International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 02 – Issue 12 || December 2017 || PP. 32-37

Deep Learning – An Overview

Gargi Joshi[1], Rina Jugale[2]
Department of Information Technology
D. Y. Patil College of Engineering Ambi Pune

Abstract: Over the past few years' deep learning has become the go-to solution for a broad range of applications in many fields such as computer vision, speech analysis, natural language processing and robotics. This paper will provide an introduction to deep learning, showing some applications where deep architectures have been successfully used for designing intelligent systems that learn from large scale datasets. With an overview of different neural-based models, ranging from basic feed-forward neural networks to convolutional neural networks and recurrent neural networks. Furthermore, will show how these models can be trained, discussing traditional algorithms for optimization (e.g., back propagation) as well as more recent technical innovations.*Deep Learning, Machine Learning, CNN , RNN, DBN*

Keywords: Deep Learning, Machine Learning, CNN, RNN, DBN

1. Introduction

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers. In a word, accuracy. Deep learning achieves recognition accuracy at higher levels than ever before. This helps consumer electronics meet user expectations, and it is crucial for safety-critical applications like driverless cars. Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images. Most deep learning methods use neural network architectures, which is why deep learning models are often referred to as deep neural networks. The term "deep" usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction. Deep learning is a specialized form of machine learning. A machine learning workflow starts with relevant features being manually extracted from images. The features are then used to create a model that categorizes the objects in the image. With a deep learning workflow, relevant features are automatically extracted from images. In addition, deep learning performs "end-to-end learning" – where a network is given raw data and a task to perform, such as classification, and it learns how to do this automatically.

2. Literature Survey

2.1 ALEX 2012

Trained the network on ImageNet data, which contained over 15 million annotated images from a total of over 22,000 categories.UsedReLU for the nonlinearity functions (Found to decrease training time as ReLUs are several times faster than the conventional tanh function).Used data augmentation techniques that consisted of image translations, horizontal reflections, and patch extractions.Implemented dropout layers in order to combat the problem of overfitting to the training data.Trained the model using batch stochastic gradient descent, with specific values for momentum and weight decay.Trained on two GTX 580 GPUs

2.2 ZF NET 2013

Very similar architecture to AlexNet, except for a few minor modifications.AlexNet trained on 15 million images, while ZF Net trained on only 1.3 million images. Instead of using 11x11 sized filters in the first layer (which is what AlexNet implemented), ZF Net used filters of size 7x7 and a decreased stride value. The reasoning behind this modification is that a smaller filter size in the first conv layer helps retain a lot of original pixel information in the input volume. A filtering of size 11x11 proved to be skipping a lot of relevant information, especially as this is the first conv layer. As the network grows, we also see a rise in the number of

filters used. Used ReLUs for their activation functions, cross-entropy loss for the error function, and trained using batch stochastic gradient descent. Trained on a GTX 580 GPU for **twelve days**.Developed a visualization technique named Deconvolutional Network, which helps to examine different feature activations and their relation to the input space. Called "deconvnet" because it maps features to pixels (the opposite of what a convolutional layer does)

2.3 GOOGLENET 2015

Used 9 Inception modules in the whole architecture, with over 100 layers in total! Now that is deep… No use of fully connected layers! They use an average pool instead, to go from a 7x7x1024 volume to a 1x1x1024 volume. This saves a huge number of parameters. Uses 12x fewer parameters than AlexNet.During testing, multiple crops of the same image were created, fed into the network, and the softmax probabilities were averaged to give us the final solution. Utilized concepts from R-CNN (a paper we'll discuss later) for their detection model. There are updated versions to the Inception module (Versions 6 and 7). Trained on "a few high-end GPUs **within a week**

