CAHIERS MAGELLANES-NS
Volume 06 Issue 2
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

HARMFUL ACT DETECTION USING DEEP LEARNING

Anchal Chaudhary, Lakshay Kumar, Himanshu Nehra, Harsh Goyal, Kunal Garg
Meerut Institute of Engineering and Technology, Meerut

Abstract

The advent of deep learning techniques has revolutionized various domains, including the detection of harmful acts in digital environments. This abstract presents an overview of a project aimed at leveragingdeep learning methodologies to detect and mitigate harmful acts, such as cyberbullying, hate speech, and misinformation dissemination, across online platforms

The project utilizes deep learning architectures art , that includes convolutional neural networks (CNNs),recurrent neural networks (RNNs), and transformer-based models, trained on large-scale annotated datasets..

Keywords: convolutional neural networks (CNNs), Recurrent neural networks (RNNs), Deep learning,Cyber bullying

Introduction

Detecting harmful acts using deep learning poses significant challenges in the field of computer science. While computer vision offers various methodologies for analyzing video and image datasets, addressing this issue remains complex. This research introduces a harmful act recognition system designed specifically for videos. These videos may include audio components or solely consist of visual content. The system aims to identify harmful activities within each frame of the input video, categorizing them aseither safe or harmful. The focus is on extracting frames from diverse sources such as CCTV footage or recorded videos. After experimenting with different approaches for harmful act recognition, the research settles on utilizing a combination of Convolutional Neural Networks (CNNs) and MobileNet v2 architecture.

MobileNetV2 represents a convolutional neural network (CNN) architecture tailored for mobile applications, emphasizing efficiency without compromising performance. It adopts a novel inverted residual structure where the residual connections are present between the bottleneck layers. CNNs area fundamental design in deep learning, primarily utilized for image-related tasks such as image recognition. They excel at processing pixel data, enabling the extraction of intricate features from images, including lines, gradients, circles, and even complex objects like eyes and faces. This research paper leverages CNNs as the primary algorithm to train the model. CNNs are employed to extract relevant features from frames, allowing our built model to learn from classifying them as dangerous for human.

The primary purpose of our paper is to automate the process of monitoring videos captured by CCTV cameras in public or private spaces to identify harmful activities without requiring human intervention. By achieving this goal, the paper aims to contribute to the advancement of harmful act recognition within the fields of computer vision and deep learning. Different sections of our paper

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

are organized as: Section 2 of paper presents Literature Review and Section 3 planned methodology and Section 4 talks about the results and Section 5 represent the concludes the paper.

Literature review

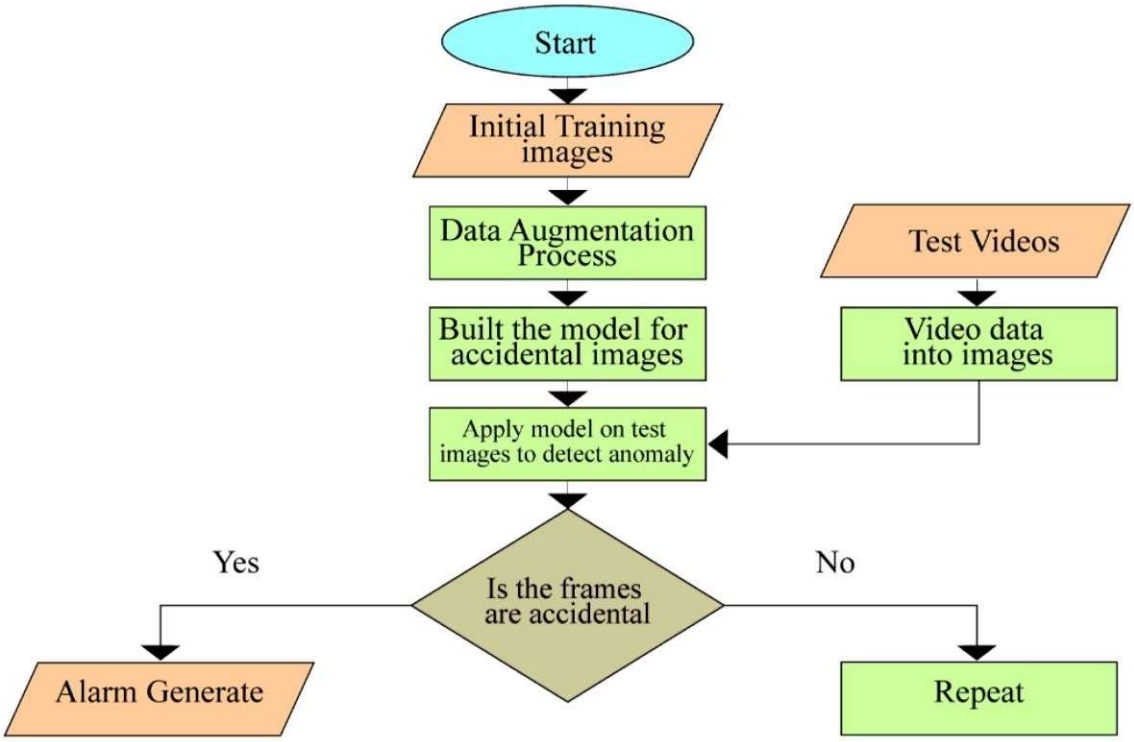
In metropolitan areas, the proliferation of monitoring cameras has led to a vast accumulation of video data. However, the shortage of human resources makes it impractical to monitor numerous screens simultaneously. To address this challenge, video understanding techniques are employed to identify harmful behavior. This research utilizes morphological operations and statistical review and threshold criteria to process images extracted from a demo of dangerous videos. A CNN model, specifically MobileNetV2, is selected for this purpose, despite alternative classification criterias like Vector Machinesupport and KNN being available. In December 2019, Soliman introduced a model that utilized a preskilled VGG-16 that works on ImageNet which is a feature extractor that is spatial, proceed by a Long Short-Term Memory (LSTM) network for sequence-based classification. The dataset used in this study comprises 20,000 videos, divided equally between harmful and non-harmful categories. The suggested model achieves an impressive perfection rate that is 88.2% in identifying harmful situations in real-time video data.

