FACE RECOGNITION BASED ATTENDANCE SYSTEM USING CONVOLUTION NEURAL NETWORK ALGORITHM WITH SAFETY MEASURES

A PROJECT REPORT

Submitted by

SHUBHAM YADAV R	**111517105088**
SHUSHIL GIRISH	**111517105089**
SUBASH KUMAR M	**111517105095**
VIGNESH V	**111517105104**

in partial fulfilment for the award of

the degree of

BACHELOR OF ENGINEERING

in

ELECTRICAL AND ELECTRONICS ENGINEERING

RMD ENGINEERING COLLEGE, KAVARAIPETTAI

ANNA UNIVERSITY: CHENNAI 600 025

SEPTEMBER 2020

CHARACTERISTICS OF COMPONENTS

TEMPERATURE SENSOR:

Temperature is the most-measured process variable in industrial automation. Most commonly, a temperature sensor is used to convert temperature value to an electrical value. Temperature Sensors are the key to read temperatures correctly and to control temperature in industrial applications. A large distinction can be made between temperature sensor types. Sensors differ a lot in properties such as contact-way, temperature range, calibrating method and sensing element. The temperature sensors contain a sensing element enclosed in housings of plastic or metal. With the help of conditioning circuits, the sensor will reflect the change of environmental temperature.

In the temperature functional module we developed, we use the LM34 series of temperature sensors. The LM34 series are precision integrated-circuit temperature sensors, whose output voltage is linearly proportional to the Fahrenheit temperature. The LM34 thus has an advantage over linear temperature sensors calibrated in degrees Kelvin, as the user is not required to subtract a large constant voltage from its output to obtain convenient Fahrenheit scaling. The LM34 does not require any external calibration.

It is easy to include the LM34 series in a temperature measuring application. The LM34 series is available packaged in hermetic TO-46 transistor packages, while the LM34C,LM34CA and LM34D are also available in the plastic TO-92 transistor package.

DESCRIPTION OF TEMPERATURE SENSOR FUNCTIONAL MODULE

The temperature sensor functional module consists of two parts: the function module box and the probe head. The LM34 temperature sensor is mounted on the probe head. By replacing the LM34 with another precision integrated-circuit temperature sensor LM35 , we can easily get an output voltage proportional to the centigrade temperature.

Pin Configuration:

Pin Number	Pin Name	Description
1	Vcc	Input voltage is +5V for typical applications
2	Analog Out	There will be increase in 10mV for raise of every 1°C. Can range from -1V(-55°C) to 6V(150°C)
3	Ground	Connected to ground of circuit

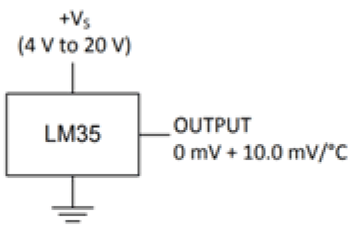
LM35 Regulator Features:

- Minimum and Maximum Input Voltage is 35V and -2V respectively. Typically 5V.
- Can measure temperature ranging from -55°C to 150°C
- Output voltage is directly proportional (Linear) to temperature (i.e.) there will be a rise of 10mV (0.01V) for every 1°C rise in temperature.
- ±0.5°C Accuracy
- Drain current is less than 60uA
- Low cost temperature sensor
- Small and hence suitable for remote applications
- Available in TO-92, TO-220, TO-CAN and SOIC package

How to use LM35 Temperature Sensor:

LM35 is a precession Integrated circuit Temperature sensor, whose output voltage varies, based on the temperature around it. It is a small and cheap IC which can be used to measure temperature anywhere between -55°C to 150°C. It can easily be interfaced with any Microcontroller that has ADC function or any development platform like Arduino.

Power the IC by applying a regulated voltage like +5V (Vs) to the input pin and connect the ground pin to the ground of the circuit. Now, you can measure the temperature in the form of voltage as shown below.



If the temperature is 0°C, then the output voltage will also be 0V. There will be a rise of 0.01V (10mV) for every degree Celsius rise in temperature. The voltage can be converted into temperature using the below formulae.

V_{OUT} = 10 mv/°C × T

where

- V_{OUT} is the LM35 output voltage
- T is the temperature in °C

LM35 Temperature Sensor Applications:

- Measuring temperature of a particular environment
- Providing thermal shutdown for a circuit/component
- Monitoring Battery Temperature
- Measuring Temperatures for HVAC applications.

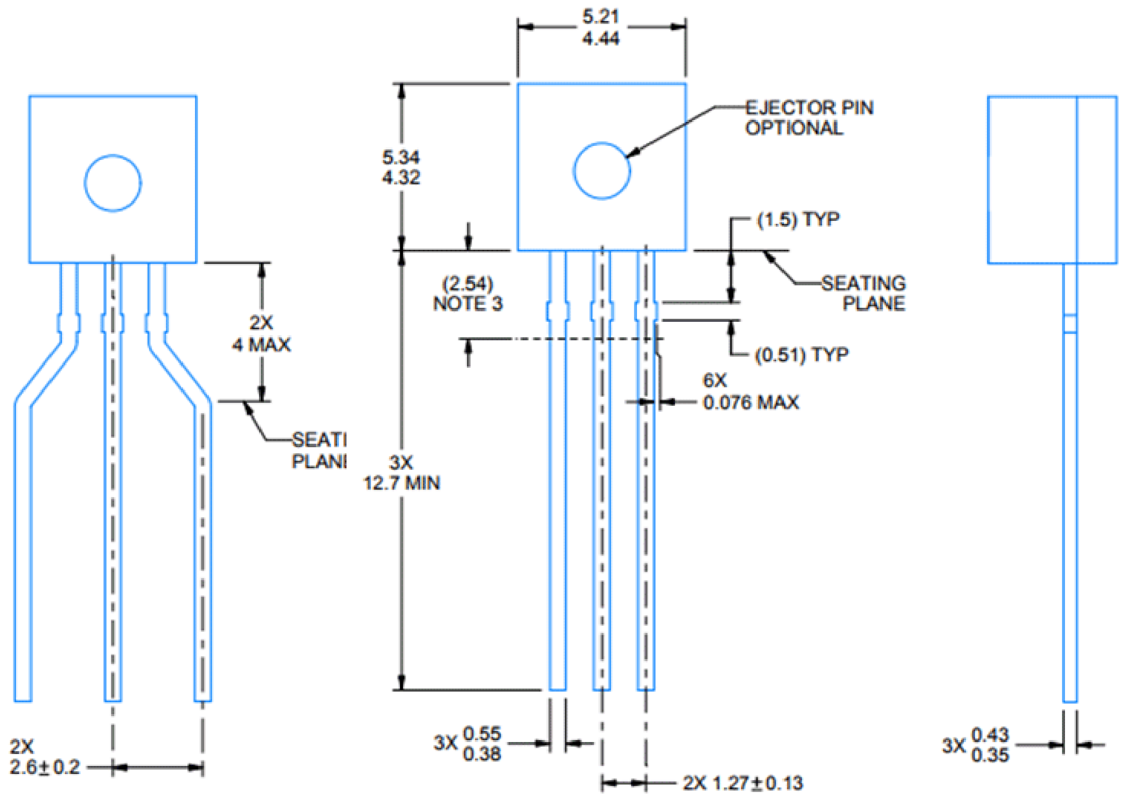
Alternative to LM35 Temperature Sensor

If you are interested in getting reading in degree Fahrenheit you can use **LM34 Temperature sensor**. It is the same as LM35 except its electrical output is proportional to Degree Fahrenheit. It has the same pin configuration and same voltage range as LM35.

You can also check out MCP9700 Temperature Sensor from microchip. MCP9700 can accurately measure temperature from -40C to +150C. The output of the MCP9700 is calibrated to a slope of 10mV/°C and has a DC offset of 500mV. The offset allows reading negative temperatures without the need for a negative supply. Basically the code for LM35 will not work for MCP9700.

If you are using a 8 bit ADC and want to measure 1°C change in temperature then you can also use MCP9701 Temperature Sensor from Microchip. The MCP9701 can accurately measure temperature from -10°C to +125°C. The output of the MCP9701 is calibrated to a slope of 19.53 mV/°C and has a DC offset of 400 mV. The offset allows reading negative temperatures without the need for a negative supply.

