Journal of Social Sciences and Economics
Vol. 1(1),2022; 42-52
ISSN (Online) 2958-1532
https://finessepublishing.com/jsse

Motorcycling-Net: A Segmentation Approach For Detecting Motorcycling Near Misses

Rotimi-Williams Bello[a], Chinedu Uchechukwu Oluigbo[a], Oluwatomilola Motunrayo Moradeyo[b], Daniel A. Olubummo[c]

a. Department of Mathematics and Computer Science, University of Africa, 561101 Sagbama, Bayelsa State, Nigeria
b. Department of Computer Science, Adeseun Ogundoyin Polytechnic, Eruwa, Oyo State, Nigeria
c. Department of Computer and Information Systems, Robert Morris University, Moon Township, Pennsylvania, United States of America

Abstract: This paper presents near misses as corrective and preventive measures to safety events. Our focus is on the risk factors of commercial motorcycling near-miss incidents, which we address by proposing a near-miss detection framework based on deep learning. Video streams of near-miss datasets containing motorcycling in different scenes were collected for the experiment. We employed the YOLOv4-DeepSort model for detection and tracking, and stored the tracked images and identity information. Every 1s, the image sequence was fed into the VGG16-BiLSTM model (VGG16 and BiLSTM were used to extract image features and recognize near misses, respectively). We evaluate the method by testing 444 sequential video frames from motorcycling near-miss incidents in an urban environment, achieving approximately 96% recognition accuracy. The results of the study indicate the practicality of automatic detection of motorcycling near miss in urban environments, and it could assist in providing a resourceful technical reference for analyzing the risk factors of motorcycling near misses.

Keywords: Accident, Computer vision, Image processing, Near miss[1]

1. Introduction:

Commercial motorcycling is one of the economic means of transportation in many countries, although many perceive it as a dangerous means of transportation, which is affirmed by the number of casualties recorded daily. This life-threatening record has greatly hindered continuous support for commercial motorcycling as an affordable means of transportation. Information retrieved from near-miss datasets can be a telltale of potential hazards and their prevention.

However, many researchers have come up with different definitions for near misses, and this has created a gap in applying the right method to near misses, thereby making it statistically difficult to address the situation for a safer commercial motorcycling. 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 the recurrence of such incident. Figure 1 shows a typical implication of ignored or unreported near miss incident. Near missed 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 preceded 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.

[1] Received 19 Oct 2022; Accepted 22 Dec 2022; Published (online) 31 Dec 2022
Finesse Publishing stays neutral regard to jurisdictional claims published maps

Corresponding email: sirbrw@yahoo.com
DOI: 10.61363/jsse.v1i1.34

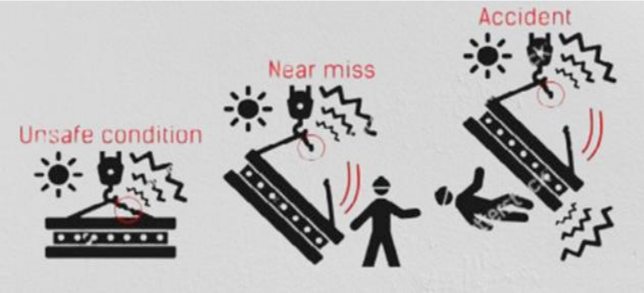

Figure 1: Unreported or ignored scenarios of near-miss incidents leading to accidents. Source: Internet

.

According to (Arribas-Bel, 2014) There is potential in the newly emerging urban data sources to handle image-related urban modeling tasks. Utilizing such data for urban scene analysis helps understand the dynamic nature of cities and control incidents related to traffic or overcrowding. Several attempts have been made by urban researchers to model cities using multi-agent models and the science of complexity and the theory of networks (Zhou & Li, 2013). The implication of these models, most of the time, is the over-simplification of the urban systems' initial settings and cities' exploration from a perspective that is one-dimensional (Batty & Torrens, 2001). The lack of large-scale, publicly accessible datasets, open-source, modifiable code, and graphics processing units (GPUs) is limiting the robustness of these models when feeding simulations of unusually large scope. However, the advent of artificial intelligence-based computer vision and image processing algorithms has changed the whole system by paving the way for data/video analytics and analytical techniques that can address urban-related problems.

