

Visual Recognition to Identify Helmet on Motorcycle Rider Using Convolutional Neural Network

Kevin Alexander^{1*}, Rayhan Ardiya Dwantara², Raihan Muhammad Naufal³, and Derwin Suhartono⁴

^{1–4}Computer Science Department, School of Computer Science, Bina Nusantara University
Jakarta 11480 Indonesia

Email: ¹kevin.alexander002@binus.ac.id, ²rayhan.dwantara@binus.ac.id, ³raihan.naufal@binus.ac.id,
⁴dsuhartono@binus.edu

Abstract—The amount of motorcycle accidents is increasing each year. The main reason is that the riders do not wear a helmet. The research aims to minimize the accident by training the machine learning using the IBM Watson Studio. It trains the data about “wearing helmet” and “not wearing helmet”. The used method is Convolutional Neural Network (CNN). About 170 image datasets are used. CNN is conducted on the input image using a kernel or filter. The filter will multiply its values with the overlapping values of the image while also sliding and adding them all to produce a single value for each of them until the entire images have passed and finished. After CNN method is done, the researchers can classify the images by using supervised learning. It can identify whether the rider is wearing a helmet or not simply by scanning a picture on the street. The result shows high accuracy of 92.87%. The method can be used to minimize the percentage of motorcycle accidents caused by not wearing a helmet.

Index Terms—Visual Recognition, Helmet, Motorcycle Rider, Convolutional Neural Network

I. INTRODUCTION

THE amount of motorcycle riders in Jakarta reaches approximately more than 18 million [1]. There were more than 84,000 cases by them or at least 74% of the traffic violations in 2019. The number of traffic violations increased up to 63.29% than in 2018 [2]. The majority of the traffic violations occur because motorcycle riders refuse to wear a helmet when they are riding. The impact can cause an accident and increase the risk of death.

Until now, the effort by the government to reduce traffic violations is patrolling by police and using cameras that can identify speed, seatbelt usage, and gadget usage while driving [3]. However, these methods are not effective enough for motorcycle riders. The used

cameras are only specialized in cars. For that reason, the research comes up with a new solution for the traffic violations that also involve the usage of the cameras. The idea is to implement a new system that can scan all motorcycle riders that do not wear a helmet. The used method is supervised data training. It means the researchers have to label the data before training them manually. With the new system, it is expected to reduce the number of traffic violations and socialize society about the importance of using helmets.

A. Related Work

Previous studies attempt to adopt machine learning to predict red-light running traffic violations using Support Vector Machine (SVM) and random forest to develop prediction models [4]. Support Vector Machine and Plain Logistic Regression methods are also used in this study with more general observed objects [5]. This SVM method works well when there is a clear margin of separation between classes, but it is not suitable for large data sets. However, the memory of SVM is intensive and a little bit harder to interpret the final model. Currently, the industry more prefers to use a random forest method rather than SVM. The other method that can be used is Scale Invariant Feature Transform (SIFT) [6, 7]. It locates specific key points and furnishes them with quantitative information. SIFT works well when the datasets consist of simpler design and fewer parameters. The drawbacks of SIFT are the poor generalization and not robust to non-linear transformations.

Nowadays, SIFT usually gets outperformed by a method called Convolutional Neural Network (CNN). The previous researchers use a method called Convolutional Neural Network (CNN) as their basic algorithm [8–11]. This method is often used for analyzing

Received: July 09, 2020; received in revised form: Oct. 20, 2020; accepted: Oct. 20, 2020; available online: Nov. 19, 2020.
*Corresponding Author

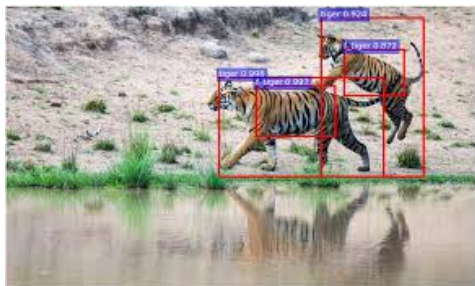


Fig. 1. The example of pattern recognition for animals.

images and improving the scene on image classification. Because of this method, the world of image recognition experiences rapid progress. These studies also make a model that can learn new stuff, such as features with input data and convolutional layers. For processing the 2D image network, it also plays an important role. CNN also only uses a small number of pre-processing steps compared to other methods. CNN advantages involve handling large datasets and classifying unstructured data, like image, speech recognition, and natural language processing. Lastly, the previous study trains a model to detect fake news using natural language processing, which is one of the derivatives of CNN specialized in word embedding for text classification [12]. This method works well within the document-intensive environment and is easy to develop. But, this method requires annotations. It is grindy if the researchers want to make a different language model.