2D model of the component (TO-92):



5.21
4.44
EJECTOR PIN
OPTIONAL
5.34
4.32
(1.5) TYP
(2.54)
NOTE 3
SEATING
PLANE
(0.51) TYP
2X
4 MAX
6X
0.076 MAX
SEATI
PLANI
3X
12.7 MIN
2X
2.6±0.2
3X 0.55
0.38
2X 1.27±0.13
3X 0.43
0.35

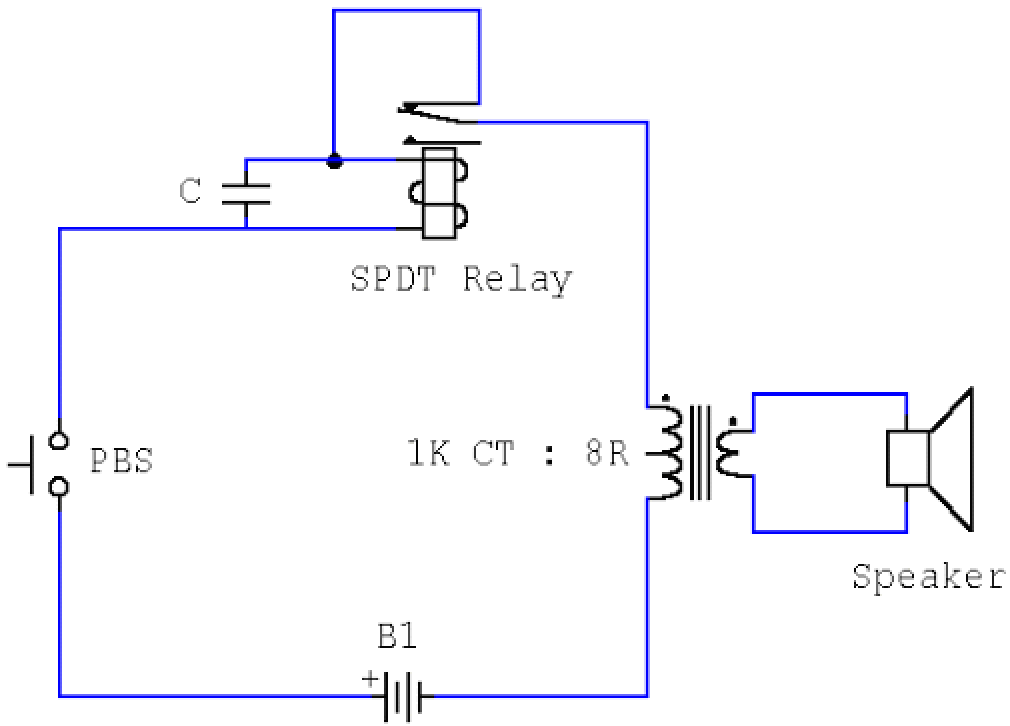
BUZZER:

A buzzer is a mechanical, electromechanical, magnetic, electromagnetic, electro-acoustic or piezoelectric audio signaling device. A Piezoelectric buzzer can be driven by an oscillating electronic circuit or other audio signal source. A click, beep or ring can indicate that a button has been pressed.

A buzzer takes some sort of input and emits a sound in response to it. They may use various means to produce the sound; everything from metal clappers to electromechanical devices. A buzzer needs to have some way of taking in energy and converting it to acoustic energy. Many buzzers are part of a larger circuit and take their power directly from the device’s power source. In other cases, however, the buzzer may be battery powered so that it will go off in the event of a mains outage.

A buzzer or beeper is a signaling device, The word "buzzer" comes from the rasping noise that buzzers made when they were electromechanical devices, operated from stepped-down AC line voltage at 50 or 60 cycles. Other sounds commonly used to indicate that a button has been pressed are a ring or a beep

BUZZER CIRCUIT



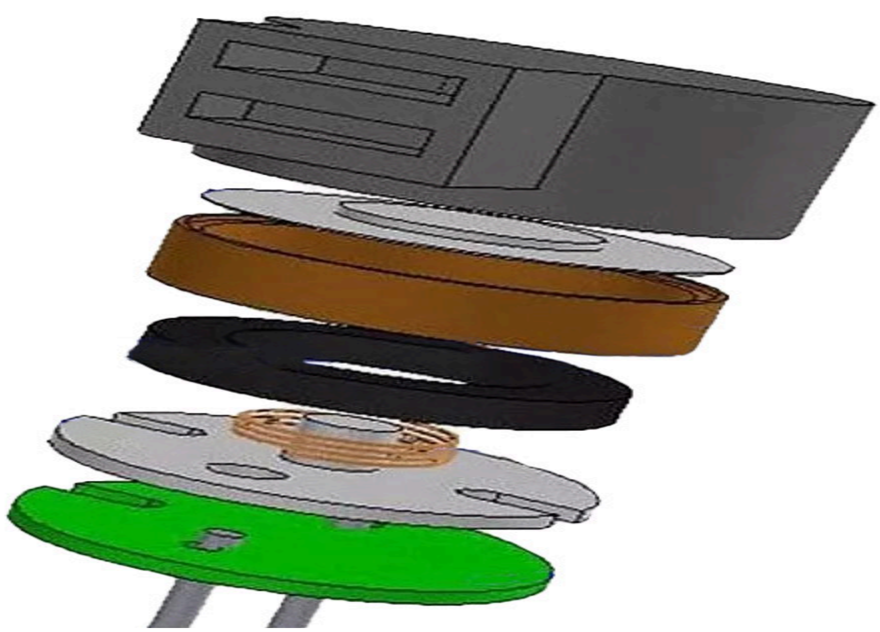
This novel buzzer circuit uses a relay in series with a small audio transformer and speaker. When the switch is pressed, the relay will operate via the transformer primary and closed relay contact. As soon as the relay operates the normally closed contact will open, removing power from the relay, the contacts close and the sequence repeats, all very quickly so fast that the pulse of current causes fluctuations in the transformer primary, and hence secondary.

The speaker's tone is thus proportional to relay operating frequency. The capacitor C can be used to "tune" the note. The nominal value is 0.001uF, increasing capacitance lowers the buzzer's tone.

TYPES OF BUZZER

ELECTROMECHANICAL

Early devices were based on an electromechanical system identical to an electric bell without the metal gong. Similarly, a relay may be connected to interrupt its own actuating current, causing the contacts to buzz. Often these units were anchored to a wall or ceiling to use it as a sounding board. The word "buzzer" comes from the rasping noise that electromechanical buzzers made.



