

Seminar report

IMAGE DETECTION AND DEEP LEARNING

October 11, 2022

Sipna college of engineering and technology

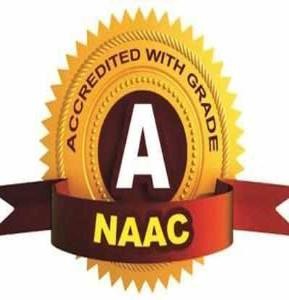
AMRAVATI

**PRATHMESH N. ADSOD**

**B.E FINAL YEAR**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**Seminar**



**“Image Detection and Deep Learning”**

IS SUBMITTED TO

**SANT GADGE BABA AMRAVATI UNIVERSITY**

IN THE PARTIAL FULFILLMENT OF THE DEGREE OF

**BACHELORS OF ENGINEERING**

**IN**

**INFORMATION TECHNOLOGY**

BY

**Mr. Prathmesh N. Adsod**

GUIDED BY

**Prof. A. B. Parandekar**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**SIPNA COLLEGE OF ENGINEERING AND TECHNOLOGY, AMRAVATI**

**(AN ISO 9001:2015 CERTIFIED INSTITUTE)**

**SANT GADGE BABA AMRAVATI UNIVERSITY, AMRAVATI**

**2022-2023**

**Sipna College of Engineering & Technology,**

**Amravati.**

**Department of Information Technology**

**CERTIFICATE**

This is to certify that **Mr. Prathmesh Narendra Adsod** has satisfactorily completed the seminar work towards the Bachelor of Engineering Degree of Sant Gadge Baba Amravati University, Amravati in Information Technology discipline on the topic entitled **“Image Detection and Deep Learning”**, during the academic year 2022-23 under my supervision and guidance.

**Date:**

**Prof A. B. Parandekar Dr. V. S. Gulhane**

**Guide Head of Dept.**

**Acknowledgement**

A moment of pause, to express deep gratitude to several individuals, without whom this seminar could not have been completed. We feel immense pleasure to express a deep sense of gratitude and indebtedness to our guide **Prof A. B. Parandekar**, for their constant encouragement and noble guidance.

We express our sincere thanks to **Dr. V. S. Gulhane**, Head, of the Department of Information Technology, and the other faculty members of the department for their kind cooperation.

We express our sincere thanks to **Dr. S. M. Kherde**, Principal, Sipna College of Engineering & Technology for his valuable guidance. We also express our sincere thanks to the library staff members of the college.

Last but not least we are thankful to our friends and parents whose best wishes are always with us.

**Student Name**

**Mr. Prathmesh Narendra Adsod**

**Abstract**

Deep learning technology is inspired by neural network present in human brain .Convolution Neural Network CNN and Recurrent Neural Network RNN are two important types or algorithms of supervised deep learning model. RNN is used for Natural Language Processing and CNN is used for Object Recognition or Image recognition . Deep learning is subset of AI/ML. ML is suitable when dataset is small but when dataset contains images or it become to large then deep learning is efficient .We can use different tools to implement our deep learning models but Tensorflow is widely used by the developers now-a-days. We can also use tools like Amazon SageMaker. For Traffic Signal Recognition project we can use deep learning and specifically CNN algorithm. It is also use in for automation. Keras is also specifically used for working with neural networks.

**Index**

|  |  |
| --- | --- |
| Sr.No. | Title Page.No. |
| 1    2  3  4  5  6  7  8  9  10  11 | Introduction 1  Machine Learning Vs Deep Learning 2-3  Architecture chart flow 4  How Deep Learning works 5-6  Deep Learning in Image Detection 7  Hardware requirements 10  Applications 11-12  How to create and train models 13  Tensorflow for deep learning 14-15  Conclusion 16  References 17 |
|  |  |
|  |  |
|  |  |

**1. Introduction**

Deep learning is a subset of [machine learning (ML)](https://www.hpe.com/in/en/what-is/machine-learning.html), which is itself a subset of [artificial intelligence (AI)](https://www.hpe.com/in/en/what-is/artificial-intelligence.html). The concept of AI has been around since the 1950s , with the goal of making computers able to think and reason in a way similar to humans. As part of making machines able to think, ML is focused on how to make them learn without being explicitly programmed. Deep learning goes beyond ML by creating more complex hierarchical models that are meant to mimic how humans learn new information.

