



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

# A Swin Transformer-Based Approach for Motorcycle Helmet Detection

M.BUVANESWARI, TAMILARASAN V, UMAR FAROOK K, VENKATACHALAM P

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

**ABSTRACT:** Visual inspection of the workplace and timely reminders of unsafe behaviors (e.g., not wearing a helmet) regarding safety helmets are very important safety management measures to avoid injuries to workers on construction sites. Video surveillance systems generate large amounts of non-structure image data on site for this purpose, but they require real-time recognition automation solutions based on computer vision. Although various deep-learning-based models have recently provided new ideas for helmet recognition in traffic monitoring, few solutions suitable for industry applications have been discussed owing to the complexity of the construction site scenarios. This paper described a fast and robust network FRSHNet based on a mutilscale Swin Transformer that was proposed to identify safety helmets at the construction site. Three key contributions of our method adopted MAE-NAS with the variant MobileNetV3's MobBlock for feature extraction, multiscale Swin Transformer Module for generating the spatial and context relationships, and efficient-RepGFPN for real-time detection. The results from the Pictor-v3 and SHWD datasets showed that the proposed method had advantages over other state-of-the-art methods.

**KEYWORDS:** MobBlock; mutilscale Swin Transformer; helmet detection

## I.INTRODUCTION

Traffic management in India faces substantial hurdles, with challenges stemming from dense populations, a surge in commuters, flawed infrastructure, and prevalent societal attitudes. Currently, many of India's city traffic systems are largely governed by manual supervision. This approach often results in significant con-gestion and errors, hindering effective traffic management and enforcement. A report from 2019 by the Road Transport and Highways Ministry highlighted the concerning statistics: 37 percent of road accident victims were two-wheeler riders. A large fraction of these accidents was attributed to riders not wearing helmets. This violation alone accounted for almost 30 percent of the total road accident deaths that year. The World Health Organization further emphasized the importance of safety measures, stating that proper helmet use could drastically reduce fatalities and injuries. They also identified the dangers of mobile phone usage while driving, noting the increased risk of accidents. However, the challenge doesn't end with identifying the problems.

The enforcement mechanism is overburdened. According to a 2017 report, there is an alarming mismatch between the number of automobiles on Indian roads (over 200 million) and the number of traffic police officers (less than 72,000). Due to this imbalance, it is difficult for the current systems to efficiently handle the variety of cars, variety of license plates, and constantly shifting traffic dynamics. Many cities continue to use antiquated traffic management systems that are unable to manage the volume of vehicles on the road today. The necessity of technology intervention is highlighted by the inefficiencies in the current system. Adopting cuttingedge technologies is essential to addressing the changing issues of traffic management and violation control. Promising solutions can be found in automated systems, especially those that incorporate AI based on video. These devices have the power to revolutionize traffic surveil-lance, identify offenses and impose fines with accuracy, and support smooth enforcement. Helmet detection systems basically use a combination of image processing and machine learning techniques to identify motorcycle riders wearing helmets. The very first step is to detect the motorcycle rider in the video frame or image. This can be done using YOLOv5 or by using the classification

methods. The system can use various features once the rider has been detected such as the shape and color of the head and the presence of a helmet visor, to classify the rider as helmeted or non-helmeted.

