Arab International Journal of Information Technology & Data
2024, No. 2 April - June 4Vol.

Building an Automatic System for Detecting Manifestations of Visual Pollution Using Geospatial Techniques and Deep Learning: An Applied Study on Abandoned Vehicles in Riyadh

Eman Rafe Alamri
Graduate Student
Department of Geography, King Saud University, Saudi Arabia
emand468@gmail.com

Prof. Mufareh Bin Daym Alquradi
Professor, Department of Geography,
King Saud University, Saudi Arabia
mufarehq@ksu.edu.sa

Dr. Duaa Hamad Alsaeed
Associate Professor, Department of Information Technology,
King Saud University, Saudi Arabia
dalsaeed@ksu.edu.sa

Abstract

Abandoned vehicles are one of the manifestations of visual pollution in Saudi cities. Some elements left by humans in urban landscape may negatively affect the view and disturb the functionality and enjoyment of a given area. Therefore, this study aimed at detecting abandoned vehicles in the city of Riyadh to reduce the manifestations of visual pollution. It sought to build an automated system to detect abandoned vehicles using Deep Learning (DL)- Based Detection Techniques in Remote Sensing (RS)Images. The study adopted an experimental approach, using Deep Learning (DL) Techniques-Based Object Detection methods. To achieve the objectives of the study, an automated system was built to detect abandoned vehicles from WorldView-3 satellite images. The automated system built adopts three-stage object detection methods, including: pre-processing the data set; training the deep learning model and refining it using Mask-RCNN algorithm; and finally classifying the detected vehicles into 'Abandoned' and 'Non-Abandoned' using intersection temporal images' analysis and matching methods. The study reached a number of results, the most important of which was that vehicle

Received: 28/12/2023 **Approved: 29/01/2024**

detector, Mask-RCNN, was able to detect vehicles in the images taken by WorldView-3 satellite with high level of accuracy (95%), recall (91%) and F1 score (93%). Moreover, the results demonstrated that the model reached a good score on the mAP@IoU scale (50-95%), amounting to 87%, which helped in tracking and identifying abandoned vehicles effectively using temporal images.

Keywords: Visual pollution; Abandoned Vehicles; Geospatial techniques; Deep learning; Object Detection; Mask-RCNN.

Introduction

Rapid urban development and the subsequent expansion of services is one of the main reasons behind recent spread of visual pollution manifestations in urban cities. Also, lack of awareness of the importance of preserving the environment is another reason for this significant increase in visual pollutants inside neighborhoods and on roadsides. However, in an attempt to limit the spread of these visual pollutants and to improve urban landscape, many countries around the world adopted a number of policies and enacted some strict laws for this purpose (Fuller et al., 2022)

The city of Riyadh was no exception. Over the past few years, it witnessed a huge increase in the number of abandoned vehicles, i.e. these are vehicles abandoned for a long period on road sides, in public places, or in vacant lands. They are usually abandoned by their owners because they are damaged or because they were stolen, causing environmental pollution and visual distortion of surrounding landscape. Furthermore, abandoned vehicles pose danger to residents due to several reasons, including: accumulation of waste and dust, their containment of dangerous materials such as gasoline and other hazardous liquids, in addition to attracting criminals and suspicious actions because they are unidentified or stolen. Therefore, adopting automated environmental monitoring systems is a proper solution towards mitigating this phenomenon.

Abandoned vehicles are considered one of the visual pollutants resulting from wrong practices, which require continuous monitoring and immediate removal. Riyadh Municipality stated that statistics of abandoned cars in the region, during the period from 2018 until 2022, signify that their number reached up to 70,195 vehicles. Therefore, it launched a campaign to remove as many of the abandoned vehicles as possible, while placing ‘Final Notice of Removal’ tickets on a number of vehicles across the city. In addition, the Council of Ministers announced adopting a number of controls for limiting the number of abandoned vehicles. The Council stated that there will be a grace period for taking necessary corrective measures before dropping these vehicles from their owners’ records. It also pointed out the need for the Ministry of Interior and the Ministry of Municipal and Rural Affairs and Housing to cooperate with Saudi Authority for Data and Artificial Intelligence to exchange information and data related to abandoned vehicles, which will inevitably create an automated link to control this phenomenon and mitigate it as much as possible.

However, supervision tours are one of the traditional methods used for monitoring abandoned vehicles. It relies, mainly, on residents' reports. Nonetheless, the number of residents' reports on abandoned cars varies from one neighborhood to another. This variation could be attributed to the extent of their awareness of the negative impacts of this phenomenon. Therefore, such reports are not sufficient to significantly limit the number of abandoned cars. Thus, it is necessary to adopt modern techniques in monitoring operations.

Recently, Machine Learning (ML) and Deep Learning (DL) techniques have facilitated various environmental monitoring processes. They have proven to be useful for solving many environmental problems; as training data and automated detection models are capable of easily and accurately identifying these environmental problems, which helped developing appropriate solutions. These techniques have also contributed to understanding environmental conditions and how to sustain them through taking the necessary measures and developing proper solutions.

Statement of Problem

Rapid urbanization and the huge increase in the number of services provided globally have led to a noticeable increase in human interaction with the environment (Bawareth, 2022). As a result, numerous cities around the world witnessed an increase in the spread of visual pollutants, including abandoned vehicles; as they are left on road sides and in public places for long periods of time; impairing the population's enjoyment of the view and lowering the aesthetic standards of the environment surrounding them. Perhaps, one of the causes leading to this problem is lack of cultural awareness and the insufficiency of monitoring tours required to address the manifestations of visual pollution. The government has recently adopted continuous automated monitoring systems; using sensors and video cameras to monitor this phenomenon. Nonetheless, despite using these advanced technological solutions, they failed to cover large areas of the urban landscape.

Study Significance

Abandoned vehicles are a form of visual pollution that is difficult to measure in urban areas (Chmielewski et al., 2015). Their spread and decline are linked to the level of culture and awareness of the population. The theoretical importance of this research lies in highlighting this phenomenon, its causes, and the extent of the danger of its spread, which will turn into an environmental and health problem. The practical importance of this research lies in proposing a modern mechanism for the automatic monitoring of abandoned vehicles using deep learning algorithms from geographic data.

Study Objectives

The study aims primarily to build an automated system for monitoring abandoned vehicles in remote sensing images using deep learning techniques. This will be done through the following sub-objectives:

1. Train a Mask-RCNN model to detect vehicles from World View-3 satellite images.
2. Evaluate and measure the performance of the vehicle detection model.
3. Build an algorithm to classify the detected vehicles as abandoned or not abandoned.
4. Evaluate the performance of the algorithm in classifying the detected vehicles as abandoned or not abandoned.

Given the difficulty of monitoring large areas of the environment using regular photography and video clips (Youme et al., 2021), the study attempted using remotely sensed data, which is characterized by its wide coverage and high accuracy. For achieving the objectives of the study, remotely sensed data were used to conduct the necessary analyzes and to build an automated system for monitoring vehicles using Deep Learning (DL)- Based Detection Techniques in Remote Sensing (RS) Images. The researcher sought to train the Mask-RCNN model to detect vehicles in World View-3 satellite imagery and building an algorithm to classify the detected vehicles into abandoned and non-abandoned vehicles.

Terminology

-Visual pollution: The distortion caused by any view that is uncomfortable to the human eye, and which causes the absence of aesthetic images of what surrounds it.

-Abandoned vehicles: These are vehicles that are fit for use and abandoned by their owners or stolen for a short or long period of time in public places and roads.

Study Area

The spatial framework of this study is limited to the geographical boundaries of the city of Riyadh, located in the center of the Kingdom of Saudi Arabia. The City of Riyadh lies between latitude 24°18' and 25°12' north; and longitude 46°65' and 46°03' east (See Figure 1). The city of Riyadh is one of the rapidly urbanizing cities; in terms of increasing number of population and services provided, which in turn has led to an increase in visual pollutants (Altuwaijri, 2017).

