**A Detailed report**

**on Deep Neural Network**

**Roopesh Bharatwaj K R**

**A00279933**

Athlone Institute of Technology

##### Declaration

I hereby certify that the material, which is submitted in this report is entirely my own work and has not been submitted for any academic assessment other than part fulfilment of assignment.

Future students may use the material contained in this report provided that the source is acknowledged in full.

**Signed:** Roopesh Bharatwaj K R

**Date:** 5th May 2021.

##### Abstract

Deep Neural Network (DNN) is one of the most trending topic around the globe. Which is a part of artificial intelligence and the topmost sought job in 2021. Perhaps, one can say that Artificial intelligence is a novel field and its currently in its pinnacle. If so, then Its wrong, as the artificial intelligence and neural networks are dated back to the mid-19th century, work has been started and most of research work has been carried out before 20th century and now we are just exploring and enhancing them., This paper talks about the detailed history of Deep Neural Networks , and its rise and fall until now. Furthermore, detailed analysis of a case study of an important library and a commercial application which are widely being used now. And a detailed information of the futuristic scope of Deep Neural Networks .

##### Table of Contents

[Declaration ii](#_Toc37502448)

[Abstract iii](#_Toc37502449)

[Table of Contents iv](#_Toc37502450)

[Chapter 1: Introduction 4](#_Toc37502453)

[Chapter 2: History of Deep Neural Networks 5](#_Toc37502454)

[Chapter 3: Case Study](#_Toc37502455) 10

[Chapter 4: The Future 15](#_Toc37502456)

[Chapter 5: Conclusion 16](#_Toc37502457)

[References 17](#_Toc37502458)

[List of Abbreviations 19](#_Toc37502460)

##### List of Figures

[Figure 1. 5](#_Toc411344353)

Figure 2….…………………………………………………………………………………………………………6

Figure 3…………………………………………………………………………………………………………...10

Figure 4 …………………………………………………………………………………………………………..11

Figure 5 …………………………………………………………………………………………………………..11

Figure 6 ………………………………………………………………………………………………………….12

# Introduction

Deep Neural Networks is the future, they are applicable for STEM (Science, Technology, Engineering, Management). Also, they can even be applied to other fields such as the Banking, Finance, Agriculture, and other day to day activities of a common person. Moreover, as the data plays the vital role in every aspects and they are processed, several data types such as Categorical, Numerical, Image, Voice are being trained. And the outcome is decided, this is one aspect, and another aspect is the complete automation task for the robots and to understand the environment and the agent’s role in reinforcement learning.

In this paper, I have discussed the brief History of the Deep Neural Network right from 1943 till 2019 with a detailed information of each paper containing the progress from each years and a case study of a well-known opensource Library’s commercial Application, and the comparison between the image classification algorithm.

# History of Deep Neural Networks

# The history of deep learning starts from the 1943, which started exactly before 78 Years before now, where the very first paper was published. Here in our paper, we will discuss about the detailed history until 2021, which brought us the extended versions (which we are currently using ) from these researchers and their contributions for the development of the Artificial Intelligence (Deep Neural Networks)

# First Artificial Neural Network in 1943, The Paper titled “*A Logical Calculus of the Ideas Immanent in Nervous Activity*” which talks about the Mathematical Model of biological neurons, though they had a very limited resources and the capability and thus there was no learning mechanism. But Still it is the root and the base layer which was strongly laid for the Future Artificial Neural Networks and Artificial Intelligence.

**Perception in 1957,** ThePapertitled “*The Perceptron: A perceiving and Recognition Automaton*”[2] , Authored by Frank Rosenblatt, where the very first perceptron was created, and it is the new and modified versions of the McCulloch-Pitts’s Neuron from First Paper. Perceptron had the capabilities to do the Binary Classification on its own, Author has also discussed with several input options such as Visual Patterns, Tone patterns, where with visual patterns input known as the PhotoPerceptron**.** Input with tone patterns is known as PhonoPerceptron, and he discusses the possibility of the Electro or RadioPerceptrons with corresponding sensory devices. This gave the strong foundation for the Perceptron and Neural Networks for the future.

