Computers, Materials & Continua
DOI: 10.32604/cmc.2022.028570
Article

Tech Science Press

Deep Learning Enabled Object Detection and Tracking Model for Big Data Environment

K. Vijaya Kumar[1], E. Laxmi Lydia[2], Ashit Kumar Dutta[3], Velmurugan Subbiah Parvathy[4], Gobi Ramasamy[5], Irina V. Pustokhina[6,*] and Denis A. Pustokhin[7]

[1]Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, 530049, India
[2]Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Visakhapatnam, 530049, India
[3]Department of Computer Science and Information System, College of Applied Sciences, AlMaarefa University, Riyadh, 11597, Kingdom of Saudi Arabia
[4]Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, 626126, Tamilnadu, India
[5]Department of Computer Science, Christ University, Bangalore, 560029, India
[6]Department of Entrepreneurship and Logistics, Plekhanov Russian University of Economics, 117997, Moscow, Russia
[7]Department of Logistics, State University of Management, Moscow, 109542, Russia
*Corresponding Author: Irina V. Pustokhina. Email: ivpustokhina@yandex.ru
Received: 12 February 2022; Accepted: 23 March 2022

Abstract: Recently, big data becomes evitable due to massive increase in the generation of data in real time application. Presently, object detection and tracking applications becomes popular among research communities and finds useful in different applications namely vehicle navigation, augmented reality, surveillance, etc. This paper introduces an effective deep learning based object tracker using Automated Image Annotation with Inception v2 based Faster RCNN (AIA-IFRCNN) model in big data environment. The AIA-IFRCNN model annotates the images by Discriminative Correlation Filter (DCF) with Channel and Spatial Reliability tracker (CSR), named DCF-CSRT model. The AIA-IFRCNN technique employs Faster RCNN for object detection and tracking, which comprises region proposal network (RPN) and Fast R-CNN. In addition, inception v2 model is applied as a shared convolution neural network (CNN) to generate the feature map. Lastly, softmax layer is applied to perform classification task. The effectiveness of the AIA-IFRCNN method undergoes experimentation against a benchmark dataset and the results are assessed under diverse aspects with maximum detection accuracy of 97.77%.

Keywords: Object detection; tracking; convolutional neural network; inception v2; image annotation

1 Introduction

Big data is a term used for massive datasets with large, distinct, and complex structure with the problems of storage, analyzing and visualizing for further processes. Big data analytics is a way of handling huge quantity of data in revealing hidden patterns and correlations. The six V's of big data is shown in Fig. 1. Visual Object Tracking (VOT) is the process of identify a random destination, represented by a Region of Interest (ROI), in a video [1]. In spite of the latest developments in the field due to Convolution Neural Network (CNN), tracking is still a crucial process in the domain of computer vision. Generally, tracking techniques produces a process of the destination ROI and utilize the generated process for searching the targets in the successive frames. The issues arise initially from the discriminative capability of the targets method that must be unique so that the targets would eliminate drifting to identical objects, whereas rest of the adaptive sufficient to the inter-frame look modifications of the targets. Besides, a tracking algorithm should be equipped to dealing with occlusion, quick targets motions and out of view scenes. Other significant feature is the tracker speed, i.e., the time period taken to locate the targets in every individual frame. The real world needs impose hard limitations on the per frame processing time of the tracker.

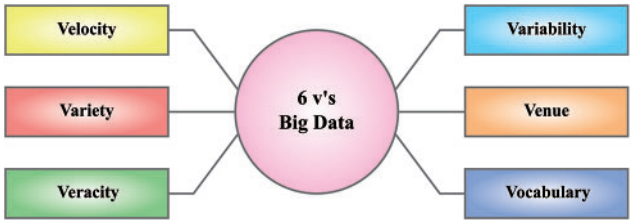

Figure 1: Six 6V's of big data

In recent times, CNNs are more effective in alternate Computer Vision tasks, like image classification, object prediction as well as semantic segmentation [2]. The effectiveness can be attributed to the semantically useful representation which is extracted from visible details. Therefore, CNNs is applied for monitoring operations. Generally, the CNN-based trackers are operated by filtering convolutional feature-based target and from the consecutive frames and cross-correlating the target method by the frame approach for locating the target. In order to report the challenging issues required by smart functions, like embedded devices and robots [3], a tracking model should force the performance tradeoff to the greater extent.

For CNN-relied trackers, it represents the present neural structures at the time of keeping the count of layers comparatively minimum. The maximum number of annotated video datasets is used like ImageNet VID or TrackingNet [4] services for training purpose. In prior to apply the traditional features, histogram-aided descriptors are applied [5,6]. Based on the encoded features, histograms are highly effective for making modifies whereas it is robust computationally. According to the proficiency of histogram based trackers, and effective representations are obtained by deep CNNs (DCNN), and 2 models are combined as a single neural structure, in which histograms are obtained from deep convolutional features, for applying optimal variables [7,8].

Massive current trackers have been employed the effective implications obtained by CNNs with maximum accuracy. A features obtained from areas are integrated with the application of fully connected (FC) layers and associated. The network undergoes training in offline, in case of regression, where it predicts the place and target size in explore area. Massive quantity of information and wide augmentation is essential for training the tracker in effectual manner. Moreover, SiamFC applies

convolutional features, instead of applying FC layers for comparing the target and explore area modules, it cross-correlates the features which is parallel to previous models have extracted features. The network is completely convolutional and trained to classified, where to differentiate among the features equivalent to target and background. A target's place in search region is selected as location with higher cross-correlation value.

