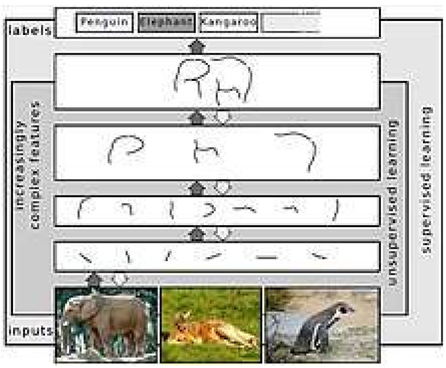
Deep learning (also known as **deep structured learning** or **hierarchical learning**) is part of a broader family of machine learning methods based on artificial neural networks. Learning can be supervised, semi-supervised or unsupervised.[1][2][3]

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts.[4][5][6]

Artificial Neural Networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.[7][8][9]

Definition



Representing Images on Multiple Layers of Abstraction in Deep Learning [10]

Deep learning is a class of machine learning algorithms that[11](pp199–200) uses multiple layers to progressively extract higher level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits or letters or faces.

Overview

Most modern deep learning models are based on artificial neural networks, specifically, Convolutional Neural Networks (CNN)s, although they can also include propositional formulas or latent variables organized layer-wise in deep generative models such as the nodes in deep belief networks and deep Boltzmann machines.[12]

In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. In an image recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face. Importantly, a deep learning process can learn which features to optimally place in which level *on its own*. (Of course, this does not completely eliminate the need for hand-tuning; for example, varying numbers of layers and layer sizes can provide different degrees of abstraction.)[1][13]

The word "deep" in "deep learning" refers to the number of layers through which the data is transformed. More precisely, deep learning systems have a substantial *credit assignment path* (CAP) depth. The CAP is the chain of transformations from input to output. CAPs describe potentially causal connections between input and output. For a feedforward neural network, the depth of the CAPs is that of the network and is the number of hidden layers plus one (as the output layer is also parameterized). For recurrent neural networks, in which a signal may propagate through a layer more than once, the CAP depth is potentially unlimited.[2] No universally agreed upon threshold of depth divides shallow learning from deep learning, but most researchers agree that deep learning involves CAP depth higher than 2. CAP of depth 2 has been shown to be a universal approximator in the sense that it can emulate any function.[14] Beyond that, more layers do not add to the function approximator ability of the network. Deep models (CAP > 2) are able to extract better features than shallow models and hence, extra layers help in learning the features effectively.

Deep learning architectures can be constructed with a greedy layer-by-layer method.[15] Deep learning helps to disentangle these abstractions and pick out which features improve performance.[1]

For supervised learning tasks, deep learning methods eliminate feature engineering, by translating the data into compact intermediate representations akin to principal components, and derive layered structures that remove redundancy in representation.

Deep learning algorithms can be applied to unsupervised learning tasks. This is an important benefit because unlabeled data are more abundant than the labeled data. Examples of deep structures that can be trained in an unsupervised manner are neural history compressors[16] and deep belief networks.[1][17]

Interpretations

Deep neural networks are generally interpreted in terms of the universal approximation theorem[18][19][20][21][22][23] or probabilistic inference.[11][12][1][2][17][24][25]

The classic universal approximation theorem concerns the capacity of feedforward neural networks with a single hidden layer of finite size to approximate continuous functions.[18][19][20][21][22] In 1989, the first proof was published by George Cybenko for sigmoid activation functions[19] and was generalised to feed-forward multi-layer architectures in 1991 by Kurt Hornik.[20] Recent work also showed that universal approximation also holds for non-bounded activation functions such as the rectified linear unit. [26]

The universal approximation theorem for deep neural networks concerns the capacity of networks with bounded width but the depth is allowed to grow. Lu et al.[23] proved that if the width of a deep neural network with ReLU activation is strictly larger than the input dimension, then the network can approximate any Lebesgue integrable function; If the width is smaller or equal to the input dimension, then deep neural network is not a universal approximator.

The probabilistic interpretation[24] derives from the field of machine learning. It features inference,[11][12][1][2][17][24] as well as the optimization concepts of training and testing, related to fitting and generalization, respectively. More specifically, the probabilistic interpretation considers the activation nonlinearity as a cumulative distribution function.[24] The probabilistic interpretation led to the introduction of dropout as regularizer in neural networks.[27] The probabilistic interpretation was introduced by researchers including Hopfield, Widrow and Narendra and popularized in surveys such as the one by Bishop.[28]

History

The term *Deep Learning* was introduced to the machine learning community by Rina Dechter in 1986,[29][16] and to artificial neural networks by Igor Aizenberg and colleagues in 2000, in the context of Boolean threshold neurons.[30][31]

The first general, working learning algorithm for supervised, deep, feedforward, multilayer perceptrons was published by Alexey Ivakhnenko and Lapa in 1965.[32] A 1971 paper described already a deep network with 8 layers trained by the group method of data handling algorithm.[33]

Other deep learning working architectures, specifically those built for computer vision, began with the Neocognitron introduced by Kunihiko Fukushima in 1980.[34] In 1989, Yann LeCun et al. applied the standard backpropagation algorithm, which had been around as the reverse mode of automatic differentiation since 1970,[35][36][37][38] to a deep neural network with the purpose of recognizing handwritten ZIP codes on mail. While the algorithm worked, training required 3 days.[39]

By 1991 such systems were used for recognizing isolated 2-D hand-written digits, while recognizing 3-D objects was done by matching 2-D images with a handcrafted 3-D object model. Weng *et al.* suggested that a human brain does not use a monolithic 3-D object model and in

1992 they published Cresceptron,[40][41][42] a method for performing 3-D object recognition in cluttered scenes. Because it directly used natural images, Cresceptron started the beginning of general-purpose visual learning for natural 3D worlds. Cresceptron is a cascade of layers similar to Neocognitron. But while Neocognitron required a human programmer to hand-merge features, Cresceptron learned an open number of features in each layer without supervision, where each feature is represented by a convolution kernel. Cresceptron segmented each learned object from a cluttered scene through back-analysis through the network. Max pooling, now often adopted by deep neural networks (e.g. ImageNet tests), was first used in Cresceptron to reduce the position resolution by a factor of (2x2) to 1 through the cascade for better generalization.

In 1994, André de Carvalho, together with Mike Fairhurst and David Bisset, published experimental results of a multi-layer boolean neural network, also known as a weightless neural network, composed of a 3-layers self-organising feature extraction neural network module (SOFT) followed by a multi-layer classification neural network module (GSN), which were independently trained. Each layer in the feature extraction module extracted features with growing complexity regarding the previous layer.[43]

In 1995, Brendan Frey demonstrated that it was possible to train (over two days) a network containing six fully connected layers and several hundred hidden units using the wake-sleep algorithm, co-developed with Peter Dayan and Hinton.[44] Many factors contribute to the slow speed, including the vanishing gradient problem analyzed in 1991 by Sepp Hochreiter.[45][46]

Simpler models that use task-specific handcrafted features such as Gabor filters and support vector machines (SVMs) were a popular choice in the 1990s and 2000s, because of artificial neural network's (ANN) computational cost and a lack of understanding of how the brain wires its biological networks.

Both shallow and deep learning (e.g., recurrent nets) of ANNs have been explored for many years.[47][48][49] These methods never outperformed non-uniform internal-handcrafting Gaussian mixture model/Hidden Markov model (GMM-HMM) technology based on generative models of speech trained discriminatively.[50] Key difficulties have been analyzed, including gradient diminishing[45] and weak temporal correlation structure in neural predictive models.[51][52] Additional difficulties were the lack of training data and limited computing power.

Most speech recognition researchers moved away from neural nets to pursue generative modeling. An exception was at SRI International in the late 1990s. Funded by the US government's NSA and DARPA, SRI studied deep neural networks in speech and speaker recognition. The speaker recognition team led by Larry Heck achieved the first significant success with deep neural networks in speech processing in the 1998 National Institute of Standards and Technology Speaker Recognition evaluation.[53] While SRI experienced success with deep neural networks in speaker recognition, they were unsuccessful in demonstrating similar success in speech recognition. The principle of elevating "raw" features over hand-crafted optimization was first explored successfully in the architecture of deep autoencoder on the "raw" spectrogram or linear filter-bank features in the late 1990s,[53] showing its superiority over the Mel-Cepstral features that contain stages of fixed transformation from spectrograms. The raw features of speech, waveforms, later produced excellent larger-scale results.[54]

Many aspects of speech recognition were taken over by a deep learning method called long short-term memory (LSTM), a recurrent neural network published by Hochreiter and Schmidhuber in 1997.[55] LSTM RNNs avoid the vanishing gradient problem and can learn "Very Deep Learning" tasks[2] that require memories of events that happened thousands of discrete time steps before, which is important for speech. In 2003, LSTM started to become competitive with traditional speech recognizers on certain tasks.[56] Later it was combined with connectionist temporal classification (CTC)[57] in stacks of LSTM RNNs.[58] In 2015, Google's speech recognition reportedly experienced a dramatic performance jump of 49% through CTC-trained LSTM, which they made available through Google Voice Search.[59]

In 2006, publications by Geoff Hinton, Ruslan Salakhutdinov, Osindero and Teh[60][61][62] showed how a many-layered feedforward neural network could be effectively pre-trained one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, then fine-tuning it using supervised backpropagation.[63] The papers referred to *learning* for *deep belief nets*.

Deep learning is part of state-of-the-art systems in various disciplines, particularly computer vision and automatic speech recognition (ASR). Results on commonly used evaluation sets such as TIMIT (ASR) and MNIST (image classification), as well as a range of large-vocabulary speech recognition tasks have steadily improved.[64][65][66] Convolutional neural networks (CNNs) were superseded for ASR by CTC[57] for LSTM.[55][59][67][68][69][70][71] but are more successful in computer vision.

