International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

A Hybrid Convolutional Neural Network-Gradient Boosted Classifier for Vehicle Classification

Marlon D. Hernandez, Arnel C. Fajardo, Ruji P. Medina

Abstract*: Vehicle tracking and classification are used for intelligent transport system to provide data in terms of traffic management, routing, vehicle volume and others. A new approach will be introduced in this paper, a hybrid classifier that would detect vehicles that would be adaptable to Philippine settings. A combination of convolutional neural network and gradient boosted classifier would boost the classifying accuracy. In the discussion, CNN has outperformed other classifier in terms of accuracy while GBC got the highest AUROC and highest accuracy in terms of classifying. Although CNN and GBC is prone to overfitting, the dataset that will be used contains 1 hour of video.

***Index Terms*: Classification, Computer Vision, Convolutional Neural Network, Gradient Boosted Tree, Vehicle detection**

I. INTRODUCTION

The detection of vehicles is still a challenging task for Intelligent Transport System (ITS) because of the different factors like extreme lightning, high speed vehicles, low resolution of camera and others. Since vision sensors is the technology used in developing traffic surveillance systems because it can store more data and it can be unsupervised. Most common vision sensors are thermal camera and closed circuit television camera which can gather information through scenes and comparing it to other sensors, vision sensors can recognize path navigation, speed, and classification of objects. Applying technology being used may help in gathering information in ITS using sensors such as magnetic sensors, loop detectors [1][2]. With some of the advantages of using computer vision in transport system like ease to install and maintenance, computer vision based system have drawbacks like identification of vehicles due to congestion, fog, weather (rain/snow) and ambient light that may decrease the performance of image drastically. There are various algorithm that was previously applied in the system for the purpose of object detection in an extent of pattern of movement in the region of computer vision. A wireless sensor network (WSN) [3][4] was used to detect the flow of traffic which resulted to efficient green time scheduling. The combination of hue, lightning and saturation (HLS) and fuzzy logic for edge detection that was used in plate detection. Chinese plates are detected using YDbDr color space that provided a better result that the conventional red, green and blue (RGB). Vertical edge detection algorithm (VEDA) was also used for plate detection which performed five times faster than Sobel filter. A number algorithm was also used to the background and the changes of appearance of the next frame. The echnical contribution of the paper is a new framework with the use of computer vision and classification algorithms, that can robust the classification that is intended for gathering data. The framework that was stablished may be adaptable to Philippine settings but there are unique transportation in each country that may not detect and the framework may have inaccurate results. Most of the common classifications that are used are cars, van, SUV, trucks, bus and motorcycles. In the Philippine settings, the unique transportations are jeepney and tricycle which are not yet included in other current frameworks. The use of hybrid of convolutional neural network (CNN) [5][6] and gradient boosted tree (GBT)[7][8] will robust the classification.

II. PROPOSED STUDY

The general objective of this study is to create a new model for ITS that will be appropriate in the Philippines with the following specific objectives. To formulate a hybrid model for classification of vehicle tracking, classifying and detecting that is suitable for Philippine settings and to determine the level of performance of the model in terms of accuracy. With the combination of CNN and GBC, the accuracy level will be higher.

III. CONVOLUTIONAL NEURAL NETWORK

In the application of machine learning in the vehicle detection and classification. Machine learning is the science of making computers learn and perform like humans by giving information and data without being explicitly programmed. One of the most famous machine learning is CNN which is a forward-structured neural network[8] which is the third phase of the conceptual framework. There are four key characteristics of CNN: weight sharing, multi-layer use, local connection and pooling[9]. Furthermore, CNN has a special weight sharing, deeper network can be formulated for better performance in the application of better complex visual task. Weight sharing decreases the quantity of adjustable parameters, thus increasing the training speed and reducing the threat of overfitting[10].

Revised Manuscript Received on 30 July 2019.

* Correspondence Author
Marlon D. Hernandez*, Graduate School, Technological Institute of the Philippines-Quezon City, Philippines and Department of Information Technology, Bulacan State University, Philippines
Arnel C. Fajardo, School of Engineering and Information Technology, Manuel L. Quezon University, Philippines
Ruji P. Medina, Graduate School, Technological Institute of the Philippines-Quezon City, Philippines.

