A data-driven approach for road accident detection in surveillance videos

Ariba Zahid[1] · Tehreem Qasim[2] · Naeem Bhatti[1] · Muhammad Zia[1]

Abstract

The use of machine learning and computer vision techniques for detecting road accidents is a challenging task due to the limited availability of accident data for training. Staging fake accidents with real cars is expensive, and car crashes are rare incidents in roadside CCTV footage. Therefore, simulating fake car crashes using computers can be a feasible option. As such, we look at the following question in this paper; *how successful can manually generated fake accident data be in terms of enabling a machine learning algorithm to detect real accidents?*. In this work, we manually construct fake accident video frames from normal video traffic footage by creating simulated accidents. We do so by following predefined principles that maintain consistency with the scene context of normal frames. In order to detect real accidents in video footage, we fine-tune pre-trained deep convolutional neural networks on the manually generated fake accident frames. We use four pre-trained models i.e., AlexNet, GoogleNet, SqueezeNet and ResNet-50 on both normal and abnormal traffic video frames during the learning phase. The experimental results show that the fine-tuned AlexNet outperforms other models providing an 80% percent true positive rate when detecting anomalies (accidents) in real-world surveillance videos of UCF-Crime dataset. This demonstrates the validity of our hypothesis that simulated accident data could be valuable for training machine learning algorithms to detect real-world accidents.

Keywords Anomaly detection · Road accident detection · Video surveillance · Transfer learning · Data preparation

1 Introduction

Various innovative technologies are being used in smart cities to improve the quality of human life. Due to the increase in population mobility, the number of moving vehicles on roads has tremendously increased. It is essential to put in place a surveillance system to

✉ Tehreem Qasim
  dr.tehreem@szabist-isb.edu.pk

[1] Department of Electronics, Quaid i Azam University, Islamabad, Pakistan

[2] Department of Computer Science, SZABIST, Islamabad, Pakistan

monitor the traffic flow on roads and to potentially detect any untoward incidents. CCTV cameras installed on roadsides are an effective tool to address this issue. However, this in turn gives rise to the challenge of monitoring the footage which is being recorded continuously. With the increasing number of CCTV camera systems being installed, it is difficult to hire sufficient human resources to keep an eye on the large volume of video footage. Computer Vision (CV) based methods [1, 2] are a suitable choice to automate the process of CCTV footage monitoring. Anomaly detection techniques based on CV are not only efficient but are also cost effective [5, 11, 14, 15, 20, 23–25, 39].

For the purpose of accident detection, ideally, each smart roadside camera should be trained with its own recorded video data because the video and environment is different for each camera. A machine learning framework trained on generalized accident videos taken from web may perform well at some locations, but might fail at other locations where the view of the scene or nature of the accident is remarkably different. We usually do not have sufficient accident data for each individual camera to train its own accident detection model.

One possible solution to this problem is to generate fake accident video frames from the normal frames obtained from a roadside CCTV camera. This way we do not have any limit to the number of generated accident frames. Also, we have the freedom of simulating different types of accidents to make the model intelligent enough to perform well in different real world situations. These are not limited to but could include: *a car rolling over, a car crashing into a tree or a wall or a pole, a car hitting a pedestrian, two cars colliding into each other, a car catching fire due to fuel tank issues, smoke emanating from a car etc.*

Focusing on road anomaly (accidents) detection in surveillance videos, we hypothesize that training a smart CCTV camera with artificial (simulated) data covering different possible types of accidents in the visible range that it covers, improves its ability to detect the real accidents in terms of accuracy. In the preliminary research in this direction, we utilize different traffic videos in the UCF Crime dataset. These videos are recorded using roadside cameras and contain short videos of each camera. Our main goal is to enable a model (that can learn the individual environment and scenery visible to each camera) to detect accidents in the scenarios where even no accident (or very few learning examples) is available. In order to achieve the set goal, we propose to prepare fake accident examples by taking some normal frames from each camera footage and then manually create accident situations at different locations in the visible view.

The experimental results show that performing the training using both, the normal and fake accident frames enables a machine learning model to detect real world accidents in the scene visible to a camera even if no prior real accident had taken place (which could have been used for training). For our experiments, we utilize popular pre-trained CNNs i.e., AlexNet, GoogleNet, SqueezeNet and ResNet-50. These CNNs are fine-tuned with two class data containing normal and accident video frames. Moreover, we observe that the AlexNet leads in terms of road accident detection accuracy.

Rest of the paper is organized as follows. In Section 2, related work is given. In Section 3, the proposed approach is described. Experimental results and discussion are given in Section 4 and conclusion is drawn in Section 5.