Deep learning is a **subset** of [machine learning](https://www.ibm.com/cloud/learn/machine-learning), which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data . While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning attempts to mimic the human brain—albeit far from matching its ability—enabling systems to cluster data and make predictions with incredible accuracy.

Deep learning drives many [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

RNN and CNN are two important algorithms. RNN is used for Natural Language Processing and CNN is used for Image Detection. We have different tools for implementing deep learning like Tensorflow , Keras , Scikit-Learn , PyTorch etc.

## 

## **2. Deep learning vs. machine learning**

If deep learning is a subset of machine learning, how do they differ? Deep learning distinguishes itself from classical machine learning by the type of data that it works with and the methods in which it learns.

Machine learning algorithms leverage structured, labeled data to make predictions—meaning that specific features are defined from the input data for the model and organized into tables. This doesn’t necessarily mean that it doesn’t use unstructured data; it just means that if it does, it generally goes through some pre-processing to organize it into a structured format.

Deep learning eliminates some of data pre-processing that is typically involved with machine learning. These algorithms can ingest and process unstructured data, like text and images, and it automates feature extraction, removing some of the dependency on human experts. For example, let’s say that we had a set of photos of different pets, and we wanted to categorize by “cat”, “dog”, “hamster”, et cetera. Deep learning algorithms can determine which features (e.g. ears) are most important to distinguish each animal from another. In machine learning, this hierarchy of features is established manually by a human expert. [1,3]

Then, through the processes of gradient descent and backpropagation, the deep learning algorithm adjusts and fits itself for accuracy, allowing it to make predictions about a new photo of an animal with increased precision.

Machine learning and deep learning models are capable of different types of learning as well, which are usually categorized as supervised learning, unsupervised learning, and reinforcement learning. Supervised learning utilizes labeled datasets to categorize or make predictions; this requires some kind of human intervention to label input data correctly. In contrast, unsupervised learning doesn’t require labeled datasets, and instead, it detects patterns in the data, clustering them by any distinguishing characteristics. Reinforcement learning is a process in which a model learns to become more accurate for performing an action in an environment based on feedback in order to maximize the reward**.**[2,4]