The formula is a feed forward neural network which is consist of three different layers: input layer, hidden layer and the output layer. Each of the layers consist of different neurons as shown in figure 9. The neuron is a computational sum of inputs and off-set of 1. The output will be

$$h_{w,b}(X) = f(W^T x) = f(\sum_{i=1}^{3} W_i X_i + b) \quad [10]$$

which the multiplication sum operation between weight and output. After the equation, the bias is added to calculate for the non-linear transformation. In figure 10, it is illustrated that a three layer feed forward neural network that has input layer, implicit layer, and an output layer. The structure of a 3-3-1 neural network is shown in the figure. In the formula below, it shows the computational process from the input layer to the output layer, which is commonly known as the forward propagation.

IV. GRADIENT BOOSTED CLASSIFIER

Gradient boosted (GB) is one of the machine learning technique that is used for prediction and classification. The construction of the GB is the ensembles the weak prediction. GB is iteratively learning an ensemble of weak classifier and uniting them to a strong classifier to deliver the final result [11]. In terms of decision trees features, weak learners are shallow trees, occasionally even small as decision stump which is composed of trees with two leaves. There several ensemble methods to construct a decision tree like bagging decision tree, random forest, rotation forest and boosted tree. Boosting decreases error mainly by reducing bias and also to some variance, by gathering the output from many models. The boosting algorithm is based on an idea of accepting multiple expert's decision rather than consulting one expert only. GB is one of the variety of boosting algorithms. The simple training prediction model is based on simple multiple model. Each model is trained for error from the previous model. The gradient boosted follows a sequence, it starts with an initial regression that will form a tree. The next regression tree is trained from the outcomes of the previous tree. As the number of trees are generated, the more accuracy of the model for predictive result is attained. Here are some of the features of decision trees:

1. The simplicity, easy to use, easy to understand and explain.

2. It do variable screening and feature selection.

3. Requires little effort to produce and prepare.

4. Can live with nonlinear parameters and relationships.

5. Can use bayesian theorem or conditional probability based reasoning.

6. Provides strategic answer to uncertain situations.

The calculations are shown [12]:

Input: loss function L(y, F(x)), training set D= {(x1,y1),…,(xn,yn)}, and the number of iteration M.

$$F_0(x) = \underset{h(x)}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_1, h(x))$$

For 1 to M

$$\gamma_{m=} \underset{y}{\operatorname{argmin}} \sum_{i=1}^{n} L\left(y_{1,} F_{m-1}(x_i) - \gamma \frac{\partial L(y_{1,} F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right)$$

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \nabla_F (y_{1,} F_{m-1}(x_i))$$

Where L(y, F(x)) is the loss function, D is the training set (x1,y1) is the first set of data, the number of iteration is M, the final model is $F_m(x)$, the initial model is $F_0(x)$ the prediction made by the model is h(x), the weight of the model mth is γ_m, and the model after iteration m is F_m.

V. DATA

The collected data was from the city traffic management office of San Jose del Monte Bulacan. Most of the barangays had set up a CCTV Cameras not only to monitor traffic but to monitor crimes as well. The extraction of the 1 hour video was be processed for the training testing. Comparing data must have a higher accuracy from the previous technology. The location is Sampol Market, San Jose del Monte Bulacan. One of the busiest streets in the municipality where all public and private transport pass by. The characteristics of each vehicle has distinct features and the extracted of each vehicle image provides the differentiation towards classification.

VI. PRE PROCESSING

From the video, the extracted images per 20 frames was conducted producing 4498 images with 1280 x 720 resolution. Selecting four region of interest (ROI) in the entire frame where most of the vehicles pass through. Having 17,992 images. Each ROI has different width and height, this is intended for capturing the detail of each vehicle to pass in the specific region. ROI A and B is intended for medium to large vehicle like car, jeep, van, SUV bus and trucks with the different position which can be traced in figure 1. ROI C and D is for small vehicles like tricycle and motorcycle. Having a large ROI for a small vehicle would be difficult to extract the small vehicles due to disturbance of the background. Cropping an image is done in images when the region is very large[13]. From each image, image subtraction is made to extract the appropriate image.

