Ministry of Higher Education and scientific Research
University of Diyala
College of Science
Department of Computer Science

Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System

A Thesis
Submitted to the Department of Computer Science\College of Science\ University of Diyala in a Partial Fulfilment of the Requirements for the Degree of Master in Computer Science

By

Mohamed Safaa Mohamed Shubber

Supervised by

Prof. Dr. Ziyad Tariq Al-Ta'i

2022 A.C. 1444 A.H

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

﴿وَلَقَدْ آتَيْنَا دَاوُودَ وَسُلَيْمَانَ عِلْمًا ۖ وَقَالَا الْحَمْدُ لِلَّهِ الَّذِي فَضَّلَنَا عَلَىٰ كَثِيرٍ مِّنْ عِبَادِهِ الْمُؤْمِنِينَ﴾

الآية 15 من سورة النمل

صَدَقَ ٱللَّهُ ٱلعَظِيمُ

الإهداء

أهدي جهدي المتواضع هذا

الى قدوتي، ومثلي الأعلى............... أبي الحبيب رحمك الله

الى منبع الايثار .... أمي الغالية رحمكِ الله

الى من دعمتني وشاركتني في السراء و الضراء....... رفيقة دربي زوجتي

الى من هم سندي وعضدي.......... اخوتي

الى من أنار لي طريق العلم والمعرفة..................... اساتذتي

محمد

Acknowledgments

Firstly, all my prayers go to (Allah), the Almighty, for the successive blessings, divine providence, and my success in this research.

I would like to extend my sincere thanks and gratitude to my supervisor **Prof. Dr. Ziyad Tariq Mustafa** for his valuable guidance, constructive suggestions, and assistance during the writing of this research.

Thanks, are also due to all staff members at computer science department in the University of Diyala for their encouragement and support.

Special thanks to my family for their patience and support that gave me the strength to accomplish this study.

Finally, I would like to thank all my friends for their support and willingness to listen and comment on different aspects of this research.

Mohamed

Linguistic Certification

This is to certify that this thesis entitled ***"Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System"*** *was prepared under my linguistic supervision. It was amended to meet the style of English language.*

Signature:

Name: **Assist. Prof. Dr. Fatima Mohamed Abood**

Date: / / 2022

Scientific Certification

I certify that the thesis entitled ***"Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System"*** *has been evaluated scientifically; therefore, it is suitable for debate by the examining committee.*

Signature:

Name: **Assist. Prof. Dr. Jamal Fadhil Tawfeeq**

Date: / / 2022

Scientific Certification

I certify that the thesis entitled ***"Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System"*** *has been evaluated scientifically; therefore, it is suitable for debate by the examining committee.*

Signature:

Name: **Prof. Dr. Saraab Majid Hamid**

Date: / / 2022

Supervisor's Certification

*I certify that the thesis entitled **"Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System"** was prepared by "**Mohamed Safaa Mohamed Shubber**" under my supervision at University of Diyala – Collage of Science – Department of Computer Science, as a partial fulfilment of the requirements for the degree of **master in computer science***

(Supervisor)

Signature:

Name: **Prof. Dr. Ziyad Tariq Mustafa Al-Ta'i**

Date: / / 2022

Approved by University of a Diyala Faculty of Science Department of Computer Science.

Signature:

Name: **Assist. Prof. Dr. Bashar Talib Hameed**

Date: / / 2022

(Head of Computer Science Department)

Examination Committee Certification

We certify that we have read the thesis entitled ***"Develop of a Hybrid Deep Learning Model for Abnormal Human Behavioral Detection System"*** *and as examination committee, examined the student* **"*Mohamed Safaa Mohamed Shubber"*** *in the thesis content and that in our opinion, it is adequate fulfill the requirement for the Degree of Master in Computer Science at the Computer Science Department, University of Diyala.*

(Chairman)

Signature:

Name: **Prof. Dr. Dhahir Abdulhade Abdulah**

Date: / / 2022

(Member)

Signature:

Name: **Assist. Prof. Dr. Adil Abdulwahhab Al-Azzawi**

Date: / / 2022

(Member)

Signature:

Name: **Assist. Prof. Dr. Mohammed S. H. Al-Tamimi**.

Date: / / 2022

(Member/Supervisor)

Signature:

Name: **Prof. Dr. Ziyad Tariq Al-Ta'i**

Date: / / 2022

Approved by the Dean of College of Science, University of Diyala.

