Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

A NOVEL DEEP LEARNING APPROACH FOR BIKE RIDER HELMET VERIFICATION

IRSAM NAGEEN
MS Scholar, Department of Computer Science, The University of Okara, Okara, Pakistan.
Email: irsamnageen7@gmail.com

SOBIA YAQOOB
Lecturer, Department of Computer Science, The University of Okara, Okara, Pakistan.
Ph.D. Scholar, Department of Computer Science, Superior University Lahore, Lahore, Pakistan.
Email: sobia.yaqoob@uo.edu.pk

NAHEED AKHTAR
Department of Computer Science, University of Education Lahore, Lahore Pakistan.
Email: naheedswl@ue.edu.pk

KHALID MAHMOOD
Assistant Professor, Division of Science and Technology, University of Education Lahore, Lahore, Pakistan.
Email: Khalid@ue.edu.pk

ARSLAN AKRAM*
SSE CS, Department of School Education (PSED) Okara, Okara, Pakistan.
Ph.D. Scholar, Department of Computer Science, Superior University Lahore, Lahore, Pakistan.
Corresponding Author Email: aaadoula11@gmail.com, phcs-f21-014@superior.edu.pk

AGHA WAFA ABBAS
Lecturer, Department of Computer Science, University of South Asia, Lahore, Pakistan.
Email: wafa.abbas@usa.edu.pk.

Abstract

Bike riders are increasing in number among other vehicles. The current traffic routine survey shows many of accidents are being caused by bike riders and many of them are dead. These factors have driven recent research in applying intelligent and computerized approaches to detect bike riders with helmet so lives of bike riders can be saved. However, these systems are not viable as they require significant time or compute power to process high-resolution images. This research objective was to investigate bike rider for helmet detection via deep learning techniques using different images from different size, angle, and scene. The intent was to reveal whether machine learning models could be developed that provide high confidence results using fractional resources by using images. A deep learning pipeline has been developed which reduces high-resolution to a sufficiently small size so they can be fed as input into a CNN for binary classification (i.e., bike rider with helmet or without). Several improvements have been implemented to boost general performance, namely supplementing the training data, and adding data augmentations. Ultimately, the developed low-resolution model is effectively skill-less for very low-resolution inputs. An observed significant decrease in model inference time is a superficial benefit given the loss in classification capability. This is a promising and expected result arising from an obvious limitation with the methodology. In conclusion, the developed deep learning pipeline is suitable as a viable helmet detection system when using the input resolutions investigated. Proposed method has performed best among state of the art methods and provide 97.34% of accuracy.

Keywords: Deep Learning; Convolutional Neural Networks; Helmet Classification; Low Resolution; Kaggle.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

1. INTRODUCTION

Machine learning systems exist for the automatic detection of helmet in digital images with high confidence. Such systems can potentially automate large portions of conventional traffic violation procedures used to identify bike rider with helmet, improving support for detection via digital second opinion or reducing cognitive load by shifting work away from security personnel. However, these current systems are complex as they often fully utilize high-resolution images with dimensions that are hundreds of thousands of pixels in width and height.

The risk of injury to motorcyclists caused by 41% of traffic incidents can be decreased by wearing a motorcycle helmet [1] This is the reason why laws utilizing compelled usage have been passed. But for motorcycles, the requirements for wearing a helmet are sometimes lax, particularly in underdeveloped nations. Seibert et al., 2019;. Governments must gather comprehensive data on the degree of helmet legislation compliance in order to launch a successful helmet campaign. However, this crucial road measuring indicator is absent in 40% of the world's minority regions. Even when the information is available, the perspective of the usage of a protective helmet is frequently studied with a limited sample size, comparing, or collecting data from only one area of the academic research community [2]. The usage of the helmet and relevant information, which employs a focus study on motorcycle helmets, is a significant factor in the general lack of information distribution.

This direct observation during a roadside survey requires resources, as they can be costly and expensive [3]–[8]. And while using video cameras that allow indirect visibility, reducing the time-consuming pressure of hat use with direct view, human managers estimate the volume of information that can be processed. As a result, there is a great need to have a system in place for the use of helmets that do not need to be seen in person. The way to get this automatic diagnosis of motorcycle helmets is ML. MLs can be made and gain a high number of diagnostic tasks related to road safety accuracy to get a general identification of pedestrians, cycles, motorcycles, and vehicles.

While the first implementation of the promise to get the use of motorcycle helmets, they could not be upgraded to make their power as high as possible. This route relies on photos and videos for the number of passengers seen, trained in selected viewing areas, or not seen passenger position. In this paper, it is proposed to examine the robust use of a helmet that uses a wide database that has significant differences in the number and traffic of visible bikes. Recent effective applications based on image processing, e.g. Image classification [9] rely heavily on the same large database used in this study. Therefore, the next section of this paper, which has a similar approach to facilitating data collection in other countries, will focus on the construction of the sub-database and its annotations. This was preceded by an article on algorithmic training and the results of previous observational studies on the use of a protective helmet.

Recent studies have demonstrated that human surveillance is useless because, as the amount of time spent watching videos grows, human mistake rates likewise rise. Researchers have tried a number of different approaches to date to identify motorcyclists

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

without helmets, but they have been unable to correctly identify defaulters under challenging conditions such occlusions, lighting, and low quality. The use of less selective presentation for text categorization and the comparison of irrelevant things with the goal of cyclist identification in earlier approaches are two factors that contributed to their poor performance. Deep networks have drawn a lot of interest due to their cutting-edge performance in challenging tasks like image categorization.

A mega neural network is a convertible neural network. It is a method of in-depth learning created for picture categorization and identification problems. It can address the issues with deep neural networks and rigorous training of numerous parameters, resulting in superior classification outcomes. Many CNNs have an input layer, a converter layer, a function pooling layer, and a fully integrated layer as part of their design (FC layer). CNNs must have local contacts and parametric availability in order to reduce the number of arguments and increase detection effectiveness.

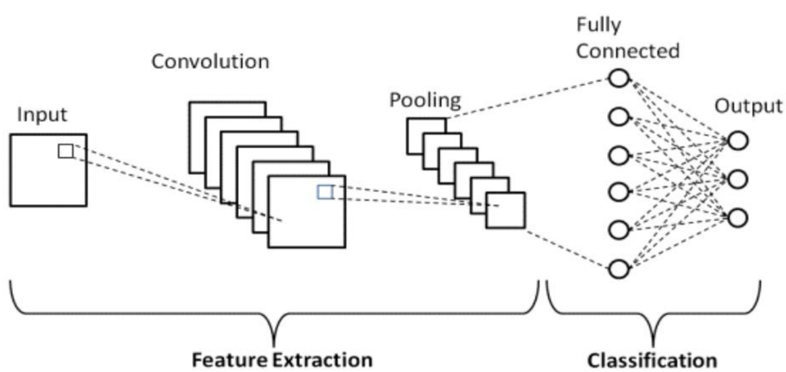

Figure 1: Traditional CNN Framework

Two main categories—two-level and single-level approaches—can be used to categories complex object detecting techniques. First, a prospective location image and item locations will be found using two-level techniques. To locate things in these locations, an object classification (using a customizable neural network) is utilized. Single-level techniques, which take longer, are less accurate than two-level ones, including such Fast R CNN [10]. Single-level techniques, in contrast, address both item identification and location concurrently.

