Combining Deep Learning and Elephant Herding Optimization for Pedestrian Detection from a Drone-based Images

Saeed Mohammed Baneamoon

Department of Computer Engineering, College of Engineering & Petroleum, Hadhramout University, Yemen

Abstract- In computer vision, detecting and classifying objects is of great importance in various fields such as transportation, healthcare, manufacturing and agriculture. On the other hand, in the field of transportation, there are different techniques are used to develop many applications related to pedestrian detection such as driver assistance system, intelligent video surveillance system, and emergency victim rescue. Therefore, this paper proposes an effective method for the problem of pedestrian detection from a drone-based images using Elephant Herding Optimization (EHO) to optimize the control and selection of the parameters and convergence speed in a deep learning architecture in order to improve finding a subset of features from a larger feature pool that provided better accurate classification and evaluation of pedestrian detection. The proposed method is evaluated by experimenting a number of pedestrian images obtained from VisDrone-Dataset and the results shows better and more effective detection and provide better and more effective classification in term of precision, 99%; recall, 98%; and F1 measure, 99%.

Keywords— Computer Vision, Object Recognition, Pedestrian detection, Deep learning, Elephant Herding Optimization (EHO)

I. INTRODUCTION

Object recognition is computer vision technique aims to identifying objects in digital images or video. Due to the attention in drones for surveillance, digital image and video capture and implementation of autonomous operations in various fields such as transportation, healthcare, manufacturing and agriculture., so many related experiments are still being conducted to solve related problems to reach optimal solutions.

Pedestrian detection is one of the most important issue in the field of transportation, although there are many applications to identify pedestrians, studies are still ongoing to develop them, especially those based on capturing digital images or video by drone in unstable weather conditions such as wind, rain, snowing and fog, or surroundings with severe lighting, or too crowded environments.

Most of these studies are based on implementing a different deep learning-based approach to pedestrian detection. Nevertheless, some limitations in these techniques still need to be developed. This motivated adopting a model for detecting pedestrian from a drone-based images.

Therefore, this paper presents a method for detecting pedestrian from a drone-based images by combining deep learning and Elephant Herding Optimization (EHO), where EHO is used to optimize the control and selection of the parameters and convergence speed in a deep learning architecture in order to improve accurate classification and evaluation of pedestrian detection.