In the research, the researchers emphasize machine learning more because it has made enormous progress compared to the past few years. Not only that, but it also makes an impact on artificial intelligence. It has also been chosen as the first choice regarding developing practical software for computer vision, robot control, speech recognition, and many more [13]. Then, data are also one of the crucial needs for machine learning to develop even further. One of the valuable usages of machine learning is high predictive accuracy [3]. These data acquired by machine learning have many applications, such as finance, bank analysis, fraud detection, stock market, manufacturing, and others.

Figure 1 shows the example of pattern recognition that has been developed until now. The process starts with an observation of the data. This observation also includes instruction and direct experience. It is to make better decision-making based on all data and to find a pattern within the data. Doing it can allow the computer to learn without the need for human interference and adjust automatically based on the

situation. However, to do so, machine learning needs a big library that can contain important data [14]. The goal of the library is to provide the researchers with a highly simple architecture and high modular for dealing with kernel algorithms. However, the kernel algorithm is not the only algorithm that works. The previous researchers find a new algorithm that can compete with the kernel algorithm, which is Gaussian and Bayesian. They have already proven that the Gaussian can be useful for experimenting with the computer. Furthermore, calibration and prediction also have been tested, and it works on both sides [15].

Machine learning also has to learn by experience. One of the methods is providing a stable framework that can support its performance. Probabilistic modeling is one of the most common and trusted ways [16]. Probabilistic modeling provides a framework for understanding what learning is supposed to be. It has become the basic foundation for practical or theoretical models. It also depends on the Bayesian method [17]. There is also a connection between Bayesian optimization and reinforcement learning [13, 18]. It happens because Bayesian optimization is a sequential decision problem that the decision does not affect the state of the system [17].

Next, there is 3D facial recognition since the research aims to determine whether motorcycle riders wear a helmet or not. The main problem of visual recognition is that a person can have the various visual appearance and many different environments [19]. The solution to this problem is by using a facial recognition algorithm that deals with expression and poses variations employing spectral conformal parameterization. By using this algorithm, it can preserve the angle of the human face. However, the downside of this algorithm is that it reduces the dimensionality of the facial model. There are also many advantages, such as combining the conformal parameterization and mean curvature, to produce a better performance than the usual 3D Scan.

For the machine to know what it will scan and compare, the researchers need to train it with the datasets. An image can be represented as a matrix. Each matrix element contains color information for a pixel. With this, the matrix will be used as input data into the neural network. Using this type of method, the researchers can scan the image from the geometrical patterns and the size of the image. Combined with the generalization, it includes invariance with a parallel position and can increase the average recognition rate to 90.70% [20].

A literature review shows that there have been several studies on the traffic violation problem using a statistical approach. However, the researchers can only find a few studies related to the research using a pattern

recognition approach. Previous studies have attempted to look for pattern recognition to identify the driving skill affecting drivers’ behavior on the street [21]. For example, a motorcycle rider has to interact with his/her surrounding, such as a vehicle, another rider, traffic, and many more. That is why the comfort and satisfaction depend not only on the tools or the vehicle but also on the dynamic interaction between the riders and how to operate the vehicle.

For this method to work, the previous researchers use traffic cameras to implement the ideas. To do this, the traffic cameras will be implemented with a performance comparison feature for classification named Histogram of Oriented Gradients, SIFT, and Local Binary Patterns. The results from the traffic cameras by using this method show an accuracy of 93.80% and a processing time of 11.58 ms per frame [22]. Other ways to use this method offers an accuracy of 94.70% by using a hybrid descriptor for the feature extraction [23].

Moreover, in the previous reserach, the lack of visibility, low camera quality, and weather conditions are the weaknesses of the traffic camera. To fix this problem, CNN can be applied for further recognitions. The datasets will be divided into two types, IITH_Helmet_1 for sparse traffic and IITH_HELMET_2 for dense traffic. The results from this method show an accuracy of 92.87% and a false alarm rate of 0.5% by using a real traffic surveillance camera [24]. To ensure that the person scanned is not wearing a helmet, K-Nearest Neighbor (KNN) can classify them. KNN is used to classify the rider’s head by using its features derived from the four sections of the segmented head region (face, nose, mouth, and eyes). However, the accuracy is very low (9%) [25].

