Robust Automatic Motorcycle Helmet Violation Detection for an Intelligent Transportation System

Duong Nguyen-Ngoc Tran, Long Hoang Pham, Hyung-Joon Jeon,
Huy-Hung Nguyen, Hyung-Min Jeon, Tai Huu-Phuong Tran, Jae Wook Jeon*

Department of Electrical and Computer Engineering,
Sungkyunkwan University, Suwon, South Korea
{duongtran, phlong, joonjeon, huyhung91, hmjeon, taithp, jwjeon}@skku.edu

Abstract

Video surveillance-based automatic detection of motorcycle helmet usage can enhance the effectiveness of educational and enforcement initiatives aimed at boosting road safety. Current detection methods, however, have room for enhancement, such as the inability to pinpoint individual motorcycles or differentiate between drivers and passengers in terms of helmet usage. This paper introduces a framework designed to detect and identify individual motorcycles while recording specific helmet usage for riders. The proposed classification approach for helmet usage demonstrates increased efficiency in comparison to previous research. Our findings highlight the exceptional accuracy of deep learning, with our method achieving a score of 0.7754 on the AI City 2023 Challenge Track 5 public leaderboard.

1. Introduction

The implementation of automated detection systems for motorcycle helmet usage through video surveillance has the potential to significantly enhance the efficacy of educational and enforcement campaigns, thereby contributing to increased road safety. Despite the potential benefits of these systems, several aspects of existing detection approaches still require improvement. For instance, current methods often need help to localize individual motorcycles within the field of view accurately, and they may fail to differentiate between drivers and passengers regarding helmet usage effectively.

In light of these challenges, this paper puts forth a comprehensive framework that aims to detect and identify individual motorcycles while concurrently registering rider-specific helmet use. By addressing the limitations mentioned above, our proposed method significantly improves the accuracy and effectiveness of helmet-use detection in real-world settings.

Through rigorous evaluation and comparison with earlier studies, we demonstrate that our helmet-use classification approach significantly enhances the efficiency of detection practices. By developing a more sophisticated and precise system for monitoring motorcycle helmet usage, we hope to bolster the overall impact of safety campaigns and ultimately contribute to a safer environment for all road users.

Thus, we introduce an innovative approach for automatically detecting helmet use in motorcyclists, leveraging the power of deep learning technology. Our method incorporates all the critical components of human-observer helmet use registration, including object detection and rider differentiation, to achieve the highest levels of accuracy and reliability. By leveraging the latest advances in deep learning, we have developed a comprehensive framework [23] that considers a wide range of variables and factors that affect helmet detection, including lighting conditions, camera angles, and rider positions. Our approach builds on and extends a previous framework, taking advantage of new techniques and algorithms to achieve even greater accuracy and precision. By incorporating state-of-the-art technologies and best practices, we have created a powerful tool for detecting helmet use in various scenarios and settings. We are confident that our approach will prove to be an invaluable resource for promoting motorcycle safety and reducing the number of preventable injuries and fatalities.

In summary, the main contributions of this paper are summarized as follows:

- We illustrate the detector and identifier for finding the helmet violation
- We introduce the data processing for improve the accuracy of framework.
- The comprehensive experiments show the efficiency of

the framework.

The rest of this paper is organized as follows. In Section 2, the related works review some method impact on the framework. The detail of the proposed method is presented in a detailed description in Section 3. In Section 4, the experiments show qualitative and benchmark results of the proposed method. Conclusions are mentioned in Section 5.

2. Related Work

In reference [7], the authors present a thorough review of the research landscape concerning motorcycle detection, focusing on both traditional techniques and deep learning approaches. Various methods have been proposed for helmet detection, relying on either video or image data. These approaches can be broadly classified into two main categories: those that employ traditional techniques and those that utilize deep learning methods.