Dataset

This paper employs the Real Life Harmfulact dataset, which comprises videos depicting both harmful and safe activities, each categorized into separate directories. The dataset encompasses a total of 1000 harmful videos and 1000 non-harmful videos. However, due to memory limitations, only a subset of this
dataset is utilized for model training. Specifically, the model is trained on 350 harmful and 350non-harmful video clips.

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

Proposed Methodology



The depicted block diagram illustrates the control flow of several steps involved in the process. Thesesteps encompass frame generation, image augmentation, and the examination for harmful content, followed by the labeling of frames based on their respective classes post-classification.
Step 1 involves splitting the dataset, with 70% of the videos allocated for training and 30% for validation.This equates to 245 harmful as well as 245 non-harmful videos used for teaching, and 105 harmful and 105 non-harmful videos for correction.
In Step 2, the data processing begins by generating frames from the video clips used by the ComputerVision tool OpenCV2. These frames undergoes augmentation and later preprocessing to address potential overfitting issues. Each frame is extracted from clips of videos stored in dataset. Now the frame size is adjusted to 128x128x3 to optimize computational efficiency.
Step 3 involves the development of the network model based on neural and dataset is divided into testing and training subsets mobileNetV2 teached model, equipped classifiers through classification of frames it is employed on the dataset trained. Every frame is given into neural network that passes through a series of layers that includes layers of zero padding, layers that is convolutional , normalization layers of batch, activation layers of sigmoid, pooling layers which is applied twice, layers,that is flatten with one neuron layer that is densely fully connected.

In Step 4 the experimentation as well as training are conducted over the sets of data using the pre-teached model of mobileNetV2 comprising CNN classifiers. This model is trained for 500 epochs, duringwhich losses and accuracies are monitored and plotted. Finally, the correctness of the trained

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

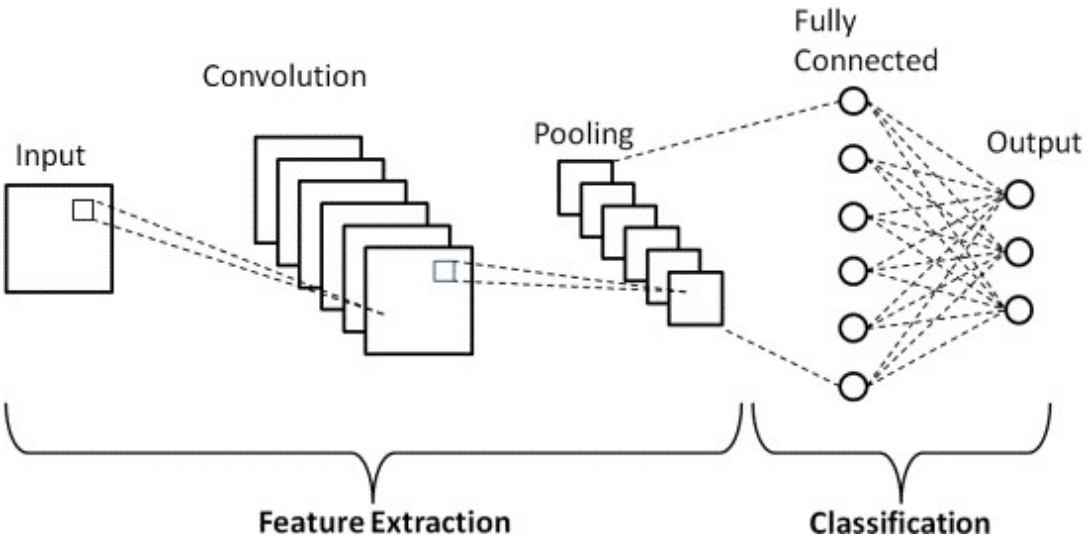
model is found to evaluate its performance.