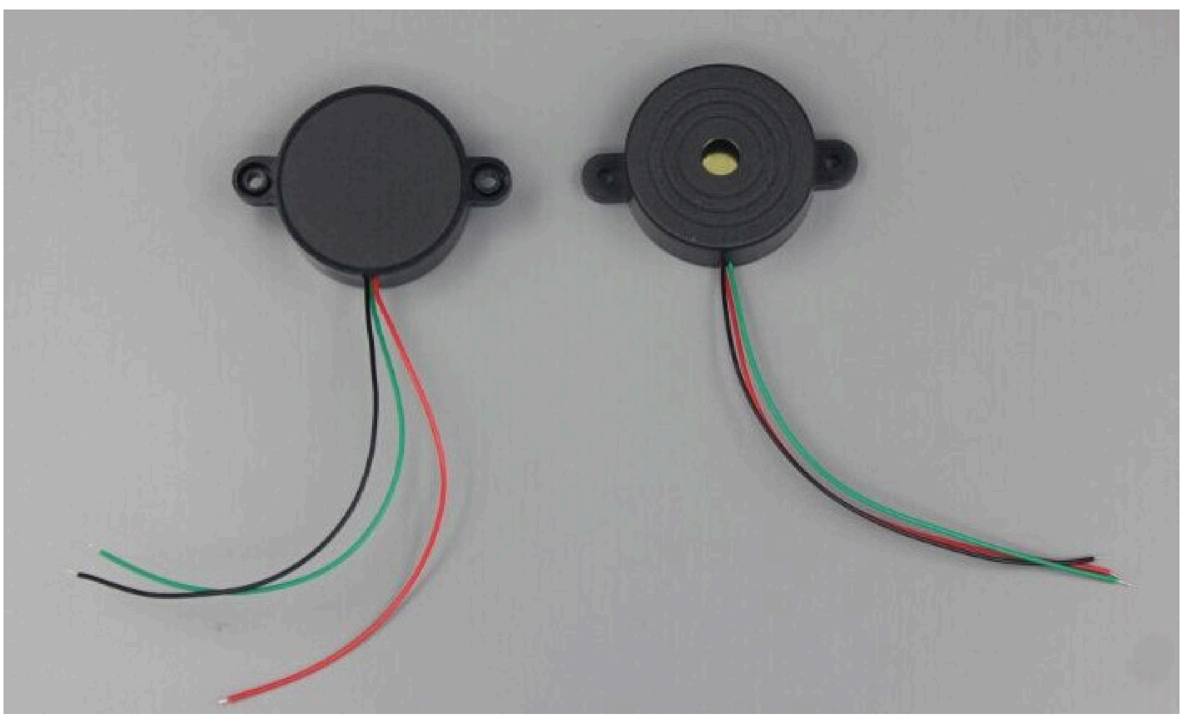
MECHANICAL

A joy buzzer is an example of a purely mechanical buzzer. They require drivers.

PIEZOELECTRIC

A piezoelectric element may be driven by an oscillating electronic circuit or other audio signal source, driven with a piezoelectric audio amplifier. Sounds commonly used to indicate that a button has been pressed are a click, a ring or a beep.

Buzzers are typically used for identification and alarm purposes across many major industries.



INDUSTRIES SERVED

- Safety and Security

ø Automotive Electronics

ø Office Automation

ø Medical Equipment

ø Industrial

ø Consumer Electronics

PIEZO VS. MAGNETIC BUZZERS

CUI's buzzer line utilizes two main technologies. Each technology has specific advantages and tradeoffs that must be taken into consideration depending on the application requirements.

PIEZO BUZZER CHARACTERISTICS

ø Wide operating voltage: 3~250V

ø Lower current consumption: less than 30mA

ø Higher rated frequency

ø Larger footprint

ø Higher sound pressure level

MAGNETIC BUZZER CHARACTERISTICS

ø Narrow operating voltage: 1~16V

ø Higher current consumption: 30~100mA

ø Lower rated frequency

ø Smaller footprint

ø Lower sound pressure level

BENEFITS OF BUZZER

Ø The use of warning systems like delta-larm or electric buzzers could be very beneficial in minimizing loss of lives during a disaster or accident. They are important devices in any building or facilities to alert and notify people if a timely evacuation is necessary.

Ø Specialized electric alarm systems could act as a warning about threatening liquid level conditions in lift pump chambers, sewage, and other non-potable water applications. It is essential that you know how efficiently your home is or business site's sewer system works and this can be gauged more accurately if it is accompanied by an electric sewage alarm.

Ø Most electromechanical buzzers are easy to set up. In fact, you don't need to hire an electrician to install it, since no hard wiring is usually needed. This means cutting down on expenditure for hiring a professional installer.

Ø In the workplace, electric buzzers, especially those with timing software, offer more benefits and features than traditional timers and expensive bells. They could be synchronized with automatic software via computer to control and switch times or channels within the day.

Ø Electric buzzers could even be used in a wide array of appliances. Some buzzers are now used in switching sounds of electric home appliances such as microwaves, washing machines, calculators, smoke detectors, and telephone ringers, transmitters, and many more.

Ø With the help of electro-mechanical buzzers, you can be notified of an automobile entering your residence or business when you're in remote areas, so you never miss a warning.

MODERN APPLICATION

Ø Novelty uses

Ø Judging Panels

Ø Educational purposes

Ø Annunciator panels

Ø Electronic metronomes

Ø Game show lock-out device

Ø Microwave ovens and other household appliances

Ø Sporting events such as basketball games

Ø Electrical alarms

Ø Joy buzzer- a mechanical buzzer used for pranks

ACTUATOR:

An actuator is a component of a machine that is responsible for moving and controlling a mechanism or system, for example by opening a valve. In simple terms, it is a "mover".

An actuator requires a control signal and a source of energy. The control signal is relatively low energy and may be electric voltage or current, pneumatic, or hydraulic fluid pressure, or even human power. Its main energy source may be an electric current, hydraulic pressure, or pneumatic pressure. When it receives a control signal, an actuator responds by converting the source's energy into mechanical motion. In the *electric*, *hydraulic*, and *pneumatic* sense, it is a form of automation or automatic control.

An actuator is a mechanism by which a control system acts upon to perform an operation or task. The control system can be simple (a fixed mechanical or electronic system), software-based (e.g. a printer driver, robot control system), a human, or any other input.

Types of Actuator:

Hydraulic:

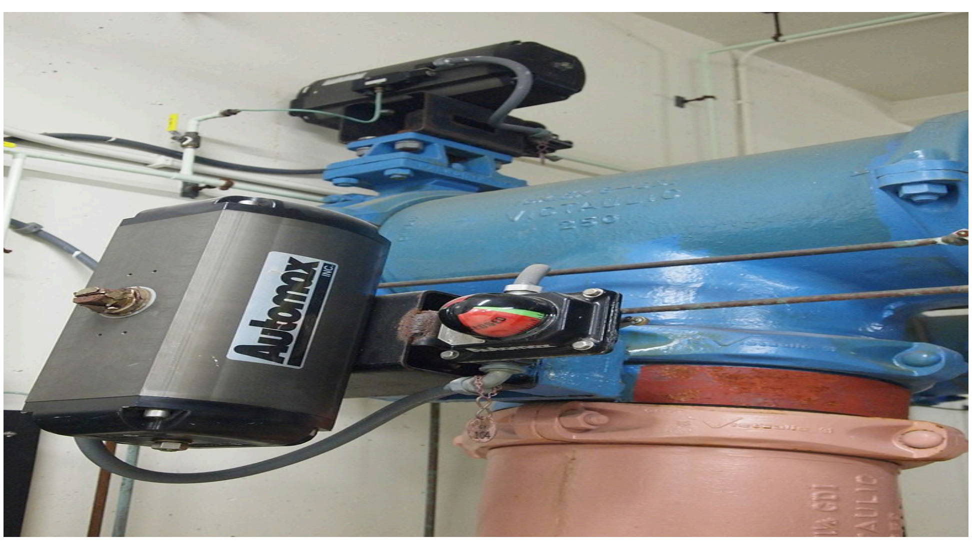
The hydraulic actuator consists of a cylinder or fluid motor that uses hydraulic power to facilitate mechanical operation. The mechanical motion gives an output in terms of linear, rotatory or oscillatory motion. As liquids are nearly impossible to compress, a hydraulic actuator can exert a large force. The drawback of this approach is its limited acceleration.

The hydraulic cylinder consists of a hollow cylindrical tube along which a piston can slide. The term *single acting* is used when the fluid pressure is applied to just one side of the piston. The piston can move in only one direction, a spring being frequently used to give the piston a return stroke. The term *double acting* is used when pressure is applied on each side of the piston; any difference in force between the two sides of the piston moves the piston to one side or the other.

Pneumatic

Pneumatic actuators enable considerable forces to be produced from relatively small pressure changes. Pneumatic energy is desirable for main engine controls because it can quickly respond in starting and stopping as the power source does not need to be stored in reserve for operation. Moreover, pneumatic actuators are cheaper, and often more powerful than other actuators. These forces are often used with valves to move diaphragms to affect the flow of air through the valve.

The advantage of pneumatic actuators consists exactly in the high level of force available in a relatively small volume. While the main drawback of the technology consists in the need for a compressed air network composed of several components such as compressors, reservoirs, filters, dryers, air treatment subsystems, valves, tubes, etc. which makes the technology energy inefficient with energy losses that can sum up to 95%.