2 Related work

Recently, there has been a growing interest among the researchers in the field of anomaly detection in road traffic surveillance videos. The presented approaches are based on both,

the classical machine learning and deep learning. We first provide a brief summary of the traditional machine learning based approaches.

In [12], Ki and Lee propose a technique to detect accidents on road by extracting position and velocity features of the vehicles. Their method detects the vehicles and computes trajectories for feature extraction. Lucas-Kanade optical flow is used in the technique proposed by Rasheed et al. [26]. Their method performs foreground detection with Gaussian mixture model before computing the optical flow. Features are extracted from the optical flow which contains information of displacements and directions related to each pixel. The computed features are fed to a feed forward neural network for classification. Huang et al. [9] use Gaussian mixture model to detect the vehicles. They use mean shift algorithm for vehicle tracking. Three different features i.e., direction, acceleration and change in position of the vehicle are utilized for the purpose of anomaly detection.

Parvathy et al. [21] propose a technique of optical flow estimation in which optical flow is used to define trajectories. They cluster the trajectories hierarchically by using time and space information to learn motion patterns. Statistical methods i.e. probability distribution are used to detect anomalies from the statistical motion patterns. Any deviation from regular motion patterns is considered as an anomaly. A spatial localization constrained sparse coding technique is introduced by Yuan et al. [38] for traffic anomaly detection. Their method spatially localizes an object using sparse reconstruction. Direction and magnitude of the object motion are adaptively weighted and fused by using a Bayesian model. This technique is useful for anomaly detection in dash-cam videos.

A fast anomaly detection method based on sparse optical flow is proposed by Tan et al. [34]. Computation of optical flow is made efficient with foreground mask and spatial sampling. Forward-backward filtering and feature selection is used to increase the robustness of optical flow. For the detection of slow speed car and static vehicles, foreground channel is added to feature vector. Vatti et al. [36] proposed a smart system to detect road accidents and inform emergency contact numbers. They used gyroscope and vibration sensors for the accident detection and GSM module to send the information of accident along with the location identified by the GPS module.

Amin et al. [32] proposed a GPS based technique to detect road accidents. They monitor the speed of the vehicles by GPS and compare it to the previous speeds of vehicle every second by the use of a micro-controller unit. Accident is reported to the service center with the location of the vehicle whenever the speed of the vehicle is less than a specified speed. A technique called textures of optical flow is proposed by Ryan et al. [27] to detect abnormalities in surveillance videos. The uniformity of the flow field is measured to detect anomalies such as vehicles, bicycles, skateboarders etc., and is combined with the spatial information to detect other anomalies. A method for the estimation of optical flow proposed by the Black and Anandan [4] is used in this framework. The algorithm proposed by Black and Anandan has a drawback for real time assessments as it does not work well with larger images. Since anomalies do not occur at pixel level rather they occur at object level so full resolution is not required. Hence, objects are identifiable even with smaller resolutions. For this reason, prior pre-processing images are down sampled in [27] to lower resolution. To detect visual anomalies, a three stage pipeline is introduced by Biradar et al. [3] to learn motion patterns in surveillance videos. First step is the identification of the motionless objects and background is estimated for this purpose from recent history frames. Normal or anomalous behavior is localized from this background image. The object of interest is detected from this estimate of background and then categorized into anomaly based on time-stamp aware abnormality detection algorithm. To remove false positives, a post-processing technique is also presented

but in some cases due to the limitations of background estimation and detectors, some false positives happen for patches, road dividers, signboards etc.

Recently, deep networks have brought about tremendous success in terms of performing different tasks in video processing e.g. action recognition, sports, health-care, robotics etc. [6, 17, 17, 17, 29]

This has led to a growing interest among the researchers to investigate the applications of deep learning in road accident detection. Singh and Mohan [28] proposed a framework for accident detection, where denoising auto-encoders in combination with support vector machines are trained on normal surveillance videos. Likelihood of deep representation and reconstruction error is used as a key to determine an accident. The performance of this framework becomes inefficient in poor lightening conditions, occlusion and due to diversity of traffic patterns. Nasaruddin et al. [19] proposed a method in which instead of using whole frame information, anomaly is detected by finding region of interest from the spatio-temporal information. Robust background extraction technique is used to extract the motion features and find the attention region. A 3D convolutional neural network is used to get the most of deep spatio-temporal information. Their method is applicable to road accidents as well as general purpose anomaly detection.

