Computers, Materials & Continua
DOI: 10.32604/cmc.2023.033038
Article

Tech Science Press

Optimal Deep Convolutional Neural Network for Vehicle Detection in Remote Sensing Images

Saeed Masoud Alshahrani[1], Saud S. Alotaibi[2], Shaha Al-Otaibi[3], Mohamed Mousa[4], Anwer Mustafa Hilal[5,*], Amgad Atta Abdelmageed[5], Abdelwahed Motwakel[5] and Mohamed I. Eldesouki[6]

[1]Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra, Saudi Arabia
[2]Department of Information Systems, College of Computing and Information System, Umm Al-Qura University, Saudi Arabia
[3]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[4]Department of Electrical Engineering, Faculty of Engineering & Technology, Future University in Egypt, New Cairo, 11845, Egypt
[5]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, AlKharj, Saudi Arabia
[6]Department of Information System, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Anwer Mustafa Hilal. Email: a.hilal@psau.edu.sa
Received: 05 June 2022; Accepted: 12 July 2022

Abstract: Object detection (OD) in remote sensing images (RSI) acts as a vital part in numerous civilian and military application areas, like urban planning, geographic information system (GIS), and search and rescue functions. Vehicle recognition from RSIs remained a challenging process because of the difficulty of background data and the redundancy of recognition regions. The latest advancements in deep learning (DL) approaches permit the design of effectual OD approaches. This study develops an Artificial Ecosystem Optimizer with Deep Convolutional Neural Network for Vehicle Detection (AEODCNN-VD) model on Remote Sensing Images. The proposed AEODCNN-VD model focuses on the identification of vehicles accurately and rapidly. To detect vehicles, the presented AEODCNN-VD model employs single shot detector (SSD) with Inception network as a baseline model. In addition, Multiway Feature Pyramid Network (MFPN) is used for handling objects of varying sizes in RSIs. The features from the Inception model are passed into the MFPN for multiway and multiscale feature fusion. Finally, the fused features are passed into bounding box and class prediction networks. For enhancing the detection efficiency of the AEODCNN-VD approach, AEO based hyperparameter optimizer is used, which is stimulated by the energy transfer strategies such as production, consumption, and decomposition in an ecosystem. The performance validation of the presented method on benchmark datasets showed promising performance over recent DL models.

Keywords: Object detection; remote sensing; vehicle detection; artificial ecosystem optimizer; convolutional neural network

1 Introduction

Object detection (OD) was considered a basic issue in remote sensing image (RSI) analysis. Latest developments in software and hardware abilities enabled the advancement of robust machine-learning (ML) related OD methods. To be specific, deep learning (DL) methods gained attention because of their higher potentiality to extract more abstract and descriptive feature data representations from original input units [1]. On the contrary utilizing shallow structures and conventional handcrafted feature processing, DL techniques present a larger array of very deep architectures on the basis of stacking layers that derives increasingly complicated and abstract features from the inputs in hierarchical and successive manner [2,3]. From this perspective, convolutional related neural methods illustrated a more generalizing power with a powerful and automated feature extracting ability, permitting them to attain an extraordinary performance and deploying as the recent several missions with regard to computer vision (CV), particularly in image classifier roles [4]. Dissimilar to the natural scene image, remotely sensed images contain greater scale variations and higher feature complications in distinct observation circumstances, which needs larger generalizing of object detectors. DNN-related detection techniques are currently being launched from the CV domain to remote sensing domain and gained high grades on multi-class OD [5]. Yet, the great object scale variation in multi-resolution RSIs till now contributes to a greater barrier for object detectors.

OD in RSIs serves a significant part in many military and civilian applications, like search-and-rescue operations, urban planning, and geographical information system upgrading [6,7]. RSI vehicle detection intends for detecting every instance of vehicles in RSIs. In previous techniques, authors generally extracted and devised vehicles feature physically and categorized them for attaining vehicle detection. The main ideology was extracting features of the vehicle and employing traditional ML techniques for categorization [8,9]. But classical target detecting techniques receive higher interest in the conclusion of RSI vehicle detection missions, and it finds it hard for balancing speed and accuracy [10]. When comparison is made to that of conventional techniques (template-matching-related techniques, knowledge-related techniques), the DL related techniques derive features automatically from raw data by shifting a load of manual feature model to the underlying learning mechanism, allowing a very strong feature depiction for extracting maximal semantic stages of feature maps. But this merit, DL related detection methods gained more achievements in both remote sensing and CV [11,12].

This study develops an Artificial Ecosystem Optimizer with Deep Convolutional Neural Network for Vehicle Detection (AEODCNN-VD) model on RSIs. The proposed AEODCNN-VD model employs single shot detector (SSD) with Inception network as a baseline model. In addition, Multiway Feature Pyramid Network (MFPN) is used for handling objects of varying sizes in RSIs. The features from the Inception model are passed into the MFPN for feature fusion which is then passed into bounding box and class prediction networks. For enhancing the detection efficiency of the AEODCNN-VD model, AEO based hyperparameter optimizer is used. The experimental result analysis of the AEODCNN-VD model is carried out using two benchmark datasets.

2 Related Works

Deng et al. [13] present a unified and effectual approach to concurrently identifying multi-class objects from RSI with huge scales of variabilities. Primarily, the researchers reform the feature extraction by implementing Concatenated ReLU and Inception elements that is improves the variation of receptive field sizes. Afterward, the recognition was executed by 2 subnetworks they are a multiscale object proposal network (MS-OPN) to object like region generation in many intermediate layers, which corresponding field equal distinct object scales, and accurate OD network (AODN) to OD dependent upon fused feature map that integrates many feature map which allows smaller and densely packed object for producing stronger response.

Zhou et al. [14] established the polar coordinate model for the DL detection for the very first time, and present an anchor free Polar Remote Sensing Object Detector (P-RSDet) that is attained competitive recognition precision utilizing easier object representation method and lesser regressing parameter. In P-RSDet approach, arbitrary-based OD is gained by forecasting the mid-point and regressing 1 polar radius and 2 polar angles. In [15], the authors present YOLOrs: a novel CNN, specially planned for real-time OD from multi-modal RSI. YOLOrs are OD at several scales, with lesser receptive domains to account for smaller targets, and forecast target orientation. Besides, the YOLOrs establish a new mid-level fusion structure which renders it applicable for multi-modal aerial imagery.

