ANOMALY DETECTION IN ROAD TRAFFIC USING VISUAL SURVEILLANCE

Report of IT499 Major Project-II
Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY
in
INFORMATION TECHNOLOGY
by

Ayush Bhandari 181IT209
Jaidev Chittoria 181IT119

under the guidance of

Dr. Sowmya Kamath S

DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE - 575025

April, 2022

DECLARATION

We hereby *declare* that the Project Work Report entitled "ANOMALY DETECTION IN ROAD TRAFFIC USING VISUAL SURVEILLANCE", which is being submitted to the **National Institute of Technology Karnataka, Surathkal**, in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Information Technology in the Department of Information Technology, is a *bonafide report of the work carried out by us.* The material contained in this Report has not been submitted at any University or Institution for the award of any degree.

Registration Number, Name & Signature of the Student(s)

(1) 181IT209 - Ayush Bhandari - Ayush

(2) 181IT119 - Jaidev Chittoria -

Department of Information Technology
Place: NITK, Surathkal
Date: 4/04/2022

CERTIFICATE

This is to *certify* that the Project Work Report entitled ”ANOMALY DETECTION IN ROAD TRAFFIC USING VISUAL SURVEILLANCE”, submitted by

Sl. No., Registration Number & Name of the Student(s)

(1) 181IT209 - Ayush Bhandari

(2) 181IT119 - Jaidev Chittoria

as the record of the work carried out by them, is *accepted as the B.Tech. Project Work report submission* in partial fulfillment of the requirement for the award of degree of **Bachelor of Technology** in Information Technology in the Department of Information Technology.

Dr. Sowmya
Kamath S

Digitally signed by Dr. Sowmya Kamath S
DN: cn=Dr. Sowmya Kamath S,
o=National Instiue of Technology
Karnataka, ou=Dept. of Information
Technology,
email=sowmyakamath@nitk.edu.in, c=IN
Date: 2022.04.04 17:32:11 +05'30'

Guide
Dr. Sowmya Kamath S

Dr. Jaidhar C. D
Chairman - DUGC

ABSTRACT

Surveillance videos are able to capture a variety of realistic anomalies that happen over the course of day . In this paper, we propose to detect anomalies caused due to human activities (i.e theft,crime,burglary ,etc) and vehicle related anomalies on road . For human activities we would learn anomalies by utilizing both normal and anomalous videos.By using multiple instance learning and instinctively providing higher score to anomalous video segments and adding smoothness in ranking for loss function to find anomaly more easily while training. For vehicle related anomalies, videos are converted into stream of images via superimposing image in such a manner that only slow moving objects are visible and after that they are treated as object detection problem with few constraint and anomalies are detected with the help of metrics model performance is judged.

.

Keywords— anomaly, multiple instance learning, superimposition.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 OVERVIEW

Rapid development in urban areas has also led to increasing use of CCTV cameras in public places i.e streets,banks,malls, etc. to increase public safety. However there is no such monitoring capability system to detect anomalous events such as traffic accidents, crimes or illegal activities by interpreting the video on its own in a feasible manner. Therefore to avoid the waste of labor and time to monitor such events ,so developing computer vision algorithms for automatic detecting anomalies based on video seems much relevant . The aim is to timely signal an activity that deviates from normal patterns and identify the time window of the occurring anomaly . Once an anomaly is detected, it can further be categorized into one of the specific activities using classification techniques. A small step towards addressing anomaly detection is to develop algorithms to detect an anomalous event i.e violence detector and traffic accident detector. [1] However, it is obvious that such solutions cannot be generalized to detect other anomalous events, therefore they render a limited use in practice. Sparse-coding based approaches are considered as representative methods that achieve state-of-the-art anomaly detection results.[2] These methods assume that only a small initial portion of a video contains normal events, and therefore the initial portion is used to build the normal event dictionary. Then, the main idea for anomaly detection is that anomalous events are not accurately reconstructable from the normal event.[3][4][5]

1.2 MOTIVATION

The increase in crime rate in public areas and availability of no proper structured system to detect such anomalous events in a efficient and feasible manner can be quite useful. Also lot of work previously done have different limitations ranging for availability of large as well as good quality video dataset to not having well defined of normal behavior and no specific well defined criteria for classification and neither dealing with the real time video detection.

Chapter 2

LITERATURE REVIEW

2.1 BACKGROUND AND RELATED WORKS

Anomaly detection is a challenging problem in a real-world scenario. The complications arise due to intricate concepts of anomaly. The world is growing, and there is an increasing demand for better security systems. Many types of research have been conducted to address the problem of anomaly detection using the development of deep learning models[6][7]. One of the domains of anomaly events in real-world scenarios, i.e., traffic anomaly, is also becoming very popular among many research groups[4][5]. Traffic anomalies include events in traffic where the vehicle's movement deviates from the regular patterns. One common strategy is modeling the expected trend in the data. Data modeling distribution will allow the anomalous events to appear as outliers, and then they can be detected by any outlier detection model.

Waqas Sultani et al. [3] proposed to learn anomalies by using both standard and anomalous videos. They propose learning anomalous events using the deep multiple instance ranking framework with weakly labelled training films, i.e., the training labels (anomalous or standard) are at the video level rather than clip level, because annotating anomalous segments in training videos is a tiresome operation. In multiple instance learning (MIL), they use their method to automatically develop a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments by treating normal and anomalous films as bags and segments as instances.

Louis Kratz et al.[1] proposed to exploit the dense activity of the crowded scene by modeling the rich motion patterns in local areas, effectively capturing the underlying intrinsic structure they form in the video. Simply expressed, they characterise the overall behaviour of the scene by modelling the motion variation of local space-time volumes and their spatial-temporal statistical characteristics. They also show that anomalous activity can be discovered naturally as statistical deviations by capturing steady-state motion behaviour with Spatio-temporal motion pattern models. In real-world scenarios with complicated behaviours that are difficult for even human observers to examine, their investigations reveal that local Spatio-temporal motion pattern modelling delivers promising outcomes.