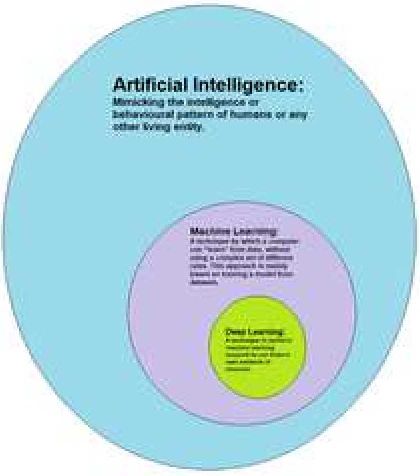
The impact of deep learning in industry began in the early 2000s, when CNNs already processed an estimated 10% to 20% of all the checks written in the US, according to Yann LeCun.[72] Industrial applications of deep learning to large-scale speech recognition started around 2010.

The 2009 NIPS Workshop on Deep Learning for Speech Recognition[73] was motivated by the limitations of deep generative models of speech, and the possibility that given more capable hardware and large-scale data sets that deep neural nets (DNN) might become practical. It was believed that pre-training DNNs using generative models of deep belief nets (DBN) would overcome the main difficulties of neural nets.[74] However, it was discovered that replacing pre-training with large amounts of training data for straightforward backpropagation when using DNNs with large, context-dependent output layers produced error rates dramatically lower than then-state-of-the-art Gaussian mixture model (GMM)/Hidden Markov Model (HMM) and also than more-advanced generative model-based systems.[64][75] The nature of the recognition errors produced by the two types of systems was characteristically different,[76][73] offering technical insights into how to integrate deep learning into the existing highly efficient, run-time speech decoding system deployed by all major speech recognition systems.[11][77][78] Analysis around 2009-2010, contrasted the GMM (and other generative speech models) vs. DNN models, stimulated early industrial investment in deep learning for speech recognition,[76][73] eventually leading to pervasive and dominant use in that industry. That analysis was done with comparable performance (less than 1.5% in error rate) between discriminative DNNs and generative models.[64][76][74][79]

In 2010, researchers extended deep learning from TIMIT to large vocabulary speech recognition, by adopting large output layers of the DNN based on context-dependent HMM states constructed by decision trees.[80][81][82][77]

Advances in hardware have enabled renewed interest in deep learning. In 2009, Nvidia was involved in what was called the “big bang” of deep learning, “as deep-learning neural networks were trained with Nvidia graphics processing units (GPUs).”[83] That year, Google Brain used Nvidia GPUs to create capable DNNs. While there, Andrew Ng determined that GPUs could increase the speed of deep-learning systems by about 100 times.[84] In particular, GPUs are well-suited for the matrix/vector computations involved in machine learning.[85][86][87] GPUs speed up training algorithms by orders of magnitude, reducing running times from weeks to days.[88][89] Further, specialized hardware and algorithm optimizations can be used for efficient processing of deep learning models.[90]

Deep learning revolution



How deep learning is a subset of machine learning and how machine learning is a subset of artificial intelligence (AI).

In 2012, a team led by George E. Dahl won the "Merck Molecular Activity Challenge" using multi-task deep neural networks to predict the biomolecular target of one drug.[91][92] In 2014, Hochreiter's group used deep learning to detect off-target and toxic effects of environmental chemicals in nutrients, household products and drugs and won the "Tox21 Data Challenge" of NIH, FDA and NCATS.[93][94][95]

Significant additional impacts in image or object recognition were felt from 2011 to 2012. Although CNNs trained by backpropagation had been around for decades, and GPU implementations of NNs for years, including CNNs, fast implementations of CNNs with max-pooling on GPUs in the style of Ciresan and colleagues were needed to progress on computer vision.[85][87][39][96][2] In 2011, this approach achieved for the first time superhuman performance in a visual pattern recognition contest. Also in 2011, it won the ICDAR Chinese handwriting contest, and in May 2012, it won the ISBI image segmentation contest.[97] Until 2011, CNNs did not play a major role at computer vision conferences, but in June 2012, a paper by Ciresan et al. at the leading conference CVPR[4] showed how max-pooling CNNs on GPU can dramatically improve many vision benchmark records. In October 2012, a similar system by Krizhevsky et al.[5] won the large-scale ImageNet competition by a significant margin over shallow machine learning methods. In November 2012, Ciresan et al.'s system also won the ICPR contest on analysis of large medical images for cancer detection, and in the following year also the MICCAI Grand Challenge on the same topic.[98] In 2013 and 2014, the error rate on the ImageNet task using deep learning was further reduced, following a similar trend in large-scale speech recognition. The Wolfram Image Identification project publicized these improvements.[99]

Image classification was then extended to the more challenging task of generating descriptions (captions) for images, often as a combination of CNNs and LSTMs.[100][101][102][103]

Some researchers assess that the October 2012 ImageNet victory anchored the start of a "deep learning revolution" that has transformed the AI industry.[104]

In March 2019, Yoshua Bengio, Geoffrey Hinton and Yann LeCun were awarded the Turing Award for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

Neural networks

Artificial neural networks

Artificial neural networks (**ANNs**) or **connectionist systems** are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn (progressively improve their ability) to do tasks by considering examples, generally without task-specific programming. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labeled as "cat" or "no cat" and using the analytic results to identify cats in other images. They have found most use in applications difficult to express with a traditional computer algorithm using rule-based programming.

An ANN is based on a collection of connected units called artificial neurons, (analogous to biological neurons in a biological brain). Each connection (synapse) between neurons can transmit a signal to another neuron. The receiving (postsynaptic) neuron can process the signal(s) and then signal downstream neurons connected to it. Neurons may have state, generally represented by real numbers, typically between 0 and 1. Neurons and synapses may also have a weight that varies as learning proceeds, which can increase or decrease the strength of the signal that it sends downstream.

Typically, neurons are organized in layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first (input), to the last (output) layer, possibly after traversing the layers multiple times.

The original goal of the neural network approach was to solve problems in the same way that a human brain would. Over time, attention focused on matching specific mental abilities, leading to deviations from biology such as backpropagation, or passing information in the reverse direction and adjusting the network to reflect that information.

Neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis.

As of 2017, neural networks typically have a few thousand to a few million units and millions of connections. Despite this number being several order of magnitude less than the number of neurons on a human brain, these networks can perform many tasks at a level beyond that of humans (e.g., recognizing faces, playing "Go"[105]).

Deep neural networks

This section may be too technical for most readers to understand. Please help improve it to make it understandable to non-experts, without removing the technical details. Learn more

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers.[12][2] The DNN finds the correct mathematical manipulation to turn the input into the output, whether it be a linear relationship or a non-linear relationship. The network moves through the layers calculating the probability of each output. For example, a DNN that is trained to recognize dog breeds will go over the given image and calculate the probability that the dog in the image is a certain breed. The user can review the results and select which probabilities the network should display (above a certain threshold, etc.) and return the proposed label. Each mathematical manipulation as such is considered a layer, and complex DNN have many layers, hence the name "deep" networks.

DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives.[106] The extra layers enable composition of features from lower layers, potentially modeling complex data with fewer units than a similarly performing shallow network.[12]

Deep architectures include many variants of a few basic approaches. Each architecture has found success in specific domains. It is not always possible to compare the performance of multiple architectures, unless they have been evaluated on the same data sets.

DNNs are typically feedforward networks in which data flows from the input layer to the output layer without looping back. At first, the DNN creates a map of virtual neurons and assigns random numerical values, or "weights", to connections between them. The weights and inputs are multiplied and return an output between 0 and 1. If the network did not accurately recognize a particular pattern, an algorithm would adjust the weights.[107] That way the algorithm can make certain parameters more influential, until it determines the correct mathematical manipulation to fully process the data.

Recurrent neural networks (RNNs), in which data can flow in any direction, are used for applications such as language modeling.[108][109][110][111][112] Long short-term memory is particularly effective for this use.[55][113]

Convolutional deep neural networks (CNNs) are used in computer vision.[114] CNNs also have been applied to acoustic modeling for automatic speech recognition (ASR).[71]

Challenges

As with ANNs, many issues can arise with naively trained DNNs. Two common issues are overfitting and computation time.

DNNs are prone to overfitting because of the added layers of abstraction, which allow them to model rare dependencies in the training data. Regularization methods such as Ivakhnenko's unit pruning[33] or weight decay (ℓ_2-regularization) or sparsity (ℓ_1-regularization) can be applied during training to combat overfitting.[115] Alternatively dropout regularization randomly omits units from the hidden layers during training. This helps to exclude rare dependencies.[116] Finally, data can be augmented via methods such as cropping and rotating such that smaller training sets can be increased in size to reduce the chances of overfitting.[117]

DNNs must consider many training parameters, such as the size (number of layers and number of units per layer), the learning rate, and initial weights. Sweeping through the parameter space for optimal parameters may not be feasible due to the cost in time and computational resources. Various tricks, such as batching (computing the gradient on several training examples at once rather than individual examples)[118] speed up computation. Large processing capabilities of many-core architectures (such as GPUs or the Intel Xeon Phi) have produced significant speedups in training, because of the suitability of such processing architectures for the matrix and vector computations.[119][120]

Alternatively, engineers may look for other types of neural networks with more straightforward and convergent training algorithms. CMAC (cerebellar model articulation controller) is one such kind of neural network. It doesn't require learning rates or randomized initial weights for CMAC. The training process can be guaranteed to converge in one step with a new batch of data, and the computational complexity of the training algorithm is linear with respect to the number of neurons involved.[121][122]

Applications

Automatic speech recognition

Large-scale automatic speech recognition is the first and most convincing successful case of deep learning. LSTM RNNs can learn "Very Deep Learning" tasks[2] that involve multi-second intervals containing speech events separated by thousands of discrete time steps, where one time step corresponds to about 10 ms. LSTM with forget gates[113] is competitive with traditional speech recognizers on certain tasks.[56]

The initial success in speech recognition was based on small-scale recognition tasks based on TIMIT. The data set contains 630 speakers from eight major dialects of American English, where each speaker reads 10 sentences.[123] Its small size lets many configurations be tried. More importantly, the TIMIT task concerns phone-sequence recognition, which, unlike word-sequence recognition, allows weak phone bigram language models. This lets the strength of the acoustic modeling aspects of speech recognition be more easily analyzed. The error rates listed below, including these early results and measured as percent phone error rates (PER), have been summarized since 1991.