The primary components are the higher layers, and they might acquire material characteristics. The max pooling can often change in response to different material properties. Max pooling can condense features and eliminate many characteristics as well as unnecessary features. Activating layers may successfully tackle nonlinear challenges by enhancing the expressive power of neural network models using nonlinear functional functions. In doing so, CNNs are able to change the original pixel values in the input pictures to the classifier trust layer.

To more properly identify the items and better distinguish their properties. Object identification, picture recognition, and image segmentation are the outputs of the in-depth learning process based on morphing neural networks. Kirshik and others the 2014

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

proposal of the R-CNN detection framework (area with CNN characteristics). SBP-Net, Fast R-CNN, and Fast R-CNN are three examples of spatial pyramid pooling networks (C). NN characteristics faster area)

Classification-based CNN the most popular techniques are object detection techniques, including quick R-CNN. However, the processing speed is poor and cannot be seen in real time. Methods of detection based on regression are becoming more and more crucial. Redman and others The YOLO (you only once) approach was suggested following the development of machine learning.

Although the SSD method is not the fastest, it is more accurate than the You Only Look Once (YOLO) technique when the source images sizes are tiny. This is true even if the algorithm's agility and precision are much slower and inferior to those of the YOLO algorithm. Although the rapid R-CNN method produces superior results, it is significantly more resource-intensive and requires 110 MS per picture. The Mobile Net model is introduced to significantly reduce size and sample thickness.

The network converter for producing score for object class instances in boxes and establishing boundary boxes of uniform sizes. To forecast the outcomes, a maximum repression approach is employed. The basic network structure for picture sorting refers to the first network layers of the SSD model. Before the categorization layers, the base network is disconnected, and the substitution layers are inserted at the detached base network. To forecast the results at various scales, the flexible size feature maps are incrementally shrunk.

This direct observation during a roadside survey requires resources, as they can be costly and expensive. And while using video cameras that allow indirect visibility, reducing the time-consuming pressure of hat use with direct view, human managers estimate the volume of information that can be processed. As a result, there is a great need to have a system in place for the use of helmets that do not need to be seen in person. The way to get this automatic diagnosis of motorcycle helmets is ML. MLs can be made and gain a high number of diagnostic tasks related to road safety accuracy to get a general identification of pedestrians, cycles, motorcycles, and vehicles.

2. LITERATURE REVIEW

This chapter discusses the first research goal and offers the theoretical framework for the investigation. It discusses the literature review for the research parameters used for this study. As shown in Table 1, several approaches have been put out so far for automatically identifying motorcycle helmet use in recorded video data.

A non-helmet motorbike detection system was created in an effort to streamline the process of identifying the traffic infraction of not wearing a helmet and obtaining the license plate number of the offending vehicle. The multiple Deep Learning Object Detection approach was the main concept. At the first layer using YOLOv2, the items identified were a person, a motorbike or scooter, a helmet, and a license plate. At the second layer using YOLOv3, the elements detected were a license plate. The registration number for the license plate was then obtained using OCR (Optical Character

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

Recognition) [11]. All these methods—especially the one that extracted license plate numbers—were subject to predetermined restrictions and circumstances. Given that video was employed as the input for this task, efficiency was crucial. Using the aforementioned techniques, they have created a full solution for both license plate number extraction and helmet detection.

In [11], authors proposed a Convolutional neural multi-task learning (MTL) approach for identifying and tracking unique motorbikes as well as logging rider-specific helmet wear. It was also made possible to access the HELMET database, which has 91,000 captioned frames from 10,066 different motorbikes at 12 monitoring station around Myanmar. Additionally, they offered a dataset and a rider detection accuracy assessment metric that could be compared with other future detection techniques as a benchmark. They showed that using MTL for simultaneous visual similarity learning and helmet use classification improved the effectiveness of their method in comparison to earlier studies, allowing system throughput of more than 8 FPS on consumer devices and a weighted mean F-measure of 67.3% for counting the number of cyclists and identifying the use of helmets on tracked motorcycles [11]. Their findings showed that deep learning may be a reasonably precise and resource-efficient method for gathering important data linked to road safety.

Using a video of activity on public roads as input, they devised a methodology for automated helmet recognition for motorcycle riders travelling without a helmet in the suggested approach [12]. The first step involved using OpenCV and Python to identify and categories the various vehicles on the road. Next, they extracted the image of the motorcycle and applied the YOLOV3 algorithm, a fully convolutional algorithm, to the extracted image to determine whether the motorcyclist was wearing a helmet or not. The license plate was then documented by the relevant authority if the result was unfavorable, i.e., not wearing a helmet.

In the research presented by [13], They demonstrated the advantages of utilizing computer vision and machine learning approaches to increase helmet compliance through computerized helmet violation detection. To find drivers and passengers who weren't wearing helmets, the system featured surveillance, classification, and cyclist detection. A single GPU server, a number of collaborative computational clients, and HTTP communication were all incorporated in the system's architecture. In a real test, the system correctly detected 97% of helmet violations whereas 15% of alerts were incorrect. The client-server architecture has a cost savings of 20–30% as compared to a baseline system.

A non-helmet rider detection method was created in an effort to automate the process of identifying the traffic infraction of failing to wear a helmet and obtaining the license plate number of the offending vehicle by [14]. The core idea was the three-level Deep Learning Object Detection technique. On the first level using YOLOv2, the items identified were a person, a motorbike or scooter, a helmet, and a license plate. At the second layer using YOLOv3, the objects detected were a license plate. OCR was then used to obtain the license plate registration number (Optical Character Recognition). All of these methods—especially the one that extracted license plate numbers—were subject to predetermined

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

restrictions and circumstances. Since this project used video just like its input, execution speed is essential. They developed a comprehensive solution including both helmet detection and license plate number extraction using the aforementioned approaches.

To overcome these challenges, a ladder-type multi-attention network (LMNet) for target identification was first created. The LMNet reduced information loss, enabling informational interactions and fusion at every level, and completely retrieved picture features. It was also suggested to use the Residual Transformer 3D-spatial Attention Module (RT3DsAM) by [15], which assimilated data from worldwide sources that is crucial for feature representation and ultimate classification detection. Additionally, it improved information correlation and self-attention and built both. Third, to restore crisp edges and more realistic texture information, the LMNet-detected rider pictures were clipped out and reassembled using the enhanced super-resolution generative adversarial networks (ESRGAN). Finally, the YOLOv5 classifier was used to categories the reconstructed photos of cyclists. According to the experiment's findings, their method improved the detection accuracy of riders' helmets in aerial photography scenes when compared to the existing approaches, with the quantitative techniques mean average precision (mAP) evaluation indicator reaching 91.67% and the image classification top 1 accuracy (TOP 1 ACC) increasing by 94.23%.

The development of a framework using sophisticated convolutional neural networks (CNNs) for identifying bicyclists who are disobeying cap laws was described by the authors. Motorcycle, detection, helmet vs. no-helmet, categorization, and method counting were all part of the system. Model was using a faster R-CNN with ResNet 50 network for the motorbike detecting phase [16]. CNN's proposed categorization approach compares wearing a helmet against not wearing one. Finally, an alarm is sounded to warn the policeman and stop a motorbike collision. They evaluated the framework's precision and speed.

