Detecting Motorcycle Crime Gangs in CCTV Video Footage Using YOLOv9 and CNN

1st Versa Syahputra Santo
School of Computing
Telkom University
Bandung, Indonesia
versasyahptr@student.telkomuniversity.ac.id

2nd Gamma Kosala
School of Computing
Telkom University
Bandung, Indonesia
gammakosala@telkomuniversity.ac.id

3rd Rifki Wijaya
School of Computing
Telkom University
Bandung, Indonesia
rifkiwijaya@telkomuniversity.ac.id

***Abstract*— Street crime, such as criminal motorcycle gangs, has become a severe problem that concerns Indonesian society, especially for those who live in big cities. Criminal Motorcycle gang members often disturb the surrounding community. One of the activities of motorcycle gangs that often cause disturbance is the activity of street convoys using motorbikes. Various solutions to reduce motorcycle gang crime have been conducted by the authorities. One of them is patrolling. However, this effort is less efficient due to limited time, workforce, and coverage of the area that can be monitored. Another preventive solution is to install CCTV. This solution also requires human resources to monitor CCTV footage. This certainly increases the possibility of human error. Some studies have been conducted to automate CCTV surveillance by detecting anomalies or crimes. In this research, motorcycle gang detection consists of three stages. The process begins with detecting and tracking motorcycles in the video using YOLOv9, which achieves an AP50 of 93.2%, alongside ByteTrack. The second step involves mapping each motorcycle's center coordinates to represent each motorcyclist's motion patterns. Finally, these patterns are examined and classified using CNN to detect motorcycle gangs. This method achieves 93.4% accuracy in detecting motorcycle gangs' presence in a video.**

Keywords*— *Video Classification, Object Tracking, Motion Analysis, CCTV, CNN

I. INTRODUCTION

Street crime is a severe problem in Indonesia, especially in big cities. One of the street crimes that disturbs the community is the crime of motorcycle gangs. Motorcycle gangs are criminal groups engaged in various criminal activities, including aggression, property destruction, robbery, and even murder [1]. These groups often exhibit aggressive driving behaviors, such as speeding and reckless maneuvers, which pose significant risks to public safety and create a sense of fear and insecurity among residents.

Various prevention efforts have been made by authorities, such as the police and local governments, to prevent motorcycle gang crimes. One of the efforts made by the police is regular patrol activities, but this solution is inefficient due to limited time, personnel, and coverage area. Another solution local governments take is installing CCTV (Closed Circuit Television) to monitor the situation on the road. Some regions in Indonesia, especially big cities, have installed many CCTVs that are monitored through a centralized control center. However, this solution has several disadvantages. First, it requires many personnel to monitor hundreds of CCTV video broadcasts. Second, the possibility of detecting anomalies such as motorcycle gangs will decrease as the number of CCTV video broadcasts increases [2].

Many studies have been conducted to automate CCTV surveillance to detect crimes. One is research conducted by Atif Jan and Gul Muhammad Khan [2] with the Quasi-3D method to detect crimes in videos. The research aims to create a system to detect malicious video events using a modified CNN (Convolutional Neural Network) filter. Event detection in videos generally uses a 3-dimensional CNN as a feature extractor. However, this method has a drawback: the large number of parameters. In the article, the author separates a 2-dimensional CNN filter to learn spatial features on video frames with a CNN filter to learn temporal features between frames. After the features are extracted, the video will be classified based on its classes: normal, fighting, shooting, and vandalism.

In contrast to previous researchers who used supervised learning methods, Samir Bouindour et al. [3] used an unsupervised learning method for anomaly detection in videos. In that study, researchers only used normal event videos as training data. They used a modified pretrained 3D residual network to extract spatio-temporal features. They proposed a new method to detect outliers based on the output vector of the 3D residual network. This method can automatically select the vector of interest to distinguish between rare and anomalous events, thus reducing false alarms.

Similar to previous research, Fath U Min Ullah et al. [4] created a system to detect crime in videos by utilizing spatiotemporal features with 3D CNN. The system consists of three stages in detecting crime. First, it detects human objects in the input video using a lightweight CNN model such as MobileNet-SSD to obtain the human bounding box and discard useless frames. Next, the human-detected frames will be fed into a 3D CNN to obtain spatio-temporal features. Finally, these features will go to the softmax classifier to classify whether there is a crime in the video.

Furthermore, Sardar Waqar Khan et al [5] utilized CCTV to detect anomalous events on the highway in the form of traffic accidents. They used Deep Learning CNN to detect anomalous events in images from videos. Unlike previous studies that used the video's spatiotemporal (space and time) features, this study focuses on using spatial features in the image to detect anomalies. The system's input is a video split frame by frame into images. Then, the image will be classified using the CNN model. If an accident is detected, the system will send an accident notification to the authorized officer.

While these methods effectively detect individual actions, they must be improved when applied to more complex scenarios like motorcycle gang activities. In such cases, the