**OBJECT DETECTION**

**A PROJECT REPORT**

**Submitted by**

**HITESH KATARA (0101CS171048)**

**MOKSH TEKAM (0101CS171062)**

**ANIKET KUMAR SINGH (0101CS171011)**

**HARSH PARTE (0101CS171045)**

**DEVANG DAMADE(0101CS171032)**

**COMPUTER SCIENCE FINAL YEAR GROUP**

**In partial fulfillment for the award**

**Of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

****

**(INDIA))**

**UNIVERSITY INSTITUTE OF TECHNOLOGY RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAY-(MADHYA PRADESH)UNIVERSITY INSTITUTE OF TECHNOLOGY RGPV BHOPAL**

****

PROJECTREPORT ON

**OBJECT DETECTION**

**Submitted by:- Guided by:-**

**MOKSH TEKAM (0101CS171062) PROF. RAJEEV PANDEY**

**HITESH KATARA (0101CS171048) PROF.UDAY CHOURASIA**

**ANIKET KUMAR SINGH (0101CS171011)**

**HARSH PARTE (0101CS171045)**

**DEVANG DAMADE (0101CS171032)**

**Submitted to:-**

**UNIVERSITYINSTITUTEOFTECHNOLOGY**

**RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYABHOPAL (M.P.)**

**DECLARATION**

I hereby declare that the project entitled **“OBJECT DETECTION** ” submitted for the B.Tech in computer science engineering is my original work.

**Date:**

**Signature of the Student:**

**CERTIFICATE**

This is to certify that the project titled “**OBJECT DETECTION APP**” is a bonafide record of work carried out under my supervision by “Hitesh Katara (0101CS171048), Moksh Tekam(0101CS171062),Aniket Kumar Singh (0101CS171011), Harsh parte (0101CS171045), DevangDamade(0101CS171032)”students of the B.TECH, Computer Science &Engineering, of **UNIVERSITY INSTITUTE OF TECHNOLOGY,RGPV BHOPAL (Madhya Pradesh)** during the Academic year 2021(jan. to june)in partial fulfillment of the requirements for the award of the B.Tech (Computer Science and Engineering).

**ACKNOWLEDGEMENT**

We would like to express my sincere thanks to **PROF. RAJEEV PANDEY** & **PROF.UDAY CHOURASIA** Lecturer Computer Science for giving me this opportunity to undertake this project and for his valuable guidance.

We would like to thank **DR. SANJAY SILAKARI**, HEAD OF DEPARTMENT(CSE) UIT RGPV (M.P.) for his support.We would like to thank the entire faculty and staff of the UIT RGPV BHOPAL for their support.

At the end we would like to express my sincere to all my classmates and others who helped me directly or indirectly during this project work. Their useful suggestions for this whole work and co- operative are sincerely acknowledged.

**CONTENT**

**ABSTRACT**

**CHAPTER 1 INTRODUCTION**

Project objective

Motivation

**CHAPTER 2 OBJECT DETECTION**

Introduction to Object Detection

Digital Image Processing

Gray scale Image

Color Image

Related Technology

Application of Object Detection

Object Detection workflow and feature extraction

**CHAPTER 3 DEEPLEARNING**

Introduction

Feed forward and feedback Network

Weighted sum

Activation function

**CHAPTER4 CONVOLUTIONAL NEURAL NETWORK**

Introduction

Convolution Neural Networks

CNN Architecture

Overall Architecture

Convolution Layer

**CHAPTER 5 OPEN COMPPUTER VISION**

Introduction

Application

Libraries in Open CV

Haar Cascde clasifier

**CHAPTER 6 RESULT AND DISCUSSION**

**CONCLUSION**

**ABSTRACT**

Real time object detection is a vast, vibrant and complex area of computer vision. If there is a single object to be detected in an image, it is known as Image Localization and if there are multiple objects in an image, then it is Object Detection. This detects the semantic objects of a class in digital images and videos. The applications of real time object detectioninclude tracking objects, video surveillance, pedestrian detection, people counting, self-driving cars, face detection, ball tracking in sports and many more. Convolution NeuralNetworks is a representative tool of Deep learning to detect objects using OpenCV(Opensource Computer Vision), which is a library of programming functions mainly aimed at realtime computer vision.

**Keywords: Computer vision, Deep Learning, Convolution Neural Networks.**

# CHAPTER 1 INTRODUCTION

#### ProjectObjective:

The motive of object detection is to recognizeand locate allknown objects in a scene. Preferably in3Dspace,recoveringposeofobjectsin3Disveryimportantforroboticcontrolsystems.

Imparting intelligence to machines and making robots moreandmore autonomous andindependenthas been a sustaining technological dream for the mankind. It is our dream to let the robots take ontedious,boring,ordangerousworksothatwecancommitourtimetomorecreativetasks.Unfortunately, the intelligent part seems to be still lagging behind. In real life, to achieve this goal,besideshardwaredevelopment, weneed the softwarethatcan enable robottheintelligence to do theworkandactindependently.Oneofthecrucialcomponentsregardingthisisvision,apartfromothertypesof intelligences such as learning andcognitive thinking.A robot cannot betoointelligent if itcannot see andadaptto a dynamicenvironment.

The searching or recognition process in real time scenario is very difficult. So far, no effective solutionhas been found for this problem. Despite a lot of research in this area, the methods developed so far arenot efficient, requirelong training time, are not suitable for real time application, and are not scalableto large number of classes. Object detection is relatively simpler if the machine is looking for detectingoneparticularobject.However,recognizingalltheobjectsinherentlyrequirestheskilltodifferentiateone object from the other, though they may be of same type. Such problem is verydifficult formachines,iftheydo notknowaboutthevariouspossibilitiesofobjects.

###### **Motivation:**

Blind people do lead a normal life with their own style of doing things. But, they definitely face troublesduetoinaccessibleinfrastructureandsocialchallenges.Thebiggestchallengeforablindperson,especially the one with the complete loss of vision, is tonavigatearoundplaces.Obviously, blindpeopleroam easily around theirhousewithout any help because they know theposition ofeverything inthe house. Blind people have a tough time finding objects around them. . So we decidedto make aREAL TIME OBJECT DETECTION System. We are interestedinthis project after we went throughfew papers in this area. As a result we are highly motivated to develop a system that recognizes objectsinthe realtimeenvironment

**CHAPTER 2**

**OBJECT DETECTION**

### INTRODUCTION TO OBJECT DETECTION

Object Detection is the process of finding and recognizing real-world object instances such as car, bike,TV,flowers,andhumansoutofanimagesorvideos.Anobjectdetectiontechniqueletsyouunderstandthe details of an image or a video as it allows for the recognition, localization, and detection of multipleobjectswithinanimage.

Itisusually utilized inapplicationslikeimageretrieval, security,surveillance,and advanceddriverassistancesystems (ADAS).ObjectDetectionisdonethrough manyways:

* FeatureBasedObjectDetection
* ViolaJonesObjectDetection
* SVMClassificationswithHOGFeatures
* DeepLearningObjectDetection

Object detection from a video in video surveillance applications is the major task these days. Objectdetection technique is used to identify required objects in video sequences and to cluster pixels of theseobjects.The detection ofanobject in video sequenceplays amajor rolein severalapplications specificallyasvideosurveillanceapplications.

Objectdetectioninavideostreamcanbedonebyprocesseslikepre-processing,segmentation,foregroundandbackgroundextraction, featureextraction.

Humans can easily detect and identify objects present in an image. The human visual system is fast andaccurate and can perform complex tasks like identifying multiple objects with little conscious thought.Withtheavailabilityoflargeamountsofdata,fasterGPUs,andbetteralgorithms,wecannoweasilytraincomputerstodetectandclassifymultipleobjectswithin animage withhigh accuracy.

### DIGITALIMAGEPROCESSING

Computerized picture preparing is a range portrayed by the requirement for broad test work to build upthe practicality of proposed answers for a given issue. A critical trademark hidden the plan of picturepreparingframeworks isthe huge leveloftestingandexperimentationthat

Typically is required before touching base at a satisfactory arrangement. This trademark infors that thecapacity to plan approaches and rapidly model hopeful arrangements by and large assumes a noteworthypartin diminishingthecostandtime required toland atasuitableframeworkexecution.

##### 

##### **WHATIS DIP?**

A picture might be characterized as a two-dimensional capacity f(x, y), where x, y are spatial directions,andtheadequacy offatany combineofdirections (x,y)isknownas thepoweror dark level of thepicturebythen.Wheneverx,yandtheabundanceestimationofarealllimiteddiscreteamounts,wecallthe picture a computerized picture. The field of DIP alludes to preparing advanced picture by methodsforcomputerizedPC.Advancedpictureismadeoutofalimitednumberofcomponents,eachofwhichhasaspecificarea andesteem. Thecomponents are calledpixels.