2.4 GAN 2014

These networks could be the next big development. Before talking about this paper, let's talk a little about adversarial examples. For example, let's consider a trained CNN that works well on ImageNet data. Let's take an example image and apply a perturbation, or a slight modification, so that the prediction error is *maximized*. Thus, the object category of the prediction changes, while the image itself looks the same when compared to the image without the perturbation. From the highest level, adversarial examples are basically the images that fool ConvNetAdversarial examples (paper) definitely surprised a lot of researchers and quickly became a topic of interest. Now let's talk about the generative adversarial networks. Let's think of two models, a generative model and a discriminative model. The discriminative model has the task of determining whether a given image looks natural (an image from the dataset) or looks like it has been artificially created. The task of the generator is to create images so that the discriminator gets trained to produce the correct outputs. This can be thought of as a zero-sum or minimax two player game. The analogy used in the paper is that the generative model is like "a team of counterfeiters, trying to produce and use fake currency" while the discriminative model is like "the police, trying to detect the counterfeit currency". The generator is trying to fool the discriminator while the discriminator is trying to not get fooled by the generator. As the models train, both methods are improved until a point where the "counterfeits are indistinguishable from the genuine articles

3. System Requirements

1. Deep learning requires large amounts of labeled data. For example, driverless car development requires millions of images and thousands of hours of video.
2. Deep learning requires substantial computing power. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from weeks to hours or less.

4. Comparison with Machine Learning and Different Neural Network Architectures

Machine learning offers a variety of techniques and models you can choose based on your application, the size of data you're processing and the type of problem you want to solve. A successful deep learning application requires a very large amount of data (thousands of images) to train the model, as well as GPUs, or graphics processing units, to rapidly process your data. When choosing between machine learning and deep learning, consider whether you have a high-performance GPU and lots of labeled data. If you don't have either of those things, it may make more sense to use machine learning instead of deep learning. Deep learning is generally more complex, so you'll need at least a few thousand images to get reliable results. Having a high-performance GPU means the model will take less time to analyze all those images Another key difference is deep learning algorithms scale with data, whereas shallow learning converges. Shallow learning refers to machine learning methods that plateau at a certain level of performance when you add more examples and training data to the network.A key advantage of deep learning networks is that they often continue to improve as the size of your data increases In machine learning, you manually choose features and a classifier to sort images. With deep learning feature extraction and modeling steps are automatic Deep Learning is primarily about neural networks, where a network is an interconnected web of nodes and edges. • Neural nets were designed to perform complex tasks, such as the task of placing objects into categories based on a few attributes. • Neural nets are highly structured networks, and have three kinds of layers - an input, an output, and so called hidden layers, which refer to any layers between the input and the output layers. • Each node (also called a neuron) in the hidden and output layers has a classifier

The **Deep Belief Network**, or DBN, was also conceived by Geoff Hinton. • Used by Google for their work on the image recognition problem. • DBN is trained two layers at a time, and these two layers are treated like an RBM. • Throughout the net, the hidden layer of an RBM acts as the input layer of the adjacent one. So the first RBM is trained, and its outputs are then used as inputs to the next RBM. This procedure is repeated until the output layer is reached. Deep Bel • DBN is capable of recognizing the inherent patterns in the data. In other words, it's a sophisticated, multilayer feature extractor. • The unique aspect of this type of net is that each layer ends up learning the full input structure. • Layers generally learn progressively complex patterns – for facial recognition, early layers could detect edges and later layers would combine them to form facial features. • DBN learns the hidden patterns globally, like a camera slowly bringing an image into focus. • DBN still requires a set of labels to apply to the resulting patterns. As a final step, the DBN is fine-tuned with supervised learning and a small set of labeled examples

The Recurrent Neural Net (RNN) is the brainchild of JuergenSchmidhuber and SeppHochreiter. • RNNs have a feedback loop where the net's output is fed back into the net along with the next input. • RNNs receive an input and produce an output. Unlike other nets, the inputs and outputs can come in a sequence. • Variant of RNN is Long Term Short Memory (LSTM) N is suitable for time series data, where an output can be the next value in a sequence, or the next several values

