A Novel Approach for Object Detection Using Optimized Convolutional Neural Network to Assist Visually Impaired People

Suraj Pardeshi[1](✉), Nikhil Wagh[1], Kailash Kharat[1], Vikul Pawar[1], and Pravin Yannawar[2]

[1] Department of MCA, Government College of Engineering, Aurangabad, Aurangabad, India
surajrp@geca.ac.in

[2] Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Aurangabad, India

Abstract. Human race is blessed with the five basic senses such as touch, taste, smell, hearing and the most important of them all 'vision or eyesight'. It is very difficult to survive without any one of them. Unfortunately a mass population across the globe suffers from the ill effects of vision, hampering their daily life. Detecting objects and providing navigational instructions in an indoor environment can considerably improve the day-to-day quality of life of visually impaired people. The motive of this research work is to propose a solution approach for assisting visually impaired population by identifying obstacles in front of them considering indoor environment. This approach focuses on feature extraction and object detection using Convolutional Neural Network (CNN) from a real time video. For this a head mounted image acquisition device may be used to detect the objects from the scene ahead and information of the detected objects is provided to the visually impaired (VI) person through the audio modality. As a first step towards the overall conceptual process, an object detection system is presented in this article, which processes the live video stream captured through the acquisition device. The video is processed frame-by-frame, treating each frame as a separate image and then using the proposed feature extraction and object detection algorithm to identify the objects.

Keywords: Vision · Convolutional Neural Network (CNN) · Visually Impaired (VI) · Feature Extraction · Object Detection · Audio Modality

List of Abbreviations Used

1) Activation Function (AF)
2) Convolutional Neural Network (CNN)
3) Discrete Fourier Transform (DFT)
4) False Discovery Rate (FDR)
5) False Negative Rate (FPR)
6) False Positive Rate (FPR)
7) Gray-Level Co-occurrence Matrix (GLCM)
8) Grey Wolf Optimization (GWO)

R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 187–207, 2023.
https://doi.org/10.2991/978-94-6463-196-8_17

9) High-Level Features (HLFs)
10) Histogram of Oriented Gradients (HOG)
11) Inertial Measurement Unit (IMU)
12) Input Image (I_{input})
13) Interplane Relationships (IPRs)
14) Inverse Discrete Fourier Transform (IDFT)
15) Local Binary Pattern (LBP)
16) Matthews's correlation coefficient (MCC)
17) Mean-Square Error (MSE)
18) Modified Sigmoid Function (MSF)
19) Negative Predictive Value (NPV)
20) Object Detection Model for Visually Impaired (ODMVI)
21) Particle Swarm Optimization (PSO)
22) Peak Signal-to-Noise Ratio (PSNR)
23) Pre-processed Image (I_{prep})
24) Region of Interest (ROI)
25) Scale-Invariant Feature Transform (SIFT)
26) Sea Lion Optimization Algorithm (SLnO)
27) Segmented Image (I_{segm})
28) Speeded-Up Robust Features (SURF)
29) Visually Impaired (VI)
30) Whale Optimization Algorithm (WOA)

1 Introduction

Having a clear vision is a valuable blessing, and it is one of the most important faculties that allow us to gain knowledge from our surroundings. Sadly, vision loss is becoming more and more common. According to the WHO report, there are 314 million people in the world who suffers from visual disabilities. Uncorrected refractive errors or eye conditions are often cited as the main causes of visual impairment. There are 314 million people who are classified as outwardly disabled, 45 million are visually impaired (blind) [1–4]. A maturing population has made visual impairments more prevalent and widespread. As people age, the danger of visual impairments increases, further complicating the challenge of independent mobility. Identifying obstacles without vision is a challenge. The VI person relies on the other senses mostly touch for movements. But in some cases the physical contact between a person and unknown object can be dangerous. A comprehensive framework is needed to help people with visual disabilities presenting with impairments or impedances [5–7]. A simple guide cane is the traditional tool mostly used by VI. The advancements in technology added considerable modifications in the cane such as sensors, vibrating pads etc. The limitation of using cane is distance. Unless the object comes in the perimeter of the cane it is not identified. A computerized system that recognizes objects without touching them and provides auditory feedback to the person giving more accurate understanding of the environment and avoiding threats involved in the overall process is needed [8–12]. Also the need of VI is not only to identify the obstacles but to gain a better understanding of the surroundings. The rapidly increasing research interests in the faculties of computer science such as image processing, machine learning, computer vision and AI is a ray of hope to overcome the problems mentioned so far [13]. Machine vision systems aims to recognize all the objects in an image and collect information about the categories and positions of those objects, so they can understand what the image shows. A number of methods were invented to address these types of situations, mostly based on the principles of computer vision and deep learning. However the overlapping objects and various lighting conditions become another hurdles to overcome [14–16].

2 Related Work

Mehta et al. [17] proposed a cost-effective mobile phone-based device solution which is both cost-effective and noise-resistant. Furthermore, features such as “Local Binary Pattern (LBP), Gabor, and Histogram-based features”, among others, can be used to differentiate between different types of obstacles present, such as a chair, vehicle, or human, to improve the efficiency of VI. The optimization algorithms [18–23] are playing a major role in the object detection approach.

Cardillo *et al.* [24] have projected a novel autonomous walking aid for the visually impaired and blind users with the aid of the electromagnetic sensor. The introduction of microwave radar to the conventional white cane aided target identification for the visually impaired while walking. Further, while compared to the state-of-art Electronic Travel Aids, the presented work had consumed fewer dimensions and had better noise tolerance with utmost better performance.

