IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 6, 2017, accepted December 3, 2017, date of publication December 11, 2017,
date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2782260

Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network

HILAL TAYARA[1], KIM GIL SOO[2], AND KIL TO CHONG[1]
[1]Department of Information and Electronics Engineering, Chonbuk National University, Jeonju 54896, South Korea
[2]Institute of International Studies, Chonbuk National University, Jeonju 54896, South Korea

Corresponding author: Kil To Chong (kitchong@jbnu.ac.kr)

This work was supported in part by BK21 PLUS, in part by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning under Grant NRF-2017M3C7A1044815, and in part by the Ministry of Trade, Industry and Energy and the Korea Institute for Advancement of Technology through the International Cooperative Research and Development Program under Grant N046200012.

ABSTRACT Vehicle detection and counting in aerial images have become an interesting research focus since the last decade. It is important for a wide range of applications, such as urban planning and traffic management. However, this task is a challenging one due to the small size of the vehicles, their different types and orientations, and similarity in their visual appearance, and some other objects, such as air conditioning units on buildings, trash bins, and road marks. Many methods have been introduced in the literature for solving this problem. These methods are either based on shallow learning or deep learning approaches. However, these methods suffer from relatively low precision and recall rate. This paper introduces an automated vehicle detection and counting system in aerial images. The proposed system utilizes convolution neural network to regress a vehicle spatial density map across the aerial image. It has been evaluated on two publicly available data sets, namely, Munich and Overhead Imagery Research Data Set. The experimental results show that our proposed system is efficient and effective, and produces higher precision and recall rate than the comparative methods.

INDEX TERMS Aerial images, convolution neural network (CNN), deep learning, regression, vehicle detection.

I. INTRODUCTION

Vehicle detection and counting is important for many applications such as surveillance, traffic management, and rescue tasks. The ability of on-line monitoring of vehicles distribution in the urban environments prevents traffic jams and congestions which in turn reduces air and noise pollution. In terms of surveillance, the accurate estimation of vehicles in parking lots or roads is essential for making right decisions. Therefore, this problem has attracted the attention of the researchers in the recent years. However, vehicle detection and counting is a challenging task due to many reasons such as: small size of the vehicles, different types and orientations, similarity in visual appearance of vehicles and some other objects (e.g., air conditioning units on the buildings, trash bins, and road marks), and detection time in very high resolution images is another challenge that researchers need to take in consideration. Fig. 1 shows vehicles in aerial images and some of the aforementioned challenges.

The solution for this task can be categorized into two groups namely fixed-ground sensors and image-based sensors. In fixed-ground sensors, traffic information and vehicle monitoring are collected efficiently using different types of fixed ground sensors such as stationary camera, radar sensors, bridge sensors, and induction loop [1], [2]. These sensors give a partial overview about vehicles density, parking lots situation, and traffic flow. However, the overall information of traffic situation will not be available, which is important for road network monitoring and planning, traffic statistics, and optimization. On the other hand, image-based sensors come from two sources: satellites and airplanes or unmanned aerial vehicles (UAV). Image-based sensors give an overall overview of traffic situation in the area of interest. This is the reason for adapting this type of sensors widely for monitoring vehicles [1], [3], [4]. Satellites provide images with submeter spatial resolution. Therefore, satellite images have been used for monitoring vehicles by many researches [2], [5], [6].

FIGURE 1. Examples of some of vehicle detection challenges (small vehicle sizes, different types and orientations, other similar objects such as air conditioning units on building, trash bins, and road marks) are shown by red ellipses.

On the other hand, aerial images captured by airplane or UAV provide a higher spatial resolution of 0.1 to 0.5m [7], [8] compared to satellite images which make them more preferable in solving vehicle detection and counting task. In addition, data acquisition is easier in UAV aerial images [9]. Thus, vehicle detection task became attainable due to high spatial resolution provided by aerial images. Nowadays, UAV are increasingly used in capturing images for vehicle detection task. The benefits of using UAV include low cost, fast acquisition of images, and environment-friendliness. The proposed algorithms for vehicle detection in the literature can be categorized into two groups: shallow-learning-based methods and deep-learning-based methods. In shallow-learning-based methods, hand-crafted features are engineered and followed by a classifier or cascade of classifiers [1], [10], [11]. However, shallow-learning-based methods do not give the desired accuracy in vehicle detection task and, recently, have been outperformed by deep learning architectures such as convolution neural network (CNN). On the other hand, deep learning-based methods have been used for vehicle detection task because of their outstanding performance in different domains such as images and sounds. More specifically, region based convolution neural network (RCNN) methods achieved outstanding performance in object detection tasks [12]. Faster RCNN [13] utilizes a fully convolution regional proposal network to generate region candidates which will be inferred by a classifier attached to region proposal network (RPN). Region proposal based networks perform better than shallow learning due to the following reasons i) CNN improves the performance because of the automatic features generation which is more powerful than hand-crafted features ii) region-based CNN model is less time consuming compared with sliding-window-based models because it reduces search space by examining hundreds of proposed objects rather than searching the whole image. Even though, region based CNNs have performed well in natural scene images, their performance in aerial images is limited due to the small size and different orientations of the vehicle, complex background in aerial images, and difficulties in fast detection due to the large size of the aerial images.

We solve the problem of vehicle detection and counting as a supervised learning problem. We try to learn a mapping function between an image $I(x)$ and a density map $D(x)$, denoted as $F : I(x) \rightarrow D(x)$ where $I \in R^{m \times n}$, and $D \in R^{m \times n}$ as shown in Fig. 2. We solve the mapping problem by utilizing the convolutional neural network (CNN) [14], [15]. A fully convolutional regression network (FCRN) has been proposed and evaluated on two publicly available datasets namely DLR Munich vehicle dataset provided by Remote Sensing Technology Institute of the German Aerospace Center [12] and Overhead Imagery Research Data Set (OIRDS) dataset [16]. The results of the proposed system have been compared with the state-of-the-art results and outperformed them.

