Average Project Rep
AY: 2020-21.

A
Major Project report
On

FOOD CLASSIFICATION USING DEEP LEARNING AND EMBEDDED SYSTEMS

Submitted in the Partial Fulfilment of the Academic Requirements for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By

ALLORI ASHA REDDY (17911A04J5)
KEMSARAM NIKITHA (16911A04L0)
SANAPATHI SAIKUMAR (17911A04N9)

Under The Esteemed Guidance Of

Mr. MD. AKRAM AHMED
Assistant Professor

Department of Electronics and Communication Engineering,
VIDYA JYOTHI INSTITUTE OF TECHNOLOGY
(An Autonomous Institute)
Accredited by NAAC & NBA, Approved By A.I.C.T.E., New Delhi, Permanently Affiliated to JNTU, Hyderabad
(Aziz Nagar, C.B.Post, Hyderabad -500075)

2017-2021

A

Major Project report

On

FOOD CLASSIFICATION USING DEEP LEARNING AND EMBEDDED SYSTEMS

Submitted in the Partial Fulfilment of the Academic Requirements for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By

ALLORI ASHA REDDY	**(17911A04J5)**
KEMSARAM NIKITHA	**(16911A04L0)**
SANAPATHI SAIKUMAR	**(17911A04N9)**

Under The Esteemed Guidance Of

Mr. MD. AKRAM AHMED

Assistant Professor

Department of Electronics and Communication Engineering,

VIDYA JYOTHI INSTITUTE OF TECHNOLOGY

(An Autonomous Institute)

Accredited by NAAC & NBA, Approved By A.I.C.T.E., New Delhi, Permanently Affiliated to JNTU, Hyderabad)

(Aziz Nagar, C.B.Post, Hyderabad -500075)

2017-2021

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology
Hyderabad-500075

VIDYA JYOTHI INSTITUTE OF TECHNOLOGY

(An Autonomous Institute)

Accredited by NAAC & NBA, Approved By A.I.C.T.E., New Delhi, Permanently Affiliated to JNTU, Hyderabad)

(Aziz Nagar, C.B.Post, Hyderabad -500075)

Department of Electronics and Communication Engineering

CERTIFICATE

This is to certify that the project report titled **"FOOD CLASSIFICATION USING DEEP LEARNING AND EMBEDDED SYSTEMS"** is being submitted by **ALLORI ASHA REDDY (17911A04J5), KEMSARAM NIKITHA (16911A04L0), SANAPATHI SAIKUMAR (17911A04N9)** of B.Tech IV-II semester of *Electronics & Communication Engineering* is a record bonafide work carried out by them. The results embodied in this report have not been submitted to any other University for the award of any degree.

INTERNAL GUIDE

Mr. MD. AKRAM AHMED

(Asst. Professor, Dept. Of ECE)

HEAD OF THE DEPARTMENT

Mr. DR. K. VASANTH

(Professor, & HOD Dept. Of ECE)

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

EXTERNAL EXAMINER

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

DECLARATION

We, the undersigned, declare that the project title "FOOD CLASSIFICATION USING DEEP LEARNING AND EMBEDDED SYSTEMS" being submitted in partial fulfilment for award of Bachelor of Technology Degree in Electronics & Communication Engineering, Vidya Jyothi Institute of Technology (*An Autonomous Institute*) is work carried out by us.

PROJECT ASSOCIATE:

ALLORI ASHA REDDY (17911A04J5)

KEMSARAM NIKITHA (16911A04L0)

SANAPATHI SAIKUMAR (17911A04N9)

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to MR. MD. AKRAM AHMED Project guide who has guided and supported us through every stage in the project.

We are grateful to MR. S. PRADEEP KUMAR REDDY Project Coordinator for his time to time, much needed valuable guidance throughout my study.

We are grateful to DR. K. VASANTH HOD ECE Dept., Vidya Jyothi Institute of Technology for his time to time, much needed valuable guidance throughout my study.

We express our heartly gratitude to DR. A. PADMAJA Principal Vidya Jyothi Institute of Technology for giving us spontaneous encouragement for completing the project.

We thank DR. P. VENUGOPAL REDDY, Dean of Examinations, Vidya Jyothi Institute of Technology for encouraging us in the completion of our project.

We thank DR. E. SAIBABA REDDY, Director of Vidya Jyothi Institute of Technology for encouraging us in the completion of our project.

It is our privilege to express our gratitude and indebtedness to SHRI. PALLA. RAJESHWAR REDDY Secretary& Correspondent of Vidya Jyothi Institute of Technology for his moral support.

We express our heartfelt thanks to Staff Of Electronics And Communication Department, Vidya Jyothi Institute of Technology for helping us in carrying out our project successfully.

PROJECT ASSOCIATE:

ALLORI ASHA REDDY (17911A04J5)

KEMSARAM NIKITHA (16911A04L0)

SANAPATHI SAIKUMAR (17911A04N9)

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075.

ABSTRACT

Foods are important things to human beings, especially for elderly and diabetics. Tradition nutrition book is not the effective way for people to use and not cover all kind of foods. Most of the food nutrition in the book focused on Western dishes not Asian dishes. This research proposed the new way to categorized food dishes and estimate nutritional value using convolutional neural networks. In this busy world, one spends least time and interest in eating healthy food and caring about our food and diet. This gradually is leading to unhealthy food practices. To provide oneself with the information regarding the estimation of nutritional values of the food consumed these networks are transferred into Raspberry-Pi 3B platform to simulate limited resources and calculation power platform likes in a mobile phone. The image of food is captured by the webcam, and it identifies the food category and ingredients in the food, and it is sent to Raspberry-Pi 3B where CNNs categorize the food dishes and estimate the nutrition value and display it on the screen. In this the diabetic patient whether to if the person is diabetic and trying to consume unhealthy food like (PIZZA) "This food is not good for diabetic patient" is displayed and if he is consuming healthy food like (APPLE) "This food is good for diabetic patient" is displayed. Convolutional neural networks are used to determine the food categories and estimate the nutritional value. CNN will take the pixel values of the image and it will compare to approximate pixel values and detect. The networks in Raspberry-Pi 3B produce good prediction accuracy but slow speed.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

LIST OF FIGURES

Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology,
Hyderabad-500075

INDEX

Department of Electronics and Communication Engineering
Vidya Jyothi Institute of Technology.
Hyderabad-500075

CHAPTER - 1

INTRODUCTION TO FOOD CLASSIFICATION USING DEEP LEARNING AND EMBEDDED

1.1 INTRODUCTION TO CONVOLUTION NEURAL NETWORK

Convolutional neural networks (CNNs) are widely used in pattern and image-recognition problems as they have several advantages compared to other techniques. This white paper covers the basics of CNNs including a description of the various layers used. CNNs are used in variety of areas, including image and pattern recognition, speech recognition, natural language processing, and video analysis. There are several reasons that convolutional neural networks are becoming important. In traditional models for pattern recognition, feature extractors are hand designed. In CNNs, the weights of the convolutional layer being used for feature extraction as well as the fully connected layer being used for classification are determined during the training process. The improved network structures of CNNs lead to savings in memory requirements and computation complexity requirements and, at the same time, give better performance for applications where the input has local correlation (e.g., image and speech).Large requirements of computational resources for training and evaluation of CNNs are sometimes met by graphic processing units (GPUs), DSPs, or other silicon architectures optimized for high throughput and low energy when executing the idea syncretic patterns of CNN computation. In fact, advanced processors such as the Tensilica Vision P5 DSP for Imaging and Computer Vision from Cadence have an almost ideal set of computation and memory resources required for running CNNs at high efficiency.

A neural network is a system of interconnected artificial "neurons" that exchange messages between each other. The connections have numeric weights that are tuned during the training process, so that a properly trained network will respond correctly when presented with an image or pattern to recognize. The network consists of multiple layers of feature-detecting "neurons". Each layer has many neurons that respond to different combinations of inputs from the previous layers. The layers are built up so that the first layer detects a set of primitive patterns in the input, the second layer detects patterns of patterns, the third layer detects patterns of those patterns, and so on. Typical CNNs use 5 to 25 distinct layers of pattern recognition.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

Layers of CNN

By stacking multiple and different layers in a CNN, complex architectures are built for classification problems. Four types of layers are most common convolution layers, pooling/sub-sampling layers, non-linear layers, and fully connected layers.

CNNs in Embedded Systems

To run CNNs on a power-constrained embedded system that supports image processing, it should fulfil the following requirements.

