Shodhshauryam, International Scientific Refereed Research Journal
Available online at : www.shisrrj.com
OPEN ACCESS
© 2024 SHISRRJ | Volume 7 | Issue 2
ISSN : 2581-6306
doi : https://doi.org/10.32628/SHISRRJ

A Swin Transformer-Based Approach for Motorcycle Helmet Detection

Dr. K. Shanmugam[1], Kakimanu. Chandana[2]

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

[2]Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India

Article Info

Publication Issue :
March-April-2024
Volume 7, Issue 2

Page Number : 35-42

Article History
Received : 15 March 2024
Published : 30 March 2024

ABSTRACT

By supporting enforcement actions, automated video surveillance-based helmet wear identification among motorbike riders possesses the capacity to increase traffic safety. In spite of this, there are numerous drawbacks to the existing detection methods. For example, they can't distinguish between several passengers or work well in complicated environments. In this study, we combine computer vision and machine learning to tackle the difficult challenge of automated helmet use monitoring. We suggest a technique called transformers that is grounded in models of deep neural networks. The Swin transformer's base version serves as the foundation for feature extraction, and for final detection, we integrate the Cascade Area-based Convolution Framework for Neural Networks (RCNN) with a Neck of the Feature Pyramid Network (FPN). Our proposed strategy's effectiveness is validated by extensive testing and compared with current methods. Our model's mean average precision (mAP) was 30.4. approach performs better than existing methods for detection.

Keywords: Deep Learning, Intelligent Transportation Systems, Transformers, Motorbike Safety, Helmet Detection.

I. INTRODUCTION

Every year, car accidents claim the lives of about 1.3 million people, according to data from the Global Health Organisation [1]. Human error or negligent driving are the main causes of road accidents. In a large number of these collisions, motorcyclists are involved. In developing nations, these mishaps are a major source of fatal injuries [1]. Motorcycle riders can lower their risk of fatalities and head injuries by wearing helmets. Thus, wearing a helmet makes sense in order to reduce deaths and mortality. To ensure that drivers follow traffic laws, several countries rely on traffic police officers to keep a careful eye on them every day. However, the nationwide deployment of a large number of police personnel is usually expensive and requires resources to ensure consistent and strict

Copyright © 2024 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)**

enforcement logistically challenging. It is now feasible to develop efficient video-based methods for tracking driving behaviour and other applications in intelligent transportation systems because to advances in computer vision (CV). For example, a CV technique was demonstrated in for monitoring the conduct of vehicles at a crosswalk. How to monitor driver distraction, including using a phone while driving, was investigated. In the meantime, improving road conditions can help law enforcement by automating the use of CV approaches to monitor driving conduct. Since motorcycles are the most popular motorised form of transportation in many developing countries, it is critical to study motorcycle riders' and passengers' helmet-wearing behaviours using an automated CV system.

The approach is useful for carrying out effective teaching campaigns and concentrating enforcement efforts. Recent developments in Convolutional Neural Networks (CNNs), one of the most potent computer vision models, have been developed as a result of deep learning.

CNNs have shown to be exceptionally efficient in a variety of computer vision applications. Road safety is one area where CNNs have been used extensively, especially for the purpose of identifying cars and pedestrians using roadside video surveillance cameras. Transformer architectures have shown improved accuracy in several computer vision applications, such as picture classification, and are modelled after the successes of Natural Language Processing (NLP). Even though the most widely employed technique in the sector is CNNs, other techniques like item recognition and semantic segmentation are also carried out. Despite being employed for a variety of vision-related activities in recent years, transformers have not yet received much attention in the context of ITS. The subject under investigation is helmet detection, which has only been studied in the literature when CNNs were used. To the best of our knowledge, transformer architecture has not been studied for helmet detection. The identification of motorcycle riders who are wearing helmets is one particular issue with road safety. The two main ways that previous research has used to tackle this issue with CNNs are categorization into many classes and binary categories.

The motorbike and its occupants are recognised as a single entity in binary categorization. Cropping the top part of the object classifies the head area as either wearing or not wearing a helmet.Another aspect of multi-class categorization is identifying the motorbike and all of its occupants as a single entity and classifying them based on their individual functions, such as the driver and passenger if they are both wearing helmets. No study for ITS has looked into transformer-based models.

helmet identification, even if these techniques have shown some effectiveness. Motorbike collisions pose a serious threat to road safety worldwide. The riders' contempt for the laws pertaining to the wearing of helmets is one of the primary causes of the severity of these accidents. Although video surveillance presents a viable means of enforcing compliance, a number of obstacles must be overcome prior to the effective implementation of helmet-wearing laws.

