Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 1, pp. 113–126
DOI : 10.47974/JIOS-1301

An effective framework for detecting the object from the video sequences by utilizing deep learning with hybrid technology

Ravi Prakash Chaturvedi *
Udayan Ghose †
University School of Information, Communication and Technology
Guru Gobind Singh, Indraprastha University
Delhi
India

Abstract

In recent technologies, object detection is considered as an effective tool for diagnosing the anonymous activities of a particular location. We can recognize the specific object from images and videos. Therefore, we can obtain essential information for developing a highly secure framework. The ODEF technique is developed for enhanced object detection and classification process. By utilizing the proposed technique, the anonymous activities in a specific region can be detected through the video. This technique detected the objects with bounding boxes; therefore, malicious activities can be shown in distinct visual. Then, the utilization of DRCNN technique provides better platform to classify the object.

Subject Classification: *68Q32, 68T07.*

Keywords: *DRCNN, Fractional calculus, GMM, Classification.*

1. Introduction

The modern image processing technique provides tremendous innovations in various domains such as medical, military, government and other industries. In image processing, the object detection is an active research topic for discovering abnormal activities of specific location [1]. The object detection process is carried out by adopting the real object samples which are collected from images or videos. The recent technology focuses on enhancement in object detection technique by incorporating automatic detection facilities for effective prediction mechanism.

* *E-mail:* **rpchaturvedi51@gmail.com (Corresponding Author)**
† *E-mail:* **udayan@ipu.ac.in**

Therefore, intelligent techniques are adopted for classifying the specific object in an automated manner with the adaptation of remote sensing technology [9]. This technique nowadays gets drastic attention in civil and military domains [2]. The automation technology provides tremendous facilities in urban location and it is gradually implemented into the rural area for its development. Moreover, this technology is adopted for road safety, ensuring the security of the environment, urban developmental process etc. However, the remote sensing may get interrupted by the disturbing components like geometric change, illumination etc. These disturbance or interpretations are becoming a serious topic for object detection enhancement [10]. It is difficult to utilize the object detection technique into the area of remote complex sense image [5]. In addition, implementing the technique in real time application is difficult because input images are large in size in complex background; therefore, it leads to false positives. For instance, a randomly textured desert area [11] [13]. Moreover, the possibility of time consumption and memory utilization is high which causes decline in execution process and becomes hard for processing in a current hardware. Moreover, detecting the small objects from low resolution images based on remote sensing images is very challenging task [4] [12][15][6]. The ability to classify the small objects is vital process based on video surveillance [7] [8].

1.1 *Literature Review*

In this section, we made a literature review work based on analysing various research articles. Here, some of the papers are discussed below.

Qibin Hou et. al. [12] gives a novel concept based HED architecture. The execution process of this technique is based on FCN based functions. Each function in this framework is significant for detecting segment. They also made a comprehensive analysis for enhancing this work for future development.

The other researchers Yuxing Tang et. al. [13] have generated a new concept based on executing visually and semantically similar categories, they focus on highlighting abnormal features than the different categories. For instance, their system is adopted to detect the object by analysing the difference between the dog and cat class. They made practical analysis based of similitude-based knowledge transfer strategy. They can prove that their strategy is very better than other technique.

The researchers Yunhang Shen et. al. [14] have provided the scheme to detect object by utilizing localization based Object Specific Pixel

Gradient (OPG) concept for localization of poorly organized objects. In OPG organize the particular object in a specific image based on image level notes. The essential process of this work is to determine the specific object image pixel's contribution. Then they are utilizing a data mining technique to fetch feature of the object. The localization concept is adopted based on medium and maximum tie layer. Moreover, CNN technique is adopted without suggestions of objects. Their method provide effective outcome with the proposed method.

Nuri Kim. et. al. [15] give concept of instance aware detection network that you can use to learn how to extract right features of nominee domain and find the similarities . Thorough experiments show that the suggested algorithm significantly improves the overlap of object recognition over existing prior art identification methods for Pascal VOC, MS COCO and Crowd Human datasets.

2. Proposed Methodology

The efficiency of the existing technology should be improved.

