Use of Optimized Convolutional Neural Network (OCNN) for Object Detection

Suraj Pardeshi[a], Pravin Yannawar[a]

[a]Vision and Intelligent System Laboratory, Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India, surajrp.cmc@gmail.com,
plyannawar.csit@bamu.ac.in

Abstract

In the daily life routine, Visually Impaired (VI) persons are mostly suffering from the trouble of physical movements. As getting the exact information regarding the objects ahead and accordingly move further without colliding or not getting harmed is a greatest challenge for them in the indoor environments also. The purpose of this research is to propose model for object detection in an indoor environment for visually impaired person. Initially, the input image from real-world scenario is given to pre-processing via wiener filtering. Subsequently, segmentation is carried out by the novel proposed Multi-Kernel K-means clustering model and SURF (Speeded-Up Robust Features), SIFT (Scale-Invariant Feature Transform), Shape based features via canny edges & Gradient features via HOG (Histogram of Oriented Gradients) were extracted from segmented image. The optimal features are selected from extracted features by a new hybrid algorithm referred as Particle Hybridized SeaLion Optimization Algorithm (PHSLA) and selected optimal features are classified using Optimized Convolutional Neural Network (OCNN) for detecting the object. It was observed from the comparative study & performance evaluation that OCNN and PHSLA received 98% accuracy in identification of object and helping system to become more stable and reliable.

Keywords: Visually Impaired (VC); Object Detection; Multi-Kernel K-means clustering; Particle Hybridized SeaLion Optimization Algorithm (PHSLA); Optimized Convolutional Neural Network (OCNN).

1. INTRODUCTION

Sense is a biological system used by an organism for sensation. Humans are blessed with *five* basic *senses*: *touch, sight, hearing, smell* and *taste*. Among these, we greatly depend on the most important sensory organ of our body: Eyes. We cannot imagine our life without eyesight. Unfortunately considerable amount of human population across the globe suffers from partial or complete loss of vision, blurred vision, night blindness, uncorrected refractive errors and many more vision impairments due to diabetes and other disorders [1]. According to the WHO (World Health Organization) report on visual impairment, 314 million of the population is suffering with the ill effects of poor vision [2-5]. As clear vision is the fundamental necessity of every human, any defect in vision spoils almost all of the daily routine of the person encumbering his confidence and enthusiasm everywhere. The absence of a suitable assistive technology makes Visually Challenged (VC) people excessively dependent on others for the simple day-to-day activities. Many applications were observed and operating in limited context based on the incorporation of sensory techniques such as use of ultrasonic sensors, radio frequencies etc. Although many major researches had contributed in the area of assistive technology for VC but still there is a need of reliable, compact, wearable, cost effective and real-time fast processing device/gadget which can help the blinds to identify obstacles in front of them while moving indoors in day to day life. In the context of recent technological advancements in the area of *image processing*, *computer vision, Convolutional Neural Network* (CNN) and *Adaptive Machine Learning* (ML) approaches, it is now possible to think of an application which combines the best of these areas and develop an interface which can mimic the actual human visual system to some extent. In this direction, this research work is presented on automatic detection and classification of indoor objects in the context of VC might come across in his/her day to day life activities. Due to the complex and diversified environment this work becomes complex and challenging because of recognizing the object in an indoor environment in the presence of overlapping of objects, partial object visibility, moving objects, different lighting conditions and so on.

2. RELATED WORK

The research in the areas of computer vision and pattern recognition have been extremely impressive and some of the remarkable contributions are required to consider while progressing in the area of development of assistive technology in the context of visually challenged person. Table 1.Summarizes, some of impressive methodology presented by researchers with considered features and challenges in the referred work in the context of an object identification and navigation for visually impaired people.

Author	Adopted Methodology	Features	Challenges
Meshram et al.[1]	NavCan	✓ Improves obstacle-free navigation performance. ✓ High usability and effectiveness	• Lower detection accuracy • Need to improve the improved user interface
Arora et al. [6]	Single-shot Multi-box Detection Framework	✓ Minimal response time ✓ Compact and portable ✓ Mobility at an affordable price	• Lower latency • Tedious • For increased precision, high-speed dynamic algorithms may be used.
Afif et al. [7]	RetinaNet	✓ Improve detection performances ✓ Improves processing time ✓ High detection precisions	• Indoor object detection accuracy can be increased • Total mean average precision (map) can be enhanced
Cardillo et al. [8]	Electromagnetic sensor	✓ Noise tolerance ✓ Reduced dimensions. ✓ Cost-effective system	• Lower stability and robustness • Echoes are audibly louder than the background noise.
Ye and Qian [9]	3-D object recognition method	✓ High success rate in object recognition ✓ Good scalability and parallelism ✓ Efficacy in detecting structural objects	• Higher computational cost and time • Weight of device is higher • The success rate of object recognition should be increased.
Krishna et al. [10]	Convolutional Neural Network (CNN)	✓ Simple ✓ accurate for navigation ✓ Enable the device to be used both indoors and outdoors.	• Obstacle avoidance algorithm for enhancing the system drastically • More lightweight neural networks can be used for accurate target detection.
Chan et al. [11]	DCNN Object Detection Processor	✓ Enhance the navigation effectiveness ✓ More precise edge detection	• Not suitable for real-time • Effectiveness of image enhancement algorithms can be improved
Jindal et al. [12]	Smart Phone based Cost-effective System	✓ High Accuracy ✓ Less prone to noise	• Applicable for less datasets • Lower sensitivity as well as specificity • Lower success rate

Table 1.Features and Challenges of Existing Object Detection Models for Assisting Visually Impaired People

3. PROPOSED OBJECT DETECTION MODEL FOR VISUALLY IMPAIRED PEOPLE (ODMVI)

The proposed model focuses on an indoor environment where the objects in the scene are stationary and moreover the scene is captured by the head mounted scene acquisition device posting data to the connected systems for the purpose of real time object detection and transliteration in audio in to ear buds of the person. In order to handle the complexity in object detection mechanism, the presented research work provides the detailed discussion over black-box functionality of the proposed system. The novel Object Detection Model for Visually Impaired persons with an objective of real-time assistance to visually impaired/challenged person through robust, fast and accurate object classification mechanism. The proposed ODMVI is designed using five major phases as *pre-processing, segmentation, feature extraction, optimal feature selection* and *object detection* as depicted in figure 1 and function of each phase discussed as,

