DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

Abnormal Behavior Detection in Video Surveillance Using Inception-v3 Transfer Learning Approaches

Sabah Abdulazeez Jebur[1], Khalid A. Hussein[2], Haider Kadhim Hoomod[3]
[1]*Department of Computer Sciences, University of Technology, Baghdad, Iraq*
[1]*Department of Computer Techniques Engineering, Imam Al-Kadhum College (IKC), Baghdad, Iraq*
[2,3]*Department of Computer Science, College of Education, Mustansiriyah University, Baghdad, Iraq*
[1]*sabah.abdulazeez@alkadhum-col.edu.iq,* [2]*dr.khalid.ali68@gmail.com,* [3]*drhjnew@gmail.com*

Abstract— *The use of video surveillance systems has increased due to security concerns and their relatively low cost. Researchers are working to create intelligent Closed Circuit Television (CCTV) cameras that can automatically analyze behavior in real-time to detect anomalous behaviors and prevent dangerous accidents. Deep Learning (DL) approaches, particularly Convolutional Neural Networks (CNNs), have shown outstanding results in video analysis and anomaly detection. This research paper focused on using Inception-v3 transfer learning approaches to improve the accuracy and efficiency of abnormal behavior detection in video surveillance. The Inception-v3 network is used to classify keyframes of a video as normal or abnormal behaviors by utilizing both pre-training and fine-tuning transfer learning approaches to extract features from the input data and develop a new classifier. The UCF-Crime dataset is used to train and evaluate the proposed models. The performance of both models was evaluated using accuracy, recall, precision, and F1 score. The fine-tuned model achieved 88.0%, 89.24%, 85.83%, and 87.50% for these measures, respectively. In contrast, the pre-trained model obtained 86.2%, 86.43%, 84.62%, and 85.52%, respectively. These results demonstrate that transfer learning using Inception-v3 architecture can effectively classify normal and abnormal behaviors in videos, and fine-tuning the weights of the layers can further improve the model's performance*.

Index Terms— *Abnormal behavior detection, Video surveillance, Deep learning, Transfer learning, InceptionV3.*

I. INTRODUCTION

The use of video surveillance systems, such as CCTV cameras, has become increasingly common in recent years due to security concerns and their relatively low cost. These cameras are utilized for monitoring indoor and outdoor environments in order to detect and analyze events. In the past, frames from these cameras were manually stored and reviewed to investigate any abnormal events that occurred [1]. However, this process was costly and prone to human error, such as fatigue and distraction. Nowadays, researchers are working on creating intelligent CCTV cameras that can automatically analyze behavior in real-time. This is an active area of research in machine vision and artificial intelligence, focusing on designing real-time systems that can detect abnormal behaviors and prevent dangerous accidents [2]. Abnormal behavior can be described as actions that deviate from what is typical in a given context. In computer vision, abnormality refers to data patterns that differ from the norm. Abnormal activity detection has been applied in various areas of video surveillance[3]. Computer vision and machine learning methods are utilized to detect and identify human activities. The use of diverse video characteristics enables an understanding of human behavior and actions, which, in turn,

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

aids in categorizing activities as either normal or abnormal [4]. However, the recognition of human activity poses challenges due to the complexity of human behavior, making it difficult to differentiate between normal and abnormal activities [5],[6]. Deep Learning (DL) approaches have demonstrated remarkable outcomes in multiple time series analyses and computer vision tasks. These methods are effective for examining video sources, making them a popular trend for anomaly detection. DL techniques have produced state-of-the-art results for problems related to anomaly detection, as well as excelling in other domains. In addition, deep learning techniques have higher accuracy than traditional methods for detecting anomalies in crowds [7],[8]. Convolutional Neural Networks (CNNs) are a type of deep learning neural network that are specifically designed for image and video recognition, as well as other types of visual data processing [9]. These networks consist of multiple layers of artificial neurons, each of which applies a set of filters to an input image, allowing the network to learn increasingly complex features of the image [10]. CNNs are particularly useful for tasks such as image classification, object detection, semantic segmentation, and many others[11]. In recent years, various CNN architectures, such as VGG16 [12], VGG19 [13],[14], Inception-v3 [15],[16], and ResNet-50 [17], have been proposed to address the problem of abnormal behavior detection in video. CNN models have been shown to be effective in detecting abnormal activities in video surveillance, such as running, climbing, falling, fighting, robbing, violence, loitering, vandalism, personal intrusion, autism, drug addiction, and reckless driving [18]. In this research, the focus was on utilizing transfer learning to develop an efficient system for detecting abnormal behavior in video surveillance. To accomplish this, the Inception-v3 network was applied in order to improve the performance of detecting abnormal human behavior in video surveillance footage. The findings of this research can be used to improve the accuracy and efficiency of video surveillance systems for detecting abnormal behavior in real-world scenarios. This paper provides the following main contribution:

- Examining the Inception-v3 advanced deep learning model for predicting abnormal behavior in surveillance cameras.
- Investigating and analyzing the results of two transfer learning approaches - pre-training and fine-tuning - using the Inception-v3 network.
- Employing several methods to ensure that the data can be readily utilized by the model.