II. RELATED WORK

In computer vision, objects recognition techniques for detecting a pedestrian from a drone-based images are more difficult to apply than other types of images. However, Many researchers have attempted to employ deep learning technology to solve problems of the pedestrian detection from a drone-based images. Marusic et al, in [1] have been demonstrated person detection from aerial perspective on, this method based on the visible spectrum using Faster R-CNN model and was tested on the HERIDAL dataset, and it was evaluated with a total of 20 images with human presence from each of 12 locations and also added 8 more on which there is no human presence which achieved 88.3% recall with precision of 67.3%. Li et al.[2]. Have been proposed deep learning-based method that employed AlexNet for Pedestrian detection, the method was evaluated with INRIA database which include 784 pedestrian training images and 218 pedestrian testing images and the experimental results demonstrated 23% miss rate of deep features as compared with 46% miss rate for hand-crafted HOG features. Zhang et al. in [3], have been used a forward deep learning for pedestrian detection, this method based on Faster R-CNN combing K-means, where R-CNN for image features extraction and K-means for cluster analysis, this method was tested on the INRIA pedestrian dataset and another pedestrian dataset which was collected in Beijing Union University these are not drone-based images and it was evaluated with a total of ten thousand training images and one thousand testing images which achieves the accuracy of 92.7% precision, and 87.60% recall. Lan et al. in [4], have been proposed a new network structure YOLO-R.First, by adding Passthrough layer to the original YOLO network which connect the shallow layer pedestrian features to the deep layer pedestrian features and link the high and low resolution pedestrian features, this development was tested on the INRIA pedestrian dataset which are not drone-based images and it was effectively improved the detection accuracy of pedestrians and the detection speed reached 25 frames per second. Yi Zhang et al. in [5], have been adopted a real-time pedestrian detection algorithm based on tiny-yolov3 to decrease loses part of the pedestrian detection accuracy in real-time, this improvement was tested on the VOC dataset to form the pedestrian dataset which are not drone-based images and the experimental results showed better accuracy as compared to the original tiny-yolov3 with 73.98% precision rate. Li et al. in [6] have been employed faster R-CNN for pedestrian detection from color or thermal images with illumination conditions, this was tested on KAIST dataset consists of 95,328 colorthermal image pairs recorded by a color and a thermal cameras mounted on the rooftop of a car at a equal frame rate of 20 fps, and the experimental results indicated the effectiveness of the proposed R-CNN. Jeon et al. in [7] have been integrated deep learning and combining a new type of local pattern with the RGB raw image with Triangular Patterns as input for pedestrian detection, this enhancement was trained and tested on the KITTI dataset and the experimental results reported effective in extracting more detailed and stable features from local regions. Burke et al. in [8] have been demonstrated that thermal infrared equipped drones have the potential to be used for search and rescue, and development of drones offer a significant advantage in reducing the risks associated with SAR, in decreasing time taken to find people in need of rescue, this pilot studied shown that in real-world search and rescue scenarios, humans were automatically detected using temperature and size thresholding method as well as thermal infrared data obtained via a drone. Bozié-S›tulié et al. in [9] have been proposed person detection in UAV aerial images for search and rescue tasks by reducing the search space through detection the salient segments in the image. this improvement was tested on the HERIDAL database the experimental results achieved a detection rate of 88.9% and a precision of 34.8%, which performed better effectiveness than the system currently used in [1]. Hung et al. in [10] have been developed a deep learning-based model for pedestrian detection from a drone-based images, in this model a Faster R-CNN was used to search for the present of a pedestrian inside the captured drone-based images, this enhancement was trained and tested on a total of 1500 images were captured by S30W drone at different places, with various weather conditions, and at daytime and night-time, and the experimental results reported 98% precision, 99% recall, and 98% F1 measure, very similar results reported also by conducted of Faster R-CNN with YOLO deep model on UAV123 publicly available dataset. Sama in [11] have been enhanced a deep learning-based model for the problem of pedestrian detection from a drone-based images, and it was evaluated with a total of 900 images were collected by S30W drone from different places, with various views and weather conditions, and at daytime and night-time, and results achieved a promising result with 96% precision, 97 % recall, and 96% F1 measure.

III. Deep Learning Algorithms

Deep learning is a class of machine learning methods that attempts to simulate the human brain behaviour and structure, it aims to enabling systems to cluster data and make predictions with high accuracy. It is based on a bio-inspired artificial neural network architecture which is essentially with three or more processing layers. In learning process a combination of data inputs, weights, and bias work together to accurately recognize, classify, and describe objects within the data. Before starting the deep learning process, the input data must be carefully selected in order to decrease the risk of overfitting during training data. For example, in image processing there are different types of input data such as a vector of pixel intensity values, a variety of edges, and regions of a certain shape [12].

The two major types of learning in deep learning are supervised learning and unsupervised learning. Supervised learning utilizes labeled datasets to categorize or make predictions; this requires some kind of human intervention to label input data correctly. In contrast, unsupervised learning doesn't require labeled datasets, and instead, it detects patterns in the data, clustering them by any distinguishing characteristics [12,13].

There are different deep learning architectures are used in computer vision field such as CNNs, auto-encoders and deep belief networks. According to the types of learning in deep learning, the CNNs is a type of Feed-Forward Neural Networks (FFNNs) use supervised learning, while other architectures are type of Feed-Backward Neural Networks (FBNNs) use unsupervised learning [13,14].

However, the deep learning architecture can be classified into three classifications are generative, discriminative, and hybrid architectures. In generative deep architectures, the high-order interconnection of the visible data is described in order to perform effective pattern analysis such as auto-encoders deep learning architecture. As for discriminative deep architectures the rear allocation of classes is described based on the visible data in order to perform pattern classification such as CNNs deep learning architecture. While hybrid deep architecture is a combination of generative and discriminative architectures, where the optimized outcomes of generative architecture help in the discrimination process. Such as the hybrid type of the deep belief network with the CNN deep learning architectures, on the other hand, a hybrid deep learning architecture is a technique that enhances analysis in image classification, segmentation and detection [12,15].