The authors in [16] enhance the YOLOv4 network and current a novel method. Primary, the authors present a classifier setting of non-maximum suppression threshold for increasing the accuracy with no effect on the speeds. Secondary, the author analysis the anchor frame allocation issue from YOLOv4 and presents 2 allocation methods. In [17], Domain Adaptation Faster R-CNN (DA Faster R-CNN) technique was presented to detect aircraft from RSI. Two domain adaptation frameworks were devised and chosen as the standards of similarity measurements among domains. Adversarial training has been implied for alleviating the shift of domain. Ye et al. [18] advance a convolution network method that has an adaptive attention fusion mechanism (AAFM). The method was suggested on the basis of the backbone networks of EfficientDet. At first, as per the object distribution characteristics in datasets, the stitcher can be implemented for making single image that has various scales objects. This process could potentially balance the proportion of multiscale objects and manages the properties of scale variables.

3 The Proposed Model

In this study, a new AEODCNN-VD approach was established for the identification of vehicles accurately and rapidly on RSIs. The proposed AEODCNN-VD model primarily utilized an SSD based object detector with Inception network as a baseline model. Besides, the features from the Inception model are passed into the MFPN to multiway and multiscale feature fusion, which are then passed into bounding box and class prediction network. At the final stage, the AEO based hyperparameter optimizer is used for Inception network.

3.1 Overview of SSD

The SSD utilizes the decreased VGG-16 as based network and added more convolutional layers to end of it. After choosing any layers of distinct sizes in more added convolutional as well as based layers for predicting score and offset to any default boxes (DB) on all the scales. These forecasts were created by single convolutional layer, and the amount of convolutional kernels to these layers are compared with the amount of DBs and the amount of forecast classifications [19]. While the RRC and RON declared, that the semantic data of shallow layers of SSD was restricted, resultant in the efficiency of shallow layers in identifying smaller objects is future weaker than the efficiency of huge ones from the deep layers. Enhance in resolution permits the system for extracting comprehensive features. Another issue is that higher resolution images require CNNs with further layers, for instance, depth scaling. The reason behind deeper network was superior receptive field is extracting similar features which contain further pixels from higher resolution images. The width of network (count of channels) is improved for acquiring the fine-grained feature from the image. Thus, the standard VGG net was exchanged by Inception network. Fig. 1 depicts the structure of SSD.

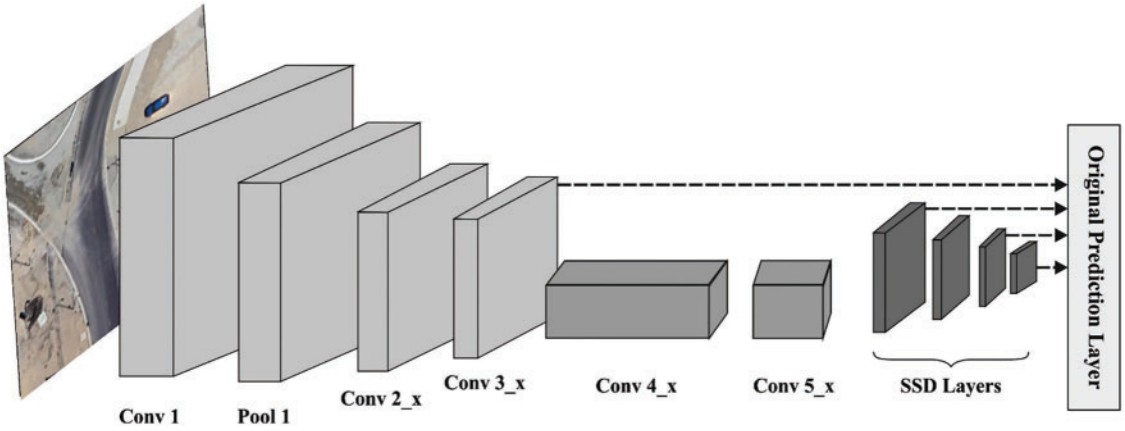

Figure 1: Structure of SSD

3.1.1 Multi-Scale Feature Maps for Detection

Here, convolution feature layers are added subsequent to truncated base network and those layers progressively reduce in size and enable prediction of detection at different scales. The convolution models to predict detection is diverse for all the feature layers and YOLO function on an individual scale feature map.

3.1.2 Convolution Predictor for Detection

Every added feature layer produces a determined set of prediction detection through a group of convolution filters. They are specified on top of SSD network structure. For a feature layer of m × n size with p channels, the building blocks to predict parameter of possible detection is a 3 × 3 × p smaller kernel that generates a score, or a shape offset in relation to DB coordinate. At every m × n position, whereas the kernel was employed, it generates a resultant value. The bounding box offset resultant value is evaluated in relation to DB location in relation to every feature map place that exploits an FC layer rather than a convolution filter.

3.1.3 Default Boxes and Aspect Ratios

In this method, relate a group of default bounding boxes with every feature map cell, for several feature maps at the top of network. The DBs tile the feature maps in a convolution process, such that the location of all the boxes in relation to their respective cell was set. In every feature map, it can be forecast the offset in relation to the DB shape in the cell, along with the per-class score indicating the existence of class sample in every box.

3.2 Base Network: Inception Model

CNN has made remarkable successes in image processing and OD. The weight sharing greatly decreases the learned free parameter count thus minimizing the storage requirement for network functioning and enabling the training of wide-ranging, more powerful network. A CNN encompasses pooling, fully connected normalization, and convolution layers. At all the layers, the $X \in \mathbb{R}^{n\times m}$ input image is convoluted by a collection of K kernels $\{W_k \in \mathbb{R}^{v\times v},\ k = 1, 2, \cdots, K\}$ and bias $\{b_k \in \mathbb{R},\ k = 1, 2, \cdots, K\}$ is added, which produces a novel X_k feature map using an element-wise nonlinear conversion σ. Likewise, the procedure is reiterated for l convolutional layers,

$$X_k^l = \sigma\left(W_k^l \otimes X^{l-1} + b_k^l\right) \tag{1}$$

From the expression, '⊗' is represented as discrete convolutional operator, and the specific type of operation has different versions like 'valid' convolution, 'extra' convolution, 'same' convolution, fractionally stridden convolution, stridden convolution, etc. Usually, convolutional layer is replaced by pooling layer where the pixel values of neighborhood are aggregated through per-mutation invariance function, generally max or average operation that provides alternate means of translational invariant.

$$S_k^{(l)} = Pooling\left(X_k^{(l)}\right) \tag{2}$$

Eventually, convolution and max-pooling layers, the highest-level reasoning in the NN has been done through FC layer, from which the weight is shared again. The considerable decline in the weighted parameter amount and the translation invariant of learned features contributed to the capability of CNN to be trained end-to-end.