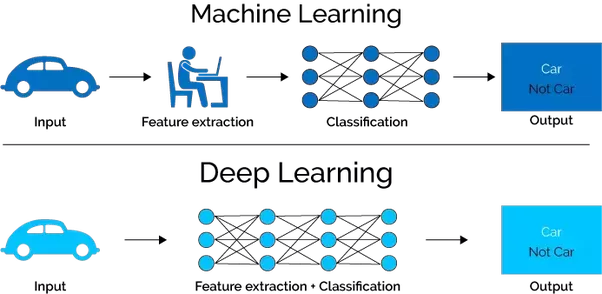


Fig. 01 Machine Learning vs Deep Learning

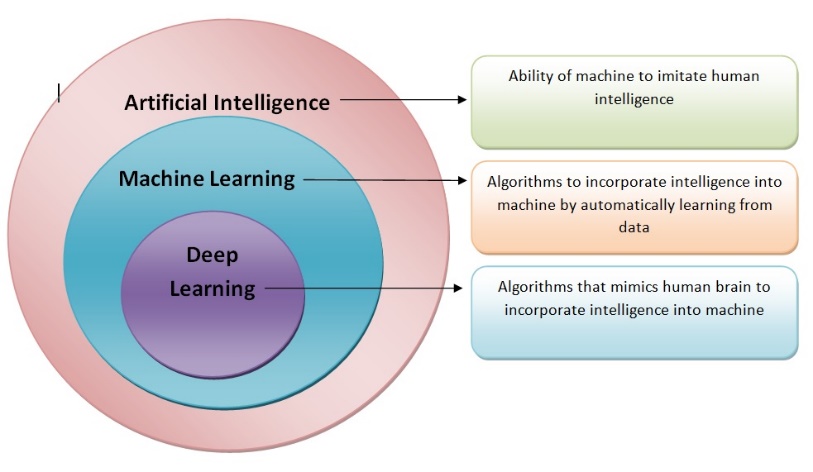


Fig. 02 AI vs ML vs DL

**3. Architecture Pattern Chart Flow**

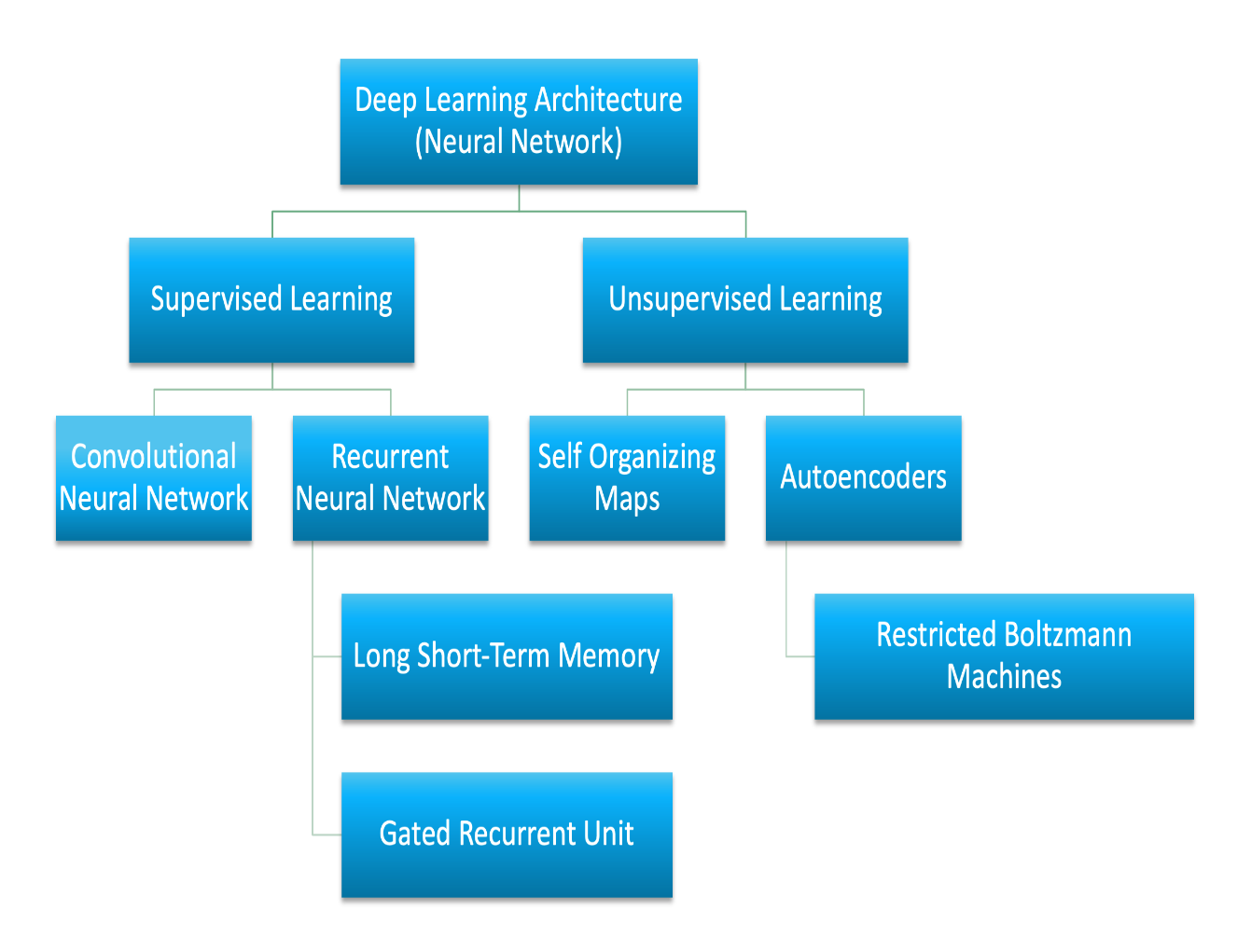


Fig. 03 Neural Networks Architecture [4]

## **4. How deep learning works**

Deep learning neural networks, or artificial neural networks, attempts to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data.

