A Seminar Report on :

"Deep Learning : Convolutional Neural Network for Image Recognition"

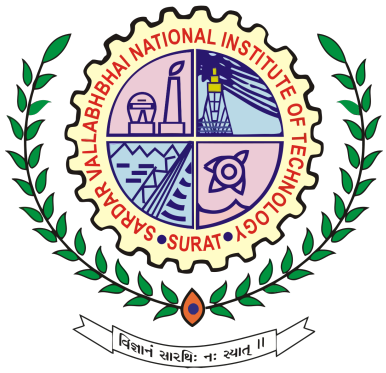
Prepared by : Hardik Sisodiya

Roll. No. : U17CO072

Class : B.Tech-IV (Computer Engineering) 7th Semester

Year : 2020-21

Guided by : Dr. Krupa N. Jariwala



Department of Computer Engineering

Sardar Vallabhbhai National Institute of Technology

Surat - 395007 (Gujarat), India

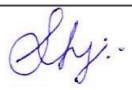
Student Declaration

This is to certify that the work described in this seminar report has been actually carried out by

Sr.	Admission No.	Student Name
1.	U17CO072	Hardik Sisodiya

Neither the source code there in, nor the content of the seminar report have been copied or downloaded from any other source. I understand that my result grades would be revoked if later it is found to be so.

Signature of the Student:

Sr.	Student Name	Signature of the Student
1.	Hardik Sisodiya	

Sardar Vallabhbhai National Institute of Technology,
Surat - 395007 (Gujarat), India

CERTIFICATE

This is to certify that the seminar report entitled **Deep Learning : Convolutional Neural Network for Image Recognition** is prepared and presented by **Mr. Hardik Sisodiya** bearing Roll. No. : **U17CO072**, 4th Year of **B. Tech (Computer Engineering)** and his work is satisfactory.

14/12/2020

GUIDE	JURY	HOD
(Prof Dr. Krupa N. Jariwala)		COED

Contents

List of Figures

List of Symbols

i	class number
K	no. of classes in multi-class classification
N	no of layers in neural network
W	weight of edge of neuron
X	Vector of input features
Y	Vector Output of model

List of Acronyms

OCR Optical Character Recognition

IR Image Recognition

PIL Python Image Library

ReLU Rectified Linear Unit

ALVINN An Autonomous Land Vehicle in Neural Network

RF Random Forest

TanH Hyperbolic Tangent

LR Logistic Regression

BLR Binary Logistic Regression

MLR Multinomial Logistic Regression

OLR Ordinal Logistic Regression

SVM Support Vector Machine

NN Neural Network

ANN Artificial Neural Network

CNN Convolutional Neural Network

GUI Graphical User Interface

WWW World Wide Web

Abstract

Abstract: Artificial Intelligence and Machine learning are growing fields in the era of new generations of computer. Both these fields are giving human a new friendly way for living and helping them in their daily life. Deep Learning is a part of Machine Learning which consists of various algorithms that deal with the unstructured and unlabeled data. Deep Learning actually helps computers to mimic the working of human brain in processing the data for use in image recognition, speech recognition, natural language processing, detection of various objects in the image. Recently, there are so many deep learning algorithms proposed to solve traditional artificial intelligence problems. This work aims to make the state-of-the-art image recognition system using various deep learning algorithm. Convolutional Neural Network is the most popular deep learning algorithm used to work with the images. This seminar work first gives the overview about Deep Learning, Image Recognition, various classification algorithms and implementation of image recognition system using convolutional neural network. Finally, the seminar summarizes applications of image recognition and advantages, disadvantages of this system.

Keywords: Deep Learning, Convolutional Neural Network, Classification Algorithm, Image Recognition, GUI.

Chapter 1

Introduction

Humans uses various sense to get the most of the information about the environment surrounding us. Using the eyes, human can get the visual information about the different-different objects and scenes in the world. By gathering visual information, we can easily make decisions like to avoid obstacles for ex lamp-post on the road. Human brain helps in understanding and making informed decisions based on the visual information captured by the eyes. So that we can use this phrase "What we see is what we get" for the human beings. Because for human brain, it is easy to process the information and get details about the objects captured by eyes. It does not take any effort in processing this type of functionality. But for machine and digital devices, this phrase cannot be used because they cannot directly understand what is captured by the camera.

Since we are living in the age of digital era, we would like to train and make the digital devices that can be used to provide the same functionality as human brain. Because of advancements in technology like Computer Vision, Machine Learning and Artificial Intelligence, it's become possible to teach the digital devices the same functionality.

Computer vision helps devices to capture the visuals like human and take decision or actions based on this visual information. This technology developed in devices is known as Image Recognition. This technology uses the pixel value of particular image and pattern analysis to recognize or classify the object in the image. Optical Character Recognition (OCR) can be one possible example of image recognition because in this, device will try to understand and extract the text from the image.

Chapter 2

What is Deep Learning?

To understand about the deep learning, first we have to go through the basics of machine learning because it is a part of machine learning. There are two definitions, one is proposed by Arthur Samuel other is defined by Tom Mitchell.

1. **Definition by "Arthur Samuel":** Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.

2. **Definition by "Tom Mitchell":** A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Deep Learning is a part of machine learning technique that enables computers to work same as how human brain naturally works. It is based on artificial neural network. This type of model can be trained by feeding large set of labeled datasets and neural architecture that contains many layers. Deep Learning mimics the human brain to give same ability to the computers. There are so many new areas where deep learning is very useful like classification of images, face recognition, speech recognition, language translation, object detection.

Now a days, there are so many advancements and researches going on in this field and that's why we can achieve the state-of-the-art model or software that sometimes gives better performance than human. This technology is growing in so many areas like autonomous vehicle, health-care, manufacturing, robotics, drones, military and surveillance.

Chapter 3

What is Image Recognition?

As we know that, Human can easily recognize the images of different things and animals like dog, cat, lion, etc. This process can be done by using eyes and brain of human. First eyes will capture the visual information and this data flows from retina to human brain via optical nerve. Optical nerve provides medium for data to flow from eyes to the brain. After that there will be some process to assemble this raw data into the facts and objects. And this is how brain actually recognize the different images.

Image recognition is the ability for computers to grab the visual raw information from the photograph like human eyes and identify or understand what's in the photograph like human brain does.

It is process of putting the image as input to some kind of model or software and getting the result of some kind of label for that image. This label of image must be the predefined into the software of model.

Chapter 4

Various Classification approaches to implement Image Recognition

4.1 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

Therefore, a RF is a classifier consisting in a collection of tree structured classifiers which uses random selection in two moments. In a first step, the algorithm selects several bootstrap samples from the historical data. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Therefore, in random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random by additionally using random thresholds for each feature rather than searching for the best possible thresholds.

4.2 Logistic Regression

Logistic regression is a statistical model that in its basic form uses a logistic regression to model a binary dependent variable.

