Deep learning and Its Development

Qianji Zhao\textsuperscript{1,a,∗, †}, Zequn Shang\textsuperscript{2,b,∗, †}

\textsuperscript{1}School of University of New South Wales, Sydney, New South Wales, 2032(post), Australia.

\textsuperscript{2}School of University of Shandong, Weihai city, Shandong province, 264209(post), China.

\textsuperscript{a}z5235513@ad.unsw.edu.au

\textsuperscript{b}201700800301@email.sdu.edu.cn

\textsuperscript{†}These authors contributed equally.

Abstract. Deep learning is a promising branch of machine learning. It is an algorithm that uses artificial neural networks as the architecture to characterize and learn data. In recent years, many companies, for example, Google, Microsoft and Baidu, have become interested in the field of deep learning and have set up many large-scale projects, such as Google’s Deepmind project, including alphago, which has achieved success in Go and e-sports. This article analyzes and summarizes each current research direction and approach of deep learning, with prospection about the future research direction and development of deep learning expounded. An overview of the three basic models of deep learning is given, namely multilayer perceptrons and perceptrons, convolutional neural networks and recurrent neural networks. The benefits and superiority of the deep learning algorithm are illustrated and compared with the conventional methodology used in the common applications. Further research on emerging types of convolutional neural networks and recurrent neural networks are introduced. An overview of the three basic models of deep learning is given, namely multilayer perceptrons and perceptrons, convolutional neural networks and recurrent neural networks. The current application of deep learning in various fields is summarized, such as artificial intelligence, computer vision and natural language processing applications, and some open problems for future research are also analyzed. Finally, the significance and purpose of deep learning are discussed.

1. Introduction

Artificial intelligence (AI) is a new technological science that studies and develops theories, methods, technologies, and application systems. AI is used to simulate, extend, and expand human intelligence, which becomes one of the most focused research directions and has been developed on a large scale all over the world. In recent years, many countries have begun to develop AI. In China, AI and related research are included in the government's "13th Five-Year Plan" [1], while in the United States, the government has included AI research into its research and development strategic planning initiative. For enterprises, especially those specializing in the Internet, AI is the field in which they must occupy a leading position. Some leading technology companies, such as Google, Microsoft, Facebook, Baidu, Tencent and Alibaba are undertaking many research projects and gradually increasing investment in this field. In the development and application of artificial intelligence in these companies, the best-known project is the Google's artificial intelligence Alphago, which has shown its amazing dominance by beating many top professional players in go and E-sports. With the research on artificial intelligence by
high-tech companies and governments, social attention is gradually deepening [2]. People from all walks of life have started to understand the large figure of applications of AI and this understanding has already started changing human and the society. Besides, it is obvious that artificial intelligence and thereby deep learning and machine learning are gradually percolating into various aspects of human daily life, and deep learning is now the most popular and important research direction in AI research.

Deep learning is an important research direction in AI and it has experienced a long history. Its earliest research can be traced back to 1943, when neuroscientist McCulloch and mathematician Pitts published the paper "Logical Calculations of Inner Thoughts in Neural Activities" and established neural networks and mathematical models, called MCP models [2]. In 1958, Rosenblatt proposed a neural network composed of two layers of neurons, called "perceptrons." This is the first time that MCP has been used for machine learning classification [3]. Perceptron has attracted a large number of scientists' interest in the study of artificial neural network, which is a milestone in the development of neural network. However, with the deepening of research, in 1969, Marvin Minsky, the father of AI, and Simon Pipert, the founder of logo language, jointly wrote a Book perceptron, in which they proved that single-layer perceptron could not solve the linear inseparable problem (such as exclusive OR problem). Because of this fatal defect and the lack of timely promotion of perceptron to multilayer neural network, in the 1970s, the artificial neural network entered the first cold winter, and people's research on neural network had stagnated for nearly 20 years. In 1986, Geoffrey Hinton invented the BP algorithm for multi-layer perceptrons [4] which perfectly solved the problem of nonlinear classification, so that the artificial neural network once again caused widespread concern; in 2006, he and his students proposed a further scheme for the above research, using unsupervised pre-training to initialize the weights and supervised training fine-tuning to solve the problem of gradient disappearance in deep network training [5], which has opened the wave of deep learning in academia and industry. In 2016, with Google's alphago, which is based on deep learning, beating Li Shishi, the world's top go player, with a score of 4:1, the popularity of deep learning has changed. Later, alphago played with many world-class go masters one after another and won all games. This also caused the world to pay attention to the deep learning technology again. In the next year, alphago 0, an upgraded version of alphago based on reinforcement learning algorithm, was born. It adopts the learning mode of "starting from scratch" and "self-taught", and easily beat the previous alphago with a score of 100-0. In addition to go, it is also proficient in chess and other board games, can be said to be a real "genius" of chess. In addition, in this year, deep learning algorithms have achieved remarkable results in medical, financial, art, driverless and other fields. Therefore, some experts regard 2017 as the year of the most rapid development of deep learning and even AI.