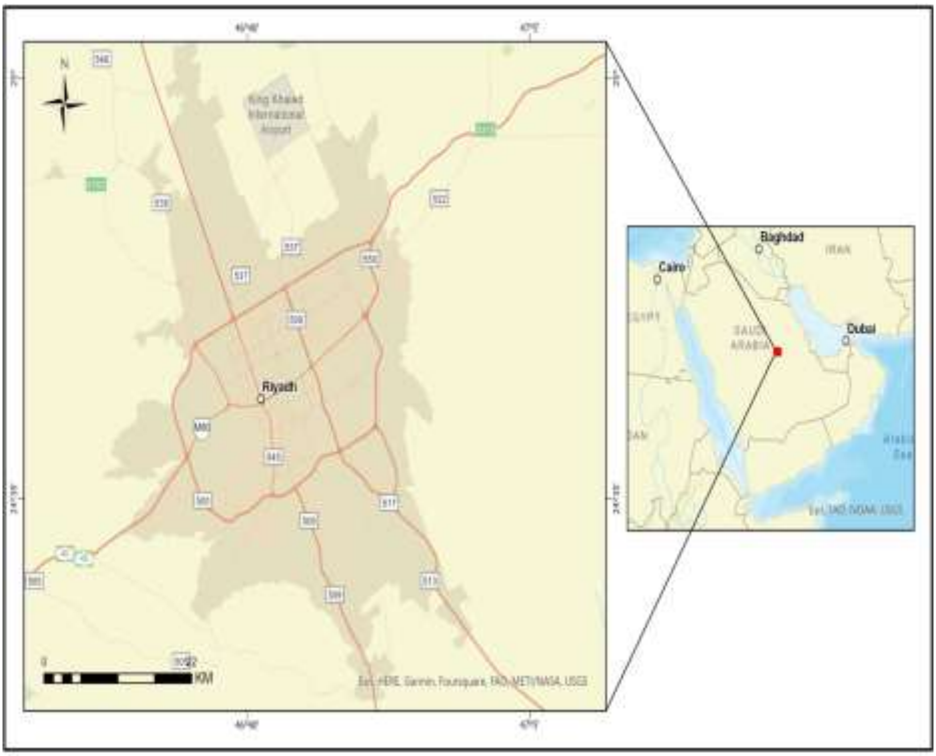

Figure (1): Riyadh City, Source: Esri

Visual Pollution

Visual pollution is one of the indicators of urbanization (Wakil et al., 2019). Many countries have adopted mechanisms and policies to control its sources. In order to achieve the sustainable development goals of improving the quality of life and the urban landscape in the natural environment, many countries have moved to reduce and mitigate this phenomenon to improve the quality of cities. However, this phenomenon still poses a major challenge due to the difficulty of measuring and evaluating it and the diversity of the general taste of the population.

Visual pollution is a relatively new concept that refers to spatial chaos and aesthetic distortion of place (Mohammed, 2023). It has been defined by (Adnan & Al, 2022) as a lack of harmony for the elements of the place or as the visual distortion of any disturbing view that a person sees in the place. The exposed wires, dilapidated buildings, damaged cars, and discarded waste form a disturbing sight for the eye and visual distortion. (Mansour et al., 2019) defined visual pollution as the unwanted change in environmental elements that destroys the aesthetic image of the natural landscape. This change includes the lack of coordination of building facades, the deterioration of sidewalks, potholes, waste, and damaged property, in addition to the spread of random commercial advertising. Likewise, (Al-Qatifi, 2022) pointed out that it represents the manifestations that violate public taste and affect the health of the city and man. It is clear from the previous definitions that visual pollution is caused by negative and distorted views of the environment that affect the quality of life and its type.

(Adnan & Al, 2022; Bawareth, 2022) pointed to the reasons for the emergence of forms of visual pollution in urban cities. These reasons include planning, economic, administrative, and behavioral reasons. Population growth and consumption are one of the most important of these reasons, which have contributed to the increase in violations such as construction waste, private property, waste, abandoned vehicles, and disabled machinery.

Human environment is an indicator of the quality of his life. Therefore, an environment free of visual pollution increases the quality of life and a sense of satisfaction (Voronych, 2013). Therefore, the topic of addressing visual pollution has attracted the attention of many researchers as attempts to reduce its effects. This study came to contribute to modern monitoring operations using geospatial technologies and deep learning.

Many studies have classified the manifestations of visual pollution into several categories such as abandoned vehicles, external advertising, exposed wires, construction waste, waste, demolished buildings, and inappropriate lighting, etc. (klei, 2019; Shrivastava & Choudhary, 2016; Wakil et al., 2019). The Ministry of Municipal and Rural Affairs and Housing in the Kingdom of Saudi Arabia classified the manifestations of visual pollution into thirty manifestations in order to improve the urban landscape and visual character in the cities and villages of the Kingdom of Saudi Arabia. In this study, of all visual pollutants, the focus was on abandoned vehicles as the most common in the city, which amounted to 70,195 abandoned vehicles according to statistics from the Riyadh Region Municipality from 2018 to 2022.

The Role of Modern Technologies in Monitoring and Controlling Aspects of Visual Pollution

Geospatial data and technologies play an effective role in alleviating the problems resulting from continuous urban growth and its negative effects (Al Hashmi, 2016). They also contribute to providing spatial solutions when planning and developing city’s life cycle through the development of urban modeling, site selection, and proper planning of facilities and infrastructure; which is a necessary step to improve urban landscape in modern cities. It is worth noting that integration of information and communications technology (ICT) and geospatial technologies play an important role in making cities smart and sustainable (Tao, 2013). Spatial information can be integrated into cities development plans through using supporting technologies, such as: networks and communications, the Internet, open data, big data, cloud computing, and community participation. The integration of geographic information systems and artificial intelligence, for example, has witnessed rapid developments that have contributed to assessing spatial issues in the environment and providing smart spatial solutions. Remote sensing (RS) and geographic information systems (GIS) also provide opportunities to monitor environmental risks associated with visual pollution to make appropriate decisions, as smart cities resort to using sensors for control and monitoring to facilitate daily life activities (Zakaria & A., 2015). The role of sensors is to provide real-time data on what is being monitored to ensure rapid intervention by governments and the concerned parties. Moreover, sensors have proven to be effective in measuring and predicting weather, traffic, and other environmental issues in the city.

Artificial intelligence (AI) is one of the branches of computer science that uses software and machines to solve problems and make decisions, mimicking human intelligence (Ongsulee, 2018). Deep learning (DL) is considered a sub-field of artificial intelligence, as it relies on algorithms that are able to learn from unstructured data and complex patterns such as speech and images (Erbe, 2021). Most deep learning models rely on an artificial neural network to transfer input data from one layer to another via communication channels using specific algorithms. These algorithms are trained on a large set of data that learns from models in order to solve problems independently without predictions.

With the rapid development of deep learning networks and GPU computing, the performance of object detectors has been greatly improved (Pal et al., 2021). Various applications and research have used computer vision and sensing methods, such as: Radar, LiDAR, and drones, to detect objects. Modern, high-quality, small size, and low-cost cameras have helped in the development of objects detection and tracking operations. Further, Graphics Processing Units (GPU) and Deep convolutional neural networks (CNNs) have led to rapid detection of objects.

Object detection technology refers to scanning and searching for objects of specific categories (such as: vehicles, buildings, and trees) in images or video (Jiao et al., 2019). Object detection models use satellite images to predict the correct category for each object in the image; placing a bounding box around the object to determine its geographical location. Object detection models can extract appearance-based features from images, i.e. focus on the visual characteristics of the objects detected, or motion-based, i.e. focus is on any change in location or movement of the objects detected. In addition, deep networks are used as a backbone in detection devices to extract features from input images and, then, analyze these features to classify objects. Objects are classified into their correct categories using a feature generation network, which is known as the basic network. Some of the most famous basic networks, used to detect objects clearly in research and applications, are: ResNet, AlexNet and VGG16.

Deep learning-based methods are divided into two categories of detectors, which include two-stage detectors and one-stage detectors, each with different characteristics (Jiao et al., 2019). Many two-stage detectors propose a set of approximate regions of objects using deep networks. These proposals are then refined and classified based on the features extracted from the proposed regions. Afterwards, bounding boxes are placed around these objects to increase detection accuracy. Meanwhile, one-stage detectors predict the bounding boxes over the objects without the region proposal, which increases the speed of detection. Therefore, single-stage detectors can be used in real-time systems that must perform analysis and make decisions quickly and effectively.

The deep learning model, Mask-RCNN, proposed for this study depends on a two-stage detector to detect and classify objects in images or video clips. This model is a development of the Faster RCNN network - which relies on proposed regions of objects and classify them based on features of these regions Figure (2). Mask RCNN adds a ResNet hierarchical network ResNet (FPN), as a basic chain to create better features and thus increase detection accuracy (Park et al., 2020). Mask-RCNN model has proven to be efficient for use by many applications used for detecting vehicles, people, animals, and objects. Although this model scored high levels of recognizing many objects from images with high accuracy, it requires a large amount of data, which makes its performance slower than one-stage detection models.

