Table of Contents

[1. Problem Statement and Motivation 1](#_Toc496047654)

[2. Dataset 1](#_Toc496047655)

[3. Data Generation 1](#_Toc496047656)

[4. Model 2](#_Toc496047657)

[5. Regularization 3](#_Toc496047658)

[Dropout 3](#_Toc496047659)

[Noise 3](#_Toc496047660)

[Image Augmentation 3](#_Toc496047661)

[6. Optimization 4](#_Toc496047662)

[Stochastic Gradient Descent 4](#_Toc496047663)

[RMSprop 4](#_Toc496047664)

[7. Performance Measurement 4](#_Toc496047665)

[8. Future Scope 4](#_Toc496047666)

[References 5](#_Toc496047667)

# Problem Statement and Motivation

Fine grain image classification using Convolutional Neural Network i.e. classify cars based on its type like Sedan, Hatchback, SUV and Convertible and identify Sedan cars.

One of the most current achievements in the IT world is the Self-Driven or Autonomous cars. These are the unmanned ground vehicles that are capable of sensing its environment and navigating without human input. While there are umpteen number of challenges in developing a self-driven car, one of those challenges would be to develop an artificial intelligence which would program the car to identity the objects in front of it, which most likely will be cars. Hence using this thought, we are training our data into classifying the cars into its types like Sedan, Hatchback, SUV and Convertible. The final aim will be to identify if the new input car is of type Sedan. Keeping the scope of the project in mind, we have limited the classification to only separating Sedan from other classes.

# Dataset

CIFAR 10 is a well-established computer vision dataset prepared as a part of a paper by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. It is a subset of the 80 million tiny images dataset and consists of 60,000 32x32 color images containing one of 10 object classes, with 6000 images per class, hence the name CIFAR 10. There is a sister dataset which has 100 classes, namely CIFAR 100. In CIFAR 10, there are 50000 training images and 10000 test images. The 10 different object classes are Airplane, Automobile, Bird, Cat, Deer, Dog, Frog, Horse, Ship and Truck.

For the purpose of this project, we have taken the automobile/car dataset. CIFAR 10 provides an image classification on 10 different classes, for this project we are aiming for a fine grain classification on one type of class from this dataset.

# Data Generation

Using 1500 images out of the 5000 under the automobile/car dataset we have taken 1000 as training data, 200 as validation data and 300 as test data. For training the data, we have labeled the 1000 images based on the different types of car, i.e. Sedan, Hatchback, SUV and Convertible. The purpose of this project is to be able to classify a new input in our system and correctly predict if the car is of type Sedan or not

# Model

Convolution Neural Network is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. It is a category of Neural Networks that have proven very effective in areas such as image recognition and classification. It have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driven cars, the motivation for our project.

There are four main operations in the Convolution Neural Network.

1. Convolution
2. Non Linearity (ReLU)
3. Pooling or Sub Sampling
4. Classification (Fully Connected Layer)

These operations are the basic building blocks of every Convolutional Neural Network. Below is a pictorial representation of the same.



The convolution operation is used to extract features from the input image. It preserves the spatial relationship between pixels by learning image features using small squares of input data. There are filters that slides over the image to capture the feature map. Convolution Neural Network learns these filters on its own during the training phase.

Next operation in our network is ReLU that stands for Rectified Linear Unit. It is used to introduce non-linearity in our system, since the real world data is mostly non-linear. In this step all the negative pixels in our feature map is replaced by zero.

Pooling or subsampling reduces the dimensionality of each feature map but retains the most important information. In our project we are using Max pooling. In this case we take a 2x2 window, and sample it with the maximum value in the given 2x2 window.

These three operations comprise of one convolutional layer. In case of multiple layers, the output of the Pooling layer is fed as the input for convolution and the same process is repeats. In our system we are using three CNNS.

Using the high level features from the multiple convolutions layers, we use the Fully Connected Layer to classify the input image into categories, in our case, Sedan or other category. The Convolution + Pooling layers act as Feature Extractors from the input image while Fully Connected layer acts as a classifier.

# Regularization

We are using multiple techniques for regularization.

## Dropout

Dropout is a regularization technique for reducing over fitting in neural networks by preventing complex co-adaptations on training data. It is a very efficient way of performing model averaging with neural networks. The term "dropout" refers to dropping out units (both hidden and visible) in a neural network. Dropout is easily implemented by randomly selecting nodes to be dropped-out with a given probability using the method *Dropout (0.25)* implemented by Keras. 0.25 here is the dropout rate.

## Noise

Injecting input noise can also be used for regularization. Addition of random noise tells the network that output is not effected around the exact input. It helps to add robustness against over fitting to the model. It complements by providing a margin around the model parameter and gives a better way to avoid over fitting by making output impervious to small perturbations in input.

## Image Augmentation

Image Augmentation is a process where we take images from the training data and manipulate the same to create altered version of the same image. By doing so we are adding versions of the same image from different angels, all belonging to the same label, hence increasing the prior knowledge in our system.

# Optimization

## Stochastic Gradient Descent

Gradient descent is one of the most popular algorithms to perform optimization and by far the most common way to optimize neural networks. Stochastic gradient descent (often shortened to SGD) is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions. In other words, SGD tries to find minima or maxima by iteration.

Unlike the gradient descent, in SGD we perform update for each training example and label.



## RMSprop

RmsProp is an optimizer that utilizes the magnitude of recent gradients to normalize the gradients. We always keep a moving average over the root mean squared (hence Rms) gradients, by which we divide the current gradient.

# Performance Measurement

We started by labeling 500 images of the car in order to construct a dataset but later on increased it to 1500 dataset in order to provide more data for proper prediction of our constructed model.

Earlier we created two convolutional layers and added dropout and max pooling at each layer. We also added 19 hidden layers. At each and every hidden layer we have used Relu activation function in order for the non-linear model.

Then we added one more convolutional layer with more feature maps.

We did parametric tuning to extract the best parameters for our model.

In order to measure performance of the model accuracy has been used as a metrics.

After doing all these we achieved a training accuracy of 62% and validation accuracy of 71%. Then we predict classes on test set and achieved an accuracy of 57%

# Future Scope

Out of 1500 extracted images, we have majority of Sedan images i.e. 730 of all. The remaining images consist of Hatchback, Convertible and SUV. As the data for other classes are less while testing the accuracy of the model it is getting biased towards Sedan. Thus, for majority of the images it is predicting it to be Sedan.

We can further improve it by adding more images for other classes as well and train the model again using Alex Net to give a better accuracy.

# Reference

1. CIFAR 10 - [*https://www.cs.toronto.edu/~kriz/cifar.html*](https://www.cs.toronto.edu/~kriz/cifar.html)
2. Convolution Neural Network - [*https://en.wikipedia.org/wiki/Convolutional\_neural\_network*](https://en.wikipedia.org/wiki/Convolutional_neural_network)
3. [Convolutional Neural Networks for Visual Recognition](http://cs231n.github.io/convolutional-networks/)
4. [An Intuitive Explanation of Convolutional Neural Networks](https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/)
5. [Dropout Regularization in Deep Learning Models With Keras](https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/)
6. Dropout - <https://en.wikipedia.org/wiki/Dropout_(neural_networks)>