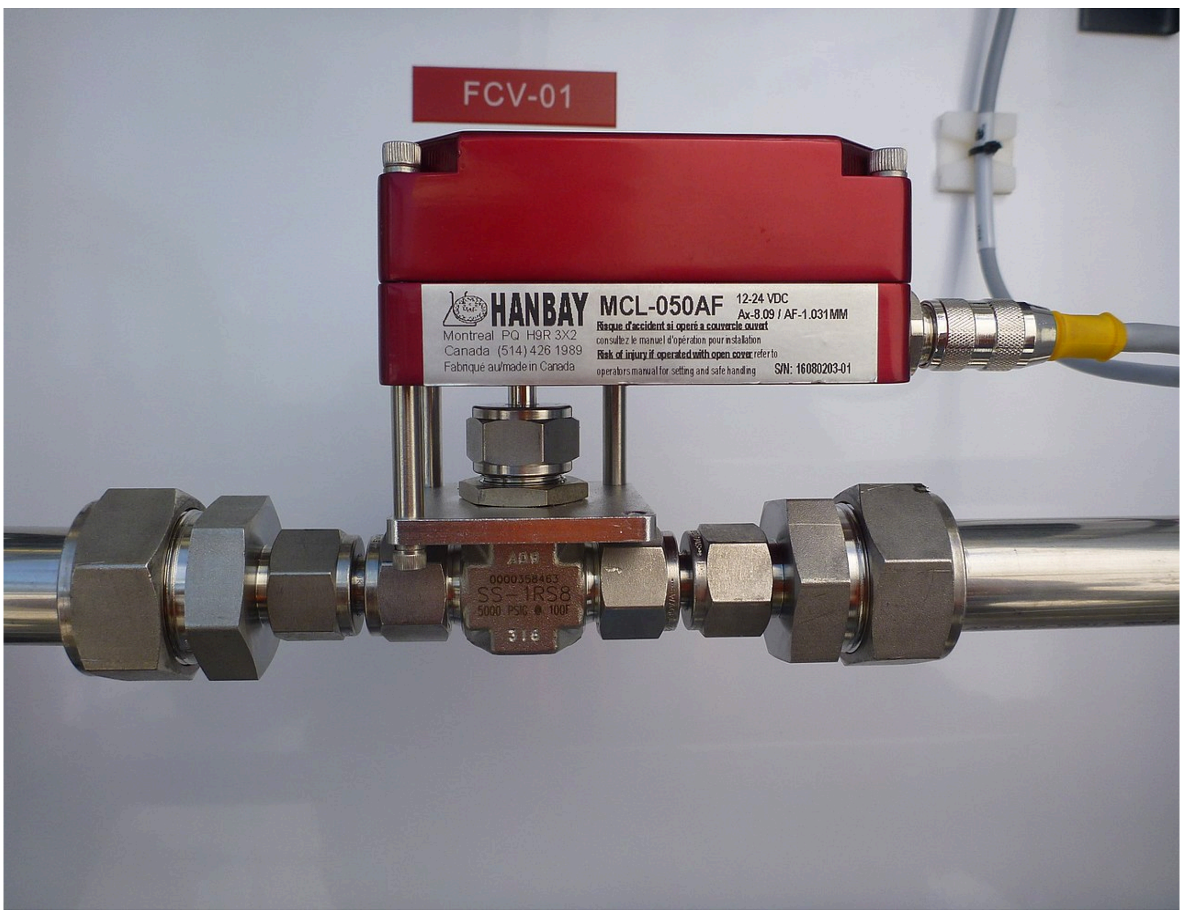


Electromechanical Actuator:

It converts the rotational force of an electric rotary motor into a linear movement to generate the requested linear movement through a mechanism either a belt (Belt Drive axis with stepper or servo) or a screw (either a ball or a lead screw or planetary mechanics)

The main advantages of electromechanical actuators are their relatively good level of accuracy with respect to pneumatics, their possible long lifecycle and the little maintenance effort required (might require grease). It is possible to reach relatively high force, on the order of 100 kN.

The main limitation of these actuators are the reachable speed, the important dimensions and weight they require.



Electro Hydraulic Actuator:

Another approach is an electrohydraulic actuator, where the electric motor remains the prime mover but provides torque to operate a hydraulic accumulator that is then used to transmit actuation force in much the same way that diesel engine/hydraulics are typically used in heavy equipment.

Electrical energy is used to actuate equipment such as multi-turn valves, or electric-powered construction and excavation equipment.

When used to control the flow of fluid through a valve, a brake is typically installed above the motor to prevent the fluid pressure from forcing open the valve. If no brake is installed, the actuator gets activated to reclose the valve, which is slowly forced open again. This sets up an oscillation (open, close, open ...) and the motor and actuator will eventually become damaged.

3D Printed Soft Actuator:

The majority of the existing soft actuators are fabricated using multistep low yield processes such as micro-moulding, solid freeform and mask lithography. However, these methods require manual fabrication of devices, post processing/assembly, and lengthy iterations until maturity in the fabrication is achieved. To avoid the tedious and time-consuming aspects of the current fabrication processes, researchers are exploring an appropriate manufacturing approach for effective fabrication of soft actuators. Therefore, special soft systems that can be fabricated in a single step by rapid prototyping methods, such as 3D printing, are utilized to narrow the gap between the design and implementation of soft actuators, making the process faster, less expensive, and simpler. They also enable incorporation of all actuator components into a single structure eliminating the need to use external joints, adhesives, and fasteners.

Shape memory polymer (SMP) actuators are the most similar to our muscles, providing a response to a range of stimuli such as light, electrical, magnetic, heat, pH, and moisture changes. They have some deficiencies including fatigue and high response time that have been improved through the introduction of smart materials and combination of different materials by means of advanced fabrication technology. The advent of 3D printers has made a new pathway for fabricating low-cost and fast response SMP actuators. The process of receiving external stimuli like heat, moisture, electrical input, light or magnetic field by SMP is referred to as shape memory effect (SME). SMP exhibits some rewarding features such as low density, high strain recovery, biocompatibility, and biodegradability.

Photopolymer/light activated polymers (LAP) are another type of SMP that are activated by light stimuli. The LAP actuators can be controlled remotely with instant response and, without any physical contact, only with the variation of light frequency or intensity.

A need for soft, lightweight and biocompatible soft actuators in soft robotics has influenced researchers for devising pneumatic soft actuators because of their intrinsic compliance nature and ability to produce muscle tension.

Polymers such as dielectric elastomers (DE), ionic polymer metal composites (IPMC), ionic electroactive polymers, polyelectrolyte gels, and gel-metal composites are common materials to form 3D layered structures that can be tailored to work as soft actuators. EAP actuators are categorized as 3D printed soft actuators that respond to electrical excitation as deformation in their shape.

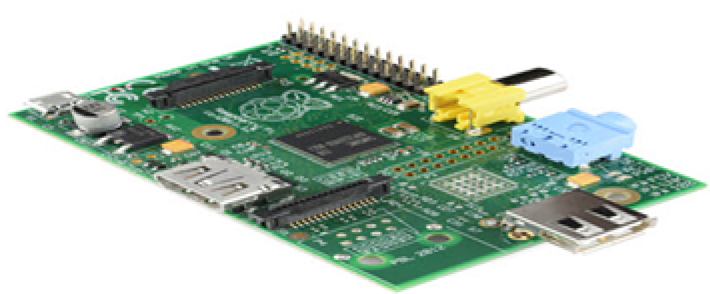
SOFTWARE IMPLEMENTATION

RASPBERRY PI

The Raspberry pi is a single computer board with credit card size, that can be used for many tasks that your computer does, like games, word processing, spreadsheets and also to play HD video. It was established by the Raspberry pi foundation from the UK. It has been ready for public consumption since 2012 with the idea of making a low-cost educational microcomputer for students and children. The main purpose of designing the raspberry pi board is, to encourage learning, experimentation and innovation for school level students. The raspberry pi board is a portable and low cost. Maximum of the raspberry pi computers are used in mobile phones. In the 2st century, the growth of mobile computing technologies is very high, a huge segment of this being driven by the mobile industries. 98% of the mobile phones were using ARM technology.

The raspberry pi comes in two models , they are model A and model B. The main difference between model A and model B is the USB port. Model board will consume less power and that does not include an Ethernet port. But, the model B board includes an Ethernet port and is designed in china.