Deep neural networks consist of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization. This progression of computations through the network is called forward propagation. The input and output layers of a deep neural network are called *visible*layers. The input layer is where the deep learning model ingests the data for processing, and the output layer is where the final prediction or classification is made. Another process called backpropagationuses algorithms, like gradient descent, to calculate errors in predictions and then adjusts the weights and biases of the function by moving backwards through the layers in an effort to train the model. Together, forward propagation and backpropagation allow a neural network to make predictions and correct for any errors accordingly. Over time, the algorithm becomes gradually more accurate.

The above describes the simplest type of deep neural network in the simplest terms.[6]

However, deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets. For example,

[***Convolutional neural networks (CNNs),***](https://www.ibm.com/cloud/learn/convolutional-neural-networks)used primarily in computer vision and image classification applications, can detect features and patterns within an image, enabling tasks, like object detection or recognition. In 2015, a CNN bested a human in an object recognition challenge for the first time.

[***Recurrent neural network (RNNs)***](https://www.ibm.com/cloud/learn/recurrent-neural-networks)are typically used in natural language and speech recognition applications as it leverages sequential or times series data.

“Deep” refers to the many layers the neural network accumulates over time, with performance improving as the network gets deeper. Each level of the network processes its input data in a specific way, which then informs the next layer. So the output from one layer becomes the input for the next.

Training deep learning networks is time consuming and requires large amounts of data to be ingested and tested against as the system gradually refines its model. Neural nets have been around since the 1950s, but only in recent years have both computational power and data storage capabilities advanced to the point where deep learning algorithms can be used to create exciting new technologies. For example, deep learning neural networks have made it possible for computers to carry out tasks such as speech recognition, computer vision, bioinformatics and medical image analysis**.**[6]

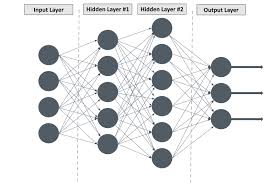
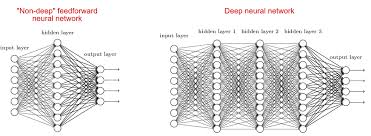


Fig . 04 Convolution Neural Network Layers



## 

Fig 05 Non-Deep neural network vs Deep Neural Network

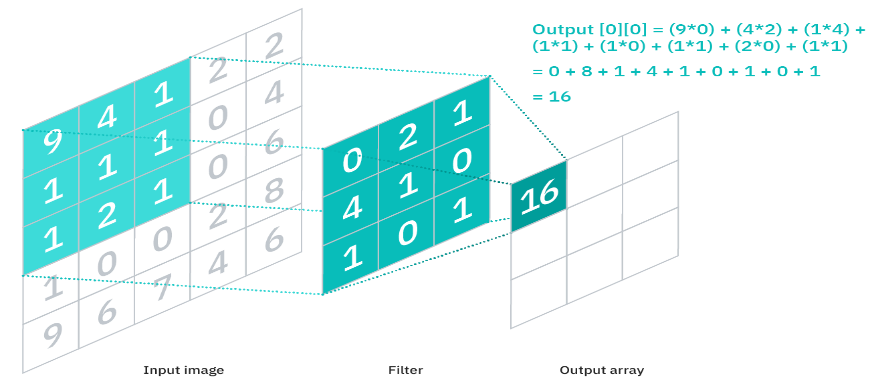


Fig 06 CNN Algorithm Working on black and white image

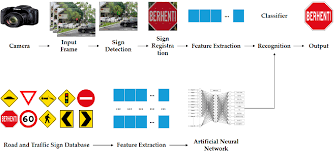
## **5. Deep Learning in Image Detection**

Convolution Neural Network is famous algorithm for working in Deep Learning.Traffic Sign Recognition project is in Deep Learning Technology.

