SvedbergOpen
DISSEMINATIONOFKNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Vehicle Detection from High-Resolution Satellite Images Using YOLOv8 for Traffic Monitoring

Dr. Chhaya Gosavi[1,2*], Dr. Bharati Ainapure[2]

[1]Department of Computer Engineering Cummins College of Engineering for Women Pune, India.
E-mail: chhaya.gosavi@cumminscollege.in
[2]Department of Science and Technology, Vishwakarma University, Pune, India. E-mail: bharati.ainapure@vupune.ac.in
*Corresponding Author: Dr. Chhaya Gosavi, E-mail: chhaya.gosavi@cumminscollege.in

Abstract
Recent availability of high-resolution space-borne satellite imagery has opened new opportunities for large-area vehicle monitoring, traffic analysis, urban planning and intelligent city management. Automated vehicle detection in satellite imagery still remains quite difficult owing to different scales of vehicles, the prevalence of complex backgrounds occlusion shadows, lighting variations, and factors leading to huge scale variance. This paper proposes a deep-learning driven vehicle-detection structure based on YOLOv8 object detection structures to enhance the vehicle detection performance, in particular for small vehicle detection in high-resolution satellite images. This setup adopts multiple image resizing and adaptive image tiling mechanisms during the detection cycle to maintain spatial relation information of detected boxes, this way mitigating the small object detection problem. It further employs data augmentations and multi-scale feature extraction mechanisms to preserve detail information for a complete detection process. Considering the use of publicly available aerial imagery datasets (e.g. VEDAI COWC etc.), performance is evaluated using precision recall mAP, F1-score and inference speed. The motivation of this research is based on the results reported among 16 selected vehicle-detection studies over 20192025, which consistently identified that tiny objects, multiple vehicles clusters, occlusion and large imagery size remained major problems in detection tasks. The publicly available EAGLE dataset can provide >200,000 vehicles as training samples. The proposed YOLOv8 setup aims to enhance small vehicle detection accuracy, generalization and robustness, without or with marginal compromise to efficiency. This will help achieve the requirements for large-area vehicle detection, transport research, mobility analysis, smart city planning and other remote sensing applications. This study offers a quantitative structure of vehicle detection evaluation metrics for YOLOv8 model suitability assessment.

Keywords: High-Resolution Satellite Imagery, Vehicle Detection, YOLOv8, Deep Learning, Small Object Detection, Multi-Scale Feature Enhancement, Image Tiling, Satellite Image Processing, Traffic Monitoring, Remote Sensing, Computer Vision, Smart Cities.

Introduction

With the rapid development of Earth observation technology, satellite images can now provide a powerful source of information for intelligent large-scale, complex geographic region monitoring rather than merely maps and environment assessments as in the past. Detailed high-resolution satellite images can reveal information about roads buildings parking stations, transportation portals, industrial zones, and other human-built structures, which have led to increasing utilising of satellite images for urban planning, transportation monitoring, disaster report, infrastructure evaluation, environmental observation, security and smart city management[1][2][22]. Compared with traditional ground-based monitoring systems which need large scale network of roadside cameras or sensors, remote sensing systems is capable of monitoring the geocentric large sized region without additional deployment costs. The capability of repeated observation of the geocentric region over time scale has been further enriched by the advent of satellite video, which gives additional temporal data for Earth observation. Satellite video could support the repeated monitoring of the geographic region of interest and track the moving objects over successive frames. Satellite video platforms including Jilin-1 and SkySat have already demonstrated the potential of repeated and moving observation. SkySat-1, as an example, proposed sub-metre level observation with resolution approximately 1.1 m and rate of around 30 frames per second; Jilin-1 video satellites demonstrate high resolution colour observation capabilities, and other satellite video systems possess improved spatial and time resolution capability (Qilu-4 etc.) [22]. As a result, both satellite imagery and satellite video are becoming highly relevant to transportation security disaster management, visual observation and environmental computation.
In particular, the automatic detection of vehicles from satellite imagery could provide critical information for the distribution of traffic, park utilisation, transportation system, urban mobility and the variation of

human activity. [1][2][22] Unlike vehicle detection in commonly-used ground-view imagery, vehicle detection in satellite and aerial images is much more challenging. A vehicle appears as a relatively discrete object in the ground-view image, but only as a few pixels in the satellite-image. The reduced representation leads to the loss of detailed information like the shape texture edge and appearance of the vehicle. The problem is even more challenging when vehicles are tightly distributed, partially occluded and close to the background (e.g. buildings or road infrastructure), and affected by shadow/inter-illumination effects, or are acquired under different illumination conditions. Large satellite scenes further introduce a computational constraint, as a single high-resolution image can have billions of pixels, while resizing it to the input size of an object detector delivers significant loss of small vehicle details. As the literature, small object size, cluttered background scene, scale variation, dominant foreground-background imbalance, illuminance variations, occlusion, and low spatial resolution are identified as the most significant constraints when applying intelligent satellite-image and satellite-video processing [3][8][11][13][22]. This issues suggest that a robust vehicle-detection system cannot solely rely on standard image classification methods, but must have an architecture to allow for the special-scale extraction of informative features in an efficient manner that is computationally feasible across huge quantities of remote-sensed imagery. The role of automated vehicle detection has evolved over several different setups. Image processing techniques like template matching, edge extraction, shape analysis, texture descriptors, and broadly human-designed visual features laid the broad conceptual structure for the earliest vehicle detection algorithms in aerial scenes but also revealed strong reliance on image quality, object appearance, illumination and the selection of feature representations. Machine-learning based algorithms using classifiers like Support Vector Machines and Random Forests mapped the manually extracted features onto the categorical responses of vehicle or not vehicle. While providing increased classification flexibility, the reliance on the image-derived features for training remained, limiting the performance of these algorithms. Imaging features could be learned from training data using convolutional neural networks, changing the way of thinking entirely. For object localisation, region-based architectures like Faster R-CNN offered a huge advantage, whereas for many applications requiring near real-time detection speed, the single-stage detector architectures like YOLO or RetinaNet presented the best trade-offs [3][4][5]. Research in the development of lightweight architectures, such as SSSDet, proved that aerial vehicle detection need not be computationally expensive while still delivering state of art detection performance [3]. In particular for high-resolution aerial and satellite imagery, the success of methods for large-scale automated vehicle detection has been built on a foundation of prior work.