The premise that caused the most worry was Object Detection Victimization. Three stages of deep learning. At first, the things that detected a victim were a human, a motorbike, or a moped. Helmet at second-level victimization, YOLOv2, Vehicle plate at the final stage of YOLOv2 victimization. After then, the license plate number was obtained OCR (Optical Character Recognition) (Optical Character Recognition) [17].

The authors of the research have suggested an approach for surveillance cameras that may automatically determine if a biker is wearing a helmet or not. They employed the Faster R-CNN model for this. They start by using the input picture that has been sent into the backbone as their beginning point for the Region Proposal Network (RPN). The Faster RCNN model was then trained using the RPN weights that had been decided upon [18]. They employed their own dataset of three separate Lahore, Pakistan, locales for training. The experimental findings show 97.26% accuracy in detecting motorcycle helmets in real-time surveillance recordings.

Authors in article presented [19] planned to resolve this issue by automation the method for identifying the cyclists wearing helmets and those who are not. The algorithm detected the moving objects in the image using a video of vehicles on an empty street as

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

information. The strategy that was suggested was based on the location of one or more bike riders who were riding sans helmets. The YOLOv3 model, a dependable variation of the YOLO model, was used early in the recommended approach to find bicycle riders. The most popular object-discrimination technique was successful in distinguishing between bikers wearing helmets and those who were not. A binary image vertical projection was used to count the riders if there were more than two.

The technique offered an intelligent surveillance video system for automatically identifying motorcycle riders wearing safety helmets and those who are not. If a motorcyclist was discovered to be riding without a helmet, the traffic police and the legal authority used the motorcyclist's license plate number (LP) to identify him or her and begin further actions, such as deducting the fine amount from the offender's account that was connected to the vehicle license and Aadhar number (Applicable to Indian Scenario). Prior to labelling, the foreground items were first segregated using a Gaussian mixture model (GMM). To confirm the existence of motorcyclists, the suggested system next adopted a faster region-based convolutional neural network (faster R-CNN) for the detection of motorbikes in the tagged foreground items [20].

In the research presented by [21], Authors created a deep learning-based approach for automatically detecting motorcycle helmet wearers. There were two phases in the procedure. The YOLOv5 detector was upgraded in the first stage to better identify bikes and riders from surveillance footage. The upgraded YOLOv5 detector was used again in the second stage to determine if the motorcyclists were wearing helmets after taking the bikes identified in the first step as input. The fusion of triplet attention and the use of soft-NMS rather than NMS are two enhancements to the YOLOv5 detector. It was suggested to create a new motorcycle helmet dataset (HFUT-MH) that is both larger and more complete than the current one, which is generated from various traffic monitoring in Chinese cities. The proposed approach was then tested in experiments and evaluated against other cutting-edge techniques. Our solution surpassed other cutting-edge detection techniques with performance metrics of mAP of 97.7%, F1-score of 92.7%, and frames per second (FPS) of 63.

The helmet detection issue was dealt by using a Single Shot Multi Box Detector (SSD) device. This model can discriminate between the bounding box areas of the motorcycle and the rider using just one CNN system by [22]. When the location was selected, they categories whether or not the rider was wearing a helmet in real-time. Convolutional Neural Network was used to identify motorcyclists among moving objects and to identify those who were not wearing helmets. They are able to identify motor motorcyclists without helmets on their license plates even further by employing the You Only Look Once (YOLO) paradigm. In all, they have used the SSD Model, the YOLO Model, and the Custom CNN Model across the framework.

Authors demonstrated a real-time automatic approach for spotting helmet-less motorcyclists in traffic surveillance films. When there were only so many computer resources available, the issue grew increasingly difficult. For the purpose of creating an automated helmet identification system, they have assembled a unique dataset [23]. The suggested technique extracted bikes from surveillance movies using a two-stage

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

classifier. Motorcycles that were discovered were then sent into a step that identified helmets. They show two algorithms—one based on manually created characteristics and the other on a deep convolutional neural network—for categorizing bikers wearing and not wearing helmets (CNN). Their tests revealed that the feature-based model provided quicker detection, while the suggested CNN model provided the greatest accuracy performance. Most crucially, only CPUs were used for all calculations to guarantee the proposed system's lightweight nature.

The study that was presented offered a methodical approach for tracking employees' hard helmets in real-time using deep learning (DL) models constructed on the You-Only-Look-Once (YOLOV5) architecture. It can reveal whether or not a worker is donning a hat. With the help of PyTorch, the suggested system utilized 7063 photos and five distinct YOLOV5 models for object identification, including the YOLOV5n, YOLOV5s, YOLOv5m, YOLOv5l, and YOLOv5x. [24]. The study's findings demonstrate that the YOLOV5x among the DL models performed well in terms of the mAP, with a performance of 95.8%, while the YOLOV5n had the fastest detection speed, with a rate of 70.4 frames per second (FPS). The suggested approach can be successfully utilized in the real world to identify a worker's hard helmet.

A super-resolution (SR) reconstruction-driven helmet recognition algorithm is used by Yicheng Liu et al. to identify helmets for monitoring activities in their end-to-end helmet monitoring system [23]. A super-resolution reconstructing module and a detecting module made up the two modules that made up the monitoring system. While the latter recognizes helmets, the former used the SR technique to get high-resolution photos. A publicly accessible database as well as a realistic dataset gathered from a real-world building site were both used for validation. The results demonstrated that the suggested solution operated well and outperformed other approaches. It will be an effective construction monitoring tool that is simple to adapt to different projects.

The presented work proposed the SAS-YOLOv3-tiny safety helmet identification algorithm to reconcile detection accuracy and model complexity [25]. The original convolution layer was replaced with a light Sandglass-Residual (SR) module based on depth wise separable convolution and channel attention techniques in order to get more informative features, enhance detection performance, and reduce the number of parameters and computation. The max-pooling layer is then swapped out for the convolution layer of stride two. Three-scale feature prediction was used instead of two-scale feature prediction to further enhance the effect of the identification of tiny items. A new spatial pyramid pooling (SPP) module has been added to the feature extraction network to collect local and global features with rich semantic information.

The authors have proposed a CNN-based multi-task learning (MTL) technique for detecting and monitoring certain motorbikes as well as registering rider-specific helmet use. There was also access to the HELMET dataset, which includes 91,000 annotated frames of 10,006 distinct motorbikes from 12 observation sites in Myanmar. Along with the dataset, they also offered an assessment measure for helmet use and rider detection accuracy that may be used as a benchmark for assessing future detection systems. [11]. They demonstrate that, when compared to earlier studies, our method is more effective

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

when MTL is used for concurrent visual similarity learning and helmet use classification. This allows for processing speeds of more than 8 frames per second. on commodity hardware and a weighted average F-measure of 67.3 percent for order to detect the number of riders and helmet use of tracked motorcycles.

Applications for automatic license plate identification and automatic helmet identification were suggested to be created using the CNN machine learning set of rules by Nishan Albert and Basavesha D. In order to find the helmet, a more effective CNN technique was applied [26].