Taking into consideration the lack of labeled training data for normal videos, Wei et al. [37] proposed a method based on background modeling to detect static vehicles which stay still for a relatively long time. In this method mixture of Gaussian (MOG2) is used for background modeling. Their method removes all moving vehicles from foreground and static vehicles are left as background. The static vehicles are then detected using Fast RCNN object detector. For these vehicles, the decision to detect an anomaly is made by using some pre-defined conditions. This method only gives a rough estimation of the start of an anomaly and is not precise in detection. Sultani et al. [30] proposed a method to avoid annotation of the abnormal segments of video frames which is a tedious task. For this purpose, they employ a multiple instance learning (MIL) technique. For training, annotation is done at video level instead of clip-level. The videos are considered as bags and segments are considered as instances in the MIL. To obtain better results for anomaly localization, sparsity and temporal smoothness constraints are introduced in ranking loss function during training.

Prabakran et al. [22] proposed novel multi-input neural network incorporating spatio-temporal features and dense flow features to detect anomalies and identify point and duration of anomaly in surveillance videos. They use optical flow for extraction of high-level information and C3D for low-level information. To learn motion-aware features a temporal augmented method is introduced by Yi and Shawn [41]. They use an attention block to incorporate temporal context into MIL. In [35], the authors used a pre-trained ResNet-50 model for feature extraction and then these features are fed to a bi-directional long-short term memory network for classification. Authors in [16, 18, 40] proposed a network based on sparse representation and dictionary learning algorithms for anomaly detection. Their proposed networks learn the dictionary of normal behaviors based on sparse representation.

We note that the discussed approaches presented in literature are all based on detection of accidents based on some real accidents that happened in the past. This is a major short-coming in situations where the accidents that happen after deployment of the model are of different nature than the ones used in training. Moreover, the discussed approaches have been tested on traditional video datasets which contain training videos and testing videos recorded at different locations. This does not guarantee that the trained model will perform with high accuracy if it is deployed at locations which are different to the ones in training data. To address these challenges, we construct fake accident video frames in this work for videos recorded with different individual cameras. The constructed fake accident frames contain

different types of accidents e.g., collisions between different vehicles etc. We utilize some popular pre-trained CNNs using the artificially generated accident frames. Experimental outcomes show that the trained models are able to detect real accidents based on the training data.

3 The proposed approach

In this section, we describe the UCF-Crime dataset, manual construction of fake accident data and the deep networks used for training with the prepared data in our framework. The proposed framework for the anomaly detection and classification is shown in Fig. 1.

3.1 UCF-Crime dataset

UCF-Crime dataset is a publicly available benchmark for anomaly detection collected by UCF (University of Central Florida) center for research in computer vision. This dataset has long untrimmed surveillance videos which cover 13 real world anomalies including abuse, arrest, arson, assault, accidents, burglaries, explosion, fighting, robberies, shootings, stealing, shoplifting and vandalism. Out of 1900 videos present in this dataset 950 videos are anomalous and other 950 are normal. The video frame size is 320×240 and frame rate of the video sequences is 30 fps. Some of the normal and abnormal frames extracted from the UCF-Crime dataset are shown in Fig. 2. To implement the proposed method, we use only the road accident videos available in the dataset. Out of these videos, we did not use the ones where:

1. The quality of the videos is poor.
2. Dash cam videos, because the focus of this research is on videos recorded with stationary CCTV cameras on the road side.

The videos used in the experiments include video number 2, 7, 27, 60, 75, 132, 141 and 144.

Choice of the dataset The dataset is suitable for our work because it contains CCTV videos recorded with different road side stationary cameras. The only caveat is these videos are short and only contain one accident in a single video. Hence, we incorporate simulated accidents to the dataset to obtain more training data for the accident class.
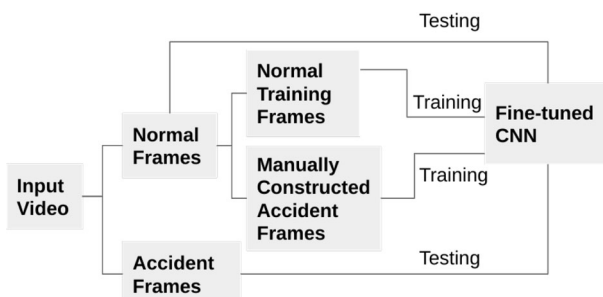