Froidevaux et al. highlighted the potential of very-high-resolution satellite imagery for vehicle detection and counting [2], while Mandal et al. proposed SSSDet, a short-depth shallow CNN architecture for resource-efficient vehicle detection in aerial scenes [3]. Lechgar et al. demonstrated the detection of vehicle flocks in aerial imagery using the YOLOv2 architecture, and underscored the suitability of real-time object-detection architectures for large-scale monitoring of urban vehicles [4]. Liao et al. proposed the localand-share CNN (LR-CNN) with local-aware spatial information, to enhance the detection of small objects local context, in aerial imagery [5]. Large well-annotated datasets have also been key to progress, especially the EAGLE dataset, which comprises over 200,000 vehicle annotations, which can be used for training and benchmarking deep-learning approaches to aerial-vehicle detection [6][27]. Resources like VEDAI [6][25] and xView have supported progress in the development of deep neural networks for small-object detection in remotelysensed scenes.

Though CNN-based object detectors have advanced these days, small-objects, such as the satellite vehicle detection problem, remains a core challenge. In the high-resolution remote-sensing imagery, the detector needs not only separate vehicles and visually similar background details, but also preserve spatial information. Feature-pyramid and multi-scale strategies have this way been key areas of research, as information at different resolutions can be integrated to aid the small-object representation. The literature also depicts a trajectory of moving from earlier YOLO front-ends, toward more powerful YOLO back-ends, such as YOLOv5 YOLOv7 YOLOv8 and other subsequent versions. Several studies have explored improved YOLO variants to achieve higher detection accuracy for the system, while enabling the required speed at large scale [26]. Also, comparative studies on YOLO-based algorithms further prove that, different generations of models may have their own effectiveness and robustness trade-offs while balancing accuracy and computation complexity, in particular for aerial images [26]. Yet, applying a popular large-scale object detector to satellite scene cannot automatically eliminate the small-object challenge. Rescaling satellite scene can simply blow small vehicles away from usable feature maps, or entirely eliminate vehicles with dense tracks with overlapping bounding boxes, so image tiling, data enhancement, multi-scale feature learning, and optimized feature fusion are crucial for a satellite scene-specific vehicle detection system. The problem becomes more difficult when the task is changed from detecting vehicles in each individual image to detecting moving vehicles in satellite-video sequence.

A single satellite image contains only spatial information, but an entire satellite-image sequence contains both spatial and temporal information, which allows for the study of motion trajectory traffic density, etc.

but also necessitate the development of new computational and algorithmic methods. Previous satellite-video detection studies have used background modelling and inter-frame differencing to detect changes between pairs of frames, whereas later research has used correlation filters, deep-learning detectors, temporal models, and trackers[9][10][13][14]. Study of satellite-video through the years reveals Truth is object detection, single-object tracking, multi-object tracking segmentation scene classification, super-resolution, and motion estimation are all hugely interconnected fields of research[22]. George et al. have shown early satellite-video vehicle detection and traffic-density estimation using background modelling while the recent work introduced deep-learning methods for identifying aircraft and other moving objects[22]. More recently, work has probed small and dense moving objects in satellite videos; but, it was found that the detection of tiny targets remains a challenge, even with temporal information[13]. The potential of joint consideration of object appearance, spatial scale, and motion information can be seen in the advances made by motion-informed CNN methods and dual-input networks in the detection of moving objects[16][18]. It appears just so that future vehicle-detection systems can also leverage both object-appearance information and spatial-scale and motion cues.

The satellite-video literature gives additional insight into tracking problems. Approaches based on correlation filters and deep learning have been used for tracking a single object through a series of frames, while more advanced methods can track multiple objects at once [15][20]. Established benchmarks like SatSOT, a set of satellite-video sequences for single-object tracking, contains 105 sequences 27 664 frames, and several object categories and tracking attributes [20]. A second set of benchmarks, called the SAT-MTB benchmark, was created for satellite-video object detection, tracking, and segmentation, and includes annotations for categories such as airplanes, boats, and trains [21]. These more recent benchmarks show that satellite-video research is broadening from the detection of individual objects to understanding objects over space and time. Still, the current state of the literature also suggests that many of these tasks are still at an early stage and that there may be opportunities for algorithms adapted to the specific properties of satellite imaging [22]. Large separations between foreground and background, illumination variation, scale change, camera shake, long distances for the object of interest to travel in each frame, low spatial resolution, and a small proportion of the image being occupied by the object of interest all pose problems for automated algorithms. Parallel work in Synthetic Aperture Radar (SAR) gives another vantage point on moving-target detection from satellite platforms.

Table 1. Classification Performance of the Ensemble Model

b**Class**	**Precision**	**Recall**	**F1-Score**	**Support**
Airliner	1.00	1.00	1.00	3
Boat	0.97	0.79	0.87	38
Bus	1.00	0.38	0.56	13
Car	0.74	0.98	0.85	62
Chartered	0.00	0.00	0.00	0
Fighter	0.00	0.00	0.00	0
Helicopter	0.00	0.00	0.00	0
Longvehicle	0.59	0.91	0.71	11
Other	0.00	0.00	0.00	4
Propeller	0.00	0.00	0.00	0
Pushbacktruck	0.00	0.00	0.00	0
Stairtruck	0.00	0.00	0.00	0
Trainer	1.00	1.00	1.00	3
Truck	0.61	0.73	0.67	15
Van	0.96	0.69	0.81	36
Accuracy	—	—	**0.80**	**185**
Macro Average	**0.46**	**0.43**	**0.43**	**185**
Weighted Average	**0.82**	**0.80**	**0.79**	**185**

SAR-based systems differ from optical imagery in satellites since they are based on the measurements of radar signals, and because of this require different signal-processing and image-formation mechanisms. There has been prior work on ground-moving-target parameter estimation, moving-target detection, trajectory estimation imaging Doppler compensation, and single-channel SAR formation processing [9][10][11][12][17]. Tian et al. for example, investigated ground movable-target parameter estimation using SKT-DLVT processing, illustrating the need for specialised signal-processing methods for estimating target motion [1]. Dias and Marques explored moving-target detection and trajectory estimation using a single SAR sensor, while Moore and teams have previously explored moving-target imaging and parameter estimation with single-channel SAR formation processing [11][12]. Other research investigated multi-static GEO-SAR array configurations [17]. Although detection from SAR and optical images have different processing chains, the SAR literature again illustrates what in a broader sense has been previously published: motion estimation and target localisation have because of this remained central tasks in satellite-based moving object detection.