IV. Elephant Herding Optimization (EHO)

Elephant Herding Optimization (EHO) is one of the inspired algorithms that optimize solutions by emulating animal herding behaviors, Elephants live as socially structured creatures of females and calves as shown in Figure 1. The elephant population consists of some clans, and the elephant clan consists of a fixed number of elephants headed by the matriarch, where the females like to live with family members, while the male members tend to live elsewhere and gradually become independent from their families until they leave their family completely as shown in Figure 2. The EHO technique proposed by Wang et al. in 2015 [16] was developed based on a study of natural elephant herding behavior.

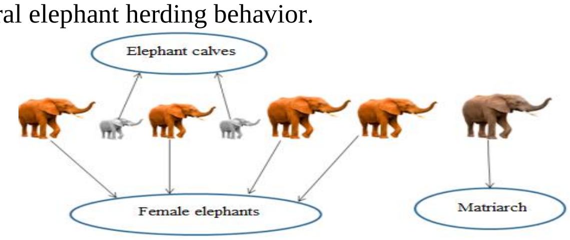

Figure1: Components of the elephant population

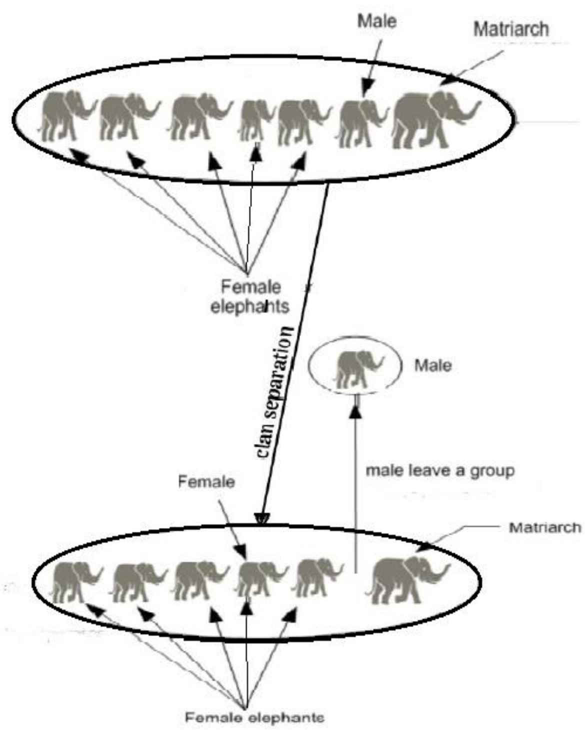

Figure2: Male elephant separation

Exploration and exploitation in EHO are achieved by the clan updating operator and the separating operator as follows:

A. Clan-updating Operator

Based on the living behavior of the elephants, each clan is led by the matriarch, therefore, the new status of each elephant *ci* is influenced by matriarch *ci*. The elephant *j* in clan *ci* can be calculated by Equation (1).

$$x_{new,ci,j} = x_{ci,j} + a \times (x_{best,ci} - x_{ci,j}) \times r \quad (1)$$

where $x_{new,ci,j}$ and $x_{ci,j}$ present the new and old position for elephant *j* in clan *ci*, respectively.

$x_{best,ci}$ is matriarch *ci* which represents the best elephant in the clan. a$\in$[0,1] indicates a scale factor, r $\in$[0,1].

The best elephant can be calculated by Equation (2).

$$x_{new,ci,j} = \beta \times x_{center,ci} \quad (2)$$

where $\beta \in$[0,1] represents a factor which determines the influence of the $x_{center,ci}$ on $x_{new,ci,j}$.

$x_{new,ci,j}$ is the new individual.

$x_{center,ci}$ is the center individual of clan *ci*. It can be calculated by Equation (3) for the *d*-th dimension.

$$x_{center,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{ci}} x_{ci,j,d} \quad (3)$$

where 1≤ d≤ D and n_{ci} indicate the number of elephants in clan *ci*,

$x_{ci,j,d}$ represent the *d*-th dimension of elephant individual $x_{ci,j}$.

$x_{center,ci}$ is the center of clan *ci* and it can be updated by Equation (3).

B. Separating Operator

When solving optimization problems, a separating operator can be designed for the separating process by which male elephants leave their family group. The separating operator is implemented by the elephant with the worst fitness in each generation, as shown in Equation (4).

$$x_{worst,ci} = x_{\min} + (x_{\max} - x_{\min} + 1) \times rand \quad (4)$$

where $x_{\max}$ represents the upper bound of the individual

$x_{\min}$ indicates lower bound of the individual.

