Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 9, No. 1, March 2023, pp. 96-106
ISSN: 2338-3070, DOI: 10.26555/jiteki.v9i1.25614

Motorcycling-Net: A Hybrid-Based Classification Method for Recognizing Motorcycling Near Misses

Rotimi-Williams Bello [1], Chinedu Uchechukwu Oluigbo [1], Oluwatomilola Motunrayo Moradeyo [2]
[1] Department of Mathematics and Computer Science, University of Africa, 561101 Sagbama, Bayelsa State, Nigeria
[2] Department of Computer Science, Adeseun Ogundoyin Polytechnic, Eruwa, Oyo State, Nigeria

ARTICLE INFO

Article history:

Received January 10, 2023
Revised February 14, 2023
Published February 15, 2023

Keywords:

Accident;
Computer vision;
Image processing;
Motorcycling;
Near misses

ABSTRACT

This article presents near misses as corrective and preventive measures to safety events. The article focuses on the risk factors of commercial motorcycling near misses, which we address by proposing a near miss detection framework based on a hybrid of YOLOv4-DeepSort and VGG16-BiLSTM models. We employed YOLOv4-DeepSort model for the detection and tracking tasks, and the tracked images and identity information were stored. The sequence of image was fetched into the VGG16-BiLSTM model for extraction of image feature information and near misses recognition respectively. Video streams of near miss datasets containing motorcycling in different scenes were collected for the experiment. We evaluate the proposed methods by testing 444 sequential video frames of motorcycling near misses in urban environment. The detection models achieved 96% accuracy for motorcycle, 89% for car, and 81% for person with lower false-positive rates on the test datasets while the tracking models achieved 34.3 MOTA on the test set and MOTP of 0.77. The results of the study indicate practicality for automatic detection of motorcycling near misses in urban environment, and it could assist in providing resourceful technical reference for analyzing the risk factors of motorcycling near misses. The research contributions are: (1) A hybrid of YOLOv4 and DeepSort model to enhance object detection and tracking in a complex environment and (2) A hybrid of YOLOv4 and DeepSort model to optimize the extraction of image feature information and near misses recognition respectively for overall system performance.

Corresponding Author:

Rotimi-Williams Bello, Department of Mathematics and Computer Science, University of Africa, 561101 Sagbama, Bayelsa State, Nigeria
Email: sirbrw@yahoo.com

1. INTRODUCTION

Although many researchers have come up with different definitions for near misses, but the most suitable definition among them all is defined as an unexpected event that results neither in dangerous hazard, damage, injury, nor death but if ignored or unreported, has the tendency to result to any of them in future. The reporting and investigation of a near miss incident by a detailed accident investigation helps in applying preventive measures to forestall reoccurrence of such incident [1]. Fig. 1 shows a typical implication of ignored or unreported near miss incident. Near miss events are not uncommon, they are more common than sickness, hazard, damage, injury and fatality any statistics may present. According to the Federal Road Safety Corps (FRSC), near miss events precede road safety incidents. The responsiveness of urban planners, practitioners, commercial motorcyclists, drivers, road users and road safety corps to near miss events can be corrective and preventive measures to future safety incidents [2].

According to [1] there is potential ability in the newly emerging urban data sources to handle image-related urban modeling tasks. Utilizing such data for urban scenes analysis helps in understanding the dynamic nature of cities and controlling incidents relating to traffic or overcrowding in such cities. Several attempts

have been made by urban researchers to model cities by multi-agent models using science of complexity and theory of network [2]. The implication of these models, most of the time, is over-simplification of the urban systems' initial settings and cities exploration from a perspective that is one-dimensional [3]. Lack of large-scale and public accessible dataset, open source modifiable code and graphic processing unit (GPU) are mitigating the robustness of these models from feeding simulations that are unusually large in scope. However, the advent of the artificial intelligence based computer vision and image processing algorithms has changed the whole systems by paving the way for data/video analytics and analytical techniques that can handle urban-related problems.

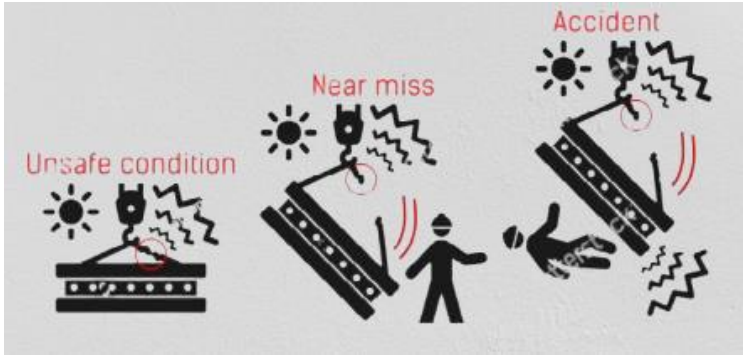

Fig. 1. Unreported or ignored scenarios of near miss incident leading to accident. Source: Internet

The emergence of computer vision and image processing, and their applications have helped in understanding the complexity of the attributes of the city dynamic for prompt detection of motorcycling near misses. Various problems confronting urban settlement can be resolved through large-scale datasets of analyzed digital images, whereby essential information and image elements can be tracked and extracted as if performed by human experts for the overall benefit of transportation industry. Some of the risks caused by the congestion and incidents of traffic to the urban dwellers are due to the nature of road and transport networks put in place. According to [4], to control these ugly incidents, for couple of years, measures were put in place such as safety awareness programmes for sensitizing the urban dwellers on road safety, and monitoring the networks of urban transport using technology-based road signs and traffic lights. However, these measures could not give detail account of the unpalatable consequences such as congestion or incidents of traffic of the agents' behavior such as motorcycles, pedestrians, and automobile, all within the city environment. Having such detail account is essential and beneficial to the motorcyclists, who are mostly exposed to road accidents, with little or no near misses data. Motorcycling is a major occupation among many youths in Nigeria and elsewhere [5]. However, incessant motorcycling road crash is worrisome; a total of 689 people were killed, over 200 injured in 1,500 road crashes involving motorcycles and tricycles between 2015 and 2019 on Lagos roads alone, according to the state government [6].