Ye and Qian [25] have implemented a “3-D object recognition method” onto the robotic navigation with the intention of assisting the blind person in indoor structural detection. The researchers have broken a point cloud into a plethora of “planar patches” and extracted the “Interplane Relationships (IPRs)” The authors have specified six “High-Level Features (HLFs)” for each of the patches based on the object’s IPRs. Then, using a “Gaussian-mixture-model-based plane classifier”, each planar patch belonging to an individual object model was categorized. At last, the classified planes were clustered into model objects using a recursive plane clustering technique. As a consequence, this method was well suited for detecting non-structural structures indoors in a precise way.

Chan *et al.* [26] have developed a new MSF framework on the basis of the “Inertial Measurement Unit (IMU)” for the migration of visually impaired persons within indoor environments. This technique was simpler since it had adopted the Modified Sigmoid Function (MSF) in estimating the blur levels of IMU. Further, the edge detections were made smoother with a moving camera in the MSF topological structure. The authors have evaluated the performance of the MSF framework by means of evaluating the object edges on “video sequences associated with IMU data”.

Jindal *et al.* [27] have designed a novel smartphone-based cost-effective system for safe walking along the roads by observing the obstacles along the paths of the visually impaired people in real-time scenarios. The video was captured with the aid of the Monocular vision approach and they have extracted the frames from the video by the meaning of neglecting the blurriness that occurred in the image due to the camera motion. Further, they have extracted the Speeded-Up Robust Features (SURF) after removing the ground plane of the non-ground area. The SURF features matched with the features of the obstacles were segmented with the aid of the active contour model from the non-ground image and these images were referred to as the region of interest (ROI). Moreover, they have verified whether the ROI belongs to an obstacle or not by means of passing the calculated Gray-Level Co-occurrence Matrix (GLCM) features onto the classification model.

Arora *et al.* [28] have developed a new prototype for "real-time multi-object detection" with the aid of the "image segmentation and deep neural network". With the proposed approach, the authors have prompted the blind persons about the entity, its location with reverence to the individual via speech stimulus. Furthermore, the authors have combined the "single-shot multibox detection system" with the "mobileNet architecture" for application construction that's also lightweight, scalable, and has a short response time. As a whole, the proposed approach has performed well in terms of accuracy and latency.

Meshram *et al.* [29] have introduced a new electronic assistive device referred to as the "NavCan" for obstacle-free paths to visually impaired people in both indoor and outdoor settings. The proposed approach provided priority information with no information overloading about the obstacles in the path. In addition, the proposed NavCane approach had also guided the users to recognize the objects in the indoor settings. The NavCane seems to be an effective tool for detecting "snags, ascending and descending flights of stairs, navigating wet floors, and object recognition in both recognized and unknown environments", according to the trial results. Similarly, when compared to a white stick, their evaluation findings show that the NavCane enhances the appearance of a snag-free pathway.

Afif *et al.* [30] have developed a novel indoor object detection system for VI people on the basis of the deep CNN "RetinaNet". The proposed model was significant in localizing as well as categorizing the indoor objects from the collected input image. This approach had gained high detection performances even under most of the challenging conditions like "extreme illumination changes, occlusion, and high inter-class and intra-class variation".

Krishna *et al.* [31] have presented a new "vision system with 3D audio feedback mechanism" to guide the VI people during their navigation. During their movement, the intuitive cognize corresponding to the localization of the object along their path was notified to the blind people via the variation in the intensity of the sounds of earphones. The three main modules of the proposed system were:" depth calculation, object detection, and 3D audio generation. The stereoscopic vision was utilized for discovering the depth map and the localization of doors indoors was detected using CNN. Further, on the basis of the object's location and depth map, the audio vector generates the 3D audio, which guides their locomotion. When tested in a real-life setting, this device proved to be reliable and effective for visually impaired navigation.

3 Architectural Description of the Proposed Object Detection Model for Visually Impaired (ODMVI)

Object Detection Model for Visually Impaired persons aims to provide reliable, fast, and accurate object recognition in real time to assist visually impaired and challenged persons. With this research work, we discuss the black-box functionality of the proposed

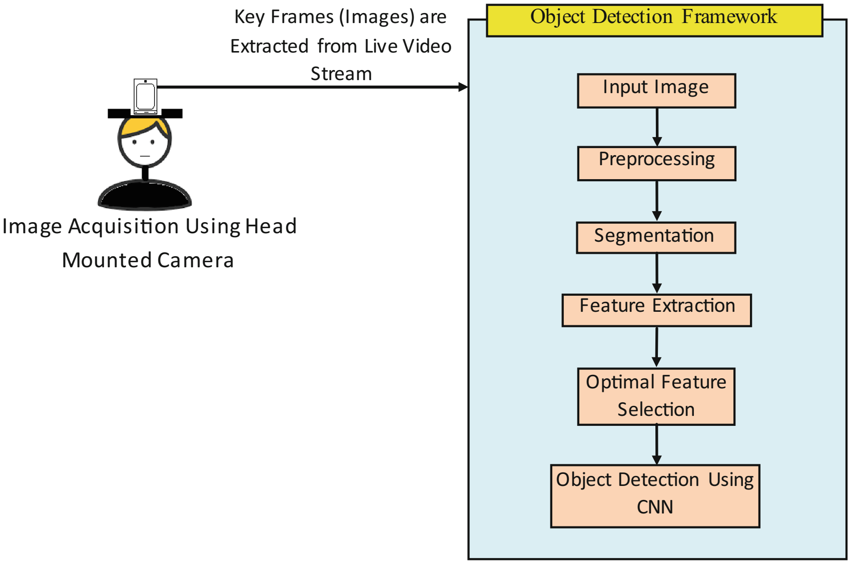
Key Frames (Images) are Extracted from Live Video Stream
Image Acquisition Using Head Mounted Camera
Object Detection Framework
Input Image
Preprocessing
Segmentation
Feature Extraction
Optimal Feature Selection
Object Detection Using CNN

Fig. 1. Architecture of the Proposed Object Detection Model-‘ODMVI’

system for handling the complexity of object detection mechanism. The model considers an indoor environment, where objects are stationary.