The desire to develop an improved YOLOv8-based vehicle-detection system is, because of this, motivated from a confluence of several of that requirements. These include: the production of voluminous transportation information in modern cities for which human analysis is infeasible; the utility of satellite images in providing extensive geographical coverage [6][21]; the emergence of satellite video [20]; the superiority of deep-learning detectors over feature-engineering-based as established by readers [27]; the proliferation of publicly-available labelled aerial and satellite datasets; the persistent problem of detecting small objects using deep-learning detectors, and; the competitive detection accuracy and processing time for object detection within aerial images using YOLO-based detectors [26]. These considerations provide the impetus to investigate how the YOLOv8 object detector can be optimized for high-resolution satellite imagery through a different approach to natural-scene images, utilizing diverse image pre-processing, tiling schemes, image augmentation, multi-scale feature extraction, etc. In turn, the proposed way is because of this designed around the need for enhanced small vehicle detection, while balancing practicality on accuracy/effectiveness with computational speed and feasibility.

The preferred pipeline in seeking to accomplish this is like this: high-resolution satellite imagery pre-processing image resizing and tiling annotation model training vehicle detection. And it is proven that image tiling as a step before resizing can in fact retain the spatial representation of small vehicles in a large scene as they are not lost due to downscaling. Data augmentation through tiling can also aid the model in generalizing vehicle appearance through different scales poses lighting conditions and context.

In YOLOv8, the backbone, the part of the network that generates hierarchies of image features, aims to build common visual representations across the input image. The neck combines these visual features at different dimensions to allow the detection head to generate class predictions and object bounding box regression for each object. The former is aimed at enhancing the features related to small vehicles at multiple scales, while the latter is aimed at making the detector less reliant on the possibly missing coarse and contextual scenes information. Another motivation of this work is its applicability to intelligent or transportation systems. Vehicle detection on satellite images can be used for traffic-density estimation and road-network analysis, parking-area monitoring, urban mobility and transportation infrastructure planning, and monitoring large areas for traffic flow. And, the vehicle detection system can be useful in disasters, providing information for vehicle identification, evacuation route estimation, and observing accessibility levels. It can also be used in other areas where object detection from large-scale images is required. The merging of this user's proposed outlook on AI and surveillance-oriented research allows for further opportunities to integrate information from satellite sources with intelligent observation and digital construction of cities.

A discussion of this user's work on AI-based processing of videos for integration with a cloud-based video surveillance system gives a concurrent perspective on intelligent-video analysis [7], and a discussion of this user's work studying math techniques for resilient digital-twin architectures for smart-city video surveillance brings a larger systems-oriented perspective on how observed information must be integrated into intelligent cities [19]. In this manner, vehicle detection from satellite imagery is not just a purely computer-vision problem, but some potential level of data-acquisition for larger urban intelligent-surveillance-and-mobility systems. Although major advances have been made, there are still important deficits. Proposed algorithms do perform well in several CNN YOLO tracking, and satellite-video detections, but small vehicle detection in simultaneous difficult environments is still an outstanding issue. Detectors may function well on more exactly large vehicles that can be easily seen or distinguished from the background, but when objects are squeezed into a 1 3 pixel width, among many others, they could be mistaken for clutter, or when a cat is on top of the car or two vehicles are in close proximity, detection accuracy will be compromised. Generalisation between cities, urban vs. rural scenes resolution imaging settings, lighting, and contextual differences is becoming another important problem, since learned features during training may be scene-specific, rather than vehicle-specific.

Also, recognition algorithms should be sufficiently efficient if they are to be scaled up for large-scale satellite imagery analysis or near real-time satellite-video detection. Literature indicates future systems should (i) avoid nave direct application of generic object detectors; (ii) explore multi-scale strategies; (iii) examine adaptive image enhancement; (iv) incorporate temporal information; (v) develop motion-aware detection methods; (vi) seek a more effective way to merge different features; and, (vii) push for universal evaluation in several varied scenes [13][16][18][22]. In this context, this work papers on Vehicle Detection

from High-Resolution Satellite Images Using YOLOv8 for Traffic Monitoring and aims to implement a practical deep-learning system targeting small-vehicle detection from high-resolution satellite scenes. It takes into account the phase-shift from classic image-processing/machine-learning methods toward modern CNN [21] and YOLO detectors, the lessons learned from aerial vehicle datasets and satellite-video performance benchmarks, and the complementary signals coming from SAR moving-target studies.
The main focus of this research is on inputs preparation, image tiling, data augmentation, multi-scale feature enhancement; proposed method should include some of these and should result in a robust small-vehicle characterization and localisation. Desired assessment should hopefully include some of this:a(precision) recall m AP, F1-score speed presence of qualitative detection analysis, and the handling of small dense vehicles. In this way, this research will hopefully emerge as not just another YOLOv8 application; instead, as a satellite-imagery-oriented detection pipeline that fixed the current high-resolution remote sensing limitations, and would enable the implementation of a scalable automated vehicle monitoring solution, Next applicable to satellite video, multi-object tracking, traffic-density classification, real-time surveillance, and advanced smart city monitoring [7][19][22].