Although other means of transportation have been encouraged globally to ease challenges of transportation and reduce air pollution caused by automobiles [7], [8], motorcycling has not been favorably considered in that category due to numerous near misses involving motorcycles as a result of little or no formal training received by the motorcyclists regarding rules guiding roads and their usage. In another perspective, motorcycling is perceived as dangerous mode of transportation in that, only a few passengers can withstand the trauma of its near misses or the risk of dodging its crash [9]. Just as found in many places of the world, motorcyclists and their passengers are not likely to get to their destinations without experiencing one form of near misses or another [10], thereby hindering the wider acceptance of commercializing motorcycling as a mode of transportation [11]-[13].

Although most of the near misses are reported, not recording them contributes to the difficulty in accessing their data as information source for investigating and identifying the associating factors responsible for the accidents risked by individual motorcyclists such as visibility, physical conditions of the roads and the motorcycles, mental and psychological state of the motorcyclists and the pedestrians. Camera-trap images and video data of commuting motorcyclists can be a good source of the data and information needed to address the aforementioned tasks, especially the video data, which provide the replica of the original near miss scenes for their features extraction [11], [12], [14], [15], [16]. In all combination, incidents of motorcycling near misses are factors- and events-based, which in actual fact, are not all the time caused by transportation alone.

In-depth understanding of these factors and events will ease the problem of analyzing motorcycling near misses. Due to the small number recorded for near misses, and the impact of prejudice on reporting data

pertaining to road crash, where only the near misses that lead to crash or death are reported while the near misses that do not lead to crash or death are either ignored or under-reported, it is an herculean task to give quantitative analysis of the risk of motorcycling [12], [17], [18]. Too many ignored or unreported near misses lead to motorcycling crash. By using this analogy, if data on these near misses can be recorded, then they can provide a rich source of information with which to study motorcyclists' crash risks and identify the factors that are most associated with them. Among the many walks-of-life that have adopted computer vision in tackling the most difficult part of their career are urban planners.

Among the catalysts that are responsible for the computer vision accuracy and efficiency in practical settings are the logics of the models; computer vision models are constructed in multiple hidden layers with high graphic processing capabilities to handle the large datasets [19]-[21]. The models logical construction enables computer vision to posses the capabilities of overcoming even the most herculean vision tasks of recognizing and extracting features from digital images more than any natural vision [22]-[24]. Urban scene elements such as those based on a collection of themes as found in our natural environment such as sky and built environment such as infrastructure need to be understood; and this can be achieved by computer vision represented by parsing and semantic segmentation for the localization of the objects in cities [25], [26].

Computer vision has shown a dramatic improvement on how cities complexity can be tackled. According to [24], computer vision as a field of artificial intelligence (AI) can be defined as an artificial method of training computers to learn to interpret and understand the features representation of the visual objects for their accurate identification and classification. Computers react to what they see, just as human eyes react, computer vision leverages artificial intelligence (AI) to enable computers to extract meaningful data from visual inputs such as images and videos. The insights gained from computer vision are then employed to take automated actions. Just like AI allows computers to possess thinking ability, computer vision gives them the ability to see. Computer vision, through the region-based convolutional neural network has been able to solve various visual issues that are related to videos and images accurately [24], [27].

We combined YOLOv4 (You Only Look Once version 4 [28] and DeepSort [29] to YOLOv4-DeepSort [30] for the detection and tracking tasks, and the tracked images and identity information were stored. Every 1s, the sequence of image was fetched into the VGG16 (Visual Geometry Group [31] and BiLSTM [32] model (VGG16 and BiLSTM were combined to VGG16-BiLSTM and used for extraction of image feature information and near misses recognition respectively). LSTM (Long Short Term Memory [33] is a unique recurrent neural network (RNN). We evaluate the method by testing 444 sequential video frames of motorcycling near misses in urban environment. The detection models achieved 96% accuracy for motorcycle, 89% for car, and 81% for person with lower false-positive rates on the test datasets while the tracking models achieved 34.3 MOTA on the test set and MOTP of 0.77. The results of the study indicate practicality for automatic detection of motorcycling near misses in urban environment, and it could assist in providing resourceful technical reference for analyzing the risk factors of motorcycling near misses. The work in this paper is a step towards alleviating near misses among motorcyclists and those that are directly affected in the complex urban environment. The research contributions are: (1) A hybrid of YOLOv4 and DeepSort model to enhance object detection and tracking in a complex environment and (2) A hybrid of YOLOv4 and DeepSort model to optimize the extraction of image feature information and near misses recognition respectively for overall system performance.

2. MATERIALS AND METHODS

In collecting and analyzing road safety data and the risk factors, there are methodological challenges that are involved [34]. The approach used in the existing methods for understanding near misses has limitations; therefore, this section presents the conceptual framework for understanding near misses occurrence and detection in urban environment using the proposed models as shown in Fig. 2.

