Thematic paper: Earth Observation for Smart City and Smart Region

Vehicle detection using panchromatic high-resolution satellite images as a support for urban planning. Case study of Prague's centre

Peter Golej[1] ✉ — Jiri Horak[1] — Pavel Kukuliac[1] — Lucie Orlikova[1]

[1]Department of Geoinformatics, Faculty of Mining and Geology, VSB-Technical University of Ostrava
✉ *peter.golej@vsb.cz*

Abstract

The optical sensors on satellites nowadays provide images covering large areas with a resolution better than 1 meter and with a frequency of more than once a week. This opens up new opportunities to utilize satellite-based information such as periodic monitoring of transport flows and parked vehicles for better transport, urban planning and decision making. Current vehicle detection methods face issues in selection of training data, utilization of augmented data, multivariate classification or complexity of the hardware. The pilot area is located in Prague in the surroundings of the Old Town Square. The WorldView3 panchromatic image with the best available spatial resolution was processed in ENVI, CATALYST Pro and ArcGIS Pro using SVM, KNN, PCA, RT and Faster R-CNN methods. Vehicle detection was relatively successful, above all in open public places with neither shade nor vegetation. The best overall performance was provided by SVM in ENVI, for which the achieved F1 score was 74%. The PCA method provided the worst results with an F1 score of 33%. The other methods achieved F1 scores ranging from 61 to 68%. Although vehicle detection using artificial intelligence on panchromatic images is more challenging than on multispectral images, it shows promising results. The following findings contribute to better design of object-based detection of vehicles in an urban environment and applications of data augmentation.

Highlights for public administration, management and planning:

- Detection of vehicle occurrence by geoinformation technologies provides essential data for planning transport infrastructure.
- Classic machine learning methods achieve similar results to advanced CNN for detection of vehicles using a panchromatic image.
- ArcGIS API for Python provides a suitable user environment and libraries for image processing using artificial intelligence
- For better CNN results, a larger amount of training data and data augmentation are recommended.

Keywords

Vehicle detection, WorldView, Segmentation, Supervised classification, Neural network, Faster R-CNN

Received:
28 September 2022

Received in revised form:
23 November 2022

Accepted:
02 December 2022

The paper was originally presented at the "GIS Ostrava 2022 Earth Observation for Smart City and Smart Region" conference held on-line in March, 2022 (https://gisak.vsb.cz/gisostrava/index.php). Selected presentations from the conference were significantly extended and are now published in this volume as thematic papers exploring various topics related to usage of Earth Observation in smart city and smart region applications.

1 Introduction

The rapidly growing number of vehicles causes various traffic problems (Tan et al. 2020), which are still difficult to measure and predict (Stuparu et al. 2020). Besides tremendous emissions caused by traffic in air, land, and water, there is a gigantic loss of time and money every day in the world due to vehicle traffic congestion (Keler et al. 2017; Stuparu et al. 2020; Tan et al. 2020). Tracked

GeoScape 16(2) — 2022: 108—119 doi: 10.2478/geosc-2022-0009 Available online at content.sciendo.com

movement of objects is nowadays widely available and used for various applications in our society (Keler et al. 2017). Such monitoring can improve the prediction of short-term and long-term traffic situations based on better understanding of traffic congestion propagation in time and space.
Assessment of vehicle occurrence can be also utilized for planning transport infrastructure, fuel demand assessments, walkability studies, evaluation of local traffic density, assessment of emissions, noise, dust, or microclimate and other parts of local environmental impact assessment (Seenouvong et al. 2016). Vehicle monitoring is essential for some financial instruments (e.g., toll tax, parking fees). The remote sensing approach can be used to localize parked vehicles in a city (Seo & Urmson 2009; Zambanini et al. 2020). Current parking monitoring usually employs surveillance cameras and sensors which enable full coverage of large areas (Stuparu et al. 2020). Remote sensing powered by deep learning technologies and appropriate financial instruments could solve the problem of how to effectively limit short/long-term parking (Hong et al. 2022). Special applications can include detection of traffic incidents (Phiri & Morgenroth 2017) or terrorist monitoring (Husain et al. 2020).
Satellite imagery provides snapshots in time that cover a relatively vast area. Compared with ground sensors, satellite remote sensing is more convenient and economic for obtaining urban traffic information (Tan et al. 2020). The progress of advanced processing techniques helps to solve the problem of big data, for which traditional techniques are not enough. The deep learning revolution has opened up an entirely novel frontier. Modern machine learning has boosted techniques such as the detection and automatic counting of objects, semantic segmentation and image classification (Merodio Gómez et al. 2021).
The accuracy of vehicle detection from high-resolution images can currently reach more than 90%. It is conditioned by good pre-processing, including feature extraction such as Hough transformation and other methods of edge detection (Eslami & Faez 2010). Current research prefers to avoid direct extraction of features and focuses instead on advanced ML methods such as RetinaNet Architecture, Faster R-CNN or YOLO (Bin Zuraimi & Kamaru Zaman 2021; Ghosh 2021; Ma et al. 2019; Stuparu et al. 2020; Tan et al. 2020), which can reach more than 94%. The accuracy is not always so high; results depend on many factors, such as environment. In urban areas, the accuracy is usually lower, even when using the same method (e.g. 65% for Faster R-CNN in Zambanini et al. 2020). Also, in the case of lower spatial resolution, results are worse (e.g. accuracy less than 70% in Chen et al. 2021).
Object-oriented classification of satellite imagery still represents a challenge Zhang et al. 2020a, partly due to the wide spectrum of required tasks such as object detection, object tracking, image segmentation and remote sensing image interpretation. Object-based methods dominate among classification techniques of remote sensing images, however, pixel-based methods or sub-pixel-based methods are also still used (Li et al. 2014). Object-based image analysis (OBIA) consists of two steps: image segmentation and object classification (Li et al. 2014; Vecer et al. 2021). The choice of segmentation parameters depends on the specialist performing the required task and the selection is usually subjective and arbitrary. Inappropriate selection of segmentation parameters can lead to undersegmentation or over-segmentation, negatively influencing the final classification (Golej et al. 2021; Zanotta et al. 2018). Object-based image analysis classification methods typically include SVM (Support Vector Machine), KNN (K-nearest neighbor classifier) or RT (Random Trees) (Abburu & Babu Golla 2015; Hossain & Chen 2019; Phiri & Morgenroth 2017; Yekkehkhany et al. 2014).
Object detection is one of the main research areas of computer vision. In the past decade, numerous deep learning based object detection methods for images in remote sensing have been designed for detection of specific objects (Hou et al. 2020; Koga et al. 2018; Xiao et al. 2020; Zhao et al. 2017). The traditional vehicle detection algorithms such as the Gaussian mixed model (GMM, Stauffer and Grimson 1999), give overall satisfactory results, however, they may not work properly when illumination changes occur or due to background cluttering (Maity et al. 2021). The deep learning methods offer an inherent extraction capability feature which makes them much more acceptable to scholars compared to traditional methods as it minimizes the errors occurring in classification tasks (Maity et al. 2021). Convolutional neural networks (CNN) are designed to learn the spatial features, such as edges, corners, or textures, that best describe the target class. The core for learning these features are manifold and successive transformations of the input data (convolutions) on different spatial scales (Kattenborn et al. 2021). The main advantage of CNN is that it automatically detects significant spatial features without any human supervision (Alzubaidi et al. 2021). As CNNs are designed to artificially replicate the functional capabilities of a human cognitive system, they perform better in var-