The rest of the paper is organized as follows: Section II lists the related works published recently in the literature. Section III describes the proposed system. Section IV introduces datasets, evaluation procedures, and experimental results. This paper is concluded in Section V.

II. RELATED WORKS

A lot of researches have been carried out on vehicle detection and counting in aerial images over the years. These works can be categorized into two main groups i.e. shallow-learning-based methods and deep-learning-based methods. In this section, we briefly introduce the latest works carried out in these two groups.

A. SHALLOW-LEARNING-BASED METHODS

The general strategy followed in this group relies on hand-crafted features extraction followed by a classifier or cascade of classifiers. Moranduzzo and Melgani [9] proposed a system for car counting in aerial images captured by UAV. They have reduced search space by selecting the regions where cars might exist using a supervised classifier then extracted feature points using scale invariant feature transform (SIFT) [17] . Then support vector machine (SVM) has been used in order to discriminate between the cars and all other objects. Four steps for car detection system have been introduced in [18]. The proposed system in [18] starts with selecting the areas that might have cars. Then, two sets of histogram of oriented gradients (HOG) features are extracted for vertical and horizontal filtering directions. The discrimination between the cars and other objects has been performed by one of three suggested techniques: mutual information measure, normalized cross correlation, and combination of the correlation measure with SVM classification. The discrimination is obtained by associating an orientation value to the points classified as cars. Finally, the points that belong to the same car are merged. Fast vehicle orientation and type detection has been introduced in [12]. The proposed system has two stages. The first stage utilizes fast binary sliding window object detector

FIGURE 2. Training procedure tries to find a mapping function that maps the input image $I(x)$ to the density map $D(x)$. (left) Represents the input image $I(x)$ whereas (right) is the density map. Each vehicle is represented by 2-D Gaussian function.

which returns bounding boxes of vehicles without orientation or type information. The second stage applies multi-class classifiers on bounding boxes in order to decide the type and the orientation of the vehicles. Moranduzzo *et al.* [19] have introduced a fast and general object detection in which non-linear filters are used to combine image gradient features at different orders which produces features vector of very high dimension. Therefore, features reduction process has been utilized. The reduced features vector has been used by Gaussian process regression (GPR) model in order to decide the presence of the target object. Finally, they have used an empirical threshold value for the final decision. Super-pixel segmentation method has been introduced for vehicle detection task in [20]. After segmentation, patches located at the center of the super-pixel have been extracted for training and detection sets; then sparse representation dictionary has been created from training set. Therefore, selected training subset enables high discriminative ability for vehicle detection. Another method has been proposed by [21] in which it integrates linear SVM classifier and Viola-Jones with HOG features. Firstly, roads and on-road vehicles are aligned to vehicle detectors by adopting a roadway orientation adjustment method. Then, HOG and SVM or V-J methods can be applied to achieve higher accuracy. Local and global information of the vehicles in high resolution images have been studied together in order to increase the accuracy of vehicle detection [22]. Two detectors have been utilized for front wind shield samples and whole vehicle samples. The combined outputs have resulted in improving the accuracy of the system. Geometric constraint function has been utilized via improved entropy rate clustering (IERC) in order to obtain more homogeneous, regular, balanced, and compact super-pixels [23]. The resulted super-pixels are considered as the seeds for training sample selection. In addition, correlation-based sequential dictionary learning (CSDL) has been constructed for fast sequential training and updating of the dictionary.

B. DEEP-LEARNING-BASED METHODS

Most of the works proposed in this category use convolution neural network for automatic features extraction. In [24], deep convolutional neural network with multi-scale spatial pyramid pooling (SPP) has been employed in extracting the target patterns with different sizes. However, input images have been pre-processed by maximum normed gradient algorithm in order to restore the edges of the objects. Another deep learning approach has been introduced by [25]. In this work, the input image has been segmented into small homogeneous regions. Then the features in the segmented regions are extracted using pre-trained convolutional neural network (CNN) by a sliding-window approach. Windows are classified using support vector machine (SVM) into car and no-car classes. Finally, post-processing is done such as morphological dilation to smooth the detected regions and fill the holes. In addition, the number of the cars detected has been determined by the estimation of the detected regions. Hyper feature map that combines hierarchical feature maps have been used in an accurate vehicle proposal network (AVPN) in [26]. Vehicle location and attributes have been extracted by the proposed coupled regional convolutional network method which merges an AVPN and a vehicle attribute learning network. Fast and Faster R-CNN have been explored in [27]. In order to overcome the limitations in Fast and Faster R-CNN, a new architecture has been proposed. They have

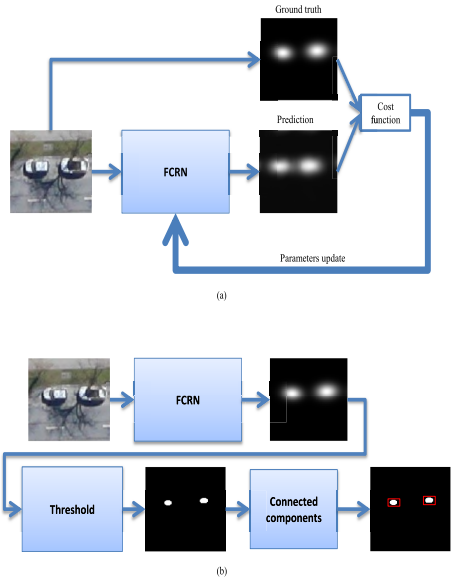

FIGURE 3. The proposed system. (a) Training phase. (b) Inference Phase.

improved the detection accuracy of the small-sized objects by using the resolution of the output of the last convolutional layer and adapting anchor boxes of RPN as feature map. Another improved detection method based on Faster R-CNN has been introduced by [28]. They have solved the limitations of Faster R-CNN by a proposing hyper region proposal network (HRPN) that extracts vehicle targets with a hierarchical feature maps. In addition, a cascade of boosted classifier has been used to classify the extracted regions.