The emergence of computer vision and image processing, and their applications, has helped in understanding the complexity of the city's dynamics for the prompt detection of motorcycling near-miss incidents. Various problems confronting urban settlements can be resolved through large-scale datasets of analyzed digital images, enabling essential information and image elements to be tracked and extracted as if performed by human experts, benefiting the transportation industry overall. Some of the risks faced by urban dwellers due to traffic congestion and incidents are due to the nature of the roads and transport networks in place. According to (Tarigan et al., 2017)To control these ugly incidents, for a couple of years, measures were put in place, such as safety awareness programs to sensitize urban dwellers on road safety and the monitoring of urban transport networks using technology-based road signs and traffic lights. However, these measures could not give a detailed account of the unpalatable consequences, such as congestion or incidents of traffic, of the agents' behavior, such as motorcycles, pedestrians, and automobiles, all within the city environment. Having such a detailed account is essential and beneficial for motorcyclists, who are mostly exposed to road accidents and have little or no near-miss data. Motorcycling is a major occupation among many youths in Nigeria and elsewhere (Dozza et al., 2017). However, incessant motorcycling road crashes are 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 (Bello et al., 2023).

Although other means of transportation have been encouraged globally to ease the challenges of transportation and reduce air pollution caused by automobiles (Pucher et al., 2010; Savan et al., 2017)Motorcycling has not been favorably considered in that category due to numerous near-miss incidents involving motorcyclists, resulting from little or no formal training in the rules governing roads and their use. In another perspective, motorcycling is perceived as a dangerous mode of transportation in that only a few passengers can withstand the trauma of its near-miss incidents or the risk of dodging its crashes (Blaizot et al., 2013). 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-miss incident or another (Aldred & Crosweller, 2015), thereby hindering the wider acceptance of commercializing motorcycling as a mode of transportation (Aldred, 2016; Winters & Branion-Calles, 2017).

Although most of the near-miss incidents are reported, not recording them contributes to the difficulty in accessing their data as an information source for investigating and identifying the associated factors responsible for the risk faced 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 valuable source of information to address these tasks, especially the video data, which provide replicas of the original near-miss scenes for feature extraction (Aldred, 2016; Beck et al., 2016; De Rome et al., 2014; Imprialou & Quddus, 2019; Teschke et al., 2014). In all combinations, incidents of motorcycling near misses are factors- and event-based, and in fact are not always caused solely by transportation.

An in-depth understanding of these factors and events will ease the problem of analyzing motorcycling near-miss incidents. Due to the small number recorded for near miss incidents, and the impact of prejudice on reporting data pertaining to road crash, where only the near miss incidents that lead to crash or death are reported while the near miss incidents 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 (Abay, 2015; Aldred, 2018; De Rome et al., 2014). Too many near-miss incidents go unreported or unaddressed, leading to motorcycling crashes. Using this analogy, if data on these near misses can be recorded, they can provide a rich source of information to study motorcyclists' crash risks and identify the factors most associated with them. Among the many walks of life that have adopted computer vision to tackle the most difficult parts of their careers are urban planners.

Among the factors that affect the accuracy and efficiency of computer vision in practical settings are the model's architecture; computer vision models are constructed with multiple hidden layers and high computational power to handle large datasets (Cordts et al., 2016; Russakovsky et al., 2015). The model's logical construction enables computer vision to overcome even the most herculean vision tasks, such as recognizing and extracting features from digital images, better than natural vision (Guo et al., 2016; LeCun et al., 2015). Urban scene elements, such as those based on a collection of themes 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 (Chaurasia & Culurciello, 2017; Zhou et al., 2017).

Computer vision has dramatically improved how the complexity of cities can be tackled. According to(LeCun et al., 2015)Computer vision, as a field of artificial intelligence (AI), is the artificial method of training computers to interpret and understand the feature representations of visual objects for 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 from computer vision are then used to automate actions. Just as AI enables computers to think, computer vision enables them 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 (He et al., 2016; LeCun et al., 2015).