(The Dean)

Signature:

Name: **Prof. Dr. Tahseen Hussein Mubarak**

Date: / / 2022

Abstract

During the recent years, creating automated systems has increased dramatically, especially the systems that are capable of keeping a track on people in public environments and identify abnormal behaviors, such as violent and suspicious incidents. The goal of violence detection is to determine whether an act of violence has taken a place. In this thesis, two proposed models for the detection of violence are proposed. First one is a pre-trained and modified deep neural network called CNN-VGG16 in which transfer learning technique is applied to take advantage of the pre knowledge of VGG16 in detecting shapes and edges. The final layers of the default VGG16 structure are modified to detect the violence. The second model is a hybrid structure consists of VGG16 and SVM, in which the VGG16 acts as a feature extractor and the SVM is the classifier. The efficiency of two the approaches is evaluated using two datasets (Automatic Violence Detection Dataset(AvdDS) and Surveillance fight dataset(SfDS)). In this thesis, the effect of applying edge detection is observed where it is noted that the accuracy is decreased slightly after applying Canny edge detector. The results show that the first proposed model has achieved an accuracy on dataset one (~95%), Precision (~95%), Recall (~95%) and F1-score (~95%), but yielded an accuracy of (91%) , Precision (~91%), Recall (~91%) and F1-score (~91%) after applying Canny filter. First model also attains an accuracy for dataset two about (~99%), Precision (~99%), Recall (~99%) and F1-score (~99%) , but (~92%), Precision (~93%), Recall (~92%) and F1-score (~92%) using Canny filter. On the other hand, the results of the second model , for dataset one is (~99%), Precision (~99%), Recall (~99%) and F1-score (~99%) , but (~95%), Precision (~95%), Recall (~95%) and F1-score (~95%) after applying Canny filter. While second model obtains for dataset two, an accuracy about (100%), Precision (100%), Recall (100%) and F1-score (100%), with and without appling the Canny filter. A violence detection is an important field because it represents the difference between death and life. However, deep learning and machine learning can be used successfully

in building violence detection models. VGG16 network is an excellent feature extractor in the field of action recognition which was used in both proposed approaches. In spite of both proposed models scored excellent classification results, SVM classifier in the second model achieved higher accuracy than the top layers of the modified VGG16 network in the first proposed model. Applying Canny edge decreased the classification accuracy, because it is eliminate large number of features. In general, accuracy results obtained from dataset 2 were higher than the accuracy results of dataset 1, because of the indoor, homogeneous, and high resolution nature of the videos in dataset 2 .

List of Contents

List of Figures

List of Tables

List of Algorithms

LIST OF ABBREVIATIONS

Abbreviations	Meaning
AdaGrad	Adaptive Gradient Algorithm
Adam	Adaptive Moment Estimation
AHE	Advanced Histogram Equalization
AI	Artificial Intelligence
ANN	Artificial Neural Network
AvdDS	Automatic Violence Detection Dataset
BiLSTM	Bidirectional Long Short Term Memory
CLAHE	Contrast Limited Advanced Histogram Equalization
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short Term Memory
CV	Computer Vision
DCNN	Deep Convolutional Neural Networks
DL	Deep Learning
FN	False Negative
FP	False Positive
FPS	Frame Per Second
GC	Google Colab
GRU	Gated Recurrent Unit
HAR	Human Action Recognition
HE	Histogram Equalization
HOG	Histogram Of Oriented Gradients
KNN	K-Nearest Neighbor
LSTM	Long Short Term Memory
MN	Max Normalization
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RGB	Red-Green-Blue
RNN	Recurrent Neural Network
SfDS	Surveillance Fight Dataset
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VGG	Visual Geometry Group

LIST OF SYMBOLS

Symbol	Meaning
$f(x)$	function, where x is the input
e	Exponential
Avg	Average
R, G, B	Red, Green, Blue
$p(x)$	Probability function, where x is the input
$\| \|$	Absolute
$\|\vec{G}\|$	Magnitude
θ	Theta
$\boldsymbol{\Sigma}$	Summation
sgn	Sign function
arc	smooth curve joining two endpoints
log	Logarithm

Chapter One

Introduction

Chapter One

Introduction

1.1 Motivation

Most of the problems that society suffers from are related to abnormal human behaviors in one way or another. Although it might be challenging to outline anomality in human behavior, it is generally simple to spot when it occurs. [1].

Violence is considered one of the most dangerous violations of the normal social habits and it may always be part of the human experience from the beginning of existence and its impact can be seen in various forms worldwide [2].

Since video surveillance equipment are usually used in public areas like banks, colleges, and train stations to monitor and control the human activity, the demands raised for abnormal human behavior detection system that automatically detects violent and abnormal incidents [3].

Due to human exhaustion and inattention, it is possible that harmful events like fights and aggressive actions won't be detected by security staff. Therefore, developing an intelligent video surveillance system that automatically identify abnormalities are crucial. Given the significance of security , research has been done in this area and several methods to identify anomalies in videos have been presented [4].

There are several applications in the field of “computer vision” (CV) field have recently seen a massive update, like activity recognition , image classification and labelling, etc. The development and introduction of a new technique for machine learning known as deep learning is considered a breakthrough and been seen in many computer vision areas. Since 2010, many researchers from computer vision area have moved from conventional handcrafted features descriptor to the learned-based features descriptor, often referred to as data-driven algorithms [5].