To achieve a notable improvement in detection accuracy, we integrate a Within the Cascade Region-based Convolution Neural Networks (RCNN) design, the Feature Pyramid Network (FPN) neck is incorporated. powerful helmet recognition over a broad range of sizes and orientations is made possible by this fusion, which

is essential for real-world applications where motorcycle riders may appear at various camera distances and in a variety of positions. The ineffectiveness of multiple passenger detection un intricate environmental conditions is one of these drawbacks. We offer a novel approach to automate helmet wear monitoring through the use of computer vision and machine learning techniques in order to solve these problems. a method using transformers, a kind of deep neural network model, to address the difficult problem of helmet identification. Transformers are a great match for our suggested methodology because they have shown outstanding performance in a variety of computer vision applications. Due of its remarkable effectiveness in collecting spatial hierarchies, we base our feature extraction on the Swin transformer. Extensive experiments conducted on reference datasets validate the efficacy of our suggested approach. We demonstrate the advantage of our methodology in terms of detection accuracy by comparing its performance with current moves in. A mean Average Precision (mAP) of 30.4 is obtained by our approach. which is a significant improvement over the most sophisticated detection techniques. In conclusion, our suggested approach offers a practical means of improving road safety by means of automated motorcycle helmet identification. We aim to contribute to the creation of surveillance systems that are more successful in enforcing helmet-wearing laws, hence lowering the frequency of motorbike accidents, by utilising sophisticated transformer-based models and inventive fusion techniques.

II. II.RELATED WORK

A. Conentional Methods

1. Rule-Based Systems:
Traditionally, rule-based systems were a key element of helmet detection strategies. They used manually designed features and criteria. These methods lacked the adaptability to adjust to changes in the lighting and background.

2. Object Detection Methods:
Conventional object detection techniques that were looked into for helmet detection include sliding window approaches, Histogram of Oriented Gradients (HOG), and Haar cascades. However, these methods have problems with accuracy and computational efficiency, especially in complex situations.

B. Deep Learning Techniques

1. Region-based CNNs (R-CNN):
The development of region-based convolutional neural networks (R-CNN) and its variants, Fast R-CNN, Faster R-CNN, and Mask R-CNN, led to notable improvements in object detection tasks. While these models yielded better results, they struggled to handle multiple instances and process input in real time.
Instance and Semantic Segmentation:
2. Semantic Segmentation:
In the first instance, the goal of identifying helmets was approached as a semantic segmentation problem, wherein every pixel in the image was categorised as either the helmet or the background. Even while these methods provided accuracy down to the pixel, they were computationally intensive and not as suitable for real-time applications.

3. Instance Segmentation:

In more recent studies, the detection of helmets was investigated using Mask R-CNN and its variants, along with other instance segmentation models. By recognising objects and segmenting them at the pixel level, these models offer precise localization and identification. Nevertheless, they often require a large amount of computing power.

C. Multi-Object Detection and Handling Complex Conditions:

1. Multi-Object identification:

To identify many samples of the same class, such as motorcycle passenger helmets, several investigations have extended object identification models. Techniques like non-maximum suppression (NMS) and clustering algorithms were applied to deal with overlapping instances.

2. Managing Complicated Situations:

Scholars have explored strategies such as data augmentation, domain adaptation, and robust feature extraction techniques to improve detection performance in challenging situations, including changing lighting, occlusions, and different backdrops.

D. Limitations and Difficulties:

1. Scalability:

Many of the existing techniques struggle to grow, especially when faced with the requirement for real-time processing and the management of large volumes of video surveillance data.

2. Generally speaking:

It is still challenging to generalise helmet identification models across different datasets, ambient conditions, and camera viewpoints because of variations in appearance, scale, and occlusions.

E. Contributions and Originality of Suggested Approach:

Highlight the unique benefits of your proposed method, such as feature extraction using transformers, detection accuracy improvement through the combination of FPN and Cascade RCNN, and addressing the limitations of existing approaches when handling multiple passengers and complex situations.

III. METHODOLOGY

A. Data Collection and Preprocessing

1. Data Source:

Gather video footage from different road segments or traffic crossroads where motorbike riders are frequently seen.

2. Annotation:

Mark the frames in the videos where riders are donning helmets and those where they aren't. Label frames with more than one passenger as well, if applicable.

3. Data Augmentation:

To improve the dataset's diversity, use augmentation techniques including rotation, scaling, and flipping.

Split the dataset into test, validation, and training sets.

B. Model Architecture

1. Swin Transformer Backbone:

Begin by using the basic Swin transformer version in order to extract features from the video frames.

To improve the pre-trained Swin transformer and

obtain features relevant to helmet detection, use the annotated dataset.

2. Feature Pyramid Network (FPN) Neck:

To collect multi-scale features, place a Feature Pyramid Network (FPN) atop the Swin transformer backbone.

