



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

∞∞∞∞

تم رفع هذه الرسالة بواسطة /صفاء محمود عبد الشافي

بقسم التوثيق الإلكتروني بمركز الشبكات وتكنولوجيا المعلومات دون

أدنى مسئولية عن محتوى هذه الرسالة.

ملاحظات: لا يوجد





**Computer Science Department
Faculty of Computer and Information Sciences
Ain Shams University**

Anomaly Detection in Crowded Scene

Thesis submitted as a partial fulfillment of the requirements for the degree of
Master of Science in Computer and Information Sciences

By

Mohamed Ali Mohamed Abd El Ghafour

Teaching Assistant at Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Supervised by

Prof. Dr. Zaki Taha Fayed

Professor in Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Maryam Nabil Al-Berry

Assistant Professor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

May – 2022
Cairo

Acknowledgment

Praise be to Allah for guiding me through my thesis path.

Thanks to supervisors, Prof. Dr. Zaki Taha Fayed for pointing me in the correct direction and Dr. Maryam Nabil Al-Berry for her never-ending knowledge and advice on how to be a better student. I'd like to express my gratitude to Dr. Ghada Hamed, TA. Nada Sherif, and TA. Dina Sherif for their assistance with the publication process.

Thanks to my mother, wife, and entire family for their unwavering support and kindness during this challenging time. Their encouragement inspired me to keep going.

Abstract

Monitoring abnormal events in a crowded scenes is essential these days, especially with the increase in surveillance cameras in most, if not all, places. This has made the field of computer vision an active field of research in recent periods. Identifying abnormal events in a crowded scene as soon as possible is essential, especially from the security side. It is a cumbersome and challenging task for humans because there are many surveillance cameras and overcrowding these days. Computer vision can be used to solve this problem and get accurate and high results. In this thesis, a convolutional autoencoder neural network architecture for anomaly detection in videos has been proposed. The proposed convolutional neural network model has been trained to recognize anomaly frames. The performance of the proposed network has been evaluated using Avenue [1] and UCSD [2] standard datasets designed to identify anomalous events in crowded scenes. Experimental results showed that the proposed model outperforms state-of-the-art methods achieving an accuracy of 71.96% and 89.52% on Avenue [1] and UCSD Peds2 [2] datasets respectively. A detailed analysis of the experiments used to choose the parameters is presented along with a comparison with other methods that used machine learning methods with an average accuracy of 69.31% and 63.90% on Avenue [1] and UCSD Peds2 [2] datasets respectively.

List of Publications

1. M. Ali, M. Al-Berry, & Z. T. Fayed (2021). Convolutional Autoencoder for Anomaly Detection in Crowded Scenes. In A. E. Hassanien, A. Haqiq, P. J. Tonellato, L. Bellatreche, S. Goundar, A. T. Azar, E. Sabir, & D. Bouzidi (Eds.), *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)* (pp. 626–633). Springer International Publishing.
https://doi.org/10.1007/978-3-030-76346-6_55
2. M. Ali, M. Al-Berry, & Z. T. Fayed (2021). Comparative Study for Anomaly Detection in Crowded Scenes. *International Journal of Intelligent Computing and Information Sciences*, 21(3), 84–94.
<https://doi.org/10.21608/ijicis.2021.84588.1112>
3. M. Ali, M. Al-Berry, & Z. T. Fayed (2021). CONVOLUTIONAL AUTOENCODER FOR ANOMALY DETECTION IN CROWDED SCENES. *Journal of Southwest Jiaotong University*, 56(6).
<https://doi.org/10.35741/issn.0258-2724.56.6.67>

Table of Contents

List of Figures	VII
List of Tables	IX
List of Abbreviations	X
Chapter 1. Introduction	2
1.1 Thesis Motivation.....	2
1.2 Thesis Objectives	3
1.3 Thesis Achievements	3
1.4 Thesis Organization	4
Chapter 2. Anomaly Detection	6
2.1 Background	6
2.1.1 Importance of Anomaly Detection Systems.....	6
2.1.2 Advantages of Anomaly Detection Systems	6
2.1.3 Anomaly Detection Systems Requirements	7
2.1.4 Anomaly Classification	7
2.1.5 Training and learning frameworks	8
2.1.5.1 Supervised video anomaly detection.....	8
2.1.5.2 Unsupervised video anomaly detection	8
2.1.5.3 Semi-supervised video anomaly detection.....	9
2.1.5.4 Active learning-based video anomaly detection	9
2.1.6 Approaches for video anomaly detection:.....	9
2.1.6.1 Accuracy-oriented approaches	10
2.1.6.2 Processing-time-oriented approaches.....	10
2.2 Literature Review	10
2.2.1 Previous Work	10
2.2.2 Datasets.....	17
2.2.3 Comparative Analysis.....	25
Chapter 3. Autoencoders.....	29
3.1 Applications of Autoencoders.....	30
3.1.1 Image Coloring	30
3.1.2 Image Enhancement	31
3.1.3 Image Compression	32
3.2 Architecture of Autoencoders	33
3.3 Properties of Autoencoders	34
3.4 Hyperparameters of Autoencoders.....	35