Deep learning is a sub-category of machine learning, which is a kind of representation learning and is introduced into machine learning to make it closer to the original goal of AI. The algorithms need to have the ability to allow the machine to learn the higher-level abstract representations of data, and provide it the avenue to automatically extract feature associated data from the raw data superset. Therefore, an external feature of deep learning is end-to-end training. That is to say, it is not to put together the parts to be debugged separately to form a system, but to train the whole system together after it is set up. Its goal is to enable the machine to have the ability of analysis and learning like human, and to recognize data such as words, images, and sounds. The information obtained in the learning process is of great help to the interpretation of data such as text, image, and sound. Deep learning is a complex machine learning algorithm, which has achieved much better results in speech and image recognition than previous related technologies. In the accumulation of deep learning ability, the ability of any deep learning model increases exponentially with the increase of depth [6]. Compared with other classical machine learning methods, the difference of deep learning lies in the inclusion of non-optimal solutions, the use of non-convex nonlinear optimization, and the courage to try unproven methods.

This article is based on the discussion of three basic deep learning models (multilayer perceptrons and perceptrons, convolutional neural networks, and recurrent neural networks), and further compared the current deep learning methods. Afterwards, the application of deep learning and its research
significance are discussed, and the possible development direction and application prospect of AI in the future are prospected.

2. Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of feed forward neural network which includes convolution calculations and has a deep structure. It is one of the representative algorithms of deep learning [7].

The earliest form of CNN is the Neocognitron model proposed by Fukushima and Miyake (1980). Neocognitron is a neural network with deep structure and one of the earliest deep learning algorithms proposed. Its hidden layer is composed of S-layer (Simple-layer) and C-layer (Complex-layer) alternately. The S-layer unit extracts image features in the receptive field, and the C-layer unit receives and responds to the same features returned by different receptive fields. Neocognitron's S-C layer combination can perform feature extraction and filtering, and partially realizes the functions of convolution layer and pooling layer in convolutional neural networks [8, 9].

The first CNN was the time delay network proposed by Alexander Waibel et al. in 1987, which was applied to speech recognition [10]. Yann LeCun also constructed a CNN applied to computer vision (CV) problems in 1989, which is the initial version of LeNet [11]. On the basis of LeNet, in 1998 Yann LeCun and his collaborators built a more complete CNN LeNet-5 and achieved success in the recognition of handwritten digits. LeNet-5 follows LeCun's learning strategy and adds a pooling layer to the original design to filter input features. LeNet-5 and its subsequent variants define the basic structure of modern convolutional neural networks. The convolutional-pooling layer that appears alternately in its construction is able to extract the translation invariant features of the input image [12].

After the deep learning theory was proposed in 2006, the representational learning capabilities of convolutional neural networks have attracted attention, and have been developed with the update of numerical computing equipment [13].