Vision is the most progressive of our sensor, so it is not amazing that picture play the absolute mostimperative part in human observation. Be that as it may, dissimilar to people, who are constrained to thevisual band of the EM range imaging machines cover practically the whole EM range, going fromgamma to radio waves. They can work likewise on pictures produced by sources that people are notacclimatedto partnerwithpicture.

There is no broad understanding among creators in regards to where picture handling stops and otherrelated territories, for example, picture examination and PC vision begin. Nowand then a qualificationis made by characterizing picture handling as a teach in which both the info and yield at a procedure arepictures. This is constraining and to some degree manufactured limit. The range of picture investigationisin the middle of picturepreparingand PCvision.

There are no obvious limits in the continuum from picture handling toward one side to finish vision atthe other. In any case, one helpful worldview is to consider three sorts of mechanized procedures in thiscontinuum: low, mid and abnormal state forms. Low-level process includes primitive operations, forexample, picture preparing to decrease commotion differentiate upgradeandpicturehoning. A low-levelprocessisdescribedbythewaythatbothits sourcesofinfoandyieldsarepictures.

Mid-levelprocessonpicturesincludesassignments,forexample,division,depictionofthat

QuestiondiminishthemtoaframereasonableforPChandlingandcharacterizationofindividualarticles

A mid-level process is portrayed by the way that its sources of info by and large are pictures however itsyields are properties removed from those pictures. At long last more clevated amount handling includes"Understanding an outlet of perceived items, as in picture examination and at the farthest end of thecontinuumplayingouttheintellectualcapacitiestypicallyconnectedwithhumanvision.Advancedpicturehandling,aseffectivelycharacterizedisutilizedeffectivelyinawidescopeofregionsofoutstandingsocialand monetaryesteem.

##### **WHATIS ANIMAGE?**

A picture is spoken to as a two dimensional capacity f(x, y) where x and y are spatial co-ordinates andtheadequacyof"T"atanymatchofdirections(x,y)isknownasthepowerofthepictureby then.



###### Fig.1.1digitalimage

**Processingon image:**

Processingonimage canbeof threetypesTheyarelow-level, mid-level,highlevel.

###### **Low-levelProcessing:**

* Preprocessingtoremovenoise.
* Contrastenhancement.
* Imagesharpening.

###### **MediumLevelProcessing:**

* Segmentation.
* Edgedetection
* Objectextraction.

###### **HighLevelProcessing:**

* Image analysis
* Scene interpretation

#### WhyImageProcessing?

Since the digital image is invisible, it must be prepared for viewing on one or more output device(laserprinter,monitorat).Thedigitalimagecanbeoptimizedfortheapplicationbyenhancingtheappearanceofthe structures withinit.

Therearethreeofimageprocessingused.Theyare

* Imageto Imagetransformation
* ImagetoInformationtransformations
* InformationtoImagetransformations

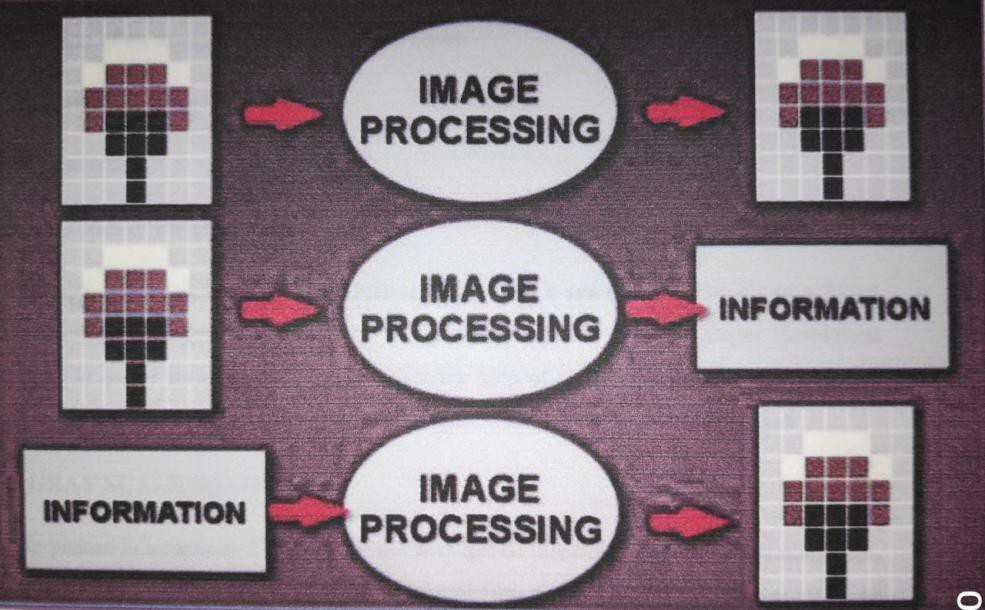


Fig.1.2Typesof ImageProcessing

###### **Pixel :**

###### Pixel is the smallest element of an image. Each pixel correspond to any one value. In an 8-bit grayscale image, the value of the pixel between 0 and 255.Each pixel store a value proportional to the lightintensityatthatparticularlocation.ItisindicatedineitherPixelsperinchorDotsperinch.

###### **Resolution:**

###### The resolution can be defined in many ways. Such as pixel resolution, spatial resolution, temporalresolution, spectral resolution.In pixel resolution, the term resolution refers tothetotalnumber ofcount of pixels in an digital image. For example, If an image has M rows and N columns, then itsresolution can be defined as MX N. Higher is the pixel resolution, the higher is the quality of theimage.

**Resolutionofanimage isof generallytwotypes.**

* LowResolutionimage
* High Resolution

Since high resolution is not a cost effective process It is not always possible to achieve high resolutionimageswithlowcost.HenceitisdesirableImaging.InSuperResolutionimaging,withthehelpofcertainmethodsandalgorithmswecanbeabletoproducehighresolutionimagesfromthelowresolutionimage fromthe lowresolutionimages.

**GRAYSCALEIMAGE**

A gray scale picture is a capacity I (xylem) of thetwo spatial directions of the pictureplane. I(x,y) istheforceofthepictureforceofpictureatthepoint(x,y)onthepictureplane.I(xylem)takenon-negativeexpectthepictureis limited byarectangle

**COLORIMAGE**

Itcanbespokentobythreecapacities,R(xylem)forred,G(xylem)forgreenandB(xylem)forblue.Apicturemightbepersistentasforthexandyfacilitatesandfurthermoreinadequacy.Changingoversuchapicturetoadvancedshaperequiresthatthedirectionsandtheadequacytobedigitized.Digitizingthefacilitate’sesteemsiscalledinspecting.Digitizingtheadequacyesteemsiscalledquantization.

**RELATED TECHNOLOGY:**

**R-CNN**

R-CNN is a progressive visual object detection system that combines bottom-up region proposals withrichoptions computed byaconvolution neuralnetwork.

R-CNN uses region proposal ways toinitial generate potential bounding boxes in a picture and thenrunaclassifieron these proposedboxes.

**SINGLE SIZE MULTI BOX DETECTOR**

SSDdiscretizestheoutputspaceofboundingboxesintoasetofdefaultboxesoverdifferentaspectratiosandscalesperfeaturemaplocation.Atthetimeofpredictionthenetworkgeneratesscoresforthe presence of each object category in each default box and generates adjustments to the box to bettermatch theobjectshape.

Additionally, the network combines predictions from multiplefeature mapswith different resolutionstonaturallyhandle objectsof varioussizes.

**YOLO**

YOLO is real-time object detection. It applies one neural network to the complete image dividing theimageinto regions andpredicts boundingboxesandpossibilities for everyregion.

Predicted probabilities are the basis on which these bounding boxes are weighted. A single neuralnetwork predicts bounding boxes and class possibilities directly from full pictures in one evaluation.Since thefulldetection pipelineisa singlenetwork,itcan be optimizedend-to-enddirectlyondetectionperformance.

**TENSORFLOW**

Tensor flow is an open source software library for high performance numerical computation. It allowssimple deployment of computation across a range of platforms (CPUs, GPUs, TPUs) due to itsversatile design also from desktops to clusters of servers to mobile and edge devices. Tensor flow wasdesigned and developed by researchers and engineers from the GoogleBrainteam at intervalsGoogle’s AI organization, it comes with robust support for machine learning and deep learning and theversatilenumerical computationcoreis usedacross several alternativescientificdomains.

To construct, train and deploy Object Detection Models TensorFlow is used that makes it easy andalso it provides a collection of Detection Models pre-trained on the COCO dataset, the Kitti dataset,and the Open Images dataset. One among the numerous Detection Models is that the combination ofSingle Shot Detector (SSDs) and Mobile Nets architecture that is quick, efficient and doesn't needhugecomputationalcapabilitytoaccomplish the objectDetection.