Method	Percent phone error rate (PER) (%)
Randomly Initialized RNN[124]	26.1
Bayesian Triphone GMM-HMM	25.6
Hidden Trajectory (Generative) Model	24.8
Monophone Randomly Initialized DNN	23.4
Monophone DBN-DNN	22.4
Triphone GMM-HMM with BMMI Training	21.7
Monophone DBN-DNN on fbank	20.7
Convolutional DNN[125]	20.0
Convolutional DNN w. Heterogeneous Pooling	18.7
Ensemble DNN/CNN/RNN[126]	18.3
Bidirectional LSTM	17.9
Hierarchical Convolutional Deep Maxout Network[127]	16.5

The debut of DNNs for speaker recognition in the late 1990s and speech recognition around 2009-2011 and of LSTM around 2003-2007, accelerated progress in eight major areas:[11][79][77]

- Scale-up/out and accelerated DNN training and decoding
- Sequence discriminative training
- Feature processing by deep models with solid understanding of the underlying mechanisms
- Adaptation of DNNs and related deep models
- Multi-task and transfer learning by DNNs and related deep models
- CNNs and how to design them to best exploit domain knowledge of speech
- RNN and its rich LSTM variants
- Other types of deep models including tensor-based models and integrated deep generative/discriminative models.

All major commercial speech recognition systems (e.g., Microsoft Cortana, Xbox, Skype Translator, Amazon Alexa, Google Now, Apple Siri, Baidu and iFlyTek voice search, and a range of Nuance speech products, etc.) are based on deep learning.[11][128][129][130]

Image recognition

A common evaluation set for image classification is the MNIST database data set. MNIST is composed of handwritten digits and includes 60,000 training examples and 10,000 test examples. As with TIMIT, its small size lets users test multiple configurations. A comprehensive list of results on this set is available.[131]

Deep learning-based image recognition has become "superhuman", producing more accurate results than human contestants. This first occurred in 2011.[132]

Deep learning-trained vehicles now interpret 360° camera views.[133] Another example is Facial Dysmorphology Novel Analysis (FDNA) used to analyze cases of human malformation connected to a large database of genetic syndromes.

Visual art processing

Closely related to the progress that has been made in image recognition is the increasing application of deep learning techniques to various visual art tasks. DNNs have proven themselves capable, for example, of a) identifying the style period of a given painting, b) Neural Style Transfer - capturing the style of a given artwork and applying it in a visually pleasing manner to an arbitrary photograph or video, and c) generating striking imagery based on random visual input fields.[134][135]

Natural language processing

Neural networks have been used for implementing language models since the early 2000s.[108][136] LSTM helped to improve machine translation and language modeling.[109][110][111]

Other key techniques in this field are negative sampling[137] and word embedding. Word embedding, such as *word2vec*, can be thought of as a representational layer in a deep learning architecture that transforms an atomic word into a positional representation of the word relative to other words in the dataset; the position is represented as a point in a vector space. Using word embedding as an RNN input layer allows the network to parse sentences and phrases using an effective compositional vector grammar. A compositional vector grammar can be thought of as probabilistic context free grammar (PCFG) implemented by an RNN.[138] Recursive auto-encoders built atop word embeddings can assess sentence similarity and detect paraphrasing.[138] Deep neural architectures provide the best results for constituency parsing,[139] sentiment analysis,[140] information retrieval,[141][142] spoken language understanding,[143] machine translation,[109][144] contextual entity linking,[144] writing style recognition,[145] Text classification and others.[146]

Recent developments generalize word embedding to sentence embedding.

Google Translate (GT) uses a large end-to-end long short-term memory network.[147][148][149][150][151][152] Google Neural Machine Translation (GNMT) uses an example-based machine translation method in which the system "learns from millions of examples."[148] It translates "whole sentences at a time, rather than pieces. Google Translate supports over one hundred languages.[148] The network encodes the "semantics of the sentence rather than simply memorizing phrase-to-phrase translations".[148][153] GT uses English as an intermediate between most language pairs.[153]

Drug discovery and toxicology

A large percentage of candidate drugs fail to win regulatory approval. These failures are caused by insufficient efficacy (on-target effect), undesired interactions (off-target effects), or unanticipated toxic effects.[154][155] Research has explored use of deep learning to predict the biomolecular targets,[91][92] off-targets, and toxic effects of environmental chemicals in nutrients, household products and drugs.[93][94][95]

AtomNet is a deep learning system for structure-based rational drug design.[156] AtomNet was used to predict novel candidate biomolecules for disease targets such as the Ebola virus[157] and multiple sclerosis.[158][159]

In 2019 generative neural networks were used to produce molecules that were validated experimentally all the way into mice [160], [161].

Customer relationship management

Deep reinforcement learning has been used to approximate the value of possible direct marketing actions, defined in terms of RFM variables. The estimated value function was shown to have a natural interpretation as customer lifetime value.[162]

Recommendation systems

Recommendation systems have used deep learning to extract meaningful features for a latent factor model for content-based music recommendations.[163] [164] Multiview deep learning has been applied for learning user preferences from multiple domains.[165] The model uses a hybrid collaborative and content-based approach and enhances recommendations in multiple tasks.

Bioinformatics

An autoencoder ANN was used in bioinformatics, to predict gene ontology annotations and gene-function relationships.[166]

In medical informatics, deep learning was used to predict sleep quality based on data from wearables[167] and predictions of health complications from electronic health record data.[168] Deep learning has also showed efficacy in healthcare.[169]

Medical Image Analysis

Deep learning has been shown to produce competitive results in medical application such as cancer cell classification, lesion detection, organ segmentation and image enhancement[170][171]

Mobile advertising

Finding the appropriate mobile audience for mobile advertising is always challenging, since many data points must be considered and assimilated before a target segment can be created and used in ad serving by any ad server.[172] Deep learning has been used to interpret large, many-dimensioned advertising datasets. Many data points are collected during the request/serve/click internet advertising cycle. This information can form the basis of machine learning to improve ad selection.

Image restoration

Deep learning has been successfully applied to inverse problems such as denoising, super-resolution, inpainting, and film colorization.[173] These applications include learning methods such as "Shrinkage Fields for Effective Image Restoration"[174] which trains on an image dataset, and Deep Image Prior, which trains on the image that needs restoration.

Financial fraud detection

Deep learning is being successfully applied to financial fraud detection and anti-money laundering. "Deep anti-money laundering detection system can spot and recognize relationships and similarities between data and, further down the road, learn to detect anomalies or classify and predict specific events". The solution leverages both supervised learning techniques, such as the classification of suspicious transactions, and unsupervised learning, e.g. anomaly detection. [175]

Military

The United States Department of Defense applied deep learning to train robots in new tasks through observation.[176]

Relation to human cognitive and brain development

Deep learning is closely related to a class of theories of brain development (specifically, neocortical development) proposed by cognitive neuroscientists in the early 1990s.[177][178][179][180] These developmental theories were instantiated in computational models, making them predecessors of deep learning systems. These developmental models share the property that various proposed learning dynamics in the brain (e.g., a wave of nerve growth factor) support the self-organization somewhat analogous to the neural networks utilized in deep learning models. Like the neocortex, neural networks employ a hierarchy of layered filters in which each layer considers information from a prior layer (or the operating environment), and then passes its output (and possibly the original input), to other layers. This process yields a self-organizing stack of transducers, well-tuned to their operating environment. A 1995 description stated, "...the infant's brain seems to organize itself under the influence of waves of so-called trophic-factors ... different regions of the brain become connected sequentially, with one layer of tissue maturing before another and so on until the whole brain is mature."[181]

A variety of approaches have been used to investigate the plausibility of deep learning models from a neurobiological perspective. On the one hand, several variants of the backpropagation algorithm have been proposed in order to increase its processing realism.[182][183] Other researchers have argued that unsupervised forms of deep learning, such as those based on hierarchical generative models and deep belief networks, may be closer to biological reality.[184][185] In this respect, generative neural network models have been related to neurobiological evidence about sampling-based processing in the cerebral cortex.[186]

Although a systematic comparison between the human brain organization and the neuronal encoding in deep networks has not yet been established, several analogies have been reported. For example, the computations performed by deep learning units could be similar to those of actual neurons[187][188] and neural populations.[189] Similarly, the representations developed by deep learning models are similar to those measured in the primate visual system[190] both at the single-unit[191] and at the population[192] levels.

Commercial activity

Facebook's AI lab performs tasks such as automatically tagging uploaded pictures with the names of the people in them.[193]

Google's DeepMind Technologies developed a system capable of learning how to play Atari video games using only pixels as data input. In 2015 they demonstrated their AlphaGo system, which learned the game of Go well enough to beat a professional Go player.[194][195][196] Google Translate uses an LSTM to translate between more than 100 languages.

In 2015, Blippar demonstrated a mobile augmented reality application that uses deep learning to recognize objects in real time.[197]

In 2017, Covariant.ai was launched, which focuses on integrating deep learning into factories.[198]

As of 2008,[199] researchers at The University of Texas at Austin (UT) developed a machine learning framework called Training an Agent Manually via Evaluative Reinforcement, or TAMER, which proposed new methods for robots or computer programs to learn how to perform tasks by interacting with a human instructor.[176] First developed as TAMER, a new algorithm called Deep TAMER was later introduced in 2018 during a collaboration between U.S. Army Research Laboratory (ARL) and UT researchers. Deep TAMER used deep learning to provide a robot the ability to learn new tasks through observation.[176] Using Deep TAMER, a robot learned a task with a human trainer, watching video streams or observing a human perform a task in-person. The robot later practiced the task with the help of some coaching from the trainer, who provided feedback such as “good job” and “bad job.”[200]

Criticism and comment

Deep learning has attracted both criticism and comment, in some cases from outside the field of computer science.

Theory

A main criticism concerns the lack of theory surrounding some methods.[201] Learning in the most common deep architectures is implemented using well-understood gradient descent. However, the theory surrounding other algorithms, such as contrastive divergence is less clear.

