AI Video Surveillance System

Exploratory Data Analysis Report



# UCF Crime Dataset Analysis

## 1. Dataset

The dataset used in this project is the UCF Crime Dataset, a large-scale dataset for anomaly detection in surveillance videos. It contains real-world crime footage spanning multiple categories. The dataset consists of:

* Total Videos: 1,500+ videos
* Categories: 14 types of crimes (Arson, Assault, Burglary, Explosion, Fighting, etc.)
* Data Type: Video files (MP4 format)
* Frame Rate: 30 fps
* Resolution: 240x320

Each video is labeled according to the crime category, making it suitable for supervised learning approaches.



## 2. Pre-processing Techniques

#### 2.1 Video to Frames Conversion

#### Processing raw video files directly is computationally expensive and inefficient. To make analysis more structured and manageable, we convert videos into individual frames. This allows frame-wise feature extraction, enabling the application of various image-processing techniques for motion and object analysis. We use OpenCV’s cv2.VideoCapture() to read each video and extract frames at every 15th interval. This frame skipping reduces redundancy, ensuring that the dataset remains computationally efficient while still capturing relevant motion and scene changes.

Relevance to the Project:

* Frame extraction enables the application of deep learning models that work on image-based inputs.
* The extracted frames are to be used for motion detection, object detection, and anomaly classification at the frame level.
* The preprocessed frames serve as the input for Vision Transformer (ViT)-based models, which analyze spatial and temporal patterns in crime detection.

#### 2.2 Optical Flow for Motion Analysis

#### Optical flow helps in capturing movement patterns between consecutive frames, which is crucial in crime detection. Criminal activities often involve sudden movements, group actions, or irregular behaviors, which can be identified through motion analysis. We use the Farneback Optical Flow algorithm (cv2.calcOpticalFlowFarneback) to compute motion vectors between consecutive frames. This generates a motion heatmap, where areas of high intensity indicate strong movement regions.

Usage in Model Training:

* The motion intensity maps will serve as an additional input channel to the model, helping in recognizing unusual activities.
* Instead of relying solely on object detection, the flow-based features help in detecting subtle and unexpected movements that could indicate crime.
* The ViT model will integrate these motion representations along with raw image frames to improve classification accuracy.

#### 2.3 Object Detection using YOLO

#### Understanding the entities present in a scene (e.g., people, vehicles, weapons) is essential for context-aware crime detection. For example, an empty street vs. a crowded fight scene will have vastly different risk levels. We use the YOLOv8 pre-trained model to detect objects in frames. The model identifies and classifies objects such as humans, vehicles, and other relevant elements by overlapping bounding boxes. This information is extracted as feature maps that represent the detected entities and their locations.

Relevance to Model Training:

* The detected object features will be fused with motion analysis to create a multi-modal input for crime classification.
* If an unusual object (e.g., a weapon) or an unusual crowd behavior (e.g., multiple people running) is detected, the system can flag it as suspicious.
* YOLO’s bounding box coordinates and class labels will be used as additional input features for the ViT model.

## Challenges in Preprocessing

While preprocessing the UCF Crime Dataset, several challenges were encountered:

##### Frame Redundancy & Selection

* + Extracting all frames increases storage and processing time. Choosing optimal frame intervals (every 15th frame) balances efficiency and performance.