This paper develops an effective deep learning (DL) based object detection and tracker model utilizing Automated Image Annotation with Inception v2 based Faster region convolution neural network (RCNN), called AIA-IFRCNN model in big data environment. The AIA-IFRCNN model comprises a novel image annotation tool utilizing Discriminative Correlation Filter (DCF) with Channel and Spatial Reliability tracker (CSR) named DCF-CSRT technique. The AIA-IFRCNN model makes use of Faster RCNN as an object detector and tracker, which includes region proposal network (RPN) and Fast R-CNN. In addition, inception v2 approach is utilized as the feature extractor to generate the feature map. At last, softmax layer is utilized to perform the task of classifying images. Extensive set of experimentation is carried out for verifying the proficient tracking performance of the AIA-IFRCNN model and the results are investigated under different dimensions.

2 Related Works

In recent times, anchor-based ROI selection has been embedded into a siamese structure for tracking, which is developed by SiamRPN [9]. The key objective of applying anchor to map bounding boxes of targets where the tracker is suitable to manage the aspect ratio modifications, while classical trackers deal with size alterations at the time of retaining a static aspect ratio. Moreover, direct bounding box regression model, where data augmentation is not essential. It is because of an anchor which is estimated for feasible position and fine-tuned to fit the ground truth bounding box. The CNN based trackers perform the tracking operation by learning a discriminative method of a target and adapt with the method in all frames. The online application of a target model renders the trackers slowly in smart graphical processing units (GPUs). Convolutional Residual Learning for Visual Tracking (CREST), is reformed Discriminative correlation filters (DCFs) as a convolutional layer, by enabling neural network (NN) and DCFs which has to be trained as a single structure. It is evolved by a sequence of alternate trackers that has integrated CNNs with DCFs. In Murugan et al. [10], an efficient region based scalable convolution neural network (RS-CNN) model has been projected for detecting anomalies in pedestrian walkways. It efficiently recognizes the anomalies at an earlier rate and carries out well with the scalability problem. Any other methods like mixture of dynamic texture (MDT) [11], mixture of optical flow (MPPCA) [12], circulant structure kernel (CSK) [13], Fast Compressive Tracking (FCT) [14], discriminative scale space tracker (DSST) [15], Convolutional Features (CF2) [16], and kernelized correlation filter (KCF) tracker [17] made to anomaly detection has been proposed in the literature. In Mehran et al. [18], an efficient social force (SF) method has been established to detection and localization of abnormal performance in crowded videos. A frame in the videos is classified into normal and abnormal ones utilizing a bag of words process.

3 The Proposed Tracker

Initially, DCF-CFRT model is used to annotate the objects exists in the image. Next, Faster RCNN is applied as an object detector, which also includes the inception v2 model as the shared CNN. Finally, softmax layer based classification process is carried out.

3.1 DCS-CSRT Model

At first, the input videos are changed into a sequence of frames and the objects in the frame are marked as objects in the creation of record files. The automated DCS-CSRT method is implemented as image annotation devices to annotate the objects occur in the input frame. It allows for annotating the objects in a single frame and generates automated annotation of the objects in each the frame exist in the video.

3.2 Faster RCNN

R-CNN is a kind of CNN that has been evolved from R-CNN [19]. When comparing to conventional selective explore technique, the Faster R-CNN is breaking the blockage issue of massive cost in calculation as the RPN creates equivalent proposal sites. So, the practical examination becomes feasible. Additionally, a Faster R-CNN is capable of adapting to arbitrary image and adjust the whole network for enhancing the accuracy of deep network recognition. A faster R-CNN model is capable of breaking the time blockage of computation, and ensures to achieve effective prediction rate. Thus, a Faster R-CNN analysis model is presented for processing feature extraction process which is carried out in insulator and the nest for identifying the destination. A Faster R-CNN technique is comprised of 2 CNN networks, as shown in Fig. 2 [19].

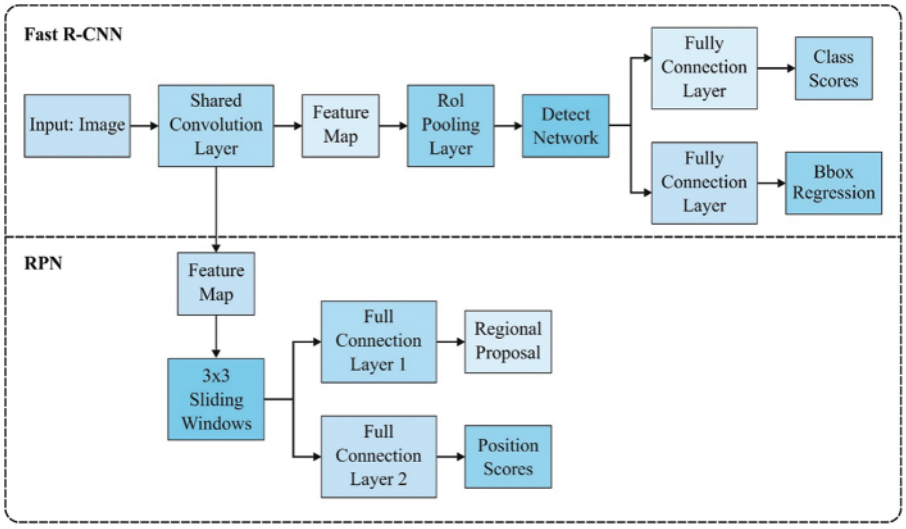

Figure 2: Process in Faster RCNN

A Fast R-CNN recognition network is present in the upper half of the flowchart and Regional Proposal Networks (RPN) are evolved in the lower half of the flowchart. A RPN samples the arbitrary area of an image as the proposal regions, and undergoes training to define with a destination. Furthermore, Fast R-CNN prediction system is again processed using the data collected by RPN system, which defines the destination type in the area.