RASPBERRY PI HARDWARE SPECIFICATIONS



The raspberry pi comes with a set of open source technologies, i.e. communication and multimedia web technologies. In the year 2014, the foundation of the raspberry pi board launched the computer module that packages a model B raspberry pi board into modules for use as a part of embedded systems, to encourage their use.

MEMORY

The raspberry pi model aboard is designed with 256MB of SDRAM and model B is designed with 51MB.Raspberry pi is a small size PC compared with other PCs. The normal PCs RAM memory is available in gigabytes. But in raspberry pi board, the RAM memory is available more than 256MB or 512MB

CPU (CENTRAL PROCESSING UNIT)

The Central processing unit is the brain of the raspberry pi board and that is responsible for carrying out the instructions of the computer through logical and mathematical operations. The raspberry pi uses ARM11 series processor, which has joined the ranks of the Samsung galaxy phone.

GPU (GRAPHICS PROCESSING UNIT)

The GPU is a specialized chip in the raspberry pi board and that is designed to speed up the operation of image calculations. This board designed with a Broadcom video core IV and it supports OpenGL

ETHERNET PORT

The Ethernet port of the raspberry pi is the main gateway for communicating with additional devices. The raspberry pi Ethernet port is used to plug your home router to access the internet.

GPIO PINS

The general purpose input & output pins are used in the raspberry pi to associate with the other electronic boards. These pins can accept input & output commands based on programming raspberry pi. The raspberry pi affords digital GPIO pins. These pins are used to connect other electronic components. For example, you can connect it to the temperature sensor to transmit digital data.

XBEE SOCKET

The XBee socket is used in raspberry pi boards for wireless communication purposes.

POWER SOURCE CONNECTOR

The power source cable is a small switch, which is placed on the side of the shield. The main purpose of the power source connector is to enable an external power source.

UART

The Universal Asynchronous Receiver/ Transmitter is a serial input & output port. That can be used to transfer the serial data in the form of text and it is useful for converting the debugging code.

DISPLAY

The connection options of the raspberry pi board are two types such as HDMI and Composite. Many LCD and HD TV monitors can be attached using an HDMI male cable and with a low-cost adaptor. The versions of HDMI are 1.3 and 1.4 are supported and 1.4 version cable is recommended.

The O/Ps of the Raspberry Pi audio and video through HDMI, but does not support HDMI I/p. Older TVs can be connected using composite video. When using a composite video connection, audio is available from the 3.5mm jack socket and can be sent to your TV. To send audio to your TV, you need a cable which adjusts from 3.5mm to double RCA connectors.

SOME USES FOR THE RASPBERRY PI

Enthusiasts around the world use the Pi for far more than its original purpose. Media centre software exists as a version of XBMC, and there are several Linux distributions that can be installed.

Retro gaming is possible (modern titles since around 2000 require far greater hardware resources) as is multimedia playback; remarkably the Pi is capable of HD video. You might also use the device as a web server, NAS controller, home security computer… the possibilities are endless!

SPECIFICATION

- 256 MB SDRAM memory
- Single 2.0 USB connector
- Dual Core Videocore IV Multimedia coprocessor
- HDMI (rev 1.3 & 1.4) Composite RCA (PAL and NTSC) Video Out
- 3.5 MM Jack, HDMI, Audio Out
- SD, MMC, SDIO Card slot on board storage
- Linux Operating system
- Broadcom BCM2835 SoC full HD multimedia processor

Ø 8.6cm*5.4cm*1.5cm dimensions

MODEL B RASPBERRY PI BOARD

The Raspberry Pi is a Broadcom BCM2835 System on Chip board. It comes equipped with a 700 MHz, 512 MB of SDRAM and ARM1176JZF-S core CPU. The USB 2.0 port of the raspberry pi boars uses only external data connectivity options.

The Ethernet in the raspberry pi is the main gateway to interconnect with other devices and the internet in model B. This draws its power from a micro USB adapter, with a minimum range of 2.5 watts (500 MA). The graphics, specialized chip is designed to speed up the manipulation of image calculations. This is built with Broadcom video core IV cable that is useful if you want to run a game and video through your raspberry pi.

STORAGE

One of the most important elements of any computer is the storage, from where the operating system is run and data stored. The Pi doesn't have a hard disk drive – instead, it is equipped with a SD card reader.

FEATURES OF RASPBERRY PI MODEL B

Ø 512 MB SDRAM memory

Ø Broadcom BCM2835 SOC full high definition multimedia processor

Ø Dual Core Videocore IV Multimedia co processor

Ø Single 2.0 USB connector

Ø HDMI (rev 1.3 and 1.4) Composite RCA (PAL & NTSC) Video Out

Ø 3.5 MM Jack, HDMI Audio Out

Ø MMC, SD, SDIO Card slot on board storage

Ø Linux Operating system

Ø On board 10/100 Ethernet RJ45 jack

APPLICATIONS OF RASPBERRY PI

The raspberry pi boards are used in many applications like

Ø Media streamer

Ø Arcade machine

Ø Tablet computer

Ø Home automation

Ø Internet radio

Ø Controlling robots

Ø Cosmic Computer

NEURAL NETWORKS

Neural networks are computing systems with interconnected nodes that work much like neurons in the human brains .Using the algorithms they can recognize hidden patterns and correlations in raw data ,cluster and classify it and overtime -continuously learn and improve

Types of neural networks:

There are different types of deep neural networks-and each has advantages and disadvantages depending on the use .

1.convolution neural networks (CNN)
2.Recurrent neural networks(RNN)
3.Feedforward neural networks
4.Autoencoder neural networks

Convolution neural networks:

It contains five types of layers: input, convolution, pooling, fully connected and output. Each layer has a specific purpose, like summarizing, connecting or activating. Convolutional neural networks have popularized image classification and object detection. However, CNNs have also been applied to other areas, such as natural language processing and forecasting.

Recurrent neural networks:

It uses sequential information such as time-stamped data from a sensor device or a spoken sentence, composed of a sequence of terms. Unlike traditional neural networks, all inputs to a recurrent neural network are not independent of each other, and the output for each element depends on the computations of its preceding elements. RNNs are used in forecasting and time series applications, sentiment analysis and other text applications.

Feedforward neural networks:

In this neural network , each perceptron in one layer is connected to every perceptron from the next layer. Information is fed forward from one layer to the next in the forward direction only. There are no feedback loops.

Autoencoder neural networks:

These neural networks are used to create abstractions called encoders, created from a given set of inputs. Although similar to more traditional neural networks, autoencoders seek to model the inputs themselves, and therefore the method is considered unsupervised. The premise of autoencoders is to desensitize the irrelevant and sensitize the relevant. As layers are added, further abstractions are formulated at higher layers (layers closest to the point at which a decoder layer is introduced). These abstractions can then be used by linear or nonlinear classifiers.

Out of all the neural networks ,CNN is found to be most suitable for detecting faces and comparatively it is found to have much higher efficiency among the other networks.

CONVOLUTION NEURAL NETWORKS:

1.ARCHITECTURE:

A convolutional neural network consists of an input layer, hidden layer and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a convolutional neural network, the hidden layers include layers that perform convolutions. Typically this includes a layer that does multiplication or other dot product, and its activation function is commonly ReLU. This is followed by other layers such as pooling layers, fully connected layers, and normalization layers.

2.CONVOLUTIONAL LAYERS:

In a CNN, the input is a tensor with a shape: (number of inputs) x (input height) x (input width) x (input channels). After passing through a convolutional layer, the image becomes abstracted to a feature map, also called an activation map, with shape: (number of inputs) x (feature map height) x (feature map width) x (feature map channels). A convolutional layer within a CNN generally has the following attributes:

- Convolutional filters/kernels defined by a width and height (hyper-parameters).
- The number of input channels and output channels (hyper-parameters). One layer's input channels must equal the number of output channels (also called depth) of its input.
- Additional hyperparameters of the convolution operation, such as: padding, stride, and dilation.

