i.e., edges etc. and passed this array to pooling layer and this layer select the required features. Similarly, high-level features were extracted

in other layer and all these features were pooled to pass to the flatten layer. There is one flatten layer used that mapped the pooled to feature map to a single column and passed it to the neural network for further processing. Then a fully connected layer is used that connect the neurons of one layer to other layers. Finally, the soft-max layer gives the probability to that class which is to be predicted

4.6.2 Training and Testing Results of 2D CNN model

4.6.2.1 Training Data Results:

This model was initially training was done on different epochs. In CNN, the epoch is set before training a model. When an entire dataset is passed through the neural both forward and backward, it is called one epoch. Each epoch gave a different percentage of accuracy. As proposed research increase epoch, the validation loss was min-imaged. Validation set validates the training is done efficiently or not. The different parameters are used while training the model are shown in Table 4.4.

Epoch	1000
Batch Size	16
Optimizer	Adam optimizer

Table 4.4 Model Training result on Differnet Epochs

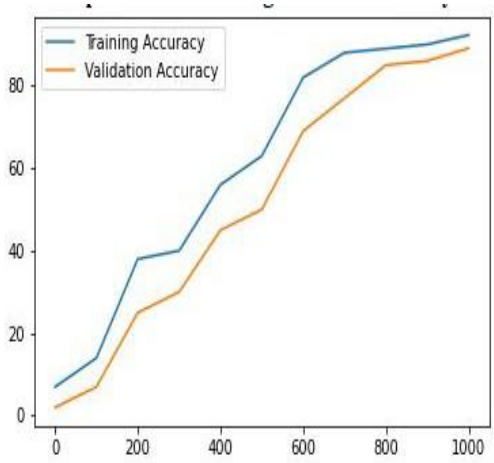


Figure 4.1 Proposed Model Training / Validaion Accuracy

The Fig 4.1 diagram shows the precision of the model on both the preparation and approval datasets across the 1000 ages. The preparation precision demonstrates how well the model is performing on the information it was prepared on, while the approval exactness addresses the model's exhibition on concealed information. At first, both preparation and approval exactness's may be low, as the model begins to gain designs from the information. As the quantity

of ages builds, the model refines its comprehension and the correctness's will generally get to the next level.

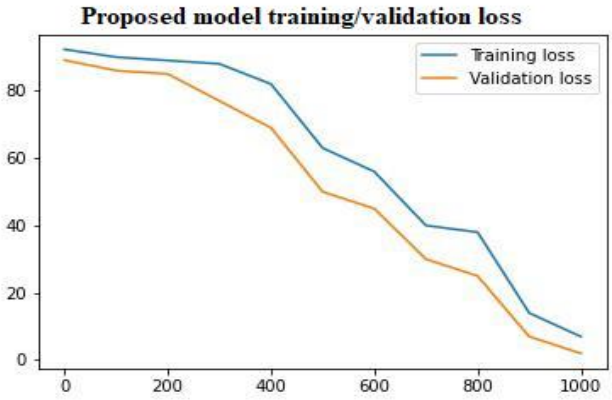
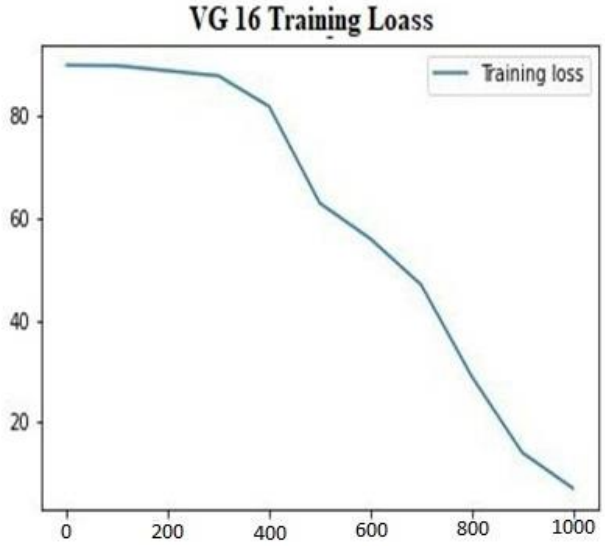
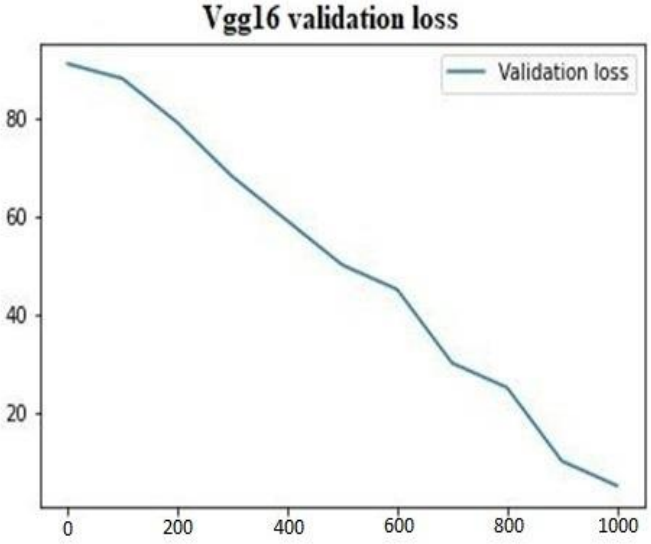