Diagram

Description automatically generated

Figure 1, First Perceptron in 1957

The author describes the three different system S-system, A-system, R-System. And talk about the Input with tone patterns is known as PhonoPerceptron, and he discusses the possibility of the Electro or RadioPerceptrons with corresponding sensory devices. This gave the strong foundation for the Perceptron and Neural Networks for the future.

**First Backpropagation in 1960,** The Papertitled “*Gradient Theory of Optimal Flight Path*” Authored by Henry. J. Kelly- His work contributed to the first ever backpropagation model [3] author presents the gradients or "method of steepest descent" for the construction of minimising the sequences of flight path by means of step wise process of descent along with the local gradient directions, which is described as a computational scheme. And his model is a context to the control theory. Which is the base foundation for the future refinement for the ANN in upcoming years.

Diagram

Description automatically generated

*Figure 2, First Perceptron of Frank Rosenblatt in 1957 at Cornell Aeronautical laboratory, INC, New York.*

# Backpropagation’s Chain Rule in 1962, The Paper titled “*The numerical solution of variational problems*”[4] Authored by Stuart Dreyfus-His work contributed to the further development of the backpropagation model that uses the simple chain rule-instead of the dynamic programming, which was used by the earlier backpropagation models ,though this is the small step in improvement, but which holds the strengths for the future.

## Group Method of Data Handling in 1968, [5] Alexy Grigoryevich Ivakhenko[6], is sometime referred as the father of deep learning, along with valentin Grigor’evich Lapa, created the Hierarchical Representation for the Neural Networks which uses the Polynomial Activation Function which are trained using the GMHD-Group Method of Data Handling, which is the family of inductive algorithm for computer-based modelling of the multiparametric datasets that features the fully automatic structural and the parametric optimization models, GMHD is widely used as the Data Mining, KDD(Knowledge Discovery), Prediction, Complex Systems, Modelling , Optimization and the Pattern Recognition , which are characterized by inductive procedure which performs the sorting-out of gradually complicated Polynomial models and selecting the best solutions by the means of the external criterion. GMHD is considered as the First ever the Multilevel Perceptron. This time is where we can refer as the birth of the Multilayer Neural Networks. Which gives the Future generation with the enhanced strong base for future. Moreover, to find the best solution of GMDH algorithms, which considers component subsets of the base function called partial models. And the Coefficients of these models are estimated by the [least squares](https://en.wikipedia.org/wiki/Least_squares) method.

**New era of Perceptron in 1969,** TheBook titled “*Perceptron’s: an introduction to computational geometry* ”Authors talks about the Rosenblatt’s perceptron’s drawbacks as it unable to solve the complicated functions like XOR (Exclusive OR) logical Gates, for which perceptron must be placed in the Multiple Hidden layers which overcomes the perceptron problem.[7,8] More over this has created Which had made a setback to AI winter : a period of reduced funding and interest in [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) research. Refer figure-1 for the single layer perceptron and how can be over came is explained by using the Hidden layers [8].

**Backpropagation in Computer Code in 1970**,Seppo Linnainmaa published overall method used for automatic differentiation for backpropagation then further implements backpropagation in computer program.

**Deeper Neural Networks in 1971,** Alexy Grigoryevich Ivakhenko [6] he continuous his work upon GMHD Group Method of Data Handling,and to furthermore leverage he introduces an 8 layered deeper level of the Neural Network using the GMHD

**Neocognitron in 1979,** A hierarchical Multi-layered neural network presented by [Kunihiko Fukushima](https://en.wikipedia.org/wiki/Kunihiko_Fukushima) [9], for the hand written character recognition and the pattern recognition and which is the starting point of inspiration for the Convoluted Neural Networks(CNN),scale-invariant feature transform (SIFT), histogram of oriented gradients (HOG)  where, Neocognitron was inspired by the model proposed by the Hubel and Wiesel in 1959 (who found the visual primary cortex’s two types of cells such as the simple cell and complex cell, and he proposed a cascading model with which the pattern recognition takes place.