Availability of high computational performance for a typical CNN implementation, billions of MACs per second is the requirement. Larger load/store bandwidth in the case of a fully connected layer used for classification purpose, each coefficient gets used in multiplication only once. So, the load-store bandwidth requirement is greater than the number of MACs performed by the processor. Low dynamic power requirement the system should consume less power. To address this issue, fixed-point implementation is required, which imposes the requirement of meeting the performance requirements using the minimum possible finite number of bits. Flexibility it should be possible to easily upgrade the existing design to new better performing design.

Since computational resources are always a constraint in embedded systems, if the use case allows a small degradation in performance, it is helpful to have an algorithm that can achieve huge savings in computational complexity at the cost of a controlled small degradation in performance. So, Cadence's work on an algorithm to achieve complexity versus a performance trade off, as explained in the previous section, has great relevance for implementing CNNs on embedded system.

In the current age, people are more conscious about their food and diet to avoid either upcoming or existing diseases. Since people are dependent on smart technologies, provision of an application to automatically monitor the individual's diet, helps in many aspects. It increases the awareness of people in their food habits and diet. Over the last two decades, research has been focused on automatically recognizing the food and their nutritional information from images capture during computer vision and machine learning techniques. In order to properly assess dietary intake, accurate estimation of calorie value of food is of paramount importance. A majority of the people are over eating and not being active enough. Given how busy and stressed people are today, it's effortless to forget to keep track of the food that they eat. This only increases the importance of proper classification of food.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

Some of the methods currently in use for dietary assessment involve self-reporting and manually recorded instruments. The issue with such methods of assessment is that the evaluation of calorie consumption by a participant is prone to bias i.e. underestimating and under reporting of food intake. In order to increase the accuracy and reduce the bias, enhancements to the current methods are required. One such potential solution is a mobile cloud computing system, which makes use of devices such as smartphones to capture dietary and calorie information. The next step is to automatically analyse the dietary and calorie information employing the computing capacity of the cloud for an objective assessment. However, users still have to enter the information manually. Over the last few years, plenty of research and development efforts have been made in the field of visual-based dietary and calorie information analysis. However, the efficient extraction of information from food images remains a challenging issue.

Convolutional neural networks (CNNs) are widely used in pattern and image-recognition problems as they have a number of advantages compared to other techniques. This white paper covers the basics of CNNs including a description of the various layers used. We outline the challenges of using CNNs in embedded systems and introduce the key characteristics of the Cadence Tensilica Vision P5 digital signal processor (DSP) for Imaging and Computer Vision and software that make it so suitable for CNN applications across many imaging and related recognition tasks.

1.2 APPLICATION:

1. Image and Pattern recognition
2. Speech recognition
3. Text/Digits recognition
4. Video analysis
5. Signal processing
6. Face detection
7. Natural object classification
8. Image processing

1.3 OBJECTIVE

An effort has been made to classify the images of food for further diet monitoring applications using convolutional neural networks (CNNs). Since the CNNs are capable of handling a large amount of data and can estimate the features automatically, they have been utilized for the task of food classification. Convolutional neural networks (CNNs) are widely used in pattern and image-recognition problems as they have a number of advantages compared to other techniques. It covers the basics of CNNs including a description of the various layers used. The device will capture the image of the food and it identifies the food category and ingredients in the food and estimate the food nutritional value and display it on the screen.

CHAPTER - 2
LITERATURE SURVEY

.1 LITERATURE SURVEY

'ood Image Recognition by Using Convolutional Neural Networks (CNNs) Yuzhen Lu

Food image recognition is one of the promising applications of visual object recognition in computer ision. In this study, a small-scale dataset consisting of 5822 images of ten categories and a five layer CNN vas constructed to recognize these images. The bag-of-features (BoF) model coupled with support vector nachine (SVM) was first evaluated for image classification, resulting in an overall accuracy of 56% while the CNN model performed much better with an overall accuracy of 74%. Data augmentation techniques based on geometric transformation were applied to increase the size of training images, which achieved a significantly improved accuracy of more than 90% while preventing the overfitting issue that occurred to the CNN based on raw training data.

A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques Raza Yunus* , Omar Arif* , Hammad Afzal* , Muhammad Faisal Amjad* , Haider Abbas.

There has been a rapid increase in dietary ailments during last few decades, caused by unhealthy food routine. Mobile-based dietary assessment systems that can record real time images of meal and analyse it for nutritional content can be very handy and improve the dietary habits, and therefore, result in healthy life. This paper proposes a novel system to automatically estimate food attributes such as ingredients and nutritional value by classifying the input image of food. Our method employs different deep learning models for accurate food identification. In addition to image analysis, attributes and ingredients are estimated by extracting semantically related words from a huge corpus of text, collected over the Internet. We performed experiments with a dataset comprising 100 classes, averaging 1000 images for each class to acquire top classification rate of up to 85 percent.

Calorie Detection of Food Image Based On SVM Algorithm

International Research Journal of Engineering and Technology (IRJET) Volume 07 , Issue 12 , Dec 2020.

The paper designs a prototype system supported the client server model. The client sends a picture detection request and processes it on the server side. The prototype system is meant with three main software components, including a pre-trained SVM model training module for classification purposes, a text data training module for attribute estimation models, and a server side module. We experimented with a spread of food categories, each containing thousands of images, and through machine learning training to realize higher classification accuracy.

Image Based Food Classification and Volume Estimation for Dietary Assessment

Frank Po Wen Lo (Student Member, IEEE), Yingnan Sun (Student Member, IEEE), Benny Lo (Senior Member, IEEE) are the authors.

They worked on how this study provides an overview of computing algorithms, mathematical models and methodologies used in the field of image-based dietary assessment. It also provides a comprehensive comparison of the state of art approaches in food recognition and volume/weight estimation in terms of their processing speed, model accuracy, efficiency and constraints. It will be followed by a discussion on deep learning method and its efficiency in dietary assessment. After a comprehensive exploration, we found that integrated dietary assessment systems combining with different approaches could be the potential solution to tackling the challenges in accurate dietary intake assessment.

Effectiveness of Food Portion Size Estimation Aids for Diet Assessment

Vidisha Sharma (PhD Scholar), Ravinder Chadha (Associate Professor) are the authors.

The work they did is a total of 1355 records were obtained through initial search, of which 27 were included in qualitative synthesis. Three categories of PSEAs were identified 2-dimensional, 3-dimensional, and technological aids. Two dimensional aids (n=9) were the most often evaluated PSEA followed by technological aids (n=8). Aspects like type of food item, its shape and food PS (small vs. large) were recognized as influencers of PS estimates. Training improved the PS estimation accuracy.

CHAPTER - 3
ANALYSIS

3.1 EXISTING SYSTEM

Estimation of food categories and calories for food photos. Since there exists strong correlation between food categories and calories in general, we expect that simultaneous training of both brings performance boosting compared to independent single training.

3.2 PROPOSED SYSTEM

The system can detect food items and give the complete nutritional values and added a feature which facilitates in recognition of diabetic friendly food and also gives an alert if excess of food is consumed by a healthy person which is not present in existing systems. But the proposed system has a drawback that it cannot detect the food items which are shapeless like liquids. In this system, the device will capture the image of the food and it identifies the food category and ingredients in the food and display it on the screen. To disseminate the nutritional values of foods by capturing the image of the food dish. Convolutional neural networks are used to determine the food categories and estimate the nutritional value. CNN will take the pixel values of the image and it will compare to approximate pixel values and detect. The first step in tracking calorie intake using images is to identify the food being consumed. The difficulty arises when one considers the various assortments of cuisines and dishes that exist in the real world. Given the size and variety of the food items in the dataset, this has proven to be quite a formidable task. The use of neural networks seems to be better to deal with the issue of scaling primarily because of their ability to learn patterns that are not linearly separable, along with the concepts of dealing with other factors such as noise in the images. Calorific Value Estimation the remaining task after the process of classification is mapping the food names to a calorific value. This can be achieved relatively easily by scraping the web for the average calorie value of food items per unit weight. The average calorie values are considered for the different classes of food, per 100g of serving. The CNN offers a cutting edge strategy for image recognition. It is a multi-layer neural organization, whose neurons take little fixes of the past layer as information. It is hearty against little moves and revolutions. A CNN framework includes a convolution layer and a pooling (or sub-sampling) layer. In the convolution layer, not at all like for general completely associated neural organizations, weights can be considered as n × (n < input size) channels. Each information convolves these channels. Each layer has numerous channels that create various yields. For the image recognition task, the various highlights are separated by these channels. The channels are regularly called (convolution) parts. The pooling layer delivers the yields by actuation over rectangular areas.There are a few actuation strategies, for example, maximum and average activation.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

This makes the CNN's yields more invariant with regard to position. A commonplace CNN contains various convolution and pooling layers, with a completely associated layer to create the eventual outcome of the undertaking. In image recognition, every unit of the last layer shows the class likelihood. A CNN has hyper boundaries that incorporate the quantity of centre layers, the size of the convolution parts, and the dynamic capacities. Due to lack of knowledge on what we are eating, we are developing health issues like increase in cholesterol which in turn leads to many other health complications like heart attacks, obesity etc. the old fashioned possible way of getting to know the nutritious value of the food being consumed is first, having the knowledge about the ingredients used in the food and then searching about their nutritional values in the traditional nutritional book which is a tedious process and given now-a -days scenario , no one would opt to follow this procedure. Aim of this research is to develop the way to extract food information from food dish images. First the Convolutional neural networks are used to determine the food categories and estimate the nutritional value and display the output.