Retrieval Number B1016078219/19©BEIESP
DOI: 10.35940/ijrte.B1016.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

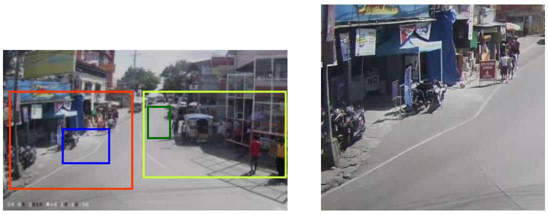


Figure 1 (left) Original Image with specific ROI (right) ROI A

Figure 2 (left) ROI A with jeep (right) extracted image

Figure 1(left) is the original image on frame number 4494 and figure 1 (right) is the region of interest (ROI) A. The coverage of ROI A is intended north bound and has the scale for bus and trucks. The separation of regions will determine each class per selected area. Figure 2 (left) is the cropped image from image number 383, resized to 120x120 pixel and converted to grayscale. Image subtraction from the original image to extract the vehicle which is found in figure 2(right).

VII. METHOD

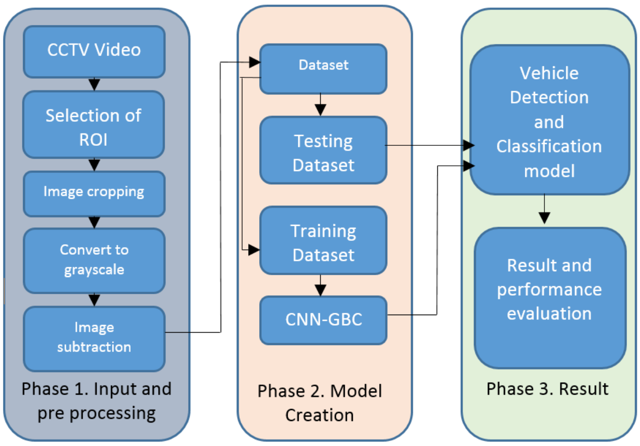


Figure 3. Concept of the study

The concept of the study has 3 different phases, Input and pre-processing, Model creation and results. Phase 1 is concentrated in the preprocessing method and strategy to gather dataset. Phase 2 includes the splitting of data set to testing dataset and training set. Model creating also test the creation of the new architecture. Phase 3 is testing the model and performance evaluation. The proposed method consist of preprocessing using CNN, and classifying using GBC found in figure 4. In this preprocess, max pooling is used with a matrix of 3x3, convolving the layer and activation of the layer. Fully connected network will be used to implement the general purpose classifier from the feature extracted by the previous layers. In this paper, the fully connected network will be replaced by GBC to forecast the labels of the input outline. Once the gradient boosted classifier is well trained, it will execute a recognition task and makes fresh decisions on testing images with such automatically extracted features.

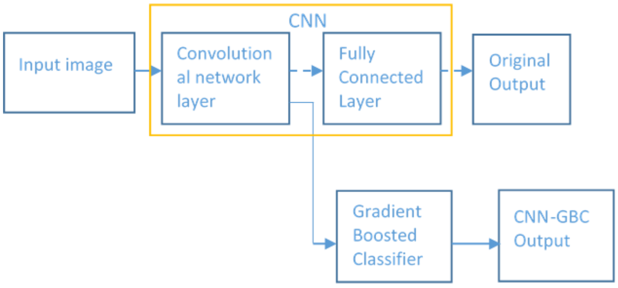


Figure 4. Structure of the hybrid CNN-GBC method

An extracted architecture of the hybrid CNN-GBC architecture that would perform the feature extraction of the image consist of convolving the layers, RELU and pooling. The classification will be handled by the gradient boosted classification instead of the traditional fully connected layers.

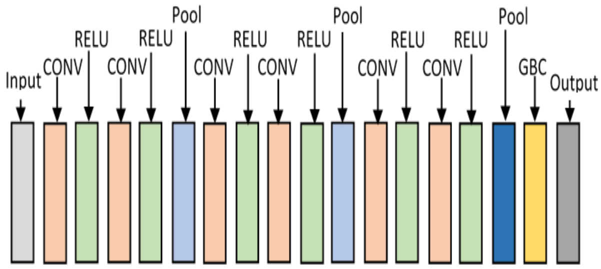
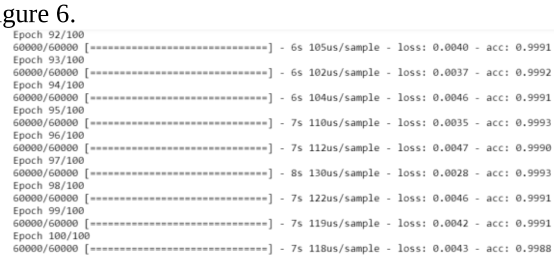


Figure 5. CNN-GBC Architecture

The GBC will be trained for the accuracy of classification, training the GBC more corresponds to more detailed classification, hence the training and testing may take some time but the result is more important. There will also be a comparison if the between the extraction layers to a more simplified extraction layers meaning some of the layers will be eliminated.