II. LITERATURE REVIEW

This section provides an overview of various studies on identifying abnormal human behavior in smart surveillance systems. In a study [19], a novel technique for identifying abnormal human behavior in smart surveillance systems. The method involves analyzing changes in static postures over time, using a deep learning model called YOLO (You Only Look Once) to detect and distinguish individuals in real-time video. The images of the identified individuals are then passed through another model called VGG-16 to classify six types of abnormal behaviors. The output is then analyzed using an LSTM (Long-Short Term Memory) model to detect abnormal behaviors as they happen. The study shows that this approach has potential and is feasible. In [20], a pre-trained CNN model called VGGNet-19 was utilized to extract features which were then inputted into a Binary Support Vector Machine classifier (BSVM) to create a binary-SVM model for abnormal behavior detection. The model's performance was tested on two datasets, UMN and UCSD-PED1, achieving high accuracy and AUC. The study also compared the performance of VGGNet-19 with other methods and pre-trained CNNs, finding that it had better performance than alternative methods and other pre-trained models like GoogleNet, ResNet50, AlexNet, and VGGNet-16. In [21], different models were used to detect and predict abnormal behavior, including LSTM, Convolutional Neural Network (CNN), and LSTM_CNN. The results showed that the LSTM_CNN combination was the most effective model, achieving high accuracy, precision, and recall. The study concluded that both temporal and spatial features are important in detecting abnormal

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

behavior, and addressed the issue of imbalanced data using the Synthetic Minority Over-sampling Technique (SMOTE). [22] presented a technique for identifying distracted driving using a hybrid CNN framework (HCF). The system employed pre-trained models such as ResNet50, Inception V3, and Xception to extract features of driver behavior via transfer learning. The resulting features are then combined to provide a comprehensive understanding and the HCF's fully connected layers are trained to eliminate anomalies and hand movements related to non-distracted driving. An enhanced dropout algorithm is also applied to avoid overfitting. The method's performance is evaluated using class activation mapping (CAM) and the results indicate that the HCF achieved a high accuracy of 96.74% in detecting distracted driving behaviors. The study [23] proposed a framework for identifying pushing behavior in crowded event entrances by utilizing top-view video recordings. The proposed method combines a deep learning model based on EfficientNet-B0, RAFT (Region-based Active Flow Transforms), and visualization techniques to extract crowd motion information and generate patches that represent pushing behavior. Additionally, the framework incorporates a patch-based approach to enhance the accuracy of the classifier and tackle the class imbalance issue. The framework demonstrated an accuracy of 88% on undistorted videos and 86% on distorted videos. [4] presented a deep learning system to identify abnormal behavior in drivers, using a new dataset that includes categories such as smoking, eating, drinking, calling, and normal behavior. The study employs pre-trained and fine-tuned CNN models, such as ResNet101, VGG-16, VGG-19, and a proposed CNN model, to analyze the results. The performance of the models was evaluated using accuracy rates, with the pre-trained models achieving 89%, 93%, 93%, and 94% respectively, and the proposed CNN model achieving an accuracy of 95%. [7] introduced a method for identifying anomalies in video frames through the use of a deep learning model called AlexNet. The effectiveness of the model was assessed through two metrics, AUC (Area Under the Curve) using ROC (Receiver Operating Characteristic) curves and overall accuracy. When tested on three benchmark datasets of video frames containing both normal and abnormal actions, the model outperformed other existing methods, achieving an AUC of 98%. Furthermore, the study also compared the performance of the AlexNet model to six other machine learning algorithms, with the softmax classifier achieving the highest accuracy of 99% on all three datasets. The proposed method also achieved the highest AUC values compared to other existing methods, reaching a maximum of 98%. [24] introduced a CNN-based model for detecting abnormal behavior in videos. The model includes two convolutional layers, and two fully connected layers, and uses Relu and sigmoid as activation functions. The model was evaluated using the ROC curve, AUC, and EER (Equal Error Rate) metrics and was found to outperform other similar methods. The model also performed well when tested on random YouTube videos of abnormal behavior. [1] proposed a deep learning approach for detecting violent content in videos. The method uses custom-designed features like visual appearance, movement speed, and key frames and inputs them into a CNN with separate streams for spatial, temporal, and combined spatial-temporal information. The spatial stream analyzes individual frames, the temporal stream examines motion patterns related to violence, and the spatiotemporal stream uses a unique representation to recognize violent actions. The CNN was trained using labeled video frames from three datasets: Hockey, Movie, and ViF. [3] proposed a new architecture called 2s-MDCN (Two streams Multi-dimensional Convolutional Network) for violence detection in videos, which utilizes both RGB frames and optical flow. The proposed method extracts both temporal and spatial information separately by using 1D, 2D, and 3D convolutions. The models are lightweight and efficient due to reduced channel capacity, but still able to extract significant spatial and temporal information. Additionally, using both RGB frames and optical flow leads to 2.2% more accuracy than using a single RGB stream.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

