EFFICIENT ANAMOLY DETECTION IN THE REAL TIME SURVEILLANCE VIDEOS

[1]Dr., C. Gulzar Mtech, Ph.D [2]A.V Hari Chandra Reddy
[1]Associate Professor [2]P.G.Scholar
[1,2,]DEPARTMENT OF CSE
[1,2]Dr. K.V SUBBA REDDY INSTITUTE OF TECHNOLOGY
Dupadu Railway Station, Lakshmipuram Post, Kurnool, Andhra Pradesh, India - 518218

ABSTRACT

Surveillance videos are able to capture a variety of real- istic anomalies. In this paper, we propose to learn anoma- lies by exploiting both normal and anomalous videos. To avoid annotating the anomalous segments or clips in train- ing videos, which is very time consuming, we propose to learn anomaly through the deep multiple instance ranking framework by leveraging weakly labeled training videos,

i.e. the training labels (anomalous or normal) are at video- level instead of clip-level. In our approach, we consider normal and anomalous videos as bags and video segments as instances in multiple instance learning (MIL), and auto- matically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. Fur- thermore, we introduce sparsity and temporal smoothness constraints in the ranking loss function to better localize anomaly during training.

We also introduce a new large-scale first of its kind dataset of 128 hours of videos. It consists of 1900 long and untrimmed real-world surveillance videos, with 13 re-alistic anomalies such as fighting, road accident, burglary, robbery, etc. as well as normal activities. This dataset can be used for two tasks. First, general anomaly de-tection considering all anomalies in one group and all normal activities in another group. Second, for recog- nizing each of 13 anomalous activities. Our experimen-tal results show that our MIL method for anomaly detec- tion achieves significant improvement on anomaly detec- tion performance as compared to the state-of-the-art ap- proaches. We provide the results of several recent deep learning baselines on anomalous activity recognition. The low recognition performance of these baselines reveals that our dataset is very challenging and opens more op- portunities for future work. The dataset is available at:
https://webpages.uncc.edu/cchen62/dataset .html

KEYWORDS: Anamoly , cnn, opencv, Machine Learning

I. INTRODUCTION

Surveillance cameras are increasingly being used in public places e.g. streets, intersections, banks, shopping malls,etc. to increase public safety. However, the monitoring ca-pability of law enforcement agencies has not kept pace. The result is that the reisaglaring deficiency in the utilization of surveillance cameras and an unworkable ratio of cameras to human monitors. One critical task in video surveillance is detecting anomalous events such as traffic accidents ,crimesor illegal activities. Generally, anomalous events rarely occur as compared to normal activities. Therefore, to alleviate the waste of labor and time, developing intelligent com-puter vision algorithms for automatic video anomaly detection is a pressing need.The goal of a practical anomaly detection system is to timely signal an activity that deviates normal patterns

and identify the time window of the occur-ring anomaly. Therefore, anomaly detection can be considered as coarse level video understanding, which filters out anomalies from normal patterns. 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 as pecific anomalous event, for example violence detector[30]and traffic accident detector[23, 34]. However, it is obvious that such solutions cannot be generalized to detect other anomalous events, there fore they render a limited use in practice.

Real-world anomalous events are complicated and diverse. It is difficult to list all of the possible anomalous events. Therefore, it is desirable that the anomaly detection algorithm does not rely on any prior information about the events. In other words, anomal y detection should bedone with minimum supervision. Sparse-coding based approaches [28,41] are considered as representative methods that achieve state-of-the-art anomaly detection results. 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 re constructable from the normal event

Motivation and contributions.

Although the above-mentioned approaches are appealing ,they are based on the assumption that any pattern that deviates from the learned normal patterns would be considered as an anomaly. However, this assumption may not hold true because it isvery difficult or impossible to define a normal event which takes all possible normal patterns/behaviors into account[9].More importantly, the boundary between normal and anomalous behaviors is often ambiguous. In addition, un-der realistic conditions, the same behavior could be a nor-mal or an anomalous behavior under different conditions. Therefore, it is argued that the training data of normal and anomalous events can help an anomaly detection system learn better .In this paper ,we propose ananomaly detection algorithm using weakly labeled training videos. That is we only know the video-level labels, i.e. a video is normal or contains anomaly somewhere, but we do not know where. This is intriguing because we can easily annotate a large number of videos by only assigning video-level labels. To formulate aweakly -superised learning approach ,were sort to multiple instance learning(MIL)[12,4].Specifically ,we propose to learn anomaly through a deep MIL frame work by treating normal and anomalous surveillance videos as bags and short segments/clips of each video as instances ina bag. Based on training videos, we automatically learn ananomaly ranking model that predicts high anomaly scoresfor anomalous segments in a video. During testing, a long-untrimmed video is divided into segments and fed into ourdeep network which assigns anomaly score for each video segment such that ananomaly can be detected.In summary,this paper makes the following contributions.