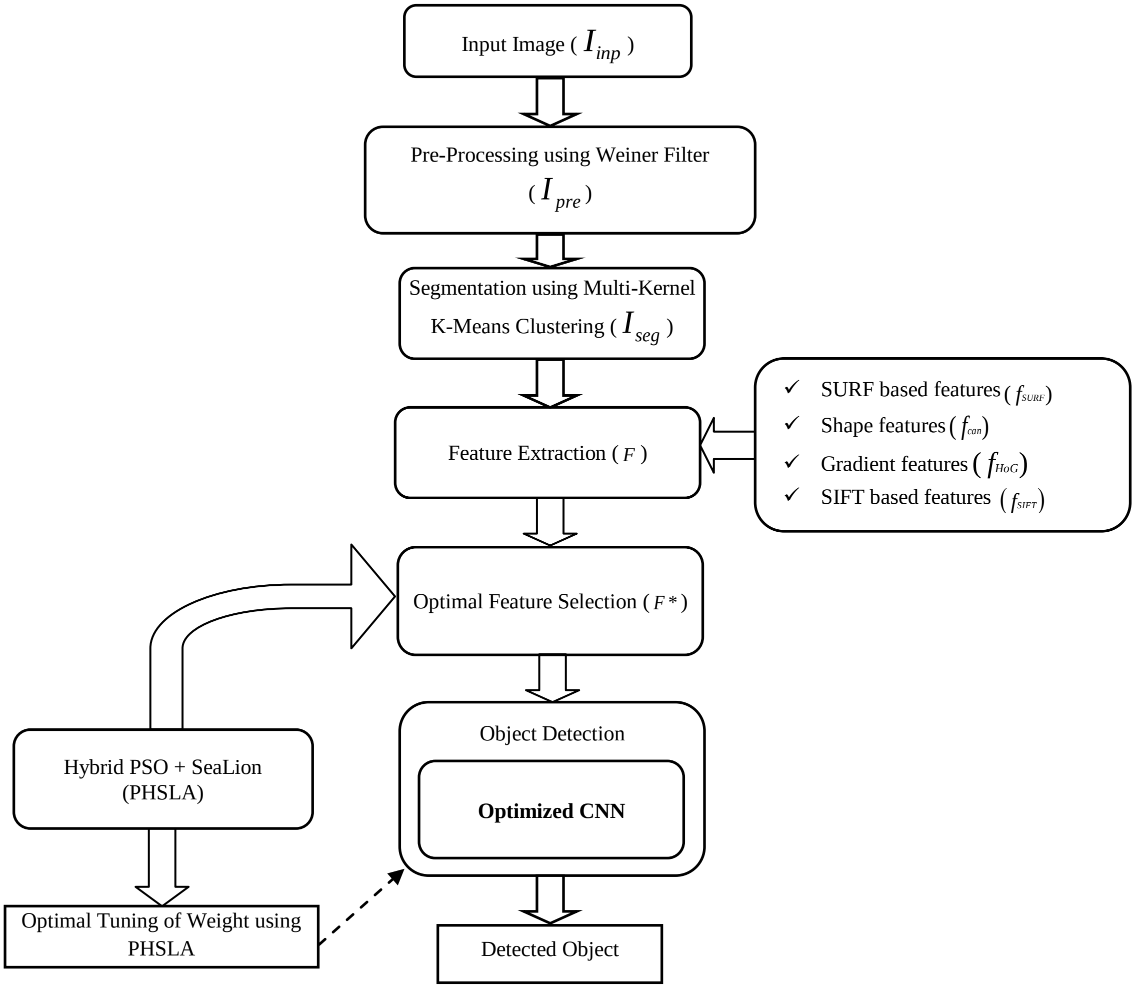

Input Image (I_{inp})
Pre-Processing using Weiner Filter (I_{pre})
Segmentation using Multi-Kernel K-Means Clustering (I_{seg})
Feature Extraction (F)
✓ SURF based features (f_{SURF})
✓ Shape features (f_{can})
✓ Gradient features (f_{HoG})
✓ SIFT based features (f_{SIFT})
Optimal Feature Selection ($F*$)
Hybrid PSO + SeaLion (PHSLA)
Optimal Tuning of Weight using PHSLA
Object Detection
Optimized CNN
Detected Object

Fig.1. Architecture of the ODMVI

3.1 Pre-Processing via Weiner Filter

Initially, the input image under consideration I_{inp} is given to pre-processing for enhancing the image structure at different scales. Here, the pre-processing is accomplished with the aid of the wiener filtering with the intent of removing the unwanted artifacts from I_{inp} image and pre-processed image is represented by I_{pre}. Unwanted artifacts remove by wiener filtering if not handled appropriately contributes in errors towards determination of feature sets. The choice of Weiner filter is due to the highest PSNR (Peak Signal-to-Noise Ratio) (in dB) thus having the lowest MSE (Mean-Square Error) (in dB), as compared to other compatible ones. Furthermore, the Weiner filter performs the optimal noise smoothing and inverse filtering tradeoffs.

The input image is the combination of the original image and the noisy image as Eq. (1),

$$I_{inp}(t,u) = org(t,u) + nos(t,u) \tag{1}$$

Where *(t,u),* represents the pixel position in Image *I*. In the input image we are using Weiner filter and obtaining the output *O(t,u)*.The standard mathematical equation for the Weiner filter is represented by Eq.(2) providing degradation function H(t,u) also called Optical Transfer function (OTF)

$$H(t,u) = \frac{P_D(t,u)}{P_D(t,u) + P_N(t,u)} \tag{2}$$

Where: $P_N(t,u)$ is the power spectrum of noise $nos(t,u)$ and $P_D(t,u)$ is the power spectrum of the original image $org(t,u)$ .For simplification we need to take derivative of Eq. (2) reduced to Eq. (3) by computing DFT of original Image *(O(t,u))* and DFT of noisy image *(N(t,u))*

$$J = E\left[\left|O((t,u)) - H((t,u)).N((t,u))\right|^2\right] \tag{3}$$

$$H(t,u) = \frac{E\left[N((t,u)).O^*((t,u))\right]}{E\left[\left|N((t,u))\right|^2\right]} \tag{4}$$

Here, the complex conjugate is denoted by $*$. In the case of the white noise availability, the numerator gets decreased as per Eq. (5). The denominator reduces as per Eq. (6). The Weiner filter's output is given as in Eq. (7) and (8), respectively. The inverse transform of DFT is IDFT (Inverse Discrete Fourier Transform). The pre-processed image I_{pre} is subjected to segmentation.

$$E\left[N(t,u).O^*(t,u)\right] = \begin{cases} = E\{\left[O(t,u)\right] + N(t,u)\} \times O^*(t,u) \\ = E\left[\left|O(t,u)\right|^2\right] \\ = P_D(t,u) \end{cases} \tag{5}$$

$$E\left[\left|O((t,u))\right|^2\right] = P_D((t,u)) + P_N((t,u)) \tag{6}$$

This helps in increasing the quality of the image, by multiplying input image with Weiner filter as per Eq. (7)and resultant image presented as $Y(t,u)$.

$$Y(t,u) = H(t,u) I_{inp}(t,u) \tag{7}$$

Finally, wiener filter need to take Inverse Discrete Fourier Transform (IDFT) as per Eq.(8).

$$y(t,u) = IDFT\left[Y(t,u)\right] \tag{8}$$

Subsequently, the resultant image of preprocessing using wiener filtering is to be passed for segmentation new Multi-Kernel K-Means clustering model

3.2. Multi-Kernel K-Means Model for Segmentation

The pre-processed images I_{pre} are segmented via the proposed Multi-Kernel K-Means Algorithm. The Multi-Kernel k-mean clustering is unsupervised. K-Means clustering algorithm is used to differentiate the interest region from the context based on the K-centroids; it clusters or partitions the given data into K-clusters or segments. It is predominantly used for clustering large sets of images into Region of Interest (ROI) and Non-ROI regions. The resolution of the image I_{pre} is*(a,b)*, which is clustered into k count of clusters. Let I_{pre} *(a,b)*be the input pixel that is to be clustered and the cluster centre is Cen_k . The steps followed in the k-means algorithm are below:

Algorithm: Multi-Kernel K-Means (I_{pre})

Step 1: Initialize the cluster centre and the count of clusters k .

Step 2: Compute the Euclidean distance $Dist$ among every pixel in I_{pre}. In fact, using the relation below, the Euclidean distance between the image's centre and each pixel is computed.

$$Dist = \| I_{pre}(a,b) - Cen_k \| \tag{9}$$

Step 3: Depending on the distance $Dist$, assign all pixels to the nearest centre.

Step 4: Recalculate the new location of the centre using the relation in Eq. (10), after all, pixels have been allocated.

$$Cen_k = \frac{1}{k}\sum_{a\in c_k}\sum_{b\in c_k} I_{pre}(a,b) \tag{10}$$

Step 5: Repeat the procedure before the tolerance or error value is met.

Step 6: Reshape the cluster pixels into the image.