3.5	Types of Autoencoders	36
3.5.1	Convolutional Autoencoders (CAE)	36
3.5.2	Sparse Autoencoders	37
3.5.3	Deep Autoencoders.....	38
3.5.4	Contractive Autoencoders	40
Chapter 4.	Proposed Method	42
	Phases of Proposed Method	42
4.1	First Phase (CAE with Modified Preprocessing).....	42
4.2	Second Phase (Proposed CAE)	46
Chapter 5.	Experimental Results	52
5.1	Dataset.....	52
5.2	Experimental Environment	53
5.3	Proposed Anomaly Detection Method Experimental Results .	54
5.3.1	First Phase (CAE with Modified Preprocessing)	54
5.3.2	Second Phase (Proposed CAE)	56
Chapter 6.	Conclusion and Future Work	62
6.1	Conclusions	62
6.2	Future Work	63
References.	64

List of Figures

Figure 1: UMN 2006 Example	20
Figure 2: PETS 2009 Example	21
Figure 3: QMUL 2009 Example.....	21
Figure 4: MIT Traffic 2009 Example	21
Figure 5: Anomalous Behavior/ York 2010 Example.....	22
Figure 6: BEHAVE 2010 Example	22
Figure 7: Violent flows 2012 Example.....	23
Figure 8: UCSD 2013 Example.....	23
Figure 9: Avenue 2014 Example	24
Figure 10: BOSS 2017 Example, (a) fighting; (b) fainting; (c) grabbing cell phone.....	24
Figure 11: Crowd Example 2018.....	25
Figure 12: Autoencoder Neural Network	29
Figure 13: Linear vs nonlinear dimensionality reduction	30
Figure 14: Image Coloring (a) Gray Image. (b) Autoencoder output. (c) Original Image	30
Figure 15: Feature Variation.....	31
Figure 16: Denoising Image	31
Figure 17: Watermark Removal	32
Figure 18: Image Compression.....	32
Figure 19: Architecture of Autoencoders	33
Figure 20: Bottleneck	34
Figure 21: Hyperparameters of Autoencoders	36
Figure 22: Convolutional Autoencoder	37
Figure 23: Sparse Autoencoder	38

Figure 24: Deep Autoencoder.....	39
Figure 25: Contractive Autoencoder	40
Figure 26: Sample results of the steps of background removal (a) The background, (b) The current frame, (c) The binarized image of $D1$, (d) The binarized image of $D2$, (e) The union of c & d, (f) The current frame after background removal.	44
Figure 27: The proposed network architecture 1 (CAE with Modified Preprocessing).....	45
Figure 28: The proposed network architecture 2 (Proposed CAE)	48
Figure 29: Representation of anomaly events in the Avenue dataset. (a) A strange action (running). (b) Walking in the wrong direction. (c) Abnormal object (bike)	52
Figure 30: Representation of anomaly events in the Peds1 (a) and the Peds2 (b) datasets. (a): A small vehicle. (b) A biker and a skater	53
Figure 31: UCSD dataset: Comparison (AUC) for (CAE with Modified Preprocessing) with the state-of-the-art models	55
Figure 32: AUC for (CAE with Modified Preprocessing) for Peds1 (a) and Peds2 (b) datasets.....	56
Figure 33: Avenue & UCSD datasets: Comparison (AUC) for (Proposed CAE) with the state-of-the-art models.....	58
Figure 34: Avenue & UCSD datasets: Comparison (AUC) for (Proposed CAE) with a volume of input frames.....	59
Figure 35: UCSD Peds2 dataset: Comparison (AUC) with a number of layers	60

List of Tables

Table 1: Comparison between the public datasets containing anomaly events in crowded scenes	17
Table 2: A brief survey on the critical studies for detecting anomalies in crowded scenes.	25
Table 3: Output Dimensions for the proposed network architecture 1 (CAE with Modified Preprocessing).....	45
Table 4: Output Dimensions for the proposed network architecture 2 (Proposed CAE).....	48
Table 5: UCSD dataset: Comparison (AUC) for (CAE with Modified Preprocessing) with the state-of-the-art models	55
Table 6: Avenue & UCSD datasets: Comparison (AUC) for (Proposed CAE) with the state-of-the-art models.....	57
Table 7: Avenue & UCSD datasets: Comparison (AUC) for (Proposed CAE) with a volume of input frames.....	58
Table 8: UCSD Peds2 dataset: Comparison (AUC) with a number of layers and total parameters	60