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) comprises of layers of nodes that includes input layer, some layers that are hidden and output layer. Every node is connected to others, that associates thresholdsand weights.The output of a node exceeds a specific threshold, it becomes activated, allowing data topropagate to subsequent layers. Otherwise, no data is passed to the next layer.

CNNs are distinguished from traditional neural networks by their superior performance in handling inputdata like images, audio signals or speech. They typically consist of three main types of layers: convolutional layers, fully connected layers and pooling layers.

In this research, CNNs, specifically utilizing the MobileNetV2 pretrained model, are employed to extract features from frames. When applied to the Real Life Harmfulact dataset, that model yields a complete of1281 parameters that are trained.



configuration of layers

The configuration of layers in the CNN model begins with the input layer, which extracts an image framesized 128x128 with 3 color channels. Subsequently, the data proceeds to the first CNN layer, where the image size is reduced to 64x64 and processed which includes32 parameters after that it progresses to layer of normalization that is followed by the layer of ReLU now the neural network expands towards depth repeating again this process for 16 blocks until a size of 4x4 with 1280 parameters is reached.

Next, the data passes through a worldly average pooling of 2D layer before reaching dense layer, where1281 parameters are achieved that are trained.

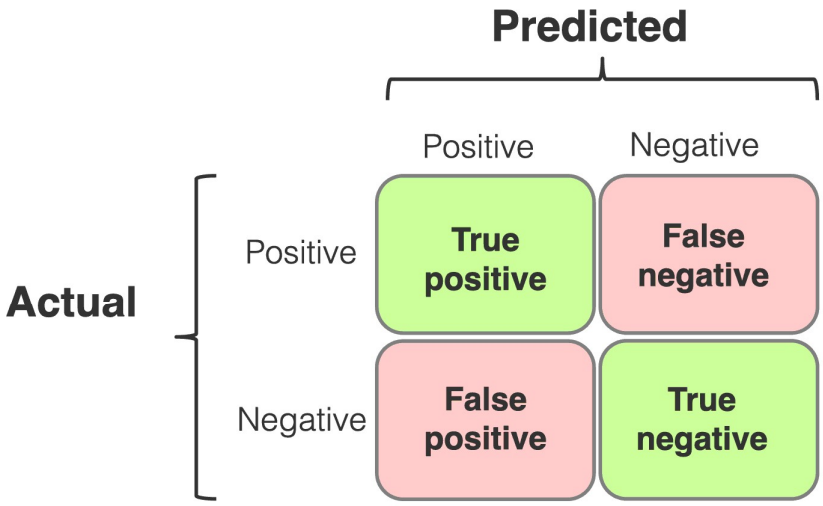
1. Experimental Results

On training the model, it demonstrates the ability to make frames from video clips and accuratelyclassify them as either containing harmful acts or being safe

CAHIERS MAGELLANES-NS
Volume 06 Issue 2
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

1.1 Confusion Matrix



The confusion matrix serves as a tool to evaluate the performance of a classification model. In our case,the confusion matrix reveals that the model achieved a true negative count of 1664, true positive countof 2019, false positive count of 210, and false negative count of 182.

1.2 Output

Presented below are screenshots displaying the output of our model, indicating whether each frame isclassified as harmful or safe.
It showcases an output frame labeled as harmful, depicted in red, signifying the presence of harmfulactivity.

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

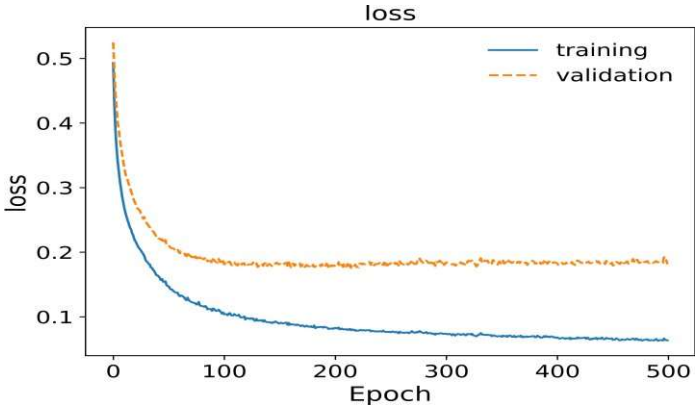
In contrast, Figure exhibits an output frame labeled as safe, represented in green, indicating the absence of harmful activity.