**Hopfield Network in 1982,** [10]John Hopfield, Created the Hopfield Network, which is an RNN (Recurrent Neural Networks and it is a type of spin glass system p, which serves as a content addressable memory system , also provides a model for the human understandable memory. Thus, lays a strong contribution for the future.

**Backpropagation in ANN in 1982,** [11]Paul Werbos **,** based on his thesis, Author proposes the backpropagation for the propagation errors when training the Neural Networks. Author’s result of the PhD thesis which will be the practical adoption of the backpropagation of the neural network in future.

**Boltzmann Machine in 1985,** [12][David H.Ackley](https://www.sciencedirect.com/science/article/abs/pii/S0364021385800124" \l "!), [Geoffrey E. Hinton](https://www.sciencedirect.com/science/article/abs/pii/S0364021385800124#!), [Terrence J. Sejnowski,](https://www.sciencedirect.com/science/article/abs/pii/S0364021385800124#!) created the Boltzmann Machine- a network of symmetrically connected, Neuron like units which can make a stochastic decisions regarding to be on or not to be off. It is also called as the Stochastic Hopfield Network with hidden units or Sherrington-Kirkpatrick model with external field or Stochastic Ising-Lenz-Little Model. Moreover, it is a simple algorithm that can discover important features in a dataset which composed of binary vector, where it has one input layer with hidden layer without output layer.

**NETtalk in 1986,** [13]Terrence Sejnowski, Charles Rosenberg creates an Artificial neural network , which has the capability to construct a simplified models that process the complexity of learning the human level cognitive tasks, and their implementation as a connectionist model that also can learn to compute comparable task, in detail: NETtalk was created to explore the mechanism of learning to correctly pronounce English text. To learn this requires a complex mechanism that involve in many parts akin to human brain-A program to learn and pronounce the English written texts by showing the text as an input and matching the phonetic transcriptions for the comparison. Moreover, NetTalk does not specifically model the image processing and letter recognition of the visual cortex , but it computes and assumes the letter that have been pre classified and recognized, and the letter sequence consist of the words are then fed into the neural networks in training time and performance testing. Then the NETtalk learns to find the correct form of the associated pronunciations with respect to the given sequence of letters based on the context in which letters appear: NETtalk learns to get the exact letter around the pronounced phoneme that gives the intended mapping.

**Learning Backpropagation by errors in 1986,** A paper titled “*Learning Representations by Back-Propagating errors”* [14], talks about the successful implementation of the backpropagation in neural network, which gave the opportunity for the training of the complex deep neural networks easily and effectively , which was one of the toughest in the previous time of the research.

**Restricted Boltzmann Machine (RBM) in 1986,** Initially invented under the name Harmonium by Paul Smolensky, it’s a variant of Boltzmann Machine[15], with the restrictions that their neurons must be a Bipartite graph- which a pair of nodes ( Visible and hidden units) will have a symmetric connection between them but there is no connection between the nodes within a group, In the Boltzmann Machines which have the connection between a group and in the Unrestricted Boltzmann Machines have the connection between the hidden units, This un-restrictedness allows the training algorithm to perform more effectively than the Boltzmann machines and in particular to the gradient contrastive divergence

**Optimal Brain Damage in 1989,** A paper titled*“Optimal Brain Damage”* by Yann Le Cun, John S. Denker and Sara A. Solla [16], Authors have introduced a new technique called Optimal Brain Damage (OBD) for reducing the size of the large networks by deleting their weights. OBD which can use both Automatic network minimization procedure and as well as an interactive tool to suggest a better architectures.

