

Vehicle Detection in Aerial Images Using Selective Search with a Simple Deep Learning Based Combination Classifier



Tanuja Tewari, Kaustubh V. Sakhare and Vibha Vyas

Abstract Contrary to the growing zest for bringing in place a complex detection methodology, aiming at improvement in performance of existing methodologies for detecting vehicles in aerial images, this novel piece of work sets forward a much simpler approach with superior results. It was found that methods that showed exemplary performance on common benchmark datasets otherwise, their performance dropped remarkably on aerial images. To achieve performance at par or comparable with the state-of-the-art methods on common benchmark datasets, several adaptations have been suggested in literature to existing methods, for detecting small vehicle instances in aerial images. This ranges from adaptations to the object proposal methods to introduction of more complex deep learning-based classifiers such as fast RCNN and faster RCNN. However, these methods have their own limitations along with the growing increase in system complexity. In this work, a novice, simple and accurate method has been proposed for the detection of small vehicles from aerial images. The experiments have been performed on the publicly accessible and diverse Vehicle Detection in Aerial Imagery (VEDAI) database. This novice technique utilizes Selective Search algorithm as the object proposal method in combination with a deep learning-based framework for classification, which comprises of a simple Convolutional Neural Network (CNN) architecture proposed in combination with a simple Deep Neural Network (DNN) architecture. The DNN utilizes Histogram of Oriented Gradients (HoG) feature input to generate output features that combine with the CNN feature map for final classification. This method is much simpler and achieves a significant accuracy of 96% in vehicle detection, which is much superior to any of the methods tried for aerial images in literature so far.

Keywords Vehicle detection · VEDAI · Object proposal method · Deep learning-based classifier · Fast RCNN · Faster RCNN · Selective search algorithm · Simple CNN architecture · Simple DNN architecture · HoG

T. Tewari (✉) · K. V. Sakhare · V. Vyas

Department of Electronics and Telecommunication, College of Engineering Pune, Pune, India
e-mail: tanujat16.extc@coep.ac.in

© Springer Nature Singapore Pte Ltd. 2019

V. Nath and J. K. Mandal (eds.), *Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems*,

Lecture Notes in Electrical Engineering 556, https://doi.org/10.1007/978-981-13-7091-5_21

1 Introduction

In this era of automatization and digitization, where there is an urge for processing huge amount of information in minimal time, aerial images have come as a savior [1, 2]. Owing to their ability to cover large areas in one go, they have been a point of growing interest among researchers for over a decade. They can access large areas altogether to determine the presence of object of interest within that range which has high utility for real-time applications such as traffic monitoring, environment pollution monitoring, inventory control, and surveillance purposes [1]. However, object detection accurately and instantly in aerial images for real time is a huge challenge. For example, for an aerial image of a traffic scenario, small instances of land vehicles such as cars, vans, and pickups will occupy a very few pixels as compared to the entire image size. Manual inspection will lead to missing of such small instances due to lack of concentration or fatigue which is obvious as human errors are inevitable in manual task. The top performing object detection methods on common standard datasets also fail to handle small instances in aerial images efficiently. This brings in place a high demand for an automatic object detection system which is robust, accurate and at the same time capable of handling small instances (Fig. 1).

Although huge work has been carried out in literature using different methods and techniques for vehicle detection in aerial images [3–14], there is always a scope



Fig. 1 Examples of aerial images in VEDAI with diverse backgrounds showing vehicles in urban and rural areas, forest, marshy, and agricultural lands

for improvement of detection accuracy, robustness and system complexity. All the methods proposed earlier were based on handcrafted features and a classifier combination which were totally dependent on human ingenuity for feature design [3–5]. These methods required manual analysis of real-world data to find an apt feature representation, thereby were not robust to meet the challenges of object detection in aerial images such as occlusion, viewpoint changes, shadow, illumination, background clutter etc. Hence, recent years see the transition to several deep learning-based frameworks such as CNNs [6, 7], which are capable of learning good features automatically from the training samples of most complex objects. However, Regional Convolutional Networks (RCNNs), the pioneers, are slow in the sense that they calculate convolutional features for each candidate region separately [8]. Fast RCNN and Faster RCNN methods, which use Object Proposal Methods to propose candidate regions to the classifier, evolved as the top performers for object detection on common benchmark datasets. However, on inspecting their potential on small occurrences, the performance was not that satisfactory [8–10, 15]. Hence, in this work, to increase the efficiency and robustness of vehicle detection system for handling small occurrences of vehicles in aerial images, a novice, efficient, simple and accurate method has been proposed, which utilizes the high-resolution proposals by Selective Search algorithm [16–18] and, a powerful complex feature representation and classification by a deep neural network framework, which is a combination of a simple CNN and DNN architecture [19–23]. The DNN architecture utilizes the HOG feature input [14], which is a powerful and classical edge information-based feature for detecting vehicles. The combination is powerful enough with its simple architecture to challenge the accuracy of highly complex deep learning-based detectors.