2.1. Datasets

The proposed framework employed two different datasets; 1) the dataset for training and testing the models, and 2) the dataset for validating the models. The datasets were labeled using LabelMe [35], which provides an online annotation tool to label image(s) for computer vision research as applied in this study for best performance. Datasets related to road users and motorcycling near misses, and risk factors (i.e., built and natural environment) were employed for the training, testing and validation of the models (YOLOv4-DeepSort and VGG16-BiLSTM models) that are proposed in this study. Both YOLOv4-DeepSort and VGG16-BiLSTM models were trained and tested on the aforementioned datasets in ratio 30:70. Fog as one of the risk factors has 628 images and 2,876 non-fog images, which were extracted from among the dataset of weather images that consist of more than 180,000 images of four classes of weather such as rainy, sunny, cloudy and foggy [36].

Moreover, the datasets are only representations of urban settlements at daytime for intensity of the clouds (other weather and visual factors are not considered in this study).

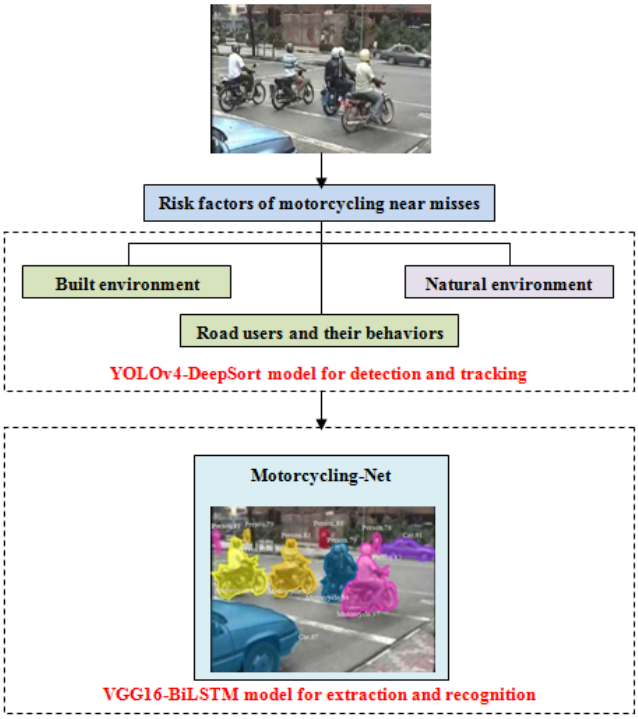

Fig. 2. Conceptual framework for understanding near misses occurrence and detection in urban environment

For the road users and motorcycling near misses, approximately 444 sequential video frames in urban environment captured by mobile and immobile cameras were employed. This is in-line with the Multi-Object Tracking (MOT) dataset [37] employed by the DeepSort method to conduct the tracking experiment. MOT dataset comprises 5,500 sequential frames of training dataset with approximately 39,905 bounding boxes, and the 5,783 sequential frames of test dataset with approximately 61,440 bounding boxes. ILSVRC CLS-LOC dataset [21] was used in training the weights of the base network of VGG16 model, and COCO dataset [20] was used in training the model by adapting the network of the last fully connected layers that were converted to layers of convolution after shortening the base network. To make up the limited datasets and improve the performance of the models, data augmentation technique was appropriately applied for two reasons; 1) for the training enhancement of the models, and the accountability for the class disparity of each model without changing the image class [24], [38]. The framework is built with one input of video frames based on the bootstrap aggregating (or bagging) technique [38] in which multi-models are trained in isolation but integrated to improve generalization.

The system specifications for carrying out the experiment are as follows: (1) Software; 64-bit Windows 10 Operating System, Jupyter IDE, and Open CV Python library, (2) Hardware; Intel Core i5 processor@2.4GHz CPU, 16 Gigabytes RAM, GeForce GTX 1080 Ti Graphics card, 2 Terabytes hard-disk, and 10.1 inch IPS HD Portable LCD Gaming Monitor PC display VGA HDMI interface for PS3/PS4/XBOx360/CCTV/Camera.

2.2. YOLOv4-DeepSort for detection and tracking

This stage comprises the detection and tracking of road users (i.e., pedestrians, automobiles, and motorcycles) and motorcycling near misses, and their risk factors (i.e., built and natural environment). The qualities possessed by YOLOv4 make it different from other approaches for detecting objects. We adopted DeepSort algorithm tracking method, which is based on SORT [39] algorithm. The simple Kalman filter was used in SORT algorithm for predicting state, and intersection over union (IOU) was used in constructing the cost matrix. Then the detection boxes and trajectory associations were made possible by using Hungarian algorithm. This algorithm in its simplicity performs excellently well in high frame rates videos.

But there is limitation to what SORT could handle; one of its limitations is ignoring the surface features of the object that was detected, its accuracy is solely depended on the low uncertainty of the state of the object. The extraction of appearance information was carried out in DeepSort, and the correlative metrics were replaced with more reliable metrics. The convolutional neural network (CNN) was trained for the extraction of appearance feature information; this increased the network robustness and greatly reduced the identification switch occurrence for an improved tracking accuracy. In this study, YOLOv4-DeepSort was employed as the multi-target tracking algorithm to track the detected road users in the video. Fig. 3 shows the flowchart of the multi-target tracking algorithm.