III. THEORETICAL BACHGROUND

A. Transfer Learning

Transfer learning is a method utilized in machine learning where a previously trained model for one particular task is utilized as a starting point for a new model designed for a second, related task. The idea is that the model has already learned features from the first task that can be useful in the second task, allowing the model to learn more quickly and effectively [25]. Transfer learning can be applied in various ways, such as using a pre-trained model as a fixed feature extractor, fine-tuning the pre-trained model, or using the pre-trained weights as initialization for training a new model [20]. A common use of transfer learning is using a pre-trained model as a feature extractor. In this method, the pre-trained model is used to extract features from the input data, and a new classifier is trained on top of these features. This allows the new classifier to learn from the already learned features, rather than starting from scratch. Fine-tuning a pre-trained model is another way to apply transfer learning. This method involves training a pre-trained model on a new dataset, to adapt the model to the new task. This is usually done by unfreezing some of the layers in the pre-trained model and retraining them with the new dataset. Another way to apply transfer learning is to use the pre-trained weights as initialization for training a new model. This method can be useful when the new task is not similar enough to the original task to use a pre-trained model directly, but the pre-trained weights can still provide a good starting point for the new model. Transfer learning is a useful method for reducing the amount of computational resources required and enhancing the performance of models when there is limited data available for training [26].

B. Inception-v3 Architecture

Inception-v3 is a deep CNN architecture designed for image classification tasks. It was introduced in the paper [27]. The Inception-v3 architecture is made up of several building blocks, including the stem, the Inception blocks, and the final layers. The stem is used to reduce the spatial resolution of the input, typically by applying a few convolutional layers. This is done to reduce the computational cost of the later layers. The Inception blocks are the core of the architecture and are used to increase the depth of the network. Each Inception block is made up of a combination of convolutional layers and Inception modules. The Inception modules are used to learn features at multiple scales, while the convolutional layers are used to increase the depth of the network. The final layers of the Inception-v3 architecture are used to reduce the spatial resolution of the feature maps and to produce the final output. These layers typically include a few convolutional layers and a global average pooling layer. The output of the final layers is then passed through a fully connected layer to produce the final classification. *Fig. 1* shows the architecture of the Inception-v3 model.

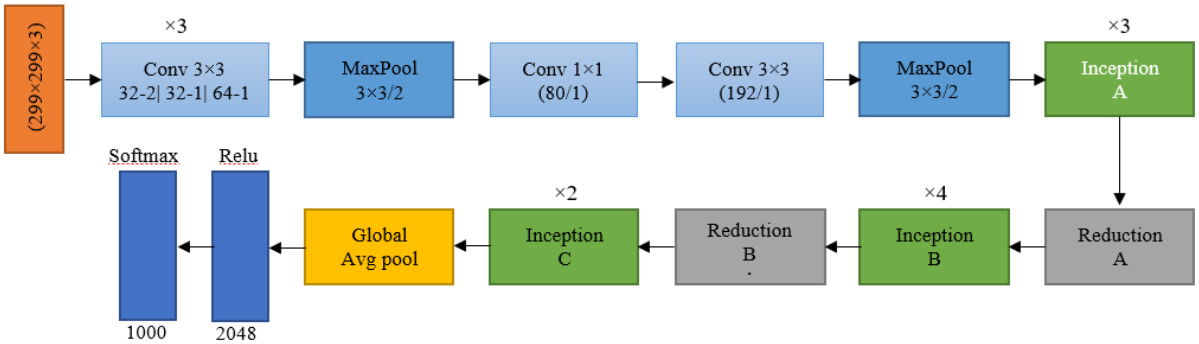

FIG. 1. THE ARCHITECTURE OF INCEPTION-V3.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

