Implementation Of Data Augmentation In Convolutional Neural Network And Gradient Boosted Classifier For Vehicle Classification

Marlon D. Hernandez, Arnel C. Fajardo, Ruji P. Medina, Jamaica T. Hernandez, Rhowell M. Dellosa

Abstract: the implementation of Gradient Boosted Classifier and Convolutional Neural Network in vehicle classification and detection for the intelligent Transport System that can provide vehicle volume, routing and traffic management. A new approach in this paper is that hybrid algorithm that caused overfitting was minimized proven thru test and comparison of the hybrid algorithm without data augmentation and with data augmentation. With still high accuracy of 98.91 and lesser loss 0.25.

Index Terms: Data Augmentation, Computer Vision, Convolutional Neural Network, Gradient Boosted Tree, Vehicle detection.

1. INTRODUCTION

Researchers are using Convolutional Neural Network (CNN) for its capability to achieve high accuracy level in terms of performance indicator [1]. Despite of this capability, it has some challenges to deal with. Some of the struggle of CNN are lack of training set reliability due to large size of network that reach millions of parameters, there is also the avoidance of adversarial attacks, and with overfitting of the generalization of abilities. One of the techniques to prevent overfitting in Convolutional Neural Network is by using Data Augmentation (DA) to increase the number of dataset [2]. DA addresses the problem thru its roots, the training dataset. It is done thru the assumption that there are more information that can be extracted using the original dataset by using data augmentation [3]. The motivation of solution to lessen overfitting is both broad and specific. Data from video and image classification contains insufficient data. This is evident in cases of medical [4], government and even surveillance which has limited access due to data privacy act. Some techniques had been developed for image processing which combined pre-trained models and expert domain knowledge. A combination of different data augmentation technique and CNN is the proposed model in this paper.

2 PROPOSED STUDY

The general objective of this study is to create a new model for Intelligent Transport System that will be appropriate in the Philippines with the specific objectives. To integrate different data augmentation technique to CNN to lessen overfitting, formulate a 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.

3 CONVOLUTIONAL NEURAL NETWORK

In the application of machine learning in the vehicle classification and detection. Machine learning (ML) is the science of allowing computers perform and learn like humans by giving data and information without being explicitly programmed. One of the most famous machine learning is CNN which is a forward-structured neural network[5] which is the third phase of the conceptual framework. There are four main features of CNN: local connection, weight sharing, multi-layer use, and pooling[6]. Deeper network is formulated to prove better performance in the application of better complicated visual task which CNN obtains the special weight sharing [1]. Weight sharing decreases the quantity of adjustable parameters, thus increasing the training speed and reducing the threat of overfitting[7]. There are three different layers in the formula of feed forward neural network[8] which is consist of three different layers: input layer, hidden layer and the output layer. Each of the layers consist of different neurons. The neuron is a computational sum of inputs and off-set of 1. The output will be

$$h_{w,b}(X) = f(W^T x) = f(\textstyle\sum_{i=1}^{3} W_i X_i + b) \text{ [7]}$$

The multiplication sum operation between weight and output. The bias is added to calculate for the non-linear transformation after the equation. In this manner, the three layer of the feed forward will be the input layer, the implicit layer and the output layer. The formula of the feed forward propagation is shows the computational process from the input layer to the output layer.

4 DATA AUGMENTATION

Data augmentation is a technique that performs increasing the dataset significantly to be used either in training and testing dataset. Cited before that overfitting as one of the problems of CNN and Deep learning due to high accuracy and is dependent to the training and testing dataset alone. Having small dataset means a high risk of overfitting. Data augmentation elaborates the data to fit in the dataset that is similar to the data with the given functions of cropping, rotating, padding and others and can be combined to create a reliable data[9]. The assumption of DA is to introduce new patterns that can be established even if the data does not fit in the original pattern. DA is widely used in image

- *Marlon D. Hernandez, Currently pursuing Doctor of Engineering at Technological Institute of the Philippines-Quezon City, Philippines and a faculty of Department of Information Technology, Bulacan State University, Philippines. Email: marrlon.hernandez@bulsu.edu.ph*
- *Arnel C. Fajardo, School of Engineering and Information Technology, Manuel L. Quezon University, Philippines. Email: acfajardo2011@yahoo.com*
- *Ruji P. Medina, Graduate School, Technological Institute of the Philippines-Quezon City, Philippines. Email ruji.medina@tip.edu.ph*
- *Jamaica T. Hernandez pursuing Master in Information Technology. Faculty of Bulacan State University, Philippines. Email: Jamaica,Tarroza@yahoo.com*
- *Rhowell M. Dellosa, Pursuing Doctor of Engineering at Technological Institue of the Philippines. Email:*

classification[10] that tends to extract original data and outperforms the existing dataset to provide a more reliable collection of data. Hybrid or combination of different DA had been used to utilize data in different forms.