3.3 Inception v2

Inception V2 was deployed by GoogLeNet. Inception V1 (or GoogLeNet) is the modern structural design at ILSRVRC 2014 [20]. It has developed the minimum error at ImageNet classifier; however, it is embedded with few effective points where the enhancement is performed to make better accuracy

and reduce the complications of the model. Previously, Inception V1 has applied the convolutions like $5 * 5$ which results in dimension reduction by enlarged margin. It tends to limit the NN accuracy. A basic structure of Inception V1 is shown in Fig. 3. A reason behind NN is malicious to data loss, while the input dimension is reduced in abundance. Besides, there is also difficulty reduction while applying higher convolutions such as 5×5 as related to 3×3.

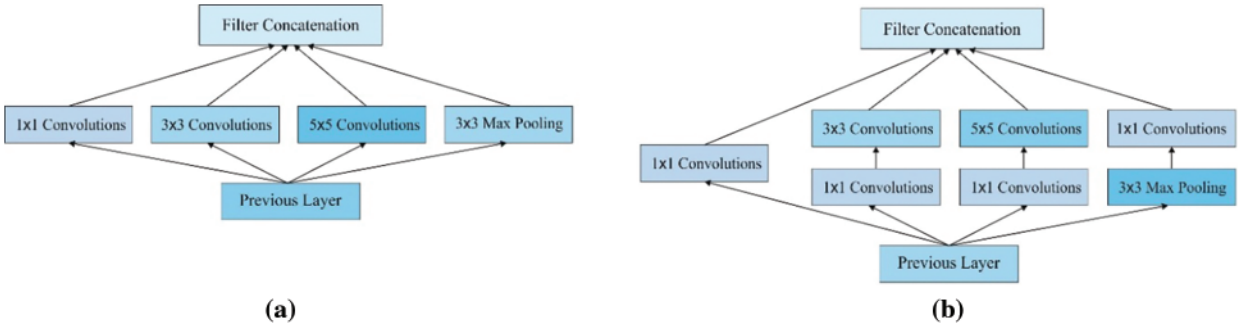

Figure 3: (a) Inception module naive version (b) Inception module with dimension reductions

The internal illustration of the testing data is in the method of normalized behind executing batch normalization (BN) into a network layer, after that a result is normalized for the normal distributing that attributes to reduce the internal covariate shift too. An internal layer in the deep CNN for adapting to the data sharing altered always that cause covariate shift. It considers that the input of a definite layer is normalization as following:

$$\hat{x} = \frac{x - E[x]}{\sqrt{Var[x] + \omega}} \tag{1}$$

where x and $\hat{x}$ indicate the input and normalization value of a definite layer correspondingly. $E[x]$ and $Var[x]$ are the expectation and difference of the input correspondingly. It is removed by implying BN, and the similar allocation of input is obtained in all layers behind normalized. For diminishing the impacts on all network layers behind normalization, parameters γ and β are contained in. An equation is illustrated as follows [21]:

$$y_i = \gamma \hat{x}_i + \beta \tag{2}$$

$$\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i} \cdot \gamma \tag{3}$$

$$\frac{\partial l}{\partial \delta^2 \theta} = \sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i} \cdot (x_i - \mu_\theta) \cdot \frac{-\left(\delta_\theta^2 + \omega\right)^{-3/2}}{2} \tag{4}$$

$$\frac{\partial l}{\partial \mu_\theta} = \left(\sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\delta_\theta^2 + \omega}} \right) + \frac{\partial l}{\partial \delta^2 \theta} \cdot \frac{-2 \sum_{i=1}^{m} (x_i - \mu_\theta)}{m} \tag{5}$$

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\delta_\theta^2 + \omega}} + \frac{\partial l}{\partial \delta^2 \theta} \cdot \frac{2(x_i - \mu_\theta)}{m} + \frac{\partial l}{\partial \mu_\theta} \cdot \frac{1}{m} \tag{6}$$

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial \gamma_i} \cdot \hat{x}_i \tag{7}$$

$$\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial \gamma_i} \cdot \gamma \tag{8}$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial \gamma_i} \tag{9}$$

where l is determined as the gradient loss of backpropagation. m is the size of mini-batch θ. x_i and y_i indicates the value of input x over the mini-batch and result behind BN model correspondingly. μ_θ and δ_θ^2 are the mean as well as variance of the mini-batch. The end result of BN network y is illustrated as follows:

$$y = \frac{\gamma x}{\sqrt{Var[x] + \omega}} + \beta - \frac{\gamma E[x]}{\sqrt{Var[x] + \omega}} \tag{10}$$

On other hand, Inception implementing BN diminishes a number of internal covariate shifts for normalizing the result in all layers. Conversely, it reduces the count of parameters and accelerating the calculating speed.