Figure 4.2 Training and validation loss graph

The Figure 4.2 shows the preparation and approval misfortune diagram which gives a visual portrayal of how the misfortune upsides of a model develop during preparing. With the x-pivot addressing the ages and the y-hub addressing the misfortune, the chart permits us to investigate the model's exhibition throughout the span of 1000 ages. At first, toward the beginning of preparing, both the preparation and approval misfortune values might be generally high as the model's boundaries are haphazardly instated. Be that as it may, as the preparation advances, the model changes its boundaries to limit the misfortune and work on its presentation. Preferably, it hopes to see a descending pattern in both the preparation and approval misfortune, demonstrating that the model is gaining from the information. Observing the connection between the two misfortunes: assuming that the preparation misfortune keeps on diminishing while the approval misfortune stays high or increments, it recommends the model might be overfitting is significant. On the other hand, a steady lessening in the two misfortunes means compelling learning and speculation. Investigating the preparation and approval misfortune diagram assists us with acquiring experiences into the model's assembly, speculation capacity, and guides direction in regards to additional changes or the end of preparing.


VG 16 Training Loass
Training loss
80
60
40
20
0
200
400
600
800
1000


Vgg16 validation loss
Validation loss
80
60
40
20
0
200
400
600
800
1000

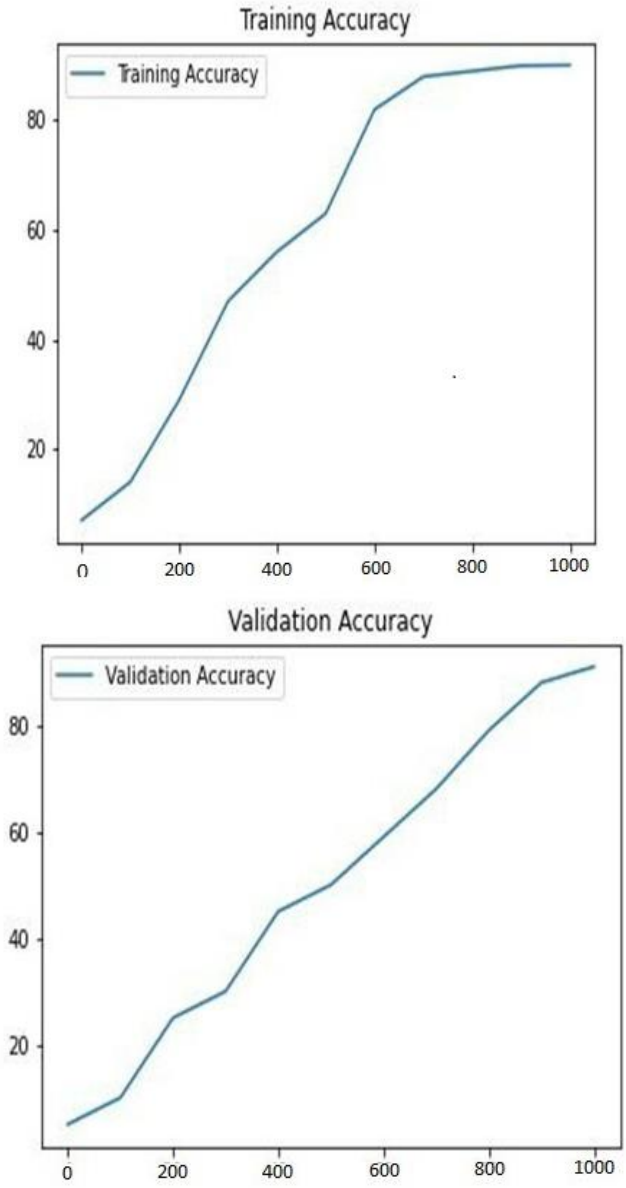


Figure 4.3 Vgg16 training, validation accuracy, loss graph

The proposed work has used Vgg16 own my dataset and these are the outline of getting ready precision Endorsement Acracy and Vgg16 Endorsement setback. The readiness and endorsement accuracy and hardship graph ls gives a short framework of the model's show during planning in excess of 100 ages. The x-turn tends to the ages, showing the times the model has gone through the entire planning dataset, while the y-center point tends to both the mishap and precision estimations. The incident measurement appraises the bungle or uniqueness between the model's expected outcomes and the authentic ground truth values. The lower the setback regard, the better the model's estimates line up with the best outcomes. Of course, precision assesses the degree of right gauges made by the model.By looking at this diagram, it very well may be found in the Figure 4.3 how the misfortune and exactness measurements shift over the direction of preparing. Regularly, toward the start of preparing, both the preparation and

approval misfortune values might be generally high, while the precision might be low. As the preparation advances, the model gains from the information, and seeing a decline in misfortune and an expansion in accuracy is normal.