Bin Zhao et al. [4] proposed a fully unsupervised dynamic sparse coding approach for detecting unusual events in videos based on sparse online re-constructibility of query signals from anatomically learned event dictionary, which forms sparse coding bases. Based on intuition, those usual events in a video are more likely to be reconstructible from an event dictionary, whereas unusual events are not. The proposed technique uses a principled convex optimization framework to jointly infer and update a sparse reconstruction code and an online dictionary. The algorithm is completely unsupervised, which means it makes no assumptions about what unexpected events might look like or how the cameras are set up. The fact that the basic dictionary is updated online as the algorithm collects more data eliminates any concept drift concerns. The suggested algorithm could reliably locate the unexpected moments in the video sequence, according to experimental results on hours of real-world surveillance video and various Youtube videos.

Huiwen Guo et al.[6] present a method to detect and localize abnormal events in crowded scenes. To describe crowd motion, most known approaches use a patch of optical flow or a human tracking-based trajectory, both of which are prone to noise. Instead, they propose using a new and efficient characteristic called short-term trajectory, which reflects the motion of the visible and constant component of the human body that moves regularly in order to describe the complex, congested image. A 3D mean-shift is utilised to smooth the video frames before a 3D seed filling technique is employed to extract the short-term trajectory. All short-term trajectories are processed as point sets and mapped into the image plane to obtain a probability distribution of normalcy for each pixel in order to detect aberrant events. To detect and localise the anomalous occurrence, cumulative energy is estimated using these probability distributions. Experiments on known crowd data sets reveal that the suggested method achieves high accuracy in anomaly detection and efficacy in anomaly localization.

This paper presented an overview of the approaches used by different teams in the AI city challenge. Milind Khapade et al.[8] methodologies were based on the basic idea of preprocessing involved background modeling, vehicle detection, road mask construction to remove stationary parked vehicles, and abnormal vehicle tracking. The winning team's dynamic tracking module, used spatio-temporal status and motion patterns to pinpoint the exact start time of the abnormalities. Further post-processing was done to fine-tune the commencement time of the traffic abnormalities. Their highest score was 0.9355, indicating that the traffic anomaly problem can be solved with present technologies. The runner-up, tracked possible Spatio-temporal anomalous tubes at the box level. Using tubes derived from background modelling and adjustments, their system can accurately detect unusual time periods. Similarly, the third-place team, used box-level and pixel-level tracking to discover abnormalities, as well as a dual-modality bi-directional tracing module to refine the time periods.

T. Kar et al.[9] proposed a new automatic cut detection method is proposed based on the local binary pattern feature. Six test movies are used to evaluate the proposed technique, and its efficacy is confirmed using a few existing popular approaches. Three alternative cut detection methods, ASHD, ECR, and PID approaches, are researched and their performances are evaluated in this paper. It has been discovered that cut detection based on ASHD beats ECR and PID. ASHD's performance can be improved by lowering the number of erroneous cuts and missed cuts. They offer a new cut detection algorithm based on a texture feature taken from a local binary pattern. In all of the test videos, the proposed technique outperforms the ECR and PID, while in most of the videos, the suggested method outperforms the ASHD, with the exception of synthetic videos and a song video from the film "Masoom."

Ankit Kariryaa et al.[10] proposed a masking method for training deep learning models from a publicly available but insufficient dataset. For example, Hamburg, Germany, has a tree index along the roadways, but this dataset is missing information on trees in residences and parks. The authors disguised the road network's street trees and aerial photos in order to build a deep learning model on such a dataset. The OpenStreetMap road network that was used to create the mask was downloaded, and it marked the area where the training data was available. The mask is one of the model's inputs, and it coatings the output. Their model learns to predict trees with 78.4 percent accuracy exclusively in the masked zone.

Kelathodi Kumaran Santhosh et al.[11] present a survey on relevant visual surveillance related researches for anomaly detection in public places, focusing primarily on roads. To begin, they review surveys conducted in this subject during the previous ten years. The authors place a greater emphasis on learning approaches applied to video sequences because learning is the basic building element of a normal anomaly detection. They also highlight the major advancements made in the field of anomaly detection over the last six years, with a particular focus on features, underlying methodologies, applicable scenarios, and types of anomalies utilising a single static camera. Finally, the authors highlight the limitations of anomaly detection systems based on computer vision, as well as some interesting future possibilities.

This paper presented an overview of the approaches used by different teams in the 2019 AI city challenge. Milind Naphade et al.[12] used foreground segmentation to reduce the search. The best performance was achieved by Team BUPT Traffic Brain using a spatio-temporal anomaly matrix. The runner-up, on the other hand, proposed a novel two-stage framework based on anomaly candidate identification and starting time estimation. The solution of the third place team was based on the second-place winning method from the 2018 AI City Challenge, with refined event recognition of stalled vehicles and back-tracking to accurately locate event occurrence.

Salisu Wada Yahaya et al.[13] proposed an approach for identifying the sources of abnormalities in human activities of daily living is proposed. Anomalies are found by using existing activity data to create a baseline model that represents an individual's normal behavioural routine. Outliers or anomalies are therefore defined as activities that deviate from the baseline. For anomaly detection, an ensemble of one-class support vector machines, isolation forests, robust covariance estimators, and local outlier factors is used, with an accuracy of 98 percent. The proposed method for finding anomaly origins is based on the concept of employing distance functions to measure similarity. One vs one similarity measure and one vs all similarity measure are two methods for measuring the pairwise distance of the features of the activity data. The proposed approach's credibility for use in an in-home monitoring system has been demonstrated through experimental evaluation of the proposed approach on activities of daily life datasets.

Shyma Zaidi et al.[14] present an automated video surveillance system for security system. In this case, we employ a single camera to detect and track human motion across time. The major purpose is to detect the item efficiently utilising background subtraction techniques, with the goal of lowering security system costs and increasing efficiency. The three video detection search directions of detection, tracking, and human motion analysis are thought to be the most relevant areas for future research. There are numerous ways for developing intelligent vision systems that strive to comprehend scenes and correct semantic interference from observable dynamics of moving targets. A database of stances from various perspectives should be created. Many assumptions of earlier methods are avoided with the multi view action recognition from a new perspective. In this method we compare the actions and recognize the changes underwent in the video frames.

