2021 International Conference on
Artificial Intelligence (ICAI)
Islamabad, Pakistan, April 05-07, 2021

Automatic Helmet Violation Detection of Motorcyclists from Surveillance Videos using Deep Learning Approaches of Computer Vision

Adil Afzal
Al-Khawarizmi Institute of Computer Science (KICS), UET Lahore
Lahore, Pakistan
adil.afzal@kics.edu.pk

Hafiz Umer Draz
Al-Khawarizmi Institute of Computer Science (KICS), UET Lahore
Lahore, Pakistan
umer.daraz@kics.edu.pk

Muhammad Zeeshan Khan
Al-Khawarizmi Institute of Computer Science (KICS), UET Lahore
Lahore, Pakistan
zeeshan.khan@kics.edu.pk

Muhammad Usman Ghani Khan
Al-Khawarizmi Institute of Computer Science (KICS), UET Lahore
Lahore, Pakistan
usman.ghani@kics.edu.pk

2021 International Conference on Artificial Intelligence (ICAI) | 978-1-6654-3293-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAI52203.2021.9445206

Abstract— **Automatic detection of helmet for motorcyclists from real-time surveillance videos is a rising application in computer science. Object detection and classification using deep learning was recently well-known over the years. Researchers used these techniques for solving several surveillance-related problems. Several deep learning models adopt for automatic detection of helmet for motorcyclists but they cannot achieve state-of-the-art results due to different difficulties such as low resolution, whether conditions, occlusion and illumination etc. In this paper, we proposed a methodology for surveillance videos that automatically detect the helmet wear by motorcyclist or not. For this purpose, we used the Faster R-CNN model. First, we apply Region Proposal Network (RPN) starting with the input image that has been delivered into the backbone. Then RPN weights are settled and proposals from RPN are utilized to train the Faster R-CNN model. For training, we used the self-generated dataset of three different locations in Lahore, Pakistan. The experimental results detect 97.26% accuracy on real-time surveillance videos for the detection of helmet for motorcyclists.**

Keywords—helmet, deep learning, detection, ITS

I. Introduction

Recognition and classification of moving objects is an active research area and it is generally utilized these days in different applications, such as human movement capturing and intelligent transportation systems (ITS). Nowadays, the intelligent transportation system has become one of the major directions of computer vision in engineering. Traffic monitoring and accident control are the main challenges in the ITS. In general, the number of two wheels transportation like motorcycles increase day by day on the roads, there has been a rapid increase in motorcycles accident due to reckless, negligence and rush of motorcycles riders [1]. Accidents due to motorcycles have been rise and this matter promotes the fundamental for addressing this issue as a major problem in developing countries like Pakistan. The negligence of not wearing a helmet by motorcyclist is a most important factor and most of them head injuries caused by not wearing a helmet. Which caused trauma to the brain and skull due to the motorcycle riders not wear a helmet [2]. In the past few decades, most of the accident death reason is injuries in the head [3]. Helmet wearing is also compulsory as per traffic rules for motorcycle riders, violation of this rule effect a massive fine. But unfortunately, there is a large number of motorcyclists that not obey this rule. In Pakistan, all major cities already deployed large scale camera networks for video surveillance purposed to keep watching on a large scale of crisis. Thus, using this already deployed system is very cost-efficient for deploying our proposed methodology in the current scenario. In video surveillance process involves a large number of humans whose performance is not durable for a long period. A recent study tells us that human surveillance is insufficient, the error made by the human increase as the duration of surveillance video increase [4]. So, a computer vision-based automated system is the need of the current scenario that detect whether the motorcyclists are with the helmet or not. This automatic detection of helmet system also reduces the burden from traffic police. The main objective of this study is to reduce the serious injuries caused by road accidents by using the computer vision-based automated detection of helmet for motorcyclists.

This research work is distributing as follows. Section-II describes a summary of related work. Section-III presents the proposed computer vision-based automated detection of wearing a helmet or not methodology. Section-IV discusses the dataset and experimental setup and performance on that dataset. Finally, Section-V enumerated the conclusion and future work.