While k-means has the advantage of being simple to be used, it does have some disadvantages. The final clustering results' consistency is determined by the random initial centroid selection. As a result, if the initial centroid is selected at random, the result would be different for different initial centers. Moreover, the K-means algorithm converges at a local minimum and it is highly computationally complex. So, we have introduced a new multi-kernel k-Means algorithm, where we've hybridized both the sigmoid and *Laplacian kernel*. Mathematically, Sigmoid and *Laplacian kernel* are shown in Eq. (11) and Eq. (12), respectively.

$$\textit{Sigmoid kernel}: k_1(a,b) = \tanh\left(\beta_0\langle a,b\rangle + \beta_1\right) \tag{11}$$

$$\textit{Laplacian kernel}: \ k_2(a,b) = \exp\left(-\frac{\|a-b\|}{\sigma}\right) \tag{12}$$

By combining Eq.(11) & (12) and by normalizing by Eq. (3) we are obtaining the cluster center using Eq. (13) and the segmented image thus obtained is denoted asI_{seg}, from which the multi-features are extracted.

$$k(a,b) = \frac{k_1 + k_2}{3} \tag{13}$$

3.3 Multi-Feature Extraction

The segmented image is denoted as I_{seg}, from which the multi-features (*SURF, shape based features via canny edge detection model, gradient features via HOG and SIFT based features*) are extracted. The features extracted from the segmented image and its relation with multi-feature extraction approach is as shown in Fig.2.

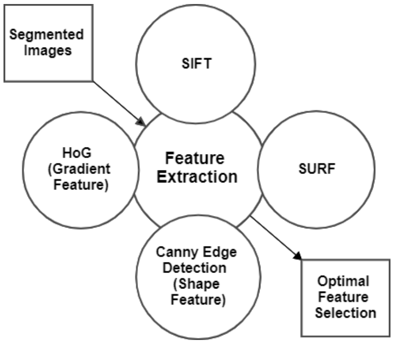


Fig. 2. Feature Extraction

These additional features using SURF (f_{SURF}), shape based features via canny edge detection model (f_{can}), gradient features via HOG (f_{HOG}) and SIFT based features (f_{SIFT})were extracted as below.

3.3.1. SURF (Speeded Up Robust Features)

The *feature description* as well as *feature extraction* is the two major steps in the SURF model. In the feature extraction phase, the *Hessian matrix approximation* has been used for detecting the interest points. In (I_{seq}), the sum of values (pixel values) is computed in an effective manner using the *Integral Image*. In addition, it also computes the average intensity with in I_{seg}at the location $R=(o,p)^T$, the Integral Image $Int_{\sum}$ (R) is computed as per Eq. (14).

$$Int_{\sum}(R)=\sum_{h=0}^{h\le o}\sum_{g=0}^{h\le p}Int(h,g) \qquad (14)$$

The SURF descriptor is generated in two steps for the image as, at first repeatable orientation using data from a circular area surrounding the key point were established. The SURF descriptor is then extracted from a square region aligned to the chosen orientation. SURF attempts to find a reproducible orientation for the interest points in order to be rotation invariant for orientation assignments. In order to perform this, SURF implies computation of Haar wavelet responses along X & Y directions of the segmented image as well as its circular neighborhood is also explored in the radius of 6's. The Hessian of this pixel (o,p) $I_{seg}$is denoted as per Eq. (15).

$$H(Fun(o,p))=\begin{bmatrix}\frac{\partial^2 fun}{\partial o^2} & \frac{\partial^2 fun}{\partial o.\partial p}\\ \frac{\partial^2 fun}{\partial o.\partial p} & \frac{\partial^2 fun}{\partial p^2}\end{bmatrix} \qquad (15)$$

Partially differentiating Eq.(15), we get Eq. (16) as follows:

The Hessian matrix $$\aleph(R,\gamma)=\begin{bmatrix}G_{oo}(R,\gamma) & G_{op}(R,\gamma)\\ G_{op}(R,\gamma) & G_{pp}(R,\gamma)\end{bmatrix} \qquad (16)$$

Here, $G_{oo}(R,\gamma)$ is the convolution corresponding to the 2nd order Gaussian derivative for I_{seg} in point R .In addition, $G_{op}(R,\gamma)$ and $G_{pp}(R,\gamma)$ points to the similarly. As a result, the wavelets are observed large at large scales due to which integral images can be used for quick filtering once more. In the second phase, the sums of *vertically* and *horizontally* wavelet responses are included in the scanning field, the scanning orientation is accordingly adjusted, and recalculate these attributes until the orientation with the highest sum value were not found, which is the key orientation of the function descriptor. The Descriptor Components represents the square area, created on the basis of centered key point aligned in the direction of the orientation. Isolated region is then divided into smaller 4* 4 square-shaped sub-regions on a constant basis. Naturally, at 5*5 regularly spaced sampling points were considered and useful in computation of sub-regions in the horizontal direction. The Haar wavelet responses referred to as d_x, while in the vertical direction it is referred to as d_y. In order to enhance the robustness against geometric deformations and localization errors, the responses d_x and d_y were weighted initially with a Gaussian (σ = 3.3s). The extracted SURF feature is denoted as (f_{SURF}).

3.3.2. SIFT

The SIFT [40] feature is extracted from preprocessed image I_{pre}. In fact, the SIFT is indeed a simple procedure. The SIFT algorithm consists primarily of four stages, at first it concentrated towards Selection of a scale-space peak where it is the possible spot for locating features selected from the segmented image and defined by Eq. (17)

$$L(c,d,\sigma) = G(c,d,\sigma) * I_{seg}(c,d) \tag{17}$$

Where, $I_{seg}(c,d)$ is the segmented image with pixels (c,d) and $G(c,d,\sigma)$ is the Gaussian variable scale. In addition, * points to the convolution operator to generate the key point in *scalespace*. However it was seen that, there are various method available for finding stable key point positions in the *scalespace* and one such technique is the difference of Gaussians, which locates *scalespace* extrema, $U(c,d,\sigma)$ by computing the difference between two images, one with scale l times that of the other.

$$U(c,d,\sigma) = O(c,d,l\sigma) - O(c,d,\sigma) \tag{18}$$

Once stable key point in *scalespace* is located there is need of second step, key point Localization where the feature key points from the selected scale-space peak are localized accurately. The key points are generated in the previous phase resulted in a large number of key points, some of them seem to be too close to the edge, or there isn't enough contrast. They aren't as useful as features in this scenario, as a result it is necessary to get rid of such key points, and we get rid of them. The method is comparable to those used to suppress edge features in the Harris Corner Detector. The extrema location Q is given as per Eq. (19).

$$Q = \frac{\partial^2 U^{-1}}{\partial c^2} \frac{\partial^2 U}{\partial c} \tag{19}$$

After key point localization it is necessary to assign orientation to Key points depending on the scale. In order to do so, a neighborhood is drawn around the key point spot, and gradient magnitude and direction of the region were determined. The result is a 360-degree orientation histogram of 36 bins and its histogram is created. It was seen that, the peak in the histogram were found at some point. Hence, the orientation ϕ is computed as per Eq. (20).

$$\phi(c,d) = \tan(O(c,d+1) - O(c,d-1) / O(c+1,d) - O(c-1,y) \tag{20}$$

The key point descriptor is required to be defined and it is done through high-dimensional vector that describes the key points. Each key point now has an attributes like *position*, *size*, and *orientation*. The next stage is to develop a descriptor for each key point's local image area which is strongly distinctive and as invariant as possible to changes in perspective as well as lighting and at last the function of key point matching is performed. The matching is performed on the basis of key point's closeness of between two images and such key points are paired. The extracted SIFT feature is denoted as (f_{SIFT}).