We combined YOLOv4 (You Only Look Once version 4 (Bochkovskiy et al., 2020) and DeepSort (Wojke et al., 2017) to YOLOv4-DeepSort (Zhang et al., 2019) for detection and tracking, and the tracked images and identity information were stored. Every 1s, the sequence of images was fed into the VGG16 (Visual Geometry Group (Simonyan & Zisserman, 2014) and BiLSTM (Hochreiter & Schmidhuber, 1997) models (which were combined into VGG16-BiLSTM and used for extracting image feature information and near-miss recognition, respectively). LSTM (Long Short Term Memory (Schuster & Paliwal, 1997) It is a unique recurrent neural network (RNN). We evaluated the method by testing 444 sequential video frames from motorcycling near-miss incidents in an urban environment, achieving approximately 96% recognition accuracy. The results of the study indicate practicality for automatic detection of motorcycling near misses in urban environments, and it could assist in providing a resourceful technical reference for analyzing the risk factors of motorcycling near misses. The work in this paper is a step towards alleviating near-miss incidents among motorcyclists and those who are directly affected in the complex urban environment.

2. Materials and Methods

In collecting and analyzing road safety data and risk factors, there are methodological challenges (Schloegl & Stuetz, 2019). The approach used in the existing methods for understanding near misses has limitations;

therefore, this section presents the conceptual framework for understanding the occurrence and detection of near misses in urban environments using the proposed models as shown in Figure 2.

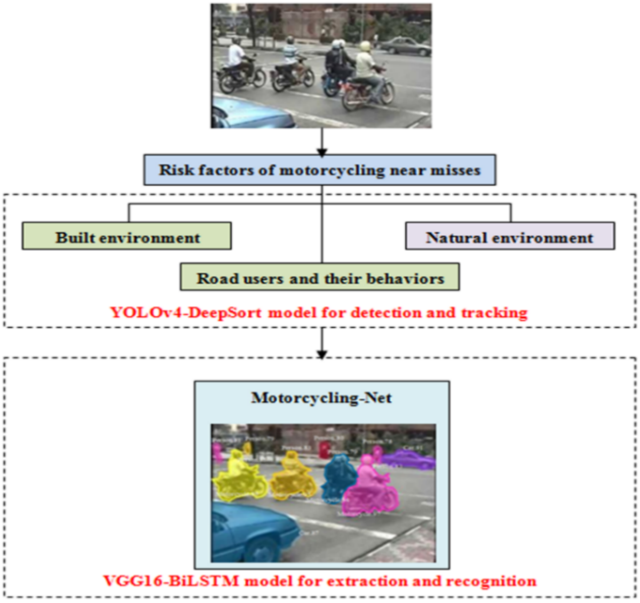
Risk factors of motorcycling near misses
Built environment
Natural environment
Road users and their behaviors
YOLOv4-DeepSort model for detection and tracking
Motorcycling-Net
VGG16-BiLSTM model for extraction and recognition

Figure 2: Conceptual framework for understanding near miss's occurrence and detection urban environment

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 (Russell et al., 2008), which provides an online annotation tool to label image(s) for computer vision research as applied in this study for the 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 a ratio of 30:70. Fog, as one of the risk factors, has 628 images and 2876 non-fog images, which were extracted from among the dataset of weather images that consists of more than 180,000 images of four classes of weather, such as rainy, sunny, cloudy, and foggy (Chu et al., 2016). Moreover, the datasets represent only daytime urban settlements and cloud intensity (other weather and visual factors are not considered in this study). For the road users and motorcycling near misses, approximately 444 sequential video frames in an urban environment, captured by mobile and immobile cameras, were employed.