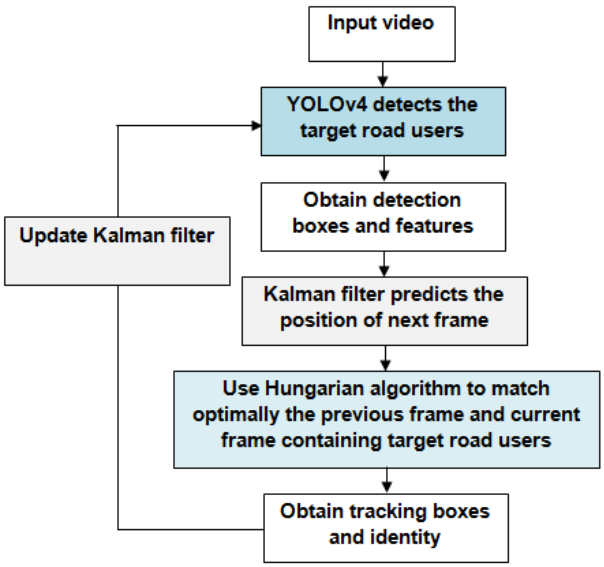

Fig. 3. Flowchart of multi-target tracking algorithm showing the contributions of the proposed models to detecting and tracking the road users for near misses analysis

In Fig. 3, the video was converted to video frames after it has been inputted into the model network, then, YOLOv4 algorithm for object detection was used to extract the deep features, followed at this stage was obtainment of candidate boxes. Non-maximum value suppression (NMS) algorithm was employed in removing the overlapping frames, thereby obtaining the final detection boxes and features. Kalman filter was used in predicting the position and state of the target road users in the next frame of the video, and the prediction result was assigned to the detection box with higher confidence after comparing the confidences of detection boxes provided by the detector. The target road users between the previous frame and the current frame were matched optimally by the Hungarian algorithm, thereby associating the tracking boxes in the previous frame with the detections in the current frame, leading to obtainment of the target trajectories in the video for the extraction of the appearance information. Concurrently, the results of the tracking were generated and the tracker's parameters were updated for target redetection.

The model's objective loss function is calculated based on the confidence loss' weighted sum and the localization loss [40]. In this study, the detection accuracy of the model is evaluated based on the centre error of performance measures of the detected object for every time-frame (1 to n). The centre error in the video for every time-frame (first frame to the last frame) is calculated based on the threshold values. The precision and recall metrics which measure the accuracy of object detection in terms of the centre error are employed. The percentage of precision, recall, false negatives and false positives is calculated as

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

Precision otherwise known as positive predictive value is the fraction of relevant objects among the total number of relevant and irrelevant retrieved objects, that is, precision is defined as the percentage of correct instances produced by a model. Recall otherwise known as sensitivity is the fraction of relevant objects that

were retrieved. A true positive is an outcome where the model correctly predicts the positive class. A false positive is an outcome where the model incorrectly predicts the positive class. A false negative is an outcome where the model incorrectly predicts the negative class. Based on the precision-recall percentile of each track object, a similarity function is employed as the metrics for evaluating the performance of the object tracking model.

The similarity function is used for evaluating tracking performance of the DeepSort in the object tracking models (YOLOv4-DeepSort). The tracking accuracy of the Deep-Sort is established if the similarity function satisfies

$$SIM\ (T_o, C_o)\ \geq\ Th_1 \tag{3}$$

where T_o is the target object and C_o is the candidate (detected) object. Th_1 is a pre-defined threshold for checking the tracking accuracy. By using Bhattacharyya coefficient, the $SIM\ (T_o, C_o)$ is calculated for computing the similarity in distance between the colour distributions of the object tracking models (YOLOv4-DeepSort) and the detected objects, the similarity function is denoted by

$$SIM\ (T_o, C_o) = \sum_{u=1}^{b} \sqrt{HT_o(u) * HC_o(u)} \tag{4}$$

where HT_o is the colour distribution of the object tracking models (YOLOv4-DeepSort) and HC_o is the colour distribution of the detected object, b denotes the total number of histogram bins. The value of threshold for occlusion detection is set between 0 and 1. Mean Average Precision (mAP) [20] is employed as the metric for evaluating the performance of the segmentation model based on the precision-recall curve of each object class. By carrying out the evaluation, the first precision-recall curve is produced, and for that particular object class, an Area Under the Curve (AUC) is calculated and referred to as Average Precision (AP). To produce the precision-recall curves, it is compulsory for the predicted instance to match with the image's ground-truth annotated object. If both the produced instance and the ground-truth instance possess the same class, and the IOU is greater in value than the predefined value, this means that there is a match between the produced instance from the model and the ground-truth instance.

The rate of overlapping between the predicted value and the ground-truth value is measured using IOU in the instance segmentation problem [41]. The IOU equation is

$$IOU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \tag{5}$$

The instance with the highest score of IOU is chosen if the instance produced by the model matches with many ground-truth values. The IOU values considered for this work is from 0.50 to 0.95 with mAP at X notation, where X is the threshold value employed in computing the metric. By removing from consideration of the ground-truth instance which matches with the produced instance, the repeated instance is penalized and considered as false positive as no other produced instance can be matched with the removed ground-truth instance object. The precision-recall is computed only after establishing all the matches for the image. Once the precision-recall points are produced using the various threshold IOU values, the average precision (AP) will be calculated. AP is calculated using

$$AP = \sum_{n=1}^{N} [\mathrm{R(n)} - \mathrm{R(n-1)}].\max \mathrm{P(n)} \tag{6}$$

where N is the number of precision-recall points produced, $P(n)$ and $R(n)$ are the precision and recall with the lowest nth recall respectively.

$$m\mathrm{AP} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} AP_i \tag{7}$$

where AP_i = the AP of class i, and N = the number of classes, and mAP = mean Average Precision.