3.3 ADVANTAGES

- If any person is allergic to any ingredient so we can avoid it.
- Helps aged people in deciding the proper diet food by classification in our project.
- Very High accuracy in image recognition problems.
- Automatically detects the important features without any human supervision.
- Weight sharing.
- The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision.
- CNN is also computationally efficient.

3.4 DISADVANTAGES

- Sometimes, it may not identify the food category and calories of the food because of image clarity.
- One of many challenges in the field of computer vision is to deal with the variance in the data present in the world.
- CNN do not encode the position and orientation of object.
- Lack of ability to be spatially invariant to the input data.
- Cannot estimate Liquid Food nutritional value.

CHAPTER - 4

SYSTEM ARCHITECTURE

4.1 Block diagram

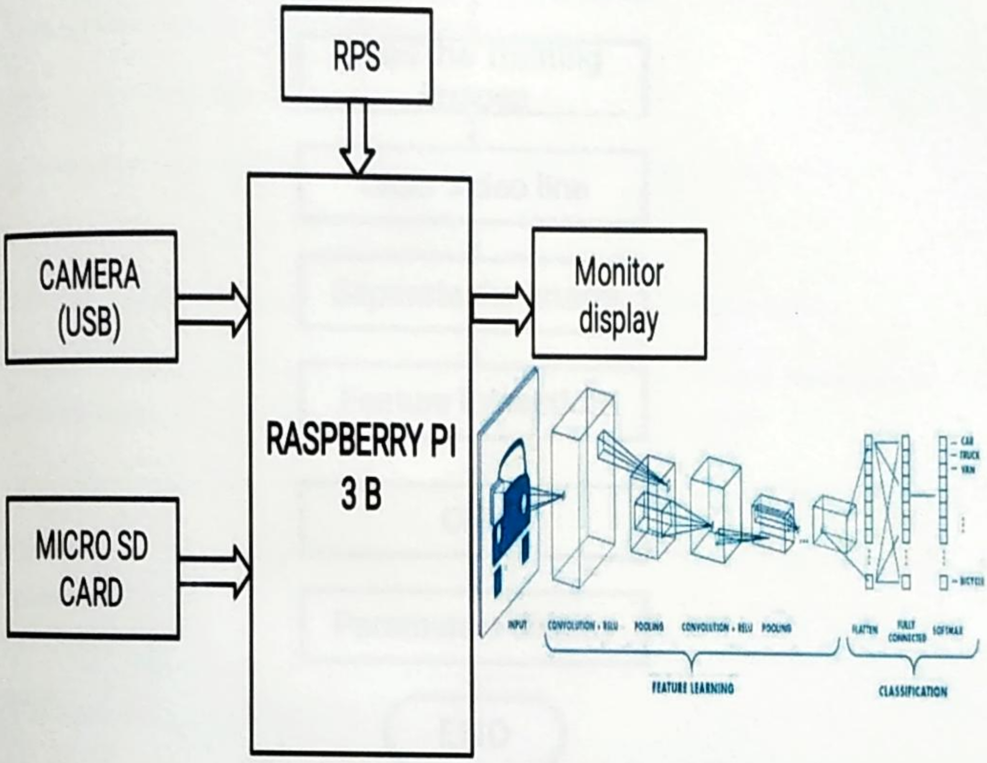


fig 4.1 block diagram of food classification using deep learning and embedded

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

4.2 User Application Based Flowchart

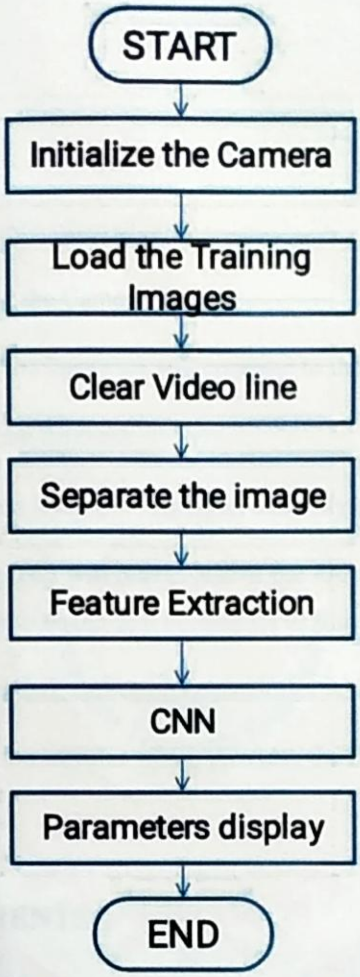


fig 4.2 user application-based flowchart of food classification

4.3 ALGORITHM

The main tools required for this project can be classified into two broad categories.

- Hardware requirement
- Software requirement

Here we are using some steps for the execution of code

STEP 1: Capturing the image by using the Camera.

STEP 2: By using the pixel values of the image we will compare to the appropriate value.

STEP 3: Load the training images to the database.

STEP 4: CNN has 3 layers convolution layer, pooling layer, and fully connected layer.

STEP 5: In layers the image of the pixels will reduce and in the fully connected layers they have different nodes connected to each nodes and gives us the best probability to find the output.

STEP 6: The parameters like proteins, fats, carbohydrates, calories which gives us the output of our image.

STEP 7: The output is displayed on the monitor which provides the estimation values of calories, proteins, fats, carbohydrates value.

4.4 SOFTWARE REQUIREMENTS

Here are the software requirements

- Raspbian OS
- Python

4.4.1 RASPBIAN OS

Raspberry Pi OS (formerly Raspbian) is a Debian-based operating system for Raspberry Pi. Since 2015, it has been officially provided by the Raspberry-Pi Foundation as the primary operating system for the Raspberry-Pi family of compact single-board computers.

Raspberry Pi OS (formerly Raspbian) is a Debian-based operating system for Raspberry-Pi. Since 2015, it has been officially provided by the Raspberry-Pi Foundation as the primary operating system for the pberry-Pi family of compact single-board computers. The first version of Raspbian was created by Mike mpson and Peter Green as an independent project. The initial build was completed in June 2012.

Raspberry-Pi OS is highly optimized for the Raspberry-Pi line of compact single-board computers with M CPUs. It runs on every Raspberry-Pi except the Pico microcontroller. Raspberry-Pi OS uses a modified DE as its desktop environment with the open box stacking window manager, along with a unique theme.

5 HARDWARE REQUIREMENTS

erating System: Windows 7/ Windows 8

ftware Setup: Python 3.7.6

chnology: Python.org

ere are the Hardware components.

- Raspberry PI 3 B
- Micro SD card
- Monitor display
- Regulated power supply
- Camera (USB)

.5.1 Raspberry-PI 3 B

Raspberry-Pi 3B Model was released in February 2016 with a 1.2 GHz 64-bit quad core ARM Cortex-A53 processor, on-board 802.11n Wi-Fi, Bluetooth and USB boot capabilities.

