IMAGE IDENTIFICATION OF PLANT DISEASES USING DEEP LEARNING APPROACHES

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

Rishav Sapahia(141310)

Under the supervision of

Dr. Pradeep Kumar Singh (Assistant Professor)

to

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Candidate's Declaration

I hereby declare that the work presented in this report entitled **" Image Identification of Plants Disease using Deep Learning Approaches"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from August 2017 to May 2018 under the supervision of D**r. Pradeep Kumar Singh** (**Assistant Professor (Senior Grade**)).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Rishav Sapahia(141310)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Pradeep Kumar Singh

Assistant Professor(Sr. Grade)

CSE

Dated:

ACKNOWLEDGEMENT

I hereby express my gratitude to my parents to raise me with great values,love and ethics and making me believe that living to impact the life of people positively is the best way to live .

Secondly ,I would like to express my gratitude towards every friend,every other person which made me a better human being.

Lastly ,I would like to express my gratitude towards every writer of the books that I have read till now,they helped me to shape my thinking greatly and enriched me with their vast knowledge and helped me to realize that not taking a risk is the greatest risk in life.

Contents

5 Conclusions **50**

List of Figures

Chapter 1
Introduction

1.1 Introduction

Agriculture is the field which is the oldest form of occupation and has been one of the pioneer in the cultivation of human society as we see now, but still agriculture is infested with traditional practices which hampers the productivity greatly. With Artificial Intelligence ,breaking the new horizons ,it could improve the traditional ongoing practices in agriculture and can improve the productivity greatly.

With the recent advances in Machine Learning ,computer vision has been able to improve at a rapid rate. Image Identification,that is detection of the required object in a picture , is one of the problem which is extremely easy for humans but difficult for computers but now with the huge onset of data ,and with inexpensive computation power available, a new Machine Learning approach-Deep Learning has evolved .Deep learning involves training a type of neural network by feeding the labeled data(supervised learning) or unlabeled data(unsupervised learning) and then getting the required result(classification,prediction).Static feature extraction is not needed in this type of ML and it is showing promising results even in the product phase which is leading to the vast development and research in the deep learning field.

AlexNet in 2012 broke all the records for how computer vision is done.Neural Networks which are deemed as an obsolete approaches paved their way again in 2012 with AlexNet and now with faster computation power and cheaper gpu's ,neural networks are breaking state of the art everyday.Now Google's Resnet is at the state of the art architecture in computer vision.

1.2 Problem Statement

Agricultural Scientists in order to identify Image based plants disease still uses primitive appraoch to check the leaf manually and then cross-checking with their memory to identify whether a disease is present in plant or not.This process is quite cumbersome and there is a possibility of human error.There is no computer/mobile based application which can help the scientist to identify the disease.Moreover many of the farmers still unable to identify the disease correctly due to lack of knowledge and awareness and results in treating the plant incorrectly or with wrong treatments.

In world today,hunger is the world's number 1 cause for the deaths.60 crores of India's population still don't get 2 square meal a day .India ranks worse than Bangladesh,Iraq to provide food security for its citizen.

Inability to correctly identify disease is still one of the biggest problem in the agriculture and if solved ,they may pave a way to improve the productivity greatly.

1.1 **Objective**

The prime objective of the project is to build a computer/mobile based system that can correctly identify the plants disease and then label the unseen data correctly.

1.2 Methodology

The main methodology used in the project is Deep Learning. Deep Learning is the subset of Machine Learning and gained momentum in 2012 by ground-breaking performance of AlexNet on ImageNet data.

Deep Learning unlike traditional Machine Learning approaches doesn't use explicit feature extraction ,the subsequent layers of the neural network does the feature extraction implicitly.

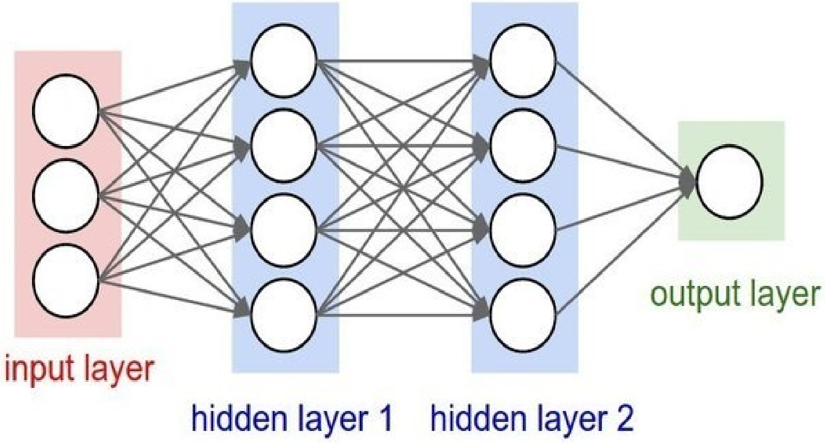


Figure 1.4 A Basic Neural Network of 2 hidden layers

The subsequent layers extract the features automatically and user don’t have to program the features by themselves.

The improvement over basic neural network is the addition of a new kind of layer called convolution layer and hence called convolution neural network. Convolution Neural Network are currently state of art in computer vision problems.

In convolution layer,A filter or a kernel traverse the image starting from top left of the image and then doing the dot product of both the values and hence the values obtained are the output of the convolution layers.

Intuitively ,the filters recognizes the simple features such as vertical,horizontal lines and then reconstruct the features and then pass them to obtain further results.

Chapter 2

Literature Survey

2.1 Feasibility Study

State of the art neural network architectures of various years were studied and the ways to optimize them using various hyper parameter tuning were observed and then applied in experiements.

Paper review: Very Deep Convolution Deep Networks for Large Scale Image Recognition

Objective of Paper:The main objective in [1],was to investigate the effect of depth of convolution networks on its accuracy in large scale image image recognition settings.The depth of the neural network was steadily increased by adding more convolution layers and keeping all the other parameters fixed.

Techniques:

Architecture:Input to convnets was 224*224 RGB image.The filter used was 3*3,padding =1 ,stride=1 and max-pooling =2*2 with a window stride of 2*2.

Dataset: Image classification result is obtained on ILSVRC-2012 dataset.The dataset includes images of 1000 classes and is split into 3 sets:training(1.3 M images),validation(50K images),testing(100K images).

Training:The training is carried out by optimising multinomial logistic regression using mini-batch gradient descent with momentum.The training was regularized by weight-decay and drop-out regularization for the first two fully-connected layers.The learning rate was initially set to 0.001.

While the input image size was 224*224,it can be cropped from any image of any size >=224.

Two approaches were used to select training scale S(Smallest side of rescaled training image from which the Convnet input is cropped)

-The first is to fix the S ,called single scale training,with value=256 and 384.

-The other approach is multi-scale training where each training image is rescaled by randomly sampling S from a certain range [S_min,S_max],where S_min=256,S_max=512 were used.

Relu activation is used to model non-linearity in images.

Results:

Classification performance is evaluated using top-1 and top-5 error.

Top-1 Error-The percentage in which the correct label is not within the highest probability class ,predicted by the network.

Top-5 Error-The percentage in which the correct label is not within the top-5 classes predicted by the network.

The classification error decreases with the increased ConvNet depth ,from 11 layers in network A to 19 layers in network E.

The top-1 % value error comes out to be 25.5 and top-5 % value error comes out to be 8% in single scale evaluation and for multiple test scale top-1 and top-5 % error comes out to be 24.8 and 7.5.

The main conclusion from VGG-16 was that the representation depth is beneficial for classification accuracy and by substantially increasing depth ,higher accuracy can be achieved using the conventional ConvNet Architecture.

Limitations:

With increasing layer,the performance saturates and then degrades rapidly .

Future Scope-

VGG-16 is being widely used in image recognition and image localization tasks by various internet based applications as it is very much acccurate in comparsion to the contemporary models.It is relatively easy to build due to its less complex architecture(only 3*3 filters).

2. Paper review: ImageNet classification with Deep Convolutional Neural Networks

Objective of Paper: The main objective described by Krizhevsky et *al.* ,is to show the enhanced performance of neural network approach ,given the availability of large labelled data and huge computation power.Earlier ,the idea of learning the neural network was seen as wrong approach in comparison to the programming approach but the work done by Krizhevsky et *al. ,*changed the status-quo of the Computer Vision research.

Techniques:

Dataset:Image classification result is obtained on ILSVRC-2010 dataset.The dataset includes images of 1000 classes and is split into 3 sets:training(1.2 M images),validation(50K images),testing(150K images).It consists of 60 million parameters.

Architecture:Input to ConvNet was 256*256 RGB pictures.The ConvNets contains eight learned layers-five convolutional and three fully-connected layers.It used ReLU for the non-linearity functions.