In our proposed work, we executed the technique for managing all these difficulties by proposing a novel and effective technique. The architecture of the proposed technique is given below (Figure 1):

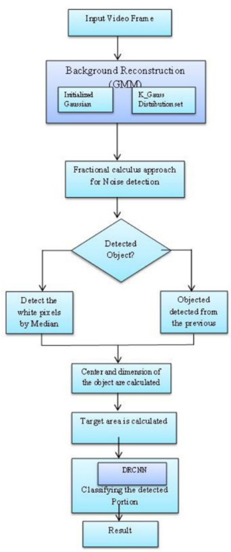

Figure 1

Architecture of Proposed Work

The object detection technique is utilizing based on images or video sequences. From the video, separating the specific object is very difficult and they require an effective technique to fetch essential components. In this work, initially, we utilized the input videos by converting the video sequence into several video frames and they are taken as input. In the pre-processing step, two kinds of technologies are adopted such as GMM and Fractional Calculus. Even if the video frames are obtained, they may not provide clear visibility for further processing [3]. Therefore, we utilized the GMM technique to reconstruct the background for the specific frame by analyzing the pixel values and fractional calculus helps to reduce the noises and make it clear for further processing. Then we execute the condition statement for checking and we utilized the median filter for removing the white pixels from the detected frame. After the completion of the above processes, the exact center and the dimension of the detected object are calculated. Finally, the target area is estimated. They utilized DRCNN for better classification and to also improve the performance of the object detection techniques.

Pre-processing

In this section the procedure of pre-processing is discussed. Pre-processing is executed for reducing the dimensionality, noise and making the dataset to be ready for next level procedure. For pre-processing, we utilized GMM and fractional calculus. It will be discussed in brief below.

GMM based background reconstruction

In Object detection based framework, the contribution of background modelling is inevitable for effectively separating the background object for predicting the specific object. For this background based process, this technique initially executes the selection process for selecting the image blocks. This process is based on analysing the image blocks. This work is based on incorporating the image boxes for background restoration process. For this process Gaussian distribution based technique is adopted, which is Gaussian mixture model. In order to determine the pixels K-gaussian distribution to determine the pixel of the particular image pixels are matched at the initial distribution process. The equation (1) and (2) executes the Gaussian GMM distribution process. They are shown below:

$$\left|D_{ij}^{m,k}\right| \leq \mu \times M_{ij}^{m-1,k} \quad (1)$$

$$D_{ij}^{m,k} = G_{ij}^{m,k} - M_{ij}^{m-1,k} \quad (2)$$

Here, $D_{ij}^{m,k}$ denotes the distance of pixel that is newly generated and Kth Gaussian distribution, μ represents the threshold of deviation, $M_{ij}^{m-1,k}$ and shows about processing image pixel frame based standard deviation. $G_{ij}^{m-1,k}$ – meant about the processing frame based standard deviation, $G_{ij}^{m,k}$ meant about executing frame of image. The distribution of background is carried out by considering the pixel values which should be matched K^{th} Gaussian distribution model while comparing with the new pixel value. Based on these results, update the value of model parameters. For updating process the equation (3) and (5) are executed.

$$M_{ij}^{m,k} = (1-p)M_{ij}^{m-1,k} + p \times G_{ij}^{m,k} \quad (3)$$

$$w_{ij}^{m,k} = (1-\alpha)w_{ij}^{m-1,k} + \alpha \quad (4)$$

$$\beta_{ij}^{m,k} = \sqrt{(1-p)(\beta_{j}^{m-1,k})^2 + \alpha} \quad (5)$$

$$P = \frac{\alpha}{w_{ij}^{m,k}} \quad (6)$$

where α symbol is utilized as learning, rate determined by the background update rate, $w_{ij}^{m,k}$, is based on executing images pixel weight and p denotes the rate of refresh of the image, which is shown in the formula (6). In case of the pixel value is not suitable with any of the Gaussian distribution, the current minimum weight is generated by a new Gaussian distribution which overcome the existing Gaussian distribution. Then create the average from the new Gaussian distribution value which is taken for pixel. For weight updating process the formula (7) is processed,

$$w_{ij}^{m,k} = (1-\alpha)w_{ij}^{m-1,k} \quad (7)$$

Once the model is generated for each pixel, the Gaussian distribution based part is remained of background and the remaining part shows for foreground. For this process, initially K-Gaussian distribution is executed based on executing ranking process in the priority of descending order and select the Bac Gaussian distribution to generate the background model. The mathematical formulation of this process is shown below in equation (8)

$$Bac = \underset{b}{\arg\min}\left(\sum_{j=1}^{b} w_{jt}^{mk} > \tau\right) \tag{8}$$

where τ is the threshold value which is adopted to find the minimum quantity of background data. During this process, if the static distribution is selected, the background intensity values are denoted by the distribution average. Otherwise, based on previous scale the background intensity is highlighted. To get a better context, the property coefficients are normalized in the framework and their mathematical formulation is given in the formula (9).