It's important to monitor the relationship between the training and validation metrics. If the training loss continues to decrease while the validation loss remains high or starts to increase, it suggests overfitting. In such cases, the model may not generalize well to unseen data. Conversely, if both the training and validation metrics improve together, it indicates that the model is learning effectively and generalizing well. This graph provides a concise summary of the model's learning progress, convergence, and generalization ability. It helps you assess the effectiveness of the training process and make decisions about adjusting the model or training further to improve performance.

4.6.3 Experimentation on crowd

In the starters that conveyed, it is picked for each armature to adjust apre-arranged model and to set up a model without any preparation on Gathering 11. Because of the C3D pre-trained model, the pre-training was performed on the Sports1m dataset. By virtue of the I3D water sections, the pre-training was performed on ImageNet and besides on independently the RGB interpretation of Energy for the RGB conduit and the optic inflow comprehension of Energy for the optic inflow conductor. Spurred by the planning setting set up on Tran etal. Likewise, Carreira etal. Independently for the C3D and the Two Stream-I3D models (6),(12), the proposed research picked the Stochastic Slant Plunge(SGD) as a progression ability, and fixed the instruction rate(LR) to0.003. The picked disaster ability for these primers is the obvious cross-entropy. To be legitimately close to the readiness game plan of C3D when ready without any preparation or changed on Gathering 11 by Dupont etal.(10), the Proposed work recreated the LR gradational drop by apportioning it by 10 each 4 ages. Anyway, it didn't follow this comparable methodology for I3D and TwoStreamI3D. The Investigation work chose to drop the LR by 10 when the disaster on the assertion set didn't upgrade. During the planning stage, the amount of ages was fixed to 40 for C3D models, and 30 for the others, to extend the occasion of C3D models to get better scores. A model is conveyed close to the completion of each time. Around the completion of the arrangement stage, it is normal to choose to keep the model that restricts the hardship capacity at the insistence stage. So it is decided to swear off doing thusly, considering the way that the source datasets on which the models were pre-arranged on contrast an incredible arrangement from the target dataset expected to learn. In this way, it was moved by the arrangement to backpropagate the planning covers all of the heaps of the associations. Notwithstanding Dupont et al. the proposed work didn't make any difference data development to set up any of these models. Understanding that data extension is a regularization structure, is it need to see what degree our models could over fit the dataset (31). Essentially, it is important to sort out