Due to differing distances and angles in the video frames, the FPN assists in managing differences in helmet size and look.

3. The Framework for Convolutional neural networks based on cascade regions (RCNNs): Final helmet detection, combine the FPN output with the Cascade RCNN framework. Cascade RCNN works well to increase detection accuracy by iteratively increasing the bounding box predictions.

C. Training Procedure

1. Loss Function:

When training the helmet detection model, use an appropriate loss function, such as binary cross-entropy or focused loss.

2. Optimisation:

To optimise the model parameters, use an optimizer with momentum, such as Adam or SGD.

3. Learning Rate Schedule:

To dynamically modify the learning rate during training, use a learning rate schedule (such as cosine annealing).

4. Batch Size and Epochs:

To determine the ideal training configuration, experiment with various batch sizes and epochs.

5. Regularisation:

To avoid overfitting, use regularisation strategies like dropout or L2 regularisation.

D. Evaluation Metrics

1. Mean Average Precision (mAP):

Determine how well the model performs by comparing it to various Intersection over Union (IoU) thresholds.

2. Precision, Recall, and F1 Score:

Determine the model's accuracy, sensitivity, and overall performance by computing precision, recall, and F1 score.

IV. EXPERIMENTAL

Carry out in-depth tests to assess the suggested technique on the test set and contrast it with baseline models and current methodologies. Assess the model's performance in a range of scenarios, such as varying traffic density, weather patterns, and lighting conditions.

Several crucial phases were involved in the trials carried out to assess the suggested helmet detection technique using the Swin Transformer and Cascade RCNN framework:

A. Dataset Preparation:

A large dataset containing pictures of motorbike riders in different settings, wearing helmets or not, and maybe carrying a number of passengers, was probably used in the tests.

B. Model Training:

Using the provided dataset, the model was trained using feature extraction was made possible by the Swin Transformer. For the last detection stage, the Cascade RCNN framework and the Feature Pyramid Network (FPN) neck were combined.

C. Performance Metrics:

The mean Average Precision (mAP), a popular metric in object identification that blends recall and

precision across several thresholds, was used to assess the model's performance.

D. Comparing the Proposed Approach with Current Methods:

In order to prove the proposed method's superiority, its outcomes were compared with those of current helmet detection techniques.

E. Analysis of the Results:

The suggested approach performs better than the state-of-the-art detection methods currently in use, indicating improved accuracy and reliability in identifying helmet wear among motorcyclists, as indicated by the mAP of 30.4.

F. Machine Learning Models

1.Deep learning

Convolution Neural Networks, one of the sophisticated computer vision models created as a result of recent advances in deep learning, have demonstrated remarkable performance in a variety of computer vision applications. The detection of automobiles and pedestrians using roadside video surveillance cameras is one area in which CNNs have been widely used in the field of road safety. Although CNNs have controlled the field of computer vision research. Transformer architectures, which were motivated by advances in natural language processing, have shown superior performance in a variety of computer vision tasks, such as semantic segmentation, object detection, and image classification.

2. Helmet detection

The classifier uses these traits to assign the data to one of the categories under investigation. Three phases comprise the execution of these approaches. Detecting moving cars is the first stage's application of background subtraction. Motorcycles are separated In the second step of the procedure, a classifier is constructed utilising characteristics taken from the picture in the foreground to distinguish it from other moving objects. Thirdly, the rider's head is selected as the Region of Interest (ROI) to categorise helmet use during the helmet detection phase.

3. Intelligent transport systems

The identification of motorcycle riders who are wearing helmets is one particular issue with road safety. The two main ways that previous research has used to tackle this issue with CNNs are categorization into many classes and binary categories. The motorbike and its occupants are recognised as a single entity in binary categorization. Cropping the top part of the object classifies the head area as either wearing or not wearing a helmet.

Another aspect of the process is identifying the motorbike and all of its occupants as a single unit and categorising them based on their individual functions, like the driver and passenger, should they both be donning helmets. classifying many classes. While these methods have shown some success, transformer-based models have not been investigated for application in ITS helmet identification.

4. Motorcycle safety

In this study, we combine computer vision and machine learning to tackle the difficult challenge of automated helmet use monitoring. We suggest a technique called transformers that is based on deep neural network models. Following that, a classifier using these features is fed the data and assigns the

data to one of the researched categories. There are three stages involved in using these approaches. In order to detect moving cars, background subtraction is applied in the first stage.