To boost target recognition on the construction site, Haikuan Wang et al. suggested a novel SHW detection technique based on improved YOLOv3 (referred to as CSYOLOv3) [27]. The darknet53 backbone network was first strengthened using the cross stage partial network (CSPNet), which decreases computation costs and speeds up training. The YOLOv3 model's spatial pyramid pooling (SPP) structure was merged with the top-down and bottom-up feature fusion approaches to perform the feature augmentation. Construction site cameras were used to produce the 10,000-picture dataset for safety helmet wearing detection, and the model training required manual annotation. Experimental results and contrastive curves demonstrate that the innovation may greatly improve performance by 6 frames per second while reducing mAP by 28 percent when compared to YOLOv3.

The authors have suggested to automate the process of recognizing the traffic offence of not wearing a helmet and acquiring the license plate number of the offending vehicle, a non-helmet rider detection system was developed [14]. The three-level Deep Learning Object Detection approach was the main concept. At the first level of the YOLOv2 algorithm, a person, a motorcycle or scooter, a helmet, and a license plate were the objects found. Using YOLOv3, items were found at the second level, including a license plate. The registration number for the license plate was then retrieved using OCR (Optical Character Recognition). All of these techniques had limitations and conditions that were preset, notably the one that retrieved license plate numbers. Because this work employs video as its input, the efficiency of the execution was crucial. They have developed a complete system for both helmet detection and license plate number extraction using the methods indicated above.

Ruiyun Cao et al [28] have proposed that YOLOV4 introduces a novel helmet identification method, achieves 95.1 percent accuracy, and adjusts and improvements in accordance with the peculiarities and challenges of this task. The programme can quickly and precisely identify if employees were wearing helmets properly, which was crucial for maintaining construction safety during actual work.

Author have proposed the capability of automated helmet violation detection using computer vision and machine learning techniques to increase helmet compliance [13]. The system, which included motorcyclist detection, classification of helmet violations, and tracking, can identify riders and passengers who were not wearing helmets. The system's architecture consisted of a single GPU server and several computational clients working together to finish the task while communicating over HTTP. In a test conducted in the

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

actual world, the system was able to identify 97 percent of helmet infractions with a false alert rate of 15%. In comparison to a base design, the client-server architecture decreases costs by 20–30%.

According to Dr. A. Anjaiah and colleagues, a system was created to identify motorcycle riders without helmets in order to automate the recognition of this traffic infraction and the extraction of the vehicle's license plate information [29]. The major element was that 3-degree Deep Learning was used for Item Discovery. A person, a motorcycle or scooter were found at the first level using YOLOv2, a helmet was discovered at the second level using YOLOv3, and a license plate was discovered at the third level using YOLOv2. The license plate's registration number was then retrieved using OCR (Optical Character Recognition). They employed more of the techniques to produce a stand-in system that can identify helmets and obscure license plate numbers.

Nikhil Chakravarthy Mallela and colleagues recommend using such approach to find the Triple Riding. The YOLO (You Only Look Once) technique, a sort of convolutional neural networks, is utilized by the deep learning framework darknet to detect the number of bike riders, and it was used to identify the triple cyclists [30]. The system determines if the vehicle is a rule-breach vehicle or not. The junctions, which act as a data center, are where the data is gathered. The image of the vehicle that was judged to have violated the guidelines was also captured, together with details such as the manufacture ID of the automobile and the speed that was communicated at that frame. The data transfer was facilitated by the use of the GSM module and the NodeMCU mounted on the vehicle. The automobile number will be verified by the transport bureau. In order for the system to function in the event of no internet access or poor internet connectivity, a GSM module was added; otherwise, the development boards placed at intersections, which serve as the main hub of the established public internetwork, could access the data related to the car. Data is sent from the car to the main computer system over the public internetwork. This was done by using the NodeMCU dynamic network configuration concept. The use of Node MCU and the public network system has greatly improved the system's viability, accessibility, and stability. Therefore, encouraging users too.

The authors provide a better helmet wearing detection approach called YOLOv5s-FCG (FourLayers, CBAM attention, GhostBottleneck) (You only look once). The network was enhanced, a shallow feature detection layer was introduced, three-scale feature detection was modified to four-scale feature detection, and the up sampling was raised by four times based on the lowest volume of YOLOv5s in the YOLOv5 series [9]. Include CBAM Attention Module; swap out fragile GhostBottleneck constructs with Bottleneck ones. Their study shows that employing YOLOv5s-FCG improves average detection accuracy (mAP) on our riding safety helmet data set by 2.0% and on the NWPU-VHR 10 public data set by 1.5%. The recommended method improved detection precision while ensuring detection rate, volume, computation, and parameter amount. Additionally, it generalized well in difficult driving situations, such as low illumination and small targets.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

Table 1: State of the art Methods

Author	Year	Method		
		Dataset	**ML Method**	**Evaluation**
[31]	2020	Self and other publicly available	Transfer Learning using Resnet50 based on ImageNet Weights	Training accuracy and training loss
[32]	2021	Automatic Constructed Images and other benchmark datasets	SR and Yolov5 Model	PSNR and SSIM for SR evaluation and Precision, Recall and AUC for detection algorithm
[33]	2021	Self and Benchmark dataset	Modification in darknet using SR	Precision and AUC was used for evaluation
[34]	2021	Self-extracted from surveillance	Modified Darknet with yolov3 modified	Model was evaluated using precision with respect to FPS
[35]	2020	Self-generated HELMET	CNN Based model with multitask learning	FPR, TPR, F measures were used for evaluation
[36]	2021	Modified benchmark dataset	YOLO v5 based deep learning model	Precision and recall were used as evaluation measure
[37]	2021	Self-generated dataset named HFUT-MH	YOLO v5 with triplet attention mechanism	Precision, Recall and F1 were used for evaluation
[38]	2020	Self-compiled dataset from available datasets from surveillance videos	Hand crafted feature approach for bike detection and CNN was used for helmet detection.	Accuracy, Recall, F1 and kappa rate were used for evaluation.
[39]	2020	Publicly available dataset was used	Ensemble algorithm for head detection and CNN for helmet detection	TPR, FPR, Accuracy and FNR were used as evaluation measures
[40]	2022	Self-generated dataset SHEL5k	Different state-of-the art methods were compared	Precision, Recall and F1 were used for evaluation

3. METHODOLOGY

The Python (≥ 3:0) programming language was selected over alternatives (e.g., Java) due to the availability of well-maintained open-source libraries for deep learning and image processing. These libraries mitigate the need for self-implementation and testing, which would hinder the feasibility of the project objectives.

This section outlines the deep learning pipeline design for the classification task. Our intention is to identify riders who are not wearing helmets, and when we are successful in doing so, we also look for the number plate with the intention of deriving the registration number from it. To start our investigation, we used video recordings that were recorded from the rear perspective of moving motorbikes in various locations. In the beginning, we got our images from OpenCV video files. The images were then tagged using the XML document generating LabelIMG programme. It was then processed after being utilized for training. We utilized the open-source object identification API Tensorflow to determine whether the rider was wearing a helmet or not. After that, we keep the registration number in a database. There are several tools available with Tensorflow. Precision and accuracy can be traded off. We'll train new convolutional neural networks on our data to categorize bike riders wearing helmets. In this chapter, high-level design decisions are explained

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

and justified, specifically regarding the deep learning pipelines developed for helmet classification. These decisions have been heavily motivated by the research conducted in the literature review. Implementation details that are peculiar given the design specification are outlined where required. This research can be divided into two parts. Firstly, we have collected dataset from different databases and preprocess it using OpenCV and then based on convolution and pooling we have trained a deep learning model to classify bikers with helmet or not.