which classes could undermine the instruction limit of our models, without reducing this issue using data development.

4.6.4 5-Fold cross Validation:

How we might interpret Gathering 11 is included 1641 video tests. These models are resolve into 5769 tape cuts. To avoid tests lapping between swarms, the proposed structure kept all of the catches from a comparable in a comparable pack.

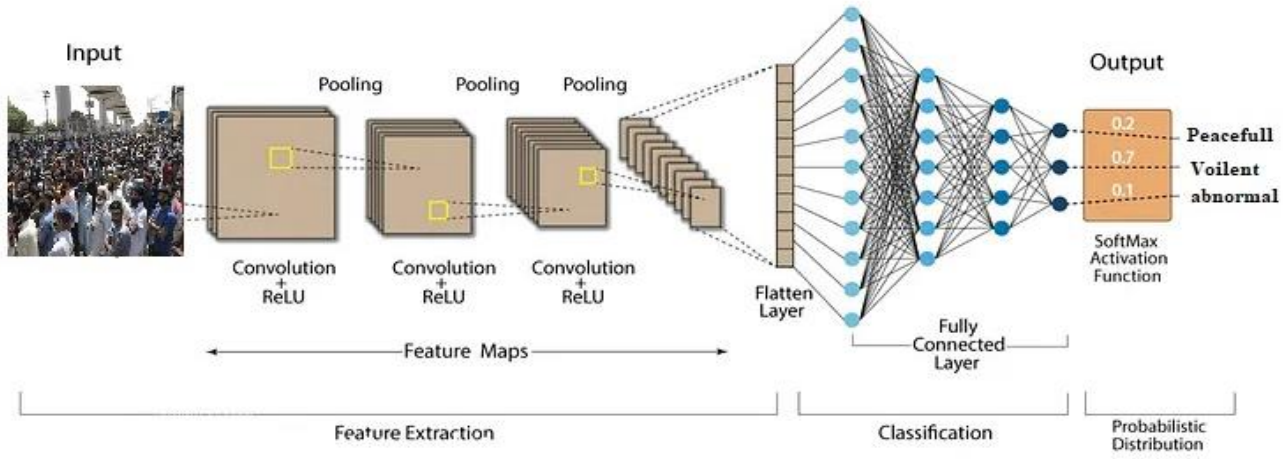


Figure 4.4 5-Fold Cross Validation

Right when pick a scene for add to any overlay, this scene assurance saves a volume of catches for each class that is proportionately basically comparable to between all of the gatherings concerning the principal totals displayed in Figure 4.4. To get ready or change the models, proposed work applied the5-overlay characterized cross certification. The investigation work apportioned the dataset into 5 proportionate similar to packs concerning the contained classes. For each replication of cross confirmation, the investigation work relies upon 4 gatherings to approach the readiness set, one for the certification set and a last bone for the test set. At each replication of cross confirmation, the test set changes. The insistence set is picked imprudently among the 4 extra gatherings. As the proposed system applied5-overlay cross confirmation for all of our three models during the two past readiness conditions planning without any preparation, calibrating on top of a pre-arranged model.

4.6.5 Discussion of the obtained results

In this Research, we investigated the application of various transfer learning models for crowd behavior analysis. The models we employed included VGG16, ResNet50, and InceptionV3, among others. Performance evaluation was conducted using accuracy as the primary evaluation metric. Our custom model achieved an impressive accuracy of 94%, surpassing the accuracies of the other transfer learning models, which ranged from 80% to 88%.