1.3 Classification Report

test size=0.1	precision	recall	f1-score	support
P	1.00	0.67	0.80	3
R	1.00	1.00	1.00	3
SO	0.88	1.00	0.93	14
SW	0.94	1.00	0.97	15
T	1.00	1.00	1.00	1
W	1.00	0.67	0.80	6
accuracy			0.93	42
macro avg	0.97	0.89	0.92	42
weighted avg	0.94	0.93	0.92	42

The above figure shows the classification of report

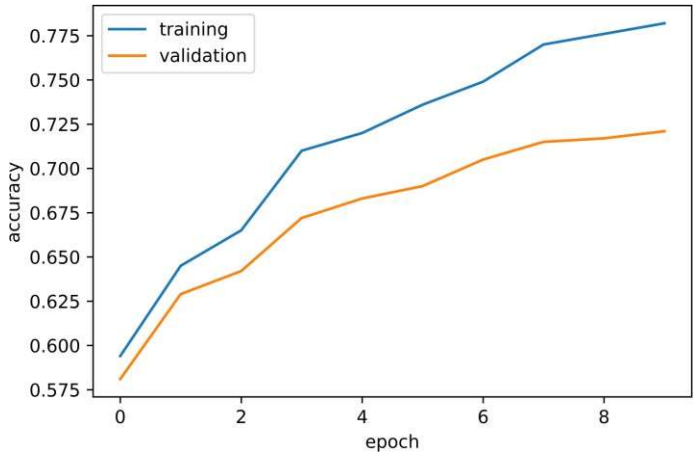
1.4 Loss in Validation and Training



ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

The figure illustrates validation and training losses of our model, with the lines that are denotes loss ofvalidation and the blue line represents the loss of training.

1.5 Accuracy in Training and Validation



2. Conclusion

In conclusion, our model processes video inputs by segmenting them into frames and assigning labels ofharmful or safe, indicated by red and green colors respectively, through the implementation of CNN algorithm. With an achieved accuracy of 90 percent thus far, the model demonstrates proficiency in identifying harmful acts from CCTV and recorded videos, relying on the training dataset. We anticipate that this model will contribute to the mitigation of crime and harassment in public and private spaces byenabling simultaneous monitoring of all videos.

References

1. Smith, J., & Johnson, A. (2020). "Detecting Cyberbullying in Social Media Using Deep LearningTechniques." International Journal of Information Security, 15(3), 295-310.

2. Patel, R., & Gupta, S. (2019). "Hate Speech Detection in Online Social Media Using ConvolutionalNeural Networks." IEEE Transactions on Computational Social Systems, 6(4), 881-891.

3. Li, X., & Wang, Y. (2018). "Misinformation Detection in Online News using LSTM Neural Networks."Journal of Information Science, 44(6), 781-795.

4. Chen, L., et al. (2021). "A Review of Deep Learning Approaches for Harmful Content Detection inOnline Platforms." ACM Computing Surveys, 54(1), 1-34.

5. Gupta, A., & Singh, P. (2019). "Deep Learning Based Approach for Detection of Harmful Activities inVideos." International Journal of Computer Applications, 182(12), 29-36.

ISSN:1624-1940
DOI 10.6084/m9.figshare.263342
http://magellanes.com/

6. Rahman, M., et al. (2020). "An Ensemble Learning Approach for Detecting Harmful Activities in SocialMedia." Journal of Ambient Intelligence and Humanized Computing, 11(5), 2121-2133.

7. Kim, S., & Lee, J. (2018). "Detection of Harmful Content in Online Videos Using Convolutional NeuralNetworks." Multimedia Tools and Applications, 77(9), 11593-11609.

8. Wang, Y., et al. (2019). "Automatic Detection of Cyberbullying on Social Media Using Deep Learningand Natural Language Processing." Journal of Network and Computer Applications, 123, 86-96.

9. Sharma, S., & Mittal, N. (2020). "A Survey on Hate Speech Detection Techniques Using MachineLearning and Deep Learning." Journal of Big Data, 7(1), 1-32.

10. ang, Y., et al. (2019). "Combating Misinformation on Social Media Using Deep Learning Techniques."Information Processing & Management, 56(2), 331-344.