4 Performance Validation

Here, the wider series of experiments are performed on 3 dataset and the attained outcomes are examined with respect to prediction accuracy, annotation time, Center Location Error (CLE) and Overlap Rate (OR). The information of a dataset, measures and results analysis is defined in the upcoming sections. In order to process the comparison task, region scalable (RS)-CNN, Fast R-CNN, MDT, MPPCA and SF were applied.

4.1 Dataset Used

Tab. 1 offers the details of the 3 test dataset. The dataset 1 is defined as a multi-object tracking bird dataset (http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html), which is composed of 99 frames with time period of 3s. The second UCSDped2 (Test004) is said to be anomalous detection dataset (http://www.svcl.ucsd.edu/projects/anomaly/dataset.html), that is comprised of 180 frames and time limit of the video is 6s. The dataset 3 contains 2 sub files such as underwater blurred as well as underwater crowded [22] that has 2875 and 4600 frames under the time limit of 575s.

Table 1: Dataset details

Samples	Dataset name	Frames	Time (s)
Dataset 1	Bird	99	3
Dataset 2	UCSDped2 (Test004)	180	6
Dataset 3	Under Water (Blurred)	2875	575
	Under Water (Crowded)	4600	575

4.2 Results Analysis on Dataset 1

Fig. 4 visualizes the detection of multiple objectives by the AIA-IFRCNN model on the applied dataset 1. The input image with the respected tracked outcome is illustrated and it is depicted that

the chick, pelican and cloud toy objects are identified by the bounding box with respective tracking accuracy.

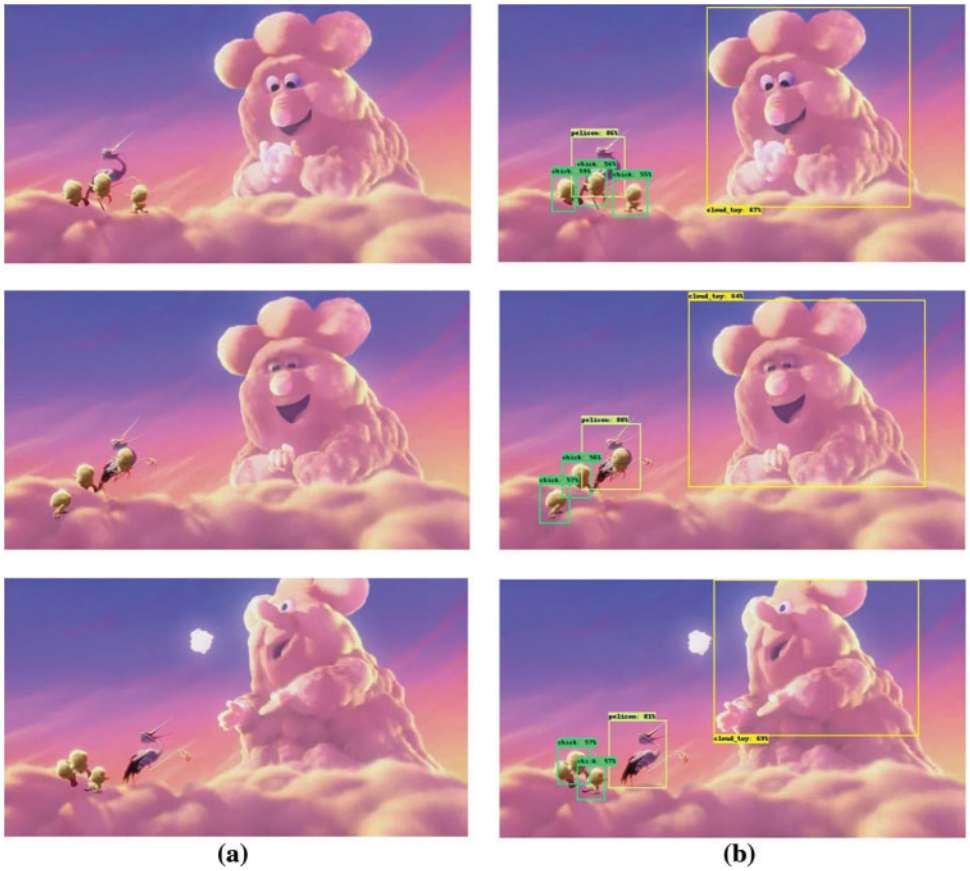

Figure 4: Visualizing objection detection of AIA-IFRCNN for Dataset 1

Fig. 5 illustrates the comparative analysis of the AIA-IFRCNN model on the applied dataset 1 interms of detection accuracy. The figure stated that the SF model has depicted as an ineffective tracker, which has reached to a least detection accuracy over the compared methods. Simultaneously, the MPPCA model has exhibited higher accuracy over SF model. On the other hand, the MDT model has tried to surpass the previous models. Besides, the Fast R-CNN has demonstrated manageable results with the moderate detection accuracy. Followed by, the RS-CNN and AIA-RFRCNN models have portrayed competitive results with the high detection accuracy. But, the AIA-IFRCNN model has achieved superior performance by attaining maximum detection accuracy.