(e.g., Does it converge? If so, how fast? What is it approximating?) Deep learning methods are often looked at as a black box, with most confirmations done empirically, rather than theoretically.[202]

Others point out that deep learning should be looked at as a step towards realizing strong AI, not as an all-encompassing solution. Despite the power of deep learning methods, they still lack much of the functionality needed for realizing this goal entirely. Research psychologist Gary Marcus noted:

> "Realistically, deep learning is only part of the larger challenge of building intelligent machines. Such techniques lack ways of representing causal relationships (...) have no obvious ways of performing logical inferences, and they are also still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful A.I. systems, like Watson (...) use techniques like deep learning as just one element in a very complicated ensemble of techniques, ranging from the statistical technique of Bayesian inference to deductive reasoning."[203]

As an alternative to this emphasis on the limits of deep learning, one author speculated that it might be possible to train a machine vision stack to perform the sophisticated task of discriminating between "old master" and amateur figure drawings, and hypothesized that such a sensitivity might represent the rudiments of a non-trivial machine empathy.[204] This same author proposed that this would be in line with anthropology, which identifies a concern with aesthetics as a key element of behavioral modernity.[205]

In further reference to the idea that artistic sensitivity might inhere within relatively low levels of the cognitive hierarchy, a published series of graphic representations of the internal states of deep (20-30 layers) neural networks attempting to discern within essentially random data the images on which they were trained[206] demonstrate a visual appeal: the original research notice received well over 1,000 comments, and was the subject of what was for a time the most frequently accessed article on *The Guardian*'s[207] web site.

Errors

Some deep learning architectures display problematic behaviors,[208] such as confidently classifying unrecognizable images as belonging to a familiar category of ordinary images[209] and misclassifying minuscule perturbations of correctly classified images.[210] Goertzel hypothesized that these behaviors are due to limitations in their internal representations and that these limitations would inhibit integration into heterogeneous multi-component artificial general intelligence (AGI) architectures.[208] These issues may possibly be addressed by deep learning architectures that internally form states homologous to image-grammar[211] decompositions of observed entities and events.[208] Learning a grammar (visual or linguistic) from training data would be equivalent to restricting the system to commonsense reasoning that operates on concepts in terms of grammatical production rules and is a basic goal of both human language acquisition[212] and artificial intelligence (AI).[213]

Cyber threat

As deep learning moves from the lab into the world, research and experience shows that artificial neural networks are vulnerable to hacks and deception.[214] By identifying patterns that these systems use to function, attackers can modify inputs to ANNs in such a way that the ANN finds a match that human observers would not recognize. For example, an attacker can make subtle changes to an image such that the ANN finds a match even though the image looks to a human nothing like the search target. Such a manipulation is termed an "adversarial attack."[215] In 2016 researchers used one ANN to doctor images in trial and error fashion, identify another's focal points and thereby generate images that deceived it. The modified images looked no different to human eyes. Another group showed that printouts of doctored images then photographed successfully tricked an image classification system.[216] One defense is reverse image search, in which a possible fake image is submitted to a site such as TinEye that can then find other instances of it. A refinement is to search using only parts of the image, to identify images from which that piece may have been taken.[217]

Another group showed that certain psychedelic spectacles could fool a facial recognition system into thinking ordinary people were celebrities, potentially allowing one person to impersonate another. In 2017 researchers added stickers to stop signs and caused an ANN to misclassify them.[216]

ANNs can however be further trained to detect attempts at deception, potentially leading attackers and defenders into an arms race similar to the kind that already defines the malware defense industry. ANNs have been trained to defeat ANN-based anti-malware software by repeatedly attacking a defense with malware that was continually altered by a genetic algorithm until it tricked the anti-malware while retaining its ability to damage the target.[216]

Another group demonstrated that certain sounds could make the Google Now voice command system open a particular web address that would download malware.[216]

In “data poisoning,” false data is continually smuggled into a machine learning system's training set to prevent it from achieving mastery.[216]

Reliance on human microwork

Most Deep Learning systems rely on training and verification data that is generated and/or annotated by humans. It has been argued in media philosophy that not only low-payed clickwork (e.g. on Amazon Mechanical Turk) is regularly deployed for this purpose, but also implicit forms of human microwork that are often not recognized as such.[218] The philosopher Rainer Mühlhoff distinguishes five types of "machinic capture" of human microwork to generate training data: (1) gamification (the embedding of annotation or computation tasks in the flow of a game), (2) "trapping and tracking" (e.g. CAPTCHAs for image recognition or click-tracking on Google search results pages), (3) exploitation of social motivations (e.g. tagging faces on Facebook to obtain labeled facial images), (4) information mining (e.g. by leveraging quantified-self devices such as activity trackers) and (5) clickwork.[218] Mühlhoff argues that in most commercial end-user applications of Deep Learning such as Facebook's face recognition system, the need for training data does not stop once an ANN is trained. Rather, there is a continued demand for human-generated verification data to constantly calibrate and update the ANN. For this purpose Facebook introduced the feature that once a user is automatically recognized in an image, they receive a notification. They can choose whether of not they like to be publicly labeled on the image, or tell Facebook that it is not them in the picture.[219] This user interface is a mechanism to generate "a constant stream of verification data"[218] to further train the network in real-time. As Mühlhoff argues, involvement of human users to generate training and verification data is so typical for most commercial end-user applications of Deep Learning that such systems may be referred to as "human-aided artificial intelligence"[218].

See also

- Applications of artificial intelligence
- Comparison of deep learning software
- Compressed sensing
- Echo state network
- List of artificial intelligence projects
- Liquid state machine
- List of datasets for machine learning research
- Reservoir computing

- Sparse coding

References

1. Bengio, Y.; Courville, A.; Vincent, P. (2013). "Representation Learning: A Review and New Perspectives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **35** (8): 1798–1828. arXiv:1206.5538 . doi:10.1109/tpami.2013.50 . PMID 23787338 .
2. Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". *Neural Networks*. **61**: 85–117. arXiv:1404.7828 . doi:10.1016/j.neunet.2014.09.003 . PMID 25462637 .
3. Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). "Deep Learning". *Nature*. **521** (7553): 436–444. Bibcode:2015Natur.521..436L . doi:10.1038/nature14539 . PMID 26017442 .
4. Ciresan, D.; Meier, U.; Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification". *2012 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3642–3649. arXiv:1202.2745 . doi:10.1109/cvpr.2012.6248110 . ISBN 978-1-4673-1228-8.
5. Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffry (2012). "ImageNet Classification with Deep Convolutional Neural Networks" (PDF). *NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada*.
6. "Google's AlphaGo AI wins three-match series against the world's best Go player" . *TechCrunch*. 25 May 2017.
7. Marblestone, Adam H.; Wayne, Greg; Kording, Konrad P. (2016). "Toward an Integration of Deep Learning and Neuroscience" . *Frontiers in Computational Neuroscience*. **10**: 94. arXiv:1606.03813 . Bibcode:2016arXiv160603813M . doi:10.3389/fncom.2016.00094 . PMC 5021692 . PMID 27683554 .
8. Olshausen, B. A. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature*. **381** (6583): 607–609. Bibcode:1996Natur.381..607O . doi:10.1038/381607a0 . PMID 8637596 .
9. Bengio, Yoshua; Lee, Dong-Hyun; Bornschein, Jorg; Mesnard, Thomas; Lin, Zhouhan (2015-02-13). "Towards Biologically Plausible Deep Learning". arXiv:1502.04156 [cs.LG].
10. Schulz, Hannes; Behnke, Sven (2012-11-01). "Deep Learning". *KI - Künstliche Intelligenz*. **26** (4): 357–363. doi:10.1007/s13218-012-0198-z . ISSN 1610-1987 .
11. Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications" (PDF). *Foundations and Trends in Signal Processing*. **7** (3–4): 1–199. doi:10.1561/2000000039 .
12. Bengio, Yoshua (2009). "Learning Deep Architectures for AI" (PDF). *Foundations and Trends in Machine Learning*. **2** (1): 1–127. CiteSeerX 10.1.1.701.9550 . doi:10.1561/2200000006 . Archived from the original (PDF) on 2016-03-04. Retrieved 2015-09-03.

13. LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (28 May 2015). "Deep learning". *Nature*. **521** (7553): 436–444. Bibcode:2015Natur.521..436L . doi:10.1038/nature14539 . PMID 26017442 .

14. Shigeki, Sugiyama (2019-04-12). *Human Behavior and Another Kind in Consciousness: Emerging Research and Opportunities: Emerging Research and Opportunities* . IGI Global. ISBN 978-1-5225-8218-2.

15. Bengio, Yoshua; Lamblin, Pascal; Popovici, Dan; Larochelle, Hugo (2007). *Greedy layer-wise training of deep networks* (PDF). Advances in neural information processing systems. pp. 153–160.

16. Jürgen Schmidhuber (2015). Deep Learning. Scholarpedia, 10(11):32832. Online

17. Hinton, G.E. (2009). "Deep belief networks". *Scholarpedia*. **4** (5): 5947. Bibcode:2009SchpJ...4.5947H . doi:10.4249/scholarpedia.5947 .

18. Balázs Csanád Csáji (2001). Approximation with Artificial Neural Networks; Faculty of Sciences; Eötvös Loránd University, Hungary

19. Cybenko (1989). "Approximations by superpositions of sigmoidal functions" (PDF). *Mathematics of Control, Signals, and Systems*. **2** (4): 303–314. doi:10.1007/bf02551274 . Archived from the original (PDF) on 2015-10-10.

20. Hornik, Kurt (1991). "Approximation Capabilities of Multilayer Feedforward Networks". *Neural Networks*. **4** (2): 251–257. doi:10.1016/0893-6080(91)90009-t .

21. Haykin, Simon S. (1999). *Neural Networks: A Comprehensive Foundation* . Prentice Hall. ISBN 978-0-13-273350-2.

22. Hassoun, Mohamad H. (1995). *Fundamentals of Artificial Neural Networks* . MIT Press. p. 48. ISBN 978-0-262-08239-6.

23. Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The Expressive Power of Neural Networks: A View from the Width . Neural Information Processing Systems, 6231-6239.

24. Murphy, Kevin P. (24 August 2012). *Machine Learning: A Probabilistic Perspective* . MIT Press. ISBN 978-0-262-01802-9.

25. Patel, Ankit; Nguyen, Tan; Baraniuk, Richard (2016). "A Probabilistic Framework for Deep Learning" (PDF). *Advances in Neural Information Processing Systems*. arXiv:1612.01936 . Bibcode:2016arXiv161201936P .

