

Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network

Romi Fadillah Rahmat^{*1}, Dennis², Opim Salim Sitompul³,
Sarah Purnamawati⁴, Rahmat Budiarto⁵

^{1,2,3,4}Department of Information Technology, Universitas Sumatera Utara,
Padang Bulan, 20155, Medan, Indonesia

⁵College of Computer Science and Information Technology, Albaha University,
Albaha, Kingdom of Saudi Arabia

*Corresponding author, e-mail: romi.fadillah@usu.ac.id

Abstract

In this paper, we propose an approach to detect and geotag advertisement billboard in real-time condition. Our approach is using AlexNet's Deep Convolutional Neural Network (DCNN) as a pre-trained neural network with 1000 categories for image classification. To improve the performance of the pre-trained neural network, we retrain the network by adding more advertisement billboard images using inductive transfer learning approach. Then, we fine-tuned the output layer into advertisement billboard related categories. Furthermore, the detected advertisement billboard images will be geotagged by inserting Exif metadata into the image file. Experimental results show that the approach achieves 92.7% training accuracy for advertisement billboard detection, while for overall testing results it will give 71.86% testing accuracy.

Keywords: advertisement billboard detection, advertisement billboard geotagging, deep convolutional neural network, image classification, transfer learning

Copyright © 2019 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Advertisement billboard is an effective commercial media for advertising information about products or services. Currently, advertisement billboard data management has been done using conventional approach, e.g. by capturing advertisement billboard using camera and writing the location details in notes. This approach is very labor-intensive and time-consuming. Hence, a faster approach is needed to extract information in advertisement billboard and the geographic location of the object to improve the data acquisition.

The advertisement billboard detection problem has been approached with many different methods in recent years, mainly for detecting advertisement billboard in sport TV. Medioni et al. [1] use interest point operator, color-based point filter, point matcher, precise lock-in using Sum of Squared Differences (SSD), and predictor using Measure of Belief to detect and substitute billboard in broadcast video. The result shows good performance in detecting advertisement billboard. Cai et al. [2] detect advertisement billboard images in sport TV using Fast Hough Transform for line detection and histogram-based analysis for segmentation. The result shows that the approach achieves 90% accuracy in detecting advertisement billboard. Another approach by Aldershoff & Gevers [3] used histogram back-projection to detect advertisement billboard in soccer broadcast video. The result shows good performance in detecting advertisement billboard. Ichimura [4] detect advertisement billboard in motor sports video using Hessian-Laplace detector and Gradient Location-Orientation Histogram (GLOH) descriptor. In addition, RANdom SAMple Consensus (RANSAC) algorithm based on homography is used to recognize multiple advertisement billboards. Watve & Sural [5] detect advertisement billboard in soccer video using hue slicing and Hough Transform. The result shows that the approach achieves 90% accuracy in detecting advertisement billboard. Orginc [6] uses homography estimation using Direct Linear Transformation (DLT), RANdom SAMple Consensus (RANSAC), and Maximally Stable Extremal

Region (MSER) to detect advertisement billboard in broadcast video. The result shows that the approach achieves high accuracy. Ordelman [7] detects advertisement billboard using template matching with Fast Fourier Transform, color matching with Euclidean distance calculation in normalized HSL (NHSL) color spaces, and neighbor voting. The result shows 36% accuracy in detecting advertisement billboards.

Occlusion and environment complexity have been the difficulties in detecting advertisement billboard in previous works. Edge and color detection techniques can only be used in advertisement billboard detection unless the environment is not complex, assuming the background is plain, and the object is not occluded. Furthermore, the previous works depend on edge in detecting the shape of the advertisement billboard. When one of the edges is covered with occlusions, the detection will fail to detect advertisement billboard. In addition, color detection approach works if the advertisement billboard has specific color and has a contrast color to the background. To overcome this problem, a supervised machine learning approach is needed to detect advertisement billboard in more complex environment, e.g. roads with many obstacles such as tree, vehicle, cable, and pole.