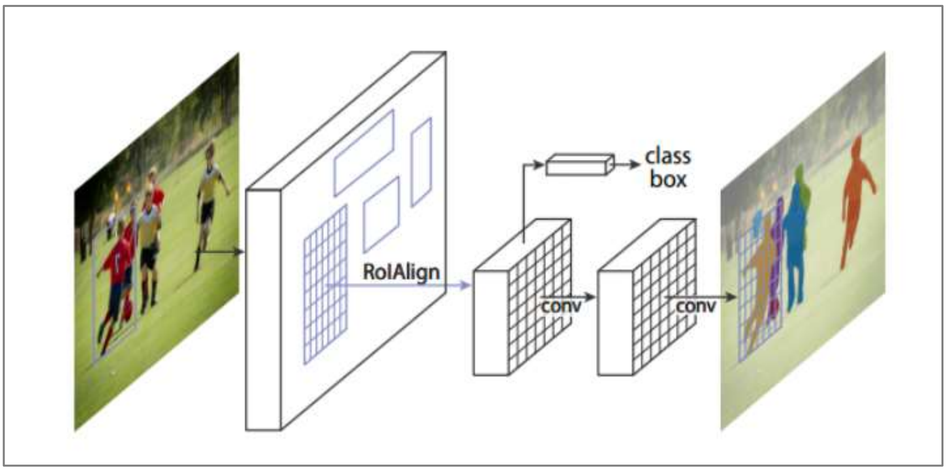

Figure (2): Mask RCNN architecture, Source: Esri

Related Work

The results of a literature review showed the superiority of deep learning techniques for detecting vehicles of all types in images taken from satellites and aircraft. The study (Yayla et al., 2022) used Mask-RCNN algorithm to detect vehicles with different colors in UAV images. Instance segmentation was adopted through different backbones such as MS Coco, Resnet-50, and vgg16 to get different weights during the first training process. Afterwards, these weights were used to monitor segmentation losses and reduce them to a fixed value at the end of the training process. The study concluded on how efficient the algorithm used is in segmenting vehicles with high accuracy in images; it explained the possibility of further enhancing the accuracy by training the model on a more diverse dataset that includes different weather conditions. Finally, the study suggested the possibility of using it in the future for detecting buildings, vehicles, and roads.

(Narcisse Nader, 2023) adopted a transfer learning approach to detect military vehicles from satellite and aerial images using Mask R-CNN algorithm. It aimed to study the potentials of instance segmentation models and evaluate their performance for detecting military vehicles by training the models on a dataset consisting of 200 aerial and satellite images containing 10,624 annotated military vehicles. The models were then tested on a separate set of unseen images to evaluate their performance. The results showed the efficiency of the proposed method in detecting military vehicles even in complex and crowded environments, through average precision (AP) metrics for each category and at varying intersection over union (IoU) thresholds (AP50 and AP75). The study recommended the importance of using machine learning and computer vision techniques for detecting and classifying military vehicles in satellite and aerial images and paving the way for further research and development in this field.

In the study of (Wu et al., 2021), the researchers proposed an improvement to the Mask R-CNN model for detecting aircrafts in remote sensing images, through using a subnetwork that helps in detecting small and mixed targets. The study used a composite dataset of satellite images containing a wide range of different aircrafts in many environments. The performance of the proposed model was evaluated on the test dataset using the mean average precision (mAP) metric as a performance measure. The results showed a detection accuracy of 95% for large targets and 85% for small targets. The study recommended improving the proposed model by increasing the diverse data and using this approach in air traffic surveillance and aviation management.

(He & Zhang, 2021) developed a smart detector that can automatically detect vehicles from satellite images, using two neural networks SegNet and Mask R-CNN, in the central area of Manchester. The accuracy of the detector was improved using features and attributes of the vehicle such as shape, size, and color. This approach has contributed to improving the results and performance efficiency in terms of facilitating the detection and identification of vehicle locations from satellite images automatically.

(Li et al., 2018) discussed the importance of detecting aircraft in satellite images using convolutional neural networks (CNNs). The study aimed to first propose candidate regions containing multiple aircraft, and then accurately detect each aircraft in these candidate regions. This was done by designing a precise and efficient detection framework consisting of two CNNs with the same structure. The first is used to propose candidate regions approximately, while the other is used for accurate detection of aircraft. This method achieved high accuracy in detecting aircraft from Google Earth images.

The study (Mansour et al., 2019) also discussed the progress made in deep learning for detecting vehicles in satellite images using the Faster and SSD CNN networks, through transfer learning, and conducting an experimental analytical comparison between the two models. The study used many satellite images such as Google Earth, JF-2, and WORLD-VIEW, to collect vehicle images from different environments and satellites with different spatial resolution (less than 1 meter). The results of the study showed the superiority of the Faster CNN network in detecting vehicles from the satellite images used.

(Torén, 2020) also compared two methods for analyzing machine learning image Faster R-CNN and SSD to detect and track ships from satellite images. The Faster R-CNN network is implemented in two stages, where interest regions are first identified, followed by detection based on these specified regions. While the SSD network is implemented in one stage, where objects are directly detected that cover a small number of cell units. A satellite image dataset was used to train and evaluate the candidate models with 5600 images taken from a variety of locations. The TensorFlow Object Detection API was used to implement the two models. This study concluded that the Faster R-CNN network is superior for identifying and tracking ships in satellite images.

(Tan et al., 2020)used a deep learning algorithm to detect vehicles in remote sensing images from high-resolution satellites to improve traffic conditions. The Faster R-CNN algorithm was used to train the deep learning model and obtain an average accuracy of 70%. This model was applied to detect vehicles in terms of flow and number of vehicles at one intersection. The results showed that the percentage of missed and false detection was very small, indicating the efficiency of the model used effectively.

(Ophoff et al., 2020) and colleagues proved the possibility of automatic detection of small objects, such as vehicles and ships, in satellite images with spatial resolution of 0.3 and 0.5 meters. Four different object detection networks such as SSD and different versions of YOLO were trained and evaluated. The results showed that the D-YOLO model outperformed with an average accuracy of 60% for vehicles and 66% for ships. Finally, the study recommended the need for explaining more large and diverse data in different environments to improve the training of the proposed models.

(Erbe, 2021) also addressed the approach of using deep learning to detect bicycles and their parking locations from aerial images and a digital surface model. The study found that satisfactory results were achieved with a detection accuracy of up to 83.2%. The study discussed some limitations such as shadows of buildings and trees that hinder the detection process from aerial images.

(Gupta et al., 2021) used Yolo deep learning algorithms for the automatic detection of vehicles from satellite images to contribute to traffic monitoring and control. The Dota available dataset was used to train the deep learning model and take advantage of the large number of data and its diversity. The study achieved good results that proved the efficiency of the proposed detector. The study also recommended developing this detector to build an intelligent automated system for traffic monitoring control.

Study Methodology

The study used experimental methodology to detect the phenomenon of visual pollution (i.e. abandoned vehicles), in the city of Riyadh. After an extensive study of vehicles' characteristics, such as color and shape, and investigating appropriate methods for detecting them according to those characteristics, the researcher designed an automated system for detecting abandoned vehicles (Abandoned Cars Automatic Detection). This system consists of three interconnected stages, as demonstrated in Figure (3):

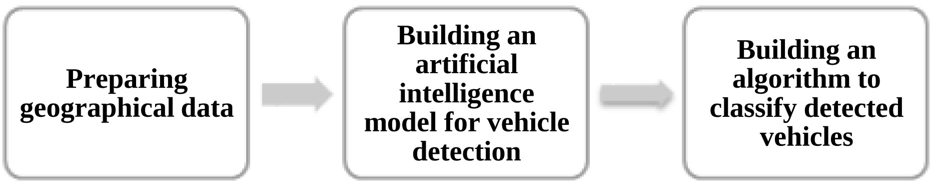

Figure (3): An illustration of the stages of building the proposed automated system for detecting abandoned vehicles (Abandoned Cars Automatic Detection)

First Stage: Preparing Geographical Data

The first stage of detecting abandoned vehicles was to provide high-resolution temporal satellite images as much as possible. The images available vary in terms of resolution and number of bands (panchromatic and multispectral), as proper satellite images are selected depending on the type of objects targeted for analysis.