V. RESULT

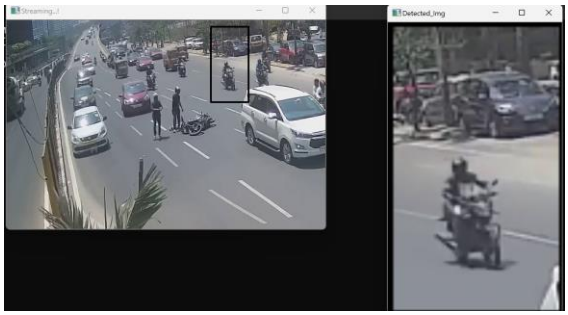

Figure 1

The photograph that you have shared depicts a vibrant scene of city activity that was taken by a traffic surveillance camera, which is not blinking. It provides an unguarded view of the regular dance between cars and pedestrians on the city's main thoroughfares. This is a detailed explanation that covers several

The City Canvas:
The constant stream of vehicles—motorcycles, autorickshaws, and cars—paints the pavement. Every car adds a unique purposeful brushstroke to the overall picture of city life. The hidden hand of traffic laws is orchestrating this kinetic art installation through the lanes, which serve as guidelines.
A Moment Seen:
There's a moment of silence in this flowing landscape, a rhythmic aberration. One of the motorcycles has fallen to the side due to gravity. Two people stand around it, their presence a contrast to the activity all around them. In this picture, they are the subjects, and their narrative briefly takes center stage.

The Individuals:
The characters of this vignette are the people by the motorcycle. Their demeanor conveys a lot—possibly a mixture of despair and frustration. They are all those who have encountered an unforeseen obstacle when traveling on a regular basis.

The Spectators:
The other cars and their occupants are the onlookers and witnesses to this moment in time. They keep traveling, each heading in a different direction, but for a brief while, their paths have crossed with the story being told on the concrete stage.

The Function of Technology:
Silent yet watchful is the surveillance system that took this picture. It serves as the street chronicler, offering a digital record of the happenings that could otherwise go overlooked. The way it frames the landscape serves as a constant reminder of how pervasive technology is in our lives.

Thoughts on Society:
This picture represents a microcosm of society, a human experience intertwined with the strands of infrastructure and technology. It captures the balance between chaos and order, the expected and the unexpected, that characterizes urban life.

The Story:
As we read more, we are able to picture both the before and after—the chaos of the early morning traffic, the unexpected swerve, the silence that follows a fall, and the final moment when the

motorcycle is straightened up and the journey continues.

The Motorcycle Fall:
The challenges we all encounter can be compared to the fallen motorcycle. The city street is a metaphor for life's journey—unexpected events might knock us off balance, but other people's assistance can get us back on track.

VI. CONCLUSION

We recommend helmet use in images to be identified with transformer models, particularly the Swin transformer. Swin transformers constitute the basis of feature extraction, which uses information extracted from the input image to identify objects. To overcome scale variance in final detection, We fuse the Cascade RCNN architecture with a Feature Pyramid Network (FPN). Top-down feature map combining is done by the FPN neck utilising up-sampled feature maps from the backbone network. methodology. The detector can thus identify objects of different sizes more precisely because of several feature maps with different scales. With the FPN's input, the Cascade RCNN architecture serves as the object detection head. makes use of multi-scale feature maps to find items. A dataset that is open to the public is used to train and evaluate the framework.

VII. REFERENCES

[1]. (Jun. 2022). Road Traffic Injuries. WHO. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[2]. H. A. Abdelali, O. Bourja, R. Haouari, H. Derrouz, Y. Zennayi, F. Bourzex, and R. O. H. Thami, "Visual vehicle tracking via deep learning and particle filter," in Advances on Smart and Soft Computing, F. Saeed, T. Al-Hadhrami, F. Mohammed, and E. Mohammed, Eds. Singapore: Springer, 2021, pp. 517–526.

[3]. H. Derrouz, A. Elbouziady, H. A. Abdelali, R. O. H. Thami, S. El Fkihi, and F. Bourzeix, "Moroccan video intelligent transport system: Vehicle type classification based on three-dimensional and two-dimensional features," IEEE Access, vol. 7, pp. 72528–72537, 2019.

[4]. Z. Charouh, A. Ezzouhri, M. Ghogho, and Z. Guennoun, "Video analysis and rule-based reasoning for driving maneuver classification at intersections," IEEE Access, vol. 10, pp. 45102–45111, 2022.

[5]. A. Ezzouhri, Z. Charouh, M. Ghogho, and Z. Guennoun, "Robust deep learning-based driver distraction detection and classification," IEEE Access, vol. 9, pp. 168080–168092, 2021.

[6]. J. Misachi. (Aug. 2019). Countries With the Highest Motorbike Usage. [Online]. Available: https://www.worldatlas.com/articles/countries-thatride-motorbikes.html

[7]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 25, Dec. 2012, pp. 1097–1105.

[8]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

[9]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.

[10]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.