VIDEO BASED CROWD ANALYSIS USING DEEP LEARNING

Submitted in partial fulfillment of the

requirements for the award of

Bachelor of Engineering degree in Computer Science and Engineering

By

GOPU SHAJI (Reg.No - 39110347)
GORANTLA GURUVARDHAN KUMAR (Reg.No - 39110348)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade “A” by NAAC | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI - 600119

APRIL - 2023

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **GOPU SHAJI (39110347)** and **GORANTLA GURUVARDHAN KUMAR (39110348)** who carried out the Project Phase-2 entitled **"VIDEO BASED CROWD ANALYSIS USING DEEP LEARNING"** under my supervision from January 2023 to April 2023.

Internal Guide

Dr. L. SUJI HELEN, M.E., Ph.D.

Head of the Department

Dr. L. LAKSHMANAN, M.E., Ph.D.

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF CSE

Submitted for Viva voce Examination held on 20.4.2023

Internal Examiner **External Examiner**

DECLARATION

I, **GOPU SHAJI (Reg.No - 39110347),** hereby declare that the Project Phase-2 Report entitled **VIDEO BASED CROWD ANALYSIS USING DEEP LEARNING"** done by me under the guidance of **Dr. L. SUJI HELEN, M.E., Ph.D.** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

DATE: 20.4.2023

PLACE: Chennai **SIGNATURE OF THECANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph. D**, **Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. L. Suji Helen, M.E., Ph.D,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-1 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

A system that analyzes crowd behavior in real time and alerts you if abnormalities occur is known as crowd behavior analysis. Pedestrian streets, supermarkets, bus stand and markets are usually crowded. It is always important to ensure the safety and comfort of large crowds, so appropriate measures must be taken to ensure their safety. This article introduces Convolutional Neural Networks and the deep learning process used to study common scenes. In these documents it's proposed to find statistics on the number of people who arrived in each group, as well as a table of the density of the group. Counting human beings in a dense crowd is an essential step in surveillance and anomaly warning. Existing population estimation methods are based on homemade capabilities such as SIFTS and HOG. Currently, the most suitable technique is for expecting the high overall performance of density and crowd estimation based on a deep understanding of networks. The purpose is to offer a complete evaluation on various kinds of crowd analysis, including crowd counting, crowd detection, anomaly detection, and human behavior. The proposed work uses a deep learning algorithm to analysis the crowd. It performs better accuracy when compared with the other algorithms.

LIST OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Automatic evaluation of crowd scene analysis refers to investigating the behavior of a large group of people sharing the same physical area. Typically, it counts the number of individuals per region, tracks the common individuals trajectories and recognizes individuals behaviors. Automatic crowd scene analysis has many essential applications.

It controls the spread of the COVID-19 virus by ensuring physical distance between individuals in stores, parks, etc. Analysis of crowd scenes of public places such as train stations, stores, and shopping malls can show the effect of crowd path or the shortcomings of the design. Mass scenes evaluation presents surveillances camera.

In addition, the evaluation of crowd scenes in public places, together with supermarkets, teach stations and shopping department shops, can display the impact of crowd motion or layout flaws. In this the project objectives to provide a perspective at the improvement of a crowd evaluation of deep learning techniques. This angle specializes in two components of crowd analysis: crowd counting and awareness of crowd action.

1.1 Crowd Analysis

- Automatic evaluation of crowd scene analysis refers to investigating the behavior of a large group of people sharing the same physical area.
- Typically, it counts the number of individuals per region, tracks the common individual trajectories and recognizes individual behaviors. Automatic crowd scene analysis has many essential applications.

1.2 Motivation of Crowd Analysis

- It controls the spread of the COVID-19 virus by ensuring physical distance between individuals in stores, parks, etc.
- Analysis of crowd scenes of public places such as train stations, stores, and shopping malls can show the effect of crowd path or the shortcomings of the design.
- Mass scenes evaluation presents surveillances camera.
- In addition, the evaluation of crowd scenes in public places, together with supermarkets, teach stations and shopping department shops, can display the impact of crowd motion or layout flaws.

1.3 Objectives of Crowd Analysis

- In this project the objectives is to provide a perspective at the improvement of a crowd evaluation of deep learning techniques. This angle specializes in two components of crowd analysis: crowd counting and awareness of crowd action.

CHAPTER 2

LITERATURE SURVEY

1. W. Liu, K. M. Lis, M. Salzmann, and P. Fua, “Geometric and physical constraints for drone-based head plane crowd density estimation,” in Proc. Int. Conf. Intell. Robots Syst., 2020.

Current methods for counting people in dense community scenes depend upon estimating the population density on the image level. This photo, whilst density is without a doubt useful for this purpose, has no instantaneous physical meaning, as it's far concern to perspective distortion.

2.Jane et al., "Discrete Residual Flow Probabilistic Prediction of Pedestrian Behavior," in Proc. CORL, 2010.

Self-using vehicles control each static and dynamic gadgets via making use of predictive behavior fashions to estimate the future location of gadgets within the environment. But destiny conduct is inherently unsure, and the motion styles that produce deterministic consequences are restricted to short time scales.

3.D. Roy, T. Ishizaka, S. K. Mohan, and A. Fukuda, "Vehicle Path Prediction at Intersections Using Interaction-Based Generative Adversarial Networks," in Proc. ITSC, 2021, pp. 2318-2323.

Vehicle trajectory prediction at intersections is an vital and tough mission for autonomous automobile navigation. This trouble is exacerbated while traffic is ruled by means of small vehicles, which frequently do no longer obey the lane, as is the case in lots of growing nations.

4.S. Marcel, H. Stefan, and D. Klaus, “Long-term occupancy grid prediction using recurrent neural networks,” in Proc. Int. Conf. Robot. Automat, 2020, pp. 9299–9305.

We are engaged in long-time period prediction of the improvement degree at the middle of a complex state undertaking for computerized using primarily based on a aggregate of lidar grids and recurrent neural networks (RNN). A chicken's eye view of the scene, such as occupancy and pace, is then handed to an RNN this is skilled to expect destiny occupancy.

5.Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Resatofigi, and S. Savarese, "Sophie: a memory gun for predicting social paths and shaping physical constraints," in Proc. Conf. Account You want. Model Review, 2020, pp. 1349-1358.

This article offers with the problem of path prediction for a couple of interacting sellers in a scene, which is an crucial step for many self sufficient systems, such as self-using motors and social robots. Introducing an interpretable framework based on a generative opposed network (GAN) that uses resources of facts, the history of all agents in the scene, and context statistics using scene photographs.

6.N. Deo and M. M. Trivedi, “Scene induced multi-modal trajectory forecasting via planning,” 2021, arXiv:1905.09949.

We convert the multimodal prediction of agent trajectories into unknown scenes and country it as a scheduling hassle. We present an method such as three fashions; a signal prediction version to pick out capacity energetic objectives, a getting to know reduction model to assist in deploying the pleasant routes to every target, and producing trajectories to derive future trajectories along planned paths. Analysis of predictions on the Stanford paint dataset suggests the generality of our method to novel scenes.

7.Q. Wang, J. Gao, W. Lin, and Yu, Yuan, “Exploring Synthetic Data for Wildlife Crowd Computing, “in Proc. Conf. Account you need. Model Review, 2021, pp.

Recently, people computing for crowd scenes is a hot topic for a wide open sotware (eg: surveillance, public protection). This work inside the wild is hard: the converting environment, the huge number of humans means that the current methods work properly. In addition, because of the shortage of information, many methods be afflicted by issue to various stages. Addressing these troubles, we first developed a statistics collector and labeler which could generate synthetic crowd scenes and examine them simultaneously without intervention.

2.1 INFERENCES FROM LITERATURE SURVEY

- From the above-mentioned literature works, it is clear that there has been effective research on how many people in room with optimal path finding and many models have been proposed.
- It is evident that the above-mentioned systems have their own pros and cons.
- While some of the recent works involve hybrid technologies and provide better accuracies, they are still far from what is needed.

2.2 EXISTING PROBLEM

- The task of crowd counting and density map estimation is riddled with many challenges such as occlusions, non-uniform density, intra-scene and inter-scene variations in scale and perspective.
- Nevertheless, over the last few years, crowd count analysis has evolved from earlier methods that are often limited to small variations in crowd density.

CHAPTER 3

REQUIREMENTS ANALYSIS

3.1 DRAWBACKS

- Data dependency: Deep learning models depend heavily on large amounts of labeled data to train and generalize well. However, collecting and annotating large datasets for crowd analysis is often expensive and time-consuming.

- Computational complexity: Deep learning models for crowd analysis can be computationally expensive and require significant computing resources to train and run. This can limit their use in real-time applications and on resource-constrained devices.

- Sensitivity to lighting and occlusion: Video-based crowd analysis systems can be sensitive to changes in lighting conditions and occlusions in the scene, which can affect the accuracy of object detection and tracking.

- Privacy concerns: Video-based crowd analysis systems that use deep learning may raise privacy concerns, as they can potentially capture and store sensitive information about individuals in public spaces.

- Interpretability: Deep learning models are often described as "black boxes" because it can be difficult to interpret how they make decisions. This can make it challenging to understand the reasoning behind the outputs of a video-based crowd analysis system.

- Generalization: Deep learning models can struggle to generalize to new or unseen scenarios that differ significantly from the training data. This can limit the applicability of a video-based crowd analysis system in diverse settings.

- Lack of diversity in training data: Many existing datasets for crowd analysis are collected in specific settings, such as train stations or sports stadiums, which may not reflect the diversity of real-world crowd scenarios.

3.2 SYSTEM REQUIREMENTS

Requirement analysis determines the requirements of a new system. This project analyses on product and resource requirement, which is required for this successful system. The product requirement includes input and output requirements it gives the wants in term of input to produce the required output. The resource requirements give in brief about the software and hardware that are needed to achieve the required functionality.