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neural processes data only for its receptive field. Although fully connected feedforward neural networks can be used to learn features and classify data, this architecture is generally impractical for larger inputs such as high resolution images. It would require a very high number of neurons, even in a shallow architecture, due to the large input size of images, where each pixel is a relevant input feature. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for *each* neuron in the second layer. Instead, convolution reduces the number of free parameters, allowing the network to be deeper. For example, regardless of image size, using a 5 x 5 tiling region, each with the same shared weights, requires only 25 learnable parameters. Using regularized weights over fewer parameters avoids the vanishing gradients and exploding gradients problems seen during backpropagation in traditional neural networks. Furthermore, convolutional neural networks are ideal for data with a grid-like topology (such as images) as spatial relations between separate features are taken into account during convolution and/or pooling.

2.a)Pooling layers:

Convolutional networks may include local and/or global pooling layers along with traditional convolutional layers. Pooling layers reduce the dimensions of data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, tiling sizes such as 2 x 2 are commonly used. Global pooling acts on all the neurons of the feature map. There are two common types of pooling in popular use: max and average. Max pooling uses the maximum value of each local cluster of neurons in the feature map, while average pooling takes the average value.

2.b)Fully connected layers:

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is the same as a traditional multi-layered perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.

2.c)Receptive field:

In neural networks, each neuron receives input from some number of locations in the previous layer. In a convolutional layer, each neuron receives input from only a restricted area of the previous layer called the neuron's *receptive field*. Typically the area is a square (e.g. 5 by 5 neurons). Whereas, in a fully connected layer, the receptive field is the *entire previous layer*. Thus, in each convolutional layer, each neuron takes input from a larger area in the input than previous layers. This is due to applying the convolution over and over, which takes into account the value of a pixel, as well as its surrounding pixels. When using dilated layers, the number of pixels in the receptive field remains constant, but the field is more sparsely populated as its dimensions grow when combining the effect of several layers.

2.d)Weights:

Each neuron in a neural network computes an output value by applying a specific function to the input values received from the receptive field in the previous layer. The function that is applied to the input values is determined by a vector of weights and a bias (typically real numbers). Learning consists of iteratively adjusting these biases and weights.

The vector of weights and the bias are called *filters* and represent particular features of the input (e.g., a particular shape). A distinguishing feature of CNNs is that many neurons can share the same filter. This reduces the memory footprints because a single bias and a single vector of weights are used across all receptive fields that share that filter, as opposed to each receptive field having its own bias and vector weighting.

3.HISTORY OF CNN:

3.a)RECEPTIVE FIELD IN THE VISUAL CORTEX:

Work by Hubell and Wiesel in the 1950s and 1960s showed that cat and monkey visual cortexes contain neurons that individually respond to small regions of the visual field. Provided the eyes are not moving, the region of visual space within which visual stimuli affect the firing of a single neuron is known as its receptive field. Neighboring cells have similar and overlapping receptive fields. Receptive field size and location varies systematically across the cortex to form a complete map of visual space. The cortex in each hemisphere represents the contralateral visual field.

heir 1968 paper identified two basic visual cell types in the brain:

- Simple cells, whose output is maximized by straight edges having particular orientations within their receptive field
- Complex cells , which have larger receptive fields , whose output is insensitive to the exact position of the edges in the field.

Hubel and Wiesel also proposed a cascading model of these two types of cells for use in pattern recognition tasks.

3.b)NECOGNITION ,ORIGIN OF THE CNN ARCHITECTURE:

It was inspired by the above-mentioned work of Hubel and Wiesel. The neocognitron introduced the two basic types of layers in CNNs: convolutional layers, and downsampling layers. A convolutional layer contains units whose receptive fields cover a patch of the previous layer. The weight vector (the set of adaptive parameters) of such a unit is often called a filter. Units can share filters. Downsampling layers contain units whose receptive fields cover patches of previous convolutional layers. Such a unit typically computes the average of the activations of the units in its patch. This downsampling helps to correctly classify objects in visual scenes even when the objects are shifted.

In a variant of the neocognitron called the cresceptron, instead of using Fukushima's spatial averaging, J. Weng et al. introduced a method called max-pooling where a downsampling unit computes the maximum of the activations of the units in its patch. Max-pooling is often used in modern CNNs.

Several supervised and unsupervised learning algorithms have been proposed over the decades to train the weights of a neocognitron. Today, however, the CNN architecture is usually trained through backpropagation

The neocognitron is the first CNN which requires units located at multiple network positions to have shared weights.

Convolutional neural networks were presented at the Neural Information Processing Workshop in 1987, automatically analyzing time-varying signals by replacing learned multiplication with convolution in time, and demonstrated for speech recognition.

3.c)Time delay neural networks:

The time delay neural networks (TDNN) was introduced in 1987 by Alex Waibell et al. and was the first convolutional network, as it achieved shift invariance. It did so by utilizing weight sharing in combination with Back propagation training. Thus, while also using a pyramidal structure as in the neocognitron, it performed a global optimization of the weights instead of a local one.

TDNNs are convolutional networks that share weights along the temporal dimension. They allow speech signals to be processed time-invariantly. In 1990 Hampshire and Waibel introduced a variant which performs a two dimensional convolution. Since these TDNNs operated on spectrograms, the resulting phoneme recognition system was invariant to both shifts in time and in frequency.

3.d)Max pooling:

In 1990 Yamaguchi et al. introduced the concept of max pooling, which is a fixed filtering operation that calculates and propagates the maximum value of a given region. They did so by combining TDNNs with max pooling in order to realize a speaker independent isolated word recognition system. In their system they used several TDNNs per word, one for each syllable. The results of each TDNN over the input signal were combined using max pooling and the outputs of the pooling layers were then passed on to networks performing the actual word classification.

3.e)Image recognition with CNNs trained by gradient descent:

A system to recognize hand-written numbers involved convolutions in which the kernel coefficients had been laboriously hand designed.

Yuan LeCun et al. (1989) used back-propagation to learn the convolution kernel coefficients directly from images of hand-written numbers. Learning was thus fully automatic, performed better than manual coefficient design, and was suited to a broader range of image recognition problems and image types.

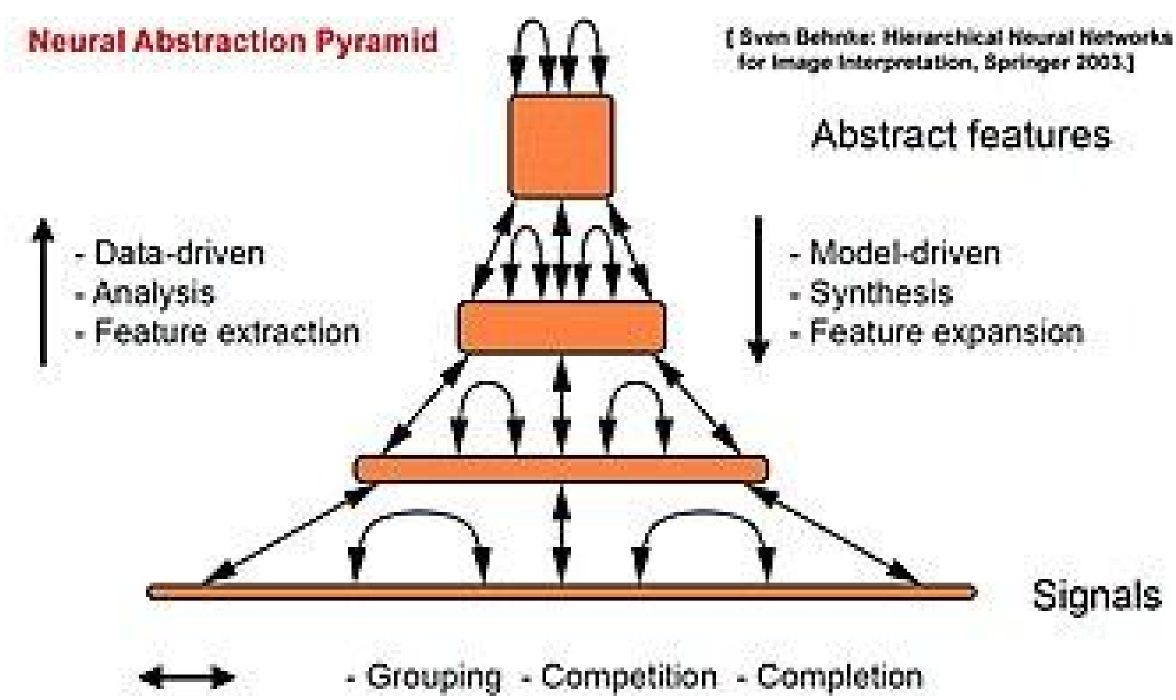
This approach became a foundation of modern computer vision..