In the proposed model, a head mounted scene acquisition device captures a stationary indoor scene and provides data (from the scene) to the connected system for the purpose of real-time object detection. We are capturing images from a live video. Checkpoints are added at specific time intervals in the video stream to get the key frames. The key frames are nothing but the two dimensional images that are provided as input to the object detection framework Depending on the specifications of the acquisition device capturing the video, the input image sizes may vary. Hence the images obtained are converted to 255 × 255 dimensions for uniformity and given to the object detection framework.

The object detection framework processes the input image through five distinct phases: “pre-processing, segmentation, feature extraction, optimal feature selection, and object detection”. Figure 1 illustrates the architecture of the proposed object recognition paradigm designed to accommodate the needs of visually impaired individuals.

3.1 Preprocessing

The purpose of image pre-processing is to improve the image data by suppressing unwanted distortions and enhances some important features of the image for further processing. In this research work, wiener filtering is applied to the input image I_{input} for removing unwanted noises. This enhances the quality of the image. The Weiner filter has the highest Peak Signal-to-Noise Ratio (PSNR) (in dB) thus having the lowest

Mean-Square Error (MSE) (in dB), as compared to other compatible ones. Furthermore, the Weiner filter performs the optimal noise smoothing and inverse filtering tradeoffs.

Weiner filtering: The Wiener filter is the most important method for separating blurred areas from input image frames (I_{input}). This distinguishes the interesting points (objects) from the rest of the scene [32]. The original image *orig(x,y)* and the noisy image *nois(x,y)* are all in the frame $I_{input}(x,y)$. The image has been degraded according to Eq. (1). The main goal of using the Weiner filter is to get the restored image *r(x,y)* from $I_{input}(x,y)$, with the stipulation that *r(x,y)* must be equal to $I_{input}(x,y)$. The pixels' positions are indicated by the notations *(x,y)*. The Weiner filler is mathematically expressed in the frequency domain as in Eq. (2). The power spectra of the *orig(x,y)* and *nois(x,y)* are denoted by $\mathrm{S_D}$(x,y) and $\mathrm{S_N}$(x,y) respectively. The solution can be found by lowering K as in Eq. (3). The discrete Fourier transforms (DFTs) of the original image and noise are represented by $O_{DFT}(x,y)$ and $N_{DFT}(x,y)$ respectively. Equation (4) is used to obtain the solution.

$$I_{input}(x, y) = orig(x, y) + nois(x, y) \tag{1}$$

The standard mathematical equation for Weiner filter is denoted by Eq. 2.

$$G(x, y) = \frac{S_D(x, y)}{S_D(x, y) + S_N(x, y)} \tag{2}$$

where: $S_N(x, y)$ is the power spectrum of noise $nois(x, y)$ and $S_D(x, y)$ is the power spectrum of the original image $orig(x, y)$. For simplification we need to take derivative of Eq. (2) reduced to Eq. (3) by computing DFT of original Image *orig(x,y)* and DFT of noisy image *nois(x,y).*

$$K = F\left[|ODFT((x, y)) - G((x, y)).NDFT((x, y))|^2\right] \tag{3}$$

$$G(x, y) = \frac{F\left[NDFT((x, y)).ODFT * ((x, y))\right]}{F\left[|NDFT((x, y))|^2\right]} \tag{4}$$

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 Eqs. (7) and (8), respectively. The inverse transform of DFT is IDFT (Inverse Discrete Fourier Transform). The pre-processed image I_{prep} is subjected to segmentation.

$$F\left[NDFT(x, y).ODFT * (x, y)\right] = \begin{cases} = F\left\{\left[ODFT(x, y)\right] + NDFT(x, y)\right\} \times ODFT * (x, y) \\ = F\left[|ODFT(x, y)|^2\right] \\ = S_D(x, y) \end{cases} \tag{5}$$

$$F\left[|ODFT((x, y))|^2\right] = S_D((x, y)) + S_N((x, y)) \tag{6}$$

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 $Z(x, y)$.

$$Z(x, y) = G(x, y).I_{input}(x, y) \tag{7}$$

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

$$z(x, y) = IDFT\big[Z(x, y)\big] \tag{8}$$

Subsequently, the resultant image of preprocessing using wiener filtering is to be passed for segmentation.

3.2 Segmentation

The process of segmenting a digital image into several distinct regions comprising each pixel (sets of pixels, also known as superpixels) with identical attributes is known as image segmentation. Objects and boundaries (lines, curves, etc.) in images are usually located using image segmentation. Image segmentation is also the process of assigning a label to each pixel in an image such that pixels with the same label share common values.