$x_{worst,ci}$ indicates the worst individual in clan *ci*.

Rand [0, 1] is a stochastic distribution between 0 and 1.

According to the description of the clan-updating operator and separating operator, the mainframe of EHO is summarized. The corresponding flowchart is shown as follows.

The corresponding flowchart of the EHO algorithm can also be seen in Figure 3.

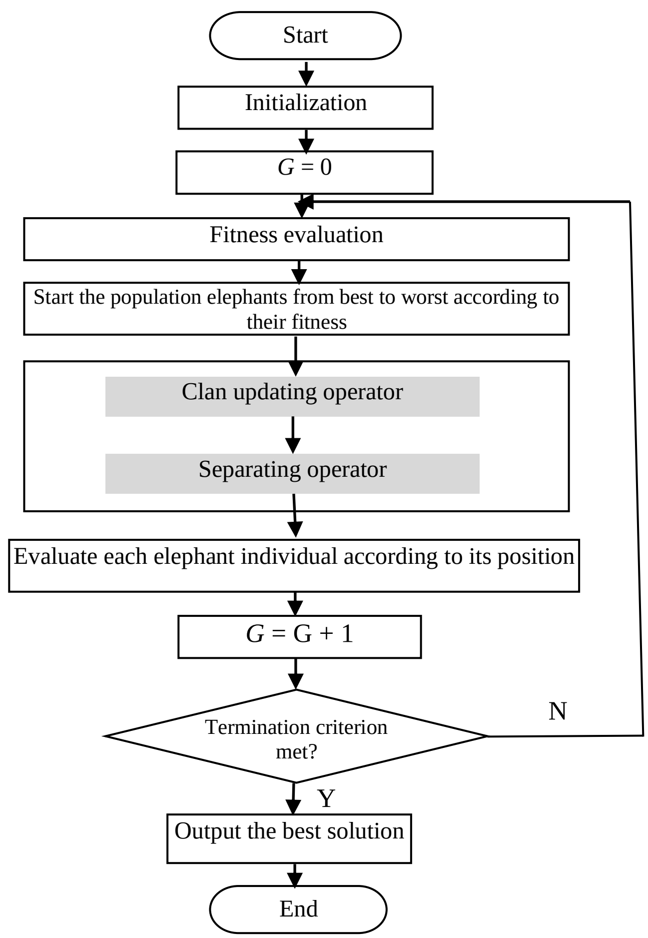

Figure3: Flowchart of the EHO algorithm

Elephant herding optimization algorithm is outlined as follow:

EHO Algorithm

Begin
Initialization. Set the initialize iterations G = 1; initialize the population P randomly; set maximum generation MaxGen.
While stopping criterion is not met do
Sort the population according to fitness of individuals.
For all clans ci do
For elephant j in the clan ci do // Updating Operator
Generate $x_{new,ci,j}$ and update $x_{ci,j}$ by Equation (1).
If $x_{ci,j} = x_{best,ci}$ then
Generate $x_{new,ci,j}$ and update $x_{ci,j}$ by Equation (2).
End if
End for
End for
For all clans ci do // Separating Operator
Replace the worst individual ci by Equation (4).
End for
Evaluate each elephant individual according to its position.
T = T + 1.
End while
End.

V. PROPOSED METHOD

In computer vision, due to the important of enhancing detection and classification of pedestrian from a drone-based images, which are affected by changes in illuminations and weather conditions, therefore, to solve this problem, the present paper introduces an effective method for the problem of pedestrian detection from a drone-based images through combining deep learning and Elephant Herding Optimization (EHO), where EHO is used to optimize the control and selection of the parameters and convergence speed in a deep learning architecture in order to improve accurate classification and evaluation of pedestrian detection.

In the proposed method, a faster-RCNN has been considered as shown in Figure 4 consists of input layer, hidden layer, and output layer, where the hidden layer contains convolutional layers, Region Proposal Network (RPN) and Fully Connected (FC).