Traditional methods [4, 5, 12, 17, 18, 20] primarily follow a similar approach, focusing on identifying moving objects and subsequently classifying helmet violations. Moving object detection typically involves several stages. First, motion segmentation techniques extract moving objects from surveillance footage. Common motion segmentation methods include optical flow, frame difference, and background subtraction [22, 24]. Second, hand-crafted feature descriptors, such as local binary pattern (LBP), histogram of oriented gradient (HOG), and scale-invariant feature transform (SIFT), are employed to extract features of motorcycles and other vehicles. Finally, binary classifiers, like support vector machine (SVM) and K-nearest neighbor (KNN), are used to classify motorcycles. Once the moving objects are identified, the driver and passenger stages are categorized. Silva et al. [4] broke down the problem of detecting helmet usage by motorcyclists into two steps. The first step involves segmenting and classifying vehicle images to detect all moving objects in the scene. The second step entails helmet detection, employing a hybrid descriptor to extract image features and a support vector machine classifier to distinguish between helmeted and non-helmeted images. Dahiya et al. [17] initially detected bike riders using background subtraction and object segmentation from surveillance footage. They then determined helmet usage based on visual features and binary classification. Similarly, Talaulikar et al. [18] utilized background subtraction techniques to identify moving vehicles and applied principal component analysis (PCA) to the extracted features. The limitations of such methods include: achieving real-time speed is challenging due to the multi-stage operation; accurately determining whether a person is not wearing a helmet becomes difficult when a motorcycle has multiple riders, particularly if the individual without a helmet is partially obscured by another rider wearing one, as classification algorithms struggle in this scenario; the performance of motion segmentation-based approaches is significantly impacted by factors such as road congestion, camera shake, branch movement, or other disturbances.

In recent years, deep learning-based methods have been proposed by researchers. In [15], the background subtraction method and the SMO classifier are employed to detect motorcycles from videos. Hand-crafted features and CNN are then used to classify helmets and no helmets, respectively, with CNN demonstrating higher accuracy than manual features. In [25], adaptive background subtraction extracts moving objects from video frames, and CNN is used to classify motorcyclists within these objects. CNN is further applied to classify the top quarter area of motorcycles to identify helmetless riders. In [27], the Gaussian mixture model (GMM) segments foreground objects, which are then labeled. A faster region-based CNN (faster R-CNN) detects motorcycles within the labeled foreground objects, ensuring the presence of motorcyclists. The faster R-CNN is also used for helmet detection. Although helmet detection in [15, 25, 27] uses deep learning, traditional background subtraction remains the technique for obtaining foreground targets in the motorcycle detection stage, which performs poorly in crowded scenes. In [10], algorithm is proposed for detecting helmet usage by motorcyclists, but motorcycle detection is not reported. In [13, 14], the YOLOv3 algorithm detects motorcycles and people in images, and the overlapping area of their bounding boxes is used to identify riders. Finally, the YOLOv3 algorithm detects helmet usage. However, in traffic monitoring, motorcyclists and motorcycles are highly overlapping, making separate detection of motorcyclists unnecessary. In [3, 6], SSD or YOLOv3 algorithms detect motorcycle areas, extract the upper part of the image, and employ classification algorithms to identify helmets and non-helmets. This approach becomes ineffective when multiple people are on a motorcycle. In [1, 11, 16], motorcycles and motorcyclists are treated as a single entity, and CNN models directly detect whether a rider is wearing a helmet. This one-step coarse-grained detection method exhibits low accuracy.

This paper presents a deep learning-based approach to detect and classify motorbike drivers and passengers, focusing on achieving high accuracy in Helmet Violation Detection. Our proposed framework utilizes a two-stage process consisting of a Detector and an Identifier, which work together to enhance the system's overall effectiveness. The Detector stage identifies motorcycles and their riders within the scene, ensuring accurate initial detection. Subsequently, the Identifier stage classifies the detected objects, determining whether the drivers and passengers are wearing helmets or not. This two-stage process allows for a more robust and comprehensive scene analysis, reducing false positives and

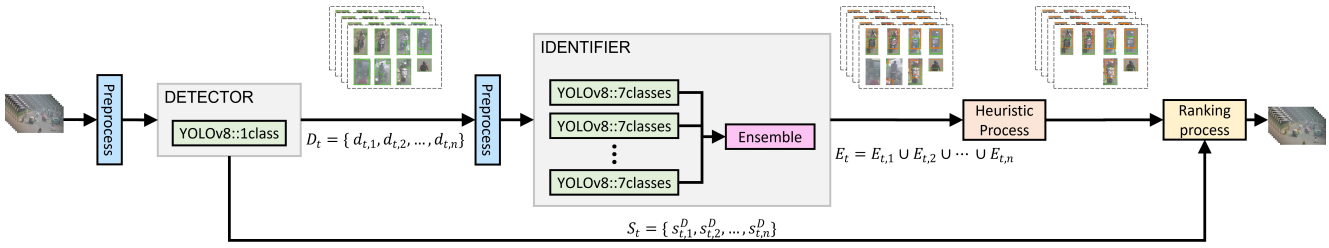

Figure 1. The pipeline of the framework includes two main stages and two processes. Initially, we employ the Detector to identify the locations of motorbikes along with their drivers and passengers. We then crop the bounding boxes and forward them to the Identifier for further information extraction. Obtaining information involves detecting motorbikes and individuals on them, either wearing helmets or not. The framework assigns confidence scores to the information gathered from the detector and identifier stages, allowing us to filter out incorrect objects and retain the correct ones for the final outcome.

negatives and improving overall performance. The experimental results obtained from this study serve as evidence of the framework's efficiency, highlighting its potential for real-world applications in traffic surveillance and safety enforcement. The combination of deep learning techniques and the two-stage approach demonstrates a powerful solution to the challenges associated with motorbike and helmet detection, paving the way for future advancements in this field.