It extends the ideas of multiple linear regression to the situation where the dependent variable, y, is discrete. In logistic regression no assumptions are made concerning the distribution of the independent variables

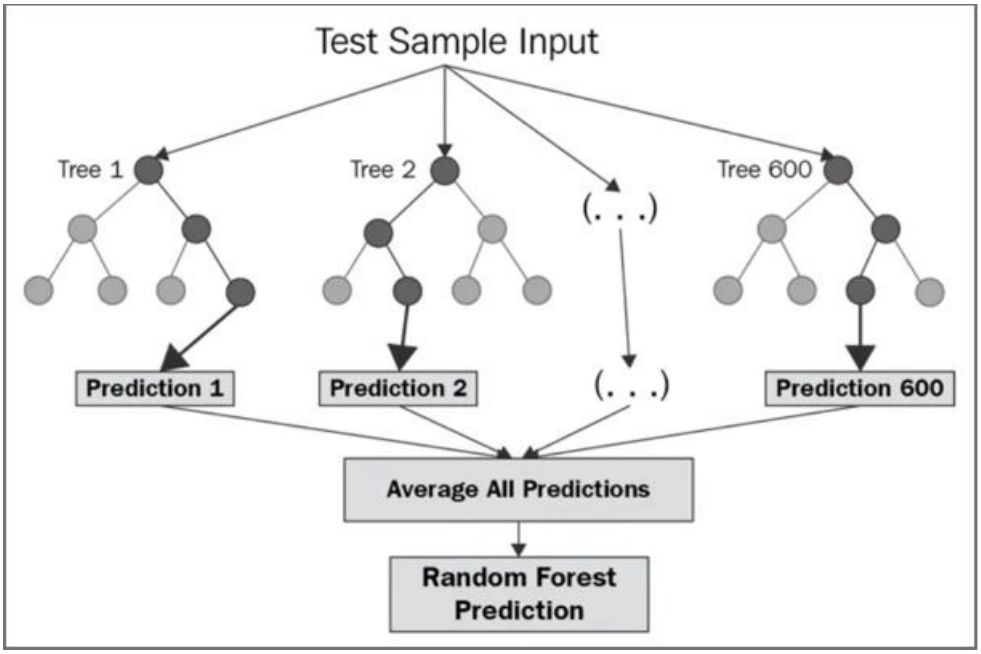


Figure 4.1: Random Forest

There are three types of logistic regression which can be listed as:

1. Binary logistic regression
2. Multinomial logistic regression
3. Ordinal logistic regression

Binary logistic regression can be used when there are only two classes in which we can distribute the images.

1. Email: Spam / Not spam
2. Tumor: Malignant / Benign

There are so many uses of multinomial logistic regression like.

1. Medical Diagrams: Not ill, cold, Flu
2. Weather: Sunny, Rain, Cloudy, Snow

In our technology of image recognition, we have to classify images into more than two classes like animal, automobile, plane, flowers, etc. So that we can use multinomial logistic regression to group the different classes of images.

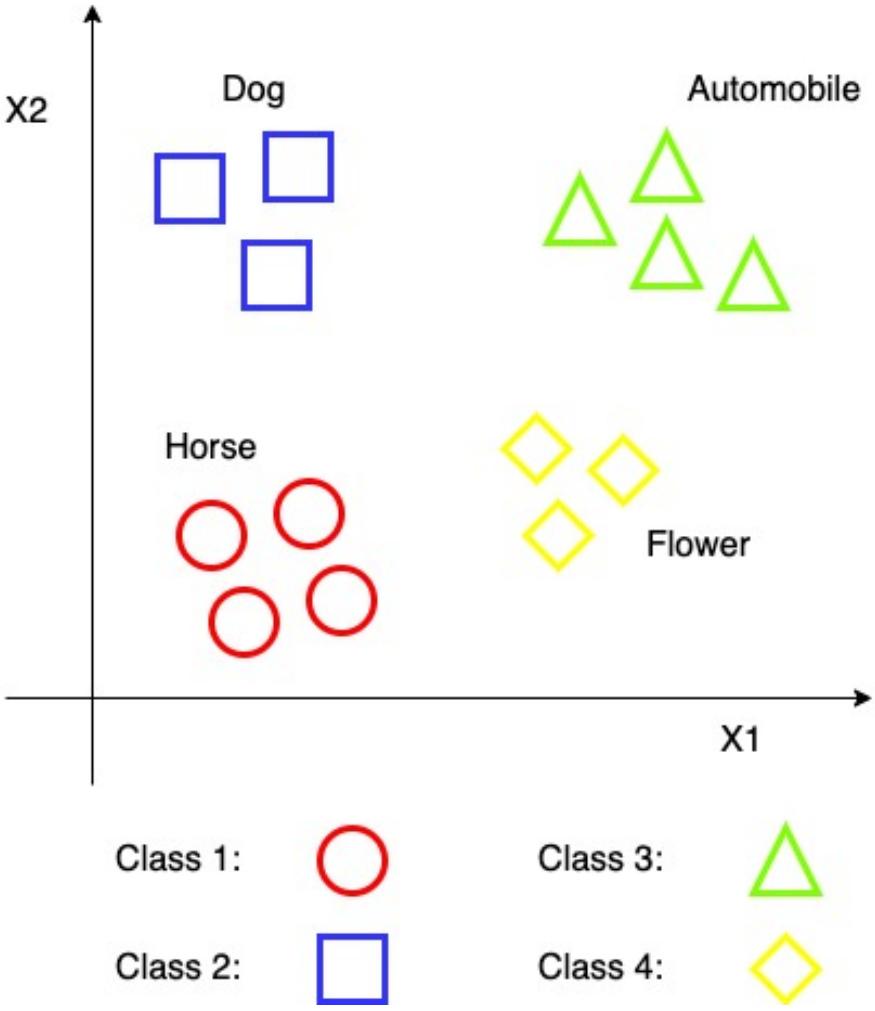


Figure 4.2: Multinomial Logistic Regression

4.2.1 Training Phase

Hypothesis can be defined as below:

$$Htheta(i)(X) = P(y = i|x;theta), where(i = 1,2,3,4) \tag{4.1}$$

We can train the multinomial logistic regression classifier Htheta(i)(X) for each class i to predict the probability that y = i.

4.2.2 Prediction

After training the machine learning model, we can feed new input image file to the model and pick the class i that maximizes the given hypothesis Htheta(i)(X).

Result is class i where

$$maxi(Htheta(i)(X)). \tag{4.2}$$

4.3 Support Vector Machine

Support Vector Machine is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems.

In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

Here we can use SVM model to choose the right hyper-plane because we are trying to separate the two types of URLs. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

In SVM, it is easy to have a linear hyper-plane between these two classes. But another burning question which arises is, should we need to add this feature manually to have a hyper- plane. No, SVM has a technique called the kernel trick. These are functions which takes low dimensional input space and transform it to a higher dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels.

It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then find out the process to separate the data based on the labels or outputs you've defined.