As the topic of the research, the researchers use this pattern recognition to identify traffic violations. There are some disadvantages in implementing pattern recognition, such as consuming more time, the need for a larger dataset to increase the accuracy, and incapability of defining whether the object is recognized or not. Additionally, the researchers want to take a new approach for traffic violations, especially on motorcycle riders that do not wear a helmet. It is to prove that pattern recognition can distinguish motorcycle riders who violate the rule.

II. RESEARCH METHOD

A. Image Datasets

Table I shows the amount of data that the researchers obtain from Google about motorcycle rider. Then, the data are divided into two sections, which are “wearing

TABLE I
IMAGE DATASETS IN THE RESEARCH.

| Data | Image |
|--------------------|-------|
| Wearing Helmet | 130 |
| Not Wearing Helmet | 40 |
| Total Dataset | 170 |

helmet” and “not wearing helmet”. The dataset consists of 170 random pictures of motorcycle riders. It includes 130 objects that wear a helmet and 40 objects that do not wear a helmet.

B. Data Processing

The researchers use IBM Watson Studio to train a model and test it into a picture. Basically, CNN has a different architecture than a Regular Neural Network (NN). Regular Neural Network transforms input by putting all data through a series of layers that are hidden. Every layer comprises a neuron set where each layer is connected directly to the previous neuron in the layer. On CNN, it is a bit different. The layers are organized in three dimensions, and only some of the neurons are connected to each layer, as seen in Fig. 2. CNN is generally composed of two parts: feature extraction and classification. Feature extraction is when the algorithm will perform a series of convolutions and pooling operations to detect features. The next major part is classification [10]. The layers are fully connected and will serve as a classifier on top of these extracted features. It will assign the probability of the inputted object, as seen in Fig. 3.

III. RESULTS AND DISCUSSION

The first step in the research is selecting all the pictures to be trained and splitting the datasets into “wearing helmet” and “not wearing helmet”. After selecting all the datasets, the process of labeling the features of the datasets helps the machine learning algorithm to extract the important feature into a model so that it can classify and detect the potential objects in the future. In conjunction with the feature extraction, CNN is conducted on the input image using a kernel or filter. The filter will multiply its value with the overlapping values of the image while also sliding and adding them all to produce a single value for each of them until the entire images have passed and finished. The whole process is called supervised learning using classification techniques. The learning algorithm takes a known set of input data and trains a model to generate reasonable predictions for responding to the new data. After the training session is done, the researchers can test the model by inputting the picture to know if it is a

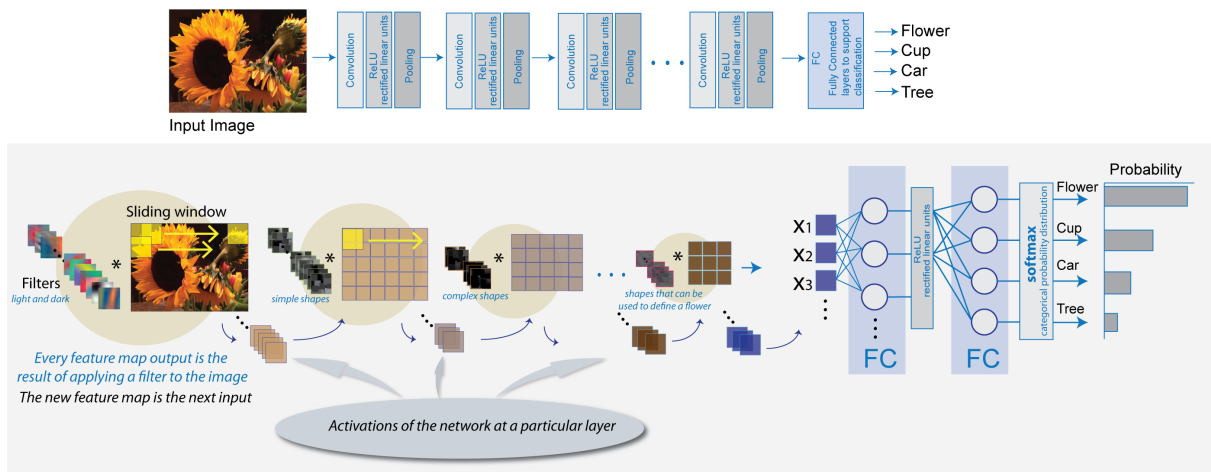


Fig. 2. Overview of Convolutional Neural Network (CNN) processing [26].

