OPTIMIZED DEEP CNN BASED OBSTACLE DETECTION FOR AIDING VISUALLY IMPAIRED PERSONS

Anamika Maurya[1], Prabhat Verma[2]
Computer Science and Engineering
HBTU Kanpur, India

ABSTRACT
A visually impaired person faces several challenges while they are moving towards unfamiliar environments. Hence, object detection approaches provide a major solution for this issue. For that, various researchers have developed obstacle detection approaches to help blind people however they have certain limitations. In this research, the optimized deep learning techniques, named Social Optimization algorithm (SOA)-based Deep Convolutional Network (Deep CNN) is developed for assisting visually damaged persons. For effective obstacle detection, the input videos are converted to multiple frames. In feature extraction, relevant features, such as Convolutional Neural Network (CNN) features, Shape Local Binary Texture (SLBT), and hierarchical skeleton features are extracted for further processing. Moreover, the object detection process is carried out using Generative Adversarial Network (GAN). In addition, the object recognition process is done by Deep CNN in which all the layers of Deep CNN are trained using SOA. In addition, the experimental result demonstrates that the developed model attained the testing accuracy, mean average precision (mAP), and recall values of 0.9485, 0.9596, and 0.9735, correspondingly.

KEYWORDS: Deep CNN, Social Optimization algorithm (SOA), Generative Adversarial Network (GAN), Obstacles detection

MSC: 68T06, 68U11

RESUMEN
Una persona con discapacidad visual se enfrenta a varios desafíos mientras se desplaza hacia entornos desconocidos. Por lo tanto, los enfoques de detección de objetos proporcionan una solución importante para este problema. Para eso, varios investigadores han desarrollado enfoques de detección de obstáculos para ayudar a las personas ciegas, sin embargo, tienen ciertas limitaciones. En esta investigación, las técnicas optimizadas de aprendizaje profundo, denominadas Red Convolucional Profunda basada en el algoritmo de Optimización Social (SOA) (Deep CNN) se desarrollan para ayudar a las personas con discapacidad visual. Para una detección efectiva de obstáculos, los videos de entrada se convierten en múltiples cuadros. En la extracción de características, las características relevantes, como las características de la red neuronal convolucional (CNN), la textura binaria local de forma (SLBT) y las características del esqueleto jerárquico se extraen para su posterior procesamiento. Además, el proceso de detección de objetos se lleva a cabo utilizando Generative Adversarial Network (GAN). Además, el proceso de reconocimiento de objetos lo realiza Deep CNN en el que todas las capas de Deep CNN se entrenan utilizando SOA. Además, el resultado experimental demuestra que el modelo desarrollado alcanzó la precisión de la prueba, la precisión media de los promedios (mAP) y los valores de recuperación de 0,9485, 0,9596 y 0,9735, respectivamente.

PALABRAS CLAVE: Deep CNN, Algoritmo de optimización social (SOA), Red generativa antagónica (GAN), Detección de obstáculos

1. INTRODUCTION

Vision acts as an important function in identifying the surrounding objects; however, blindness makes daily life more challenging. According to the World Health Organization (WHO), approximately 285 million people are affected by visual diseases around th[1]e globe. From that, 39 million people are totally blind. Thus, the obstacle detection and providing alertness to the people makes the detection system more challenging. Several supporting systems have been modeled for assisting visually damaged people, which makes their life happy. However, all these systems have certain limitations and complexity while utilizing that by the people. In [12], an ultrasonic-based headset was modeled for helping blind people, which detects an obstacle from both left as well as right sides using two sensors. Nevertheless, this technique failed to detect an obstacle in various directions. An assistive supporting approach provides the visual

[1] anamikamaurya22@gmail.com

information into audio format for going forward to the blind people, which is more effective than human guidance and white cane. An assistive scheme in [11] utilized two ultrasonic sensors in which one for detecting water sources and the other one for tracking objects based on the Global Positioning System (GPS) module. However GPS-assisted approaches are not suitable for obstacle detection in urban environments [18].