Available online at content.sciendo.com GeoScape 16(2) — 2022: 108—119 doi: 10.2478/geosc-2022-0009

ious computer vision tasks compared to the traditional methods (Benjdira et al. 2019; Maity et al. 2021). Currently, the most popular target detection algorithms include RCNN, Fast RCNN, and Faster RCNN (Hou et al. 2020; Rawat 2019; Zhang et al. 2020b).

Faster R-CNN is the state-of-the-art algorithm used for generic object detection and it has been successfully adapted to account for many recognition problems. Faster R-CNN has two predecessors: R-CNN and Fast R-CNN. R-CNN (Region-based Convolutional Network) is one of the primary deep neural networks which are designed to perform object detection (Maity et al. 2021). A selection search algorithm extracts around 2000 region proposals which might contain objects (Girshick et al. 2014; Maity et al. 2021). Each of these region proposals or Regions of Interest (RoIs) is processed through a convolutional neural network (CNN) to obtain feature maps. The feature maps are then classified using the SVM model and a bounding box regressor is used to obtain bounding boxes (Girshick 2015; Girshick et al. 2014; Maity et al. 2021). However, this method was criticized for high complexity and the resulting time demands which limits its real applications (Girshick 2015; Maity et al. 2021). Fast R-CNN was developed by (Girshick 2015) and was intended to accelerate processing. In this model, instead of feeding each of the 2000 regions to separate CNNs, the whole image is fed to a single CNN. An RoI pooling layer is used to extract and resize the feature maps of all the region proposals to the same size (Maity et al. 2021). This is then passed on to fully connected layers consist of two branches: a softmax classifier to give probabilities for each class, and a bounding box regressor for precise bounding box coordinates (Girshick 2015). The RoI pooling layer speeds up the object detection of Fast R-CNN compared to R-CNN. However, the problem of inaccurate region proposals still exists in Fast R-CNN due to the non-learning capability of the selective search algorithm (Maity et al. 2021).

Faster R-CNN is divided into two modules: the Region Proposal Network (RPN) and a Fast R-CNN detector. Faster R-CNN improves the object detection architecture by replacing the selection search algorithm in Fast R-CNN with a Region Proposal Network (RPN) (Benjdira et al. 2019; Li et al. 2020). RPN is a fully convolutional network used to generate region proposals, but also can simultaneously propose object bounds and object scores at each position (Li et al. 2020; Maity et al. 2021). The rest of the model architecture remains the same as Fast R-CNN. This design increases the speed of detection and brings it closer to real time (Benjdira et al. 2019; Maity et al. 2021). The system overview of Faster R-CNN is given in Fig. 1.

In order to detect vehicles more accurately and provide accurate traffic information effectively and timely, this study uses object-oriented classifications and also deep learning algorithm to detect vehicles in panchromatic images from high-resolution satellite remote sensing technology. The aim of the paper is to demonstrate differences and issues in vehicle detection using a panchromatic satellite image of the old city centre, and to compare usage of the selected traditional machine learning methods with the advanced Faster R-CNN method.

The paper is organized as follows: after the introduction, the study area and data sources are described. Further, processing in selected SW tools is explained focusing on the explanation of hyperparameters used for processing settings. Next, the obtained results and calculated quality parameters describing the success of vehicle detection methods are provided. Finally, results and various limitations of these methods are discussed.

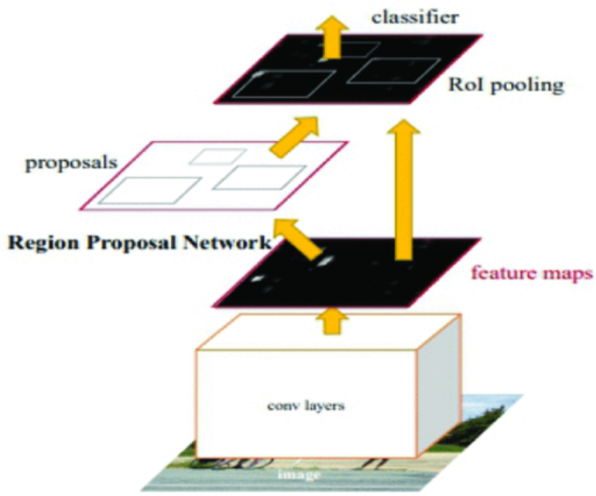

Fig. 1 System overview of Faster-RCNN (Maity et al. 2021)

2 Methods

2.1 Study area and data sources

The study area (Fig. 2) is situated in the centre of Prague around the Old Town Square, with an approx. size of 460 x 300 m. This area represents a historical centre with both narrow and curved old streets, newly built boulevards and various squares, including vehicles on streets as well as in parking lots.

Among various high resolution satellite imagery, one scene of WorldView3 was obtained. Spatial resolution of the panchromatic and multispectral WorldView3 images is 0.3 m and 1.6 m, respectively. The full scene (July 23, 2019, morning) covers 25 km^2 from which a small subset of the panchromatic image was selected to establish the study area. As appropriate auxiliary data, vector data were obtained for buildings in Prague (Prague Geoportal 2021). This data enabled the creation of a mask to improve the quality of detection.
Assessment of the accuracy of vehicle detection requires knowing the ground truth. On-line manual vectorization of vehicles in the study area was performed by a specialist. Three types of vehicles were distinguished – dark, bright, and vehicles in shadow (Table 1); the total number of vehicles is 264. The average size of vectorized vehicles is 7.89 m2 which corresponds to the average size of the top 10 best-selling cars in the Czech Republic, which is 7.94 m2 (ePojisteni.cz 2022). As expected, the size of vehicles in shadow is smaller due to worse conditions for detection.

