IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, June 2022
Impact Factor: 6.252

A Survey on Deep Learning Concepts and Techniques

Vysakh V Mohan[1], Pradeep Prakash[2], Resmi K R[3]
Student, Computer Science, Santhigiri College of Computer Sciences, Thodupuzha[1, 2]
Assistant Professor, Computer Science, Santhigiri College of Computer Sciences, Thodupuzha[3]

Abstract:*For the last few decades, the deep learning (DL) computing paradigm has been held to be the benchmark in the machine learning (ML) community. As well, it has gradually become the most widely used computational approach in the field of ML, thus achieving outstanding results on several complex cognitive tasks, matching or even beating those provided by human performance. One of the benefits of DL is the ability to learn massive amounts of data. The DL field has grown fast in the last few years and it has been extensively used to successfully address a wide range of traditional applications. This review paper represents the major concepts in deep learning and the use of the neural network, the major applications of deep learning such as in object detection, visual object recognition, speech recognition, face recognition, vision for driverless cars, virtual assistants, and many other fields such as genomics and drug discovery. Finally, this paper also showcases the current developments and challenges in deep neural networks training.*

Keywords: Deep Learning, Machine Learning, Convolution Neural Network (CNN), Deep Neural Network Architectures, Deep Learning Applications, Image Classification, Transfer Learning, Medical Image Analysis, Supervised Learning, etc.

I. INTRODUCTION

Recently, machine learning (ML) has become terribly widespread in analysis and has been incorporated into a spread of applications, as well as text mining, spam detection, video recommendation, image classification, and multimedia system conception retrieval. Among the various metric capacity unit algorithms, deep learning (DL) is the most generally used in these applications. Another name for deciliter is illustration learning (RL)[1]. deciliter springs from the standard neural network however significantly outperform its predecessors. Moreover, deciliter employs transformations and graph technologies at the same time to make up multi-layer learning models. the foremost recently developed. deciliter techniques have obtained sensible outstanding performance across a spread of applications, as well as audio and speech process, visual processing, and linguistic communication process (NLP), among others[2][3].

Usually, the effectiveness of the associate metric capacity unit rule is extremely enthusiastic about the integrity of the input-data illustration. it's been shown that an acceptable information illustration provides improved performance compared to a poor information illustration. Thus, a major analysis trend in a metric capacity unit for several years has been feature engineering, which has upon varied analysis studies. This approach aims at constructing options from the information. additionally, it's very field-specific and often needs sizable human effort. as an example, many sorts of options were introduced and compared within the pc vision context, like bar graph of oriented gradients (HOG), scale-invariant feature rework (SIFT), and a bag of words (BoW). Before long as a unique feature is introduced and is found to perform well, it becomes a brand-new analysis direction that's pursued over multiple decades. comparatively speaking, feature extraction is achieved mechanically throughout the deciliter algorithms[4].

This encourages researchers to extract discriminative options victimization of the tiniest attainable quantity of human effort and field data. These algorithms have a multi-layer information illustration design, during which the primary layers extract the low-level options whereas the last layers extract the high-level options. Note that AI (AI) originally impressed this kind of design, which simulates the method that happens in core sensory regions inside the human brain. victimization totally different scenes, the human brain will mechanically extract information illustration. Additionally, specifically, the output of this method is the classified objects, whereas the received scene data represents the input. This method stimulates the operating methodology of the human brain. Thus, it emphasizes the most advantage of deciliter.

II. BACKGROUND OF DEEP LEARNING

DL, a subset of ML, is inspired by the information processing patterns found in the human brain. DL does not require any human-designed rules to operate; rather, it uses a large amount of data to map the given input to specific labels. DL is designed using numerous layers of algorithms (artificial neural networks, or ANNs), each of which provides a different interpretation of the data that has been fed to them. Achieving the classification task using conventional ML techniques requires several sequential steps, specifically pre-processing, feature extraction, wise feature selection, learning, and classification. Furthermore, feature selection has a great impact on the performance of ML techniques.

