Deep Learning – A Review

Vamsi Madhav Kota¹, V Manoj Kumar² and C Bharatiraja³

¹ Post-Graduate Student, Department of Mechanical Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.
²Assistant Professor (Sr.G), Department of Mechanical Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.
³Associate Professor, Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.

E-mail: km2237@srmist.edu.in

Abstract. In recent years, tech giants in various parts are showing Curiosity on Artificial Intelligence by investment on the project that can be a game-changer for both corporate and researchers. A company such as Google, Baidu, and Yandex have already started their multimillion-dollar project in this pitch. This article presents the latest progress and also tries to paint a predictive picture of future research directions and developments in the domain of deep learning. Each of the said research directions and avenues is analyzed and summarized in a brief yet concise manner in this article. Initially, an outline of the three elementary models of deep learning that including multilayer perceptions and perceptrons, convolutional neural networks and recurrent neural networks. Building on the bases of foundation, further analyses of the emerging new types of convolutional neural networks and recurrent neural networks are also undertaken in this current study. This article then summarizes deep learning and its applications in the domain of artificial intelligence, counting speech processing, computer vision, and natural language processing. Finally, the purpose of deep learning is discussed. The current article also delves a little deeper into the inner workings of the neural networking architecture associated with object detection and computer vision.

1. Introduction

Artificial Intelligence is a well-known word today, it has become so popular that people from all walks of life and not just those who are directly associated with this field, can comprehend its meaning to various extents. In March 2016, artificial intelligence and the associated research were incorporated into the Chinese government’s 13th Five-Year Plan. In October 2016, the US government issued several notices which incorporated artificial intelligence research into the Research and Development Strategic Planning initiative. Corporations such as Google, Microsoft, Facebook, Baidu, Tencent, and Alibaba have also stepped up their efforts and investment associated with the domain of artificial intelligence and is currently undertaking many projects in this area of research. A variety of artificial intelligence start-ups are also emerging, and this has led to people from all walks of life understanding the vast number of applications of artificial intelligence and this understanding has already started changing the nature of human life and it is quite clear that artificial intelligence and thereby deep learning and machine learning are gradually percolating into all aspects of everyday human life. Deep learning is presently the most significant explored areas which come under the purview of artificial intelligence. Deep learning and machine learning can be utilized in many areas and therefore possess
innumerable submissions to include speech processing, computer vision, natural language processing, etc. These all began in 1943 when McCulloch and Pitts [1] anticipated a mathematical model of MP (multiple perceptron's) type neurons. In the year 1958, the initial generation of the neural network with a single-layer perceptron was developed via Rosenblatt [2]. These initial generations of neural networks could distinguish basic shapes such as triangles and squares. The discovery that a class of machines could distinguish between different shapes made scientists realize that it was possible to invent devices i.e., intelligent machines, that can actually perceive, learn, and remember data and information. But the first-generation neural networks had many limitations which started discouraging further research in this sector. This situation persisted until 1969 when Minsky published a paper on perceptron [3]. Minsky stated that a single-layer perception processor cannot solve the XOR problem. Minsky’s article also said that these first-generation neural networks were inconsistent with the definition of truly smart machines. In 1986, Hinton et al. [4] proposed a subsequent neural network generation, in which the fixed original single-layer perceptron processor was replaced with a processor possessing several hidden layers and where the activation of these hidden layers was accomplished with the help of the sigmoid function, using an Error back-propagation training algorithm which assisted in training the model to give it the ability to effectually resolve nonlinear classification problems.

