IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 1, Ser. 5 (Jan. – Feb. 2025), PP 01-09
www.iosrjournals.org

Mission Vision: From Concept To Reality Going The Feature Pyramid Network Way.

Neena Ahuja, Khwaja Mohammad Rafi
Department Of EEE
Mewat Engineering College, Nuh, India

Abstract
Impossible seems to be blurring away in the days of technological growth. Women in the privileged as well as the less privileged class undergo mistreatment leading to causes of congenital imperfections, eyes being primarily effected. Eye fusion in childhood together with progressive myopia could be a leading cause of worry to parents throughout the world. Muscle spasm or stretched eye muscles, irregular shape of eyes could be one of the leading cause of ocular fusion or squint in kids or adults leading to persistent stress to parenting their future. Insecurities could lead from macular degeneration, to blindness. To avoid setback of any nature the best proposition is to be acquainted with the worst scenario and research on the solutions available. The worst outcome in case of any eye disease is blindness partial or total. People at one point either suffer from mind complex under societal pressure for aesthetic looks or from setbacks because of sight loss. The first line of solution is to get oneself acquainted with the solutions in hand. The second line of solution is to develop products that are economic as well as set up that infrastructure within ones reach. The third line of solution is to develop Artificial Intelligence algorithms that can help us with smart kitchens, bathrooms, bedrooms, libraries, transportation for the visually impaired. The fourth line of solution is deeply enhance the neurological capabilities of people with high ocular degeneration to develop images within the brain based on the limited focus of their eye muscles or the neuroreceptors associated with the art of visualization.

Keywords: *fusion; macula; myopia; neuroreceptors; artificial intelligence*

Date of Submission: 15-02-2025 Date of Acceptance: 25-02-2025

I. Introduction

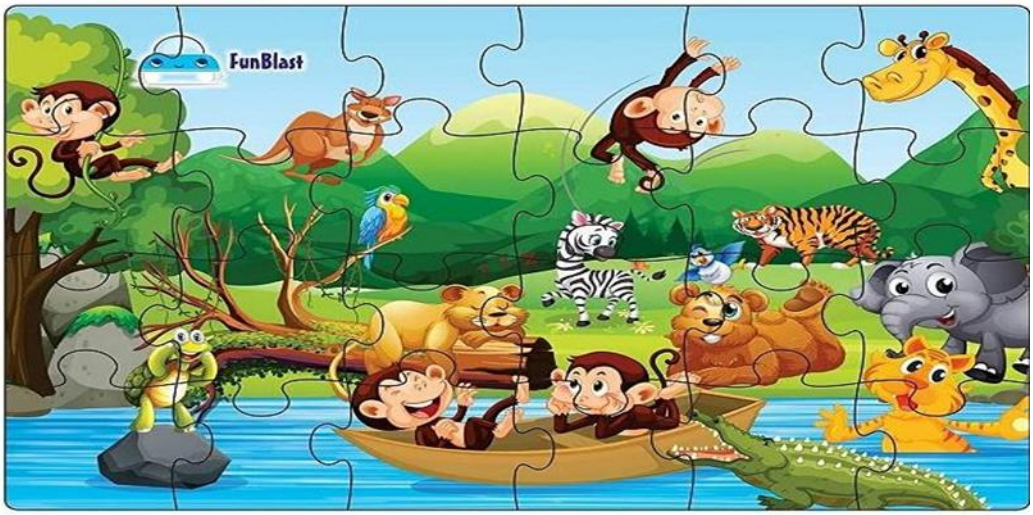
We talk about a net that would capture the characteristics of human beings/objects ; sensing through the computer eye and processing through its central nervous system called the Feature Pyramid Network [1]. Feature Pyramid Network started taking its roots back in the 1950's when the conceptualization of computer vision was envisaged in the 1950's. In its primitive years computer vision was meant to classify objects into simple categories like circle and square and detect an objects edges. In its early seventies Computer vision advanced to interpret written text both typed and handwritten for the blind using optical character reader; thus commercialising the concept.

Vision Maturity in the nineties made face recognition specific in videos and pictures due to large data sets of images available online for analysis ; this adding to the commercialisation of the concept through face recognition programs [2].

Object identification and image processing is actually incorporation of the puzzle solving mechanism. Assembly of visual images is carried on by the processor in similar ways as we put together the pieces of a puzzle [2].