Biased feature selection may lead to incorrect discrimination between classes. Conversely, DL has the ability to automate the learning of feature sets for several tasks, unlike conventional ML methods. DL enables learning and classification to be achieved in a single shot. DL has become incredibly. Fig 1 shows how deep learning is related to AI and Machine learning as it comes under as the subset of the techniques.

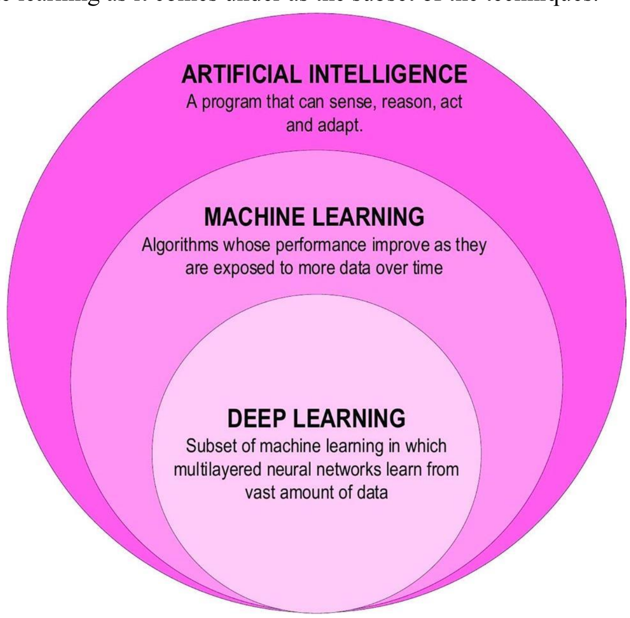

Figure 1: Deep learning subset of artificial intelligence and Machine learning.

III. NEURAL NETWORKS

The most illustrious styles of deep learning networks square measure mentioned during this section: these embrace algorithmic neural networks (RvNNs), RNNs, and CNNs. Evans and RNNs were concisely explained during this section whereas CNNs were explained in deep because of the importance of this kind. Moreover, it's the foremost utilized in many applications among different networks [5].

A. Recursive Neural Network (RNN)

RNN can do predictions in an exceeding data structure and additionally classify the outputs utilizing integrative vectors[6]. algorithmic auto-associative memory (RAAM) is the primary inspiration for RvNN development. The RNN design is generated for process objects, that have randomly formed structures like graphs or trees. This approach generates a fixed-width distributed illustration from a variable-size recursive-data structure. The network is trained mistreatment Associate in Nursing introduced back-propagation through structure (BTS) learning system. The BTS system tracks a similar technique because the general-back propagation algorithmic rule has the flexibility to support a branchy structure. Auto-association trains the network to regenerate the input-layer pattern at the output layer[7]. RNN is extremely effective within the natural language processing context. Socher et al. introduced RNN design designed to method inputs from a range of modalities. Figure 2 shows typical RNN.