WorldView-3 satellite images, with a spatial resolution of 30 cm, which were taken on three different dates during the 3rd, 5th and 7th month of 2020, were obtained from multiple sources; such as: Apollo mapping and King Abdulaziz City for Science and Technology. WorldView-3 satellite images were selected in order to facilitate sampling and training a deep learning model to detect abandoned vehicles. Pre-processing operations, such as geometric correction, atmospheric correction, and orthorectification, were performed.

Second Stage: Vehicle Detection Model

The study relied on the deep learning (DL) algorithm (Mask RCNN) available in the ArcGIS Pro environment to build an artificial intelligence (AI) model for vehicle detection through five basic stages that included: collecting training samples, pre-processing the data, training the model, testing the model on new images, and evaluating model’s accuracy, Figure (4).

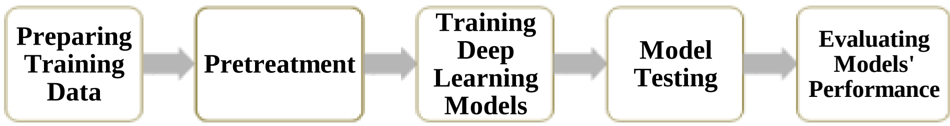

Figure (4): Deep learning model for vehicle detection

First: Preparing Training Data

The training samples included all the vehicles in the temporal images to train a influential model capable of detecting them later, as shown in Table (1). A regular scanning process was performed to review the images used in a regular pattern from left to right and vice versa in search of target vehicles, then came the step of numbering all the vehicles in the images using vehicle bounding box, Figure (5).

Table (1): Training Samples

Date of Satellite Images	Satellite	Training Samples
03-2020	WorldView-3	1249
05-2020	WorldView-3	1414
07-2020	WorldView-3	109

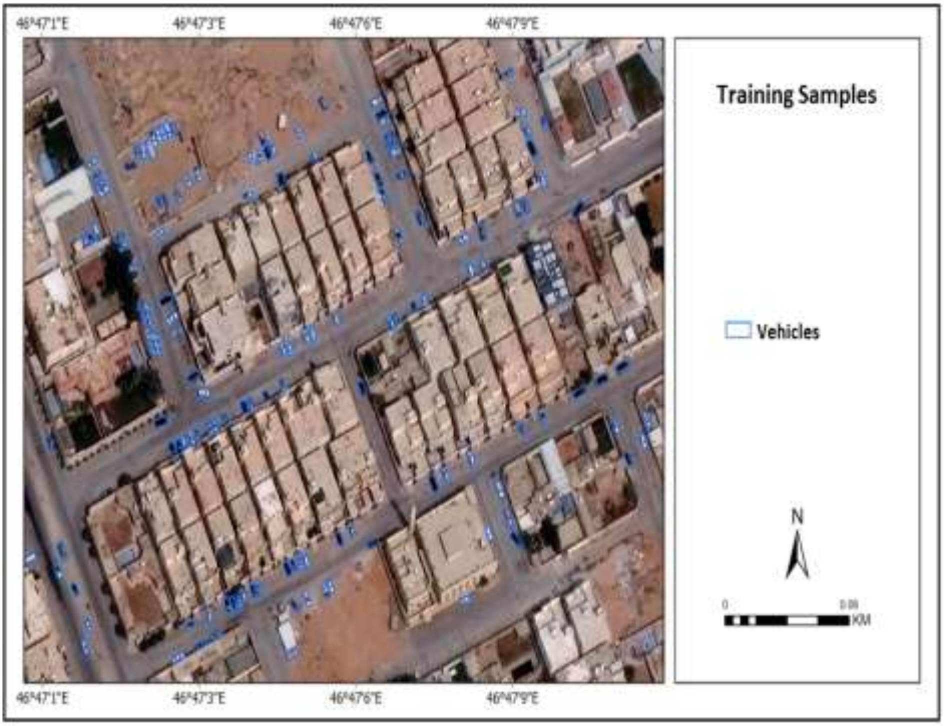

Figure (5): Creating Training Samples

Second: Preprocessing

1. Removal of Buildings

This step of image pre-processing aimed at removing everything that negatively affects the performance of the model. This stage is considered part of the integrated system for detecting abandoned vehicles. Therefore, due to the great similarity in formal features between overhead water tanks and vehicles, all buildings in the input images were removed before exporting training data, Figure (6).

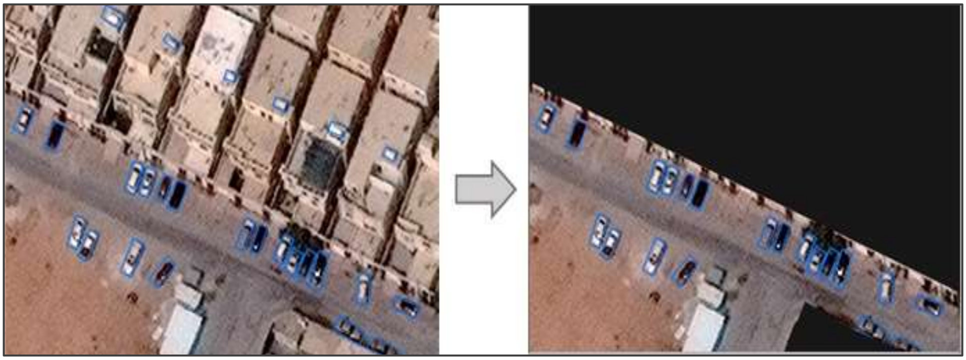

Figure (6): Removing buildings from all images used in the study

2. Label Data

To perform the process of labeling and annotating data in the images under-study, samples were first identified by placing bounding boxes on vehicles, as in the previous stage. Then, relevant spatial and descriptive information necessary to train the deep learning (DL) model was filled in.

3. Exporting Training Data

Using previous samples, image slices were created by cropping satellite images under-study into several slices. This process, called (exporting training samples), is necessary to train the deep learning model, as shown in Figure (7) and Table (2), which demonstrate the number of training and validation data, and the objects within each image.

Figure (7) Crop satellite images to export training data.

Table (2) Training Data Set

Date of Satellite Images	Satellite	Training Data 90%	Validation Data 10%	Number of Objects in Each Image
03-2020	WorldView-3	545	54	5359
05-2020	WorldView-3	549	54	6321
07-2020	WorldView-3	377	37	2215

Third: The Deep Learning Model Training

To train the model to recognize vehicles in satellite images, training data set in the previous stage (i.e. the vehicles) was used to identify the characteristics of these vehicles. The training data was divided into 90% for training and 10% for verification. The study used a convolutional neural network to detect objects (Mask R-CNN for Object Detection). This network is one of the supervised methods used in image processing in the field of deep learning (DL). It can be used in this study to identify vehicles in remote sensing images. It was chosen because it places a specific mask on the detected object, ignoring the background, which will help in tracking abandoned vehicles in temporal images when making intersections between detected vehicles. The trained model was then adjusted using a pre-trained model to utilize the huge amount of data available, which was used in training it. The model was retrained using more than one pre-trained model and the model parameters were changed to obtain the best model with acceptable accuracy.

Fourth: Testing the Model

After completing training, the best model which achieved the highest accuracy and lowest loss of data was chosen to be applied to three new sites. Object Detection tool was used to detect vehicles in the input images. Deep Learning Toolkit contains tools which detect features from images using multiple layers in neural networks. This is an Object Detection tool, which runs a trained deep learning model on cellular data to produce categories for the features of the objects it contains, by placing boxes around the detected object or dots at the center of the object detected. The detection threshold for all images was set at 0.1 because the size of the vehicles is very small in the images. Therefore, the detection threshold must be reduced to ensure achieving largest number of detections. After conducting the vehicle detection process, false and repeated discoveries that did not contain values were removed by sorting the descriptive data for each detected vehicle.

Fifth: Evaluating the Model's Performance

After completing the application of the model on input images, the criteria adopted to evaluate the accuracy of deep learning (DL) models for detecting objects in images were measured by comparing the detected objects with reference samples. In this phase, five standard criteria were relied upon to evaluate and compare detection accuracy quantitatively. These criteria include: calculating IoU, Precision, Recall, F1 Score, and Average Precision; in addition to Mean Average Precision. Finally, sorting true and false positives and false negatives was done (Gupta et al., 2021; Tan et al., 2020).