**APPLICATION OF OBJECT DETECTION**

ThemajorapplicationsofObjectDetectionare:

**FACIALRECOGNITION**

“Deep Face” is a deep learning facial recognition system developed to identify human faces in a digitalimage. Designed and developed by a group of researchers in Facebook. Google also has its own facialrecognition system inGooglePhotos,which automatically separatesall thephotos according to theperson intheimage.

There are various components involved in Facial Recognition or authors could sayit focuses onvariousaspectslike theeyes,nose, mouthandtheeyebrows for recognizingafaces.

**PEOPLECOUNTING**

People counting is also a part of object detection which can be used for various purposes like findingpersonoracriminal;itisusedforanalysingstoreperformanceorstatisticsofcrowdduringfestivals.Thisprocessisconsidereda difficult one aspeople moveout oftheframequickly.

**INDUSTRIALQUALITYCHECK**

Object detection also plays an important role in industrial processes to identify or recognize products.Finding a particular object through visual examination could be a basic task that's involved in multipleindustrialprocesses like sorting, inventory management,machining, quality management,packagingandsoon.Inventorymanagementcanbeterriblytoughasthingsarehardtotraceinrealtime.Automaticobjectcountingand localization permitsimprovinginventoryaccuracy.

**SELFDRIVINGCARS**

Self-driving is the future most promising technology to be used, but the working behind can be verycomplex as it combines a variety of techniques to perceive their surroundings, including radar, laserlight, GPS, odometer, and computer vision. Advanced control systems interpret sensory info to allownavigation methods to work, as well as obstacles and it. This is a big step towards Driverless cars as ithappensatveryfastspeed.

**SECURITY**

ObjectDetectionplaysavitalroleinthefieldofSecurity;ittakespartinmajorfieldssuchasfaceIDof Apple or the retina scan used in all the sci-fi movies. Government also widely use this application toaccess the security feed and match it with their existing database to find any criminals or to detectingobjectslikecar numberinvolvedincriminal activities.Theapplicationsarelimitless.

**OBJECTDETECTIONWORKFLOWANDFEATUREEXTRACTION**

Every ObjectDetection Algorithm works on the sameprincipleand it’s just theworking thatdiffersfrom others. They focus on extracting features from the images that are given as the input at hands andthenituses thesefeaturesto determine theclassoftheimage.

##### 

##### **CHAPTER 3**

##### **DEEP LEARNING**

##### **INTRODUCTION**

##### Deep learning is a machine learning technique. It teaches a computer to filter inputs through layers tolearn how to predict and classify information. Observations can be in the form of images, text, or sound.The inspiration for deep learning is the way that the human brain filters information. Its purpose is tomimic how the human brain works to create some real magic. In the human brain, there are about 100billion neurons. Each neuron connects to about 100,000 of its neighbors. We’re kind of recreating that,but in a way and at a level that works for machines. In our brains, a neuron has a body, dendrites, andan axon. The signal from one neuron travels down the axon and transfers to the dendrites of the nextneuron. That connection where the signal passes is called a synapse. Neurons by themselves are kind ofuseless.Butwhenyouhavelotsofthem,theyworktogethertocreatesomeseriousmagic.That’stheidea behind a deep learning algorithm! You get input from observation and you put your input into onelayer. That layer creates an output which in turn becomes the input for the next layer, and so on. Thishappens over and over until your final output signal! The neuron (**node**) gets a signal or signals ( **inputvalues**),which passthrough theneuron.Thatneurondeliversthe**outputsignal**.

##### Think of the input layer as your senses: the things you see, smell, and feel, for example. These areindependent variables for one single observation. Thisinformation isbroken down into numbers andthe bits of binary data that a computer can use. You’ll need to either standardize or normalize thesevariables so that they’re within the same range. They use many layers of nonlinear processing units forfeature extraction and transformation. Each successive layer uses the output of the previous layer for itsinput. What they learn forms a hierarchy of concepts. In this hierarchy, each level learns to transform itsinputdataintoamoreandmoreabstractandcompositerepresentation.Thatmeansthatforanimage,for example, the input might be a matrix of pixels. The first layer might encode the edges and composethe pixels. The next layer might compose an arrangement of edges. The next layer might encode a noseandeyes.Thenext layermightrecognize thattheimage containsaface, and soon.

##### **What happensinside theneuron?**

##### The input node takes in information in a numerical form. The information is presented as an activationvalue where each node is given a number. The higher the number, the greater the activation. Based ontheconnectionstrength(weights)andtransferfunction,theactivationvaluepassestothenextnode.Each of the nodes sums the activation values that it receives (it calculates the**weighted sum**) andmodifiesthatsumbasedonitstransferfunction.Next,itappliesanactivationfunction.Anactivation functionisafunctionthat’sappliedtothisparticularneuron.Fromthat,theneuronunderstandsifitneedsto passalonga signalornot.

##### Each of the synapses gets assigned weights, which are crucial to **Artificial Neural Networks** (ANNs).Weights are how ANNs learn. By adjusting the weights, the ANN decides to what extent signals getpassedalong. Whenyou’retrainingyour network,you’re decidinghowtheweightsareadjusted.

##### The activation runs through the network until it reaches the output nodes. The output nodes then give usthe information in a way that we can understand. Your network will use a cost function to compare theoutput and the actual expected output. The model performance is evaluated by the cost function. It’sexpressedasthedifferencebetween theactualvalueand the predictedvalue.

##### There are many different cost functions you can use, you’re looking at what the error you have in yournetwork is.You’reworking tominimizeloss function.(Inessence, the lower theloss function, thecloser it is to your desired output). The information goes back, and the neural network begins to learnwiththegoalofminimizingthecostfunctionby tweaking the weights. This process iscalled**backpropagation**.

##### In **forward propagation**, information is entered into the input layer and propagates forward throughthe network to get our output values. We compare the values to our expected results. Next, we calculatethe errors and propagate the info backward. This allows us to train the network and update the weights.(Backpropagationallowsustoadjustalltheweightssimultaneously.)Duringthisprocess,becauseofthe way the algorithm is structured, you’re able to adjust all of the weights simultaneously. This allowsyoutoseewhichpart oftheerroreachof yourweightsin theneural networkisresponsiblefor.

##### **Howdoesanartificialneural networklearn?**

##### There are two different approaches to get a program to do what you want. First, there’s the specificallyguided and hard-programmed approach. You tell the program exactly what you want itto do.Thenthere are **neural networks**. In neural networks, you tell your network the inputs and what you want fortheoutputs,andthen you letitlearn onitsown.

##### By allowing the network to learn on its own, you can avoid the necessity of entering in all of the rules.You can create the architecture and then let it go and learn. Once it’s trained up, you can give it a newimageand itwillbe abletodistinguish output.

##### **Feedforwardandfeedbacknetworks**

##### A **feedforward** network is a network that contains inputs, outputs, and hidden layers. The signals canonly travel in one direction (forward). Input data passes into a layer where calculations are performed.Each processing element computes based upon the weighted sum of its inputs. The new values becomethe new input values that feed the next layer (feed-forward). This continues through all the layers anddeterminestheoutput. Feedforwardnetworksare oftenusedin, forexample,datamining.

##### A**feedbacknetwork**(forexample,arecurrentneuralnetwork)hasfeedbackpaths.Thismeansthatthey can have signals traveling in both directions using loops. All possible connections betweenneurons are allowed. Since loops are present in this type of network, it becomes a non-linear dynamicsystem which changes continuously until it reaches a state of equilibrium. Feedback networks are oftenused in optimization problems where the network looks for the best arrangement of interconnectedfactors.

##### **WeightedSum**

##### Inputs to a neuron can either be features from a training set or outputs from the neurons of a previouslayer. Each connection between two neurons has a unique synapse with a unique weight attached. Ifyou want to get from one neuron to the next, you have to travel along the synapse and pay the “toll”(weight). The neuron then applies an activation function to the sum of the weighted inputs from eachincoming synapse. Itpasses the result onto all the neurons in the next layer.Whenwe talk aboutupdatingweightsin a network, we’retalkingaboutadjustingthe weights onthesesynapses.

##### Aneuron’sinputisthesumofweightedoutputsfromalltheneuronsinthepreviouslayer.Eachinputismultiplied by theweightassociated with the synapse connectingthe input tothecurrentneuron.Ifthere are 3 inputs or neurons in the previous layer, each neuron in the current layer will have 3 distinctweights:oneforeachsynapse.

##### The activation function (or transfer function) translates the input signals to output signals. It maps theoutput values on a range like 0 to 1 or-1 to 1. It’s an abstraction that represents the rate of actionpotential firing in the cell. It’s a number that represents the likelihood that the cell will fire. At it’ssimplest, the function is binary: **yes** (the neuron fires) or **no** (the neuron doesn’t fire). The output can beeither0or1(on/offoryes/no),oritcanbeanywhereinarange.Ifyouwereusingafunctionthatmapsa range between 0 and 1 to determine the likelihood that an image is a cat, for example, an output of 0.9wouldshowa 90% probabilitythatyourimage is, infact, acat.