III. THE PROPOSED SYSTEM

In this section, we introduce the architecture of the proposed system, ground truth preparation, and implementation details.

A. FULLY CONVOLUTIONAL REGRESSION NETWORK (FCRN)

We propose to solve vehicle detection and counting problem by a fully convolutional regression network (FCRN). Fig 3 illustrates the proposed system. During training, an input image and its corresponding ground truth are given to the FCRN where the goal is to minimize the error between the ground truth and predicted output. During inference, the output of the trained model goes under an empirical thresholding after which a simple connected component algorithm is used for returning the count and the location of the detected vehicles.

In the proposed architecture, we build an auto-encoder-like network as shown in Fig. 4. FCRN has two paths: down-sampling path and up-sampling path. The down-sampling path is the pre-trained VGG-16 network [29]. This path consists of repeated padded 3×3 convolutions followed by rectified linear unit (ReLU) and a max pooling operation. VGG-16 network has been trained on ImageNet large scale visual recognition challenge (ILSVRC) dataset [30]. We use

TABLE 1. Detailed architecture of the proposed system (FCRN).

Name	Configuration
conv1	[Conv $3 \times 3 \times 64$ - ReLU] $\times 2$ Max-pooling
conv2	[Conv $3 \times 3 \times 128$ - ReLU] $\times 2$ Max-pooling
conv3	[Conv $3 \times 3 \times 256$ - ReLU] $\times 3$ Max-pooling
conv4	[Conv $3 \times 3 \times 512$ - ReLU] $\times 3$ Max-pooling
conv5	[Conv $3 \times 3 \times 512$ - ReLU] $\times 3$ Up-Sampling
D1	[Conv $3 \times 3 \times 256$ - Batch normalization - ReLU] $\times 2$
D2	[Conv $3 \times 3 \times 128$ - Batch normalization - ReLU] $\times 2$
D3	[Conv $3 \times 3 \times 64$ - Batch normalization - ReLU] $\times 2$
deconv1	Concatenate [D1_output, Cov5_output] [Conv $3 \times 3 \times 256$ - Batch normalization - ReLU] $\times 2$ Up-Sampling
deconv2	Concatenate [D2_output, deconv1_output] [Conv $3 \times 3 \times 256$ - Batch normalization - ReLU] $\times 2$ Up-Sampling
deconv3	Concatenate [D3_output, deconv2_output] [Conv $3 \times 3 \times 256$ - Batch normalization - ReLU] $\times 2$ Up-Sampling
deconv4	[Conv $3 \times 3 \times 256$ - Batch normalization - ReLU] $\times 2$
deconv5	[Conv $1 \times 1 \times 1$ - linear activation]

the layers up to 'conv5' from VGG-16 network. Removing the remaining layers reduces the number of the parameters significantly. In the up-sampling path, the symmetry has been broken because of asymmetric nature of image regression problem (the input is an image and the output is a density map). We have used a skip connections in order to merge fine, shallow, appearance information and coarse, deep, semantic information. Therefore, accurate vehicles detection and localization has been achieved. The detailed architecture is given in Table 1.

B. GROUND-TRUTH PREPARATION

A two-dimensional elliptical Gaussian function has been utilized for generating the ground-truth from the dataset as depicted in the general equation

$$f(x, y) = Ae^{(-(a(x-x_0)^2+2b(x-x_0)(y-y_0)+c(y-y_0)^2))} \quad (1)$$

where $a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2}$, $b = \frac{-\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}$, and $c = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2}$ are the elements of the positive-definite matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ and used for generating rotated ground-truth. We set $A = 1$, σ_x, and σ_y are inferred from the width and height of the vehicle, and θ is the orientation of the vehicle. Width, height, and orientation of the vehicle are taken from the bounding box ground-truth annotation provided by the dataset. Fig. 5 shows an example of a generated ground-truth. Fig. 5(a) shows