3.1 Data Set

This section describes the data set used and explains any manipulations of this data set to assist model training and evaluation. Since there are no benchmarked datasets available, we have created a fresh dataset for the project. Images from Google are retrieved and manually cropped and tagged for the databases used for system testing and development. The dataset includes a total of 1082 two-wheelers, of which 428 are driven by people wearing helmets and 654 are not. Here, we have included all motorized two-wheelers under the umbrella term "motorcycle." Scooters and motorcycles are part of it. The system consists of 654 helmet users, 429 single riders without helmets, and the remaining 429 riders. In order to create a balanced class, we utilized an equal number of cyclists wearing and not wearing helmets. Figure 11 presents images from the test dataset. Using the method outlined above, all of the photos utilized in the experiment were extracted. The remaining photos are utilized for training, while 30% are preserved for testing.

Figure 2: Sample Images from Dataset

Dataset provided is separated in two main categories. So, we have divided it into two different categories training and validation to feed it to convolutional neural network. Dataset was splinted into two sets as training and validation respectively.

ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

Figure 3: Training and Validation Split of Dataset

The labels for the output classes are provided in categorical string format, but they were converted to numerical format, which is required for model training and inference. As the model performs binary classification, a simple binary encoding of the target classes has been used (see Table 2).

Table 2: Table showing binary encoding of the output classes

Category Label	Binary Encoding
'Without Helmet (or 'None' or 'Negative')	0
'With Helmet (or 'Positive')	1

3.2 Data Preparation:

Any data augmentation used is applied to the training images; however, data preparation is performed on all images (i.e., train, validate, and test) as a pre-processing step immediately before feeding the images as model input. Data preparation consists of image resizing and normalization.

Image resizing involves converting the low-resolution .PNG images into target low-resolution dimensions desired for investigation. As an initial choice, image sizes of 128 x 128 pixels were used. This is the smallest input size permitted by the chosen CNN architecture. Based on the assumption that the model will likely perform worse as image resolution is decreased (discussed in the context survey - see chapter 2), investigating the input dimensions of 128 x 128 pixels will establish a lower-bound for model performance. From this, the input resolution is explored further within these bounds with the intent to find an optimal resolution that maximizes prediction confidence whilst minimizing resource requirements.

To handle increasing input images resolutions (above the original 128 x 128 pixels), whilst maintaining the same batch size (at 16), an accumulated gradient approach was implemented for use with larger image resolutions that would otherwise create out-of-memory issues. The accumulated gradient approach essentially sets the batch size as 1 so that individual images are loaded into memory. However, the network is only updated once the required batch size number of images (i.e., 16) have had their losses computed.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

This is effectively the same as just using the required batch size, but memory management is utilized for practical purposes.

Data augmentation techniques are used to provide fabricated instances of real-world data in order to produce more input samples for model training. To provide a more reliable dataset for the classifier during training, data augmentation is used for datasets with sparse data. This is often advantageous for training models with limited data. By minimally altering the current data, data augmentation methods are used to expand the amount of input data.

Important deep characteristics can be learned by deep neural network models utilizing a scaled-down training set. The scaling factors for the X and Y directions are represented by Gx and Gy in the operation G, which can be done in a variety of directions. Scaling can produce usable augmented pictures for training since tumour sizes vary. In order to retain the dimensions of the supplied image, scaling is coupled with cropping. Only the essential portions of a picture can be cropped.

$$Z = \begin{pmatrix} G_x & 0 \\ 0 & G_y \end{pmatrix}$$

The original image is mirrored along the axes when randomly flipped. Natural pictures may often be flipped along the horizontal axis but not the vertical axis because up and down components of an image are not always "interchangeable." A brain comprises two hemispheres in the axial plane, and the brain is typically thought of as anatomically symmetrical. When you rotate along the horizontal axis, the left and right hemispheres are switched, and vice versa. A picture can be usefully rotated in this situation around the center pixel. After that, the original picture size is fitted using the proper interpolation. When zero padding is employed to fill up the gaps between pixels, the rotation operation Z is widely utilized in:

$$G = \begin{pmatrix} \cos\alpha - \sin\alpha \\ \sin\alpha \cos\alpha \end{pmatrix}$$

3.3 Model Training

This section outlines the design pertaining to model training carried out using the pre-processed images.

3.3.1 Deep CNN Architecture

In the field of visual categorization, CNN has surpassed several traditional machine learning methods. Although they maintain the spatial dimension of a picture, they are comparable to neural networks. Deep learning has to have millions of parameters tuned, hence there must be a lot of training instances. Since CNN is translational and perspective invariant, we have enhanced datasets using techniques like flipping, rotation, etc. Each category has 1500 photos in the final dataset. The CNN network receives the split ROIs directly for classification. In this case, hand-crafted features are not required because CNN is capable of acquiring the characteristics required for class differentiation during training. Low-level characteristics like corners, aligned edges, etc. are learned by

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

lower layers of a CNN architecture from an input picture. Convolutional layers of higher order are in charge of learning more intricate properties like texture. The widely used CNN image classification models have intricate topologies since they are designed to distinguish between a huge numbers of classes. When using pretrained models for straightforward predictions, there is a considerable risk of overfitting. Additionally, for real-time applications, they need a GPU-based system and vast amounts of processing power. In contrast to expensive GPU-based systems, we intend for our final application to run in real-time on a cost-effective embedded device with less computing power. The maximum number of computations that may be performed without sacrificing accuracy is therefore strictly capped. We were inspired by this to create our own CNN for helmet identification. Our CNN model uses a 6464 RGB picture as its basis. Fig. 4 depicts the flowchart of our Custom CNN. There are 2 maxpool layers and 7 convolution layers in our custom CNN network. Convoluting 32 filters of size 3 3 allows for the capture of low-level characteristics from the picture. ReLU activation then happens after that. The size of the feature map is then cut in half via maxpooling. After another convolution layer, the feature size is once more cut in half. There are now 64 feature maps in the next three convolution layers. The next two convolution layers bring the number of activation maps back down to 32. Strided convolutions are used in every convolution layer to make the feature sizes smaller. Two completely linked layers come after the convolution layer. After each convolutional layer and to the first fully connected layer, the ReLU activation function is applied. Additionally, to avoid overfitting in the first completely linked layer, a drop out value of 0.5 is applied. The final layer with all connections makes use of the sigmoid activation function. With the quantity of trainable parameters and losses in early epochs in mind, we arrived at this setup by trial and error. Table 3 displays the Custom CNN model's architecture, which was utilized to classify helmets. Compared to Alexnet's employed 62.3 million parameters, our network employs just 99,654 trainable parameters. As a result, when compared to training from scratch, less computational resources are needed for our Custom CNN during both training and testing. When installing a CNN network on an embedded platform, this is a crucial feature.