26. Sonoda, Sho; Murata, Noboru (2017). "Neural network with unbounded activation functions is universal approximator". *Applied and Computational Harmonic Analysis*. **43** (2): 233–268. arXiv:1505.03654 . doi:10.1016/j.acha.2015.12.005 .

27. Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. (2012). "Improving neural networks by preventing co-adaptation of feature detectors". arXiv:1207.0580 [math.LG].

28. Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning* (PDF). Springer. ISBN 978-0-387-31073-2.

29. Rina Dechter (1986). Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory.Online

30. Igor Aizenberg, Naum N. Aizenberg, Joos P.L. Vandewalle (2000). Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer Science & Business Media.

31. Co-evolving recurrent neurons learn deep memory POMDPs. Proc. GECCO, Washington, D. C., pp. 1795-1802, ACM Press, New York, NY, USA, 2005.

32. Ivakhnenko, A. G.; Lapa, V. G. (1967). *Cybernetics and Forecasting Techniques*. American Elsevier Publishing Co. ISBN 978-0-444-00020-0.

33. Ivakhnenko, Alexey (1971). "Polynomial theory of complex systems" (PDF). *IEEE Transactions on Systems, Man and Cybernetics*. SMC-1 (4): 364–378. doi:10.1109/TSMC.1971.4308320.

34. Fukushima, K. (1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biol. Cybern.* **36** (4): 193–202. doi:10.1007/bf00344251. PMID 7370364.

35. Seppo Linnainmaa (1970). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's Thesis (in Finnish), Univ. Helsinki, 6-7.

36. Griewank, Andreas (2012). "Who Invented the Reverse Mode of Differentiation?" (PDF). *Documenta Mathematica* (Extra Volume ISMP): 389–400. Archived from the original (PDF) on 2017-07-21. Retrieved 2017-06-11.

37. Werbos, P. (1974). "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences". *Harvard University*. Retrieved 12 June 2017.

38. Werbos, Paul (1982). "Applications of advances in nonlinear sensitivity analysis" (PDF). *System modeling and optimization*. Springer. pp. 762–770.

39. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, 1, pp. 541–551, 1989.

40. J. Weng, N. Ahuja and T. S. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," *Proc. International Joint Conference on Neural Networks*, Baltimore, Maryland, vol I, pp. 576-581, June, 1992.

41. J. Weng, N. Ahuja and T. S. Huang, "Learning recognition and segmentation of 3-D objects from 2-D images ," *Proc. 4th International Conf. Computer Vision*, Berlin, Germany, pp. 121-128, May, 1993.

42. J. Weng, N. Ahuja and T. S. Huang, "Learning recognition and segmentation using the Cresceptron ," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 105-139, Nov. 1997.

43. de Carvalho, Andre C. L. F.; Fairhurst, Mike C.; Bisset, David (1994-08-08). "An integrated Boolean neural network for pattern classification". *Pattern Recognition Letters*. **15** (8): 807–813. doi:10.1016/0167-8655(94)90009-4 .

44. Hinton, Geoffrey E.; Dayan, Peter; Frey, Brendan J.; Neal, Radford (1995-05-26). "The wake-sleep algorithm for unsupervised neural networks". *Science*. **268** (5214): 1158–1161. Bibcode:1995Sci...268.1158H . doi:10.1126/science.7761831 . PMID 7761831 .

45. S. Hochreiter., "Untersuchungen zu dynamischen neuronalen Netzen ," *Diploma thesis. Institut f. Informatik, Technische Univ. Munich. Advisor: J. Schmidhuber*, 1991.

46. Hochreiter, S.; et al. (15 January 2001). "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies" . In Kolen, John F.; Kremer, Stefan C. (eds.). *A Field Guide to Dynamical Recurrent Networks*. John Wiley & Sons. ISBN 978-0-7803-5369-5.

47. Morgan, Nelson; Bourlard, Hervé; Renals, Steve; Cohen, Michael; Franco, Horacio (1993-08-01). "Hybrid neural network/hidden markov model systems for continuous speech recognition". *International Journal of Pattern Recognition and Artificial Intelligence*. **07** (4): 899–916. doi:10.1142/s0218001493000455 . ISSN 0218-0014 .

48. Robinson, T. (1992). "A real-time recurrent error propagation network word recognition system" . *ICASSP*. Icassp'92: 617–620. ISBN 9780780305328.

49. Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K. J. (March 1989). "Phoneme recognition using time-delay neural networks" (PDF). *IEEE Transactions on Acoustics, Speech, and Signal Processing*. **37** (3): 328–339. doi:10.1109/29.21701 . hdl:10338.dmlcz/135496 . ISSN 0096-3518 .

50. Baker, J.; Deng, Li; Glass, Jim; Khudanpur, S.; Lee, C.-H.; Morgan, N.; O'Shaughnessy, D. (2009). "Research Developments and Directions in Speech Recognition and Understanding, Part 1". *IEEE Signal Processing Magazine*. **26** (3): 75–80. Bibcode:2009ISPM...26...75B . doi:10.1109/msp.2009.932166 .

51. Bengio, Y. (1991). "Artificial Neural Networks and their Application to Speech/Sequence Recognition" . McGill University Ph.D. thesis.

52. Deng, L.; Hassanein, K.; Elmasry, M. (1994). "Analysis of correlation structure for a neural predictive model with applications to speech recognition". *Neural Networks*. **7** (2): 331–339. doi:10.1016/0893-6080(94)90027-2 .

53. Heck, L.; Konig, Y.; Sonmez, M.; Weintraub, M. (2000). "Robustness to Telephone Handset Distortion in Speaker Recognition by Discriminative Feature Design". *Speech Communication*. **31** (2): 181–192. doi:10.1016/s0167-6393(99)00077-1 .
54. "Acoustic Modeling with Deep Neural Networks Using Raw Time Signal for LVCSR (PDF Download Available)" . *ResearchGate*. Retrieved 2017-06-14.
55. Hochreiter, Sepp; Schmidhuber, Jürgen (1997-11-01). "Long Short-Term Memory". *Neural Computation*. **9** (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735 . ISSN 0899-7667 . PMID 9377276 .
56. Graves, Alex; Eck, Douglas; Beringer, Nicole; Schmidhuber, Jürgen (2003). "Biologically Plausible Speech Recognition with LSTM Neural Nets" (PDF). *1st Intl. Workshop on Biologically Inspired Approaches to Advanced Information Technology, Bio-ADIT 2004, Lausanne, Switzerland*. pp. 175–184.
57. Graves, Alex; Fernández, Santiago; Gomez, Faustino (2006). "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks". *Proceedings of the International Conference on Machine Learning, ICML 2006*: 369–376. CiteSeerX 10.1.1.75.6306 .
58. Santiago Fernandez, Alex Graves, and Jürgen Schmidhuber (2007). An application of recurrent neural networks to discriminative keyword spotting . Proceedings of ICANN (2), pp. 220–229.
59. Sak, Haşim; Senior, Andrew; Rao, Kanishka; Beaufays, Françoise; Schalkwyk, Johan (September 2015). "Google voice search: faster and more accurate" .
60. Hinton, Geoffrey E. (2007-10-01). "Learning multiple layers of representation" . *Trends in Cognitive Sciences*. **11** (10): 428–434. doi:10.1016/j.tics.2007.09.004 . ISSN 1364-6613 . PMID 17921042 .
61. Hinton, G. E.; Osindero, S.; Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets" (PDF). *Neural Computation*. **18** (7): 1527–1554. doi:10.1162/neco.2006.18.7.1527 . PMID 16764513 .
62. Bengio, Yoshua (2012). "Practical recommendations for gradient-based training of deep architectures". arXiv:1206.5533 [cs.LG].
63. G. E. Hinton., "Learning multiple layers of representation ," *Trends in Cognitive Sciences*, 11, pp. 428–434, 2007.
64. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; Kingsbury, B. (2012). "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups". *IEEE Signal Processing Magazine*. **29** (6): 82–97. doi:10.1109/msp.2012.2205597 .

65. Deng, Li; Hinton, Geoffrey; Kingsbury, Brian (1 May 2013). "New types of deep neural network learning for speech recognition and related applications: An overview". *Microsoft Research*. CiteSeerX 10.1.1.368.1123 – via research.microsoft.com.

66. Deng, Li; Li, Jinyu; Huang, Jui-Ting; Yao, Kaisheng; Yu, Dong; Seide, Frank; Seltzer, Michael; Zweig, Geoff; He, Xiaodong; Williams, Jason; Gong, Yifan; Acero, Alex (2013). "Recent advances in deep learning for speech research at Microsoft". *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 8604–8608. doi:10.1109/icassp.2013.6639345. ISBN 978-1-4799-0356-6.

67. Sak, Hasim; Senior, Andrew; Beaufays, Francoise (2014). "Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling" (PDF). Archived from the original (PDF) on 2018-04-24.

68. Li, Xiangang; Wu, Xihong (2014). "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition". arXiv:1410.4281 [cs.CL].

69. Zen, Heiga; Sak, Hasim (2015). "Unidirectional Long Short-Term Memory Recurrent Neural Network with Recurrent Output Layer for Low-Latency Speech Synthesis" (PDF). *Google.com*. ICASSP. pp. 4470–4474.

70. Deng, L.; Abdel-Hamid, O.; Yu, D. (2013). "A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion" (PDF). *Google.com*. ICASSP.

71. Sainath, Tara N.; Mohamed, Abdel-Rahman; Kingsbury, Brian; Ramabhadran, Bhuvana (2013). "Deep convolutional neural networks for LVCSR". *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 8614–8618. doi:10.1109/icassp.2013.6639347. ISBN 978-1-4799-0356-6.

72. Yann LeCun (2016). Slides on Deep Learning Online

73. NIPS Workshop: Deep Learning for Speech Recognition and Related Applications, Whistler, BC, Canada, Dec. 2009 (Organizers: Li Deng, Geoff Hinton, D. Yu).

74. Keynote talk: Recent Developments in Deep Neural Networks. ICASSP, 2013 (by Geoff Hinton).

75. D. Yu, L. Deng, G. Li, and F. Seide (2011). "Discriminative pretraining of deep neural networks," U.S. Patent Filing.

76. Deng, L.; Hinton, G.; Kingsbury, B. (2013). "New types of deep neural network learning for speech recognition and related applications: An overview (ICASSP)" (PDF).