3. Methodology

In this section, we illustrate the framework we use for detect the violation the helmet rule. Because of the various scene of both training and testing datasets, there are some challenges in the traffic surveillance system (as shown in Figure 3). Firstly, the change in illumination by daytime and nighttime leads to false detection (e.g., the shadow along with the motor in the clear sky weather and the headlight of both motor and car). Secondly, the haze weather will blur the object in the whole scene. From one camera perspective, we have to detect the violation of wearing a helmet by the motorbike driver and the passengers on it. Therefore, the possibility of detecting the whole image at once lowers the accuracy.

At the outset of our research, we aimed to develop a model that would enable the simultaneous detection of the driver, passenger, and motorbike in a given image. However, we encountered a significant challenge that had the potential to adversely impact the accuracy of the model. Despite the increased complexity of the model, we observed that it was unable to accurately localize all the driver and motorbike components. Subsequently, we undertook a more comprehensive analysis of the model's performance and discovered that it was more effective at detecting the entire motorbike, including the person riding it, as a single object. Based on this observation, we revised our approach to focus on first localizing the entire motorbike with a person on it, and then identifying the individual components of the object, including the motorbike, driver, and passenger, and whether or not they were wearing a helmet.

3.1. Pipeline

The pipeline of the Automatic Motorcycle Helmet Violation Detection framework is shown in Figure 1. As can be seen, in the beginning, we use the detector to find the location of the motorbike with the driver and passengers on them. We crop the bounding boxes and send them to the identifier to extract the information on them. We run obtaining the information process by finding the motorbike and the person on it with or without a helmet. The framework ranks the information by the confidence score of the detector and the identifier; we can filter out the wrong object and keep the right one for the final result.

3.2. Detector

In order to accurately detect motorbikes within a given scene, we employ the YOLOv8 [9] algorithm, specifically its largest version (at the time we write framework, YOLOv8x6 is the best one). This state-of-the-art object detection model has proven to be highly effective in identifying various objects, including motorbikes, within complex environments. Upon detecting motorbikes in the scene, YOLOv8 provides not only the cropped images of the bounding boxes containing the motorbikes but also the corresponding coordinates on the input image.

$$D_t = \{d_{t,1}, d_{t,2}, ..., d_{t,n}\} \tag{1}$$

where detected bounding box $d_(t, i)$ has the values:

$$d_{t,i} = \{x^D_{t,i,c}, y^D_{t,i,c}, w^D_{t,i}, h^D_{t,i}, s^D_{t,i}\} \tag{2}$$

where $\{x^D_{t,i,c}, y^D_{t,i,c}\}$ represents the center point, $\{w^D_{t,i,c}, h^D_{t,i,c}\}$ illustrates the width and height, and $s^D_{t,i}$ is the confidence score of the bounding box of the Detector at index i in time t. We can consider the confidence

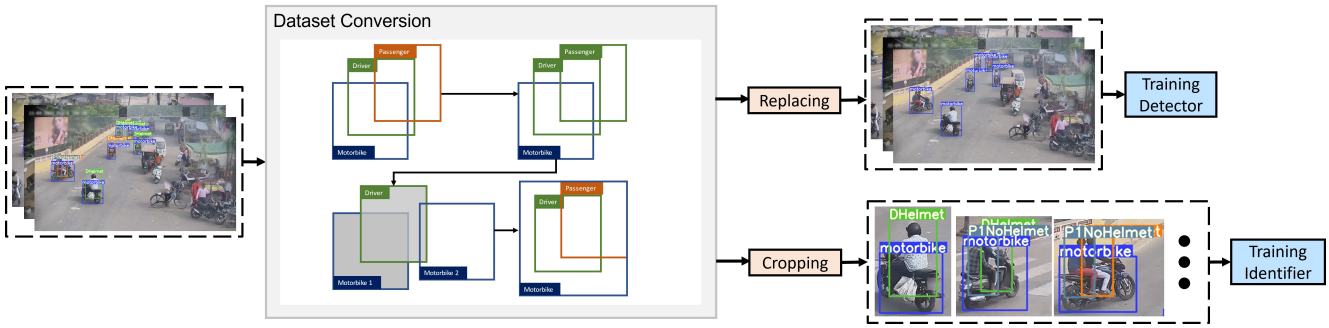

Figure 2. The diagram of data conversion for training both the Detector and Identifier. The input is the groundtruth of the given dataset with 7 classes and 1920x1080 resolution; we convert them into two new datasets, including 1 class with 1920x1080 image size and 7 classes with the cropped image.