The first convolution layer filter the 224*224*3 input image with 96 kernels of size 11*11*3 with a stride of 4 pixels.The second convolutional layer takes as input the output of the fixed convolution layer and filter it with 256 kernels of size 5*5*48.The third ,fourth and fifth convolution layers are connected to one another without any intervening pooling or normalization layers.The third convolution layer has 384 kernels of size 3*3*256 connected to the outputs of the second convolutional layer.The fourth convolution layer has 384 kernels of size 3*3*192 and the fifth convolution layer has 256 kernels of size 3*3*192.The fully -connected layer have 4096 neurons each.

In order to reduce overfitting ,data augmentation and drop-out is used.

***Results*-**

Classification performance is evaluated using top-1 and top-5 error.

Top-1 Error-The percentage in which the correct label is not within the highest probability class ,predicted by the network.

Top-5 Error-The percentage in which the correct label is not within the top-5 classes predicted by the network.

AlexNet achieved top-1 and top-5 tests-set rate of 37.5% and 17.0% respectively.

It is notable that our networks's performance degrades if a single convolutional layer is removed.

Limitations:

Training accuracy was improved further by VGG-16[2].

Future Scope-

Alexnet was the revolutionary step in the computer vision and made all the giants like Google,Facebook,Microsoft,Baidu switched to deep neural network approach.

3.**Paper review:Deep Residual Learning for Image Recognition**

***Objective**:*The main objective described by He *et al. In their* ConvNet ,also called ResNet,is to address the degradation problem of the ConvNets-performance of ConvNets was saturating after stacking more layers and then degrading rapidly ,by introducing a deep residual learning framework.

Techniques:

Architecture:

The He *et al.* ,proposed a hypothesis that stacking layers should not degrade a network because ,say,if the final layer output is x then if they starts increasing layer by identity mapping -x+1,x+2,x+3 ,then the performance should ideally remain same but on experimenting,it was observed that it was degrading .So ,instead of learning an mapping from x to H(x),learn the difference between the two(called residual).

F(x)=H(x)-x

Now the ConvNets are learning the F(x)+x instead of x.Now ,these residual blocks are easy to train than the identity functions and the identity functions can be achieved by using the weight decay on F(x).So adding residual blocks does not harm the networks performance and can even better the network performance .Also while in backprop,it prevents the problem of vanishing gradients by providing a skip-route to the gradients.

-224*224 crop is randomly sampled from an image.

-It was ultra deep with 152 layers.

-Batch Normalisation and color augmentation was used.

-SGD with mini-batch of 256 was used.

-Learning rate started from 0.1 and was divided by 10 ,when the error plateaues.

-Weight decay of 0.0001 and a momentum of 0.9 is used.

-No drop-out is used.

Results:

Classification performance is evaluated using top-1 and top-5 error.

Top-1 Error-The percentage in which the correct label is not within the highest probability class ,predicted by the network.

Top-5 Error-The percentage in which the correct label is not within the top-5 classes predicted by the network.

152-layer ResNet achieved top-1 and top-5 % error of 19.38 and 4.49% respectively. It was the best performance till date achieved by any ConvNet so far and was the winner of ILSVRC-2015 classification task and COCO object detection and Segmentation task.

Limitations:

On exploring a deeper model of 1202 layers,the testing result of this network is worse than the 110-layer network.Although both have similar training error.This is speculated to be due to overfitting but needs to be explored further.

4.**Paper review:Going Deeper with Convolutions**

Objective:The main objective behind Szegedy et al., is to design a network ,codenamed GoogleNet ,which is computationally efficient,that is ,it should work on low computational resources.They introduced a inception module to achieve this goal.

Techniques:

In inception module,several different kind of filters operations (1*1,3*3&5*5)are applied parallely and then filter outputs are concatenated depth wise.In addition to this,additional 1*1 convolutions(network in network) are used to for dimension reduction to avoid limiting the size of network.This architecture helps in a way that network in network is able to extract information about very minute details while the 5*5 filters is able to cover a large input and pooling operation helps to reduce spatial size and prevents overfitting

-No fully connected layers instead average pool was used .

-12 times less parameters then Alexnet.

-22 layers deeper.

-Dropout layer with 70% ratio of dropped outputs.

-Linear layer with softmax loss as the classfier

-Was trained on a few high end GPUs within a Week.

Results:

Classification performance is evaluated using top-1 and top-5 error.

Top-1 Error-The percentage in which the correct label is not within the highest probability class ,predicted by the network.

Top-5 Error-The percentage in which the correct label is not within the top-5 classes predicted by the network.

GoogleNet achieved a top-5 error rate of 6.67% .

Future Scope-

GoogleNet was one of the first module that showed that creative structuring of layers can lead to improved performace and computationally effieciency and acted as a pioneer for the amazing future architectures.

5 Paper review:Xception:Deep Learning with Depthwise Separable Convolutions

***Objective*:**The main objective by F.Chollet in paper was to introduce a ConvNet inspired by Inception where Inception module have been replaced with depthwise separable convolutions.

Techniques:

In Xception,instead of partitioning input data into several compressed chunks,it maps the spatial correlations for each output channel separately and then performs a 1*1 depthwise convolution to capture cross-channel correlation.

Dataset:Comparison of ConvNet is done on two image classification tasks

-ImageNet 1000-class single label classification task

-17,000-class multi-label classification task on large scale JFT dataset.

Architecture:

Different optimization configurations was used for ImageNet and JFT.

On ImageNet:

-Optimizer:SGD

-Momentum:0.9

-Initial Learning Rate:0.045

-Learning rate decay:decay of rate 0.94 every 2 epochs

-Dropout used.

-Weight Decay was used.

On JFT:

-Optimizer:RMSProp

-Momentum:0.9

-Initial Learning Rate:0.001

-Learning rate decay:decay of rate 0.9 every 3,000,000 samples.

-No Dropout.

-Weight Decay was used.

Result:

On ImageNet dataset,**Xception achieved Top-1 accuracy of 0.790 and top-5 accuracy of 0.945 which** is slightly better than the Inception.While on JFT dataset ,Xception achieved a greater accuracy than the Inception.

Future Scope-

After the results of Xception ,Depthwise separable convolutions are expected to become a cornerstone of convolution neural network architecture design in future.

6 Paper review:Generative Adversial Networks

*Objective:*The main obejctive by Goodfellow et. al.,is to introduce train two models-discriminative model and generative model,in which discriminative model identifies whether an image is artificially created or natural while generative model generates images so that adversial model can train better in identifying the images.

Techniques:

In Gans,there are 2 networks

Discriminative model(D):Discriminative models can be a multilayer perceptron.It identifies whether a sample is from the model distribution or the data distribution

Generative Models(G):In the paper,the generative models passes random noise through a multilayer perceptrons and then discriminative model judges the sample.

D is trained to maximize the probability of assigning the correct label to both training examples and samples from G while G is simultaneously trained to generate images that can beat D.

-Generative and Discriminative models ,are both multilayer perceptrons.

-Generative nets uses a mixture of relu and sigmoid activations.

-Discriminator nets uses maxout activations.

-Dropout is applied in training the discriminator nets.

-Adversial nets are tested on MNIST,TFD,and CIFAR-10 dataset.

-Markov chains are never needed in it ,only backprop is used to obtain gradient.

Limitations:

During trainin ,D and G must be synchronized well to avoid collapsing of values by G to have enough diversity to the models.

Future Scope:

A conditional generative model p(x|c) can be obtained by adding c as input to both G and D.As the discriminator model is aware of the internal representations of the data ,it can be used as a feature extractor that can be used in a CNN.

7 Paper review:Deep Visual Semantic Alignments for Generating Image Descriptions

Objective:The objective of the paper is to build a model that aligns the visual and textual data such that the textual data describes about the visual input.The alignment model is based on combination of Convolution Neural Networks over image regions,bidirectional Recurrent Neural Networks over sentences and structured objective that aligns the two models through multimodal embeddings.

Techniques:

In it,the training examples have weak labels that is they hhave segments of sentence refers to parts of image whose location is unknown.

It consists of 2 components-

Alignment Model-The goal of this model is to be able to align the visual and textual data.The model works by accepting an image and a sentence as input ,where the output is a score for how well they match.

Now ,the image is represented by first feeding the image into an R-CNN in order to detect the individual objects,this R-CNN is trained on ImageNet data.The top-19 object regions are embedded into a 500 dimensional space.Next step,is to collect informationa about the sentence is by embedding words into same multimodal space which is done by using bi-directional recurrent neural network.

Generation Model-The objective model of generation model is to learn from the dataset created by the alignment model in order to generate descriptions given in an image.

Dataset:Flickr8K,Flickr30k,MSCOCO datasets are used which contains 8000,31000,123000 images.