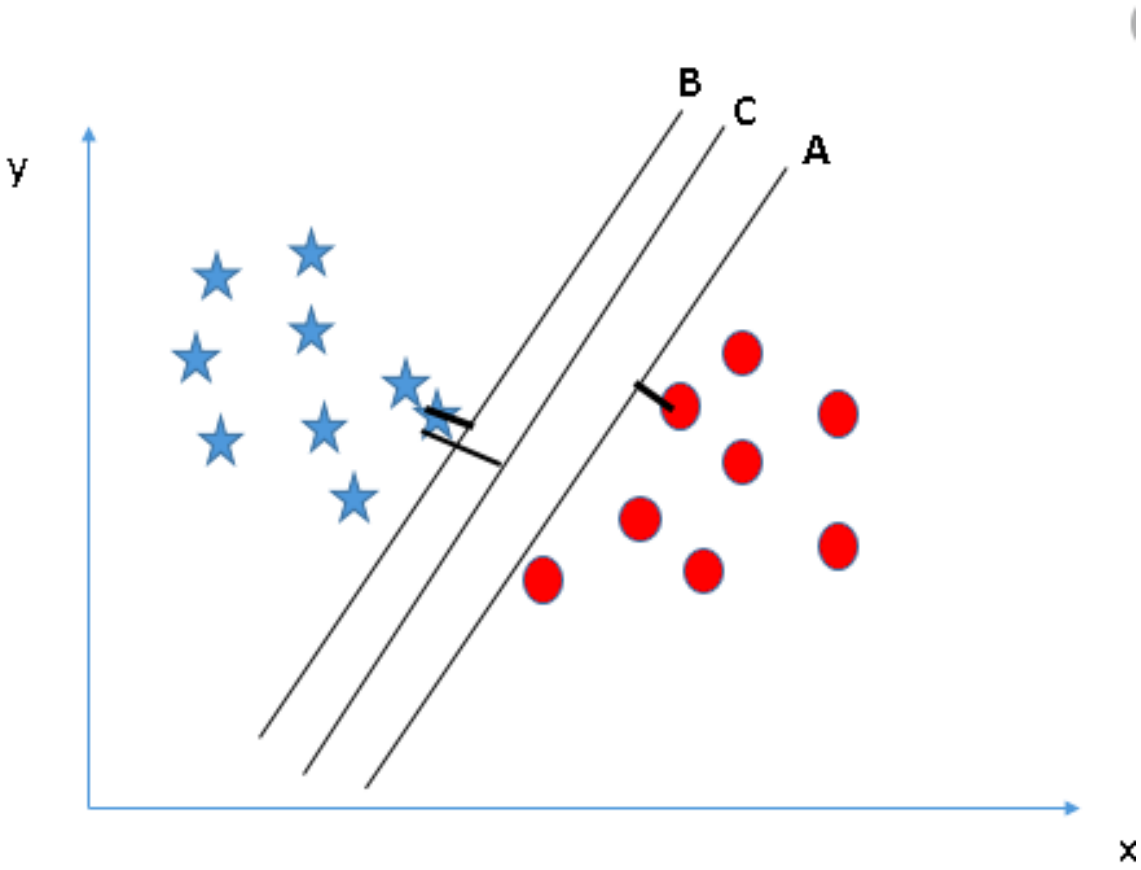


Figure 4.3: Support Vector Machine

4.4 Neural Network

A neural network is made up of individual nodes called neurons. A group or series of these neurons is known as layers. In this network, neurons of each layer are connected with neurons of following layer. Data flows from input layer to output layer along these connections. Usually there is one input layer and one output layer. More the number of hidden layers better the result. Simple representation of neural network can be shown in following figure 4.4.

The neural network takes set of input values in the input layer. And after that in training, each neuron is trained to perform simple mathematical calculation and then feed the results to all the neurons it's connected.

At each node, function is implemented on the input value in order to generate the output. This function is known as activation function. Activation function decides which inputs from previous layers are important enough to feed to the next layer. There are so many activation functions available according to developer's need. Number on each edge represents the weight of the edge. Input of neurons first will be multiplied by the weight of respective input neurons and then it goes into the activation function.

There are mainly three types of activation functions.

1. Binary step function
2. Linear Activation function
3. Non-linear Activation function
    (a) Sigmoid / logistic
    (b) TanH
    (c) ReLU
    (d) Leaky ReLU
    (e) Parametric ReLU
    (f) SoftMax
    (g) Swish

There is another function called as loss function. Loss function is used to calculate the loss which is nothing but a prediction error of neural network.

A neural network is known as one of most popular supervised machine learning algorithms

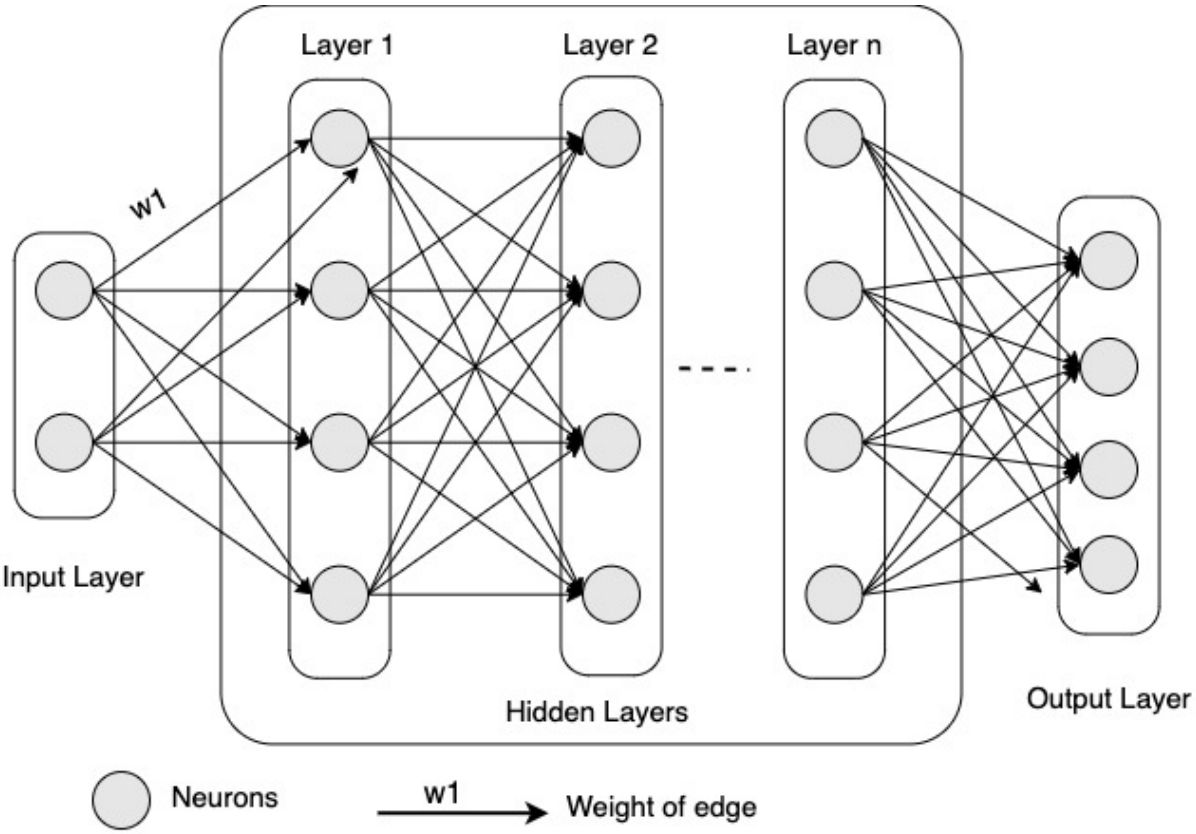


Figure 4.4: Neural Network Architecture

N = total number of layers, K = no. of classes in multi-class classification

Y (output) can be one of these [1 0 0 0] - Dog, [0 1 0 0] - Horse [0 0 1 0] - Flower, [0 0 0 1] - Automobile

to work with the images because neural network is very efficient in computation of higher dimensional data. Image can be represented as two-dimensional array of numbers (pixel values). So that neural network can be the best classification algorithm to implement the image recognition.
We can classify the neural network into main six types on the basis of mathematical operations and a set of parameters required to determine the output.