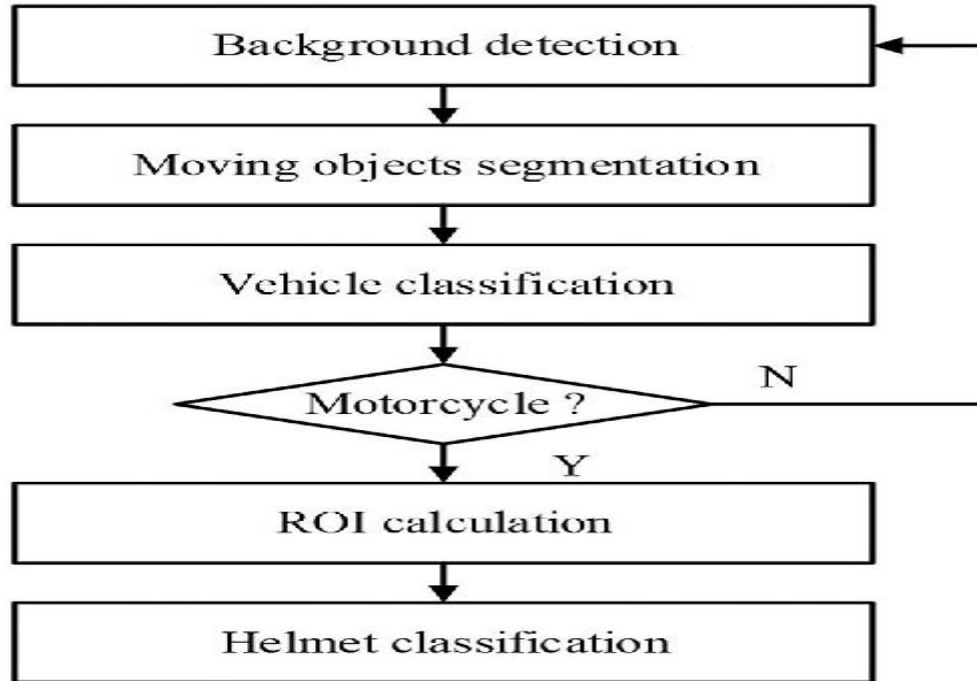


Fig 1: Real time Helmet Classification

Number plate extraction systems typically use a combination image processing and machine learning techniques to identify and extracting the number plate from video frame or an image. The first step is the detection of the number plate in the image or video frame. We can use various techniques such as Optical Character Recognition, Neural Networks etc. to extract the number and transform it into a readable form once the number plate from the image or video has been detected. Even with the advancements in helmet detection and number plate extraction, certain issues remain that require attention. For example, helmet detection systems can be challenging in nighttime conditions and low light. Furthermore, when the number plate is covered in dirt, debris, or other things, number plate extraction devices may provide difficulties. In the following sections of this survey paper, we will review the state of the art in helmet detection and number plate extraction. We will discuss the various challenges that need to be addressed and the future directions of research in this field.

## II.LITERATURE SURVEY

This paper [1] investigates one such method that uses an improved YOLOv5 (you only look once) model for helmet identification. The main duties performed by this approach are the adjustment of the initial candidate frame, loss function, and input size of the original YOLOV5 model. In order to enhance its suitability for helmet wearer identification, the YOLOV5 model has been improved. A camera and extra hardware are added to the learned model to form a complete real-time detection system. High accuracy at real-time detection rates is guaranteed by the YOLOv5 identification model. YOLOV5 has a fast detection rate and good detection accuracy, but in order to create a better model that can better satisfy the requirements of real-time detection and yield a better detection result, we still need to optimize the original YOLOv5 and continually alter the super parameters.

In the paper, an improvement to create anchor boxes was suggested. It's possible that the generation ratio of the anchor box will directly affect the model's speed. This study uses the K-Means clustering technique to investigate the labeled target



box’s width inside the training set. Determine the nine cluster centers’ wide and high dimensions, which correspond to the network profile’s anchor parameter values. The Safety-Helmet-Wearing dataset clustering, an open-source dataset, serves as the foundation for the finest anchors. This is carried out in compliance with the YOLOV5 model network structure’s characteristics. This type produces a high level of accuracy and precision. Additionally, the model generates findings quickly. However, because this model uses facial features to identify the region of interest, technical parameters like resolution and camera angle are crucial to its proper functioning.

This paper [2] addresses several of the shortcomings of current models, including their incapacity to monitor bikes over numerous frames and identify motorcycle riders from their passengers in order to inspect them for helmets. There are three main phases in the methodology that this paper discusses. Motorcycle detection is the initial phase, and it may be accomplished in two ways: either using a single-stage model or a twostage model. Because the two-stage model iterates through the picture twice rather than once, it is more accurate than the one-stage model but requires more computing power. Single stage models operate at a faster pace since they only go through the model once. In contrast to other common object identification models like YOLO, the authors of this research have chosen to employ a single stage method with a single-stage technique that is fine-tuned and pre-trained RetinaNet, which has a greater accuracy but a slower speed.

### III.METHODS

When the visibility complexity of a construction site is generated by multiple dimensions, including intricate construction equipment, workers’ postures and positions, and wearing helmets of different colors and sizes, the performances of helmet detection can be upgraded using the relationships between the worker and helmet. In the studies of, the effectiveness of the Transformer based on the self-attention mechanism when modeling various temporal and spatial position relationships was confirmed. However, when the transformer was applied to vision tasks, its internal spatial self-attention operations led to a significant increase in computational complexity. To this end, the Swin Transformer proposed a novel windowing strategy that grouped the spatial dimensions of the input into multiple non-overlapping windows. In this way, the model only needs to calculate spatial self-attention within each local window, rather than in the entire input space.