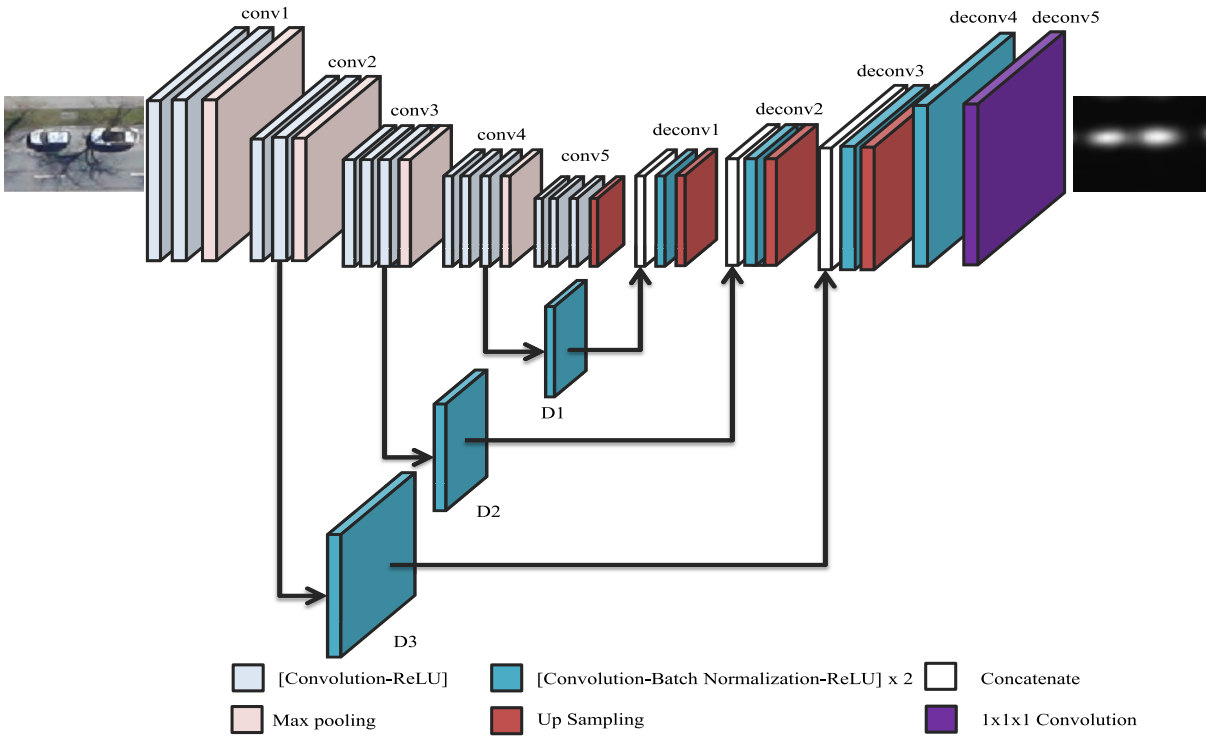

FIGURE 4. The architecture of the proposed fully convolutional regression network (FCRN).

an input image with the yellow bounding boxes ground-truth annotation, Fig. 5(b) shows the generated ground-truth, and Fig. 5(c) shows a 2D visualization of the generated ground-truth.

C. IMPLEMENTATION DETAILS

The implementation of the proposed architecture is based on Tensorflow [31]. During training phase, 224x224 random patches were selected from the aerial image. The selected patch contains at least one vehicle. Thus, patches with no vehicles were not chosen during training. In order to increase the amount of training examples, data augmentation techniques were utilized such as rotation, horizontal and vertical flipping and shifting. The mean square error target function

$$I(\Phi, X) = \frac{1}{M}\sum_{i=1}^{M}(Y_T - Y_P)^T(Y_T - Y_P) \qquad (2)$$

is used. In (2), X is the input patch with M samples, Φ are all trainable parameters, Y_P is the predicted density map, and Y_T is the ground truth annotation. RMSprop optimizer has been used for updating the parameters values [32]. Parameters in the down-sampling path have been initialized with the parameters of VGG-16 networks and fixed during training. However, the parameters of the up-sampling path and skip connections have been initialized using "He" initialization method [33] and updated during training. Gaussian annotation of the ground-truth is scaled to 255 in order to make training easier. Initial learning rate is 0.01 and decayed exponentially. The number of epochs is set to 200. The empirical threshold is set to 120.

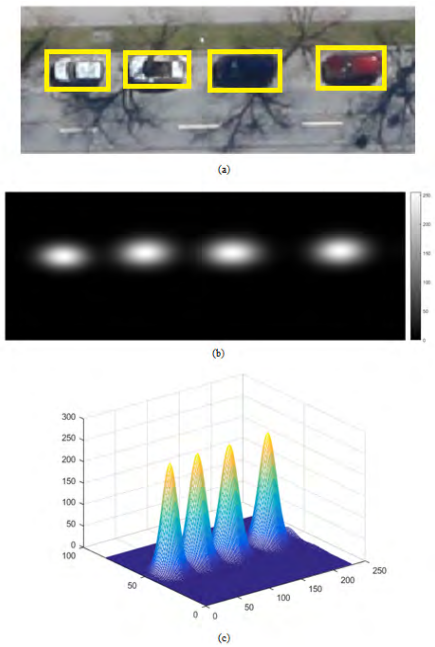

FIGURE 5. Example of ground truth preparation. (a) Input image, (b) generated ground truth, (c) 2D visualization of the ground truth.

FIGURE 6. Examples of aerial images in Munich dataset (first row) and OIRDS dataset (second row).

IV. EXPERIMENTAL RESULTS

In this section, we introduce the datasets used for training the proposed system and the results of the proposed system with comparison with the state-of-the-art methods.

A. DATASETS DESCRIPTION

The proposed system has been evaluated on two public datasets namely DLR Munich vehicle dataset provided by Remote Sensing Technology Institute of the German Aerospace Center [12] and Overhead Imagery Research Data Set (OIRDS) dataset [16]. Munich dataset contains 20 images (5616 x 3744 pixels) taken by DLR 3K camera system at a height of 1000 m above the ground over the area of Munich, Germany. GSD is 13 cm approximately. This dataset contains 3418 cars and 54 trucks annotated in the training image set and 5799 cars and 93 trucks annotated in testing image set. This dataset is challenging because of existence of many disturbance factors such as trees, streets, roads, and similar objects. The images in this dataset were captured in Munich city therefore the vehicle detection task is more challenging than the case of rural areas. To further evaluate the performance of our proposed system, OIRDS dataset has been used. This dataset contains 907 aerial images with approximately 1800 annotated vehicles. The images in this dataset have been taken in suburban areas. Vehicles are occluded partially or totally by trees, buildings, or other objects. Thus, this dataset is equally challenging. Fig. 6 shows few examples of Munich and OIRDS datasets.