Yashswi Jain et al.[15] propose PoseCVAE, employing the concept of generative modelling to develop a novel abnormal human activity detection strategy. They use a hybrid training technique that combines self-directed and undirected learning. Their goal was to develop a conditional posterior distribution that reflects normal training data and train our framework to predict future pose trajectory given a normal track of past poses. To do this, the authors developed PoseCVAE, an unique modification of a conditional variational autoencoder (CVAE). If the given poses are sampled from a distribution that differs from the learned posterior, as is the case with aberrant behaviours, future pose prediction will be incorrect. To further separate the abnormal class, they imitate abnormal poses in the encoded space by sampling from a distinct mixture of Gaussians (MoG). They use a binary cross-entropy (BCE) loss as a novel addition to the standard CVAE loss function to achieve this. Finally they test their framework on three publicly available datasets and achieve comparable performance to existing unsupervised methods that exploit pose information.

2.2 OUTCOME OF LITERATURE REVIEW

It is evident that anomaly detection in real-world scenarios is an important and cumbersome problem. There has been much research that has been done to address the problem of anomaly detection. One research presents learning anomalous events using the deep multiple instance ranking framework with weakly labeled training films. Some authors addressed the problem using foreground segmentation and two complementary predictors to cover all the anomalous events. One of the research investigations reveals that local Spatio-temporal motion pattern modeling delivers promising outcomes.

Different methods have been employed to deal with sparseness. Some authors have introduced short-term trajectory-based detection to solve the complex trajectory of anomalous events. In video-based anomaly detection, the dissimilarity between frames is recognized using a cut detection-based approach. Some research has also been done to identify anomalies in human-based events; authors have introduced approaches using changes in human pose behavior, differences in changed behavior, and trajectory of motion.

Lastly, in the AI city challenge, the problem of detecting anomalous events in traffic roads has been addressed. Participants have used many approaches like foreground segmentation, post-processing the frames to reduce false positives, a novel two-stage framework based on anomaly candidate identification, box level tracking feature, and Spatio-temporal tubes. All of these works have done a great job indeed to solve the problem of anomaly detection.

In this work, we intend to perform experiments to detect anomalies in traffic videos combining various approaches with road masking techniques to get better performance.

2.3 PROBLEM STATEMENT

The aim is to build an algorithm to detect anomalous events i.e that deviates from normal patterns and on anomaly detection then further categorizing into one of the specific activities using classification techniques.

2.4 RESEARCH OBJECTIVES

- To detect anomalous events that deviates from normal patterns.
- To compare different techniques used to detect human activities and vehicle activities in the video.
- To develop a model which can not only find the anomaly but can also predict

the time on anomaly event in efficient manner.

Chapter 3

PROPOSED METHODOLOGY

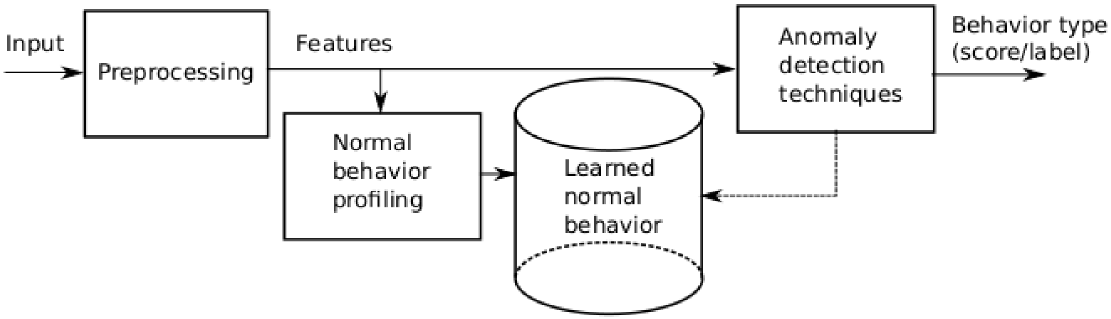

Input
Preprocessing
Features
Normal behavior profiling
Learned normal behavior
Anomaly detection techniques
Behavior type (score/label)

Figure 3.0.1: Detailed Workflow

As anomalous activities are the events/behavior which deviates from normal behavior with no proper well defined definition. So for human activities anomalies (i.e abuse, arrest, assault, burglary, fighting, robbery, shooting, shoplifting, stealing) we propose an anomaly detection algorithm in which weakly labeled training videos are used such that we know that either the video is normal or an anomaly is present in video but at does not at what time instant. This can be quite helpful as one can easily interpret large set of videos and assign labels to videos. Since it a weakly supervised learning ,one can use multiple instance learning[16] as it is quite useful approach in such situation. So in order to carry out such task deep multiple instance learning will be used such that normal and videos containing anomaly will be considered as set of bags and instances are defined as continuous short time intervals of each video . Based on training videos, an anomaly ranking model that predicts high anomaly scores for anomalous segments in a video which will help us in differentiating it from others and while testing, video is divided into segments and fed into deep network which assigns an anomaly score for each video segment such that an anomaly can be detected.

The flow diagram of the anomaly detection approach. Given the positive (containing anomaly somewhere) and negative(containing no anomaly) videos. We divide each of them into multiple temporal video segments. Then, each video is represented