This is in line with the Multi-Object Tracking (MOT) dataset (Leal-Taixé et al., 2015) employed by the DeepSort method to conduct the tracking experiment. The MOT dataset comprises 5500 sequential frames of the training dataset with approximately 39,905 bounding boxes, and 5,783 sequential frames of the test dataset with approximately 61,440 bounding boxes. ILSVRC CLS-LOC dataset (Russakovsky et al., 2015) was used in training the weights of the base network of the VGG16 model, and the COCO dataset (Lin et al., 2014) was used to train the model by adapting the network, converting the last fully connected layers into convolutional layers after shortening the base network. To make up for the limited datasets and improve the performance of the models, a data augmentation technique was appropriately applied for two reasons: 1) for the training enhancement of the models, and the account for the class disparity of each model without changing the image class (LeCun et al., 2015; Ortega et al., 2021). The framework is built with one input of video frames based on the bootstrap aggregating (or bagging) technique (Ortega et al., 2021) 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), motorcycling near miss, and their risk factors (i.e., built and natural environments). The qualities of YOLOv4 set it apart from other object detection approaches. We adopted the DeepSort algorithm tracking method, which is based on SORT (Bewley et al., 2016) algorithm. The simple Kalman filter was used in the SORT algorithm to predict the state, and intersection over union (IoU) was used to construct the cost matrix. Then, the detection boxes and trajectory associations were computed using the Hungarian algorithm. This algorithm, despite its simplicity, performs excellently on high-frame-rate videos.

But there is a 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 dependent on the low uncertainty of the state of the object. The extraction of appearance information was performed in DeepSort, and the corresponding metrics were replaced with more reliable ones. The convolutional neural network (CNN) was trained to extract appearance features; this increased the network's robustness and greatly reduced the occurrence of identification switches, thereby improving tracking accuracy. In this study, YOLOv4-DeepSort was employed as the multi-target tracking algorithm to track the detected road users in the video. Figure 3 shows the flowchart of the multi-target tracking algorithm.

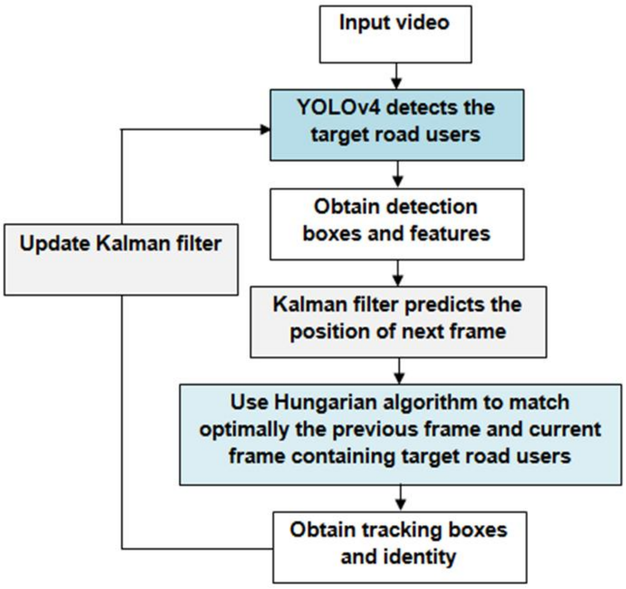
Input video
YOLOv4 detects the target road users
Obtain detection boxes and features
Update Kalman filter
Kalman filter predicts the position of next frame
Use Hungarian algorithm to match optimally the previous frame and current frame containing target road users
Obtain tracking boxes and identity

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

In Figure 3, the video was converted to frames after it was input into the model network, then the YOLOv4 algorithm for object detection was used to extract deep features, followed by the generation of candidate boxes. The Non-Maximum Suppression (NMS) algorithm was employed to remove overlapping frames, thereby obtaining the final detection boxes and features. A Kalman filter was used to predict the position and state of target road users in the next frame of the video, and the prediction results were assigned to the detection boxes with higher confidence after comparing the detector's confidence scores. The target road users between the previous and current frames were matched optimally using the Hungarian algorithm, thereby associating the tracking boxes in the previous frame with the detections in the current frame, leading to the target trajectories in the video for the extraction of appearance information. Concurrently, the tracking results were generated, and the tracker's parameters were updated for target redetection. The model's objective loss function is calculated as the weighted sum of the confidence loss and the localization loss (Liu et al., 2017). In this study, the model's detection accuracy is evaluated based on the centre error of the performance measures for the detected object at each time frame (1 to n). The centre error for each time frame (from the first to the last) 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

$$\text{Precision} = (\text{True Positive})/(\text{True Positive} + \text{False Positive}) \tag{1}$$

$$\text{Recall} = (\text{True Positive})/(\text{True Positive} + \text{False Negative}) \tag{2}$$