B. QUANTITATIVE EVALUATION AND COMPARISON

We have adopted the following evaluation criteria in vehicle detection: recall rate, precision rate, and F1-score. Recall rate is given in (3), precision rate is defined by (4), and F1-score is given by (5).

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5}$$

Where TP is true positive, FP is false positive, and FN is false negative. The comparison between the proposed system and the state-of-the-art detection methods is given in the Table 2. For Munich dataset, the following methods have been used for the comparison.

1) AGGREGATED CHANNEL FEATURES (ACF) DETECTOR [34]

This detector is used as a baseline [26] and has been used by the work proposed by [12].

2) ACF WITH FAST R-CNN [35]

ACF is used for extracting region of interest (ROI) which will be input to Fast R-CNN network for classification.

3) SELECTIVE SEARCH (SS) WITH FAST R-CNN

Selective search [36] is used for predicting the regions of the all object classes. These regions are fed into Fast R-CNN for classification.

TABLE 2. Performance comparison between the proposed method and the state-of-the-art methods.

Method	Ground Truth	True Positive	False Positive	Recall	Precision	F1 score	Time
[12]	5892	4085	619	69.3%	86.8%	0.77	4.40s
ACF detector	5892	3078	4062	52.24%	43.31%	0.47	4.37s
ACF+fast R-CNN	5892	2583	1540	43.84%	62.65%	0.52	6.29 s
SS+fast R-CNN	5892	3287	15012	55.79%	17.96%	0.27	87.84 s
Faster R-CNN	5892	4050	503	68.74%	88.95%	0.78	3.84 s
AVPN_basic [26]	5892	4454	729	75.59%	85.93%	0.80	3.65 s
AVPN_basic+fast R-CNN [26]	5892	4403	384	74.73%	91.98%	0.82	4.05 s
AVPN_large [26]	5892	4538	630	77.02%	87.81%	0.82	3.65 s
H-Fast [28]	5892	4363	696	74.00%	86.2%	0.80	3.65 s
HPRN + Cascade Classifier [28]	5892	4615	560	78.3%	89.2%	0.83	3.93 s
Our proposed system	5892	**5333**	**383**	**90.51**%	**93.30**%	**0.92**	9.7 s

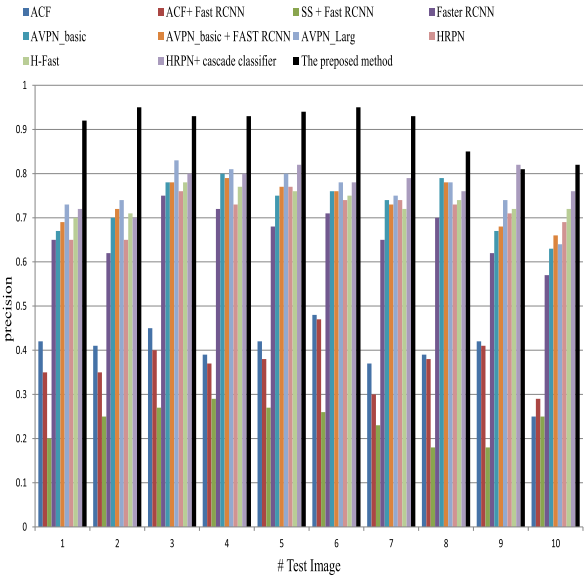

FIGURE 7. Performance comparisons with different methods in terms of precision values for 10 test images in Munich dataset.

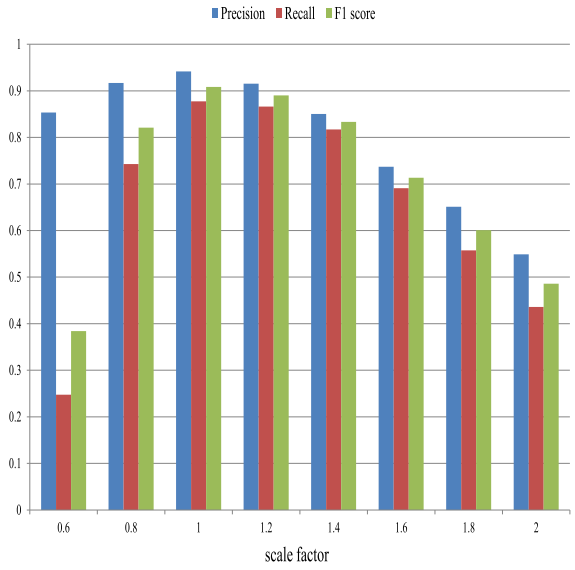

FIGURE 8. Performance of the proposed system after rescaling the test image 2 with different factors.

4) FASTER R-CNN

This network combines region proposal network (RPN) with Fast R-CNN [13]. It performs better than SS with Fast R-CNN.

5) ACCURATE VEHICLE PROPOSAL NETWORK (AVPN) [26]

AVPN combines heretical feature maps which helps in detecting small-sized objects .

6) HYPER REGION PROPOSAL NETWORK (HRPN) WITH FAST R-CNN AND CASCADE CLASSIFIER [28]

HRPN has been used for improving the recall by using a technique similar to [26]. Then, they have used cascade classifier by replacing the one after RPN for reducing the false alarm.

From Table 2, It can be seen that the proposed system outperforms the aforementioned methods in terms of F1-score, precision, and recall. More precisely, we achieve 9%, 1.32%, and 12.2% improvements in terms of F1 score, precision rate, and recall rate, respectively. It can be also observed that our method has achieved the lowest false positive rate and the highest true positive rate. However, the average time spent by FCRN for processing a large scale image from Munich dataset is 9.7 sec. on TitanX GPU with 12 GB memory due to the auto-encoder-like architecture used in the proposed system.