The operation of CNN is very similar to a standard neural network. However, the key difference is that each unit in the CNN layer is a two-dimensional (or high-dimensional) filter that is convolved with the input of the layer [14]. The structure of CNN includes input layer, Rectified Linear Units layer (ReLU layer), convolutional layer (CONV layer), pooling layer and fully connected layer (FC layer).

The input layer of the CNN can handle one to four-dimensional arrays according to its dimensionality. Because the gradient descent algorithm is used for learning, the input features of the CNN need to be normalized and de-averaged [15].

The core of the ReLU layer is the Rectified Linear Unit, which is an excitation function commonly used in artificial neural networks [16]. The form of the ramp function is:

$$f(x) = \max(0, x)$$

In the neural network, linear rectification is used as the activation function of the neuron, which defines the nonlinear output result of the neuron after the linear transformation. Therefore, for the input vector from the previous neural network, the output result of the neuron using the linear rectification activation function is:

$$f(x) = \max(0, w^T x + b)$$

The result of this output will be input to the next layer of neurons or as the output of the entire neural network.

Other variants similar to ReLU include Leaky ReLU (LReLU), Parametric ReLU (PReLU), Randomized ReLU (RReLU), and Exponential Linear Unit (ELU) [17].

The CONV layer is the core layer for constructing a convolutional neural network, which generates most of the calculations in the network. The function of the convolutional layer is to extract features from the input data, and it contains multiple convolution kernels to form a convolution kernel collective. Each element corresponds to a weight coefficient and a bias vector, similar to the neuron of a feedforward neural network. Each neuron in the convolutional layer is connected to multiple neurons in the area close to the previous layer, which is called the receptive field. When the convolution kernel is
working, it will scan the input features regularly, do matrix element multiplication and summation of the input features in the receptive field and superimpose the deviation amount:

$$Z^{t+1}(i, j) = [Z^t \otimes \theta^{t+1}](i, j) + b = \sum_{k=1}^{K} \sum_{j=1}^{f} \sum_{i=1}^{f} \left[ Z^t_{ij} (s_{ij} + x, s_{ij} + y) \theta^{t+1}_k (x, y) \right] + b$$

(3)

$$L_{t+1} = \frac{L_j + 2p - f}{s_0} + 1$$

(4)

The above formula takes a two-dimensional convolution kernel as an example. The working method of a one-dimensional or three-dimensional convolution kernel is similar to that [16].

The parameters of the CONV layer include the size of the convolution kernel, the step size and the padding. The three parameters which are the hyperparameters of the CNN together determine the size of the output feature map of the convolutional layer. The size of the convolution kernel can be specified as any value smaller than the size of the input image. The larger the convolution kernel, the more complex the input features that can be extracted. The convolution step length defines the distance between the position of the convolution kernel when it scans the feature map twice. When the convolution step is 1, the convolution kernel will scan the elements of the feature map one by one. When the step is n the next time Scan skip n-1 pixels. Padding can be divided into four categories according to the number of layers and purpose:

- **Valid padding:** No padding is used at all, and the convolution kernel is only allowed to access the position of the feature map that contains the complete receptive field. All pixels of the output are a function of the same number of pixels in the input.
- **Same/half padding:** Only make enough padding to keep the same size of feature maps of the output and input.
- **Full padding:** Make enough padding so that each pixel is accessed the same number of times in each direction.
- **Arbitrary padding:** between valid padding and full padding [18].

The pooling layer is periodically inserted between successive convolutional layers. Its function is to gradually reduce the spatial size of the data volume, so that the number of parameters in the network can be reduced, therefore, it reduces the consumption of computing resources and can effectively control over-fitting. The data processed in the pooling layer can extract key information by filtering redundant information without CNN changing its characteristics [19].

The FC layer in the is equivalent to the hidden layer in the traditional feedforward neural network. The FC layer is in the last part of the hidden layer of the CNN and only transmits signals to other fully connected layers. The feature map loses the spatial topology in the fully connected layer, which is expanded into a vector and passes the activation function [20].