##### **Activationfunction**

##### Ina nutshell,theactivationfunctionofa nodedefines theoutputofthat node.

##### The activation function (or transfer function) translates the input signals to output signals. It maps theoutput values on a range like 0 to 1 or-1 to 1. It’s an abstraction that represents the rate of actionpotential firing in the cell. It’s a number that represents the likelihood that the cell will fire. At it’ssimplest, the function is binary: **yes** (the neuron fires) or **no** (the neuron doesn’t fire). The output can beeither0 or1 (on/off or yes/no),or itcan beanywherein arange.

##### Whatoptionsdowehave?Therearemanyactivationfunctions,butthesearethefourverycommonones:

###### **Thresholdfunction**

Thisisastepfunction.Ifthesummedvalueoftheinputreachesacertainthresholdthefunctionpasseson0.Ifit’sequaltoormorethanzero,thenitwouldpasson1.It’saveryrigid,straightforward,yesorno function.

###### **Sigmoidfunction**

Thisfunctionisusedinlogisticregression.Unlikethethresholdfunction,it’sasmooth,gradualprogressionfrom0to 1.It’susefulintheoutputlayerandis usedheavilyforlinear regression.

###### **HyperbolicTangentFunction**

Thisfunctionisverysimilartothesigmoidfunction.Butunlikethesigmoidfunctionwhichgoesfrom0 to 1, the value goes below zero, from -1 to 1. Even though this isn’t a lot like what happens in a brain,this function gives better results when it comes to training neural networks. Neural networks sometimesget “stuck” during training with the sigmoid function. This happens when there’s a lot of stronglynegativeinputthatkeepsthe output nearzero,which messeswiththelearningprocess.

###### **Rectifierfunction**

Thismightbethemostpopularactivationfunctionintheuniverseofneuralnetworks.It’sthemostefficient andbiologicallyplausible.Eventhoughit hasa kink,it’ssmoothandgradualafterthekinkat

0. This means, for example, that your output would be either “no” or a percentage of “yes.” Thisfunctiondoesn’trequirenormalization orothercomplicatedcalculations.

The field of artificial intelligence is essential when machines can do tasks that typically require humanintelligence. It comes under the layer of machine learning, where machines can acquire skills and learnfrom pastexperiencewithoutany involvementofhuman.Deeplearningcomesundermachinelearning where artificial neural networks, algorithms inspired by the human brain, learn from largeamounts of data. The concept of deep learning is based on humans’ experiences; the deep learningalgorithm would perform a task continuously so that it can improvetheoutcome.Neural networkshavevarious(deep)layersthatenablelearning.Anydrawbackthatneeds“thought”toworkoutcouldbea drawbackdeep learningcan learntounravel.

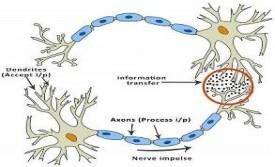
**CHAPTER 4**

**CONVOULUTION NEURAL NETWORKS**

**INTRODUCTIONTOCONVOLUTIONALNEURALNETWORKS(CNN)**

### ArtificialNeural Networks

TheideaofANNsisbasedonthebeliefthatworkingofhumanbrainbymakingtherightconnectionscan be imitated usingsiliconand wiresaslivingneurons anddendrites.



###### Fig:4.1

The human brain is composed of 86 billion nerve cells called neurons. They are connected to otherthousandcellsbyAxons.Stimulifrom externalenvironmentorinputsfrom sensoryorgansareaccepted by dendrites. These inputs create electric impulses, whichquicklytravel through theneural network. A neuron can then send the message to other neuron to handle the issue or does notsenditforward.

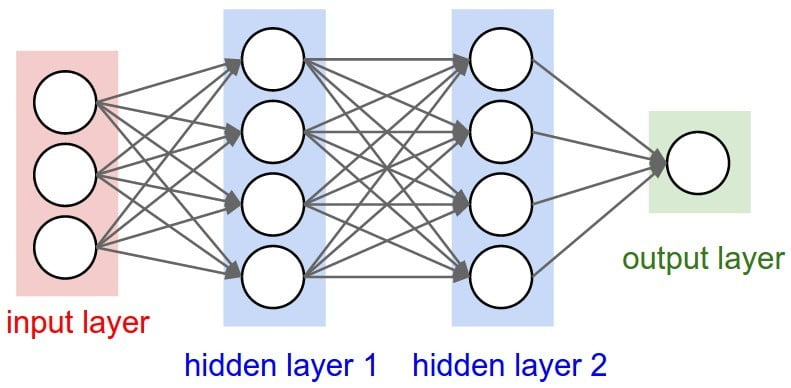
ANNsare composed of multiple nodes, which imitatebiologicalneurons of humanbrain.Theneurons are connected by links and they interact with each other. The nodes can take input data andperform simple operations on the data. The result of these operations is passed to other neurons. Theoutput at each node is called its activation or node value. Each link is associated with weight. ANNsarecapableoflearning, which takesplace byalteringweightvalues.

#### Neuralnetwork:

Aneuralnetwork isanetwork orcircuitofneurons,orinamodernsense, anartificialneuralnetwork, composed of artificial neurons or nodes. Thus a neural networkiseithera biologicalneuralnetwork,madeup ofrealbiologicalneurons,oranartificialneural network, forsolvingartificialintelligence(AI)problem.Theconnectionsofthebiologicalneuronaremodeledasweights. A positive weight reflects an excitatory connection, while negative values mean inhibitoryconnections.Allinputsaremodifiedbyaweightandsummed.Thisactivityisreferredasalinear combination. Finally, an activation function controls the amplitude of the output.

For example, anacceptablerange ofoutputis usuallybetween0 and 1,or itcouldbe -1 and 1.

Theseartificialnetworksmay be used forpredictive modeling, adaptive controland applicationswhere they can be trained via a dataset. Self-learning resulting from experience can occur withinnetworks, whichcanderiveconclusions fromacomplexandseeminglyunrelatedset ofinformation.

****

###### Fig. 4.2 Asimple neuralnetwork

Adeepneuralnetwork(DNN)isanartificialneuralnetwork(ANN)withmultiplelayersbetweenthe input and output layers. The DNN finds the correct mathematical manipulation to turn the inputintotheoutput, whetheritis alinearrelationshipor anon-linearrelationship.

**CONVOLUTIONALNEURAL NETWORKS:**

Convolution Neural Networks are very similar to ordinary Neural Networks from the previouschapter: they are made up of neurons that have learnable weights and biases. Each neuron receivessomeinputs, performsa dotproductand optionallyfollows itwith anon-linearity.

ConvolutionNeuralNetworks(CNNs)areanalogoustotraditionalANNsinthattheyarecomprisedofneuronsthatself-optimizethroughlearning.Eachneuronwillstillreceiveaninputand perform an operation (such as a scalar product followed by a non-linear function) - the basis ofcountless ANNs. From the input rawimagevectors tothefinaloutputof theclassscore, theentireofthenetworkwillstillexpressasingle perceptive score function (the weight). The last layer will contain loss functions associated with theclasses,andalloftheregulartipsandtricksdevelopedfortraditionalANNsstillapply.

TheonlynotabledifferencebetweenCNNsandtraditionalANNsisthatCNNsareprimarilyusedinthe fieldof patternrecognitionwithinimages. This allows us toencodeimage-specific features intothearchitecture,making thenetworkmoresuitedfor image-focusedtasks-whilst further reducingtheparameters requiredtosetup themodel.Oneofthelargestlimitationsoftraditional formsofANNisthattheytendtostrugglewiththecomputationalcomplexityrequiredtocomputeimagedata.CommonmachinelearningbenchmarkingdatasetssuchastheMNISTdatabaseofhandwrittendigitsaresuitableformostformsofANN,duetoitsrelativelysmallimagedimensionalityofjust28×28.Withthisdatasetasingleneuroninthefirsthiddenlayerwillcontain784weights(28×28×1where1bearinmindthatMNISTisnormalizedtojustblackandwhitevalues),which is manageable for most forms ofANN. Ifyou consider a more substantial coloredimage input of 64 × 64, the number of weights on just a single neuron of the first layer increasessubstantiallyto12,288.Alsotakeintoaccountthattodealwiththisscaleofinput,thenetworkwillalso need to be a lot larger than one used to classify colour-normalised MNIST digits, then you willunderstandthedrawbacksofusingsuchmodels.

**CNNARCHITECTURE:**

CNNs are feedforward networks in that information flow takes place in onedirectiononly, fromtheir inputs to their outputs. Just as artificialneural networks (ANN) are biologically inspired, soare CNNs. The visual cortex in the brain, which consists of alternating layersof simple andcomplexcells(Hubel&Wiesel, 1959,1962), motivatestheirarchitecture.