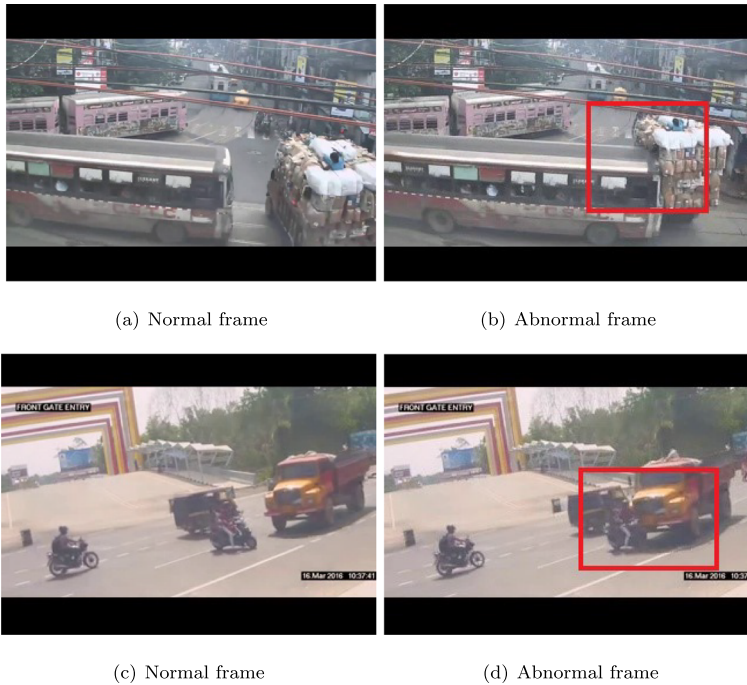**Fig. 1** Diagramatic representation of the proposed framework

(a) Normal frame   (b) Abnormal frame

(c) Normal frame   (d) Abnormal frame

Fig. 2 **a**, **b**, **c** and **d** normal and anomalous frames taken from different videos of of UCF-Crime dataset

3.2 Manual construction of fake accident data

UCF-Crime dataset contains total 150 videos containing road accident anomalies. These videos contain traffic accidents involving vehicles, pedestrians or cyclists. In some videos, the accidents are not clearly visible due to the camera angle and video quality. As discussed in Section 3.1, in this work, we utilize selected accident videos in the UCF-Crime dataset which contain a better view of the scene. Each video is captured from a different camera at a different location. There is no single video in the whole dataset with large footage on a single road junction which creates the problem of insufficient data for the training of a deep network. Moreover, there is only one accident event in each video. If we use the frames of that event for training then there is no accident data left for testing at that particular camera location. It is desirable to have many accident events at each camera location so that we may be able to train a deep network with both normal and accident frames and leave out some accident events for testing.

To overcome this problem of insufficient abnormal training data, we construct fake abnormal data frames from the normal ones. All these realistic hand-crafted anomalous data samples are constructed with great precaution. This way we have both normal frames and multiple fake accident frames for the purpose of training. We leave out the real accident event in each video for testing. While manually constructing the fake accident frames, we adopt some principles specific to the adopted dataset. Note that these principles can be adopted for any dataset in general. These are listed as follows:

1. The resolution of the vehicle (taken from an external image) inserted into a video frame to create a crash scene, should be similar to the rest of the vehicles in the frame.

2. The vehicle inserted into a normal video frame to simulate a crash scene should be taken from the time of the day similar to the original normal frame.
3. The vehicle should be taken from an external image source for which the camera is at a similar position and angle as the one used to capture the original video frames.
4. The vehicle that is inserted into a normal frame to create a crash scene should have same distance from the camera as the other car involved in the crash.

Figure 3 shows some of the manually constructed abnormal frames and corresponding normal frames.

3.3 Deep networks used

In the proposed framework, we use well known convolutional neural networks (AlexNet, GoogleNet, SqueezeNet and ResNet-50) and train them on manually constructed frames (containing fake accidents). We adopt transfer learning with two class data. Class one contains normal frames which are extracted from real world road accident videos of the UCF-Crime dataset. Class two contains manually constructed abnormal frames. These frames are constructed by using the normal traffic flow videos. Architectures of the employed CNNs are briefly described in the following.