The 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 ROI and Non-ROI regions. The resolution of the image is I_{prep}(x,y), which is clustered into k- the count of clusters. Let I_{prep}(x,y) be the input pixel that is to be clustered and the cluster centre is C_k the steps followed in the k-means algorithm is depicted below:

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

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

$$E_{Dist} = \left|\left|I_{prep}(x, y) - \mathrm{C_K}\right|\right| \tag{9}$$

Step 3: Depending on the distance $\mathrm{E_{Dist}}$, 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.

$$C_k = \frac{1}{k}\sum_{x \in c_k}\sum_{y \in c_k} Iprep(x, y) \tag{10}$$

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

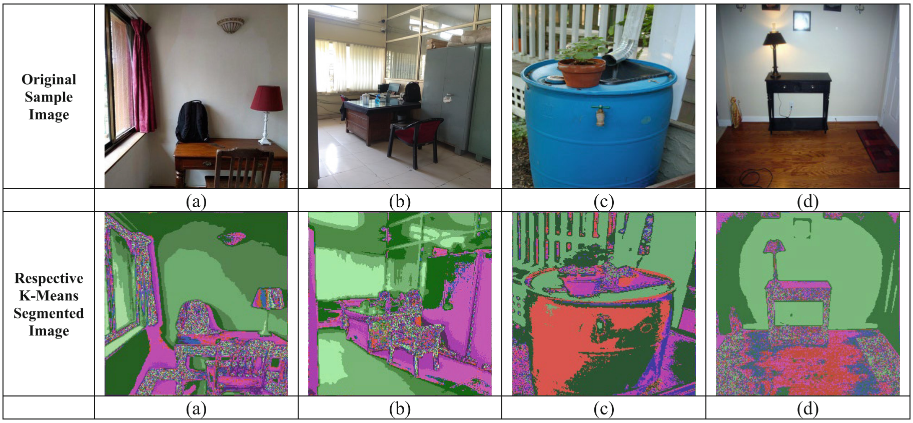

Fig. 2. Sample and Segmented Images

result would be different for different initial centres. 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 Eqs. (11) and Eq. (12), respectively.

$$\text{Sigmoid kernel}: Sk_1(x, y) = \tanh(\beta_0 \langle x, y \rangle + \beta_1) \tag{11}$$

$$\text{Laplacian kernel}: Lk_2(x,y) = \exp\left(-\frac{\|x - y\|}{\sigma}\right) \tag{12}$$

$$Mk(x, y) = \frac{Sk_1 + Lk_2}{3} \tag{13}$$

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

The sample and segmented images used for evaluation are shown in following Fig. 2.

3.3 Feature Extraction

The segmented image is denoted by I_{segm} from which multiple features (SURF, Scale-Invariant Feature Transform (SIFT), Shape features via Canny edge and gradient features via Histogram of Oriented Gradients (HOG)) are extracted as shown in Fig. 3.

3.3.1 SURF

The SURF feature is extracted from I_{segm}. The SURF method is indeed a stable and accurate technique towards describing as well as contrasting images in a local, similarity invariant manner. The prominent feature of the SURF technique has been its ability

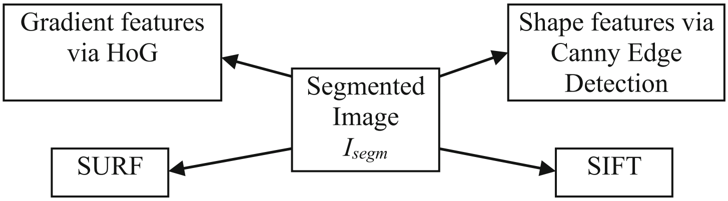
Gradient features via HoG
Shape features via Canny Edge Detection
Segmented Image I_{segm}
SURF
SIFT

Fig. 3. Multiple Feature Extraction from the Segmented Image

to compute operators quickly using box filters, thereby supporting real-time applications including surveillance and object recognition. The "Feature Description as well as Feature Extraction" are the two major steps in the SURF model [33].

In the feature extraction phase, the "Hessian matrix approximation" has been used for detecting the interest points.

The "SURF descriptor" is generated in two steps: The very first step is to establish a repeatable orientation using data from a circular area surrounding the keypoint. The SURF descriptor is then extracted from a square region aligned to the chosen orientation.

Descriptor Components: (a) The square area creation is the first step, and here the square area is created in the form, which is centered on the keypoint and aligned in the direction of the orientation. (b) The region is then divided into smaller 4 * 4 square-shaped sub-regions on a constant basis. At 5 * 5 regularly spaced sampling points, we compute a few basic features for each sub-region. (c) In the horizontal direction, The Haar wavelet responses referred to as dx, while in the vertical direction it is referred to as *dy*. In order to enhance the robustness against geometric deformations and localization errors, the responses *dx* and *dy* were weighted initially with a Gaussian (σ = 3.3 s). The extracted SURF feature is denoted as 'f_{SURF}'.

3.3.2 SIFT

The SIFT feature is extracted from I_{segm}. In fact, the SIFT is indeed a simple procedure. The SIFT algorithm consists primarily of four stages:

(a) Selection of a scale-space peak: A possible spot for locating features is selected from the segmented image. The scale shape is defined as per Eq. (14).

$$Q(x, y, \sigma) = Ggaus(x, y, \sigma) * Iseg(x, y) \tag{14}$$

Here, I_{seg}*(x,y)* is the segmented image with pixels *(x,y)* and $Ggaus(x, y, \sigma)$ is the Gaussian variable scale.

(b) Keypoint Localization: the feature keypoints from the selected scale-space peak are localized accurately. The keypoints generated in the previous phase result in a large number of keypoints. 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 these scenarios. As a result, we get rid of them. The method is comparable to those used to suppress edge features in the Harris Corner Detector. The extrema location L is given as per Eq. (15).

$$L = \frac{\partial^2 W^{-1}}{\partial d^2} \frac{\partial^2 W}{\partial d} \tag{15}$$

(c) Assigning Orientation to Keypoints: Depending on the scale, a neighborhood is drawn around the keypoint spot, and in this region the gradient magnitude and direction are determined. The result is a 360-degree orientation histogram of 36 bins. Then, the histogram is created. There would be a peak in the histogram at some point. The orientation φ is computed as per Eq. (16).

$$\varphi(x, y) = \tan(O(x, y + 1) - O(x, y - 1)/O(x + 1, y) - O(x - 1, y) \tag{16}$$

(d) Keypoint descriptor: A high-dimensional vector that describes the keypoints. Each keypoint now has a position, size, and orientation. The next stage is to develop a descriptor for each keypoint's local image area which is strongly distinctive and as invariant as possible to changes in perspective as well as lighting. (e) Keypoint Matching: The closest neighbors of two images' keypoints are identified and paired [34].