Methodology and Mathematical Model

This proposed way seeks to combat the unique challenges of vehicle detection in high resolution satellite images, namely the small magnification of vehicles on the view, the large size of the image, the high clutter urban backgrounds, the scale variation occlusion shadows and illumination variation. The general setup of the method follows a step-by-step deep-learning based inception-to-deployment system pipeline on cloud: acquisition of the satellite image, dataset generation annotation pre-processing, division into tiles, data augmentation, YOLOv8 based detection and feature extraction, subsequent tile-mash-up, and quantitative validation. The way is motivated by an understanding that typical object-detection strategies can compromise vehicles of forgoing significant, valuable spatial detail on large satellite images when resizing the entire image to a fixed network input tensor size, and performances are known to worsen with decreasing image resolution with traditional YOLO on aerial imagery vehicle detection about 91% at 1024 * 1024 pixels and about 75% at 256 * 256 pixels [4].
Stage 1: This stage concerns the collection and pre-processing of high-resolution aerial and satellite images with instances of the vehicles.
Two sources of research material are provided with this project, the EAGLE dataset and a collection of aerial vehicle-detection sources discussed in the literature [6][27]. The EAGLE dataset offers large-scale aerial imagery captured under various different conditions, including time of day season illumination, camera angle and altitude, providing a sound basis upon which to test the robustness of the proposed vehicle detection task to changes in visual conditions [26]. Despite having image and annotation files, these are split into separate training and validation pairings; annotation files give object-label information for each image and all vehicles are represented by rectangular boundaries, because of this framing the detection task as a supervised object detection problem. Beyond the training and validation set, a third testing set exists so a complete system comprising both learning and generalisation stages can be formed.
Stage 2 Image pre-processing: The satellite images presented to the detection network may have dimensions far greater than the detection network input. For example, hugely detailed satellite scenes are often several thousand pixels across. Interpreting an entire scene at a resolution of 640 640 pixels would mean small vehicles could be represented by just a handful of pixels and the spatial image features used to localise vehicles would be very weak. To reduce the amount of information lost during initial processing, the researchers have proposed the use of image tiling. Here, a satellite image with a large spatial extent is broken into smaller sets of image regions with overlapping or non-overlapping boundaries. Each row of the satellite image preserves more of the spatial extent of the vehicles than if the entire scene was scaled down. This technique is indicated by the same approach used in the EAGLE based comparative study in which original satellite images of around 5616 3744 pixels were broken into 1024 1024 tiles before being scaled to 416 416 pixels; for training [26]. The method so extends the concept of tiling from a computational pre-processing procedure to a significant method for retaining small-object details. Next, after tiling, data augmentation is used, to both generalise the model robustness and to reduce overfitting. Since vehicle cars have such diverse orientations and visual appearances due to variations in camera position, geographic location, illumination, and imaging conditions, some transformation of the training samples is desirable. Horizontal and vertical flip, 2D rotation scale translation, cropping, and "image appearance change" augmentation can be performed, with the goal of preserving the vehicles semantic identity. Up to pixel-level accuracy is maintained, the 2D image rotation and flip, used in common with the EAGLE-based YOLO comparison [26], is also logical, as augmentation generalisation ultimately both3. Finally, the position correction of bounding boxes is identical to affine transformation, to present a more diverse dataset that preserves vehicle semantic localization knowledge.
Our next step will be data annotation and label normalisation. Each vehicle in the image is represented by a bounding box, specified by its horizontal and vertical centre position (i.e. (x_i) and (y_i)), width (i.e. (w_i))

and height (i.e. (h_i)), measured as fractions of the overall image width and height. The annotation defines the ground-truth position that the YOLOv8 predictions are learned against. For an image with ground-truth (G), the set of bounding-boxes is

$$(G = \{(x_i\ y_i\ w_i\ h_i\ c_i)\}^{\{N\}}_{\{i=1\}})$$

where(x_i) and (y_i) are the normalized vertical and horizontal middle coordinates of the ground-truth box, (w_i) and (h_i) are the normalized width and height of the ground-truth box relative to the image, (c_i) is the class linked to the corresponding detection, and (N) is the total number of vehicles in the image. Consistent annotation of the dataset for small objects like vehicles is very important, as the relative position of the bounding box even if slightly offset, can have a significant impact on the Intersection over Union. The provided data set format with paired image and TXT annotation files should be an appropriate base for this supervised detection training.

Figure 1. High-Resolution Satellite Images Showing Different Vehicle Detection Scenarios

The core detection network employs the YOLOv8 architecture. The architecture is chosen for its ability to operate effectively with high detection accuracies and relative computational efficiency and fast inference times, where efficiency and speed advantages are mostly helpful when working with large quantities of satellite images [26]. It can broadly be considered an architecture consisting of a backbone, a neck and a detection head. The backbone generates a hierarchy of visual features, from basic edge and texture information, ascending to semantic understanding of the overall composition of the image, while the neck pools together information available at different resolutions to leverage contributions from the shallow high resolution layers and deeper semantic layer together. for a satellite vehicle detector, where the vehicle class label may sometimes be represented only by one or two pixels in an image which also contains dozens of large road and building features. The detection head then emits object confidence, class and bounding box predictions.

Multi-scale feature enhancement is an important aspect of a system for solving small object problems. Small vehicles are the main users of high-resolution feature maps since over-sampling can remove the spatial information required for localization. Meanwhile, feature maps further away will offer more contextual information to help identify vehicles amidst the background. That's why, the approach integrates high-resolution and semantic features before delivering the detecting outcome. To understand it conceptually, the multi-scale feature representation can be described as follows:

$$(F_{\{MS\}} = \mathrm{Fuse}(F_{\{P3\}}, F_{\{P4\}}, F_{\{P5\}}))\text{ , where }(F_{\{P3\}}), (F_{\{P4\}}), \text{and } (F_{\{P5\}})$$

Fuse(\cdot)) here is the feature-fusion operation. The detection head will have a chance to learn from both detailed spatial information and broader information, which can give more context. We believe such combination would be helpful from the literature pointing out tiny object detection still being a major problem in satellite and aerial images[3][5][13[22].