CNN architectures come in several variations; however, ingeneral, they consist of convolutionaland pooling (or subsampling) layers, which are grouped into modules. Either one or more fullyconnected layers, as in a standard feedforward neural network, follow these modules. Modules areoften stacked on top of each other to form a deep model. It illustratestypical CNN architecture for atoyimageclassificationtask.Animageisinputdirectlytothenetwork,andthisisfollowedbyseveral stages of convolution and pooling. Thereafter, representations fromthese operations feedoneormorefullyconnectedlayers.

Finally, the last fully connected layer outputs the class label. Despite thisbeing themost popularbase architecture found in the literature, several architecture changes have been proposed in recentyears with the objective of improving image classification accuracy or reducing computation costs.Althoughfortheremainderofthissection,wemerelyfleetinglyintroducestandardCNNarchitecture.

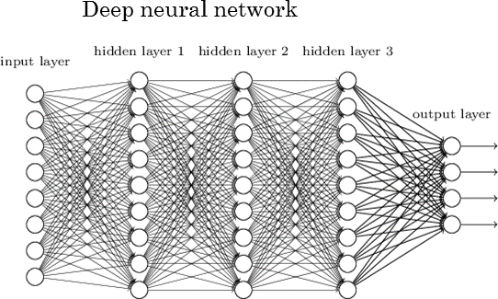


Fig:4.3

**OVERALL ARCHITECTURE:**

CNNs are comprised of three types of layers. These are convolutional layers, pooling layers and fully-connected layers. When these layers are stacked, a CNN architecture has been formed. A simplifiedCNN architecture for MNIST classification is illustrated in Figure 2. input09convolution w/ReLupooling output fully-connected w/ ReLu fully-connected ... Fig. 2: An simple CNN architecture,comprised of just five layers The basic functionality of the example CNN above can be broken downinto four key areas. 1. As found in other forms of ANN, the inputlayerwillhold the pixel values oftheimage. 2. The convolutional layer will determine the output of neuronsof which are connected tolocal regions of the input through the calculation of the scalar product between their weights and theregion connected to the input volume.

The rectified linear unit (commonly shortened to ReLu) aims toapply an ’elementwise’activation function such as sigmoidto the output of the activation producedby the previous layer. 3. The pooling layer will then simply perform downsampling along the spatialdimensionality of the given input, further reducing the number of parameters within that activation. 4.The fully-connected layers willthen perform thesame duties found in standard ANNs and attempt toproduce class scores from the activations, to be used for classification. It is also suggested that ReLumay be used between these layers,asto improve performance.

Through this simple method oftransformation, CNNsare able to transformthe original input layer by layer using convolutional anddownsamplingtechniquestoproduceclassscoresforclassificationandregressionpurposes.However,itisimportanttonotethatsimply understandingtheoverallarchitectureofaCNNarchitecturewillnotsuffice.Thecreationandoptimisation of these models can take quite some time, and can be quite confusing. We will nowexploreindetailtheindividuallayers,detailingtheirhyperparameters andconnectivities.

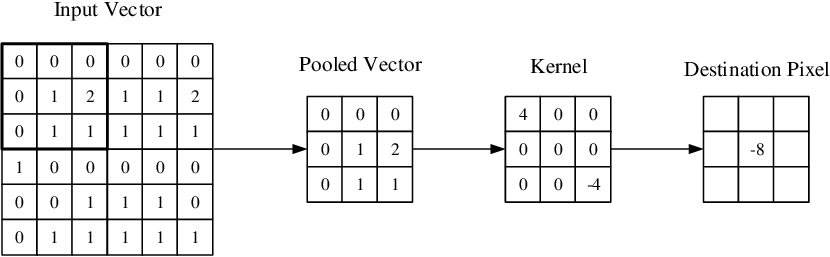
**CONVOLUTIONALLAYERS:**

The convolutional layers serve as feature extractors, andthus theylearn the feature representationsof their input images. The neurons in the convolutional layers are arranged into feature maps. Eachneuron in a feature map has a receptive field, which is connected to a neighborhood of neurons inthe previous layer via a set of trainable weights, sometimes referred to as a filter bank. Inputs areconvolved with the learned weights in order to compute a new featuremap,and the convolvedresultsare sentthrough anonlinear activationfunction.

All neurons within a feature map have weights that are constrained to be equal; however, differentfeaturemapswithinthesameconvolutionallayerhavedifferentweightssothatseveralfeaturescanbeextracted ateach location.

As the name implies, the convolutional layer plays a vital role in how CNNs operate. The layersparametersfocusaroundthe use oflearnable kernels.

These kernels are usually small in spatial dimensionality, but spreads along theentirety of thedepthoftheinput.Whenthedatahitsaconvolutionallayer,thelayerconvolveseachfilteracrossthe spatial dimensionality of theinput to produce a 2D activation map. These activation maps canbevisualised.

Aswe glide through the input, the scalarproductiscalculated foreach value in thatkernel.Fromthisthenetwork will learnkernelsthat ’fire’when they seea specific featureat agivenspatialpositionoftheinput.Theseare commonlyknownasactivations.

###### Fig:4.4Visual representationofaconvolutionallayer**z**

The centre element of the kernel is placed over the input vector, of which is then calculated andreplacedwith a weightedsumof itselfand anynearbypixels. As we alluded to earlier, training ANNs on inputs such as images results in models of which are toobigtotraineffectively.ThiscomesdowntothefullyconnectedmannerofstanardANNneurons,so to mitigate against this every neuron in a convolutional layer is only connected to small region oftheinputvolume.Thedimensionalityofthisregioniscommonlyreferredtoasthereceptivefieldsize of the neuron. The magnitude of the connectivity throughthe depth is nearly always equal tothe depthoftheinput.

Forexample,iftheinputtothenetworkisanimageofsize64×64×3(aRGBcolouredimagewitha dimensionality of 64 × 64) and we set the receptive field size as 6 × 6, we would have a total of108 weights on each neuron within the convolutional layer. (6 × 6 × 3 where 3 is the magnitude ofconnectivity across the depth of the volume) To put this into perspective, a standard neuron seen inotherforms ofANNwouldcontain 12, 288 weightseach.

Wearealsoabletodefinethestrideinwhichwesetthedeptharoundthespatialdimensionalityofthe input in order to place the receptive field. For example if we were to set a stride as 1, then wewouldhaveaheavilyoverlappedreceptivefieldproducingextremelylargeactivations.Alternatively,setting the strideto agreaternumberwillreducetheamountofoverlapping andproduceanoutputof lowerspatialdimensions.

Zero-padding is the simple process of padding the border of the input, and is an effective method togivefurther controlasto thedimensionalityoftheoutputvolumes.

As a result of this as the backpropagation stage occurs, each neuron in the output will represent theoverall gradient of which can be totalled across the depth - thus only updating asingleset ofweights,as opposedto everysingleone.

**PoolingLayers**

The purpose of the pooling layers is to reduce the spatial resolution of the feature maps and thusachievespatialinvariancetoinputdistortionsandtranslations.Initially,itwascommonpracticetouse average pooling aggregation layers to propagate the average of all the input values, of a smallneighbourhoodofanimagetothenextlayer.However,inmorerecentmodels,maxpoolingaggregationlayerspropagatethemaximum valuewithinareceptivefieldtothenextlayer.

Pooling layersaim togradually reduce thedimensionality ofthe representation,andthusfurtherreducethe number ofparameters andthecomputationalcomplexityofthemodel.

The pooling layer operates over each activation map in the input, and scales its dimensionality usingthe“MAX”function.InmostCNNs,these comein theformofmax-poolinglayerswith kernelsof a dimensionality of 2 × 2 applied with a stride of 2 along the spatial dimensions of the input. Thisscales the activation map down to 25% of the original size- whilstmaintaining the depth volume toitsstandardsize.

, having a kernel size above 3 will usually greatly decrease theperformanceof themodel.

It is also important to understand that beyond max-pooling, CNN architectures may contain general-pooling.Generalpoolinglayersarecomprisedofpoolingneuronsthatareabletoperformamultitude of common operations including L1/L2-normalisation,andaveragepooling. However,thistutorialwillprimarilyfocus ontheuseofmax-pooling

**FullyConnectedLayers**

Several convolutional and pooling layers are usually stacked on top of each other to extract moreabstract feature representations in moving through the network. The fullyconnected layers thatfollowtheselayersinterpretthesefeaturerepresentationsandperformthefunctionofhigh-levelreasoning. . For classification problems, it is standard to use the softmax operator on top of a DCNN.Whileearlysuccesswasenjoyedbyusingradialbasisfunctions(RBFs),astheclassifierontopofthe convolutional towers found that replacing the softmax operator with a support vector machine(SVM)leadsto improved classificationaccuracy.

The fully-connected layer contains neurons of which are directly connected to the neurons in the twoadjacent layers, without being connected to any layers within them. This is analogous to way thatneuronsare arranged intraditionalforms ofANN.