Moreover, we compare the precision performance of our proposed method with the above mentioned methods for all test images in Munich dataset. It is clearly seen that our proposed system outperforms the comparative methods in all test images as shown in Fig. 7.

In addition, we test the ability of the proposed method for vehicle detection and counting in aerial images with different scales. In this case, we resized the test image only without performing training on the new scales. Fig. 8 shows the detection results of the scaled test image 2. It can be seen from the Fig. 8 that the proposed system performs best on the same scale as it was trained. However, the performance decreases remarkably when increasing or decreasing the resolution with

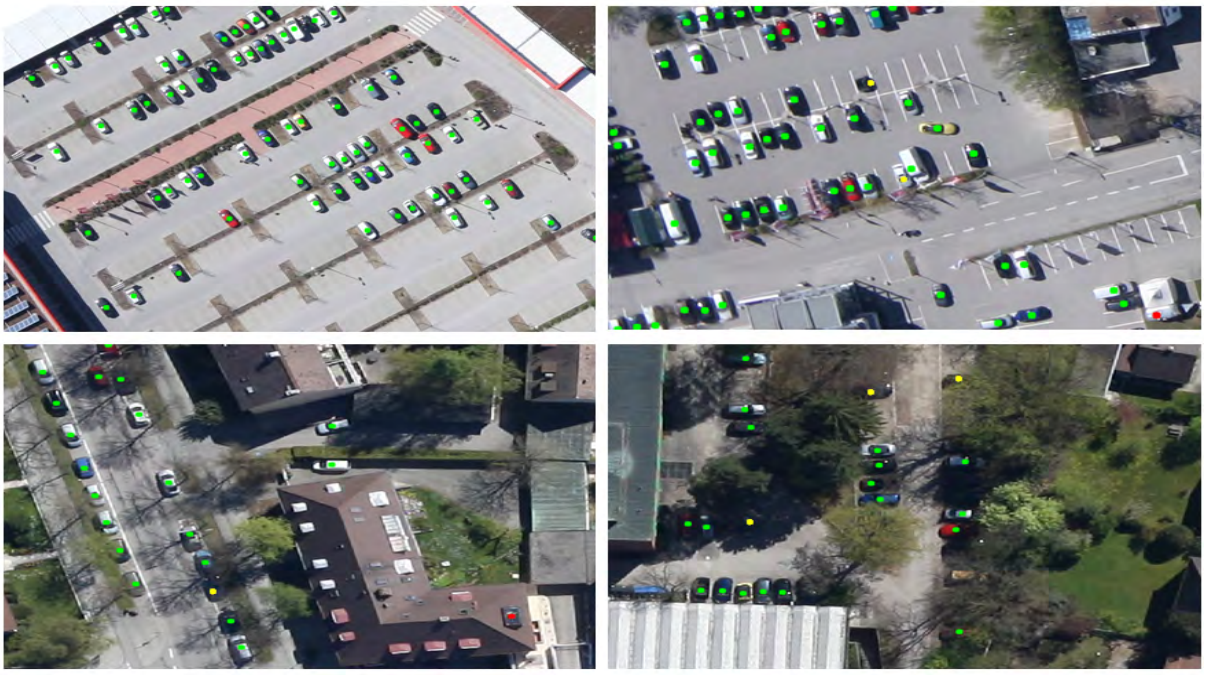

FIGURE 9. Examples of output results of the proposed system on Munich dataset. Green represents true positive cases, yellow represents false negative cases, and red represents false positive cases.

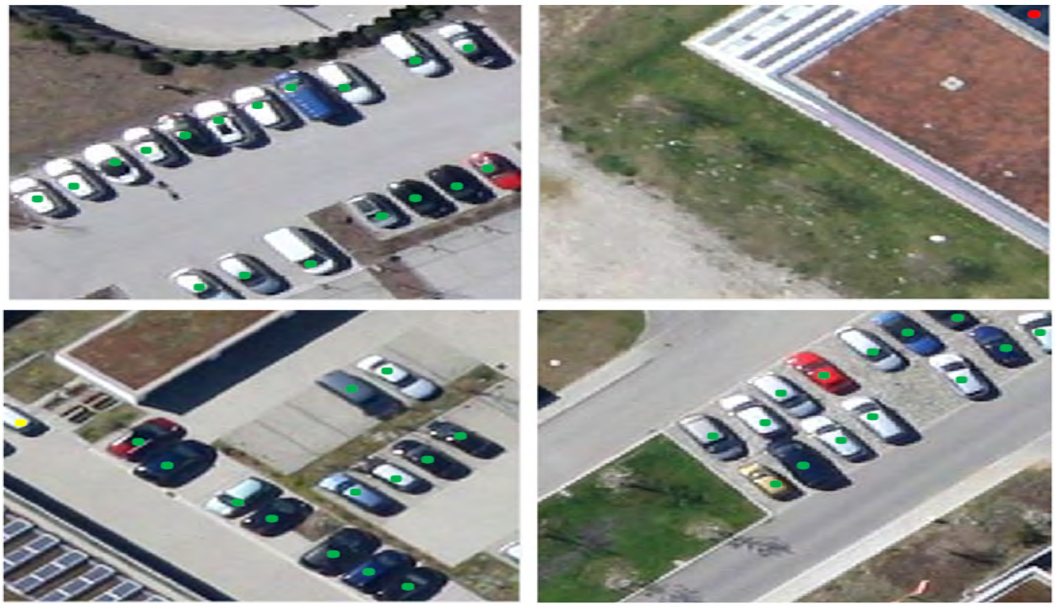

FIGURE 10. Examples of output results of the proposed system on OIRDS dataset. Green represents true positive cases, yellow represents false negative cases, and red represents false positive cases.

a large scale factor. On the other hand, the performance is comparable when increasing or decreasing the resolution slightly.