Recently, the field of machine learning has made tremendous progress on addressing classification with complex objects. Researchers have found that using a pre-trained neural network model with large learning capacity, e.g. QuocNet, AlexNet, and GoogleNet could improve image classification accuracy. In addition, a Deep Convolutional Neural Network (DCNN) has been proven to solve image classification on hard visual recognition tasks [8].

For advertisement billboard acquisition, geotagging plays an important role in obtaining the geographical location of the detected advertisement billboard. Several approaches for geotagging object using smartphones has been done. Macias et al. [9] adds geotag information in video using angular values from smartphone GPS, 3G networks, and WiFi. The result shows the geotagging achieves high accuracy and consumes a little bandwidth. Sahu and Chakraborty [10] add geotag information by inserting Exchangeable image file (Exif) metadata into the image. The result shows that the geotagging achieves up to 30-meter accuracy from the maps. In addition to overcome the geotagging problem, Rahmat et al [11] improve the geotagging accuracy by using perspective projection. Overall, the geotagging process depends on network connection.

We have reviewed several related works that use Deep Convolutional Neural Network (DCNN) for large-scale image classification. Lin & Chen [12] detect pedestrians using GoogleNet Two Parallel DCNN. The results show 19.57% regression in pedestrian detection. Zhang et al. [13] classify makers and models of cars image using pre-trained DCNN with transfer learning. The results show 79% accuracy in classifying cars. Yanai & Kawano [14] use DCNN for food image classification. The results show 78.77% accuracy for Top-1 prediction using UEC-FOOD100 testing dataset and 67.57% accuracy for Top-1 prediction using UEC-FOOD256 testing dataset. Yan et al. [15] classify objects into 1000 categories from ILSVRC 2012 dataset using Hierarchical DCNN (HD-CNN). The results show 36.66% error-rate for Top-1 prediction and 15.80% for Top-5 prediction. Pasquale et al. [16] identify 50 objects from iCubWorld dataset using DCNN. The results show 86% for identifying 50 objects. Jung et al. [17] recognize traffic signs using Le Net-5 CNN. The results show that the methods can classify 16 street signs. Ouyang et al. [18] recognize common objects using Deformable Deep Convolutional Neural Networks (DeeplD-Net). The classification results in 50.3% accuracy. Li et al. [19] use GoogleNet Inception with 22 convolutional layers for classifying 1000 objects. The classification results show 89.45% for Top-1 prediction. Martinson & Yalla [20] use Krizhevsky's DCNN with 5 convolutional layers, 3 pooling layers, and 3 fully-connected layers for classifying 1000 objects. The results show 90.1% under structured light sensor in open lab space, 86.7% under structured light sensor in home environment, 80.4% under stereo camera in an office environment, and 73.9% under time-of-flight camera in a home environment.

In this paper, we propose a method to detect advertisement billboard by using Deep Convolutional Neural Network (DCNN). First, a pre-trained neural network architecture from AlexNet is used for image classification. To improve the advertisement billboard detection performance, we retrain the network using transfer learning approach. Several other informations such as billboard categories and names also propagated to the system. In addition, for the geotagging process, an Exif metadata insertion is used to store geographical location of the detected advertisement billboard.

2. Methodology

In this research, the method for detecting and do geotagging actions for billboard image by using smartphone can be described in several phases. The phases can be described as follows. Image Pre-Training process which will be combined with inputted image and retraining process using transfer learning method. After this process finished and the system filled with the trained DCNN, then it can be used to identify a new image taken by Android device in real-time mode. Thus, Fine Tuning Process will categorize six classes, they are scoreboard, monitor, television, cinema, and web site into one class named as billboard. The next phase is feature extraction which can detect billboard image frame by frame in DCNN, then automatic image storage process to the database with minimum accuracy, then we apply geotagging process with the metadata steganography process in Exif file. In the postprocessing phase, we will add more important informations of the billboard image and do validation process. All these processes are described and shown in Figure 1.