Several deep learning and other techniques for obstacle detection have been invented by researchers recently. In [12], region-based CNN has been introduced to identify and recognize objects from real-world environments. In this method, the sharpening, as well as blurring filters was utilized to improve the feature depiction. In [16], Robot Operating System (ROS) with machine vision approaches have been introduced to assist blind people in identifying the obstacles. Manhole, as well as staircase detection, approaches utilized ultrasonic sensors for detecting the obstacles from videos. Moreover, the deep learning of obstacles detection techniques takes less time for processing. Generally, Computer Vision-based Several Electronic Travel Aid (ETA) approaches are broadly employed to detect objects for a visually impaired person, but there exists a difficulty whether to recognize an object or detect an obstacle. In [12], the hybrid obstacle detection approach was invented by the researchers where the hybrid techniques were obtained by joining one ultrasonic sensor as well as a computer vision approach. In addition, the regression-based approach was introduced to detect the object without the necessity of excess information, which utilizes the movement of the camera for object detection. Moreover, smartphone apps have been projected to lead to visually damaged people. Moreover, the object detection approaches have been processed in two modes, such as online mode as well as offline mode. The online mode of object recognition is quicker than the offline mode of object recognition [4].

The main intention of this research is the development of SOA-based Deep CNN. In this research, object detection process is carried out using GAN and the SOA-based Deep CNN is employed to detect an object from videos. Here, the objects are recognized using Deep CNN and the training process is carried out based on SOA.

The organization of this research is explained below. Section 2 describes the literature survey, section 3 portrays the proposed methodology, section 4 explains the result and discussion, and then section 5 concludes the research.

2. LITERATURE SURVEY

The eight literature surveys based on obstacle detection approaches are portrayed below. Saumya Yadav *et al.* [18] developed the fusion-based obstacle detection approach for helping the visually impaired person. Here, the obstacle identification algorithm was introduced based on threshold values. In this method, various kinds of obstacles, such as upstairs, ascending slope, downstairs, and so on were detected using a threshold value. However, this method did not attain maximum recognition accuracy. Steccanella *et al.*[13] introduced the waterline and object detection approach for monitoring the environment. This method is comprised of two steps, like pixel-wise segmentation and object detection. Here, a pixel-wise segmentation approach was employed to detect the binary map for isolating the water and non-water regions. Although this method was effective, the computational complexity of this method was high. Supriyadi *et al.* [15] modeled the obstacle detection approach for visually impaired people. Here, the color filter was utilized to remove the color objects from images, and then a feed-forward neural network was employed to detect the obstacles. However, this method failed to process with bad environments. George Dimas *et al.* [4] introduced the Generative Adversarial Network (GAN) and Fuzzy Set for detecting the obstacle from videos to help blind persons. Here, the GAN model was utilized to detect the saliency, feature sets were employed to translate the spatial information to linguistic values, and then the fuzzy operators were utilized to merge the spatial information with linguistic values. However, the training time of this method was high. In [18], the assistive device can be developed to improve the performance of developed obstacle detection schemes. In addition, a quicker machine learning strategy can be introduced to attain appropriate guidance for visually impaired people. In [4], the developed obstacle detection technique can be extended by forming a multi-dimensional membership function for the risk-level interpretation with speed and the location of an obstacle. In addition, different fuzzy set-based schemes can be intended to perform image fusion for creating an extended dataset [4]. Obstacle detection is a complex process due to the difficulty of features in the image, such as different shapes, sizes, colors and locations. In [2], the deep learning method was developed for obstacle recognition; however, this method did not provide a better result with facial language features.