II. Approaching Visual Dynamics:

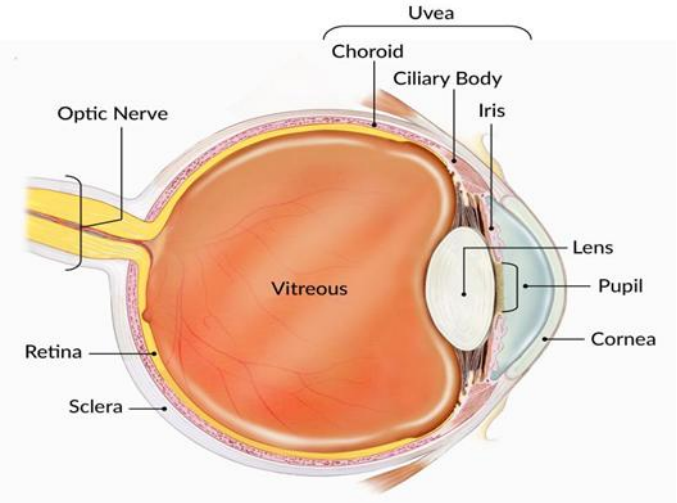
An animal kingdom jigsaw puzzle is to be pieced together right to build the foundation of feature extraction [2].



Picture 1. Funblast wild animals jigsaw puzzle [3].

The piecing together exercise consists of distinguishing each of the components that make up the puzzle, identifying all the the edges of the elements/components, locating their place values and finally converging the photograph in two dimensional plane in the retina of the computer . The art of piecing together the different pieces of a network consists of distinguishing each of the components that make up the puzzle[2].

Artificial intelligence models the human brain by comprehending the image and taking measures based on it. Light is made to dance through various steps as it passes through the eye to reach the retina where the final image is formed. The test begins : First light has to pass through the Cornea, the clear front layer of the eye. Dome shaped cornea helps with good distribution of air and light; primarily it helps light to bend and focus [4].



Picture 2: Eye optics, National Eye Institute.

A portion of this light enters the eye through an opening called the pupil. The controlling station named as Iris (coloured eye portion) decides how much light enters the pupil. Next pass the light takes is through the lens which is also the clear portion of the eyes. The lens and the cornea together play the trick to focus the light on the retina which is a light sensitive tissue at the back of one's eye. Another pass the light gets is when the photoreceptors (special cells) convert light into electrical signals . These electrical signals travel up the optic nerve to the brain where the image of the animal/thing is formed [4].

The space between the eyeball and the retina is filled with a clear gel called vitreous humour that provides shape and nutrients to the eyeball[5]. There is no space occupied by vacuum in the eye. The Ciliary Body secretes a water like liquid called aqueous humour with less protein otherwise having a composition similar to blood plasma in the space between the lense and the retina [6]. Nourishment and oxygen to the outer retina is supplied by choroid.

Forward and backward movement of the retina is accomplished by choroidal fluctuations such that the photoreceptors align to the focus line[7].The supporting wall of the eyeball is the sclera. It preserves the shape of the eyeball and protects it from injury [8].

III. Addressing Artificial Intelligencia.

An artificial eye can be modelled using artificial neural networks. Eye as we know only senses the light, however the image is formed in the brain. The brain actually processes the electrical signal that reaches the retina. Say an image is passed with NXN pixels, each pixel fed to first layer of neurons as the input [9].

Channels connect neurons in each subsequent layer and each channel has a specific weight associated to it. These weights are multiplied to the input neurons ; their computed sum added to a bias value linked to the neurons is sent to the second layer viz the hidden layer. The activation function passes this computed value through itself transferring data to the next later.

The neuron with the highest value determines the output. This is the sequence by which data propagates through the network [9].