Fig. 6 investigates the average detection accuracy of the presented AIA-IFRCNN model with compared methods on the applied dataset 1. The figure depicted that the SF model has offered a detection accuracy of 67.01%, which is lower than the performance attained by other methods. Followed by, the MPPCA and MDT models have achieved slightly higher detection accuracy of 73.12% and 77.70% respectively. Besides, the Fast R-CNN model has the ability to attain moderate detection accuracy of 86.91% whereas even better results are offered by the RS-CNN method with the detection accuracy of 93%. Though the AIA-RFRCNN model has achieved a considerable detection accuracy

of 94.67%, the presented AIA-IFRCNN model has resulted to effective performance with the higher detection accuracy of 95.62%.

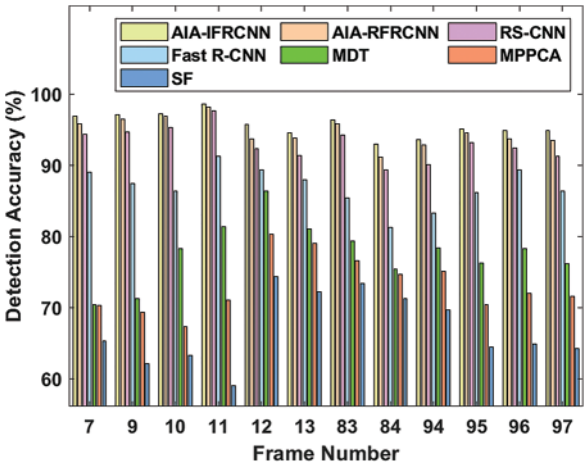

Figure 5: Detection accuracy analysis of AIA-IFRCNN model on dataset 1

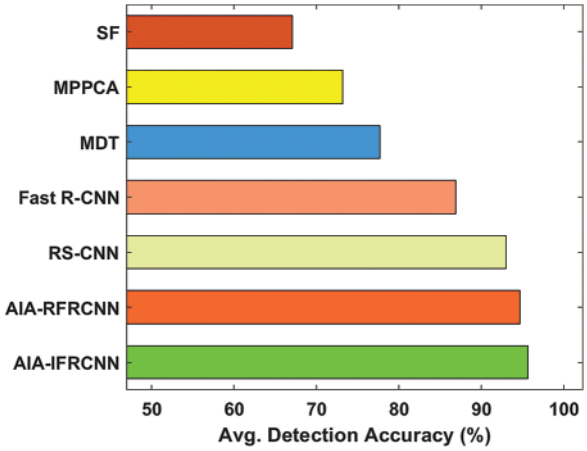

Figure 6: Average detection accuracy analysis of AIA-IFRCNN model on dataset 1

4.3 Results Analysis on Dataset 2

Fig. 7 imagines the prediction of diverse objectives by the AIA-IFRCNN method on the given dataset 2. The input image with corresponding tracked results are depicted that the person, car, truck, and skater objects are found by bounding box with parallel tracking accuracy.

Fig. 8 depicts the relative analysis of the AIA-IFRCNN approach on the applied dataset 2 with respect to detection accuracy. The figure implied that the SF approach has showcased an worst tracker, that has attained lower prediction accuracy than the previous technologies. At the same time, the MPPCA scheme has represented maximum accuracy when compared to SF model. Followed by, the MDT scheme has attempted to perform well than the existing methodologies. Then, the Fast R-CNN and RS-CNN technologies have depicted considerable outcomes with the better detection accuracy. Next, the IRS-CNN and AIA-RFRCNN approaches have depicted competing results with higher

detection accuracy. However, the AIA-IFRCNN technique has accomplished supreme function by reaching optimal detection accuracy.