Alexnet is trained on imagenet dataset set and has the ability to classify images into 1,000 objects categories [13]. Alexnet has 5 convolutional layers, 3 max-pooling layers, 2 normalization layers and 2 fully connected layers. Softmax is used for the final decision
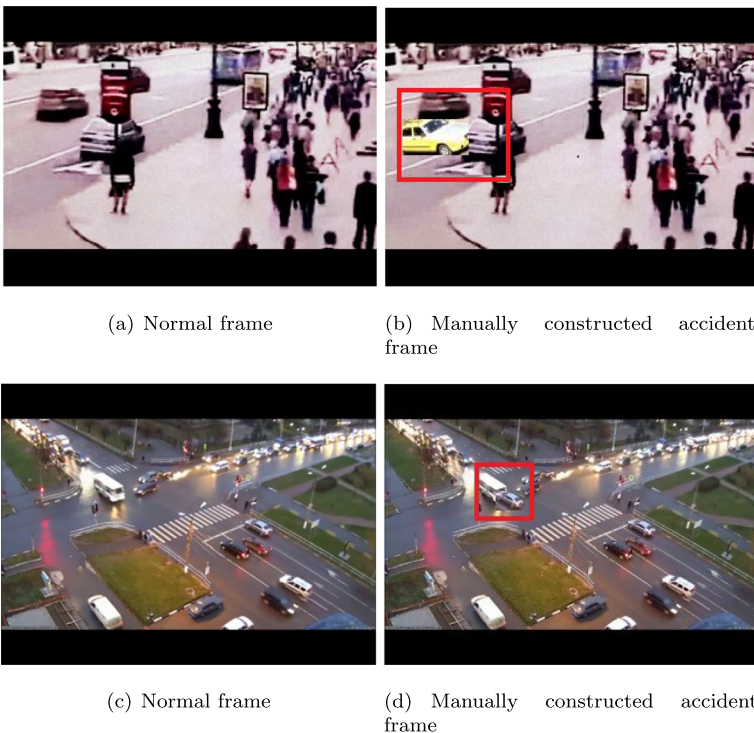

(a) Normal frame   (b) Manually constructed accident frame

(c) Normal frame   (d) Manually constructed accident frame

Fig. 3 Manually constructed accident frames

making. Alexnet uses ReLU as the activation function. Input images of size $227 \times 227 \times 3$ are used. The number of parameters utilized by AlexNet is over 60 million.

SqueezeNet is an 18 layers deep CNN [10]. It uses 1×1 filters instead of 3×3. It is trained on image-net data set set. It takes an input image of size 227×227. SqueezNet has an initial standalone convolution layer (conv1). Next, there are 8 fire modules. In the end there is a final conv layer (conv10). SqueezeNet includes max-pooling with a stride of 2 and ReLU is used as an activation function.

GoogleNet is trained on imagenet dataset and classifies images into 1,000 different object categories [33]. GoogleNet is 22 layers deep. It includes 27 pooling layers. GoogleNet also contains 9 inception modules which are connected to the global average pooling layer. It uses ReLU activation functions and softmax for classification.

ResNet-50 is 50 layers deep and contains 48 convolutional layers [8]. It also contains 1 max-pooling and 1 average pool layer. It is trained on image-net data set. It can classify images into 1,000 object classes. It takes an image input image of size of 224×224.

4 Experiments and results

In this section we present the performance evaluation of the proposed framework and a discussion on the experimental findings.

4.1 Performance evaluation

We evaluate the performance of the proposed anomaly detection framework using a subset (containing accident videos) of the UCF-crime dataset [31] which contains real world surveillance videos.

In this section, we provide the experimental results of the proposed approach. The test frames are extracted from the real accident videos of the UCF-Crime dataset. We train and test four different pre-trained networks: AlexNet, GoogleNet, SqueezeNet and ResNet-50 and present the performance comparison. In the visual results shown in Fig. 4, it can be observed that our method accurately detects the vehicle accidents on the road. The results shown in Fig. 4 are the test results of AlexNet. We perform a two class classification (normal and accident) using the fine tuned deep network, which provides a probability of prediction for each class. The normal detected frames are indicated by score value 0 and accident frames are with score value 1.

In order to evaluate the performance of our approach quantitatively, we use four empirical measures which are computed at frame level. These measures are given as:

True Positive (TP): A frame is said to be true positive, when the detection algorithm marks it as anomaly and it is annotated as anomaly in the ground-truth.

False Positive (FP): A frame is said to be false positive, when the detection algorithm marks it as anomaly but it is not annotated as anomaly in the ground-truth.

True Negative (TN): A frame is said to be true negative, when the detection algorithm mark it as normal and it is annotated as normal frame in the ground-truth.

False Negative (FN): A frame is said to be false negative, when the detection algorithm marks it as normal but it is annotated as anomalous in the ground-truth.