2.3. Motorcycling-Net (VGG16-BiLSTM model)

The different environmental factors with potential to influence the motorcyclist safety and cause near misses were addressed at this stage using the proposed method with a sensor-based detector for sensing the qualitative measures, which are associated with the built environment and natural environment such as fog and road infrastructure. A fog is an atmospheric environment in which visibility is reduced because of a cloud of some substance. The framework proposed in this study for this stage is according to the work in [42], which relies on 3D-Convolutional Neural Network (3D-CNN) and VGG16-BiLSTM model based computer vision and image processing for extracting the information pertaining to risk factors (i.e., fog, bad road infrastructure,

carefree motorcyclist and pedestrian) from road-captured images using a merged method. The classification of the risk factors is also carried out at this stage irrespective of the foggy and visibility conditions. The convolution neurons of the model were trained using error back-propagation with a batch size of 32, an initial learning rate of 0.001, momentum of 0.9, 50 epochs, and Adam optimizer. Fig. 4 shows the random samples for foggy type of weather carefully acquired to suit the purpose of the study.

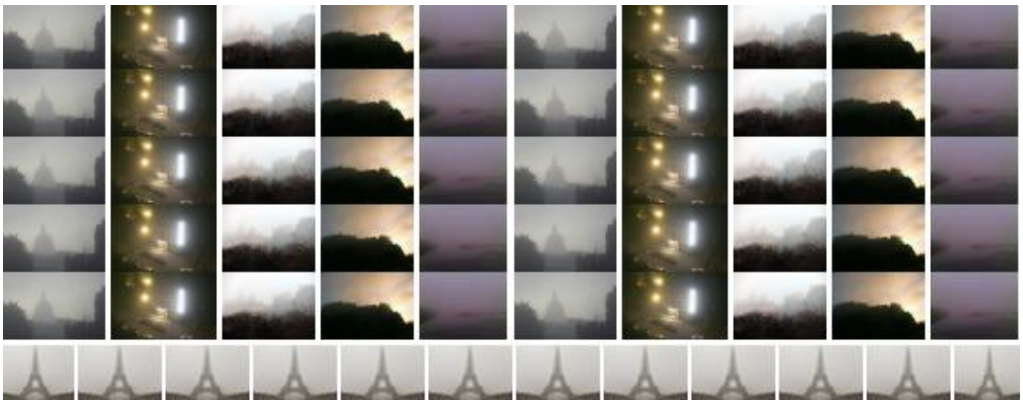

Fig. 4. Random samples for foggy weather type

This stage provides a proposed model called Motorcycling-Net for the extraction and segmentation of the detected road users and motorcycling near misses from the generated video. Motorcycling-Net is a model that is based on computer vision which itself is based on structure of convolution embedded with VGG16-BiLSTM model blocks for recognizing near miss actions from scene images. BiLSTM model was employed for the recognition of near misses in this study. BiLSTM has a better effect when it comes to time series data processing by combining the forward LSTM and backward LSTM. In this study, the pre-trained VGG16 model for Image-Net was used for extracting the image sequence features.

The feature sequences are inputted into the network model of BiLSTM after normalizing them for model effects testing. To recognize near misses, the model needs to learn some elements such as the relative motions of the objects in the scene, and the recognition of past events. The finalized hyper-parameters of the model after many experiments are a batch size of 32, dropout of 0.5, decay of 0.00005, hidden unit of 256, an initial learning rate of 0.001, momentum of 0.9, 50 epochs, and Adam optimizer. The primary motive behind the proposed model is to serve as an information generator from which important conclusions can be drawn with respect to how motorcycling behaves in urban areas, for the overall benefit of policy makers and urban planners in understanding what is required for safety measures during the course of designing urban infrastructure. Fig. 5 is an image sample of urban environment showing road users and built environment.

Fig. 5. An image sample of urban environment showing road users and built environment

3. RESULTS AND DISCUSSION

This section presents and discusses the results of the experiments conducted in this study where the first stage was responsible for detection and tracking of road users and motorcycling near misses, and risk factors

for the extraction and classification of the detected objects at the second stage. As shown in Table 1, the detection models achieved 96% accuracy for motorcycle, 89% for car, and 81% for person with lower false-positive rates on the test datasets based on the aforementioned parameters used in training the CNN model. Likewise, Table 2 shows the result achieved by YOLOv4-DeepSort model for fog detection. Fig. 6 shows the visual result of the detection experiment conducted on image sample of road users and motorcycling near misses, and risk factors (i.e., built and natural environment).