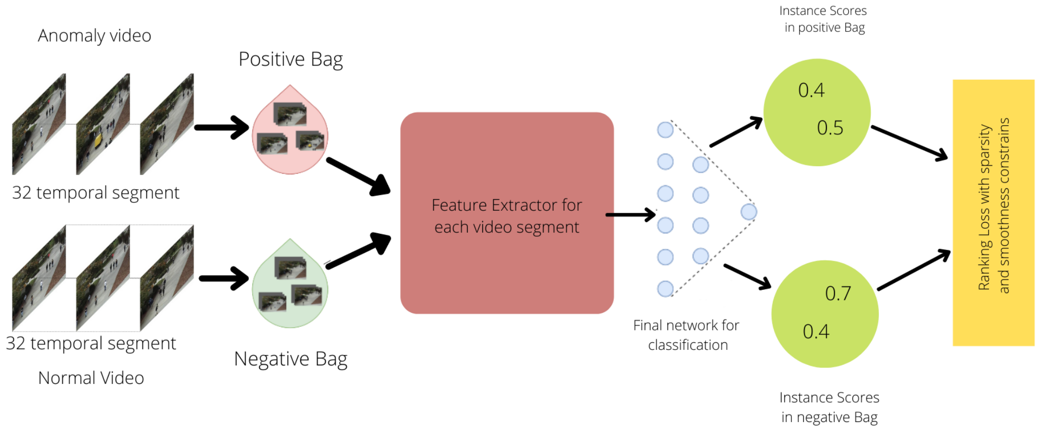

Anomaly video
32 temporal segment
Positive Bag
32 temporal segment
Normal Video
Negative Bag
Feature Extractor for each video segment
Final network for classification
Instance Scores in positive Bag
0.4
0.5
0.7
0.4
Instance Scores in negative Bag
Ranking Loss with sparsity and smoothness constrains

Figure 3.0.2: Classification of human action flow

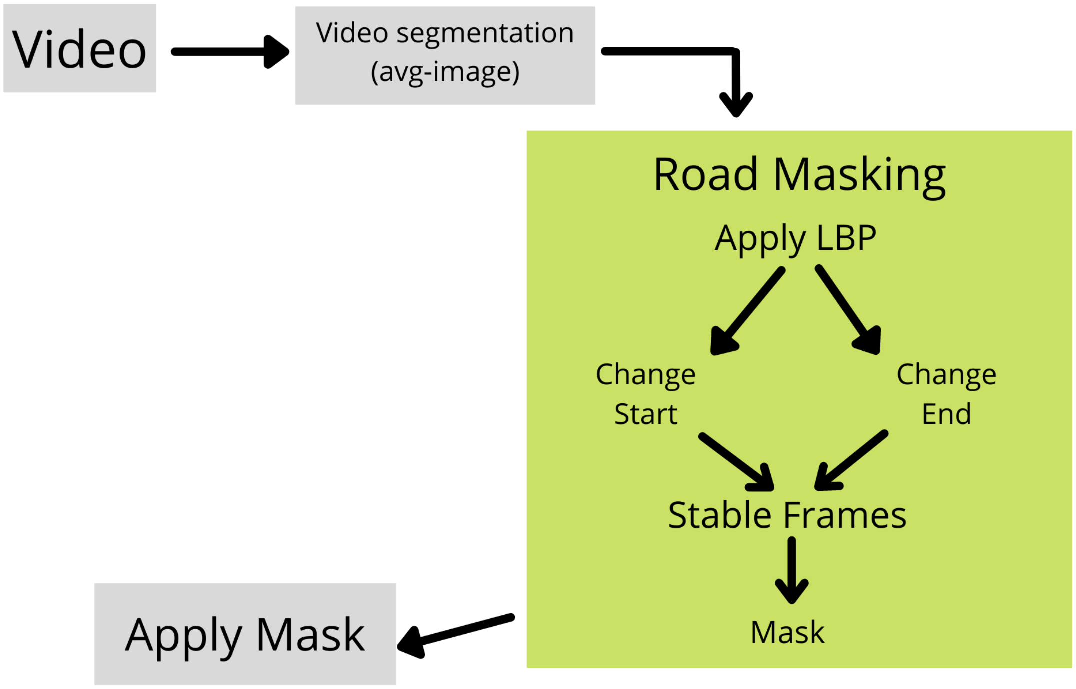


Figure 3.0.3: Masking

as a bag and each temporal segment represents an instance in the bag.

After extracting features for video segments, we train a fully connected neural network by utilizing a ranking loss function which computes the ranking loss between

6	5	2
5	6	3
9	7	8

1	0	0
0		0
1	1	1

1	2	4
128		8
64	32	16

Figure 3.0.4: Local binary pattern (LBP)

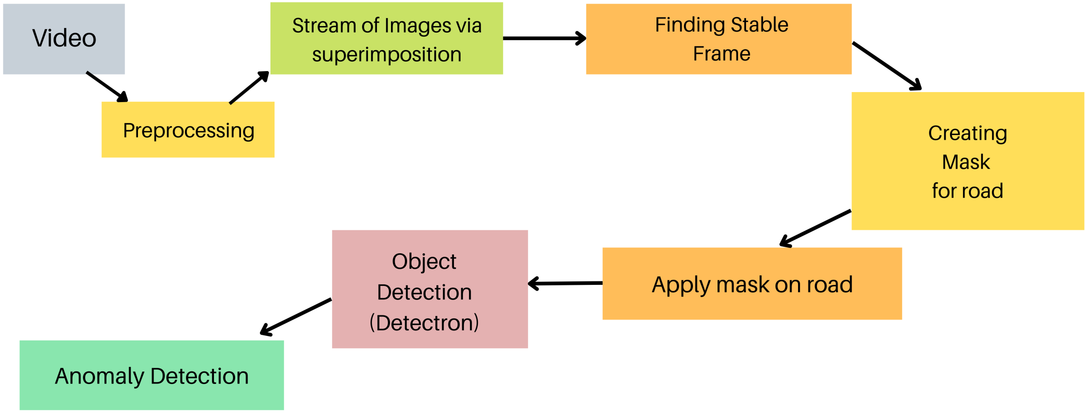


Figure 3.0.5: Methodology for vehicle activities

the highest scored instances in the positive bag and the negative bag.

Now for vehicle related anomaly i.e vehicles at stall or crashed, either a vehicle is at stall for long period of time or there is sudden change in trajectory of vehicle in short span of time. In order to find such features we have first preprocess the video then further divided it into continuous stream of superimposed images where each image contains the information of 30 frames merged in one i.e 30fps in which only slow moving objects will be considered to subjected to anomaly as stall and sudden stop after crash are the main reason for anomaly.