How do we understand whether a person passing by on the street is an acquaintance or a stranger (complications like short-sightedness aren’t included)? We look at them, subconsciously analyze their appearance, and if some inherent features – face shape, eye color, hairstyle, body type, gait, or even fashion choices – match with a specific person we know, we recognize this individual. This brainwork takes just a moment.

So, to be able to recognize faces, a system must learn their features first. It must be trained to predict whether an object is X or Z. Deep learning models learn these characteristics in a different way from machine learning (ML) models. That’s why model training approaches are different as well.

**Each layer of nodes trains on the output (feature set) produced by the previous layer. So, nodes in each successive layer can recognize more complex, detailed features – visual representations of what the image depicts. Such a** [**hierarchy**](https://skymind.ai/wiki/neural-network)**of increasing complexity and abstraction” is known as feature hierarchy.**



## 

Fig . 07 Recognition of Traffic Sign by Neural Network

**Layers used to build CNN**

A CNN is a sequence of layers, and every layer transforms one volume to another through a differentiable function.

**Types of layers:**   
Let’s take an example by running a CNN on of image of dimension 32 x 32 x 3. 

1. **Input Layer:**

 This layer holds the raw input of the image with width 32, height 32, and depth 3.

1. **Convolution Layer**

This layer computes the output volume by computing the dot product between all filters and image patches. Suppose we use a total of 12 filters for this layer we’ll get output volume of dimension 32 x 32 x 12.

1. **Activation Function Layer**

 This layer will apply an element-wise activation function to the output of the convolution layer. Some common activation functions are RELU: max(0, x), Sigmoid: 1/(1+e^-x), Tanh, Leaky RELU, etc. The volume remains unchanged hence output volume will have dimension 32 x 32 x 12.

1. **Pool Layer**

 This layer is periodically inserted in the CNN and its main function is to reduce the size of volume which makes the computation fast reduces memory and also prevents overfitting. Two common types of pooling layers are **max pooling** and **average pooling**. If we use a max pool with 2 x 2 filters and stride 2, the resultant volume will be of dimension 16x16x12.

5. **Fully-Connected Layer** This layer is a regular neural network layer that takes input from the previous layer and computes the class scores and outputs the 1-D array of size equal to the number of classes. 

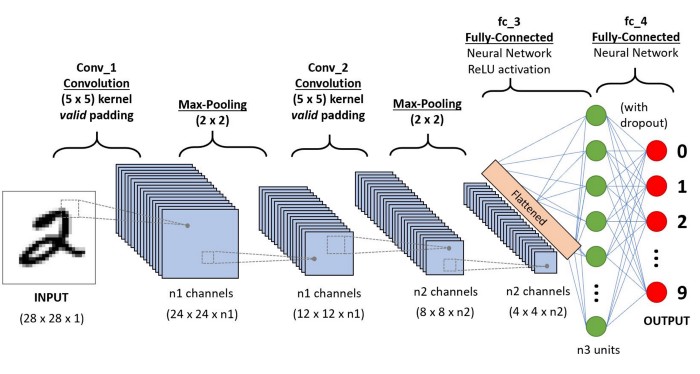


Fig.08 Working of CNN on Colorful Image [7]