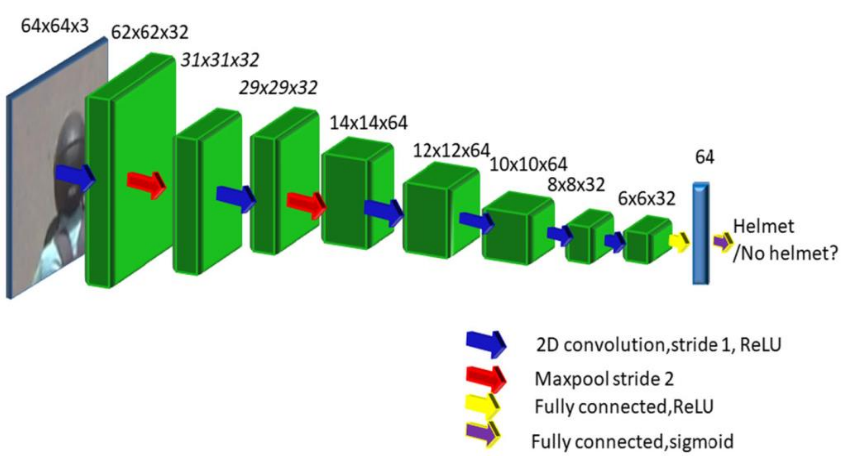

Figure 4: Flow diagram of Custom CNN for helmet classification

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

Table 3: Customized CNN Architecture

Layer	Input	Kernal Size	Kernals	Stride
Conv	64 x 64	3 x 3	16	1
Pool	62 x 62	2 x 2		2
Conv	32 x 32	3 x 3	32	1
Pool	31 x 31	2 x 2		2
Conv	16 x 16	3 x 3	64	1
Pool	15 x 15	2 x 2		2
Conv	8 x 8	3 x 3	64	1
Pool	7 x 7	2 x 2		2
Flatten	576			
Dense	64			
Dropout	64			
Dense	32			
Dropout	32			
Dense	1			

3.3.2 Loss Function

The selected loss function is binary cross entropy loss (also known as log loss), which is a standard machine learning choice for binary classification tasks (Litjens et al., 2017). Binary cross entropy compares each predicted output probability to its corresponding actual output. An overall score is calculated which penalizes the predicted probabilities based on distance from their expected output. Usefully, loss increases exponentially the more incorrect (greater the distance) a predicted probability is from the expected output (i.e., large penalties deter the model from predicting low probabilities for positive class samples).

3.3.3 Optimizer

Due to the complexity involved with the deep nature of the CNN model, it is key to minimize the number of hyper-parameters to tune. Adaptive learning rate optimization algorithms provide good out-of-the-box performance via fast convergence and little parameter fine-tuning when compared with traditional optimization algorithms, such as stochastic gradient decent (SGD). The most generally used adaptive optimizer is adaptive moment estimation (Adam), which combines momentum (for larger steps in the direction of the steepest gradient) and root mean square propagation (for higher acceleration on steep slopes). Adam was selected as a suitable optimizer for these qualities, though many different optimizers and many variations of Adam itself exist. An improvement would be to perform an empirical investigation as to which optimizer yields the best performance for the data set. This was not prioritized due to the complex interplay of model variables, where a change to one variable may alter the best choice for all other variables, so great effort would need to be expended for what would be tantamount to a limited gain in terms of the project objectives.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

4. RESULTS AND DISCUSSIONS

In this section, the metrics used to evaluate the low-resolution models are defined and the corresponding results for the whole-slide classification task are presented and analyzed.

4.1 Evaluation Metrics

The main metric used to evaluate this project's low-resolution models was used by us, allowing for meaningful performance comparisons to state of the art involved. Additional metrics are used for resource requirement comparisons, or to support further analysis of the developed models. For example, secondary metrics are formed from the confusion matrices of the binary classifiers (precision, recall, etc.) to analyses the limitations of the classification models, where appropriate.

The area under the receiver operating characteristic curve serves as the measure (i.e., Area under ROC or AUC). The AUC score will be used to describe this. The true positive rate (also known as sensitivity) vs the percentage of false positives for all categorization threshold levels are plotted on the ROC curve. Accuracy is a measure of the model's efficiency across all classes. It is useful after all classes are given equal weight. It is calculated by multiplying the total number of predictions by the number of accurate forecasts.

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$

Cross-entropy is the typical number of bits needed to transmit a message from distributed A to distribution B. When algorithms are developed to make predictions from the model, cross entropy is a notion utilized. The model is built using a comparison of the actual and predicted results. Cross-entropy may be visualized mathematically as follows:

$$H(p, q) = -\sum_{i \epsilon n} p(i). \log q(i)$$

Let T1 represent the proportion of cases with helmets correctly predicted, T2 represent the proportion of cases without helmets correctly predicted, T3 represent the proportion of cases with helmets incorrectly identified as cases without helmets by our algorithm, and T4 represent the proportion of cases with helmets correctly identified as cases with helmets by our algorithm.

Precision (P) is the proportion of instances accurately classified as helmet cases to all cases classified as helmet.

$$Precision = \frac{T1}{T1 + T3}$$

The proportion of accurately anticipated helmet cases to all helmet cases is known as recall (R). Additional name: True Positive Rate (TPR).

$$Recall = \frac{T1}{T1 + T4}$$

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 03-2023
DOI 10.17605/OSF.IO/T6YFB

The weighted sum of recall and accuracy is known as the F1 Score (F1).

$$F1 = \frac{2 * \ Recall * \ Precision}{Recall + Precision}$$

4.2 Classification Results

This section details the results generated by the low-resolution models developed for the classification task.

Figure 6 shows training and validation accuracies with respect to each epoch with learning rate of 0.001. Model outperformed on this challenge dataset of images. We have used traffic images 70% of images for training. Figure 14 shows division of dataset. CNN model was evaluated using binary cross entropy and results are shown in following figure 6.

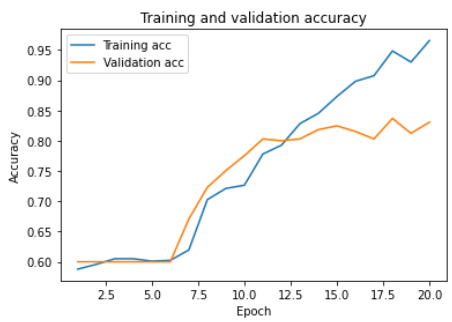

Figure 5: Training Accuracy

The over-arching result of this project is the developed deep learning pipeline for binary classification is effectively skill-less at the investigated low input resolutions, even when additional techniques were implemented to improve general capability.

This project's deep learning pipeline may have increased proficiency at larger input resolutions than those investigated, possibly revealing some optimal resolution that captures sufficient prediction performance and average inference time. However, this seems unlikely given that there appears to be an inherent trade-off between the two desired attributes, requiring an unsatisfactory decision to prioritize either model performance or inference speed to define the input resolution to use. Currently, this project's models suffer from the 'garbage-in, garbage-out' characteristic attributed to the black box nature of deep learning models. The low-resolution input images, are significantly blurred, making the identification of necessary features for helmet detection questionable even by trained security personnels.

Other parameters were also being used to better evaluate the deep learning framework. Figure 7 shows performance results of different parameters, as we can see model has outperformed among all the tests decided.