After superimposing images we required to find stable frames in order to remove area that is not a part of road since it can lead to false positive detection as vehicle can be parked , filling gas ,etc . And in order to identify the road region in image we can use a simple fact the movement / change in pixel will be the most if we considering a set of continouse superimposed images , so in such case Local Binary Pattern is one of the feature which can be used for texture representation. This featured histogram of the texture information will have different histogram even if the gray-level histogram are same for two different scenes. So, using local binary pattern it can be identified whether a vehicle on the road is stable or not. Local binary pattern (LBP) is one of the important features used for texture analysis. Most of the literature used circular neighborhood structure to evaluate the binary pattern[17]. Ojala et.al [15] used 8 neighboring pixels for evaluation of LBP to capture the texture feature. We have used a 3x3 neighborhood structure as shown in Fig. 3.4 is considered for the evaluation of local binary pattern(LBP) feature. After using LBP with two different threshold we would get the change start and change end i.e frame by frame significant change and frame by frame significantly small change with this for each video time range instances will be created which will help us identify frames with lesser movement(stable frames) and once we find such intervals in the video we can again use LBP to get the idea of the road region in image and store such region in file (.npy) and late will be used to mask the road to avoid rising of false positive while detecting anomaly. After applying masking on image we would not only had eliminated the region of false positive but also had changed the problem to object detection in superimposed image with some constraints which makes the case for anomalous event . And for object detection we would be using Detectron [18] which is a Facebook AI Research's software system that implements state-of-the-art object detection algorithms, including Mask R-CNN. It is written in Python and powered by the Caffe2 deep learning framework. The goal of Detectron is to provide a high-quality, high-performance codebase for object detection research. It is designed to be flexible in order to support rapid implementation and evaluation of novel research. Detectron includes implementations of the following object detection algorithms: Masked R-CNN, Retinanet, Faster-RCNN using many backbone architectures like Resnext versions, Resnet version and vggnet. Also we had hard coded the day and night videos scene in file which will help in vehicle detection including small vehicles. Now for each video we would be having anomalous event we would not only be having bounding box, but start and end time of anomaly i.e accuracy of predicting anomaly in video but also predicting the time error if anomaly predicted correctly.

3.1 Experimentation

Experimenting with super-imposition of frames rather than taking the average of all 16-frame clip features within that segment. Experiment with I3D model (Two-Stream Inflated 3D ConvNets).
Experimentation with different networks for final classification.

1. ResNet-50, 101..
2. ImageNet
3. VGGNet
4. Faster R-CNN
5. EfficientNet models

Experimentation with post-processing techniques such as query expansion. Two complementary predictors: one works on the normal scale of videos, while the other works on a magnified scale on videos missed by the first predictor. Box-level tracking of the potential spatio-temporal anomalous tubes. Dual-mode(static and dynamic) analysis method that integrates background modeling, vehicle detection and segmentation using Mask R-CNN, followed by outlier filtering. Foreground segmentation to reduce the search. Perform Real-Time Detection of anomalies.

3.2 Datasets

1. UCF-Crime dataset (https://www.kaggle.com/mission-ai/ucfdatasetforanomaly)
   (a) The UCF-Crime Dataset is a collection of 1900 videos, 950 of them composed of normal events and 950 of abnormal events.
   (b) Abnormal events are also labeled in 13 different classes, so the data set can be used both for detecting anomalies and for classifying behaviors.

   (c) This dataset is especially relevant due to its large size, consisting of more than 13 million frames and a duration of more than 128 hours.

2. AI city traffic anomaly dataset

   (a) AI city traffic anomaly dataset is collection of 100 videos of 15 minutes in length.

   (b) The anomalies are the event which are either stalled vehicle or due to car crashes or stalled vehicles.

3.3 Performance Metrics

1. Accuracy: Binary classification

2. Confusion matrix:

   (a) Precision: Fraction of elements well classified from among those classified as positive class.

   (b) F1-score: It is defined as the harmonic mean between the TPR and the precision. It is a measure of the quality of the model w.r.t a certain class.

3. Area Under Curve (AUC-ROC) : It measures the capability of distinguishing the different classes by the model and can be created by plotting False Positive Rate(x-axis) vs True PositiveRate(y-axis). Values ranges [0,1] and more the value closer to 1 the better the model have performed.
   True Positive Rate=TP/(TP+FN)
   False Positive Rate=FP/(FP+TN)

4. NRMSEt = min(RMSEi)/300
   It is the normalized root mean square error of the predicted anomaly time.

5. Score= F1 X (1-NRMSEt)
   It is considering two factors that how precisely the anomaly is detect and how much was the error time i.e actual time of anomaly event and predicted time of anomaly event.

Credit is provided for detecting each anomaly in video recordings with numerous ground-truth abnormalities. Many erroneous predictions in a single video clip, on the other hand, are considered multiple false alarms. We only evaluate the anomaly with the smallest detection time error if numerous anomalies are supplied within the time frame of a single ground-truth anomaly. We expect all anomalies to be effectively identified, and the F1 component of the S4 assessment score penalises missed and erroneous detections. For all genuine pos- itives, we calculate the detection time error as the RMSE between the ground-truth anomaly start time and anticipated start time. We calculated N RM SEt as the normalised detection time RMSE between 0 and 300 frames (for movies of 30 frames per second, this amounts to 10 seconds) to give a normalised evaluation score, which reflects a tolerable range of RMSE values for the anomaly detection job.

Chapter 4

RESULTS AND ANALYSIS

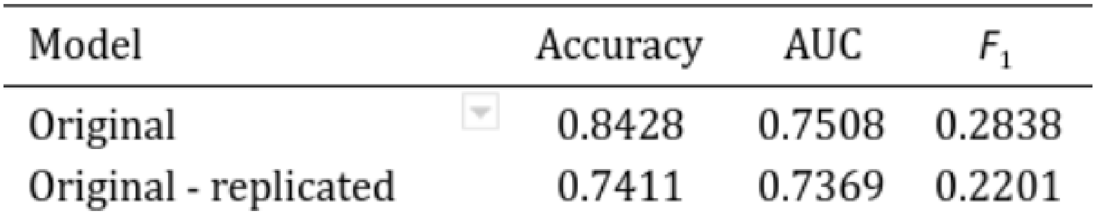
1. UCF Dataset

   (a) Preprocessing: Since UCF dataset had 1800 videos so the training set we have 800 normal and 810 anomaly videos whereas testing set have 150 normal and 140 anomalous videos.Also in both set it contains all different kind of human activity anomalies at various time instances in the videos and some having multiple instances of anomaly in the video.