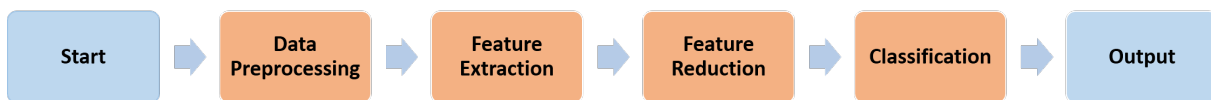


Fig. 3. Processing step of Convolutional Neural Network (CNN).

traffic violation or not (wearing a helmet or not). If the results are still far from what the researchers expect, the previous model will be retrained by adding new datasets into the collection.

Using this program can train all the data by labeling the datasets that the researchers input manually. After labeling the datasets, the researchers can determine whether the riders use a helmet or not because the researchers already pre-set the label in the beginning process. This process is called supervised learning because the researchers have to set and determine a few datasets manually. Later, this program can scan the data and determine whether the motorcycle riders use a helmet or not.

Besides machine learning and visual recognition, the researchers have to think about the security of the system. An adversarial example is the malicious inputs designed to fool the machine learning model. Adversarial often attacks the black box without the knowledge of the target model parameters. One of the methods used to prevent or minimize the chance of being attacked is the one-step target class method. This method works by using Fast Gradient Step Method (FGSM) to finds adversarial perturbations, which increase the value of the loss function. FGSM works by using the gradients of the neural network to create an adversarial example. For an input image, the method

TABLE II
THE RESULTS OF ANALYSIS.

| Method | Score | |
|--|----------|------------------|
| | Accuracy | False Alarm Rate |
| CNN | 92.87% | 0.5% |
| Local Binary Pattern, Histogram of Oriented Gradients, and Hough transform descriptors | 93.80% | 9% |
| Local Binary Pattern, Histogram of Oriented Gradients, and Hough transform with a hybrid descriptor for the feature extraction | 94.70% | 9% |

uses the gradients of the loss concerning the input image to create a new image that maximizes the loss. The method can minimize the data to be transferred between networks, which provides indirect robustness against the black box.

CNN uses image recognition and classification to detect the objects. They are made of neurons with learnable weights and biases. Each specific neuron receives numerous inputs and takes a weighted sum over them. It passes through an activation function, a repetition of the same filter to an input results, and responds with an output. The data in Table II are the percentage of methods used. It shows the

result with the confidence level at 92.87%. The result can be considered pretty high, and it has a false alarm rate of 0.5%. The result can go higher with the larger datasets, but for now, the researchers will leave the value as it is. However, using Local Binary Pattern, Histogram of Oriented Gradients, and Hough Transform descriptors, the result is slightly higher with the accuracy of 93.80%. Moreover, adding a hybrid descriptor for feature extraction has a higher accuracy of 94.70%. Even though the results are higher, but the percentages of false alarm are also higher, which is 9%. When the results are compared to the previous research, coincidentally, the results are similar.

From this data, the researchers can conclude that the result is more stable by using CNN compared to Local Binary Pattern, Histograms of Oriented Gradients, and Hough and Transform descriptors. There are also some errors in some of the results, caused by a lack of datasets. The common error is that the algorithm determines the wrong place for the label and misinterprets the picture. The implementation of supervised learning using the classification techniques shows that the algorithm will receive a set of data. Then, it will train the model to generate a prediction as accurate as it can according to the data that have been selected before. This method is also applied to the result of Table II. From the results, machine learning can distinguish the difference from the requirement given. It gives a result and decides which picture meets the requirement of "wearing helmet" and "not wearing helmet".

IV. CONCLUSION

The research presents a method that can be used to identify whether the motorcycle riders wear a helmet or not. The used method is CNN. The method can be implemented to minimize the motorcycle incidents that may happen and prevent unwanted accidents on the road. The result shows high accuracy of 92.87%. Although the result is not higher than other methods, it is more stable.

The weakness of the research is due to using a street camera as the vision and database. The research cannot scan an object that is covered or behind by another object. For future research, the researchers would like to find how to implement this method with a street camera without having the problem of an object that is out of reach.