DespitetherelativelysmallnumberoflayersrequiredtoformaCNN,thereisnosetwayofformulating a CNN architecture. That being said, it would be idiotic to simply throw a few of layerstogether and expect it to work. Through reading of related literature it is obvious that much like otherformsofANNs,CNNstendtofollowacommonarchitecture.Thiscommonarchitectureisillustratedin Figure 2, where convolutional layers are stacked, followed by pooling layers in a repeated mannerbeforefeedingforwardtofully-connectedlayers. Convolutional Neural Networks differ to other forms of Artifical Neural Network in that instead offocusingontheentiretyoftheproblemdomain,knowledgeaboutthespecifictypeofinputisexploited.Thisinturnallows fora much simpler networkarchitecturetobe setup.

This paper has outlined the basic concepts of Convolutional Neural Networks, explaining the layersrequiredtobuild oneanddetailinghowbesttostructure thenetworkinmostimageanalysistasks.

#### Training

CNNsand ANN in generaluselearningalgorithmsto adjust their free parameters in orderto attainthe desired network output. The most common algorithm used for this purpose is backpropagation.Backpropagationcomputesthegradientofanobjectivefunctiontodeterminehowtoadjustanetwork’s parameters in order to minimize errors that affect performance. A commonly experiencedproblemwithtrainingCNNs,andinparticularDCNNs,isoverfitting,whichispoorperformanceona held-out test set after the network is trained on a small or even large training set. This affects themodel’s ability to generalize on unseen data and is a major challenge for DCNNsthat can beassuagedby regularization.

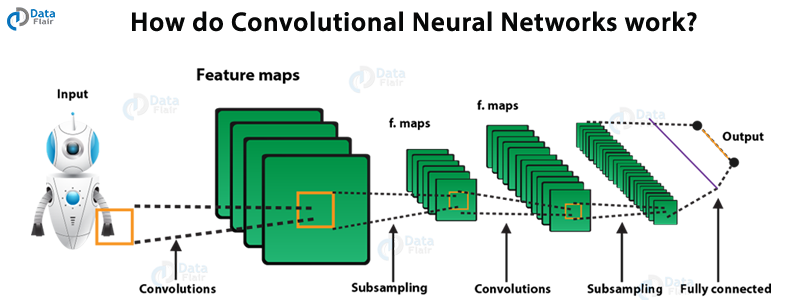


Fig 4.5

**CaffeModel**

CaffeisaframeworkofDeepLearninganditwasmadeusedfortheimplementationandtoaccessthefollowingthings inan objectdetectionsystem.

* Expression:Modelsandoptimizationsaredefinedasplaintextschemasinthecaffemodelunlikeotherswhich use codesforthispurpose.
* Speed: for research and industry alike speed is crucial for state-of-the-art models and massive data
* Modularity:Flexibilityandextensionismajorlyrequiredforthenewtasksanddifferentsettings.
* Openness: Common code, reference models, and reproducibility are the basic requirements ofscientificand appliedprogress.

#### TypesofCaffeModels :

###### **OpenPose**

Thefirst real-timemulti-personsystemisportrayedbyOpenPosewhichcancollectivelysighthumanbody, hand,andfacial keypoints(intotal130keypoints) onsinglepictures.

**FullyConvolutionalNetworksforSemanticSegmentation**

Intheabsolutelyconvolutionalnetworks(FCNs)FullyConvolutionalNetworksarethereferenceimplementationofthemodelsand codeforthe withinthe PAMIFCNand CVPRFCNpapers.

###### **Cnn-vis**

Cnn-visisanopen-sourcetool thatletsyouuseconvolutionalneuralnetworkstogenerateimages.It has taken inspirationfromthe Google'srecentInceptionismblogpost.

###### **SpeechRecognition**

SpeechRecognitionwiththecaffedeeplearningframework.

###### **DeconvNet**

LearningDeconvolutionNetworkforSemanticSegmentation.

###### **CoupledFaceGeneration**

This is the open source repository for the Coupled Generative Adversarial Network (CoupledGAN orCoGAN) work.These models are compatible with Caffe master, unlike earlier FCNs that required a pre-release branch (note: this reference edition of the models remains ongoing and not allof themodelshaveyetbeen portedtomaster).

**CodesforFast ImageRetrieval**

Tocreatethehash-likebinarycodesit provideseffectiveframeworkforfastimage retrieval.

**SegNetandBayesianSegNet**

SegNetisreal-timesemanticsegmentationarchitectureforsceneunderstanding.

**DeepHand**

It givespre-trainedCNNmodels.

**DeepYeast**

Deep Yeast may be an 11-layer convolutional neural network trained on biaural research pictures ofyeast cellscarryingfluorescentproteinswithtotallydifferent subcellularlocalizations.

PythonVSotherlanguagesforObjectDetection:Objectdetectionmaybeadomain-specificvariation of themachine learning predictiondrawback.Intel’sOpenCV library that is implementedin C/C++ has its interfaces offered during a} very vary of programming environments like C#,Matlab,Octave,R,Pythonandthenon. Why Pythoncodesaremuchbetter optionthanotherlanguagecodesfor objectdetectionare more compactand readablecode.

###### Python uses zero-based indexing.

###### Dictionary(hashes)supportprovided.

###### SimpleandelegantObject-orientedprogrammingFreeand open

###### Multiplefunctionscanbepackage inonemodule

MorechoicesingraphicspackagesandtoolsetsSupervisedlearningalsoplaysanimportantrole.

The utility of unsupervised pre-training is usually evaluated on the premise of what performance isachievedwhensupervisedfine-tuning.Thispaperreviewsanddiscussesthefundamentalsof learning as well as supervised learning for classification models, and also talks about the mini batch stochasticgradient descentalgorithmthatisused tofine-tunemanyof the models.

Object Classification in Moving Object Detection Object classification works on the shape, motion,color and texture. The classification can be done under variouscategorieslikeplants, objects,animals, humans etc. The key concept of object classification is tracking objects and analysing theirfeatures.

**Shape-Based**

A mixture of image-based and scene based object parameters such asimageblob (binary largeobject) area, the as pectration of blob bounding box and camera zoom is given as inputto thisdetection system. Classification is performed on the basis of the blob at each and every frame. Theresultsare keptin thehistogram.

**Motion-Based**

When an easy image is given as an input with no objects in motion,this classification isn't required.In general, non- rigid articulated human motion shows a periodicproperty;therefore this has beenused as a powerful clue for classification of moving objects. based on this useful clue,human motionis distinguished from different objects motion. ColorBased- though colorisn'tan applicable livealone for police investigation and following objects, but the low process value of the colour primarilybased algorithms makes the coloura awfully smart feature to be exploited. As an example, the color-histogrambasedtechniqueisemployedfordetectionofvehiclesinperiod.Colorbarchartdescribesthecolourdistributioninaverygiven regionthatispowerfulagainstpartial occlusions.

**Texture-Based**

The texture-based approaches with the assistance of texture pattern recognition work just like motion-based approaches. It provides higher accuracy, by exploitation overlapping native distinction socialcontrol however might need longer, which may be improved exploitation somequicktechniques. I.proposedWORKAuthorshaveappliedperiodobjectdetectionexploitationdeeplearningandOpenCV to figure to work with video streams and video files.

This will be accomplished using thehighly efficient open computer vision. Implementation of proposed strategy includes caffe- model basedonGoogleImageScenery;Caffeoffersthemodeldefinitions,optimizationsettings,pre-trainedweights[4]. Prerequisite includesPython 3.7,OpenCV 4 packagesand numpy to completethis task ofobject detection.NumPy is the elementary package for scientific computing with Python.It containsamong other things: a strong N-dimensional array object, subtle (broadcasting) functions tools forintegratingC/C++ andfortran code,helpfullinearalgebra, Fouriertransform, andrandom

CHAPTER 5

OPEN COMPUTER VISION

##### **5.1INTRODUCTION**

OpenCV stands for Open supply pc Vision Library is associate open supply pc vision and machinelearningsoftwaresystemlibrary.ThepurposeofcreationofOpenCVwastoproduceastandardinfrastructure for computer vision applications and to accelerate the utilization of machine perceptionwithin the business product [6]. It becomes very easy forbusinesses to utilize and modify the codewith OpenCV as it isa BSD-licensedproduct. Itisa richwholesomelibraby asit contains 2500optimizedalgorithms,whichalsoincludesacomprehensivesetofbothclassicandprogressivecomputervisionandmachinelearningalgorithms.Thesealgorithmsisusedforvariousfunctionssuchas discover and acknowledging faces. Identify objects classify human actions. In videos, track cameramovements, track moving objects. Extract 3D models of objects, manufacture 3D purpose clouds fromstereo cameras, sew pictures along to provide a high-resolution image of a complete scene, find similarpictures from a picture information, remove red eyes from images that are clickedwith the flash,followeyemovements,recognizesceneryandestablishmarkerstooverlayitwithaugmentedreality.