The most common form of CNN is to put some CONV layers and ReLU layers together, followed by pooling layers, and then repeats these until the image is spatially reduced to a small enough size, and finally connects to the FC layer to get the output [21]. The most common CNN structure is as follows:

$$0 < X \leq 3$$

$$Y \geq 0$$

$$0 \leq Z < 3$$

(5)
One of the advantages of CNN is that it can effectively reduce the dimensionality of a picture with a large amount of data into a small amount of data. Before the advent of CNN technology, the amount of data that should be processed when processing a picture with many pixels, resulting in high cost and low efficiency. CNN can reduce many parameters to a small number of parameters, and then perform processing without changing the nature of the image [22]. At the same time, CNN can preserve image features. For example, when the position of the image is changed, the essence of the image does not change, but the data obtained by traditional processing methods will be quite different from the original data. CNN solves this problem; it preserves the image in a visual-like manner [17]. Therefore, it can accurately recognize images that undergo transformations which do not change their essence.

3. Recurrent Neural Network
Recurrent neural network (RNN) is a type of recurrent neural network, which takes sequence data as input, recursively in the evolution direction of the sequence, and all nodes (recurrent units) are connected in a chain [23]. Bidirectional recurrent neural network (Bi-RNN) and long short-term memory network (LSTM) are common recurrent neural networks.

In 1982, John Hopfield used binary nodes to establish a neural network with content-addressable memory capability, which named Hopfield neural network [24]. In 1986, Michael I. Jordan proposed the Jordan network base on the theory of parallel distributed processing. Each hidden layer node of the Jordan network is connected to a state unit to implement the delay input and uses the logistic function as the activation function [25]. In 1989, Ronald Williams and David Zipser proposed the Real-Time Recurrent Learning (RTRL). In 1990, Jeffrey Elman proposed the first fully connected RNN, which named Elman network. In 1997, M. Schuster and K. Paliwal proposed a bidirectional RNN (BRNN) which has a deep structure and conducted a speech recognition experiment on it. In this year, Hochreiter and Schmidhuber proposed a long short-term memory (LSTM) network [26]. And it has created accuracy records in multiple application fields. After the 21st century, with the advent of deep learning theory and the improvement of numerical computing capabilities, RNNs with higher complexity have begun to attract attention in natural language processing problems.
The formulas given above are the explain of the Figure 2. X is the vector representing the value of the input layer, S is the vector representing the value of the hidden layer, U is the weight matrix from the input layer to the hidden layer, O is the vector representing the value of the output layer, and V is the weight matrix from the hidden layer to the output layer. The weights matrix W is the value of the hidden layer in last time which is the weight of the input in this time.

RNN is prone to extreme non-linear behavior after the error gradient is back propagated after multiple time steps, including gradient vanishing and gradient explosion. In order to solve these problems, the methods of gradient truncation, regularization, layer normalization, reservoir computing, skip connection, leaky unit and gated unit can be used to solve the errors generated during deep learning.

Algorithmically, the early Simple Recurrent Network includes the Elman network and the Jordan network. The recursive method of Elman network is [28]:

\[
\begin{align*}
    h_t &= \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \\
    y_t &= \sigma_y(W_y h_t + b_y)
\end{align*}
\]

(8)

The recursive method of Jordan network is [29]:

\[
\begin{align*}
    h_t &= \sigma_h(W_h x_t + U_h y_{t-1} + b_h) \\
    y_t &= \sigma_y(W_y h_t + b_y)
\end{align*}
\]

(10)

LSTM and Gated Recurrent Unit networks (GRU) are RNN gating algorithms. They use BPTT and RTRL for learning. The recursive method of LSTM is:

\[
\begin{align*}
    h_t &= g_o f_h s_t \\
    s_t &= g_f s_{t-1} + g_i f_s (wh_{t-1} + u_X X_t + b) \\
    g_i &= \text{sigmoid}(w_i h_{t-1} + u_i X_t + b_i) \\
    g_f &= \text{sigmoid}(w_f h_{t-1} + u_f X_t + b_f) \\
    g_o &= \text{sigmoid}(w_o h_{t-1} + u_o X_t + b_o)
\end{align*}
\]

(12)
\(f_s\) and \(f_i\) are the activation functions of the system state and internal state, \(g\) is the gating updated with time step, which is essentially a feedforward neural network with the Sigmoid function as the activation function.

