# Overview of deep learning theory and its application

**Authors:** Xiaoning Liu

**Date:** 2024-01-06T00:00:00+00:00

## Abstract

Deep learning represents a novel research direction within machine learning, introduced to advance the field toward its original goal of artificial intelligence (AI). It discovers the underlying structures and hierarchical representations within sample data. The knowledge acquired through these learning processes proves highly valuable for interpreting various data modalities, including text, images, and audio. Its ultimate objective is to enable machines to possess analytical and learning capabilities comparable to humans, allowing them to recognize and process data such as text, images, and audio. As a sophisticated machine learning methodology, it has achieved remarkable performance in speech and image recognition, substantially surpassing prior approaches. This success extends to numerous domains, including information retrieval, data mining, machine translation, natural language processing, multimedia learning, speech processing, recommendation systems, personalization technologies, and other related areas. This article examines the theoretical foundations of deep learning and investigates its algorithmic applications across diverse fields, aiming to provide a valuable reference for deep learning research.

## Full Text

## Abstract

Deep learning represents a novel research direction within machine learning that aims to advance the field closer to its original goal of achieving artificial intelligence. It seeks to discover the inherent laws and hierarchical representations within sample data. The information acquired through this learning process proves highly valuable for interpreting various data types, including text, images, and audio. The ultimate objective is to enable machines to analyze and learn with human-like capability, recognizing text, images, and sounds effectively. As a sophisticated machine learning algorithm, deep learning has demonstrated performance in speech and image recognition that far surpasses previous technologies. Its applications span numerous domains, including search technology, data mining, machine translation, natural language processing, multimedia learning, speech processing, recommendation systems, and personalized technologies. This paper examines the theoretical foundations of deep learning and investigates its algorithmic applications across diverse fields, providing a reference for future research.

**Keywords:** Deep learning, natural language processing

## Chapter 1 Fundamentals

### 1.1 Theoretical Background

Deep learning encompasses pattern analysis methods that generally fall into three categories: (1) neural network systems based on convolutional operations, namely Convolutional Neural Networks (CNN); (2) multi-layer neural networks for self-encoding, including Autoencoders and Sparse Coding; and (3) Deep Belief Networks (DBN), which employ unsupervised pre-training using multi-layer self-encoding neural networks followed by supervised fine-tuning of network weights using label information. Through multi-level processing, deep learning progressively transforms initial "low-level" feature representations into "high-level" feature representations, enabling complex classification and other learning tasks to be accomplished using relatively simple models. Consequently, deep learning can be understood as "feature learning" or "representation learning." Traditionally, when applying machine learning to real-world tasks, features describing samples required manual design by human experts—a process known as "feature engineering." Since feature quality critically impacts generalization performance and designing effective features proves challenging for human experts, feature learning automates this process by generating high-quality features through machine learning itself, advancing the field toward fully automatic data analysis. In recent years, researchers have increasingly combined these methods to improve prediction accuracy. Compared with conventional shallow learning approaches, deep learning methods assume substantially more model parameters, making training more demanding. According to statistical learning theory, greater model complexity necessitates larger training datasets. During the 1980s and 1990s, limited computing power and data availability constrained deep learning' s performance in pattern analysis. Since 2006, Hinton and colleagues' introduction of fast training algorithms for Restricted Boltzmann Machines (RBM) using the contrastive divergence (CD-k) algorithm has established RBM as a powerful tool for increasing neural network depth. This breakthrough led to the development of deep architectures such as the widely adopted DBN, which Hinton developed and which Microsoft and other companies have successfully applied to speech recognition. Concurrently, sparse coding has been integrated into deep learning due to its ability to automatically extract features from data, while convolutional neural networks based on local receptive fields have also received extensive attention.

### 1.2 Core Concepts

Deep learning is a subset of machine learning, which itself constitutes an essential pathway to artificial intelligence. The concept originated from artificial neural network research, with multi-layer perceptrons containing multiple hidden layers representing a fundamental deep learning architecture. Deep learning combines low-level features to construct more abstract high-level representations of attribute categories, thereby discovering distributed feature representations of data. The motivation for deep learning research lies in establishing neural networks that simulate the human brain' s analytical learning mechanisms to interpret data such as images, sounds, and text.