score of each bounding box is the first rank of the object in the final result. We have the list of confidence score:

$$S_t = \{s^D_{t,1}, s^D_{t,2}, ..., s^D_{t,n}\} \tag{3}$$

This information, which includes the detection scores and filtered results, is vital for the subsequent stages of the framework. It enables precise tracking and classification of motorbike riders and their helmets, allowing the system to assess compliance with helmet regulations accurately. By incorporating this information into the tracking and classification processes, the framework can maintain high accuracy while minimizing false positives and negatives, leading to more reliable and efficient helmet violation detection. Furthermore, filtering results based on user-defined criteria can improve the system's overall performance by focusing on the most relevant objects and events within the scene. This targeted approach increases the system's effectiveness in identifying helmet violations and conserves computational resources, making it more suitable for deployment in real-world traffic monitoring applications.

We prune each image bounding boxes result as input of the Identifier. The cropped image is small enough and does not include any motor or overlap as less as possible.

3.3. Identifier

Upon obtaining the results from the detector stage, we proceed to employ YOLOv8 [9] as the identifier to differentiate between motorbikes, drivers, and passengers. This continuation with YOLOv8 ensures that the identifier is capable of effectively analyzing and classifying the distinct elements within the input images. Process of each bounding box result $d_{t,i}$ of the Detector give us the result of the Identifier:

$$E_{t,i} = \{e_{t,i,1}, e_{t,i,2}, ..., e_{t,i,m}\} \tag{4}$$

where detected bounding box $e_{t,i,j}$ has the values:

$$e_{t,i,j} = \{x^E_{t,i,j,c}, y^E_{t,i,j,c}, w^E_{t,i,j}, h^E_{t,i,j}, s^E_{t,i,j}\} \tag{5}$$

where $\{x^E_{t,i,j,c}, y^E_{t,i,j,c}\}$ represents the center point, $\{w^E_{t,i,j,c}, h^E_{t,i,j,c}\}$ illustrates the width and height, and $s^E_{t,i,j}$ is the confidence score of the bounding box of the Detector at index i, j in time t. We can consider the confidence score of each bounding box is the second rank of the object in the final result. After getting the result, we continue to save the result with the new confident score, which is considered as the second rank of objects.

$$S_{t,i} = \{s^D_{t,i,1}, s^D_{t,i,2}, ..., s^D_{t,i,m}\} \tag{6}$$

After the whole process of the Identifier, we get the second rank of each object in cropped images.

$$E_t = E_{t,1} \cup E_{t,2} \cup ... \cup E_{t,n} \tag{7}$$

We can use them for the filter in the following process, in which the object is kept or not based on the threshold.

3.4. Ranking

To minimize the number of false positives in our results, we have implemented a filtering mechanism that targets lower rank objects. The ranking system is based on two key factors: the first rank, which is generated by our detector algorithm and reflects the likelihood that an object is present in the image, and the second rank, generated by our identifier algorithm and indicates the confidence level that the object belongs to a particular class. We obtain the object's overall rank by multiplying these two ranks.

$$r_{t,i,j} = s^D_{t,i} \cdot s^E_{t,i,j} \tag{8}$$

To guarantee the accuracy and precision of our results, we have incorporated a threshold-based filtering mechanism

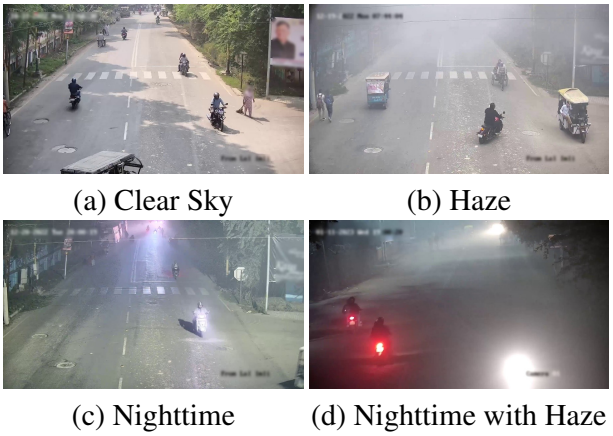

Figure 3. Visualization of various outside environment on one location in dataset.