5. Deep Learning for Object Classification Using CNN

One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images.CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pertained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification..CNNs learn to detect different features of an image using tens or hundreds of hidden layers. Every hidden layer increases the complexity of the learned image features. For example, the first hidden layer could learn how to detect edges, and the last learns how to detect more complex shapes specifically catered to the shape of the object we are trying to recognize

CNN inspired by the Visual Cortex. • CNNs are deep nets that are used for image, object, and even speech recognition. • Pioneered by YannLecun (NYU) • Deep supervised neural networks are generally too difficult to train. • CNNs have multiple types of layers, the first of which is the convolutional layer. A series of filters forms layer one, called the convolutional layer. The weights and biases in this layer determine the effectiveness of the filtering process. • Each flashlight represents a single neuron. Typically, neurons in a layer activate or fire. On the other hand, in the convolutional layer, neurons search for patterns through convolution. Neurons from different filters search for different patterns, and thus they will process the input differently

The three most common ways people use deep learning to perform object classification are:

1. Training from Scratch

To train a deep network from scratch, you gather a very large labeled data set and design a network architecture that will learn the features and model. This is good for new applications, or applications that will have a large number of output categories. This is a less common approach because with the large amount of data and rate of learning, these networks typically take days or weeks to train.

2. Transfer Learning

Most deep learning applications use the transfer learning approach, a process that involves fine-tuning a pretrained model. You start with an existing network, such as AlexNet or GoogLeNet, and feed in new data containing previously unknown classes. After making some tweaks to the network, you can now perform a new task, such as categorizing only dogs or cats instead of 1000 different objects. This also has the advantage of needing much less data (processing thousands of images, rather than millions), so computation time drops to minutes or hours.Transfer learning requires an interface to the internals of the pre-existing network, so it can be surgically modified and enhanced for the new task.

3. Feature Extraction

A slightly less common, more specialized approach to deep learning is to use the network as a feature extractor. Since all the layers are tasked with learning certain features from images, we can pull these features out of the network at any time during the training process. These features can then be used as input to a machine learning modelFeature engineering is the process of using domain knowledge of the data to create features that

make machine learning algorithms work. • "When working on a machine learning problem, feature engineering is manually designing what the input x's should be." "Coming up with features is difficult, time consuming, requires expert knowledge.

4. Deep Learning is Hierarchical Feature Learning

In addition to scalability, another often cited benefit of deep learning models is their ability to perform automatic feature extraction from raw data, also called feature learning.Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features.

Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features..The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning.

The quintessential example of a deep learning model is the feedforward deep network or multilayer perceptron (MLP)a kind of learning where the representation you form have several levels of abstraction, rather than a direct input to output Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory. description of "deep" to describe their approach to developing networks with many more layers than was previously typical.We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.backpropagation through deep autoencoders would be very effective for nonlinear dimensionality reduction, provided that computers were fast enough, data sets were big enough, and the initial weights were close enough to a good solution. All three conditions are now satisfied.The descriptions of deep learning in the Royal Society talk are very backpropagation centric as you would expect. Interesting, he gives 4 reasons why backpropagation (read "deep learning") did not take off last time around in the 1990s. The first two points match comments by Andrew Ng above about datasets being too small and computers being too slow.