II. Literature Review

Many researchers have proposed different methodologies to solve the automatic detection of helmet wearing problem in

978-1-6654-3293-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.

real-time traffic scenario. These methodologies are discussed in the below section.

In the early stages, researchers used machine learning approaches such as HOG, SURF, SIFT for computer vision to automatic detection of helmet wear or not for motorcyclists. RRV e Silva et al. [5] applied the HOG (Histogram of Oriented Gradients) descriptor for extracting the image attributes and then they used a multiplayer perception classifier for helmet detection. They collect data from traffic images having 255 images of which 151 images with a helmet and 104 without a helmet. And they get 91.37% accuracy rate of helmet detection. K Dahiya et al. [6] proposed an approach for detecting motorcyclists from surveillance videos. They have used object segmentation and background subtraction than by using SVM they detect helmet to wear or not. They collect their dataset. Their accuracy 93.80% on real-world surveillance data.

A few years ago, the field of computer vision is moving from machine learning to deep learning neural network techniques [7] [8]. There are as yet many challenging issues to illuminate in computer vision. In any case, deep learning techniques such as YOLO V1, YOLO V2, YOLO V3, VGG16, VGG19, CNN, Fast R-CNN, Faster R-CNN etc. are accomplishing state-of-the-art results on some particular issues. It isn't only the exhibition of deep learning models on benchmark issues that is most fascinating; the reality is that deep learning model can perform vision takes after learning from images, forestalling the requirement for a pipeline of particular and hand-created techniques. In computer vision, many problems, such as classification of images, detection of objects, object segmentation, image colourization etc. can be solved by using deep learning techniques. This research work is a part of the object detection problem in computer vision. In object detection, we detect motorcyclists with a helmet or not with a helmet.

C. Vishnu et al. [9] proposed a method to detect motorcyclists without a helmet in traffic videos using CNN (Convolutional Neural Network). First author used the adaptive background subtraction method on videos for getting objects. Second, they apply CNN for the detection of motorcycle and finally, they again apply CNN on the upper one-fourth part for helmet detection. They used IITH Helmet 1 and IITH Helmet 2 dataset. They achieved 92.87% accuracy on their experiments. Dharma Raj KC et al. [10] used deep learning for helmet violation processing. First, they detect motorcycle using HOG and then extract the region of interest. After they apply CNN for the detection of helmet and number plate detection and recognition. They used 3 types of datasets (public, private, public + private). They get maximum mAP (%) 90.5% on the public dataset. J Mistry et al. [11] proposed a solution for helmet detection by using a convolutional neural network (YOLO V2) and they also extract number plate by using open ALPR. They get mAP (%) 94.70 accuracy for helmet detection. Fan Wu et al. [12] proposed a system for helmet detection based on an improved YOLO V3 deep model. They used YOLO V3 and YOLO-dense backbone for experimentation and their dataset include images taken from the internet and a small part taken by themselves. The training dataset includes 402 images. They get maximum mAP (%) with YOLO V3 have 95.15 and YOLO-Dense backbone have 97.59.4

Narong Boonsirisumpun et al. [13] proposed a method for the detection of helmet for motorcyclists using deep learning algorithms. They used VGG16, VGG19, GoogleNet and MobileNet Models for experimental purpose and compare their results. Dataset used by Narong Boonsirisumpun at el. having a total of 493 images in which 335 images have "Bike_with_helmet" and 157 images have "Bike_with_no_helmet". They get maximum accuracy 85.19% on MobileNet model. B. Yogameena et al. [14] proposed a system for solving the automatic detection of wear helmet by using the deep-learning-based method (R-CNN), B. Yogameena [14] proposed a system that performs foreground segmentation on video frames using Gaussion Mixture Model (GMM) and blobs are labelled, then they use Faster R-CNN for detection of motorcycle from foreground regions. Then they detect helmet for the motorcyclists using the Faster R-CNN. Finally, they detect the number plate of non-helmet motorcyclists using the CNN (Character Sequences Encoding) model and ST (Spatial Transformation). They use low resolution, blur, occlusion, bald head and person with different colour of hair. They used TCE1 (2014), TCE2 (2017), Bangalore1, Bangalore2 datasets to train and evaluate the Faster R-CNN model and get maximum mAP (%) on TCE2 dataset with 79.5 on motorcyclists without helmet and 77.5 on motorcyclists with a helmet.