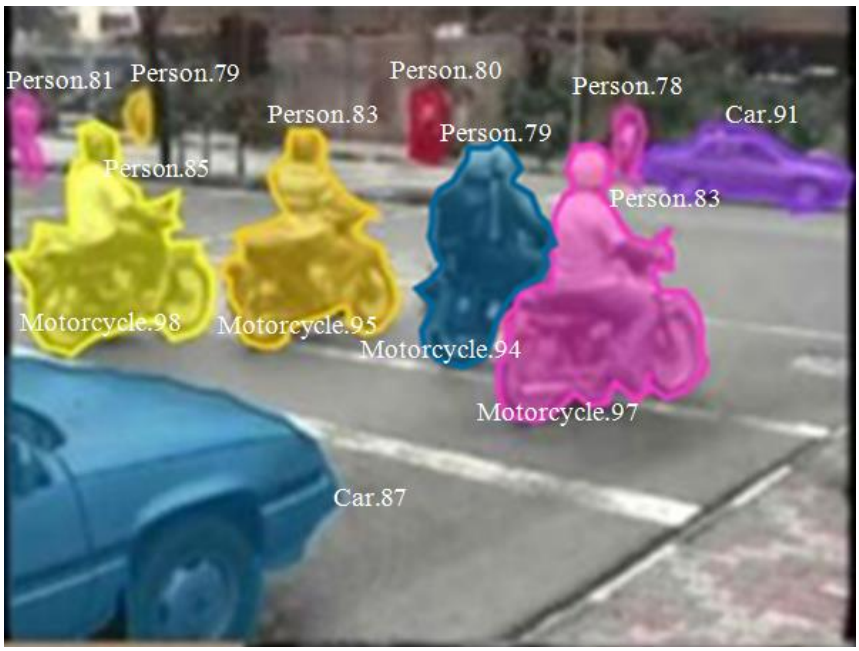

Fig. 6. Sample of testing images showing segmentation of cars, motorcycles, persons as road users

After the testing stage, we evaluated our models by comparing our results with the results achieved by other related methods. [43] achieved overall score of 0.91 by using CNN-LSTM model to detect fog and four classes of weather (rainy, sunny, cloudy, snowy); they could not detect night time and glare. [44] achieved overall score of 0.80 by using different types of CNN models to detect fog and two classes of weather (snowy and rainy); they could not detect night time and glare. [45] achieved overall score of 0.93 by using multiple residual deep models to detect the following; night time, glare, fog and weather classes (clear, rain, snow).

The proposed models are limited in performance in some instances when compared with the existing work; for example, the unavailability and inconsideration of other classes of weather datasets such as weather at night-time, snow, rain and glare images in this study affected the general performance of the models. However, narrowing the data acquisition to only road users and motorcycling near misses dataset, and fog dataset as one of the risk factors under weather condition makes the proposed models essential in addressing the current challenges reported in the existing work and for the analysis of the variations in images of urban scenes by computer vision and deep learning, which may assist city planners. Fig. 7 shows the average precision result of using VGG16-BiLSTM model for detecting (a) Motorcycle, (b) Car, and (c) Person.

Table 1. Object detection result using VGG16-BiLSTM model

	Class		
Model	**AP for Motorcycle**	**AP for Car**	**AP for Person**
VGG16-BiLSTM	0.94	0.87	0.78
	0.95	0.91	0.79
	0.97	-	0.79
	0.98	-	0.80
	-	-	0.81
	-	-	0.83
	-	-	0.83
	-	-	0.85
mAP	**0.96**	**0.89**	**0.81**

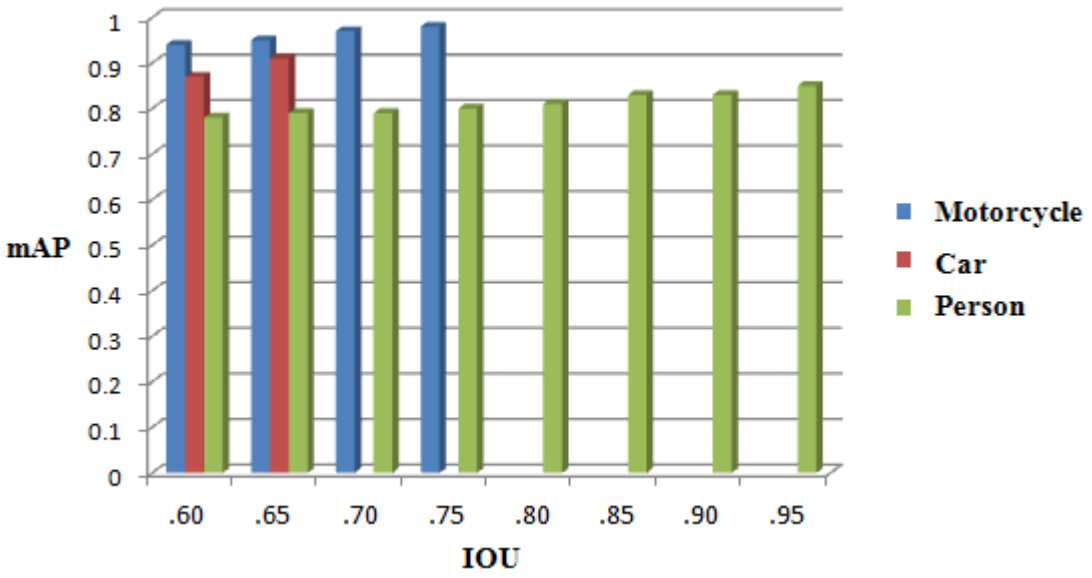

Fig. 7. Average precision result of using VGG16-BiLSTM model for detecting (a) Motorcycle, (b) Car, and (c) Person. Motorcycle shows higher recognition accuracy of 96% than others.

Table 2. Fog detection result using YOLOv4-DeepSort model

Model	Loss (cross entropy)	Accuracy (%)	Precision	True positive	False positive
YOLOv4-DeepSort	0.88	96	0.96	0.95	0.28

The tracking models achieved 34.3 Multi-Object Tracking Accuracy (MOTA) on the test set and Multi-Object Tracking Precision (MOTP) of 0.77.