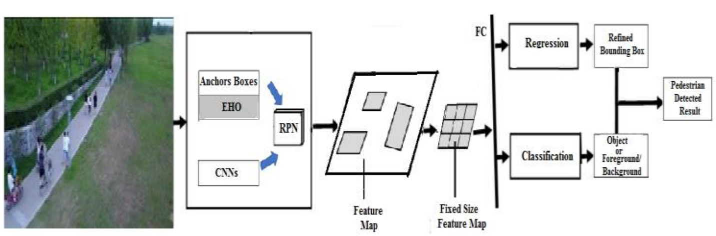

Figure 4. The architecture of deep learning model

In CNNs stage, the CNN is initialized by a pre-trained model, where the images input of the input layer that are input of deep convolutional layers to extract deep image features (feature extraction). In the proposed method, Inception-v3 is used, it is a convolutional neural network that is 48 layers deep and the output is of this stage is a feature map for possible candidate foreground/background objects. furthermore, Anchor Boxes are the most important components of RPN, they are a set of predefined bounding and fixed-size boxes for the entire image, these boxes are defined in different sizes and different aspect ratios to capture different types of objects. On the other hand, to improve detection accuracy and ensure faster pedestrian detection, pedestrian features are enriched by using EHO to find the most suitable anchor boxes on pedestrian dataset, so that the more semantic information can be extracted. Moreover, RPN component takes all the anchor boxes as input and then is generating the objectness score for each candidate of target object locations, where a K predefined candidate boxes are generated from K anchor boxes at different scales and aspect-ratio, here K=9 due the chose 3 scale and 3 aspect-ratio, these 3 scales are 128x 128, 256×256 and 512×512 and the 3 aspects-ratio are 1:1, 2:1 and 1:2, as the result RPN predicts a total of 9 boxes for each candidate location associated with a probability of foreground /background. The last stage in deep learning network is FC for classification and localization the target object inside the input image, where linear regressor is used for refining and finding exact location of the classified object and soft max classifier is used for the classification. Finally, precision rate, recall rate, and F1 score measures are calculated by Equation (5), Equation (6) and Equation (7) consequently to ensure the effectiveness of the proposed method.

$$Precision = \frac{TruePositives}{TruePositives + FalsePsitives} \quad (5)$$

where *Precision* represents the ratio of true pedestrian detections (*TruePositive*) to the total number of pedestrian that the classifier predicted (*TruePositive* and *FalsePositive*)
True Positive represents the number of samples that successfully detected pedestrian as pedestrian
False Positive represents the number of samples that wrongly detected pedestrian as pedestrian

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (6)$$

where *Recall* represents the ratio of true pedestrian detections (*TruePositive*) to the total number of pedestrian in the data set (*TruePositive* and *FalseNegative*)
FalseNegative represents the number of samples that pedestrian is not detected as pedestrian

$$F1Score = 2 * \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (7)$$

where *F1 Score* represents the Harmonic Mean between precision and recall

On the other hand, the accuracy of target detection results is evaluated by calculating Overlap Coefficient (OC) by Equation (8) that indicates the ratio of pixels that belong to the target object (pedestrian) or background where the higher the value of OC, the better the detection performance.

$$OC = \frac{A_T \cap A_D}{A_T \cup A_D} \quad (8)$$

where A_T and A_D represent the target bounding box of the ground truth and detected from the model respectively.

VI. Experimental Results

All experiments of the proposed method were preformed on the platform with Matlab 2018b on an Intel(R) Core(TM) i7 CPU 3.0 GHz, 8GB RAM, Windows 10 Professional OS system. The images are obtained from VisDrone-Dataset including 1500 images were split as 1000 training images for DCNN and 500 testing images set to assess its performances. In order to evaluate the proposed method, the results obtained by combining deep learning and EHO is compared with the results produced without EHO.

Parameters for the EHO algorithm were set by conducting preliminary computational experiments. The elephant population size was set to 100 and it was divided into 4 clans. Maximal number of the fitness function evaluation was set to 1400. The influence of the matriarch (parameter α) was set to 0.6, while parameter β controlling the influence of the clan's center was 0.1. Search range was [-100,100] and 10-dimensional problems were considered. Fitness function was calculated as the accuracy of 10-fold cross validation.

In the proposed method, Anchor boxes is integrated with EHO, where the modified dataset of each box acts as clan for EHO and is used to evaluate the fitness of each elephant. After the fitness have been calculated clan updating operator and separating operator are implemented. The best fit elephants act as clan for the next generation. Using the best fit elephants the fitness of clan is again evaluated and classification accuracy is calculated.