   (b) Feature Extraction: We extract visual features from the fully connected layer of the C3D network[10]. Before computing features, we re-size each video frame to 240 × 320 pixels and fix the frame rate to 30 fps. We compute C3D features[5] for every 16-frame video clip followed by L2 normalization. To obtain features for a video segment, we take the average of all 16-frame clip features within that segment.

   (c) Baseline Inference: We use ReLU[4] activation and Sigmoid activation for the first and the last FC layers respectively, and employ Adagrad [10] optimizer with the initial learning rate of 0.001. The parameters of sparsity and smoothness constraints in the ranking loss are set to 1 =2 = 8 × 105 for the best performance. We divide each video into 32 non-overlapping segments and consider each video segment as an instance of the bag. The number of segments (32) is empirically set. We also experimented with multi-scale overlapping temporal segments but it does not affect detection accuracy. We randomly select 30 positive and 30 negative bags [13] as a mini-batch. We compute gradients by reverse mode automatic differentiation on computation graph using Theano[14]. Specifically, we identify set of variables on which loss depends, compute gradient for each variable and obtain final gradient through chain rule on the computation graph. Each video passes through the network and we get the score for each of its temporal segments.

2. AI city traffic anomaly dataset



Model	Accuracy	AUC	F_1
Original	0.8428	0.7508	0.2838
Original - replicated	0.7411	0.7369	0.2201

Figure 4.0.1: Model results

(a) Video segmentation : Each video is around 15 minutes long and anomaly is considered to be object at stop for long period or the object which had change in trajectory due to accident . Therefore emphasis should be done on filtering fast moving objects from the video as they don't come under classification of anomaly.So we interpret a video as continuous array of superimposed images as:

$$s\_img(1), s\_img(2), s\_img(3), ..., s\_img(n)$$

where s_img(1)= 1st frame and s_img(i) = prev_s_img*(1-k) + k*current_frame(i) , k is constant with value 0.01 since it signify the importance of previous frame more wrt to current frame so only those object which are slow motion object will be captured.

(b) Road Masking : Since object away from road can be stationary doesn't indicate any anomaly as may be parked so to remove detection of these as a anomaly masking of road is required . Masking will help us to narrow down the region of importance for detection of anomaly as well as help us to only identify the anomaly that are taken place on road. To identify the road we must find frames which not much of movement in video.

i. Local Binary Pattern : While using pattern we have classified the change in frames into two part as change start and change end. For change start threshold has been set as 7000 and for change stop threshold has been set as 2000. And on combining the result of change start and end we get the time range for each video when the movement on video is minimum.

ii. Masking : For each video in minimum movement frames comparsion of differences between two consecutive frames is done and the regions in which changes exceeds the threshold of 2000 is considered the area of movement which is basically the road. To create the a raw motion mask sum up of changes between consecutive is done and finally smoothing is done by adding very small increase in mask. All the mask are saved in .npy format for each respective video.

iii. Applying Masking : For each video their respective road mask are applied on the array of average images to remove the area other than road.

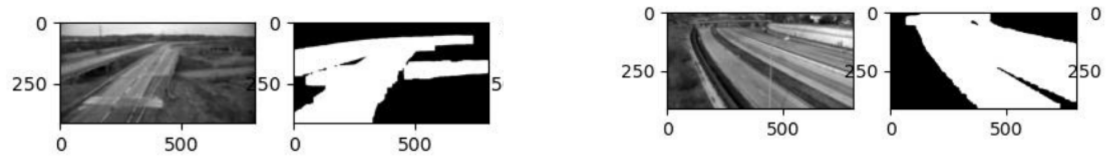


Figure 4.0.2: Road with their respective mask pattens

Table 4.0.1: Comparision between models

Model	F1 score	RMSE	Score
Baseline	0.9412	11.2556	0.9059
Without Mask	0.9126	14.2517	0.8692
With Mask	0.9468	14.0283	0.9025

While comparing with baseline we were closely able to detect the start time in case of anomaly videos but had some variation in without masking time detection because of the fact of false positives detection of anomaly in some cases started early or in some case it got delayed by 5 to 10%.