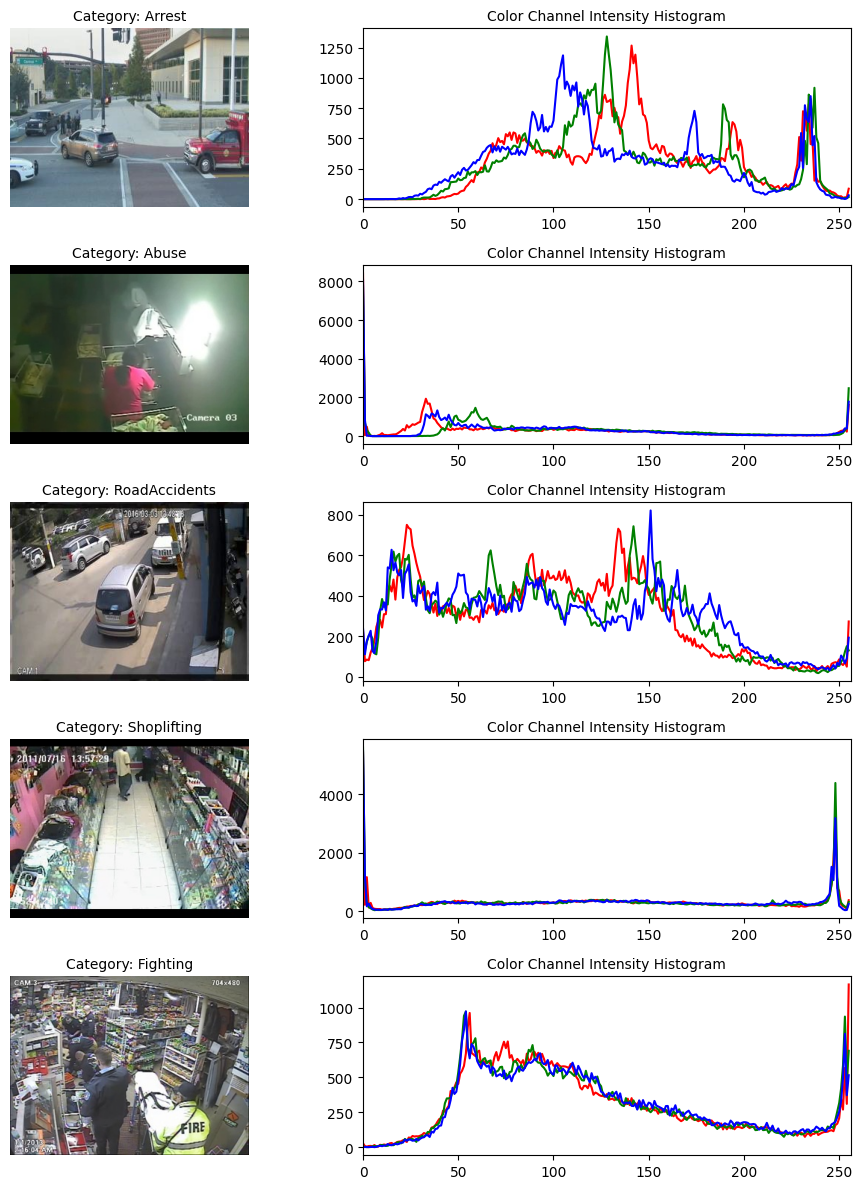
##### Computational Cost

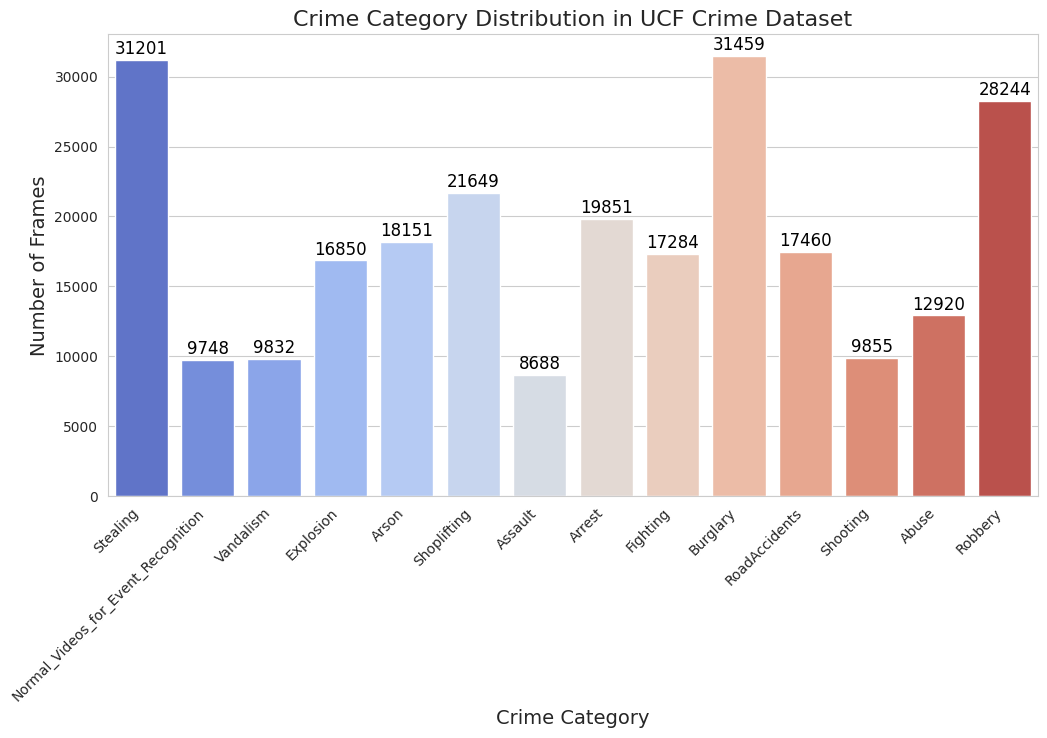
* + Processing optical flow, YOLO, and motion tracking for every frame is computationally expensive. Strategies such as frame skipping (every 15th frame) and batch-wise processing helped optimize runtime. Also used GPUs for processing