In the next ten years, various shallow machine learning models were proposed successively, including the one proposed by Cortes and Vapnik [5] in the year 1995. The two of them successfully designed a support vector machine. In 2006, Hinton et al. [6] analyzed the human mind and came up with graphical models of the brain and tried to apply their findings in order to design a new type of neural network. The model proposed by them introduced an encoder (aka an authenticator) to reduce the dimensionality of the data. They also came up with the concept of pre-training and they proposed to use this concept pre-training in order to quickly and efficiently train the deep learning network to suppress and thus overcome the vanishing gradient problem. Choromanska et al. [7] and Dauphin et al. [10] discussed that the local minima problem was not big problem as the evolution of deep learning. Bengio [7] and others proved that the pre-exercise method was also valid to unsupervised autoencoders in addition to supervised learning. Poultney et al. utilized energy-based models to efficiently make machines learning sparse means. These documents put the foundation for deep learning and also ushered in an era of rapid development in the domains of machine learning and artificial intelligence.

Deep learning is actually a subclass of machine learning. And machine learning takes skilled two waves in history and evolved since shallow learning to deep machine learning. There are a few noteworthy differences that differentiate the deep learning model from its shallow learning model counterparts. Shallow learning models don’t utilize disseminated representations (distributions), besides features, need to be artificially extracted. Moreover, the prototypical structure itself only classifies or forecasts features of the measurement or the parameter and the quality of the artificially extracted features largely determines the efficacy of this entire system. Feature extraction necessitates professional level field knowledge, and feature mining and associated practices take a lot of time. Deep learning is a kind of representational learning and therefore the algorithms need to have the ability to allow the machine to learn the higher-level abstract representations of data, and provide it the avenue to be able to automatically extract feature associated data from the raw data superset. There is also the matter of the concealed layer in deep learning that requires elaboration. This hidden layer is equivalent to the participated features associated with linear grouping, while the weight among the hidden layer and the contributed layer is comparable to the contributed features associated with weights in linear assortments. In accumulation with the capabilities of deep learning, models vary with depth in that the capability of any deep learning model increases exponentially with increasing depth in figure 1.
2. Basic Network Structure

2.1. Multilayer Perceptron
A multilayer perceptron (MLP) [11] is addressed as a forward propagation network and a deep feed-forward network. It’s mostly basic deep learning network construction MLPs consist of numerous layers, and apiece layer contains numerous neurons that require activation in order to accomplish machine learning. A multi-layer perceptron with radial basis functions is called radial basis networks. The forward propagation function of the multilayer perceptron is publicized in Figure 2.

![Multi-Layer Perceptron](image)

**Figure 2.** Multi-Layer Perceptron during forward propagation.
The fundamental governing equation for input $x_j$ to $x_z$ as shown in equation 1:

$$W_i \cdot X = w_{ij}x_j + w_{ik}x_k + \cdots + w_{iz}x_z = \sum_{j=1}^{z} w_j x_j$$  \hspace{1cm} (1)$$

The forward propagation formulae associated with MLPs are shown with the help of equations (2) and (3):

$$z_{j}^{l+1} = \sum_{j=1}^{z} w_j x_j + b_j$$  \hspace{1cm} \text{Where } b \text{ is bias} \hspace{1cm} (2)$$

$$y_j^z = f(z_j^{l+1})$$  \hspace{1cm} (3)$$

Among them, $y_1^1$ is the output associated with the first neuron in the first layer, whereas $z_i^{l+1}$ is the output associated with the neuron belonging to the $(l+1)^{th}$ layer. This function gives the value of the $i$th neuron before being activated by the activation function. The weight amongst the $j^{th}$ neuron and the $i^{th}$ neuron in the $(l+1)^{th}$ layer is $b_j^l$ and it serves as a bias; $f(\cdot)$ is a non-linear initiation function, and common radial basis functions, ReLU, PReLU, Tanh, Sigmoid, etc [11] are utilized to generate suitable outputs.