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, also known as sensitivity, is the fraction of relevant objects 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 for each track object, a similar function is used as the metric for evaluating the object tracking model's performance. The similarity function is used to evaluate the tracking performance of the DeepSort in the object-tracking model (YOLOv4-DeepSort). The tracking accuracy of the Deep-Sort is established if the similarity function satisfies.

$$\text{SIM}\ (\text{T}_\text{o}, \text{C}_\text{o}) \geq \text{Th}_1 \tag{3}$$

where To is the target object and Co is the candidate (detected) object. Th1 is a predefined threshold used to check tracking accuracy. By using the Bhattacharyya coefficient, the SIM (To, Co) 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

$$\text{SIM}\ (\text{T}_\text{o}, \text{C}_\text{o}) = \sum_{u=1}^{b} \sqrt{\text{HTo(u)} * \text{HCo(u)}} \tag{4}$$

where HTo is the colour distribution of the object tracking models (YOLOv4-DeepSort) and HCo is the colour distribution of the detected object, b denotes the total number of histogram bins. The value of the threshold for occlusion detection is set between 0 and 1. Mean Average Precision (mAP) (Lin et al., 2014) is used as the metric for evaluating the segmentation model's performance, based on the precision-recall curve for 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 overlap between the predicted value and the ground-truth value is measured using IOU in the instance segmentation problem (He et al., 2017). The IOU equation is

$$\text{IOU} = (\text{Area of Intersection})/(\text{Area of Union}) \tag{5}$$

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 0.60 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 that matches the produced instance, the repeated instance is penalized and considered as a false positive, as no other produced instance can be matched with the removed ground-truth instance object. Precision-recall is computed only after all matches for the image are established. Once the precision-recall points are produced using the various threshold IOU values, the average precision (AP) will be calculated. AP is calculated using

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

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.

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

where 〖AP〗_i= the AP of class i, 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 (Kamangir et al., 2021), which relies on a 3D Convolutional Neural Network (3D-CNN) and a VGG16-BiLSTM model based on computer vision and image processing to extract information on risk factors (i.e., fog, poor road infrastructure, careless motorcyclists, and pedestrians) from road-captured images using a merged approach. The classification of risk factors is also carried out at this stage, regardless of fog or visibility conditions. The convolutional neurons of the model were trained using backpropagation with a batch size of 32, an initial learning rate of 0.001, a momentum of 0.9, 50 epochs, and the Adam optimizer. Figure 4 shows the random samples for foggy type of weather carefully acquired to suit the purpose of the study.

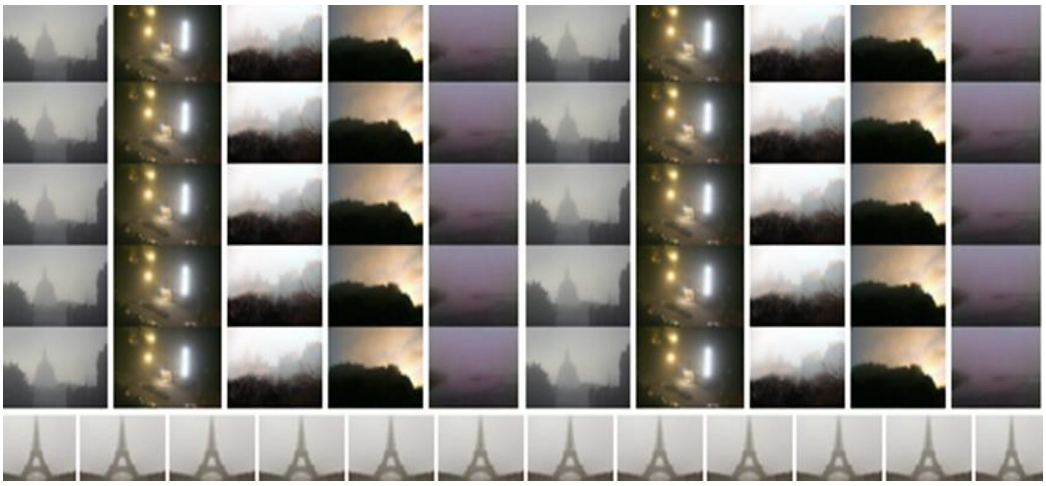

Figure 4: Random samples for foggy weather type

This stage proposes a model, Motorcycling-Net, for extracting and segmenting detected road users and motorcycling near misses from the generated video. Motorcycling-Net is a model based on computer vision, which is itself built on the VGG16-BiLSTM architecture for recognizing near-miss actions from scene images.