List of Abbreviations

<u>Abbreviation</u>	<u>Stands for</u>
AUC	Area Under the Curve
BiGRU	Bi-directional Gated Recurrent Unit
BSO	Bird Swarm Optimization
CAE	Convolutional Autoencoder
CBE	Crowds By Examples
CCTVs	Closed-Circuit Televisions
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
EMD	Earth Mover Distance
FPR	False Positive Rate
GMM	Gaussian Mixture Model
IR	Information Retrieval
MAP	Maximum A Posteriori
MBSO	Monarch Bird Swarm Optimization
MHOF	Multi Histogram Optical Flow
MRF	Markov Random Field
MSE	Mean Squared Error
OF	Optical Flow
OSVM	Optimal Support Vector Machine
PCA	Principal Component Analysis
PCANet	Principal Component Analysis Network
PPN	Position Projection Network
QMUL	Queen Mary University of London
ROC	Receiver Operator Characteristic
SI	Saliency Information
STIPs	Spatio-Temporal Interest Points
SVM	Support Vector Machines
TM	Trademark
TPR	True Positive Rate
UCSD	University of California, San Diego
UMN	University of Minnesota

Chapter 1

Introduction

Chapter 1. Introduction

1.1 Thesis Motivation

Now, Closed-Circuit Televisions (CCTVs) are everywhere, especially in crowded places. So, it is essential to use machine learning to help in detecting anomalies in those places. Without the help of artificial intelligence, this task will require a lot of human efforts to monitor many screens and discover these cases, knowing that they are rare.

There is no fixed definition of anomalous events because the event varies according to the event's location. For example, running in parks is normal, but running in a restaurant is strange. The event, which is traction, changes its evaluation according to the place in which it occurred. Those differences add more challenges for machine learning to discover these events in applications in the daily life. Other challenges include environmental conditions (presence of shadow, different lighting, the appearance of obstructions, background problems, etc.), crowd density, and the complex nature of human behavior. These things add further challenges to machine learning to discover anomalies.

In literature, some recent researches deal with the problem of discovering anomalies as a binary classification problem (normal and anomaly), and some other works focus on classifying each event separately, as each moving object is tracked and the behavior is studied to determine if it is normal or not.

Fixed surveillance cameras can be used, such as surveillance cameras in the streets or front of stores, or mobile surveillance cameras such as front car cameras or police cameras. However, to the best of our knowledge, there is little research devoted to detecting video anomalies using mobile cameras, so this research will focus on machine learning methods to detect anomalies using fixed cameras.

1.2 Thesis Objectives

This research aims to propose a fast and accurate method that can detect anomaly events in crowded scenes without any human interaction. The objectives of this research can be summarized in the following points:

1. Studying different anomaly detection methods.
2. Analyzing the already existing related studies to get the crucial problems to be considered.
3. Implementing relevant techniques and comparing the performance of implemented techniques using a benchmark dataset to choose an appropriate technique.
4. Enhancing the chosen technique and analyzing the performance after enhancement.
5. Comparing the proposed method against similar studies in the literature regarding the overall and average accuracy and reliability.

1.3 Thesis Achievements

The contributions of the work can be summarized as follows:

1. Providing a comprehensive survey on the existing anomaly detection techniques in crowded scenes and introducing a

comparative analysis for the crucial issues to be considered in this thesis.

2. Modifying existing preprocessing steps by using a better way for background removal.
3. Modifying an existing method by using better models for autoencoders.
4. Providing evidence on the robustness of the proposed method by evaluating it using three datasets.

1.4 Thesis Organization

This thesis is organized into six chapters, including this one. Their contents are described briefly as follows:

Chapter 2 (Titled: Anomaly Detection) is divided into two sections. Section one is info about Anomaly Detection Systems, training and learning frameworks and approaches for video anomaly detection. The second section is the literature review of some previous work. It presents a survey on different datasets and a comparative analysis.

Chapter 3 (Titled: Autoencoder) provides the explanation and details for machine learning method “Autoencoder”.

Chapter 4 (Titled: Proposed Method) describes the algorithms used for the proposed anomaly detection in crowded scenes in detail.

Chapter 5 (Titled: Results and Discussion) provides the used datasets and the experimental results for the proposed anomaly detection method in detail.

Chapter 6 (Titled: Conclusions and Future Work) presents the conclusions and future work.