1. Feedforward Neural Network
2. Convolutional Neural Network
3. Recurrent Neural Network
4. Modular Neural Network
5. Multilayer Perceptron
6. Radial Basis Function Neural Network

Chapter 5

Step-wise Implementation of Image Recognition Technology

Every machine learning project can be implemented by following the given steps:

1. Import the dataset
2. Clean the data
3. Split dataset into training and testing
4. Create a model
5. Train the model
6. Make predictions
7. Evaluate the model

5.0.1 Complete workflow of Image recognition using machine learning

Workflow of image recognition mostly follows all the steps of machine learning project which are mentioned as above. Complete workflow of image recognition technology is shown in figure 5.1 given below.

Now for this particular implementation, we can use CIFAR-10 dataset which is publicly available on the internet. It is set of 60k color images with the size of 32*32 in 10 classes. This dataset is already labeled with class name of images. There is total 6000 images of one image in the dataset. This dataset contains the classes of image like automobile, dog, horse, airplane. We can split this dataset into the training and testing dataset to improve the accuracy of the model by using the different libraries like TensorFlow and Keras framework. By importing the cifar10 module, a part of keras.datasets package, we can load this dataset directly in form of higher-dimensional array of pixel intensity of this images. After that we can feed these numerical values to train the model. But before that we will split data into training and testing dataset.

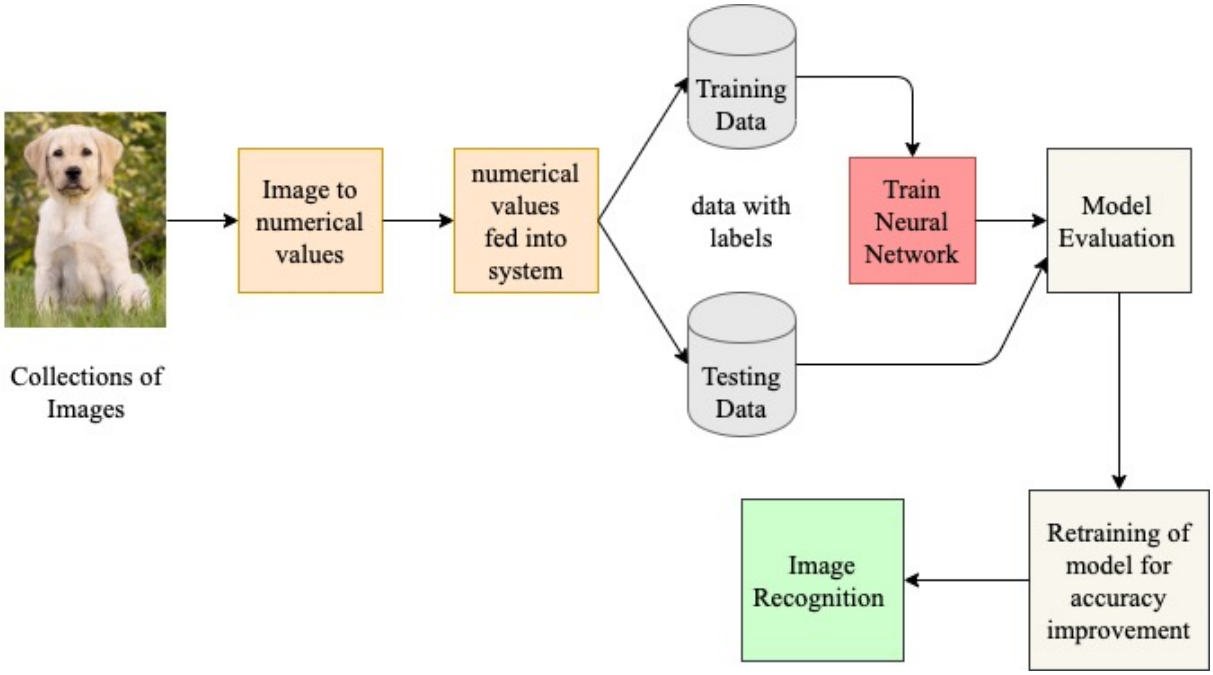


Figure 5.1: Work-Flow of Image Recognition

Only training data can be used to train the model. We can use mean squared error (MSE) function to find error in the predictions with respect to original result. Finally, we can use different percentage of split of training and testing dataset to retrain the model and to improve the accuracy model.

5.0.2 Difference in visual perception of image

Human can directly interpret the objects in the image with the help of brain. But for device it cannot understand the visual information directly.

Digital image is composed of element known as pixels, each with some discrete properties of numerical representation for its intensity or grey scale. Camera captures the visuals in terms of two-dimensional pixel intensity values. Computer or digital device will recognize each image by its pixel values. Machine learning model uses these numerical values in order to find out the similar pattern and regularities between the images. So, the difference between interpretation of eye and camera in terms of visual information is shown in figure 5.2 below.

As there is wide range of values in pixel intensity (0-255), we can normalize these values after dividing by 255. So, all values will be mapped in the range of (0, 1) and it will be easy to understand these values.

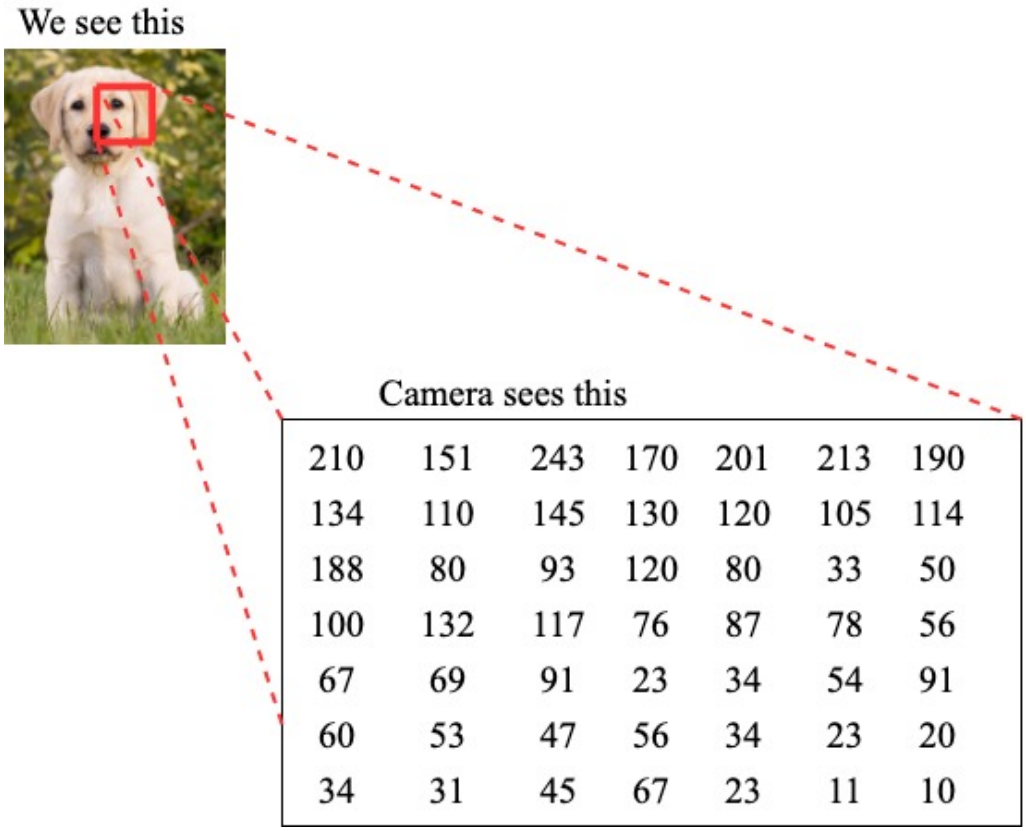


Figure 5.2: Difference between interpretation of visual information by human and camera

5.0.3 Architecture of convolutional neural network

Now after importing and cleaning the data in the workflow, it's time to create a machine leaning model. This technology requires the complex calculations in order to find the pattern because of high-dimensional data. Neural network will be the best choice of classification algorithm to develop the machine learning model.