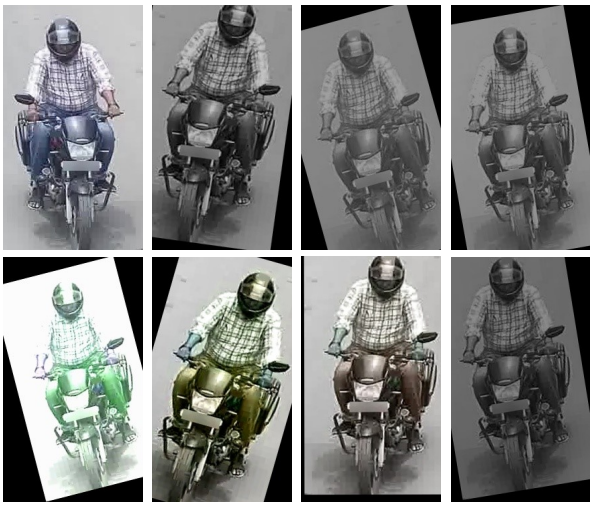

Figure 4. Example of data augmentation in training dataset for Identifier.

that eliminates any bounding boxes with scores falling below a specified level. This approach serves as a critical tool in maintaining the integrity of our object detections by effectively removing potential false positives and retaining only the most relevant and accurate detections. By implementing this threshold-based filtering, our helmet violation detection framework can better focus on the most pertinent objects within a given scene, consequently improving the overall performance and reliability of the system. We ascertain the final position of the motorbike and individual by employing D_t and E_t. This approach is especially beneficial in complex traffic scenarios where numerous objects are.

3.5. Data Processing

3.5.1 Data Conversion

In order to effectively develop the detector and identifier components of our framework, it is necessary to create two separate training datasets tailored to the specific requirements of each stage. These datasets must be generated from the original AI City Challenge Track 5 dataset through a transformation process that yields two new datasets, each featuring distinct labels and image dimensions. The transformation process is crucial for optimizing the performance of the detector and identifier stages, as it ensures that each component is provided with the most relevant and suitable data for its respective task. By carefully modifying the original dataset to produce these two distinct training sets, we can effectively train each stage to deliver optimal results in the context of the overall framework. The process of transforming the original dataset involves adjusting the labels to reflect the specific objectives of each stage but also resizing the images to accommodate the input requirements of the detector and identifier. By doing so, we ensure that our framework can efficiently process and analyze the input data, ultimately leading to more accurate and reliable helmet violation detection in real-world traffic scenarios.

For the detector method, it is essential to have labels with rectangles that encapsulate the entire motorbike, driver, and passenger. As a result, it is necessary to transform the initial dataset, which contains 7 classes, into a detector training dataset consisting of just 1 class. The process of converting the original dataset to suit the detector's requirements is demonstrated in Figure 2.

On the other hand, the identifier method relies on training with cropped images, which necessitates matching the image size to the bounding box dimensions outputted by the detector. To accomplish this, the original dataset, comprised of full-size images, must be transformed into an identifier dataset featuring smaller-sized images. Figure 2 illustrates the transformation process required to generate the identifier dataset, which is specifically tailored to the identifier method's input and training needs.

By creating these two separate training datasets, we can effectively train the detector and identifier for their respective tasks. The detector will be optimized for detecting motorbikes along with their drivers and passengers, while the identifier will be fine-tuned for recognizing and differentiating between specific instances of motorbikes and their riders. This combination of methods allows for a more robust and accurate system in addressing the unique challenges of the AI City Challenge 2023 Track 5.

3.5.2 Data Augmentation

As illustrated in Figure 3, the camera perspective encompasses a diverse array of outdoor environments, encompass-

Model	Image size	mAP
YOLOv8-6e6	1280	0.3814
YOLOv8-6e6	1536	0.3917
YOLOv8-6e6	1920	0.3823

Table 1. Ablation study of image size in the Detector.

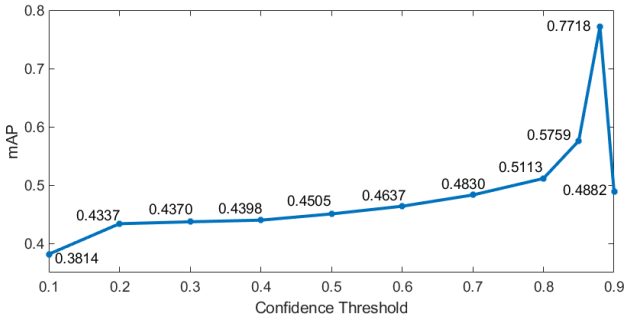

Figure 5. Ablation study of confidence score in Final rank.

ing clear skies, nighttime settings, and adverse weather conditions such as haze. To enhance the model's accuracy and mitigate overfitting, we have employed a dataset augmentation strategy [2] (as shown in Figure 4). This method generates additional training examples by applying transformations and perturbations to the original dataset, including rotations, flips, and alterations in lighting and contrast.