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_{segm}. The edge detection phenomena are carried out to estimate the shape of the objects. For this, we've used the canny detection operation. The Canny edge detector is indeed an edge detection operator that recognizes a large variety of edges in images using a multi-stage algorithm. A multi-stage edge detector seems to be the Canny filter [35]. To compute the intensity corresponding to the image gradients, it employs a filter dependent on the derivative of a Gaussian. The Gaussian filter eliminates the influence of image noise. Then, 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 [36]. The general criteria for edge detection include (a) Edge detection with a low error rate, which ensures that the detection can recognize as much of the image's edges as possible. (b) The operated sensed edge point should be effective in locating the edge's center. (c) Image noise does not produce "false edges". The extracted shape base features via the Canny edge detector are denoted as 'f_{canny}'.

3.3.4 Gradient Features via HoG

In I_{segm}, the HOG (feature descriptor) identifies the homogeneous identical area [37, 38]. The steps followed in HoG feature extraction is depicted below:

(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 group of adjacent 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 normalization of histograms. 'f_{HoG}' represent the extracted HOG characteristics.

4 Optimal Feature Selection

To reduce the computational costs of modeling as well as, in some cases, improve the performance of the model, it is desirable to reduce the number of input variables (features) [39]. All feature subset selection systems use two main components: the search strategy used to choose the subsets of features and the evaluation method employed to measure the quality of those subsets [40–42]. In general, there are four main steps involved in a feature selection procedure. They are (a) generation of the subset; (b) evaluation of the subset; (c) stopping criteria for the procedure; and (d) validation. The step (a) involves selecting subsets based on the approach used for searching. Search direction and research methodology typically determine the approach. Several parameters are taken into consideration in Step (b) such as distance, dependency, consistency, etc. The stopping criteria in step (c) are dependent on other criteria (e.g., less error than required/chosen, complete the search, etc.) In step (d), advanced AI/ML algorithms are used to validate selected attributes. Genetic algorithms (GAs) provide a simple, general, and powerful framework for selecting good subsets of features, leading to improved detection rates.

The overall extracted features are denoted as $F = (f_{SURF}) + (f_{SIFT}) + (f_{canny}) + (f_{HoG})$. The best features ‘$F^*$’ can be derived from the overall extracted features ‘F’. The extracted best features are $F^* = (f_{SURF})^* + (f_{SIFT})^* + (f_{canny})^* + (f_{HoG})^*$, and are fed as input to optimized CNN, for training purposes.

5 Object Detection Using CNN

CNN Architecture: Convolutional neural network extracts features from an input image and provides learnable parameters to efficiently do the classification, detection, and many other tasks of an image. The proposed architecture contains five convolutional layers and five pooling layers.

5.1 Convolution Layer

The convolutional layers create a convolution kernel that is convolved with the layer input to produce outputs. The input image is convoluted using filters by using convolution operation. The kernal size is (3,3), specifying the height and width of the 2D convolution window. The convolutional phase will apply the filter on a small array of pixels within the picture. The filter will move along the input image with a shape of 3×3. It means the network will slide these windows across all the input image and compute the convolution. The output matrix is the result of the element-wise operation between the image matrix and the filter. At the end of the convolution operation, the output is subject to an activation function to allow non-linearity. The activation function for our model is the Relu. All the pixel with a negative value will be replaced by zero.

5.2 Pooling Layer

The convolutional layer is followed by pooling layer. The purpose of the pooling is to reduce the dimensionality of the input image. The steps are done to reduce the computational complexity of the operation. By diminishing the dimensionality, the network has lower weights to compute, so it prevents overfitting. In order to down sample images while preserving information, we use pooling layers, we have two types of pooling layers which are max-pooling and average pooling. In our model we are using max-pooling with pool size (2,2). Pool size is nothing but the factor by which to downscale. (2,2) will have the input in both spatial dimensions.

5.3 Fully Connected Layer

Convolution and max pooling is applied to the data set, before sending it to the output layer the model is flattened. Dropout is applied to prevent overfitting of images. The convolutional layers apply different filters on a subregion of the picture. The Relu activation function adds non-linearity, and the pooling layers reduce the dimensionality of the features maps. All these layers extract essential information from the images. At last, the features maps are fed to a primary fully connected layer with a softmax function to make a prediction. We connect all neurons from the previous layer to the next layer. We have used 'softmax' activation function to classify the input image. The overall process is depicted in following Fig. 4.

As discussed in the earlier sections, the extracted $F*$ are given as input to optimized CNN 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 CNN to enhance the detection accuracy.