Table 1 Ground truths of observed vehicles in the study area

	Count	Average vehicle size (m^2)
Dark vehicles	114	7.89
Bright vehicles	104	8.25
Vehicles in shadow	46	7.17

3 Methodology

First, the image was masked. The mask filtered out places where no vehicles or people were expected. Masking limits the detection of objects to the road network, and public spaces such as streets or squares, which provides better classification results.
Three different software packages were used for vehicle detection, image segmentation, modelling and classifications – ENVI, CATALYST Pro, and ArcGIS Pro. The results from all three systems were compared and evaluated.

3.1 Segmentation and classification in ENVI

ENVI is a software dedicated to image processing and analysis, integrated in Esri´s ArcGIS platform. ENVI version 5.6 was used in the research for this paper. ENVI offers two algorithms for segmentation settings. The first one is the edge detection algorithm which is suitable for clearly defined objects. The second is the intensity algorithm, which performs best for images with subtle gradients such as digital elevation models. ENVI also offers two algorithms for merge settings. In this study, the Full Lambda Schedule algorithm was used because it is more efficient in heterogeneous urban areas with a variable texture. It is necessary to set up the merge value appropriately to avoid over-segmentation or under-segmentation. The Texture Kernel Size can be specified as an odd number between 3 and 19. In our case, The value was set to 3 to assure minimal segment areas for such variable urban conditions. After image segmentation, training data were obtained. In total, eight training classes were created - road, shadow, vegetation, crosswalk, sidewalk, bright vehicles, dark vehicles and vehicles in shadow. The classes of crosswalk and sidewalk were added due to the similarity of their pixels to one of the three vehicle types. The second step of OBIA processing is image classification. ENVI provides the following classification methods: KNN, SVM and PCA. For all three methods, after preliminary testing, the threshold value was kept at the default value 5. Each segment was assigned to the class with the highest confidence value. Segments with class confidence values below the threshold value were left unclassified. In the case of the K-NN method, K (number of neighbors) was chosen to be 5. This is because an odd integer ranging from one to a value less than or equal to the total number of training classes for all classes needed to be set. SVM offers four kernel types. The Radial Basis function (Thurnhofer-Hemsi et al. 2020) was selected for this SVM classification. Further, spectral, textural, and spatial attributes for classification must be chosen. The best combination of attributes recommended by ENVI was selected for KNN and SVM, while for PCA all attributes were used.

3.2 Segmentation and classification in CATALYST Pro

CATALYST Professional is a geospatial desktop suite specializing in Remote Sensing, Photogrammetry & Earth Observation Science for optical and SAR imagery. CATALYST PRO version 2222.0.3 was used in this study. Segmentation in CATALYST Pro requires setting three main segmentation parameters: scale (corresponding to the object size – the larger the scale, the larger the output object will be), shape (representing the weight of the shape for segmentation in balance with the colour weight) and com-

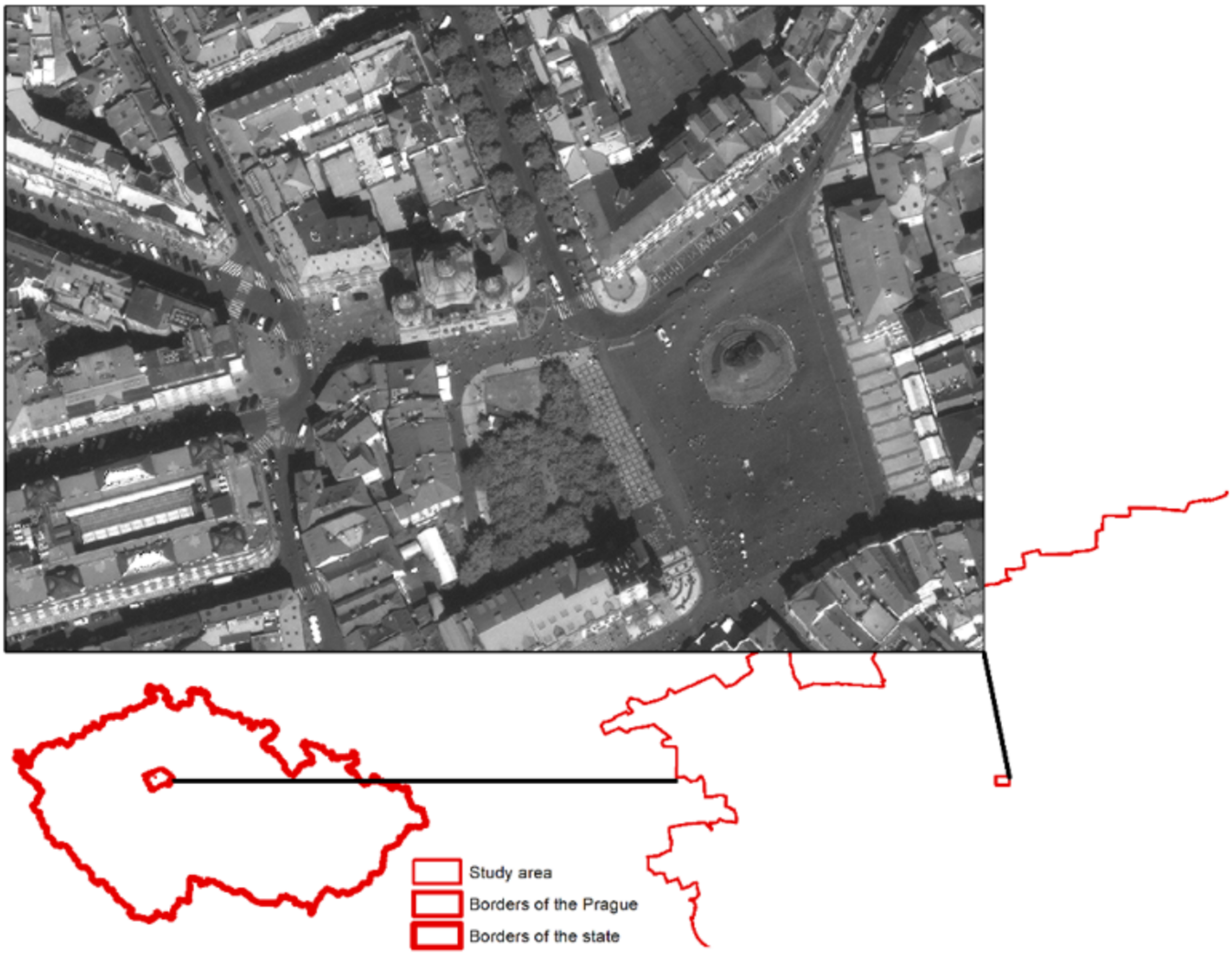

Fig. 2 Study area (Old Town Square in Prague, WorldView3, panchromatic image, resolution of pixel is 30x30 cm)

pactness (denoting the compactness of the shape for segmentation). A higher compactness value generates more compact object boundaries, suitable for, i.e., crop fields. For the second step of OBIA, CATALYST Pro offers two classification methods: RT and SVM (again with four kernel types). Similarly to ENVI's settings, the Radial Basis kernel was selected for SVM. CATALYST Pro feeds the segmentation process with many enumerated attributes organized into the following groups: statistical, geometric, vegetation indices, textural, and polarimetric. For vehicle detection, the following attributes were selected: the mean from the statistical group, and the compactness (unitless geometric measure of a shape, independent of scale and rotation, Hage and Hamade 2016), elongation (ratio between the major axis and the minor axis, Fernandes et al. 2014), circularity (Xu et al. 2017), and rectangularity (the ratio of the segment area to the area of its minimum bounding rectangle, Guindon et al. 2004) from the geometric group.