REFERENCES

- [1] G. Y. N. N. Adi, M. H. Tandio, V. Ong, and D. Suhartono, "Optimization for automatic personality recognition on Twitter in Bahasa Indonesia," *Procedia Computer Science*, vol. 135, pp. 473–480, 2018.
- [2] R. Kurniawan. (2019) Angka kecelakaan lalu lintas di 2019 meningkat. [Online]. Available: <https://otomotif.kompas.com/read/2019/12/30/172100015/angka-kecelakaan-lalu-lintas-di-2019-meningkat>
- [3] E. Alpaydin, *Introduction to machine learning*. Cambridge, Massachusetts: MIT Press, 2020.
- [4] A. Jahangiri, H. A. Rakha, and T. A. Dingus, "Adopting machine learning methods to predict red-light running violations," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Las Palmas, Spain: IEEE, Sept. 15–18, 2015, pp. 650–655.
- [5] L. J. Li, H. Su, Y. Lim, and L. Fei-Fei, "Object bank: An object-level image representation for high-level visual recognition," *International Journal of Computer Vision*, vol. 107, no. 1, pp. 20–39, 2014.
- [6] C. Cusano and P. Napoletano, "Visual recognition of aircraft mechanical parts for smart maintenance," *Computers in Industry*, vol. 86, no. April, pp. 26–33, 2017.
- [7] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, 2014.
- [8] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. Turin, Italy: IEEE, June 29–July 3, 2015, pp. 1–6.
- [9] X. Sun, J. Shi, J. Dong, and X. Wang, "Fish recognition from low-resolution underwater images," in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. Datong, China: IEEE, Oct. 15–17, 2016, pp. 471–476.
- [10] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–11, 2016.
- [11] E. Aguilar, M. Bolaños, and P. Radeva, "Food recognition using fusion of classifiers based on CNNs," in *International Conference on Image Analysis and Processing*. Catania, Italy: Springer, Sept. 11–15, 2017, pp. 213–224.
- [12] S. Gilda, "Evaluating machine learning algo-

- rithms for fake news detection,” in *2017 IEEE 15th Student Conference on Research and Development (SCORED)*. Putrajaya, Malaysia: IEEE, Dec. 13–14, 2017, pp. 110–115.
- [13] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [14] D. E. King, “Dlib-ml: A machine learning toolkit,” *The Journal of Machine Learning Research*, vol. 10, no. 60, pp. 1755–1758, 2009.
- [15] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” in *Advances in Neural Information Processing Systems 25*, Lake Tahoe, Nevada, USA, Dec. 3–6, 2012, pp. 2951–2959.
- [16] Z. Ghahramani, “Probabilistic machine learning and artificial intelligence,” *Nature*, vol. 521, no. 7553, pp. 452–459, 2015.
- [17] M. Feurer, A. Klein, K. Eggenberger, J. Springenberg, M. Blum, and F. Hutter, “Efficient and robust automated machine learning,” in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, Quebec, Canada, Dec. 7–12, 2015, pp. 2962–2970.
- [18] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, “TensorFlow: A system for large-scale machine learning,” in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, Savannah, GA, USA, Nov. 2–4, 2016, pp. 265–283.
- [19] B. A. Echeagaray-Patron, D. Miramontes-Jaramillo, and V. Kober, “Conformal parameterization and curvature analysis for 3D facial recognition,” in *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*. Las Vegas, NV, USA: IEEE, Dec. 7–9, 2015, pp. 843–844.
- [20] A. E. Pandelea, M. Budescu, and G. Covatariu, “Image processing using Artificial Neural Networks,” *Buletinul Institutului Politehnic din Iasi. Sectia Constructii, Arhitectura*, vol. 61, no. 4, pp. 9–21, 2015.
- [21] Y. Zhang, W. C. Lin, and Y. K. S. Chin, “A pattern-recognition approach for driving skill characterization,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 905–916, 2010.
- [22] 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)*. Vancouver, BC, Canada: IEEE, July 24–29, 2016, pp. 3046–3051.
- [23] J. Mistry, A. K. Misraa, M. Agarwal, A. Vyas, V. M. Chudasama, and K. P. Upla, “An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using Convolutional Neural Network,” in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. Montreal, QC, Canada: IEEE, Nov. 28–Dec. 1, 2017, pp. 1–6.
- [24] C. Vishnu, D. Singh, C. K. Mohan, and S. Babu, “Detection of motorcyclists without helmet in videos using Convolutional Neural Network,” in *2017 International Joint Conference on Neural Networks (IJCNN)*. Anchorage, AK, USA: IEEE, May 14–19, 2017, pp. 3036–3041.
- [25] P. Doungmala and K. Klubsuwan, “Helmet wearing detection in Thailand using haar like feature and circle hough transform on image processing,” in *2016 IEEE International Conference on Computer and Information Technology (CIT)*. Nadi, Fiji: IEEE, Dec. 8–10, 2016, pp. 611–614.
- [26] MathWorks. What is deep learning? How it works, techniques & applications. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html?s_eid=PSM_da