$$W_{ij} = \frac{w_{ij}}{\sum_{m=1}^{k} w_{mj}}, i = 1, 2, 3 \ldots, k \tag{9}$$

where, $W_{ij}$of latest pixel.

Mathematical strategy based on fractional calculus for noise removal

In our work, we utilized a mathematical method which is executed based on fractional calculus for noise removal. It will be exampled in details below.

In image processing, it has been observed that a Fractional calculus has many advantages over an integer calculation. Fractional calculus is therefore considered as a global operator[5]. The fraction operator can rely on more information about neighbouring pixels and bring out more details about the texture of the frames.

This method offers better noise reduction performance than traditional noise reduction algorithms. Another advantage of fractions is that they can be used very well in directional noise applications. The reduction of noise in the frames can be performed in the desired direction. The noise image consists of signal points and noise points, the signal points are divided into a flat area of the multi-frame and the edge of the multi-frame. To better see the noise in the frames, the multiform is divided into three areas into flat areas, edges and noise points. It is very important to accurately identify the noise points in the frames.

$$E(a) = \min_{a} \int_{\Omega} \left| L^{\alpha} a \right| + \lambda \int_{\Omega} \frac{(f-a)^2}{a}, \tag{10}$$

where, $L^{\alpha} a = (L_x^{\alpha} a, L_y^{\alpha} a), \left| F_u^{\alpha} \right| = \sqrt{(L_x^{\alpha} a)^2 + (L_y^{\alpha} a^2)}$

and $1 \leq \alpha \leq 2$. Term regularization call J(u) .and then we call data fitting term.

$$\partial(J(a)) + \lambda\left(1 - \frac{f^2}{a^2}\right) = 0 \quad (11)$$

where, $\partial(J(a)) = \frac{d}{d\tau}\int_{\Omega}\left|L^{\alpha}a + \tau L^{\alpha}\eta\right|_{r=0}$

2.1 *White Pixel Removal with median filter*

In this section, the white pixel removal technique is discussed based on the utilization of median filter technique. It treats the distorted frames by initially diagnosing the white pixel noise. The processing pixel is considered for checking the pixels in the images are noisy or it is noise less.pixel remain unchanged if pixel lies between maximum and minimum value of grey level values. The Median filter technique is adopted for overcoming from the issues of noise by considering the grey level of both maximum or minimum. Initially the size of the window is selected as 3 X 3 then X_{ij} is denoted for pixel processing. Based on executing the condition such as 0<255 is to prove X_{ij} is uncorrupted pixel otherwise the pixels are corrupted. In the window, median value is given for changing the X_{ij} and replace the value by current pixel with final pixel, where O is less than 255.

$$f(x) = 2x / (2\sigma^2 + x^2) \quad (12)$$

Where σ point of rejection (outlier), is given by,

$$\sigma = \frac{\tau_s}{\sqrt{2}} \quad (13)$$

where τ_s higher value of outlier,

$$\tau_s = \varsigma\sigma_N \quad (14)$$

Where σ_N is standard deviation and D is a levelling factor.

$$s1 = \sum_{l \in L} \frac{pixel(l) \times f(x)}{x} \quad (15)$$

$$s2 = \sum_{l \in L} \frac{f(x)}{x} \quad (16)$$

s1 and s2 estimated values of pixel.

2.2 *Deep learning-based classification (DRCNN)*

In this article, we employed a modified DRCNN for image recognition. A large separable convolution is used to create a thin feature map which greatly reduces the output channels. The thin functional map is then used to positionally distort the RoI, increase accuracy and reduce computational time for training and reflection. The modified DRCNN's configuration parameters are matched to the image. Detection cells are divided into different categories. In general, the possibilities of obtaining the three categories of noises such as noise coordinates of bound box, false negative and false positive. In this work, DRCNN provides a framework to classifying object by its activity and forecast it as malicious or normal.