3.3 Convolutional neural networks

Finally, vehicle detection using artificial intelligence was performed, specifically a convolutional neural network and its Faster R-CNN algorithm. The ArcGIS Pro environment using the ArcGIS API for Python was used to automate the detection process.

Contrary to detections in ENVI or CATALYSTO Pro, the detection with the Faster R-CNN algorithm applied a mask of buildings at the final stage, after the training and testing phases. Testing of the use of a preliminary mask at the beginning of the classification process with Faster R-CNN failed probably due to insufficient training data. Also, all three types of vehicles were joined into one type.

First, it was necessary to create an image that contained three bands (such as RGB), because neural network algorithms work better with such an image type (Kumar et al. 2018). The "composite bands tool" was used to create an artificial image with 3 bands from the panchromatic image. Next, it was

GeoScape 16(2) — 2022: 108—119 doi: 10.2478/geosc-2022-0009 Available online at content.sciendo.com

necessary to export the training data. The online manually vectorised data were used as training data (Table 1). In the training phase, it was necessary to set the following parameters: the tile size (the size of the image chips) was set to 128. A smaller size did not improve the results, and a larger size was limited by the current hardware performance. A stride size (the distance to move in the X and Y when creating the next image chip; when stride is equal to half of the tile size, there will be 50% overlap) was set to 64. The metadata format of training data was set to PASCAL Visual Object Classes in .xml format, which is suitable for the further processing. The maximum number of training epochs was set to 50. Faster R-CNN was selected with the maximal available batch size 32. After optimization, the backbone model was set to ResNet-50 providing the best results. The percentage of training cases for validation was set to 30%. In the testing phase, the probability threshold for the detection of objects (vehicles) was set to 0.65, and the batch size to 32.

After the training and testing phase, the mask of buildings was used as a final step. It helped to eliminate a number of False Positive vehicles detected on rooves.

4 Results

All methods were applied to detect vehicles in the pilot area. The number of detected vehicles varied between 142 (Faster R-CNN) and 596 (PCA-e) (Table 2). As it is shown in the table below, the number of vehicles from PCA-e is overestimated (533 vehicles detected). Processing speed is relatively high, especially in the case of ENVI and CATALYST Pro. Detection using neural networks compared to OBIA is more time-consuming. However, the most time-consuming was the vectorization of vehicles by a specialist and the creation of training samples for classification. This was mainly due to the differentiation of vehicle types, differentiation of vehicles from other classes such as shadow, or the selection of representative segments for training classes.

Table 2 Number of detected vehicles

Vehicles	KNN-e	SVM-e	PCA-e	SVM-c	RT-c	Faster R-CNN	Vectorization
Bright	119	88	55	54	97	-	104
Dark	217	172	533	88	111	-	114
In shadow	20	20	8	11	31	-	46
Total	356	280	596	153	239	142	264

KNN-e, SVM-e, and PCA-e denote methods implemented in ENVI; SVM-c and RT-c methods implemented in CATALYST Pro

Results were compared with the outputs of online vectorization which is considered as a ground truth (Table 3). It enables the evaluation of the number of vehicles successfully detected by the given method (True Positives – TP), the number of non-vehicle objects that are falsely detected as vehicles, (False Positives – FP) and the number of vehicles which the method did not recognize as vehicles but actually were vehicles (False Negatives – FN) (Benjdira et al. 2019).

Table 3 Evaluation parameters of using methods

Method	True Positives	False Positives	False Negatives
KNN-e	193	163	71
SVM-e	200	80	64
PCA-e	143	453	121
SVM-c	128	25	136
RT-c	169	70	95
Faster R-CNN	127	15	137

KNN-e, SVM-e, and PCA-e denote methods implemented in ENVI; SVM-c and RT-c methods implemented in CATALYST Pro

The behaviour of each method is exemplified in the following figures showing the area of Kaprova street. The main issue in classification seems to be an occurrence of mixed pixels. Especially, the KNN method in ENVI was unable to detect many vehicles (Fig. 3a left). Better results can be seen in the case of SVM, but still classification deteriorates for vehicles with a high frequency of mixed pixels (Fig. 3a right). The main problem could be the similarity of the mixed pixels of the vehicle and the mixed pixels of the sidewalks, or the similarity of the mixed pixels of sidewalk and crosswalk, which causes worse vehicle detection results. Similar issues with "mixed" pixels are found in the results of the RT method in CATALYST Pro (Fig. 3b left). A further issue is the detection of vehicles in shadows (both tree and building shadows).

The SVM method in CATALYST Pro (Fig. 3b right) could not detect a vehicle in the shadow.
The PCA method in ENVI in this situation achieved the worst result (Fig. 3c left). This method failed to detect vehicles in shadow, bright vehicles, and vehicles that contain “mixed” pixels. The PCA method also was not able to detect bright vehicles on Kaprova street. Conversely, Faster R-CNN was able to detect vehicles in shadow, and dark as well as bright vehicles, so it detected all the vehicles in the selected area of Kaprova street (Fig. 3c right). False positive vehicles were detected in ENVI and CATALYST Pro SW mainly in shadow areas and in road areas, due to pixel similarity to darker vehicles (Figure 3d).
An issue in the usage of the Faster R-CNN method is false classification of some parts of rooves as vehicles. This is due to the fact that the building mask was not used before the detection (Fig. 3e left). That is why masking was applied after the vehicle detection and the number of FP vehicles was markedly reduced. Resting FP vehicles were mainly detected on roads or in empty parking lots containing “mixed” pixels – usually the result of horizontal traffic signs (Fig. 3e right).
After processing the panchromatic image, three quality indicators were enumerated to evaluate the accuracy of the vehicle detection - Precision, Recall and F1 score (Table 4) (Benjdira et al. 2019). In order to calculate these indicators, it was necessary to determine TP, FP and FN (Table 3).
The best true detection of vehicles was obtained using the SVM method in ENVI. Oppositely, the Faster R-CNN algorithm and SVM method in CATALYST Pro did not recognize the majority of vehicles, so the resulting FN value was the highest of all. The highest FP indication was provided by the PCA method in ENVI. Faster R-CNN detected the least FP vehicles after using the mask.