5 RELATED WORKS

Several studies shows the effectiveness of using Image Augmentation (IA) in image dataset that resulted to standard outcome. Some of the image dataset are tiny-imagenet-200, CIFAR-10/100, MIT-Adobe 5K dataset, MNIST, Stanford Cars and others [3]. The performance of image augmentation has different techniques which has basically the basic form of random cropping, color space augmentation and flipping horizontal. The use of IA and combining it to another technique such as flipping, random erase, cropping, and color shift can result enormously exaggerated dataset size. With this note, this is not a guarantee that it would be an advantage. The application of selection through various models must be determined to have a high accuracy but lesser loss.

Table 1 *The result of Data Augmentation Experiment by Taylor and Nitchke [11].*

DA	Top-1 accuracy (%)	Top-5 accuracy (%)
Baseline	48.13 ± 0.42	64.50 ± 0.65
Flipping	49.73 + 1.13	67.36 ± 138
Rotating	50.80 ± 0.63	69.41 ± 0.48
Cropping	61.95 + 1.01	79.10 ± 0.80
Color jittering	49.57 ± 0.53	67.18 ± 0.42
Edge Enhancement	49.29 + 1.16	66.49 + 0.84
Fancy PCA	49.41 ± 0.84	67.54 ± 1.01

Based from the experiment of Taylor and Nitchke from the Caltech101 dataset on table 1, the comparison of different augmentation was performed with different parameters. For the geometric transformation the first was flipping horizontal and vertical, rotating starts from +30° to -30°, and cropping. For the color space transformation, were color jittering, edge enhancement and Fancy PCA. The Caltech101 dataset was 8421 images with 256x256 size was tested with 4-fold cross-validation. The static image augmentation technique is applicable in time series that is applied in video cameras and surveillance cameras. However, the procedures shown accuracy based on the performance of still images and transpose the dataset to compare to video file images. This allows delay for the detection of the vehicle with the simulations process and as well as the classification. CNN requires big data to have a more accurate classification, thus having small data may result to overfitting. Some of the methods used to prevent this are dropouts [12], batch normalization [13], transfer learning [14], Pre-training [15], one-shot and zero-shot learning [16]. Another process was proposed by Liang by using Generative Adversarial Networks (GAN). It is a powerful method to execute unsupervised generation of fresh images for training. It was also proven enormously effective in big data generation, an example is novel paragraph generation [17].

6 TRADITIONAL TRANSFORMATION

Traditional transformation consist of basic augmentation such as cropping, flipping and rotating. Advance traditional transformation is the combination of two or more ordinary augmentation. In the figure below, each augmented image that is defined as cropped, cropped and flipped, and cropped and rotate in +30° and -30°. Both the augmented images are fed into the model. For the image dataset size of M, we generated a dataset of 4M size.

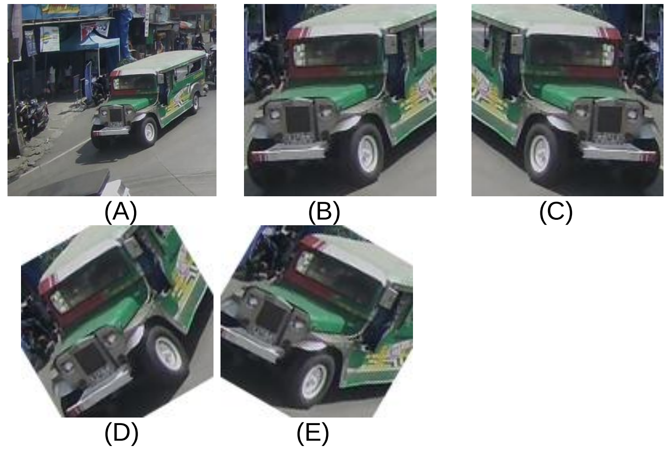

Figure 1*. The transformation of original data to augmented. (A) Original image (B) Cropped image (C) Cropped Image + flipped horizontal (D) cropped horizontal + rotate +30° (E) cropped image + rotate -30°.*