Officially launchedin1999theOpenCVprojectwasinitiallyan IntelResearch initiative toadvance CPU-intensive applications, part of a series of projects including real-time ray tracing and 3Ddisplay walls The main contributors to the project included a number of optimization experts in IntelRussia,aswell asIntel'sPerformanceLibrary Team.Intheearly daysof OpenCV, thegoals of theprojectweredescribedas:

* Advance vision research by providing not only open but also optimized code for basic visioninfrastructure.No morereinventingthewheel.
* Disseminate vision knowledge by providing a common infrastructure that developers could buildon,sothatcode would bemore readilyreadable andtransferable.
* Advance vision-based commercial applications by making portable, performance-optimized codeavailable forfree–with alicensethatdidnotrequirecodetobeopenorfreeitself.

The first alpha version of OpenCV was released to the public at the IEEE Conference on ComputerVisionandPatternRecognitionin2000,andfivebetaswerereleasedbetween2001and2005.Thefirst 1.0versionwasreleasedin2006.Aversion1.1"pre-release" was releasedinOctober 2008.

ThesecondmajorreleaseoftheOpenCVwasinOctober2009.OpenCV2includesmajorchangestotheC++interface,aimingateasier,moretype-safepatterns,newfunctions,andbetter

implementationsforexisting ones in termsofperformance(especially onmulti-coresystems).Official releasesnowoccur every six months and development is now done by an independent Russian team supported bycommercialcorporations.

OpenCV (Open source computervision) isa library ofprogramming functionsmainly aimedat real-timecomputervision.OriginallydevelopedbyIntel,itwaslatersupportedbyWillowGaragethenItseez(whichwaslateracquiredbyIntel).Thelibraryis cross-platform andfreefor use undertheopen-source BSDlicense.

It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and MacOS. OpenCV leans mostly towards real-time vision applications andtakes advantageofMMX andSSEinstructionswhenavailable.Afull-featuredCUDAandOpenCLinterfacesarebeingactivelydevelopedrightnow.

There are over 500 algorithms and about 10 times as many functions that compose or support thosealgorithms. OpenCV is written natively in C++ and has a templated interface that works seamlesslywithSTLcontainers.

###### **OpenCV'sapplicationareas include:**

* 2Dand3Dfeaturetoolkits
* Egomotionestimation
* Facialrecognitionsystem
* Gesturerecognition
* Human–computerinteraction(HCI)
* Mobilerobotics
* Motion understanding
* Objectidentification
* Segmentationandrecognition
* Stereopsisstereovision:depthperceptionfrom2cameras
* Structurefrommotion(SFM)
* Motiontracking
* Augmented reality

**To supportsomeoftheaboveareas,OpenCVincludesastatisticalmachinelearning librarythatcontains:**

* BoostingDecisiontreelearning
* Gradientboostingtrees
* Expectation-maximizationalgorithm
* k-nearestneighboralgorithm
* NaiveBayesclassifier
* Artificialneuralnetworks
* Randomforest
* Randomforest
* Support vectormachine(SVM)
* Deepneuralnetworks(DNN)

**Library in Open cv**

###### **Numpy:**

NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open source extensionmoduleforPython,whichprovidesfastprecompiledfunctionsformathematicalandnumericalroutines.Furthermore,NumPyenrichestheprogramminglanguagePythonwithpowerfuldatastructures for efficient computation of multi-dimensional arrays and matrices. The implementationiseven aiming at huge matrices and arrays. Besides that the module supplies a large library of high-levelmathematicalfunctionsto operate on these matricesandarrays.

ItisthefundamentalpackageforscientificcomputingwithPython.Itcontainsvariousfeaturesincludingtheseimportantones:

* ApowerfulN-dimensionalarrayobjectSophisticated(broadcasting)functions
* ToolsforintegratingC/C++andFortrancode
* Usefullinearalgebra,FourierTransform,andrandomnumbercapabilities.

###### **NumpyArray:**

Anumpyarrayisagridofvalues,allofthesametype,andisindexedbyatupleofnonnegativeintegers. The number of dimensions is the rank of the array; the shape of an array is a tuple of integersgivingthe size ofthearrayalongeachdimension.

###### **SciPy:**

SciPy(ScientificPython)isoftenmentionedinthesamebreathwithNumPy.SciPyextendsthecapabilitiesofNumPywithfurtherusefulfunctionsforminimization,regression,Fourier-transformation and manyothers.NumPy isbased ontwo earlierPythonmodules dealing with arrays.One of these is Numeric. Numeric is like NumPy aPython module for high-performance, numericcomputing,butitisobsoletenowadays.AnotherpredecessorofNumPyisNumarray,whichisacompleterewriteofNumericbutisdeprecatedaswell.NumPyisamergerofthosetwo,i.e.itisbuildonthecodeof Numeric andthefeatures ofNumarray.

**ThePythonAlternativeToMatlab:**

Python in combination with Numpy, Scipy and Matplotlib can be used as a replacement for MATLAB.The combination of NumPy, SciPy and Matplotlib is a free (meaning both "free" as in "free beer" and"free"asin"freedom")alternativetoMATLAB.EventhoughMATLABhasahugenumberofadditional toolboxes available, NumPy has the advantage that Python is a more modern and completeprogramming language and - as we have said already before - is open source. SciPy adds even moreMATLAB-like functionalities to Python. Python is rounded out in the direction of MATLAB with themoduleMatplotlib, whichprovidesMATLAB-like plottingfunctionality.

**HaarCascadeClassifierinOpenCv**

The algorithm needs a lot of positive images (images of faces) and negative images (images withoutfaces) to train the classifier. Then we need to extract features from it. For this, haar features shown inbelow imageareused. They are justlikeour convolutionalkernel. Each featureisa singlevalueobtainedbysubtractingsumofpixelsunderwhiterectanglefromsumofpixelsunderblackrectangle.

Now all possible sizes and locations of each kernel is used to calculate plenty of features. (Just imaginehow much computation it needs? Even a 24x24 window results over 160000 features). For each featurecalculation,we needtofindsumofpixels underwhiteand blackrectangles.

To solve this,they introducedtheintegralimages.Itsimplifiescalculationofsumofpixels,howlargemaybethenumberofpixels,toanoperationinvolvingjustfourpixels.Nice,isn‟tit?Itmakesthingssuper-fast.

Butamong allthesefeatureswecalculated,mostof them areirrelevant.For example,considertheimagebelow.Toprowshowstwogoodfeatures.Thefirstfeatureselectedseemstofocusontheproperty that the region of the eyes is often darker than the region of the nose and cheeks. The secondfeature selected relies on the property that the eyes are darker than the bridge of the nose. But the samewindowsapplyingoncheeksoranyotherplaceisirrelevant.Sohowdoweselectthebestfeaturesoutof160000+features? Itisachieved byAdaboost.

Final classifier is a weighted sum of these weak classifiers. It is called weak because it alone can‟tclassify the image, but together with others forms a strong classifier. The paper says even 200 featuresprovide detection with 95% accuracy. Their final setup had around 6000 features. (Imagine a reductionfrom160000+featuresto 6000 features.Thatis a biggain).

Sonow you takean image. Takeeach 24x24 window.Apply 6000 features to it. Check if it is face ornot. Wow..Wow..Isn‟tit alittleinefficient and timeconsuming?Yes,it is. Authorshaveagoodsolutionforthat.

Inanimage,mostoftheimageregionisnon-faceregion.Soitisabetterideatohaveasimplemethodto check if a window is not a face region. If it is not, discard it in a single shot. Don‟t process it again.Instead focus on region where there can be a face. This way, we can find more time to check a possiblefaceregion

ForthistheyintroducedtheconceptofCascadeofClassifiers.Insteadofapplyingallthe6000featuresonawindow,groupthefeaturesintodifferentstagesofclassifiersandapplyone-by-one.(Normallyfirstfewstageswillcontainverylessnumberoffeatures).Ifawindowfailsthefirststage,discardit.Wedon‟tconsiderremainingfeaturesonit.Ifitpasses,applythesecondstageoffeaturesand continuetheprocess.The windowwhich passes all stages is a face region. Haar-likefeatures are digital image features used in object recognition. They owe their name to their intuitivesimilaritywith Haarwavelets and wereused inthefirstreal-timefacedetector.