VIII. RESULTS AND ANALYSIS

Evaluating the results of the approach, CNN run through 100 epoch with high results. The result is then passed through GBC for further classification such as different classes of jeep (owner type jeep and passenger jeepney) to be labeled as jeepney or car (sedan, hatchback, and others) as car. After training the model for accuracy at 100 epochs using Anaconda and Jupyter notebook. The results show that the highest accuracy rate was noted at 95 and 97 epochs found in figure 6.



```
Epoch 92/100
60000/60000 [==============================] - 6s 105us/sample - loss: 0.0040 - acc: 0.9991
Epoch 93/100
60000/60000 [==============================] - 6s 102us/sample - loss: 0.0037 - acc: 0.9992
Epoch 94/100
60000/60000 [==============================] - 6s 104us/sample - loss: 0.0046 - acc: 0.9991
Epoch 95/100
60000/60000 [==============================] - 7s 110us/sample - loss: 0.0035 - acc: 0.9993
Epoch 96/100
60000/60000 [==============================] - 7s 112us/sample - loss: 0.0047 - acc: 0.9990
Epoch 97/100
60000/60000 [==============================] - 8s 130us/sample - loss: 0.0028 - acc: 0.9993
Epoch 98/100
60000/60000 [==============================] - 7s 122us/sample - loss: 0.0046 - acc: 0.9991
Epoch 99/100
60000/60000 [==============================] - 7s 119us/sample - loss: 0.0042 - acc: 0.9991
Epoch 100/100
60000/60000 [==============================] - 7s 118us/sample - loss: 0.0043 - acc: 0.9988
```

Figure 6. Model Testing

After reaching 0.993 values in accuracy, the sudden drop of values in the next iteration and the values goes up again through machine learning as shown in figure 7.

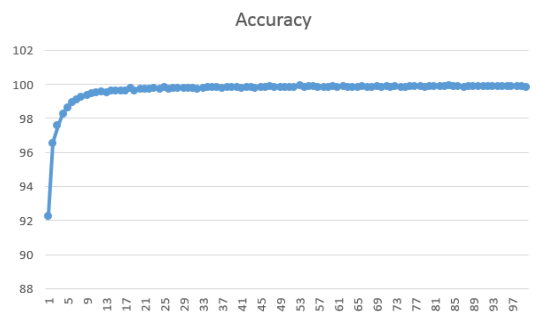


Figure 7. Test of Accuracy

Having high accuracy, tend to have loss to optimize the algorithm of machine learning.

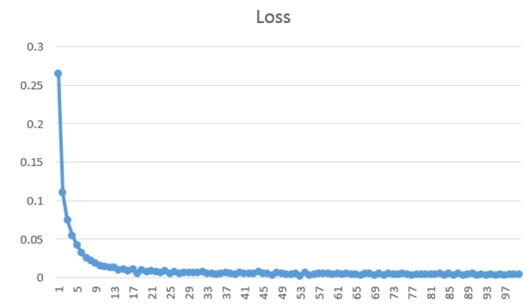


Figure 8. Test of Loss

The measurement of loss of 0.26 at the start of the epoch, the lowest measurement is 0.0028. Like in the accuracy, loss also has a fluctuating value found in figure 8.

IX. CONCLUSION AND FUTURE WORKS

The paper was able to present a high accuracy based on the dataset used for vehicle classification on a Philippine environment. The new collection of dataset was formulated to have a precise distinction of jeepney and tricycle which are not yet included from the previous works. The new framework for vehicle classification showed that this can be a reference of collecting data from particular location for strategic traffic management and can be used in intelligent transport system. The future works may focus on overfitting that CNN and GBC is prone due to high accuracy.

ACKNOWLEDGMENT

The authors would like to thank Technological Institute of the Philippines, Bulacan State University and Manuel L. Quezon University.