Convolutional kernel defines the entire function map in such a way that r$^{\text{th}}$ layer of convolutional layer is mapped with z$^{\text{th}}$ feature map as well feature values in the location

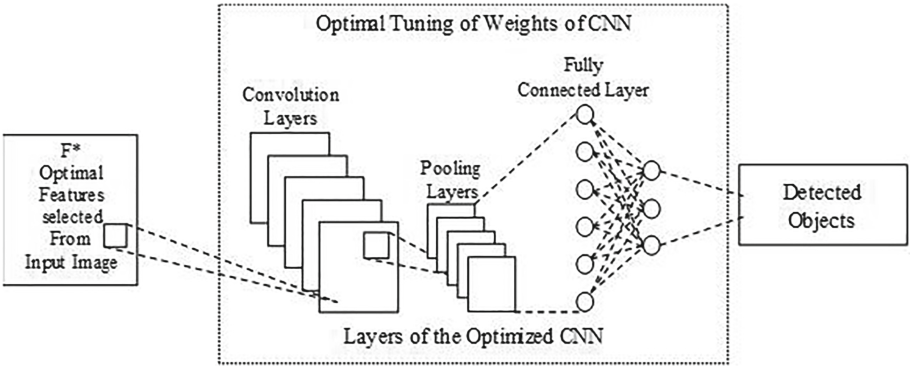
Optimal Tuning of Weights of CNN
Convolution Layers
Fully Connected Layer
Pooling Layers
F* Optimal Features selected From Input Image
Detected Objects
Layers of the Optimized CNN

Fig. 4. Schematic Diagram of Optimized CNN for ODMVI

provided to the CNN and being determined by Eq. (17).

$$S^r_{e,x,z} = W^r_z F *^r_{e,x} + B^r_z \tag{17}$$

$$AF^r_{e,x,z} = AF\left(S^r_{e,x,z}\right) \tag{18}$$

$$O^r_{e,x,z} = pool\left(AF^r_{e,x,z}\right), \forall (c, r) \in \Re_{e,x} \tag{19}$$

where W^r_z and B^r_z are the optimum weight vector and bias, which provides the optimal tuning of the weights. Similarly, activation function $AF(\cdot)$ provides prediction of non-linear features of multilayer networks and here it is used to achieve non-linearity and presented in Eq. (18), when processed provides activation value. The Shift variance in the pooling layer is handled by Eq. (19) and it deals with decreasing the resolution of induced feature map by local neighbourhood and presented by *pool* (). The down sampling operations were also conducted by the pooling layer in CNN with the result 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

$$Loss = \frac{1}{Num} \sum_{t=1}^{ms} G\left(\varsigma; V^{(t)}, OUT^{(t)}\right) \tag{20}$$

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 Eq. (21).

$$Obj = \min(Loss) \tag{21}$$

In order to have minimization of loss, the efforts were taken for fine-tuning the weight– 'W' of CNN.

6 Dataset

We have used '*ImageNet*' dataset for the simulation of the proposed model [43]. 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 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.). The dataset images were separated in to training dataset and testing datasets taking 80%-20% ratio (8000 images for training and 2000 images for testing). After the complete analysis of CNN over the dataset, the model was trained to identify the objects with remarkable accuracy and precision.

7 Results and Discussion

7.1 Simulation Procedure

The proposed model (ODMVI) is evaluated over the existing models like CNN+ PSO (Particle Swarm Optimization), CNN+ WOA (Whale Optimization Algorithm), CNN+GWO (Grey Wolf Optimization) & CNN+SLnO (Sea Lion Optimization Algorithm) 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. The positive measures like "accuracy, specificity, sensitivity, and precision" are ought to be sustained at a higher level, for the most favorable results. The error measures or native measures are False Positive Rate (FPR), False Negative Rate (FNR), and False Discovery Rate (FDR), which need to be as low as possible. The F1-score (harmonic mean of precision and recall), Matthews's correlation coefficient (MCC), and Negative Predictive Value (NPV) are additional value-added indicators that exhibit the supremacy of the proposed work.

7.2 Convergence Analysis

The proposed model should exhibit higher convergence towards the defined objective function in order to better understand its performance. Figure 5 summarizes the results of the convergence analysis of both proposed and existing models for each iteration. Initially, both the proposed and existing models have higher convergence at the lowest iteration count (at 0th iteration). When the number of iterations grows, the cost of the proposed and existing models comes down as they go through more iteration. Comparing the proposed method with the traditional method, the proposed one eventually achieves minimum fitness values. Furthermore, the proposed method achieves fewer fitness values at a maximum of 50 iterations. Thus, the ODMVI model can reduce the CNN loss, representing the superiority of the proposed model in terms of detection accuracy.

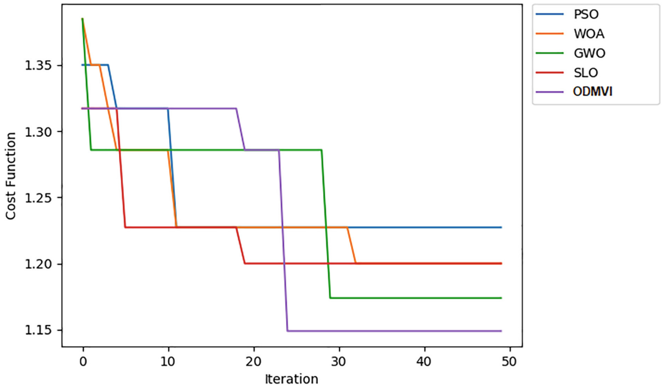

Fig. 5. Convergence Analysis

7.3 Performance Evaluation of ODMVI

The results for accuracy and precision are shown in Table 1, Fig. 6 and Table 2, Fig. 7 respectively as shown below. The obtained results indicate that the ODMVI is very beneficial, since it has indexed the highest percentage for every variation in learning. By focusing on the most important metric, accuracy, the ODMVI shows its superiority. Interestingly, the accuracy of planned work is found higher with any shift in the learning percentage. Additionally, the planned work was at its best even at the highest learning percentage (80 percent).The accuracy of the ODMVI at learning percentage = 80 is 76.78%, which is better than the existing models like CNN +PSO = 65.67%, CNN + WOA = 49%, CNN +GWO = 53.17%, CNN + SLnO = 61.50%.