Results:

R@Kscale is used(Recall @K)-High is good.

Med r:Median Rank-Low is good.

R@10 was 61.4 for the model while med was 4.8.

Limitations:

The model can only generate a description of one input array of pixels at a fixed resolution

Future Scope:

This model uses different CNN and RNN models to create a very useful applications that combines the field of computer vision and natural language processing which can act as a pioneer step for dealing with tasks that cross different fields.

8 Paper review:Spatial Transformer Networks

***Objective**:*The main objective by Jaderberg *et. al is* to introduce spatial invariance and pose normalization by introducing a spatial transformer module.

Techniques:

Spatial Transformer Module transformes the input image such that the subsequent layers have an easier time making a classification.Earlier ,the traditional CNN deals with spatial invariance by maxpooling layer which is static as it depends on receptive field,while the spatial transformer is dynamic such that it will produce different distortions/transformations for each input image.It consists of -

-Localization network:It takes in input volume and outputs parameters (theta)of the spatial transformation that should be applied.The size of theta depends upon the transformation type that is parameterised.

-Parameterised Sampling grid:The creation of sampling grid that is the result of warping the regular grid with the theta created in the localization network.

-Sampler:Responsible for warping of the input feature map.

-Computationally fast and does not degrade the training speed.

-MNIST dataset was used which was distorted in various ways like rotation ,scale and translation,elastic warping.

Result:

The network achieved 0.65 error in comparison to traditional CNN of 0.8%.

Limitations:

The spatial transformation networks were only applied to the images of the fixed resolution and only limited scope is covered in terms of the data available.

Future Scope:

This paper demonstrated that the improvements in CNN doesn't need to come from some drastic changes in neural network architecture but can be easily improved by just applying transformations to the network.

9 Paper review:Rich Feature Hierarchies for accurate object detection and semantic segmentation

Objective:The objective of the paper is to build a Convnet that can solve the problem of object detection by providing a simple and scalable detection algorithm.

Techinques:

Detection requires a localization objects within an image,which by traditional CNN aproach used,the sliding window approach but precise localization within the sliding window is still a significant problem.

Instead ,Girshick *et. al.* ,used a "reco*gnition* using regions" which is succesful for object detection and semantic segmentation.

It consists of 3 parts-

Region proposals-Any of the methods generating category-independent region proposals can be used.Selective Search[12] is used in this case and it generates 2000 different regions that have the highest probability of containing an object.

Feature extraction- The region proposals are then warped into an image which are fed into a trained CNN,Alexnet[2] in this case,which then extracts a feature vector for each region.

SVM-The vector obtained in above step is fed into SVMs that are trained for each class and outputs a classification.Also it is fed into a bounding box regressor to obtain the most accurate coordinates.

Dataset: R-CNN was tested on 200-class ILSVRC 2013 detection dataset

Results:

The mean average precision(mAp) obtained by R-CNN was 31.4% which was better than any models present at that time.

Limitations:

-Training took multiple stages so another improved model was needed.

-Computationally expensive

-Training was extremely slow,it was taking 53 seconds per image.

Future Scope:

R-CNN paved the way for improved and faster models like Fast R-CNN and Faster R-CNN.

10 Paper review:Fast R-CNN

Objective: The objective of the paper is to improve upon the previous work [9] to efficiently classify objects using deep convolutional networks.It employs several methods to improve training and testing speed and detection accuracy.

Techniques:

Fast R-CNN takes as input the region of proposals and the image.The network then produces a convolution feature map,then for each object proposal ,A Region of Interest pooling layer extracts a fixed-length feature vector which instead of feeding into SVM is fed into fuly connected layers which uses a softmax classifier to output the probability estimates over classses and another layer that outputs four real-valued numbers for each of the K object classes.

The prime reason why Fast R-CNN faster than the R-CNN is because R-CNN features are first computed for the entire image and then a subset of those numbers are passed through SVM or softmax .As many object proposals will be overlapping ,it avouds computing CNN features for the common portions of the proposals multiple times.

The ROI pooling layer uses max pooling layer to convert the features into a feature map.

-Softmax layer is used instead of SVM.

-Pre-trained ImageNet models were used ,Alexnet[2],VGG[3] and a deeper VGG16[3] models were used.

-Fast R-CNN trained all the network weights with backpropagation.

Results:

It achieved fast training and testing in comparison to R-CNN[9].It achieved the mAP of 65.7% on VOC12 dataset.

Future Scope:

*I*t leads to other improved models called Faster R-CNN which was more effiecient and faster than the contemporary models.

11 Paper review:Visualizing and Understanding Convolution Networks

Objective:The main objective by Zeiler *et. al.,*is to provide great intuition to how the ConvNets works and illustrated more ways to improve performance.The visualization approach helped to explain the inner working of CNNs.

Techniques:

-Similar approach to AlexNet[2] ,except for few minor changes.

-AlexNet[2] trained on 15 million images while ZFNet trained on 1.3 million images.

-Used Relu for the activation functions,cross-entropy loss for the error functions and trained using batch stochastic gradient descent.

-ZFNet used filters of size 7*7 and a decreased stride value because smaller filter size helps to retain a lot of pixel information in the input volume

-Trained on a GPU for 12 days.

-Devised a new visualization technique named Deconvolution(deconvnet) Network which examines different feature activations and their relation to the input space.

At every layer of the trained CNN,we attach a deconvnet which stores the path back to image pixels.

Dataset:The model was trained on ImageNet 2012 training set of 1.3 million images and 1000 classes,50K,100K validation/test examples

Result:

A test error rate of 11.2 % was obtained and was the winner of the ILSVRC 2013 challenge.

Limitations:

ZFNet although showing an intuitive approach ,was slow in computations and was subsequently overthrown by the other ConvNets in subsequent years.

Future Scope:

The visualizing techniques used in ZFNet helped the researcher to visualize the CNN correctly and helps what was going inside the architecture.

12 Paper Review:Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Objective:The main objective behind Srivastava *et. al.,*is to address the problem of overfitting in large neural network by introducing a new regularization method called dropout.

Techniques:

Dropout is a technique in which some of the units of neural network is dropped out from calculating the result which results in reducing the dependency on any of the units to adapt too much which in turn reduces overfitting.The choice of which units is dropped is random and is generally a probability is attached to each units.A sparse or thin network is obtained after applying dropout.Dropout neural networks are trained using stochiastic gradient descent in a manner similar to standard neural nets.

Dropout neural networks were for classification on different datasets-

-MNIST

-TIMIT

-CIFAR-10,CIFAR-100

-Street View House Numbers Data Set

-ImageNet

-Reuters-RCV1

-Alternating Splicing Data Set

Large diversity of datasets was choose to avoid typecasting dropout technique in a particular single genre.

Result:

Dataset	Accuracy Without Dropout	Accuracy With Dropout
MNIST	1.60%	1.35%
SVHN	3.95%	2.55%
CIFAR-10	14.98%	12.61%
CIFAR-100	43.48%	37.20%
ImageNet	26%	16%
TIMIT	23.4%	21.8%

Limitations:

It increases training time and generally takes 2-3 times longer to train than a standard neural network of the same architecture.

13 Paper Review:Batch Normalisation:Accelerating Deep Network Training by Reducing Internal Covariate Shift

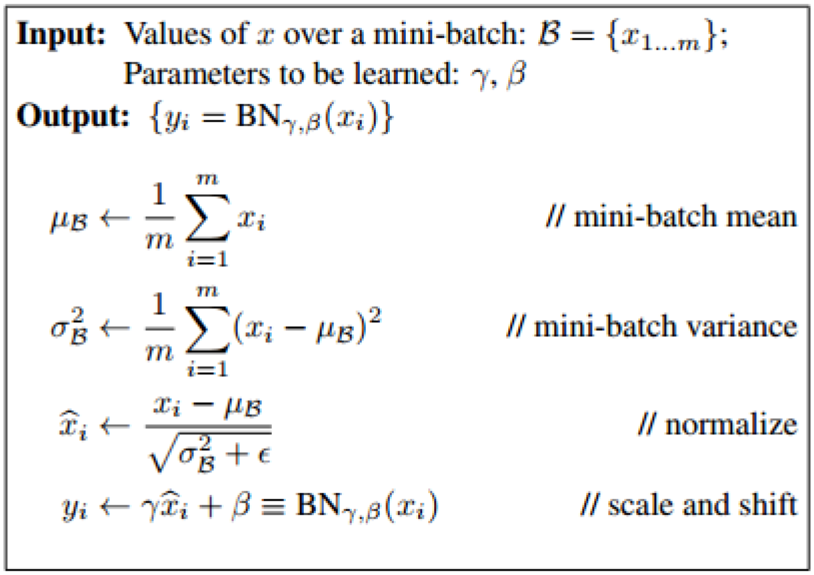
***Objective**:*The main objective by Ioffe *et.al,* is to resolve the problem of internal covariate shift by normalizing layer inputs and in turn allows to use higher learning rates.