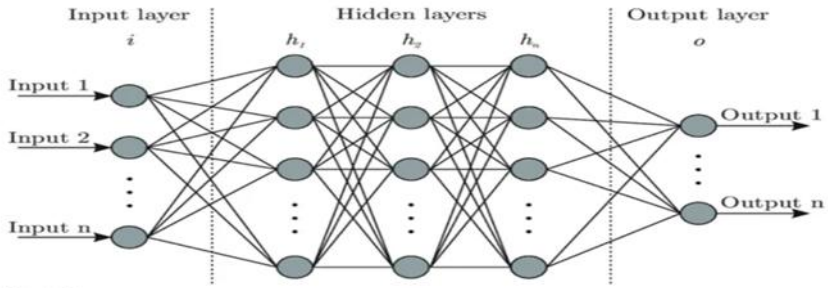


Figure 1

Training a single neuron using Neural network can be simulated using Microsoft Excel. Three input columns viz; X0,X1 and X2 are chosen X0 is assigned a common number 1 ; while the second input is fed with 0,0,1,1 sequence and X2 is fed with 0,1,0,1 sequence. Then we select weights viz; weight 0 (w0), weight 1 (w1) and weight 2 (w2). We give an initial bias of 1 to all the three weights. The weighted sum i.e. net is calculated as summation of weight times it's input respectively. The output ' Y' is calculated using ;'if then 'statement for eg. If weighted sum is going to be less than or equal to zero put a 1 and if not then put a 0 [10].

Now the big question is how do we train this single neuron through artificial neural network as per the dynamics of the problem. For this we select 3 variables namely change in weight wo (changewo), change in weight w1 (changew1), change in weight w2 (change w2).Firstly we need to recognize the desired output. Say the desired output is 0,0,0,1. Then we go ahead choosing a coefficient of learning (usually a fixed value) 0.5. The change in weight w0 being calculated as the learning coefficient times the input X0 times the difference between the value desired and that obtained i.e the error and so on on similar lines for change in weight of w1 and w2.

IV. How And Whys's Of Computer Vision.

Why a computer engineer needs to restore the vision of a computer is because he needs to study the intricacies of designs , plans, diseases,documentation,objects for further analysis, treatments and conclusions . He starts by segmenting the image ; observes them as squares of Cadbury dairy milk for individual examination.

When he goes about detecting object/objects in an image , say for example table tennis court , the net ,an offensive player, a defensive player , the ball , the bats, the table and so on he uses XY coordinate geometry to bound the whole lot.

He then goes about Facial identification by not only identifying the face of a human,but also of a specific individual. Next he goes about better understanding of an image , by studying the boundary of the object . Another attempt is taken up to identify colours,repeated shapes and other visual indicators in images which he terms as pattern recognition ; the purpose being to classify images in different categories; this termed as image classification.

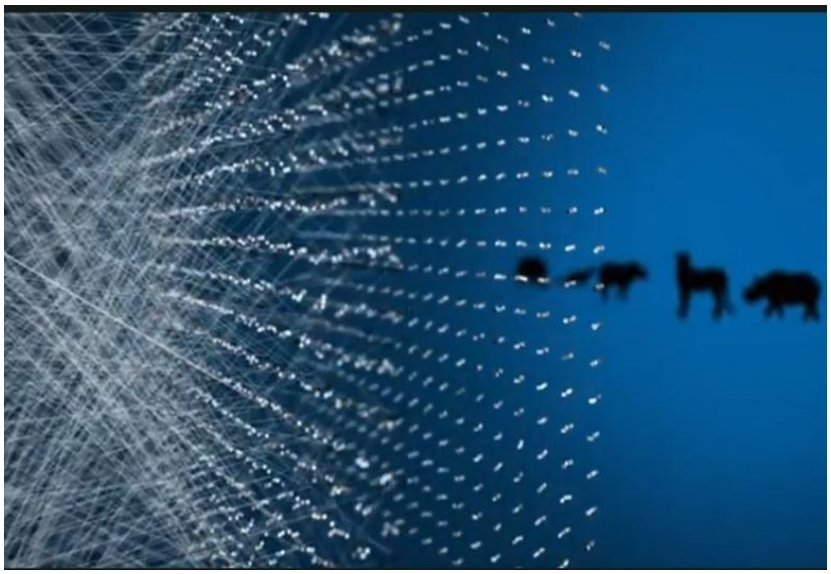
Finally we put together pattern recognition and image classification to arrive at feature matching of images [2]. The how's of computer vision starts by questioning how to acquire Images whether in large sets or small in real time through 3D imaging , video or photo analysis. The exercise of deep learning/ processing the image requires feeding the computer with numerous tagged images that are pre - identified, interpreted, and /or clasified [2].

Picture 3: Image display on a monitor

This image is displayed on a screen of a computer. A computer does not specifically see these colours, it instead sees pixels ; each having a value [Picture 3]. The neural networks does fancy computations and reasoning using these values at the back of the computer to develop the picture or the object [Picture 4,5,6,7].