5. Deep Learning as Scalable Learning across Domains

Deep learning excels on problem domains where the inputs (and even output) are analog. Meaning, they are not a few quantities in a tabular format but instead are images of pixel data, documents of text data or files of audio data. YannLeCun is the director of Facebook Research and is the father of the network architecture that excels at object recognition in image data called the Convolutional Neural Network (CNN). This technique is seeing great success because like multilayer perceptron feed forward neural networks, the technique scales with data and model size and can be trained with back propagation. This biases his definition of deep learning as the development of very large CNNs, which have had great success on object recognition in photographs. deep learning [is] … a pipeline of modules all of which are trainable. … deep because [has] multiple stages in the process of recognizing an object and all of those stages are part of the training" is the father of another popular algorithm that like MLPs and CNNs also scales with model size and dataset size and can be trained with back propagation, but is instead tailored to learning sequence data, called the Long Short-Term Memory Network (LSTM), a type of recurrent neural network. We do see some confusion in the phrasing of the field as "deep learning". In his 2014 paper titled "Deep Learning in Neural Networks: An Overview" he does comment on the problematic naming of the field and the differentiation of deep from shallow learning. He also interestingly describes depth in terms of the complexity of the problem rather than the model used to solve the problem. To achieve this, we developed a novel agent, a deep Q-network (DQN), which is able to combine reinforcement learning with a class of artificial neural network known as deep neural networks. Notably, recent advances in deep neural networks, in which several layers of nodes are used to build up progressively more abstract representations of the data, have made it possible for artificial neural networks to learn concepts such as object categories directly from raw sensory data. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Later the multi-layered approach is described in terms of representation learning and abstraction. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. […] The key aspect of deep learning is that these layers

of features are not designed by human engineers: they are learned from data using a general-purpose learning paradigm

6. Advantages

Robust No need to design the features ahead of time – features are automatically learned to be optimal for the task at hand Robustness to natural variations in the data is automatically learned •Generalizable The same neural net approach can be used for many different applications and data types Scalable Performance improves with more data, method is massively parallelizable

7. Disadvantages

Deep Learning requires a large dataset, hence long training period. In term of cost, Machine Learning methods like SVMs and other tree ensembles are very easily deployed even by relative machine learning novices and can usually get you reasonably good results. • Deep learning methods tend to learn everything. It's better to encode prior knowledge about structure of images (or audio or text). • The learned features are often difficult to understand. Many vision features are also not really human-understandable (e.g, concatenations/combinations of different features). • Requires a good understanding of how to model multiple modalities with traditional tools

8. Applications

Deep learning applications are used in industries from automated driving to medical devices.

Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

Aerospace and Defense: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.

Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

9. Conclusions

This paper expresses the importance of deep learning technology applications. In recent years, the technology of deep learning in image classification, object detection and face identification and many other computer vision tasks have achieved great success. Compares deep learning with machine learning and other architectures focus on usage of deep learning for object classification and give a brief description for DFN, CNN , RNN deep learning architectures and state advantage, disadvantage and applications of deep learning in various fields such as automated driving, aerospace defense, medical research, industrial automation and electronics.

References

[1]. Hinton, G. E., Osindero, S., and Teh, Y. (2006). A fast learning algorithm for deep belief nets. Neural Computa on, 18:1527–1554.

[2]. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2012). Improving neural networks by preven ng co-adapta on of feature detectors.CoRR, abs/1207.0580.

[3]. Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computa on, 9(8):1735–1780.

[4]. Lecun, Y., Boћou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recogni on. In Proceedings of the IEEE, pages 2278–2324.

[5]. Martens, J. (2010). Deep learning via hessian-free op miza on. In Proceedings of the 27th Interna onal Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa,

[6]. R. Bellman, Dynamic Programming. Princeton, NJ: Princeton Univ. Press, 1957.

[7]. R. Duda, P. Hart, and D. Stork, Pattern Recognition, 2nd ed. New York: Wiley-Interscience, 2000.

[8]. T. Lee and D. Mumford, "Hierarchical Bayesian inference in the visual cortex," J. Opt. Soc. Amer., vol. 20, pt. 7, pp. 1434–1448, 2003.
[9]. T. Lee, D. Mumford, R. Romero, and V. Lamme, "The role of the primary visual cortex in higher level vision," Vision Res., vol. 38, pp. 2429–2454, 1998
[10]. G. Wallis and H. Bülthoff, "Learning to recognize objects," Trends Cogn.Sci., vol. 3, no. 1, pp. 23–31, 1999.
[11]. G. Wallis and E. Rolls, "Invariant face and object recognition in the visual system," Prog.Neurobiol., vol. 51, pp. 167–194, 1997.
[12]. Y. Bengio, "Learning deep architectures for AI," Found. Trends Mach. Learn., vol. 2, no. 1, pp. 1–127, 2009.
[13]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.