Techniques:

During training time,the distribution of each layer's input changes during training ,as the previous layer's parameters changes which slows down the training by requiring lower learning rates.,phenomenon called internal covariate shift.

Instead of having all the input layers ,zero mean and unit variances,normalization will be done of each scalar feature as normalizing each layer is costly

Algorithm:



```
Input: Values of $x$ over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
       Parameters to be learned: $\gamma$, $\beta$
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

    $\mu_\mathcal{B} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$    // mini-batch mean
    $\sigma_\mathcal{B}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_\mathcal{B})^2$    // mini-batch variance
    $\widehat{x}_i \leftarrow \frac{x_i - \mu_\mathcal{B}}{\sqrt{\sigma_\mathcal{B}^2 + \epsilon}}$    // normalize
    $y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i)$    // scale and shift
```

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

Batch Normalization enable higher learning rates as it prevents small changes in layer parameters to amplify

Result:

The models with batch normalization were trained with Inception,architecture and on ImageNet and MNIST dataset and learning rate was improve drastically.

Future Scope:

Batch Normalisation on RNN has yet to be explored and will be explored in future models

14 Paper Review:Microsoft COCO:Common Objects in Context

Objective:The main goal by the authors is to introduce a new large-scale dataset that helps to resolve the three core research problems of detecting non-iconic views of objects,contextual reasoning between objects and precise 2D localization of objects.

Techniques:

To achieve the goals stated,Amazon Mechanical Turk was used to gather data.A large set of images containing contextual relationship and non-iconic object views were collected.

The Microsoft COCO contains 91 common object categories,with 2500000 labeled instances in 328000 images.

Several sources to collect entry-level object categories of “things.” We first compiled a list of categories by combining categories from PASCAL VOC,and a subset of the 1200 most frequently used words that denote visually identifiable objects.To further augment our set of candidate categories, several children ranging in ages from 4 to 8 were asked to name every object they see in indoor and outdoor environments.The final selection of categories attempts to pick categories with high votes, while keeping the number of categories per supercategory (animals, vehicles, furniture, etc.) balanced. Categories for which obtaining a large number of instances (greater than 5,000) was difficult were also removed. To ensure backwards compatibility all categories from PASCAL VOC are also included

The next goal was to collect the set of candidate images.Iconic images have the benefit that they may be easily found by directly searching for specific categories using Google or Bing image search. While iconic images generally provide high quality object instances, they can lack important contextual information and non-canonical viewpoints.

Then correspondingly images were annoted by Instance segmentation and Instance spotting.

Result:

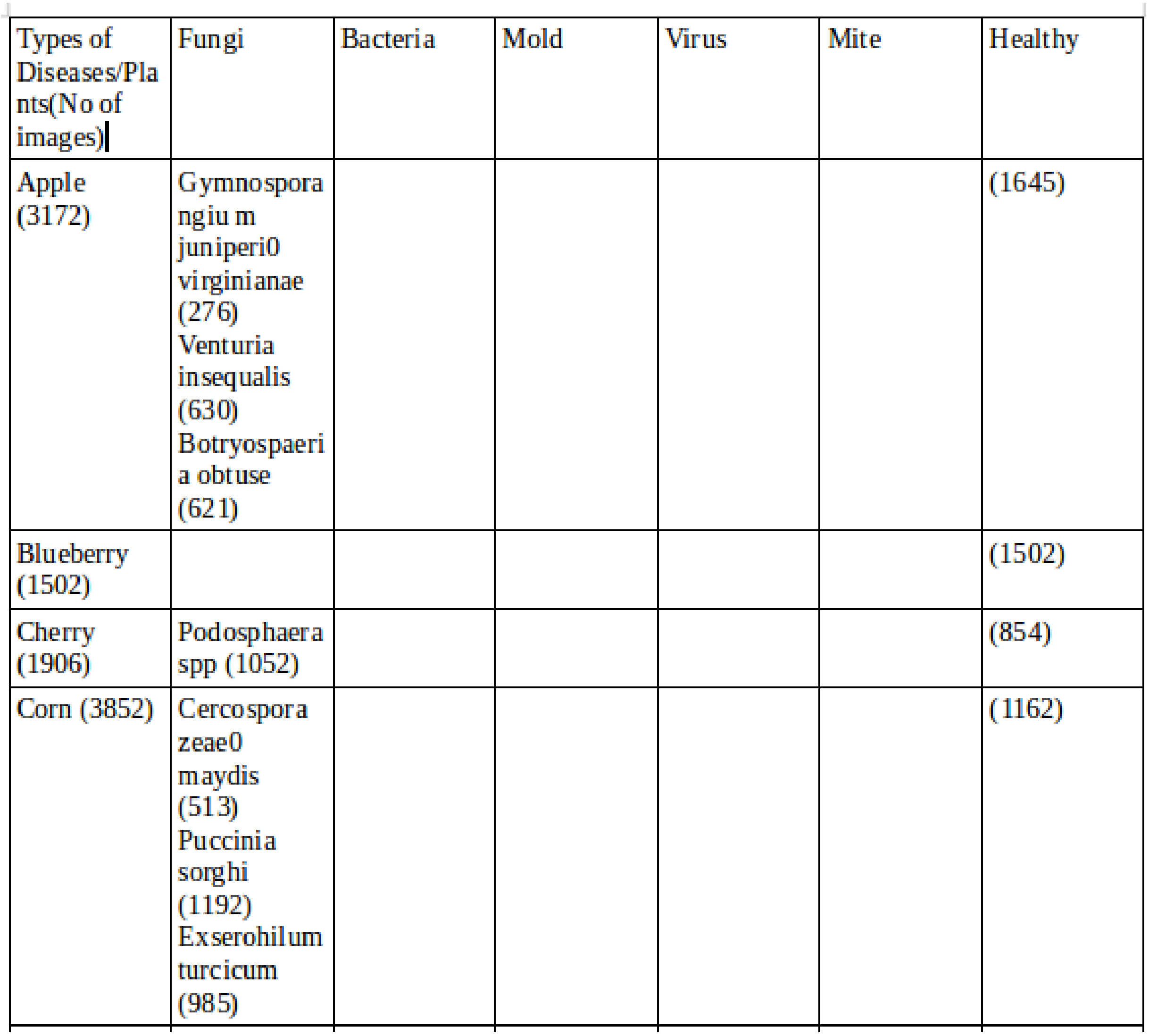
A large dataset to help resolve out research problems was made succesfully and is now used as one of the benchmarks to perform well.

Chapter 3

System Development

3.1 **Dataset**

Dataset obtained from Plant Village platform[15] were used. Dataset contains 54,309 images ranging for 14 plant species.



Types of Diseases/Plants(No of images)	Fungi	Bacteria	Mold	Virus	Mite	Healthy
Apple (3172)	Gymnosporangiu m juniperi0 virginianae (276) Venturia insequalis (630) Botryospaeria obtuse (621)					(1645)
Blueberry (1502)						(1502)
Cherry (1906)	Podosphaera spp (1052)					(854)
Corn (3852)	Cercospora zeae0 maydis (513) Puccinia sorghi (1192) Exserohilum turcicum (985)					(1162)

Grape (4063)	Guignardia bidwellii (1180) Phaeomoniella-- spp. (1384) Pseudocercospor a vitis (1076)					(423)
Orange (5507)		Candidatus Liber ibacter (5507)				(5507)
Peach (2657)		Xanthomonas campestris (2291)				(360)
Bell Pepper (2475)		Xanthomonas campestris (997)				(1478)
Potato (2152)	Alternaria solani (1000)		Phytophthora Infestans (1000)			(152)
Raspberry (371)						(371)
Soybean (5090)						(5090)
Squash (1835)	Erysiphe cichoracearum / Sphaerotheca fuliginea (1835)					1835
Strawberry (1565)	Diplocarpon earlianum (1109)					(456)
Tomato (18,162)	Alternaria solani (1000) Septoria lycopersici (1771) Corynespora cassiicola (1404) Fulvia fulva (952)	Xanthomonas campestris pv. Vesicatoria (2127)	Phytophthora Infestans (1910)	Tomato Yello Leaf Curl Virus (5357) Tomato Mosaic Virus (373)	Tetranychus urticae (1676)	(1592)

Fig3.1 Dataset obtained from Plant Village[15]

3.2 Software Used-

Keras: Keras is a high level python based API for building neural network.It is build on top of CNTK,Theano,tensorflow.By default,tensorflow is the backend on keras on linux while on Mac OS ,the defualt is Theano.It is easy to build neural network and is suitable for easy prototyping.