The recursive method of GRU is:

\[
h^t = g_u^{t-1}h^{t-1} + (1 - g_u^{t-1})f_h(wh^{t-1} + uX^t g_s^t + b)
\]

\[
g'_s = \text{sigmoid}(w_s h^{t-1} + u_s X^t + b_s)
\]

\[
g'_i = \text{sigmoid}(w_i h^{t-1} + u_i X^t + b_i)
\]

Compared with LSTM, GRU simplifies based on LSTM, ignoring gates with less contribution. Since its gate control does not form a self-loop, but directly recurs between system states, the update equation does not include the internal state \(s^t\) [29].

The depth algorithm of RNN is deepened from single-layer structure to multi-layer. The most common form is stacked recurrent neural network (SRNN). RNN algorithms with multi-layer structure can also be superimposed to further deepen learning.

RNN has memory, parameter sharing and Turing completeness, therefore it has certain advantages for learning nonlinear features of sequences. It can also be combined with CNN to process CV problems that include sequence input [30].

4. Perceptron and multi-layer perceptron

4.1. Perceptron

Perceptron is an early neural network model, which was proposed by American scholar F. Rosenblatt in 1957 [31]. The concept of learning was first introduced into perceptron, which makes the learning function of human brain simulate to a certain extent in the mathematics based on symbol processing, so it has attracted extensive attention.

Perceptron is a kind of simulated neuron structure in neural network, including input, output, weight, feed forward, activation function and so on. Single layer perceptron can simulate logic and, logic or, logic not and logic and not operations, but it can’t realize logic XOR [32].

The activation function of perceptron can be expressed as:

\[
y = f(x) = f(wx + b)
\]

Where \(x\) is the input, \(W\) is the weight vector corresponding to the input \(x\), \(b\) is the offset, and \(y\) is the expected calculation result; generally, the offset \(b\) is taken as a value of \(w\), and a corresponding analog input \(x\) is added to make \(x\) equal to 1 to simulate the expression, which is convenient for matrix operation.

The sum of the values in brackets is called feedforward operation of perceptron, that is, the result that has not entered the activation function for calculation denoted as logit or Logits.

A simple single layer perceptron can be shown as follows:
The limitation of the perceptron to linear indivisible problem determines that it has poor inductive ability, and usually needs long off-line learning to achieve the effect.

In order to realize some function of forward neural network model, it must be trained, let it learn to do gradually, and memorize the knowledge learned in the weight of the network. The weight of artificial neural network is not determined by calculation, but through the network training itself [32]. This is also the biggest difference between the artificial neural network and other methods in solving problems. With the help of computer, hundreds of times or even thousands of network weight training and adjustment process can be completed in a short time [32].

The training process of perceptron is as follows:

Under the action of input vector p, calculate the actual output a of the network, and compare with the corresponding target vector t, check whether a equals T, and then adjust the weight and deviation according to the learning rules with the error after comparison; recalculate the input of the network under the new weight, repeat the weight adjustment process until the output a of the network equals the target vector t or the training times reach to the training ends when the maximum value is set in advance.

If the network training is successful, the network after training can generate a set of expected output for each group of input vectors under the effect of network weight; if the network fails to achieve the goal of a = t under the given input vector p within the maximum training times, the new initial weight and deviation can be used and longer output can be adopted train the number of times, or analyze whether the problem to be solved belongs to the kind of problem that cannot be solved due to the limitation of perceptron itself [33].

