Motorcycle Recognition System Using Convolutional Neural Network

Jess Z. Masula[1], Reimar R. Tingga[2], Julie Ann A. Salido[3], Jing Chor T. Española[4], Gazle Kent Gillesania[5]*

Abstract: Vehicle detection is becoming increasingly important for highway management, particularly in Aklan, where motorcycles and tricycles are the predominant modes of transportation. Due to their diverse designs, accurate detection of these vehicles remains challenging. This study addresses this issue by developing a vision-based motorcycle recognition system using a Convolutional Neural Network (CNN) implemented in MATLAB. A new high-definition dataset, comprising 34,002 annotated instances from 17,785 motorcycle images and 16,217 tricycle images, was created. The dataset was collected from various locations around Ibajay, Aklan, using regular cameras and smartphones. The images were classified into motorcycles and tricycles and then used to train the CNN model, which was compared with the You Only Look Once (YOLO) version 2 network. Experimental results indicate that the proposed CNN-based motorcycle recognition system achieves high detection accuracy, comparable to the YOLOv2 model. This system contributes to Sustainable Development Goal (SDG) 11 by improving sustainable urban transportation systems and traffic management. Additionally, it holds significant commercial value for motorcycle brands through enhanced vehicle tracking and market analysis.

Keywords: Motorcycle detection, Convolutional Neural Network, Vehicle recognition, YOLO, Traffic management

1. Introduction

Machine learning involves using algorithms to create models that allow computers to learn from data without explicit programming. This technology is now applied to large datasets, driving innovations like

[1] BS Computer Science Program, Aklan State University – Ibajay Campus, Ibajay, Aklan, Philippines
Email: jzmasula@asu.edu.ph
[2] BS Computer Science Program, Aklan State University – Ibajay Campus, Ibajay, Aklan, Philippines
Email: tingga.reimar@asu.edu.ph
[3] College of Computer Studies, Aklan State University – Kalibo Campus, Kalibo, Aklan, Philippines
Email: jasalido@asu.edu.ph
[4] BS Computer Science Program, Aklan State University – Ibajay Campus, Ibajay, Aklan, Philippines
Email: espanola.jingchor@asu.edu.ph
[5]* BS Computer Science Program, Aklan State University – Ibajay Campus, Ibajay, Aklan, Philippines
Email: gillesania.gazlekent@asu.edu.ph (Corresponding Author)

Received [July 11, 2024]; Revised [August 15, 2024]; Accepted [September 2, 2024]

Published by InnoCon Publishing
ISSN 2704-4440

autonomous vehicles. By imitating how humans and animals learn from experiences, it enables machines to do the same. Instead of relying on fixed equations, machine learning algorithms analyze data to uncover patterns and insights. As they process more data, their accuracy and performance improve over time.

Vehicle detection and statistics in highway monitoring video scenes are of considerable significance to intelligent traffic management and control of the highway, especially in the Philippines, where in 2019 there were 121,771 reported vehicular accidents. On average, there were 334 reported accidents per day, with one resulting in fatality, 56 being non-fatal, and 276 leading to damages to property, and 35% of vehicular accidents are from motorcycles [1]. With the popular installation of traffic surveillance cameras, CCTV's, a vast database of traffic video footage has been obtained for analysis. The object size of the vehicle changes greatly at any angle, and the detection accuracy of a small object far away from the road is low. In the face of complex camera scenes, it is essential to effectively solve the above problems and further apply them. This can further be applied in the vehicle detection results to multi-object tracking and vehicle counting.

There are a lot of studies on face recognition nowadays, but it is very difficult to implement because of ethical considerations. Studies of face recognition are one of the most important image processing research topics and are widely used in personal identification, verification, and security applications [2]. It deals with a face recognition system based on principal component analysis (PCA), and the feedforward neural network is developed. The system consists of two phases, which are the PCA preprocessing phase and the neural network classification phase. Automatic recognition of people is a challenging problem that has gained significant attention in recent years due to its broad applications in various fields [3]. Face recognition is one of those challenging problems, and up to date, there is no technique that provides a facial recognition, which is then performed by a probabilistic decision rule. This approach yields highly promising outcomes for facial recognition, effectively handling variations in lighting, facial poses, and expressions. Face recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities [4]. This study proposes a Face Recognition System for personal identification and verification, utilizing Principal Component Analysis (PCA) in conjunction with Back Propagation Neural Networks (BPNN).