Tensorflow:Tensorflow is low level based python based API to develop complex neural network .Tensorflow is build by Google AI team under the supervision of Jeff Dean ,the lead Google AI scientist which were responsible to build the framework.It is under the competiton of facebook's pytorch.

3.3 Hardware Used:

Amazon Web Services EC2 instances.

G2.2 x large.

Four NVIDIA GRID

GPUs, each with 1,536 CUDA cores-32

vCPUs.-60 GiB of memory-240 GB (2 x 120) of SSD storage

CHAPTER-4

RESULT AND PERFORMANCE ANALYSIS

4.1-Convolution Neural Network

Convolution Neural network is neural network with convolutional layers which helps in enhancing the performance of network as the convolution layer does the implicit feature extraction on its own.

Basic terminology

Convolution-Convolution is the operation which occurs in convolution layer where a filter or kernel is hovered around a receptive field and then the element wise dot product is obtained which is then given as output.

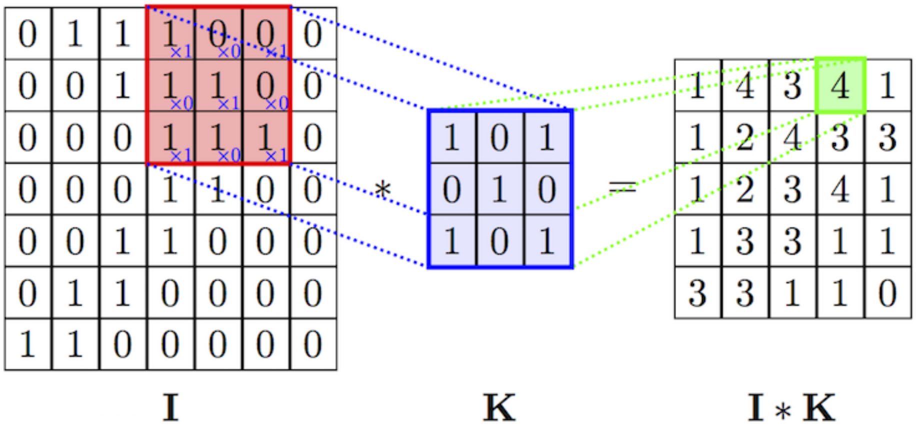
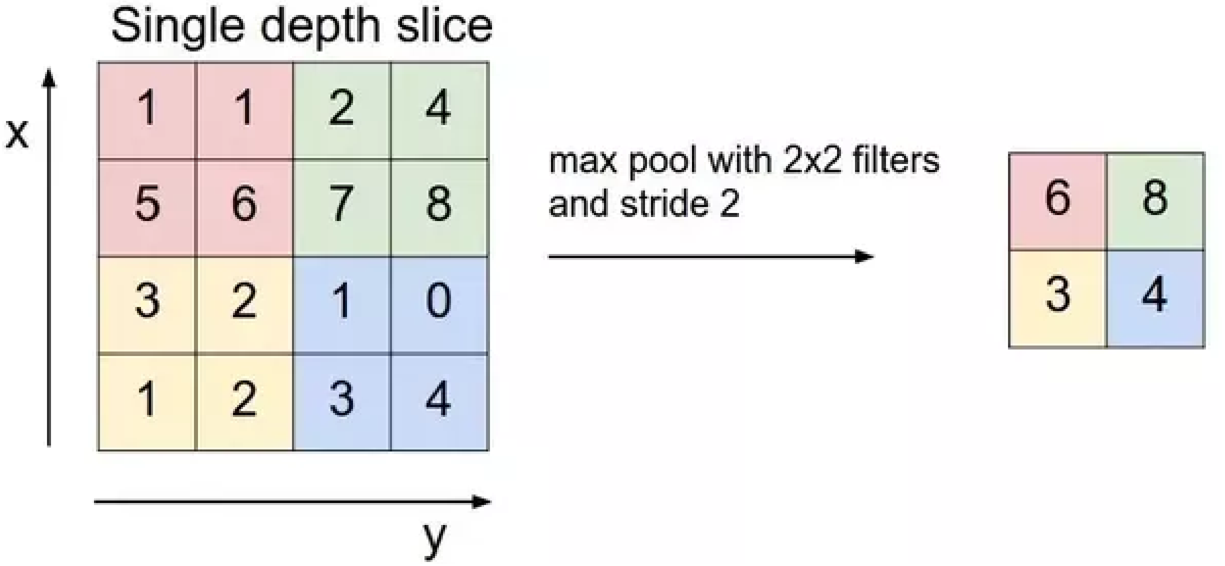


Fig 4.1:Convolution Operation

Max Pooling- Max pooling is the operation where the maximum of the values in the field is obtained and the rest is discarded. It is useful to magnify the features which are prominent .



1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

6	8
3	4

Fig 4.2-Max Pooling

Max Pooling leads to loss in information as the most prominent value is only considered while the rest of the values are lost in pooling which may leads to loss of some data which can be resolved out using padding layers as they pad the values with additional bits.

Padding-In Padding operation,extra bits are added at the edges of the images to maintain the effect of top most and bottom layers of images to retain their significance ,otherwise the output will be more biased to the values present in between.

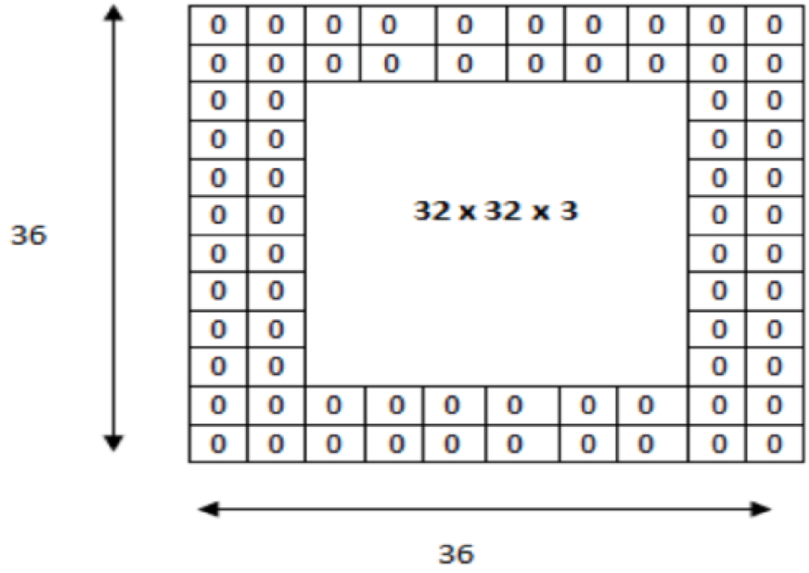


Fig:4.3-Padding Operation

4.2 Gradient Descent-Gradient Descent is most popularly used algorithm used in neural network to optimize the cost function. The main intuitive idea behind it is that suppose if a person is walking down a valley then he/she have to walk through steepest slope in order to reach the bottom fastely and accurately.

Pseudocode-

```
Initialize randomly.
Calculate the gradients.
Update the weights.
Repeat until the required value is obtained.
```