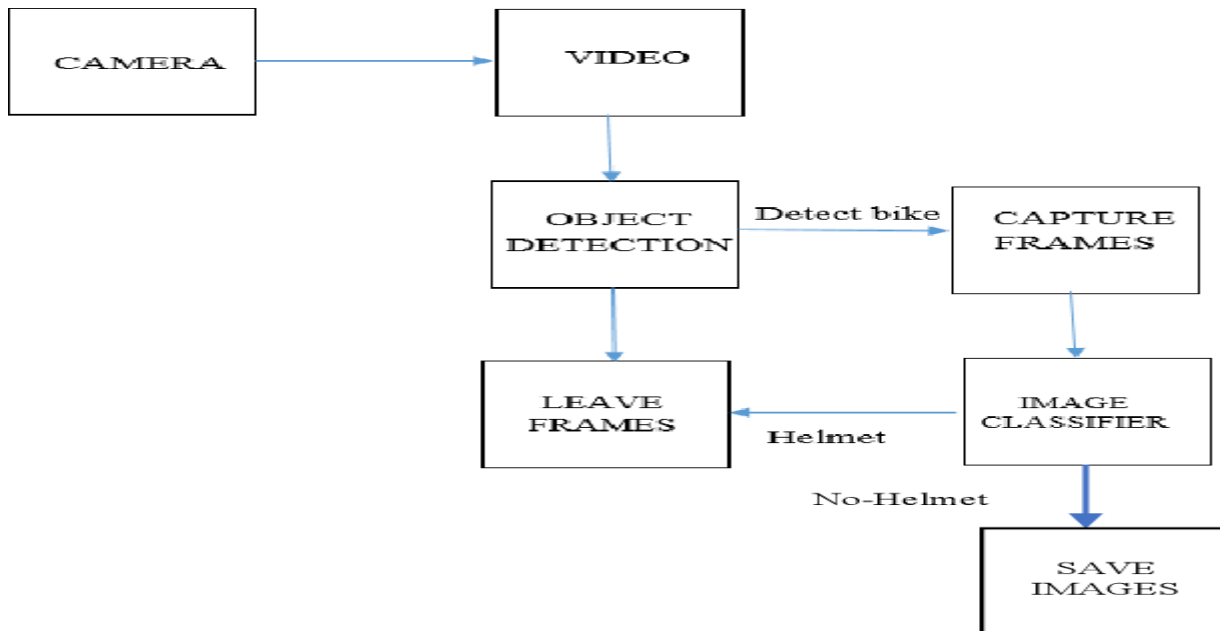


Fig 2: Work Flow

Concretely, a patch splitting module was first applied to each feature of the multi-level features to finely extract compact non-overlapping patches. This process ensures that each patch contains meaningful semantic information and is treated as an independent “token”. Next, to better process these extracted patches, each patch was transformed through a linear embedding layer. This embedding process enabled each patch to easily be projected to any output dimension, providing greater flexibility for subsequent processing. Subsequently, in order to capture complex patterns and relationships at different scales, three tokens were separately input to the Swin Transformer block, which consisted of different layers of (shifted) window-based multi-head self attention ((S)W-MSA) and multi-layer perceptron blocks. The advantage of SW-MSA is that it can weigh information at different spatial positions and scales using a weight adjustment mechanism.

### RESULT ANALYSIS

proposes an enhanced YOLOv5m and LPRNet model-based automobile license plate identification technique. Based on an analysis of the YOLOv5m algorithm and the image features of the vehicle license plate, the algorithm is refined in three ways: the matching degree between the anchor frame and the detection target is enhanced via the K-means++ algorithm; the NMS method is refined via the DIOU loss function; and the feature map with  $20 \times 20$  is eliminated to decrease the number of detection layers. Without character segmentation, license plate character recognition is achieved using a lightweight LPRNet network. A license plate recognition system based on the IYOLOv5m-LPRNet model is built by combining the enhanced YOLOv5m algorithm with the LPRNet network. The authors suggested a few methods to increase license plate detection accuracy. The multidimensional clustering of label data sets using the K-means++ method can significantly shorten the time it takes the model to locate the anchor.



**Fig 3: A Swin Transformer-Based Approach**

Furthermore Given that NMS has a tendency to miss detection while screening the real target frame, the DIOU-NMS method is created by re-fining the post-processing technique of NMS using the concept of the DIOU loss function. The license plate in the traffic scenario is a tiny target, therefore fewer detection layers are needed. This means keeping the feature map of the scale with  $20 \times 20$  and conserving the feature map of the scale with  $80 \times 80$  and  $40 \times 40$ . In addition to satisfying license plate detection requirements, the enhanced network lowers network complexity and increases target detection and network running speeds. Plate identification is quicker and more accurate using this approach. For character recognition, it also makes use of a lightweight LPRNet network, which provides resilience and adaptability in challenging settings. A automobile license plate recognition system (IYOLOv5m-LPRNet model) is developed using this enhanced algorithm and LPRNet, and it is validated through rigorous training and simulations. Significant improvements in identification accuracy (99.49%), recall (98.79%), mAP (98.56%), and recognition speed (42 images per second) are demonstrated by the testing findings. The technique meets the requirements for complex license plate recognition with high accuracy, speed, resilience, and real-time performance.