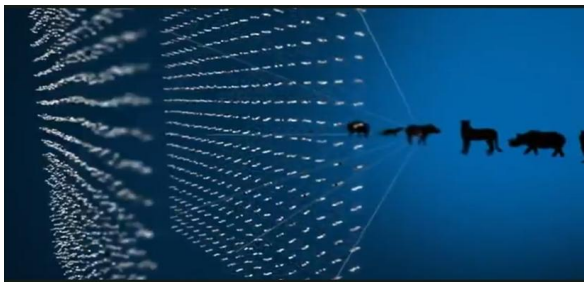
Picture 4: Flow of information in the computer

Convolutional Neural Networks (CNN's) are based on modern computer vision algorithms. Advantage being dramatic performance improvement as compared to image processing the traditional way [11].



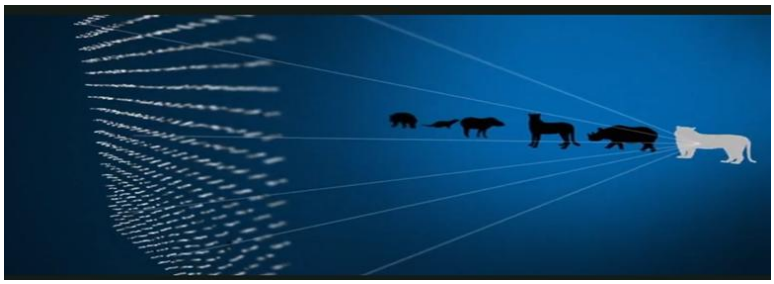
Picture 5: Propagation of signal through neural network

CNN architecture is multilayered architecture, multilayered to reduce data and calculations to the most relevant set. Comparision of the relevant set to a known set is done to identify and eventually classify the data input [11]



Picture 6; Convergence Of Signal To Form Image

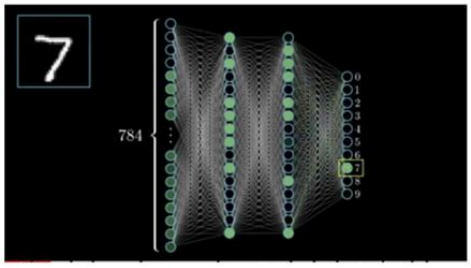
Detailing the above, let's start with the top- down approach. In the top-down approach, instruction flows from the top management to the bottom workforce. This is the traditional approach to computer vision , in which the top management (human assistance) always dictates the computer or its fellow associates as to 'what should be there ' in the organization/image as compared to the CNN/modern computer vision in which the bottom to top approach is followed which speaks about training the neural network as to 'what is there' in the image through deep learning.



Picture 7: Focus adjustment by neural networks for sharp image formation.

Speaking about the the top down approach , the traditional way of assisting the brain to visualise the intricacies of the object or the animal , the object/animal is understood by, examined and stored in the system/brain of the computer to be recollected later . The brain of the computer, stores the image/features of the object/animal which include but not limited to its eyes, ears, nose, head, legs, tail etc as pixels for its recognition which further trigers an output. At a later stage given animal/object to be recognised , the digital image is subdivided into numerous images to be matched with that stored into the memory . The method turned out to be inefficient as the system would fail to recognise if there was even a slight mismatch in the colour or the brightness of the two images .Secondly, it was burdensome and càumbersome requiring lot of storage memory as every detail of the object/animal/human/landscape/seascape needed to be stored in the computer for later reference and recognition. Manual crafting of rules for feature detection proved to be laborious as well as time consuming [11].

When modern engineering moved towards the bottom top approach we started visualizing and conceptualizing biology as mathematics. For example we started understanding a neuron as a number and only a number, between 0 and 1. We wanted the computer to interpret the image of a number fed to a square grid of 28 x 28 pixels, equivalent to 784 neurons which comprise the first layer of the neural network.



Picture 8: Neural network [11]

Picture 8: Neural network [12]

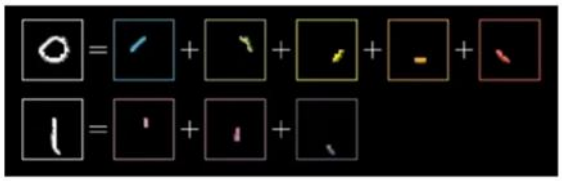
The grey pixels are fed with a value between 0 and 1 , called as activation while the black pixels acquire a value of zero,the lighted neuron acquires a value close to or equivalent to 1 when activated. [1.