Third Stage: Classification of the Detecting Vehicles

In the previous stage, the preparation of deep learning (DL) model was completed and its efficiency in detecting vehicles was evaluated. After accepting the results of performance metrics of the proposed model, the final stage started. In this stage, the researcher sought to distinguish abandoned vehicles from non-abandoned vehicles conducting the following three steps:

First: Removing Vehicles Parked in Front of Residential Buildings

The study used images with a spatial resolution of 30 cm, as mentioned previously, and the accuracy of the model's performance demonstrated its suitability for accurately detecting and identifying vehicles in different circumstances and situations. In order to distinguish between abandoned and non-abandoned vehicles, the study assumed that all vehicles parked in front of residential buildings are currently active, non-abandoned vehicles. This was done after visually tracking them in time-lapse satellite images. Therefore, the focus of this study was on the vehicles detected on commercial roadsides, un-built lands, mosque parking lots, and commercial complexes.

At this stage, all vehicles parked in front of residential buildings were identified using the 'Select by Location' tool at a distance of three meters from the residential building layer, in order to ensure that all vehicles in front of residential buildings were identified and removed from the list of vehicles detected in all the temporal images under study.

Second: Calculating Intersections

This stage focused on identifying intersecting vehicles in the temporal images on the sides of commercial roads, un-built lands, and public parking lots, in preparation for matching and detecting the reference data of abandoned vehicles in the next stage. Several attempts were made to attribute the intersection between the vehicles detected in the images taken on different dates at several different thresholds. Meanwhile, the amount of displacement between the temporal images was taken into account, so the following values were tried (25%, 50%, 75%, and 99%) to ensure obtaining the largest number of intersections. After conducting several experiments, the value was set to 50%; ignoring the lower values, as it was found that the vehicles intersect at rates less than 50% are not identical. So, they were excluded from the analysis.

Third: Matching the Results

All vehicles intersecting at a percentage of 50% or more were identified in the previous stage in order to match them with reference abandoned vehicles to determine whether the vehicle was abandoned or not. At this stage, manual matching was performed for each of the two intersecting vehicles in the two temporal images and compared to several images taken on different dates using Google Earth Pro to increase reliability.

Results and Discussion

A Mask-RCNN detector was created, tested and proved to detect vehicles in satellite images effectively, with a spatial resolution of 30 cm, using 1471 training images and pre-trained models. Many experiments were conducted using several variants of the model to reach the best accuracy level, with lowest loss rate. The results reached in various stages of the automated system developed are summed up below:

First: Results of Training a Vehicle Detection Model

Pre-processing operations were carried out on the training data set and input images to improve training the model to recognize vehicles in satellite images. The four most famous neural networks were also examined as they are all based on the same technology and differ in structure and number of layers. The networks examined in the course of this study were: (Resnet152 - Resnet101 - Resnet34 - Resnet50). Table (3) demonstrates the results of deep learning (DL) models were trained. The second model was found to be the most accurate one and therefore it was adopted in vehicle detection.

Table (3): The best training models for vehicle detection

Run	Samples Size	Backbone	Freeze Model	#epoch	Pre-trained Model	Accuracy
1	1249	Resnet50	Yes	30	No	0.81%
2	**2663**	**Resnet50**	**No**	**30**	**Yes**	**0.83%**
3	2772	Resnet50	Yes	20	Yes	0.82%
4	2772	Resnet34	Yes	50	No	0.36%
5	2772	Resnet101	Yes	50	No	0.60%
6	2772	Resnet152	Yes	50	No	0.30%

The first deep learning model was trained using a local data set without using a pre-trained model. Model parameters were adjusted and input size was set to 256×256; while the volume was 32 images in each iteration, as demonstrated in the previous table (3). Its accuracy level reached up to 81%. Figure (8) reveals training performance of the first model, where training loss was evident at the beginning of training, then, decreased as training progressed (number of iterations: epoch). It was also found that validation loss rate for test data is slightly higher than training data loss rate; which indicated that model’s performance was starting to over-fit and under-fit to the training data, which is a common problem in deep learning (DL) models which need to fine-tune the data or stop early when its performance improves.

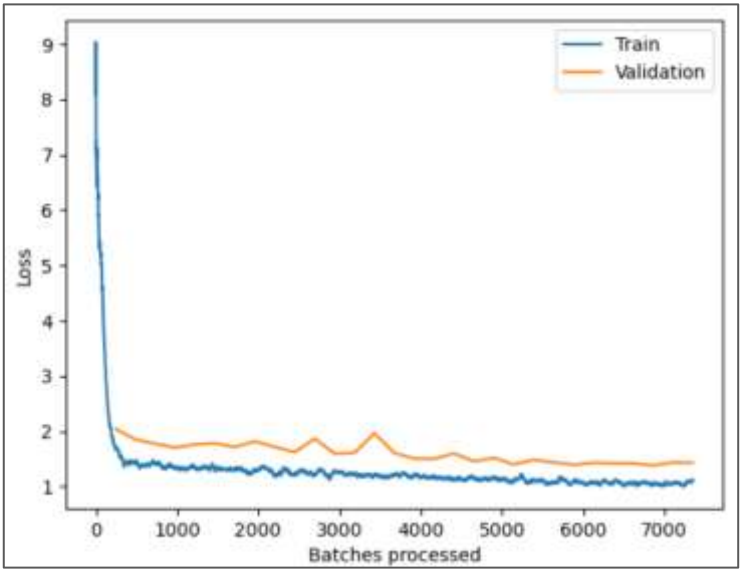

Figure (8): Performance of the first training model 'Mask R-CNN'

In the second model, Transfer Learning (TL) technology was used. It is a new artificial intelligence technology and is considered one of the most successful technologies, especially in cases where a large data set is not available to train the model well. It relied on using a model that was pre-trained on general images; as the basis for a new model. In other words, the new model was built by retraining the pre-trained model on images that are more specific and relevant to the current problem of the study. Thus, the researcher managed, in terms of this model, to benefit from the experiences and knowledge which the model has acquired from previous learning - instead of building and training the model from scratch, which is useless if there was limited data for retraining it on the data set relevant to the topic under research. Therefore, adjusting this model's variables and improving its performance, so that it would have learning experiences for more specific features of the problem under research was the optimal option.

This type of learning may also increase the speed of learning by focusing only on new features that differ from the pre-trained model. Therefore, a previously trained model with 81% accuracy, one of the deep learning models available at Esri, was used. It was developed for car detection in high-resolution aerial photographs and the images taken by drones. Using the pre-trained model contributed to raising the accuracy of the model used by 2% over the first model, with an increase in the training time taken, as shown in the previous table (3).

Figure (9) illustrates the training performance of the second model, where training loss and validation loss decrease as the training progresses, which is a positive indicator of learning. It was found that training loss decreases more consistently compared to validation loss, which had some fluctuations. But, in general, it took a downward turn. This may indicate that the model's performance has begun to over-perform on the validation data. This happens when the model learns so well that it cannot generalize to new data. However, Figure (10) indicates achieving good detection when comparing the training samples to the vehicles detected for this model. Overall, the transfer learning approach using pre-trained models made it possible to take advantage of the vast amount of available data to create a more efficient detection model.