The steps of perceptron design training can be summarized as follows:

1) For the problems to be solved, the input vector p and the target vector t are determined, and the dimensions of each vector and the number of neurons of the network structure are determined: R, s and Q;

2) Parameter initialization:
   a) The random non-zero initial value of weight vector w is given in (- L, 1);
   b) Give the maximum training cycle Max_epoch;

3) Network expression: calculate the output vector a according to the vector p and the latest weight vector w;

4) Check: check whether output vector a is the same as target vector t, if yes, or the training of maximum cycle times has been reached, otherwise, turn to 5);

5) Learning: adjust the weight vector according to the learning rules of (4.5) perceptron and return to 3) [34].
4.2. Multi-layer Perceptron

In Minsky and Papert's monograph perceptron, it is analyzed that perceptron can only solve the so-called first-order predicate logic problems: and, or, etc., but not XOR, etc. Although a single perceptron cannot solve the XOR problem, it can realize the segmentation of complex space by combining multiple perceptrons [35]. As shown in the figure below:

![Multi-layer Perceptron Diagram]

The two perceptrons are combined according to a certain structure and coefficient. The first perceptron implements two linear classifiers and divides the feature space [36]. Then, a layer of perceptron is added to the output of the two perceptrons to realize XOR operation.

In other words, it is composed of multiple perceptrons:

\[ y = \theta(\sum_{j=1}^{m} v_j \theta(\sum_{i=1}^{n} w_{ij} x_i + w_{0j}) + v_0) \]  

To realize the nonlinear classification surface, where \( \theta(\cdot) \) represents a step function or a sign function.

In fact, the above model is the basic model of MLP neural networks. Each node in the neural network is a perceptron. The basic function of the neuron in the model biological neural network is: the electrical signal from the outside (environment or other cells) is transmitted to the neuron through the synapse [37]. When the total signal received by the cell exceeds a certain threshold, the cell is activated and sends the electrical signal to the next cell through the axon to complete the processing of the external information.

However, the learning algorithm of perceptron cannot be directly applied to the parameter learning of multi-layer perceptron model. Therefore, the original learning scheme is: in addition to the last neuron, all the weights of all other neurons are fixed in advance [37]. The learning process only uses perceptron learning algorithm to learn the weight coefficient of the last neuron. In fact, this is equivalent to transforming the original feature space into a new feature space by the first layer of neurons. Each neuron in the first layer forms one dimension of the new space, and then constructs a linear classifier by perceptron learning algorithm in the new feature space [38]. Obviously, because the weights of the first layer of neurons need to be given artificially, the performance of the model depends largely on whether the appropriate first layer neuron model can be designed, which depends on the understanding of the problems and data faced, and there is no method to solve the first layer of semimeric parameters for any problem.
5. Discussion

5.1. Image Processing
The image recognition deep learning represented by the high-precision AlexNet image recognition network proposed by Geoffrey Hinton in 2012 has achieved great breakthroughs [39]. The application of this model shows that performance improvement depends on the deep method, while the high computation causes the high cost. However, due to the usage of graphics processing units (GPU) in the training process, its calculation is feasible.

The current application and development of deep learning in image processing are mainly classified into three aspects: image transformation, image recognition and image generation. Image transformation refers to routine operations performed on pictures, including simple operations such as image scaling and copying, as well as common operations such as denoising and super-resolution enhancement. The purpose is to improve the quality of the picture to obtain an ideal target picture. As a common target in life, image has always been a research hotspot in CV direction [40]. CV is an important development direction in the field of deep learning. Its main content is target recognition. The usual method of using deep learning for image recognition is to build a neural network that recognizes images. Therefore, this network can achieve high accuracy and low computing resource consumption. Image generation refers to learning features from known images and then combining them. The generated image is a combination of all the learned image features. The Deep Dream algorithm developed by Google is an example of image generation [41].

5.2. Text Recognition
The most common application of text recognition is translation. The main methods of translation are text translation and speech translation. At present, there is also a way of text recognition by reading lip language. The most common translation software is Google Translate.