4. CONCLUSION

A hybrid-based classification method for recognizing motorcycling near misses has been proposed in this study. YOLOv4-DeepSort was employed for the detection and tracking of road users (i.e., pedestrians, automobiles, and motorcycles) and motorcycling near misses, and their risk factors (i.e., built and natural environment). The qualities possessed by YOLOv4 make it different from other approaches for detecting objects. The extraction and recognition experiments were conducted by using VGG16-BiLSTM model for Motorcycling-Net. While the detection models achieved 96% accuracy for motorcycle, 89% for car, and 81% for person with lower false-positive rates on the test datasets, the tracking models achieved 34.3 MOTA on the test set and MOTP of 0.77. Even though these results justify the objectives of the research, in our future work, we intend to utilize more datasets of different classes of weather and other risk factors of near misses with their agents.

REFERENCES

[1] J. Kandt, and M. Batty, "Smart cities, big data and urban policy: Towards urban analytics for the long run," *Cities*, vol. 109, pp. 1-10, 2021, https://doi.org/10.1016/j.cities.2020.102992.
[2] Z. Lv, K. Ota, J. Lloret, W. Xiang, and P. Bellavista, "Complexity Problems Handled by Advanced Computer Simulation Technology in Smart Cities 2021," *Complexity*, vol. 2022, pp. 1-3, 2022, https://doi.org/10.1155/2022/9847249.
[3] S. G. Ortman, J. Lobo, and M. E. Smith, "Cities: Complexity, theory and history," *Plos one*, vol. 15, pp. 1-24, 2020, https://doi.org/10.1371/journal.pone.0243621.
[4] P. Marzuki, A. R. Syafeeza, Y. C. Wong, N. A. Hamid, A. N. Alisa, and M. M. Ibrahim," A design of license plate recognition system using convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 9, pp. 2196-2204, 2019, https://doi.org/10.11591/ijece.v9i3.pp2196-2204.
[5] M. Dozza, A. Schwab, and F. Wegman, "Safety science special issue on cycling safety," *Saf. Sci*., vol. 92, pp. 262-263, 2017, https://doi.org/10.1016/j.ssci.2016.06.009.
[6] https://www.vanguardngr.com/2020/01/689-dead-over-250-injured-in-over-1500-okada-tricycle-accidents-within-four-years-reports/

[7] M. Winters, R. Buehler, and T. Götschi, "Policies to promote active travel: evidence from reviews of the literature," *Current environmental health reports*, vol. 4, pp. 278-285, 2017, https://doi.org/10.1007/s40572-017-0148-x.
[8] B. Savan, E. Cohlmeyer, and T. Ledsham, "Integrated strategies to accelerate the adoption of cycling for transportation," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 46, pp. 236-249, 2017, https://doi.org/10.1016/j.trf.2017.03.002.
[9] J. Eriksson, A. Niska, and A. Forsman, "Injured cyclists with focus on single-bicycle crashes and differences in injury severity in Sweden," *Accident Analysis & Prevention*, vol. 165, pp. 1-7, 2022, https://doi.org/10.1016/j.aap.2021.106510.
[10] K. Gildea, D. Hall, and C. Simms, "Configurations of underreported cyclist-motorised vehicle and single cyclist collisions: Analysis of a self-reported survey," *Accident Analysis & Prevention*, vol. 159, pp. 1-17, 2021, https://doi.org/10.1016/j.aap.2021.106264.
[11] R. von Stülpnagel, and J. Lucas, "Crash risk and subjective risk perception during urban cycling: Evidence for congruent and incongruent sources," *Accident Analysis & Prevention*, vol. 142, pp. 1-12, 2020, https://doi.org/10.1016/j.aap.2020.105584.
[12] L. B. Meuleners, M. Fraser, M. Johnson, M. Stevenson, G. Rose, and J. Oxley, "Characteristics of the road infrastructure and injurious cyclist crashes resulting in a hospitalization," *Accident Analysis & Prevention*, vol. 136, pp. 1-9, 2020, https://doi.org/10.1016/j.aap.2019.105407.
[13] M. Winters, and M. Branion-Calles, "Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia," *J. Transp. Health*, vol. 7, pp. 48-53, 2017, https://doi.org/10.1016/j.jth.2017.02.010.
[14] M. Hosseinpour, T. K. O. Madsen, A. V. Olesen, and H. Lahrmann, "An in-depth analysis of self-reported cycling injuries in single and multiparty bicycle crashes in Denmark," *Journal of safety research*, vol. 77, pp. 114-124, 2021, https://doi.org/10.1016/j.jsr.2021.02.009.
[15] M. Imprialou, and M. Quddus, "Crash data quality for road safety research: Current state and future directions," *Accid. Anal. Prev.* vol. 130, pp. 84-90, 2019, https://doi.org/10.1109/CVPR.2018.00823.
[16] J. Fischer, T. Nelson, K. Laberee, and M. Winters, "What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city, *Accident Analysis & Prevention*, vol. 145, pp. 1-8, 2020, https://doi.org/10.1007/978-3-319-10602-1_48.
[17] R. Aldred, "Inequalities in self-report road injury risk in Britain: A new analysis of National Travel Survey data, focusing on pedestrian injuries," *J. Transp. Health*, vol. 9, pp. 96-104, 2018, https://doi.org/10.1016/j.jth.2018.03.006.
[18] D. Shannon, F. Murphy, M. Mullins, and J. Eggert, "Applying crash data to injury claims-an investigation of determinant factors in severe motor vehicle accidents," *Accident Analysis & Prevention*, vol. 113, pp. 244-256, 2018, https://doi.org/10.1016/j.aap.2018.01.037.
[19] Y. Chen, W. Li, and L. Van Gool, "Road: Reality oriented adaptation for semantic segmentation of urban scenes," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7892-7901, 2018, https://doi.org/10.1109/CVPR.2018.00823.
[20] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common Objects in Context," *in: Computer Vision – ECCV, Springer*, pp. 740-755, 2014, https://doi.org/10.1007/978-3-319-10602-1_48.
[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211-252, 2015, https://doi.org/10.1007/s11263-015-0816-y.
[22] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, pp. 1-14, 2018, https://doi.org/10.1155/2018/7068349.
[23] C. C. J. Kuo, M. Zhang, S. Li, J. Duan, and Y. Chen, "Interpretable convolutional neural networks via feedforward design," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 346-359, 2019, https://doi.org/10.1016/j.jvcir.2019.03.010.
[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015, https://doi.org/10.1038/nature14539.
[25] A. Chaurasia, and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *IEEE Vis. Commun. Image Process. VCIP*, pp. 1-4, 2017, https://doi.org/10.1109/VCIP.2017.8305148.
[26] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene Parsing through ADE20K Dataset," *in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5122-5130, 2017, https://doi.org/10.1109/CVPR.2017.544.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016, https://doi.org/10.1109/CVPR.2016.90.
[28] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, pp. 1-17, 2020, https://arxiv.org/abs/2004.10934.
[29] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *In IEEE international conference on image processing (ICIP)*, pp. 3645-3649, 2017, https://doi.org/10.1109/ICIP.2017.8296962.