The Raspberry-Pi 3B Model is the third generation Raspberry-Pi. This powerful credit-card sized single board computer can be used for many applications and super seeds the original Raspberry-Pi Model B+ and Raspberry-Pi 2B Model. Whilst maintaining the popular board format the Raspberry-Pi 3B Model brings you a more powerful processer, 10x faster than the first generation Raspberry-Pi. Additionally it adds wireless LAN & Bluetooth connectivity making it the ideal solution for powerful connected designs.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

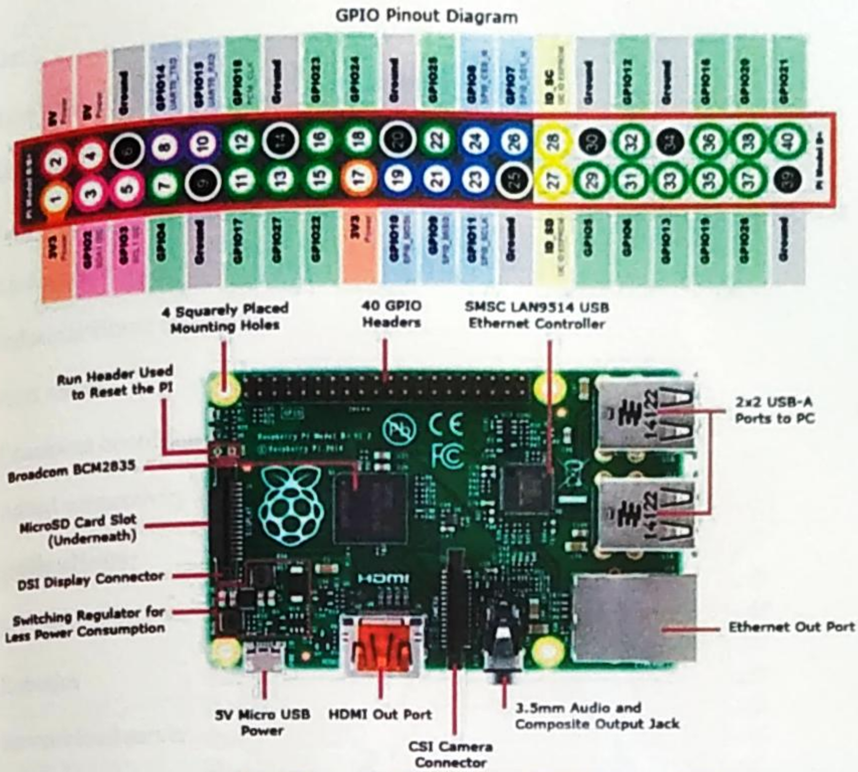


fig 4.5.1 Raspberry-pi 3B

Features:

- Quad Core 1.2GHz Broadcom BCM2837 64bit CPU
- 1GB RAM
- BCM43438 wireless LAN and Bluetooth Low Energy (BLE) on board
- 100 Base Ethernet
- 40-pin extended GPIO
- 4 USB 2 ports
- 4 Pole stereo output and composite video port
- Full size HDMI
- CSI camera port for connecting a Raspberry Pi camera
- DSI display port for connecting a Raspberry Pi touchscreen display
- Micro SD port for loading your operating system and storing data
- Upgraded switched Micro USB power source up to 2.5A

Benefits:

- Low cost
- 10x faster processing
- Low cost PC/tablet/laptop
- Media centre
- Industrial/Home automation
- Print server
- Consistent board format
- Added connectivity

Key Applications:

- IoT applications
- Robotics
- Server/cloud server
- Security monitoring
- Gaming
- Web camera
- Wireless access point
- Environmental sensing/monitoring (e.g. weather station)

4.5.2 Micro SD card

Micro Card may support various combinations of the following bus types and transfer modes. The SPI bus mode and one-bit SD bus mode are mandatory for all SD families, as explained in the next section. Once the host device and the SD card negotiate a bus interface mode, the usage of the numbered pins is the same for all card sizes.

SPI bus mode Serial Peripheral Interface. Bus is primarily used by embedded microcontrollers. This bus type supports only a 3.3-volt interface. This is the only bus type that does not require a host license. All SD cards support this mode. UHS-I and UHS-II require this bus type two differential lines SD UHS-II mode: Uses two low-voltage differential interfaces to transfer commands and data. UHS-II cards include this interface in addition to the SD bus modes. The physical interface comprises 9 pins, except that the mini SD card adds two unconnected pins in the centre and the micro SD card omits one of the two VSS (Ground) pins.

4.5.3 Monitor display

A display monitor is an electronic device used to display video output from computers. Display monitors are used in many computer devices, ranging from personal computers (PC) and laptops to small handheld mobile devices, like cell phones and MP3 players.

fig 4.5.3 Monitor

4.5.4 Regulated power supply

Regulated power supply in new smart substations and retrofitted unattended substations of power systems, some smart-monitoring technologies like helicopters, drones, robots, etc., equipped with cameras, take high-definition videos and infrared thermal images, to achieve efficient and rapid substation inspection. These massive media data streams provide a database for the image-based methods for power equipment state recognition. However, due to the particularity of the power equipment itself and the operating environment, it is not reasonable to use the object recognition in computer vision field to power equipment directly. In this way, it is of great significance to propose a method applicable to power edge image recognition. There are now mainly two kinds of object recognition in power systems: manual feature extraction-based method and automatic feature extraction-based method.

Based on deep learning, automatic feature extraction-based method can effectively reduce the deviation of the model, and has higher accuracy because of the large volume of data used and deep feature extraction. Based on the infrared images, a two stage method for current transformer fault location is proposed.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

4.5.5 Camera (USB)

USB Cameras are imaging cameras that use USB 2.0 or USB 3.0 technology to transfer image data. A USB webcam is a camera that connects to a computer, usually through plugging it in to a USB port on the machine. USB3 Vision cameras are an excellent tool for a variety of applications. Especially their bandwidth that effectively closes the speed gap between Camera Link and GigE interfaces, their simple plug and play functionality and their Vision Standard compliance make them suitable for industrial applications. In addition, the USB 3.0 is perfectly tailored for the latest generation of CMOS sensors, with the architecture and bandwidth to take advantage of all that new technology has to offer.

fig 4.5.5 USB Camera

CHAPTER - 5

MODULE DESCRIPTION

5.1 CONVOLUTION NEURAL NETWORK

Convolutional neural networks (CNNs) are widely used in pattern and image-recognition problems as they have a number of advantages compared to other techniques. This white paper covers the basics of CNNs including a description of the various layers used.

A neural network is a system of interconnected artificial "neurons" that exchange messages between each other. The connections have numeric weights that are tuned during the training process, so that a properly trained network will respond correctly when presented with an image or pattern to recognize. The network consists of multiple layers of feature-detecting "neurons". Each layer has many neurons that respond to different combinations of inputs from the previous layer. By stacking multiple and different layers in a CNN, complex architectures are built for classification problems. Four types of layers are most common convolution layers, pooling/subsampling layers, non-linear layers, and fully connected layers.

Neural networks are inspired by biological neural systems. The basic computational unit of the brain is a neuron and they are connected with synapses. Compares a biological neuron with a basic mathematical model.

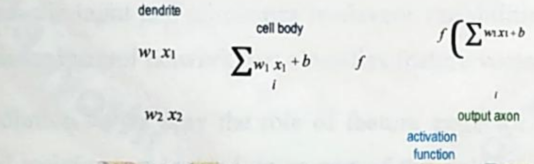


fig 5.1 Mathematical model of neuron

In a real animal neural system, a neuron is perceived to be receiving input signals from its dendrites and producing output signals along its axon. The axon branches out and connects via synapses to dendrites of other neurons.

When the combination of input signals reaches some threshold condition among its input dendrites, the neuron is triggered and its activation is communicated to successor neurons.

In the neural network computational model, the signals that travel along the axons (e.g.0) interact multiplicand- timely (e.g., w0x0) with the dendrites of the other neuron based on the synaptic strength at that synapse (e.g., w0). Synaptic weights are learnable and control the influence of one neuron or another. The dendrites carry the signal to the cell body, where they all are summed. If the final sum is

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology,
Hyderabad-500075

above a specified threshold, the neuron fires, sending a spike along its axon. In the computational model, it is assumed that the precise timings of the firing do not matter and only the frequency of the firing communicates information. Based on the rate code interpretation, the firing rate of the neuron is modelled with an activation function f that represents the frequency of the spikes along the axon. A common choice of activation function is sigmoid. In summary, each neuron calculates the dot product of inputs and weights, adds the bias, and applies non-linearity as a trigger function (for example, following a sigmoid response function).