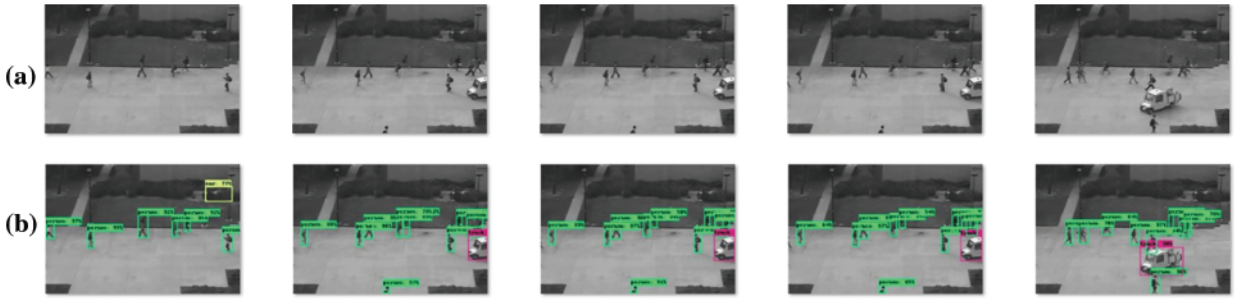

Figure 7: Visualizing objection detection of AIA-IFRCNN for Dataset 2

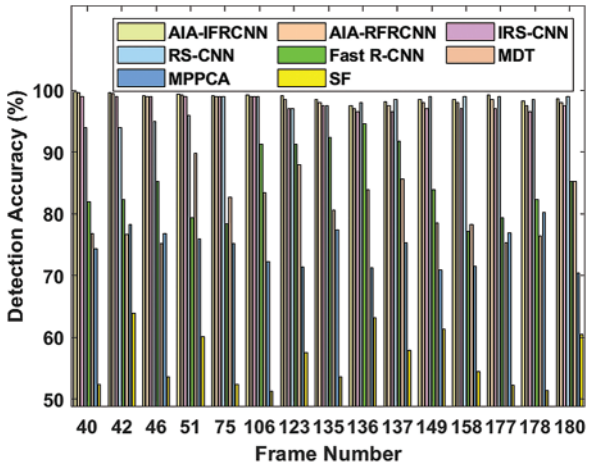

Figure 8: Detection accuracy analysis of AIA-IFRCNN model on dataset 2

Fig. 9 examines the average detection accuracy of the projected AIA-IFRCNN scheme with previous models on the applied dataset 2. The figure portrayed that the SF approach has provided a detection accuracy of 56.38%, that is minimum than the function performed by alternate models. Besides, the MPPCA and MDT methodologies have achieved accomplished moderate detection accuracy of 74.56% and 81.11% correspondingly. Followed by, the Fast R-CNN scheme is capable of reaching considerable detection accuracy of 85.10% while acceptable results are generated by the RS-CNN method with the detection accuracy of 97.50%. Although the IRS-CNN and AIA-RFRCNN approaches have attained a reasonable detection accuracy of 97.77% and 98.43%, the projected AIA-IFRCNN scheme has resulted to effectual performance with the maximum detection accuracy of 98.85%.

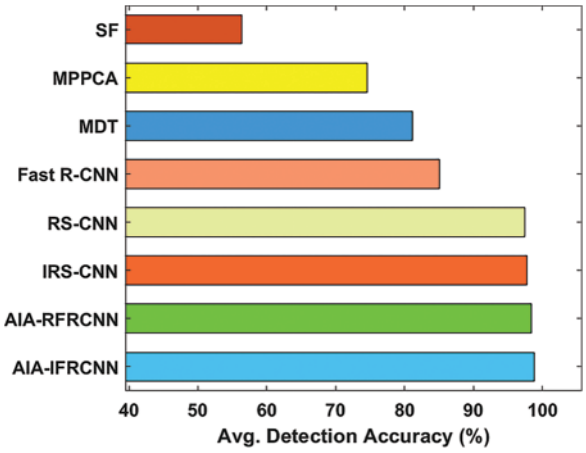

Figure 9: Average detection accuracy analysis of AIA-IFRCNN model on dataset 2

4.4 Results Analysis on Dataset 3

Fig. 10 showcases the detection of various objectives by the AIA-IFRCNN method on the applied dataset 3. The input image with the given tracked result is demonstrated and it is illustrated that the fish objects are discovered by the bounding box with corresponding tracking accuracy.

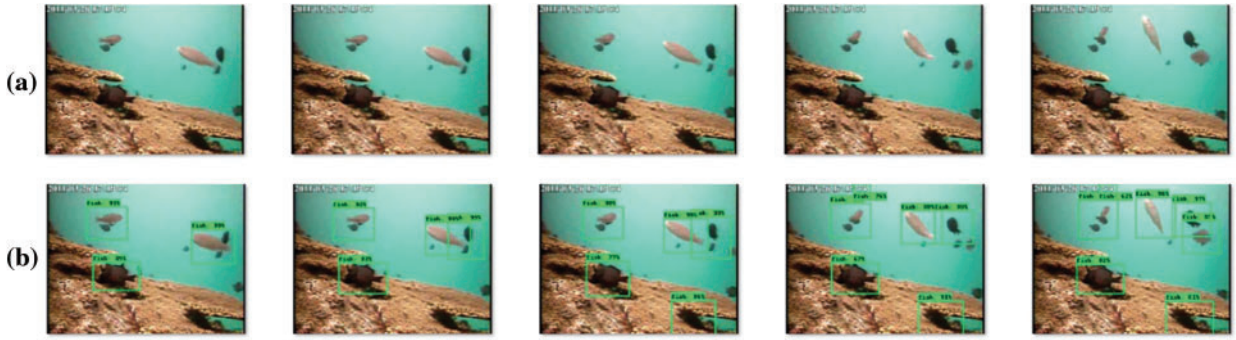

Figure 10: Visualizing objection detection of AIA-IFRCNN for Dataset 3 (Crowded)

Fig. 11 depicts the competing analysis of the AIA-IFRCNN method on the applied dataset 3 in light of detection accuracy. The figure depicted that the SF approach has showcased an inferior tracker that has accomplished minimum detection accuracy than the former models. Concurrently, the MPPCA scheme has shown maximum accuracy than the SF model. Besides, the MDT approach has managed to outperform the traditional methods. On the other hand, the Fast R-CNN has illustrated considerable results with the considerable detection accuracy. Then, the RS-CNN and AIA-RFRCNN methodologies have implied competing results with the higher detection accuracy. However, the AIA-IFRCNN technology has accomplished supreme function by achieving best detection accuracy.