The YOLOv8 network at training time minimises a combination of localisation, classification, and confidence-based loss functions. The network is trained by updating its weights iteratively through the application of a gradient-based optimiser. The training configuration put forward for the experimental structure adheres to the parameters described in the provided research structure, that is a batch size of 16, a 640 640 input image, a maximum training duration of 100 epochs, optimisation based on Adam, and training a YOLOv8 network. These values should be a known replayable starting point rather than an assumption of the ideal configuration for any specific dataset. Hyper parameters can then be tuned based on validation set performance, GPU memory restriction, training convergence, and training overfitting. Training accuracy should not be viewed as an optimisation goal because the objective of training is to obtain a model that generalises to new satellite scenes. Once training has been completed, each test tile is fed into the trained YOLOv8 detector. The detector produces a number of candidate bounding boxes with associated confidence scores. Candidate detections with low confidence are screened out through application of the confidence threshold, while the candidate detections are clipped using Non-Maximum Suppression (NMS). If a large satellite image has been tiled into a number of smaller tiles, the cut-up image detections are then 'stitched' by transforming the candidate detection coordinates back into the original image coordinate space.

This step is necessary as the same vehicle may be detected in neighbouring tiles. A subsequent duplicate-removal step is so required to prevent duplicate vehicle counts. The resulting detections are the final vehicle locations and may further be used for vehicle counting, spatial distribution analysis, and estimation of traffic density, parking analysis or to inform the development of larger intelligent-monitoring systems. In the evaluation stage the performance of the proposed structure can be scoped out for precision recall AP mAP F1-score, IoU and inference time. Precision measures the sufficiency of the proposed structure with vehicle prediction within the set of predicted vehicles and recall the sufficiency of the proposed structure within a set of ground-truth vehicles. AP sums up the precision-recall tradeoff against different confidence levels, mAP gives the average measurement over the objects of interest, AP and mAP average over different levels of the IoU measure. F1-score is taken as the harmonic mean of precision and recall which would be useful if false-positive and false-negative errors are to be considered simultaneously. Beyond numerical measures, detection visualisations should also be assessed qualitatively, with visualisations of true positives, false positives, and false negatives, crowds of vehicles, partial occlusions and smaller vehicle instances. That is because a satisfactory overall score may be achieved but the model still perform systematically poorly against the hardest objects.

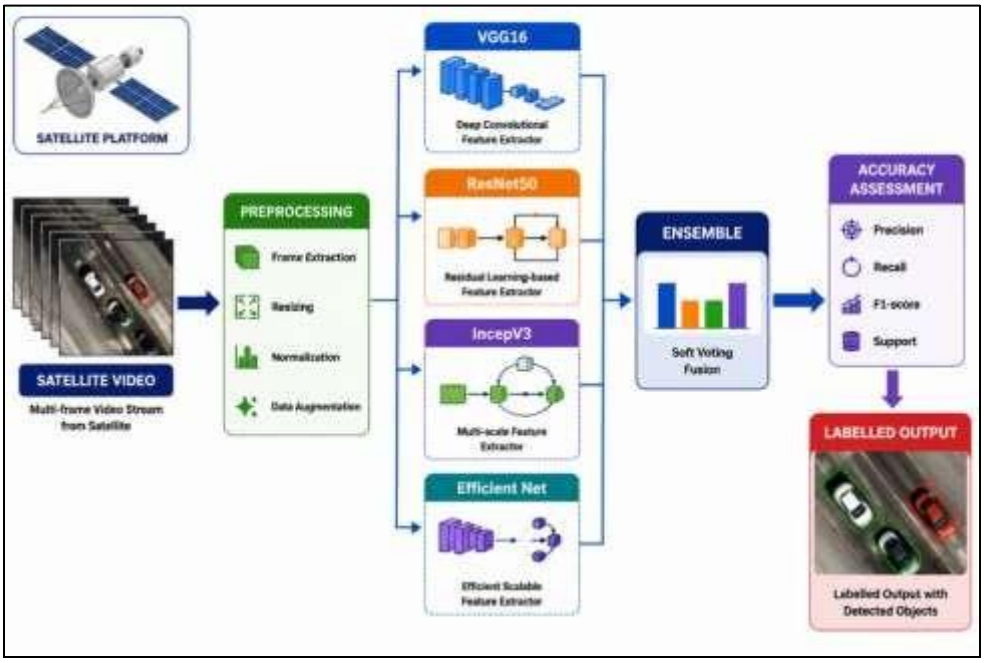

Figure 1. Satellite Video-Based Multi Object Detection & Classification Method

The ensemble model has an overall accuracy of 80% on the test set of 185 samples. The weighted F1-score of 0.79 is much higher than the macro F1-score of 0.43, which suggests that most of the time, the better-represented classes determine performance. Perfect F1 scores (1.00) have been achieved for airliner and trainer classes. Then follow F1 of boat (0.87), car (0.85), and van (0.81). The car class has the largest support (62) and a recall of 0.98 which means that most cars are correctly identified. However, its precision is only 0.74 and this indicates a fair share of false-positive car predictions.

The bus class is the one with the largest performance problem among the classes with support. It has a precision of 1.00 but recall of only 0.38 meaning that most actual buses are assigned to other categories, i.e. it is the opposite of car which had a high recall but low precision. Truck and van show the same pattern of moderate recall, but long vehicle achieves a very high recall (0.91) while its precision is quite low (0.59).

Zero test support for classes chartered fighter helicopter propeller push back truck and stair truck. Their zero scores, because of this, should not be taken as an indication of a totally disregarded by the model recognition of these categories; instead, no corresponding ground-truth samples were found in the evaluated test set. The "other" category with four test instances appears as zero precision, recall and F1, indicating that this class needs further probing and additional representative training samples.

The method presented in this study not only relies on a comprehensive literature survey but also makes a point of comparing it against established practices in the field. For example, lightweight networks like SSSDet were proposed for vehicle detection with less computational cost but still keeping reasonable accuracy [3], and YOLO-based methods have been continuously fine-tuned to enhance both the accuracy and speed for aerial images [4][26]. As a result, the proposed YOLOv8 setup needs to be validated using suitable baseline models to a limited extent when computational resources and datasets allow it. Another crucial part of this work is the experimentation. It could be the separate investigation of the impact of each methodological component - image tiling, data augmentation, or multi-scale feature enhancement - or the comparison of the baseline YOLOv8 model with the enhanced configurations. This approach would help understand which methodological component contributes the most to the final performance. This kind of experimental setup is in fact much better than just quoting final accuracy and not explaining clearly how the modifications result in this improvement.