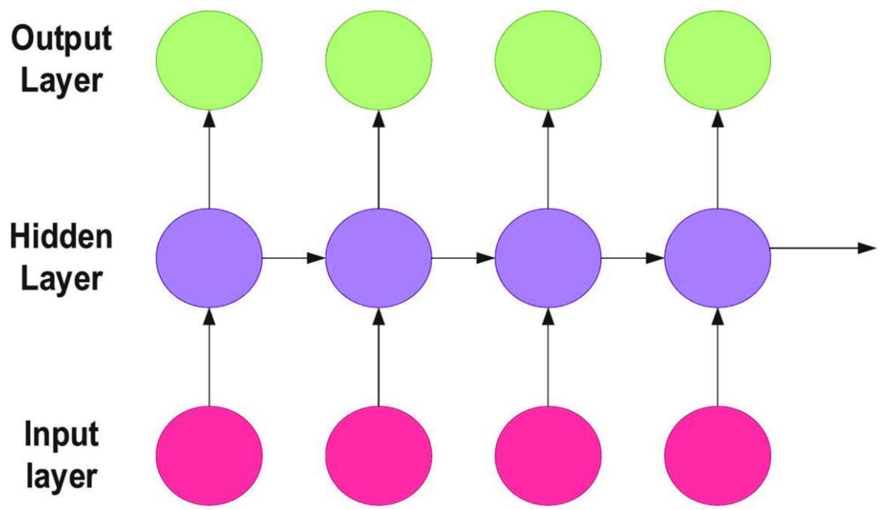

Figure 2: Typical unfolded RNN Diagram

B. ***Convolutional Neural Networks***

In the field of metric capacity unit, the CNN is that the most renowned and ordinarily utilized rule. the most advantage of CNN compared to its predecessors is that it mechanically identifies the relevant options with none human supervising. CNNs are extensively applied in a very vary of various fields, together with pc vision, speech process, Face Recognition, etc. The structure of CNNs was impressed by neurons in human and animal brains, almost like a standard neural network. additional specifically, in a very cat's brain, a fancy sequence of cells forms the visual cortex; this sequence is simulated by the CNN. Goodfellow et al. known 3 key advantages of the CNN: equivalent representations, thin interactions, and parameter sharing.

Figure 3 shows CNN. Unlike conventional absolutely connected (FC) networks, shared weights and native connections in the CNN are utilized to create full use of 2nd input-data structures like image signals. This operation utilizes an especially tiny variety of parameters, that each simplifies the coaching method and races the network. this can be constant as within the cortical area cells. Notably, solely tiny regions of a scene ar detected by these cells instead of the whole scene (i.e., these cells spatially extract the native correlation out there within the input, like native filters over the input). An ordinarily used variety of CNN, that is comparable to the multi-layer perceptron (MLP), consists of diverse convolution layers preceding sub-sampling (pooling) layers, while the ending layers are FC layers. associate degree example of CNN design for image classification is illustrated in Figure. 3.

The input x of every layer in a very CNN model is organized in 3 dimensions: height, width, and depth, or m × m×r, wherever the peak (m) is adequate to the breadth. The depth is also named because of the channel variety. for instance, in an associate degree RGB image, the depth (r) is equal to 3. many kernels (filters) out there in every convolutional layer are denoted by k and even have 3 dimensions (n×n×letter of the alphabet), almost like the input image; here,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, June 2022
Impact Factor: 6.252