3. PROPOSED SOA-BASED DEEP CNN FOR OBSTACLE DETECTION

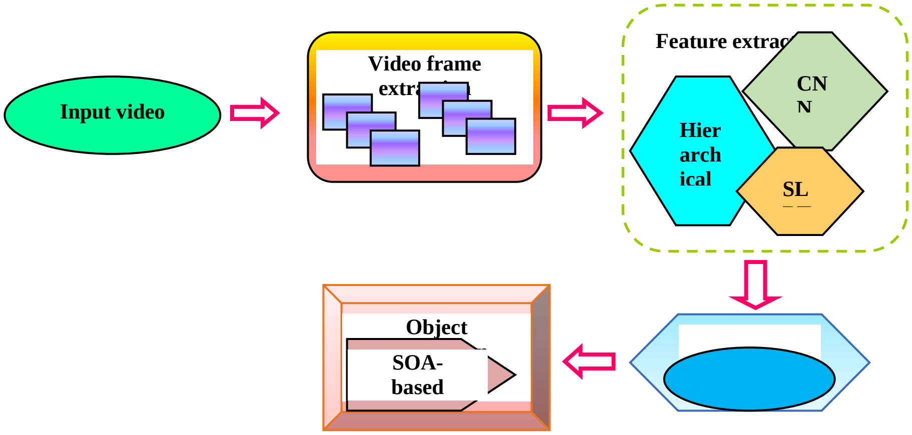

Figure 1. Proposed SOA-based Deep CNN for obstacle detection

This section describes the developed SOA-based Deep CNN for obstacle detection to help visually impaired people. The series of steps followed by the developed SOA-based Deep CNN algorithm are video frame extraction, feature extraction, object detection and object recognition. Initially, the input video is acquired from the video object tracking dataset, which is transformed into a series of frames using video frame extraction method. After that, the features, such as CNN features, SLBT and hierarchical skeleton features are extracted by the feature extraction process to improve the recognition process. Moreover, the object detection process is completed by GAN, and then the object recognition process is carried out using SOA-based Deep CNN. The schematic view of developed SOA-based Deep CNN is depicted in figure 1.

Let us assume the input video database as Ω, which contains o number of videos and is expressed as,

(1)

Here, K denotes the videos in the dataset, K_d represents the d^{th} input video and w indicates the total number of input videos. Here, the input video K_d is selected as input for further processing.

3.1 Video frame extraction

The video frame extraction method converts the input video into multiple numbers of frames for further processing. Here, the input video is acquired from the dataset, which is represented as and then the input video K is converted into several frames, which is represented as,

(2)

where, V_x denotes the x^{th} frame, n represents the total number of frames, which is subjected to the feature extraction process for further processing.

3.2 Feature extraction

In the feature extraction process, three features, such as CNN, SLBT and hierarchical skeleton features are taken out from the extracted frames V_x as input.

i) CNN features: The advantage of selecting the CNN feature is that it attained better detection performance with high convergence speed. CNN features are extracted from the CNN network, which comprises three layers, such as convolution (conv) layer, pooling layer (pool), and fully connected (FC) layer. Here, the CNN features are acquired from FC layers, and then the extracted features are indicated as M.

ii) SLBT feature: SLBT [14] is an effective feature for both the shape and texture-based detection process. The integration of both shapes as well as texture feature is attained by estimating the weights of every shape feature. Moreover, the texture, global and local shapes are obtained using equation (3).

(3)

Where, J specifies the SLBT constraints, denotes the eigenvector, $\bar{h}$ specifies the mean of h .

iii) Hierarchical skeleton features:

Skeleton pruning is an approach, which is employed for extracting the hierarchical skeleton region and removing the skeleton branches or visually inappropriate regions using the boundary abstraction method, namely Discrete Curve Evolution (DCE). Moreover, the expression for hierarchical skeleton feature [10] is represented as,

(4)

where, represents the angle of corner, W indicates hierarchical skeleton feature.

Feature vector representation: The total feature vector is formed by integrating CNN feature M , SLBT feature J and hierarchical skeleton feature W . Thus, the mathematical notation of total feature vector is represented as $\Re$ and then the expression becomes,

(5)

Here, $\Re$ represents the total feature vector and its dimension is .

3.3 Object detection using GAN

The object detection process is carried out using GAN where the input of GAN is considered as the extracted feature $\Re$. The advantage of GAN is that it has fewer training constraints and the training time of this method is minimal as compared to existing. The structural design of GAN is described in the following section.

3.3.1 Structure of GAN

GAN [9] comprises two components, such as generator and discriminator . Here, the generator H is modeled as fully conventional, which does not have any dense layers. The generator assumes an extracted feature $\Re$ as input with dimension , and it is forwarded to the convolutional layer. The outcome of the convolutional layer is subjected to the rectified linear unit (ReLU). In this model, five ReLU and batch normalization layers are followed by the convolutional layers. Moreover, the final outcome is obtained from the layer. After that, the discriminator operation takes place. The discriminator considers the output of layer as input. Here, five convolution layers are followed by a leaky ReLU layer. The output is obtained from the final layer of discriminator M , which is termed as FC. Moreover, the loss function of GAN is represented as,

(6)

Here, H and M are generator and discriminator components, and $\Re$ denotes input. The output obtained from the GAN is represented as M .

3.4 Object recognition using developed SOA-based Deep CNN

This section explains the object recognition process using Deep CNN in which all the layers of Deep CNN are trained using SOA. The input of Deep CNN is illustrated as M .The Deep CNN classifier classifies whether the detected object or obstacle from the video frame is a table or chair or any other object. The structure of Deep CNN is portrayed as below.