The organization of the paper is as follows. Section 2 comprises of the brief review of prior works and optimal object proposal method in literature. Section 3 discusses the database, performance evaluation parameters, HoG features used as input to the proposed DNN classifier, the proposed deep learning-based classifier combination (CNN + DNN) and the process flow for the novel detection methodology. Section 4 carries the experimentation results and comparison with prior methods whereas Sect. 5 derives the experiment-based conclusion and anticipates the future scope.

2 Related Work

2.1 *Detection of Vehicles in Aerial Images*

The advent of Unmanned Air Vehicles (UAVs) and drones have resulted in the upsurge of aerial photography which has accounted for numerous works on vehicle detection from aerial images for over a decade.

The wide-ranging survey by Cheng and Han [3], reviews traditional methods used for the detection of vehicles in aerial images and anticipates the much-needed tran-

sition to deep learning-based methods in this area. The initial methods in literature were generally based on mechanically derived features such as HoG, Texture, Bag-of-Words (BoW), Sparse Representation, Haar-like features, etc., and classifiers such as Support Vector Machines (SVMs), Adaboost, Artificial Neural Networks, and k-Nearest Neighbor, using a sliding window approach to generate candidate regions. The published work by Moranduzzo and Melgani [4], reviews the performance of various handcrafted features and classifier combination for detecting cars in UAV images. Razakarivony and Jurie [5] used handcrafted features and classifier combination such as (i) Haar wavelets and cascade of boosted classifiers (ii) HoG with SVM classifiers (iii) BoW model and (iv) Deformable Parts Model to counter the challenges in detecting vehicles in aerial images. However, these proposed methods might fail to give powerful feature representation for more complex instances.

Recent times perceive sliding window approach for generating candidate regions being replaced by object proposal methods which can generate regions with high objectness [16, 24]. This brought about a huge revolution in detecting objects in aerial images too, permitting the use of more complex classifiers like RCNNs [7, 20–23]. Thereafter, Fast RCNN [9] and Faster RCNN [10] emerged as the faster variants of RCNN. Instead of generating feature maps for all the proposals separately, they shared convolutional feature map for the entire image among the generated proposals. However, although they emerged to be the top performers on the common standard non-aerial datasets, their performance on detecting small- and medium-sized objects is found to be questionable [8, 15].

2.2 Object Proposal Methods

Among the state-of-the-art Object Proposal Methods based on handcrafted feature-based techniques, Selective Search [17, 18] has shown outstanding performance among all the other methods on common standard datasets as well as aerial datasets with some adaptations [8]. This method with its diverse grouping strategies¹ and hierarchical grouping is powerful enough to capture all the regions with high objectness.

$$s(r1, r2) = a_1s_{colour}(r1, r2) + a_2s_{texture}(r1, r2) + a_3s_{size}(r1, r2) + a_4s_{shape}(r1, r2) \quad (1)$$

$s(r1, r2)$ gives the final similarity measure used by Selective Search for grouping any two segmented regions $r1$ and $r2$ which is the weighted combination of color, texture, size, and shape based similarity measures.

For common benchmark dataset like PASCAL VOC, Selective Search gives a high recall of **99%** with Mean Average Best Overlap (MABO) as **0.879** [17]. On aerial images dataset such as VEDAI 1024, the performance of original Selective Search algorithm [17] drops. However, with the adaptation of parameters like initial segmentation size and minimum proposal width, recall value for Selective Search Algorithm

¹Selective Search uses different similarity measures for hierarchical grouping of segmented regions.



Fig. 2 Illustration of proposals generated by selective search from a VEDAI 1024 image

reaches close to 1 [8]. Hence adapted Selective Search algorithm is selected for generation of well captured and localized proposals to capture small instances in this work. The adaptations made to the algorithm to handle small instances in aerial images are discussed in detail in the subsequent section (Fig. 2).