2.1. DCNN Training with ILSVRC 2012

First, we use Inception-v3 as a pre-trained model for image classification with 1.2 million dataset images and 1000 different classes from ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. This model uses AlexNet's DCNN architecture with 60 million parameters and 650.000 neurons, consists of five convolutional layers, three pooling layers, and three fully-connected layers. The output of the model is 1000-way softmax resulting in five top predictions of the object. The network contains eight layers with weights; the first five are convolutional and the remaining three are fully-connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253, 440-186, 624-64, 896-64, 896-43, 264-4096-4096-1000.

Table 1 shows the DCNN process in every layer. The first convolutional layer filters the 224x224x3 input image with 96 kernels of size 11x11x3 with a stride of 4 pixels. The output of the first convolutional layer will be used as an input to the second convolutional layer with 256 kernels of size 5x5x48. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size 3x3x256 connected to the outputs of the second convolutional layer. The fourth and fifth convolutional layer has 384 kernels of size 3x3x192. The fully-connected layers have 4096 neurons each.

Table 1. DCNN Layer Details

Layer	Size	Kernel	Neuron
Input: (Resized Image)	224 x 224 x 3 x 1	-	150528
Layer 1: Convolution + Max Pool	55 x 55 x 48 x 2	96	290400
Layer 2: Convolution + Max Pool	27 x 27 x 128 x 2	256	186624
Layer 3: Convolution	13 x 13 x 192 x 2	384	64896
Layer 4: Convolution	13 x 13 x 192 x 2	384	64896
Layer 5: Convolution + Max Pool	13 x 13 x 128 x 2	256	43264
Layer 6: Fully-Connected	2048 x 2	-	4096
Layer 7: Fully-Connected	2048 x 2	-	4096
Layer 8: Softmax Output	1000 x 1	-	1000

2.2. DCNN with Transfer Learning

To improve DCNN features, we retrain the network by adding more advertisement billboard images into the pre-trained model using inductive transfer learning approach. Transfer learning is a process which will modify a value from previously trained dataset, with additional data propagated to the DCNN. The first process is by inputting 300 additional billboard image taken from Google image randomly, and 84 billboard images taken from Android smartphone in realtime mode to the Inception-v3 model and do DCNN Retraining Process. Next process involves a bottleneck process where all the images used for training will be analyzed and calculated. The bottleneck values contain a meaningful and compact summary of the images. By using transfer learning, we can train dataset faster than training from scratch. After we finish the bottleneck process then we start using 8000 training steps in the DCNN which resulted 5 highest predicted objects with the accuracy itself.

Machine Learning: Deep Convolutional Neural Network (DCNN)