Inception is a model in GoogleNet and it was approved to be better in complex image classification tasks [20]. It has multiscale convolutional kernel for extracting the feature from the input image by incrementing the number of convolution kernels and introducing multiscale convolution kernel. The inception model had been enhanced with respect to accuracy and speed. There exist different editions of Inception: Inception ResNet, Inception V1, V2, V3, and V4, which is an iterative evolution of the preceding version. Generally, a lesser variant of Inception model works better in classification tasks. As demonstrated, 1 × 1, 3 × 3, and 5 × 5 convolution kernels are utilized for convoluting the output of upper layer simultaneously to form a multiple branch structure. Feature maps attained from the dissimilar branches are later concatenated for obtaining classification feature of an input image. Processing the operation simultaneously and integrating each result will lead to better image depiction. For making the feature maps have a similar size, all the branches adopt the similar padding mode using the stride of 1. The 1 × 1 convolutional process is utilized beforehand 3 × 3 and 5 × 5 and afterward Max-pooling to decrease the calculation count. Fig. 2 depicts the process of Inception model.

3.3 Multiway Feature Pyramid Network

In SSD, when the feature extracting stage, it can attain a feature layer of sizes $x \times y$ with n channels (8 × 8 or 12 × 12 or superior) [21]. After that 3 × 3 convolutional was executed on $x \times y \times n$ feature layer for obtaining fused features in several scales. It can be changed this further feature layer with MFPN which groups feature at distinct resolutions. MFPN permits the detection features for flow from several paths for obtaining superior fused features. The feature detection at several resolutions does not continuously pay a similar weightage for outcome of method. Further weights were allocated for all the input layers that the network acquires the importance of all the filter fusion procedures. Rather than standard convolutional, it can be executed depth-wise detachable convolutional. Steps to fuse lower-level features with higher-level features were provided under:

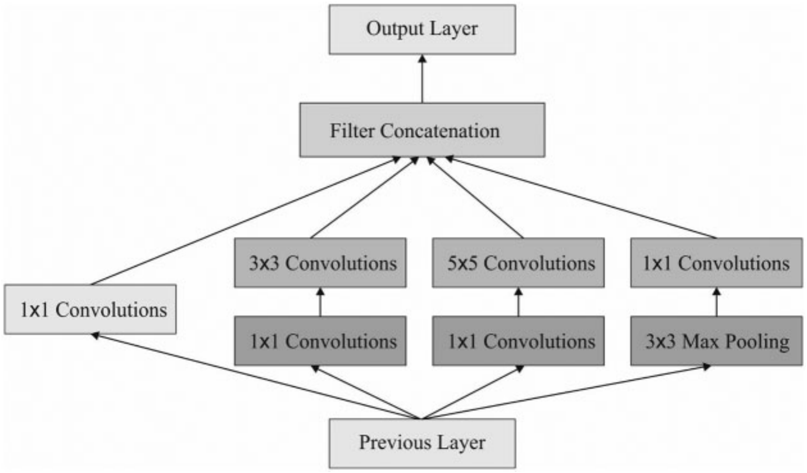

Figure 2: Process of Inception model

(i) The nodes with one input edge don't require some feature fusion. Thus, the nodes were distant.

(ii) When the input as well as output nodes are at a similar level, then a further edge was additional among them.

(iii) The 2-way path was constructed that is repeated several times for obtaining superior feature fusion.

(iv) Execute weight fusion provided as:

$$O = \sum\nolimits_m \frac{w_m}{\in + \sum\nolimits_n w_n} . \ I_m, \tag{3}$$

whereas I_m implies the input features at level m and w_m signifies the learnable weighted input features at level m. Value of $\in$ has smaller arbitrary value nearby and superior to 0.

(v) Combine MFPN multiscale connection with weight fusion as provided under:

$$F_{inter}^{n} = Convol\left(\frac{w_1 \cdot F_{in}^{n} + w_2 \cdot \ \mathrm{Res}ize\left(F_{in}^{n+1}\right)}{w_1 + w_2 + \in}\right) \tag{4}$$

$$F_{out}^{n} = Convol\left(\frac{w_1' \cdot F_{in}^{n} + w_2' \cdot F_{inter}^{n} + w_3' \cdot \ \mathrm{Res}ize\left(F_{out}^{n-1}\right)}{w_1' + w_1' + w_1' + \in}\right) \tag{5}$$

In where F_{inter}^{n} denotes the feature at middle level n on top-down paths of MFPN and F_{out}^{n} implies the resultant feature at level n on bottom-up paths of MFPN.

3.4 Hyperparameter Optimization

In order to effectually choose the hyperparameter values of the Inception model [22–24], the APO algorithm is utilized. AEO is based on the three energy transfer techniques including decomposition, consumption, and production in the ecosystem [25]. In the production method, the operator enables AEO to randomly generate a new individual that replaces the prior one among an individual $(r\,(U - L) + L)$ and the best individual (x_n) randomly produced in the searching area and it is formulated by the subsequent Eq. (6):

$$x_1(t+1) = \left(1-\left(1-\frac{t}{T}\right)r_1\right)x_n(t) + r_1\left(1-\frac{t}{T}\right)(r\,(U-L)+L) \tag{6}$$

Let, the (n) and T be the size of population and the highest iteration count, correspondingly, the variables (U) and (L) correspondingly represent the lower and upper limits. Further, (r) and (r_1) illustrates a random vector and number lies within zero and one. In addition, $(r(U-L)+L)$ and $((1-t/T)\ r_1)$ demonstrates a position of an individual and linear weight coefficient, that is randomly produced in the searching region correspondingly. During the consumption model, Levy flight is included in nature-inspired algorithm that effectively explores the searching area. Parameter-free random walk named consumption factor using the features of Levy flight.