77. Yu, D.; Deng, L. (2014). *Automatic Speech Recognition: A Deep Learning Approach (Publisher: Springer)*. ISBN 978-1-4471-5779-3.

78. "Deng receives prestigious IEEE Technical Achievement Award - Microsoft Research" . *Microsoft Research*. 3 December 2015.

79. Li, Deng (September 2014). "Keynote talk: 'Achievements and Challenges of Deep Learning - From Speech Analysis and Recognition To Language and Multimodal Processing' " . *Interspeech*.

80. Yu, D.; Deng, L. (2010). "Roles of Pre-Training and Fine-Tuning in Context-Dependent DBN-HMMs for Real-World Speech Recognition" . *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*.

81. Seide, F.; Li, G.; Yu, D. (2011). "Conversational speech transcription using context-dependent deep neural networks" . *Interspeech*.

82. Deng, Li; Li, Jinyu; Huang, Jui-Ting; Yao, Kaisheng; Yu, Dong; Seide, Frank; Seltzer, Mike; Zweig, Geoff; He, Xiaodong (2013-05-01). "Recent Advances in Deep Learning for Speech Research at Microsoft" . *Microsoft Research*.

83. "Nvidia CEO bets big on deep learning and VR" . Venture Beat. April 5, 2016.

84. "From not working to neural networking" . *The Economist*.

85. Oh, K.-S.; Jung, K. (2004). "GPU implementation of neural networks". *Pattern Recognition*. **37** (6): 1311–1314. doi:10.1016/j.patcog.2004.01.013 .

86. "A Survey of Techniques for Optimizing Deep Learning on GPUs ", S. Mittal and S. Vaishay, Journal of Systems Architecture, 2019

87. Chellapilla, K., Puri, S., and Simard, P. (2006). High performance convolutional neural networks for document processing. International Workshop on Frontiers in Handwriting Recognition.

88. Cireşan, Dan Claudiu; Meier, Ueli; Gambardella, Luca Maria; Schmidhuber, Jürgen (2010-09-21). "Deep, Big, Simple Neural Nets for Handwritten Digit Recognition". *Neural Computation*. **22** (12): 3207–3220. arXiv:1003.0358 . doi:10.1162/neco_a_00052 . ISSN 0899-7667 . PMID 20858131 .

89. Raina, Rajat; Madhavan, Anand; Ng, Andrew Y. (2009). "Large-scale Deep Unsupervised Learning Using Graphics Processors". *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09. New York, NY, USA: ACM: 873–880. CiteSeerX 10.1.1.154.372 . doi:10.1145/1553374.1553486 . ISBN 9781605585161.

90. Sze, Vivienne; Chen, Yu-Hsin; Yang, Tien-Ju; Emer, Joel (2017). "Efficient Processing of Deep Neural Networks: A Tutorial and Survey". arXiv:1703.09039 [cs.CV].

91. "Announcement of the winners of the Merck Molecular Activity Challenge" .

92. "Multi-task Neural Networks for QSAR Predictions | Data Science Association" . *www.datascienceassn.org*. Retrieved 2017-06-14.

93. "Toxicology in the 21st century Data Challenge"

94. "NCATS Announces Tox21 Data Challenge Winners" .

95. "Archived copy" . Archived from the original on 2015-02-28. Retrieved 2015-03-05.

96. Ciresan, D. C.; Meier, U.; Masci, J.; Gambardella, L. M.; Schmidhuber, J. (2011). "Flexible, High Performance Convolutional Neural Networks for Image Classification" (PDF). *International Joint Conference on Artificial Intelligence*. doi:10.5591/978-1-57735-516-8/ijcai11-210 .

97. Ciresan, Dan; Giusti, Alessandro; Gambardella, Luca M.; Schmidhuber, Juergen (2012). Pereira, F.; Burges, C. J. C.; Bottou, L.; Weinberger, K. Q. (eds.). *Advances in Neural Information Processing Systems 25* (PDF). Curran Associates, Inc. pp. 2843–2851.

98. Ciresan, D.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. (2013). "Mitosis Detection in Breast Cancer Histology Images using Deep Neural Networks". *Proceedings MICCAI*. Lecture Notes in Computer Science. **7908** (Pt 2): 411–418. doi:10.1007/978-3-642-40763-5_51 . ISBN 978-3-642-38708-1. PMID 24579167 .

99. "The Wolfram Language Image Identification Project" . *www.imageidentify.com*. Retrieved 2017-03-22.

100. Vinyals, Oriol; Toshev, Alexander; Bengio, Samy; Erhan, Dumitru (2014). "Show and Tell: A Neural Image Caption Generator". arXiv:1411.4555 [cs.CV]..

101. Fang, Hao; Gupta, Saurabh; Iandola, Forrest; Srivastava, Rupesh; Deng, Li; Dollár, Piotr; Gao, Jianfeng; He, Xiaodong; Mitchell, Margaret; Platt, John C; Lawrence Zitnick, C; Zweig, Geoffrey (2014). "From Captions to Visual Concepts and Back". arXiv:1411.4952 [cs.CV]..

102. Kiros, Ryan; Salakhutdinov, Ruslan; Zemel, Richard S (2014). "Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models". arXiv:1411.2539 [cs.LG]..

103. Zhong, Sheng-hua; Liu, Yan; Liu, Yang (2011). "Bilinear Deep Learning for Image Classification". *Proceedings of the 19th ACM International Conference on Multimedia*. MM '11. New York, NY, USA: ACM: 343–352. doi:10.1145/2072298.2072344 . ISBN 9781450306164.

104. "Why Deep Learning Is Suddenly Changing Your Life" . *Fortune*. 2016. Retrieved 13 April 2018.

105. Silver, David; Huang, Aja; Maddison, Chris J.; Guez, Arthur; Sifre, Laurent; Driessche, George van den; Schrittwieser, Julian; Antonoglou, Ioannis; Panneershelvam, Veda (January 2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*. **529** (7587): 484–489. Bibcode:2016Natur.529..484S . doi:10.1038/nature16961 . ISSN 1476-4687 . PMID 26819042 .

106. Szegedy, Christian; Toshev, Alexander; Erhan, Dumitru (2013). "Deep neural networks for object detection". *Advances in Neural Information Processing Systems*: 2553–2561.
107. Hof, Robert D. "Is Artificial Intelligence Finally Coming into Its Own?". *MIT Technology Review*. Retrieved 2018-07-10.
108. Gers, Felix A.; Schmidhuber, Jürgen (2001). "LSTM Recurrent Networks Learn Simple Context Free and Context Sensitive Languages". *IEEE Transactions on Neural Networks*. **12** (6): 1333–1340. doi:10.1109/72.963769. PMID 18249962.
109. Sutskever, L.; Vinyals, O.; Le, Q. (2014). "Sequence to Sequence Learning with Neural Networks" (PDF). *Proc. NIPS*. arXiv:1409.3215. Bibcode:2014arXiv1409.3215S.
110. Jozefowicz, Rafal; Vinyals, Oriol; Schuster, Mike; Shazeer, Noam; Wu, Yonghui (2016). "Exploring the Limits of Language Modeling". arXiv:1602.02410 [cs.CL].
111. Gillick, Dan; Brunk, Cliff; Vinyals, Oriol; Subramanya, Amarnag (2015). "Multilingual Language Processing from Bytes". arXiv:1512.00103 [cs.CL].
112. Mikolov, T.; et al. (2010). "Recurrent neural network based language model" (PDF). *Interspeech*.
113. "Learning Precise Timing with LSTM Recurrent Networks (PDF Download Available)". *ResearchGate*. Retrieved 2017-06-13.
114. LeCun, Y.; et al. (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*. **86** (11): 2278–2324. doi:10.1109/5.726791.
115. Bengio, Yoshua; Boulanger-Lewandowski, Nicolas; Pascanu, Razvan (2013). "Advances in optimizing recurrent networks". *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 8624–8628. arXiv:1212.0901. CiteSeerX 10.1.1.752.9151. doi:10.1109/icassp.2013.6639349. ISBN 978-1-4799-0356-6.
116. Dahl, G.; et al. (2013). "Improving DNNs for LVCSR using rectified linear units and dropout" (PDF). *ICASSP*.
117. "Data Augmentation - deeplearning.ai | Coursera". *Coursera*. Retrieved 2017-11-30.
118. Hinton, G. E. (2010). "A Practical Guide to Training Restricted Boltzmann Machines". *Tech. Rep. UTML TR 2010-003*.
119. You, Yang; Buluç, Aydın; Demmel, James (November 2017). "Scaling deep learning on GPU and knights landing clusters". *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '17*. SC '17, ACM. pp. 1–12. doi:10.1145/3126908.3126912. ISBN 9781450351140. Retrieved 5 March 2018.
120. Viebke, André; Memeti, Suejb; Pllana, Sabri; Abraham, Ajith (March 2017). "CHAOS: a parallelization scheme for training convolutional neural networks on Intel Xeon Phi". *The*

Journal of Supercomputing. **75**: 197–227. arXiv:1702.07908 . Bibcode:2017arXiv170207908V . doi:10.1007/s11227-017-1994-x .