From the literature survey despite the fact that video analytics exits for motorcyclists with helmet or not with helmet detection still needs a lot of enhancement in real-world scenario. Especially for the Pakistan scenario, the advancement required for motorcyclist with helmet or not with helmet detection. Especially in different variation of an object, low-resolution video, illumination, profile and frontal view, occlusion and scale variations. The key contribution in this paper is that we make own dataset from Pakistan scenarios videos, this dataset comprises the real-world challenges of helmet detection such as blur images, cap wear images, low resolution, frontal view images, profile view images, back view images, occlusion, helmet carry in hand. This dataset and experimentation using Faster R-CNN is the main contribution of the proposed work.

III. PROPOSED METHADOLOGY

In this paper, we proposed a methodology for helmet detection with motorcycle. For this purpose, we used the Faster R-CNN model of deep learning. The proposed model has two phases, the first phase is the detection of the helmet by using Region Proposal Network (RPN) and the second phase is recognition of the detected helmet. Figure 1 shown the system diagram of the proposed methodology.

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.

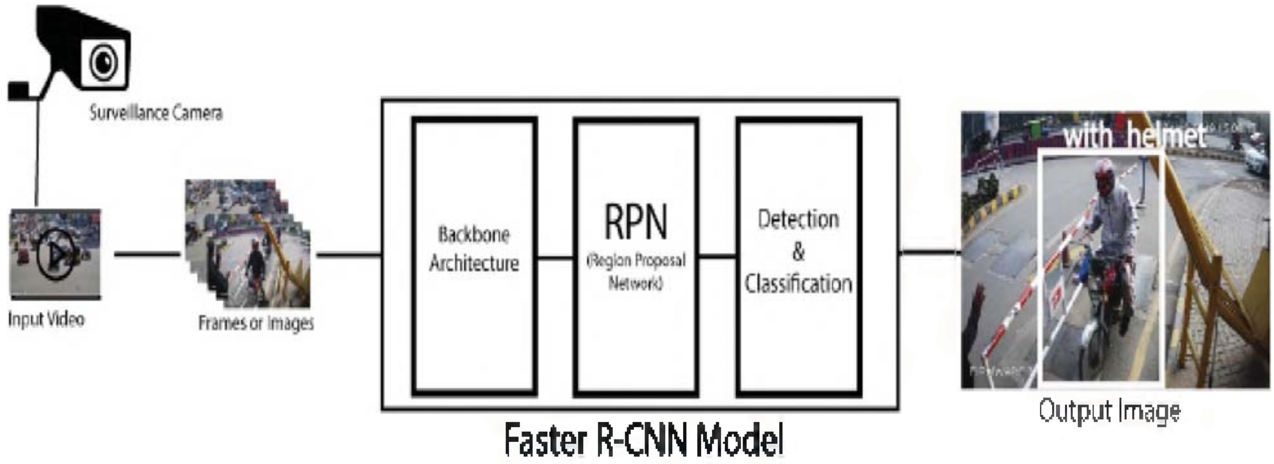

Figure 1: Proposed System Methodology

The proposed network architecture is shown in figure 2.