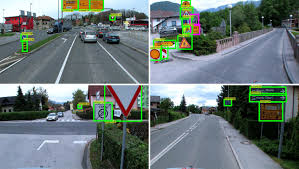


Fig. 09 Traffic Light Recognising by the algorithms

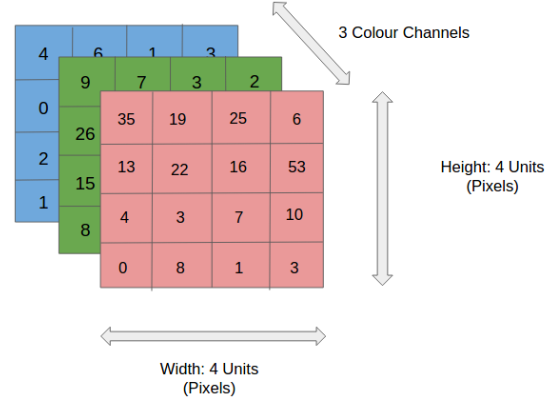


Fig.10 RGB Image in 4\*4\*3 dimension

## **6. Hardware requirements**

Deep learning requires a tremendous amount of computing power. High performance [graphical processing units *(GPUs)*](https://www.ibm.com/cloud/blog/video-what-is-a-gpu) are ideal because they can handle a large volume of calculations in multiple cores with copious memory available. However, managing multiple GPUs on-premises can create a large demand on internal resources and be incredibly costly to scale**.**[4]

Training a deep learning model can take a long time, from days to weeks. Using GPU acceleration can speed up the process significantly. Using MATLAB with a GPU reduces the time required to train a network and can cut the training time for an image classification problem from days down to hours. In training deep learning models, MATLAB uses GPUs (when available) without requiring you to understand how to program GPUs explicitly.[4]

CPUs are latency optimized while GPUs are bandwidth optimized. You can visualize this as a CPU being a Ferrari and a GPU being a big truck. The task of both is to pick up packages from a random location A and to transport those packages to another random location B. The CPU (Ferrari) can fetch some memory (packages) in your RAM quickly while the GPU (big truck) is slower in doing that (much higher latency). However, the CPU (Ferrari) needs to go back and forth many times to do its job (location A $\rightarrow$ pick up 2 packages $\rightarrow$ location B ... repeat) while the GPU can fetch much more memory at once (location A $\rightarrow$ pick up 100 packages $\rightarrow$ location B ... repeat).

So, in other words, the CPU is good at fetching small amounts of memory quickly (5 \* 3 \* 7) while the GPU is good at fetching large amounts of memory (Matrix multiplication: (A\*B)\*C). The best CPUs have about 50GB/s while the best GPUs have 750GB/s memory bandwidth. So the more memory your computational operations require, the more significant the advantage of GPUs over CPUs. But there is still the latency that may hurt performance in the case of the GPU. A big truck may be able to pick up a lot of packages with each tour, but the problem is that you are waiting a long time until the next set of packages arrives. Without solving this problem, GPUs would be very slow even for large amounts of data.

**Minimum Hardware Requirements –**

* GRAPHICS CARD:- 1650 or 1660TI
* CPU:- i5 9th or 10th Gen.
* RAM:- 16GB

**7**. **Applications**

Real-world deep learning applications are a part of our daily lives, but in most cases, they are so well-integrated into products and services that users are unaware of the complex data processing that is taking place in the background. Some of these examples include the following:

### **Traffic Sign Recognition**

Using deep learning technology we can recognize sign of traffic rules . There are multiple rules of Traffic signals so we can also recognize it so what this signal or image saying.

### **Object Detection Using Deep Learning**

In addition to [object recognition](https://www.mathworks.com/solutions/image-video-processing/object-recognition.html), which identifies a specific object in an image or video, deep learning can also be used for **object detection**. [Object detection](https://www.mathworks.com/discovery/object-detection.html) algorithms like YOLO You Only Look Once can recognize and locate the object in a scene, and can locate multiple objects within the image.

**Law enforcement**

Deep learning algorithms can analyze and learn from transactional data to identify dangerous patterns that indicate possible fraudulent or criminal activity. Speech recognition, computer vision, and other deep learning applications can improve the efficiency and effectiveness of investigative analysis by extracting patterns and evidence from sound and video recordings, images, and documents, which helps law enforcement analyze large amounts of data more quickly and accurately.

**Financial services**

Financial institutions regularly use predictive analytics to drive algorithmic trading of stocks, assess business risks for loan approvals, detect fraud, and help manage credit and investment portfolios for clients.

**Customer service**

Many organizations incorporate deep learning technology into their customer service processes. [Chatbots](https://www.ibm.com/cloud/learn/chatbots-explained)—used in a variety of applications, services, and customer service portals—are a straightforward form of AI. Traditional chatbots use natural language and even visual recognition, commonly found in call center-like menus. However, more [sophisticated chatbot solutions](https://www.ibm.com/products/watson-assistant) attempt to determine, through learning, if there are multiple responses to ambiguous questions. Based on the responses it receives, the chatbot then tries to answer these questions directly or route the conversation to a human user.