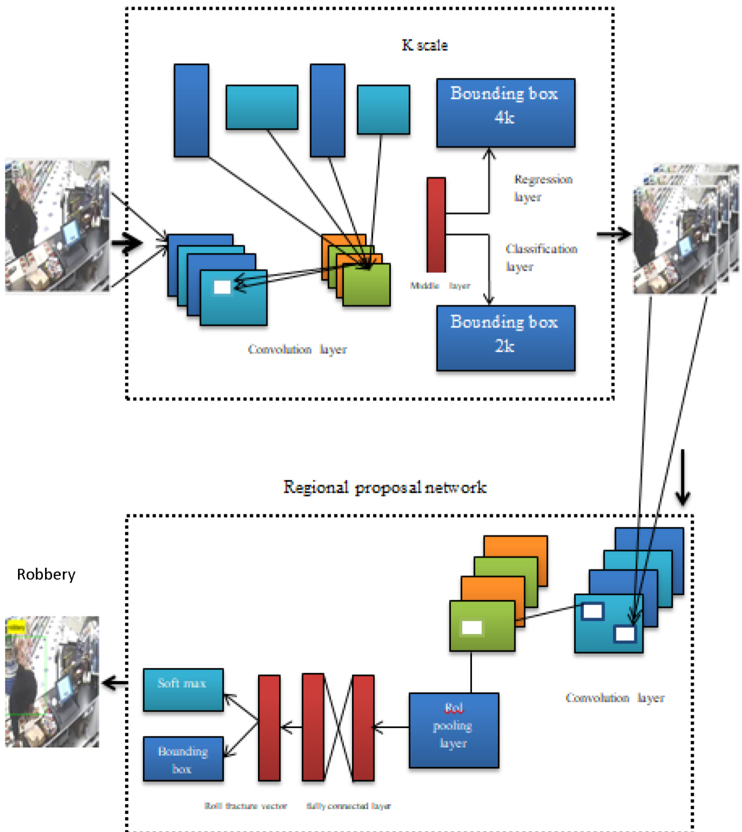

Figure 2

Shows the DRCNN based proposed work

DRCNN is designed to allow the generating noisy block notes in a mining learning environment. The RPN loss function is as follows:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{17}$$

i is anchor index and pi is anchor probability of predicted object. If ground truth box overlap maximum with anchor box then pi* is 1 otherwise 0.Lcis cross entropy

$$L_{cls} = -p_1^* \log(p_i)$$

The ultimate purpose of utilizing the DRCNN technique in our work is to classify the activity based on monitoring particular location such as robbery, pistol, fighting scenario and normal. Our technique provides the activity result based on analysing situation. In Figure 2, the executing process of proposed technique carried out for classifying the objects by taking input video frames, executing regional proposal network, finally it shows the classification results based on checking whether the activity is normal, abnormal, or what kind of activity is taken place.

3. Result and Discussion

UCF dataset includes the multiple anomaly videos like threat, shooting and robbery surveillance videos. In these dataset every video has the three dimensions in each single frame. This dataset can be used for different Real-time researches. In this work, the proposed work is applied in the UCF dataset. Moreover, our proposed methodology implemented using MATLAB 2020 a software.

In this section, we described object detection in surveillance videos. The Input video must be ".mp4" format. Initially, we have separated the video into several frames. Then we applied the pre-processing technique to convert frames into foreground images and eliminate the noise from Input frames by applying the fraction-calculus method. In addition, by applying the Gaussian mixture model, we have utilized the essential features from execution video and detect the object with the help of these features. Figures (3) & (4) shows the Input video and extracted frame folder.

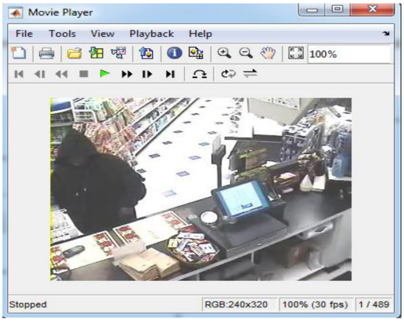

Figure 3
Input video

Figure 4
Extracted frame folder

In Figure (5) & (6) shows the output of noisy image and filtered image. In this Input frame, we have applied the white pixel noise for remove the unwanted pixel. After this process; we have applied the median filter to remove the noise from images.