The end result is always kept in mind.The last layer is supposed to show the digit which has been fed to the first neuron. The two hidden layer is chosen arbitrarily of 16 neurons.When one layer activates; it activates the other layer[12].

Picture 9: Fragmenting a number 9into image constituents

Whenever we feed a number to the input layer , the number is broken down into smaller, simpler components for the neural network to absorb as to what has been fed to it and provide a result that helps a computer brain analogous to a human brain to recognize the feed. Next every loop is sub- divided into various edges.

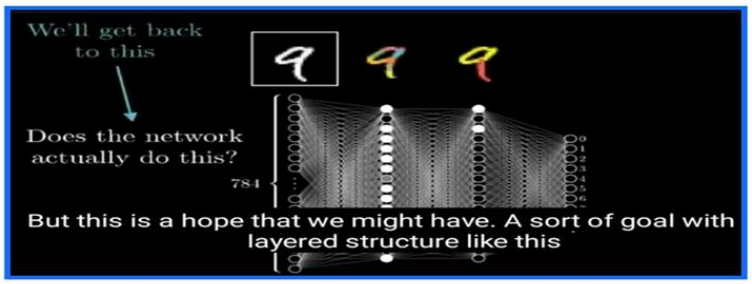
Let's take for example the number 9 and divide it into two components. These two components are again divided into various smaller fragments also called as elements that when put together will sum up to the number 9 back again.Now this is no mathematics . Here there are no summation formulae to be upplied but pure human reasoning as to how to identify the images that shall randomly occupy the neurons in the layers and then place the eight subcomponents in such a manner that the digit 9 is retrieved by the time the output neural layer is reached [12]



Picture 10: Further dividing number 9into still smaller segments

The first hidden layer of the neural network lights up when 8-10 little edges are identified. Which sub components of the components correspond to the input digits is the task that the neural network has to identify. Once the edges are identified that make up the component the neurons are lighted up.The first layer consisting of say 784 pixels is fed to the first layer i.e the input layer to be read. These 784 pixels depict the handwritten number 9.

The second layer brights up the pixels corresponding to the edges (8-10) in number which correspond to the number 9. These in turn fire the third layer which correspond to the loop and the line understanding some sort of pattern that make up the number 9.Finally the last layer identifies the number 9 and only one neuron gets lightened up that represents the number 9.



Picture 11: Finally we ignite the neuron in the last layer that dictates the number 9

Neural network thus can be visualized as a motivational exercise in which I feed an input to the first layer triggering some neurons to fire which in turn activates the second layer of neurons to fire, which in turn motivates the third layer to fire and finally the final answer is lit up in the fourth output layer.

We go on to choose neural networks to help visually impaired see because of its deep learning properties. With a little training , we can systematically develop the model from unsupervised learning, to supervised learning and enhance it with reinforced learning capabilities, the way we have in Google maps guiding us through unknown routes.

While the Convolutional Neural Network helps identify images; the Long short term memory network is good for speech recognition[12].

V. Data Analytics For The Visually Impaired

Actionable insights by studying raw data is the main purpose of Data Analytics. Data analytics helps improve decision making, study processes, deeply study and understand customer requirements, their service issues , their problems to provide the requisite customer support. Visually impaired require deep understanding of their surroundings and environment they move in.Insights we gain through data analytics of their habitat will help companies develop products customised for their requirements[13].

Famine in data visualization in visually impaired scholars and readers was addressed by SAS Graphic Accelerator that addressed graphs and charts through verbal mode and sonification.

Interactive sonification helps explore charts ,graphs, scatter maps, time series plots, heat maps, histograms

When sound travels from the left ear to the right ear the user interprets as a mobile X axis wherein a bar graph is being represented.The pitch of the sound represents the length of the bar graph.The higher the pitch the taller the bar graph[14].

Literature reviewed suggests that a number of applications have been developed to guide the blind through their daily activities and chores.

Firstly, blind man sees only black. The first help with the smart phone is that we can zoom the picture and get verbal help with the aplication that has been clicked , say for example ' this is Google Chrome', 'this is Gmail'., ' this is Message folder' etc [15]. This is how Android phones can be helpful.