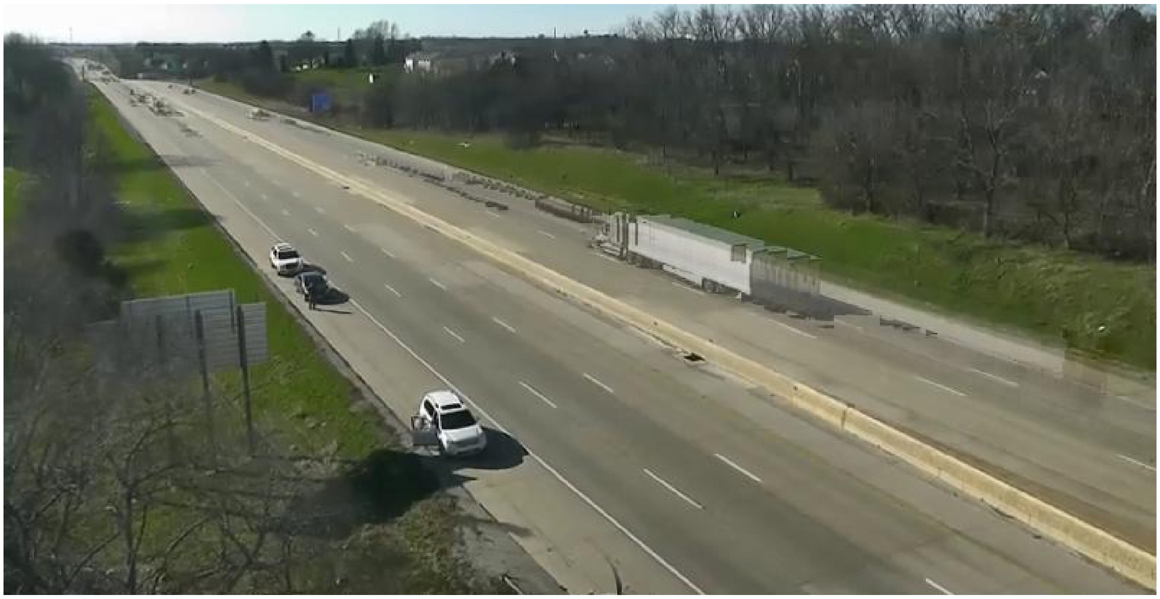


Figure 4.0.3: Superimposed Image

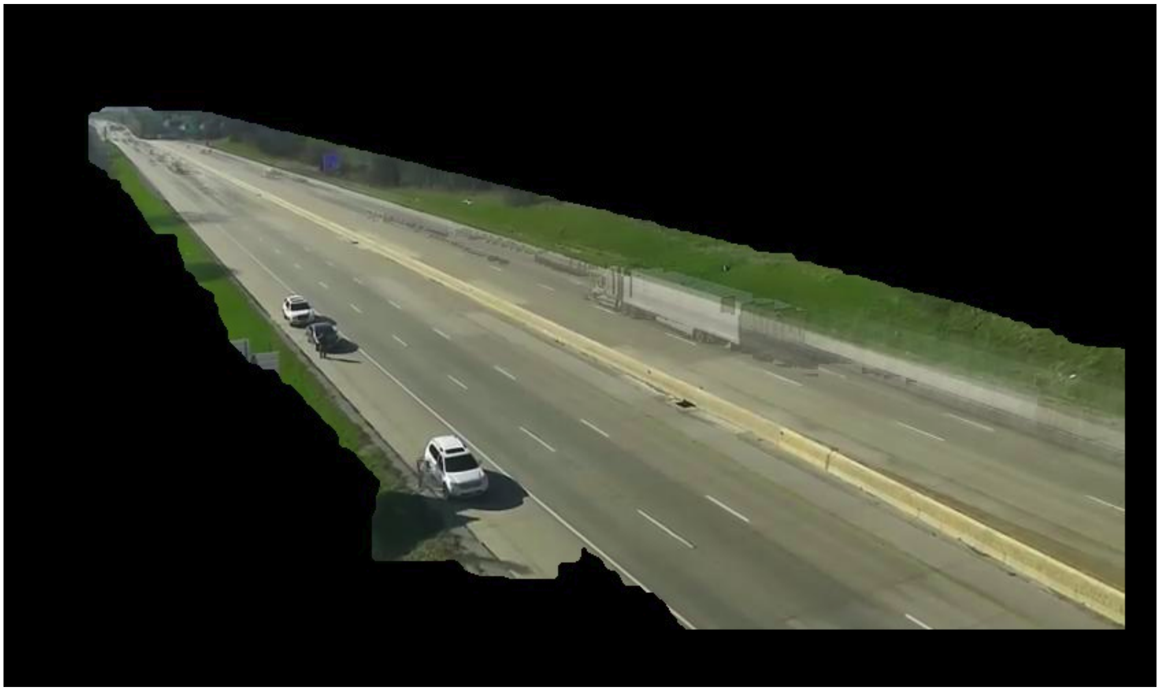


Figure 4.0.4: Masked image

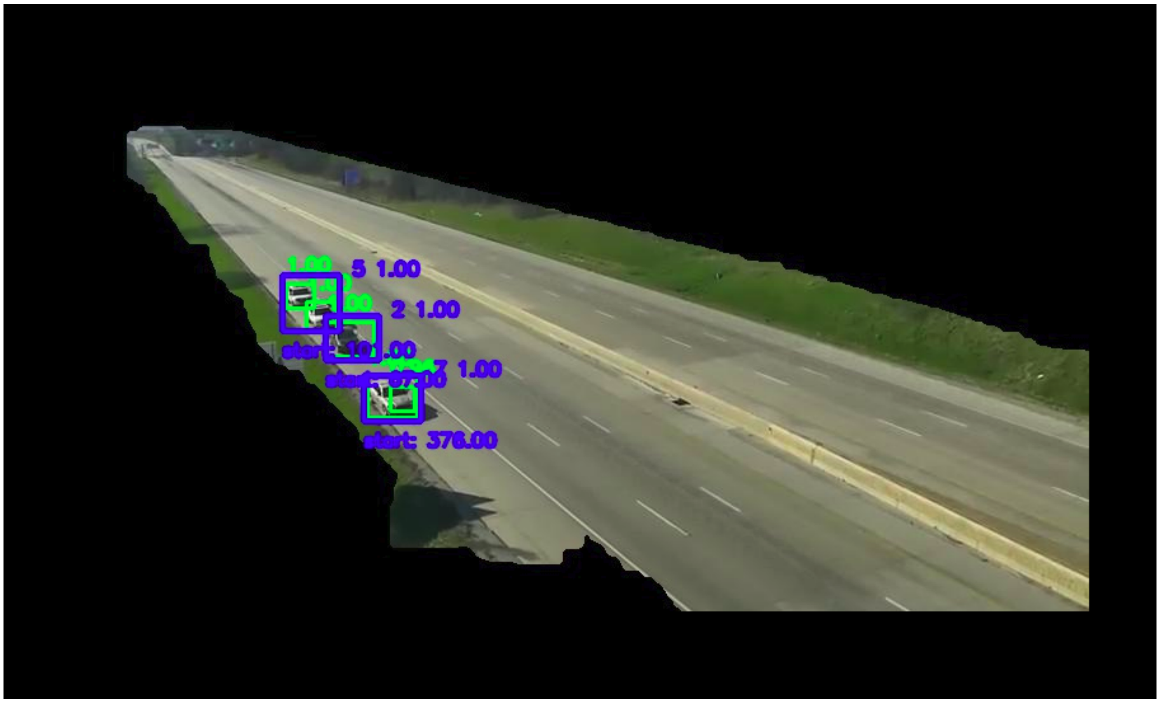


Figure 4.0.5: Anomaly Detection in form of bounding box with start time

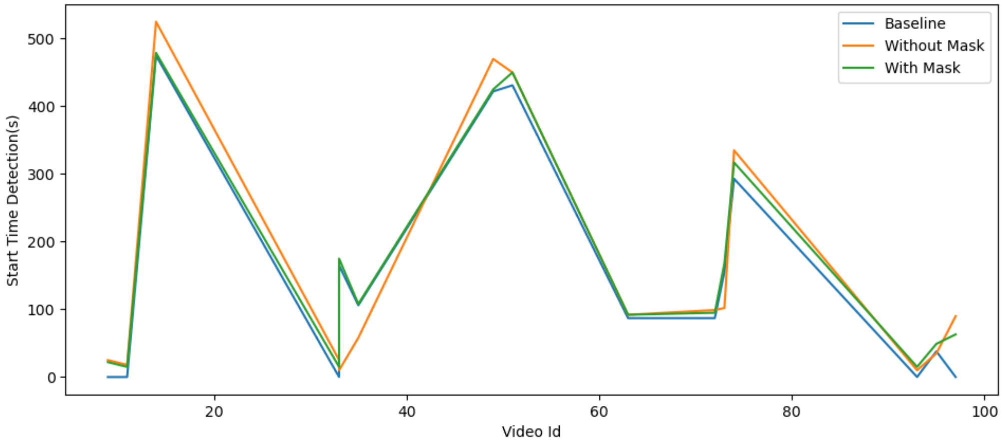


Figure 4.0.6: Comparison of start time detection of Some Videos for baseline , without masking and masking

Chapter 5

CONCLUSION AND FUTURE WORK

Real world anomalous events are complicated and diverse and is quite challenging to list all the possible anomalous events. So an attempt to exploit both normal and anomalous surveillance videos is done. We have implemented superimpose of images to reduce redundant data. Further road masking technique has been used to reduce false positives. While applying masking on images. The local binary pattern(LBP) for the given region exceeds the threshold then this region is not a part of the road and previously in the baseline the region which is not part of the road are considered which could lead to increase in false positives. As in the replicated model, the accuracy achieved was around 74%. On applying masking and comparing with non masking on the given threshold we have detected around 3-6% frames in each video which may had lead to false positive count and would had led to increase in f1 score which could had reduced the score for identifying efficiency of model. After generating masked superimposed images we finally localized the anomaly then we used detectron with fast-RCNN to detect the stalled vehicles or the anomalies in the frames. We detected anomalies in both day and night videos. The results achieved are quite promising with an F1-score of 0.9126 without applying the mask and 0.9468 when applied mask. The results achieved by our methodology are quite comparable with the baseline model. We found out that applying masking on superimposed frames reduces false positive rate. Further on, detecting anomalous events in real time video might be a complex problem to solve which will be addressed in our future work.

REFERENCES

[1] Louis Kratz and Ko Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1446–1453, 2009.

[2] Dan Xu, Elisa Ricci, Yan Yan, Jingkuan Song, and Nicu Sebe. Learning deep representations of appearance and motion for anomalous event detection. 09 2015.

[3] Waqas Sultani and Jin Young Choi. Abnormal traffic detection using intelligent driver model. In *2010 20th International Conference on Pattern Recognition*, pages 324–327, 2010.

[4] Bin Zhao, Li Fei-Fei, and Eric P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In *CVPR 2011*, pages 3313–3320, 2011.