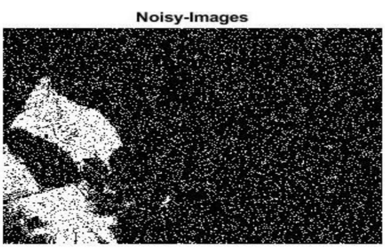

Figure 5
Noisy Image

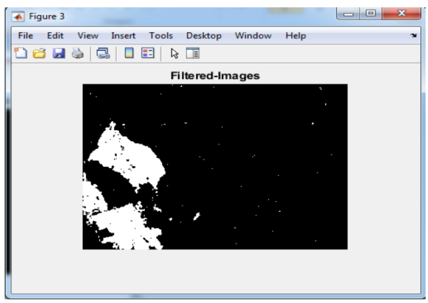

Figure 6
Filtered Image

Figure (7) displays the output image for foreground captured image. In foreground detection, we have detected the foreground objects from images based on background object.

Figure 7

Foreground-image

In Figure (8), the detected part from the video is displayed in bounding box using the Gaussian Mixture Method (GMM)

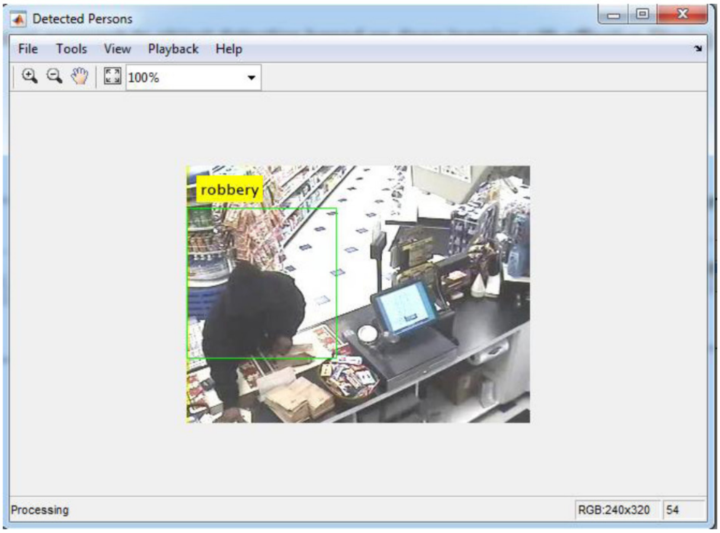

Figure 8

Detected object

The Motion Detection Performance gives better results when comparing to the existing method.

In the following table, the results obtained based on analysing various performance analysis are shown that the proposed work is very effective than existing technique. We compare the accuracy and precision, recall specificity, sensitivity and F-measure.

S. No	Proposed					
	Accuracy	Precision	Recall	`specificity	Sensitivity	F-Measure
1	0.9756	0.81	0.758	0.85	0.9	0.758
CNN						
2	0.95	0.75	0.71	0.79	0.83	0.72
ANN						
3	0.93	0.68	0.65	0.71	0.78	0.7

In the following section, we provide the comparative results of proposed and existing technology to prove that the proposed work is effective than the existing technique

4. Comparative Results analysis

The obtained results in terms of calculating the value such as accuracy, precision, re-call, F-measure, sensitivity and specificity. In the above section, the results are tabulated according to its results such as accuracy, precision, recall, f-measure, specificity and sensitivity. The obtained results have shown that the proposed technique scored higher results than other existing techniques such CNN and ANN. The accuracy results of proposed work are 0.9756. Therefore, the results have proven that the ODEF is better than existing techniques. In Figure(9) the graphical results of the proposed work are shown.

From the comparative analysis of this work, the proposed work is considered as an effective technique than other existing techniques. In next section, the conclusion part of the work is discussed.