$$C = \frac{1}{2}\frac{v_1}{|v_2|}v_1 \approx N(O,1), v_2 \approx N(0,1) \tag{7}$$

In Eq. (7), standard distribution with mean (0) and standard deviation (1) can be represented as $N\,(O,1)$. Thus, the consumption factors might help various kinds of consumers for adopting three consumption approaches. The initial process is Herbivore, where the consumer consumes the producer. Such behaviors are modeled by the following expression:

$$x_i(t+1) = x_i(t) + C\cdot(x_i(t)-x_1(t))\,, i \in [2,\ldots n] \tag{8}$$

The next process is Carnivore, where the consumer eats a consumer in a random manner using the highest level of energy. Such behaviors are modelled by Eq. (9):

$$\begin{cases} x_i(t+1) = x_i(t) + C\cdot(x_i(t)-x_1(t))\,,\ i \in [3,\ldots n] \\ j = r\,([2i-1]) \end{cases} \tag{9}$$

The secondary process is Omnivore, where the consumer eats producer and consumer at a random manner with the highest level of energy and it is given in the following:

$$\begin{cases} x_i(t+1) = x_i(t) + C\cdot(r_2\cdot x_i(t)-x_1(t)) \\ +(1-r_2)\left(x_j(t)-x_j(t)\right),\ i=3,\ldots nj = r\,([2i-1]) \end{cases} \tag{10}$$

During decomposition process, the *i-th* individual x_i location in a population is upgraded, with the decomposer x_n location through the weight coefficient $(r_3\cdot r\,([12])-1$ and $2\cdot r_3-1)$ and the decomposition factor $=3u, u\approx N(O,1)$. Therefore, the process demonstrates exploitation to certain degree since it allows the following location of every individual to spread around the optimal individual and it is formulated by:

$$\begin{cases} x_i(t+1) = x_n(t) + 3u\cdot((r_3\cdot r\,([12])-1)\cdot x_n(t) \\ -(2\cdot r_3-1)\cdot x_j(t)), i=1,\ldots n, u\approx N(O,1) \end{cases} \tag{11}$$

AEO is initiated by producing g a population in a random manner. The initial search individual, at all the iterations, updates their location based on Eq. (6), but, there is an equal probability, for other individuals, to select between Herbivore, Carnivore, and Omnivore as in Eqs. (8)–(10) for updating the position. It is accepted if a better function value is attained using an individual. Next, the location of all the individuals is upgraded based on Eq. (11). In the updating procedure, an individual is arbitrarily generated in the searching space, if it is farther from the lower or upper bounds. Until the AEO process with an end condition is fulfilled, every single update are interactively carried out. At last, the solution of best individual is attained.

Algorithm 1: Pseudocode of AEO

```
Arbitrarily initializing an ecosystem $X_i$ (solutions) and compute the fitness $Fit_i$, and $X_{best}$, $=$ an optimum solution initiates so far.
While the end condition is not fulfilled do
For individual $X_1$, upgrade their solution.
For individual $X_i$ $(i = 2, \ldots, n\ )$,
If $rand < 1/3$ then upgrade their solution,
Else If $\frac{1}{3} \leq rand \leq 2/3$ then upgrade their solution
Else upgrade their solution,
End If.
End lf.
Compute the fitness of all the individuals.
Upgrade the optimal solutions establish so far $X_{best}$.
Upgrade the place of all the individuals.
Compute the fitness of all the individuals.
Upgrade an optimum solution initiate so far $X_{best}$,
End While.
Return $Xbest$.
```

4 Result and Discussion

In this section, the vehicle detection performance of the AEODCNN-VD model is tested using two datasets namely VEDAI dataset and VEDAI512 dataset [26]. The VEDAI is a dataset for vehicle recognition from the aerial images offered as a tool to benchmark automatic target detection techniques from unconstrained environments and gathered in Utah, USA.

Tab. 1 and Fig. 3 illustrate the vehicle detection results of the AEODCNN-VD model on VEDAI dataset. The experimental results reported that the AEODCNN-VD model has recognized the vehicles properly under every run.

Table 1: Result analysis of AEODCNN-VD approach with distinct runs under VEDAI dataset

VEDAI dataset			
No. of runs	Precision	Recall	F1-score
Run-1	95.59	97.68	94.15
Run-2	95.21	97.28	94.50
Run-3	96.49	95.90	95.62
Run-4	98.00	95.45	97.07
Run-5	96.16	95.45	97.37
Run-6	95.29	97.07	97.24
Run-7	96.63	96.34	95.33
Run-8	96.94	95.72	95.69
Run-9	95.46	97.87	94.22
Run-10	95.89	96.05	94.71
Average	96.17	96.48	95.59