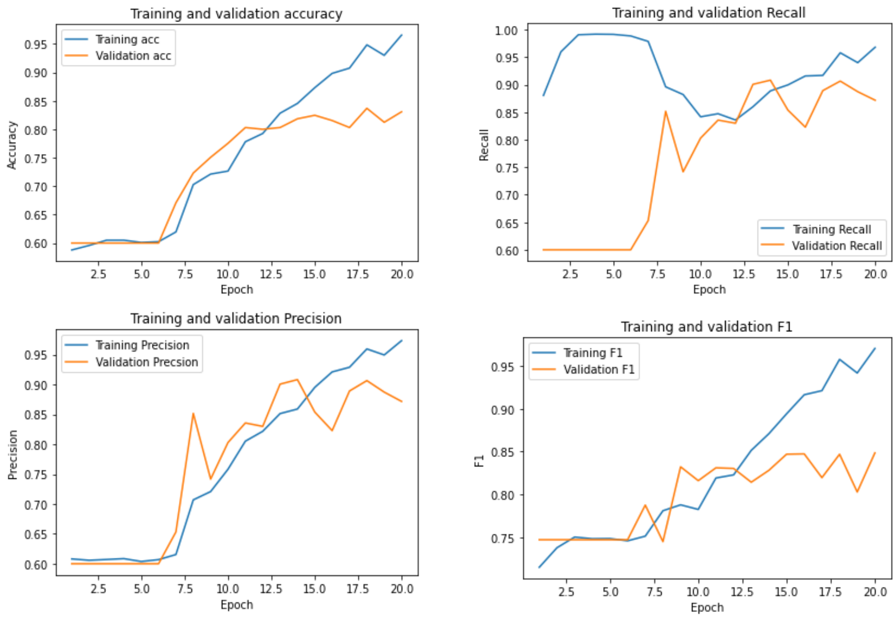

Figure 6: Epoch wise Performance graphs for Helmet Classification

4.3 Comparative Analysis with State of the Art Methods

In this subsection, we have reported comparative analysis of proposed framework with different state of the art methods. Table 4 shows results and citations of previous studies.

Table 4: Comparative analysis with state-of-the-art methods

Citation	Performance Results			
	Accuracy	**Precision**	**Recall**	**F1**
[33]	78.2	71.6	80.9	75.2
[13]	---	96.00	61.00	78.00
[15]	---	83.84	80.54	---
Proposed Customized CNN	**96.57**	**97.34**	**96.81**	**97.02**

In summary, the developed deep learning pipeline has been shown to be sufficient even when the input resolution has been increased to 128 x 128 pixels. A slight observed increase in accuracy score (as the input resolution has been increased) indicates that the models are becoming less incapable at separating the outputs, suggesting that some as yet uninvestigated minimum resolution is required for class separation and better than chance predictions. However, the benefit of further investigation to find such an input resolution is questionable given the increasing in-feasibility of training CNNs using these larger resolution images entirely as input. For example, attempts to investigate 2048 x 2048-pixel inputs were fruitless as the model did not reach convergence after several days of training. Although, the feasibility of model training depends on available equipment, allocated training time, and perceived benefit of the resulting model, so no claims can be made that future work investigating larger input resolutions would not be beneficial. Furthermore, trained model has achieved 96.57% of training and 83.08% of testing accuracy at its best model after training by 20 epochs so far.

5. CONCLUSION

Overall, the primary achievement of this project is in establishing the following conclusion: the developed deep learning pipeline is sufficient at performing binary helmet classification at the low input resolutions investigated, even when additional techniques were added to improve general performance. As a result, the deep learning pipeline is shown to be suitable for use as a form of filter to quickly classify helmet vs non-helmet bike rider. This method shows promising as expected. This method also demonstrated that there is an inverse relationship between classification performance measures like accuracy and loss used in the developed pipeline, which is a suspected intuition. This has been shown in the most extreme regard, where the lowest input resolutions result in effectively skill-less models. An optimal, higher input resolution with an acceptable balance of this relationship may exist, but this would require careful definition of what is sufficient prediction confidence and inference time for a helmet detection system.

6. FUTURE WORK

Finally, this project aimed to improve on the viability of the Helmet detection challenges. Optimistically, this project's developed models aimed to replace them. More realistically, this project aimed to create a model that could first be applied to filter out easily classifiable cases before using some existing models for the remaining difficult cases. In such a scenario, the rapidness of using a low-resolution model to filter out most cases could justify the longer inference times incurred by the high-performing submissions. However, neither of these scenarios were realized, emphasizing the conclusion that future work is required investigating multi-resolution approaches. These approaches encapsulate the described need for swift low-resolution processing of easy pathology cases in tandem with high-confidence, high-resolution processing for difficult cases.

References

[1] Y. Liu, Z. Li, B. Zhan, J. Han, and Y. Liu, "A Super-Resolution Reconstruction Driven Helmet Detection Workflow," *Appl. Sci.*, vol. 12, no. 2, p. 545, Jan. 2022, doi: 10.3390/app12020545.

[2] A. F. Qasim, R. Aspin, F. Meziane, and P. Hogg, "ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR Images," *Multimed. Tools Appl.*, vol. 78, no. 12, Art. No. 12, Jun. 2019, doi: 10.1007/s11042-018-7029-7.

[3] ARSLAN AKRAM et al, "A ROBUST AND SCALE INVARIANT METHOD FOR IMAGE FORGERY CLASSIFICATION USING EDGE WEIGHTED LOCAL TEXTURE FEATURES," *JJU*, vol. 41, no. 12–2022, pp. 330–344, Dec. 2022, doi: 10.17605/OSF.IO/MVWDU.

[4] A. Akram, S. Ramzan, A. Rasool, A. Jaffar, U. Furqan, and W. Javed, "Image splicing detection using discriminative robust local binary pattern and support vector machine," *World J. Eng.*, vol. 19, no. 4, pp. 459–466, Jun. 2022, doi: 10.1108/WJE-09-2020-0456.

[5] Asis Jamal, Sarah Javed, Arslan Akram, and Shahzaib Jamal, "Recovery Method for Disasters of Network Servers by Using POX controller in Software defined Networks.," *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 4, pp. 45–52, Dec. 2019, doi: 10.54692/lgurjcsit.2019.030492.

[6] MUHAMMAD USMAN TARIQ et al, "REAL TIME AGE AND GENDER CLASSIFICATION USING VGG19," *JJU*, vol. 41, no. 12–2022, pp. 641–655, Dec. 2022, doi: 10.17605/OSF.IO/BKJWH.

[7] MUHAMMAD SALMAN ALI et al., "CHEST X-RAY BASED PNEUMONIA CLASSIFICATION USING VGG-19," *JJU*, vol. 42, no. 01–2023, pp. 118–136, Jan. 2023, doi: 10.17605/OSF.IO/M2FPD.

[8] SOBIA YAQOOB et al, "A NOVEL METHOD FOR IMAGE BASED BREAST CANCER CLASSIFICATION," *JJU*, vol. 41, no. 12–2022, pp. 470–484, Dec. 2022, doi: 10.17605/OSF.IO/JZFRM.

[9] P. Wang, H. Huang, M. Wang, and B. Li, "YOLOv5s-FCG : An Improved YOLOv5 Method for Inspecting Riders' Helmet Wearing," *J. Phys. Conf. Ser.*, vol. 2024, no. 1, p. 012059, Sep. 2021, doi: 10.1088/1742-6596/2024/1/012059.