Fig. 4 Test results of AlexNet for video 75

Table 1 Quantitative results for video 2

	Video No:2					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	15	84	2	0	0.88	0
GoogleNet	16	63	1	21	0.94	0.25
ResNet-50	7	75	0	4	1	0.05
SqueezeNet	23	83	11	0	0.68	0

The the above mentioned variables are used to compute true positive rate (TPR) and false positive rate (FPR). The TPR and FPR are given as:

$$TPR = \frac{TP}{TP + FN} \qquad (1)$$

$$FPR = \frac{FP}{TN + FP} \qquad (2)$$

4.2 Discussion

The true positive rates and false positive rates of the employed CNNs for different videos are given in Table 1, 2, 3, 4, 5, 6, 7, and 8. The best over all results are achieved for AlexNet in the experiments. AlexNet detects accidents with minimum number of false positives and false negatives. Other networks are able to detect accidents for some videos, but with greater number of false positives and false negatives.

In Table 1, we can see that for video 1 of accident category in UCF-Crime dataset, AlexNet performs better with TPR of 0.88 while FPR is zero. ResNet-50 also shows good results for video 2, but it has an FPR of 0.05 which is greater than that of AlexNet. Googlenet has higher TPR and FPR values. The TPR of the SqueezeNet is less than all other networks but, it also has an FPR less than GoogleNet and ResNet.

Table 2 shows the test results of the networks for video 7 of the UCF-Crime dataset. For this particular video all the networks show a hundred percent TPR, which means accident is successfully detected by all the networks. Their performance differs on the basis of the FPR values. AlexNet performs better than GoogleNet and SqueezeNet. ResNet-50 shows best results for this video with an FPR of 0.08.

Table 3 shows the results for the video number 27. AlexNet detects accident with a TPR of 70 percent and FPR of 0. The GoogleNet also shows good results with a TPR of 0.94, but

Table 2 Quantitative results for video 7

	Video No:7					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	6	75	0	10	1	0.12
GoogleNet	6	72	0	13	1	0.15
ResNet-50	5	79	0	7	1	0.08
SqueezeNet	5	70	0	16	1	0.19

Table 3 Quantitative results for video 27

	Video No:27					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	12	74	5	0	0.70	0
GoogleNet	16	65	1	9	0.94	0.12
ResNet-50	1	87	3	0	0.25	0
SqueezeNet	11	67	5	9	0.7	0.11

Table 4 Quantitative results for video 60

	Video No:60					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	12	24	15	0	0.44	0
GoogleNet	11	14	14	10	0.44	0.41
ResNet-50	0	9	3	16	0	0.64
SqueezeNet	15	18	12	6	0.45	0.25

Table 5 Quantitative results for video 75

	Video No:75					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	78	31	1	0	0.99	0
GoogleNet	0	14	0	0	0	0
ResNet-50	0	0	0	0	0	0
SqueezeNet	0	31	79	0	0	0

Table 6 Quantitative results for video 132

	Video No:132					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	4	33	4	0	0.5	0
GoogleNet	5	33	1	0	0.8	0
ResNet-50	0	33	3	0	0	0
squeezenet	2	33	7	0	0.20	0

Table 7 Quantitative results for video 141

	Video No:141					
Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	5	55	10	0	0.33	0
GoogleNet	10	40	10	15	0.5	0.27
ResNet-50	11	48	0	6	1	00.11
SqueezeNet	9	44	1	12	0.9	0.21

Table 8 Quantitative results for video 144

Network Name	True Positive	True Negative	False Negative	False Positive	TPR	FPR
AlexNet	14	28	0	0	1	0
GoogleNet	14	16	1	11	0.93	0.40
ResNet-50	14	24	0	4	1	0.14
SqueezeNet	0	28	14	0	0	0

it has a higher FPR in comparison with AlexNet. ResNet-50 completely fails to detect the accident. The SqueezeNet has a TPR lower than alexnet and and also has a higher FPR.

The results shown in Table 4 are for the video number 60. All networks except ResNet-50 detect accidents but with a very low TPR. Test results of AlexNet for this video are also better than others as it has an FPR of 0.

The results for the video number 75 are illustrated in Table 5. AlexNet outperforms other CNNs with a TPR of 99 and an FPR which is 0. All other three networks fail to detect the accidents.

GoogleNet shows better results than AlexNet for video number 132 as shown in Table 6. GoogleNet has higher TPR than AlexNet and SqueezeNet. ResNet-50 failed to detect accident. All these results are shown numerically in Table 6.

The results for video number 141 are shown in Table 7. A TPR value of hundred percent and an FPR value of eleven percent is achieved by ResNet-50. AlexNet detects anomaly with a smaller FPR value. GoogleNet and SqueezeNet detect anomaly but thier FPR is quite high.