3.3 SOFTWARE REQUIREMENTS

The software requirements are the specification of the system. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team's and tracking the team's progress throughout the development activity.

- Operating system : Windows 7/8/10/11
- Coding Language : Python 3.8 Version

3.4 HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the systems do and not how it should be implemented.

- Processor : intel i5/i7/i9
- Disk space : 3 GB.
- Ram : 4GB.

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

4.1 PROPOSED SYSTEM

- In this proposed system deep learning algorithm is used to detect the crowd analysis.

- The proposed work are divided into two modules: crowd counting and crowd action recognition.

- Crowd scene datasets are used in crowd action recognition.

- This proposed algorithm will predict the crowd analysis with high accuracy.

Crowd Counting

Crowd counting is a method of counting or estimating the number of humans in an picture. Accurately estimating the variety of human beings/objects in a single image is a tough but critical mission that is used in many packages together with city making plans and public protection. Population counting stands out a number of the diverse objects of counting tasks because of its special significance for security and social improvement. Before we move on, let's assessment the numerous businesses that the common photo analysis organization considers.

Kinds of crowd quantity

There are sorts of crowd counting
1. Traditional method
2. Normal arrival

Traditional approach

1. Object degree detection
2. Image / pixel detection
3. Mesh / leak detection

4.2 PROPOSED ARCHITECTURE DIAGRAM

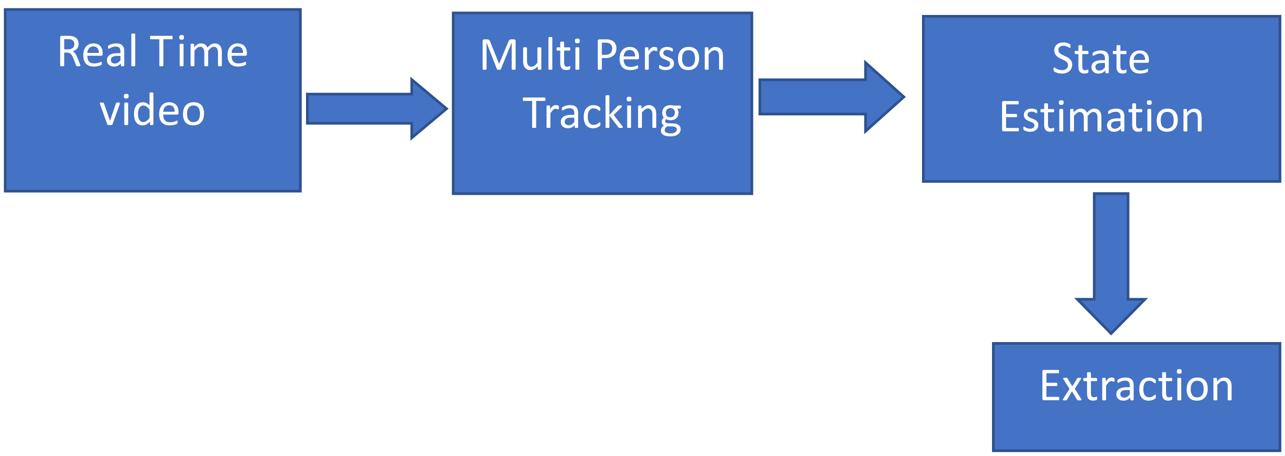


Fig 4.1 Proposed Architecture Diagram of the video human analysis

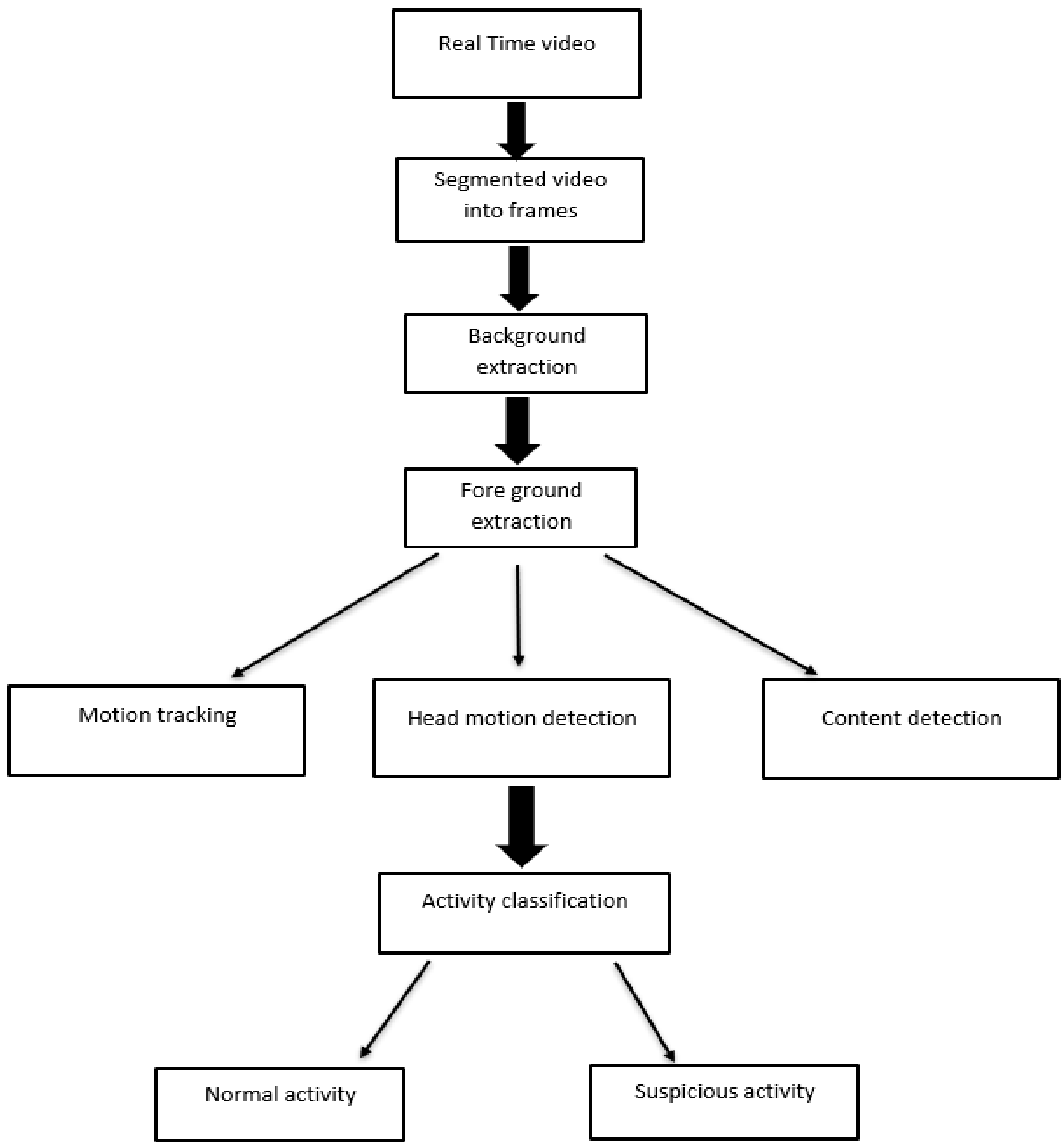


Fig 4.2 System Model of the analysis

4.3 FLOW DIAGRAM OF PROPOSED SYSTEM

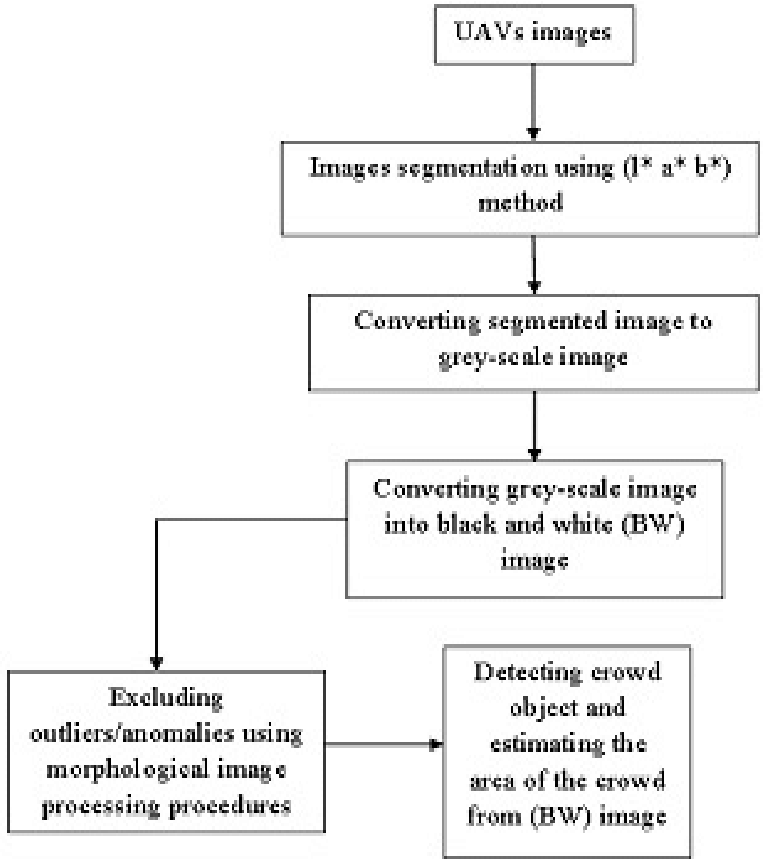


Fig 4.3 Flow Diagram of Proposed System

4.4 ALGORITHM AND METHODS

Modules

1. Collect the data set
2. Multi person tracking
3. State estimation
4. Feature extraction

1. Collect the data set :

- Collect the information and data by using the real time video. The data's are object and people.