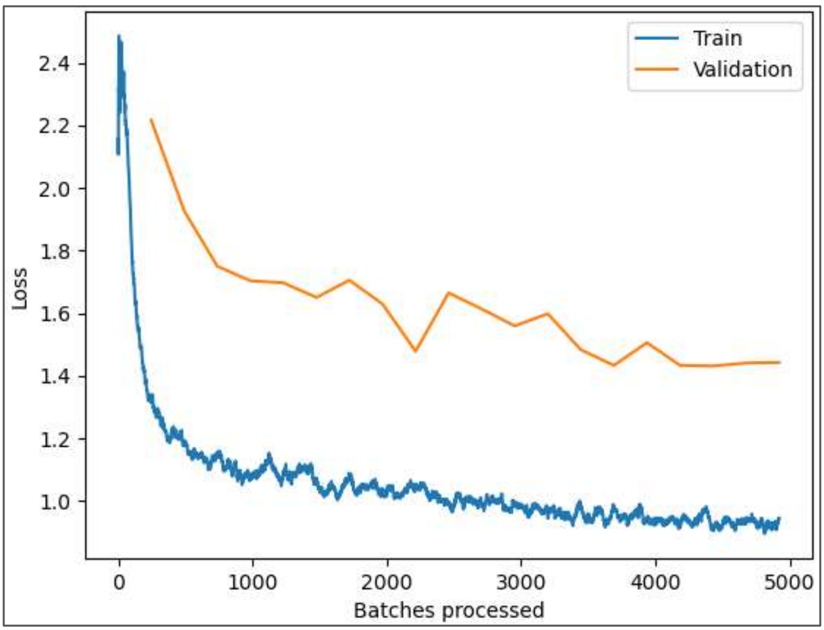

Figure (9): Performance of the second training model 'Mask R-CNN'

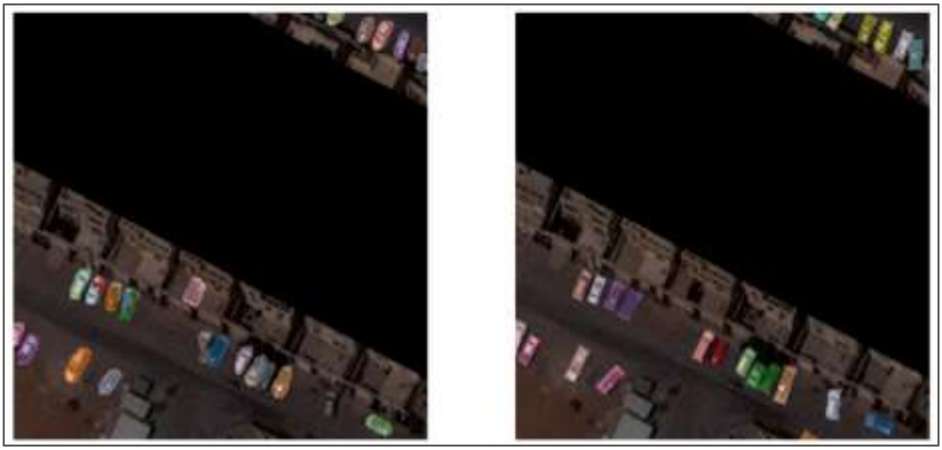

Figure (10): Comparing training samples (A) with the vehicles detected (B) in the second model

In the third model, another previously trained model with a high accuracy of 91% was used on a new additional data set; so as to help improving the current detection model. It was noted that the accuracy did not change much from the second model. As for the fourth, fifth, and sixth model, the Backbone model was modified, which is a basic model used as a basis for a new model which was previously trained on a very large data set. Thus, Resnet34, Resnet101, and Resnet152 were tried, but the model's performance did not improve. In fact, it decreased by different percentages, as demonstrated in the previous table (3). Therefore, the second model was the optimal model to use for all temporal satellite images used in the study.

Second: Test Results of the Vehicle Detection Model

Initially, the second deep learning model, which achieved the best performance results with 83% accuracy, was applied to the input images to verify the detection performance and compare it with the reference data. After several experiments, a detector threshold of 0.9 was used for each detection process to ensure that only reliable detections were obtained. It was then applied to a new site that was not shown to the model during training, which is Al-Munsiyah neighborhood in the city of Riyadh, where detection results proved the efficiency of the model used in identifying vehicles accurately, as in the Figure (11).

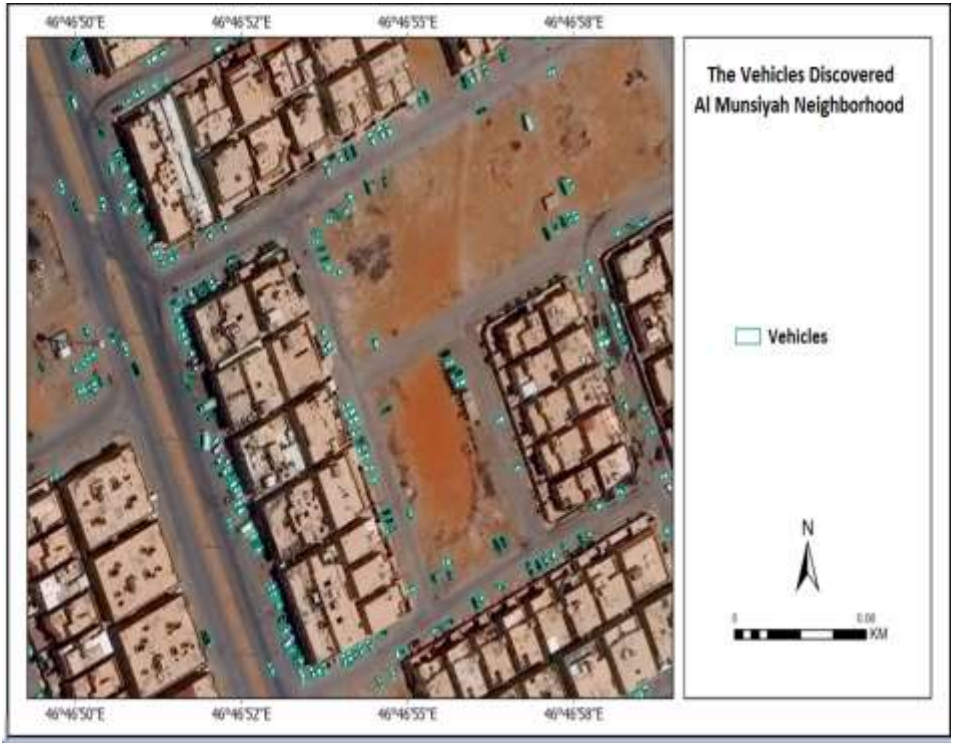

Figure (11): Applying a deep learning model in Al Munsiyah neighborhood

Third: Evaluation of Vehicle Detection Model

Vehicle detection methods were evaluated on the study data set. The analysis of the results was based on five standard criteria commonly used in previous studies to compare detection accuracy quantitatively. The following Table (4) demonstrates the percentages of detection accuracy after using deep learning model in Al Munsiyah neighborhood, according to the criteria have been previously specified.

In order to judge the accuracy of the model in detecting vehicles within images correctly, the detection accuracy must be evaluated through the percentage of overlap between detected vehicles and real vehicles (IoU). Worth noting that this model scored good values on the mAP @ IoU (50-95%) scale, with a percentage of 87 %. As for accuracy level, it was also high, with values of 95%, 91% and 93% for precision, recall and F1 score, respectively.

Table (4): Evaluation of the deep learning model applied in Al-Munsiyah neighborhood

Mask-RCNN Model	mAP @ IoU (50-95%)	Precision	Recall	F1score	Ap
Al Munsiyah Neighborhood	87%	95%	91%	93%	88%

Thus, it became clear, based on the above, how efficient this model is in detecting vehicles in satellite images, despite the low resolution and quality of the images; compared to other models that use drone data and aerial images of few centimeters. The model used proved to be capable of predicting vehicles in different conditions, such as vehicles parked in the shade, vehicles which are covered with dust and parked in open areas, tightly packed vehicles, or vehicles which only a small part of them was visible because they were located under a tree or behind a barrier. This indicates the applicability of the model in other locations with high accuracy. Nonetheless, the model did not perform well in some difficult cases, such as in case of sparsely vegetated lands, images of the remains of liquid building materials which usually appear as a dark color on the ground, images of water tanks on buildings' rooftops, and images of some damaged structures lying around which the model predicted as positive cases. However, these are few wrong cases in comparison with multiple efficient positive results.

The overall performance of the deep learning model used on the proposed data set was demonstrated, which achieved more than 87% in both precision and recall. Further, loss curves showed a training which is close to stability and the model trained well on training data set and gave good detection results.

Fourth: Evaluating the Classification of the Detected Vehicles

In the previous stage, a deep learning model was built capable of detecting small vehicles in satellite images and detection accuracy was calculated based on several standard criteria. Evaluation of the results achieved by the model demonstrated its efficiency and suitability for detecting all vehicles in the images. This leads us to the final stage in which those vehicles are classified. In the final stage, the vehicles detected were distinguished to identify abandoned vehicles after matching and tracking them in various temporal images and determining their locations.