3.3.3. Shape features via canny edge detection

The shape based features are extracted from I_{pre}. This computation is required to determine the shape of an object under observations and this can be effectively done with the use of Canny edge detection [41]. The Canny edge detector recognizes a large variety of edges in images using a multi-stage algorithm (Canny Filter). To compute the intensity corresponding to the image gradients, Canny, employs a filter dependent on the derivative of a Gaussian. The Gaussian filter eliminates the influence of image noise by eliminating *non-maximum pixels* of the gradient *magnitude*, possible edges are thinned down to 1-pixel curves. Finally, using hysteresis thresholding on the gradient magnitude, edge pixels are retained or deleted. The general criteria for edge detection include edge detection with a low error rate, which ensures that the detection can recognize as much of the image's edges as possible. Similarly the operated sensed edge point should be effective in locating the edge's centre and moreover image noise does not produce false edges in this case. The extracted shape base features via the Canny edge detector are denoted as (f_{can}).

3.3.4. Gradient features via HOG

InI_{seg}, the HOG [42] (feature descriptor) identifies the homogeneous identical area and the features are computed using following steps

(a) Calculate the *histogram of gradient directions or edge orientations* of each pixel in each cell by dividing the pre-processed image into smaller related regions (referred to as cells).
(b)Each cell is discretized into angular bins using gradient orientation.
(c)Each cell's pixel about its angular bin receives a weighted gradient.
(d) Consider a spatial region to be a collection of neighbouring cells (blocks).

The block histogram is formed by representing the Normalized type of histograms and is referred to as the descriptor. It is dependent on the classification and normalisation of histograms. The f_{HOG} represents the extracted HOG characteristics.The overall extracted features are constituted in feature matrix denoted as

$$F = (f_{SURF}) + (f_{can}) + (f_{HOG}) + (f_{SIFT}).$$

The resultant matrix of feature is composed by the model vectors of $(f_{SURF}), (f_{can}), (f_{HOG}), (f_{SIFT})$, therefore in this case the problem of *curse of dimensionality* needs to be handled appropriately, so that the best optimal features from F can be derived and represented as F^* .
In order to address *curse of dimensionality*, a novel hybrid algorithm referred as PHSLA[41] is used to achieve this optimum feature selection and these computed optimal features F^* are given as input to Optimized Convolutional Neural Network (CNN). The weights of CNN are optimally tuned using the new PHSLA model to make target recognition more accurate. The PHSLA incorporates the principles of both PSO (Particle Swarm Optimization) and SLnO on I_{seg}, segmented image.

4. DATASET

The dataset used for the simulation of the proposed model is *ImageNet*. To overcome the problem of overlapping objects as well as objects being camouflaged in the background it is necessary to enhance the performance of underlying algorithm to handle such complexity in object detection and classification ImageNet dataset is found to be very useful. The data in this dataset is available for free to researchers for non-commercial use. ImageNet dataset has 100,000 images across 200 classes. Each class has 500 training images, 50 validation images, and 50 test images provided with the labeling of images. The proposed ODMVI model is trained and evaluated over ImageNet dataset. 10000 images of over 25 different categories were selected from the dataset for training the model. Initially most commonly used indoor objects were targeted and organized in to respective directories for training. The images were labeled with a string starting with 'n' preceded by a sequence of eight integers e.g. 'n03950228'. Every directory contained 400 different images of a particular object. The images were labeled with the name of directory preceded by underscore and sequence number e.g. 'n03950228_1, n03950228_2 etc). After the complete analysis of CNN over the dataset, the model was trained to identify the objects with remarkable accuracy and precision.

5. OBJECT DETECTION VIA OPTIMIZED CNN USING PHSLA

The Deep neural networks have demonstrated their capacity to solve classification problems using a hierarchical model, there are millions of parameters, and learning with big datasets of ImageNet. In order to utilize the multiple features set as well optimal features the CNN and PHSLA needs to be configured as below

5.1 CNN Architecture

Convolutional neural networks (CNN) are a special class of deep neural networks that consist of several convolutions, pooling, and fully connected layers; this has proven tobe a robust method for image or video processing, classification, and pattern recognition. Figure 3 below depicts the schematic diagram of a basic convolutional neural network using proposed PHSLA for feature extraction as well as optimal tuning of weights of CNN for ODMVI.

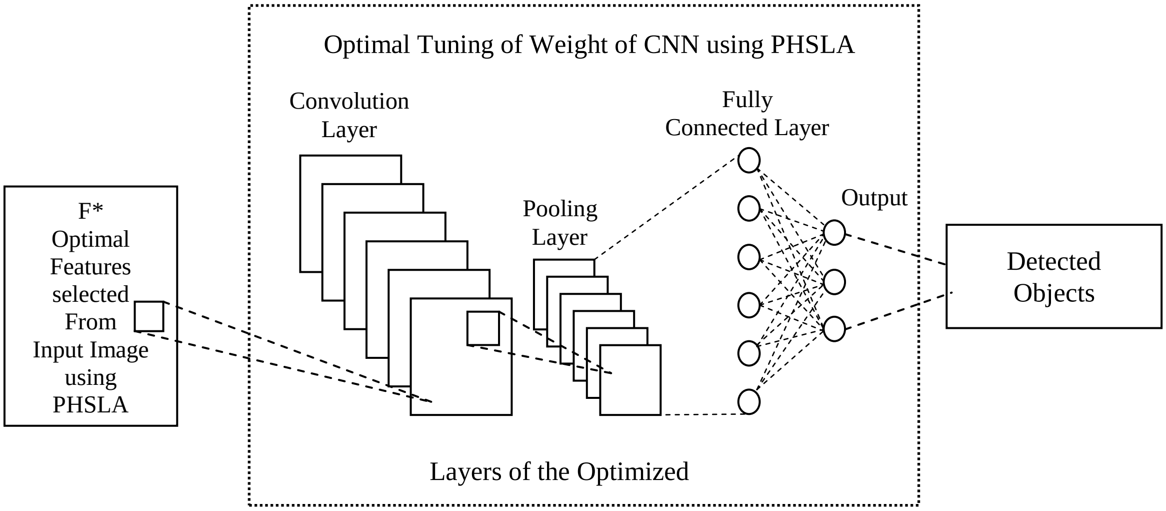

Optimal Tuning of Weight of CNN using PHSLA
Convolution Layer
Fully Connected Layer
Pooling Layer
Output
F* Optimal Features selected From Input Image using PHSLA
Detected Objects
Layers of the Optimized

Fig. 3.Schematic Diagram of Optimized CNN+ PHSLA

As discussed in the earlier sections, the extracted $F*$ are given as input to optimized CNN [35 - 37] for classifying the objects. Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that measure the weighted number of several inputs and emit an activation value, similar to their biological counterparts. Each neuron's action is determined by its weight. Therefore, we are fine-tuning the weights of Convolutional Neural Network using the proposed PHSLA model to enhance the detection accuracy. The convolutional layer is pooling layer which is fully connected layer and has convolutional kernels in place. These convolutional kernel defines the entire function map in such a way that rth layer of convolutional layer is mapped with zth feature map as well feature values in the location provided to the CNN and being determined by Eq. (21).

$$S_{e,x,z}^{r} = W_{z}^{r} F *_{e,x}^{r} + B_{z}^{r} \tag{21}$$

$$AF_{e,x,z}^{r} = AF\left(S_{e,x,z}^{r}\right) \tag{22}$$

$$O_{e,x,z}^{r} = pool\left(AF_{e,x,z}^{r}\right), \forall\left(c,r\right) \in \Re_{e,x} \tag{23}$$

Where W_z^r and B_z^r are the optimum weight vector and bias, which provides the optimal tuning of the weights performed using the adopted hybrid optimization scheme. Similarly, $AF\left(\bullet\right)$ is activation function provides prediction of nonlinear features of multilayer networks and here it is used to achieve non-linearity and presented in Eq. 22, when processed provides activation value. The Shift variance in the pooling layer is handled by Eq. (23) and it deals with decreasing the resolution of induced feature map by local neighbourhood and presented by $pool\left(\ \right)$. The down sampling operations were also conducted by the pooling layer in CNN with the resultant collected from the convolutional layers. Additionally, maximum pooling and average pooling were also explored. The higher value was observed in the max-pooling; nevertheless, the average value was observed in the average pooling. Function loss be determined by using CNN as

$$Loss = \frac{1}{Num} \sum_{t=1}^{ms} G\left(\varsigma; V^{(t)}, OUT^{(t)}\right) \tag{24}$$

Where, (ς) is the constraints of CNN are associated with required input-output relations and to be operated in limits of $\left\{\left(U^{(t)},V^{(t)}\right); t\in[1,\cdots,IO]\right\}$ and furthermore, output of CNN, t^{th} input data, and the related target values are determined as $OUT^{(t)}$, $U^{(t)}$ and $V^{(t)}$, correspondingly.