7 DATASET AND FEATURES

There are six sets of vehicle classification in the Philippines: bus, truck, van, car, jeep and tricycle. The images where extracted from a video file that has a 1280 x 720 pixels and is at 20 frames per second. The image will have 4 Region of Interest (ROI) that will be used to capture small vehicle from north bound (one is for large vehicles and one for small vehicles) and from the south bound (with the same features as north bound). Each ROI has 4498 of images, the images where cropped for each ROI which has a size of 120x120 pixels which produces 17,992 images. Data scrubbing was manually introduced so that the data set will be refined. The data was manually labeled and added to the dataset. The dataset will pass through a series of traditional augmentation, cropped + flipped, and cropped + rotation. The total data of 6000 because each category was limited to 1000 images and was multiplied by 4. The new data set is now 24,000 due to image augmentation.

Figure 2 *(left) Original Image with specific ROI (right) ROI A*

Figure 2(left) is the original image on frame number 4494 and figure 2 (right) is the region of interest (ROI) A. The region is allocated to capture large vehicles passing through north bound. Each bounding box represents different ROI.

8 DATA COLLECTION

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.

9 METHOD

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 data augmented dataset. From the selected ROI, the extraction of image was cropped. The cropped image was then saved to the dataset and re-augmented by flipping, rotating to +30° and -30°. All of the output image data was also saved to the dataset. The original image is 6000, after different augmentation and combination, the total images was 24,000. The increase number of image dataset, would improve the accuracy of the model and preventing the tendency of overfitting. Thus, some of the augmentation techniques are used for different purposes, such approach are more significant for image analization.

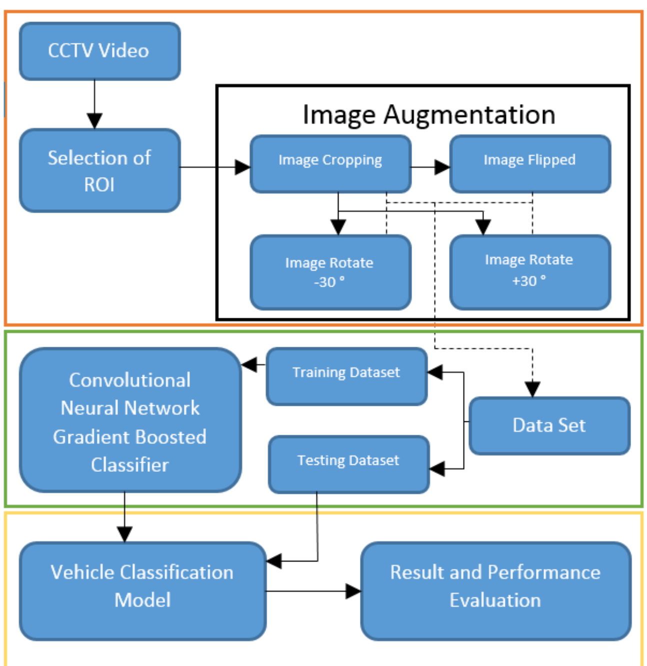

Figure 3. *Concept of the study*

Phase 2 includes the splitting of data set to testing dataset and training set. The study will use 20 % for the training and 80% for the testing and will be compared from the previous model to evaluate the accuracy and preventing the overfitting of the hybrid algorithm. Combining the convolutional neural net and the classifying principle of gradient boosted classifier, the paper assumes that the overfitting and the accuracy will be higher than the previous studies. Phase 3 is testing the proposed model using data augmentation. The comparison of original dataset and augmented data and evaluating the overfitting decrease in the introduced method. The proposed method used the Convolutional Neural Network and Gradient Boosted tree that is illustrated in figure 4. The process of neural network is to determine the classification of the vehicle that will be boosted by the gradient classifier. By applying the gradient boosted in the convolutional neural network, the last phase of the neural net which is the fully connected layer will be replaced by the injected gradient boosted classifier to forecast the labels from the dataset. Once the model is fully trained, the execution of the recognition and fresh decision on the testing images will have a high result. Using the image augmentation in convolutional neural net and gradient boosted classifier will solve the over fitting issue of the model.