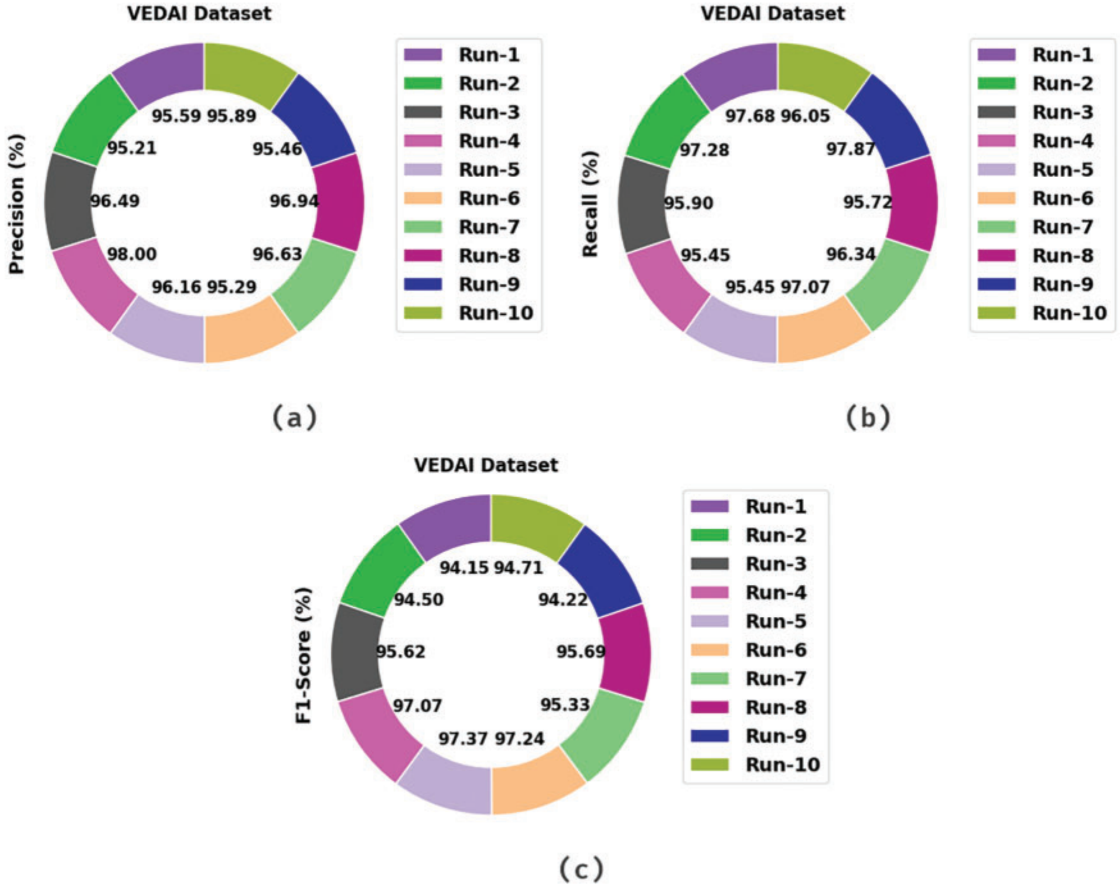

Figure 3: Result analysis of AEODCNN-VD approach under VEDAI dataset (a) $Prec_n$, (b) $Reca_l$, and (c) $F1_{score}$

For instance, with run-1, the AEODCNN-VD model has offered $prec_n$ of 95.59%, $reca_l$ of 97.68%, and $F1_{score}$ of 94.15%. Meanwhile, with run-4, the AEODCNN-VD approach has obtainable $prec_n$ of 98%, $reca_l$ of 95.45%, and $F1_{score}$ of 97.07%. Eventually, with run-8, the AEODCNN-VD system has accessible $prec_n$ of 96.94%, $reca_l$ of 95.72%, and $F1_{score}$ of 95.69%. At last, with run-10, the AEODCNN-VD algorithm has obtainable $prec_n$ of 95.89%, $reca_l$ of 96.05%, and $F1_{score}$ of 94.71%.

The training accuracy (TA) and validation accuracy (VA) achieved by the AEODCNN-VD approach under VEDAI dataset is demonstrated in Fig. 4. The experimental outcome represented that the AEODCNN-VD algorithm has gained increased values of TA and VA. In specific, the VA has appeared that superior to TA.

The training loss (TL) and validation loss (VL) reached by the AEODCNN-VD approach under VEDAI dataset are established in Fig. 5. The experimental outcome revealed that the AEODCNN-VD methodology has been able least values of TL and VL. In specific, the VL seemed to be lower than TL.

A detailed ROC investigation of the AEODCNN-VD approach under VEDAI dataset is represented in Fig. 6. The figure portrayed that the AEODCNN-VD approach has resulted in proficient results with maximal ROC value.