121. Ting Qin, et al. "A learning algorithm of CMAC based on RLS." Neural Processing Letters 19.1 (2004): 49-61.
122. Ting Qin, et al. "Continuous CMAC-QRLS and its systolic array ." Neural Processing Letters 22.1 (2005): 1-16.
123. *TIMIT Acoustic-Phonetic Continuous Speech Corpus* Linguistic Data Consortium, Philadelphia.
124. Robinson, Tony (30 September 1991). "Several Improvements to a Recurrent Error Propagation Network Phone Recognition System". *Cambridge University Engineering Department Technical Report*. CUED/F-INFENG/TR82. doi:10.13140/RG.2.2.15418.90567 .
125. Abdel-Hamid, O.; et al. (2014). "Convolutional Neural Networks for Speech Recognition" . *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. **22** (10): 1533–1545. doi:10.1109/taslp.2014.2339736 .
126. Deng, L.; Platt, J. (2014). "Ensemble Deep Learning for Speech Recognition" (PDF). *Proc. Interspeech*.
127. Tóth, Laszló (2015). "Phone Recognition with Hierarchical Convolutional Deep Maxout Networks" (PDF). *EURASIP Journal on Audio, Speech, and Music Processing*. **2015**. doi:10.1186/s13636-015-0068-3 .
128. "How Skype Used AI to Build Its Amazing New Language Translator | WIRED" . *Wired*. 2014-12-17. Retrieved 2017-06-14.
129. Hannun, Awni; Case, Carl; Casper, Jared; Catanzaro, Bryan; Diamos, Greg; Elsen, Erich; Prenger, Ryan; Satheesh, Sanjeev; Sengupta, Shubho; Coates, Adam; Ng, Andrew Y (2014). "Deep Speech: Scaling up end-to-end speech recognition". arXiv:1412.5567 [cs.CL].
130. "Plenary presentation at ICASSP-2016" (PDF).
131. "MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges" . *yann.lecun.com*.
132. Cireşan, Dan; Meier, Ueli; Masci, Jonathan; Schmidhuber, Jürgen (August 2012). "Multi-column deep neural network for traffic sign classification". *Neural Networks*. Selected Papers from IJCNN 2011. **32**: 333–338. CiteSeerX 10.1.1.226.8219 . doi:10.1016/j.neunet.2012.02.023 . PMID 22386783 .
133. Nvidia Demos a Car Computer Trained with "Deep Learning" (2015-01-06), David Talbot, *MIT Technology Review*
134. G. W. Smith; Frederic Fol Leymarie (10 April 2017). "The Machine as Artist: An Introduction". *Arts*. **6** (4): 5. doi:10.3390/arts6020005 .

135. Blaise Agüera y Arcas (29 September 2017). "Art in the Age of Machine Intelligence". *Arts*. **6** (4): 18. doi:10.3390/arts6040018 .

136. Bengio, Yoshua; Ducharme, Réjean; Vincent, Pascal; Janvin, Christian (March 2003). "A Neural Probabilistic Language Model" . *J. Mach. Learn. Res.* **3**: 1137–1155. ISSN 1532-4435 .

137. Goldberg, Yoav; Levy, Omar (2014). "word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method". arXiv:1402.3722 [cs.CL].

138. Socher, Richard; Manning, Christopher. "Deep Learning for NLP" (PDF). Retrieved 26 October 2014.

139. Socher, Richard; Bauer, John; Manning, Christopher; Ng, Andrew (2013). "Parsing With Compositional Vector Grammars" (PDF). *Proceedings of the ACL 2013 Conference*.

140. Socher, Richard (2013). "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank" (PDF).

141. Shen, Yelong; He, Xiaodong; Gao, Jianfeng; Deng, Li; Mesnil, Gregoire (2014-11-01). "A Latent Semantic Model with Convolutional-Pooling Structure for Information Retrieval" . *Microsoft Research*.

142. Huang, Po-Sen; He, Xiaodong; Gao, Jianfeng; Deng, Li; Acero, Alex; Heck, Larry (2013-10-01). "Learning Deep Structured Semantic Models for Web Search using Clickthrough Data" . *Microsoft Research*.

143. Mesnil, G.; Dauphin, Y.; Yao, K.; Bengio, Y.; Deng, L.; Hakkani-Tur, D.; He, X.; Heck, L.; Tur, G.; Yu, D.; Zweig, G. (2015). "Using recurrent neural networks for slot filling in spoken language understanding". *IEEE Transactions on Audio, Speech, and Language Processing*. **23** (3): 530–539. doi:10.1109/taslp.2014.2383614 .

144. Gao, Jianfeng; He, Xiaodong; Yih, Scott Wen-tau; Deng, Li (2014-06-01). "Learning Continuous Phrase Representations for Translation Modeling" . *Microsoft Research*.

145. Brocardo, Marcelo Luiz; Traore, Issa; Woungang, Isaac; Obaidat, Mohammad S. (2017). "Authorship verification using deep belief network systems". *International Journal of Communication Systems*. **30** (12): e3259. doi:10.1002/dac.3259 .

146. "Deep Learning for Natural Language Processing: Theory and Practice (CIKM2014 Tutorial) - Microsoft Research" . *Microsoft Research*. Retrieved 2017-06-14.

147. Turovsky, Barak (November 15, 2016). "Found in translation: More accurate, fluent sentences in Google Translate" . *The Keyword Google Blog*. Retrieved March 23, 2017.

148. Schuster, Mike; Johnson, Melvin; Thorat, Nikhil (November 22, 2016). "Zero-Shot Translation with Google's Multilingual Neural Machine Translation System" . *Google Research Blog*. Retrieved March 23, 2017.

149. Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". *Neural Computation*. **9** (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276.

150. Felix A. Gers; Jürgen Schmidhuber; Fred Cummins (2000). "Learning to Forget: Continual Prediction with LSTM". *Neural Computation*. **12** (10): 2451–2471. CiteSeerX 10.1.1.55.5709. doi:10.1162/089976600300015015. PMID 11032042.

151. Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le, Quoc V; Norouzi, Mohammad; Macherey, Wolfgang; Krikun, Maxim; Cao, Yuan; Gao, Qin; Macherey, Klaus; Klingner, Jeff; Shah, Apurva; Johnson, Melvin; Liu, Xiaobing; Kaiser, Łukasz; Gouws, Stephan; Kato, Yoshikiyo; Kudo, Taku; Kazawa, Hideto; Stevens, Keith; Kurian, George; Patil, Nishant; Wang, Wei; Young, Cliff; Smith, Jason; Riesa, Jason; Rudnick, Alex; Vinyals, Oriol; Corrado, Greg; et al. (2016). "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation". arXiv:1609.08144 [cs.CL].

152. "An Infusion of AI Makes Google Translate More Powerful Than Ever." Cade Metz, WIRED, Date of Publication: 09.27.16. https://www.wired.com/2016/09/google-claims-ai-breakthrough-machine-translation/

153. Boitet, Christian; Blanchon, Hervé; Seligman, Mark; Bellynck, Valérie (2010). "MT on and for the Web" (PDF). Retrieved December 1, 2016.

154. Arrowsmith, J; Miller, P (2013). "Trial watch: Phase II and phase III attrition rates 2011-2012". *Nature Reviews Drug Discovery*. **12** (8): 569. doi:10.1038/nrd4090. PMID 23903212.

155. Verbist, B; Klambauer, G; Vervoort, L; Talloen, W; The Qstar, Consortium; Shkedy, Z; Thas, O; Bender, A; Göhlmann, H. W.; Hochreiter, S (2015). "Using transcriptomics to guide lead optimization in drug discovery projects: Lessons learned from the QSTAR project". *Drug Discovery Today*. **20** (5): 505–513. doi:10.1016/j.drudis.2014.12.014. PMID 25582842.

156. Wallach, Izhar; Dzamba, Michael; Heifets, Abraham (2015-10-09). "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery". arXiv:1510.02855 [cs.LG].

157. "Toronto startup has a faster way to discover effective medicines". *The Globe and Mail*. Retrieved 2015-11-09.

158. "Startup Harnesses Supercomputers to Seek Cures". *KQED Future of You*. Retrieved 2015-11-09.

159. "Toronto startup has a faster way to discover effective medicines".

160. Zhavoronkov, Alex (2019). "Deep learning enables rapid identification of potent DDR1 kinase inhibitors". *Nature Biotechnology*. **37** (9): 1038–1040. doi:10.1038/s41587-019-0224-x. PMID 31477924.

161. Gregory, Barber. "A Molecule Designed By AI Exhibits 'Druglike' Qualities". *Wired*.

162. Tkachenko, Yegor (April 8, 2015). "Autonomous CRM Control via CLV Approximation with Deep Reinforcement Learning in Discrete and Continuous Action Space". arXiv:1504.01840 [cs.LG].

163. van den Oord, Aaron; Dieleman, Sander; Schrauwen, Benjamin (2013). Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K. Q. (eds.). *Advances in Neural Information Processing Systems 26* (PDF). Curran Associates, Inc. pp. 2643–2651.

164. X.Y. Feng, H. Zhang, Y.J. Ren, P.H. Shang, Y. Zhu, Y.C. Liang, R.C. Guan, D. Xu, (2019), "The Deep Learning–Based Recommender System “Pubmender” for Choosing a Biomedical Publication Venue: Development and Validation Study", *Journal of Medical Internet Research*, 21 (5): e12957

165. Elkahky, Ali Mamdouh; Song, Yang; He, Xiaodong (2015-05-01). "A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems". *Microsoft Research*.

166. Chicco, Davide; Sadowski, Peter; Baldi, Pierre (1 January 2014). *Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions*. *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14*. ACM. pp. 533–540. doi:10.1145/2649387.2649442. hdl:11311/964622. ISBN 9781450328944.

167. Sathyanarayana, Aarti (2016-01-01). "Sleep Quality Prediction From Wearable Data Using Deep Learning". *JMIR mHealth and uHealth*. **4** (4): e125. doi:10.2196/mhealth.6562. PMC 5116102. PMID 27815231.

168. Choi, Edward; Schuetz, Andy; Stewart, Walter F.; Sun, Jimeng (2016-08-13). "Using recurrent neural network models for early detection of heart failure onset". *Journal of the American Medical Informatics Association*. **24** (2): 361–370. doi:10.1093/jamia/ocw112. ISSN 1067-5027. PMC 5391725. PMID 27521897.

169. "Deep Learning in Healthcare: Challenges and Opportunities". *Medium*. 2016-08-12. Retrieved 2018-04-10.

170. Litjens, Geert; Kooi, Thijs; Bejnordi, Babak Ehteshami; Setio, Arnaud Arindra Adiyoso; Ciompi, Francesco; Ghafoorian, Mohsen; van der Laak, Jeroen A.W.M.; van Ginneken, Bram; Sánchez, Clara I. (December 2017). "A survey on deep learning in medical image analysis". *Medical Image Analysis*. **42**: 60–88. arXiv:1702.05747. Bibcode:2017arXiv170205747L. doi:10.1016/j.media.2017.07.005. PMID 28778026.

171. Forslid, Gustav; Wieslander, Hakan; Bengtsson, Ewert; Wahlby, Carolina; Hirsch, Jan-Michael; Stark, Christina Runow; Sadanandan, Sajith Kecheril (2017). "Deep Convolutional Neural Networks for Detecting Cellular Changes Due to Malignancy". *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. pp. 82–89. doi:10.1109/ICCVW.2017.18. ISBN 9781538610343.