IV. PROPOSED METHODOLOGY

This research employed transfer learning on the Inception-v3 architecture, utilizing both pre-training and fine-tuning approaches to extract features from the input data and develop a new classifier. The framework of the proposed system and the stages of the research methodology are illustrated in *Fig. 2*. The proposed system utilizes Inception-v3 transfer learning to classify the keyframes of a video as either normal or abnormal behaviors. The fundamental steps will be discussed in the following sections.

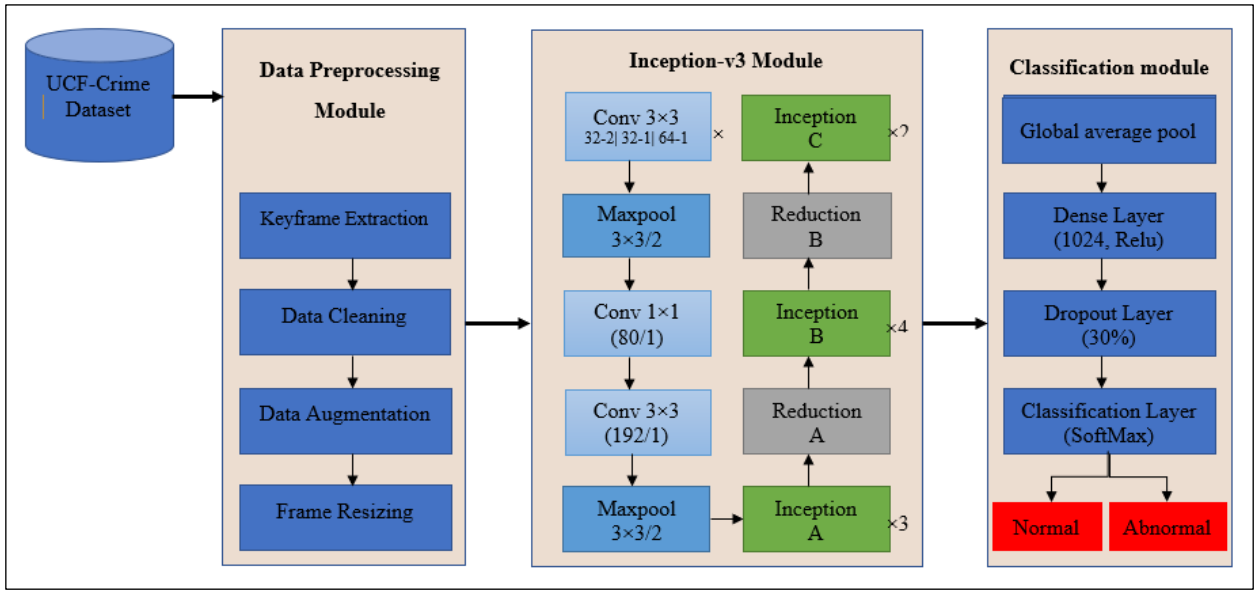

FIG. 2. PROPOSED FRAMEWORK FOR ABNORMAL BEHAVIOR DETECTION.

A. UCF-Crime Dataset

The UCF-Crime dataset is a benchmark dataset used in studies related to criminal activity recognition. It was created by the University of Central Florida (UCF) and is publicly available for researchers and practitioners to use for their studies. The dataset's videos were collected from various sources, including YouTube, TV news, and documentaries, and have varying quality and resolution. It comprises 1900 long, untrimmed surveillance videos, totaling 128 hours of footage. The videos depict 13 real-world criminal activities, including abuse, arrest, arson, assault, road accidents, burglary, explosions, fighting, robbery, shooting, stealing, shoplifting, and vandalism [28]. These activities were selected due to their significant impact on public safety. Additionally, the UCF-Crime dataset poses several challenges, such as variations in camera viewpoint and motion, cluttered backgrounds, and variations in the scale of the foreground. Furthermore, some of the video clips contain low quality, noise, and changes in illumination. In addition to the challenges mentioned earlier, the UCF-Crime dataset also has variations in the number of people involved in the activities, their clothing, and the type of weapons used, among other factors. These variations make the dataset a challenging benchmark for developing robust and accurate criminal activity recognition systems. For the specific study, the proposed system aims to recognize the normal and fighting classes in the UCF-Crime dataset. To achieve this, 70 video clips were selected for the normal class, divided into 50 videos for training and 20 videos for testing. Similarly, the fighting class comprises 50 video clips divided into 40 videos for training and 10 videos for testing. *Fig. 3* shows examples of different fighting videos in the UCF-crime dataset. The data is split in this way to establish a balanced dataset, as the video clips have varying duration times and therefore different numbers of frames.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

FIG. 3. EXAMPLES OF DIFFERENT FIGHTING VIDEOS IN THE UCF-CRIME DATASET.