**Approximators Theorem in 1989,** A paper titled*“Approximation by superpositions of a sigmoidal function”* by [17],George Cybenko, demonstrated the finite linear combinations of t compositions of a fixed , univariate and a set of affine functions, which can perform uniformly approximate any continuous functions of a n real variables with support in the hypercube. Also, he has discussed the approximation properties of the other possible nonlinearities which are implemented in the artificial neural networks .Moreover Author proves the feed forward neural network with single hidden unit containing the finite number of neurons, which can approximate any continuous functions .Which add the strong foundation to the Deep Learning.

**Vanishing Gradient Problem in 1991,** A paper titled **“”** by Sepp Hochreiter[18], identifies the problem of the vanishing Gradients which make the deep neural network to get extremely slow and impossible.

**Optical Character Recognition in 1993,** John S. Denker proposed the Optical Character Recognition

**Era of LSTM in 1997,** A paper titled*“Long Short-Term Memory ” [19]* bySepp Hochreiter and Jurgen Schmidhuber a new type of RNN (Recurrent Neural Network) Architecture was proposed. The idea was to store information over the extended period intervals via Recurrent Backpropagation which takes a very long time and moreover, mostly due to the insufficient decaying error backflow. Authors briefly analysed the Hochreiter's 1991problem of the vanishing gradients problem , then address it in a novel and efficient way, knows as the LSTM. This Truncates the Vanishing Gradient problem and it will not harm it , LSTM can learn to bridge minimal time lags more than 1000 discrete time steps by enforcing the constant error flow through “*Constant error carousels*” within special units.

**Deep Belief Network in 2006,** A paper titled*“A fast learning Algorithm for deep belief nets“* authored by Geoffrey Hinton , Simon Osindero and Yee-Whye Author shows how to use the complementary priors to eliminate the explaining away effects that make difficulty for the densely connected belief networks which has many hidden layers. From which authors derive a fast and greedy algorithm that can learn deep, directional belief networks.[ 20]

**GPU Revolution in 2008,** Andrew Ng at Stanford starts the advocating for the use of the GPU for training the Deep Neural Networks to speed up training process and reduce time it takes to train

**ImageNet in 2009,** A paper Titled *“ImageNet: A Large-Scale Hierarchical Image Database”* [21]A new database for the Image has been created and known as the ImageNet A large scale ontology of images built upon the backbone of the WordNet structure, at that time ImageNet aims to populate the majority of the 80,000 synsets of wordnet. And its database has the 12 subtrees with 7247 synsets and 3.2 Million Images in total in 2009.

**Combat for Vanishing Gradient in 2011,** A paper Titled *“Deep Sparse Rectifier Neural Networks”*[22], Authored by Xavier Glorot, Antoine Bordes, Yoshua Bengio, Author presents the rectified Neurons and states they are more performing than the hyperbolic tangent networks despite of the hard non-linearity and non-differentiability at zero, creating sparse representations with true zeros, which seem remarkably suitable for naturally sparse data.

**Start of AlexNet Deep Learning in 2012, Alex Krizhevsky, [23]** AlexNet is the GPU based CNN Model which won the ImageNet’s Image classification, and which acquired an accuracy of 84% which had a great jump over 75 % Accuracy over the earlier models developed

**GAN in 2014,** A paper titled*“Generative Adversarial Nets”*by Ian Goodfellow [24], Author proposes a new framework for estimating generative models via an Adversarial process, where they simultaneously train the two models (G, D) G- a generative model that captures the data distribution and d- discriminative model that estimates the probability that a sample came from the training data rather than G. Both the G and D are defined by Multi-layered Perceptron and the entire system is trained by the backpropagation there is no need for the Markov chains or unrolled approximate interference networks during either training or the generation of samples.

**AlphaGo in 2016,** A deep mind’s Deep reinforcement Learning models won against the Human Champion of a complex game called “Go” which is more complex than the chess , this gives the entire words the essence of the Artificial Intelligence and the Future of the Deep Learning **.**

**Turning Award in 2019,** Geoffrey Hilton, Yoshua Bengio, Yann LeCun won the Turning Award for their contribution right from 1970 on Neural Network and Deep learning. after the entire research has been moved away because of AI winter right from 1970.