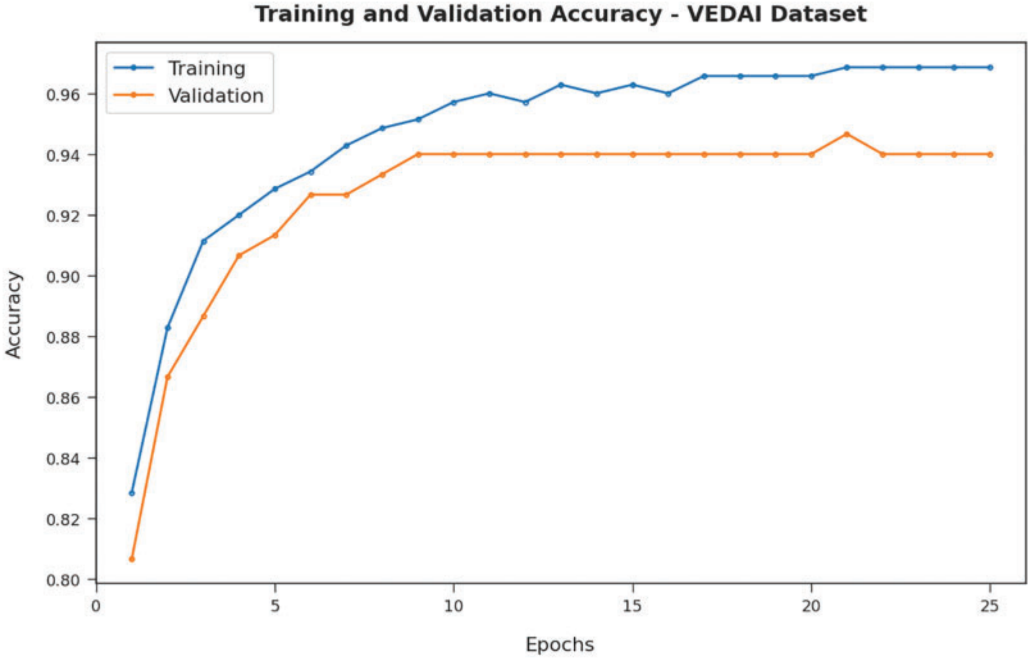

Figure 4: TA and VA analysis of AEODCNN-VD approach under VEDAI dataset

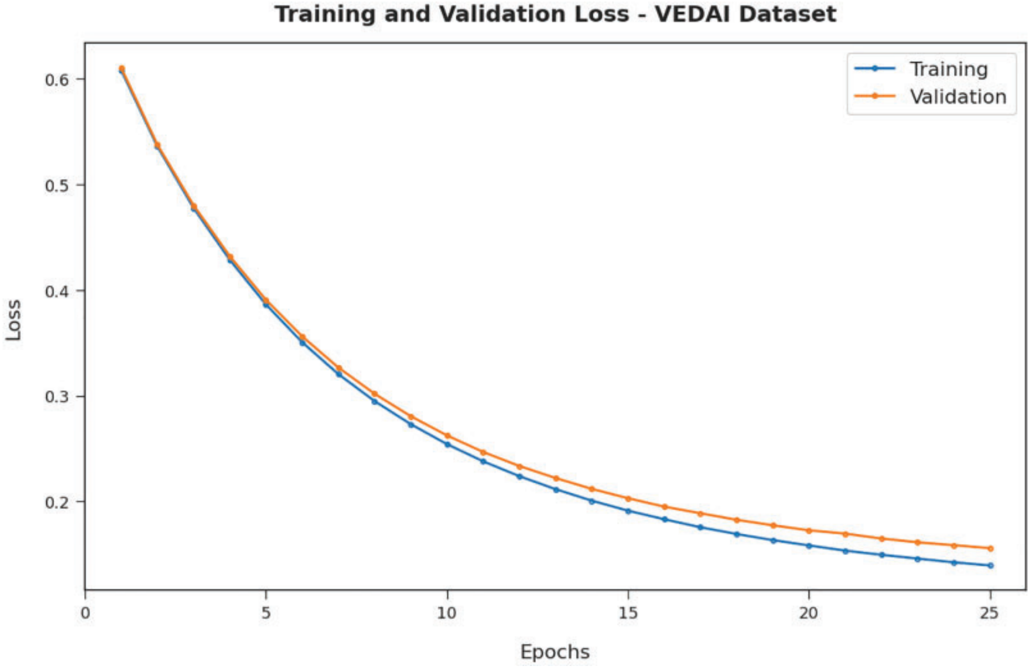

Figure 5: TL and VL analysis of AEODCNN-VD approach under VEDAI dataset

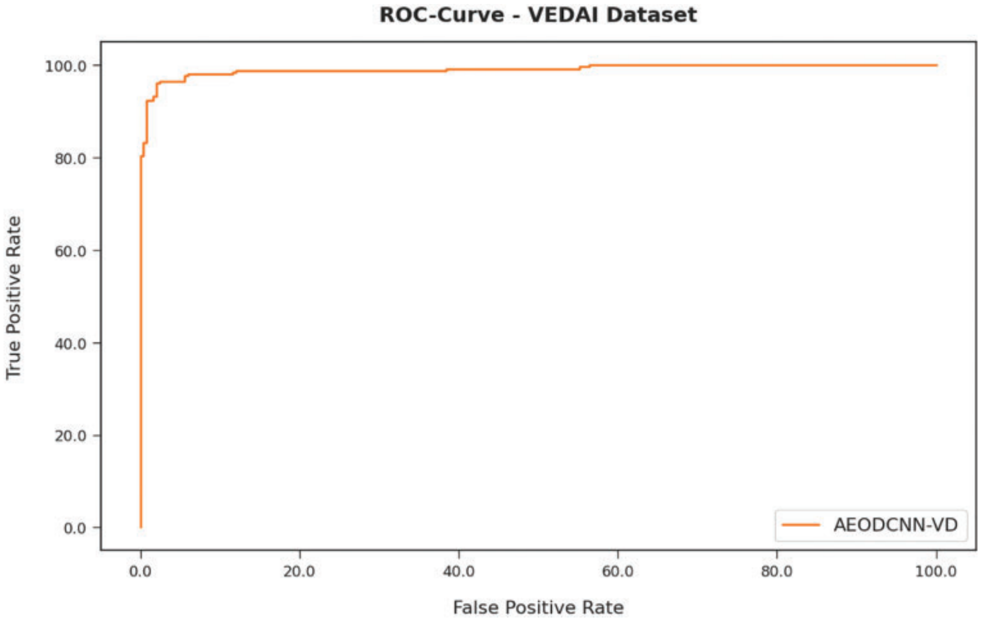

Figure 6: ROC analysis of AEODCNN-VD approach under VEDAI dataset

To assure the enhanced detection efficiency of the AEODCNN-VD model, a comparison study is made with existing models on VEDAI dataset as illustrated in Tab. 2 [27]. Fig. 7 illustrates a comparative $prec_n$ investigation of the AEODCNN-VD model with recent algorithms on VEDAI dataset. The figure implied that the SSD512 approach has shown least performance with $prec_n$ of 76.96%. Besides, the FRCNN and CRCNN systems have demonstrated somewhat enhanced $prec_n$ of 81.86% and 83.59% respectively. In line with, the FRCNN-FPN, CRCNN-FPN, RetinaNet, FCOS, FoveaBox, and MA-FPN models have reached reasonable $prec_n$ of 89.23%, 88.51%, 87.26%, 86.90%, 86.46%, and 89.72% respectively. But the AEODCNN-VD model has outperformed other models with maximum $prec_n$ of 96.17%.

Table 2: Comparative analysis of AEODCNN-VD approach with existing algorithms under VEDAI dataset

Methods	Precision	Recall	F1-score
AEODCNN-VD	96.17	96.48	95.59
FRCNN	81.86	87.69	84.54
FRCNN-FPN	89.23	91.87	90.21
CRCNN	83.59	86.78	85.19
CRCNN-FPN	88.51	89.93	89.25
SSD512	76.96	91.08	83.88
RetinaNet	87.26	94.13	91.70
FCOS	86.90	93.50	89.53
FoveaBox	86.46	90.37	88.14
MA-FPN	89.72	94.17	93.07