B. Data Preprocessing Module

To ensure that the dataset used in this work is in a format that can be easily used by the model, several steps have been taken:

1. Keyframe Extraction

When analyzing frames that have been extracted from video clips, it was observed that many consecutive frames were identical. This repetition of frames increases the complexity of the model and its computational requirements. To address this issue, the keyframe extraction technique was employed. Keyframe extraction is a technique used to identify important frames within a video or series of images. These keyframes are typically selected based on visual characteristics such as changes in motion, color, or content. In this research, a threshold-based method was used to eliminate the repetition of consecutive frames in the video. This method works by identifying changes in the video stream and selecting keyframes at the points where these changes occur. In threshold-based methods, a threshold is set to distinguish between consecutive shots based on a particular feature. The equation for this method can be represented as follows:

$$\textit{If } F(x,y) - F(x',y') > T, \textit{ then } (x,y) \textit{ and } (x',y') \textit{ belong to different shots} \qquad (1)$$

where F(x,y) and F(x',y') are the feature values at the two frames, and T is the threshold value.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

2. Data cleaning

Data cleaning involves identifying and removing or modifying data that is not relevant to the task at hand. In the context of video data, this can include removing frames that do not contain useful information. In this work, when extracting frames from videos of fighting, some frames may not actually depict the fight itself and may be more similar to frames from a normal video. These frames, as well as blank and noisy frames, have been removed manually during the data cleaning process to improve the overall quality and relevance of the data being used for training.

3. Data Augmentation

In the data preprocessing module, a data augmentation technique was used to boost the diversity of the training dataset and enhance the performance of the proposed model. This process entails artificially expanding the amount of training data by generating modified versions of existing data, achieved through techniques such as rotation, flipping, shifting, and scaling. By employing data augmentation methods, it is possible to artificially balance the dataset and enhance the model's capability to generalize to new examples.

4. Frame Resizing

Since the Inception-v3 network processes image inputs of size 299×299 for RGB images, the video input of the aforementioned dataset was transformed into image frames and cleaned of unrelated and noisy frames. These frames were then resized to a shape of 299×299×3 (height, width, channels) and fed as input to the Inception-v3 network for feature extraction. The image frames utilized for training were segregated into two categories, namely abnormal (representing fighting) and normal, with an almost equal number of images for both groups to ensure unbiased training. Table I presents details about the data division followed in this work.

TABLE I. DATA DIVISION USED IN THE EXPERIMENTS

Class name	Group name	No. of clips	No. of frames
Normal	Training	50	4612
	Testing	20	1636
Fighting	Training	40	4641
	testing	10	1836

C. Inception-v3 Module

To train the proposed system and extract features, the Inception-v3 network was utilized in this module for extracting features from input frames as follows:

First, three convolution layers with 3×3 filters and one max pooling layer are used to extract low-level features of the input image. in addition, two convolution layers with 1×1 and 3×3 filters and another max pooling layer are used to further feature extraction, resulting in a feature map of size 35×35×192.

Next, nine inception layers and two reduction layers are used to extract high-dimensional features, resulting in a feature map of size 8×8×2048.

D. Classification Module

The feature classification module is used to train the weights of the feature vectors to identify human behavior as either normal or anomalous. This module includes the following layers:

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

- **A global average pooling layer** is used to reduce the spatial dimensions of the input feature maps and obtains a fixed-length feature vector. Here is the equation for a global average pooling layer:

$$G_i = \frac{1}{H \times w} \times \sum \sum f(i,j,k) \qquad (2)$$

where the summation is taken over all the spatial dimensions of the feature map, i.e., j = 1 to H and k = 1 to W.

Given a feature map F of size HxWxC, where H is the height, W is the width, and C is the number of channels, the output of the global average pooling layer G is a vector of length C, where each element i is the average of all the activations in feature map F across the spatial dimensions.