From the various types of neural networks, Convolutional neural network in particular can be used to work with images because it contains several learnable image filters that help in analysis of similar patterns in the image.

For this experiment, given figure 5.3 shows the entire architecture of the convolutional neural network.

5.0.3.1 Input and Output Layer

There will be one input layer and one output layer. Input layer contains the array of pixel intensity values that we have derived by pre-processing the image. Here we are dealing with multi-class classification problem so that output layer contains the 10 neurons (nodes) because there are 10 pre-defined classes in the dataset. Output layer will generate the output in terms

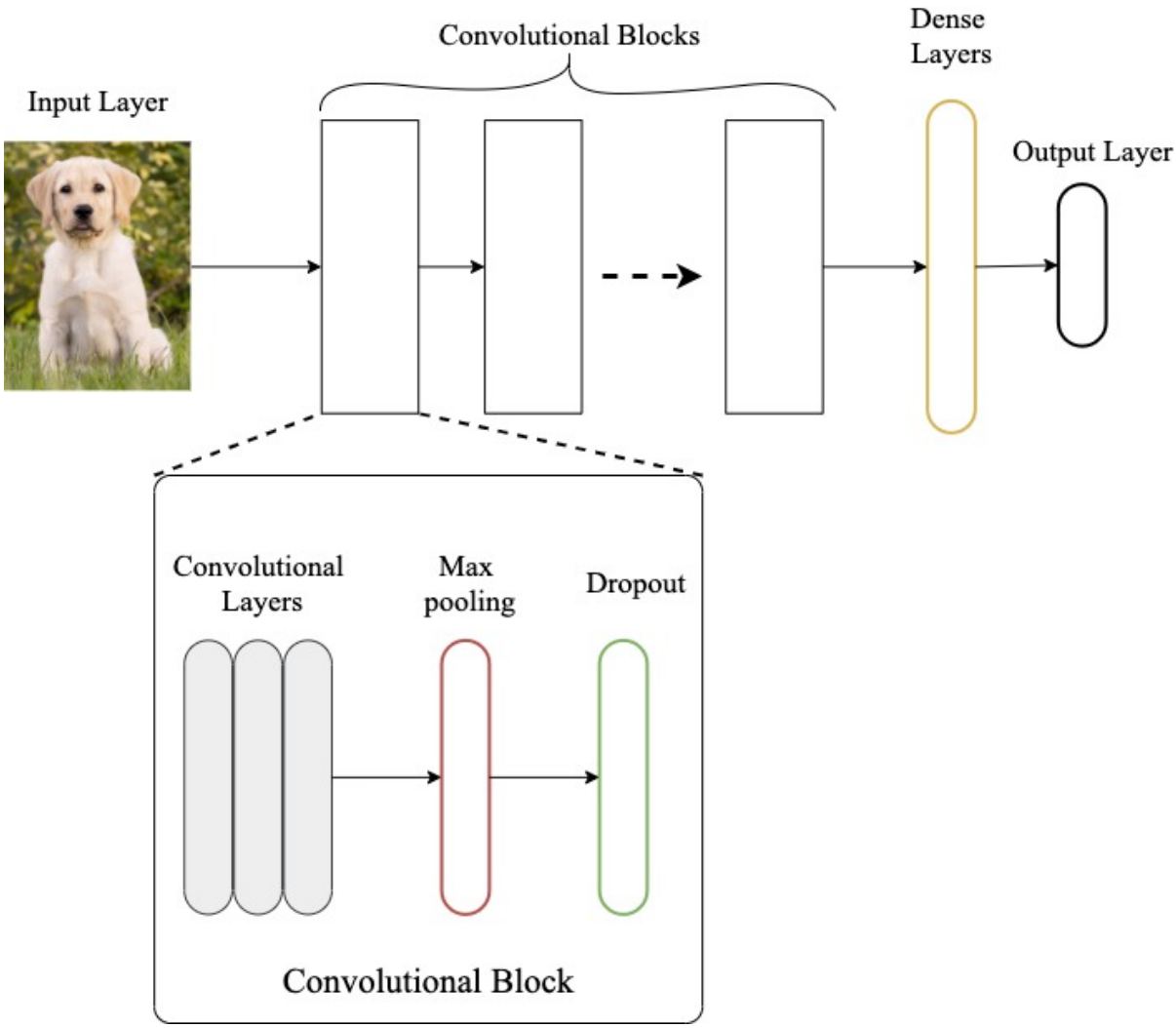


Figure 5.3: Convolutional Neural Network Architecture

of vector form and only 1 element of the vector will be 1 and other will be 0. Neuron with value 1 can be used for indexing in order to find the class name from the array of 10 names of pre-defined classes.

5.0.3.2 Hidden Layers

Other than input and output layer, we can add more than one convolution blocks to increase the efficiency of pattern analysis as you can see in the figure. Convolutional block consists of convolutional layers, max pooling layers, dropout layers, etc.
We can use neural network that contains only dense layers for image recognition but it won't be efficient in accuracy because objects like dog, cat, automobile can appear in lots of different places in the image. One solution is that we can add one or more convolutional layers.

1. **Convolutional Layers:** Convolutional layer helps detect the object no matter where object appears in image. It gives exact pixel location of object in the image using the