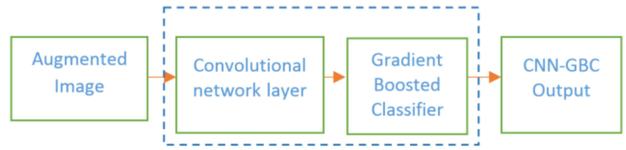

Figure 4. *Structure of the hybrid CNN-GBC method*

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.

9 RESULT AND ANALYSIS

Using the 80 percent as training resulting to 22,500 images in the dataset and 20 percent as testing, the evaluation before the data augmentation shows evidences of overfitting thus obtaining 0.9994 on the training for the highest accuracy computed and 0.9761for the testing. Having the same iteration shows that the loss has discrepancy having 0.0028 for the training and 0.2448 for the testing. The projection of the loss and the inconsistency is shown in the figure 5. Where the projected values clearly state the presence of overfitting.

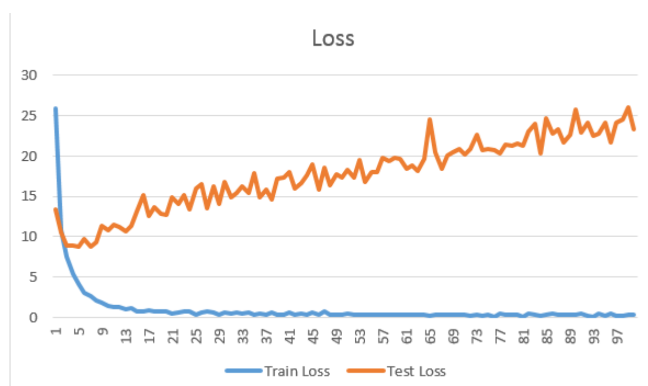

Figure 5. *Test of Loss with Overfitting*

The accuracy also showed difference in projected values showing evidences of overfitting that is shown in figure 6. The evaluation of the approach was tested with the same parameters and epochs of 100, the augmented dataset pass through the hybrid of CNN and GBC. The dataset contains 7 categories which are labeled as cars (sedan, hatchback, and others), motorcycle, van, trucks, bus, jeepney and tricycle.

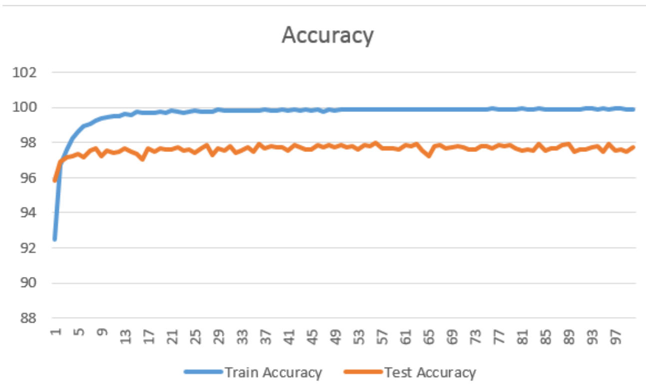

Figure 6. *Test of Accuracy with Overfitting*

After the testing of the new dataset using Anaconda Jupyter, figure 7 shows the graph of training loss and testing loss of the hybrid algorithm proving that overfitting was lessen but not eliminated.

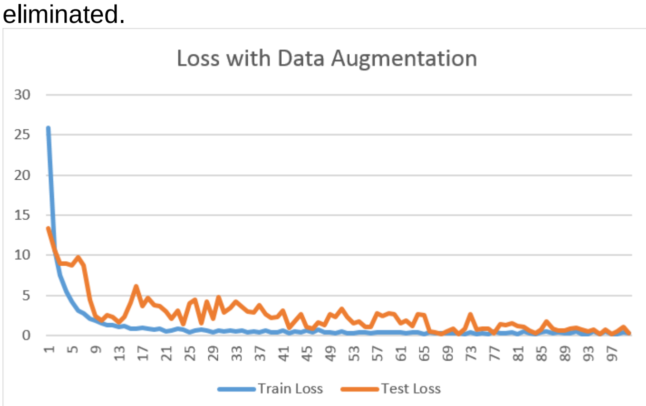

Figure 7. *Test of Loss with Data Augmented Images*

Comparing the values from the training loss value of 0.3 at epochs 100 and 0.25 for the testing loss shows that overfitting for loss has tremendously decrease.