We propose a MIL solution to anomaly detection byleveraging only weakly labeled training videos.We pro-pose a MIL ranking loss with sparsity and smoothness con-straints for a deep learning network to learn anomaly scores for video segments. To the best of our knowledge, we are t he first to formulate the video anomaly detection problem in the context of MIL.

We introduce a large-scale video anomaly detection dataset consisting of 1900 real-world surveillance videos of 13 different anomalous events and normal activities cap-tured by surveillance cameras. It is by far the largest dataset with more than 15 times videos than

existing anomaly datasets and has a totalof128hoursofvideos.

II. RELATEDWORK

Anomaly detection .Anomaly detection is one of themost challenging and long standing problems in computer vision[39,38,7,10,5,20,43,27,26,28,42,18,26]. For

video surveillance applications, there are several attempts to detect violence or aggression [15, 25, 11, 30] in videos.Datta et al. proposed to detect human violence by exploit-ingmotion and limbs orientation ofpeople.Kooijetal.[25]employedvideoandaudiodatatodetectaggressiveactionsinsurveillancevideos.Gaoetal.proposedviolentflowdescriptorstodetectviolenceincrowdvideos.Morerecently,Mohammadi et al. [30] proposed a new behavior heuristic based approach to classify violent and non-violentvideos.

Beyond violent and non-violent patterns discrimination,authorsin[38,7]proposed to use tracking to model the nor-mal motion of people and detect deviation from that normal motion as an anomaly. Due to difficulties in obtaining re-liable tracks, several approaches avoid tracking and learnglobal motion patterns through histogram-based methods[10],topicmodeling[20],motionpatterns[31],socialforcemodels[29],mixturesofdynamictexturesmodel[27],Hid-den Markov Model (HMM) on local spatio-temporal vol-umes[26],andcontext-drivenmethod[43].Giventhetrain-ingvideosofnormalbehaviors,theseapproachesl earndis-tributions of normal motion patterns and detect low proba-ble patterns as anomalies.

Following the success of sparse representation and dic-tionary learning approaches in several computer vision problems, researchers in[28,42]used sparse representation to learn the dictionary of normal behaviors. During testing, the patterns which have large re1 construction errors are con-sidered as anomalous behaviors. Due to successful demon-stration of deep learning for image classification, several ap-proaches have been proposed for video action classification [24,36].However,obtainingannotationsfortraini ngisdif-ficultandlaborious,specificallyforvideos.

Recently,[18,39] used deep learning based autoen-coders to learn the model of normal behaviors and em-ployed reconst ruction loss to detect anomalies.Our ap-proach not only considers normal behaviors but also anoma-lousbehaviorsforanomalydetection,usingonlyweaklyla-beledtrainingdata.

Ranking .Learning to rank is an activere search are ain machine learning. These approaches mainly focused on improving relative scores of the items instead of individ-ual scores. Joachimset al. [22] presented rank-SVM to improve retrieval quality of search engines. Bergeron etal. [8] proposed an algorithm for solving multiple instance ranking problems using successive linear programming and demonstrated its application in hydrogen abstraction prob-lem in computational chemistry.Recently, deep ranking networks have been used in several computer vision appli-cations and have shown state-of-the-art performances.Theyhavebeenusedforfeaturelearning[37],highlightdetection[40], Graphics Interchange Format (GIF) generation [17],facedetectionandverification[32],personre-identification[13], place recognition [6], metric learning and image re-trieval[16]. All deep ranking methods a vast amount of annotations of positive and negative samples.

In contrast to the existing methods,we formulate anomaly detection as a regression problem in the ranking framework by utilizing normal and anomalous data. To alleviate the difficulty of obtaining precise segment-level labels relies on weakly labeled data (i.e. video-level labels – nor-mal or abnormal which are much easier to obtain than tem-poral annotations) to learn the anomaly model and

detect video segment level anomaly during testing.