[5] Yingying Zhu, Nandita M. Nayak, and Amit K. Roy-Chowdhury. Context-aware activity recognition and anomaly detection in video. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):91–101, 2013.

[6] Huiwen Guo, Xinyu Wu, Nannan Li, Ruiqing Fu, Guoyuan Liang, and Wei Feng. Anomaly detection and localization in crowded scenes using short-term trajectories. In *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 245–249, 2013.

[7] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. *CoRR*, abs/1801.04264, 2018.

[8] Milind Naphade, Shuo Wang, David C. Anastasiu, Zheng Tang, Ming-Ching Chang, Xiaodong Yang, Yue Yao, Liang Zheng, Pranamesh Chakraborty, Anuj Sharma, Qi Feng, Vitaly Ablavsky, and Stan Sclaroff. The 5th AI city challenge. *CoRR*, abs/2104.12233, 2021.

[9] T. Kar and P. Kanungo. A texture based method for scene change detection. In *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, pages 72–77, 2015.

[10] Ankit Kariryaa. Maskit: Masking for efficient utilization of incomplete public datasets for training deep learning models. *CoRR*, abs/2006.12004, 2020.

[11] Santhosh Kelathodi Kumaran, Debi Prosad Dogra, and Partha Pratim Roy. Anomaly detection in road traffic using visual surveillance: A survey. *CoRR*, abs/1901.08292, 2019.

[12] Milind Naphade, Shuo Wang, David C. Anastasiu, Zheng Tang, Ming-Ching Chang, Xiaodong Yang, Liang Zheng, Anuj Sharma, Rama Chellappa, and Pranamesh Chakraborty. The 4th AI city challenge. *CoRR*, abs/2004.14619, 2020.

[13] Salisu Wada Yahaya, Ahmad Lotfi, and Mufti Mahmud. A consensus novelty detection ensemble approach for anomaly detection in activities of daily living. *Applied Soft Computing*, 83:105613, 2019.

[14] Shyma Zaidi, B Jagadeesh, K V Sudheesh, and Arlene A Audre. Video anomaly detection and classification for human activity recognition. In *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pages 544–548, 2017.

[15] Du Tran, Lubomir D. Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. C3D: generic features for video analysis. *CoRR*, abs/1412.0767, 2014.

[16] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[17] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.

[18] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.

Appendix A

BIO-DATA

Name: Ayush Bhandari
Address: National Institute of Technology Karnataka, Surathkal
Email: ayushbhandari.181it209@nitk.edu.in
Contact No.: +91-7457871967

Name: Jaidev Chittoria
Address: National Institute of Technology Karnataka, Surathkal
Email: jaidevchittoria.181it119@nitk.edu.in
Contact No.: +91-9611401222