REFERENCES

1. S. S. Sarikan, A. M. Ozbayoglu, and O. Zilci, "Automated Vehicle Classification with Image Processing and Computational Intelligence," *Procedia Comput. Sci.*, vol. 114, pp. 515–522, 2017.
2. A. Ambardekar, G. Bebis, and M. Nicolescu, "Visual traffic surveillance framework : classification to event detection Visual traffic surveillance framework :," 2013.
3. J. J. Anaya, E. Talavera, F. Jiménez, F. Serradilla, and J. E. Naranjo, "Ad Hoc Networks Vehicle to Vehicle GeoNetworking using Wireless Sensor Networks," vol. 27, pp. 133–146, 2015.
4. H. Mostafaei and M. Menth, "Journal of Network and Computer Applications Software-de fi ned wireless sensor networks : A survey," vol. 119, no. June, pp. 42–56, 2018.
5. A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, and D. Ruiz-Hernández, "A survey of artificial neural network in wind energy systems," *Appl. Energy*, vol. 228, no. July, pp. 1822–1836, 2018.
6. G. Peiyuan, B. Man, Q. Shiha, and C. Tianhua, "Detection of meat fresh degree based on neural network," *Proc. 2007 IEEE Int. Conf. Mechatronics Autom. ICMA 2007*, pp. 2726–2730, 2007.
7. H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput. J.*, vol. 74, pp. 634–642, 2018.
8. D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of Co-salient Objects by Looking Deep and Wide," *Int. J. Comput. Vis.*, no. April, 2016.
9. L. Zhang, M. Song, Y. Yang, Q. Zhao, C. Zhao, and N. Sebe, "Weakly Supervised Photo Cropping," no. c, pp. 1–13, 2013.
10. L. Zhao, Y. Zhou, H. Lu, and H. Fujita, "Parallel computing method of deep belief networks and its application to traffic flow prediction," *Knowledge-Based Syst.*, 2018.
11. L. Guelman, "Gradient boosting trees for auto insurance loss cost modeling and prediction," *Gradient Boost. trees auto Insur. loss cost Model. Predict.*, vol. 39, no. 3, pp. 3659–3667, 2012.
12. Y. Song, H. Zhou, P. Wang, and M. Yang, "Prediction of clathrate hydrate phase equilibria using gradient boosted regression trees and deep neural networks," *J. Chem. Thermodyn.*, vol. 135, pp. 86–96, 2019.
13. A. V Kavitha, A. Srikrishna, and C. Satyanarayana, "Crop image classification using spherical contact distributions from remote sensing images," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, pp. 0–11, 2019.

AUTHORS PROFILE

Marlon D. Hernandez obtained his Bachelor of Science degree in Computer Engineering from Bulacan State University. In 2004, he obtained a separate 2 year course Cyber programming with e-commerce from System Technology Institute Malolos. He finished his Master's degree in Engineering with specialization in Computer Engineering in 2016 from Bulacan State University. He is currently taking up Doctor of Engineering with specialization in Computer Engineering from the Technological Institute of the Philippines in Quezon City, Philippines. His research interests include Image Processing, Artificial Intelligence, Network Programming, 3D animation and Information Technology education.

Arnel C. Fajardo received his B.S. Electrical Engineering degree in 1991 from the Mapúa Institute of Technology in Intramuros, Manila. He obtained his M.S. Computer Science degree in 1999 from De La Salle University Manila. He obtained his Ph.D. in Computer Engineering degree from Hanbat National University, Daejeon, South Korea in 2014. He is the Senior Assistant Vice President and Research Director of Manuel L. Quezon University and Professorial lecturer of various universities in the Philippines. He is an author of more than 30 articles in journals and conference proceedings. His research interest includes speech recognition, image processing, artificial intelligence and engineering education

Ruji P. Medina is Dean of Graduate Programs of the Technological Institute of the Philippines in Quezon City, Philippines. He received his B.S. Chemical Engineering degree from the University of the Philippines – Diliman, Quezon City in 1992. He graduated summa cum laude from the Mapúa Institute of Technology in Intramuros, Manila in 2000 with a M.S. in Environmental Engineering degree. He holds a Ph.D. in Environmental Engineering from the University of the Philippines with a sandwich program at the University of Houston, Texas. He counts among his expertise environmental and mathematical modeling, urban mining, and nanomaterials. Apart from his active role in research and graduate engineering education, he is an