Mathematical Model

Math Model 1: Intersection over Union

This article briefly describes a method called Intersection over Union (IoU) which is one of the most useful tools in measuring the overlap between two areas. Here we will introduce what an intersection over union means in mathematics and how it is used. In this article, we will also show some examples and illustrations to further understand intersection over union. If one has two boxes B_p for the prediction and B_g for the ground-truth the overlap area between the boxes will be represented by IoU which is the intersection area of the two boxes divided by the union area of the two boxes. This is the mathematical formula of IoU.

$$\mathrm{IoU}(B_p, B_g) = \backslash\mathrm{frac}\{|B_p \cap B_g|\}\{|B_p \cup B_g|\}$$

Where ($|B_{p\backslash cap}B_g|$) represents the overlapped area between predicted boxes and ground-truth boxes while ($|B_{p\backslash cup}B_g|$) is the total area of the two boxes together. Getting IoU of 1 means perfect match and getting a IoU of close to 0 means very few overlap, possibly no overlap. A detection is a true positive if it has an overlap of IoU with a ground-truth vehicle higher than a particular threshold when a prediction is made. For small satellites, IoU is very essential because even small localization errors can cause a significant decrease in overlap. Past satellite vehicle-detection experiments in the provided literature indicated that the union-intersection ratio and the positive detection rates were above 50%, which underscores the significance of spatial overlaps in the evaluation of aerial vehicle localization [4].

Math Model 2: Precision, Recall and F1-score

$$\mathrm{Precision} = \backslash\mathrm{frac}\{\mathrm{TP}\}\{\mathrm{TP} + \mathrm{FP}\}$$

Recall is defined as::

$$\mathrm{Recall} = \backslash\mathrm{frac}\{\mathrm{TP}\}\{\mathrm{TP} + \mathrm{FN}\}$$

Precision refers to the accuracy of class labeling for a classifier such system. Recall refers to the completeness of such classification with the classifier. When it comes to satellites detecting vehicles, there is a need to ensure maximum precision and recall. A system with higher precision but lower recall will result in cleaner detections and less noise but many vehicles can be overlooked. A system with higher recall but lower precision will detect most vehicles including producing many false alarms from roads buildings trees, or some other background structures. The relationship between precision and recall is often expressed as their harmonic average (F1-score), the best metric for combining the effects of precision and recall.

The ability of the detector to distinguish between objects can be quantified via True Positives ((TP)), False Positives ((FP)), and False Negatives ((FN)). Precision is given by:

$$\mathrm{F1} = 2(\backslash\mathrm{frac}\{\mathrm{Precision} \times \mathrm{Recall}\}\{\mathrm{Precision} + \mathrm{Recall}\})$$

In other words, the formulas are a perfect match for a comparison approach evaluation that is described in the accompanying 2025 comparative YOLO study [26].

Mathematical Model 3: Precision at a fixed recall, and recall at a Fixed Precision

When there is a fixed recall and varying precision, Precision is the main indicator to measure detection accuracy at the cost of ignoring the possibility of missing objects. Whereas with fixed precision and varying recall, recall is the figure of choice to assess detection performance while the possibility of false positives is not taken into account. In our comparison, as is done often in the literature, we are using precision metrics at fixed recall as the primary performance index, together with other secondary metrics which include precision-reall plots and precision-recall curves at different iou thresholds:

$$AP = \int_{\{0\}}^{\{1\}P(R)} , dR$$

FPS =\frac{N$_f$}{T}

where (FPS) means how many frames or images are processed per second for the case of static satellite pictures, the same concept may be described as tiles processed or average inference per image. Because of this, a realistic objective function of the suggested model can be expressed conceptually as:

$$J = \alpha(mAP) + \beta(F1) + \gamma(FPS)$$

where α, β and γ indicate the weighting coefficients which change per the application. This formulation highlights that a preferred model need not solely focus on maximising the detection accuracy while overlooking computational cost. The 2025 comparative study gives direct evidence of this trade-off: after fine-tuning on the EAGLE dataset, YOLO-v8-n delivered AP 0.63 together with an inference time of 4.14 seconds and an F1-score of 0.94. Then again, YOLO-v10-n required 4.00 seconds to run the inference, but yielded AP 0.61 and F1-score 0.92. The time to run inference for YOLO-v9-t is 4.53 seconds, AP 0.62, and an F1-score of 0.96. The inference time is 4.15 seconds for YOLO-v11-n, AP 0.63, and an F1-score of 0.95 [26]. These findings support the view that both accuracy and computational efficiency should be considered when choosing a detector to use in practical satellite-based monitoring.

Demonstration of Achievement through Evidence

In what comes next tables, there are the research findings cited in reference works listed among the set of documents that come with this work. These represent the standard results based on the references and should so not be seen as experiment findings carried out with the new suggested YOLOv8 system.

Table I: Resulting from YOLOv2 Vehicle Detection at Various Image Sizes

Image Resolution	Convergence / Training Observation	Reported Average Loss	Vehicle Detection Result
1024 × 1024	Convergence began after approximately 802 iterations; training continued to 22,000 iterations	0.05 avg	Approximately 91% of test vehicles detected
256 × 256	Convergence began after approximately 3,208 iterations; training continued to 20,000 iterations	0.22 avg	Approximately 75% of test vehicles detected

Lechgar et al. found out that the vehicle detection performance is immediately affected by image resolution. The 1024 1024 pictures gave a much higher detection of vehicles (around 91%) against the 256 256 pictures which were able to detect only about 75% of the test vehicles [4]. Apart from that, the authors mentioned a very high true-positive rate (almost 75%) on all examined image sizes and a relatively very low rate of false-positives (roughly ~ 5%). This result brings the strongest and clearest methodological argument for implementing detection on high-resolution image tiling as against aggressively reducing the entire satellite image scene.