Table 4 Evaluation parameters of using methods

Method	Precision	Recall	F1 score
KNN-e (%)	54.21	73.11	62.26
SVM-e (%)	71.42	75.76	73.53
PCA-e (%)	23.99	54.17	33.26
SVM-c (%)	83.66	48.48	61.39
RT-c (%)	70.71	64.02	67.19
Faster R-CNN (%)	89.44	48.11	62.56

KNN-e, SVM-e, and PCA-e denote methods implemented in ENVI; SVM-c and RT-c methods implemented in CATALYST Pro

As can be seen from Table 3, evaluation parameters show that, Faster R-CNN and SVM in CATALYST Pro have a high precision rate (89.44% for Faster R-CNN and 83.66 for SVM). This high value indicates that when they classify an object as a vehicle, it is very highly probable that this object is in fact a vehicle. So, the ability of these algorithms and methods to detect real vehicles is quite high. But, when comparing the recall, it is apparent that the best results are achieved by SVM in ENVI and then KNN in ENVI (75.76% and 73.11). The recall measures the ability of the algorithm to detect all the instances of vehicles in an image. Faster R-CNN and SVM in CATALYST Pro achieve the worst results (48.11% and 48.48%). SVM and KNN in ENVI are more capable of extracting almost all instances of vehicles, while Faster R-CNN and SVM in CATALYST Pro miss the majority of vehicles. Considering F1 score, which is a harmonic mean of the precision and recall that gives an assessment of the algorithms’ effectiveness, it is possible to summarize that the best results were achieved by SVM in ENVI. Of course, the suitability depends on which type of errors the end-user prefers to eliminate.

5 Discussion

The Faster R-CNN algorithm detected the least number of vehicles (143). The PCA method achieved the largest number of detected vehicles (596), but in many cases they were FPs. The SVM-c method detected a similar number of vehicles to Faster R-CNN. Approximately the same number of vehicles compared to vectorization was achieved by the RT-c (239) and SVM-e (280) methods. The KNN-e method detected a total of 356 vehicles.
High FP detection was usually caused by mixed pixels, mainly on roads (especially caused by horizontal traffic signs) but also on crosswalks or sidewalks with bright vehicles or vehicles in the shadow. The PCA method has the largest number of FPs (almost double the total number of vehicles), which confirms the inappropriateness of using this method for classification. The least FPs were detected by Faster R-CNN.
The main problem is segmentation, where some vehicles are lost due to their low contrast. Eikvil et al. (2009) presented an automatic approach consisting of a segmentation step followed by object classification to detect vehicles in high-resolution satellite images with 0.6m resolution. Approximately 80% of the vehicles obtained by vectorization were detected. Similar results were also obtained in our case, mainly using the KNN and SVM methods in ENVI. Eslami and Faez (2010) presented a framework to detect vehicles from high-resolution panchromatic images (0.6m) in non-urban areas.