3 Evaluation

3.1 Database for Aerial Images

All the experiments pertaining to vehicle detection in aerial images have been performed on VEDAI 1024 database. Contrary to prior databases, VEDAI, introduced for aerial images, is specially tailored to detect very small instances and to remove the drawbacks of earlier databases [5]. The dataset contains images for detection of small vehicles in an unrestricted environment, miscellaneous background conditions, images of vehicles affected by occlusions or masks and different orientations [8]. Collected over Utah in U.S., the images consist of varying backgrounds such as urban and rural; forests, marshy and agricultural lands. The images have a resolution of **1024 × 1024 pixels** and Ground Sampling Distance (GSD) of **12.5 cm** per pixel. Ground Truth annotations are available for all the classes in VEDAI. The classes selected for this experimentation are cars, vans, and pickups, since sufficient number of ground truths are available for these classes for training the classifier and performance evaluation. **70%** of the database is used for training, **20%** for validation, and the rest **10%** for testing. The available ground truths were aligned in a way for overlap calculation with proposal bounding boxes generated by the Selective Search algorithm (Table 1).

Table 1 Characteristics of VEDAI database used for the experiments in terms of number of images, size of images, GSD, number of objects, average number of objects per image and the approximate object sizes

Dataset characteristics	VEDAI
Number of images	1268
Size of image	1024 × 1024
GSD in cm per pixel	12.5
Number of objects	2950
Number of objects per image	2.88
Bounding box average width	33.40 ± 11.33
Bounding box average height	33.47 ± 11.68

3.2 Performance Evaluation

The performance of object detection depends on the performance of the Object Proposal Method as well as the classifier. How well the Selective Search algorithm captures the cars, vans, and pickups without missing them is determined by its Recall value and, how well the vehicles are localized in the proposals is determined by MABO. Adaptation of initial segmentation size in Selective Search as per the small size vehicles in VEDAI 1024 renders Recall value close to 1 and MABO close to 0.8 [8]. Intersection over Union (IoU) determines the measure of overlap between the ground truths and the proposals and is given by Eq. 2.

$$IoU = \frac{Ar_{proposal} \cap Ar_{Ground\ Truth}}{Ar_{proposal} \cup Ar_{Ground\ Truth}} \quad (2)$$

Here, $Ar_{proposal}$ is the area of the proposal bounding box and $Ar_{Ground\ Truth}$ is the area of the bounding box of Ground Truth annotation. Aligned ground truths which have an IoU greater than 0.5 are considered as covered in case of VEDAI according to PASCAL VOC criterion. [8, 25]. MABO gives the measure of localization of vehicle objects in the generated proposals, which is given by Eq. 3.

$$ABO = \frac{1}{|G|} \sum_{gt_i} \max IoU(gt_i, l_j); l_j \in L \quad (3)$$

It is calculated by finding the average of the best overlap between each Ground Truth annotation $gt_i \in G$ where G is the set of ground truths and the corresponding object proposals $l_j \in L$ where L is the set of object proposals corresponding to gt_i . Here $|G|$ stands for the total number of Ground Truth annotations.

The classifier performance (combination of CNN and DNN using HoG features) is measured by the **Average Precision** metric, which gives the ratio or percentage of ground truth annotations, predicted as vehicles by the Selective Search and the deep learning-based classifier.

3.3 HoG Features

Features that can distinguish vehicles from most of the objects in the background, are most prominently the edges. Vehicles seem to contain more edges than the surrounding natural objects in the background [14]. Hence features based on edges can be exploited to train the DNN classifier to distinguish vehicles from most of the objects in diverse background scenarios. Many works have also utilized morphological features to enable classifier to distinguish vehicles from the background, but they failed in case of unconstrained background and worked well for a specific background [11–13]. HoG features have been selected as an input to DNN classifier owing to its ability to describe the edge directions and intensity gradient distribution in a localized area which works well to classify the vehicles and non-vehicles. This combination with CNN classifier has shown to improve the accuracy further by introducing feature engineering to assist the classifier (Fig. 3).