Vision-based vehicle object detection is categorized into traditional machine vision techniques and advanced deep learning approaches. Traditional machine vision techniques rely on vehicle motion to distinguish it from a static background. These techniques are further classified into three types: background subtraction, continuous video frame differencing, and optical flow methods [5]. The use of deep convolutional networks (CNNs) has achieved amazing success in the field of object detection. CNNs have a strong ability to learn image features and can perform multiple related tasks, such as classification and bounding box regression.

This study focus on a viable solution to design and develop a motorcycle recognition model and system using a deep convolutional neural network.

Generally, the objective of the study aims to design and develop a motorcycle recognition model and system using deep convolutional neural network.

Specifically, the study aimed to:

1. Collect motorcycle images for machine learning;
2. Build a motorcycle model as an image classifier;

3. Design and develop a motorcycle recognition system; and
4. Determine the accuracy of the model and system.

2. Review of Related Literature

Vehicle detection systems have evolved significantly with the advent of deep learning, particularly Convolutional Neural Networks (CNNs). Traditional methods like background subtraction and optical flow struggled with environmental variations, while CNNs have demonstrated superior performance in recognizing vehicles, especially in real-time traffic systems [6].

CNN-based models such as YOLO (You Only Look Once) are widely used due to their ability to detect objects in a single pass, ensuring both speed and accuracy, which is essential for traffic monitoring [7]. These models can detect multiple objects, including motorcycles, which are harder to identify due to their smaller size and varied designs [8].

Motorcycle recognition plays a critical role in traffic safety, especially in regions like Southeast Asia, where motorcycles are prevalent. Accurate detection of motorcycles helps in reducing traffic accidents and improving road safety [9]. CNN-based systems have shown great promise in identifying motorcycles in real-time, contributing to better traffic management and smart city applications [10].

3. Methodology

This section outlines the process used to develop a Motorcycle Recognition System based on Convolutional Neural Network (CNN) techniques. The methodology is divided into several phases: data collection, preprocessing, model training, and performance evaluation. The process flow is illustrated in Figure 1, which provides an overview of the system architecture and the steps involved in creating and deploying the motorcycle recognition model.

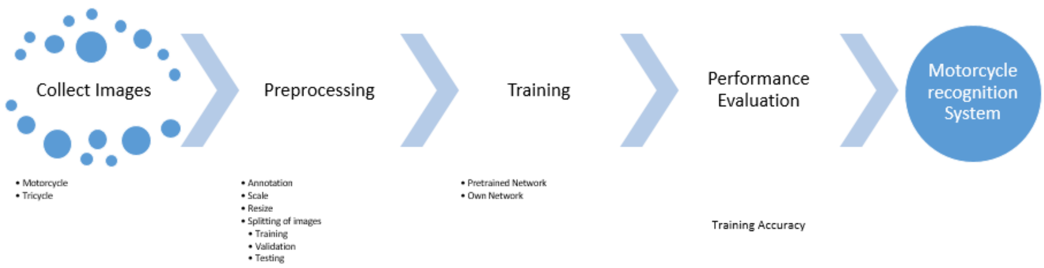

Figure 1. The Approach Used to Create the Motorcycle Recognition System

3.1 Vehicle Dataset

The first step in creating the system is collecting images for tricycles and motorcycles, since there is no known open-source dataset for these. The dataset consists of a total of 34,002 high-resolution images in the classification dataset acquired at different times of the day and different periods of the year by a regular camera and smartphones. Those images have been selected to cover a wide range of challenges and are representative of typical visual data captured today in urban and rural traffic scenarios. Each moving object has been carefully identified and annotated by nearly 100 people to enable a quantitative comparison and ranking of various vehicles. This dataset aims to provide a rigorous benchmarking facility for training and testing existing and new algorithms for the classification of motorcycle and tricycle vehicles in traffic and street scenes. Each dataset image contains one vehicle for training the model of the characteristics and features of each vehicle classification. A new high-definition highway motorcycle and tricycle dataset with a total of 34,002 annotated instances in 17,785 motorcycle images and 16,217 tricycle images. All images that contained more than one vehicle were used in validation and testing; only images that contained a single class of vehicle, such as a motorcycle and tricycle, were used for training. For the recognition system to learn the features of each class.