Human action recognition (HAR) is a crucial yet difficult aspect of analyzing human behavior. This issue involves the monitoring of differences in human movement and the recognition of activity by using machine learning algorithms [6]. A few difficulties arise in automatic detection of violence or in general aggressive behavior due to its subjective nature which imposes some obstacles in outlining what should be considered as violence. Besides that, some of human behaviors, might be misclassified which appear very similar to aggressive actions. [7]

This is where this thesis comes in which proposes two models to detect violent activity automatically. The results show that the first proposed model has achieved an accuracy on dataset one (~95%), Precision (~95%), Recall (~95%) and F1-score (~95%), but (91%) , Precision (~91%), Recall (~91%) and F1-score (~91%) after applying Canny filter. First model also attains an accuracy for dataset two about (~99%), Precision (~99%), Recall (~99%) and F1-score (~99%) , but (~92%), Precision (~93%), Recall (~92%) and F1-score (~92%) using Canny filter. On the other hand, the results of the second model , for dataset one is (~99%), Precision (~99%), Recall (~99%) and F1-score (~99%) , but (~95%), Precision (~95%), Recall (~95%) and F1-score (~95%) after applying Canny filter. While second model obtains for dataset two, an accuracy about (100%), Precision (100%), Recall (100%) and F1-score (100%), with and without applying the Canny filter.

1.2 Related Work

Different researches in the major of violence detection are developed. The present survey includes previous work related to this thesis:

- **A.S. Keçeli , et. al. in 2017** [8] presented a model that is using optical flow and a technique based on transfer learning for detecting violence. Firstly, they develop a 2D templates using optical flow patterns, and the value of the magnitudes and the velocities . Those instance are presented as input to a CNN that has been previously trained, and deep features are retrieved. Finally, To predict violent actions, two forms of classifiers are developed using these attributes. Results demonstrate that the high-level characteristics are quite good at detecting violence. The proposed methodology was evaluated using three distinct datasets and the accuracy results were as follow HocDS 94.40% ,MovDS 96.50% and VifDS 80.90%.
- **A. Hanson, et. al. in 2019** [9], presented a reduced spatial encoder and a spatio-temporal encoder architecture for supervised violence detection. They made a contribution by creating and encodings that is a bidirectional temporal , which were then max-pooled elementwise to improve representations which is based on the context. As a result, their Bidirectional ConvLSTM behaved more accurately than the ConvLSTM model for more complicated and mixed datasets like the VifDS dataset. Their accuracy results were 96.96% on HocDS and 100% on MovDS and on VifDS was 92.18%.
- **F. U. M. Ullah et.al. in 2021** [10], presented a reliable violence detection approach which can detect violent events in surveillance footages. The main parts of the proposed method were preprocessing, feature extraction, and learning the action sequence. Initially, a CNN was utilized during the phase extracting the feature to collect the data and attributes. Afterwards, the resulting feature map was constructed by concatenating remaining optical flow CNN features with high-level features from Darknet19 model. In the last stage, LSTM network acquired

and learnt the sequence characteristics for violence detection. upon final layer of LSTM, the outcome (violent or not) is determined. On the VifDS, HocDS, and SfDS datasets, the suggested technique yielded accuracy of 98.21%, 98.8%, and 74%, respectively.

- **M. S. Kang, et. al. in 2021** [11], proposed a novel violence detection pipeline that can be combined with the conventional 2-dimensional Convolutional Neural Networks. On top of that, they presented temporal along with spatial attentions modules which are low complexity but persistently boost the efficiency of violence behavior recognition. They achieved an accuracy of 99.6%, 100%, ,98.0% , 92.0% , 97.8% and 92%respectively on HocDS, MovDS, VifDS, SfDS , RlvDS and RwfDS datasets.
- **M. Haque, et. al. in 2022** [12], Using the Gated Recurrent Unit (GRU),they designed a novel Deep Convolutional Neural Network (DCNN) called "BrutNet", The network is intended to function based on the patterns that are present in a video over several frames. For every frame in the timely-distributed layer, convolutional layers were used to obtain the features and pattern of the image. In order to acquire a collection of 512 features for each frame, the model converts the data from 4D to 2D and encodes it. The GRU layer then extracts the temporal character of these frames as a 1-dimensional vector, and then processed by many fully-connected(dense) layers. As a result, a classification is carried out, with the results identifying the content as either violent or non-violent. On AvdDS, the model achieved an accuracy rating of 90% during the testing process.

1.3 Problem Statement

Abnormal human behavior that comes in the shape of violence in public or private areas is a big problem, since it deals with human safety. Normally, this violence is observed and recorded by surveillance cameras and analysis of these footages for the purpose of detecting these activities. This analysis and the decision for the specified footage if it is violent or nonviolent activities are the problem of this thesis.

1.4 Aim of the Thesis

This thesis aims to design two models for detecting violence behavior in surveillance videos from different datasets. One model based on deep learning methodology; the other model depends on machine learning techniques. Also, this work aims to compare between the two proposed

1.5 Outline of Thesis

The other chapters included in this thesis are outlined as follow:

Chapter Two: Theoretical Background

This chapter gives the background of violence detection models.

Chapter Three: **The Proposed Models**

This chapter presents in detail the proposed models which are used to detect violence.

Chapter Four: Results

This chapter explains the results that are obtained from implementing the proposed models.

Chapter Five: Conclusions and Suggestions for Future work

This chapter presents the conclusions of this work. Furthermore, it provides suggestions for future work.