If the mean square error is used, the loss function becomes

$$J = \frac{1}{2} \sum_j (y_j^z - y_j)$$  \hspace{1cm} (4)$$

2.2. Convolutional Neural Networks

Convolutional Neural Network (CNN) [12] stand suitable for dispensation spatial data. This type of neural network architecture is commonly used in the field of computer vision. The one-dimensional convolutional neural network is acknowledged as time-delayed neural networks. This variant of CNN can be employed to progression in 1-D data. The idea associated with the design of CNNs was Inspired by neuroscience, especially the segment of neuroscience associated with vision and the neuronal architecture affiliated with processing visual stimuli. This type of neural network is mainly composed of convolution layers and pooling layers. The convolutional layers help with building up a spatial continuity, also possess the ability to extract local features of the image. The pooling layer can use the maximum pooling (mean-poling) or average pooling (mean-poling) functions to enable the pooling layer to decrease the dimensions of the intermediate hidden layer(s) and decrease the calculation time of the subsequent layers. Moreover, this also helps generate rotation invariance. Convolution and pooling operations are described in Figure 3.
Figure 3. An Illustration of Convolution progression and pooling progression.

As described above the figure has a $3 \times 3$ convolution kernel and a $2 \times 2$ polling kernel is used in this type of neural network architecture. The structure of the proposed Lenet-5 network architecture is shown in Figure 4.

Figure 4. LeNet-5 Architectural Outline.

2.3. Recurrent Neural Networks
Recurrent Neural Networks (RNN) [13] stand suitable for dispensing time series data and are extensively used in speech processing and natural language processing. Human speech and language are inherently sequential. The algorithm associated with the construction of RNNs and its expanded
view is shown in Figure 5. Pascanu et al.[14] and Sutskever et al.[15] Amended the training technique of RNN.

![Recurrent Neural Networks](image)

**Figure 5.** Recurrent Neural Networks.

3. **Improvements of Network Structure**

3.1. **Improved Convolutional Neural Network**

ImageNet [8], a periodical Contest (Imagine Nettle measure visual cognition competition) has significantly sponsored the development of convolutional neural networks. The afresh developed convolutional neural networks starting from AlexNet [16], which was developed in 2012, QFNet, which came about in 2013, and the VGGNet [17] which was developed in 2014, these were followed by GoogLeNet [18,19], and ResNet [20] in 2016; the design and development of these newer variants of CNNs vastly improved the performance of ImageNet [8]. It was noted that the amount of network layers is constantly growing and the model’s abilities are also growing. The first exhibition of AlexNet proved and ability the power of deep learning. QFNet, on the other hand, exhibited the visual sympathetic capabilities of convolutional neural networks. The network, VGGnet demonstrated that the system empathetic can altogether improve deep learning capabilities. Initial in the first time itself, GoogLeNet (a.k.a. Inception V1) broke the limitations associated with the convolution layer-pooling layer stacking model. ResNet magnificently trained a neural network with a depth of 152 layers. The mainstream methods of CNN aimed at object detection are R-CNN, Fast R-CNN, Faster R-CNN and finally, Mask R-CNN were some of its breakthroughs. The improvement process is actually causing the replacement of deep machine learning models with the newer deep learning models. This foundation of this phenomenon lies in end-to-end training, and the speed of neural networks is increasing at a rapid rate. In addition, the innovative new network-in-network structuring idea allows for nesting networks inside a pre-existing neural network. The spatial alteration network model expressions that the improvements achieved by a certain neural network model over its peers and predecessors don’t essentially pose a requirement to modify the network architecture. This kind of thing is possible because of the realization that network structures could also be transformed by inputting data.
3.2. Improved Recurrent Neural Network:
There most significant problem associated with RNNs [13] is the problem of gradient disappearance or explosion. RNNs are incapable to take advantage of previous information beyond duration of time, i.e., this type of neural network has a limited potential to remember things that it learned in the past. RNNs are generally associated with complex training algorithms and a large number of calculations needing to be performed by the models. This type of network does not need to repeatedly calculate the gradient to achieve high accuracy outputs. Training provides new ideas and dimensions to these models. Recurrent neural networks lack the inference function and can’t comprehensive the tasks that require cognitive. In order to solve this issue, the neural network needs to pass the Turing test. The addition of a memory unit also assists to resolve this problem.