3.2 Preprocessing

The collected images were annotated, meaning they were labeled according to their classification as either a motorcycle or tricycle. Each training image contained only a single vehicle per frame. The images were then scaled down because they were captured using different devices, leading to variations in their original sizes. To ensure consistency, the images were resized equally for training, as the model requires all input images to be of the same size. Specifically, 70% of the images were randomly selected for training, 10% for validation, and 20% for testing. The images were later preprocessed by resizing them to [224 224 3], where 224 refers to the height and width in pixels, and 3 represents the number of color channels (RGB: Red, Green, and Blue). This consistent resizing ensures that the CNN can process each image uniformly, capturing both spatial and color information necessary for effective classification.

3.3 Training

The motorcycle recognition system using a convolutional neural network classifier consists of two primary components: training and classification. Then image resizing is done to fit the input parameters of the YOLOv2 model and our own model. The two primary categories are motorcycle and tricycle. For the training of images, 70% of the dataset was augmented using rotate, translate, and reflection.

For comparison, we used the YOLOv2 network and compared it with our architecture. The YOLOv2 network was implemented using MATLAB. The yolov2Layers function creates a YOLOv2 network architecture for object detection, and the trainYOLOv2ObjectDetector function was used to train the network [11].

The yolov2Layers uses a pretrained neural network as the base network, to which it adds a detection subnetwork required for creating a YOLO v2 object detection network. In YOLOv2, the yolov2Layers function removes all layers after the feature layer in the base network and attaches a detection subnetwork. This detection subnetwork consists of multiple convolution, ReLU, and batch normalization layers connected in sequence. Additionally, the YOLOv2 transform and output layers are appended to this detection subnetwork. If the 'ReorgLayerSource' parameter is specified, the YOLOv2 network concatenates the output of the reorganization layer with the output of the feature layer.

The images were resized based on the input size parameter of each CNN architecture for both architectures, which is [224 x 224 x 3]. For YOLOv2, it has 9 layers, and our own architecture has also 9 layers.

Convolutional Neural Network (CNN) is a widely used neural network for solving complex deep learning problems. It is designed to autonomously and adaptively learn spatial hierarchies of features through backpropagation, using several building blocks [12]. Its strong classification ability based on contextual information has proven useful in both image processing and natural language processing.

The standard structure of a CNN includes four key components: (a) convolution layer, (b) pooling layer, (c) activation function, and (d) fully connected layer.

The CNN model used in this study consisted of 8 convolution layers. Convolutional layers are the layers where multiple filters are applied to the input image to extract features. In our model, 32, 64, 128, 128, 256, 256, 32, and 64 filters were utilized on each convolutional layer, respectively. Each convolution layer is followed by normalization and a max-pooling layer, which takes the maximum value in a certain filter region. To introduce non-linearity in our CNN, all the layers in the network usually have Rectified Linear Unit (ReLu) activation that replaces all negative pixel values in the feature map by zero. After that, it is followed by two fully connected layers: a dropout layer of 50% and a softmax layer. The output from the convolutional and pooling layers captures the high-level features of the input image. The fully connected layer then utilizes these features to classify the input image into different categories, based on the information learned from the training dataset.