3.4 Deep Learning Framework for Classification

Machine learning based classifiers are dependent on manually designed features for classification which aim to reduce the complexity of the data so that the distinguishable patterns become more visible to the classifier [3, 4]. For classification of small instances (approximately 100 to 2000 pixels) in case of VEDAI 1024, merely human ingenuity does not suffice to derive distinguishable features. Hence, here deep

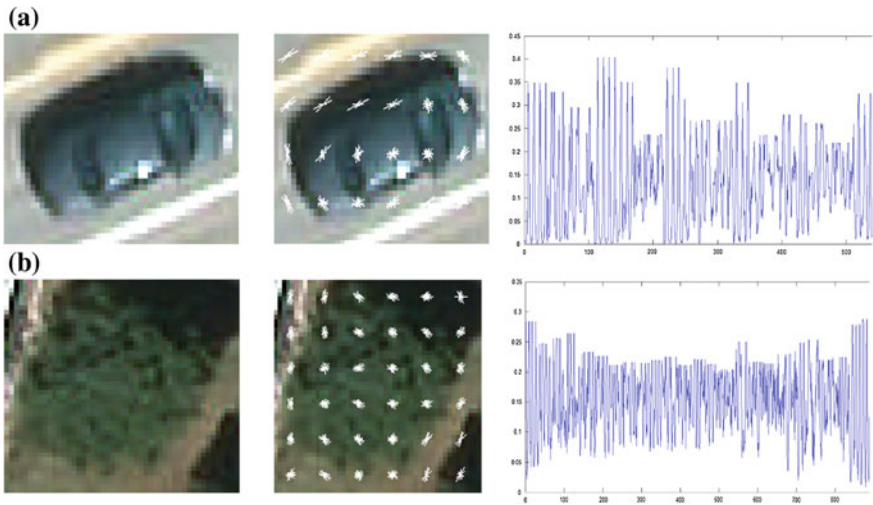


Fig. 3 **a** HoG feature visualization and plot of a vehicle image. **b** HoG feature visualization and plot of a non-vehicle image

learning-based classifiers come for rescue. Deep neural networks like CNN and DNN can extract high-level features to avoid underfitting of such complex sizes.

While CNNs are self-capable of extracting low-level features like edges and gradually construct more complex features, DNNs require low-level features to build up high-level features from them. Although the existing neural networks like RCNN [6, 7] and their new variants Fast RCNN [9] and Faster RCNN [10] have shown remarkable performance in classifying objects in common standard datasets, yet their performance is limited when it comes to small- and medium-sized objects owing to the low resolution of their feature map which leads to missed detections [8, 15]. Also, their complex and deep architecture require high-end machines and GPUs (Graphical Processing Units) for execution. In this work, the effort has been made to exploit the benefits of both feature engineering and deep learning by designing a simple but effective deep learning-based classifier combination that gives accuracy comparable to and above the existing methods. Also owing to its simple architecture, it does not require very high-end machines and GPU's for its execution. The combination

Table 2 Proposed CNN architecture summary for handling small vehicle sizes. Input size is 32×32 pixels

Layer type	Size of kernel	Pad, stride	Output shape
Conv2D Relu	$3 \times 3 \times 32$ –	0, 1 –	(No. of samples, 30, 30, 32) (No. of samples, 30, 30, 32)
Conv2D Relu MaxPool2D Dropout	$3 \times 3 \times 32$ – 2×2 –	0,1 – 0, None –	(No. of samples, 28, 28, 32) (No. of samples, 28, 28, 32) (No. of samples, 14, 14, 32) (No. of samples, 14, 14, 32)
Conv2D Relu	$3 \times 3 \times 64$ –	0, 1 –	(No. of samples, 12, 12, 64) (No. of samples, 12, 12, 64)
Conv2D Relu MaxPool2D Dropout	$3 \times 3 \times 64$ – 2×2 –	0,1 – 0, None –	(No. of samples, 10, 10, 64) (No. of samples, 10, 10, 64) (No. of samples, 5, 5, 64) (No. of samples, 5, 5, 64)
Flatten	–	–	(No. of samples, 1600)
FC Relu Dropout	–	–	(No. of samples, 256) (No. of samples, 256) (No. of samples, 256)
FC	–	–	(No. of samples, 2)

Table 3 Proposed DNN architecture summary for handling small vehicle sizes. Input is 324 HoG features corresponding to each CNN input of 32×32 pixels

Layer (type)	Output shape
FC	(No. of samples, 1024)
Relu	(No. of samples, 1024)
FC	(No. of samples, 256)
Relu	(No. of samples, 256)
FC	(No. of samples, 256)
Relu	(No. of samples, 256)
FC	(No. of samples, 2)

consists of a simple CNN architecture combined with a simple DNN architecture that uses the mechanically engineered HoG features as its low-level feature input (Tables 2 and 3).