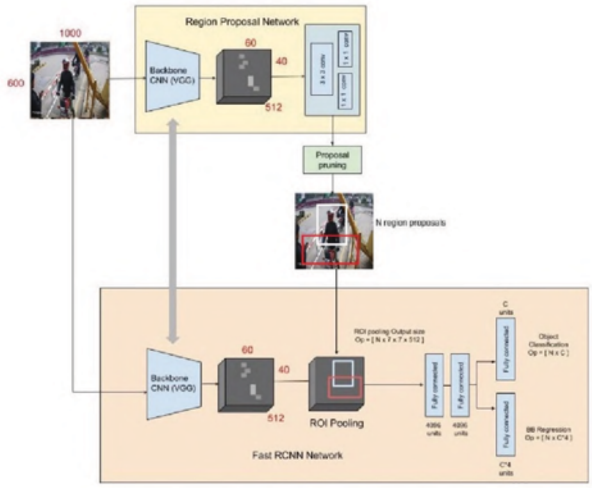

Figure 2:Network Architecture

A. Helmet Detection

Region Proposal Network (RPN) draw anchors on the image to detect the helmet and the one that most feasible consist the objects refer as output. Figure ---- shown an image completely drawn by anchor boxes. Based on scores top 2000 anchors from multiple anchors have been selected on the image. RPN is a thin weight network that sweeps the images with the assistance of the moving windows over the anchors, then it restores the anchors that have the highest possibility of consisting of the objects as represented in figure 3.

Figure 3: Anchors on the image

The intersection over union overlay is utilized to verify the closeness of the specific bounding box with the other bouncing boxes. Intersection over union calculated by utilizing equation 1.

$$\mathrm{IOU} = \frac{\text{Region of overlay of compared boxes}}{\text{Region of Union of compared boxes}} \tag{1}$$

To calculate the loss of RPN's at the time of training, the binary label is assigned to each anchor that drew on the image. The certain values are appointed to the anchor who has the values of intersection over union cover larger than 0.7. While negative value is appointed to those anchors, has intersection over union values lower than 0.3. Those anchors that have not positive nor negative values didn't participate during training time. Equations 2 described the loss function from these definitions.

$$Loss\,(M_k\,,N_k) = \frac{1}{n_{class}} \sum_k P_{class}\,(M_k, M_k^*) + \lambda \frac{1}{n_{reg}} \sum_k M_k^* P_r eg\,(N_k\,, N_k^*) \tag{2}$$

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.

Here k illustrates the number of objects. M_k illustrates the chances conclude at the network of the anchors having the objects, while M^*_k is the base truth chances whatever is 1 for and 0 as the negative case. N_k illustrates 4 directions of the image where the object is available dependent on the network expectation and N^*_k related with the ground–truth estimation at the bounding boxes. Value of $M^*_k P_r eg(N_k, N^*_k)$ portrays regression loss is simulated when the anchor's values are true else it remains deactivated.

B. Backbone Architecture

Our proposed model takes input image of 224*224*3 cropped helmet image. Backbone architecture consists of total of eight blocks, in which five blocks are convolutional layers and other remaining three blocks are connected layers. Every convolutional layer is pursued by non-linearities like as Max Pooling and rectification layer (ReLU). So, it is having three fully connected layers, from those layers the first two connected layers outcomes are 4049 dimensional and the last one fully connected layer has N = 2622 as an output depends on the class present in dataset. For handling the un-normalized vectors, we used SoftMax layer that located after the second fully connected layer and predictions come in the form of probabilistic. The probabilistic score can be carryout by utilizing equation 3.

Probabilistic score = $s_y = \frac{e^{s_y}}{\sum_x e^{s_x}}$ for all y in {1,2, ……, n}
(3)

IV. Data Acquisition

As we know that, deep learning models are data-hungry, their accuracy depends on how much data used for experimentation purpose. The more data there is, the better performance will be performed by deep learning models. So, we need a large amount of dataset to train and evaluate the deep learning model. For this purpose, we have collected the dataset from the 3 different locations. The first dataset collected from surveillance videos of an institute. For this purpose, we have placed camera at the main gate of an institute and gathered 70 videos of back and frontal view motorcyclists passing through the main gate. Each video length is 15 minutes long. Second dataset collected from a dense intercross section area in Lahore, Pakistan and gathered 40 videos of profile views of vehicles passing through the public road. Each video length is 10 minutes. Third dataset also collects from dense one-way traffic area of Lahore, Pakistan and get 30 videos of aerial view from drone camera. Each video length is 10 minutes. Total videos length is 1750 from 3 datasets. Then these videos converted into frames with the rate of 25 frames per second and collect 43750 frames from 140 videos dataset. After that, we clean dataset and takes those images in which motorcyclist appeared and skip other vehicle images. After cleaning dataset, the remaining size of images is 16422. Then we make two repositories one known as "motorcyclists with helmet" and the other is "motorcyclists without helmet". In which "motorcyclists with helmet" contained 8566 images and "motorcyclists without helmet" contains 7856 images respectively. Finally, we manually annotate these images with 2 labels first label name as "with_helmet" and the other label is without_helemt. After annotation split all images into "train" and "test" repositories with the ratio of 8:2 and 13137 images in the training repository and 3285 images in the test repository respectively. The training sample comprises the real-world challenges of helmet detection such as cap wear images, blur images, frontal view images, low resolution, back view images, profile view images, helmet carry in hand, bald head images, aerial view images and occlusion images. Figure 4 and 5 illustrate the statistics and samples of dataset respectively.