1. DATASET

Due to the limitations of previous datasets ,we construct an ewlarge-scale data set to evaluate our method.It consists of long untrimmed surveillance videos which cover13real-world anomalies, including Abuse, Arrest, Arson, Assault ,Accident, Burglary, Explosion, Fighting, Robbery, Shoot-ing, Stealing, Shoplifting, and Vandalism. These anomalies are selected because they have a significant impact on pub-lic safety. We compare our dataset with previous anomalydetectiondatasetsinTable1.

Video collection.To ensure the quality of our dataset,wetraintenannotators(havingdifferentle velsofcomputervision expertise) to collect the dataset.We search videos on YouTube 1 and Live Leak 2 using text search queries(withslight variations e.g. “car crash”, “road accident”) of each anomaly.In order to retrieve as many videos as possible,wealsousetextqueriesindifferentlangu ages(e.g.French,Russian,Chinese,etc.)foreacha nomaly,thankstoGoogletranslator.Weremovevi deoswhichfallintoanyofthefol-lowing conditions: manually edited, prank videos, not cap-tured by CCTV cameras, taking from news, captured usinga hand-held camera, and containing compilation.We alsodiscard videos in which the anomaly is not clear. With the above video pruning constraints, 950 unedited real-world surveillance videos with clear anomalies are collected. Us-ing the same constraints, 950 normal videos are gathered ,leading to a total of 1900 videos in our dataset.InFig-ure2, we show four frames of an example video from each anomaly.

An notation.For our anomaly detection method, onlyvideo-level labels are required for training. However,in order to evaluate its performance on testing videos, we need to know the temporal annotations, i.e. the start and ending frames of the anomalous event in each testing anomalous video. To this end, we assign the same videos to multi-ple an not atorstolabel the temporal extent of each anomaly .The final temporal annotations are obtained by averaging an notations of different an notators.The complete data set is finalized after in tense efforts of several months. Training and testing sets.We divide our dataset into two parts:the trainingset consisting of 800 normal and 810 anomalous videos (detailsshowninTable2) and the testing set including the remaining 150 normal and 140 anomalous videos. Both training and testings ets contain all 13 anoma-lies at various temporal locations in the videos.Further-more,some of the videos have multiple anomalies .Thedis-tribution of the training videos inter ms of length (inminute)is shown in Figures 3. The number of frames and percent-age of anomaly in each testing video are presented in Fig-ures4and5,respectively.

2. EXPERIMENTS

2.1. IMPLEMENTATIONDETAILS

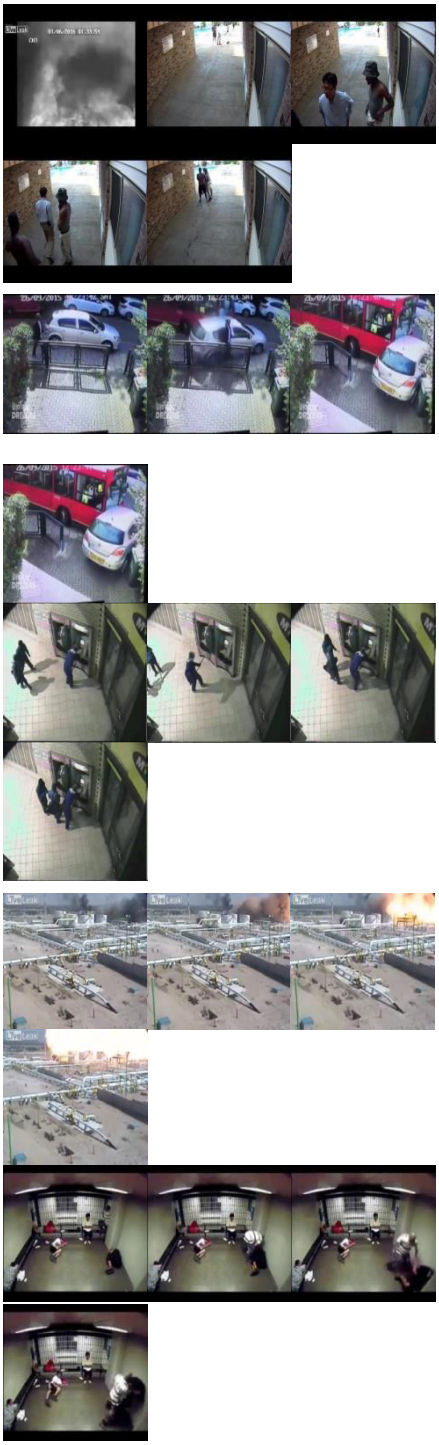
We extract visual features from the fully connected(FC)layerFC6oftheC3Dnetwork[36]. Beforecomputingfea-tures, we re-size each video frame to 240320 pixels andfix the frame rate to 30 fps. We compute C3D features forevery16-frame video clip followed by l2 normalization.To obtain features for a video segment ,we take the average of all 16-frame clip features within that segment.We in-put these features (4096D) to a 3-layer FC neural network. The first FClayer has 512 units followed by 32units and 1 unit FC layers.60% dropout regularization [33] is used between FC layers.We experiment with deeper networks butdo not observe better detection accuracy. We use ReLU[19] activation and Sigmoid activation for the first and thelast FC layers respectively, and employ Adagrad[14] opti-izer with the initial learning rate of 0.001. The parameters of sparsity and smoothness constraints in the MIL ranking