The results obtained from the pooling layer are usually given as an entry to the completely fully - connected, and hence the inputs are associated with both layers. The fully connected layer in the work appears at the output of the CNN system. The output layer of CNN is the final layer, and it includes the *softmax* function for performing precise final detection of objects in the images (targets). CNN's loss function $(Loss)$ must be minimised in order to get the best result as per equation (25).

$$Obj = \min(Loss) \tag{25}$$

In order to have minimization of loss, the efforts were taken for fine-tuning the weight of CNN - W by the proposed PHSLA model. The obtained solution fed as input to PHSLA model with inputs of weight of CNN and extracted features, the method of solution encoding for PHSLA is as depicted in Fig.4.

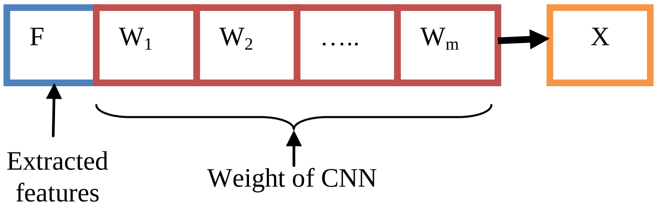


Fig. 4. Solution Encoding

5.2. PHSLA

The PSO [32] [34] is a Particle Swarm Optimization which is inspired by the activities populace of random solutions, and the quest for the best one is done by updating generations. The particles in this PSO are mainly engaged in a random location and are relocated in random definite directions inside the quest space... The selection of PSO has been good in solving complex optimization-related issues with higher accuracy and convergence rate. On the other hand, the SLnO algorithm is an iterative optimization technique based developing extremely sensitive features [33] that have higher memory and at the same time, they search for the best solutions within the search space without getting caught into local optima. Therefore, in the presented research work, these two algorithm were blended together to form Particle Hybridized Sea Lion Optimization Algorithm (PHSLA) to update the solutions with utmost precisions and higher convergence.

The steps followed in the PHSLA model is depicted below:

Step 1: The population P of search agents (sea lion and flock of birds) is initialized. In addition, the current iteration and the maximal iteration are denoted as itr , $\max^{itr}$, respectively.

Step 2: Initialize the N search agents with random velocities $V_i; i = 1,2,...N$ and position $X_i; i = 1,2,...N$

Step 3: Compute the fitness of the search agent using Eq. (25).

Step 5: Initialize $pbest$ with $X_i; i = 1,2,...N$

Step 6: Initialize $gbest$ with the best fitness search agent of $X_i; i = 1,2,...N$

Step 7: while $itr < \max^{itr}$ do

Step 8: update V_i^{itr} using Eq. (26).

$$V_i^{itr} = V_i^{itr} + 2*rand*(pbest - X_i^{itr}) + 2*rand*(gbest - X_i^{itr}) \tag{26}$$

Step 9: The proposed contribution is in this phase. In fact, the current position X_i^{itr} of the search agent has updated the standard PSO based positional update. Since the positional update with standard PSO sometimes gets traps into the local optima. Hence, we have updated the position of the search agent using a new update expression, which is based on SLnO's position update model. Thereby, the proposed algorithm has been named PHSLA. The proposed positional update of the search agent is given in Eq. (27).

$$X_i^{itr+1} = X_i^{itr} + ran_1 * (pbest - \left|X_i^{itr}\right|) + ran_2 * (gbest - \left|X_i^{itr}\right|) \tag{27}$$

Here, ran_1 and ran_2 are two random numbers within [0, 1].

Step 10: In addition, we have introduced arithmetic cross-over operation, which is the second contribution. Following the velocity and position updates, each iteration undergoes a breeding phase. A crossover operator is being used to choose two particles at random for breeding at a user-specified breeding probability. Among the population, we have selected two particles A, B for performing the crossover operation. The position corresponding to X_A^{itr} and X_B^{itr} is replaced by the offspring. This is mathematically shown in Eq. (28) and Eq. (29), respectively.

$$X_A^{itr+1} = rad^{itr} * X_A^{itr} + (1 - rad^{itr}).X_B^{itr} \tag{28}$$

$$X_B^{itr+1} = rad^{itr} * X_B^{itr} + (1 - rad^{itr}).X_A^{itr} \tag{29}$$

Here, rad is a random number generated within the range [0,1]. The velocities of these A, B particles are shown in Eq. (30) and Eq. (31), respectively.

$$V_A^{itr+1} = \frac{V_A^{itr} + V_B^{itr}}{\left\|V_A^{itr} + V_B^{itr}\right\|}\left\|V_A^{itr}\right\| \tag{30}$$

$$V_B^{itr+1} = \frac{V_A^{itr} + V_B^{itr}}{\left\|V_A^{itr} + V_B^{itr}\right\|}\left\|V_B^{itr}\right\| \tag{31}$$

Step 9: Compute the fitness of the search agent using Eq. (18). If the termination criterion is not met, increment the iteration itr by 1.

Step 10: Return the best solution

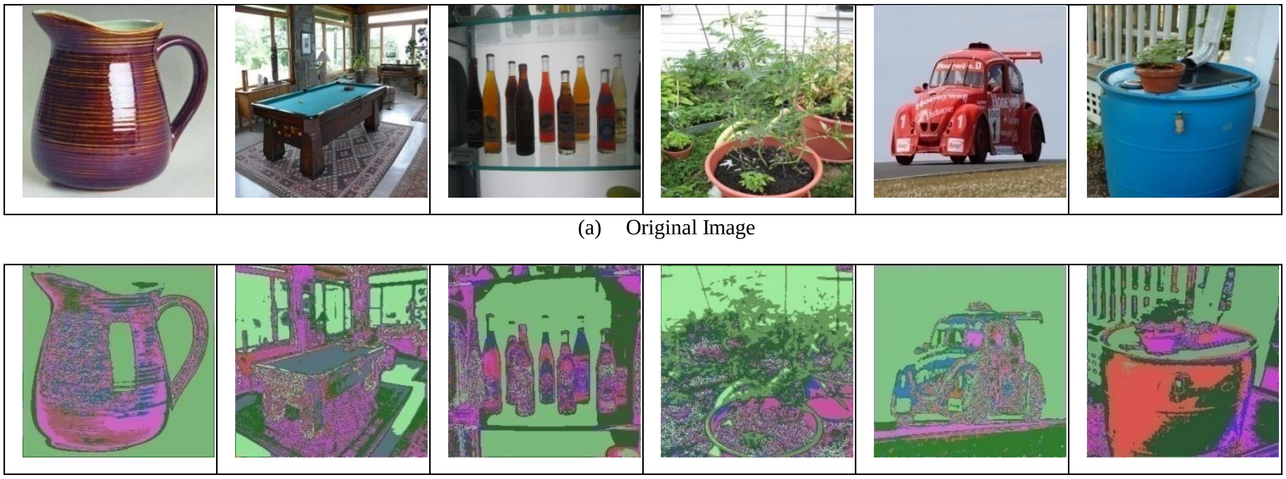
Step11: Terminates

Algorithm 1: Pseudo code of PHSLA

```
Step1: Initialize $P$ , $\max^{itr}$ , $itr$
Step2: Initialize the $N$ search agents with random velocities $V_i; i = 1,2,...N$ and position $X_i; i = 1,2,...N$
Step 3: Compute the fitness of the search agent using Eq. (25).
Step 4: Initialize $pbest$ with $X_i; i = 1,2,...N$
Step 5: Initialize $gbest$ with the best fitness search agent of $X_i; i = 1,2,...N$
Step 6: while $itr < \max^{itr}$ do
Update $V_i^{itr}$ using Eq. (26).
Step 7: Revise the place of the solutions by the proposed update equation as in Eq. (27)
Step 8: Perform arithmetic cross over operation
Step 9: Compute the fitness of the search agent using Eq. (25).
If the termination criterion is not met, increment the iteration $itr$ by 1.
Step 10: Return the best solution
End while
```