3.f)Shift-invariant neural network:

Similarly, a shift invariant neural network was proposed by W. Zhang et al. for image character recognition in 1988.The architecture and training algorithm were modified in 1991 and applied for medical image processing and automatic detection of breast cancer in mammograms.

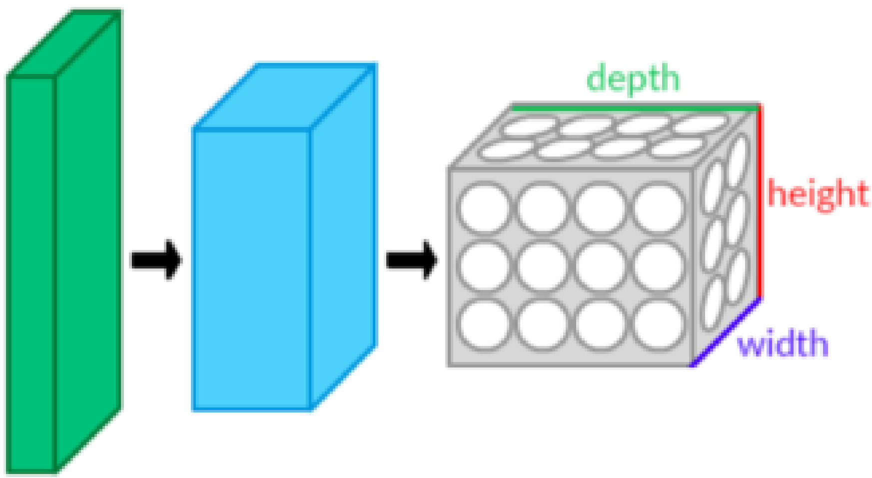
A different convolution-based design was proposed in 1988 for application to decomposition of one-dimensional electromyography convolved signals via de-convolution. This design was modified in 1989 to other de-convolution-based designs.

Neural abstraction pyramid:



The feed-forward architecture of convolutional neural networks was extended in the neural abstraction pyramid by lateral and feedback connections. The resulting recurrent convolutional network allows for the flexible incorporation of contextual information to iteratively resolve local ambiguities. In contrast to previous models, image-like outputs at the highest resolution were generated, e.g., for semantic segmentation, image reconstruction, and object localization tasks.

4.DISTINGUISHING FEATURES:



In the past, traditional multilayer perceptron (MLP) models were used for image recognition. However, the full connectivity between nodes caused the curse of directionality , and was computationally intractable with higher resolution images. A 1000×1000-pixel image with RGB color channels has 3 million weights, which is too high to feasibly process efficiently at scale with full connectivity.For example, in CIFAR-10, images are only of size 32×32×3 (32 wide, 32 high, 3 color channels), so a single fully connected neuron in the first hidden layer of a regular neural network would have 32*32*3 = 3,072 weights. A 200×200 image, however, would lead to neurons that have 200*200*3 = 120,000 weights.

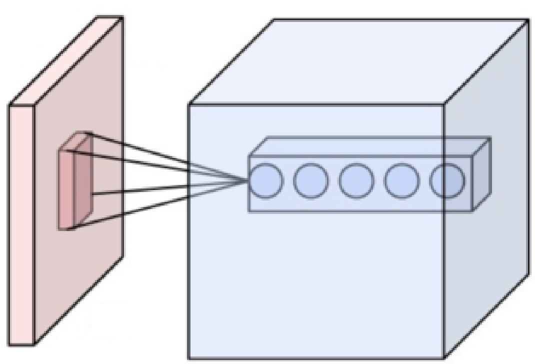
Also, such network architecture does not take into account the spatial structure of data, treating input pixels which are far apart in the same way as pixels that are close together. This ignores locality of reference in data with a grid-topology (such as images), both computationally and semantically. Thus, full connectivity of neurons is wasteful for purposes such as image recognition that are dominated by spatially local input patterns. Convolutional neural networks are variants of multilayer perceptrons, designed to emulate the behavior of a visual cortex. These models mitigate the challenges posed by the MLP architecture by exploiting the strong spatially local correlation present in natural images. As opposed to MLPs, CNNs have the following distinguishing features:

- 3D volumes of neurons. The layers of a CNN have neurons arranged in 3 dimensions: width, height and depth. Where each neuron inside a convolutional layer is connected to only a small region of the layer before it, called a receptive field. Distinct types of layers, both locally and completely connected, are stacked to form a CNN architecture.

- Local connectivity: following the concept of receptive fields, CNNs exploit spatial locality by enforcing a local connectivity pattern between neurons of adjacent layers. The architecture thus ensures that the learned "filters" produce the strongest response to a spatially local input pattern. Stacking many such layers leads to non-linear filters that become increasingly global (i.e. responsive to a larger region of pixel space) so that the network first creates representations of small parts of the input, then from them assembles representations of larger areas.
- Shared weights: In CNNs, each filter is replicated across the entire visual field. These replicated units share the same parameterization (weight vector and bias) and form a feature map. This means that all the neurons in a given convolutional layer respond to the same feature within their specific response field. Replicating units in this way allows for the resulting activation map to be equivariant under shifts of the locations of input features in the visual field, i.e. they grant translational equivariance - given that the layer has a stride of one.
- Pooling: In a CNN's pooling layers, feature maps are divided into rectangular sub-regions, and the features in each rectangle are independently down-sampled to a single value, commonly by taking their average or maximum value. In addition to reducing the sizes of feature maps, the pooling operation grants a degree of local translation invariance to the features contained therein, allowing the CNN to be more robust to variations in their positions.

Together, these properties allow CNNs to achieve better generalization on visual problems. Weight sharing dramatically reduces the number of free parameters learned, thus lowering the memory requirements for running the network and allowing the training of larger, more powerful networks.

5.**BUILDING BLOCKS:**



A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume (e.g. holding the class scores) through a differentiable function. A few distinct types of layers are commonly used. These are further discussed below.

5.a)**CONVOLUTION LAYER:**

The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the filter entries and the input, producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input.Stacking the activation maps for all filters along the depth dimension forms the full output volume of the convolution layer. Every entry in the output volume can thus also be interpreted as an output of a neuron that looks at a small region in the input and shares parameters with neurons in the same activation map.

5.b)LOCAL CONNECTIVITY:

When dealing with high-dimensional inputs such as images, it is impractical to connect neurons to all neurons in the previous volume because such a network architecture does not take the spatial structure of the data into account. Convolutional networks exploit spatially local correlation by enforcing a sparse local connectivity pattern between neurons of adjacent layers: each neuron is connected to only a small region of the input volume.

The extent of this connectivity is a hyperparameter called the receptive field of the neuron. The connections are local in space (along width and height), but always extend along the entire depth of the input volume. Such an architecture ensures that the learnt filters produce the strongest response to a spatially local input pattern.

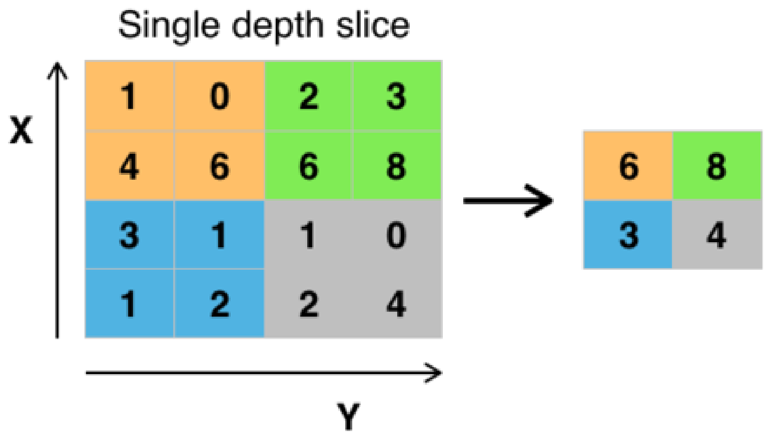
5.c)PARAMETER SHARING:

A parameter sharing scheme is used in convolutional layers to control the number of free parameters. It relies on the assumption that if a patch feature is useful to compute at some spatial position, then it should also be useful to compute at other positions. Denoting a single 2-dimensional slice of depth as a depth slice, the neurons in each depth slice are constrained to use the same weights and bias.