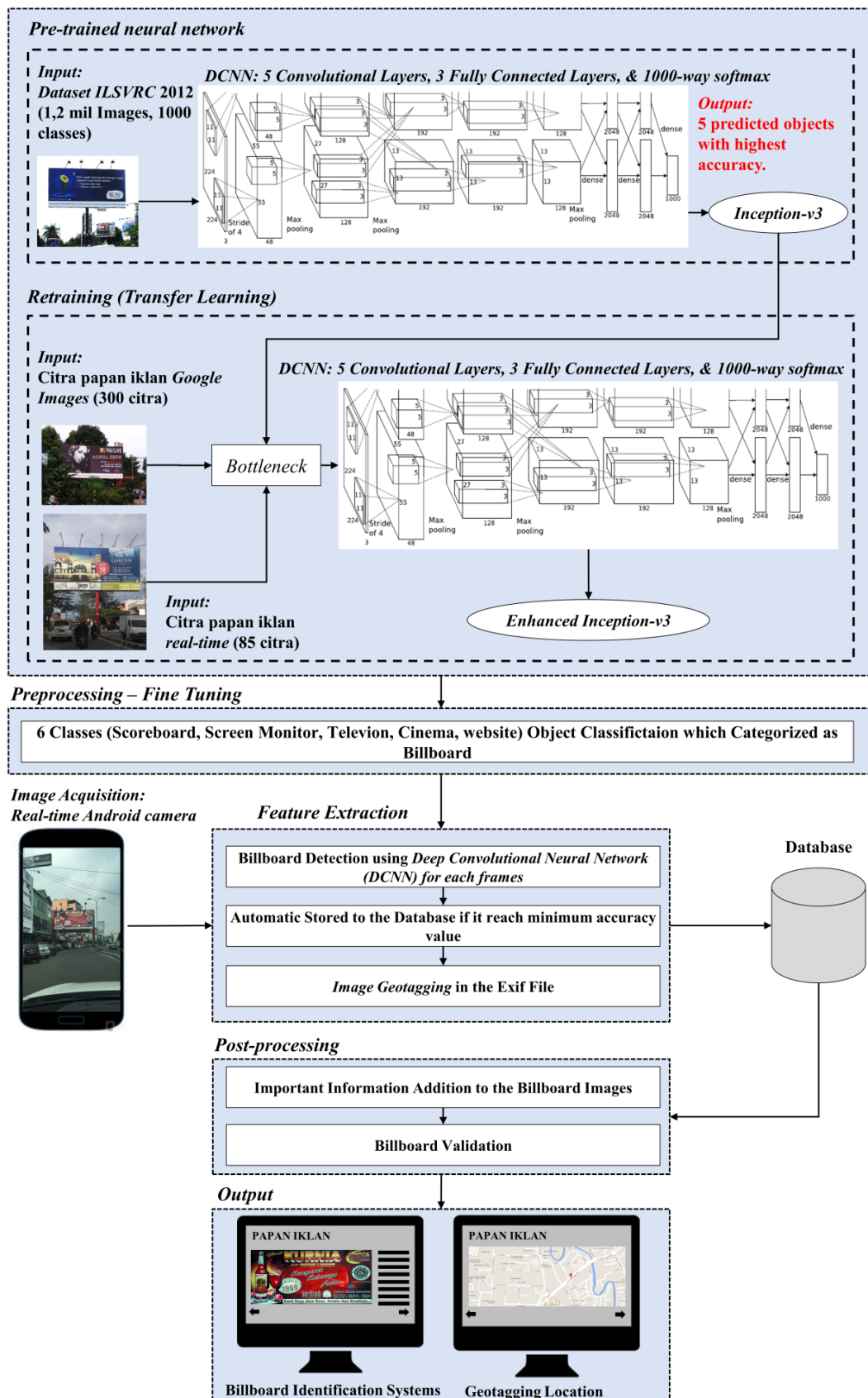


Figure 1. Advertisement billboard detection and geotagging system architecture

2.3. Fine-tuning

We fine-tuned the DCNN with 1000 categories using transfer learning approach into advertisement billboard related categories, such as scoreboard, screen, monitor, television, and cinema. Figure 2 shows an example of fine-tuning process showing top five object predictions. The highest advertisement billboard related prediction will be selected as the final output. There are 2 related categories related to advertisement billboard, which are scoreboard with 72.6% accuracy and cinema with 0.548% accuracy. Since the scoreboard accuracy has higher accuracy, the accuracy will be chosen as the final accuracy for the detected object.