A completely voice controlled Smart Phone RealSAM Pocket is ideal for blind people because of its simplicity of usage .

Picture 12: Smart phone for the visually impaired with large font

Voice to text, voice navigation are the two features together with large screens that can be modified for sharpness and size of texts for ease of reading with little button use. This simplicity of single button navigation was introduced in phones fo the visually impaired, prime feature being large button where the phone toes not have the touch feature facility.

In the economy class, we have Jitterbug smart3 phone with medical alert capabilities. Priced at $113 approximately , it's voice to text feature helps one to dictate messages , perform searches online and text emails. The Urgent Response Button helps the elderly with low vision or with no vision to contact a medical staff or representative and the representative will get back with the required help, care, advice etc. It's wide range of data plan in addition to hearing aid helps one to watch movies , surf the web, understand a landscape or seascape by asking the phone to describe it[16]

Another IPhone SE not only offers one to enlarge photos and texts for better visibility but allows very good voice commands. Siri , the inbuilt voice assistant helps with calling, texting,browsing,setting reminders and making appointments, the only disadvatage being it's 4.7 inch screen[16]

Another cell phone available for the visually impaired is the Minivision 2+, ; it's smart characteristics include an alarm, camera, detection system for light, colour and money , voice notes, flashlight, calculator,

weather and 'where am I' feature including recorder , making calls and texting messages. Elevated lettering in the keyboards, three bumps in the power supply and alarm buttons helps the visually impaired touch and feel the button and select the right one providing confidence of sight. The Inherent Voice guide of the Minivision 2+ makes it a 'talking phone' . Every move of the keyboard is spoken , if one is writing every letter is spoken , if one has deleted a letter it will express itself as 'letter s removed'. The phone will speak it's status as 'battery low' or 'today's date and time is ' etc. The phone supports 19 languages. All premium voice modules that are ones favourite can be downloaded. Again, calls and emails can be ,made directly pressing the OK button and voicing 'Call Mr. Chatterjee' or ' send message to Mr. Chatterjee'. The call history will let one know the time and date of the call as well as make calls directly from the call log if directed through verbal assistance, as well as the audio guide will analyse graphical data and provide the results.

Without any voicemail or applications or options that could confuse we have the Raz Memory Cell Phone with the 911 emergency dial and an additional Raz emergency dial (that needs to be subscribed) with six photos and names (choice extending upto thirty).Screen being only one that is called the primary screen.The simplicity in design is primarily to aid patients with Alzheimer's, Dementia and visually impaired.
In addition there are numerous applications that can be downloaded that are of help to the blind.

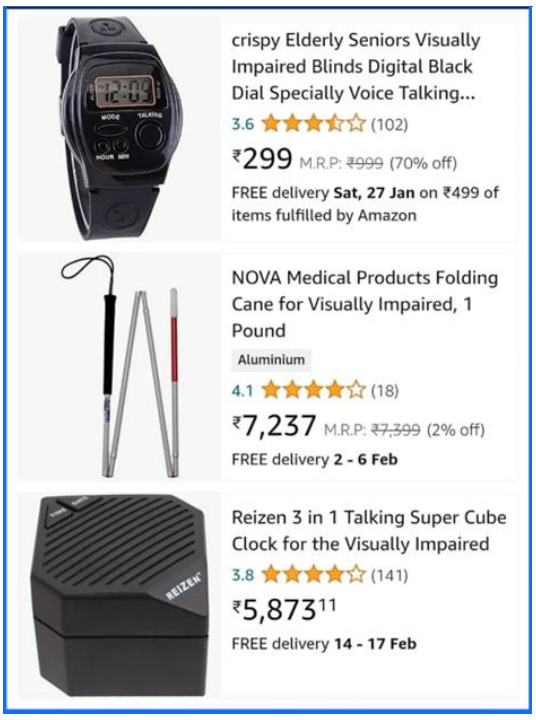
Volunteers registered with applications like 'Be my Eyes', 'Envision', 'NEI See what I see', help not only the blind but also people with a range of eye diseases like Glucoma, Cataract, Macular Degeneration, Diabetic Retinopathy, progressive myopia recognise objects and help them perform day to day activities.