A CNN is a special case of the neural network described above. A CNN consists of one or more convolutional layers, often with a subsampling layer, which are followed by one or more fully connected layers as in a standard neural network. The design of a CNN is motivated by the discovery of a visual mechanism, the visual cortex, in the brain. The visual cortex contains a lot of cells that are responsible for detecting light in small, overlapping sub-regions of the visual field, which are called receptive fields. These cells act as local filters over the input space, and the more complex cells have larger receptive fields. The convolution layer in a CNN performs the function that is performed by the cells in the visual cortex. A typical CNN for recognizing traffic signs. Each feature of a layer receives inputs from a set of features located in a small neighbourhood in the previous layer called a local receptive field. With local receptive fields, features can extract elementary visual features, such as oriented edges, end-points, corners, etc., which are then combined by the higher layers.

In the traditional model of pattern/image recognition, a hand-designed feature extractor gathers relevant information from the input and eliminates irrelevant variabilities. The extractor is followed by a trainable classifier, a standard neural network that classifies feature vectors into classes.

In a CNN, convolution layers play the role of feature extractor. But they are not hand designed. Convolution filter kernel weights are decided on as part of the training process. Convolutional layers are able to extract the local features because they restrict the receptive fields of the hidden layers to be local.

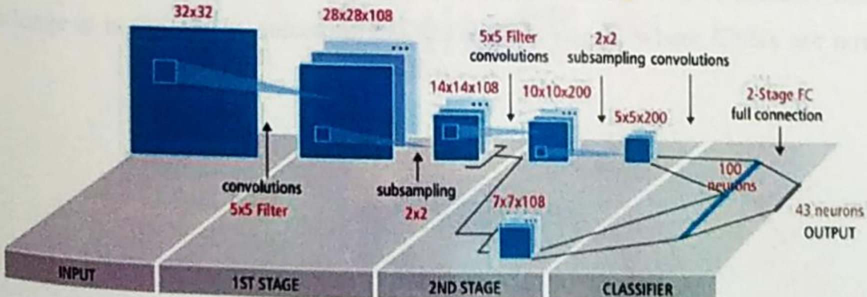


fig 5.1 Typical block diagram of CNN

CNNs are used in variety of areas, including image and pattern recognition, speech recognition, natural language processing, and video analysis. There are a number of reasons that convolutional neural networks are becoming important. In traditional models for pattern recognition, feature extractors are hand designed. In CNNs, the weights of the convolutional layer being used for feature extraction as well as the fully connected layer being used for classification are determined during the training process. The improved network structures of CNNs lead to savings in memory requirements and computation complexity requirements and, at the same time, give better performance for applications where the input has local correlation (e.g., image and speech). Large requirements of computational resources for training and evaluation of CNNs are sometimes met by graphic processing units (GPUs), DSPs, or other silicon architectures optimized for high throughput and low energy when executing the idiosyncratic patterns of CNN computation. In fact, advanced processors such as the Tensilica Vision P5 DSP for Imaging and Computer Vision from Cadence have an almost ideal set of computation and memory resources required for running CNNs at high efficiency.

In pattern and image recognition applications, the best possible correct detection rates (CDRs) have been achieved using CNNs. For example, CNNs have achieved a CDR of 99.77% using the MNIST database of handwritten digits, a CDR of 97.47% with the NORB dataset of 3D objects, and a CDR of 97.6% on ~5600 images of more than 10 objects. CNNs not only give the best performance compared to other detection algorithms, they even outperform humans in cases such as classifying objects into fine-grained categories such as the particular breed of dog or species of bird.

A typical vision algorithm pipeline, which consists of four stages: pre-processing the image, detecting regions of interest (ROI) that contain likely objects, object recognition, and vision decision making. The pre-processing step is usually dependent on the details of the input, especially the camera system, and is often implemented in a hardwired unit outside the vision subsystem. The decision making at the end of pipeline typically operates on recognized objects—It may make complex decisions, but it operates on much less data, so these decisions are not usually computationally hard or memory intensive problems. The big challenge is in the object detection and recognition stages, where CNNs are now having a wide impact.

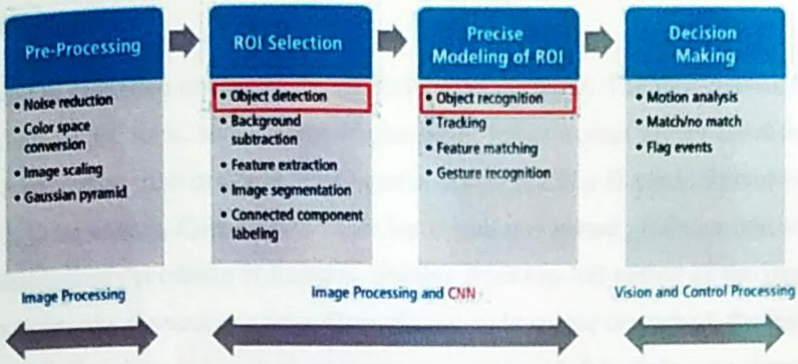


fig 5.1 Vision algorithm pipeline

5.2 Layers of CNN Structure

Four types of layers are most commonly used

1. Convolution
2. Pooling
3. Dense layers
4. Fully connected layers

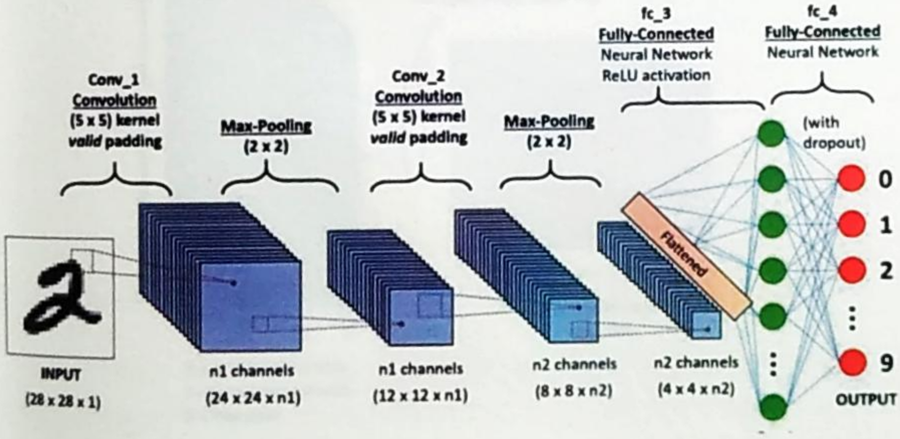


Fig 5.2 layers of CNN

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

Convolution layers

The convolution operation extracts different features of the input. The first convolution layer extracts low-level features like edges, lines, and corners. Higher-level layers extract higher-level features. Illustrates the process of 3D convolution used in CNNs. The input is of size N x N x D and is convolved with H kernels, each of size k x k x D separately. Convolution of an input with one kernel produces one output feature, and with H kernels independently produces H features. Starting from top-left corner of the input, each kernel is moved from left to right, one element at a time. Once the top-right corner is reached, the kernel is moved one element in a downward direction, and again the kernel is moved from left to right, one element at a time. This process is repeated until the kernel reaches the bottom-right corner.

For the case when N = 32 and k = 5 , there are 28 unique positions from left to right and 28 unique positions from top to bottom that the kernel can take. Corresponding to these positions, each feature in the output will contain 28x28 (i.e., (N-k+1) x (N-k+1)) elements. For each position of the kernel in a sliding window process, k x k x D elements of input and k x k x D elements of kernel are element-by- element multiplied and accumulated. So to create one element of one output feature, k x k x D multiply-accumulate operations are required.