The results of Table 8 show that AlexNet detects accidents in video 144 with a higher TPR and minimum FPR. ResNet also detects anomalous events, but it shows a higher FPR. SqueezeNet fails to detect anomaly for this video, whereas GoogleNet has a higher FPR value.

From the results of fine tuned networks given above, we conclude that over all best results for accident detection are achieved by AlexNet CNN. For some videos, the fine-tuned AlexNet performs less well due to low quality of the videos. The best results are achieved for video 144 because this video is captured in good quality. The frames are visually clear and do not have any occlusion. This video is captured by the camera mounted at an elevated position, which is why all the vehicles are of the same size in a video frame. The experimental results indicate that manually constructed fake accident frames successfully enable the trained CNNs to detect real accidents.

5 Conclusion

In this paper, we present road accident detection using fine-tuned CNNs. In the proposed approach, pre-trained neural networks are trained on manually constructed image data using transfer learning in a data-driven paradigm. Abnormal frames of fake accidents are constructed using the road traffic videos from UCF-Crime dataset. This helps to overcome the shortage of footage of road accidents on a single road junction for the training of the neural networks. The trained models are tested to detect real accident frames in the UCF-Crime dataset. In the experimental evaluation, encouraging results are achieved for seven videos. We also observed that out of the four pre-trained neural networks, AlexNet performs best with higher true positive rate.

In the presented work, fake accidents are generated in a rough manner manually. In future we have an insight to use deep networks like GANs to simulate more real looking accidents. By using GANs for this purpose we may be able to have a large enough data to train the neural network for a practical real world application. In this approach, we have only used spatial data (individual frames). In future, we aim to use use fake but real looking temporal data (video sequences) to make predictions before an accident occurs.

Declarations

References

1. Bhatti UA, Yu Z, Li J, Nawaz SA, Mehmood A, Zhang K, Yuan L (2020) Hybrid watermarking algorithm using Clifford Algebra with arnold scrambling and chaotic encryption. IEEE Access 8:76386–76398
2. Bhatti UA, Yu Z, Yuan L, Zeeshan Z, Nawaz SA, Bhatti M, Mehmood A, Ain QU, Wen L (2020) Geometric algebra applications in geospatial artificial intelligence and remote sensing image processing. IEEE Access 8:155783–155796
3. Biradar K, Gupta A, Mandal M, Vipparthi S (2019) Challenges in time-stamp aware anomaly detection in traffic videos. ArXiv abs/1906.04574
4. Black MJ, Anandan P (1996) The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Computer vision and image understanding 63(1):75–104
5. Chang Y, Tu Z, Xie W, Luo B, Zhang S, Sui H, Yuan J (2022) Video anomaly detection with spatio-temporal dissociation. Pattern Recognition 122:108213
6. Gupta A, Muthiah SB (2022) Learning cricket strokes from spatial and motion visual word sequences. Multimedia Tools and Applications 1–23
7. https://www.crcv.ucf.edu/projects/real-world/. Accessed 20 Sept 2022
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp 770–778)
9. Huang X, He P, Rangarajan A, Ranka S (2020) Intelligent intersection: two-stream convolutional networks for real-time near-accident detection in traffic video. ACM Transactions on Spatial Algorithms and Systems (TSAS) 6(2):1–28
10. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. Preprint at http://arxiv.org/abs/1602.07360
11. Ilyas Z, Aziz Z, Qasim T, Bhatti N, Hayat MF (2021) A hybrid deep network based approach for crowd anomaly detection. Multimedia Tools and Applications 80(16):24053–24067
12. Ki Y-K, Lee D-Y (2007) A traffic accident recording and reporting model at intersections. IEEE Transactions on Intelligent Transportation Systems 8(2):188–194. https://doi.org/10.1109/TITS.2006.890070
13. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25
14. Li D, Nie X, Li X, Zhang Y, Yin Y (2022) Context-related video anomaly detection via generative adversarial network. Pattern Recognition Letters 156:183–189
15. Li N, Zhong J-X, Shu X, Guo H (2022) Weakly-supervised anomaly detection in video surveillance via graph convolutional label noise cleaning. Neurocomputing 481:154–167
16. Lu C, Shi J, Jia J (2013) Abnormal event detection at 150 fps in matlab. In: Proceedings of the IEEE International Conference on Computer Vision. pp 2720–2727
17. Ma X, Zhang Z (2022) Research on sports health care information system based on computer deep learning algorithm, Computational Intelligence and Neuroscience

18. Mohammadi S, Perina A, Kiani H, Murino V (2016) Angry crowds: detecting violent events in videos. In: European Conference on Computer Vision. Springer, pp 3–18