The computation involved in generating outputs from inputs can be represented using flow graphs—graphical structures where each node represents a basic computation and its computed value, with results applied to child nodes. By defining permissible computations at each node and possible graph structures, a family of functions can be specified. Input nodes have no parents, while output nodes have no children. A key property of such flow graphs is depth, defined as the length of the longest path from any input to any output. Traditional feedforward neural networks can be viewed as having depth equal to their number of layers.

Artificial intelligence research has historically followed two approaches: "expert systems" based on extensive "If-Then" rules represent a top-down methodology, while Artificial Neural Networks embody a bottom-up alternative. Although neural networks lack a rigorous formal definition, their fundamental characteristic involves mimicking the brain' s patterns of information transmission and processing between neurons.

[Figure 1: see original paper] A Deep Learning Model with Multiple Hidden Layers [Figure 2: see original paper] The Training Process

### 1.3 Distinguishing Characteristics

Compared with traditional shallow learning, deep learning exhibits several key characteristics. First, deep learning emphasizes model depth, typically comprising five, six, or even ten or more hidden layers. Second, it clarifies the importance of feature learning—through layer-by-layer feature transformation, sample representations from the original space are mapped into a new feature space that facilitates easier classification or prediction. Unlike manually engineered rule-based features, learning features from large datasets better captures the rich intrinsic information within data.

By designing and establishing an appropriate number of neural computing nodes and hierarchical operational layers, selecting suitable input and output layers, and performing network learning and optimization, deep learning establishes a functional relationship from input to output. While this relationship cannot be determined absolutely, it can approximate the true underlying correlations with high accuracy. Successfully trained network models thereby enable automation of complex processing tasks.

# Chapter 2 Typical Models

## 2.1 Convolutional Neural Network Model

Before the advent of unsupervised pre-training, training deep neural networks proved exceptionally difficult, with convolutional neural networks representing a notable exception. Inspired by the visual system' s architecture, the first computational model of a convolutional neural network emerged from Fukushima' s neocognitron. Based on local neuronal connections and hierarchical image transformations, this approach applies identical parameters to neurons at different positions in the previous layer, yielding a translation-invariant network structure. Subsequently, LeCun et al. designed and trained convolutional neural networks using error gradient-based methods, achieving superior performance on various pattern recognition tasks. To date, pattern recognition systems based on convolutional neural networks rank among the best-performing implementations, demonstrating extraordinary capabilities particularly in handwritten character recognition.

Connections between convolutional layers in CNNs are sparse, meaning that unlike fully connected feedforward networks, neurons in convolutional layers connect only to a subset of neurons in adjacent layers rather than to all of them. Specifically, any pixel (neuron) in the l-th layer feature map represents a linear combination of pixels within the receptive field defined by the convolutional kernel in the (l-1)-th layer. This sparse connectivity acts as a regularizer, improving network stability and generalization while preventing overfitting. Additionally, sparse connections reduce the total number of weight parameters, accelerating learning and decreasing computational memory requirements.

In convolutional neural networks, all pixels within the same feature map channel share a single set of convolutional kernel weights—a property known as weight sharing. This characteristic distinguishes CNNs from other locally connected networks that, while sparse, maintain distinct weights for different connections. Like sparse connectivity, weight sharing reduces the total parameter count and provides regularization benefits.

Viewed from the perspective of fully connected networks, the sparse connections and weight sharing in convolutional neural networks can be interpreted as two infinitely strong priors: all weight coefficients for a hidden layer neuron outside its receptive field are constrained to zero, while the receptive field itself can shift across space, and all neurons within a channel share identical weight coefficients.

## 2.2 DBN Model

Deep Belief Networks can be interpreted as Bayesian probabilistic generative models comprising multiple layers of stochastic latent variables. The top two layers maintain undirected symmetric connections, while lower layers receive top-down directed connections from the layer above. The lowest-level units represent visible input data vectors. DBNs consist of a stack of two structural units, typically Restricted Boltzmann Machines, where each RBM unit' s visible layer neuron count equals the hidden layer neuron count of the preceding RBM unit. Following the deep learning mechanism, input examples train the first RBM layer, and the resulting outputs train the second RBM layer. Additional layers can be stacked to enhance model performance. During unsupervised pre-training, DBNs encode inputs to the top-level RBM and decode the top-level state back to the lowest-level units to achieve input reconstruction. As structural units of DBNs, RBMs share parameters across each DBN layer.