Since all neurons in a single depth slice share the same parameters, the forward pass in each depth slice of the convolutional layer can be computed as a convolution of the neuron's weights with the input volume.[nb 2] Therefore, it is common to refer to the sets of weights as a filter (or a kernel), which is convolved with the input. The result of this convolution is an activation map, and the set of activation maps for each different filter are stacked together along the depth dimension to produce the output volume. Parameter sharing contributes to the translation invariance of the CNN architecture.[64]

Sometimes, the parameter sharing assumption may not make sense. This is especially the case when the input images to a CNN have some specific centered structure; for which we expect completely different features to be learned on different spatial locations. One practical example is when the inputs are faces that have been centered in the image: we might expect different eye-specific or hair-specific features to be learned in different parts of the image. In that case it is common to relax the parameter sharing scheme, and instead simply call the layer a "locally connected layer".

5.d)POOLING LAYER:



Another important concept of CNNs is pooling, which is a form of non-linear down-sampling. There are several non-linear functions to implement pooling, where max pooling is the most common. It partitions the input image into a set of rectangles and, for each such sub-region, outputs the maximum.

Intuitively, the exact location of a feature is less important than its rough location relative to other features. This is the idea behind the use of pooling in convolutional neural networks. The pooling layer serves to progressively reduce the spatial size of the representation, to reduce the number of parameters, memory footprint and amount of computation in the network, and hence to also control overfitting. This is known as down-sampling It is common to periodically insert a pooling layer between successive convolutional layers (each one typically followed by an activation function, such as a ReLU layer) in a CNN architecture. While pooling layers contribute to local translation invariance, they do not provide global translation invariance in a CNN, unless a form of global pooling is used[4][61]. The pooling layer commonly operates independently on every depth, or slice, of the input and resizes it spatially. A very common form of max pooling is a layer with filters of size 2×2, applied with a stride of 2, which subsamples every depth slice in the input by 2 along both width and height, discarding 75% of the activations:

$$f_{X,Y}(S) = \max_{a,b=0}^{1} S_{2X+a,2Y+b}.$$

In this case, every max operation is over 4 numbers. The depth dimension remains unchanged (this is true for other forms of pooling as well).
In addition to max pooling, pooling units can use other functions, such as average pooling or ℓ2-norm pooling. Average pooling was often used historically but has recently fallen out of favor compared to max pooling, which generally performs better in practice.

Due to the effects of fast spatial reduction of the size of the representation,there is a recent trend towards using smaller filters or discarding pooling layers altogether.

5.e)ReLU LAYER:

ReLu applies the non-saturating activation function f(x)=max(0,x).It effectively removes negative values from an activation map by setting them to 0.It introduces the nonlinearnities to the decision function and in the overall network without affecting the receptive fields of the convolution layers.

5.f)FULLY CONNECTED LAYER:

After several convolutional and max pooling layers ,the final classification is done via fully connected layers.Neurons in a fully connected layer have connections to all activations in the previous layer,as seen in regular artificial neural networks .Their activations can thus be computed as an affine transformation,with matrix multiplication followed by a bias offset.

5.g)LOSS LAYER:

The loss layer or loss function specifies how training penalizes the deviation between the predicted output of the network and the true data labels .Various loss functions can be used,depending on the specific task.The Softmax loss function is used for predicting a single class of K mutually exclusive classes.Sigmoid cross-entropy loss is used for predicting K independent probability values in [0,1].Euclidean loss is used for regressing to real-valued labels.

6.CHOOSING HYPERPARAMTERS:

CNNs use more hyperparameters than a standard multilayer perceptron .While the usual rules for learning rates and regularization constants still apply,the following should be kept in mind when optimizing.

6.a)NUMBER OF FILTERS:

Since feature map size decreases with depth ,layers near the input layer tend to have fewer filters while higher layers can have more.To equalise computation at each layer,the product of feature values Va with pixel position is kept roughly constant across layers.Preserving more information about the input would require keeping the total number of activations non-decreasing from one layer to the next.The number of feature maps directly controls the capacity and depends on the number of available examples and task completely.

6.b)FILTER SIZE:

Common filter sizes found in the literature vary greatly and are usually chosen based on the data set.The challenge is to find the right level of granularity so as to create abstraction at the proper scale ,given a particular data set and without overfitting.

6.c)POOLING TYPE AND SIZE :

In modern CNNs,max pooling is typically used and often of size of 2*2 ,with a stride of 2.This implies that the input is drastically downsampled ,further improving the computational efficiency.Very large input volumes may warrant 4*4 pooling in the lower layers.However,choosing larger shapes will dramatically reduce the dimension of the signal and may result in excess information loss.Often ,non-overlapping pooling windows perform best.

7.TRANSLATIONAL EQUIVARIANCE:

It is commonly assumed that CNNs are invariant to shifts to the inputs .However,convolution or pooling layers within a CNN that do not have a stride greater than one are equivariant ,as opposed to invariant,to translation of the input.Layers with a stride greater than one ignores the Nyquist-Shannon sampling theorem,and leads to aliasing of the input signal which breaks the equivariance property.Furthermore ,if a CNN makes use of fully connected layers,translation equivariance ,as the fully connected translation invariance is avoiding any down-sampling throughout the network and applying global average pooling at the last layer.Additionally ,several other partial solutions have been proposed such as anti-aliasing ,spatial transformer networks,data augmentation,subsampling combined with pooling and capsule neural networks.

8.REGULARIZATION METHODS:

8.a)EMPIRICAL:

Dropout:

Because a fully connected layer occupies most of the parameters ,it is prone to overfitting .1 method to reduce overfitting is dropout .At each training stage ,individual nodes are either dropped out of the net with probability 1-p or kept with probability p,so that a reduced network is left ,incoming and outgoing edges to a dropped out node are also removed.Only the reduced network is trained on the data in that stage.The removed nodes are then reinserted into the network with their original weights.In the training stages p is usually 0.5,for input nodes it is typically much higher because information is directly lost when input node are ignored.

At testing time after training has finished ,we would ideally like to find sample,average of all possible 2^n dropped out networks unfortunately this is unfeasible for larger values of n.However,we can find an approximation by using the full network with each node;s output weighted by a factor of p,so the expected value of the output of any node is the same as in the training stages.By avoiding training all nodes on all training data dropout decreases overfitting.The method also significantly improves training speed.The technique seems to reduce node interactions,leading them to learn more robust features that better generalize to new data.

DropConnect:

DropConnect is the generalization of dropout in which each connection rather than each output unit,can be dropped with probability 1-p.Each unit thus receives input from a random subsets of units in the previous layer.It is similar to dropout as it introduces dynamic sparsity within the model,but differs in that sparsity is on the weights ,rather than the output vectors of a layer.In other words,the fully connected layer with DropConnect becomes a sparsely connected layer in which the connections are chosen at random during the training stage.

Stochastic pooling:

A major drawback to Dropout is that it does not have the same benefits for convolutional layers,where the neurons are not fully connected.In scholastic pooling ,the conventional deterministic pooling operations are replaced with a stochastic procedure ,where the activation within each pooling region is picked randomly according to a multinomial distribution,given by the activities within the pooling region.This approach is free of hyperparameters and can be combined with other regularization approaches,such as dropout and data augmentation.An alternate view of stochastic pooling is that it is equivalent to standard max pooling but with many copies of an input image ,each having small local deformations.This is similar to explicit elastic deformation of the input images,which delivers excellent performance on the MNist data set.Using the stochastic pooling in a multilayered model gives an exponential number of deformations since the selections in higher layers are independent of those below.

ARTIFICIAL DATA:

Because of the degree of model overfitting of a network is to simply stop the training before overfitting has a chance to occur.It comes with the disadvantage that the learning process is halted.