19. Nasaruddin N, Muchtar K, Afdhal A, Dwiyantoro APJ (2020) Deep anomaly detection through visual attention in surveillance videos. Journal of Big Data 7(1):1–17

20. Nguyen T-N, Meunier J (2019) Anomaly detection in video sequence with appearance-motion correspondence. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp 1273–1283

21. Parvathy R, Thilakan S, Joy M, Sameera K (2013) Anomaly detection using motion patterns computed from optical flow. In: 2013 Third International Conference on Advances in Computing and Communications. pp 58–61. https://doi.org/10.1109/ICACC.2013.18

22. Prabakaran A, Voleti MR, Lakshya, Manurkar PS (2019) A multi-input neural network with dense flow and spatio-temporal features for anomaly detection. In: 2019 Fifteenth International Conference on Information Processing (ICINPRO). pp 1–6. https://doi.org/10.1109/ICInPro47689.2019.9092161

23. Qasim T, Bhatti N (2019) A hybrid swarm intelligence based approach for abnormal event detection in crowded environments. Pattern Recognition Letters 128:220–225

24. Qasim T, Bhatti N (2019) A low dimensional descriptor for detection of anomalies in crowd videos. Mathematics and Computers in Simulation 166:245–252

25. Qasim T, Fisher RB, Bhatti N (2021) Ground-truthing large human behavior monitoring datasets. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, pp 2763–2770

26. Rasheed N, Khan SA, Khalid A (2014) Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks. In: 2014 28th International Conference on Advanced Information Networking and Applications Workshops. pp 61–66. https://doi.org/10.1109/WAINA.2014.18

27. Ryan D, Denman S, Fookes C, Sridharan S (2011) Textures of optical flow for real-time anomaly detection in crowds. In: 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp 230–235. https://doi.org/10.1109/AVSS.2011.6027327

28. Singh D, Mohan CK (2019) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. IEEE Transactions on Intelligent Transportation Systems 20(3):879–887. https://doi.org/10.1109/TITS.2018.2835308

29. Singh K, Malhotra J (2022) Smart neurocare approach for detection of epileptic seizures using deep learning based temporal analysis of EEG patterns. Multimed Tools Appl 1–32

30. Sultani W, Chen C, Shah M (2018) Real-world anomaly detection in surveillance videos. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA. pp 6479–6488. https://doi.org/10.1109/CVPR.2018.00678. https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00678

31. Sultani W, Chen C, Shah M (2018) Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 6479–6488

32. Syedul Amin M, Jalil J, Reaz MBI (2012) Accident detection and reporting system using GPS, GPRS and GSM technology. In: 2012 International Conference on Informatics, Electronics Vision (ICIEV). pp 640–643. https://doi.org/10.1109/ICIEV.2012.6317382

33. Szegedy C, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, Dumitru E, Vincent V, Andrew R (2015) "Going deeper with convolutions." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 1–9

34. Tan H, Zhai Y, Liu Y, Zhang M (2016) Fast anomaly detection in traffic surveillance video based on robust sparse optical flow. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp 1976–1980. https://doi.org/10.1109/ICASSP.2016.7472022

35. Ullah W, Ullah A, Haq IU, Muhammad K, Sajjad M, Baik SW (2021) CNN features with bi-directional lstm for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications 80(11):16979–16995

36. Vatti NR, Vatti PL, Vatti R, Garde C (2018) Smart road accident detection and communication system. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). pp 1–4. https://doi.org/10.1109/ICCTCT.2018.8551179

37. Wei J, Zhao J, Zhao Y, Zhao Z (2018) Unsupervised anomaly detection for traffic surveillance based on background modeling. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society, Los Alamitos, CA, USA. pp 129–1297. https://doi.org/10.1109/CVPRW.2018.00025. https://doi.ieeecomputersociety.org/10.1109/CVPRW.2018.00025

38. Yuan Y, Wang D, Wang Q (2017) Anomaly detection in traffic scenes via spatial-aware motion reconstruction. IEEE Transactions on Intelligent Transportation Systems 18(5):1198–1209. https://doi.org/10.1109/TITS.2016.2601655

39. Zhang Q, Feng G, Wu H (2022) Surveillance video anomaly detection via non-local u-net frame prediction. Multimed Tools Appl 1–16

40. Zhao B, Fei-Fei L, Xing EP (2011) Online detection of unusual events in videos via dynamic sparse coding. In: CVPR 2011. IEEE, pp 3313–3320
41. Zhu Y, Newsam S (2019) Motion-aware feature for improved video anomaly detection. In: BMVC