The results of TP rate versus FP rate with/without EHO has been recorded by reducing the threshold value by 0.1 where the heights set to 0.9 and it decrease iteratively to 0.1 as shown in Table 1, it is mentioned that without EHO at higher threshold value (0.5) only 13 wrongly detected pedestrian and 10 are not detected as pedestrian, while it is mentioned that at higher threshold value (0.9) only 2 wrongly detected pedestrian and 22 are not detected as pedestrian. On the other hand, it is mentioned that with EHO at higher threshold value (0.5) only 5 wrongly detected pedestrian and 3 are not detected as pedestrian, while it is mentioned that at higher threshold value (0.9) no wrongly detected pedestrian and 13 are not detected as pedestrian.

Table 1: Detection outcomes at each threshold value with and without EHO

	Threshold	True Positive	False Positive	False Negative
Deep Learning (Without EHO)	0.5	772	13	10
	0.9	760	2	22
Combining Deep Learning and EHO	0.5	782	5	3
	0.9	772	0	13

The results were achieved at threshold of 0.9 shown in Table 2 reported without EHO that precision of 99.74%, recall of 97.18%, and F1 score measures of 98.44, where decreasing the threshold value to 0.5, the precision rate decreases from 99.74% to 97.60% rate, while with EHO that precision of 100%, recall of 98.34 %, and F1 score measures of 99.16, where decreasing the threshold value to 0.5, the precision rate decreases from 100% to 99.37% rate.

Table 2: Results of Precision, Recall and F1 Score with and without EHO

	Threshold	Precision %	Recall %	F1 Score
Deep Learning (Without EHO)	0.5	97.60	98.72	98.17
	0.9	99.74	97.18	98.44
Combining Deep Learning and EHO	0.5	99.37	99.62	99.49
	0.9	100.00	98.34	99.16

Finally, the performances of the proposed method on the training and testing images with and without set are shown in Table 3 where shows the results of precession rate, recall rate, and F1 score are very close.

Table 3: Performances on Training / Testing images with and without EHO

	Training/Testing images	Precision %	Recall %	F1 Score
Deep Learning (Without EHO)	Testing Images	97.60	98.72	98.17
	Training Images	98.06	95.68	98.11
Combining Deep Learning and EHO	Testing Images	99.37	99.62	99.49
	Training Images	99.20	98.12	99.01

In this paper, precision, recall and F1 Score are used to indicate the performance of the proposed method. The comparative experiment was carried out with and without EHO. As is shown in Figure 5, the accuracy of precision, recall and F1 Score with EHO is significantly higher than that without EHO.

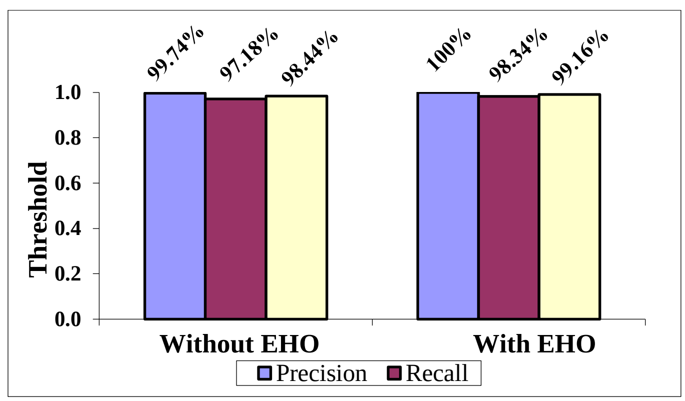

Figure 5. Comparison of the proposed method with and without combining of EHO

VII. Conclusions

In this paper, elephant herding optimization algorithm is combined with deep learning to optimize the control and selection of the parameters and convergence speed for enhancing detection and classification of pedestrian from a drone-based images. In order to prove the good performance of the proposed method, it was tested on VisDrone-Dataset. Based on the experimental and computational results, it can be concluded that EHO in the proposed method has good characteristics as optimization algorithm and it perform better for pedestrian detection and classification in term of precision, recall and F1 measure.