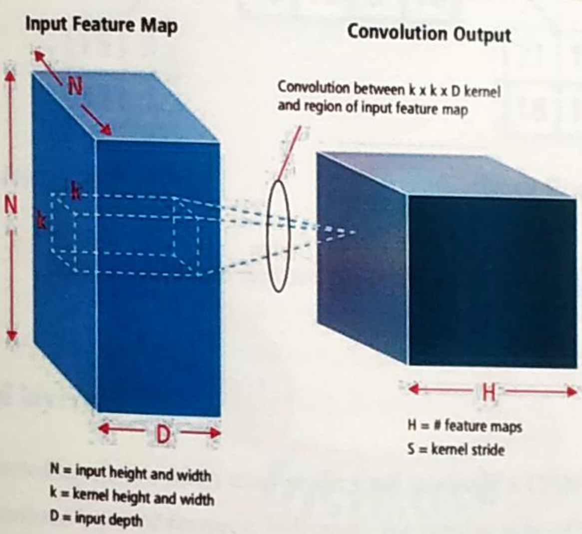


fig 5.1.1 Pictorial representation of convolution process

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

Pooling/subsampling layers

The pooling/subsampling layer reduces the resolution of the features. It makes the features robust against noise and distortion. There are two ways to do pooling: max pooling and average pooling. In both cases, the input is divided into non-overlapping two-dimensional spaces. Layer 2 is the pooling layer. Each input feature is 28x28 and is divided into 14x14 regions of size 2x2. For average pooling, the average of the four values in the region are calculated. For max pooling, the maximum value of the four values is selected and elaborates the pooling process further. The input is of size 4x4. For 2x2 subsampling, a 4x4 image is divided into four non-overlapping matrices of size 2x2. In the case of max pooling, the maximum value of the four values in the 2x2 matrix is the output. In case of average pooling, the average of the four values is the output. Please note that for the output with index (2,2), the result of averaging is a fraction that has been rounded to nearest integer.

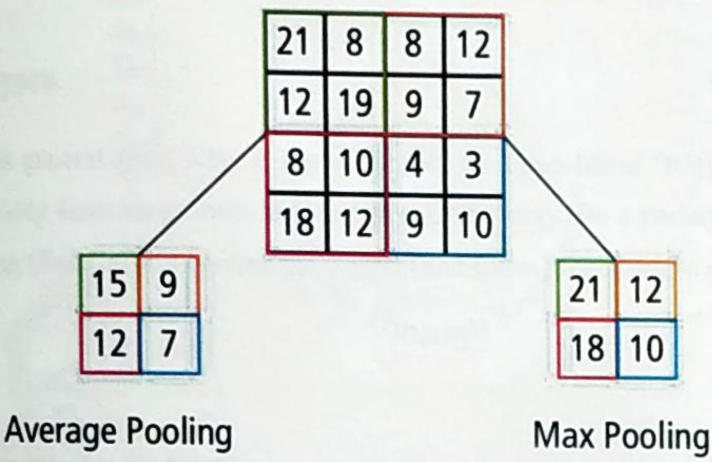


fig 5.1.2 Pictorial representation of max pooling and average pooling

Fully connected layers

Fully connected layers are often used as the final layers of a CNN. These layers mathematically sum a weighting of the previous layer of features, indicating the precise mix of "ingredients" to determine a specific target output result. In case of a fully connected layer, all the elements of all the features of the previous layer get used in the calculation of each element of each output feature. It explains the fully connected layer L. Layer L-1 has two features, each of which is 2x2, i.e., has four elements. Layer L has two features, each having a single element.

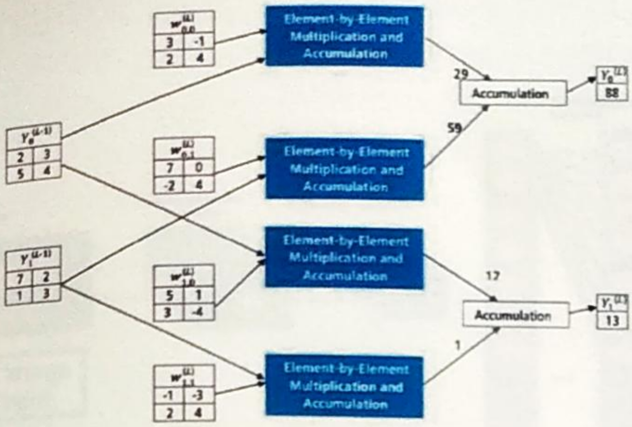


fig 5.1.3 Processing of a fully connected layer

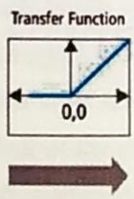
Non- linear Layers

Neural networks in general and CNNs in particular rely on a non-linear "trigger" function to signal distinct identification of likely features on each hidden layer. CNNs may use a variety of specific functions such as rectified linear units (ReLUs) and continuous trigger (non-linear) functions to efficiently implement this non-linear triggering.

ReLU

A ReLU implements the function y = max(x, 0), so the input and output sizes of this layer are the same. It increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer. In comparison to the other non-linear functions used in CNNs (e.g., hyperbolic tangent, absolute of hyperbolic tangent, and sigmoid), the advantage of a ReLU is that the network trains many times faster. ReLU functionality with its transfer function plotted above the arrow.

15	20	-10	35
18	-110	25	100
20	-15	25	-10
101	75	18	23



15	20	0	35
18	0	25	100
20	0	25	0
101	75	18	23

fig 5.1.4 Pictorial of RELU functionality

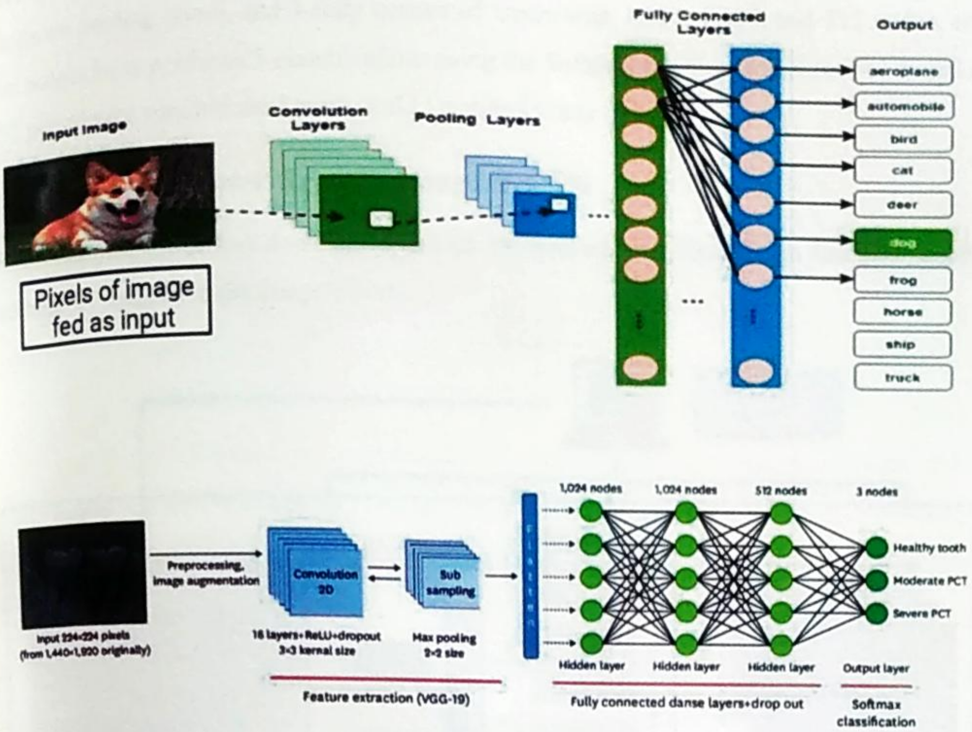


fig 5.2 Layers of CNN

The large amount of data collected by the tools for food quality and safety evaluation can be successfully processed by CNN. If suppose a picture with a dog and a grass background so if we provide this image to a deep learning architecture it will classify as a dog.

They are a kind of deep neural network which were designed. Researches were found how a human will perceive an image into the beam in different layers and that how this CNN is designed and it has been time and very efficient for all the image processing recognition.

Firstly an input image will go through a bunch of layers, first is Convolution layer. It will capture pixels of image from input to convolution layer and apply filter cl has a set of filter which are applied to a given input image that processed filter goes to the another layer is pooling layer which is a nonlinear down sampling layer which is the image has been shortened to half of the size and that same pixel will go to the layer. In pooling layer the nonlinear which is introduced information from that image is downed sampled layers the network connected eye are called fully connected layer which implies the layer where each and every node is connected to the next node in the coefficient. The coefficient will give the best cases for the object so at the end of fully connected layer it will convey with the probabilities and sum let's say like cat or deer.

Overall architecture of the deep CNN model. The dataset for the PCT images (224×224 pixels) is labelled as the input. Each of the convolutional layers is followed by a ReLU activation function, dropout, maximum pooling layers, and 3 fully connected layers with 1,024, 1,024, and 512 nodes, respectively. The final output layer performs 3 classifications using the Softmax function. CNN convolutional neural network, PCT periodontal compromised tooth, ReLU rectified linear unit.

How Does A Computer Read an Image in CNN

The image is broken down into 3 colour-channels which is Red, Green and Blue. Each of these colour channels are mapped to the image's pixel.