[30] X. Zhang, X. Hao, S. Liu, J. Wang, J. Xu, and J. Hu, "Multi-target tracking of surveillance video with differential YOLO and DeepSort," *In: Proceedings of 11th International Conference on Digital Image Processing*, vol.11179, pp. 701-710, 2019, https://doi.org/10.1117/12.2540269.
[31] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, pp. 1-14, 2014, https://arxiv.org/abs/1409.1556.
[32] G. S. Chadha, A. Panambilly, A. Schwung, and S. X. Ding, "Bidirectional deep recurrent neural networks for process fault classification," *ISA transactions*, vol. 106, pp. 330-342, 2020, https://doi.org/10.1016/j.isatra.2020.07.011.
[33] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735-1780, 1997, https://doi.org/10.1109/78.650093.
[34] M. Schlögl, and R. Stütz, "Methodological considerations with data uncertainty in road safety analysis," Accid. Anal. Prev. vol. 130, pp. 136-150, 2019, https://doi.org/10.1016/j.aap.2017.02.001.
[35] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, pp. 157-173, 2008, https://doi.org/10.1007/s11263-007-0090-8.
[36] W. T. Chu, X. Y. Zheng, and D. S. Ding, "Image2Weather: A large-scale image dataset for weather property estimation," *in: IEEE Second International Conference on Multimedia Big Data (BigMM)*, pp. 137-144, 2016, https://doi.org/10.1109/BigMM.2016.9.
[37] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a benchmark for multi-target tracking," *arXiv preprint arXiv:1504.01942*, pp. 1-15, 2015, https://arxiv.org/abs/1504.01942.
[38] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, *Adaptive computation and machine learning series*, MIT Press, 2016, https://books.google.co.id/books?id=omivDQAAQBAJ.
[39] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," *IEEE Int. Conf. Image Process. ICIP*, pp. 3464-3468, 2016, https://doi.org/10.1109/ICIP.2016.7533003.
[40] L. Liu, E. A. Silva, C. Wu, and H. Wang, "A machine learning-based method for the large-scale evaluation of the qualities of the urban environment," *Comput. Environ. Urban Syst.*, vol. 65, pp. 113-125, 2017, https://doi.org/10.1016/j.compenvurbsys.2017.06.003.
[41] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 386-397, 2020, https://doi.org/10.1109/TPAMI.2018.2844175.
[42] H. Kamangir, W. Collins, P. Tissot, S. A. King, H. T. H. Dinh, N. Durham, and J. Rizzo, "FogNet: A multiscale 3D CNN with double-branch dense block and attention mechanism for fog prediction," *Machine Learning with Applications*, vol. 5, pp. 1-17, 2021, https://doi.org/10.1016/j.neucom.2018.09.048.
[43] B. Zhao, X. Li, X. Lu, and Z. Wang, "A CNN-RNN architecture for multi-label weather recognition," Neurocomputing, vol. 322, pp. 47–57, 2018, https://doi.org/10.1016/j.neucom.2018.09.048.
[44] J. C. Villarreal Guerra, Z. Khanam, S. Ehsan, R. Stolkin, and K. McDonald-Maier, "Weather classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of convolutional neural networks," *in: NASA/ESA Conference on Adaptive Hardware and Systems (AHS), IEEE*, pp. 305–310, 2018, https://doi.org/10.1109/AHS.2018.8541482.
[45] M. R. Ibrahim, J. Haworth, and T. Cheng, "WeatherNet: Recognising weather and visual conditions from street-level images using deep residual learning," *ISPRS Int. J. Geo-Inf.*, vol. 8, pp. 549, 2019, https://doi.org/10.3390/ijgi8120549.