2. Multi person tracking :

- It takes detection, prediction, and data association to track many things. Find interesting objects in a video frame using detection. Prediction: Predict where the objects will be in the following frame. Data association: Create tracks by connecting detections across frames using the expected positions.

3. State estimation :

- The aim of state estimation is to adjust the model results such that they are closer to observed values, thus generating improved ocean re analyses.

4. Feature extraction :

- Feature extraction helps to reduce the amount of redundant data from the data set. In the end, the reduction of the data helps to build the model with less machine effort and also increases the speed of learning and generalization steps in the machine learning process.

List of Modules

- Data pre-processing
- Feature extraction
- Build model
- Result

- **Data Wrangling**

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

- **Data collection**

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms and Support vector classifier (SVC) are applied on the Training set and based on the test result accuracy, Test set prediction is done.

- **Data-Preprocessing**

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done.

- **Building the classification model**

The predicting the Human activity recognition, decision tree algorithm prediction model is effective because of the following reasons: It provides better results in classification problem.

It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.

It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

- **Construction of a Predictive Model**

Machine learning needs data gathering have lot of past data's. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can't be used directly. It's used to preprocess then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

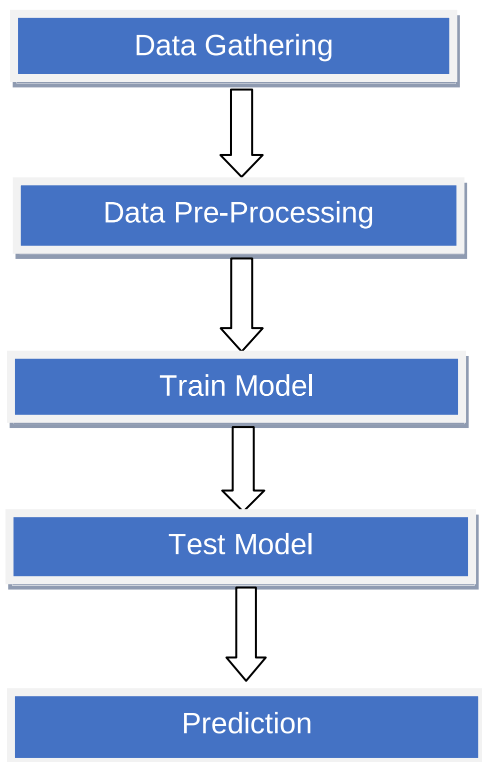


Fig 4.4 Process of dataflow diagram

ALGORITHM

CNN ALGORITHM

- A CNN is a deep learning algorithm that is used for image recognition.

- Convolutional Neural Network is one of the main categories to do image classification and image recognition.

- Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as **h * w * d**, where h= height w= width and d= dimension. For example, An RGB image is **6 * 6 * 3** array of the matrix, and the grayscale image is **4 * 4 * 1** array of the matrix.

Convolution Layer

Convolution layer is the first layer to extract features from an input image. By learning image features using a small square of input data, the convolutional layer preserves the relationship between pixels. It is a mathematical operation which takes two inputs such as image matrix and a kernel or filter.

- The dimension of the image matrix is $\mathbf{h \times w \times d}$.
- The dimension of the filter is $\mathbf{f_h \times f_w \times d}$.
- The dimension of the output is $\mathbf{(h-f_h+1) \times (w-f_w+1) \times 1}$.

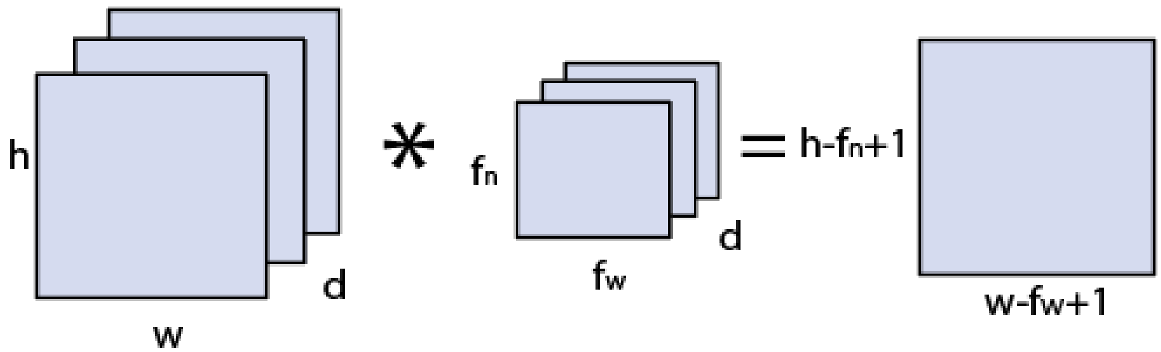


Image matrix multiplies kernl or filter matrix

Let's start with consideration a 5*5 image whose pixel values are 0, 1, and filter matrix 3*3 as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

5×5 — Image Matrix $\quad$ 3×3 — Filter Matrix

The convolution of 5*5 image matrix multiplies with 3*3 filter matrix is called "**Features Map**" and show as an output.

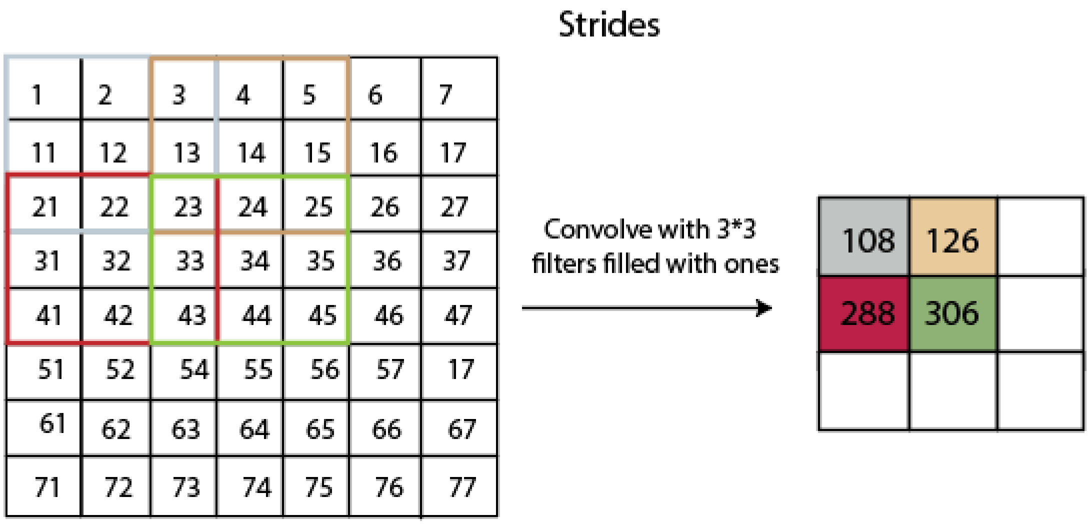
$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

Convolved Feature

Convolution of an image with different filters can perform an operation such as blur, sharpen, and edge detection by applying filters.

Strides

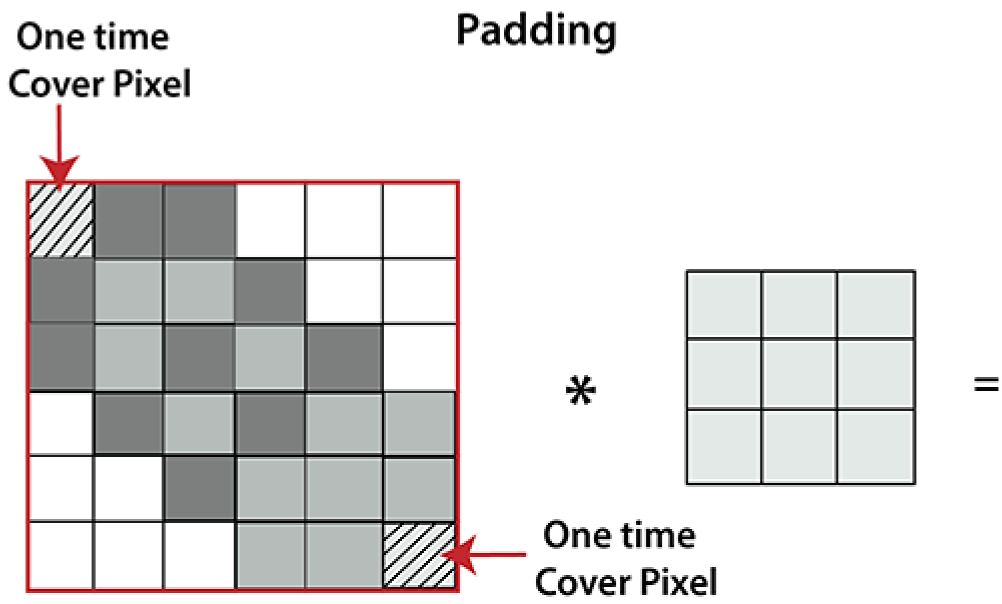
Stride is the number of pixels which are shift over the input matrix. When the stride is equal to 1, then we move the filters to 1 pixel at a time and similarly, if the stride is equal to 2, then we move the filters to 2 pixels at a time. The following figure shows that the convolution would work with a stride of 2.



Padding

Padding plays a crucial role in building the convolutional neural network. If the image will get shrink and if we will take a neural network with 100's of layers on it, it will give us a small image after filtered in the end.

If we take a three by three filter on top of a grayscale image and do the convolving then what will happen?



It is clear from the above picture that the pixel in the corner will only get covers one time, but the middle pixel will get covered more than once. It means that we have more information on that middle pixel, so there are two downsides:

- Shrinking outputs
- Losing information on the corner of the image.

