Paper Three summary

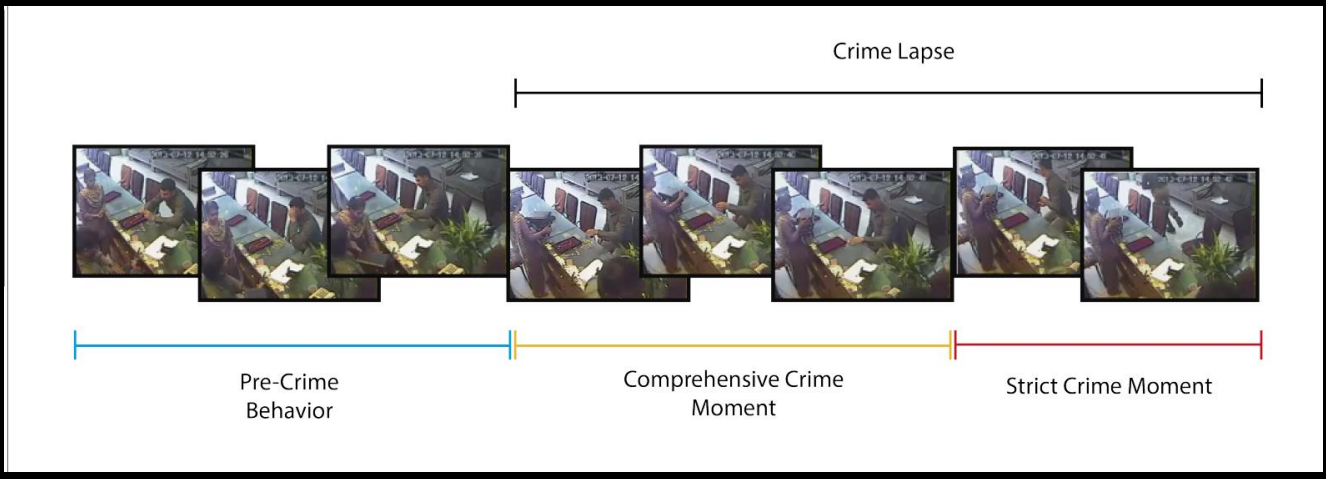
Suspicious Behavior Detection on Shoplifting Cases for Crime Prevention by Using 3D Convolutional Neural Networks.

Description of the Dataset

We use the UCF-Crime dataset, proposed by , to analyze suspicious behavior before a shoplifting crime. The dataset consists of 1900 real-world surveillance videos and provides around 129 hours of videos (with a resolution of 320x240 pixels and not normalized in length). The dataset includes scenarios from several locations and persons that are grouped into thirteen classes such as 'abuse', 'burglary', and 'explosion', among others. From those classes, we extracted the samples used in this work from the 'shoplifting' and 'normal' classes.

3D Convolutional Neural Networks

We use a 3DCNN for feature extraction and classification. We choose a basic structure to explore the performance of the 3DCNN for suspicious behavior detection task. The architecture of the model consists of four Conv3D layers, two max-pooling layers, and two fully connected layers. As a default configuration, in the first pair of Conv3D layers, we apply 32 filters, and for the second pair, 64 filters. All kernels have a size of 3×3×3, and the model uses an Adam optimizer and cross-entropy for loss calculation. At the end of the model, it has two dense layers with 512 and 2 neurons, respectively. The output is binary, 1 for Suspicious Behavior and 0 for Normal Behavior. This architecture is selected because it has been used for similar applications , and seems suitable as a first approach for behavior detection in surveillance videos. For handling the model training, we use Google Colaboratory . This free cloud tool allows to write and execute code in cells, runs from a browser, and uses a GPU to train deep learning models. We can upload the datasets to a storage service, link the files, prepare the training environment, and save considerable time during the model training, using a virtual GPU.



Pre-Crime Behavior

To detect suspicious behavior, the proposed model must focus on what happens before a shoplifting crime is committed. For this purpose, we propose a new method to process surveillance videos. Before we explain our proposal, we introduce some concepts, which are listed below.

• Strict Crime Moment (SCM). In a surveillance video, and after being reviewed by a human, the SCM is the segment of video where a person commits shoplifting crime. This moment is the primary evidence to accuse a person of committing a crime.

• Comprehensive Crime Moment (CCM). it is the precise moment when an ordinary person can detect the suspect's intentions. He/she started to watch out to go unnoticed and looks for the best moment to commit the crime. Other CCM examples are unsuccessful attempts or reorder things to distract attention. If we isolate this moment, we can doubt the suspect in the video, but there is no clear evidence to know if the suspect steals something.

• Crime Lapse (CL). In a video, the CL is the entire segment where a crime takes place. If we remove the CL from the video, it will be impossible to determine that there is a criminal act in the video. The CCM supports the beginning of the CL. It is essential not to leave any trace of the crime to avoid biasing the training.

• Pre-crime Behavior (PCB). The PCB contains what happens from the first appearance of the suspect until the CCM begins. These samples have different sizes since each video shows many behaviors. We can find more than one CL per video. The next PCB will start where the previous CL ends and until the next CCM. The result is a video segment in which an ordinary person may not detect that a crime will occur, but we are sure that the sample comes from a video where criminal activity was present.

Experiments and Results

3DCNN is a recent approach for Spatio-temporal analysis, showing a remarkable performance by processing videos in different areas, such as moving objects action recognition , gesture recognition and action recognition. We decided to implement 3DCNN in a more challenging context, such as the search for patterns in criminal samples, which lack suspicious and illegal visual behavior. In this section, we present the proposed experiments and their results. The initial experiment aims at exploring the impact of different values for the parameters of the system. The second experiment focuses on obtaining statistical support that the best configurations obtained from the first experiment are useful for further testing in different situations.

Table 8: Confusion Matrix of best results

Dataset: **unb_60s120n_30t_10f_80x60**
Accuracy: **92.5%**

	Suspicious	Normal	Accuracy
Suspicious	18	0	100%
Normal	4	32	88.9%

Dataset: **bal_240_30t_10f_80x60**
Accuracy: **91.6%**

	Suspicious	Normal	Accuracy
Suspicious	36	0	100%
Normal	6	30	83.3%

Dataset: **unb_60s120n_30t_10f_80x60**
Accuracy: **90.7%**

	Suspicious	Normal	Accuracy
Suspicious	18	0	100%
Normal	5	31	86.0%

Dataset: **bal_240_30t_10f_80x60**
Accuracy: **90.2%**

	Suspicious	Normal	Accuracy
Suspicious	36	0	100%
Normal	7	29	80.6%