- **A flatten layer**, to convert the output of the previous layer into a 2048-feature vector that can be fed into a fully connected layer.
- **A dense (fully connected) layer** is consisting of 1024 neurons and a Relu activation function. Each neuron connects to 2048 nodes of the feature vector.
- **A dropout layer** is used to reduce overfitting and improve the model's ability to handle noise and generalize to new data. The dropout rate used was 0.3, meaning that 30% of the neurons were randomly dropped out during training.
- **A Softmax layer** is used to classify the features and determine the final probability of the two types of human behavior in the UCF-Crime dataset. The definition of the softmax activation function is as follows:

$$softmax(c_i) = \frac{e^{ci}}{\sum_{j=1}^{k} e^{ci}} \qquad (3)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experiment Setup

The experiments were conducted using the Python programming language, and the Keras library was utilized with Tensorflow as the underlying platform. Keras is a Python-based high-level API for neural networks, which can run on top of TensorFlow. The hardware setup used for training and testing the system had the following specifications:

CPU: 12th Generation Intel Core i7-1265H (2.30 GHz, 10 cores).

GPU: NVIDIA GeForce RTX 3070 (8 GB).

B. Hyper-parameters Tuning

Deep learning models contain several hyperparameters such as the activation function, optimizer, number of neurons per layer, number of epochs, and batch size. Due to the high-dimensional nature of this configuration space, determining the optimal parameter settings is a challenging task. Table II lists the hyper-parameters used in this research.

TABLE II. HYPER-PARAMETERS USED IN THE EXPERIMENTS

Parameters	Values
Number of epochs	30
Batch size	32
learning rate	0.0001
Dropout rate	0.3
No. of neurons in the dense layer	1024

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

C. Performance Evaluation Metrics

The proposed system's performance is evaluated based on several measures to demonstrate its effectiveness. These measures include accuracy, precision, recall, and F1-score. These metrics are defined below [18][29].

- **Accuracy** is a metric that evaluates the successful classification of both normal and anomaly instances in relation to the entire dataset. It can be a valuable metric when the dataset has a balanced distribution of instances (Eq. 1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- **Precision** is a measure that assesses the accuracy of true positive results by comparing the number of actual anomalies detected to the overall number of anomalies present. It gauges the accuracy of the true positive results (Eq. 2).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- **Recall** is a metric that calculates the ratio of detected anomalies to the total number of anomalies. It is frequently utilized when it is necessary to correctly categorize events that have already taken place (Eq. 3).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

- **F1-Score** is a metric that determines the harmonic mean of recall and precision rates. The higher the F1-Score, the better the model's performance. It is often employed when the distribution of classes is uneven (Eq. 4).

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

D. Results Analysis of Pre-Trained and Fine-Tuned Inception-v3 approaches

In this study, the researchers applied transfer learning techniques to the Inception-v3 architecture to extract features from the input data and train a new classifier to classify normal and abnormal behaviors in the UCF-Crime dataset. The pre-trained and fine-tuned models both utilized a Softmax classifier to classify the features and determine the final probability of normal and abnormal behaviors in the dataset. The pre-trained model had all layers frozen, meaning that the weights of the layers were not changed during the training process. In contrast, the fine-tuned model had all layers unfrozen, and the weights of the layers were updated during the training process. This approach allowed the fine-tuned model to adapt the features learned by the pre-trained model to the specific characteristics of the UCF-Crime dataset. Table III displays the results obtained by implementing the pre-trained and fine-tuned Inception-v3 network approaches in the terms of accuracy, recall, precision, and F1-score. The research results showed that both pre-trained and fine-tuned Inception-v3 models were effective in classifying the normal and abnormal behaviors in videos, but the fine-tuned model outperformed the pre-trained model in terms of accuracy, recall, precision, and F1 score. Specifically, the fine-tuned model achieved an accuracy of 88.0%, a recall of 89.24%, a precision of 85.83%, and an F1 score of 87.50%. These results indicate that the fine-tuned model correctly identified the class of 88% of the instances in the dataset, correctly identified 89.24% of the positive instances in the dataset, and correctly classified 85.83% of the instances it identified as positive. The F1 score of 87.50% provides a single score that balances both recall and precision, indicating that the model achieved a good balance between correctly identifying positive instances and avoiding false positives. On the other hand, the pre-trained model achieved an accuracy of 86.2%, a recall of 86.43%, a precision of 84.62%, and an F1 score of 85.52%.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

These results were still relatively good, but they were lower than those of the fine-tuned model, suggesting that fine-tuning the weights of the layers can lead to improved performance in criminal activity recognition systems. Overall, these results demonstrate that transfer learning with the Inception-v3 architecture is an effective approach for classifying criminal activities in surveillance videos and fine-tuning the weights of the layers can further improve the performance of the model.