### 2.3 Stacked Autoencoder Network Model

The architecture of stacked autoencoder networks resembles that of DBNs, comprising several stacked structural units. The key distinction lies in using autoencoder models rather than RBMs as the fundamental building blocks. An autoencoder is a two-layer neural network where the first layer serves as the encoding layer and the second as the decoding layer.

## Chapter 3 Applications

Deep learning finds extensive application across computer vision, speech recognition, and natural language processing. The Multimedia Laboratory of The Chinese University of Hong Kong pioneered deep learning applications in computer vision research among Chinese institutions, once surpassing Facebook to win the Large-scale Facial Recognition Competition (LFW), marking the first instance where AI recognition capabilities exceeded human performance in this domain. Microsoft researchers, in collaboration with Hinton, were the first to integrate RBMs and DBNs into speech recognition acoustic model training, achieving remarkable success in large-vocabulary speech recognition systems and reducing word error rates by 30 percent. However, DNNs currently lack efficient parallel algorithms, prompting many research institutions to leverage GPU platforms and large-scale corpora to improve training efficiency. Internationally, companies including IBM and Google have rapidly advanced DNN speech recognition research, while domestic organizations such as Alibaba, Baidu, and the Institute of Automation of the Chinese Academy of Sciences have also launched deep learning initiatives for speech recognition. Numerous institutions have pursued natural language processing research. In 2013, Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean published "Efficient Estimation of Word Representations in Vector Space," establishing the word2vec model. Compared with traditional bag-of-words models, word2vec more effectively captures grammatical information. Deep learning primarily applies to natural language processing, machine translation, and semantic mining. In 2020, deep learning began accelerating innovation in semiconductor packaging and testing. For reducing repetitive labor, improving yield, controlling accuracy and efficiency, and lowering inspection costs, AI-driven AOI (Automated Optical Inspection) based on deep learning holds broad market prospects, though implementation challenges remain. On April 13, 2020, Swiss scientists published a medical AI study in *Nature Machine Intelligence* introducing an AI system capable of scanning cardiovascular blood flow within seconds. This deep learning model may enable clinicians to observe real-time blood flow changes during MRI scans, thereby optimizing diagnostic workflows.

## References

[1] Max C, Christian V, Tobias S, et al. Deep Concrete Flow: Deep learning based characterisation of fresh concrete properties from open-channel flow using spatio-temporal flow fields[J]. Construction and Building Materials, 2024, 411134809.

[2] Xiaofei Y, Enrique C R D, Yang Z, et al. UAV-deployed deep learning network for real-time multi-class damage detection using model quantization techniques[J]. Automation in Construction, 2024, 159105254.

[3] Jiyu X, Mitsuyoshi A, M.DF. Autonomous detection of steel corrosion spatial variability in reinforced concrete using X-ray technology and deep learning-based semantic segmentation[J]. Automation in Construction, 2024, 158105252.

[4] Sheng J, Xinming W, Qinghao M. Spatial memory-augmented visual navigation based on hierarchical deep reinforcement learning in unknown environments[J]. Knowledge-Based Systems, 2024, 285111358.

[5] Xukai Z, Yuxing L, Guangsi L. An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images[J]. Engineering Applications of Artificial Intelligence, 2024.

[6] Zhaohui L, Longyan W, Jian X, et al. A deep learning framework for reconstructing experimental missing flow field of hydrofoil[J]. Ocean Engineering, 2024, 293116605.

[7] Ruoning Y, Wenyi R, Man Z, et al. Transparent objects segmentation based on polarization imaging and deep learning[J]. Optics Communications, 2024, 555130246.

[8] Mina K, Angelo F, Ludovic L. Improving deep-learning methods for area-based traffic demand prediction via hierarchical reconciliation[J]. Transportation Research Part C, 2024, 159104410.

*Note: Figure translations are in progress. See original paper for figures.*