4. Potential Applications of Deep Learning

4.1. Computer Vision
Deep learning is typically utilized in different errands related with computer vision, including communication, sign detection and classification, facial recognition, face detection, image classification, multi-scale transform fusion image processing, object detection, image description, semantic segmentation, instantaneous multi-person pose estimation, pedestrian detection, scene recognition, object tracking, end-to-end video classification, human motion recognition and so on.

In accumulation, there are some fascinating applications, for example, automatically differentiates between color and black and white photos, turning graffiti into art, art style transfer, remove the mosaic in pictures, etc. Oxford University and Google DeepMind [23] correspondingly projected an arrangement acknowledged as LipNet to read lip language accurately. That has an accuracy rate reached 93%, far beyond the average level of 52% which is the level at which humans can read lips.

4.2. Speech Processing
Deep learning made its first breakthrough in the domain of speech processing. And the size of the dataset didn’t affect the ability of the model to process the speech associated information. Here are binary two major tasks in the field of speech processing: speech recognition and speech synthesis. Deep learning is commonly used in speech recognition today. Google familiarized an end-to-end speech recognition system a few years ago. This was emulated by other corporations such as Baidu, which launched its own speech recognition system Deep Speech [21] in 2016. Microsoft achieved a 5.9% accuracy in speech recognition while using daily conversational data. This ensured that deep learning reached a level where it could be applied for interacting with and assisting humans for the first time in history. Major companies including Google, Apple, HKUST Xunfei, etc., also use deep learning to implement speech synthesis [22]. Google came up with DeepMind [24] which is a parallel WaveNet model for speech synthesis. Later, Baidu introduced a product that could perform real-time speech synthesis (DeepVoice3).

5. Current Problems and Additional Research Directions of Deep Learning
Although deep learning has made advances in many areas there are still some problems preventing this field of research from progressing further and making additional breakthroughs. Further research is required to address the problems encountered in this field of research. This section of the current article exists for the sole purpose of identifying, understanding and exploring the various issues associated with deep learning and its models and tries and come up with innovative solutions in order to overcome the problems listed. The problems plaguing this field are divided into the following categories, namely, training problems, landing problems, functional problems, and domain related questions.
Deep learning [7] is now being employed for image perception and image recognition tasks. While deep learning models are able to accomplish most of these tasks adequately, they do have issues when it comes to image understanding such as understanding and characterizing the visual relationship between different pictures. There are not many achievements in the domain of deep learning when it comes to content question answering and visual attention point prediction. Deep learning models need to understand that you need to detect crucial objects first, and then foresee the relationships between objects. But unfortunately, current deep learning models aren’t quite capable in this endeavor.

6. Conclusion
Deep learning is an emerging but exciting direction of research in the domain of machine learning. Deep learning basically involves a cascade of layers that have the ability to process large amounts of data in a nonlinear manner and learn several levels of data depictions. For a very long time, researchers and corporations working on the machine-learning sector have tried to identify all possible simple and complex patterns and data representations from raw data. This way of machine learning is termed as representation learning. Representation learning differs from conventional machine-learning and data mining techniques in that it allows deep learning models to be bright to generate extremely sophisticated data depictions from immense volumes of raw data. Hence, deep learning has revolutionized this area of research as it has provided many scientists solutions for solving a multitude of real-world problems and it has made researchers and the common public aware of the huge set of current and potential applications of deep learning in all walks of life. The current article begins with an attempt to present a brief history of artificial neural networks and deep learning and proceeds to describe and survey the conversant algorithms and techniques currently extant in the field of deep learning and their applications. This section is followed by a concise description of techniques of deep learning are discussed in addition to providing descriptions of how the traditional neural networks and several supervised deep learning algorithms, including recurrent and convolutional neural networks function. The current article also explains how the concept of the key algorithms and frameworks associated with improvised recurrent and convolutional neural networks and how this can be applied to real-world problems are also briefly discussed. This article then proceeds to discuss the problems and challenges that deep learning needs to overcome in the future and also suggest some possible solutions to champion these challenges.