3.4.1 Structure of Deep CNN

Deep CNN [8] is comprised of three layers, such as conv, pool and FC layers and each of them performs separate operations. The object recognition process is carried out based on the feature map generated by Deep CNN. In Deep CNN, the feature map is generated, sampled and then fed to the pooling layer. The advantage of Deep CNN is fast processing with better performance. The input of Deep CNN is illustrated as M .

(i) Conv layers:

Here, the feature map is generated based on the input M . The conv layer acts as an input and produced the output based on centred unit and is portrayed as,

(7)

where, $*$ signifies the convolutional operator, represents the rigid feature map of conv layers, denotes the rigid feature map of previous conv layers which is centered at , denotes the bias at conv layers and portrays the kernel function in conv layers.

(ii) Pooling layers: The pooling layer is a non-parametric layer without having any weights or bias, hence the processes are done successfully.

(iii) FC layers: The data acquired from pooling layer is forwarded to the input of FC layer for performing the object recognition, and its expression is given by,

(8)

Hence, the output attained through DCNN is a table or chair or other obstacles, which is represented as .

3.4.2 Structure optimization using SOA

This section explains the training process of Deep CNN structure using SOA. Deep CNN comprises various layers, such as conv layer, activation layer, max pooling layer and dropout layer. Here, the SOA [7] is used to train each layer of Deep CNN individually. SOA is derived based on the principle of community and population in a society. The advantage of SOA algorithm is that it has a high convergence speed and provides optimal performance. The algorithmic procedure for SOA scheme is described as follows.

The algorithmic process of developed SOA is given as,

1) Initialization and fitness computation: SOA operates based on the principle of opportunity equality, which describes every person having their own right to shift from one occasion to another occasion as their wish. The initialization function of society is expressed as,

(9)

Where, indicates the position of an individual in society, v_b depicts the objective function of every human being in the community.

2) Estimating Update equation: The update function of SOA algorithm is evaluated depending on the equality of opportunity strategy. In this stage, the subsequent location of individuals are updated, which is portrayed as,

(10)

where, $rand$ shows the random number , F specifies the optimal position of a person in the society, and G indicates the coefficient of individual choice .

3) Empty_point estimation: For resolving the inequality, there is a need to evaluate the density point, which is based on the empty_point and is expressed as,

(11)

where, symbolizes the location of a person, v_b indicates the objective function of every person. Here, if the closeness among the empty point as well as best point is fewer, then the society is named as a desirable society. Consequently, the expression becomes,

(12)

4.) Estimating density _point: If the society with all individuals shares their positive vibes, the optimal location is determined based on the density point and is portrayed as,

(13)

where, symbolizes the position of a person, v_b signifies the objective function of each person in the society.

5) Re-evaluation of solution with error: The error is re-examined based on the fitness function described in equation (8). The solution makes a minimal error, which is utilized for obstacle detection using Deep CNN.

6) Estimate optimal weight with proposed soa and terminate: The error with all solutions is re-evaluated using SOA so that solution having a lesser error is considered to train Deep CNN. The optimum weights are achieved frequently till maximum iterations are reached.

4. RESULTS AND DISCUSSION

This section describes the results and discussion of developed SOA-based Deep CNN for obstacle detection to help visually impaired people based on accuracy, mean average precision and recall.

4.1 Experimental setup and Performance metrics

The execution of developed SOA-based Deep CNN method is carried out in python tool with Keras tool using PC along with Windows 10 OS and Inter i3 core processor. The dataset employed for the developed object detection algorithm is the video object tracking dataset [17], which contains 13 directories and 1 file. The performance metrics employed for the developed SOA-based Deep CNN for obstacle detection are testing accuracy, mean average precision and recall.

4.2 Experimental outcome

This section demonstrates the experimental outcome of developed SOA-based Deep CNN for obstacle detection. Figure 2 a) shows the extracted input frame from the video and figure 5 b) demonstrates the detected obstacle from the frame.