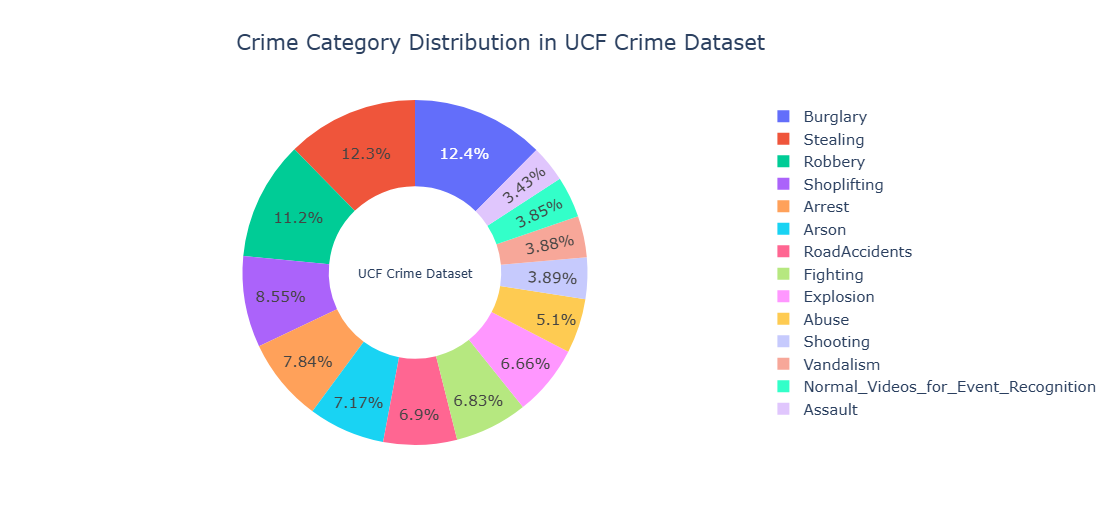
## 3. Visualizations

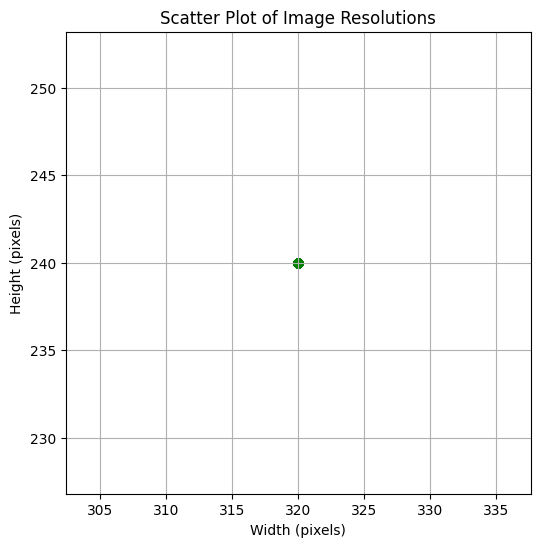
Below are some key visualizations from the dataset processing:

* Color ChanneL Intensity histograms
* Class Distribution: A bar chart showing the frequency of different crime categories.
* Dataset Resolution Distribution
* Object Detection Results: Annotated frames showing detected objects in crime scenes.
* Optical Flow Visualizations: Flow fields highlighting movement patterns.



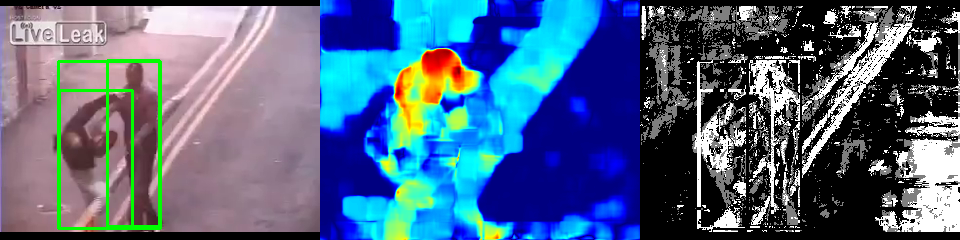


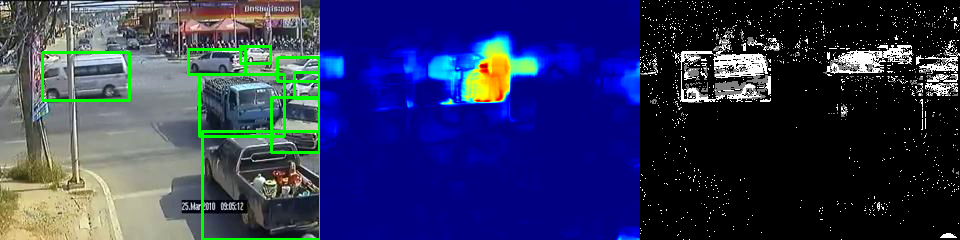




### Preprocessing Visualization:

3 techniques explored. The first image is objects identified through YOLO. second image is motion heatmap using optical flow and third image is obtained through background subtraction





## 

## 

## 4. Future Work

Moving forward, the cleaned and preprocessed dataset will be used for training a Vision Transformer (ViT) model for crime detection. The next steps include:

* Fine-tuning ViT: Pre-training and fine-tuning on UCF Crime dataset.
* Data Splitting: Using an 80-10-10 split for training, validation, and testing.
* Model Evaluation: Measuring performance using accuracy, precision, recall, and F1-score.

The extracted features from frame images, motion heatmaps, and object detection will be combined into a multi-channel input representation. The ViT model will be fine-tuned on these preprocessed inputs to:

* Recognize crime-related movements (from optical flow).
* Identify relevant entities (from YOLO).
* Learn spatial and temporal patterns from frame sequences.

By integrating these preprocessing techniques, the model leverages both spatial and motion-based information to enhance crime detection accuracy. By fine-tuning ViT, we aim to improve the accuracy of crime classification and detection in real-world scenarios. Will try to deploy a 2 layered approach that is As part of future work, we plan to implement a two-layered crime detection approach to enhance accuracy and flexibility. The first model will analyze CCTV footage to determine if any crime is occurring. If a crime is detected, a second model will classify the specific type of crime. This method allows for adjustable detection thresholds, making the system more adaptable to different scenarios while balancing accuracy and computational efficiency.