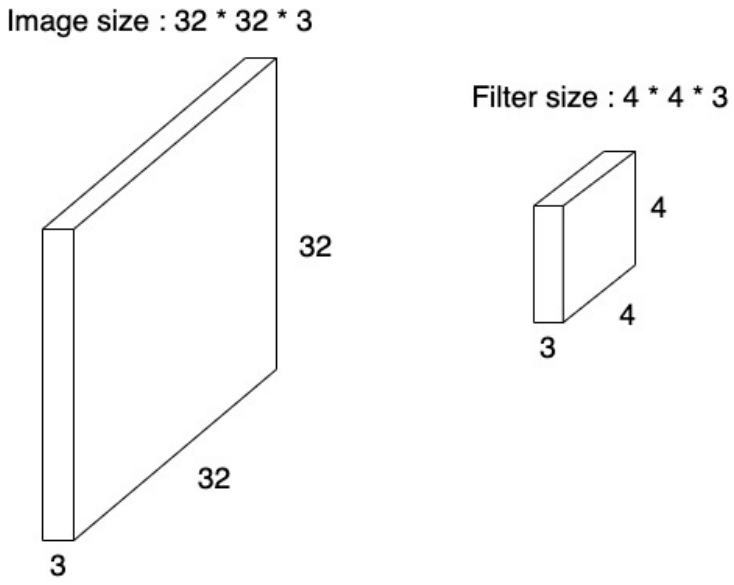


Figure 5.4: Filter of size 4*4*3 dividing the image

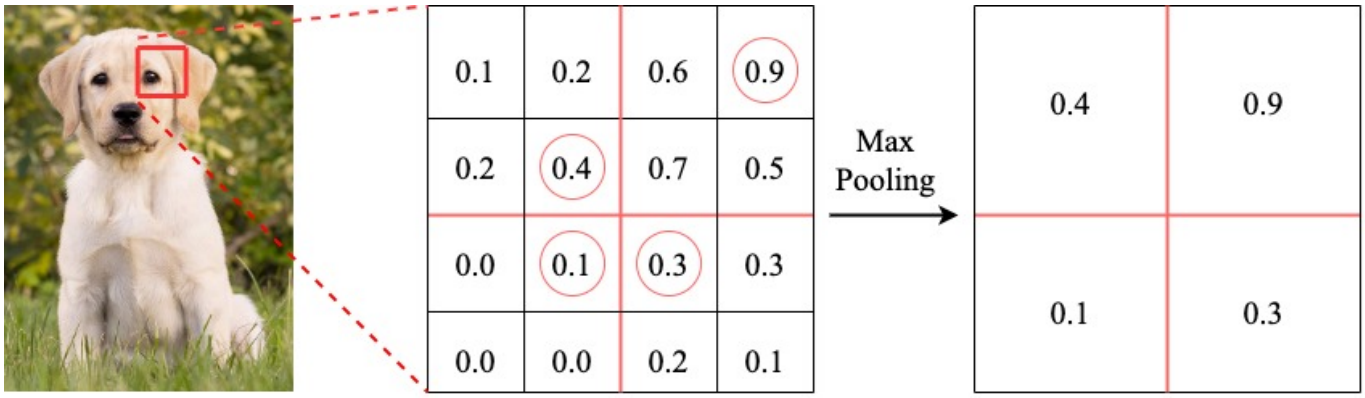


Figure 5.5: Max pooling Layers

filter because filter scans the whole image pixels and gathers information in the batch of photos. These layers look for the particular patterns in the image and record whether they found those patterns in each part of the image. Functionality of filters in convolutional layer is shown in figure 5.4

2. **Max pooling Layers:** As we know, convolutional layer gives exact location of object in image but there is no need of specific location. What is we need is rough location of object so that we can increase the efficiency of model because there will be less work for computer to find exact location. This technique is called max pooling.
   As shown in figure 5.5, camera captures the pixel intensity values into grid of just eye of dog and these values depend upon the actual image of eye.

   Idea of max pooling is that we divide the whole grid values into 2*2 squares and after that

we just select the most important bits from the squares. This selection of bits depends upon the pooling function which can be either max or average of all bits. After the max pooling of data, we just create the new array, store them and we still get the rough idea of location of object in image but with only 1/4th of the data. So, we are getting the same result but work for computer will be less in following layers of neural network.

In short, pooling operation uses sliding a two-dimensional filter over the image and summarizing the object lying within the region covered by filter. Pooling layer helps in reducing the number of parameters and complex computation by down-sampling.

3. **Dropout Layers:** There is one more way by which we can make our neural network more robust. It is known as Dropout. Problem with NNs is that they tend to memorize the input numerical data instead of actually learning how to tell different objects apart. The solution is that we can dropout layers between other layers that is used to throw away some of the input data passing through it by cutting the some of the connections in the NNs. Because of this, NN is forced to learn the actual patterns in the images than just memorizing the input data. It is known as dropout because we are letting some of the data to drop out of network randomly.

   After summarizing the all layers, we can say that convolutional layers add translational invariance, max pooling layers down the sample data, dropout layers forces NN to learn in a more robust way.

4. **Dense Layers:** This is the simplest regular type of layers that any NN contains. In the dense layers, each neuron (node) of current layer is connected with each and every node of previous layers. As these layers are densely connected, they are known as dense layers. In this NN, dense layer maps the output of previous layers to the output layer which helps in prediction of class of the image.

We can create this type of architecture of convolutional neural network using the different libraries like TensorFlow and Keras framework. To create the same architecture shown in figure 5.3 above, we can use the following Python code which is shown in the figure 5.6 below.

```
# Create a model and add layers
model = Sequential()

model.add(Conv2D(32, (3, 3), padding='same', input_shape=(32, 32, 3), activation="relu"))
model.add(Conv2D(32, (3, 3), activation="relu"))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Conv2D(64, (3, 3), padding='same', activation="relu"))
model.add(Conv2D(64, (3, 3), activation="relu"))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(512, activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(10, activation="softmax"))
```

Figure 5.6: Sample code to build CNN

Chapter 6

Graphical User Interface of technology

Some advanced technology can be developed but it is very complex to use for the naïve users. Like in image recognition, model can be trained using the advanced libraries but to use this technology without any interface, it will be very difficult because users have to give input in terms of image and without the knowledge of libraries, one cannot easily use this technology.

Machine learning model can be trained using the libraries like TensorFlow and Keras framework to classify the images. But to provide the model in terms of interface that is easily accessible to the user, we can use different libraries like TKinter, Pathlib, PIL to develop the whole graphical user interface.

6.1 Libraries used for creating GUI

1. **TKinter:** This library provides the functionality to create and develop graphical user interface.
2. **PathLib:** It is used to access the file system of a particular digital device like mobile, computer and upload the image by selecting it.
3. **PIL:** This library provides the functions to display and work with images in Python.

First of all, we can train the neural network model using TensorFlow and Keras by splitting the data into training and testing datasets. After that user can upload the image that he/she wants to classify by selecting particular image file with the help of Pathlib library and after that model will show the result on the GUI as shown in below figure 6.1.

As shown in figure 6.1, Class of the image will be shown in the GUI itself and also prints it into the terminal which is Airplane.