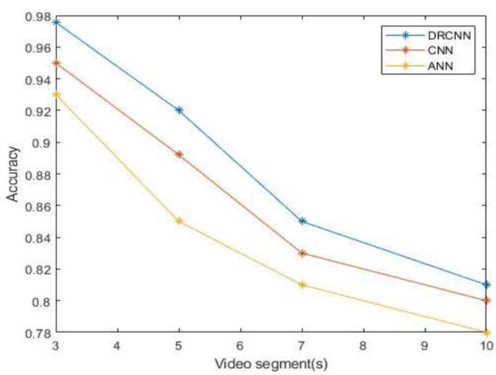

Figure 9

Shows the graphical representation of accuracy

5. Conclusion

In this paper, the proposed ODEF technique depends on hybrid techniques (GMM and DRCNN). In this proposed method we have detected the malicious activity from Video by capturing the region in the Input Videos. Initially, we loaded the Input videos. After that, we converted the frame from videos and reconstruct each frame by applying the Gaussian mixture model (method). Then classified the object by using the Deep learning regions with convolutional neural networks (DRCNN). This method frequently gives sufficient outcomes for all the treated test images and gives improved and boundaries length of the object in videos. The Experimental Result of this method validates that the proposed method captures the malicious activity in bad brightness videos and impressive to capture the object in various sizes. The proposed method reaches a better performance measurement compared to the existing method. In the future, we will concentrate on improving the detecting object size and increase the processing time in the algorithm.

References

[1] Borji, Ali, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang and Jia Li. "Salient object detection: A survey." *Computational Visual Media* (2019): 1-34.

[2] Chen, Yuhuan, Wenbin Zou, Yi Tang, Xia Li, Chen Xu and Nikos Komodakis. "SCOM: Spatiotemporal constrained optimization for salient object detection." *IEEE Transactions on Image Processing* 27, no. 7 (2018): 3345-3357.

[3] Han, Junwei, Dingwen Zhang, Gong Cheng, Nian Liu and Dong Xu. "Advanced deep-learning techniques for salient and category-specific object detection: a survey." *IEEE Signal Processing Magazine* 35, no. 1 (2018): 84-100.

[4] Hua, Xia, Xinqing Wang, Ting Rui, Haitao Zhang and Dong Wang. "A fast self-attention cascaded network for object detection in large scene remote sensing images." *Applied Soft Computing* 94 (2020): 106495.

[5] Varghese, Emmanu, Jaison Mulerikkal and Amitha Mathew. "Video anomaly detection in confined areas." *Procedia Computer Science* 115 (2017): 448-459.

[6] Wang, Fengchao, Zhewei Xu, Yanfen Gan, Chi-Man Vong and Qiong Liu. "SCNet: Scale-aware Coupling-structure Network for Efficient Video Object Detection." *Neurocomputing* (2020).

[7] Huang, Kan, Ge Li and Shan Liu. "Learning Channel-wise Spatio-temporal Representations for Video Salient Object Detection." *Neurocomputing* (2020).

[8] Nikita Sareen, Anuradha Chug & Amit Prakash Singh (2022) An image based prediction system for early blight disease in tomato plants using deep learning algorithm, *Journal of Information and Optimization Sciences*, 43:4, 761-779.

[9] Neeraj Kumar, K. Sudha & Kusum Tharani (2022) Wind power prediction analysis by ANFIS, GA-ANFIS and PSO-ANFIS, *Journal of Information and Optimization Sciences*, 43:3, 481-486.

[10] Nitendra Kumar, Khursheed alam and Abul Hasan siddiqi. "Wavelet Transform for Classification of EEG Signal using SVM and ANN." *Biomedical and Pharmacology Journal* 10(4), 2061-2069.

[11] Giordano, P. & Gaeta, Matteo & Salerno, Saverio & Iovane, G.. (2004). Image mining and massive analysis for large vision machines in Astronomy. *Journal of Information & Optimization Sciences*. 25. 557-588. 10.1080/02522667.2004.10699628.

[12] Hou, Qibin, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu and Philip HS Torr. "Deeply supervised salient object detection with short connections." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, PP.3203-3212,2017.

[13] Tang, Yuxing, Josiah Wang, Xiaofang Wang, Boyang Gao, Emmanuel Dellandréa, Robert Gaizauskas and Liming Chen. "Visual and semantic knowledge transfer for large scale semi-supervised object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol.40, No.12, PP.3045-3058, 2017.

[14] Shen, Yunhang, Rongrong Ji, Changhu Wang, Xi Li and Xuelong Li. "Weakly supervised object detection via object-specific pixel gradient." *IEEE Transactions on Neural Networks and Learning Systems* Vol. 29, No. 12, PP. 5960-5970, 2018.

[15] Kim, Nuri, Donghoon Lee and Songhwai Oh. "Learning instance-aware object detection using determinantal point processes." *Computer Vision and Image Understanding* Vol. 201 PP. 103061, 2020.