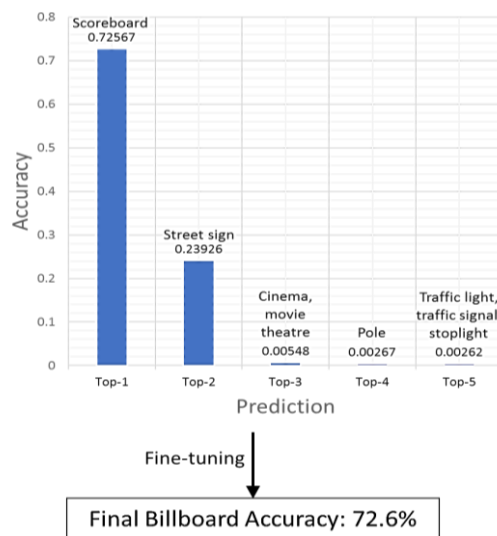


Figure 2. Fine-tuning example

2.4. Real-time Advertisement Billboard Detection

Advertisement billboard detection is conducted using Android smartphone in real-time road condition based on frame by frame images. The detected billboard image will be automatically geotagged and saved into the database.

2.5. Geotagging

We use Android smartphone GPS and 4G networks to obtain geotag information. Then, Exchangeable Image File (Exif) metadata is used to store geographical location, e.g. latitude and longitude value of the detected advertisement billboard. Furthermore, the address of the advertisement billboard image can be extracted using reverse geocoding. Reverse geocoding represents a location in more readable forms, e.g. street name, place name, county, state, country, and postal code.

2.6. Post-processing

After saving image into database, we add additional information to advertisement billboard data, such as advertisement billboard name and category. Advertisement billboard is divided into several categories, e.g. food and beverage, telecommunication, bank, insurance, transportation, real estate, education, cigarette, event, campaign, promotion, home products, electronic products, media. These categories is useful for commercial purpose. In addition for advertisement billboard verification, we need to ensure if the advertisement billboard collected is not expired yet. Hence, a routine update is necessary to check whether the advertisement billboard is still active or expired.

3. Training Results

During training process, each step chooses ten images at random from the training set, find their bottleneck values, and use them into the final layer to get predictions. Those

predictions are then compared against the actual labels to update the final layer's weights through the back-propagation process.

The training is conducted in two experiments, such as experiment with 4000 training steps and 8000 training steps. Table 2 shows several parameters used for training with 4000 and 8000 training steps. The results of the training process show that the experiment with 4000 training steps result in 89.3% accuracy, whereas the experiment with 8000 training steps result in 92.7% accuracy. These show that more training steps result in higher accuracy. Since the training accuracy is higher, we will use the model with 8000 training steps for the detection process. Table 3 shows the advertisement billboard images training results with 8000 training steps. Overall, the training results achieve 92.7% accuracy.

Table 2. Parameter used for Training Dataset

Parameter	Experiment 1	Experiment 2
Training steps	4000	8000
Learning rate	0.01	0.01
Testing percentage	10%	10%
Validation percentage	10%	10%
Eval step interval	10	10
Train batch size	100	100
Validation batch size	100	100

Table 3. Advertisement Billboard Training Results with 8000 Training Steps Summary

Training Steps	Training Accuracy (%)	Cross Entropy	Validation Accuracy (%)
0	60.0	0.644521	56.0
500	98.0	0.085934	71.0
1000	100.0	0.050542	76.0
1500	100.0	0.025906	74.0
2000	100.0	0.026487	79.0
2500	100.0	0.017426	85.0
3000	100.0	0.014568	85.0
3500	100.0	0.016349	81.0
4000	100.0	0.013454	86.0
4500	100.0	0.007575	83.0
5000	100.0	0.008694	82.0
5500	100.0	0.007368	84.0
6000	100.0	0.008262	84.0
6500	100.0	0.006660	87.0
7000	100.0	0.006335	81.0
7500	100.0	0.008083	83.0
7999	100.0	0.005449	75.0
Final Accuracy			92.7%

4. Testing Result

The detection process is conducted during a day and in real-time condition. Table 4 shows that the testing results achieve 72.0%. We evaluate several problems affecting the accuracy of DCNN in detecting advertisement billboard. First, given the testing results in Table 3, we conclude that brightness affects the detection accuracy. Another factors also affect the detection process, which are occlusions, e.g. noise, electrical cable, tree, pole, street sign, and other objects that cover the advertisement billboard area. In addition, shooting distance affects the advertisement billboard accuracy.