By augmenting the dataset in this fashion, we can expose the model to an expanded range of scenarios and conditions, ultimately bolstering its robustness and generalization capabilities. Rigorous experimentation and analysis have verified the effectiveness of this strategy, and we are confident that it offers significant potential for future research in this domain. The incorporation of dataset augmentation improves the model's performance in diverse real-world conditions and contributes to the ongoing advancement of object detection and classification techniques.

4. Experiments

4.1. Implementation Details

4.1.1 Training Phase

In the training phase, we must train both the Detector and the Identifier simultaneously. Because there are two distinguished datasets after conversion, and the time for the challenge is constrained, we have to carefully select the hyperparameter for training to get a high score.

In the detector stage, despite the original video resolution being 1920, we observed that training with resolutions of 1280, 1536, and 1920 achieved nearly the same accuracy for detecting entire motorbikes. Consequently, to optimize time efficiency, we decided to train our model at a resolution

Trainning Size						Inference Size	mAP
256	320	384	448	512	576		
	✓	✓	✓	✓		384	0.5861
	✓	✓	✓	✓		448	0.5888
	✓	✓	✓	✓		512	0.7269
	✓	✓	✓	✓	✓	512	0.7754
✓	✓	✓	✓	✓	✓	512	0.7718

Table 2. Ablation study of ensemble in the Identifier.

of 1280 using the largest version of YOLOv8 [9]. Another reason for choosing this resolution is the considerable size of the training dataset, which has been expanded due to the augmentation process. By selecting a lower resolution for training without compromising detection accuracy, we are able to reduce computational complexity and training time.

In order to obtain detailed information about the individuals on motorbikes for the Identifier stage, we employ an ensemble method [19] to enhance the accuracy of our model. Since we are working exclusively with cropped images derived from the bounding box results of the Detector, we need to select the training resolutions carefully. For this purpose, we consider several resolutions, including 256, 320, 384, 448, 512, and 576. The ensemble approach combines multiple models' outputs, improving the overall predictive performance and mitigating the risk of overfitting.

Moreover, we have trained the original dataset with 7 classes for object detection over images to compare results with the proposed framework.

4.1.2 Testing Phase

In the testing phase, according to the guideline of AI City Challenge 2023, we extract the whole image from all videos by using FFmpeg [21] to maintain both the time frame and quality of the image while inferring.

In the detection stage, we employ the YOLOv8 model, which has been trained for 189 epochs. The confidence threshold is set to a low value of 0.1, and the inference resolution is 1280. By keeping the confidence threshold low, we are able to retain a large number of bounding boxes for subsequent processing in the next stage.

For the identification stage, we experiment with various combinations of YOLOv8 models to determine the most suitable ensemble configuration (as demonstrated in Table 2). Our experiments range from using a single model to multiple models and from lower to higher resolutions. The confidence threshold for the identifier is set even lower at 0.1. Once the identification process is complete, objects are filtered out based on their confidence scores, which helps enhance the overall accuracy of the final results. This ap-

Figure 6. Visualization of results with 35 different scenes of 100 videos of the test set. Various scenes have different outside environments, including overcast, clear-sky, haze, and nighttime.

proach ensures that our framework delivers optimal performance while maintaining high precision in detecting and classifying motorbike drivers and passengers.

Moreover, we also run the YOLOv8 model for detecting all defined objects for comparison with our framework.

4.2. Datasets

In developing countries such as India, motorcycles are a popular means of transportation, but riders face a higher risk of accidents due to their lack of protection. To promote traffic safety, it is mandatory for motorcycle riders to wear helmets, and strict enforcement is necessary. A training dataset comprising 100 videos, each 20 seconds long and recorded at 10 fps with a resolution of 1920x1080, includes bounding box annotations for each motorcycle and its up to three riders, specifying whether they are wearing helmets or not.

4.3. Evaluation Metrics

The Average Precision (AP) is calculated by plotting a precision-recall curve for each object class and then computing the area under the curve (AUC), as defined in PASCAL VOC 2012 competition [8]. The curve is created by varying the confidence thresholds of the predicted bounding boxes. A higher AUC indicates a better performance of the model in detecting objects of that particular class. The mean Average Precision (mAP) is then computed by taking the mean of AP values across all object classes. This provides a single, unified score that can be used to compare the performance of different object detection models. Higher mAP values indicate better overall performance in detecting objects across all classes. It is important to note that mAP is sensitive to the choice of IoU thresholds, so when comparing different models, it is essential to use the same set of thresholds for a fair comparison.