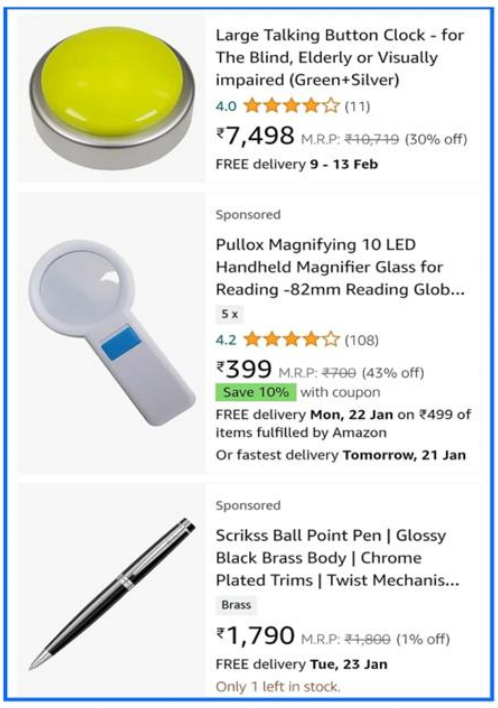
Priced at 600 pounds the smart cane is a great support model for the visually impaired. Initially the cane used to rattle, while the latest cane is the silencer cane that makes no noise or is noise free. It is the most reliable and handy technology for the blind with a cost price of 100 pounds. It helps a blind man smoothly tread the road and navigate unknown territories way in line the campass operates.

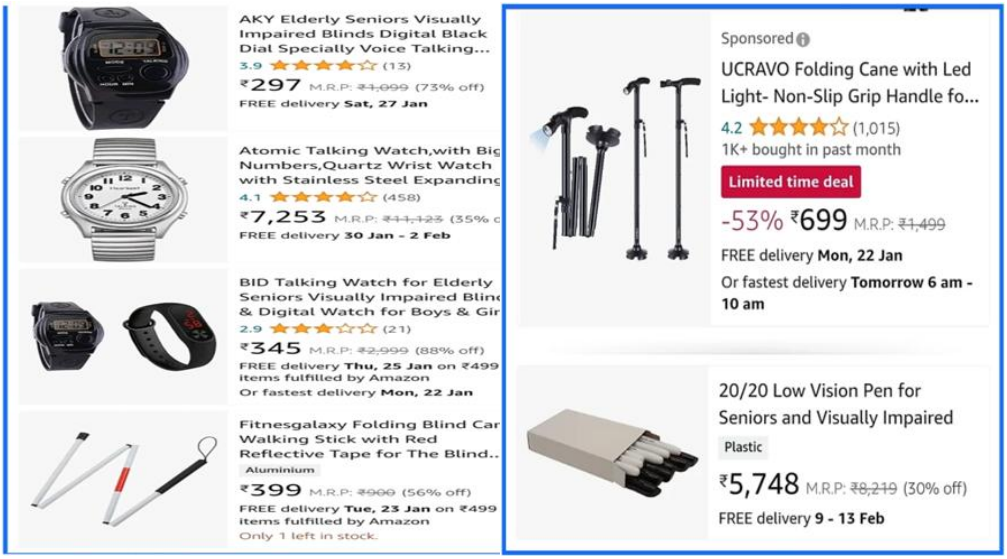
The most reliable tread though is a dog friend who assists the blind 24x7x365. In addition the assistance of a smart glass does the trick. Just click the side of the spectacle handle and it will take the image of the surrounding / the magazine / the commodity kept in the kitchen and guide you through the kitchen chores. One could as well on the click of a button on the handle of the spectacle call your general practitioner/ opthalmologist/ volunteer to guide one through the household chores if one gets stuck or cannot figure out the route [17], talking glasses being the focus of interest.

VI. Economic Concerns

Science has made mission impossible possible. Literature reviewed has generated a confidence that light prevails even in darkness. Blind men can "see" - technology has made it possible. The only concern being whether this technology is feasible for the common man or only affordable to the rich and people in power. Range of products and their prices that empower the blind and visually impaired are summarized[Screenshot 1],[Screenshot 2],[Screenshot 3],[Screenshot 4]







Right from Smart Kitchen for the blind with the smart watch and the smart spectacle guiding you through the kitchen helping you surf shelves to exact location of the comodities/spices/condiments to locating of the gas stove and lighting it is a perfect example of engineering a smart phone for the benefit of the blind.