For OIRDS dataset, the proposed system has been fine-tuned and then tested on 385 images which contains 351 vehicles. The true positive rate is 329, false positive rate is 17. Therefore, the precision rate and recall rates are 95.09% and 93.72%, respectively. These results outperform the works proposed by [23] and [20] where the detection deteriorates when the recall rate is less than 0.7.

C. QUALITATIVE RESULTS

In order to illustrate the effectiveness of the proposed system qualitatively, some of the vehicle detection results are shown

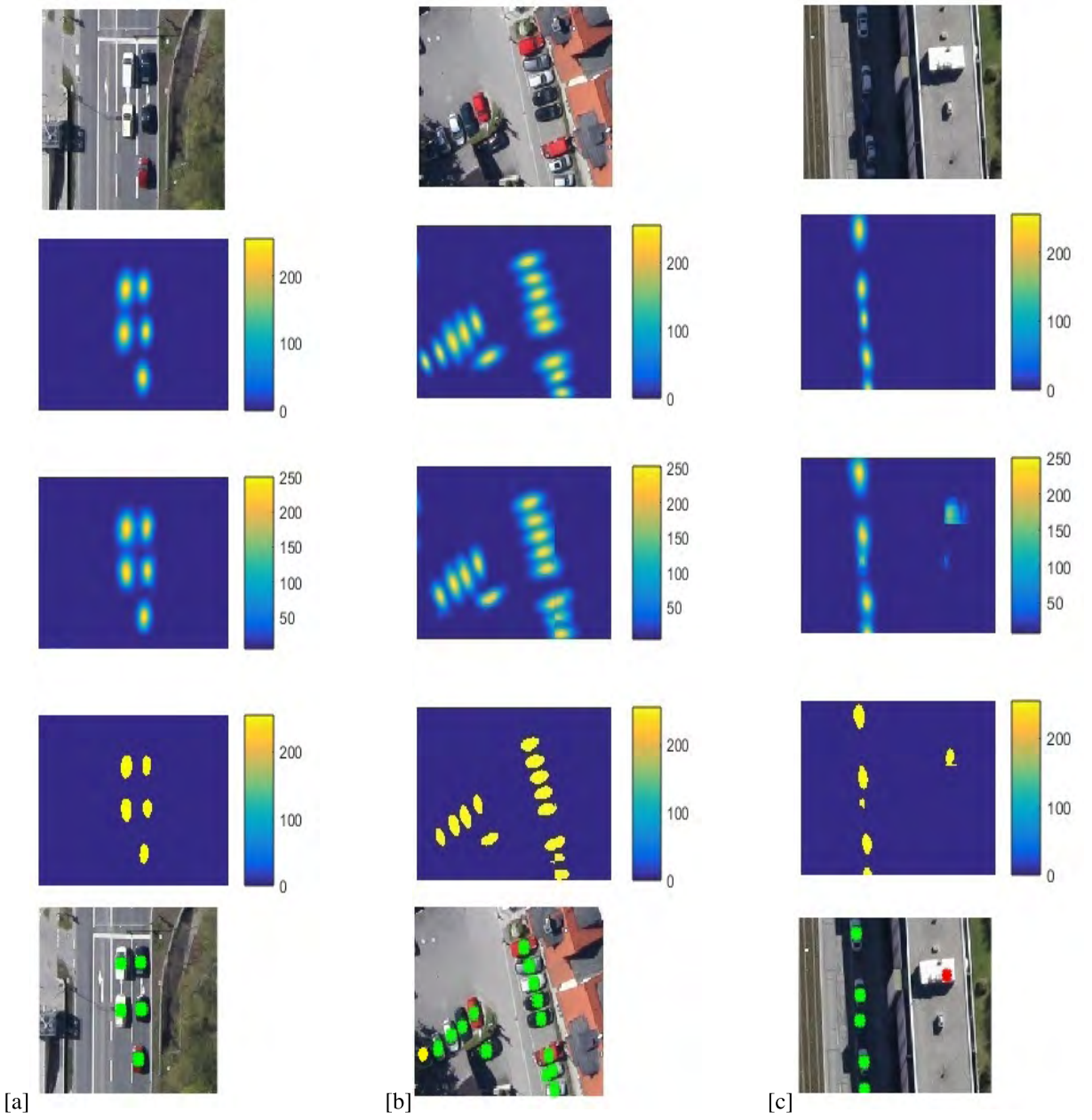

FIGURE 11. Examples illustrate the steps that the proposed system follows: First row represents the input patch, Second row shows the ground truth, Third row shows the predicted density map, Fourth row shows the result of applying thresholding and connected component algorithm, Last row shows the final output of the proposed system. (a) An example of finding all vehicles successfully, (b) an example of false negative which is marked by yellow dot, (c) an example of false positive case which is marked by a red dot.

in Fig. 9 (Munich dataset) and Fig. 10 (OIRDS dataset). In Fig. 9 and Fig. 10, green dots represent true positive cases, red and yellow dots represent false positive and false negative cases, respectively. From the output results, we can see that the proposed system can detect the vehicles in the aerial images accurately. In addition, very low rate of false positive and false negative has been achieved. False negative cases mainly occurs when the vehicles are completely occluded by trees or shadow of the buildings . On the other hand, false positive cases have been reduced as shown in Fig. 9 and Fig. 10. More detailed examples are shown in Fig. 11. First row represents an input patch, second row shows the ground-truth whereas the predicted density map is shown in the third row. The result of applying thresholding and connected component analysis is shown in the fourth row. Finally, fifth row shows the final output of the proposed system. In the Fig. 11, first column shows a case where the proposed model finds all vehicles successfully, second column gives an example of false negative case where missed cars are marked by yellow dot, third column shows a false positive case which is marked by a red dot.