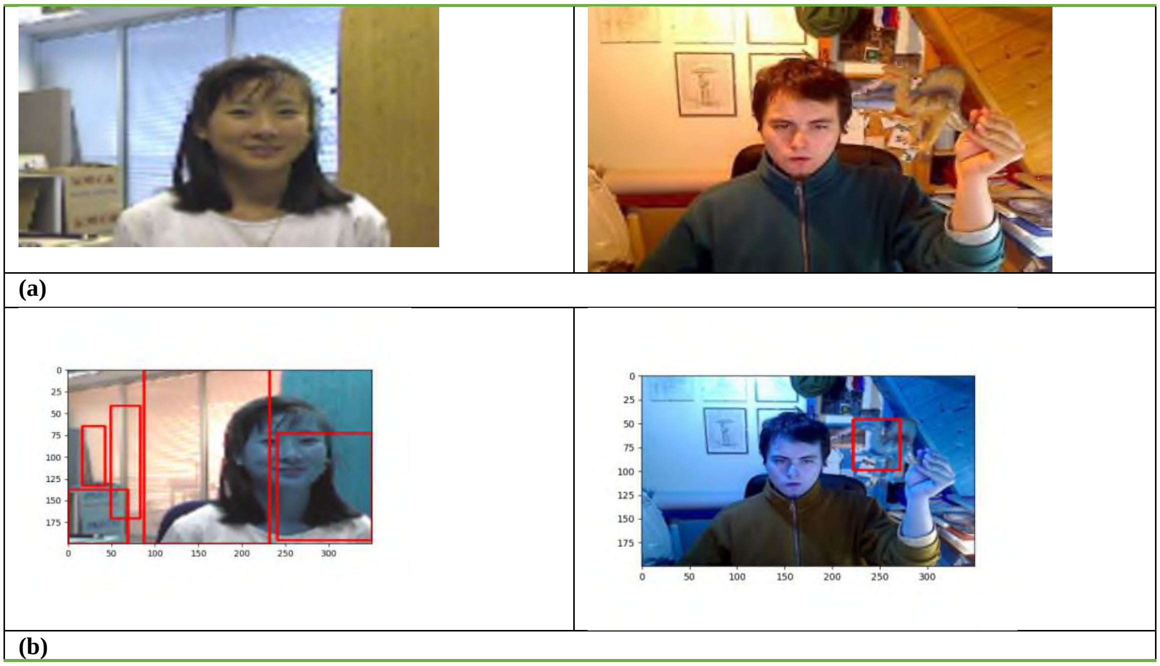

Figure 2. Experimental outcome of developed SOA-based Deep CNN for obstacle detection a) Extracted frames from input video, b) Detected obstacle

4.3 Comparative techniques

The performance of the developed SOA-based Deep CNN method is assessed by comparing the obtained result with existing methods, such as LSTM [6], DisNet [1], DNN [3] and HGSO+Deep CNN. The table 1 shows the comparative assessment of the developed model by varying the percentage of training data and k-fold value.

Table 1. Comparative discussion

Variations	Metric	LSTM	DisNet	DNN	HGSO+Deep CNN	Developed SOA-based Deep CNN
Training data	Testing accuracy	0.8014	0.8241	0.8847	0.9125	0.9374
	mAP	0.7954	0.8124	0.8541	0.8965	0.9541
	Recall	0.8041	0.8354	0.8745	0.8954	0.9685
K-fold value	Testing accuracy	0.8140	0.8456	0.8772	0.9293	0.9485
	mAP	0.8142	0.8745	0.8985	0.9365	0.9596
	Recall	0.7754	0.8254	0.9075	0.9325	0.9735

5. CONCLUSION

This section describes the developed SOA-based Deep CNN for obstacle detection to help visually impaired people. The processing steps employed by the developed SOA-based Deep CNN algorithm are video frame extraction, feature extraction, object detection and object recognition. The video frame extraction method transforms the captured videos into multiple frames and the feature extraction method extracts the relevant features, like CNN features, SLBT and hierarchical skeleton features. In addition, the object detection process is done by SeGAN, where the process is done using a generator and discriminator. In SeGAN, the detected object is acquired from the tanh layer. Moreover, Deep CNN recognizes the obstacle from frames with high testing accuracy. In the developed model, the training process of various layers that exists in Deep CNN is completed using the SOA algorithm, which operates based on the principle of community and population in a society. Here, the video object tracking dataset is utilized for the experimentation of the developed model. Furthermore, the experimental result demonstrates that the developed model attained the testing accuracy, mean average precision (mAP) and recall values of 0.9485, 0.9596 and 0.9735, correspondingly.

RECEIVED: AUGUST, 2022.
REVISED: MARCH, 2023.