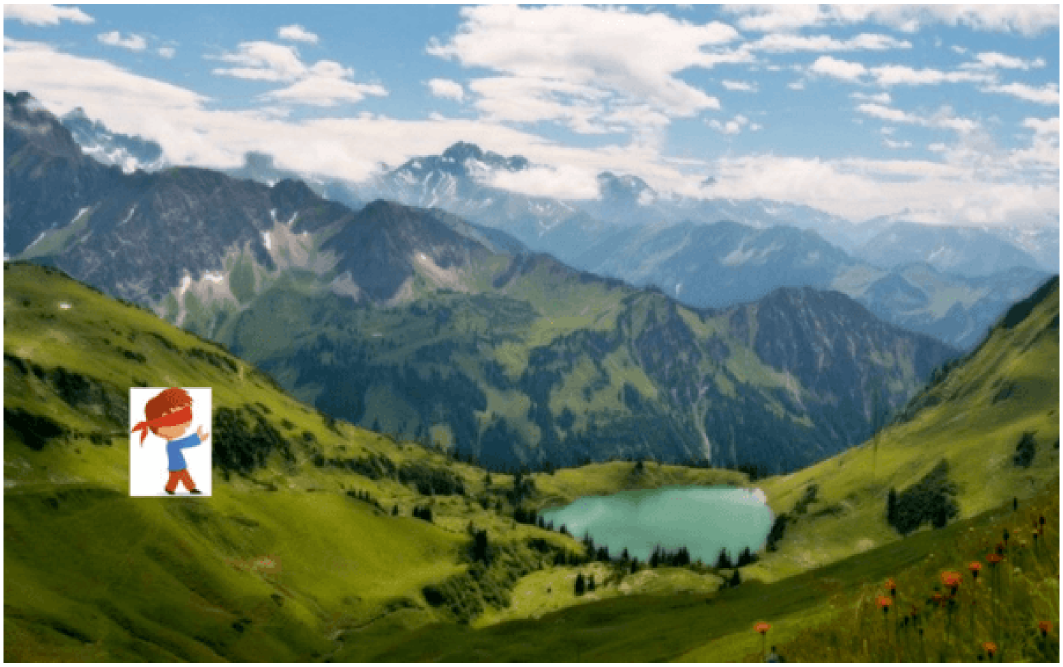


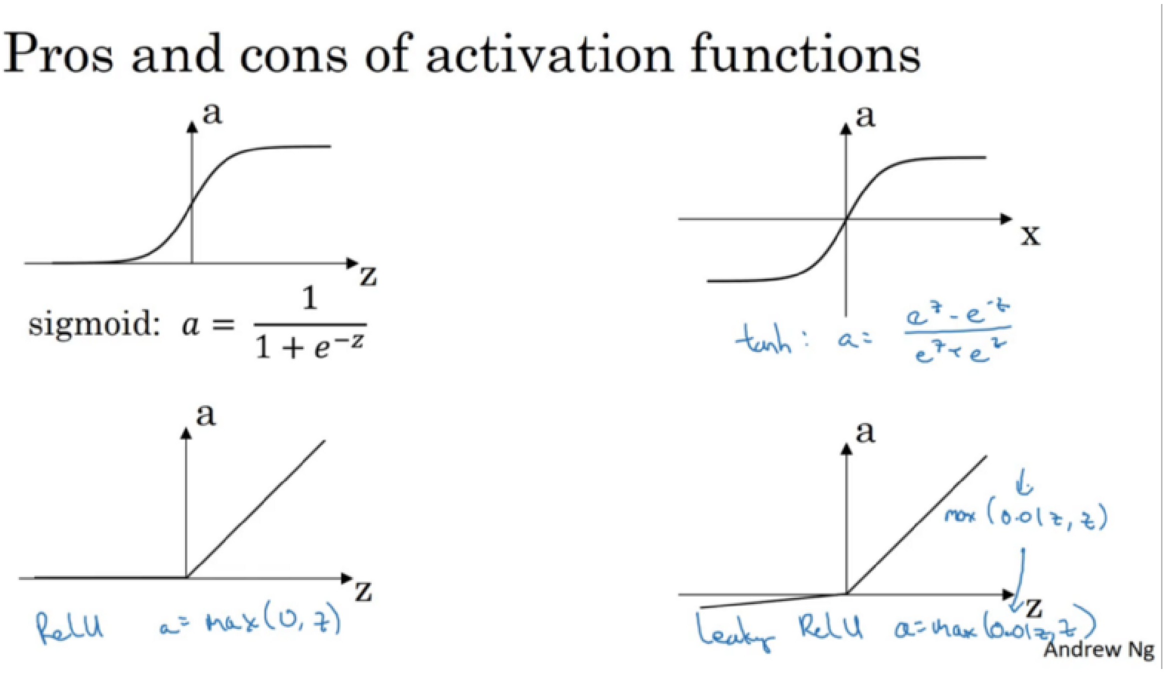
Fig 4.4-Intuitive Idea behind gradient Descent

4.3 Activation Functions-Activation functions are the type of functions which takes the output of the calculated previous values and then instigate the values after they exceeds a particular value. It is similar to firing up of neurons in brain cells,when a certain level of electric charge is received.

There are mainly four types of activation functions which are used-

- Sigmoid
- Tanh
- Relu
- Leaky Relu

Fig 4.5-Pros and Cons of Activation Functions



Pros and cons of activation functions

Fig 4.5 :Different Types of Activation Functions

Sigmoid functions is used mainly for binary classification .In other cases ,relu is preferred as they provide much accurate results.

4.4 Dataset-

Dataset is generally divided into Training and Dev/Test set in the ratio of 80,20 % respectively.

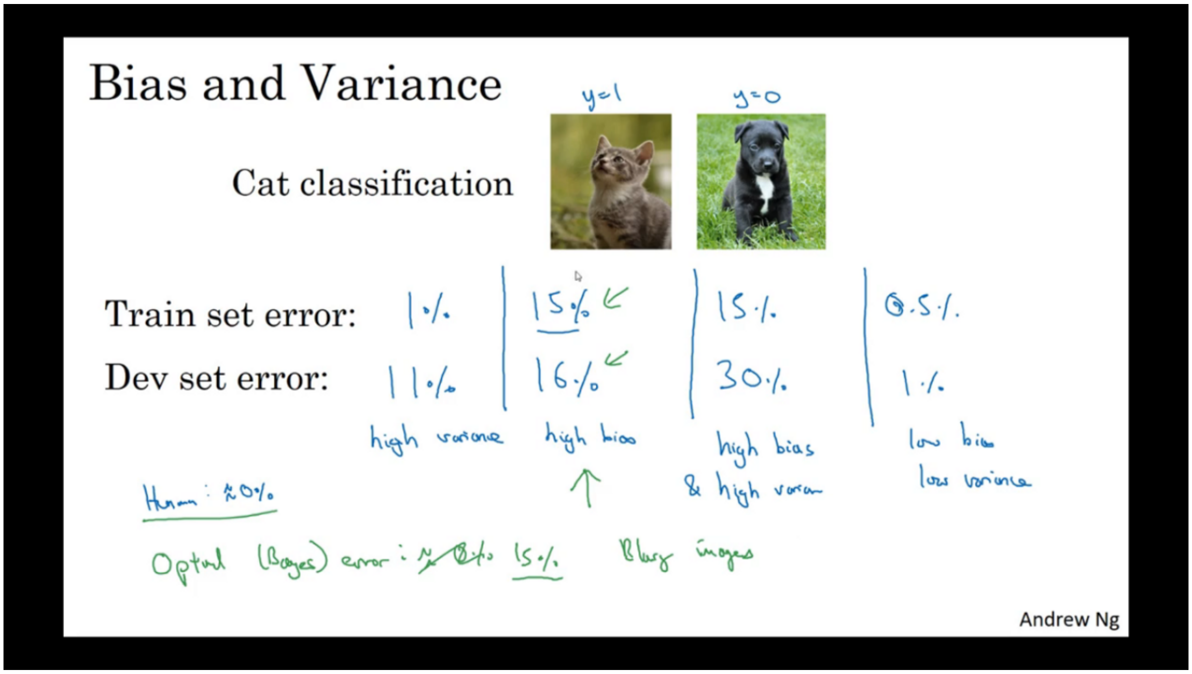
Train Set-Algorithm is trained on this dataset and the result is optimized accordingly.

Dev Set-Most of the times confused with Test set,but is used to tune the hyperparameters.

Test Set-Test set is the unseen data on which the performance of algorithm is observed and the performance of the algorithm on it is pretty much essential.

4.5 Bias and Variance Problem-

When there is a significant difference between train set and dev set then the problem is called of high variance where as when the train/dev set is more than the human error ,then it is bias problem.



Bias and Variance

Cat classification

Train set error:	1%	15%	15%	0.5%
Dev set error:	11%	16%	30%	1%
	high variance	high bias	high bias & high variance	low bias low variance

Human : ≈0%

Optimal (Bayes) error : ~~≈0%~~ 15% Blurry images

Andrew Ng

Fig 4.6 Bias and Variance Problem on Cat classification problem

Regularization method are used to remove the bias and variance and bias problem..

4.6 Regularization-

There are various techniques which are employed to remove the high bias and high variance problem.

Plenty of techniques are employed in order to solve the problem of bias and variance.

Overfitting is solved by manipulating the neural network models's complexity by introducing a new term in the cost function to chastise the much greater weights.This will make the model to be less complex where complexity is defined by greater weights and overfitting is done by larger weights

There are other regularization techniques as L1 regularization or Lasso Regularization,which promotes its model's parameters to become void and hence the model becomes less dense in comparison to the models using other form of regularization.

4.7 Dropout-Dropout is another regularization technique in which each node in the neural network have an equal chance of staying or getting removed which reduce the dependence of network on a particular node and hence reduce overfitting.

Apart from these techniques,data augmentation by rotating or cropping images are done to increase the randomness in the dataset and hence ,it acts as a real world sandbox of actual data.

4.8 Hyperparameter Tuning-

Hyperparameters are those variables whose value has to be adjusted before training of neural network starts.The value of these are calibrated accordingly and then passed .