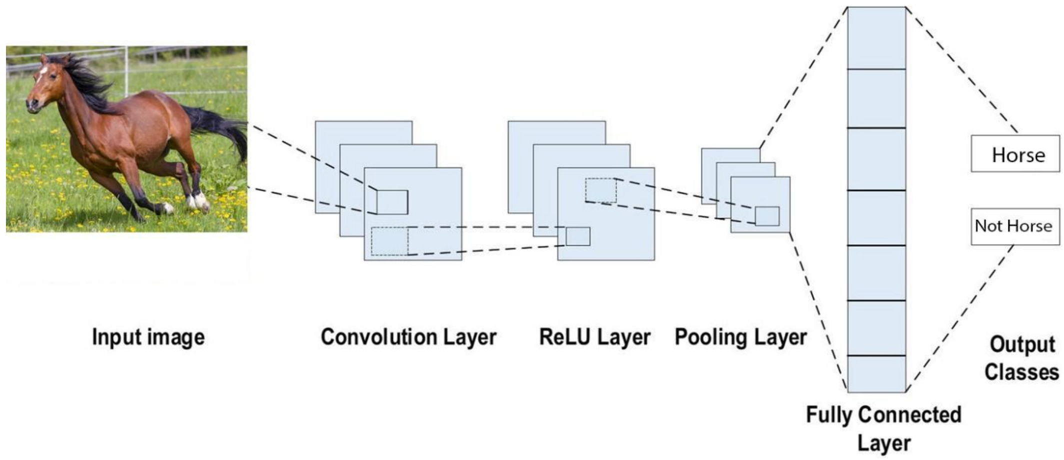

Figure 3: Example of CNN architecture for image classification

However, n should be smaller than m, whereas Q is either adequate to or smaller than r. additionally, the kernels area unit the idea of the native connections, that share similar parameters (bias metal and weight Wk) for generating k feature maps hk with a size of (m − n − one) every and area unit convolved with input, as mentioned higher than. The convolution layer calculates a real between its input and also the weights as in relative atomic mass. 1, kind of like information science, however the inputs area unit undersize areas of the initial image size. Next, by applying the nonlinearity or associate activation performed to the convolution-layer output, we tend to obtain the following: the succeeding step is down sampling each feature map within the sub-sampling layers.

This ends up in a discount within the network parameters, that accelerates the coaching method and successively permits the handling of the overfitting issue. For all feature maps, the pooling performs (e.g., liquid ecstasy or average) is applied to associate adjacent space of size p × p wherever p is the kernel size. Finally, the FC layers receive the mid-and low-level options and make the high-level abstraction, that represents the last-stage layers as in a very typical neural network. The classification scores area unit generated victimization of the ending layer [e.g., support vector machines (SVMs) or softmax]. For a given instance, each score represents the likelihood of a particular category.

C. CNN Layers

The CNN design consists of a variety of layers (or questionable multi-building blocks). every layer within the CNN design, together with its operation, is delineated intimately below.

1. Convolutional Layer: In CNN design, the foremost major factor is that the convolutional layer. It consists of a group of convolutional filters (so-called kernels). The input image, expressed as N-dimensional metrics, is convolved with these filters to get the output feature map.

- Kernel definition: A grid of distinct numbers or values describes the kernel. every worth is termed the kernel weight. Random numbers square measure appointed to act because the weights of the kernel at the start of the CNN coaching method. additionally, there square measure many totally different strategies accustomed initialize the weights. Next, these weights square measure adjusted at every coaching era; therefore, the kernel learns to extract important options.
- Convolutional Operation: ab initio, the CNN input format is delineated. The vector format is that the input of the standard neural network, whereas the multichannel image is that the input of the CNN. for example, monaural is that the format of the gray-scale image, whereas the RGB image format is three-channeled. to grasp the convolutional operation, allow us to take the Associate in Nursing example of a 4×4 gray-scale image with a 2×2 random weight-initialized kernel. First, the kernel slides over the complete image horizontally and vertically. additionally, the scalar product between

the input image and therefore the kernel is set, wherever their corresponding worth's square measure increased then summed up to form one scalar value, calculated at the same time. the complete method is then perennial till no additional slippery is feasible. Note that the calculated scalar product values represent the feature map of the output [8]. Figure eight diagrammatically illustrates the first calculations dead at every step. during this figure, the sunshine inexperienced color represents the 2×2 kernel, whereas the sunshine blue color represents the similar size space of the input image. each square measure multiplied; the tip result once rundown the ensuing product represents an entry value to the output feature map.

- However, artefact to the input image isn't applied within the previous example, whereas a stride of 1 (denoted for the chosen step-size over all vertical or horizontal locations) is applied to the kernel. Note that it's conjointly attainable to use another stride worth. additionally, a feature map of lower dimensions is obtained as a result of increasing the stride worth.
- On the opposite hand, artefact is very important to determinative border size data associated with the input image. against this, the border side-features moves anxious in no time [9]. By applying artefact, the dimensions of the input image. Typical calculation of CNN is shown in Figure 4.