To overcome this, we have introduced padding to an image. **"Padding is an additional layer which can add to the border of an image."**

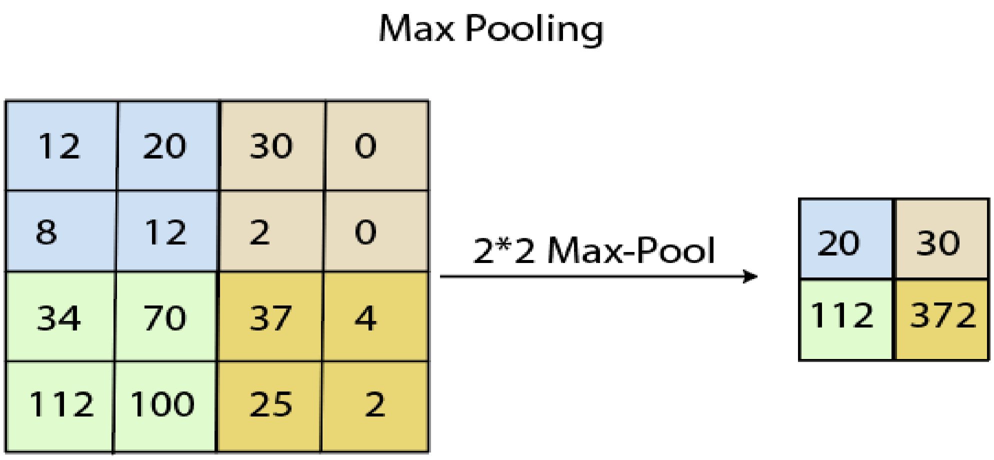
Pooling Layer

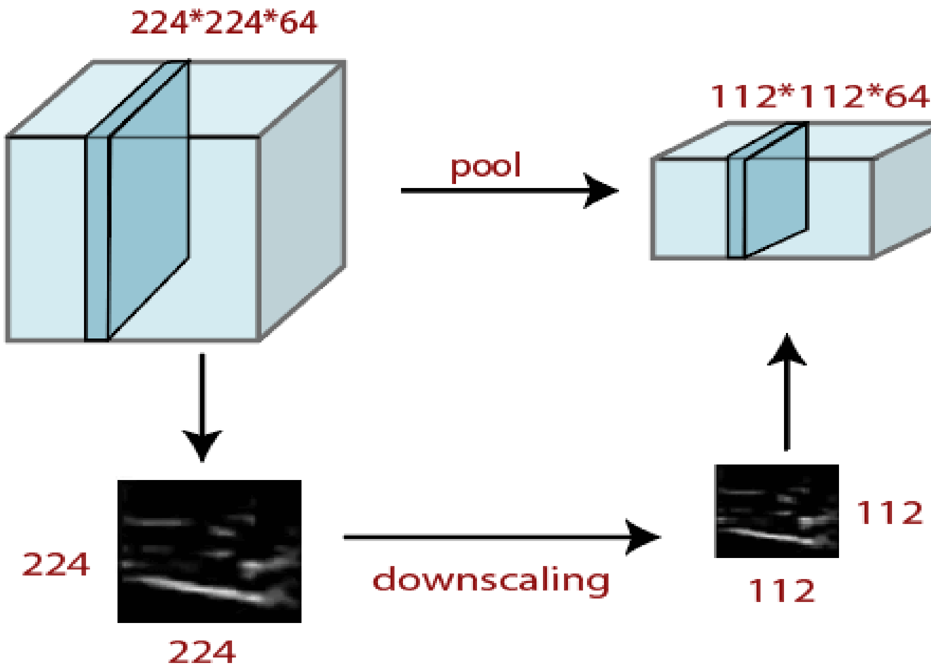
Pooling layer plays an important role in pre-processing of an image. Pooling layer reduces the number of parameters when the images are too large. Pooling is "**downscaling**" of the image obtained from the previous layers. It can be compared to shrinking an image to reduce its pixel density. Spatial pooling is also called downsampling or subsampling, which reduces the dimensionality of each map but retains the important information. There are the following types of spatial pooling.

Max Pooling

Max pooling is a **sample-based discretization process**. Its main objective is to downscale an input representation, reducing its dimensionality and allowing for the assumption to be made about features contained in the sub-region binned.

Max pooling is done by applying a max filter to non-overlapping sub-regions of the initial representation.





Average Pooling

Down-scaling will perform through average pooling by dividing the input into rectangular pooling regions and computing the average values of each region.

Syntax

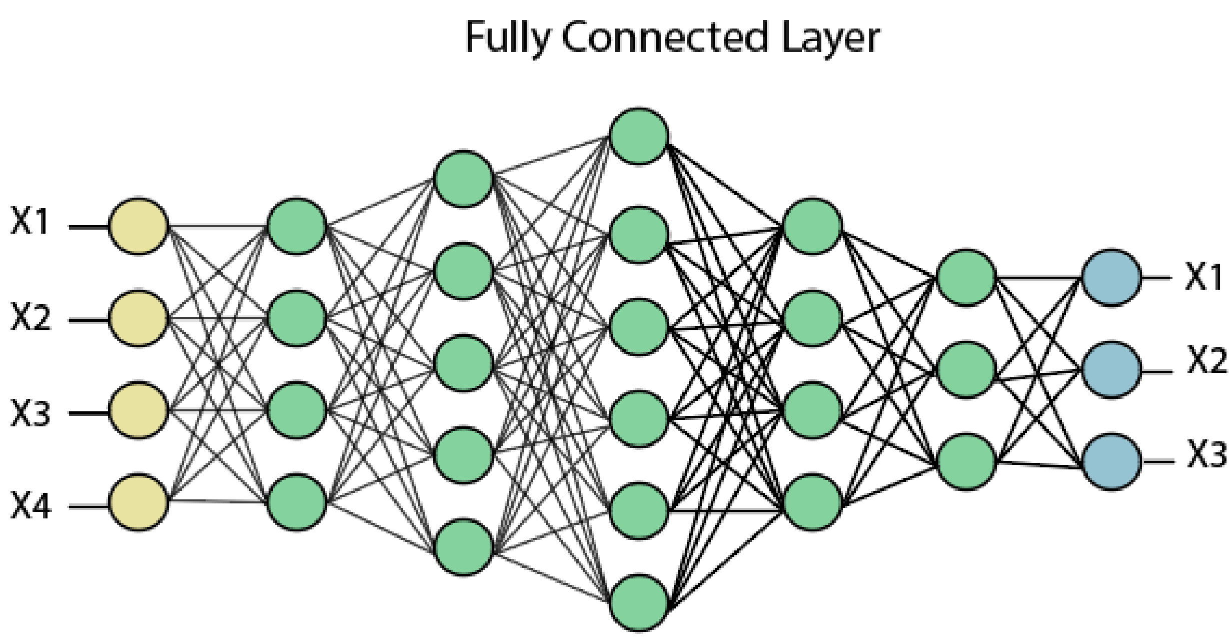
layer = averagePooling2dLayer(poolSize)
layer = averagePooling2dLayer(poolSize,Name,Value)

Sum Pooling

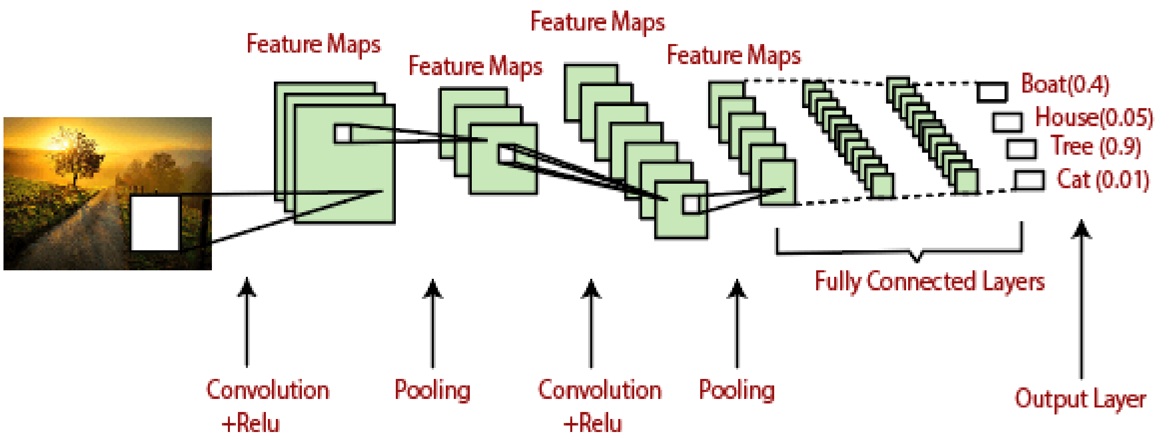
The sub-region for **sum pooling** or **mean pooling** are set exactly the same as for **max-pooling** but instead of using the max function we use sum or mean.

Fully Connected Layer

The fully connected layer is a layer in which the input from the other layers will be flattened into a vector and sent. It will transform the output into the desired number of classes by the network.



In the above diagram, the feature map matrix will be converted into the vector such as **x1, x2, x3... xn** with the help of fully connected layers. We will combine features to create a model and apply the activation function such as **softmax** or **sigmoid** to classify the outputs as a car, dog, truck, etc.