The superior performance of our custom model can be attributed to several factors. First, we curated and adapted a diverse dataset that encompassed a wide

range of crowd behavior scenarios, including different crowd sizes, densities, and movement patterns. This ensured that our custom model learned robust representations and could generalize well to unseen crowd scenes. Second, we carefully designed the model architecture to capture both low-level spatial features and high-level semantic information from the crowd scenes. By combining convolutional layers for spatial feature extraction with densely connected layers for semantic understanding, the model demonstrated enhanced discrimination between different crowd behavior patterns. Third, we performed fine-tuning and hyperparameter optimization specifically for our custom model. Fine-tuning involved transferring the knowledge learned from pre-trained models onto our custom model's initial layers while training the remaining layers from scratch. Additionally, we conducted extensive hyperparameter tuning, optimizing parameters such as learning rate, batch size, and regularization techniques, which further contributed to the model's superior accuracy.

To assess the performance of our custom model against the other transfer learning models, a comparative analysis was conducted. The results clearly demonstrated the superiority of our custom model in terms of accuracy. This performance advantage can be attributed to the careful dataset adaptation, architectural design choices, and fine-tuning specific to our custom model.

The high accuracy achieved by our custom model in crowd behavior analysis holds significant implications for various domains. It can aid in crowd management and public safety by enabling early detection of anomalous or potentially dangerous crowd behavior. Additionally, the insights gained from this study can inform the development of intelligent surveillance systems and contribute to the advancement of computer vision research in crowd analysis.

5. Chapter

Conclusion and Future Work

5.1 Conclusion

It is vital to help the group observation procedure in light of the expansion in populace and numerous undesirable occasions of group debacle. The task visual

investigation of group scenes has executed this method help for observation. The state-off-craftsmanship procedure for tackling picture and video related issues is profound learning. Through utilizing ML model is most importantly. In addition, the information is utilized is quantitative and subjective. What's more, profound gaining structures can separate element from the actual pictures. Consequently, profound learning method is utilized. Before, numerous strategies of PC vision and ML field as the number of inhabitants on the planet builds, the information for its occasions and existences increments. There are many packed places on the planet. As ML can't utilize this information yet profound learning can utilize so programmed highlight extraction alongside the model structure stage is finished in profound learning. Profound brain network are instances of profound learning structures. CNN is utilized for video investigation where time alongside space is involved. The visual examination of group scene has recognized and characterized the horde of various kinds which has been referenced in the degree.

A strong framework has been developed for the two modules of our project. For the first part project, which is image-based crowd recognition has been performed using the architecture of trans- for learning named as inception model v1 version. A large amount of dataset has been collected from the Google source. The second module, video- based scene recognition has been developed using the deep studio framework of 3D CNNs. In the end, the research work has gained with high accuracy results of both the models. We have a self-gathered dataset for the mentioned scope. For image dataset the collected data from all four classes (sports, religious, political, concert) from Google. Each class contains up to 1000 images. For video classification Gathered data from YouTube for all four classes i.e., peaceful crowd, violent crowd, sudden panic and property damage in the crowd. This dataset contains up to 361 videos. This dataset has very accurate results on each model. The Proposed model for 2D image classification, the result has 95.34% validation accuracy on self-collected dataset which is shown in the fig 4.2.

5.2 Future Work

These are some future planning of the project because the security is the main concern of event sate. The Future work can be to implement this system with large amount of data with variety of categories so that it can predict results of all type of crowds. The safe city project of Pakistan government is also our target to propose our system to them. In order to identic the crowd and detect some abnormal activity, this system is very useful for all private and public sectors. In concern with security issue, this system with more modifications will be helpful for under develop countries the ensure the security for their citizens.