Chapter 3: Case Study on Tesla self-Driving Application using Pytorch

**Pytorch History:**

Pytorch is a Facebook‘s indigenous creation, it’s a free opensource library for the machine learning, computer vision tasks, Natural Language Processing it’s based on the torch library and was developed by Facebook AI research lab team (FAIR) in 2016, first official release was alpha 1 in September 2016. Python Interface is the most widely used for the Pytorch Development, but it has also the C++ Interface for development. With latest version of Pytorch 1.8.0 released on 4th March2021.[24]

**Famous Applications Built using the Pytorch as on date:**

The Most famous Deep Learning Application built under the top of Pytorch

1. Tesla Autopilot
2. Uber’s Pyro
3. Hugging Face’s Transformers
4. Pytorch Lightning and Catalyst.
5. Lyft Self driving cars

**Case Study Details :**

As part of the case study, we shall discuss in detail about the deep learning-based Autopilot software application which is built on the top of the Pytorch.

**Tesla Autopilot :**

Tesla is an Automobile Manufacture, who introduced the electric car to the world and has the fully capability of self-driving application for the car, which is based on the Deep Neural Networks built in the Python, using Pytorch library.

**Data Collection:**

Training is the one of the major factors responsible for the performance of the deep neural networks and other deciding factors such as the Network Architectures and the optimizing algorithms. In a brief How heavy is the training data its directly proportional to the performance accuracy in results. Tesla is the one of the best autonomous vehicles at present in the world.

A picture containing text, night sky

Description automatically generated

Figure 3, Represents the tesla Fleet of daily Travelling data.

Where the Half a Million vehicles on road has been equipped with the tesla driving hardware , and tesla’s driving fleet travels daily around 15 Million Miles a day and it is expected to be 5.4 Billion miles a year in terms of the data gathering.[25]

**Background Training:**

Tesla Uses the Cutting-edge research on the Deep Neural Network for the control of the entire operation of the self-driving , problems such as the perception to the control entire has been taken care by the DNN , camera performs the Semantic Segmentation, Object Detection and Monocular Depth Estimator. Tesla’s Bird eye-view networks take the video from all the cameras to output the road layout [26], on a Top-Down view of the static objects , infrastructures, road layout and other 3d objects. A fully trained Autopilot Neural Network takes around 70,000 GPU Hours to train, and it involved 48 Networks and they output 1000 different tensor: Predictions , at each timestep.

**Neural Network Architecture:**

Tesla Neural Network uses the HydraNet as the Background Architecture and it is a shared backbone. HydraNets are the are the wide Networks which contains the distinct components that are specialized to compute the several features for the visual similar classes, perhaps they retain the efficiency by only dynamic selection of the small number of components to evaluate anyone input image. Figure-4 shows the backbone overview.

Akin to the transfer learning where you will be having a common block and the train specified blocks for specified task. HydraNets have the backbones that trained on all the objects and trained on the specific task, which improves the efficiency speed and the training speed.

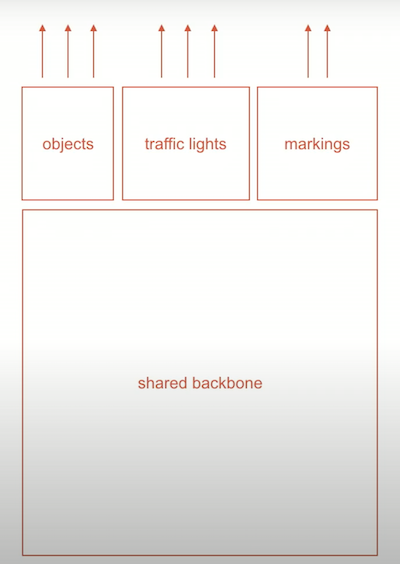


Figure 4, Shared Back Bone classification (HydraNet) Architecture in Tesla

If we look Inside the Tesla’s Neural Network , we can see that they are trained using the Pytorch .,Input image have the dimensions about 1280, 960,3 is passed through the specific neural network and the backbone is modified of the ResNet 50 and the specific Modification is used of the Dilated Convolutions, Heads are based on the sematic segmentation FPN/Deep Lab/UNet architectures, but still, this seems to be a “end task as the conversation between the 2D pixels and the 3D pixels are completely prone to pile up errors.