Moreover, the ODMVI's precision, sensitivity, and specificity increase with an increase in learning percentage. Also the precision outcomes of the model are also exciting. The precision of the ODMVI at learning percentage 80 is 65.67%, which is better than the existing models like CNN+ PSO = 32.33%, CNN +WOA = 0%, CNN+ GWO = 7.33%, CNN+SLnO = 24%. 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.

In addition, the ODMVI has archived the least error measures. The False Discovery Rate - FDR (as shown in Table 3) of the ODMVI is 0.32, which is the least value when compared to traditional works like CNN+ PSO = 0.65, CNN +WOA = 0.99, CNN+GWO = 0.90, CNN +SLnO = 0.74. The False Negative Rate - FNR (as shown

Table 1. Performance Analysis of Proposed and Conventional Work in terms of Accuracy

Learning Rate	CNN +PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
60	0.698333	0.656667	0.656667	0.601111	0.656667
70	0.573333	0.61963	0.684444	0.61963	0.730741
80	0.656667	0.49	0.531667	0.615	**0.767778**

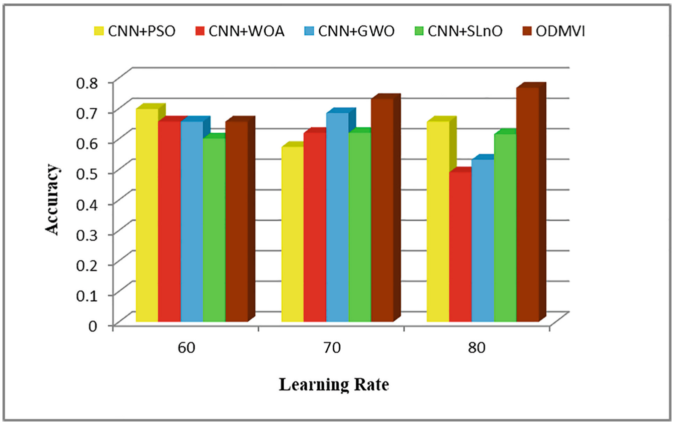

Fig. 6. Performance of Adopted Method Over Extant Models for Accuracy

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

Learning Rate	CNN+PSO	CNN+WOA	CNN +GWO	CNN+SLnO	**ODMVI**
60	0.406667	0.323333	0.323333	0.406667	0.49
70	0.156667	0.434444	0.378889	0.434444	0.601111
80	0.323333	0	0.073333	0.24	**0.656667**

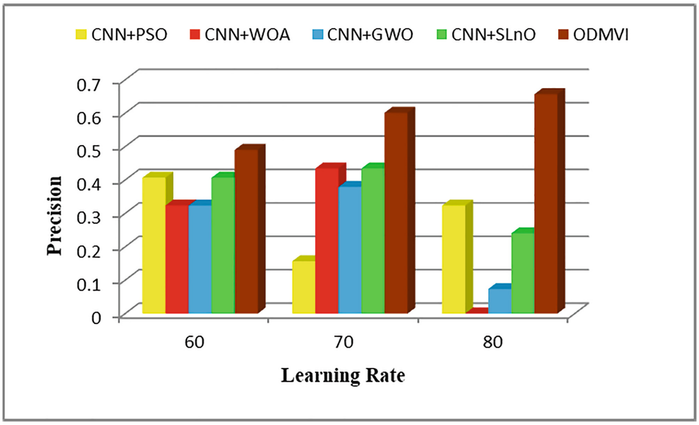

Fig. 7. Performance of Adopted Method Over Extant Models for Precision

in Table 4) of the ODMVI is 0.31 (least value) at 80th learning iteration, which is better than the existing works like CNN+PSO = 0.646, CNN +WOA = 0.98, CNN +GWO = 0.896, CNN +SLnO = 0.73. The False Positive Rate-FPR (as shown in Table 5) of the ODMVI is 0.156 (least value) at 80th learning iteration, which is better than the existing

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

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
60	0.573333	0.656667	0.656667	0.573333	0.49
70	0.823333	0.545556	0.601111	0.545556	0.378889
80	0.656667	0.99	0.906667	0.74	**0.323333**

works like CNN+ PSO = 0.212, CNN +WOA = 0.323, CNN + GWO = 0.295, CNN +SLnO = 0.240.

Moreover, F1-score (harmonic mean between precision and recall) = 0.656 (as shown in Table 6), Matthews Correlation Coefficient (MCC) = 0.49 (as shown in Table 7), and Negative Predictive Value (NPV) = 0.823 (as shown in Table 8) are found to be higher with the ODMVI for every variation in the learning percentage. From the overall evaluation, it is clear that the ODMVI had achieved the optimal values; thereby the ODMVI had become much sufficient for detecting the objects.