Dataset	Institution Main Gate	Dense Traffic Area	Aerial View of Traffic
Frame			
Frames Per Second	25 fps	25 fps	25 fps
Total Videos	70	40	30
Each Video Length	15 Minute	10 Minute	10 Minute
Total Frames	26250	10000	7500
Frame Format	.jpg	.jpg	.jpg

Figure 4: Dataset Acquisition

Various Conditions	Dataset Images Samples for Motorcyclists with_helmet or without_helmet
Front view	
Back view	
Profile view	
Aerial view	

Figure 5: Surveillance videos challenging environments dataset samples

V. Experimental Results and Discussion

Our proposed model has been utilized a tensor flow deep learning framework. NVIDIA 1080 Ti GPU has been used for training purpose which has 11Gb of memory capacity. Approximately system takes 16 hours for completed architecture training. As discuss in dataset section, dataset has been divided into 2 repositories. The training repository contains 13137 and the test repository contain 3285 images. Table 1 shown the training and testing images statistics.

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.

Table 1: TRAINING AND VALIDATION DATASET STATISTICS

Training	Validation
13137	3285

The main logic to use the cross-validation used to prevent the model farm from over-fitting the validation set by discovering the error rate over validation set. During the training of deep learning models to prevent over-fitting one of the main problems. if the model is over-fitted then it also affects the overall accuracy of the model. The over-fitting occurs when model is learned noise from the training dataset. But have irregular chances to happen noise due to data points that don't show true properties of dataset. On the chance of over-fitting we able to develop the more malleable model by training those parameters. So, to dishearten the way toward learning complex data points and develop the model depends on these parameters, regularization has been used to maintain a strategic distance from the procedure of over-fitting. In equation 4 represented a simple linear regression.

$$x = \cong \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots\ldots + \beta_n x_n \quad (4)$$

In the above equation, x represented the learning relation and Beta (symbol) is representing the assessment of coefficients on predictor x. Loss function involves in the over-fitting process, thus we have utilized the mean square error loss function as illustrated in equation 5.

$$Mean\, square\, error = \frac{1}{j}\sum_{i=1}^{j}(x_i^*, x_i)^2 \quad (5)$$

In above equation x^* illustrate the vectors after the addition of β values for n (number) that declared in advance and x represents the true value across each sample. It is exposed to estimate loss depends on predictions related to the actual value. We have set 0.7 beta as a value for regularization and 0.01 for learning rate. For the escalation of weights, we have been used Adam optimization. We have been training our proposed model till the 200,000 epochs. At that point, training stopped because we achieve the maximum value of loss convergence and figure 6 shown clearly. Our proposed model is train till 200,000 epochs and we calculated accuracy and loss function after an interval of 25,000 epochs. Figure 7 illustrate that our accuracy of the test data after an interval of each 25,000 epochs. After training our proposed model get a maximum accuracy of 97.26% on the test dataset.

We also train our dataset on different deep learning approaches. Figure 8 shows the comparison of other deep learning approaches and their accuracy on our self-generated dataset and evaluate that our proposed model gets the maximum accuracy on our self-generated dataset.