Virtual assistants like Apple's Siri, Amazon Alexa, or Google Assistant extends the idea of a chatbot by enabling speech recognition functionality. This creates a new method to engage users in a personalized way.

**Healthcare**

The healthcare industry has benefited greatly from deep learning capabilities ever since the digitization of hospital records and images. Image recognition applications can support medical imaging specialists and radiologists, helping them analyze and assess more images in less time. **[6]**



Fig . 11 One Traffic Sign Fig. 12 Multiple Traffic Signs

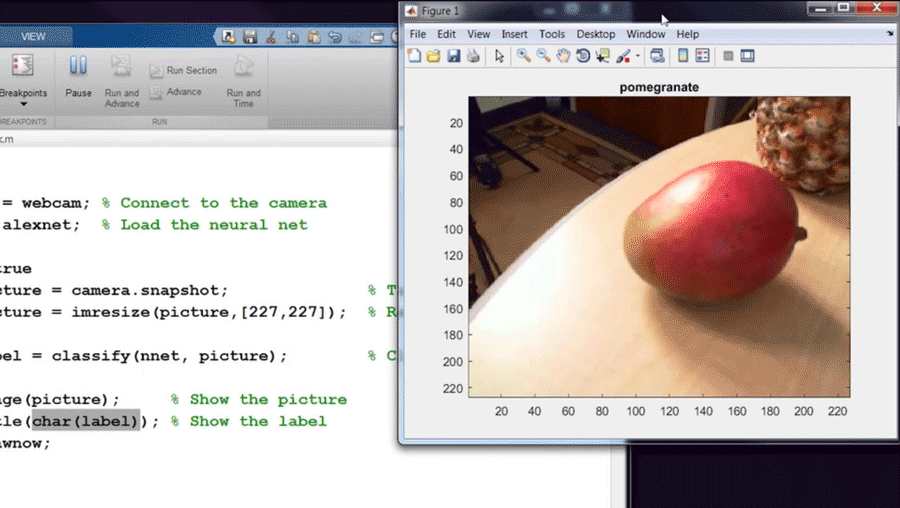


Fig 13 Object Detection Fig 14 Fruit Detection

### **8. How to Create and Train Deep Learning Models**

The three most common ways people use deep learning to perform object classification are:

**Training from Scratch**

To train a deep network from scratch, you gather a very large labeled data set and design a network architecture that will learn the features and model. This is good for new applications, or applications that will have a large number of output categories. This is a less common approach because with the large amount of data and rate of learning, these networks typically take days or weeks to train.[5]

**Transfer Learning**

Most deep learning applications use the [transfer learning](https://www.mathworks.com/discovery/transfer-learning.html) approach, a process that involves fine-tuning a pretrained model. You start with an existing network, such as AlexNet or GoogLeNet, and feed in new data containing previously unknown classes. After making some tweaks to the network, you can now perform a new task, such as categorizing only dogs or cats instead of 1000 different objects. This also has the advantage of needing much less data (processing thousands of images, rather than millions), so computation time drops to minutes or hours.[5]

Transfer learning requires an interface to the internals of the pre-existing network, so it can be surgically modified and enhanced for the new task. [MATLAB®](https://www.mathworks.com/products/matlab.html) has tools and functions designed to help you do transfer learning.

**Feature Extraction**

A slightly less common, more specialized approach to deep learning is to use the network as a **feature extractor**. Since all the layers are tasked with learning certain features from images, we can pull these features out of the network at any time during the training process. These features can then be used as input to a [machine learning model](https://www.mathworks.com/solutions/machine-learning.html) such as [support vector machines (SVM)](https://www.mathworks.com/discovery/support-vector-machine.html).[**5**]

**9. Deep Learning with Tensorflow**

 Fig. 15 TensorFlow Logo

TensorFlow is a framework created by Google for creating Deep Learning models. Deep Learning is a category of machine learning algorithms that use multi-layer neural networks.

