Paper one summary

Shoplifting Detection Using Hybrid Neural Network CNN-BiLSMT and Development of Benchmark Dataset

introduction

Each year goods worth billions of dollars are shoplifted globally. Shoplifting (also known as retail theft) is a criminal act that reduces the profitability of businesses. It involves stealing merchandise from a retail establishment without paying for it while avoiding detection. Shoplifting is a crime that receives significant attention among the various criminal activities that occur in stores. It also involves stealing merchandise from a retail establishment by hiding the items in one's clothing, pockets, or bag and leaving the store without paying for them. Shoplifting poses a significant challenge for store owners and other entities such as law enforcement, the government, and the justice system. A study conducted by the National Association for Shoplifting Prevention indicates that one out of every 11 individuals engages in shoplifting. Additionally, it has been reported that these thieves are apprehended only once in every 48 instances of theft

Contribution of This Study

The main contribution of this study is as follows: (1) a large dataset of 900 videos has been developed, (2) baseline methods 2D and 3D CNN have been implemented to evaluate the dataset, (3) proposed shoplifting detection technique using hybrid neural network CNN-BiLSTM (4) make a comparison of deep learning approaches 2D and 3D CNN with the proposed technique.

Methods used in this paper

[1] Baseline methods

CNNs have been demonstrated to have exceptional performance in a computer vision tasks, One variant of CNNs, 2D and 3D CNNs have been developed to extract spatial and temporal features from videos.

[1] Two Dimensional Convolutional Neural Networks

[2] Three Dimensional Convolutional Neural Networks

In 3D CNN, we have utilized a 3D model to train a dataset that identifies shoplifting. Our training set consists of 81,000 images with (frame sequence = 90) for each movie chosen for training input out of a total of 900 videos from 2 classes. During training, the images are inputted with a size of 600 × 600 and are passed from the input layer. In the next step, 8 Convolutional layers were used to extract low-level features for learning from the sequence of frames. After each 3D convolutional layer, an activation layer was added to the network image with the RELU activation function. In the final step, the softmax activation function was used for the final prediction. The incorporation of an extra-temporal dimension in 3D CNNs raises the training and inference computing costs. Due to this, training big 3D CNNs on hardware with limited resources might be difficult. Since volumetric data is processed using 3D CNNs, a sizable amount of memory is needed to store the data and intermediate feature maps

[2] Proposed Method

CNN has proven to be a highly effective method for image classification, outperforming other approaches. When it comes to video classification, the features of each frame in a sequence need to be extracted and inputted into a Bidirectional Long Short-Term Memory. Since CNN can identify hidden patterns in individual frames and changes in a sequence of frames, it is the best option for feature extraction. However, training a CNN from scratch requires a large dataset of images, substantial computational resources, and significant time investment for training and testing the model, making it a costly approach. Transfer learning can be used to mitigate this problem, where the last hidden layers of a pre-trained CNN can be removed, and only the image features are extracted.

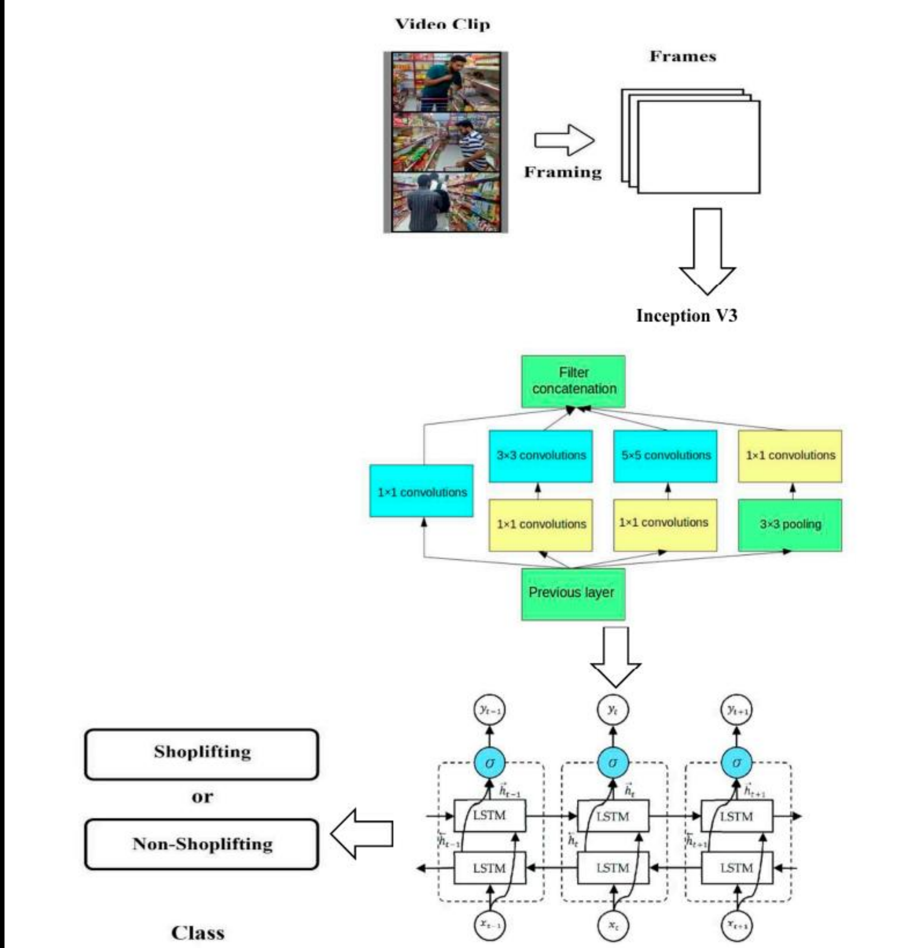

Figure 7. Proposed architecutre.

Results and Discussion

[1] Comparison of baseline methods and proposed method in terms of accuracy

Methods	Convolutional layers	Activation units	Training accuracy	Validation accuracy	Recall	F1
2D CNN	4	ReLU Softmax	50.00	45.00	50.00	50.40
3D CNN	8	ReLU Softmax	60.85	55.38	58.80	61.80
Proposed Method	16	ReLU Softmax	82.01	81.00	78.40	83.01

[2] Confusion Matrix of 2D CNN

Table 3. Confusion Matrix of 2D CNN.

		Predicted	
		Shoplifting	Non-Shoplifting
Actual	Shoplifting	230	220
	Non-Shoplifting	230	220

[3] Confusion Matrix of 3D CNN

Table 4. Confusion Matrix of 3D CNN.

		Predicted	
		Shoplifting	Non-Shoplifting
Actual	Shoplifting	300	150
	Non-Shoplifting	210	240

[4] Confusion Matrix of proposed methods

Table 5. Confusion Matrix of Proposed Model.

		Predicted	
		Shoplifting	Non-Shoplifting
Actual	Shoplifting	400	50
	Non-Shoplifting	110	340