REFERENCES

[1] Z. Marusic, D. Bozic-S›tulic, S. Gotovac, T. Marusié, "*Region Proposal Approach for Human Detection on Aerial Imagery*," in Proceedings of the 3rd International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 2G—29 June 2015, pp. 1—6.

[2] H. Li, Z. Wu, J. Zhang, "*Pedestrian detection based on deep learning model*," in 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2016, pp. 796-800.

[3] H. Zhang, Y. Du, S Ning, Y. Zhang, S. Yang, C. Du, "*Pedestrian Detection Method Based on Faster R-CNN*," in Proceedings of the 13th International Conference on Computational Intelligence and Security (CIS), Hong Kong, China, 15–18 December 2017, IEEE, pp. 427-430.

[4] W. Lan, J. Dang, Y. Wang, S. Wang, "*Pedestrian Detection Based on YOLO Network Model*," in Proceedings of the International Conference on Mechatronics and Automation (ICMA), 2018, IEEE, pp. 1547-1551.

[5] Yi, Zhang, S. Yongliang, J. Zhang, "*An improved tiny-yolov3 pedestrian detection algorithm*," Optik, Vol. 183, 2019. 183: pp. 17-23.

[6] C. Li, et al., "*Illumination-aware faster R-CNN for robust multispectral pedestrian detection. Pattern Recognition*," 2019. 85: pp. 161-171.

[7] H. Jeon, V. D. Nguyen, J. W. Jeon, "*Pedestrian Detection Based on Deep Learning*," IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 144-151,

[8] C. Burke, P.R. McWhirter, J. Veitcli-Michaelis, O. McAree, H. A. Pointon, S. Wield, "Longmore, S. "*Requirements and Limitations of Thermal Drones for Effective Search and Rescue in Marine and Coastal Areas*," Drones 2019, J, 78. [CrossRef]

[9] D. Bozié-S›tulié, Z. Marusic, S. Gotovac, "*Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions*," fnf. J. Comput. Vis. 2019, 127, 125G-127S.

[10] G. L. Hung, M. S. B. Sahimi, H. Samma et al. "*Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images*," SN Computer Science, 1, 116 (2020), pp. 1-9.

[11] A. S. Bin Sama, "*Pedestrian Detection from Drone Images Using Deep Learning Model*,", In Journal of Xidian University (JXU), Vol.15, Issue.6, 2021, pp. 298- 304.

[12] W. Liu, Z. Wang, X Liu, N. Zeng, Y. Liu, FE. Alsaadi, "*A Survey of Deep Neural Network Architectures and Their Applications*," Neurocomputing. 2017;234: pp.11–26.

[13] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, MP. Reyes, ML. Shyu, SC. Chen, S. Iyengar, "*A Survey on Deep Learning: Algorithms, Techniques, and Applications*," ACM Comput Surv (CSUR). 2018;51(5):PP.1–36.

[14] MZ Alom, TM Taha, C Yakopcic, S. Westberg, P. Sidike, MS Nasrin, M Hasan, BC Van Essen, AA Awwal, VK Asari, "*A State-of-the-Art Survey on Deep Learning Theory and Architectures*," Electronics. 2019;8(3):292.

[15] L. Alzubaidi, J. Zhang, AJ. Humaidi, et al. "*Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions*," J Big Data 8, 53 (2021).

[16] G. G.Wang, S. & d. S. Deb, L. Coelho, "*Elephant Herding Optimization*," in Computational and Business Intelligence (ISCBI), 2015 3rd International Symposium on, 1–5 (IEEE, 2015).

[17] Gai-Ge Wang, *"Elephant Herding Optimization (EHO),"* (https://www.mathworks.com/matlabcentral/fileexchange/53486-elephant-herding-optimization-eho), MATLAB Central File Exchange. Retrieved January 19, (2022).

AUTHORS

First Author – Saeed Mohammed Baneamoon
Department of Computer Engineering,
College of Engineering & Petroleum,
Hadhramout University,
Mukalla, Hadhramout, Yemen .
Email: sbaneamoon@hu.edu.ye
/OR/
saeed_banaimoon @yahoo.com

Correspondence Author – Saeed Mohammed Baneamoon
Department of Computer Engineering,
College of Engineering & Petroleum,
Hadhramout University,
Mukalla, Hadhramout, Yemen .
Email: sbaneamoon@hu.edu.ye
/OR/
saeed_banaimoon @yahoo.com