CHAPTER 5

RESULT AND DISCUSSION

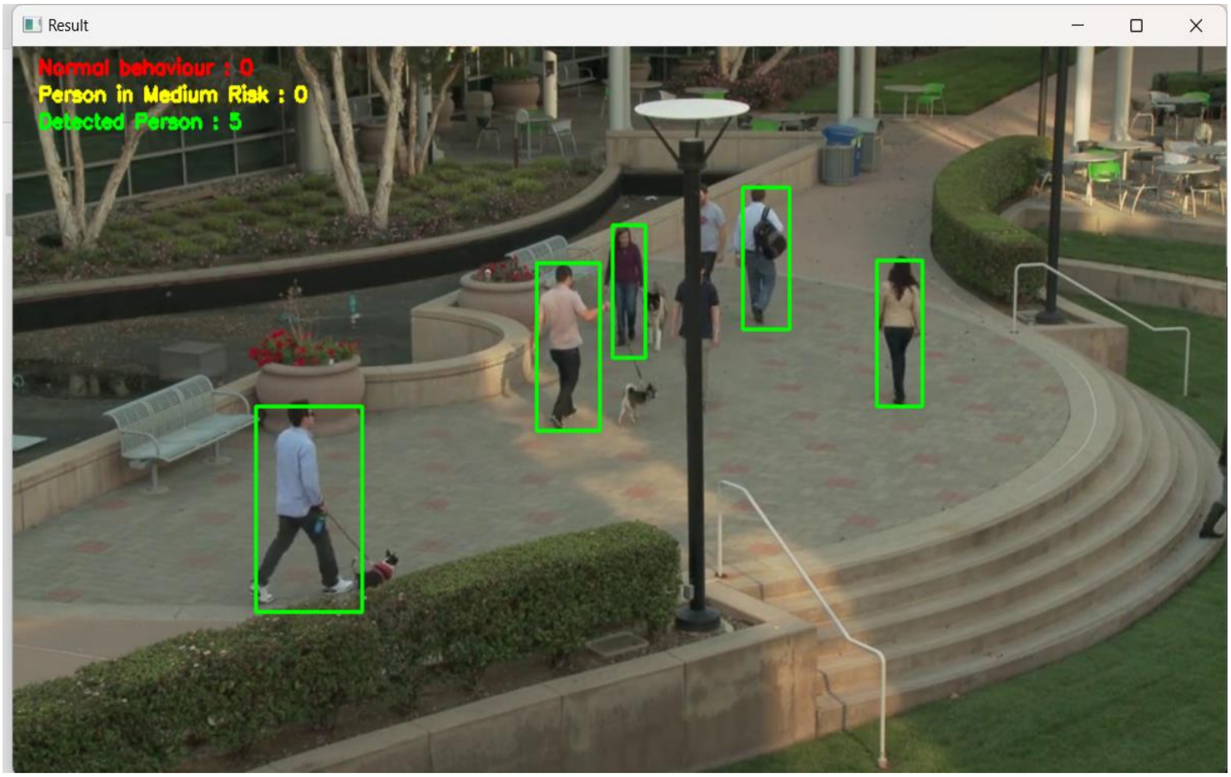
OUTPUT :

Fig 5.1 Screenshot of the output

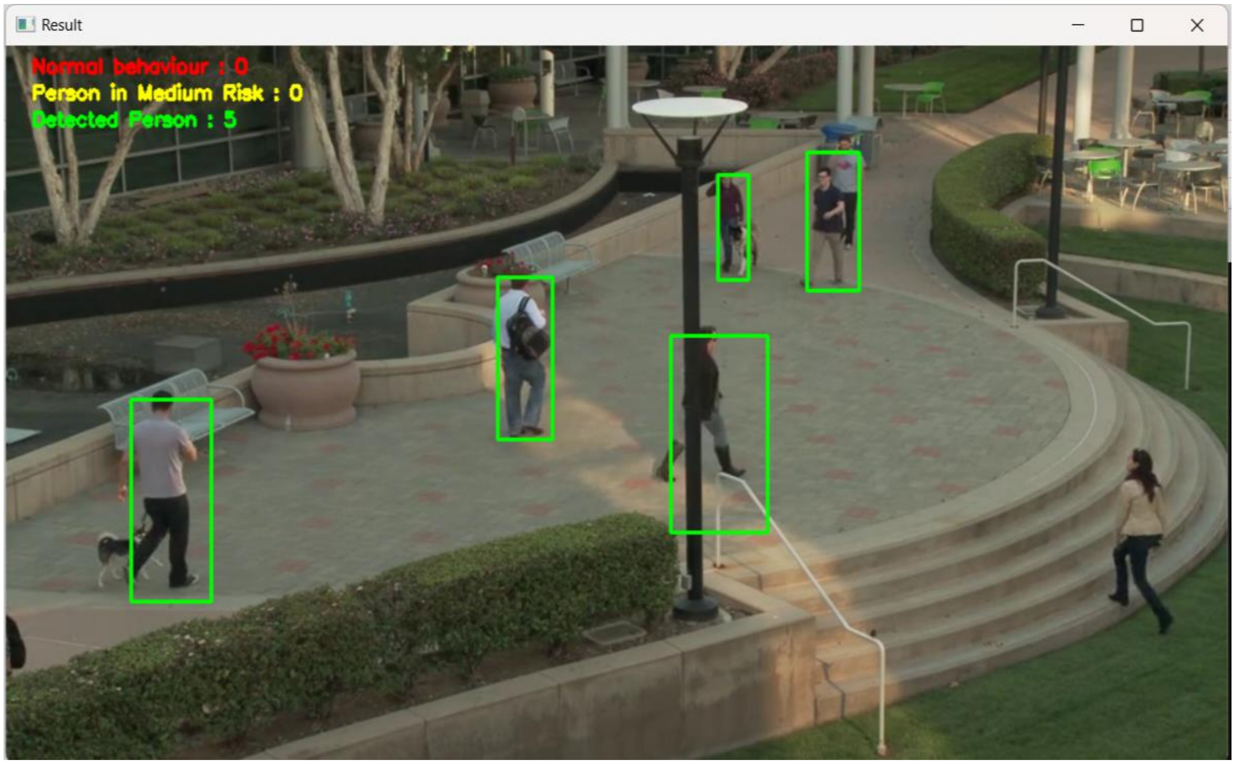
- shows the peoples count and how the object is tracked. Table 1 shows the proposed work is compared with the existing work. The proposed work has high accuracy and detect the people accurately.

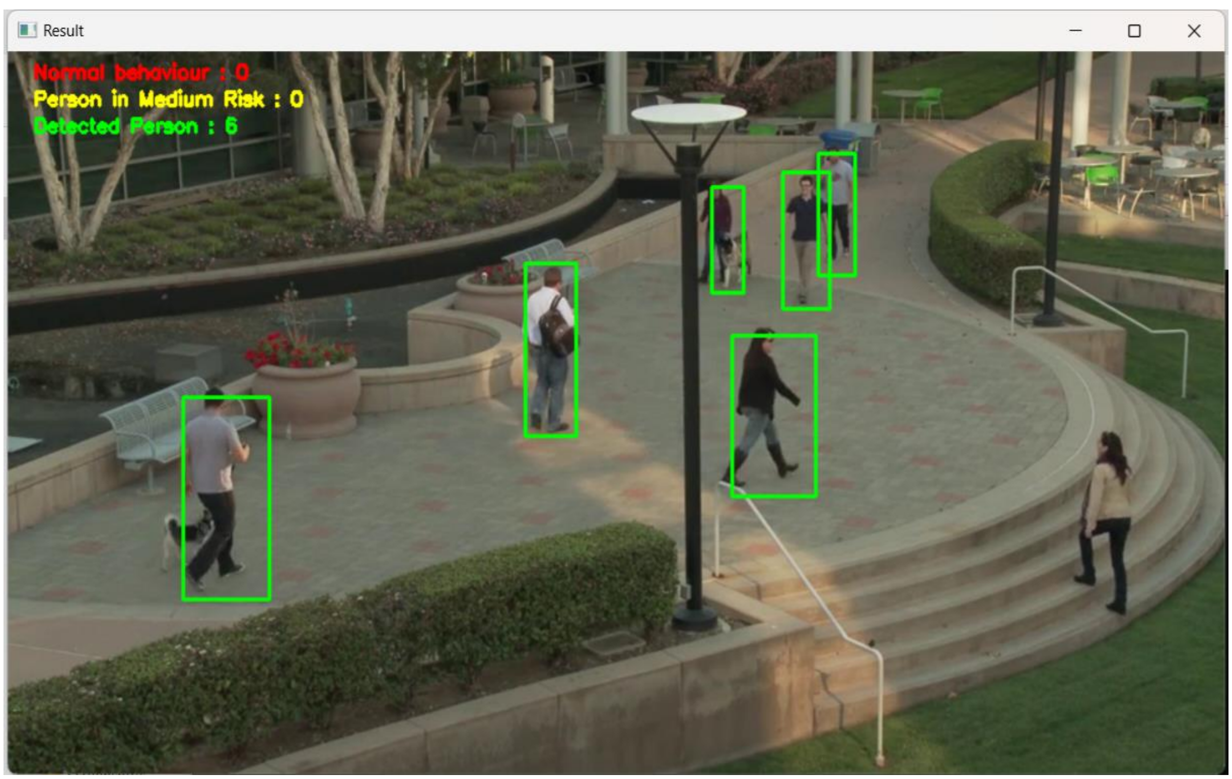
Table 1 Comparison Chart

Category	Accuracy	Observations
Detection-based	Low	Low accuracy for highly crowded scenes
Regression-based	Medium	Less interpretable, lacks location information
Density Map Estimation	High	Low accuracy in low crowd scenes
Deep Learning Algorithms	High	high accuracy in low and high crowd scenes


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 5


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 5


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 5


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 6


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 11
30% off
sweaters
GAP


Result
Normal behaviour : 0
Person in Medium Risk : 0
Detected Person : 14
30% off
sweaters
GAP


Result
Normal behaviour : 4
Person in Medium Risk : 0
Detected Person : 14


Result
Normal behaviour : 2
Person in Medium Risk : 2
Detected Person : 17


Result
Normal behaviour : 2
Person in Medium Risk : 0
Detected Person : 15


Result
Normal behaviour : 4
Person in Medium Risk : 4
Detected Person : 17

- In this project we are able to detect how may humans are in the video and can able to check if there is any abnormal activity that they are any risk detected. These are the some of Screenshots of the outputs in the green it detects that the how many persons are they in the video and red detects that the persons are near by to each other and there behavior whether is it suspisious or not if there is any suspisious then it change into yellow that there is some medium risk in that crowd.