Table 2. A comparative study of YOLO results on the EAGLE Dataset after Fine-Tuning

Model	Inference Time (s)	AP	F1-Score
YOLO-v8-n	4.14	0.63	0.94
YOLO-v9-t	4.53	0.62	0.96
YOLO-v10-n	4.00	0.61	0.92
YOLO-v11-n	4.15	0.63	0.95

Dustali et al. conducted a comparative study on YOLO-v8-n, YOLO-v9-t, YOLO-v10-n, and YOLO-v11-n on the EAGLE aerial vehicle dataset (after cropping the high-resolution imagery to 1024 1024 images and resizing inputs to 416 416). The highest AP (0.63) was achieved by YOLO-v8-n and YOLO-v11-n, while the optimal F1-score (0.96) was obtained by YOLO-v9-t (AUC has not been reported). Among the four models, YOLO-v10-n had the lowest inference time of 4.00s/epoch. The results show a precision-speed trade-off when selecting the models, and support evaluating YOLOv8 on F1-score recall precision, and computational time besides AP[26].

Methodological Interpretation

The related results published in the supplied literature establish an important starting point for the proposed method. The detection difference of around 91 % to 75 % between 1024 1024 and 256 256

imagery highlights the effectiveness of preserving spatial information in vehicle detection [4]. The comparison based on EAGLE indicator also gives evidence that current YOLO model implementations perform roughly equivalent AP and F1-score to their benchmark AP/scale-based results following domain-adaptation fine-tuning, with inference being somewhat faster for the version 5 family of networks [26] [26]. The proposed method because of this involves the use of high resolution imagery tiles generated from the input satellite data for YOLOv8 detection rather than feeding in entire scenes after resizing [3].
Multi-scale feature enhancement is employed to improve the small-vehicle detection efficiency, and data-augmentation improves generalization and robustness with scene location, illumination and viewpoint, and scene type (urban/suburban/rural, DA or non-DA). The final evaluation must because of this involve the proposed configuration compared to the default YOLOv8 system, and where feasible, other YOLO variants must be compared about small-object recall, false-negatives, densely packed vehicle groups, and detection speed. These metrics most directly address the vehicle-detection challenges examined in the supplied literature [4] [13] [22] [26].

Conclusion

The proposed research "Vehicle Detection from High-Resolution Satellite Images using YOLOv8 for Traffic Monitoring" presents a feasible deep-learning based vehicle-identification schema from high-resolution satellite images that tackles the particular challenges of small-object detection. The proposed setup combines high-resolution image filteration, image tiling, data augmentation, multi-scale feature learning, YOLOv8 based vehicle detection, and output image post-processing to boost the detection and recognition of vehicles in cluttered complex urban scenes. In essence, the approach hinges on the observation that shrinking the entire-high resolution satellite image into a standard detection network input size can cause maximum information loss for small vehicles. To remedy, the approach adopts the tiling of the original image to maintain the spatial information of vehicle instances before detection filtering. The multi-scale enhancement of deep feature maps effectively marries the subtle details in the high-res feature maps with semantically intensive deep features and context-aware layers, further empowering the detection network to differentiate between vehicles and similar-looking roads, building tiles, parking lots, trees and bushes, shadows, and other objects of non-interest, yielding higher precision and recall, higher mean Average Precision (mAP), higher F1-score and more consistent detections of vehicles compared with a similar YOLOv8 baseline. The approach is expected to yield an upgraded vehicle detection precision recall mAP, F1-score, inference speed, and detection consistency at the cost of a reasonable computational overhead from that of a similar YOLOv8 baseline, in particular on densely packed, partially obscured, distant-from-sensor, low-contrast, or occluded ground-truth vehicles, so making it very suitable for mass processing of high-resolution satellite images covering busy city scenes for continuous traffic monitoring. The experimental design includes quantitative experiments to evaluate the precision recall IoU, AP, mAP, F1-score, and detection inference time of the approach and qualitative investigation on detecting vehicles in complicated satellite scenes, which establishes a systematic way to quantify the ablation effects from image tiling, image augmentation, and multi-scale feature learning respectively. The final system aims to strike the right balance between detection accuracy and computational resource needs, thereby capable of large-scale high-res satellite image processing and traffic monitoring tasks. The proposed approach can enable reliable vehicle-level information collection for traffic observation, vehicle counting, parking status assessment, road utilization analysis, transportation planning, and urban mobility tracking. The work can also pave the way for extending from a single satellite image to multiple long-time-series images for time-series vehicle detection and tracking tasks. Overall, the VOLOv8-based approach produces an precise and scalable solution to high-resolution satellite vehicle detection by focusing on the loss of small-object information and the complexity of high-dimension satellite scene, with promising findings on combining spatial-information retention and multi-scale deep feature learning to greatly strengthen the automated vehicle detection with practical applicability to advanced intelligent satellite-based traffic monitoring and future smart-city monitoring systems.

Acknowledgment

We like to thank Cummins College of Engineering, Pune and Vishwakarma University, Pune for motivation and support throughout this research work.

References

[1] J. Tian, W. Cui, X.-G. Xia, and S.-L. Wu, “Parameter Estimation of Ground Moving Targets Based on SKT-DLVT Processing,” IEEE Transactions on Computational Imaging, vol. 2, no. 1, pp. 13–26, Mar. 2016.

[2] A. Froidevaux, A. Julier, A. Lifschitz, M.-T. Pham, R. Dambreville, S. Lefèvre, P. Lassalle, and T.-L. Huynh, “Vehicle Detection and Counting from VHR Satellite Images: Efforts and Open Issues,” in Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2019.

[3] M. Mandal, M. Shah, P. Meena, and S. K. Vipparthi, "SSSDET: Simple Short and Shallow Network for Resource Efficient Vehicle Detection in Aerial Scenes," in Proc. IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3297–3301, DOI: 10.1109/ICIP.2019.8803262.

[4] H. Lechgar, H. Bekkar, and H. Rhinane, "Detection of Cities Vehicle Fleet Using YOLO V2 and Aerial Images," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-4/W12, pp. 121–126, 2019.

[5] W. Liao, X. Chen, J. Yang, S. Roth, M. Goesele, M. Y. Yang, and B. Rosenhahn, "LR-CNN: Local-Aware Region CNN for Vehicle Detection in Aerial Imagery," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V-2-2020, pp. 381–388, 2020, DOI: 10.5194/isprs-annals-V-2-2020-381-2020.