V. CONCLUSION

A novel vehicle detection and counting method has been introduced using convloutional regression neural network. In the proposed system, we have used regression model in order to predict the density map of the input patches. Then, the output of FCRN goes under empirical threshold which results a binary image. Finally, a simple connected component algorithm is used for finding the locations and count of the blobs that represent the detected vehicles. The results of the proposed architecture outperforms the state-of-the-art methods. We have achieved the highest true positive rate and the lowest false alarm rate. In addition, the F1 and precision scores are better than the state-of-the-art methods. However the proposed system consumes more time during inference compared with the other systems. The future work will be focusing on a much faster model with better performance.

REFERENCES

[1] Z. Zheng, X. Wang, G. Zhou, and L. Jiang, "Vehicle detection based on morphology from highway aerial images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 5997–6000.

[2] J. Leitloff, S. Hinz, and U. Stilla, "Vehicle detection in very high resolution satellite images of city areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2795–2806, Jul. 2010.

[3] X. Jin and C. H. Davis, "Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks," *Image Vis. Comput.*, vol. 25, no. 9, pp. 1422–1431, 2007.

[4] R. Ruskone, L. Guigues, S. Airault, and O. Jamet, "Vehicle detection on aerial images: A structural approach," in *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 3. Aug. 1996, pp. 900–904.

[5] B. Salehi, Y. Zhang, and M. Zhong, "Automatic moving vehicles information extraction from single-pass worldview-2 imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 135–145, Feb. 2012.

[6] W. Liu, F. Yamazaki, and T. T. Vu, "Automated vehicle extraction and speed determination from quickbird satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 75–82, Mar. 2011.

[7] A. Kembhavi, D. Harwood, and L. S. Davis, "Vehicle detection using partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1250–1265, Jun. 2011.

[8] H. Grabner, T. T. Nguyen, B. Gruber, and H. Bischof, "On-line boosting-based car detection from aerial images," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 3, pp. 382–396, 2008.

[9] T. Moranduzzo and F. Melgani, "Detecting cars in UAV images with a catalog-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 886–893.

[11] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[12] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[16] F. Tanner *et al.*, "Overhead imagery research data set—An annotated data library & tools to aid in the development of computer vision algorithms," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, Oct. 2009, pp. 1–8.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[18] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1635–1647, Mar. 2014.

[19] T. Moranduzzo, F. Melgani, Y. Bazi, and N. Alajlan, "A fast object detector based on high-order gradients and Gaussian process regression for UAV images," *Int. J. Remote Sens.*, vol. 36, no. 10, pp. 2713–2733, 2015.

[20] Z. Chen *et al.*, "Vehicle detection in high-resolution aerial images via sparse representation and superpixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 103–116, Jan. 2016.

[21] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on Viola–Jones and hog + SVM from UAV images," *Sensors*, vol. 16, no. 8, p. 1325, 2016, doi: 10.3390/s16081325.

[22] J. Zhang, C. Tao, and Z. Zou, "An on-road vehicle detection method for high-resolution aerial images based on local and global structure learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1198–1202, Aug. 2017.

[23] X. Zhang, H. Xu, J. Fang, and G. Sheng, "Urban vehicle detection in high-resolution aerial images via superpixel segmentation and correlation-based sequential dictionary learning," *J. Appl. Remote Sens.*, vol. 11, no. 2, p. 026028, 2017, doi: 10.1117/1.JRS.11.026028.

[24] T. Qu, Q. Zhang, and S. Sun, "Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21651–21663, 2016.

[25] N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan, and M. Zuair, "Deep learning approach for car detection in UAV imagery," *Remote Sens.*, vol. 9, no. 4, p. 312, 2017.

[26] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017.

[27] L. W. Sommer, T. Schuchert, and J. Beyerer, "Fast deep vehicle detection in aerial images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 311–319.

[28] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, p. 336, 2017.

[29] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[30] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[31] M. Abadi *et al.* (May 2016). "Tensorflow: A system for large-scale machine learning." [Online]. Available: https://arxiv.org/abs/1605.08695

[32] T. Tieleman and G. Hinton, "Lecture 6.5—RMSPROP," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[34] P. Dollár, *Piotr's Computer Vision Matlab Toolbox (PMT)*. [Online]. Available: https://github.com/pdollar/toolbox

[35] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Dec. 2015, pp. 1440–1448. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.169

[36] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013.

HILAL TAYARA received the B.Sc. degree in computer engineering from Aleppo University, Aleppo, Syria, in 2008, and the M.S. degree in electronics and information engineering from Chonbuk National University, Jeonju, South Korea, in 2015. He is currently a Researcher with Chonbuk National University. His research interests include machine learning and image processing.

KIM GIL SOO received the Ph.D. degree from the Graduate School, Dongguk University. He was a Lecturer with the Public Administration Department, Chonbuk National University, for 20 years. In 2012, he was an Assistant Professor with industry-university cooperation, Chonbuk National University, where he is currently an Assistant Professor with the Institute of International Affairs. His research interests include image processing and machine learning.

KIL TO CHONG received the Ph.D. degree in mechanical engineering from Texas A&M University in 1995. He is currently a Professor with the School of Electronics and Information Engineering, Chonbuk National University, Jeonju, South Korea, and the Head of the Advanced Research Center of Electronics. His research interests are in the areas of machine learning, signal processing, motor fault detection, network system control, and time-delay systems.