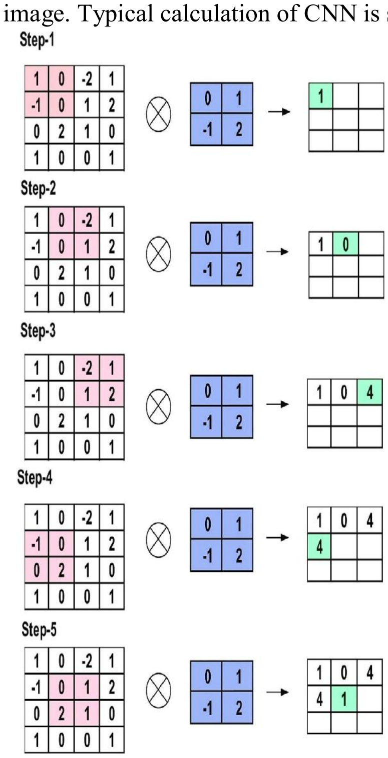

Figure 4: The initial calculations executed at each step of the convolutional layer

- Sparse Connectivity: every nerve cell of a layer in FC neural networks links with all neurons within the following layer. in contrast, in CNNs, solely some weights ar out there between 2 adjacent layers[10]. Thus, the number of needed weights or connections is little, whereas the memory needed to store these weights is additionally small; thence, this approach is memory-effective. additionally, operation is computationally rather more expensive than the dot (.) operation in CNN.
- Weight Sharing: There aren't any allotted weights between any 2 neurons of neighboring layers in CNN, because the whole weights operate with one and every one pixels of the input matrix. Learning one cluster of weights for the full input can considerably decrease the desired coaching time and varied prices, because it isn't necessary to find out further weights for every nerve cell.

2. Pooling Layer: the most task of the pooling layer is the sub-sampling of the feature maps. These maps are generated by following the convolutional operations. In alternative words, this approach shrinks large-size feature maps to form smaller feature maps. at the same time, it maintains the bulk of the dominant data (or features) in each step of the pooling stage[11]. During a similar manner to the convolutional operation, each stride and also the kernel are at first size-assigned before the pooling operation is dead. Many forms of pooling ways are out there for utilization in varied pooling layers. These ways embrace tree pooling, gated pooling, average pooling, min pooling, scoop pooling, international average pooling (GAP), and international scoop pooling. the foremost acquainted and often used pooling ways at the scoop, min, and GAP pooling. Figure nine illustrates these 3 pooling operations.

3. Activation operate (non-linearity) Mapping the input to the output is that the core operates of every type of activation operate all told forms of neural network. The input price is decided by computing the weighted summation of the nerve cell input beside its bias (if present). this suggests that the activation operate makes the choice on whether or not or to not hearth a nerve cell with respect to a specific input by making the corresponding output.
4. Absolutely Connected Layer: usually, this layer is found at the tip of every CNN design. within this layer, every nerve cell is connected to all or any neurons of the previous layer, the questionable absolutely Connected (FC) approach. it's used because the CNN classifier. It follows the fundamental methodology of the standard multiple-layer perceptron neural network, because it could be a style of feed-forward ANN. The input of the FC layer comes from the last pooling or convolutional layer. This input is within the sort of a vector, that is made from the feature maps when flattening. The output of the FC layer represents the ultimate CNN output, as illustrated in Fig. 10.
5. Loss Functions: The previous section has conferred varied layer forms of CNN design. additionally, the ultimate classification is achieved from the output layer, that represents the last layer of the CNN design. Some loss functions are used within the output layer to calculate the expected error created across the coaching samples within the CNN model. This error reveals the distinction between the particular output and also the foreseen one. Next, it'll be optimized through the CNN learning method.

   However, 2 parameters are employed by the loss operate to calculate the error. The CNN calculable output (referred to because the prediction) is that the initial parameter. the particular output (referred to because the label) is that the second parameter. many forms of loss operate are used in varied drawback varieties. the subsequent shortly explains a number of the loss operate varieties.

IV. FUTURE AND DRAWBACKS OF DEEP LEARNING

Purely supervised learning had in turn turned up through a system in spite of the chemical change effects of unattended learning. But there has been an enormous increase in the importance of unattended learning for an extended amount of your time. Animals and individuals are principally subjected to unattended learning as a result of the observation is completed for locating the structure of the globe instead of telling the names of objects.