CHAPTER 6

CONCLUSION

6.1 CONCLUSION:-

It examines deep getting to know-primarily based methods for studying interventions. The strategies are subdivided into crowd counting and crowd interest reputation. Crowd counting techniques aim to estimate the wide variety of people in a bodily place. Crowd hobby popularity methods determine the pastime of a group of humans or specific suspicious activity. For th sake of completeness, this review considers conventional computer visoin methods for crowd scene evaluation. Clearly, superior technological know-how-primarily based techniques are superior to standard laptop imaginative and prescient strategies within the frequency evaluation stage. Based on this overview, GAN frameworks and context-sensitive methods are promising instrutions in crowdsourced scene evaluation.

6.2 FUTURE WORK:-

Now the project works on only realtime videos. In further, the project will be develop on live video with more acuracy that this can be able to detect human behavior and crowd counting that can used which it can be used for public saftey purposes.

REFERENCES

1. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In international Conference on computer vision & Pattern Recognition (CVPR'05), vol. 1, pp. 886-893. IEEE Computer Society, 2005.
2. D. Roy, T. Ishizaka, C. K. Mohan, and A. Fukuda, “Vehicle trajectory prediction at intersections using interaction based generative adversarial networks,” in Proc. ITSC, 2021, pp. 2318–2323.
3. Gavrila, Dariu M., Jan Giebel, and Stefan Munder. "Vision-based pedestrian detection: The protector system." In Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy. 2004.
4. Jain et al., “Discrete residual flow for probabilistic pedestrian behavior prediction,” in Proc. CoRL, 2020.
5. Mahadevan, Vijay, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. "Anomaly detection in crowded scenes." In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1975-1981. IEEE, 2010.
6. Mehran, R.; Moore, B.E.; Shah, M. A streak line representation of flow in crowded scenes. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2010; pp. 439–452.
7. N. Deo and M. M. Trivedi, “Scene induced multi-modal trajectory forecasting via planning,” 2021, arXiv:1905.09949.
8. Q. Wang, J. Gao, W. Lin, and Y. Yuan, “Learning from synthetic data for crowd counting in the wild,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2021, pp. 8198–8207.
9. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, “Sophie: An attentive gan for predicting paths compliant to social and physical constraints,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1349–1358.
10. S. Marcel, H. Stefan, and D. Klaus, “Long-term occupancy grid prediction using recurrent neural networks,” in Proc. Int. Conf. Robot. Automat, 2020, pp. 9299–9305.
11. Tuzel, Oncel, Fatih Porikli, and Peter Meer. "Pedestrian detection via

classification on riemannian manifolds." IEEE transactions on pattern analysis and machine intelligence 30, no. 10 (2008): 1713-1727.

12. Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.
13. Wang, Lu, and Nelson HC Yung. "Crowd counting and segmentation in visual surveillance." In 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 2573-2576. IEEE, 2009.
14. W. Liu, K. M. Lis, M. Salzmann, and P. Fua, “Geometric and physical constraints for drone-based head plane crowd density estimation,” in Proc. Int. Conf. Intell. Robots Syst., 2020.
15. Wu, Bo, and Ramakant Nevatia. "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors." In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, pp. 90-97. IEEE, 2005.

APPENDIX

A. SOURCE CODE

```
### REFERENCES
### https://www.pyimagesearch.com/2017/09/11/object-detection-with-deep-
learning-and-opencv/
### https://docs.opencv.org/4.3.0/d6/d0f/group__dnn.html
### https://www.ebenezertechs.com/mobilenet-ssd-using-opencv-3-4-1-deep-
learning-module-python/
### https://www.pyimagesearch.com/2020/06/01/opencv-social-distancing-detector/

import cv2
import numpy as np
from math import pow, sqrt

#Constant Values
preprocessing = False
calculateConstant_x = 300
calculateConstant_y = 615
personLabelID = 15.00
debug = True
accuracyThreshold = 0.4
RED = (0,0,255)
YELLOW = (0,255,255)
GREEN = (0,255,0)
write_video = False

# I used CLAHE preprocessing algorithm for detect humans better.
# HSV (Hue, Saturation, and Value channel). CLAHE uses value channel.
# Value channel refers to the lightness or darkness of a colour. An image without
hue or saturation is a grayscale image.
```

```
def CLAHE(bgr_image: np.array) -> np.array:
    hsv = cv2.cvtColor(bgr_image, cv2.COLOR_BGR2HSV)
    hsv_planes = cv2.split(hsv)
    clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8, 8))
    hsv_planes[2] = clahe.apply(hsv_planes[2])
    hsv = cv2.merge(hsv_planes)
    return cv2.cvtColor(hsv, cv2.COLOR_HSV2BGR)

def centroid(startX,endX,startY,endY):
    centroid_x = round((startX+endX)/2,4)
    centroid_y = round((startY+endY)/2,4)
    bboxHeight = round(endY-startY,4)
    return centroid_x,centroid_y,bboxHeight

def calcDistance(bboxHeight):
    distance = (calculateConstant_x * calculateConstant_y) / bboxHeight
    return distance

def drawResult(frame,position):
    for i in position.keys():
        if i in highRisk:
            rectangleColor = RED
        elif i in mediumRisk:
            rectangleColor = YELLOW
        else:
            rectangleColor = GREEN
        (startX, startY, endX, endY) = detectionCoordinates[i]

        cv2.rectangle(frame, (startX, startY), (endX, endY), rectangleColor, 2)
```

```
if __name__== "__main__":

    caffeNetwork = cv2.dnn.readNetFromCaffe("./SSD_MobileNet_prototxt.txt",
"./SSD_MobileNet.caffemodel")
    cap = cv2.VideoCapture("./video.mp4")
    fourcc = cv2.VideoWriter_fourcc(*"XVID")
    output_movie = cv2.VideoWriter("./result.avi", fourcc, 24,
(int(cap.get(cv2.CAP_PROP_FRAME_WIDTH)),
int(cap.get(cv2.CAP_PROP_FRAME_HEIGHT))))


    while cap.isOpened():

        debug_frame, frame = cap.read()
        highRisk = set()
        mediumRisk = set()
        position = dict()
        detectionCoordinates = dict()

        if not debug_frame:
            print("Video cannot opened or finished!")
            break

        if preprocessing:
            frame = CLAHE(frame)

        (imageHeight, imageWidth) = frame.shape[:2]
        pDetection = cv2.dnn.blobFromImage(cv2.resize(frame, (imageWidth,
imageHeight)), 0.007843, (imageWidth, imageHeight), 127.5)

        caffeNetwork.setInput(pDetection)
```

```
    detections = caffeNetwork.forward()

    for i in range(detections.shape[2]):

        accuracy = detections[0, 0, i, 2]
        if accuracy > accuracyThreshold:
            # Detection class and detection box coordinates.
            idOfClasses = int(detections[0, 0, i, 1])
            box = detections[0, 0, i, 3:7] * np.array([imageWidth, imageHeight,
imageWidth, imageHeight])
            (startX, startY, endX, endY) = box.astype('int')

            if idOfClasses == personLabelID:
                # Default drawing bounding box.
                bboxDefaultColor = (255,255,255)
                cv2.rectangle(frame, (startX, startY), (endX, endY), bboxDefaultColor,
2)
                detectionCoordinates[i] = (startX, startY, endX, endY)

                # Centroid of bounding boxes
                centroid_x, centroid_y, bboxHeight = centroid(startX,endX,startY,endY)
                distance = calcDistance(bboxHeight)
                # Centroid in centimeter distance
                centroid_x_centimeters = (centroid_x * distance) / calculateConstant_y
                centroid_y_centimeters = (centroid_y * distance) / calculateConstant_y
                position[i] = (centroid_x_centimeters, centroid_y_centimeters, distance)

    #Risk Counter Using Distance of Positions
    for i in position.keys():
        for j in position.keys():
            if i < j:
```

```
                distanceOfBboxes = sqrt(pow(position[i][0]-position[j][0],2)
                                  + pow(position[i][1]-position[j][1],2)
                                  + pow(position[i][2]-position[j][2],2)
                                  )
                if distanceOfBboxes < 150: # 150cm or lower
                    highRisk.add(i),highRisk.add(j)
                elif distanceOfBboxes < 200 > 150: # between 150 and 200
                    mediumRisk.add(i),mediumRisk.add(j)


        cv2.putText(frame, "Normal behaviour : " + str(len(highRisk)) , (20, 20),
cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 0, 255), 2)
        cv2.putText(frame, "Person in Medium Risk : " + str(len(mediumRisk)) , (20,
40), cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 255, 255), 2)
        cv2.putText(frame, "Detected Person : " + str(len(detectionCoordinates)), (20,
60), cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 255, 0), 2)

        drawResult(frame, position)
        if write_video:
            output_movie.write(frame)
        cv2.imshow('Result', frame)
        waitkey = cv2.waitKey(1)
        if waitkey == ord("q"):
            break

    cap.release()
    cv2.destroyAllWindows()
```

B. RESEARCH PAPER

Video based Crowd Analysis Using Deep Learning

Gorantla Guruvardhan kumar[1]
Department of CSE
Sathyabama Institute of Science and Technology , Chennai

Gopu Shaji[2]
Department of CSE
Sathyabama Institute of Science and Technology, Chennai

L.Sujihelen[3]
Department of CSE
Sathyabama Institute of Science and Technology, Chennai

gopushajiblr@gmail.com ***vardhangorantla143@gmail.com*** ***sujihelen@gmail.com***

Abstract

A system that analyzes crowd behavior in real time and alerts you if abnormalities occur is known as crowd behavior analysis. Pedestrian streets, supermarkets, bus stand and markets are usually crowded. It is always important to ensure the safety and comfort of large crowds, so appropriate measures must be taken to ensure their safety. This article introduces Convolutional Neural Networks and the deep learning process used to study common scenes. In these documents it's proposed to find statistics on the number of people who arrived in each group, as well as a table of the density of the group. Counting human beings in a dense crowd is an essential step in surveillance and anomaly warning. Existing population estimation methods are based on homemade capabilities such as SIFTS and HOG. Currently, the most suitable technique is for expecting the high overall performance of density and crowd estimation based on a deep understanding of networks. The purpose is to offer a complete evaluation on various kinds of crowd analysis, including crowd counting, crowd detection, anomaly detection, and human behavior. The proposed work uses a deep learning algorithm to analysis the crowd. It performs better accuracy when compared with the other algorithms.