4.4. Ablation Study

This section presents an ablation study to demonstrate the efficiency of the proposed framework.

Initially, we examine the impact of image size on both the detector and identifier processes. We trained and tested the detector model at resolutions of 1280, 1536, and 1920. As indicated in Table 1, the results are nearly identical across these resolutions. Consequently, we opt for a 1280 resolution to enhance speed performance and reduce training time.

In the case of the identifier, we experimented with various image sizes for running a single model, including 256, 320, 384, 448, 512, and 576. Additionally, to boost the identifier's accuracy, we explore the use of ensemble models. As depicted in Table 2, we implemented ensemble models by selecting combinations of different image sizes. This approach allows for further optimization of the framework's performance, ensuring a balance between speed and accuracy while maintaining the system's overall effectiveness in detecting and identifying objects.

We have test server threshold for final rank, as shown in Figure 5. We test by increasing the threshold from 0.1 with step 0.1. As can be seen, the mAP score rises to the peak at 0.8 and 0.9. We use binary search for the range 0.8 and 0.9

Rank	Team ID	Score
1	58	0.8340
2	**33 (Ours)**	**0.7754**
3	37	0.6997
4	18	0.6422
5	16	0.6389
6	45	0.6112
7	192	0.5861
8	55	0.5569
9	145	0.5474
10	11	0.5377

Table 3. Leaderboard of Detecting Violation of Helmet Rule for Motorcyclists. 0.7754 is the final score of Dataset A in 2023 AI City Challenge Track 5.

to find the most suitable threshold with the highest score. Finally, the best one is 0.88. One confidence threshold is not fixed for every dataset. Therefore, depending on the dataset, we must try several confidence thresholds for better accuracy.

At the time of the contest, there were two new YOLO versions, including YOLOv7 [26] and YOLOv8 [9]. According to the training process, the YOLOv8 trained faster than YOLOv7; we chose the faster one for the main backbones.

4.5. Quantitative Result

Table 3 presents the final ranking results for the test sequence, with our result emphasized in bold text. A week prior to the leaderboard's finalization, 50% of the test set was utilized for evaluation, during which our method occupied the top position. Nevertheless, in the last two days leading up to the conclusion of the competition, our performance ranking experienced a decline. The final score registered was 0.7754. Given the unlikelihood that our framework overfits the dataset, a plausible explanation for this performance decrease could be the experimental data's unequal distribution.

5. Conclusions

Video surveillance-based automated detection of motorcycle helmet usage can enhance the effectiveness of education and enforcement initiatives, leading to improved road safety. However, current detection methods have room for growth, including issues with pinpointing individual motorcycles and differentiating between helmet-wearing drivers and passengers. This paper presents a framework to detect and identify separate motorcycles while tracking rider-specific helmet use. Our helmet-use classification approach shows increased efficiency compared to previous studies. Highlighting the high accuracy of deep learning, our technique achieved a score of 0.7754 on the AI City 2023 Challenge Track 5 public leaderboard. In the future, we will add the tracker to the main framework for ensembled information from several frames to better classify each rider.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2021-0-01364, An intelligent system for 24/7 real-time traffic surveillance on edge devices)

References

[1] Narong Boonsirisumpun, Wichai Puarungroj, and Phonratichi Wairotchanaphuttha. Automatic Detector for Bikers with no Helmet using Deep Learning. In *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, pages 1–4, Chiang Mai, Thailand, Nov. 2018. IEEE. 2

[2] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020. 6

[3] Aphinya Chairat, Matthew N. Dailey, Somphop Limsoonthrakul, Mongkol Ekpanyapong, and Dharma Raj K.C. Low Cost, High Performance Automatic Motorcycle Helmet Violation Detection. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3549–3557, Snowmass Village, CO, USA, Mar. 2020. IEEE. 2

[4] Dhwani Contractorr, Ketki Pathak, Sonali Sharma, Shreya Bhagat, and Tanu Sharma. Cascade classifier based helmet detection using opencv in image processing. 05 2016. 2

[5] Kunal Dahiya, Dinesh Singh, and C. Krishna Mohan. Automatic detection of bike-riders without helmet using surveillance videos in real-time. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3046–3051, Vancouver, BC, Canada, July 2016. IEEE. 2

[6] Madhuchhanda Dasgupta, Oishila Bandyopadhyay, and Sanjay Chatterji. Automated Helmet Detection for Multiple Motorcycle Riders using CNN. In *2019 IEEE Conference on Information and Communication Technology*, pages 1–4, Allahabad, India, Dec. 2019. IEEE. 2