The nearer the shooting distance to the object, the higher the accuracy achieved. Figure 3 shows two advertisement billboard images with two different shooting distances in which the far shooting distance results in 28.1% accuracy and the near shooting distance results in 87.3%. This shows that the accuracy is higher when the object is nearer. Shooting angle also affects detection accuracy. Figure 4 shows two advertisement billboard images with two different shooting angles in which shooting from front angle achieves 86.7% accuracy and the shooting from right angle achieves 66.1%. This shows that the accuracy is higher when the object is detected in front angle.

In the future, several phase will be added into the system. We will train the advertisement billboard images with deeper DCNN architecture, e.g. GoogleNet's DCNN

with 22 hidden layers [21] or with better machine learning approach, such as Deep Residual Network (DRN) [22]. In addition, we will add more advertisement billboard images at night time to improve the detection at night. Furthermore, we will add optical character recognition (OCR) to extract text in advertisement billboard automatically [23]. Moreover, we will use estimation algorithm to measure the distance and size of the advertisement billboard to improve the accuracy of geotagging process [24, 25].

Table 4. Advertisement Billboard Testing Results During Day

Advertisement Billboard Image	Accuracy (%)	Advertisement Billboard Image	Accuracy (%)
1	93.9	16	44.5
2	76.5	17	50.2
3	87.3	18	67.5
4	70.1	19	74.6
5	32.2	20	49.7
6	54.7	21	59.1
7	71.9	22	80.8
8	86.7	23	81.9
9	56.2	24	88.3
10	85.6	25	87.9
11	97.1	26	90.5
12	76.7	27	65.1
13	65.1	28	61.5
14	90.3	29	72.9
15	69.0	30	68.0
Average Testing Accuracy		71.86%	

Shooting distance: far



Accuracy: 28.1%

Shooting distance: near



Accuracy: 87.3%

Shooting angle: front



Accuracy: 86.7%

Shooting angle: right



Accuracy: 66.1%

Figure 3. Two advertisement billboard images with different shooting distances

Figure 4. Two advertisement billboard images with different shooting angles

4. Conclusion

In this paper, we present Deep Convolutional Neural Network (DCNN) to detect advertisement billboard. The DCNN training results achieve 92.7% accuracy. In addition, the DCNN real-time testing results achieve 71.86% accuracy during day and 16.98% at night. This shows that the detection performs better during day time. For geotagging process, most of the saved images contain its geotagging location. However, the images are taken from a car, thus will have 1-10 metres difference from exact location. It means the real accuracy is degraded because of real situation in the field. The optimal range of the system achieves high accuracy under 4G networks.

References

- [1] G Medioni, G Guy, H Rom, A François. *Real-Time Billboard Substitution in a Video Stream*. Proceedings of the 10th Tyrrhenian International Workshop on Digital Communications. 1998: 1-14.