172. De, Shaunak; Maity, Abhishek; Goel, Vritti; Shitole, Sanjay; Bhattacharya, Avik (2017). "Predicting the popularity of instagram posts for a lifestyle magazine using deep learning". *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*. pp. 174–177. doi:10.1109/CSCITA.2017.8066548 . ISBN 978-1-5090-4381-1.

173. "Colorizing and Restoring Old Images with Deep Learning" . *FloydHub Blog*. 2018-11-13. Retrieved 2019-10-11.

174. Schmidt, Uwe; Roth, Stefan. *Shrinkage Fields for Effective Image Restoration* (PDF). Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on.

175. Czech, Tomasz. "Deep learning: the next frontier for money laundering detection" . *Global Banking and Finance Review*.

176. "Army researchers develop new algorithms to train robots" . *EurekAlert!*. Retrieved 2018-08-29.

177. Utgoff, P. E.; Stracuzzi, D. J. (2002). "Many-layered learning". *Neural Computation*. **14** (10): 2497–2529. doi:10.1162/08997660260293319 . PMID 12396572 .

178. Elman, Jeffrey L. (1998). *Rethinking Innateness: A Connectionist Perspective on Development* . MIT Press. ISBN 978-0-262-55030-7.

179. Shrager, J.; Johnson, MH (1996). "Dynamic plasticity influences the emergence of function in a simple cortical array". *Neural Networks*. **9** (7): 1119–1129. doi:10.1016/0893-6080(96)00033-0 . PMID 12662587 .

180. Quartz, SR; Sejnowski, TJ (1997). "The neural basis of cognitive development: A constructivist manifesto". *Behavioral and Brain Sciences*. **20** (4): 537–556. CiteSeerX 10.1.1.41.7854 . doi:10.1017/s0140525x97001581 . PMID 10097006 .

181. S. Blakeslee., "In brain's early growth, timetable may be critical," *The New York Times, Science Section*, pp. B5–B6, 1995.

182. Mazzoni, P.; Andersen, R. A.; Jordan, M. I. (1991-05-15). "A more biologically plausible learning rule for neural networks" . *Proceedings of the National Academy of Sciences*. **88** (10): 4433–4437. Bibcode:1991PNAS...88.4433M . doi:10.1073/pnas.88.10.4433 . ISSN 0027-8424 . PMC 51674 . PMID 1903542 .

183. O'Reilly, Randall C. (1996-07-01). "Biologically Plausible Error-Driven Learning Using Local Activation Differences: The Generalized Recirculation Algorithm". *Neural Computation*. **8** (5): 895–938. doi:10.1162/neco.1996.8.5.895 . ISSN 0899-7667 .

184. Testolin, Alberto; Zorzi, Marco (2016). "Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired

Neurocognitive Functions". *Frontiers in Computational Neuroscience*. **10**: 73. doi:10.3389/fncom.2016.00073. ISSN 1662-5188. PMC 4943066. PMID 27468262.

185. Testolin, Alberto; Stoianov, Ivilin; Zorzi, Marco (September 2017). "Letter perception emerges from unsupervised deep learning and recycling of natural image features". *Nature Human Behaviour*. **1** (9): 657–664. doi:10.1038/s41562-017-0186-2. ISSN 2397-3374. PMID 31024135.
186. Buesing, Lars; Bill, Johannes; Nessler, Bernhard; Maass, Wolfgang (2011-11-03). "Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons". *PLOS Computational Biology*. **7** (11): e1002211. Bibcode:2011PLSCB...7E2211B. doi:10.1371/journal.pcbi.1002211. ISSN 1553-7358. PMC 3207943. PMID 22096452.
187. Morel, Danielle; Singh, Chandan; Levy, William B. (2018-01-25). "Linearization of excitatory synaptic integration at no extra cost". *Journal of Computational Neuroscience*. **44** (2): 173–188. doi:10.1007/s10827-017-0673-5. ISSN 0929-5313. PMID 29372434.
188. Cash, S.; Yuste, R. (February 1999). "Linear summation of excitatory inputs by CA1 pyramidal neurons". *Neuron*. **22** (2): 383–394. doi:10.1016/s0896-6273(00)81098-3. ISSN 0896-6273. PMID 10069343.
189. Olshausen, B; Field, D (2004-08-01). "Sparse coding of sensory inputs". *Current Opinion in Neurobiology*. **14** (4): 481–487. doi:10.1016/j.conb.2004.07.007. ISSN 0959-4388. PMID 15321069.
190. Yamins, Daniel L K; DiCarlo, James J (March 2016). "Using goal-driven deep learning models to understand sensory cortex". *Nature Neuroscience*. **19** (3): 356–365. doi:10.1038/nn.4244. ISSN 1546-1726. PMID 26906502.
191. Zorzi, Marco; Testolin, Alberto (2018-02-19). "An emergentist perspective on the origin of number sense". *Phil. Trans. R. Soc. B*. **373** (1740): 20170043. doi:10.1098/rstb.2017.0043. ISSN 0962-8436. PMC 5784047. PMID 29292348.
192. Güçlü, Umut; van Gerven, Marcel A. J. (2015-07-08). "Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream". *Journal of Neuroscience*. **35** (27): 10005–10014. arXiv:1411.6422. doi:10.1523/jneurosci.5023-14.2015. PMC 6605414. PMID 26157000.
193. Metz, C. (12 December 2013). "Facebook's 'Deep Learning' Guru Reveals the Future of AI". *Wired*.
194. "Google AI algorithm masters ancient game of Go". *Nature News & Comment*. Retrieved 2016-01-30.
195. Silver, David; Huang, Aja; Maddison, Chris J.; Guez, Arthur; Sifre, Laurent; Driessche, George van den; Schrittwieser, Julian; Antonoglou, Ioannis; Panneershelvam, Veda; Lanctot, Marc;

Dieleman, Sander; Grewe, Dominik; Nham, John; Kalchbrenner, Nal; Sutskever, Ilya; Lillicrap, Timothy; Leach, Madeleine; Kavukcuoglu, Koray; Graepel, Thore; Hassabis, Demis (28 January 2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*. **529** (7587): 484–489. Bibcode:2016Natur.529..484S . doi:10.1038/nature16961 . ISSN 0028-0836 . PMID 26819042 .

196. "A Google DeepMind Algorithm Uses Deep Learning and More to Master the Game of Go | MIT Technology Review" . *MIT Technology Review*. Retrieved 2016-01-30.
197. "Blippar Demonstrates New Real-Time Augmented Reality App" . *TechCrunch*.
198. A.I. Researchers Leave Elon Musk Lab to Begin Robotics Start-Up
199. "TAMER: Training an Agent Manually via Evaluative Reinforcement - IEEE Conference Publication" . *ieeexplore.ieee.org*. Retrieved 2018-08-29.
200. "Talk to the Algorithms: AI Becomes a Faster Learner" . *governmentciomedia.com*. Retrieved 2018-08-29.
201. Marcus, Gary (2018-01-14). "In defense of skepticism about deep learning" . *Gary Marcus*. Retrieved 2018-10-11.
202. Knight, Will (2017-03-14). "DARPA is funding projects that will try to open up AI's black boxes" . *MIT Technology Review*. Retrieved 2017-11-02.
203. Marcus, Gary (November 25, 2012). "Is "Deep Learning" a Revolution in Artificial Intelligence?" . *The New Yorker*. Retrieved 2017-06-14.
204. Smith, G. W. (March 27, 2015). "Art and Artificial Intelligence" . ArtEnt. Archived from the original on June 25, 2017. Retrieved March 27, 2015.
205. Mellars, Paul (February 1, 2005). "The Impossible Coincidence: A Single-Species Model for the Origins of Modern Human Behavior in Europe" (PDF). Evolutionary Anthropology: Issues, News, and Reviews. Retrieved April 5, 2017.
206. Alexander Mordvintsev; Christopher Olah; Mike Tyka (June 17, 2015). "Inceptionism: Going Deeper into Neural Networks" . Google Research Blog. Retrieved June 20, 2015.
207. Alex Hern (June 18, 2015). "Yes, androids do dream of electric sheep" . *The Guardian*. Retrieved June 20, 2015.
208. Goertzel, Ben (2015). "Are there Deep Reasons Underlying the Pathologies of Today's Deep Learning Algorithms?" (PDF).
209. Nguyen, Anh; Yosinski, Jason; Clune, Jeff (2014). "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images". arXiv:1412.1897 [cs.CV].
210. Szegedy, Christian; Zaremba, Wojciech; Sutskever, Ilya; Bruna, Joan; Erhan, Dumitru; Goodfellow, Ian; Fergus, Rob (2013). "Intriguing properties of neural networks".

arXiv:1312.6199 [cs.CV].

211. Zhu, S.C.; Mumford, D. (2006). "A stochastic grammar of images". *Found. Trends Comput. Graph. Vis*. **2** (4): 259–362. CiteSeerX 10.1.1.681.2190. doi:10.1561/0600000018.
212. Miller, G. A., and N. Chomsky. "Pattern conception." Paper for Conference on pattern detection, University of Michigan. 1957.
213. Eisner, Jason. "Deep Learning of Recursive Structure: Grammar Induction".
214. "Hackers Have Already Started to Weaponize Artificial Intelligence". *Gizmodo*. Retrieved 2019-10-11.
215. "How hackers can force AI to make dumb mistakes". *The Daily Dot*. 2018-06-18. Retrieved 2019-10-11.
216. "AI Is Easy to Fool—Why That Needs to Change". *Singularity Hub*. 2017-10-10. Retrieved 2017-10-11.
217. Gibney, Elizabeth (2017). "The scientist who spots fake videos". *Nature*. doi:10.1038/nature.2017.22784.
218. Mühlhoff, Rainer (2019-11-06). "Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning". *New Media & Society*: 146144481988533. doi:10.1177/1461444819885334. ISSN 1461-4448.
219. "Facebook Can Now Find Your Face, Even When It's Not Tagged". *Wired*. ISSN 1059-1028. Retrieved 2019-11-22.

Further reading

- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). *Deep Learning*. MIT Press. ISBN 978-0-26203561-3, introductory textbook.