The various hyperparameters are-

- Layers-No of layers are one of the important hyperparameters.
- Hidden Units-No of neurons decides whether the result will be overfitted or perfectly achieved.
- Learning Rate-The intuitive idea behind the learning rate is the speed and size of the steps someone take to correctly reach at the end of the value as observed in gradient descent method.
- Mini Batch Size-Mini batch size represents size of the batch,it depends on memory available.

To tune hyperparameters ,arbitrary values are suggested to use rather than using a well known method of using a grid.But it can lead to large sample space ,in order to eradicate it ,a known checkpoint can be observed by randomly using any values and then steering the values around it.

A proper measure to pick hyperparameter is preferred like an exponential or logarithmic scale instead of linear scale.

Testing a model multiple times when plenty of resources are available and then choosing the one which is best fit and if only few resources are there ,testing a model continously after regular interval of time.

Greater training times can be largely avoided by using appropriate hyperparameters.

4.9 Architecture-

Various current architectures were used to test on dataset and their values are evaluated.

AlexNet-

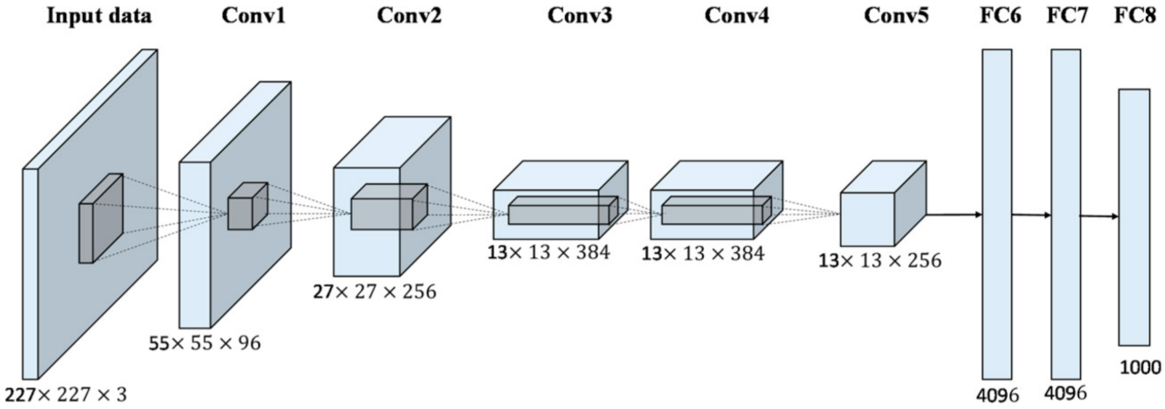


Fig 4.7-AlexNet

AlexNet consists of 8 layers ,out of which 5 are convolutional layers while rest are fully connected layers.

VGG-16-

VGG-16 investigated the effect of depth of convolution networks on its accuracy in large scale image image recognition settings.The depth of the neural network was steadily increased by adding more convolution layers and keeping all the other parameters fixed.

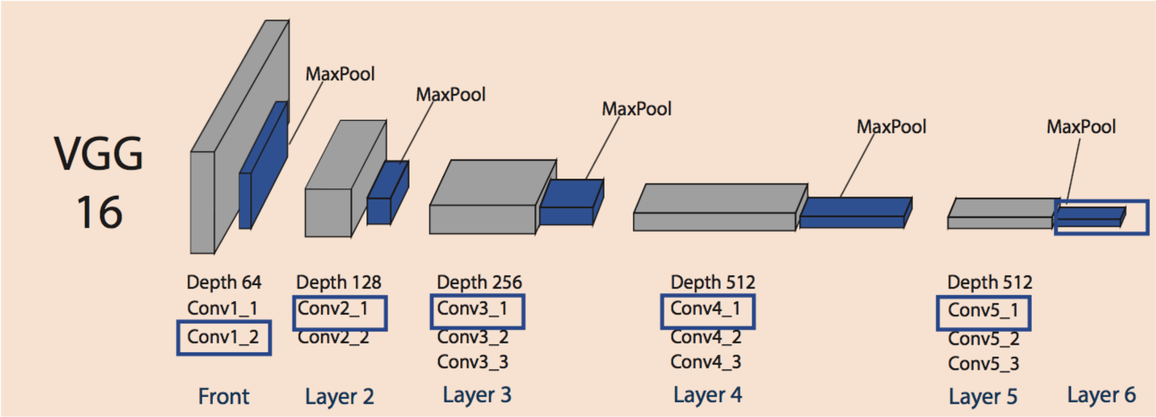


Fig 4.7-VGG16

t proved that depth have a positive effect on the accuracy of convolution layers.

It extensively used pooling layers to increase the effect of required features and hence resulted in more accurate prediction.

Padding was also used to increase the effect of it and hence it was a great success in 2012 Image Recognition Challenge.

ResNet-ResNet was invented by Google and it introduced the notion of residual network which introduced feed forward networks .

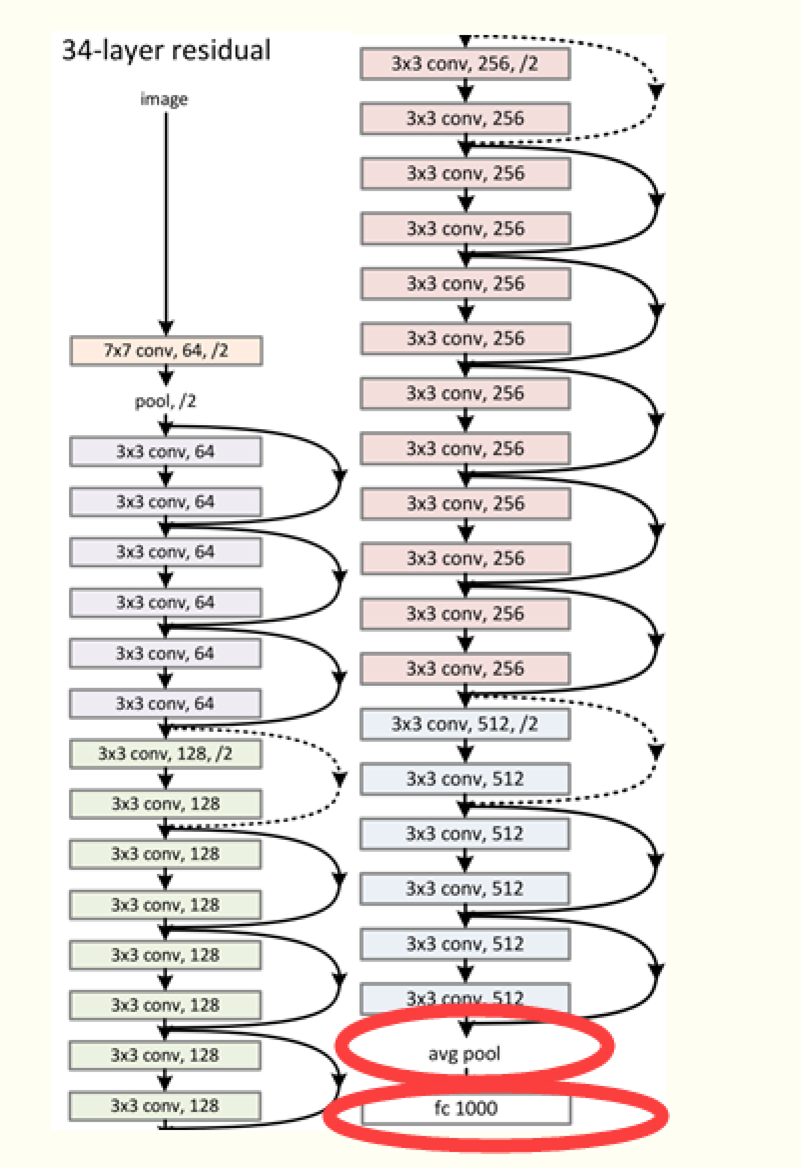


Fig 4.8:ResNet Architecture

CHAPTER-5

CONCLUSION

On running the algorithm for the sample of 600 images on tomato plants images,training accuracy of 76% was obtained and the validation accuracy of 36% was obtained.

The complete dataset was not used due to the limited computation power available at hand.

5.1 Conclusions

Training Accuracy-The training accuracy is the accuracy of the convolutional neural network on the training set.