6. RESULTS AND DISCUSSION

The proposed model of object classification is primarily based on (CNN+PHSLA) and is evaluated over the existing models like CNN+PSO, CNN+WOA (Whale Optimization Algorithm), CNN+GWO (Grey Wolf Optimization), CNN+SLnO, CNN, and SVM (Support Vector Machine) in terms of *positive*, *negative* and *other measures*. This evaluation is carried out by varying the learning percentage from 60 (40% of data was used for training), 70 (30% of data was used for training), and 80 (20% of data was used for training), respectively. All the positive measures viz accuracy, specificity, sensitivity and precision needs to be persistent at the elevated levels for the most encouraging results. The FPR (False Positive Rate), FNR (False Negative Rate) and FDR (False Discovery Rate) needs to be as low as possible. The F1-score, MCC (Matthews's Correlation Coefficient), and NPV (Negative Predictive Value) are supplementary value-added factors that demonstrate the superiority of the planned work. The sample image collected for evaluation is shown in Fig.5 and the efficacy is described using convergence and performance analysis



(a) Original Image

(b) Corresponding Multi-kernel K-means Segmented Image

Fig. 5. (a) and (b) represent same class sample and its corrosponding segmented image

6.1. Convergence Analysis

In order to understand the performance of the proposed model, it is expected that, proposed model has higher convergence towards the defined objective function. The outcomes of the convergence analysis of the proposed as well as existing models for each iteration are shown in Fig.6. Initially, at the least iteration count (at 0th iteration), the convergence of the proposed, as well as existing models, is higher. Then, as the count of iteration gets increased gradually, the cost of the proposed, as well as existing models tends to be decreased. Further, at the highest iteration count (i.e. at 50th iterations), the PHSLA has achieved the least cost function (minimization of loss). At the 50th iteration it was observed that, the cost function of the PHSLA is 7.2%, 1.9%, 0.9%, and 2.1% which is found better than PSO, WOA, GWO, and SLnO, respectively. Thus, from the evaluation, it is seen that the PHSLA can lessen the CNN loss, which shows the supremacy of the model to increase the detection accuracy.

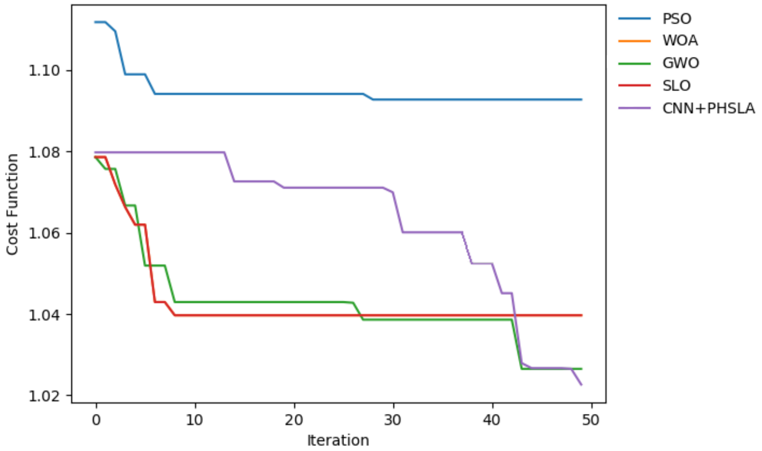


Fig. 6. Convergence Analysis

6.2. Performance Evaluation of CNN+PHSLA

As discussed earlier, proposed model's outcome (CNN+PHSLA) is assessed using *positive, negative*, and *other measures*. Fig. 7(i & ii) depicts the results obtained in terms of positive measures accuracy and precision. Based on the obtained results, the CNN+PHSLA is considered to be very beneficial, as it has archived the highest value for every variation in the learning percentage. The accuracy, which is the most important metric, reveals the superiority of the CNN+PHSLA. Remarkably, with any shift in the learning percentage, the accuracy of the planned work has been shown to be higher. Furthermore, the planned work had attained the best accuracy even at the highest learning percentage (i.e. 80 percent). The accuracy of the CNN+PHSLA (90.80%) at learning percentage=80 is 90.80%, which is better than the existing models like CNN+PSO=83.59%, CNN+WOA=83.81%, CNN+GWO=83.86%, CNN+SLnO=82.50%, CNN=83.70% and SVM=8.50%.

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	CNN	SVM	**CNN+PHSLA**
60	0.843409	0.844318	0.842045	0.829318	0.8405	0.853182	**0.896**
70	0.842121	0.833333	0.837273	0.838182	0.844	0.836061	**0.905333**
80	0.835909	0.838182	0.838636	0.825	0.837	0.825	**0.908**

Table 2.Performance Analysis of Proposed and Conventional Work in terms of Accuracy

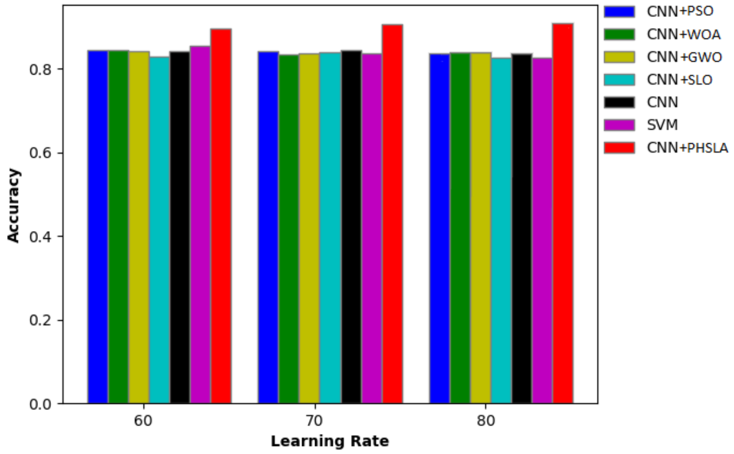


Fig. 7(i) Performance of Adopted Method Over Extant Models for Accuracy

In addition, the *precision*, *sensitivity* as well as specificity of the CNN+PHSLA are found to be higher for every variation in learning percentage. Among all these an interesting outcome has been recorded in case of precision. The precision of the proposed model is 79% at the highest learning percentage, which is indeed a favorable score. At 80th learning percentage, the precision of the work with fusion approach is found to be 9.75%, 11.00%, 11.25%, 3.75%, 18.50%,3.75% and 79% for CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN, SVM, and CNN+PHSLA respectively. The performance of CNN+PHSLA is observed to be outperforming than the existing works reported combinations of CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN and SVM. In addition, the sensitivity of the proposed work at 70th learning is found to be 77.66 where are other fusion performance are as 13.16%, 8.33%, 10.50%, 11%, 22%, 9.83%, for CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN, and SVM, respectively. Thus, from the evaluation, it's clear that the proposed work had archived maximal values in terms of positive performance measures, and this is said to be the most favorable outcome.