3.5 Vehicle Detection

The classes in VEDAI 1024 considered for experimentation are cars, vans, and pick-ups due to their sufficient number of annotations available to train the CNN classifier and for performance evaluation. In Selective Search algorithm, which is a grouping-based object proposal method, the size of initial segmentation k which determines approximately the height or width of the vehicles of interest to be captured is adjusted to capture even the small-sized vehicles (<200 pixels). The vehicles of interest in VEDAI 1024 ranges from 100 to 2000 pixels. The proposals generated even smaller than the size to be captured are eliminated by adjusting the value of *minBoxWidth*. Also, a variable *maxBoxWidth* is introduced in the existing algorithm to set the upper limit. It does away with the proposals which are much larger than the vehicle sizes of interest in aerial images. The value of k , *minBoxWidth* and *maxBoxWidth*, play a vital role in capturing well-localized vehicles in the proposals which is essential for classifier’s performance in terms of *Average Precision*, since vehicles missed and with poor localization will definitely lead to missed detection at the classifier stage. For the training of the CNN classifier to classify the proposals generated from Selective Search, positive samples (vehicles) are obtained from the ground truth annotations available and positive proposals through Selective Search. The negative samples (non-vehicles) are obtained from the negative proposals generated by the Selective Search algorithm. The ratio of positive to negative samples is kept 1:3, since the background of all the images in VEDAI is very diverse so to train the classifier on maximum diversity, negative samples are kept 3 times the positive ones. Out of the entire ground truth annotations available, eliminating the Ground Truths of the test images, the rest are used as positive samples for training the CNN and its validation. 70% of the remaining Ground Truths are used as positive samples for the training dataset and 30% are used as positive samples for the validation dataset. The positive and negative samples to train and validate the classifier are resized to input

size of **32 × 32 pixels** which is approximately equal to the sizes of small vehicles to be detected in the VEDAI dataset. The proposals generated by Selective Search for test images are also resized to **32 × 32 pixels** before feeding to the classifier for classification. Bounding boxes using the coordinates of the generated proposal are constructed if they are qualified as vehicles by the classifier.

To improve the validation accuracy of CNN classifier by reducing the data complexity, some feature engineering was done by introducing HoG features derived from the resized positive and negative samples of the training dataset along with their corresponding labels to train the proposed simple DNN architecture. After training, the validation accuracy of the DNN classifier is computed by feeding HOG features of the positive and negative samples in the corresponding validation dataset. Both the proposed CNN and DNN are trained until their best validation accuracy is obtained without overfitting. At this stage, the final softmax layer is removed and the final activations before the softmax layer from both the networks are combined. Using these activations, the final softmax layer is trained. The validation accuracy obtained from this combined model is much improved which gives an outstanding test accuracy of **96%** on the test dataset.

4 Results and Discussion

The train and validation accuracy of the proposed CNN and DNN model individually as well as that of the combined model is given in Table 4.

The combined CNN and DNN model is effective enough to give an Average Precision of **96%** which is above the accuracy obtained by the classifiers in aerial imagery vehicle detection so far. Examples of vehicles detected by the proposed method are shown in Fig. 4. Examples of missed detections and false positives are shown in Figs. 5 and 6. Table 5 shows a comparison of results of various detection models and the proposed model.

Table 4 Results of the experiments done for detecting vehicles in VEDAI database

Experiments	CNN classifier	DNN classifier	CNN + DNN combination classifier
Training accuracy	0.978	0.983	0.9971
Validation accuracy	0.956	0.9627	0.9709
Validation loss	0.1168	0.129	0.0936
Average precision (CNN + DNN classifier)			0.96



Fig. 4 Examples of small size vehicles detected in VEDAI images by the proposed architecture. The detected vehicles are bounded by red boxes



Fig. 5 Examples of missed detections caused due to partial occlusion or shadows



Fig. 6 Examples of false positives caused by objects of similar shapes such as houses, solar cells, trailers etc

Table 5 Comparison results for various detection models

Methods	Test accuracy (%)	System complexity	Database
SVM classifier with HoG + LBP [5]	76.8	Low	VEDAI 1024
Soft cascade structure of ICF + Adaboost classifier [26]	86.8	Low	DLR 3 K Munich
Faster RCNN with VGG-16 [8]	92.8	High. Expensive GPU machines required	VEDAI 1024
Proposed method CNN + DNN classifier with selective search	96	Low	VEDAI 1024