According to 'Municipal Services, Planning, and Urban Development Sector', the definition of 'Abandoned Vehicles' is: "The vehicles in which some part has been damaged and cannot be used unless they are repaired. Thus, they were left for a long time in open areas causing distortion of the general view." Now, these characteristics can be distinguished in images and video clips more accurately than in satellite image which require expertise for interpretation. In general, monitoring a phenomenon using satellite images can be affected by several factors, especially if the purpose is time tracking. The study found that it was not possible to obtain two or more images under same shooting conditions, as different angle of view may affect the appearance of the shape of a single phenomenon. Therefore, the vehicle which appears in an image, which was taken at a different side angle, looks different in other images were taken at a high or low angle, Figure (12). During analysis, a small amount of displacement was found for some intersecting vehicles, and this problem was overcome by increasing the threshold of intersection between vehicles by 50%; to ensure obtaining the largest number of vehicles that are believed to be abandoned.

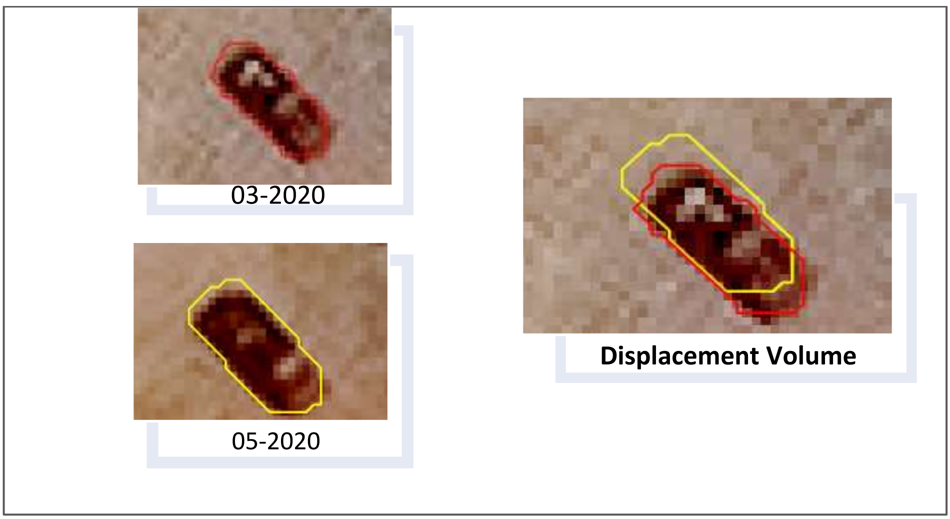

Figure (12): the displacement in two time frames.

At this stage, an intersection analysis was conducted for vehicles that intersected at a rate of 50% or more in the temporal images, taking into account that not every two intersecting vehicles are considered an abandoned vehicle. Rather, matching and temporal tracking must be performed depending on reference data to classify the vehicle correctly. So, verification of vehicles' type was enhanced by examining intersecting vehicles visually using Google Earth Pro images that were taken on more than one date to verify that they are the same abandoned vehicles, Figure (13).

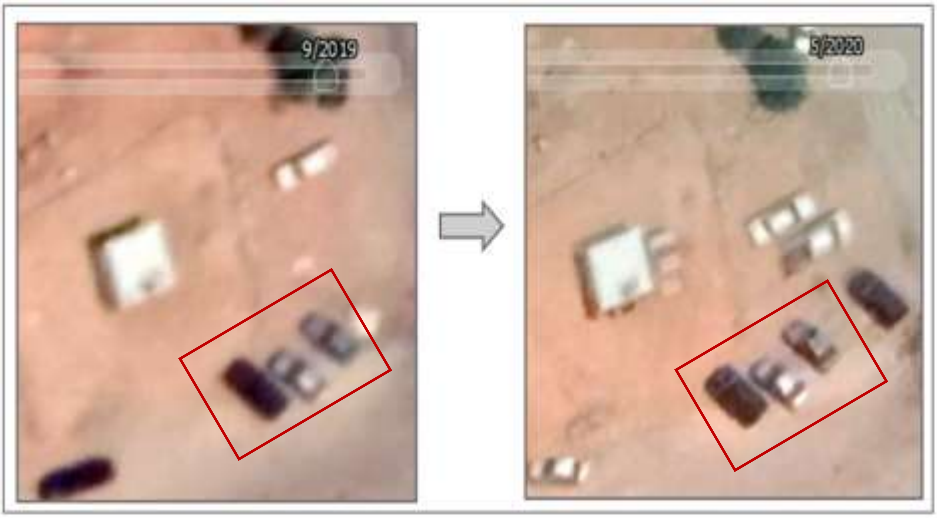

Figure (13): Matching intersecting vehicles using Google Earth Pro images covering eight months period.

Based on the matching results, 25 vehicles were monitored and verified to be abandoned in Al Munsiyah neighborhood. These results highlight the efficiency of vehicle detection model, which facilitated monitoring and tracking abandoned vehicles faster than traditional methods, Figure (14).

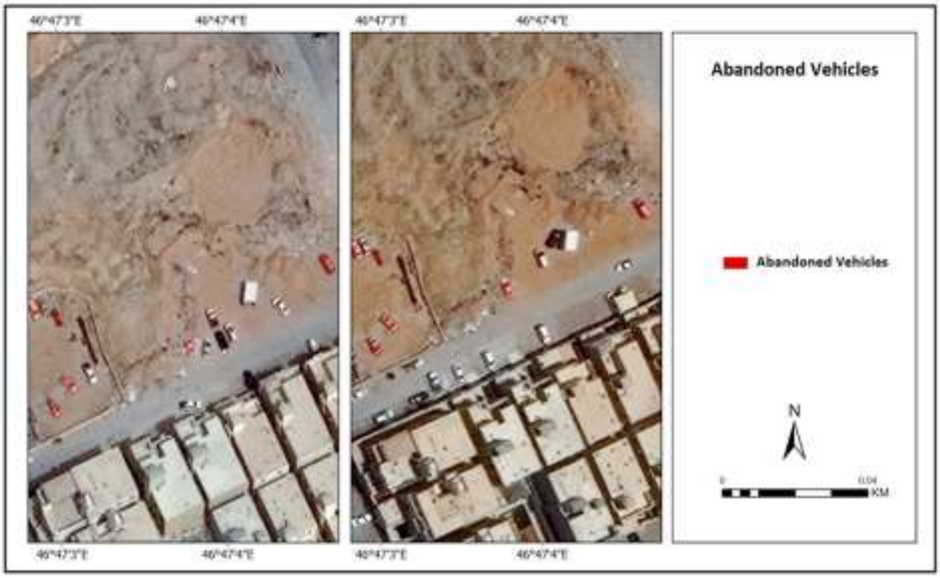

Figure (14) Detection of abandoned vehicles in two temporal satellite images taken on 3-2020 and 5-2020

Conclusion

This study attempted contributing to the efforts made for developing monitoring operations aiming at mitigating the phenomenon of visual pollution (caused by abandoned vehicles) in the city of Riyadh; using geospatial techniques and deep learning. The study designed an automated system that managed to detect abandoned vehicles in WorldView-3 satellite's temporal images of different areas efficiently. The system consisted of three basic stages to achieve the objectives of the study. The first stage focused on pre-processing study data set, while the second stage, which is the most important stage, focused on training the deep learning model and refining it using Mask-RCNN algorithm; to detect vehicles in satellite images. The last stage sought to distinguish abandoned vehicles from other actively used vehicles, using intersection analysis and matching techniques.

The results of the proposed automated system highlighted the extent of its effectiveness, ensuring that it can contribute to automated monitoring of abandoned vehicles and help improve decision-making process. Employing this system may improve urban landscape in the city. The proposed deep learning model, Mask-RCNN, also proved to be highly efficient in identifying vehicles in general. Moreover, it helped in classifying detected vehicles into abandoned and non-abandoned vehicles, using intersection analysis and matching techniques to analyze various temporal images. To sum up, this model proved to be effective and can be further developed and applied in other important fields, such as: vehicle counting, traffic density analysis, and prediction of congestion and traffic accidents.

REFERENCES

Adnan, S., & Al, M. (2022). Visual pollution in Fallujah city.

Al Hashmi. (2016). Spatial Data Infrastructure for smart cities – NSDI Oman – Geospatial World. https://www.geospatialworld.net/article/spatial-data-infrastructure-sdi-smart-cities/

Al-Qutaifi, Nasser Badr. (2022) Visual pollution in Kuwait City - a field analytical study according to the new Environmental Protection Law No. 42 of 2014.