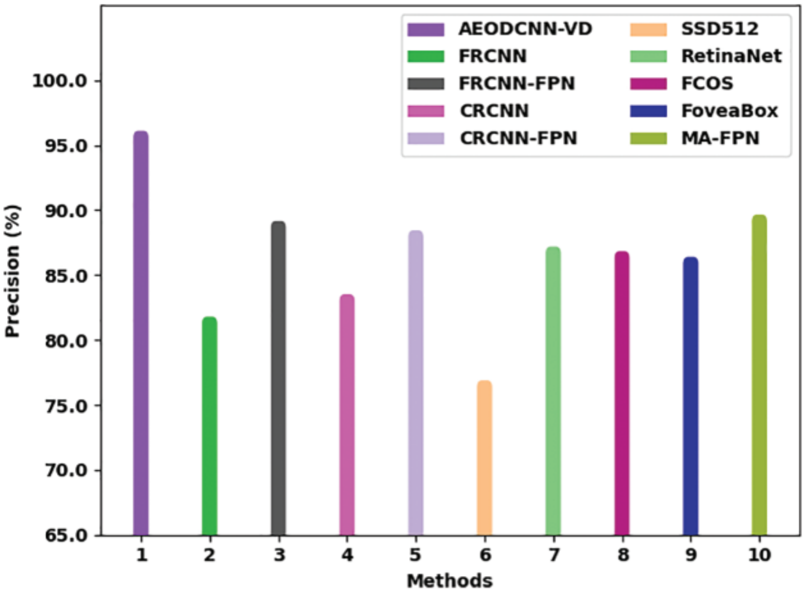

Figure 7: $Prec_n$ analysis of AEODCNN-VD approach under VEDAI dataset

Fig. 8 depicts a comparative $reca_l$ investigation of the AEODCNN-VD method with recent methodologies on VEDAI dataset. The figure exposed that the SSD512 approach has shown least performance with $reca_l$ of 91.08%. Besides, the FRCNN and CRCNN techniques have exhibited somewhat enhanced $reca_l$ of 87.69% and 86.78% correspondingly. In addition, the FRCNN-FPN, CRCNN-FPN, RetinaNet, FCOS, FoveaBox, and MA-FPN models have reached reasonable $reca_l$ of 91.87%, 89.93%, 94.13%, 93.50%, 90.37%, and 94.17% correspondingly. At last, the AEODCNN-VD algorithm has depicted other models with maximal $reca_l$ of 96.48%.

Fig. 9 demonstrates a comparative $F1_{score}$ examination of the AEODCNN-VD approach with recent approaches on VEDAI dataset. The figure revealed that the SSD512 model has outperformed least performance with $F1_{score}$ of 83.88%. Followed by, the FRCNN and CRCNN approaches have demonstrated somewhat maximal $F1_{score}$ of 84.54% and 85.19% correspondingly. Along with that, the FRCNN-FPN, CRCNN-FPN, RetinaNet, FCOS, FoveaBox, and MA-FPN methodologies have reached reasonable $F1_{score}$ of 90.21%, 89.25%, 91.70%, 89.53%, 88.14%, and 93.07% correspondingly. Eventually, the AEODCNN-VD algorithm has portrayed other models with higher $F1_{score}$ of 95.59%.