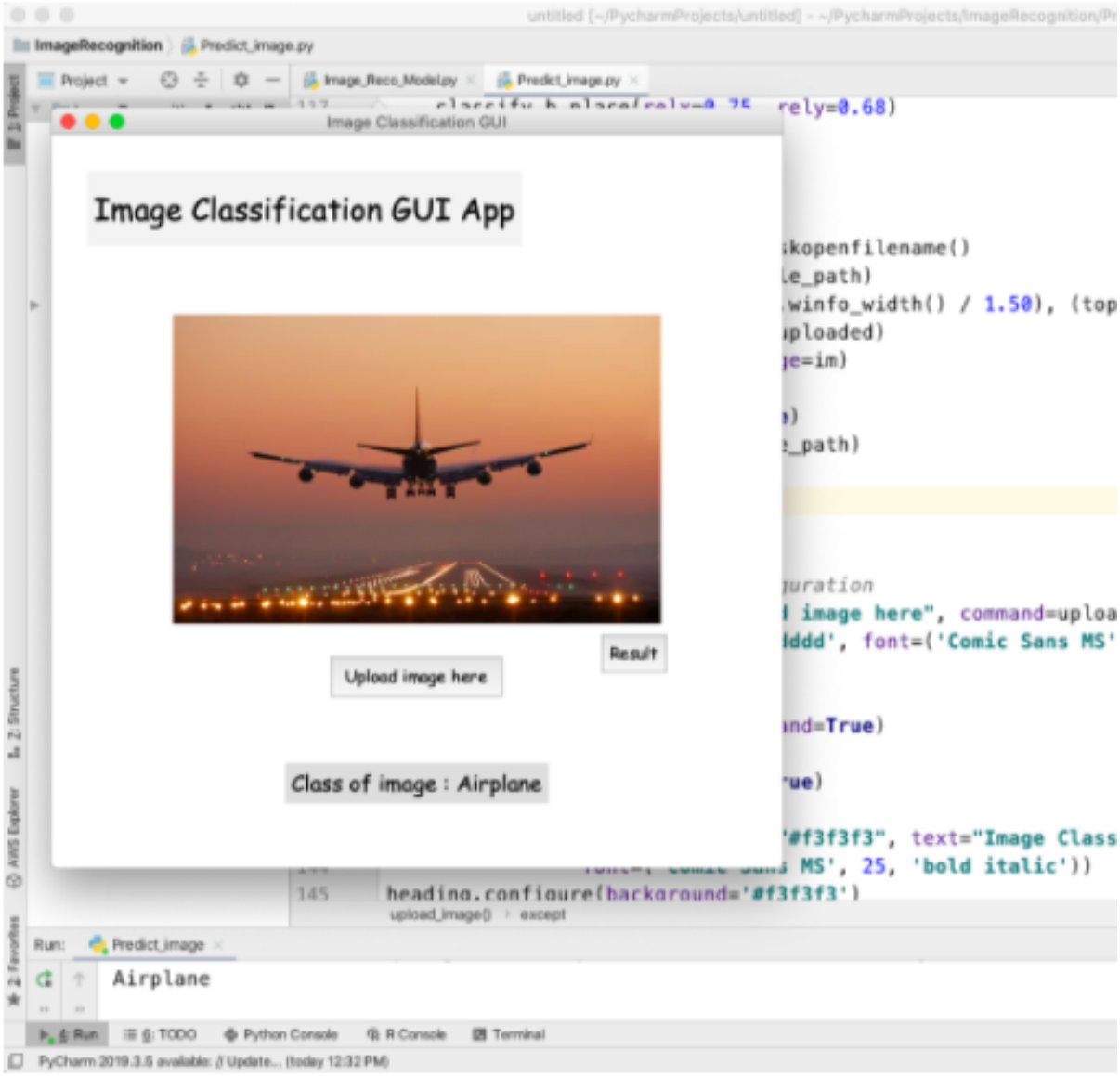


Figure 6.1: Graphical User Interface for Image Recognition

Chapter 7

Applications of Image Recognition

7.1 Google Lens

Google Lens is the most powerful image recognition technology developed by Google using the modern approaches like Machine Learning. In this technology, visual analysis is done using the neural network algorithm. It is used to retrieve the information relevant to the object that is captured by the camera. Sometimes human can visualize the object but don't know much about the image like person may not know the name of the flower or place by seeing it. So, to overcome this day-to-day difficulty, there are so many features in the Google Lens. Location of the device plays major role in order to generate the best results.

Google Lens works by following these steps:

1. Device camera will capture the image.
2. After that Lens compares the object or scene of this image to other images that is stored on web.
3. Lens will rank this image based on the similarities of other images on web.
4. Sometimes Lens generates the many possible results and ranks of the probable result. It may possible that it will narrow it to the single result.

Google Lens is used to

1. Copy some text on some card, paper by scanning it using camera and paste it anywhere in our device.
2. Explore the nearby places and see the rating of place, historical facts about places.
3. Identify plants and animals.
4. Search the results of math, biology, physics by capturing the equation or text.
5. Discover the useful products.

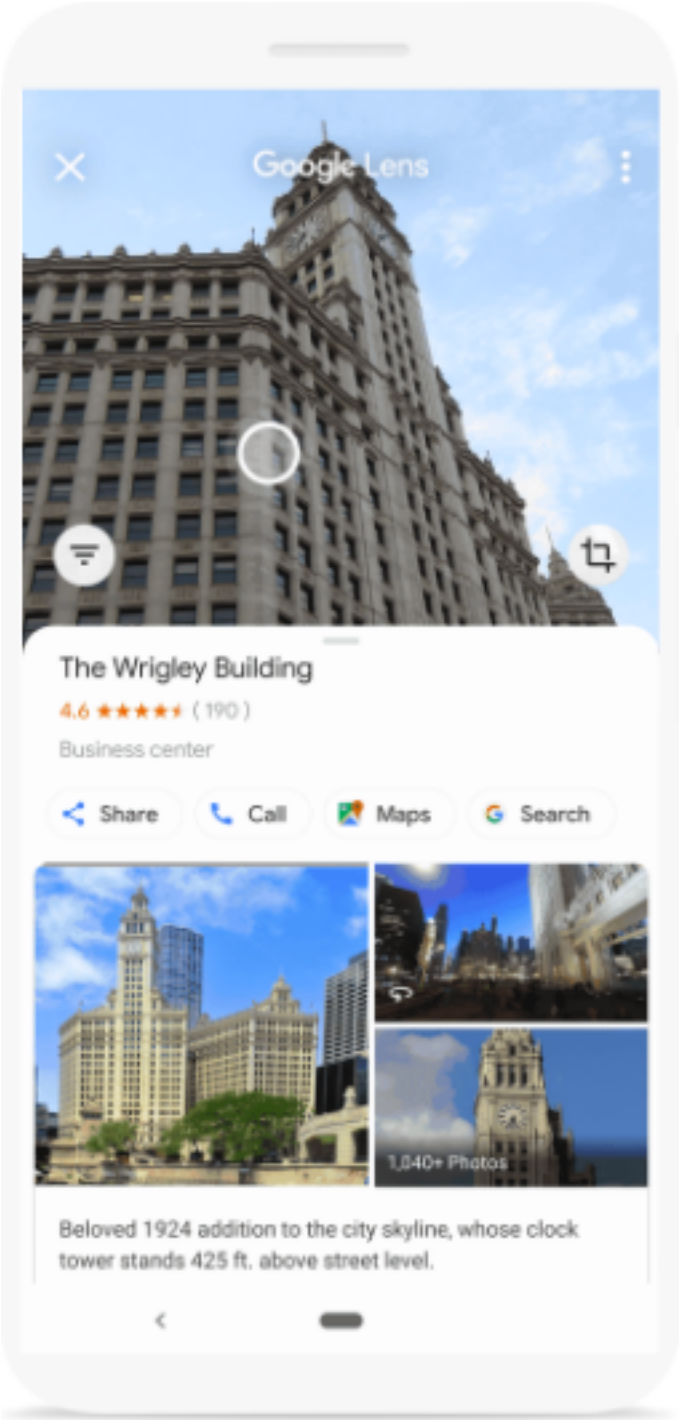


Figure 7.1: Google Lens - An example of Image Recognition Technology

7.2 Autonomous Vehicles

There are so many researches is going on currently to find better solution for autonomous vehicles. These vehicles can use the image recognition functionality by capturing the video image of the road ahead using camera mounted on the vehicle. Image recognition is used to identify the pathway, person, vehicles, sharp turns, moving objects. After recognizing these objects, autonomous vehicle can take necessary actions. The technology is even capable of reading the road signs and stop lights. Ex. ALVINN is a self-driving car which uses the neural network algorithm to produce the results and take actions after feeding the images of road ahead into the network.

7.3 Robotics and Manufacturing

Robotics field can use image recognition technology to identify some defects or evaluating critical points on the inspection line. Mini robots can be used to identify the defect in the production line and separate that items. There are so many logistics industries which is using the image recognition technology.