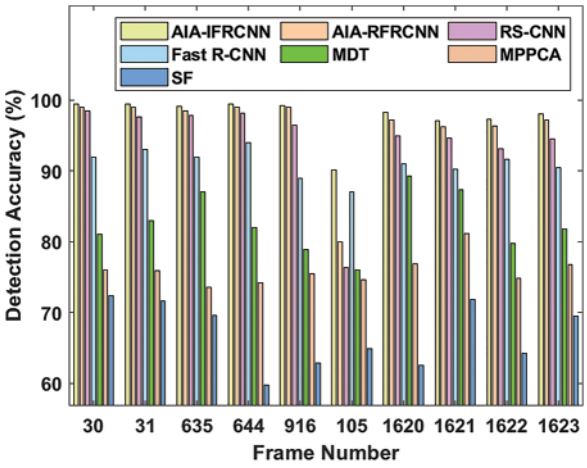

Figure 11: Detection accuracy analysis of AIA-IFRCNN model on dataset 3

Fig. 12 examines the average detection accuracy of the presented AIA-IFRCNN model with previous models on the applied dataset 3. The figure demonstrated that the SF technology has provided a detection accuracy of 66.94%, that is minimal than the performance achieved by alternate schemes. Then, the MPPCA and MDT methodologies have reached acceptable detection accuracy of 75.95% and 82.62% correspondingly. Followed by, the Fast R-CNN technology has the potential to obtain better detection accuracy of 91.05% while even better results can be generated by the RS-CNN scheme with the detection accuracy of 94.23%. Even though the AIA-RFRCNN approach has attained a reasonable detection accuracy of 96.15%, the projected AIA-IFRCNN scheme has provided efficient function with the maximum detection accuracy of 97.77%.

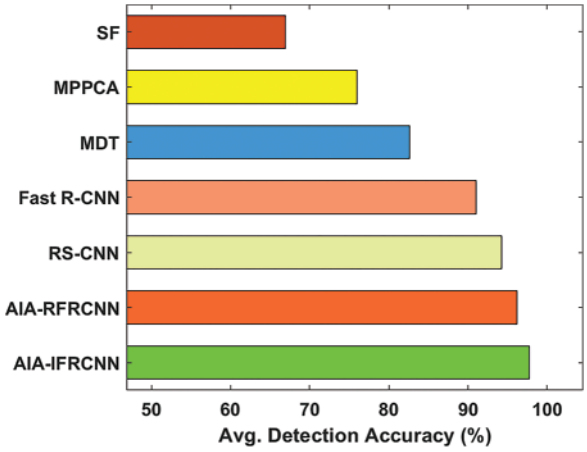

Figure 12: Average detection accuracy analysis of AIA-IFRCNN model on dataset 3

4.5 Analysis of Average CLE

Fig. 13 investigates the performance of the AIA-IFRCNN model on the applied dataset interms of average CLE. On analyzing the average CLE results in the dataset 1, the presented AIA-IFRCNN model shows its effectiveness by attaining a minimum average CLE of 4.16. At the same time, the

other methods such as AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF models have resulted to a higher average CLE of 5.67, 7.49, 17.38, 90.30, 56.72, 38.50 and 45.57. On determining the average CLE results in the dataset 2, the newly developed AIA-IFRCNN method showcases the effectiveness efficiency by obtaining least average CLE of 5.78. Meanwhile, the alternate models like AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF approaches have achieved maximum values of CLE of 6.89, 9.21, 19.48, 15.86, 58.31, 40.58 and 47.20. On investigating the average CLE results in the dataset 3, the proposed AIA-IFRCNN scheme implies the supremacy by acquiring lower average CLE of 3.54. Simultaneously, the other schemes like AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF models have generated maximum average CLE of 4.32, 12.88, 29.40, 20.79, 63.84, 48.76 and 50.42.

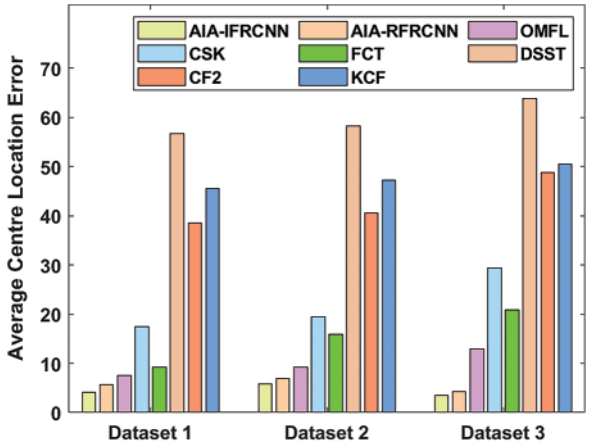

Figure 13: Average CLE analysis of AIA-IFRCNN model

4.6 Analysis of Overlap Rate

An analysis of overlap rate by the AIA-IFRCNN model has been made with the existing methods on the applied dataset, and the outcomes are depicted in Fig. 14. The figure showcased that the AIA-IFRCNN model has achieved superior results over the existing methods by attaining maximum overlap rate. On determining the overlap rate in dataset 1, the projected AIA-IFRCNN model has achieved a higher overlap rate of 0.92 whereas the AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF models have displayed lower overlap rates of 0.89, 0.78, 0.71, 0.74, 0.52, 0.68 and 0.63 respectively. On analyzing the overlap rate in dataset 2, the newly developed AIA-IFRCNN method has reached a maximum overlap rate of 0.90 while the AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF methodologies have showcased least overlap rates of 0.86, 0.77, 0.68, 0.73, 0.49, 0.65 and 0.59 correspondingly. On investigating the overlap rate in dataset 3, the presented AIA-IFRCNN scheme has reached a greater overlap rate of 0.94 and the AIA-RFRCNN, OMFL, CSK, FCT, DSST, CF2 and KCF technologies have exhibited minimum overlap rates of 0.91, 0.72, 0.64, 0.69, 0.46, 0.62 and 0.51 correspondingly.