loss are set toλ1=λ2=8×10−5for the best performance. Wedivide each video into 32 non-over lapping segments and consider each video segment as an instance of the bag. The number of segments (32)is empirically set. Wealsoex-perimentedwithmulti-scale over lapping temporal segment s but it does not affect detection accuracy.We randomly se-lect30 positive and 30 negative bags asamini-batch.We compute gradients by reverse mode automatic differentia-tion on computation graph using Theano[35]. Specifically,we identify set of variables on which loss depends, compute gradient for eachvariable 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 it stem-poral segments.Then we computelossasshowninEq.6and Eq. 7 and back-propagate the loss for the whole batch.EvaluationMetric. Following previous works on anomaly detection[27],we use frame based receiver op-erating characteristic(ROC) curve and corresponding area under the curve(AUC) to evaluate the performance of our method. We do not use equal error rate(EER)[27]as it does not measure anomaly correctly, specifically if only a small portion of along video contains anomalous behavior

.

	#ofvideos	Averageframe s	Datasetlength	Exampleanomalies
UCSDPed1[27]	70	201	5min	Bikers,smallcarts,walkingacrosswalkways
UCSDPed2[27]	28	163	5min	Bikers,smallcarts,walkingacrosswalkways
SubwayEntrance[3]	1	121,749	1.5hours	Wrongdirection,Nopayment
SubwayExit[3]	1	64,901	1.5hours	Wrongdirection,Nopayment
Avenue[28]	37	839	30min	Run,throw,newobject
UMN[2]	5	1290	5min	Run
BOSS[1]	12	4052	27min	Harass,Disease,Panic
Ours	**1900**	**7247**	**128hours**	**Abuse,arrest,arson,assault,accident,burglary,fighting,robbery**

Table1. Acomparison of anomaly datasets .Our dataset contains large number of longer surveillance videos with mo rerealisti can omalies.