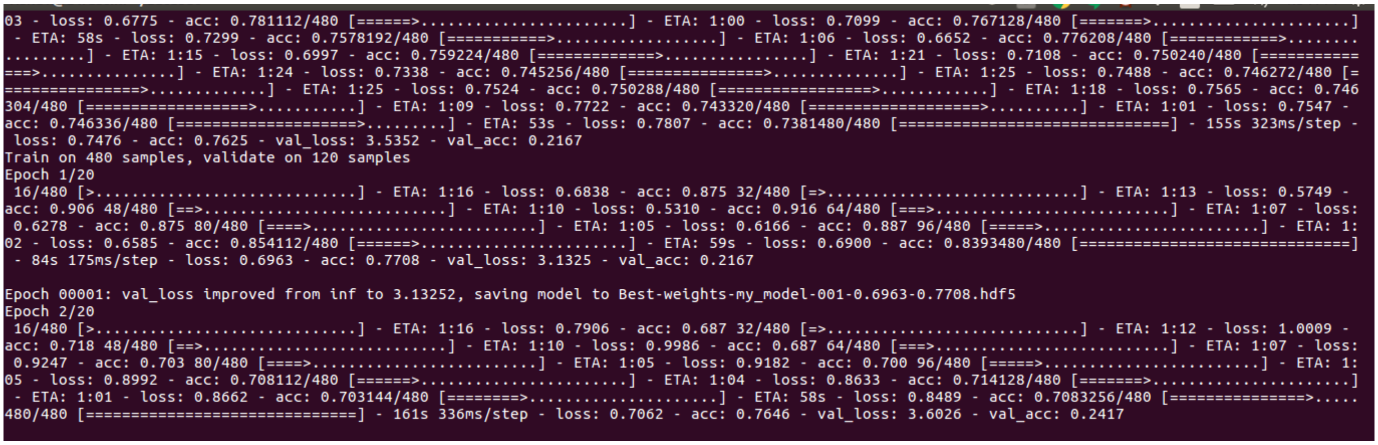
Training Loss-Trainig Loss is the error on the training set.Various loss functions like cross-entropy loss,mean-squared loss are calculated.The tendency is to minimize the loss and maximize the training accuracy.

Validation Accuracy-The validation accuracy is the accuracy of the convolutional neural network on the validation/test set.Performance of algorithm depends on this benchmark as it represents the performance on unseen data.

Validation Loss-Validation Loss is the error on the validation set/test set of the data.The tendency is to minimize the test loss and maximize the validation accuracy.

Result:

Training Accuracy	76%
Training Loss	0.7062
Validation Accuracy	36%
Validation Loss	3.6026



```
03 - loss: 0.6775 - acc: 0.781112/480 [======>.......................] - ETA: 1:00 - loss: 0.7099 - acc: 0.767128/480 [=======>......................]
 - ETA: 58s - loss: 0.7299 - acc: 0.7578192/480 [===========>..................] - ETA: 1:06 - loss: 0.6652 - acc: 0.776208/480 [============>........
.........] - ETA: 1:15 - loss: 0.6997 - acc: 0.759224/480 [=============>................] - ETA: 1:21 - loss: 0.7108 - acc: 0.750240/480 [===========
===>...............] - ETA: 1:24 - loss: 0.7338 - acc: 0.745256/480 [===============>..............] - ETA: 1:25 - loss: 0.7488 - acc: 0.746272/480 [=
===============>.............] - ETA: 1:25 - loss: 0.7524 - acc: 0.750288/480 [=================>............] - ETA: 1:18 - loss: 0.7565 - acc: 0.746
304/480 [==================>...........] - ETA: 1:09 - loss: 0.7722 - acc: 0.743320/480 [===================>..........] - ETA: 1:01 - loss: 0.7547 -
acc: 0.746336/480 [====================>.........] - ETA: 53s - loss: 0.7807 - acc: 0.7381480/480 [==============================] - 155s 323ms/step -
 loss: 0.7476 - acc: 0.7625 - val_loss: 3.5352 - val_acc: 0.2167
Train on 480 samples, validate on 120 samples
Epoch 1/20
 16/480 [>.............................] - ETA: 1:16 - loss: 0.6838 - acc: 0.875 32/480 [=>............................] - ETA: 1:13 - loss: 0.5749 -
acc: 0.906 48/480 [==>...........................] - ETA: 1:10 - loss: 0.5310 - acc: 0.916 64/480 [===>..........................] - ETA: 1:07 - loss:
 0.6278 - acc: 0.875 80/480 [====>.........................] - ETA: 1:05 - loss: 0.6166 - acc: 0.887 96/480 [=====>........................] - ETA: 1:
02 - loss: 0.6585 - acc: 0.854112/480 [======>.......................] - ETA: 59s - loss: 0.6900 - acc: 0.8393480/480 [==============================]
 - 84s 175ms/step - loss: 0.6963 - acc: 0.7708 - val_loss: 3.1325 - val_acc: 0.2167

Epoch 00001: val_loss improved from inf to 3.13252, saving model to Best-weights-my_model-001-0.6963-0.7708.hdf5
Epoch 2/20
 16/480 [>.............................] - ETA: 1:16 - loss: 0.7906 - acc: 0.687 32/480 [=>............................] - ETA: 1:12 - loss: 1.0009 -
acc: 0.718 48/480 [==>...........................] - ETA: 1:10 - loss: 0.9986 - acc: 0.687 64/480 [===>..........................] - ETA: 1:07 - loss:
 0.9247 - acc: 0.703 80/480 [====>.........................] - ETA: 1:05 - loss: 0.9182 - acc: 0.700 96/480 [=====>........................] - ETA: 1:
05 - loss: 0.8992 - acc: 0.708112/480 [======>.......................] - ETA: 1:04 - loss: 0.8633 - acc: 0.714128/480 [=======>......................]
 - ETA: 1:01 - loss: 0.8662 - acc: 0.703144/480 [========>.....................] - ETA: 58s - loss: 0.8489 - acc: 0.7083256/480 [===============>.....
480/480 [==============================] - 161s 336ms/step - loss: 0.7062 - acc: 0.7646 - val_loss: 3.6026 - val_acc: 0.2417
```

Fig 5.1-Screenshots of the Epochs

Training Accuracy of 76% is fairly impressive as if trained on complete dataset of 64,309 images.

Validation accuracy of 36% is fairly low but will improve accurately when trained on complete dataset.

5.2 Future Scope

This project can be used extensively by agricultural scientist to increase their productivity and result in increasing the yield of farmers.

Agriculture field can benefits itself greatly with the advances happening in the field of Artificial Intelligence. With the advent of deep learning ,the accuracy of image classification and image detection has been improved greatly and if properly applied in the field of agriculture can improve the farmer's productivity and hence reduce the vast amount of losses every year farmers faces due to diseased crops. Earlier the accuracy of image classification reached to an optimal due to limits of Support Vector Machines but now with Deep learning and with huge amount of data available and inexpensive computational resources available ,We can easily deploy the efficient Deep Learning models in the productions phase and harbor the benefits from it.

REFERENCES

[1] K. Simonyan , A. Zisserman Very Deep Convolution Networks for Large Scale Image Recognition,Semanticscholar

[2] A.Krizhevsky,I.Sutskever,G. Hinton,ImageNet Classification with deep convolutional neural networks,Communications of the ACM,Vol. 60,No. 6,pp. 84-90,2017

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition

"*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778.

[4] C. Szegedy ,W.Liu,Y. Jia,P. Sermanet,S. Reed,D. Anguelov,Dumitru Erhan,V. Vanchouke,A. Rabinovich ",Going deeper with convolutions,"*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9. doi: 10.1109/CVPR.2015.7298594

[5] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions,"*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1800-1807.

[6] Ian J. Goodfellow,Jean Pougat-Abadie,Mehdi Mirza,Bing Xu,David Warde-Farley,Sherji Ozair,Yoshua Bengio,"Generative Adversial Networks",NIPS 2014

[7] A. Karpathy and L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 664-676, April 1 2017.

[8] M. Jaderberg,K. Simonayan,A. Zisserman,K. Kavukcuoglu,"Spatial Transformer Networks",Neural Information Processing Systems,2015

[9] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448.

[10] Zeiler M.D., Fergus R. (2014) Visualizing and Understanding Convolutional Networks. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham

[11] Zeiler M.D., Fergus R. (2014) Visualizing and Understanding Convolutional Networks. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham

[12] N. Srivastava,G. Hinton,A. Krizhevsky,I. Sutskever,R. Salakhutdinov,”Dropout:A Simple Way to Prevent Neural Networks from Overfitting”,The Journal of Machine Learning Research,Volume 15 Issue 1.January 2014,Pages 1929-1958

[13] S. Ioffe,C. Szegedy,”Batch Normalisation:Accelerating Deep Network Training by Reducing Internal Covariate Shift”,ICML15,Vol 37,Pages 448-456,July,2015

[14] T. Lin,M. Maire,S. Belongie,J. Hays,P. Perona,D.Ramanan,P.Dollar,C. Lawrence Zitnick,”Microsoft COCO”,Computer Vision-EECV 2014,VOL. 8693,Springer Cham

[15] D. Hughes ,M. Salathe,”An open access repsository of images on plant health to enable the development of mobile disease diagnostics.”.arxiv.org