Al-Tuwaijri, Hamad bin Ahmed. «The Urban Expansion of Riyadh)1987-2017(Study Using Remote Sensing Techniques and Geographic Information Systems». Architecture and Planning Journal, pp. 195-.(2018) .213

Bawareth. (2022). Study of the manifestations of visual pollution in some neighborhoods of Jeddah city: Comparative study. Journal of Humanities and Social Sciences, 6(2), 170–198. https://doi.org/10.26389/ajsrp.w251021

Erbe, K. (2021). Detection of Bicycle Racks from Geodata Using Deep Learning. October.

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., … Asia, A. (2022). Pollution and health: a progress update. Global Alliance on Health and Pollution. https://doi.org/10.1016/S2542-5196(22)00090-0

Gupta, H., Jindal, P., & Verma, O. P. (2021). Automatic Vehicle Detection from Satellite Images Using Deep Learning Algorithm (pp. 551–562). https://doi.org/10.1007/978-981-16-1696-9_52

He, K., & Zhang, L. (2021). VEHICLE DETECTION IN SATELLITE IMAGES WITH DEEP NEURAL NETWORKS AND VEHICLE SHAPE FEATURES. International Geoscience and Remote Sensing Symposium (IGARSS), 2021-July, 4212–4215. https://doi.org/10.1109/IGARSS47720.2021.9553697

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., & Qu, R. (2019). A survey of deep learning-based object detection. IEEE Access, 7, 128837–128868. https://doi.org/10.1109/ACCESS.2019.2939201

klei. (2019). The Role of Geospatial Techonologies in Building Smarter Cities.

Li, X., Wang, S., Jiang, B., & Chan, X. (2018). Airplane detection using convolutional neural networks in a coarse-to-fine manner. Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2017, 2018-January, 235–239. https://doi.org/10.1109/ITNEC.2017.8284943

Mansour, A., Hassan, A., Hussein, W. M., & Said, E. (2019). Automated vehicle detection in satellite images using deep learning. IOP Conference Series: Materials Science and Engineering, 610(1). https://doi.org/10.1088/1757-899X/610/1/012027

Mohammed, O. A. (2023). A Study of The Phenomena of Visual Pollution in Halabja and its Effects on The Psychology of Citizens. January. https://doi.org/10.24042/ajp.v5i2.14979

Narcisse Nader. (2023). Military Vehicles Detection in Satellite & Aerial Imagery using Instance Segmentation | MVD-Satellite-Imagery – Weights & Biases. https://wandb.ai/nadernarcisse/MVD-Satellite-Imagery/reports/Military-Vehicles-Detection-in-Satellite-Aerial-Imagery-using-Instance-Segmentation--VmlldzoyNzE1MTY0

Ongsulee, P. (2018). Artificial intelligence, machine learning and deep learning. International Conference on ICT and Knowledge Engineering, 1–6. https://doi.org/10.1109/ICTKE.2017.8259629

Ophoff, T., Puttemans, S., Kalogirou, V., Robin, J. P., & Goedemé, T. (2020). Vehicle and vessel detection on satellite imagery: A comparative study on single-shot detectors. Remote Sensing, 12(7). https://doi.org/10.3390/rs12071217

Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: state of the art. Applied Intelligence, 51(9), 6400–6429. https://doi.org/10.1007/S10489-021-02293-7.PDF

Park, H., Park, S., & Joo, Y. (2020). Detection of Abandoned and Stolen Objects Based on Dual Background Model and Mask R-CNN. IEEE Access, 8, 80010–80019. https://doi.org/10.1109/ACCESS.2020.2990618

Shrivastava, A. T., & Choudhary, A. (2016). MODEL TO MITIGATE VISUAL POLLUTION BY ADS AND SIGNAGE FOR URBAN UNISON. International Journal of Engineering Research and General Science, 4(3). www.ijergs.org

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. https://doi.org/10.1109/ACCESS.2020.3017894

Tao, W. (2013). Interdisciplinary urban GIS for smart cities: advancements and opportunities. Http://Www.Tandfonline.Com/Action/JournalInformation?Show=aimsScope&journalCode=tgsi20#.VsXpLiCLRhE, 16(1), 25–34. https://doi.org/10.1080/10095020.2013.774108

Torén, R. (2020). Comparing CNN methods for detection and tracking of ships in satellite images. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1448055

Voronych, Y. (2013). Visual pollution of urban space in Lviv. Undefined.

Wakil et al. (2019). The Assessment and Mapping of Urban Visual Pollution through an Assembly of Open Source Geospatial Tools. https://doi.org/10.21625/ARCHIVE.V3I2.500

Wu, Q., Feng, D., Cao, C., Zeng, X., Feng, Z., Wu, J., & Huang, Z. (2021). Improved mask r-cnn for aircraft detection in remote sensing images. Sensors, 21(8). https://doi.org/10.3390/s21082618

Yayla, R., Albayrak, E., & Yüzgeç, U. (2022). Vehicle Detection from Unmanned Aerial Images with Deep Mask R-CNN. Computer Science Journal of Moldova, 30(2), 148–169. https://doi.org/10.56415/csjm.v30.09

Youme, O., Bayet, T., Dembele, J. M., & Cambier, C. (2021). Deep Learning and Remote Sensing: Detection of Dumping Waste Using

UAV. Procedia Computer Science, 185, 361–369. https://doi.org/10.1016/j.procs.2021.05.037

Zakaria, N., & A., J. (2015). Smart City Architecture: Vision and Challenges. International Journal of Advanced Computer Science and Applications, 6(11), 246–255. https://doi.org/10.14569/ijacsa.2015.061132

بناء نظام آلي لكشف مظاهر التلوث البصري باستخدام التقنيات الجيومكانية والتعلم العميق: دراسة تطبيقية على المركبات المهجورة في مدينة الرياض

إيمان العمري
باحثة دراسات عليا
قسم الجغرافيا، كلية العلوم الانسانية، جامعة الملك سعود، مدينة الرياض، المملكة العربية السعودية
emand468@gmail.com

أ.د. مفرح القرادي
أستاذ بقسم الجغرافيا، كلية العلوم الانسانية، جامعة الملك سعود، مدينة الرياض، المملكة العربية السعودية
mufarehq@ksu.edu.sa

د. دعاء السعيد
أستاذ مشارك بقسم تقنية المعلومات،
كلية علوم الحاسب الآلي والمعلومات، جامعة الملك سعود،
مدينة الرياض، المملكة العربية السعودية
dalsaeed@ksu.edu.sa

المستخلص:

تعد المركبات المهملة إحدى مظاهر التلوث البصري في المدن السعودية، حيث يشير التلوث البصري إلى التأثير الجمالي السلبي للعناصر التي خلفها الإنسان في المشهد الحضري، وتهدف هذه الدراسة إلى كشف المركبات المهملة في مدينة الرياض كأحد مظاهر التلوث البصري وذلك من خلال بناء نظام آلي للكشف عنها في صور الاستشعار عن بعد باستخدام تقنيات التعلم العميق. تم اعتماد المنهج التجريبي في هذه الدراسة وتمثلت استراتيجيته في تطبيق أسلوب كشف الأجسام **Object Detection** بواسطة تقنيات التعلم العميق. تعتبر تقنية كشف الأجسام من التقنيات المتقدمة في مجال رؤية الحواسيب والذكاء الاصطناعي. وتهدف

إلى التعرف على وتصـنيف الأجسـام المختلفة في الصـور أو مقاطع الفيديو. لتحقيق أهداف الدراسـة تم بناء نظام آلي لكشـف المركبات المهملة من صـور زمنية مختلفة عبر القمر الصناعيWorld View-3 ، وقد بني هذا النظام الآلي ليعمل على ثلاث مراحل أساسية شملت المعالجة المسبقة لمجموعة بيانات الدراسة وتدريب نموذج التعلم العميق وصقله باستخدام خوارزمية Mask-RCNN واخيرا تصنيف المركبات المكتشفة إلى مهملة وغير مهملة باستخدام تحليل التقاطع والمطابقة باستخدام الصور الزمنية. وخرجت الدراسة بعدد من النتائج أهمها قدرة كاشف المركبات Mask-RCNN على كشف المركبات من صور القمر الصناعي World View-3بمقاييس مرتفعة بلغت 95%، 91% و93% لكلا من الدقة والاستدعاء ودرجة F1 على التوالي كذلك أظهر هذا النموذج قيمة جيدة في مقياس (50-95%) mAp @ IoU بلغت 87%، والذي ساعد بدورة في تتبع المركبات المهملة من الصور الزمنية وتحديدها بدقة.

الكلمات المفتاحية: التلوث البصري؛ المركبات المهجورة؛ اكتشاف الأشياء