Table 4. Performance Analysis of Proposed and Conventional Work in terms of FNR

Learning Rate	CNN+PSO	CNN+WOA	CNN +GWO	CNN+SLnO	**ODMVI**
60	0.563333	0.646667	0.646667	0.563333	0.48
70	0.813333	0.535556	0.591111	0.535556	0.368889
80	0.646667	0.98	0.896667	0.73	**0.313333**

Table 5. Performance Analysis of Proposed and Conventional Work in terms of FPR

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
60	0.184444	0.212222	0.212222	0.281667	0.24
70	0.267778	0.267778	0.193704	0.267778	0.184444
80	0.212222	0.323333	0.295556	0.24	0.156667

Table 6. Performance Analysis of Proposed and Conventional Work in terms of F1-Score

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
60	0.406667	0.323333	0.323333	0.406667	0.49
70	0.156667	0.434444	0.378889	0.434444	0.601111
80	0.323333	NaN	0.073333	0.24	**0.656667**

Table 7. Performance Analysis of Proposed and Conventional Work in terms of MCC

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
60	0.212222	0.101111	0.101111	0.115	0.24
70	−0.12111	0.156667	0.175185	0.156667	0.406667
80	0.101111	−0.34333	−0.23222	−0.01	**0.49**

Table 8. Performance Analysis of Proposed and Conventional Work in terms of NPV

Learning Rate	CNN+PSO	CNN+WOA	CNN+GWO	CNN+SLnO	**ODMVI**
0.795556	0.767778	0.767778	0.698333	0.74	0.74
0.712222	0.712222	0.786296	0.712222	0.795556	0.795556
0.767778	0.656667	0.684444	0.74	0.823333	**0.823333**

8 Conclusion and Future Scope

In this proposed model the discussed ODMVI architecture will be working in the background of the final developed system for assisting visually impaired people. The object detection is thus achieved through the major phases comprising pre-processing, segmentation, feature extraction, optimal feature selection and object detection.

The optimal features extracted from the overall features were fed as input to the optimized convolutional network for detecting multiple objects from the image. In order to attain the maximum accuracy and precision in the results, the weights of CNN were optimally tuned. The accuracy (76.78%) and the precision (65.67%) of the proposed ODMVI model at the 80th learning rate was found to be better than the existing models like CNN + PSO, CNN+WOA, CNN +GWO, CNN+SLnO. Thus from the overall performance analysis it can be concluded that the ODMVI had been proven to be more effective for object identification. The proposed work emphasizes object detection in an indoor environment for visually impaired people to assist them to live day-to-day life more easily.

In future the work will be extended by emphasizing on the optimization of time factor of the detection activity and providing the user with notification regarding the detected objects in an audio form. An android application may be developed to capture the video through smartphone and a backend server application may be developed and used to detect real-time objects.

References

1. Seiffert Simões, 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/10.1016/j.ifacol.2016.11.149

2. 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/10.1016/j.procs.2016.09.132
3. 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/10.1016/j.matpr.2017.12.344
4. 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
5. 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/10.1016/j.procs.2019.04.058
6. 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/10.1016/j.asoc.2020.106130
7. 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
8. 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/10.1016/j.simpat.2019.102015
9. 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
10. Chen, X., Xu, J., & Yu, Z. (2019). A 68-mw 2.2 Tops/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
11. 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/10.1016/j.engappai.2019.02.016
12. 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
13. 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.
14. 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
15. 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/10.1016/j.patrec.2019.03.008
16. Manjari, K., Verma, M., & Singal, G. (2020). A survey on Assistive Technology for visually impaired. *Internet of Things*, *11*, 100188. https://doi.org/10.1016/j.iot.2020.100188
17. 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/10.1016/j.procs.2017.01.190

18. Tanweer, M. R., Suresh, S., & Sundararajan, N. (2015). Self regulating particle swarm optimization algorithm. *Information Sciences*, *294*, 182–202. https://doi.org/10.1016/j.ins.2014.09.053
19. 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
20. 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
21. Darekar Raviraj Vishwambhar, 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
22. 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
23. 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
24. 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
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. 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/10.1016/j.cam.2017.12.026
33. SURF feature, from : “https://medium.com/data-breach/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e “, Access Date: 2021–0–17
34. SIFT feature, from :”https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40“, Access Date: 2021–0–17

35. Canny edge detection, from: "https://docs.opencv.org/master/da/d22/tutorial_py_canny.html", Access Date: 2021–0–17
36. 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
37. Chandrakala, M., & Durga Devi, P. (2021). Two-stage classifier for face recognition using HOG features. *Materials Today: Proceedings*, *47*, 5771–5775. https://doi.org/10.1016/j.matpr.2021.04.114
38. Salve P., Sardesai M., Manza R., Yannawar P. (2016) Identification of the Plants Based on Leaf Shape Descriptors. In: Satapathy S., Raju K., Mandal J., Bhateja V. (eds) Proceedings of the Second International Conference on Computer and Communication Technologies. Advances in Intelligent Systems and Computing, vol 379. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2517-1_10.
39. S. Gaikwad, B. Gawali, P. Yannawar and S. Mehrotra, "Feature extraction using fusion MFCC for continuous marathi speech recognition," 2011 Annual IEEE India Conference, 2011, pp. 1-5, doi: https://doi.org/10.1109/INDCON.2011.6139372.
40. K. D. Kharat, V. J. Pawar and S. R. Pardeshi, "Feature extraction and selection from MRI images for the brain tumor classification," *2016 International Conference on Communication and Electronics Systems (ICCES)*, 2016, pp. 1-5, doi: https://doi.org/10.1109/CESYS.2016.7889969.
41. Pawar, Vikul & Kharat, Kailash & Pardeshi, Suraj. (2019). Enhancement in Brain Tumor Diagnosis Using MRI Image Processing Techniques: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I. https://doi.org/10.1007/978-981-13-3140-4_59.
42. Vivek H. Mahale, Mouad M.H. Ali, Pravin L. Yannawar, Ashok T. Gaikwad, Image Inconsistency Detection Using Local Binary Pattern (LBP), Procedia Computer Science, Volume 115, 2017, Pages 501–508, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.09.097. https://www.sciencedirect.com/science/article/pii/S187705091731921X)
43. Dataset link: https://www.kaggle.com/c/imagenet-object-localization-challenge/data?select=imagenet_object_localization_patched2019.tar.gz