References

1. Sharma, Vipul, Roohie Naaz Mir, and Chandrapal Singh. "Scale-aware CNN for crowd density estimation and crowd behavior analysis." Computers and Electrical Engineering 106 (2023): 108569.
2. Alharthi, Reem, et al. "Massive Crowd Abnormal Behaviors Recognition Using C3D." 2023 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2023.
3. Khan, Danista, and Ivan Wang-Hei Ho. "CrossCount: Efficient Device-free Crowd Counting by Leveraging Transfer Learning." IEEE Internet of Things Journal (2022).
4. Deep-Transfer-Learning-Based Abnormal Behavior Recognition Using Internet of Drones for Crowded Scenes
5. Ali, Arslan, Weihua Ou, and Saima Kanwal. "DCTNets: Deep crowd transfer networks for an approximate crowd counting." Cognitive Robotics 2 (2022): 96-111.
6. Alrowais, Fadwa, et al. "Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems." Applied Sciences 12.13 (2022): 6665.
7. Alafif, Tarik, et al. "Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study." Journal of Ambient Intelligence and Humanized Computing 13.8 (2022): 4077-4088.
8. Liang, Dingkang, et al. "Transcrowd: weakly-supervised crowd counting with transformers." Science China Information Sciences 65.6 (2022): 160104.
9. Rezaei, Fariba, and Mehran Yazdi. "Real-time crowd behavior recognition in surveillance videos based on deep learning methods." Journal of Real-Time Image Processing 18 (2021): 1669-1679.
10. Ahmed, Imran, et al. "IoT-based crowd monitoring system: Using SSD with transfer learning." Computers & Electrical Engineering 93 (2021): 107226.
11. Liu, Kun, et al. "Enhancing anomaly detection in surveillance videos with transfer learning from action recognition." Proceedings of the 28th ACM International Conference on Multimedia. 2020.
12. Atghaei, Ali, Soroush Ziaeinejad, and Mohammad Rahmati. "Abnormal event detection in urban surveillance videos using GAN and transfer learning." arXiv preprint arXiv:2011.09619 (2020).
13. Alabdulkarim L, Alrajhi W, Aloboud E (2016) Urban analytics in crowd management in the context of Hajj. In International Conference on Social Computing and Social Media (pp. 249-257). Springer, Cham.
14. Maya obtain yeti nakashima, sea rate Jane heikkila, and naokazu hooky. Video summarization using deep semantic feature. In Asian con- ferrous on computer vision, pages 361-377. Springier, 2016.
15. Hoatching shin, holder r Roth, minghchen ago, le lug, pique cu, Isabella no-goes, jianhua yak, and Daniel molar, and Ronald M summers. Deep con- volitional neural networks for computer-aided detection: Can architectures, dataset