For the active processes of human vision, the optical arrays are sampled in tasks specific and intelligent ways that employ a smaller, high resolution forever with enormous, lower resolution surroundings [12]. By combining with repeated neural networks, reinforcement learning is employed for creating calls wherever to visualize and is trained from finish to finish. during this paper, a comprehensive progressive review is accomplished within the current situation. though deep learning and straightforward reasoning are used for speech and handwriting recognition for an extended time, new paradigms are required to interchange rule-based manipulation of symbolic expressions by operations on giant vectors. the subsequent are the drawbacks of deep learning:

- **a.** Large quantity of knowledge is needed for performing better than the other technique
- **b.** Due to advanced knowledge modules, it's terribly high ticket to train
- **c.** Deep learning wants many machines and expensive GPUs for process knowledge and therefore it's costly.
- **d.** It needs a classifier for comprehending the output based on mere learning.

V. CONCLUSION

In this paper, a structured and comprehensive read of deep learning technology is explored. It starts with a history of artificial neural networks and moves to recent deep learning techniques and breakthroughs in several applications. Then, the key algorithms during this space, further as deep neural network modelling in varied dimensions are explored. Deep learning, not like ancient machine learning and data processing algorithms, will turn out extraordinarily high-level knowledge representations from monumental amounts of data. As a result, it's provided a wonderful answer to a spread of real-world issues. Deep learning has shown to be helpful in an exceedingly wide selection of applications and analysis areas like tending, sentiment analysis,visual recognition, business intelligence, cybersecurity, and lots of that area unit summarized within the paper.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Proceedings of the Advances in Neural Information Processing Systems, vol. 25, pp. 1090–1098, Lake Tahoe, NV, USA, December 2012.

[2] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1915–1929, 2013.

[3] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in Proceedings of the Advances in Neural Information Processing Systems, vol. 27, pp. 1799–1807, Montreal, Canada, December 2014.

[4] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," 2014, http://arxiv.org/abs/1409.4842.

[5] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Cernocky, "Strategies for training large scale neural network language models," in Proceedings of the Automatic Speech Recognition and Understanding, pp. 196–201, Waikoloa, HI, USA, December 2011.

[6] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[7] T. Sainath,A.-R.Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in Proceedings of the Acoustics, Speech and Signal Processing, Vancouver, Canada, May 2013.

[8] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, "Deep neural nets as a method for quantitative structureactivity relationships," Journal of Chemical Information and Modeling, vol. 55, no. 2, 263–274.

[9] T. Ciodaro, D. Deva, J. de Seixas, and D. Damazio, "Online particle detection with neural networks based on topological calorimetry information," Journal of Physics: Conference Series, vol. 368, 12030, 2012.

[10] K. Higgs, Boson Machine Learning Challenge, Kaggle, San Francisco, CA, USA, 2014, https://www.kaggle.com/c/higgsboson.

[11] M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk, "Connectomic reconstruction of the inner plexiform layer in the mouse retina," Nature, vol. 500, no. 7461, pp. 168–174.

[12] M. K. K. Leung, H. Y. Xiong, L. J. Lee, and B. J. Frey, "Deep learning of the tissue-regulated splicing code," Bioinformatics, vol. 30, no. 12, pp. i121–i129, 2014.

BIOGRAPHY

Resmi K R, received PhD degree in computer science from Mahatma Gandhi university Kerala,India in 2021.She is currently working as Asst. Professor, Computer Science, Santhigiri College of Computer Sciences, Thodupuzha, India. Her research interests include biometric, pattern recognition, image processing and computer vision.

Vysakh V Mohan is studying Master of Computer Applications in Santhigiri College of Computer Sciences, Vazhithala, Idukki, Kerala. He has completed his Bachelor of Computer Applications from Mahatma Gandhi University, Kerala.

Pradeep Prakash is studying Master of Computer Applications in Santhigiri College of Computer Sciences, Vazhithala, Idukki, Kerala. He has completed his Bachelor of Computer Applications from Mahatma Gandhi University, Kerala.