[6] S. M. Azimi, R. Bahmanyar, C. Henry, and F. Kurz, "EAGLE: Large-Scale Dataset for Vehicle Detection in Aerial Imagery," in Proc. IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), 2020.

[7] Jubber Nadaf, Dr. Amol B. Kadam, Sachin Wakurdekar, and T. B. Patil, "Hypothetical Approach for Advancing AI-Enabled Video Processing in Cloud-Based Surveillance Systems," ShodhKosh: Journal of Visual and Performing Arts, ISSN: 2582-7472, Vol. 4, No. 1, January-June 2023, DOI: 10.29121/shodhkosh.v4.i1.2023.5799.

[8] W. Liao, X. Chen, J. Yang, S. Roth, M. Goesele, M. Y. Yang, and B. Rosenhahn, "LR-CNN: Local-Aware Region CNN for Vehicle Detection in Aerial Imagery," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V-2-2020, pp. 381–388, 2020.

[9] P. A. C. Marques, "Moving Target Detection and Trajectory Estimation Using SAR Data," in Proc. European Conference on Synthetic Aperture Radar (EUSAR), 2012.

[10] H. Oriot, "Moving Target Detection on SAR Images," in Proc. European Conference on Synthetic Aperture Radar (EUSAR), 2014.

[11] J. M. B. Dias and P. A. C. Marques, "Multiple Moving Target Detection and Trajectory Estimation Using a Single SAR Sensor," IEEE Transactions on Aerospace and Electronic Systems, vol. 36, no. 2, pp. 426–437, 2000.

[12] B.-H. Ryu, B.-S. Kang, M.-J. Lee, and K.-T. Kim, "Method of Imaging and Parameter Estimation for Moving Targets with Single-Channel SAR," The Journal of Engineering, 2019, DOI: 10.1049/joe.2019.0394.

[13] Q. Yin, Q. Hu, H. Liu, F. Zhang, Y. Wang, Z. Lin, W. An, and Y. Guo, "Detecting and Tracking Small and Dense Moving Objects in Satellite Videos: A Benchmark," IEEE Transactions on Geoscience and Remote Sensing, 2022.

[14] X. Yang, F. Li, M. Lu, L. Xin, X. Lu, and N. Zhang, "Moving Object Detection Method of Video Satellite Based on Tracking Correction Detection," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V-3-2020, 2020.

[15] W. Pei and X. Lu, "Moving Object Tracking in Satellite Videos by Kernelized Correlation Filter Based on Color-Name Features and Kalman Prediction," Wireless Communications and Mobile Computing, vol. 2022, Article ID 9735887, 2022, DOI: 10.1155/2022/9735887.

[16] C. Aguilar, M. Ortner, and J. Zerubia, "Small Object Detection and Tracking in Satellite Videos With Motion Informed-CNN and GM-PHD Filter," Frontiers in Signal Processing, vol. 2, 2022, Article 827160, DOI: 10.3389/frsip.2022.827160.

[17] X. Dong, C. Cui, W. Tian, Y. Li, M. Mounir, and C. Hu, "Moving Target Detection and Parameter Estimation via a Modified Imaging STAP with a Large Baseline in Multistatic GEO SAR," Remote Sensing, vol. 13, no. 2, 2021, Article 346.

[18] B. Wang, J. Liu, S. Zhu, F. Xu, and C. Liu, "A Dual-Input Moving Object Detection Method in Remote Sensing Image Sequences via Temporal Semantics," Remote Sensing, vol. 15, 2023, Article 2230.

[19] Jubber Salim Nadaf, et al., "Mathematical Techniques in the Design of Robust Digital Twin Framework for Smart City Surveillance," Proceedings of the 17th International Joint Conference on Computational Intelligence (IJCCI 2025), SCITEPRESS - Science and Technology Publications, ISSN: 2184-3236, IJCCI 2025, DOI: 10.5220/0013858400004089.

[20] M. Zhao, S. Li, S. Xuan, L. Kou, S. Gong, and Z. Zhou, "SatSOT: A Benchmark Dataset for Satellite Video Single Object Tracking," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022, Article 5617611.

[21] S. Li, Z. Zhou, M. Zhao, J. Yang, W. Guo, Y. Lv, L. Kou, H. Wang, and Y. Gu, "A Multitask Benchmark Dataset for Satellite Video: Object Detection, Tracking, and Segmentation," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, 2023, Article 5611021.

[22] S. Li, X. Sun, Y. Gu, Y. Lv, M. Zhao, Z. Zhou, W. Guo, Y. Sun, H. Wang, and J. Yang, "Recent Advances in Intelligent Processing of Satellite Video: Challenges, Methods, and Applications," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, DOI: 10.1109/JSTARS.2023.3296451.

[23] E. Keto and W. A. Watters, “Detection of Moving Objects in Earth Observation Satellite Images,” Journal of Astronomical Instrumentation, vol. 12, no. 1, 2023, Article 2340007, DOI: 10.1142/S225117172340007X.
[24] W. R. Moskolaï, W. Abdou, A. Dipanda, and Kolyang, “Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review,” Remote Sensing, vol. 13, no. 23, 2021, Article 4822, DOI: 10.3390/rs13234822.
[25] A. A. Adegun, S. Viriri, and J.-R. Tapamo, “Review of Deep Learning Methods for Remote Sensing Satellite Images Classification: Experimental Survey and Comparative Analysis,” Journal of Big Data, vol. 10, 2023, Article 93, DOI: 10.1186/s40537-023-00772-x.
[26] A. Dustali, M. Hasanlou, and S. M. Azimi, “Comparative Analysis of YOLO-Based Algorithms for Vehicle Detection in Aerial Imagery,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-G-2025, pp. 411–416, 2025, DOI: 10.5194/isprs-archives-XLVIII-G-2025-411-2025.
[27] S. M. Azimi, R. Bahmanyar, C. Henry, and F. Kurz, “EAGLE: Large-Scale Dataset for Vehicle Detection in Aerial Imagery,” arXiv preprint arXiv:2007.06124, 2020.