Another smart countouring method is to map the visually impaired in your framework and see if he is meeting an obstacle if yes caution or place the image in another location for smart turning. Smart photovoltaics are assisting the blind men tour their territory as well". Photovoltaics for helping the blind men see. solar panels into smallest but not the last we can have plenty of volunteers in the technology of Batman or heman or spiderman who have always been our heroes helping everyone out through obstacles big or small.

References:

[1] Lin Et.Al, Feature Pyramid For Object Detection, IEEE, Https://Paperswithcode.Com/Method/Fpn
[2] Sas, Computer Vision ,What It Is And Why It Matters. Https://Www.Sas.Com/En_In/Insights/Analytics/Computer-Vision.Html
[3] Funblast Wild Animals Jigsaw Puzzle-24 Multicolour., Firstcry.
[4] National Eye Institute, Https://Www.Nei.Nih.Gov/Learn-About-Eye-Health/Healthy-Vision/How-Eyes-Work#:~:Text=The%20cornea%20is%20shaped%20like,Inner%20part%20of%20the%20eye
[5] Vitreous Body, Https://En.M.Wikipedia.Org/Wiki/Vitreous_Body
[6] Aqueous Humour, Https://En.M.Wikipedia.Org/Wiki/Aqueous_Humour
[7] Choroid, Https://Byjus.Com/Neet/Choroid/#:~:Text=Function%20of%20Choroid&Text=Apart%20from%20its%20vascular%20functions,To%20the%20plane%20of%20focus.
[8] Sclera, Https://Www.Google.Com/Search?Q=Inportance+Of+Sclera+In+The+Eyes&Sca_Esv=559959589&Sxsrf=AB5stBgpgJWGchuY8wFbuuLfr0tEbR_Aiq%3A1692950660209&Source=Hp&Ei=Hgdozovib46-Hwpihogqcq&Oq=Inportance+Of+Sclera+In+The+Eyes&Gs_Lp=Ehftb2jpbgutz3dzlxdpei1occigaw5wb3j0yw5jzsbvzibzy2xlcmegaw4gdghligv5zxmybxahgkabgaoycbahgbyyhhgdsl6gavcrt1jhnafwangakaeamagfa6abuukqaqyyltmwljo4aqpiaqd4aqgoag_Cagcqixjqahgnwgieecmyj8iccxaagiaegledgimbwgifec4ygatcaheqlhikbrixaxidarjharjra8iccxaagiofgledgimbwgifeaaygatcaggqabiabbixa8icer augiaegledgimbgmcbgnedwgilec4ygaqyxweyrwhcagcqabiabbgkwgikec4ygaqyahjlacicchaagiaegaiyywhcagoqabiabbixaxgkwgineaaygaqysqmygweycsiccbaagiofgledwgimeaaygaqychgcgmsbwgineaaydriabbixaxidacicbxaaga0ygatcagyqabgwgb7caggqabgwgb4yd8iccbaagiofgiydwgieeceyfcicchahgbyyhhgpgb0&Sclient=Mobile-Gws-Wiz-Hp#Irss-Feed
[9] Vivek Bhagat, Overview Of Convoutional Neural Networks, Https://Www.Topcoder.Com/Thrive/Articles/Overview-Of-Convolutional-Neural-Networks?Utm_Source=Thrive&Utm_Campaign=Thrive-Feed&Utm_Medium=Rss-Feed
[10] Scott Turner, Artificial Neural Network. Training A Single Neuron With Excel, Youtube.
[11] Pinakin Ariwala, Modernising Computer Vision With The Help Of Neural Networks. Https://Marutitech.Com/Computer-Vision-Neural-Networks/
[12] 3blue1brown," But What Is Neural Network", Chapter 1, Deep Learning
[13] What Is Data Analytics, Https://Aws.Amazon.Com/What-Is/Data-Analytics/#:~:Text=Data%20analytics%20converts%20raw%20data,Making%2C%20and%20foster%20business%20growth.
[14] SAS Enables Visually Impaired To Visualise Data, Https://Insidebigdata.Com/2017/03/18/Sas-Enables-Visually-Impaired-Visualize-Data/
[15] TCS International Voice Centre,Tata An Adaptive Enterprise.
[16] Macularevive, Https://Lowvisionaids.Org/Blog/Best-Phones-For-Visually-Impaired/
[17] Can AI Smart Glasses Make Blind Men More Independent, BBC News