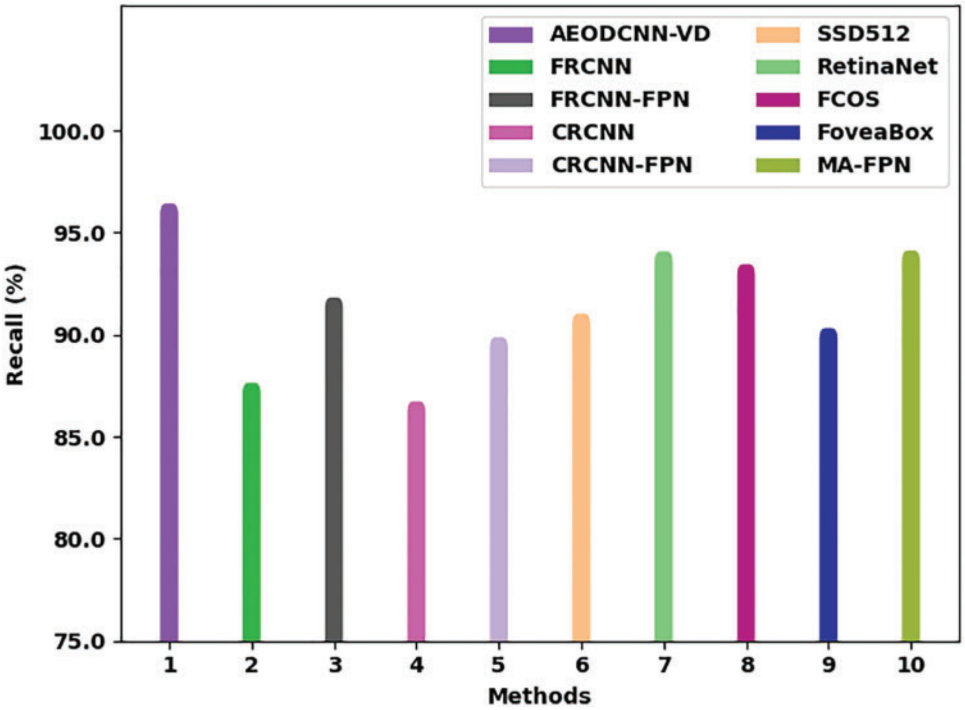

Figure 8: $Reca_l$ analysis of AEODCNN-VD approach under VEDAI dataset

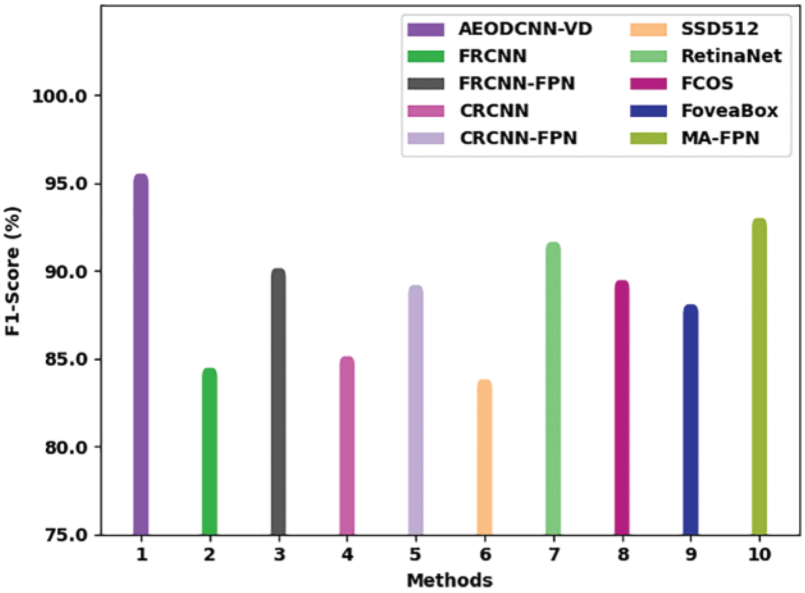

Figure 9: $F1_{score}$ analysis of AEODCNN-VD approach under VEDAI dataset

5 Conclusion

In this study, a novel AEODCNN-VD algorithm was introduced for the identification of vehicles accurately and rapidly on RSIs. The proposed AEODCNN-VD model primarily utilized an SSD based object detector with Inception network as a baseline model. Besides, the features from the Inception model are passed into the MFPN for multiway and multiscale feature fusion, which are then passed into bounding box and class prediction networks. At the final stage, the AEO based hyperparameter optimizer is used for Inception network. The experimental result analysis of the AEODCNN-VD model is carried out using two benchmark datasets. The extensive comparative study of the proposed AEODCNN-VD model showed a promising performance over recent DL models. In future, advanced DL models can be applied to boost detection efficiency.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4210118DSR28). The author would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. Li, G. Wan, G. Cheng, L. Meng and J. Han, “Object detection in optical remote sensing images: A survey and a new benchmark,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 296–307, 2020.

[2] S. Gadamsetty, R. Ch, A. Ch, C. Iwendi and T. R. Gadekallu, “Hash-based deep learning approach for remote sensing satellite imagery detection,” *Water*, vol. 14, no. 5, pp. 707, 2022.

[3] G. Cheng and J. Han, “A survey on object detection in optical remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.

[4] K. Li, G. Cheng, S. Bu and X. You, “Rotation-insensitive and context-augmented object detection in remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2337–2348, 2017.

[5] X. Li, J. Deng and Y. Fang, “Few-shot object detection on remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[6] S. A. Fatima, A. Kumar, A. Pratap and S. S. Raoof, “Object recognition and detection in remote sensing images: A comparative study,” in *2020 Int. Conf. on Artificial Intelligence and Signal Processing (AISP)*, Amaravati, India, pp. 1–5, 2020.

[7] I. Abunadi, M. M. Althobaiti, F. N. Al-Wesabi, A. M. Hilal, M. Medani *et al.,* “Federated learning with blockchain assisted image classification for clustered UAV networks,” *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1195–1212, 2022.

[8] F. N. Al-Wesabi, M. Obayya, M. Hamza, J. S. Alzahrani, D. Gupta *et al.,* “Energy aware resource optimization using unified metaheuristic optimization algorithm allocation for cloud computing environment,” *Sustainable Computing: Informatics and Systems*, vol. 35, pp. 100686, 2022.

[9] X. Sun, P. Wang, Z. Yan, F. Xu, R. Wang *et al.,* “FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 184, pp. 116–130, 2022.

[10] F. Xiaolin, H. Fan, Y. Ming, Z. Tongxin, B. Ran *et al.,* “Small object detection in remote sensing images based on super-resolution,” *Pattern Recognition Letters*, vol. 153, pp. 107–112, 2022.

[11] J. Liu, D. Yang and F. Hu, “Multiscale object detection in remote sensing images combined with multi-receptive-field features and relation-connected attention,” *Remote Sensing*, vol. 14, no. 2, pp. 427, 2022.

[12] Y. Liu, S. Zhang, Z. Wang, B. Zhao and L. Zou, "Global perception network for salient object detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[13] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei *et al.,* "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 3–22, 2018.

[14] L. Zhou, H. Wei, H. Li, W. Zhao, Y. Zhang *et al.,* "Arbitrary-oriented object detection in remote sensing images based on polar coordinates," *IEEE Access*, vol. 8, pp. 223373–223384, 2020.

[15] M. Sharma, M. Dhanaraj, S. Karnam, D. G. Chachlakis, R. Ptucha *et al.,* "YOLOrs: Object detection in multimodal remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1497–1508, 2021.

[16] Z. Zakria, J. Deng, R. Kumar, M. S. Khokhar, J. Cai *et al.,* "Multiscale and direction target detecting in remote sensing images via modified YOLO-v4," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1039–1048, 2022.

[17] J. Chen, J. Sun, Y. Li and C. Hou, "Object detection in remote sensing images based on deep transfer learning," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12093–12109, 2022.

[18] Y. Ye, X. Ren, B. Zhu, T. Tang, X. Tan *et al.,* "An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images," *Remote Sensing*, vol. 14, no. 3, pp. 516, 2022.

[19] L. Zheng, C. Fu and Y. Zhao, "Extend the shallow part of single shot multibox detector via convolutional neural network," in *Tenth Int. Conf. on Digital Image Processing (ICDIP 2018)*, Shanghai, China, pp. 141, 2018.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Thirty-First AAAI Conf. on Artificial Intelligence*, California, USA, pp. 4278–4284, 2017.

[21] P. Kaur, B. S. Khehra and A. P. S. Pharwaha, "Deep transfer learning based multiway feature pyramid network for object detection in images," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–13, 2021.

[22] K. Shankar, E. Perumal, M. Elhoseny, F. Taher, B. B. Gupta *et al.,* "Synergic deep learning for smart health diagnosis of COVID-19 for connected living and smart cities," *ACM Transactions on Internet Technology*, vol. 22, no. 3, pp. 16: 1–14, 2022.

[23] D. K. Jain, Y. Li, M. J. Er, Q. Xin, D. Gupta *et al.,* "Enabling unmanned aerial vehicle borne secure communication with classification framework for industry 5.0," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5477–5484, 2022.

[24] M. Y. Sikkandar, B. A. Alrasheadi, N. B. Prakash, G. R. Hemalakshmi, A. Mohanarathinam *et al.,* "Deep learning based an automated skin lesion segmentation and intelligent classification model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3245–55, 2021.

[25] R. M. R. Allah and A. A. El-Fergany, "Artificial ecosystem optimizer for parameters identification of proton exchange membrane fuel cells model," *International Journal of Hydrogen Energy*, vol. 46, no. 75, pp. 37612–37627, 2021.

[26] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016.

[27] X. Li, F. Men, S. Lv, X. Jiang, M. Pan *et al.,* "Vehicle detection in very-high-resolution remote sensing images based on an anchor-free detection model with a more precise foveal area," *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, pp. 549, 2021.