7. References
[1] McCulloch W S and Pitts W 1943 A logical calculus of the ideas immanent in nervous activity Bull. Math. Biophys. 5 115–33
[2] Rosenblatt F 1958 The perceptron: A probabilistic model for information storage and organization in the brain Psychol. Rev.
[3] Minsky M and Papert S 1969 Perceptron: an introduction to computational geometry MIT Press. Cambridge, Expand. Ed. 19 88
[4] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors Nature 323 533–6
[5] Cortes C and Vapnik V 1995 Support-Vector Networks Mach. Learn
[6] Hinton G E and Salakhutdinov R R 2006 Reducing the dimensionality of data with neural networks Science 313 504–7
[7] Lecun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44
[8] Krizhevsky A, Sutskever I and Geoffrey E H 2012 Imagenet Adv. Neural Inf. Process. Syst. 25
[9] Choromanska A, Henaff M, Mathieu M, Arous G Ben and LeCun Y 2015 The loss surfaces of multilayer networks J. Mach. Learn. Res. 38 192–204
[10] Dauphin Y N, Pascanu R, Gulcehre C, Cho K, Ganguli S and Bengio Y 2014 Identifying and attacking the saddle point problem in high-dimensional non-convex optimization Adv. Neural Inf. Process. Syst. 4 2933–41
[11] Lek S and Park Y S 2008 Multilayer Perceptron Encyclopedia of Ecology, Five-Volume Set
[12] Rocco I, Arandjelovic R and Sivic J 2017 Convolutional neural network architecture for geometric matching Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition

[13] Caterini A L and Chang D E 2018 Recurrent neural networks Springer Briefs in Computer Science

[14] Pascanu R, Mikolov T and Bengio Y 2013 On the difficulty of training recurrent neural networks Proc. 30th International Conference on Machine Learning

[15] Martens J and Sutskever I 2012 Training deep and recurrent networks with hessian-free optimization Lect. Notes Comput. Sci.

[16] Krizhevsky A, Sutskever I and Hinton G E 2012 AlexNet Adv. Neural Inf. Process. Syst.

[17] Chung C, Patel S, Lee R, Fu L, Reilley S, Ho T, Lionetti J, George M D and Taylor P 2018 VGGNET Am. J. Heal. Pharm.

[18] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2014 GoogLeNet Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.

[19] Zeng G, He Y, Yu Z, Yang X, Yang R and Zhang L 2016 InceptionNet/GoogLeNet - Going Deeper with Convolutions Cypr

[20] He K, Zhang X, Ren S and Sun J 2016 ResNet Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.

[21] Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Casper J, Catanzaro B, Cheng Q, Chen G and Chen J 2016 Deep speech 2: End-to-end speech recognition in english and mandarin Proc. InInternational conference on machine learning 173-182

[22] Arik SO, Chrzanowski M, Coates A, Diamos G, Gibiansky A, Kang Y, Li X, Miller J, Ng A, Raiman J and Sengupta S 2017 Deep voice: Real-time neural text-to-speech arXiv preprint arXiv

[23] Powles J and Hodson H Google DeepMind and healthcare in an age of algorithms Health and technology 7 351-67

[24] Oord A, Li Y, Babuschkin I, Simonyan K, Vinyals O, Kavukcuoglu K, Driessche G, Lockhart E, Cobo L, Stimberg F and Casagrande N Parallel wavenet: Fast high-fidelity speech synthesis Proc. InInternational conference on machine learning 3918-3926

[25] Zhang H, Wang A, Li D and Xu W 2018 Deepvoice: A voiceprint-based mobile health framework for parkinson's disease identification. Proc. In2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) 214-217