A BiLSTM model was employed for the recognition of near misses in this study. BiLSTM performs better for time series data processing by combining forward and backward LSTM layers. In this study, the pre-trained VGG16 model for ImageNet was used to extract image sequence features.

The feature sequences are input to the BiLSTM network after normalization for model effects testing. To recognize near misses, the model needs to learn elements such as the relative motions of objects in the scene and the recognition of past events. The finalized hyperparameters of the model, after many experiments, are a batch size of 32, dropout of 0.5, decay of 0.00005, hidden units of 256, an initial learning rate of 0.001, momentum of 0.9, 50 epochs, and the Adam optimizer.

The primary motive behind the proposed model is to serve as an information source from which important conclusions can be drawn about how motorcycling behaves in urban areas, for the overall benefit of policymakers and urban planners in understanding what is required for safety measures during the design of urban infrastructure. Figure 5 is an image sample of an urban environment showing road users and the built environment.

Figure 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. Figure 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).

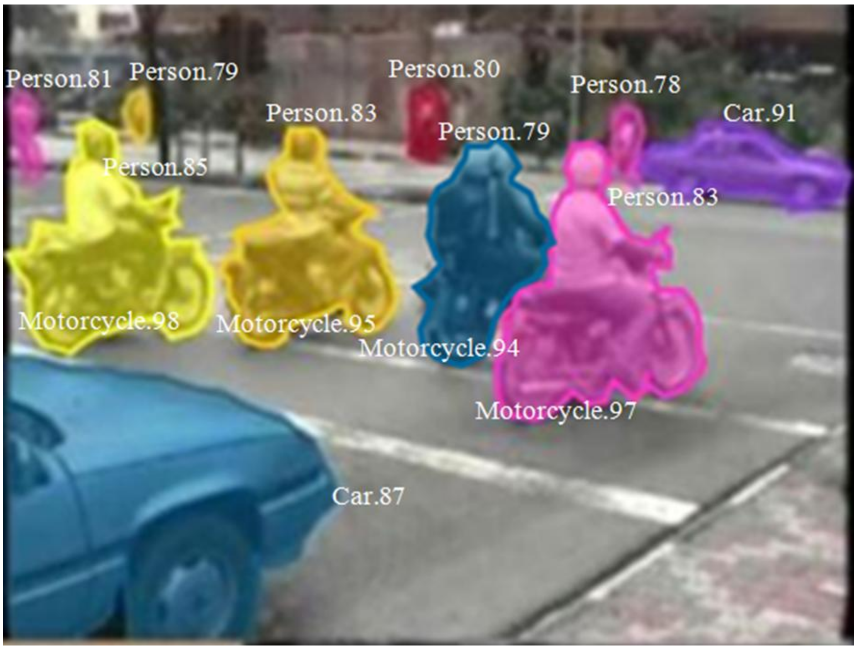

Figure 6: Sample of testing images showing segmentation of cars, motorcycles, persons as road users

After the testing stage, we evaluated our models by comparing our results with those of other related methods(Zhao et al., 2018). achieved an overall score of 0.91 by using a CNN-LSTM model to detect fog and four classes of weather (rainy, sunny, cloudy, snowy); they could not detect nighttime and glare (Guerra et al., 2018). They achieved an 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 nighttime and glare. (Ibrahim et al., 2019) achieved an overall score of 0.93 by using multiple residual deep models to detect the following: nighttime, glare, fog, and weather classes (clear, rain, snow).