5 Future Scope and Conclusion

The scope of deep learning is ever progressing, but along with it, even classifier complexity is increasing by leaps and bounds. In the midst of this, this work proposes a Selective Search and deep learning classifier combination-based method for vehicle detection in aerial images, which gives a simplified architecture without compromising with accuracy, compared to prior methods. The proposed method can be generalized to any vehicle aerial images, by merely varying the value of initial segmentation size k , minimum proposal size *minBoxWidth* and maximum proposal size *maxBoxWidth* in the Selective Search algorithm, depending upon the resolution of the input image and the class of vehicles to be detected. The architecture is simple enough to be executed on low-end machines. The experiments are performed on *Intel Core i5 processor, 4 GB RAM*. Simulation tools are *MATLAB 8.2* for generating Selective Search proposals, *Python 3.6* for CNN model designing and *Python Flask* for demonstration. Further taking the proposed method to even low-end GPU machines, can subsequently reduce the time complexity for real-time applications. The future might look forward to the creation of publicly available and diverse databases like VEDAI, with much more increased number of samples of small instances or even consider data augmentation. This would further improve the performance of the proposed combination neural network which increases with the scale of data.

References

1. <http://www.environmentalscience.org/principles-applications-aerial-photography>
2. <http://sciencing.com/difference-satellite-imagery-aerial-photography-8621214.html>
3. Cheng G, Han J (2016) A survey on object detection in optical remote sensing images. *ISPRS J Photogr Remote Sens* 117:11–28
4. Moranduzzo T, Melgani F (2013) Comparison of different feature detectors and descriptors for car classification in UAV images. In: *IEEE international geoscience and remote sensing symposium-IGARSS*, pp 204–207
5. Razakarivony S, Jurie F (2016) Vehicle detection in aerial imagery: a small target detection benchmark. *J Vis Commun Image Represent* 34:187–203
6. Konoplich GV, Putin EO, Filchenkov AA (2016) Application of deep learning to the problem of vehicle detection in UAV images. In: *XIX IEEE international conference on soft computing and measurements (SCM)*, pp 4–6
7. Chen X, Xiang S, Liu C-L, Pan C-H (2014) Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci Remote Sens Lett* 11(10):1797–1801
8. Sommer L, Schuchert T, Beyerer J (2017) Fast deep vehicle detection in aerial images. *IEEE winter conference on applications of computer vision (WACV)*, pp 311–319
9. Girshick R (2015) Fast R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, pp 1440–1448
10. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, pp 91–99
11. Ajmal A, Hussain IM (2010) Vehicle detection using morphological image processing technique. In: *IEEE international conference on multimedia computing and information technology*

12. Pawar BD, Humbe VT (2015) Morphology based composite method for vehicle detection from high resolution aerial imagery. *VNSGU J Sci Technol* 4(1):50–56
13. Zheng H (2006) Automatic vehicle detection from high resolution satellite imagery using morphological neural networks. In: 10th WSEAS international conference on computers, Athens, Greece, pp 608–613
14. Bharathi TK, Yuvaraj S, Steffi DS, Perumal SK (2012) Vehicle detection in aerial surveillance using morphological shared-pixels neural (MSPN) networks. In: IEEE fourth international conference on advanced computing
15. Zhang L, Lin L, Liang X, He K (2016) Is faster R-CNN doing well for pedestrian detection? In: European conference on computer vision, pp 443–457. Springer
16. Hosang J, Benenson R, Dollár P, Schiele B (2016) What makes for effective detection proposals? *IEEE Trans Pattern Anal Mach Intell* 38(4):814–830
17. Uijlings JR, van de Sande KE, Gevers T, Smeulders AW (2013) Selective search for object recognition. *Int J Comput Vis* 104(2):154–171
18. van de Sande KE, Uijlings JR, Gevers T, Smeulders AW (2011) Segmentation as selective search for object recognition. *IEEE international conference on computer vision*
19. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
20. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *Neural information processing systems*, pp 1106–1114
21. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: *International conference on learning representations*
22. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
23. <https://keras.io/>
24. Alexe B, Deselaers T, Ferrari V (2010) What is an object? In: *Proceedings of IEEE conference on computer vision and pattern recognition*, pp 73–80
25. Everingham M, Van Gool L, Williams CK, Winn J, Jisserman A (2010) The Pascal visual object classes (VOC) challenge. *Int J Comput Vis* 88(2):303–338
26. Liu K, Mattyus G (2015) Fast multiclass vehicle detection in aerial images. *Geosci Remote Sens Lett* 1–5. IEEE