Tesla Has 8 cameras, with 16 time-steps , batch size of 32 (8\*16\*32) = 4096 images in a single forward pass,

which requires a combination of both data parallel and the model parallel for the training purpose. On overall we would be required to hold the 4096 images of memory with their activations. For example : As every forward pass there are 4096 images are processed, even the top performing computers like Mac Book Pro cannot even handle this type of computation , why not even 2 GPU cannot handle this. To overcome such problem HydraNets are made to use.[27],[28]

**HydraNet in Tesla:** In figure For the Road Layout Estimation, there are eight different cameras, with each has a separate HydraNet, for eight different task instances, which all produces different kinds of the intermediate predictions at the initial round and moreover the HydraNet go into the second round of processing at this stage every neurons are combined into a middle neural network. which is the potential Recurrent and thereby we have the several output from top-down view.

**Neural Network Regression :** Tesla uses the Regression for the calculation of the distance in an effective manner, for example Depth estimation is done using Stereo cameras, having 2 cameras helps to track the distance better. Tesla uses the neural networks with the regression on the depth. As it does not use the LIDAR to use this to track the distance between the objects. One advantage Is that the camera has not the same lens quality: on the right, further distances appear much closer.

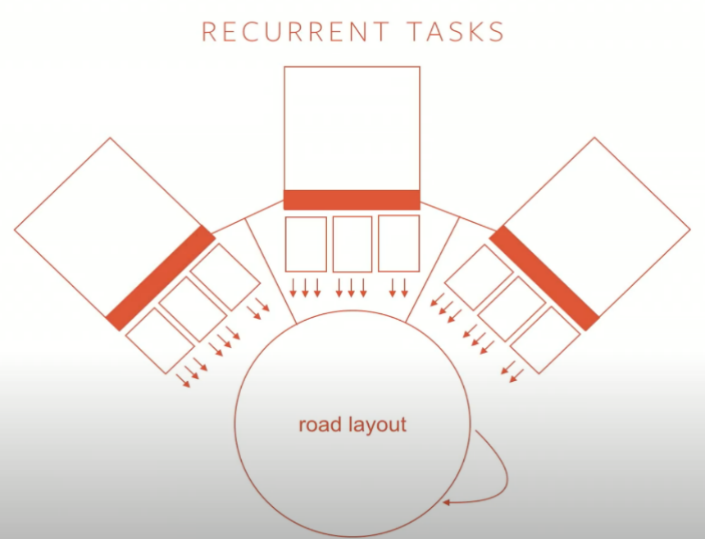
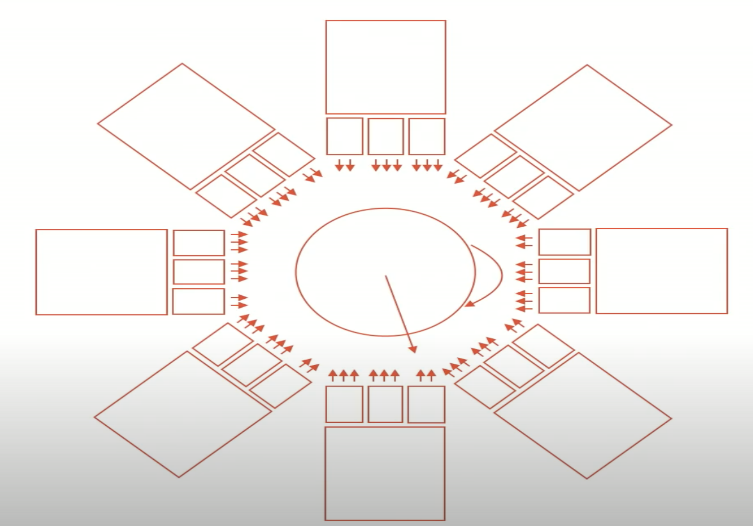


Figure 5, Recurrent Task for road layout with less camera.