- Learning Rate: $1^{e^{-3}}$
- Optimizer: sgdm (*i.e.,* stochastic gradient descent with momentum)
- Loss: Categorical Cross-Entropy
- Epoch: 400

3.4 Performance evaluation

The classifier model trained using the training images will be used for testing and validation. The model created can identify if the image is that of a motorcycle or tricycle. The classifier model will be evaluated using the standard accuracy formula depicted in equation 1 [13].

$$Accuracy = \frac{(TN+TP)}{(TN+TP+FN+FP)} \tag{1}$$

where TP is the true positive, FN is the false negative, TN is the true negative, and FP is the false positive. The following categories are utilized in deriving our formula for these evaluation metrics. True Positives (TP) are items where the true label is positive and whose class is correctly predicted to be positive. False Positives (FP) are items where the true label is negative and whose class is incorrectly predicted to be positive. True Negatives (TN) are items where the true label is negative and whose class is correctly predicted to be negative. False Negatives (FN) are items where the true label is positive and whose class is incorrectly predicted to be negative.

3.2 Motorcycle Recognition System

The end product of the training is the motorcycle model, which can be implemented to recognize a motorcycle from a tricycle vehicle. Figure 2 shows the flowchart of the motorcycle recognition system. A graphical user interface is used as a way to capture images from a live camera, and then a motorcycle recognition system using the model that is created in the training is used to identify if an image contains a motorcycle or tricycle. The classification is motorcycle or tricycle.

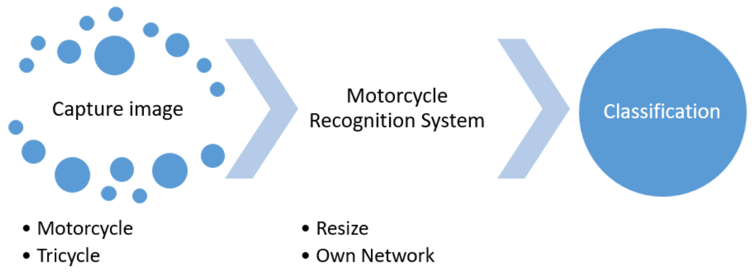

Figure 2. The Approach Used to Test the Motorcycle Recognition System

4. Results and Discussion

All the computations and implementations were performed on a 2.8 GHz Intel(R) Core (TM) i7-7700HQ CPU with 64-bit operating system, 20 GB RAM, and NVDIA GeForce GTX 1050 video RAM, running on a Windows 10 Home edition. The software used for implementation is MATLAB with license number 40766207 of R2020a.

For the preprocessing results, some sample images are shown in Figure 3. The first row shows the motorcycle images; the second row shows the tricycle images.

Figure 3. Sample Images from the Dataset

The performance of the classification is measured using the accuracy formula. Table 1 summarizes the performance accuracies obtained by each deep learning model. All models obtain excellent performance with above 99% accuracy. Of the two, our own architecture exhibited the lowest memory allocation needed for both classifications, with 21MB only compared to the 26MB of YOLOv2, having 99.56% accuracy for the motorcycle recognition system and only a 0.11% difference higher than the present network. It is already comparable since both of them are excellent in performance. Based on the experiments, both architectures have poor recognition for small objects. And as tested with the designs of tricycles, it fails to properly detect objects of new or uncommon shapes, as it does not generalize well beyond the bounding boxes in the training set.

Table 1. Comparison of YOLOv2 and Custom CNN Architecture in Terms of Accuracy and Model Size

Network	Accuracy	Size
YOLOv2 (9 layers)	99.67	26MB
Own Architecture (9 layers)	99.56	21MB

5. Conclusion and Recommendations

This study established a motorcycle and tricycle vehicle object dataset from the perspective of regular cameras and smartphones. The YOLOv2 and our own motorcycle recognition systems, object detection algorithms, obtained the end-to-end vehicle detection model based on the annotated vehicle object dataset. The experimental results confirmed that the proposed vehicle detection and tracking method for highway surveillance videos demonstrates strong performance and practical applicability. In comparison to traditional hardware-based vehicle traffic monitoring methods, the approach presented in this paper is more cost-effective, highly stable, and requires minimal modifications or installation of existing monitoring systems. Additionally, it demands little human intervention. Based on the findings of this research, the surveillance camera can be further calibrated to determine its internal and external parameters. This paper has general practical significance for the management and control of highway scenes and traffic management.