- [2] G Cai, L Chen, J Li. *Billboard advertising detection in sport TV*. Proceedings of the 7th International Symposium on Signal Processing and Its Applications. 2003: 537-540.
- [3] F Aldershoff, T Gevers. *Visual Tracking and Localisation of Billboards in Streamed Soccer Matches*. Storage and Retrieval Methods and Applications for Multimedia. 2004; 5307: 408-416.
- [4] N Ichimura. *Recognizing multiple billboard advertisements in videos*. Pacific-Rim Symposium on Image and Video Technology (PSIVT). 2006; 1(4319): 463-473.
- [5] AK Watve, S Sural. *Detection of on-field advertisement billboards from soccer telecasts*. IET International Conference on Visual Information Engineering (ICVIE). 2006: 12-17.
- [6] B Orginc. *A vision-based system for detecting and tracking of advertisement billboards*. Thesis. Ljubljana: University of Ljubljana. 2009.
- [7] J H S Ordelman. *Automatic Billboard Detection*. Thesis. Twent: University of Twente; 2010.
- [8] A Krizhevsky, I Sutskever, G E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. Advances in Neural Information Processing System (NIPS). 2012; 12(25): 1097-1105.
- [9] E Macias, H Abdelfatah, A Suarez, A Canovas. *Full Geo-localized Mobile Video in Android Mobile Telephone*. *Macrothink Institute Academic Journal of Network Protocols and Algorithms*. 2011; 3(1): 64-81.
- [10] I Sahu, I Chakraborty. *Understanding Location Manager in Android and Implementing an Optimal Image Geotagging Application*. *International Journal of Computer Trends and Technology (IJCTT)*. 2013; 4(6): 1682-1686.
- [11] RF Rahmat, S Purnamawati, H Saito, Ichwan MF, Lubis TM. *Android-based automatic detection and measurement system of highway billboard for tax calculation in Indonesia*. *Indonesian Journal of Electrical Engineering and Computer Science*. 2019; 14(2): 877-886.
- [12] BY Lin, CS Chen. *Two Parallel Deep Convolutional Neural Networks for Pedestrian Detection*. International Conference on Image and Vision Computing New Zealand (IVCNZ). 2015: 1-6.
- [13] F Zhang, X Xu, Y Qiao. *Deep Classification of Vehicle Makers and Models: The Effectiveness of Pre-training and Data Enhancement*. IEEE International Conference on Robotics and Biomimetics (ROBIO). 2015: 231-236.
- [14] K Yanai, Y Kawano. *Food image recognition using deep convolutional neural network with pre-training and fine-tuning*. IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2015: 1-6.
- [15] Z Yan, H Zhang, R Piramuthu, V Jagadeesh, D Decoste, W Di, Y Yu. *HD-CNN: Hierarchical Deep Convolutional Neural Networks for Large Scale Visual Recognition*. IEEE International Conference on Computer Vision (ICCV). 2015: 2740-2748.
- [16] G Pasquale, C Ciliberto, L Rosasco, L Natale. *Object identification from few examples by improving the invariance of a Deep Convolutional Neural Network*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2016: 4904-4911.
- [17] S Jung, U Lee, J Jung, DH Shim. *Real-Time Traffic Sign Recognition System with Deep Convolutional Neural Network*. International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). 2016; 13(1): 31-34.
- [18] W Ouyang, X Zeng, X Wang, S Qiu, P Luo, Y Tian, H Li, S Yang, Z Wang, H Li, CC Loy, K Wang, J Yan, X Tang. *DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection*. IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2403-2412.
- [19] H Li, P Su, Z Chi, J Wang. *Image Retrieval and Classification on Deep Convolutional SparkNet*. IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). 2016: 1-6.
- [20] E Martinson, V Yalla. *Augmenting Deep Convolutional Neural Networks with Depth-Based Layered Detection for Human Detection*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2016: 1073-1078.
- [21] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, A Rabinovich. *Going Deeper with Convolutions*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
- [22] K He, X Zhang, S Ren, J Sun. *Deep Residual Learning for Image Recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 770-778.
- [23] K Hosozawa, R H Wijaya, T D Linh, H Seya, M Arai, T Maekawa, K Mizutani. *Recognition of Expiration Dates Written on Food Packages with Open Source OCR*. *International Journal of Computer Theory and Engineering*. 2018; 10(5): 170-174.
- [24] L Bazzani, A Bergamo, D Anguelov, L Torresani. *Self-Taught Object Localization with Deep Networks*. IEEE Winter Conference on Applications of Computer Vision (WACV). 2016: 1-9.
- [25] S Laotrakunchai, A Wongkaew, K Patanukhom. *Measurement of Size and Distance of Objects Using Mobile Devices*. International Conference on Signal-Image Technology & Internet-Based System. 2013: 156-161.