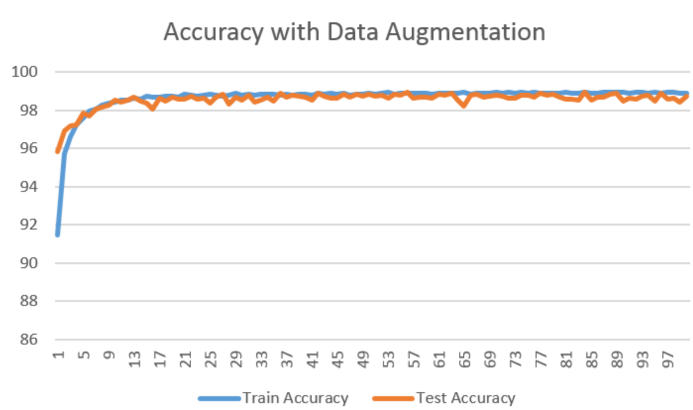

Figure 8. *Test of Accuracy with Data Augmented Images*

The measurement of accuracy of 98.91 at epoch 100 for the training and the accuracy of testing is 98.76 at the same epochs.

9 CONCLUSION AND FUTURE WORKS

The paper presented showed evidences of minimizing overfitting with the use of data augmentation in hybrid algorithms such as Convolutional Neural Network with Gradient Boosted Classifier. The collection of the new dataset and the augmented images had proven that the larger the dataset, the less overfitting may occur. The framework can be a reliable model for the traffic management of City of San Jose del Monte proving its consistency and accuracy. The future works may focus on larger dataset and using different pooling methods

ACKNOWLEDGMENT

The authors would like to thank Bulacan State University , Technological Institute of the Philippines, and Manuel L. Quezon University. For the Video file used for the Dataset, Civil Security Unit of the City of San Jose del Monte.

REFERENCES

[1] M. D. Hernandez, A. C. Fajardo, and R. P. Medina, “A Hybrid Convolutional Neural Network-Gradient Boosted Classifier for Vehicle Classification,” IJRTE J., no. 2, pp. 213–216, 2019.

[2] A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” 2018 Int. Interdiscip. PhD Work., pp. 117–122, 2018.

[3] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” J. Big Data, 2019.

[4] J. Wang and L. Perez, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning.”

[5] 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.

[6] L. Zhang, M. Song, Y. Yang, Q. Zhao, C. Zhao, and N. Sebe, “Weakly Supervised Photo Cropping,” no. c, pp. 1–13, 2013.

[7] 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.

[8] W. Cao, X. Wang, Z. Ming, and J. Gao, “A review on neural networks with random weights,” Neurocomputing, vol. 0, pp. 1–10, 2017.

[9] J. Lv, X. Shao, J. Huang, X. Zhou, and X. Zhou, “Data augmentation for face recognition,” Neurocomputing, vol. 230, no. December 2016, pp. 184–196, 2017.

[10] 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.

[11] L. Taylor and G. Nitschke, “Improving Deep Learning using Generic Data Augmentation,” 2015.

[12] G. Hinton, “Dropout : A Simple Way to Prevent Neural Networks from Overfitting,” J. Mach. Learn. Res. 15 1929-1958, vol. 15, pp. 1929–1958, 2014.

[13] S. Ioffe and C. Szegedy, “Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift,” 2015.

[14] K. Weiss, T. M. Khoshgoftaar, and D. Wang, A survey of transfer learning. Springer International Publishing, 2016.

[15] D. Erhan, A. Courville, and P. Vincent, “Why Does

Unsupervised Pre-training Help Deep Learning ?," J. Mach. Learn. Res. 11 625-660, vol. 11, pp. 625–660, 2010.
[16] M. Palatucci, G. Hinton, D. Pomerleau, and T. M. Mitchell, "Zero-Shot Learning with Semantic Output Codes," pp. 1–9.
[17] C. Gan and E. P. Xing, "Recurrent Topic-Transition GAN for Visual Paragraph Generation," 2017.