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLO	CNN	SVM	**CNN+PHSLA**
60	0.13875	0.14375	0.13125	0.06125	0.2025	0.1925	**0.73**
70	0.131667	0.083333	0.105	0.11	0.22	0.098333	**0.776667**
80	0.0975	0.11	0.1125	0.0375	0.185	0.0375	**0.79**

Table 3.Performance Analysis of Proposed and Conventional Work in terms of Precision

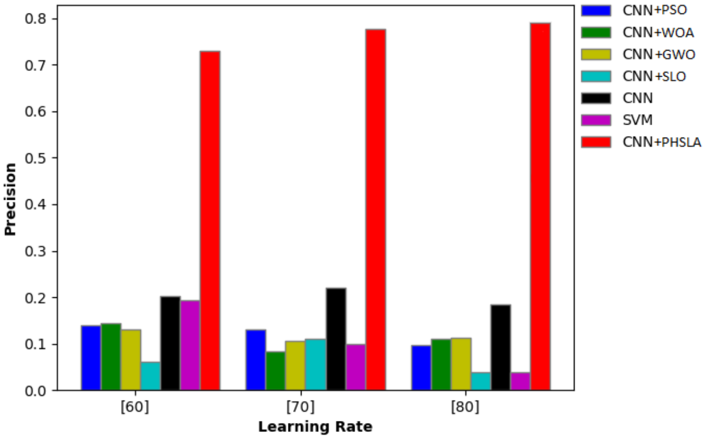


Fig. 7(ii) Performance of Adopted Method Over Extant Models for Precision

6.3 Overall Performance Analysis

The accuracy of the CNN+PHSLA at 80th learning percentage is 0.908, which's the highest value when compared to existing works like CNN+PSO=0.835909, CNN+WOA=0.838182, CNN+GWO=0.838636, CNN+SLnO=0.825, CNN=0.837 and SVM=0.825. In addition, the precision of the CNN+PHSLA at 80th learning percentage is0.79, which is 87.7%, 86%, 85.7%, 95.2%, 76.5%, and 95.23% better than the existing works like CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN and SVM, respectively. In addition, the CNN+PHSLA has archived the least error measures. On the other hand, the FNR of the CNN+PHSLA is 0.27 (least value), which is 68.7%, 68.4%, 68.9%, 71.23%, 66.14%, and 66.5% better than the existing works like CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN and SVM, respectively. Moreover, F1-score, MCC, and NPV are found to be higher with the CNN+PHSLA for every variation in the learning percentage. In addition, the rand index of the CNN+PHSLA at 80th learning percentage is 0.978775, which is 6.6%, 6.4%, 6.437%, 7.2%, 6.52%, and 7.20% better than the existing works like CNN+PSO, CNN+WOA, CNN+GWO, CNN+SLnO, CNN and SVM, respectively. From the overall evaluation, it is clear that the CNN+PHSLA had achieved the optimal values, thereby the CNN+PHSLA had become much sufficient for detecting the objects.Table below summerizes the performance analysis of CNN+PHSLA at the 80th learning percentage for the measures accuracy and precision.

Method	Accuracy	Precision
CNN+PSO	0.835909	0.0975
CNN+WOA	0.838182	0.11
CNN+GWO	0.838636	0.1125
CNN+SLO	0.825	0.0375
CNN	0.837	0.185
SVM	0.825	0.0375
CNN+PHSLA	0.908	0.79

Table 4.Performance analysis of CNN+PHSLA for Accuracy and Precision

8. CONCLUSION AND FUTURE SCOPE

In this research work, the proposed model will act as black box for the system developed on top for assisting visually challenged. The proposed work meticulously consider the development of system in major phases including pre-processing, segmentation, feature extraction, optimal feature selection and object detection. . The optimal features were mined from the extracted features using a new hybrid algorithm (PHSLA). These optimal features were fed as the input to Optimized CNN for object detection mechanism. In order to make the object detection more precise, the weights of CNN are optimally tuned using a new hybrid model. The proposed hybrid model referred to as PHSLA, which combines the power of both PSO and SLnO, respectively. The accuracy of the CNN+PHSLA (90.80%) at learning percentage80and which is better than the existing models like CNN+PSO (83.59%), CNN+WOA(83.81%), CNN+GWO(83.86%), CNN+SLnO(82.50%), CNN(83370%) and SVM(82.50%). From the overall evaluation, it is clear that the CNN+PHSLA had achieved the optimal values; thereby the CNN+PHSLA had become much sufficient for detecting the object.The current research work has emphasized assisting visually impaired people by helping them identify the object in their environment. The work will be extended towards detection of multiple objects in unconstrained environment or in the vicinity of the stationary viewer and giving them aural notifications regarding the multiple objects [42].

9. REFERENCES

1) Meshram, V. V., Patil, K., Meshram, V. A., &Shu, F. C. (2019). An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People. *IEEE Transactions on Human-Machine Systems*, *49*(5), 449–460. https://doi.org/10.1109/THMS.2019.2931745

2) SeiffertSimões, W. C. S., & de Lucena, V. F. (2016). Indoor Navigation Assistant for Visually Impaired by Pedestrian Dead Reckoning and Position Estimative of Correction for Patterns Recognition. *IFAC-PapersOnLine*, *49*(30), 167–170. https://doi.org/https://doi.org/10.1016/j.ifacol.2016.11.149

3) Khenkar, S., Alsulaiman, H., Ismail, S., Fairaq, A., Jarraya, S. K., & Ben-Abdallah, H. (2016). ENVISION: Assisted Navigation of Visually Impaired Smartphone Users. *Procedia Computer Science*, *100*, 128–135. https://doi.org/https://doi.org/10.1016/j.procs.2016.09.132

4) Siddhartha, B., Chavan, A. P., & Uma, B. V. (2018). An Electronic Smart Jacket for the Navigation of Visually Impaired Society. *Materials Today: Proceedings*, *5*(4, Part 3), 10665–10669. https://doi.org/https://doi.org/10.1016/j.matpr.2017.12.344

5) Connier, J., Zhou, H., Vaulx, C. De, Li, J., Shi, H., Vaslin, P., &Hou, K. M. (2020). Perception Assistance for the Visually Impaired Through Smart Objects: Concept, Implementation, and Experiment Scenario. *IEEE Access*, *8*, 46931–46945. https://doi.org/10.1109/ACCESS.2020.2976543