7.4 Drones and Military Surveillance

Drones with mounted camera can use image recognition technology to provide based on automatic monitoring, inspection, identify the assets located in the remote area. In the military surveillance, machinery can detect the unusual activities of the enemy on the border by identifying the image of border and use decision making capabilities to stop the infiltration of enemy and save many lives.

7.5 Gaming Industry

Gaming industry has done so much progress after the concept of augmented reality came into the picture. Image recognition technology with augmented reality helps the gamers to enjoy the realistic experience. These games use the computer vision and identify the movements of user in real time and according to that it provides the new user experiences and interfaces. Also, image recognition technology can be used to create the fictional characters and environments. Ex. China's video game "Honour of kings" uses this technology to recognize the age of the gamer by using face recognition functionality.

7.6 Google Photos

Google Photo is an application to share and store the photos online developed by Google. This application uses the image recognition technology to group the photos automatically uploaded

by the user. It will try to find similar patterns and classify the photos into the albums. It will separate photos into albums like people, places, things, pets, etc.

7.7 Visual Search Engines

Search engine can be used by users to retrieve some best results by entering the text in search engine. But in the visual search engines like Google and Bing, user can search required results by entering the image into the search engine. This technology of visual searching uses the image recognition to provide the best results.

7.8 E-commerce Websites

Most of the e-commerce websites adopts the image recognition technology to provide better search results and advertising. These websites can keep track of interest and behavior of customers by learning the patterns of photos uploaded by customers. Image recognition plays the major role in the predicting the consumer's behavior and interest. Ex. CamFind API by Image searcher Inc. uses the image recognition technology to give better user interface for e-commerce. This API identifies objects like hat, watches, bags, jeans and provides the purchasing options to the customer.

7.9 Medical services

Image recognition can also be used in the medical services and health-care. There are so many diseases in which doctor uses x-ray to find out if there is infection or not. Now we can use image recognition technology because here we have to deal with classification of x-ray images. We can train a model with the large data of x-ray images and it will generate the results.

Other applications of image recognition can be found in areas like iris recognition development, education, social media platforms, security industry, etc.

Chapter 8

Advantages and Disadvantages of Image Recognition

Every technology that grows in the market has its advantages and disadvantages. For image recognition, some of the pros and cons are listed as below.

8.1 Advantages

1. This technology can be very useful in area of security and surveillance. Best use of this feature is in identifying the criminal by feeding the photo of criminal into the surveillance devices. Because the device will capture all the humans that comes in the range and after that device will compare the faces of all the humans with the input photo of criminal.

2. Applications containing this technology like Google Lens provide the better and friendly experience to retrieve the information about the images captured by camera of mobile device.

3. There are many benefits to image recognition systems such as its connivence and social acceptability because it is so easy to use and access

4. Image recognition has also improved the online market platforms like e-commerce websites like Amazon, Flipkart because this technology provides improved results than traditional approaches. Ex. Amazon Rekognition image is based on this technology which is used to extract the patterns and text, to recognize the celebrities, to detect the object and inappropriate content in images.

8.2 Disadvantages

1. At this development level of image recognition, sometime software or application using this technology (IR) can be less reliable because modern approaches like machine learning model even lacks in the accuracy of prediction because of different scale of image,

deformed image, different angle of object in image, variation in size of image.

2. Development of this type of technology uses millions of images and video from the world wide web (WWW) uploaded by the users. That's why this technology has also changed the future privacy of data. So that users have to be more aware what type of data that they upload over the world wide web.

3. For working of this technology, we have to feed millions of different - different images as historical data to train the machine learning model. If we don't provide the image of any particular class like flower, animal, automobile then even trained model will not be able to predict images of that class.

Chapter 9

Limitations of Image Recognition

Some applications or developed machine learning model can even give better accuracy than human can identify because sometimes historical details are known to the model because developer has feed it to the model but possibly human doesn't know the historical facts about the image.

Image recognition technology is very useful to the human and growing very fast but there are some limitations of this technology. Machine learning model clearly depends on the dataset. If the dataset given to the model is fine-grained then it will classify the images more accurately. But just because it performs well on this dataset, doesn't mean it will work in all the cases.
Some limitation of this technology can be listed as below.

9.1 Limitations

1. As image holds more memory than simple text data, there is huge requirement of storage to store all the images in the system for feeding it to the neural network.

2. Because of limited hardware, it is difficult to process huge number of image data available.

3. Complexity of this particular model is very high because of higher-dimensional data. So, time to train the model will be very high. There are libraries like TensorFlow, Keras which provide more control and make development simpler but these libraries are complicated in nature and lack flexibility in its usage.

4. Difficulty in interpreting the model since the vague nature of the models prohibits its application in a number of areas.

5. This model lacks in the accuracy when there is some abnormality in the input image like different arrangement of object, occlusion, different shades of lighting, different angle of objects, smaller or larger scale of object in the image.

6. Accuracy of this technology also depends upon the quality of image, size of image because these factors play major role in the prediction of class of an image.

7. This technology learns a specific shape or pattern of an object and percepts that particular object can be in that shape only but in the real world, object can be deformed but still it belongs to same class of that object. But this technology lacks in prediction of deformed images.

Chapter 10

Conclusion

In this seminar work, we have explored a technology that is being used in every aspects of day-to-day life of human. Because image recognition is used in various areas like e-commerce websites, drones and robotics, visual search engine, health-care industry, military services, autonomous vehicles, etc. In this topic, first we have explored the modern approaches like deep learning and various algorithms used in classification of images based on machine learning. After understanding the basics, we have discussed the implementation of this technology using the convolutional neural networks and created a GUI application to use this technology without any prior knowledge of libraries of Python to work with images.

We came to know about detailed description of layers of CNN and why we are using it to work with images. We proposed a simple workflow to implement this technology. In this seminar, it is clear from the work, that this current implementation of image recognition is well suited to certain classes of images. And finally, we summarized topic by discussing about the various necessary parts of technology like advantages, disadvantages and in particular, the limitations of image recognition.

References

[1] Xin Jia, ”Image Recognition Method Based on Deep Learning”

[2] https://en.wikipedia.org/wiki/Machine-learning

[3] Saad ALBAWI , Tareq Abed MOHAMMED, ”Understanding of a Convolutional Neural Network”, ICET,2017.

[4] Chao Ma, Kianyong Yi, Chenglong Yu, Shuo Xu, Linyi Li, ”Research on Image Classification Method Based on DCNN”, ICCEA, 2020.

[5] https://www.mygreatlearning.com/blog/image-recognition

[6] https://lens.google.com/howlensworks

[7] CIFAR-10 Dataset, https://www.cs.toronto.edu/ kriz/cifar.html

[8] https://www.tensorflow.org/apidocs/python/tf/keras/Model

[9] https://www.einfochips.com/blog/understanding-image-recognition-and-its-uses

[10] https://www.datasciencesociety.net/image-recognition-applications-7-essential-future-uses

Acknowledgement

I would like to express my deep gratitude to my seminar guide, Dr. Krupa N. Jariwala, Professor, Computer Engineering Department, SVNIT Surat for her valuable guidance, useful feedback and co-operation with kind and encouraging attitude at all stages of the experimental work for the successful completion of this work.

I am also thankful to SVNIT Surat and its faculty for providing this opportunity which helped me gain sufficient knowledge to make my seminar work successful.