## V.CONCLUSION

In this paper we have explained how important it is to wear a helmet and how can wearing a helmet can prevent major accidents. We now know why it's necessary to have a system that can recognize whether or not a person is wearing a helmet when riding a motorcycle. In this paper we also reviewed various technologies and algorithms that mainly involve Deep Learning algorithms like CNN and YOLO which are useful to detect helmets in the given footage. We also have further seen on how we can isolate the number plate from the vehicle from the given footage using technologies like edge detection and algorithms such as Support Vector Machines etc. We have also seen that we can train an advanced Optical Character Recognition Model that can be used to extract characters from the image extracted. Furthermore, we have also seen on how the proposed algorithms perform un-der given conditions and on how improving and custom training some algorithms improve the efficiency of model. We have also seen how to get better results, keeping efficiency and latency in first priority. As we can see that this work highly depends on efficiency and speed, so it is important to maintain them. To summarize this paper provides the overview of the technologies used and methodologies proposed and how we can use them to help reduce the accidents that can occur by not wearing a helmet.

## REFERENCES

1. Wu, H.; Zhao, J. An intelligent vision-based approach for helmet identification for work safety. *Comput. Ind.* **2018**, *100*, 267–277. [[Google Scholar](#)] [[CrossRef](#)]
2. Yu, F.; Wang, X.; Li, J.; Wu, S.; Zhang, J.; Zeng, Z. Towards Complex Real-World Safety Factory Inspection: A High-Quality Dataset for Safety Clothing and Helmet Detection. *arXiv* **2023**, arXiv:2306.02098. [[Google Scholar](#)]
3. Chen, C.; Wu, W. Color pattern recognition with the multi-channel non-zero-order joint transform correlator based on the HSV color space. *Opt. Commun.* **2005**, *244*, 51–59. [[Google Scholar](#)] [[CrossRef](#)]
4. Kelm, A.; Laußat, L.; Meins-Becker, A.; Platz, D.; Khazaei, M.J.; Costin, A.M.; Helmus, M.; Teizer, J. Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. *Autom. Constr.* **2013**, *36*, 38–52. [[Google Scholar](#)] [[CrossRef](#)]
5. Li, Y.; Wei, H.; Han, Z.; Huang, J.; Wang, W. Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Network. *Adv. Civ. Eng.* **2020**, *2020*, 10. [[Google Scholar](#)] [[CrossRef](#)]
6. Rajaraman, V. Radio frequency identification. *Reson* **2017**, *22*, 549–575. [[Google Scholar](#)] [[CrossRef](#)]
7. Dolez, P.I. Chapter 3.6—Progress in Personal Protective Equipment for Nanomaterials. In *Nanoengineering*; Dolez, P.I., Ed.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 607–635. [[Google Scholar](#)] [[CrossRef](#)]
8. Swain, M.J.; Ballard, D.H. Indexing via Color Histograms. In *Proceedings of the Active Perception and Robot Vision*, Maratea, Italy, 16–29 July 1989; Sood, A.K., Wechsler, H., Eds.; Springer: Berlin/Heidelberg, Germany, 1992; pp. 261–273. [[Google Scholar](#)]
9. Žunić, J.; Hirota, K.; Rosin, P.L. A Hu moment invariant as a shape circularity measure. *Pattern Recognit.* **2010**, *43*, 47–57. [[Google Scholar](#)] [[CrossRef](#)]
10. Pietikäinen, M. Local Binary Patterns. *Scholarpedia* **2010**, *5*, 9775. [[Google Scholar](#)] [[CrossRef](#)]
11. Zhigang, L.; Wenzhong, S.; Qianqing, Q.; Xiaowen, L.; Donghui, X. Hierarchical support vector machines. In *Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '05)*, Seoul, Republic of Korea, 20 July 2005; Volume 1, p. 4. [[Google Scholar](#)] [[CrossRef](#)]
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[Google Scholar](#)]
13. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271. [[Google Scholar](#)]
14. Jocher, G.; Stoken, A.; Borovec, J.; NanoCode012, C.; Changyu, L.; Laughing, H. ultralytics/yolov5: v3.0. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 20 December 2020).
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37. [[Google Scholar](#)]



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details