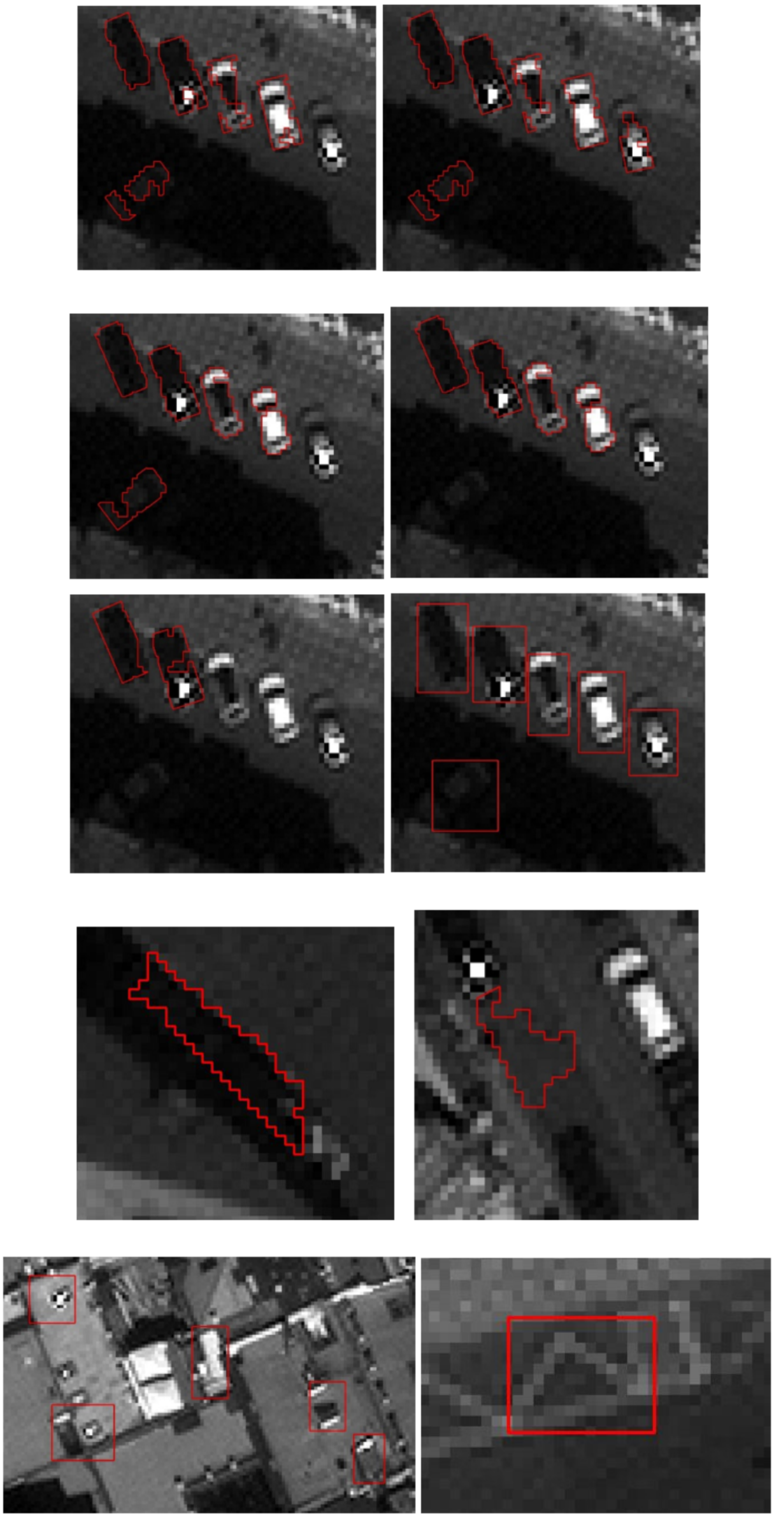

Fig. 3 Detection of vehicles by KNN (left) and SVM (right) methods in ENVI (resolution of pixel is 30x30 cm); b: Detection of vehicles by RT (left) and SVM (right) methods in in CATALYST Pro (resolution of pixel is 30x30 cm); c: Detection of vehicles by PCA method in ENVI (left) and using Faster R-CNN (right)(resolution of pixel is 30x30 cm); d: Incorrectly detected vehicles using KNN method in ENVI (image on the left - shadow area; image on the right - road area. Resolution of pixel is 30x30 cm); e: Incorrectly detected vehicles using Faster R-CNN algorithm (resolution of pixel is 30x30 cm)

Available online at content.sciendo.com GeoScape 16(2) — 2022: 108—119 doi: 10.2478/geosc-2022-0009

They used feature extraction and image processing techniques like Hough transformation, gradient, and thresholding operation. They detected about 94% of total vehicles.

With the development of deep learning and artificial neural networks, selected deep learning methods are also being used in high-resolution remote sensing based vehicle detection in recent years (Ma et al. 2019). As stated by Stuparu et al. (2020), an F1 score of 94.48% was achieved when using the RetinaNet Architecture for object detection and the Cars Overhead With Context dataset (COWC) with a 0.15m resolution. Chen et al. (2021) utilized multitemporal Planet satellite image data (3 m) to develop a vehicle detection method. This vehicle detection algorithm achieved an F1 score of 68% due to low spatial resolution. Zambanini et al. (2020) presented an approach to localize parked vehicles in a city, where a Faster R-CNN detector was trained and applied to stereo satellite images (WorldView 3) to discriminate static from moving vehicles. This model achieved an average precision of 65% for parked vehicle detection. Arora et al. (2022) detected vehicles using Fast R-CNN technology for efficient detection on the road, in both day and night mode. They performed automatic detection of moving vehicles on the road using real-time video of front-side vehicles. Precision was 90%. Ghosh (2021) proposed a method for on-road vehicle detection in varying weather conditions using Faster R-CNN. Three different public datasets, DAWN, CDNet 2014 and LISA, were used for detection. Precision reached the level of 89.48% for DAWN, 91.20% for CDNet 2014 and 95.16% for LISA. Tahir et al. (2022) presented an approach for detection of aircraft objects from Google Earth images. They used Faster R-CNN, YOLO and SSD with accuracy 95.31%, 94.20% and 84.61%, respectively. As Rawat, 2019 describes, Faster R-CNN excels in its task of object detection. He recommends enhancing the model performance by generating higher quality images using data augmentation. To detect various objects (ships, planes, storage, bridges and harbours) in a panchromatic image, Hou et al. (2020) used Faster R-CNN, SSD and YOLOv2 algorithms where the results were very good mainly for ship, storage and bridge detection. The main detection problem was the detection of planes because they have extremely small sizes and plenty of intraclass, which is similar to vehicles detection.

The procedure for vehicle detection using Faster R-CNN proposed in this paper uses very high-resolution satellite images, namely panchromatic images with a resolution of 0.3m. The Faster R-CNN algorithm used achieves an F1 score of 62.56%. Stuparu et al., 2020 used the RetinaNet architecture and achieved an F1 score of 94%. This may be due to having more data and better data resolution (0.15m). Although Chen et al. (2021) achieved a better F1 score (68%) with lower resolution satellite images, this value may be affected by more available training data. Precision in our case reached almost 90%. Arora et al. (2022) detected vehicles using images created from videos, where precision was also around 90%. Ghosh (2021) achieved precision values ranging from 89% to 95% on a more robust dataset. Tahir et al. (2022) detected aircraft objects using Google Earth images with better resolution than our satellite images. The Faster R-CNN algorithm achieved an F1 score of 95.31%. In each of these cases, a larger number of satellite images or images were provided (in most cases with better resolution), which resulted in a larger amount of training data, therefore, the results of the individual algorithms were better. As Rawat (2019) notes, it will be necessary to carry out data augmentation in the future for a larger amount of training data, which will help us achieve better results.

The current hardware limitation (Intel Core i7-8700 CPU 3.20GHz; NVIDIA GEFORCE GTX 1070, and 16 GB RAM) does not enable full optimization of the advanced machine learning process. CUDA (Gavali & Banu 2019) was installed to improve the calculation. CUDA divides more complex calculations into simpler ones and distributes them in parallel among threads in a Graphic Processing Unit (GPU). The improvement in GPU performance over CPU performance is usually 10-20:1 (Alzubaidi et al. 2021)

The recommendation of the best method for vehicle detection depends on the user's preferences. If the goal is to detect as many vehicles as possible, the Faster R-CNN algorithm (89%) seems to be the best choice, followed by SVM in CATALYST Pro (84%). If the objective is to minimize missed vehicles, scholars should prefer SVM in ENVI (76%) followed by KNN in ENVI (73%). According to the F1 score, the SVM method in ENVI appears to be the best (74%), followed by the RT method in CATALYST Pro (67%).

Object-oriented classification has its limitations. Advanced image processing methods are available in relatively expensive commercial software or in free open-source software requiring knowledge of programming languages. Another limitation can be the loss of target objects during segmentation, e.g., of dark vehicles in the shadows. An important step that can improve segmentation is the creation of a mask to filter out areas where vehicles cannot be located, however, such data is not always

available. Other factors that make detection of vehicles challenging are complex backgrounds, varying illumination and differences in the vehicles' types, appearances, and orientations. Another limitation is the spatial resolution of the available satellite images. All these factors reduce the efficacy of segmentation.
It is necessary to note, the results depend on local conditions such as lighting, shadows or the phenological phase of vegetation. The assessment of classification accuracy is affected by the quality of manual classification or vectorization, because the appearance of observed objects can be ambiguous. Classification results can be further improved, especially in the case of the Faster R-CNN algorithm. E.g., data augmentation seems to be a promising technique for improvement of this algorithm.