Keywords: Crowd analysis, pre-processing, object tracking, event behavior recognition.

I Introduction

Automatic evaluation of crowd scenes is used to take a look at the behavior of big groups of humans residing in the identical bodily area[1]. Typically, it counts the number of humans in a country, tracks the overall trajectories of people, and recognizes human being's conduct. Thus, the automatic evaluation of the hundreds has many and vital applications. The unfold of the COVID-19 virus is managed by way of ensuring physical distance between human beings in shops, parks, and so on. To provide for the protection of mass activities inclusive of ball video games, carnivals , annual celebrations, and Muslim journey is every other utility of spontaneous crowd evaluation[2]. Mass scene evaluation presents surveillance camera structures with the potential to hit upon ordinary conduct of big businesses of people. In addition, the evaluation of crowd scenes in public places, together with teach stations, supermarkets, and shopping department shops, can display the impact of crowd motion or layout flaws. Therefore, those research are gaining fitness considerations[3][4]. Due to the importance of crowd evaluation, as proven above, diverse evaluate papers have been proposed. However, the present papers reviewed both used conventional laptop techniques for frequency analysis, or taken into consideration only one component of crowd evaluation, together with crowd counts[5]. Therefore, this evaluation article objectives to provide a perspective at the improvement of a crowd evaluation of complete level strategies for the most current deep learning techniques[6]. This angle specializes in two main components of crowd analysis: (1) crowd counting and (2) awareness of crowd action. In addition, this paper proposes an statistics theory-stimulated scoring matrix, which proposes common variance (CD) methods for crowd scene analysis. Compared to famous estimation matrices, consisting of the mean squared errors (MSE and the imply absolute mistakes (MAE)), CD measures how carefully the anticipated crowd length distribution is to the real distribution. The proposed metric calculates the amount of discrepancy between actual and predicted numbers: The contribution of this paper includes three components: • a top-level view of deep mastering-based totally methods for the analysis of frequency distribution, • an overview of the available frequency characteristics, and • a. Population variance thought (CD) for the unique evaluation of the techniques. The relaxation of this assessment is organized as follows: Section 2 discusses the approach of counting mass gatherings Section three on methods to become aware of the activities of mass gatherings Section four considers datasets of common scenes The assessment matrix for the frequenting approach proposed in Article five Section 6 incorporates the discussion of the document, Section 7 concludes our evaluation paper and similarly part pre it will be

II Literature survey

Current methods for counting people in dense community scenes depend upon estimating the population density on the image level. This photo, whilst density is without a doubt useful for this purpose, has no instantaneous physical meaning, as it's far concern to perspective distortion. This is a trouble with drone photos because the factor of view modifications frequently. This distortion is commonly removed implicitly, both by means of searching at regarded scale invariants or by way of estimating the density in unevenness of various sizes, however neither approach considers that the dimensions modifications should be constant all through the scene. In this article, we express the styles of modifications and ratios primarily based on the range of human beings per rectangular meter. We display that the transmission perspective version lets in for consistency at a global scale, and that this version can be completed with drone sensors at the fly[7]. In addition, it also lets in us to use bodily consistency to temporal constraints, which does no longer ought to be discovered. This effects in an set of rules this is advanced to cutting-edge methods for determining crowd density from a transferring drone digicam, specifically with strong angle consequences.

Self-using vehicles control each static and dynamic gadgets via making use of predictive behavior fashions to estimate the future location of gadgets within the environment[8]. But destiny conduct is inherently unsure, and the motion styles that produce deterministic consequences are restricted to short time scales. It is extraordinarily difficult to predict human's conduct. In this work, we propose a Discrete Residual Flow Network (DRF-Net), a convolutional neural network for human motion prediction that captures the inherent uncertainty of lengthy-time period motion prediction. In particular, our educated community successfully captures multimodal posterior facts approximately destiny human movement by way of predicting and updating the discrete distribution in local regions. We examine our model with several sturdy competition and show that our model outperforms all benchmarks[9].

Vehicle trajectory prediction at intersections is an vital and tough mission for autonomous automobile navigation. This trouble is exacerbated while traffic is ruled by means of small vehicles, which frequently do no longer obey the lane, as is the case in lots of growing nations. The current macro techniques the hassle of trajectory prediction for lane site visitors, which can't consider the massive discrepancy in vehicle sizes and conduct among one-of-a-kind sorts of motors. Hence, we advocate an technique to automobile trajectory prediction that fashions the interplay between unique types of motors with very unique using patterns. These interactions are embedded in a shape of social context embedded in a generative adversarial network (GAN) to predict the trajectory of any car at a signed or unsigned intersection[10]. The GAN version creates a desired future trajectory among many options that matches the behavior of past activities including the trajectories of close by motors. We evaluate the proposed method on aerial pictures of intersections taken in China, where cars do not follow site visitors rules. The proposed GAN-based technique indicates a 6.4% relative development in trajectory prediction over the nation of the art[11].

We are engaged in long-time period prediction of the improvement degree at the middle of a complex state undertaking for computerized using primarily based on a aggregate of lidar grids and recurrent neural networks (RNN). A chicken's eye view of the scene, such as occupancy and pace, is then handed to an RNN this is skilled to expect destiny occupancy[12]. The predictive nature permits numerous hours of studying data to be generated without a whole lot manual work. Thus, the studying strategy and the loss feature are designed for long collection of real information (unmatched, changing conditions, fake labels, etc.)[13]. The deep array structure carries Convolutional Long Short-Term Memory (Conv LSTM) to separate static from dynamic regions and are expecting dynamic items in future frames. The new links show the capability to expect the skip's ability to predict small closed objects including pedestrians and closed static regions. Spatiotemporal correlations among cells are used to predict future multimodal pathways and interactions between items. The experiments also make bigger improvements to our preceding networks, the Monte Carlo technique, and the literature.

This article offers with the problem of path prediction for a couple of interacting sellers in a scene, which is an crucial step for many self sufficient systems, such as self-using motors and social robots. Introducing an interpretable framework based on a generative opposed network (GAN) that uses resources of facts, the history of all agents in the scene, and context statistics using scene photographs. To are expecting an agent's future course, both physical and social facts ought to be used. Previous paintings on co-simulation of bodily and social interactions has now not been successful. Our technique combines the mechanism of social interest with physical interest, a version of which allows to recognize in which to look in a large scene and to focus on the main elements of the road-related picture. As a part of social interest, it gathers statistics approximately the agent's diverse interactions and extracts key trajectory functions from surrounding friends. So additionally enables GANs to create melioristic models and mirror the uncertainty of destiny paths via their model distribution. All those mechanisms allow our technique to are expecting street managers socially and bodily and acquire state-of-the-art consequences in several one-of-a-kind course prediction benchmarks[14].

We convert the multimodal prediction of agent trajectories into unknown scenes and country it as a scheduling hassle. We present an method such as three fashions; a signal prediction version to pick out capacity energetic objectives, a getting to know reduction model to assist in deploying the pleasant routes to every target, and producing trajectories to derive future trajectories along planned paths. Analysis of predictions on the Stanford paint dataset suggests the generality of our method to novel scenes[6].

Recently, people computing for crowd scenes is a hot topic for a wide open software (eg surveillance, public protection[15]). This work inside the wild is hard: the converting environment, the huge number of humans means that the current methods cannot work properly. In addition, because of the shortage of information, many methods be afflicted by issue to various stages. Addressing these troubles, we first developed a statistics collector and labeler which could generate synthetic crowd scenes and examine them simultaneously without intervention[10]. From it we build a large-scale, various artificial dataset. Second, we advise strategies that we use artificial records to enhance the populace this is counted within the wild: 1) pre-computing the population on artificial statistics, and then nice-tuning it the usage of real information, which substantially improves the model's performance on real records. . ; 2) To advocate a crowd-sourced counting approach tailored with the aid of area, that can supply from big facts annotations. Extensive trying out shows that the primary approach outperforms our baseline on four actual schedules, at the same time as the second one method outperforms our baseline. The drawbacks in existing system is:

- The mission of crowd calculation and density estimation refers to many issues, along with blockages, non-uniform density, intra- and inter-degree differences in length and perspective.
- However, during the last few years, crowd analysis has evolved from previous techniques that have been regularly limited to small changes in density in frequency.

III Proposed system

This article discusses techniques for accumulating the most often used scenes in deep studying. Considered procedures are divided into two businesses: crowd counting and crowd reputation activities .In addition, crowdsourced datasets are checked.In popular video, this metric measures the discrepancy among the estimate and the real crowd.