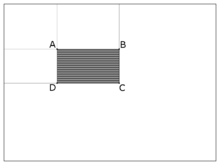
Historically, working with only image intensities (i.e., the RGB pixel values at each and every pixel ofimage)madethetaskoffeaturecalculationcomputationallyexpensive.ApublicationbyPapageorgiouet al. discussed working with an alternate feature set based on Haar wavelets instead of the usual imageintensities. Paul Viola and Michael Jones adapted the idea of using Haar wavelets and developed the so-called Haar-like features. A Haar-like feature considers adjacent rectangular regions at a specific locationin a detection window, sums up the pixel intensities in each region andcalculatesthe differencebetweenthesesums.Thisdifferenceisthenusedtocategorizesubsectionsofanimage.Forexample,with a human face, it is a common observation that among all faces the region of the eyes is darker thanthe region of the cheeks.Therefore, a common Haar feature for face detection is a set oftwo adjacentrectangles that lie above the eye and the cheek region. The position of these rectangles is defined relativetoa detection windowthatactslikea boundingboxtothetargetobject(theface inthis case).

In the detection phase of the Viola–Jones object detection framework, a window of the target size ismovedovertheinputimage,andforeachsubsectionoftheimagetheHaar-likefeatureiscalculated.Thisdifferenceisthencomparedtoalearnedthresholdthatseparatesnon-objectsfromobjects.BecausesuchaHaar-likefeatureisonlyaweaklearnerorclassifier(itsdetectionqualityisslightlybetterthanrandomguessing)alargenumberofHaar-likefeaturesarenecessarytodescribeanobjectwith sufficientaccuracy. In theViola–Jonesobject detection framework, the Haar-likefeaturesarethereforeorganizedinsomethingcalledaclassifiercascadetoformastronglearnerorclassifier.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to theuseofintegralimages,aHaar-likefeatureofanysizecanbecalculatedinconstanttime(approximately60 microprocessorinstructionsfor a2-rectanglefeature).

#### RectangularHaar-likefeatures

AsimplerectangularHaar-likefeaturecanbedefinedasthedifferenceofthesumofpixelsofareasinside the rectangle, which can be at any position and scale within the original image. This modifiedfeaturesetiscalled2-rectanglefeature.ViolaandJonesalsodefined3-rectanglefeaturesand4-rectangle features. The values indicate certain characteristics of a particular area of the image. Eachfeature type can indicate the existence (or absence) of certain characteristicsintheimage, such asedgesor changesin texture.For example, a2-rectanglefeature can indicate where theborder liesbetweena darkregion anda lightregion



Sum=I(C)+I(A) –I(B)–I(D)

###### Fig: 5.12-Rectanglefeature

###### OpenCV has a modular structure, which means that the package includes several shared or static libraries. The following modules are available:

* [**Core functionality**](https://docs.opencv.org/master/d0/de1/group__core.html) (**core**) - a compact module defining basic data structures, including the dense multi-dimensional array Mat and basic functions used by all other modules.
* [**Image Processing**](https://docs.opencv.org/master/d7/dbd/group__imgproc.html) (**imgproc**) - an image processing module that includes linear and non-linear image filtering, geometrical image transformations (resize, affine and perspective warping, generic table-based remapping), color space conversion, histograms, and so on.
* [**Video Analysis**](https://docs.opencv.org/master/d7/de9/group__video.html) (**video**) - a video analysis module that includes motion estimation, background subtraction, and object tracking algorithms.
* [**Camera Calibration and 3D Reconstruction**](https://docs.opencv.org/master/d9/d0c/group__calib3d.html) (**calib3d**) - basic multiple-view geometry algorithms, single and stereo camera calibration, object pose estimation, stereo correspondence algorithms, and elements of 3D reconstruction.
* [**2D Features Framework**](https://docs.opencv.org/master/da/d9b/group__features2d.html) (**features2d**) - salient feature detectors, descriptors, and descriptor matchers.
* [**Object Detection**](https://docs.opencv.org/master/d5/d54/group__objdetect.html) (**objdetect**) - detection of objects and instances of the predefined classes (for example, faces, eyes, mugs, people, cars, and so on).
* [**High-level GUI**](https://docs.opencv.org/master/d7/dfc/group__highgui.html) (**highgui**) - an easy-to-use interface to simple UI capabilities.
* [**Video I/O**](https://docs.opencv.org/master/dd/de7/group__videoio.html) (**videoio**) - an easy-to-use interface to video capturing and video codecs.
* some other helper modules, such as FLANN and Google test wrappers, Python bindings, and others.

**CHAPTER 6**

**RESULT AND DISCUSSION**

### INTRODUCTIONTOIMPLEMENTATIONOFPROBLEM

#### TheModel

Deep learning is a popular technique used in computer vision. We chose Convolutional Neural Network(CNN) layers as building blocks to create our model architecture. CNNs are known to imitate how thehumanbrain works when analyzingvisuals.

A typical architectureof aconvolutionalneuralnetwork containan inputlayer,someconvolutionallayers, some dense layers (aka. fully-connected layers), and an output layer . These are linearly stackedlayersorderedinsequence.

**Input Layer**

The input layerhas pre-determined, fixeddimensions, so the imagemustbe pre-processedbefore it canbefedintothelayer.WeusedOpenCV,acomputervisionlibrary,forobjectdetectioninthevideo.

The OpenCV contains pre-trained filters and uses Adaboost to quickly find and crop the object. Thecropped object is then converted into gray scale using cv2.cvtColor and resized to 48-by-48 pixels withcv2.resize. This step greatly reduces the dimensionscompared to the original RGB format with threecolourdimensions (3, 48, 48). The pipeline ensuresevery imagecan be fed into the input layer as a(1,48, 48)numpyarray.

**ConvolutionalLayers**

The numpy array gets passed into the Convolution2D layer where we specify the number of filters asoneofthehyperparameters.Thesetoffiltersareuniquewithrandomlygeneratedweights.Eachfilter,(3,3)receptivefield,slidesacrosstheoriginalimagewithsharedweightstocreateafeaturemap.

Convolution generates feature maps that represent how pixel values are enhanced, for example,edgeandpatterndetection.Afeaturemapiscreatedbyapplyingfilter1acrosstheentireimage.Otherfiltersareappliedone afteranother creatinga setof featuremaps.

Pooling is a dimension reduction technique usually applied afterone or several convolutional layers. ItisanimportantstepwhenbuildingCNNsasaddingmoreconvolutionallayerscangreatlyaffectcomputational time. We used a popular pooling method called MaxPooling2D that uses (2, 2) windowsacross the feature map only keeping the maximum pixel value. The pooled pixels form an image withdimentionsreduced by4.

**DenseLayers**

The dense layer (aka fully connected layers), is inspired by the way neurons transmit signals through thebrain. It takes a large number of input features and transform features through layers connected withtrainableweights.

These weights are trained by forward propagation of training datathen backward propagation of itserrors.Backpropagationstartsfromevaluatingthedifferencebetweenpredictionandtruevalue,andback calculatestheweightadjustmentneeded toevery layerbefore.We can control the training speedand the complexity of the architecture by tuning the hyper-parameters, suchaslearning rate andnetwork density. As we feed in more data, the network isabletograduallymake adjustments untilerrors are minimized. Essentially, the more layers/nodes we add to the network the better it can pick upsignals.

Asgood as it may sound, the model also becomes increasingly prone to overfitting the training data.One method to prevent overfitting and generalize on unseen data is to apply dropout. Dropout randomlyselectsa portion (usually less than 50%) of nodes tosettheirweights to zeroduring training. Thismethod can effectively control the model's sensitivity to noise during training while maintaining thenecessarycomplexityofthearchitecture.

**Outputlayer**

The output layer in a CNN as mentioned previously is a fully connected layer, where the input from theotherlayersisflattenedandsentsoasthetransformtheoutputintothenumberofclassesasdesiredbythenetwork.

## RESULTS

## Input Output

## stop sign.jpgimage.jpg

## football.jpg

## Football.jpg





**CONCLUSION:**

Deep learning based Real time object detection has been a research hotspot in recent years. This project starts ongenericobjectdetectionpipelineswhichprovidebasearchitecturesforotherrelatedtasks.Withthehelpofthis thethree other commontasks,namely object detection, facedetection andpedestriandetection,can be accomplished. Authors accomplished this by combing two things: Object detection with deeplearning and OpenCV and Efficient, threaded video streams with OpenCV. The camera sensor noise andlightening conditioncanchangetheresultas itcancreateproblem in recognizing the object. The endresultisa deep learning-based objectdetector thatcanprocessaround 6-8FPS.

REFERENCES:-

1. Geethapriya S, N. Duraimurugan, S.P. Chokkalingam, “Real-Time Object Detection with Yolo”, International Journal of Engineering and Advanced Technology (IJEAT)

2. Abdul Vahab, Maruti S Naik, Prasanna G Raikar an Prasad S R4, “Applications of Object Detection System”, International Research Journal of Engineering and Technology (IRJET)

3. Hammad Naeem, Jawad Ahmad and Muhammad Tayyab, “Real-Time Object Detection and Tracking”,IEEE .

4. Meera M K, & Shajee Mohan B S. 2016, "Object recognition in images", International Conference on Information60 Science (ICIS)