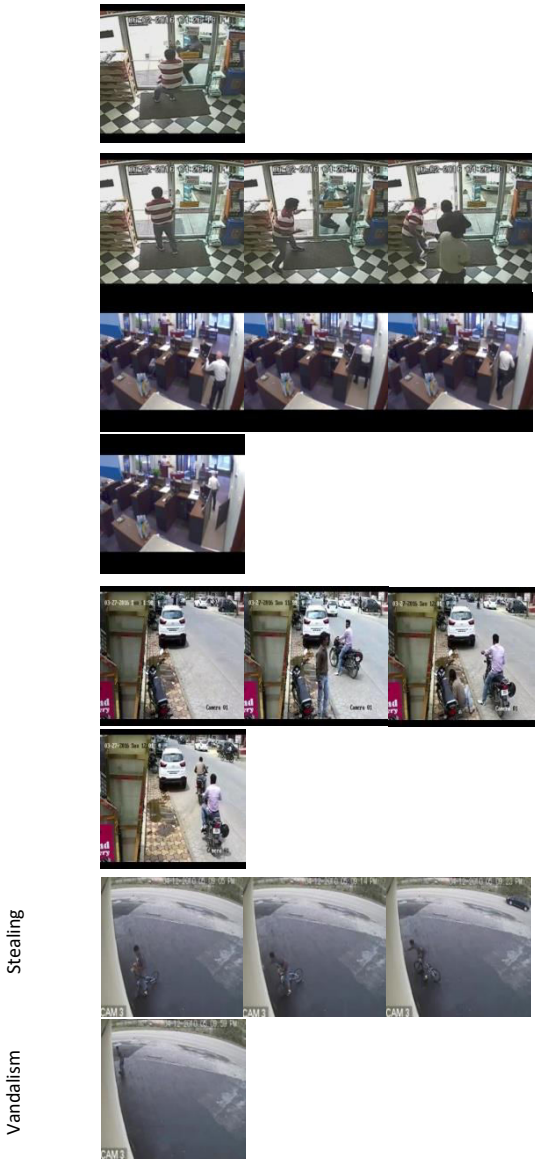


Figure2.Examplesofdifferentanomaliesfromthetrainingandtestingvideosinourdataset.

III. CONCLUSIONS

We propose a deep learning approach to detect real-world anomalies in surveillance videos.Due to the com-plexity of these realistic anomalies, using only normal dataalone may not be optimal for anomaly detection.We at-tempt to exploit both normal and anomalous surveillance videos.To avoid labor-intensive temporal annotations of anomalous segments in training videos, we learn a general model of anomaly detection using deep multiple instance ranking framework with weakly labeled data.To validate the proposed approach, a new large-scale anomaly dataset consisting of a variety of real-world anomalies is intro-duced.The experimental results on this dataset show thatour proposed anomaly detection approach performs sig-nificantly better than baseline methods.Furthermore, wed emonstrate the usefulness of our dataset for the second task of anomalous activity recognition.

IV. References

[1] http://www.multitel.be/image /research-development/research-projects/boss.php.

[2] Unusualcrowdactivitydatasetofuniversityofminnesota.In *http://mha.cs.umn.edu/movies/crowdactivity-all.avi.*

[3] A. Adam,E. Rivlin,I. Shimshoni,and D. Reinitz.Ro-bust real-time unusual event detection using multiple fixed-locationmonitors.*TPAMI*,2008.

[4] S.Andrews,I.Tsochantaridis,andT.Hofmann.Suppor tvec-tor machines for multiple-instance learning. In *NIPS*, pages577–584,Cambridge,MA,USA,2002.MITPress.

[5] B.AntiandB.Ommer.Videoparsingforabnormalitydetec-tion.In*ICCV*,2011.

[6] R.Arandjelović,P.Gronat,A.Torii,T.Pajdla,andJ.Sivic. NetVLAD: CNN architecture for weakly supervised placerecognition.In*CVPR*,2016.

[7] A.Basharat,A.Gritai,andM.Shah.Learningobject motionpatterns for anomaly detection and improved object detec-tion.In*CVPR*,2008.

[8] C. Bergeron, J. Zaretzki, C. Breneman, and K. P. Bennett.Multipleinstanceranking.In*ICML*,2008.

[9] V. Chandola, A. Banerjee, and V. Kumar.Anomaly detection:Asurvey.*ACMComput.Surv.*,2009.

[10] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas. Abnormal detectionusinginteractionenergypotentials.In*CVPR*, 2011.

[11] A. Datta, M. Shah, and N. Da Vitoria Lobo.Person-on-personviolencedetectioninvideodata.In*ICPR*,2002.

[12] T.G.Dietterich,R.H.Lathrop,andT.Lozano-Pérez.Solv- ing the multiple instance problem with axis-parallel rectangles.*ArtificialIntelligence*,89(1):31–71,1997.

[13] S.Ding,L.Lin,G.Wang,andH.Chao.Deepfea-ture learning with relative distance comparison for personre-identification.*Pattern Recognition*, 48(10):2993–3003,2015.

[14] J. Duchi, E. Hazan, and Y. Singer.Adaptive subgradientmethods for online learning and stochastic optimization.*J.Mach.Learn.Res.*,2011.

[15] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu. Violence detectionusingorientedviolentflows.*ImageandVisionCom-puting*,2016.

[16] A.Gordo,J.Almazán,J.Revaud,andD.Larlus.Deepimage retrieval: Learning global representations

for image search.In*ECCV*,2016.

[17] M. Gygli, Y. Song, and L. Cao. Video2gif: Automatic generationofanimatedgifsfromvideo.In*CVPR*,June2016.

[18] M. Hasan,J. Choi,J. Neumann,A. K. Roy-Chowdhury,and L. S. Davis.Learning temporal regularity in video sequences.In*CVPR*,June2016.