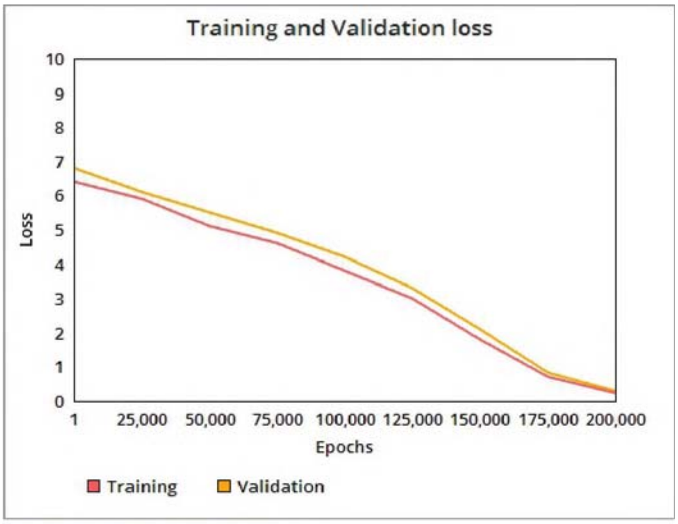

Figure 6: Loss convergence

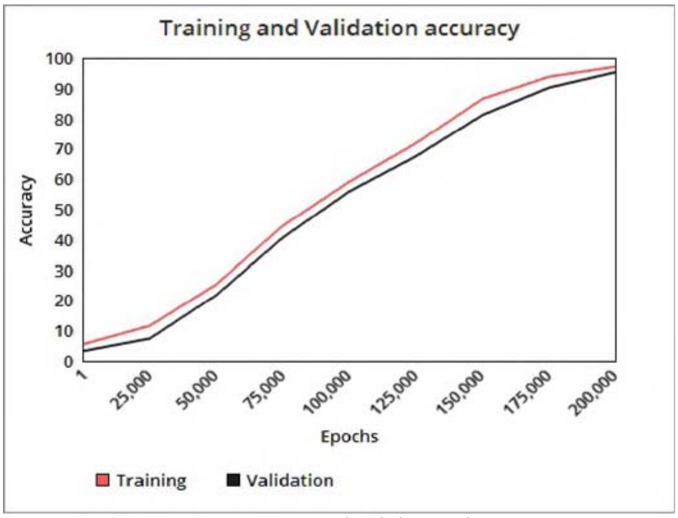

Figure 7: Accuracy on training and validation dataset

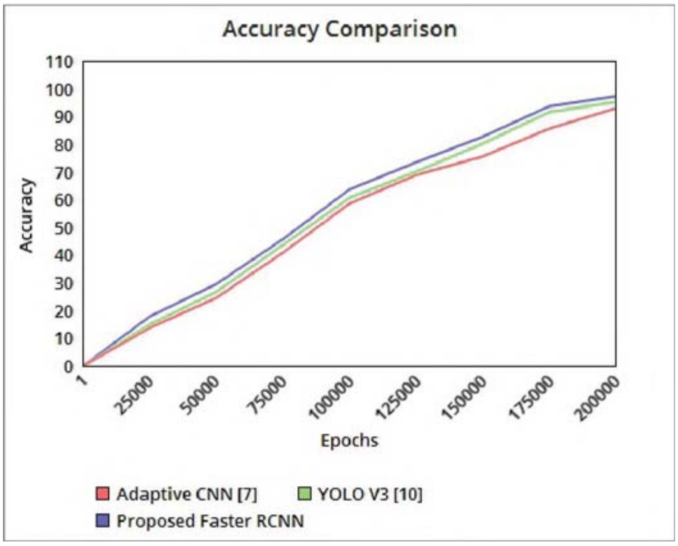

Figure 8: Accuracy comparison on different deep learning approaches

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.

The total time required for our system for processing is 0.45 millisecond to 1 second. Here we show some results images in figure 9 taken from run time implementation of model on surveillance cameras in a different location in Lahore, Pakistan.

Figure 9: Resulted images from run time surveillance video camera

VI. CONCLUSION

In this paper, we proposed our experimental results on Deep Learning based Faster R-CNN model which classifies motorcyclists with helmet and without a helmet. This proposed system efficiently detects motorcyclists wear a helmet or not. The proposed Faster R-CNN based system performed state-of-the-art results on complex video surveillance scenario such as profile view images, frontal view images, cap wear images, blur images, low resolution, bald head images, helmet carry in hand, back view images, occlusion images and aerial view images. The proposed system has been compared with another deep learning-based model such as Adaptive CNN, YOLO V3 and we get the maximum accuracy of 97.26% on our proposed model.