6 Conclusion

Regular assessment of vehicle occurrence in urban areas contributes to better transport and supports evidence-based planning policy and decision making. Short-term periodic assessment enables mapping the distribution of vehicles, and better understanding of traffic flow, parking behaviour, relationships to trees, shade and surrounding targets. It brings a unique opportunity to evaluate temporal variability of these factors and monitor seasonal changes, the influence of meteorological conditions and societal context. The location of parking lots may be optimized as well as limited to short/long-term parking using financial instruments thanks to detail long-term evidence of parking. Improved local environmental impact assessment may better evaluate local (street-based) quality of life including more precise modelling of noise, dust, emissions and microclimate conditions. Urban walkability analysis may take into account temporary barriers consisting of parked vehicles. Security analysis may utilize the parking situation and usual traffic flows for planning protection for special objects as well as secured escape ways. Remote detection methods offer effective data collection and processing methods to establish a new data source for urban planners and municipal managers on their mission to build smart cities.
Vehicle detection from the high-resolution WorldView image was tested in the urban area using three SW packages: ENVI, CATALYST Pro and AcrGIS Pro. The following methods for classification were applied: SVM, KNN and PCA in ENVI SVM and RT in CATALYST Pro, and Faster R-CNN in ArcGIS Pro. For all methods, a mask of buildings created from data provided by the city was used.
The results varied according to the methods used. The best overall performance seems to be provided by SVM in ENVI, where the achieved F1 score was 73.53%. The KNN method achieved 62.26%. Also, CATALYST Pro with the methods it offers achieved similar results to ENVI, the SVM method achieved 61.39%, and RT 67.19%. On the other hand, Faster R-CNN achieved comparable results to classic machine learning methods, where the F1 score was more than 62%. The results of the Faster R-CNN algorithm were mainly influenced by the amount of training data. For better results, it is advisable to use a larger amount of training data or a larger number of images, which is, however, more challenging due to the price of the image, or applied data augmentation.
For vehicle detection using a panchromatic image, based on the results obtained in this paper, it is possible to use classic machine learning methods, as well as advanced CNN. It is also important to choose appropriate and representative training samples for classification. However, it is important to note that CNN requires powerful hardware as well as a larger amount of training data. Future research should focus on testing other algorithms (SSD, U-net or Mask R-CNN) on panchromatic images but also on RGB images.

Acknowledgements

The authors gratefully acknowledge the financial support from grant SP2021/36 of the Faculty of Mining and Geology of the Technical University of Ostrava. Data courtesy of WorldView 3 was provided by the DigitalGlobe.

References

Abburu S, Babu Golla S (2015) Satellite Image Classification Methods and Techniques: A Review. *Int. J. Comput. Appl* 119: 20–25. doi: 10.5120/21088-3779.

Alzubaidi L, Zhang J, Humaidi A, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel M, Al-Amidie M, Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8:53. doi: 10.1186/s40537-021-00444-8.

Arora N, Kumar Y, Karkra R, Kumar M (2022) Automatic vehicle detection system in different environment conditions using fast R-CNN. *Multimed. Tools Appl* 81:18715–18735. doi: 10.1007/s11042-022-12347-8.

Benjdira B, Khursheed T, Koubaa A, Ammar A, Ouni K (2019) Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3. In: *2019 1st International*

Conference on Unmanned Vehicle Systems-Oman (UVS). Presented at the 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS pp. 1–6. doi: 10.1109/UVS.2019.8658300.

Bin Zuraimi M, Kamaru Zaman F (2021) Vehicle Detection and Tracking using YOLO and DeepSORT. In: *2021 IEEE 11th IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE). Presented at the 2021 IEEE 11th IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE* pp. 23–29. doi: 10.1109/ISCAIE51753.2021.9431784.

Chen Y, Qin R, Zhang G, Albanwan H (2021) Spatial Temporal Analysis of Traffic Patterns during the COVID-19 Epidemic by Vehicle Detection Using Planet Remote-Sensing Satellite Images. *Remote Sens* 13:208. doi: 10.3390/rs13020208.

Eikvil L, Aurdal L, Koren H (2009) Classification-based vehicle detection in high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens* 64:65–72. doi: 10.1016/j.isprsjprs.2008.09.005.

ePojistenicz (2022) The best-selling vehicles in the Czech Republic [WWW Document. Available at: <https://www.epojisteni.cz/aktuality-nejprodavanejsi-znacky-a-modely-aut-v-cesku-ktere-to-jsou/>.

Eslami M, Faez K (2010) Automatic Traffic Monitoring from Satellite Images Using Artificial Immune System. In: RC W, T U, I E, F, (eds.) *in*. Springer, Hancock, E.R., Wilson, pp. 170–179. doi: 10.1007/978-3-642-14980-1_16.

Fernandes M, Aguiar F, Silva J, Ferreira M, Pereira J (2014) Optimal attributes for the object based detection of giant reed in riparian habitats: A comparative study between Airborne High Spatial Resolution and WorldView-2 imagery. *Int. J. Appl. Earth Obs. Geoinformation* 32:79–91.

Gavali P, Banu J (2019) Chapter 6 - Deep Convolutional Neural Network for Image Classification on CUDA Platform. In: Ed, (ed.) *Sangaiah, A.K.* Academic Press, pp. 99–122. doi: 10.1016/B978-0-12-816718-2.00013-0.

Ghosh R (2021) On-road vehicle detection in varying weather conditions using faster R-CNN with several region proposal networks. *Multimed. Tools Appl* 80:25985–25999. doi: 10.1007/s11042-021-10954-5.

Girshick R (2015) Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision* pp. 1440–1448.

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 580–587.

Golej P, Orlikova L, Horak J, Linhartova P, Struhar J (2021) DETECTION OF PEOPLE AND VEHICLES USING VERY HIGH-RESOLUTION SATELLITE IMAGES. In: *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM* pp. 283–291. doi: 10.5593/sgem2021/2.1/s10.65.

Guindon B, Zhang Y, Dillabaugh C (2004) Landsat urban mapping based on a combined spectral–spatial methodology. *Remote Sens. Environ* 92:218–232. doi: 10.1016/j.rse.2004.06.015.

Hage I, Hamade R (2016) Geometric-attributes-based segmentation of cortical bone slides using optimized neural networks. *J. Bone Miner. Metab* 34:251–265. doi: 10.1007/s00774-015-0668-0.

Hong S, Shin H, Choi J, Park N (2022) Prediction-based One-shot Dynamic Parking Pricing. In: *Proceedings of the 31st ACM International Conference on Information Knowledge Management. Presented at the CIKM '22: The 31st ACM International Conference on Information and Knowledge Management*ACM, Atlanta GA USA, pp. 748–757. doi: 10.1145/3511808.3557421.

Hossain M, Chen D (2019) Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogramm. Remote Sens* 150:115–134. doi: 10.1016/j.isprsjprs.2019.02.009.