The proposed models perform less well in some instances compared with the existing work; for example, the unavailability and lack of consideration of other weather datasets, such as night-time, snow, rain, and glare images, in this study affected the models' overall performance. 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 conditions 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. Figure 7 shows the average precision result of using the VGG16-BiLSTM model for detecting (a) Motorcycle, (b) Car, and (c) Person

Table 1: Object detection result using the 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**

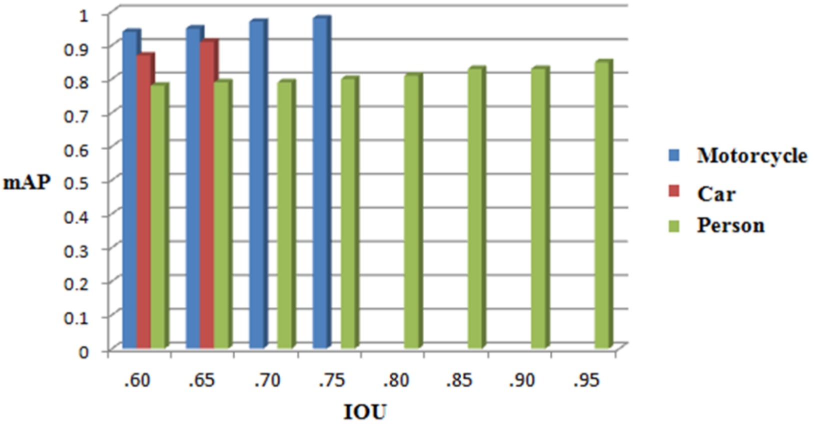

Figure 7: Average precision result of using the VGG16-BiLSTM model for detecting (a) Motorcycle, (b) Car, and (c) Person. Motorcycle shows higher recognition accuracy of 96% than others.

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 segmentation approach for detecting 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), motorcycling near misses, and their risk factors (i.e., built and natural environments). The qualities of YOLOv4 set it apart from other object detection approaches. The extraction and recognition experiments were conducted by using Motorcycling-Net (VGG16-BiLSTM model). While the detection models achieved 96% accuracy for motorcycles, 89% for cars, and 81% for people, with lower false-positive rates on the test datasets, the tracking models achieved 34.3 MOTA and a MOTP of 0.77. Although these results justify the research objectives, we intend to use additional datasets from different weather classes and other risk factors associated with near misses involving their agents in future work.

References

Abay, K. A. (2015). Investigating the nature and impact of reporting bias in road crash data. *Transportation Research Part A: Policy and Practice*, *71*, 31-45.

Aldred, R. (2016). Cycling near misses: Their frequency, impact, and prevention. *Transportation Research Part A: Policy and Practice*, *90*, 69-83.

Aldred, R. (2018). Inequalities in self-report road injury risk in Britain: A new analysis of National Travel Survey data, focusing on pedestrian injuries. *Journal of Transport & Health*, *9*, 96-104.

Aldred, R., & Crosweller, S. (2015). Investigating the rates and impacts of near misses and related incidents among UK cyclists. *Journal of Transport & Health*, *2*(3), 379-393.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *applied Geography*, *49*, 45-53.

Batty, M., & Torrens, P. M. (2001). Modelling complexity: the limits to prediction. *Cybergeo: European Journal of Geography*.

Beck, B., Stevenson, M., Newstead, S., Cameron, P., Judson, R., Edwards, E. R., Bucknill, A., Johnson, M., & Gabbe, B. (2016). Bicycling crash characteristics: An in-depth crash investigation study. *Accident Analysis & Prevention*, *96*, 219-227.

Bello, R.-W., Oluigbo, C. U., Moradeyo, O. M., & Olubummo, D. A. (2023). Motorcycling-Net: A Segmentation Approach for Detecting Motorcycling Near Misses. *Journal of Social Sciences and Economics*, *2*(1), 20-30.