VII. FUTURE WORK

This paper proposed a deep learning-based Faster R-CNN model. In future, we focus on generated more dataset from the complex scenario of traffic in Pakistan and also performed experimentation on future deep learning models for improving the detection of helmet wear or not by motorcyclists. Moving forward we will add automatic number plate detection and recognition and automatic challan (in term of Pakistan) generated system for non-helmet wearing motorcyclists.

VIII. ACKNOWLEDGEMENT

This work sponsored by National Center of Artificial Intelligence (*NCAI*), Pakistan funds. We would also appreciate KICS, UET Lahore, Pakistan (colleagues and organization) for their support.

REFERENCES

[1] Gopalakrishnan S. A public health perspective of road traffic accidents. Journal of family medicine and primary care. 2012 Jul;1(2):144

[2] Ravikumar R. Patterns of head injuries in road traffic accidents involving two wheelers: An autopsy study. Journal of Indian Academy of Forensic Medicine. 2013;35(4):349-52.

[3] Behera C, Rautji R, Lalwani S, Dogra TD. A comprehensive study of motorcycle fatalities in South Delhi. Journal of Indian Academy of Forensic Medicine. 2009;31(1):6-10.

[4] Hu W, Tan T, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2004 Jul 12;34(3):334-52.

[5] e Silva RR, Aires KR, Veras RD. Helmet detection on motorcyclists using image descriptors and classifiers. In2014 27th SIBGRAPI Conference on Graphics, Patterns and Images 2014 Aug 26 (pp. 141-148). IEEE.

[6] Dahiya K, Singh D, Mohan CK. Automatic detection of bike-riders without helmet using surveillance videos in real-time. In2016 International Joint Conference on Neural Networks (IJCNN) 2016 Jul 24 (pp. 3046-3051). IEEE.

[7] Khan MZ, Jabeen S, Khan MU, Saba T, Rehmat A, Rehman A, Tariq U. A realistic image generation of face from text description using the fully trained generative adversarial networks. IEEE Access. 2020 Aug 10.

[8] Khan MZ, Harous S, Hassan SU, Khan MU, Iqbal R, Mumtaz S. Deep unified model for face recognition based on convolution neural network and edge computing. IEEE Access. 2019 May 23;7:72622-33.

[9] Vishnu C, Singh D, Mohan CK, Babu S. Detection of motorcyclists without helmet in videos using convolutional neural network. In2017 International Joint Conference on Neural Networks (IJCNN) 2017 May 14 (pp. 3036-3041). IEEE.

[10] Raj KD, Chairat A, Timtong V, Dailey MN, Ekpanyapong M. Helmet violation processing using deep learning. In2018 International Workshop on Advanced Image Technology (IWAIT) 2018 Jan 7 (pp. 1-4). IEEE.

[11] Mistry J, Misraa AK, Agarwal M, Vyas A, Chudasama VM, Upla KP. An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network. In2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA) 2017 (pp. 1-6). IEEE.

[12] Wu F, Jin G, Gao M, Zhiwei HE, Yang Y. Helmet Detection Based On Improved YOLO V3 Deep Model. In2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC) 2019 May 9 (pp. 363-368). IEEE.

[13] Boonsirisumpun N, Puarungroj W, Wairotchanaphuttha P. Automatic Detector for Bikers with no Helmet using Deep Learning. In2018 22nd International Computer Science and Engineering Conference (ICSEC) 2018 Nov 21 (pp. 1-4). IEEE.

[14] Yogameena B, Menaka K, Perumaal SS. Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system. IET Intelligent Transport Systems. 2019 Apr 1;13(7):1190-8.

Authorized licensed use limited to: BOURNEMOUTH UNIVERSITY. Downloaded on June 30,2021 at 07:36:33 UTC from IEEE Xplore. Restrictions apply.