[7] Jorge E. Espinosa, Sergio A. Velastin, and John W. Branch. Detection of Motorcycles in Urban Traffic Using Video Analysis: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 22(10):6115–6130, Oct. 2021. 2

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 7

[9] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, 1 2023. 3, 4, 6, 8

[10] Fahad A Khan, Nitin Nagori, and Ameya Naik. Helmet and Number Plate detection of Motorcyclists using Deep Learning and Advanced Machine Vision Techniques. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 714–717, Coimbatore, India, July 2020. IEEE. 2

[11] Hanhe Lin, Jeremiah D. Deng, Deike Albers, and Felix Wilhelm Siebert. Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning. *IEEE Access*, 8:162073–162084, 2020. 2

[12] Valanukonda Lakshmi Padmini, G. Krishna Kishore, Ponnuru Durgamalleswarao, and Parasa Teja Sree. Real Time Automatic Detection of Motorcyclists With and Without a Safety Helmet. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1251–1256, Trichy, India, Sept. 2020. IEEE. 2

[13] C A Rohith, Shilpa A Nair, Parvathi Sanil Nair, Sneha Alphonsa, and Nithin Prince John. An Efficient Helmet Detection for MVD using Deep learning. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 282–286, Tirunelveli, India, Apr. 2019. IEEE. 2

[14] Apoorva Saumya, V Gayathri, K Venkateswaran, Sarthak Kale, and N Sridhar. Machine Learning based Surveillance System for Detection of Bike Riders without Helmet and Triple Rides. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 347–352, Trichy, India, Sept. 2020. IEEE. 2

[15] Linu Shine and Jiji C. V. Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN. *Multimedia Tools and Applications*, 79(19-20):14179–14199, May 2020. 2

[16] Felix Wilhelm Siebert and Hanhe Lin. Detecting motorcycle helmet use with deep learning. *Accident Analysis & Prevention*, 134:105319, Jan. 2020. 2

[17] Romuere Silva, Kelson Aires, Thiago Santos, Kalyf Abdala, Rodrigo Veras, and Andre Soares. Automatic detection of motorcyclists without helmet. In *2013 XXXIX Latin American Computing Conference (CLEI)*, pages 1–7, Caracas (Naiguata), Venezuela, Oct. 2013. IEEE. 2

[18] Romuere R. V. e Silva, Kelson R. T. Aires, and Rodrigo de M. S. Veras. Detection of helmets on motorcyclists. *Multimedia Tools and Applications*, 77(5):5659–5683, Mar. 2018. 2

[19] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107:104117, Mar. 2021. 6

[20] Abhijeet S. Talaulikar, Sanjay Sanathanan, and Chirag N. Modi. An Enhanced Approach for Detecting Helmet on Motorcyclists Using Image Processing and Machine Learning Techniques. In *Advanced Computing and Communication Technologies*, volume 702, pages 109–119. Springer Singapore, Singapore, 2019. 2

[21] Suramya Tomar. Converting video formats with ffmpeg. *Linux Journal*, 2006(146):10, 2006. 6

[22] Duong Nguyen-Ngoc Tran, Tien Phuoc Nguyen, Tai Nhu Do, and Synh Viet-Uyen Ha. Subsequent Processing of Background Modeling for Traffic Surveillance System. *International Journal of Computer Theory and Engineering*, 8(3):235–239, June 2016. 2

[23] Duong Nguyen-Ngoc Tran, Long Hoang Pham, Hyung-Joon Jeon, Huy-Hung Nguyen, Hyung-Min Jeon, Tai Huu-Phuong Tran, and Jae Wook Jeon. A Robust Traffic-Aware City-Scale Multi-Camera Vehicle Tracking Of Vehicles. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3149–3158, New Orleans, LA, USA, June 2022. IEEE. 1

[24] Synh Viet-Uyen Ha, Duong Nguyen-Ngoc Tran, Tien Phuoc Nguyen, and Son Vu-Truong Dao. High variation removal for background subtraction in traffic surveillance systems. *IET Computer Vision*, 12(8):1163–1170, Dec. 2018. 2

[25] C. Vishnu, Dinesh Singh, C. Krishna Mohan, and Sobhan Babu. Detection of motorcyclists without helmet in videos using convolutional neural network. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3036–3041, Anchorage, AK, USA, May 2017. IEEE. 2

[26] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. 8

[27] B. Yogameena, K. Menaka, and S. Saravana Perumaal. Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system. *IET Intelligent Transport Systems*, 13(7):1190–1198, July 2019. 2