6) Garcia-Macias, J. A., Ramos, A. G., Hasimoto-Beltran, R., &Pomares Hernandez, S. E. (2019). Uasisi: a modular and adaptable wearable system to assist the visually impaired. *Procedia Computer Science*, *151*, 425–430. https://doi.org/https://doi.org/10.1016/j.procs.2019.04.058
7) Dourado, A. M. B., &Pedrino, E. C. (2020). Multi-objective Cartesian Genetic Programming optimization of morphological filters in navigation systems for Visually Impaired People. *Applied Soft Computing*, *89*, 106130. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106130
8) Gharani, P., &Karimi, H. (2017). Context-aware obstacle detection for navigation by visually impaired. *Image and Vision Computing*, *64*. https://doi.org/10.1016/j.imavis.2017.06.002
9) Zhu, J., Hu, J., Zhang, M., Chen, Y., & Bi, S. (2020). A fog computing model for implementing motion guide to visually impaired. *Simulation Modelling Practice and Theory*, *101*, 102015. https://doi.org/https://doi.org/10.1016/j.simpat.2019.102015
10) Cordeiro, N., &Pedrino, E. (2019). A new methodology applied to dynamic object detection and tracking systems for visually impaired people. *Computers & Electrical Engineering*, *77*, 61–71. https://doi.org/10.1016/j.compeleceng.2019.05.003
11) Chen, X., Xu, J., & Yu, Z. (2019). A 68-mw 2.2 Top s/w Low Bit Width and Multiplierless DCNN Object Detection Processor for Visually Impaired People. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(11), 3444–3453. https://doi.org/10.1109/TCSVT.2018.2883087
12) Cordeiro, N. H., &Pedrino, E. C. (2019). Collision risk prediction for visually impaired people using high level information fusion. *Engineering Applications of Artificial Intelligence*, *81*, 180–192. https://doi.org/https://doi.org/10.1016/j.engappai.2019.02.016
13) Jimenez, M., Mello, R., Freire, T., &Frizera, A. (2020). Assistive Locomotion Device with Haptic Feedback For Guiding Visually Impaired People. *Medical Engineering & Physics*, *80*. https://doi.org/10.1016/j.medengphy.2020.04.002
14) Arora, A., Grover, A., Chugh, R., &Reka, S. S. (2019). Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector. *Wireless Personal Communications*, *107*(1), 651–661. https://doi.org/10.1007/s11277-019-06294-1
15) Afif, M., Ayachi, R., Said, Y., Pissaloux, E., &Atri, M. (2020). An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation. *Neural Processing Letters*, *51*(3), 2265–2279. https://doi.org/10.1007/s11063-020-10197-9
16) Cardillo, E., Di Mattia, V., Manfredi, G., Russo, P., De Leo, A., Caddemi, A., &Cerri, G. (2018). An Electromagnetic Sensor Prototype to Assist Visually Impaired and Blind People in Autonomous Walking. *IEEE Sensors Journal*, *18*(6), 2568–2576. https://doi.org/10.1109/JSEN.2018.2795046
17) Ye, C., &Qian, X. (2018). 3-D Object Recognition of a Robotic Navigation Aid for the Visually Impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(2), 441–450. https://doi.org/10.1109/TNSRE.2017.2748419
18) Aakash Krishna, G. S., Pon, V. N., Rai, S., &Baskar, A. (2020). Vision System with 3D Audio Feedback to assist Navigation for Visually Impaired. *Procedia Computer Science*, *167*, 235–243. https://doi.org/10.1016/j.procs.2020.03.216
19) Chan, K. Y., Engelke, U., &Abhayasinghe, N. (2017). An edge detection framework conjoining with IMU data for assisting indoor navigation of visually impaired persons. *Expert Systems with Applications*, *67*, 272–284. https://doi.org/10.1016/j.eswa.2016.09.007
20) Jindal, A., Aggarwal, N., & Gupta, S. (2018). An Obstacle Detection Method for Visually Impaired Persons by Ground Plane Removal Using Speeded-Up Robust Features and Gray Level Co-Occurrence Matrix. *Pattern Recognition and Image Analysis*, *28*(2), 288–300. https://doi.org/10.1134/S1054661818020086
21) Mehta, U., Alim, M., & Kumar, S. (2017). Smart Path Guidance Mobile Aid for Visually Disabled Persons. *Procedia Computer Science*, *105*, 52–56. https://doi.org/https://doi.org/10.1016/j.procs.2017.01.190
22) Guimares, C., Henriques, R., & Pereira, C. (2016). Tracking System Proposal of Walking Sticks Aiming the Orientation and Mobility of the Visually Impaired. *IFAC-PapersOnLine*, *49*. https://doi.org/10.1016/j.ifacol.2016.11.147
23) Bauer, Z., Dominguez, A., Cruz, E., Gomez-Donoso, F., Orts-Escolano, S., &Cazorla, M. (2020). Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors. *Pattern Recognition Letters*, *137*, 27–36. https://doi.org/https://doi.org/10.1016/j.patrec.2019.03.008
24) Manjari, K., Verma, M., &Singal, G. (2020). A survey on Assistive Technology for visually impaired. *Internet of Things*, *11*, 100188. https://doi.org/https://doi.org/10.1016/j.iot.2020.100188
25) Beno, M., R, V., M, S., &Rajakumar, B. (2014). Threshold Prediction for Segmenting Tumour from Brain MRI Scans. *International Journal of Imaging Systems and Technology*, *24*. https://doi.org/10.1002/ima.22087
26) Canny edge detection, from:" https://docs.opencv.org/master/da/d22/ tutorial_py_canny.html", Access Date: 2021-0-17
27) Chandrakala, M., &Durga Devi, P. (2021). Two-stage classifier for face recognition using HOG features. *Materials Today: Proceedings*, *47*, 5771–5775. https://doi.org/https://doi.org/10.1016/j.matpr.2021.04.114
28) Swamy, S. M., Rajakumar, B. R., &Valarmathi, I. R. (2013). Design of hybrid wind and photovoltaic power system using opposition-based genetic algorithm with Cauchy mutation. *IET Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2013)*, 504–510. https://doi.org/10.1049/ic.2013.0361

29) Brammya, G., Praveena, S., NinuPreetha, N. S., Ramya, R., Rajakumar, B. R., &Binu, D. (2019). Deer Hunting Optimization Algorithm: A New Nature-Inspired Meta-heuristic Paradigm. *The Computer Journal*. https://doi.org/10.1093/comjnl/bxy133

30) Rajakumar, B. R., & George, A. (2012). A new adaptive mutation technique for genetic algorithm. *2012 IEEE International Conference on Computational Intelligence and Computing Research*, 1–7. https://doi.org/10.1109/ICCIC.2012.6510293

31) Tanweer, M. R., Suresh, S., &Sundararajan, N. (2015). Self regulating particle swarm optimization algorithm. *Information Sciences*, *294*, 182–202. https://doi.org/https://doi.org/10.1016/j.ins.2014.09.053

32) Masadeh, R., Mahafzah, B., &Sharieh, A. (2019). Sea Lion Optimization Algorithm. *International Journal of Advanced Computer Science and Applications*, *10*, 388–395. https://doi.org/10.14569/IJACSA.2019.0100548

33) Rewadkar, D., &Doye, D. (2017). FGWSO-TAR: Fractional glowworm swarm optimization for traffic aware routing in urban VANET. *International Journal of Communication Systems*, *31*, e3430. https://doi.org/10.1002/dac.3430

34) DarekarRavirajVishwambhar, D. A. P. (2019). Emotion Recognition from Speech Signals Using DCNN with Hybrid GA-GWO Algorithm. *Multimedia Research*, *2*(4), 12–22. https://doi.org/10.46253/j.mr.v2i4.a2

35) Sammulal, M. G. & K. M. C. &. (2019). Enhanced Crow Search Optimization Algorithm and Hybrid NN-CNN Classifiers for Classification of Land Cover Images. *Multimedia Research*, *2*(3), 12–22. https://doi.org/10.46253/j.mr.v2i3.a2

36) G.Gokulkumari. (2020). Classification of Brain Tumor using Manta Ray Foraging Optimization-based DeepCNN Classifier. *Multimedia Research*, *3*, 32–42. https://doi.org/10.46253/j.mr.v3i4.a4

37) Li, F., Lv, X.-G., & Deng, Z. (2018). Regularized iterative Weiner filter method for blind image deconvolution. *Journal of Computational and Applied Mathematics*, *336*, 425–438. https://doi.org/https://doi.org/10.1016/j.cam.2017.12.026

38) SURF feature, from : “https://medium.com/data-breach/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e “, Access Date: 2021-0-17

39) SIFT faeture, from :”https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40”, Access Date: 2021-0-17

40) Dataset link: https://www.kaggle.com/c/imagenet-object-localization-challenge/data?select= imagenet_object_ localization_patched2019.tar.gz

41) Suraj Pardeshi, Pravin Yannwar, Ramesh Manza, Bharti Gawali, & Filbert Hilman Juwono. (2022). Particle Hybridized Sealion Optimization Algorithm (PHSLA) - A Novel Object Detection Method. 17(08), 1154–1175. https://doi.org/10.5281/zenodo.7005884

42) Pardeshi S.R., Pawar, V.J., Kharat, K.D., Chavan, S. (2021). Assistive Technologies for Visually Impaired Persons Using Image Processing Techniques – A Survey. In: Santosh, K.C., Gawali, B. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2020. Communications in Computer and Information Science, vol 1380. Springer, Singapore. https://doi.org/10.1007/978-981-16-0507-9_9