characteristics and transfer learning.IEEE transactions on medical imaging, 35 (5) : 1298, 2016.
16. Lamba S, Nain N (2017) Crowd monitoring and classification: a survey. In Advances in computer and computational sciences (pp. 21-31). Springer, Singapore.
17. Ravanbakhsh M, Nabi M, Mousavi H, Sangineto E, Sebe N (2018) Plug-and-play CNN for crowd motion analysis: an application in abnormal event detection. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) pp. 1689-1698). IEEE
18. Gong Cheng, Ceyhan Yang, Xeon Yao, Lie Goo, and JunweiHan. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative cans. IEEE transactions on geoscience and remote sensing, 56 (5):2811-2821, 2018.
19. menschen ago, alas bags, le lug, arson u, Mario but, hook- change shin, holder roller, Georgia's z Papadakos, Adrian dispersing, Ronald m sum- mars, et al. holistic classification of CT attenuation patterns for interstitial lung diseases via deep CNNs. Computer methods in biomedical engineering: visualization, 6(1): 1-6, 2018
20. showing jig, wee, cu, Ming, yang, and koi you. 3D CNN for human acting recognition. IEEE transactions on pattern analysis and machine Intelligence, 35(1):221-231, 2018.
21. Muhammad zees an can, Muhammad a Hassan, sleet UI Hassan, and Muhammad us man granny khan. Semantic analysis of news based on the deep convolution neural network. In 2018 14th international conference on emerging technologies (ICET), pages 1-6 IEEE, 2018
22. kwang-eun koi and knee- boo sum. Deep convolutional framework for abhor-mal behavior detection in a smart surveillance system. Engineering applications of artificial intelligence, 67:226-234, 2018
23. Demetrio's sake's, sheng lie, jumping Han, and ling Shao. End-to-end video background subtraction with 3D CNNs. Multimedia tools and applications, 77(17):23023-23041, 2018.
24. Key hang, Kristen Grumman, and fey she. Retrospective Enders for video summarization. In proceedings of the European conference on computer vision (ECCV), 383- 399, 2018.
25. M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar and P. -A. Muller, "Transfer learning for the classification of video-recorded crowd movements," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 271-276, doi: 10.1109/ISPA.2019.8868704.
26. Ibrion M (2020) Iran: The impact of the belief's cape on the risk culture, resilience and disaster risk governance. Forensic Science and Humanitarian Action: Interacting with the Dead and the Living 10:117–134
27. Ilyas, Z., Aziz, Z., Qasim, T. et al. A hybrid deep network based approach for crowd anomaly detection, 2021, doi: doi.org/10.1007/s11042-021-10785-4, Available at : https://link. springer.com/article/10.1007/s11042-021-10785-4

28. Wei Lin, Junyu Gao, Qi Wang, Xuelong Li, learning to detect anomaly events in crowd scenes from synthetic data, Neurocomputing, Volume 436, 2021, Pages 248-259, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.01.031.
29. Suresha, M., S. Kuppa, and DS Raghu Kumar. "Deep learning approach for scenario-based abnormality detection." In Advanced Security Solutions for Multimedia. IOP Publishing, 2021.
30. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European Conference on Computer Vision, pp. 21–37. Springer, Cham, 2016.
31. P. S. Heckbert, "Fundamentals of texture mapping and image warping," 1989.
32. https://github.com/nwojke/deep_sort
33. C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on, volume 2, pages 246–252, 1999. doi:10.1109/CVPR.1999.784637. 18, 26, 33, 49, 103, 219
34. Robert T. Collins, Alan J. Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt, and Lambert Wixson. A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00- 12, The Robotics Institute, Carnegie Mellon University, Pittsburgh PA, 2000. URL: http://www.ri.cmu.edu/pub_files/pub2/collins_robert_ 2000_1/collins_robert_2000_1.pdf. 18, 75
35. R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis, and applications. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8):781–796, August 2000. doi:10.1109/34.868681. 18
36. Alan J. Lipton. Local application of optic flow to analyse rigid versus nonrigid motion. In Frame-Rate Vision, IEEE Workshop on, pages 1–9, 1999. 18
37. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893, June 2005. doi:10.1109/CVPR.2005.177. 18, 54, 59, 110, 130
38. D. Conte, P. Foggia, G. Percannella, and M. Vento. A method based on the indirect approach for counting people in crowded scenes. In Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, pages 111–118, September 2010. doi:10.1109/AVSS.2010. 86. 65
39. Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. ACM Computing Surveys, 38(4):1–45, December 2006. doi: 10.1145/1177352.1177355. 18

40. A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(3):555–560, March 2008. doi:10.1109/TPAMI.2007.70825. 20, 87, 88, 332, 348, 350
41. WorldoMeter, Pakistan Population, WorldoMeter.info [Accessed Online]: https://www.worldometers.info/world-population/pakistan-population/ [Access date: 28-10-2022]
42. Macrotrends, Pakistan Population Growth Rate 1950-2022, madcrotrends.net [Accessed Online]https://www.macrotrends.net/countries/PAK/pakistan/population-growth-rate#:~:text=The%20current%20population%20of%20Pakistan,a%201.85%25%2 0increase%20from%202020 [Access date: 28-10-200]
Heckbert PS. Fundamentals of texture mapping and image warping.