[10] K. Dahiya, D. Singh, and C. K. Mohan, "Automatic detection of bike-riders without helmet using surveillance videos in real-time," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 3046–3051.

[11] H. Lin, J. D. Deng, D. Albers, and F. W. Siebert, "Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning," *IEEE Access*, vol. 8, pp. 162073–162084, 2020, doi: 10.1109/ACCESS.2020.3021357.

[12] N. Charlie, "Automatic Helmet Detection System on Motorcyclists Using YOLOv3," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, no. 5, pp. 2763–2766, May 2020, doi: 10.22214/ijraset.2020.5464.

[13] A. Chairat, M. N. Dailey, S. Limsoonthrakul, M. Ekpanyapong, and D. Raj K.C., "Low Cost, High Performance Automatic Motorcycle Helmet Violation Detection," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Snowmass Village, CO, USA, Mar. 2020, pp. 3549–3557. Doi: 10.1109/WACV45572.2020.9093538.

[14] UG student Sri Jayachamarajendra College of Engineering, Mysore. *et al.*, "Detection of Non-Helmet Riders and Extraction of License Plate Number using Yolo v2 and OCR Method," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 2, pp. 5167–5172, Dec. 2019, doi: 10.35940/ijitee.B6527.129219.

[15] S. Chen, J. Lan, H. Liu, C. Chen, and X. Wang, "Helmet Wearing Detection of Motorcycle Drivers Using Deep Learning Network with Residual Transformer-Spatial Attention," *Drones*, vol. 6, no. 12, p. 415, Dec. 2022, doi: 10.3390/drones6120415.

[16] S. Eliyas, K. Swaathi, D. P. Ranjana, and A. Harshavardhan, "Helmet, Violation, Detection Using Deep Learning," *Clin. Med.*, vol. 07, no. 02, 2020.

[17] M. J. Prajwal, K. B. Tejas, V. Varshad, M. M. Murgod, and R. Shashidhar, "Detection of non-helmet riders and extraction of license plate number using Yolo v2 and OCR method," *Int. J. Innov. Technol. Explor. Eng. IJITEE*, vol. 9, no. 2, pp. 5167–5172, 2019.

[18] A. Afzal, H. U. Draz, M. Z. Khan, and M. U. G. Khan, "Automatic Helmet Violation Detection of Motorcyclists from Surveillance Videos using Deep Learning Approaches of Computer Vision," in *2021 International Conference on Artificial Intelligence (ICAI)*, Islamabad, Pakistan, Apr. 2021, pp. 252–257. Doi: 10.1109/ICAI52203.2021.9445206.

[19] A. Saumya, V. Gayathri, K. Venkateswaran, S. Kale, and N. Sridhar, "Machine Learning based Surveillance System for Detection of Bike Riders without Helmet and Triple Rides," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, Sep. 2020, pp. 347–352. Doi: 10.1109/ICOSEC49089.2020.9215266.

[20] B. Yogameena, K. Menaka, and S. Saravana Perumaal, "Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system," *IET Intell. Transp. Syst.*, vol. 13, no. 7, pp. 1190–1198, 2019.

[21] W. Jia *et al.*, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," *IET Image Process.*, vol. 15, no. 14, pp. 3623–3637, 2021.

[22] H. Nagoriya, "Live Helmet Detection System for Detecting Bikers without Helmet," Sep. 2020, doi: 10.5281/ZENODO.4050483.

[23] L. Shine and J. C. V., "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN," *Multimed. Tools Appl.*, vol. 79, no. 19–20, pp. 14179–14199, May 2020, doi: 10.1007/s11042-020-08627-w.

[24] Kisaezehra, M. Umer Farooq, M. Aslam Bhutto, and A. Karim Kazi, "Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites," *Intell. Autom. Soft Comput.*, vol. 36, no. 1, pp. 911–927, 2023, doi: 10.32604/iasc.2023.031359.

[25] R. Cheng, X. He, Z. Zheng, and Z. Wang, "Multi-Scale Safety Helmet Detection Based on SAS-YOLOv3-Tiny," *Appl. Sci.*, vol. 11, no. 8, Art. no. 8, 2021.

[26] N. Albert and D. Basavesha, "An Automatic Helmet Detection System using Convolution Neural Network," *Int. J. Adv. Sci. Innov.*, vol. 4, no. 3, 2022.

[27] H. Wang, Z. Hu, Y. Guo, Z. Yang, F. Zhou, and P. Xu, "A Real-Time Safety Helmet Wearing Detection Approach Based on CSYOLOv3," *Appl. Sci.*, vol. 10, no. 19, p. 6732, Sep. 2020, doi: 10.3390/app10196732.

[28] R. Cao, H. Li, B. Yang, A. Feng, J. Yang, and J. Mu, "Helmet wear detection based on neural network algorithm," *J. Phys. Conf. Ser.*, vol. 1650, no. 3, p. 032190, Oct. 2020, doi: 10.1088/1742-6596/1650/3/032190.

[29] A. Anjaiah, P. P. L. L. Reddy, V. Keerthika, N. S. Nayak, and K. Rakesh, "Automatic Number Plate Detection for Motorcyclists Riding Without Helmet," *J. Algebr. Stat.*, vol. 13, no. 3, pp. 2158–2165, 2022.

[30] N. C. Mallela, R. Volety, S. P. R., and N. R. K., "Detection of the triple riding and speed violation on two-wheelers using deep learning algorithms," *Multimed. Tools Appl.*, vol. 80, no. 6, pp. 8175–8187, Mar. 2021, doi: 10.1007/s11042-020-10126-x.

[31] F. W. Siebert and H. Lin, "Detecting motorcycle helmet use with deep learning," *Accid. Anal. Prev.*, vol. 134, p. 105319, Jan. 2020, doi: 10.1016/j.aap.2019.105319.

[32] Y. Liu, Z. Li, B. Zhan, J. Han, and Y. Liu, "A Super-Resolution Reconstruction Driven Helmet Detection Workflow," *Appl. Sci.*, vol. 12, no. 2, p. 545, 2022.

[33] R. Cheng, X. He, Z. Zheng, and Z. Wang, "Multi-Scale Safety Helmet Detection Based on SAS-YOLOv3-Tiny," *Appl. Sci.*, vol. 11, no. 8, p. 3652, 2021.

[34] L. Huang, Q. Fu, M. He, D. Jiang, and Z. Hao, "Detection algorithm of safety helmet wearing based on deep learning," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 13, p. e6234, 2021.

[35] F. W. SIEBERT, "Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning".

[36] K. Han and X. Zeng, "Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features," *IEEE Access*, vol. 10, pp. 718–729, 2021.

[37] W. Jia *et al.*, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," *IET Image Process.*, vol. 15, no. 14, pp. 3623–3637, 2021.

[38] L. Shine and J. CV, "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN," *Multimed. Tools Appl.*, vol. 79, no. 19, pp. 14179–14199, 2020.

[39] Z. Fan, C. Peng, L. Dai, F. Cao, J. Qi, and W. Hua, "A deep learning-based ensemble method for helmet-wearing detection," *PeerJ Comput. Sci.*, vol. 6, p. e311, 2020.

[40] M.-E. Otgonbold *et al.*, "SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet Detection," *Sensors*, vol. 22, no. 6, p. 2315, Mar. 2022, doi: 10.3390/s22062315.