Machine Learning has enabled us to build complex applications with great accuracy. Whether it has to do with images, videos, text or even audio, Machine Learning can solve problems from a wide range. Tensorflow can be used to achieve all of these applications.

The reason for its popularity is the ease with which developers can build and deploy applications. The GitHub projects which we’ll look closer at due the next parts are very powerful but also so easy to work with. Moreover, Tensorflow was created with processing power limitations in mind. The library can be ran on computers of all kinds, even on smartphones (yes, even on that overpriced thing with half an apple on it). Working on a Intel Core i3 with 8 GB of RAM, you won’t have performance issues.

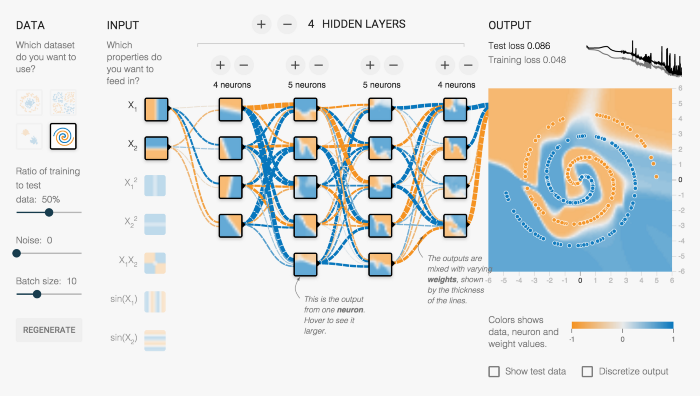


Fig. 16 Deep learning concept of Tensorflow

**How can we make our machines “think”?**

The human brain consists of billions of neurons which are interconnected by synapses. If “enough” synaptic inputs to the neuron fires, then the neuron will also fire. This process is called thinking. To replicate that process on computers, we need machine learning and neural networks. [5]

**What is Keras and why it is used?**

Keras is a high level API designed by Google engineers for neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.

**10. Conclusion**

By this way we got covered how we are using deep learning in our project . Traffic sign detection is one of the good way to study CNN and deep learning . By this way we learn how our CNN algorithm will directly digest the images and we do not need to pre-process data. We have learnt how deep learning work internally and got idea about how our model will be working in this process .We also got idea about different layers in cnn.

We have taken different different use cases to know more about Deep Learning and learnt about how layer by layer our model gets build . We also covered deep learning and machine learning difference.

**11. References**

[1] Research Paper – Unified Real Time Object Detection

Joseph Redmon, Santosh Divvala, Ross Girshick , Ali Farhadi University of Washington, Allen Institute for AI , Facebook AI Research http://pjreddie.com/yolo/ arXiv:1506.02640v5 [cs.CV] 9 May 2016

[2] Reasearch Paper – Convolution Network for BioMedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany ronneber@informatik.uni-freiburg.de , WWW home page: http://lmb.informatik.uni-freiburg.de/

[3]Hawlette Packwet blog <https://www.hpe.com/in/en/what-is/deep-learning.html?jumpid=ps_jwzma4und1_aid-520061736&ef_id=Cj0KCQjwhY-aBhCUARIsALNIC05xkpMWRafxXAwPRqVvNOrhcAYxFi0rHhBJDOTlqQ3vAIvFpbUFNggaAlBqEALw_wcB:G:s&s_kwcid=AL!13472!3!595029415351!e!!g!!how%20does%20deep%20learning%20work!17064341482!133355606862&>

[4] IBM Blog - <https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/#:~:text=Deep%20learning%20is%20represented%20by,permit%20consideration%20of%20previous%20inputs>.

[5] Towardsdatascience blog - <https://towardsdatascience.com/deep-learning-with-tensorflow-part-1-b19ce7803428>

[6]IBM Blog - <https://www.ibm.com/cloud/learn/deep-learning#:~:text=Deep%20learning%20is%20a%20subset,from%20large%20amounts%20of%20data>.

[7]towardsdatascienceblog - https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53