Advantages

These days, when the facility ought to observe COVID-19 potential limits and social distancing suggestions similarly to emergency and fireplace protection requirements, this is specially essential.

Crowd Counting

Crowd counting is a method of counting or estimating the number of humans in an picture. Accurately estimating the variety of human beings/objects in a single image is a tough but critical mission that is used in many packages together with city making plans and public protection. Population counting stands out a number of the diverse objects of counting tasks because of its special significance for security and social improvement. Before we move on, let's assessment the numerous businesses that the common photo analysis organization considers.

Kinds of crowd quantity

There are sorts of crowd counting

1. Traditional method
2. Normal arrival

Traditional approach

1. Object degree detection
2. Image / pixel detection
3. Mesh / leak detection

Block diagram

Fig.1 and Fig.2 shows the architecture diagram. The real time video is taken for crowd analysis. Each person is identified and list out the number of persons identified.

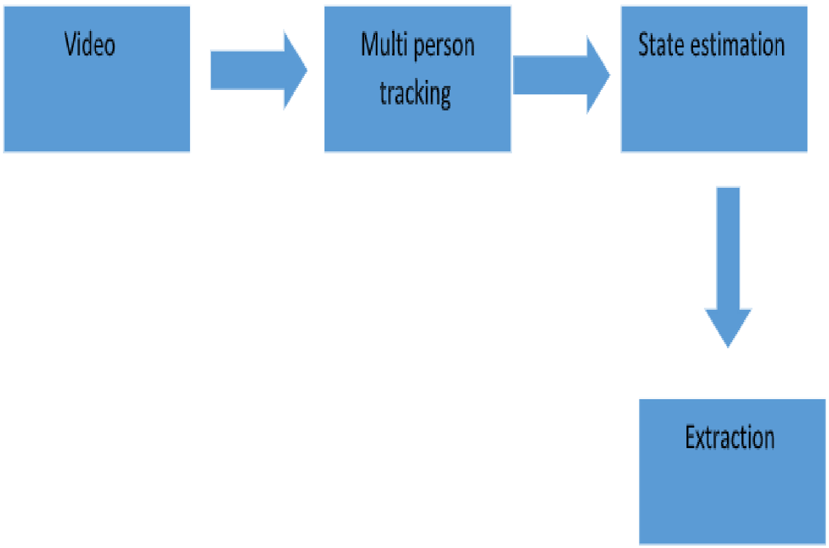
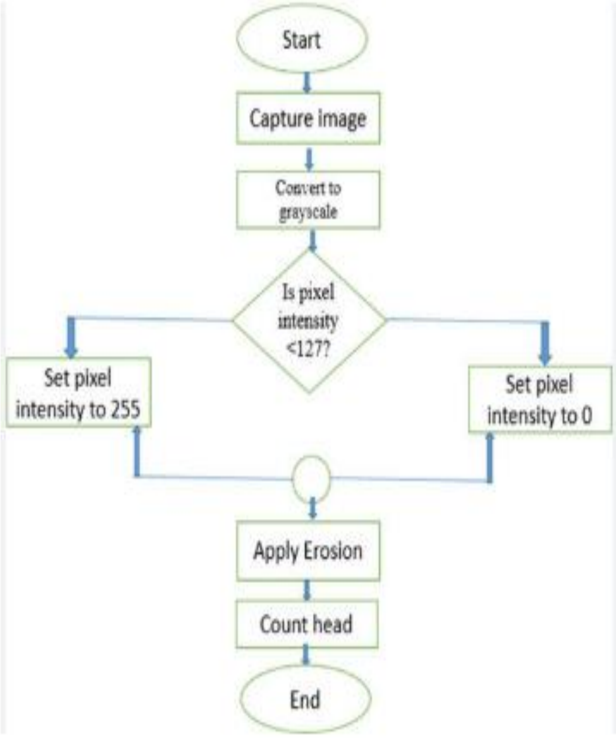


Fig . 1 Architecture diagram



Each image is captured and converts to grayscale. If the pixel intensity is less than 127 then set pixel intensity to 255 otherwise set pixel intensity to 0. After applying erosion and count head.

Fig.2 Flow diagram of the proposed work

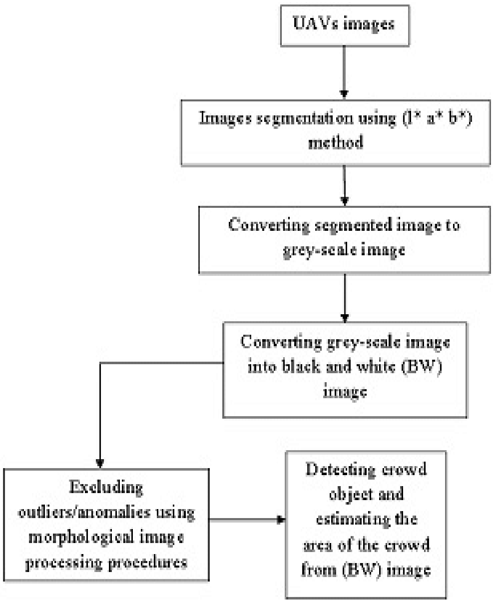


Fig.3 Algorithm used in this one

CNN

A cluster is a sort of network architecture for deep studying algorithms that are used for image recognition and other duties that require the processing of pixel facts. There are other sorts of neural networks in deep getting to know, but neural networks are the preferred structure for gaining knowledge of and reputation shown in Fig.3.

Modules used

1. Collect the data set
2. Multi person tracking
3. State estimation
4. Feature extraction

1. Collect the data set

Information and data using real time video. It is given to the concern and the human beings.

2. Multi person tracking

Tracking loads of things requires finding, predicting and correlating records. Find thrilling things on the map the use of video detection. Prediction: Predict wherein things may be in the next table. Data association: create traces connecting detections among tables the use of expected positions!

3. State estimation

The reason of the general public assessment is to alter the model consequences to be toward the located values, for this reason providing a higher reanalysis of the ocean.

4. Feature extraction

Feature extraction enables to reduce the quantity of redundant information in a dataset. After all, facts discount helps to construct a model with less system attempt and additionally will increase the speed of reputation and generalization ranges in the device getting to know method.

IV Results and Discussion

Fig.4 shows the peoples count and how the object is tracked. Table 1 shows the proposed work is compared with the existing work. The proposed work has high accuracy and detect the people accurately.

Fig.4 Results

Table 1 Comparison Chart

Category	Accuracy	Observations
Detection-based	Low	Low accuracy for highly crowded scenes
Regression-based	Medium	Less interpretable, lacks location information
Density Map Estimation	High	Low accuracy in low crowd scenes
Deep Learning Algorithms	High	high accuracy in low and high crowd scenes

V Conclusion

This article examines deep getting to know-primarily based methods for studying interventions. Survey strategies are subdivided into crowd counting and crowd interest reputation. Crowd counting techniques aim to estimate the wide variety of people in a bodily place. Crowd hobby popularity methods determine the pastime of a group of humans or specific suspicious activity. For the sake of completeness, this review considers conventional computer vision methods for crowd scene evaluation. Clearly, superior technological know-how-primarily based techniques are superior to standard laptop imaginative and prescient strategies within the frequency evaluation stage. In addition, a brand new overall performance metric, CD, i.E., is proposed to offer an accurate and dependable evaluation degree for crowd assessment. This is finished by way of measuring the difference between the real trajectory/amount and the expected trajectory amount. Based on this overview, GAN frameworks and context-sensitive methods are promising instructions in crowdsourced scene evaluation.

References

1.W. Liu, K. M. Lis, M. Salzmann, and P. Fua, “Geometric and physical constraints for drone-based head plane crowd density estimation,” in Proc. Int. Conf. Intell. Robots Syst., 2020.

2.Jain et al., “Discrete residual flow for probabilistic pedestrian behavior prediction,” in Proc. CoRL, 2020.

3.D. Roy, T. Ishizaka, C. K. Mohan, and A. Fukuda, “Vehicle trajectory prediction at intersections using interaction based generative adversarial networks,” in Proc. ITSC, 2021, pp. 2318–2323.

4.S. Marcel, H. Stefan, and D. Klaus, “Long-term occupancy grid prediction using recurrent neural networks,” in Proc. Int. Conf. Robot. Automat, 2020, pp. 9299–9305.

5.Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, “Sophie: An attentive gan

for predicting paths compliant to social and physical constraints,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1349–1358.
6.N. Deo and M. M. Trivedi, “Scene induced multi-modal trajectory forecasting via planning,” 2021, arXiv:1905.09949.
7.Q. Wang, J. Gao, W. Lin, and Y. Yuan, “Learning from synthetic data for crowd counting in the wild,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2021, pp. 8198–8207.
8. Mehran, R.; Moore, B.E.; Shah, M. A streak line representation of flow in crowded scenes. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2010; pp. 439–452.
9.Wang, Lu, and Nelson HC Yung. "Crowd counting and segmentation in visual surveillance." In 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 2573-2576. IEEE, 2009.
10.Gavrila, Dariu M., Jan Giebel, and Stefan Munder. "Vision-based pedestrian detection: The protector system." In Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy. 2004.
11.Mahadevan, Vijay, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. "Anomaly detection in crowded scenes." In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1975-1981. IEEE, 2010.
12.Tuzel, Oncel, Fatih Porikli, and Peter Meer. "Pedestrian detection via classification on riemannian manifolds." IEEE transactions on pattern analysis and machine intelligence 30, no. 10 (2008): 1713-1727. 13.Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.
14.Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In international Conference on computer vision & Pattern Recognition (CVPR'05), vol. 1, pp. 886-893. IEEE Computer Society, 2005.
15. Wu, Bo, and Ramakant Nevatia. "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors." In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, pp. 90-97. IEEE, 2005.