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. 2016 IEEE international conference on image processing (ICIP),

Blaizot, S., Papon, F., Haddak, M. M., & Amoros, E. (2013). Injury incidence rates of cyclists compared to pedestrians, car occupants and powered two-wheeler riders, using a medical registry and mobility data, Rhône County, France. *Accident Analysis & Prevention*, *58*, 35-45.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Chaurasia, A., & Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE visual communications and image processing (VCIP),

Chu, W.-T., Zheng, X.-Y., & Ding, D.-S. (2016). Image2weather: A large-scale image dataset for weather property estimation. 2016 IEEE Second International Conference on Multimedia Big Data (BigMM),

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. Proceedings of the IEEE conference on computer vision and pattern recognition,

De Rome, L., Boufous, S., Georgeson, T., Senserrick, T., Richardson, D., & Ivers, R. (2014). Bicycle crashes in different riding environments in the Australian capital territory. *Traffic injury prevention*, *15*(1), 81-88.

Dozza, M., Schwab, A., & Wegman, F. (2017). Safety science special issue on cycling safety. *Safety science*, *92*, 262-263.

Guerra, J. C. V., Khanam, Z., Ehsan, S., Stolkin, R., & McDonald-Maier, K. (2018). Weather Classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of Convolutional Neural Networks. 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS),

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27-48.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. Proceedings of the IEEE international conference on computer vision,

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Ibrahim, M. R., Haworth, J., & Cheng, T. (2019). WeatherNet: Recognising weather and visual conditions from street-level images using deep residual learning. *ISPRS International Journal of Geo-Information*, *8*(12), 549.

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis & Prevention*, *130*, 84-90.

Kamangir, H., Collins, W., Tissot, P., King, S. A., Dinh, H. T. H., Durham, N., & Rizzo, J. (2021). FogNet: A multiscale 3D CNN with double-branch dense block and attention mechanism for fog prediction. *Machine Learning with Applications*, *5*, 100038.

Leal-Taixé, L., Milan, A., Reid, I., Roth, S., & Schindler, K. (2015). Motchallenge 2015: Towards a benchmark for multi-target tracking. *arXiv preprint arXiv:1504.01942*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. European conference on computer vision,

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, environment and urban systems*, *65*, 113-125.

Ortega, R. F., Irurita, J., Campo, E. J. E., & Mesejo, P. (2021). Analysis of the performance of machine learning and deep learning methods for sex estimation of infant individuals from the analysis of 2D images of the ilium. *International Journal of Legal Medicine*, *135*(6), 2659-2666.

Pucher, J., Dill, J., & Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive medicine*, *50*, S106-S125.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., & Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211-252.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, *77*(1), 157-173.

Savan, B., Cohlmeyer, E., & Ledsham, T. (2017). Integrated strategies to accelerate the adoption of cycling for transportation. *Transportation research part F: traffic psychology and behaviour*, *46*, 236-249.

Schloegl, M., & Stuetz, R. (2019). Methodological considerations with data uncertainty in road safety analysis. *Accident Analysis & Prevention*, *130*, 136-150.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, *45*(11), 2673-2681.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tarigan, J., Diedan, R., & Suryana, Y. (2017). Plate recognition using backpropagation neural network and genetic algorithm. *Procedia computer science*, *116*, 365-372.

Teschke, K., Frendo, T., Shen, H., Harris, M. A., Reynolds, C. C., Cripton, P. A., Brubacher, J., Cusimano, M. D., Friedman, S. M., & Hunte, G. (2014). Bicycling crash circumstances vary by route type: a cross-sectional analysis. *BMC public health*, *14*(1), 1205.

Winters, M., & Branion-Calles, M. (2017). Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia. *Journal of Transport & Health*, *7*, 48-53.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. 2017 IEEE international conference on image processing (ICIP),

Zhang, X., Hao, X., Liu, S., Wang, J., Xu, J., & Hu, J. (2019). Multi-target tracking of surveillance video with differential YOLO and DeepSort. Eleventh international conference on digital image processing (ICDIP 2019),

Zhao, B., Li, X., Lu, X., & Wang, Z. (2018). A CNN–RNN architecture for multi-label weather recognition. *Neurocomputing*, *322*, 47-57.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. Proceedings of the IEEE conference on computer vision and pattern recognition,

Zhou, W., & Li, Q. (2013). Complexity and dynamic modeling of urban system. *International Journal of Machine Learning and Computing*, *3*(5), 440.