Figure 6, Recurrent Task for 8 Cameras.

**Overview of HydraNet Architecture which is used in Tesla.**

Below is the Architecture Diagram of the HydraNet, the word Hydra implies a system with several heads.

Diagram

Description automatically generated

Figure 6, HydraNet Architecture. [29]

The HydraNet template, shown in Figure 6, has four major components.

**Branches**: specialized for computation of the visually similar classes. We view computing features relevant to a subset of the network inputs as a subtask of the larger classification task.

**Stem**: A computes features, which is used by all branches and in deciding which subtasks to perform for an input.

**Gating Mechanism :**which decides what branches to execute at inference by using features from the stem.

**Combiner:** which aggregates features from multiple branches to make final predictions. Realizing the HydraNet template requires partitioning the classes into visually similar groups that the branches specialize for, an accurate and cost-effective gating mechanism for choosing branches to execute given an input, and 8081 a method for training all the components. The following sections describe how we address these key questions.

**A comparison and cost benefit analysis between the rivals of Tesla vs Waymo Architecture for the Image Classifications**

**Waymo :**

In 2019, A paper titled *“ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst”* Authored by Mayank Bansal, Alex Krizhevsky, Abhijit Ogale under the Google Brain and Waymo Research presented the new algorithm called ChauffeurNet to train a policy for autonomous driving via imitation learning that is robust enough to drive a real vehicle. [30]

Waymo car are very different unlike the tesla where they use the LIDAR for the distance measurement, its physical inputs are Lidar, Camera, Radar, Sensor data and it uses the TensorFlow

**Building Blocks: [32]**

**Perceptron:** to find the road paths , traffic light, obstacles-which are leveraged network architecture search (NAS) to find out best architecture models**.**

**Behaviour Predictions:** Google map using the train agents through the ChauffeurNets which is used to estimate the trajectories in a simulated environment

**Planning:** Generate the Trajectories through the ChauffeurNets based on the feasibility on the road conditions : collision avoidance,

**Control optimizer:** Throttle and the Steering

For the computer vision in the Waymo [31] they use the Network Architecture search and the Auto Machine Learning .

**Diagram

Description automatically generated**

Figure 7, Network Architecture Search used in Waymo.

Diagram

Description automatically generatedFigure 8, Auto Ml in Waymo.

It is a Building blocks of the Neural Networks which is a repetition of the larger neural networks akin to the ResNets, this concept has been taken and adapted to a extended version called the Auto ML , Where the estimation of the Algorithm takes place for the best fitted one and with the top performance.

**Unlike AutoML in Waymo , Tesla Uses the HydraNets:**

**HydraNets Cost Calculation Analysis :**

Hierarchical classification uses the gating function to classify among the subtasks and thereby then branches classifying among the classes within a subtask.

Text, letter

Description automatically generatedFor the Dynamic execution to be accurate both the grating functions and the chosen branch need to be performed with respective classification task. Moreover, the accuracy of the grating function is dependent on the capacity ( computational cost and parameters) of the gating component.

Figure 9, cost Analysis of the HydraNet.

# The Future

Future of the Artificial Intelligence is focussed on the Deep Neural Networks and its vast applications such as the Self -Driving, Image classification, Natural Language processing , Object Detection , Machinery equipment’s, robotics , solving problem which cannot be solved by human capability . And the Defence purposes. DNN evolve into a higher-level algorithm that can solve most problems using Neural Networks.