Hou B, Ren Z, Zhao W, Wu Q, Jiao L (2020) Object Detection in High-Resolution Panchromatic Images Using Deep Models and Spatial Template Matching. *IEEE Trans. Geosci. Remote Sens* 58:956–970. doi: 10.1109/TGRS.2019.2942103.

Husain A, Maity T, Yadav R (2020) Vehicle detection in intelligent transport system under a hazy environment: a survey. *IET Image Process* 14:1–10. doi: 10.1049/iet-ipr.2018.5351.

Kattenborn T, Leitloff J, Schiefer F, Hinz S (2021) Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens* 173:24–49. doi: 10.1016/j.isprsjprs.2020.12.010.

Keler A, Krisp J, Ding L (2017) Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points. *Geo-Spat. Inf. Sci* 20:333–344. doi: 10.1080/10095020.2017.1399672.

Koga Y, Miyazaki H, Shibasaki R (2018) A CNN-Based Method of Vehicle Detection from Aerial Images Using Hard Example Mining. *Remote Sens* 10:124. doi: 10.3390/rs10010124.

Kumar J, Huan T, Li X, Yuan Y (2018) Panchromatic and Multispectral Remote Sensing Image Fusion Using Particle Swarm Optimization of Convolutional Neural Network for Effective Comparison of Bucolic and Farming Region.

Li M, Zang S, Zhang B, Li S, Wu C (2014) A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information. *Eur. J. Remote Sens* 47:389–411. doi: 10.5721/EuJRS20144723.

Li M, Zhang Z, Lei L, Wang X, Guo X (2020) Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD. *Sensors* 20:4938. doi: 10.3390/s20174938.

Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson B (2019) Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens* 152:166–177. doi: 10.1016/j.isprsjprs.2019.04.015.

Maity M, Banerjee S, Sinha Chaudhuri S (2021) Faster R-CNN and YOLO based Vehicle detection: A Survey. In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC). Presented at the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC* pp. 1442–1447. doi: 10.1109/ICCMC51019.2021.9418274.

Merodio Gómez P, Juarez Carrillo O, Kuffer M, Thomson D, Olarte Quiroz J, Villaseñor García E, Vanhuysse S, Abascal A, Oluoch I, Nagenborg M, Persello C, Brito P (2021) Earth Observations and Statistics: Unlocking Sociodemographic Knowledge through the Power of Satellite Images. *Sustainability* 13:12640. doi: 10.3390/su132212640.

Phiri D, Morgenroth J (2017) Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sens* 9:967. doi: 10.3390/rs9090967.

Prague Geoportal (2021) Prague Geoportal.

Rawat S (2019) Airplanes Detection for Satellite using Faster RCNN [WWW Document. *Medium. URL*. Available at: <https://towardsdatascience.com/airplanes-detection-for-satellite-using-faster-rcnn-d307d58353f1>.

Seenouvong N, Watchareeruetai U, Nuthong C, Khongsomboon K, Ohnishi N (2016) A computer vision based vehicle detection and counting system. In: *2016 8th International Conference on Knowledge and Smart Technology (KST). Presented at the 2016 8th International Conference on Knowledge and Smart Technology (KST* pp. 224–227. doi: 10.1109/KST.2016.7440510.

Seo YW, Urmson C (2009) Utilizing prior information to enhance self-supervised aerial image analysis for extracting parking lot structures. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. Presented at the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* pp. 339–344. doi: 10.1109/IROS.2009.5354405.

Stauffer C, Grimson W (1999) Adaptive background mixture models for real-time tracking. *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). Presented at the Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat* 2, (PR00149)):246–252. doi: 10.1109/CVPR.1999.784637.

Stuparu DG, Ciobanu RI, Dobre C (2020) Vehicle Detection in Overhead Satellite Images Using a One-Stage Object Detection Model. *Sensors* 20:6485. doi: 10.3390/s20226485.

Tahir A, Munawar H, Akram J, Adil M, Ali S, Kouzani A, Mahmud M (2022) Automatic Target Detection from Satellite Imagery Using Machine Learning. *Sensors* 22:1147. doi: 10.3390/s22031147.

Tan Q, Ling J, Hu J, Qin X, Hu J (2020) Vehicle Detection in High Resolution Satellite Remote Sensing Images Based on Deep Learning. *IEEE Access* 8:153394–153402. doi: 10.1109/ACCESS.2020.3017894.

Thurnhofer-Hemsi K, López-Rubio E, Molina-Cabello M, Najarian K (2020) Radial basis function kernel optimization for Support Vector Machine classifiers.

Vecer M, Horak J, Golej P, Orlikova L (2021) Segmentation and Object-Based Land Cover Classification of Airborne Images in Kraliky County. In: *2021 8th International Conference on Military Technologies, ICMT 2021 - Proceedings*. doi: 10.1109/ICMT52455.2021.9502817.

Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, Lan X (2020) A review of object detection based on deep learning. *Multimed. Tools Appl* 79:23729–23791. doi: 10.1007/s11042-020-08976-6.

Xu Y, Carlinet E, Géraud T, Najman L (2017) Hierarchical Segmentation Using Tree-Based Shape Spaces. *IEEE Trans. Pattern Anal. Mach. Intell* 39:457–469. doi: 10.1109/TPAMI.2016.2554550.

Yekkehkhany B, Safari A, Homayouni S, Hasanlou M (2014) A COMPARISON STUDY OF DIFFERENT KERNEL FUNCTIONS FOR SVM-BASED CLASSIFICATION OF MULTI-TEMPORAL POLARIMETRY SAR DATA. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 2, (W3):281–285. doi: 10.5194/isprsarchives-XL-2-W3-281-2014.

Zambanini S, Loghin AM, Pfeifer N, Soley E, Sablatnig R (2020) Detection of Parking Cars in Stereo Satellite Images. *Remote Sens* 12:2170. doi: 10.3390/rs12132170.

Zanotta D, Zortea M, Ferreira M (2018) A supervised approach for simultaneous segmentation and classification of remote sensing images. *ISPRS J. Photogramm. Remote Sens* 142:162–173. doi: 10.1016/j.isprsjprs.2018.05.021.

Zhang D, Liu Z, Shi X (2020a) Transfer learning on EfficientNet for remote sensing image classification. In: *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering*ICMCCE, pp. 2255–2258. doi: 10.1109/ICMCCE51767.2020.00489.

Zhang Q, Chang X, Bian S (2020b) Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN. *IEEE Access* 8:6997–7004. doi: 10.1109/ACCESS.2020.2964055.

Zhao W, Du S, Emery W (2017) Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens* 10:3386–3396. doi: 10.1109/JSTARS.2017.2680324.