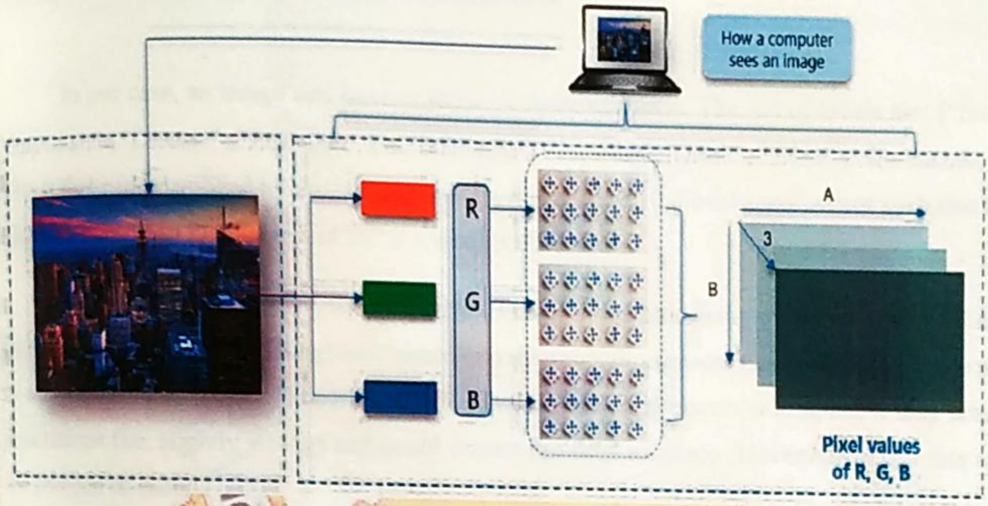


fig 5.2.1 Read an image in CNN

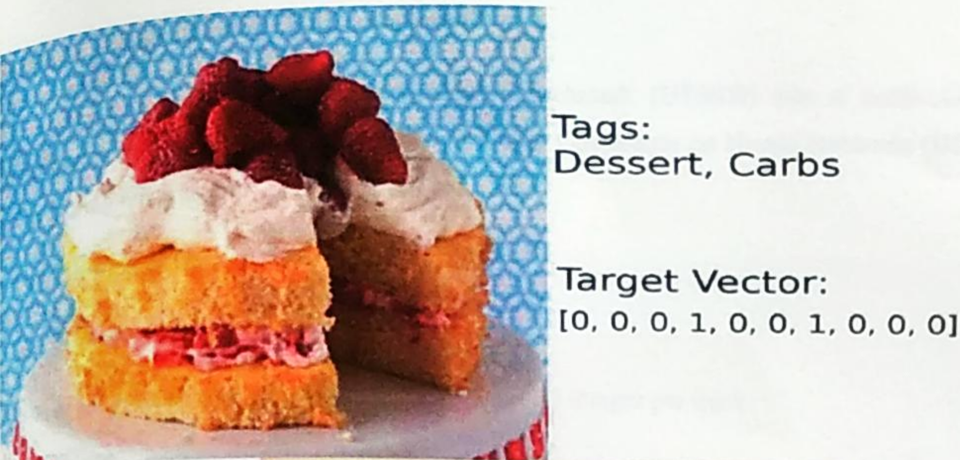
Then, the computer recognizes the value associated with each pixel and determine the size of the image. However, for black-white images, there is only one channel and the concept is the same.

HOW IT WORKS USING CNN

The accuracy of CNNs in image classification is quite remarkable and its real-life applications through APIs quite profound. The first step is to collect and clean the data. I sampled around 2000 images from the Recipes 5k dataset and resized them to size 224 x 224.

The original dataset had annotations of the ingredients of a food item. However, there were more than 1000 possible ingredients (i.e. labels) and this would create highly sparse label vectors. Hence, I created my own set of annotations for the same images.

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology,
Hyderabad-500075



In our case, an image can have at most 10 possible labels. The list of labels are: ["Soups", "Mains", "Appetizer", "Dessert" ,"Protein", "Fats", "Carbs", "Healthy", "Junk", "Meat"]. To encode the labels in a format that can be utilized by the neural network, we create a 10 dimensional vector such that there is a "1" if a label is present in the image and "0" if a label is absent.

To make the annotation process simpler, I made some bold assumptions such as: "All cake images are Desserts and have Carbs". This greatly simplified the annotation process, and i wrote a simple python function to carry most of the heavy lifting. While this strategy makes the process simpler, it may create some noisy annotations (i.e. slightly wrong) and could impact the final accuracy. Nevertheless, for this toy experiment, we proceed as such.

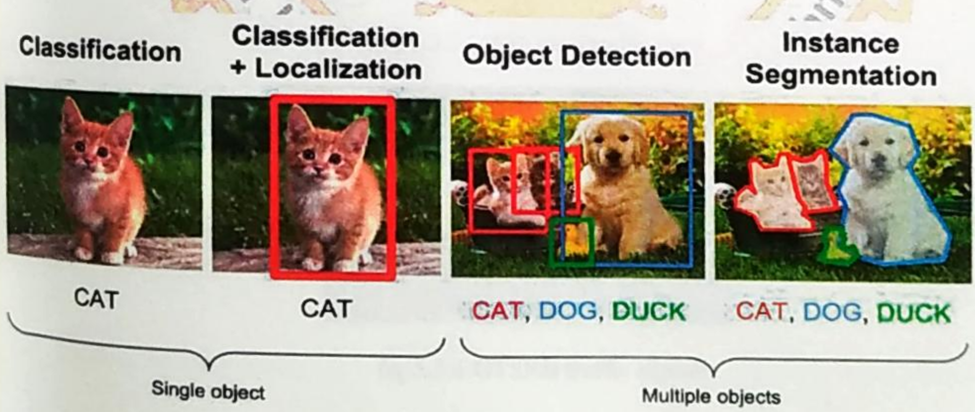


fig 5.3 Segmentation of CNN

Recognition Algorithm for GTSRB Dataset

The German Traffic Sign Recognition Benchmark (GTSRB) was a multi-class, single-image classification challenge held at the International Joint Conference on Neural Networks (IJCNN) 2011, with the following requirements:

• 51,840 images of German road signs in 43 classes

• Size of images varies from 15x15 to 222x193

• Images are grouped by class and track with at least 30 images per track.

• Images are available as colour images (RGB), HOG features, Har features, and colour histograms.

• Competition is only for the classification algorithm; algorithm to find region of interest in the frame is not required.

• Temporal information of the test sequences is not shared, so temporal dimension cannot be used in the classification algorithm.

fig 5.3.1 GTSRB ideal traffic signs

fig 5.3.1 GTSRB traffic signals

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology
Hyderabad-500075

5.3 DATABASE

We have collected the data from the google from different websites to find the estimation of values.

Here are some database we have collected.

fig 5.4 Database collected from different food items

Prediction:
[Healthy, Mains, Protein]

Actual:
[Healthy, Mains, Protein]

Prediction:
[Dessert, Carbs]

Actual:
[Dessert, Carbs]

Prediction:
[Healthy, Mains, Protein]

Actual:
[Healthy, Mains, Carbs, Protein]

Head of the Department
Department of Electronics and Communication Engg.
Vidya Jyothi Institute of Technology.
Hyderabad-500075

CHAPTER - 6

METHODOLOGY

6.1 OVERVIEW

In this section, we provide the details of the deep neural network method used in our system. The first step in our approach is to generate a pre-trained model file with the help of CNN network. This is performed by initially capturing a set of images of one particular class (e.g. 50 images of apple class) and then labelling them with object name-set (object being apple). These images will be considered as the set of relevant (positive) images and are used to train the system. In the second step of the training, we re-train the system with the set of negative images (images that do not contain the relevant object). In our case, we trained the system with the background images, so it does not categorize them as part of the apple class. Once the model file is generated from the training, we load it into the application and test it against the images captured and submitted by the user. The system then performs the image recognition process and generates a list of probabilities against the label name. The label with the highest probability is prompted to the user in the dialog box, to confirm the object name. Once the object name is confirmed, the system performs the calorie computation part by calculating the size of the food item with respect to the finger in the frame. It finally prints the output to the user with the required calorie.