REFERENCES

[1] AADI, F.Z., SADIQ, A. (2020): Proposed real-time obstacle detection system for visually impaired assistance based on deep learning, **International Journal of Advanced Trends in computer science and Engineering,** vol.9, no.4.
[2] ASHRAF, A., NOOR, S.B., FAROOQ, M.A., ALI, A., and HASHAM, A. (2020): Iot Empowered Smart Stick Assistance For Visually Impaired People. **International Journal of Scientific & Technology Research,** 9, 356-360.
[3] BASHIRI, F.S., LAROSE, E., BADGER, J.C., D'SOUZA, R.M., YU, Z., and PEISSIG, P. (2018): Object Detection to Assist Visually Impaired People: A Deep Neural Network Adventure. In:, **Advances in Visual Computing. ISVC 2018. Lecture Notes in Computer Science,** vol 11241. Springer, Cham. https://doi.org/10.1007/978-3-030-03801-4_44
[4] DIMAS, G., NTAKOLIA, C. and IAKOvidis, D.K. (2019): Obstacle detection based on generative adversarial networks and fuzzy sets for computer-assisted navigation, In **Proceedings of International Conference on Engineering Applications of Neural Networks**, 533-544.
[5] HABIB, A., ISLAM, M.M., KABIR, M.N., MREDUL, M.B. and HASAN, M. (2019): Staircase Detection to Guide Visually Impaired People: A Hybrid Approach, **Revue d'Intelligence Artificielle., .**33, 327-334
[6] HENGLE, A., KULKARNI, A., BAVADEKAR, N., KULKARNI, N., and UDYAWAR. R. (2020) : Smart Cap: A Deep Learning and IoT Based Assistant for the Visually Impaired, In **Proceedings of 2020 Third International Conference on Smart Systems and Inventive Technology** (ICSSIT), 1109-1116.
[7] KARIMI, N., and KHANDANI, K. (2020): Social optimization algorithm with application to economic dispatch problem, **International Transactions on Electrical Energy Systems**, 30, 12593.
[8] LAkshmiprabha, N.S. and MAJUMDER, S. (2012): Face recognition system invariant to plastic surgery, In **Proceedings of 2012 12th International Conference on Intelligent Systems Design and Applications** (ISDA), 258-263.
[9] PASCUAL, S., BONAFONTE, A. and SERRA, J. (2017): SEGAN: Speech enhancement generative adversarial network.
[10] REN, S., HE, K., GIRSHICK, R. and SUN, J. (2015): Faster r-cnn: Towards real-time object detection with region proposal networks, **Advances in Neural Information Processing Systems**, 28, 91-99.
[11] SAHOO, N., LIN, H.W. and Chang, Y.H. (2019): Design and implementation of a walking stick aid for visually challenged people, **Sensors**, 19, 130
[12] SHEEBA, P. T., and MURUGAN, S. (2022): Fuzzy dragon deep belief neural network for activity recognition using hierarchical skeleton features. **Evolutionary Intelligence**, 15, 907-924.
[13] STECCANELLA, L., BLOISI, D.D., CASTELLINI, A. and FARINELLI, A. (2020): Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring, **Robotics and Autonomous Systems,** 124, 103346.
[14] SUGAVE, S. and JAGDALE, B. (2020): Monarch-EWA: Monarch-Earthworm-Based Secure Routing Protocol in IoT, **The Computer Journal**, .63, 817-831.
[15] SUPRIYADI, T., SETIADI, B. and Nugroho, H. (2020) : Pedestrian lane and obstacle detection for blind people, In **Proceedings of Journal of Physics: Conference Series**, 1450, 012036.
[16] SURESH, A., ARORA, C., LAHA, D., GABA, D. and BHAMBRI, S.(2017): Intelligent smart glass for visually impaired using deep learning machine vision techniques and robot operating system (ROS), In **Proceedings of International Conference on Robot Intelligence Technology and Applications**, 99-112.
[17] VIDEO OBJECT TRACKING DATASET : taken from, **"https://www.kaggle.com/kmader/videoobjecttracking**", accessed on 2021
[18] YADAV, S., JOSHI, R.C., DUTTA, M.K., KIAC, M. and SIKORA, P. (2020): Fusion of Object Recognition and Obstacle Detection approach for Assisting Visually Challenged Person, In **Proceedings of 2020 43rd International Conference on Telecommunications and Signal Processing (TSP)**, 537-540.