As future works, to address the problem of the small object detection and the multi-scale variation of the object, the road surface area could be determined in the future and a proximal area. The two road areas of each frame for detection to obtain good vehicle detection results in the monitoring could be done after. The position of the object in the image could be predicted for the vehicle trajectory, which could be obtained by tracking multiple objects for a vehicle counting system. To collect the data under the current highway traffic scene, such as driving direction, vehicle type, and vehicle number for added features.

References

[1] Metro Manila Development Authority, "*2019 Metro Manila Accident Reporting and Analysis System (MMARAS)*," https://mmda.gov.ph/images/Home/FOI/MMARAS/MMARAS_Annual_Report_2019.pdf.

[2] A. Eleyan, H. Demirel, M. Anbarjafari, "*Face Recognition System Based on PCA and Feedforward Neural Networks*," in *Computational Intelligence and Bioinspired Systems*, J. Cabestany, A. Prieto, F. Sandoval, (Eds), IWANN 2005, Lecture Notes in Computer Science, vol. 3512, Springer, Berlin, Heidelberg, https://doi.org/10.1007/11494669_115.

[3] J. Nagi, G. Duc Pham, and M. B. Ain, "*A MATLAB Based Face Recognition System Using Image Processing and Neural Networks*," in Proceedings of the 4th IEEE International Colloquium on Signal Processing and its Applications, Kuala Lumpur, Malaysia, March 7-9, 2008, pp. 83-88, ISBN: 978-983-42747-9-5.

[4] B. Sugandi, I. Dewita, R. P. Hudjajanto, "*Face Recognition Based on PCA and Neural Network*," in Proc. 2019 2nd International Conference on Applied Engineering (ICAE), Batam, Indonesia, October 2-3, 2019, pp. 1-5, https://doi.org/10.1109/ICAE47758.2019.9221831.

[5] M. Al-Smadi, K. Abdulrahim, R. A. Salam, “*Traffic Surveillance: A Review of Vision-based Vehicle Detection, Recognition, and Tracking,*” International Journal of Applied Engineering Research, vol. 11, no. 1, February 2016, pp. 713–726, EISSN: ISSN 0973-9769.

[6] E. Shelhamer, J. Long, T. Darrell, “*Fully Convolutional Networks for Semantic Segmentation*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, April 2017, pp. 640–651, https://doi.org/10.1109/TPAMI.2016.2572683.

[7] J. Redmon and A. Farhadi, “*YOLO9000: Better, Faster, Stronger*,” in the Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 31-26, 2017, https://doi.org/10.1109/CVPR.2017.690.

[8] H. Song, H. Liang, H. Li, Z. Dai, X. Yun, “*Vision-based Vehicle Detection and Counting System Using Deep Learning in Highway Scenes*,” European Transport Research Review, vol. 11, December 2019, https://doi.org/10.1186/s12544-019-0390-4.

[9] M. T. Yousif, A. F. M. Sadullah, K. A. A. Kassim, “*A Review of Behavioural Issues Contribution to Motorcycle Safety*,” IATSS Research, vol. 44, no. 2, July 2020, pp. 142-154, https://doi.org/10.1016/j.iatssr.2019.12.001.

[10] K. Simonyan and A. Zisserman, “*Very Deep Convolutional Networks for Large-Scale Image Recognition*,” arXiv:1409.1556, April 2015, https://arxiv.org/abs/1409.1556.

[11] K. T. Islam, R. G. Raj, S. M. S. Islam, S. Wijewickrema, M. S. Hossain, T. Razmovski, S. O’Leary, “*A Vision-Based Machine Learning Method for Barrier Access Control Using Vehicle License Plate Authentication*,” Sensors, vol. 20, no. 12, June 2020, https://doi.org/10.3390/s20123578.

[12] J. A. A. Salido and C. Ruiz, “*Using Deep Learning to Detect Melanoma in Dermoscopy Images*,” International Journal of Machine Learning and Computing, vol. 8, no. 1, February 2018, pp. 1-5, https://doi.org/10.18178/ijmlc.2018.8.1.664.

[13] S. R. Gowtham, “*Confusion Matrix to No Confusion Matrix in Just 5mins*,” Towards AI, June 2022, https://pub.towardsai.net/confusion-matrix-179b9c758b55.