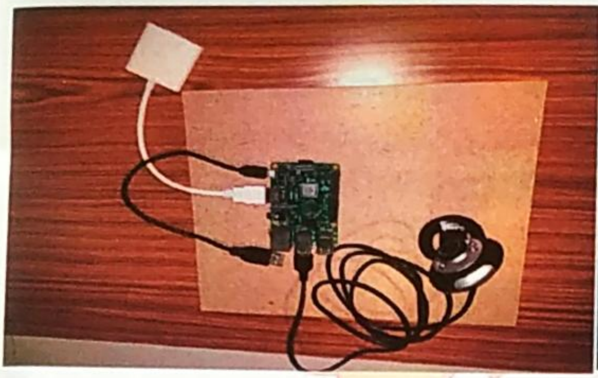
6.2 HISTORY

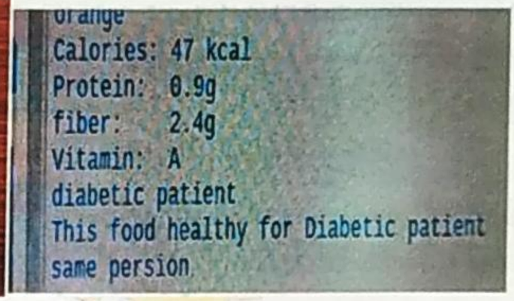
Deep Convolutional Neural Networks have been shown to be very useful for visual recognition tasks. Alex Net won the ImageNet Large Scale Visual Recognition Challenge in 2012, spurring a lot of interest in using deep learning to solve challenging problems. Since then, deep learning has been used successfully in multiple fields like machine vision, facial recognition, voice recognition, natural language processing etc. These range from being able to recognize had written digits to classifying plants using the images of leaves. The closest example of a food-related task we could find was a restaurant classification problem from Yelp. In their scenario, they want to classify the images of restaurants along some business attributes (e.g. restaurant is kid friendly, has table service etc.). Classifying food dishes from images presents several distinct characteristics. Image classification is the task of assigning a single label to an image (or rather an array of pixels that represents animate) from a fixed set of categories.

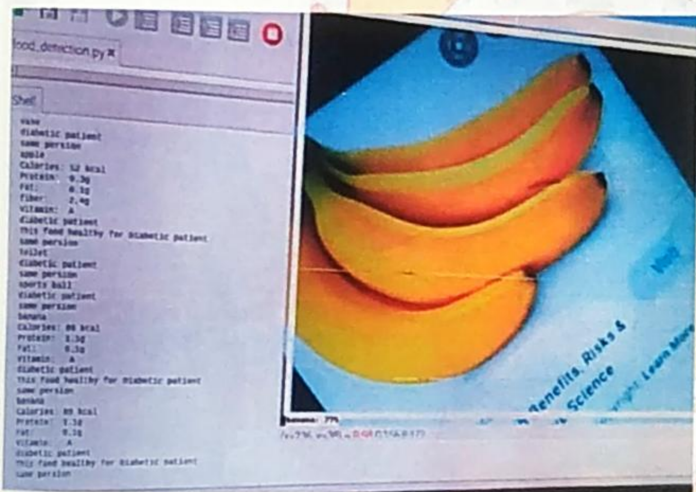
CHAPTER -7

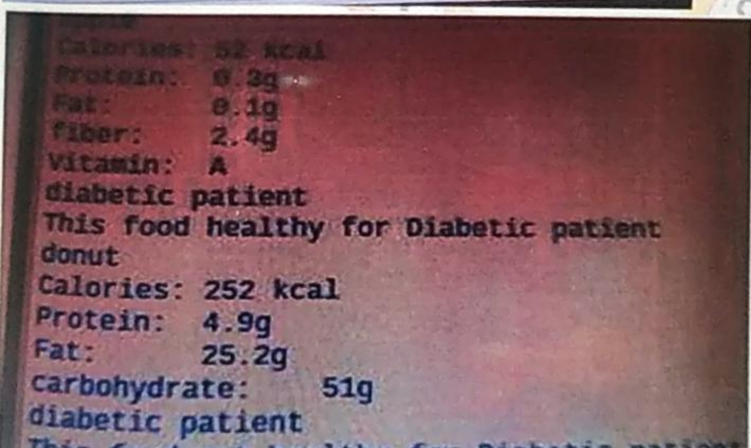
RESULT

7.1 OUTPUT SCREEN









Head of the Department
Department of Electronics and Communication
Vidya Jyothi Institute of Technology
Hyderabad-500075

The input image(food image) is fed to Convolutional neural network where it undergoes through different layers like convolutional layer, pooling layer, fully connected layer(fc layer) and a dense layer to categorize the dish and to find nutritional value of the food. These networks are trained using tensor flow framework. We trained up to 6 food items, they are Banana, Apple, Donut, Pizza, Sandwich, and Orange. Several images of each dish are used for training CNNs. Food dishes are categorized by CNNs and complete nutritional value is estimated.

The nutritional value which we got is compared with the nutritional value from WHO (world health organization). It also shows whether the food is good for diabetic patient or not. It warns the people if they are over consuming the food through LEDs their by reducing the risk of becoming obese. It shows the nutritional value of orange and if person is diabetic it displays "This food is healthy for diabetic patient.

The proposed system can detect food items and give the complete nutritional values and added a feature which facilitates in recognition of diabetic friendly food and also gives an alert if excess of food is consumed by a healthy person which is not present in existing systems. But the proposed system has a drawback that it cannot detect the food items which are shapeless like liquids.

CHAPTER - 8
CONCLUSION

This project shows that convolutional neural network can successfully solve food image classification problems to relatively small number of classes. Food according to the nutrients composition so that they may have a clear understanding of their diet and make necessary decisions to improve their health. To disseminate the nutritional values of foods by capturing the image of the food by using camera. Load the training images to the database. Convolutional neural networks are used to determine the food categories and estimate the nutritional value. CNN will take the pixel values of the image and it will compare to approximate pixel values and detect and classify the nutritional values to the input image and checks whether the diabetic person should be consume or not, if the diabetic patient should consume the food by taking nutritional values correctly then the output of the nutritional value will be displayed on monitor. Food calorie data is to help people easily to keep a daily food diary and calorie count and enabling them to better control their food intake and stick to a diet. For a diabetic patients to know the nutritional values like calories, fats, proteins, carbohydrates. Using these nutritional value we will maintain our health in a proper intake.

CHAPTER - 9

FUTURE WORK

As the field of Data Sciences and Embedded Systems are vastly expanding giving room to new technologies, there is always a possibility of improving this project using the facilities. The use of Convolutional Neural Network is the basis of recognition and classification of food images to discern patterns from images of food. For the future work, the model is to be developed and implemented according to the framework and datasets as stated. We shall use Convolution Neural Network as the algorithm to conduct the training and implementation of the deep learning model to make predictions on food nutrients composition. After data pre-processing, the chosen datasets, Food-101 UEC-256, and Yummly-28k, shall be used to train the model where patterns and characteristics of the food images are distinguished over multiple passes of the neural network. Once that is completed, testing is done to ensure the model achieves accuracy in prediction and fit for deployment in a food nutrients classifier web application.

REFERENCES

- Yanai, K., Kawano, Y., 2015. Food image recognition using deep convolutional network with pre-training and fine-tuning, in Multimedia and Expo Workshops (ICMEW), 2015 IEEE International Conference on.
- Kagaya, H., Aizawa, K., Ogawa, M., 2014. Food detection and recognition using convolutional neural network, in Proc. of ACM International Conference Multimedia, pp. 1085-1088.
- Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y., 2016. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment, in International Conference on Smart Homes and Health Telematics, pp. 37-48.
- Hou, L., Wu, Q., Sun, Q., Yang H., Li, P., 2016. Fruit Recognition Based On Convolution Neural Network, in 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 18-22.
- Krizhevsky, A., Sutskever I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in Advances in Neural Information Processing Systems.
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition, in Proceedings of the IEEE 86(11): 2278-2324.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.
- P. Pandey, A. Deepthi, B. Mandal, and N. B. Puhan, "Foodnet: Recognizing foods using ensemble of deep networks," IEEE Signal Process. Lett., vol. 24, no. 12, pp. 1758–1762, Dec. 2017.
- J. Chae et al., "Volume estimation using food specific shape templates in mobile image-based dietary assessment," in Proc. Comput. Imag. IX, 2011, vol. 7873, Art. no. 78730K.
- R. Joseph, D. Santosh, G. Ross and F. Ali., "You Only Look Once: Unified, Real-Time Object Detection". Conference: 2016 IEEE.