# Conclusion

This Paper has the in a base reference article for the future students and the researchers who are interested in studying the Deep Neural Network, I have discussed the history of the DNN right from the scratch with detailed references and a trending case study of the self-driving car Architecture of both rivals (Tesla and Waymo) in which I have discussed the potential difference in the architecture with concise difference between them.

##### Reference

[1] **1943:** **https://link.springer.com/article/10.1007/BF02478259**

[2] **1957:** https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf

[3] **1960:** https://www.gwern.net/docs/statistics/decision/1960-kelley.pdf

[4] **1962:** https://www.sciencedirect.com/science/article/pii/0022247X62900045

[5] **1968:** https://en.wikipedia.org/wiki/Group\_method\_of\_data\_handling

[6] **Alexey\_Ivakhnenko:**  https://en.wikipedia.org/wiki/Alexey\_Ivakhnenko

[7] **Perceptron Book:** https://en.wikipedia.org/wiki/Perceptrons\_(book)

[8] **Solving XOR problem:** http://www.niser.ac.in/~smishra/teach/cs460/lectures/lec19/

[9] **Neocognitron:** https://link.springer.com/article/10.1007%2FBF00344251

[10]**Hopfield Network:** https: //www.sciencedirect.com/topics/computer-science/hopfield-network

[11] <https://search.proquest.com/docview/302768593>

[12] **Boltzmann Machine:** <https://www.sciencedirect.com/science/article/abs/pii/S0364021385800124>

[13] **NETtalk:** https://en.wikipedia.org/wiki/NETtalk\_(artificial\_neural\_network)

[14] **Backpropagation by Errors:** <https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf>

[15] **Restricted Boltzmann:** https://en.wikipedia.org/wiki/Restricted\_Boltzmann\_machine

[16] **Optimal Brain Damage:** <https://www.researchgate.net/publication/221618539_Optimal_Brain_Damage>

[17] **Approximator Theorem:** <https://web.njit.edu/~usman/courses/cs675_fall18/10.1.1.441.7873.pdf>

[18] **Vanishing Gradients:** <https://people.idsia.ch//~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf>

[19] **LSTM:** <https://www.bioinf.jku.at/publications/older/2604.pdf>

[20] **Deep Belief Networks:** <https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf>

[21] **ImageNet:** <https://www.image-net.org/static_files/papers/imagenet_cvpr09.pdf>

[22] **RELU:** <http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf>

[23] **AlexNet:** <https://www.image-net.org/static_files/papers/imagenet_cvpr09.pdf>

[24] **GAN:** <https://arxiv.org/abs/1406.2661>

[25] https://en.wikipedia.org/wiki/PyTorch

[26] https://towardsdatascience.com/teslas-deep-learning-at-scale-7eed85b235d3

[27] <https://www.tesla.com/en_IE/autopilotAI>

[28] <https://medium.com/data-science-bootcamp/tesla-and-pytorch-pytorch-developer-conference-highlights-part-3ed36f2c9d5e>

[29]<https://openaccess.thecvf.com/content_cvpr_2018/papers/Mullapudi_HydraNets_Specialized_Dynamic_CVPR_2018_paper.pdf>

[30] <https://arxiv.org/search/cs?searchtype=author&query=Ogale%2C+A>

[31] <https://heartbeat.fritz.ai/how-googles-self-driving-cars-work-c77e4126f6e7>

[32] <https://asiliconvalleyinsider.com/2020/03/08/waymo-chauffeurnet-versus-telsa-hydranet/>

[33] https://machinelearningknowledge.ai/brief-history-of-deep-learning/

##### List of Abbreviations

DNN- Deep Neural Networks

ANN- Artificial Neural Networks

NN- Neural Networks

LIDAR-

CNN- Convoluted Neural Networks

GPU - Graphics processing unit

GMHD-Group Method of Data Handling

LSTM- Long Short-Term Memory

SIFT- Scale-Invariant Feature Transform

OBD- Optimal Brain Damage