TABLE III. THE RESULTS OBTAINED BY THE PRE-TRAINED AND FINE-TUNED APPROACHES

Approach Name	Accuracy	Recall	Precision	F1 score
Pre-trained Inception-v3	86.20%	86.43%	84.62%	85.52%
Fine-tuned Inception-v3	88.0%	89.24%	85.83%	87.50%

Table IV presents a comparison between the two proposed frameworks and other contemporary frameworks, using the UCF-crime dataset, in terms of accuracy. It should be noted that the fine-tuned Inception-v3 model achieves a very competitive level of accuracy compared to the results reported in [30] and is better than other methods.

TABLE IV. COMPARISON OF OUR PROPOSED METHODS WITH OTHER CONTEMPORARY METHODS THAT USED THE UCF-CRIME DATASET IN THE TERM OF ACCURACY

Ref., year	Method	Accuracy %
[31], 2018	VGG19- FlowNet	76.66
[32], 2019	CNN-LSTM	85.0
[30], 2020	InceptionV3-VGG16	88.74
[33], 2021	Resnet50- LSTM	79.81`
Our proposed model	**Pre-trained Inception-v3**	**86.20**
Our proposed model	**Fine-tuned Inception-v3**	**88.0**

VI. CONCLUSIONS

In conclusion, this study demonstrated the effectiveness of using transfer learning and the Inception-v3 network to develop a system for detecting abnormal human behavior in video surveillance footage. Both pre-training and fine-tuning approaches were employed to extract features and develop a new classifier. The proposed system was evaluated using the benchmark UCF-Crime dataset, and the results showed that transfer learning with Inception-v3 architecture accurately classifies normal and abnormal behaviors in videos. Moreover, the fine-tuned Inception-v3 model outperformed the pre-trained model in terms of accuracy, recall, precision, and F1 score. Specifically, the accuracy of the fine-tuned model was 88.0%, with a recall of 89.24%, a precision of 85.83%, and an F1-score of 87.50%. These results clearly demonstrate the benefits of fine-tuning the weights of the layers for improved performance. Overall, this research provides a valuable framework for using transfer learning in video surveillance, which can have important implications for public safety and security. The improved performance of the fine-tuned Inception-v3 model indicates that this approach may be particularly useful in real-world applications of video surveillance for abnormal behavior detection.

REFERENCES

[1] S. M. Mohtavipour, M. Saeidi, and A. Arabsorkhi, “A multi-stream CNN for deep violence detection in video sequences using handcrafted features,” Vis. Comput., pp. 1–16, 2022.

[2] M. A. Ali, A. J. Hussain, and A. T. Sadiq, “Deep Learning Algorithms for Human Fighting Action Recognition.,” Int. J. Online Biomed. Eng., vol. 18, no. 2, 2022.

[3] D. K. Ghosh and A. Chakrabarty, “Two-stream Multi-dimensional Convolutional Network for Real-time Violence Detection,” arXiv Prepr. arXiv2211.04255, 2022.

Received 19/February/2023; Accepted 22/March/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

[4] H. A. Abosaq et al., "Unusual driver behavior detection in videos using deep learning models," Sensors, vol. 23, no. 1, p. 311, 2022.

[5] R. A. Lateef and A. R. Abbas, "Human Activity Recognition using Smartwatch and Smartphone: A Review on Methods, Applications, and Challenges," Iraqi J. Sci., pp. 363–379, 2022.

[6] M. A. Ali, A. J. Hussain, and A. T. Sadiq, "Detection And Count of Human Bodies In a Crowd Scene Based on Enhancement Features By Using The YOLO v5 Algorithm," IRAQI J. Comput. Commun. Control Syst. Eng., vol. 22, no. 2, pp. 125–134, 2022.

[7] A. A. Khan et al., "Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model," Electronics, vol. 11, no. 19, p. 3105, 2022.

[8] M. M. Mahmoud and A. R. Nasser, "Dual Architecture Deep Learning Based Object Detection System for Autonomous Driving," Iraqi J. Comput. Commun. Control Syst. Eng, vol. 21, no. 2, pp. 36–43, 2021.

[9] T. A. Jaber, "Artificial intelligence in computer networks," Period. Eng. Nat. Sci., vol. 10, no. 1, pp. 309–322, 2022.

[10] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," J. big Data, vol. 8, no. 1, pp. 1–74, 2021.