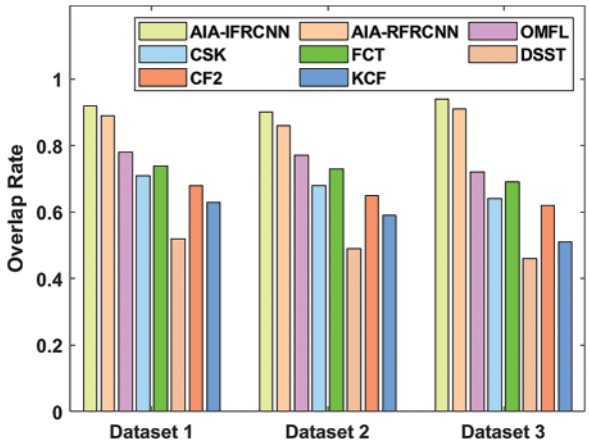

Figure 14: Overlap rate analysis of AIA-IFRCNN model

5 Conclusion

This paper has developed a novel DL based object detection and tracker model utilizing AIA-IFRCNN model. Initially, DCF-CFRT model is used to annotate the objects exists in the image. Next, Faster RCNN is used to recognize objects, which also includes the inception v2 model as the shared CNN. Finally, softmax layer classifier is exploited. The effectiveness of the AIA-IFRCNN method undergoes experimentation against a benchmark dataset and the results are assessed in diverse aspects. The experimental outcome confirmed the effective tracking performance of the projected method with the in future; the presented model can be deployed in concurrent surveillance cameras for detecting and tracking abnormalities. The experimental outcome indicated that the AIA-IFRCNN model has surpassed the compared techniques with increased detection accuracy of 95.62%, 98.85% and 97.77% on datasets I, II and III respectively. In future, hybrid DL models can be included to improve detection performance.

Funding Statement: The authors received no specific funding for this study.

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

References

[1] B. Deori and D. M. Thounaojam, "A survey on moving object tracking in video," *International Journal on Information Theory*, vol. 3, no. 3, pp. 31–46, 2014.

[2] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, no. 3, pp. 85–117, 2015.

[3] S. Kamate and N. Yilmazer, "Application of object detection and tracking techniques for unmanned aerial vehicles," *Procedia Computer Science*, vol. 61, no. 25-25, pp. 436–441, 2015.

[4] W. Fan, X. Xu, X. Xing, W. Chen and D. Huang, "LSSED: A large-scale dataset and benchmark for speech emotion recognition," in *ICASSP, 2021–2021 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, pp. 641–645, 2021.

[5] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[6] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, "TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021. (Article in press). https://doi.org/10.1109/TITS.2021.3130403.
[7] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 92, no. 6, pp. 1–16, 2021.
[8] H. Huang, G. Liu, Y. Liu and Y. Zhang, "GRSiamFC: Group residual convolutional siamese networks for object tracking," in *2021 Int. Conf. on Control, Automation and Information Sciences (ICCAIS)*, Xi'an, China, pp. 614–619, 2021.
[9] B. Li, J. Yan, W. Wu, Z. Zhu and X. Hu, "High performance visual tracking with siamese region proposal network," in *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, pp. 8971–8980, 2018.
[10] B. S. Murugan, M. Elhoseny, K. Shankar and J. Uthayakumar, "Region-based scalable smart system for anomaly detection in pedestrian walkways," *Computers & Electrical Engineering*, vol. 75, no. 3, pp. 146–160, 2019.
[11] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 909–926, 2008.
[12] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 2921–2928, 2009.
[13] A. R. Pathak, M. Pandey and S. Rautaray, "Application of deep learning for object detection," *Procedia Computer Science*, vol. 132, no. 6, pp. 1706–1717, 2018.
[14] K. Zhang, L. Zhang and M. H. Yang, "Fast compressive tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 2002–2015, 2014.
[15] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. of the British Machine Vision Conf. 2014*, Nottingham, pp. 65.1–65.11, 2014.
[16] C. Ma, J. B. Huang, X. Yang and M. H. Yang, "Robust visual tracking via hierarchical convolutional features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2709–2723, 2019.
[17] C. C. Ukwuoma and C. Bo, "Deep learning review on drivers drowsiness detection," in *2019 4th Technology Innovation Management and Engineering Science Int. Conf. (TIMES-iCON)*, Bangkok, Thailand, pp. 1–5, 2019.
[18] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 935–942, 2009.
[19] X. Lei and Z. Sui, "Intelligent fault detection of high voltage line based on the Faster R-CNN," *Measurement*, vol. 138, no. 1, pp. 379–385, 2019.
[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 2818–2826, 2016.
[21] X. Zhu, M. Zhu and H. Ren, "Method of plant leaf recognition based on improved deep convolutional neural network," *Cognitive Systems Research*, vol. 52, no. 1, pp. 223–233, 2018.
[22] N. Krishnaraj, M. Elhoseny, M. Thenmozhi, M. M. Selim and K. Shankar, "Deep learning model for real-time image compression in Internet of Underwater Things (IoUT)," *Journal of Real-Time Image Processing*, vol. 17, no. 6, pp. 2097–2111, 2020.