[11] S. A. Jebur, M. A. Mohammed, and A. K. Abdulhassan, "Covid-19 detection using medical images," in AIP Conference Proceedings, 2023, vol. 2591, no. 1, p. 30030.

[12] A. Ravichandran and S. Sankaranarayanan, "Anomaly Detection in Videos Using Deep Learning Techniques," in Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020, 2021, pp. 263–275.

[13] U. M. Butt, S. Letchmunan, F. H. Hassan, S. Zia, and A. Baqir, "Detecting video surveillance using VGG19 convolutional neural networks," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 2, 2020.

[14] M. Rajeshwari and C. H. MallikarjunaRao, "Detecting anomalous road traffic conditions using VGG19 CNN Model," in E3S Web of Conferences, 2021, vol. 309, p. 1226.

[15] Y. Zahid, M. A. Tahir, and M. N. Durrani, "Ensemble learning using bagging and inception-V3 for anomaly detection in surveillance videos," in 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 588–592.

[16] Y. Zahid, M. A. Tahir, N. M. Durrani, and A. Bouridane, "Ibaggedfcnet: An ensemble framework for anomaly detection in surveillance videos," IEEE Access, vol. 8, pp. 220620–220630, 2020.

[17] G. Pang, C. Yan, C. Shen, A. van den Hengel, and X. Bai, "Self-trained deep ordinal regression for end-to-end video anomaly detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12173–12182.

[18] S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, "Review on Deep Learning Approaches for Anomaly Event Detection in Video Surveillance," Electronics, vol. 12, no. 1, p. 29, 2022.

[19] K.-E. Ko and K.-B. Sim, "Deep convolutional framework for abnormal behavior detection in a smart surveillance system," Eng. Appl. Artif. Intell., vol. 67, pp. 226–234, 2018.

[20] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "Transfer deep learning along with binary support vector machine for abnormal behavior detection," IEEE Access, vol. 8, pp. 61085–61095, 2020.

[21] M. Zerkouk and B. Chikhaoui, "Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models," Sensors, vol. 20, no. 8, p. 2359, 2020.

[22] C. Huang, X. Wang, J. Cao, S. Wang, and Y. Zhang, "HCF: A hybrid CNN framework for behavior detection of distracted drivers," IEEE access, vol. 8, pp. 109335–109349, 2020.

[23] A. Alia, M. Maree, and M. Chraibi, "A hybrid deep learning and visualization framework for pushing behavior detection in pedestrian dynamics," Sensors, vol. 22, no. 11, p. 4040, 2022.

[24] R. Lalit, R. K. Purwar, S. Verma, and A. Jain, "Crowd abnormality detection in video sequences using supervised convolutional neural network," Multimed. Tools Appl., vol. 81, no. 4, pp. 5259–5277, 2022.

[25] L. R. Ali, S. A. Jebur, M. M. Jahefer, and B. N. Shaker, "Employing Transfer Learning for Diagnosing COVID-19 Disease.," Int. J. Online Biomed. Eng., vol. 18, no. 15, 2022.

[26] S. H. Ameen, "Detection and Classification of Leaf Disease Using Deep Learning for a Greenhouses' Robot," IRAQI J. Comput. Commun. Control Syst. Eng., vol. 21, no. 4, pp. 15–28, 2021.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[28] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6479–6488.

[29] M. A. Ali, A. J. Hussain, and A. T. Sadiq, "Human Fall Down Recognition Using Coordinates Key Points Skeleton.," Int. J. Online Biomed. Eng., vol. 18, no. 2, 2022.

[30] S. Majhi, R. Dash, and P. K. Sa, "Two-Stream CNN architecture for anomalous event detection in real world scenarios," in Computer Vision and Image Processing: 4th International Conference, CVIP 2019, Jaipur, India, September 27–29, 2019, Revised Selected Papers, Part II 4, 2020, pp. 343–353.

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.23.2.16

[31] K. Biradar, S. Dube, and S. K. Vipparthi, "DEARESt: deep Convolutional aberrant behavior detection in real-world scenarios," in 2018 IEEE 13th international conference on industrial and information systems (ICIIS), 2018, pp. 163–167.

[32] M. R. Anala, M. Makker, and A. Ashok, "Anomaly detection in surveillance videos," in 2019 26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW), 2019, pp. 93–98.

[33] Z. K. Abbas and A. A. Al-Ani, "Anomaly detection in surveillance videos based on H265 and deep learning," Int. J. Adv. Technol. Eng. Explor., vol. 9, no. 92, p. 910, 2022.

Received 19/February/2023; Accepted 22/March/2023
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*