The core technology of Google Translate is "statistical machine translation". The basic idea is constructing a statistical translation model through statistical analysis of many parallel corpus. When Google Translate generates a translation, it would look for various models in many manually translated documents, make reasonable guesses, and obtain a more appropriate translation after a lot of learning. The basis for adopting the "statistical translation model" is Google's cloud computing architecture. Machine translation requires massive data storage space and efficient computing power. Google has Google MapReduce (distributed computing system) and BigTable (distributed storage system) to solve these two requirements respectively. It can be known from the principle that the more artificial translation documents which Google Translate can analyze for a specific language, the higher the quality of the translation [42].

5.3. Robotic
The deep learning is an active agent for the robot that can act and interact with the physical real world. It perceives the world through different sensors, builds a continuous world model, and updates this model over time, but ultimately the intelligent robot must make decisions, plan actions, and perform these actions to complete useful tasks. Robot vision has brought many research challenges from three aspects: learning, embodiedness, and understanding. Learning challenges include uncertainty estimation, unknown identification, incremental learning, class incremental learning, and active learning. The embodied challenges include understanding and using time and space embodiedness to improve perception, allowing robot vision to perform active vision, and targeted manipulation of the environment. Understanding challenges include understanding the semantics of objects and scenes, understanding the geometry of objects and scenes, and understanding the combination of the two. Deep learning can effectively deal with these challenges. In this regard, deep learning can also be applied to unmanned driving technology [43].
5.4. Limitation
The essence of deep learning is to vectorize the input. Each layer in the deep learning model would perform a simple geometric transformation on the data passing through it. The transformation changes according to the weight parameters of different layers, and these layers perform iterative updates based on the current execution level of the model, and finally one space is mapped to another space. However, this transformation has limitations. Extending the current deep learning technology by stacking more layers and using more training data can only surface alleviate some of the problems. Deep learning would be difficult to complete any program design which requires reasoning and the application of scientific methods. The current deep learning has a low degree of recognition of things. When two things are relatively similar, deep learning will confuse the judgment of two things. Deep learning is also far away from the aim which is similar to humans, it cannot obtain output from relatively simple information and generate a prediction of subsequent output. When the input has a small deviation, the output always has a huge difference if it is obtained through deep learning. So far, deep learning has used continuous geometric transformations to map space X to space Y successfully. However, in this process, a large amount of artificially annotated data needs to be given, which is far from realizing AI.

6. Conclusion
Deep learning is a new but rapidly developing research direction in the field of AI. In this review, the research status of AI technology and deep learning is summarized. Firstly, this paper reviews the development history of AI and deep learning, and summarizes their basic concepts, research status and related applications. After this part, it analyzes and summarizes the current research directions and methods of deep learning, and introduces the basic principles and applications of three important models, namely multilayer perceptron and perceptron, convolutional neural network and recurrent neural network. This paper also discusses three important application directions of deep learning, namely, the application of text recognition, image recognition and robot, and analyzes the limitations and shortcomings of deep learning, and puts forward feasible suggestions for these problems.

References
[1] Outline of the 13th Five-Year Plan for National Economic and Social Development of the People's Republic of China. http://www.gov.cn/xinwen/2016-03/17/content_5054992.htm
[2] McCulloch W S and Pitts W 1943 A logical calculus of the ideas immanent in nervous activity Bull. Math. Biophys. 5 115–33
[3] Rosenblatt F 1958 The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev.
[4] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back propagating errors. Nature 323 533–6
[5] Hinton G E and Salakhutdinov R R 2006 Reducing the dimensionality of data with neural networks Science 313 504–7
[6] Dauphin Y N, Pascanu R, Gulcehre C, Cho K, Ganguli S and Bengio Y 2014 Identifying and attacking the saddle point problem in high-dimensional non-convex optimization Adv. Neural Inf. Process. Syst. 4 2933–41
[7] Convolutional Neural Networks (LeNet) - DeepLearning 0.1 documentation. DeepLearning 0.1. LISA Lab. [31 August 2013].
[8] Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4), 193-202.
[9] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, 85-117.
[10] Waibel, A., 1987. Phoneme recognition using time-delay neural networks. Meeting of the Institute of Electrical, Information and Communication Engineers (IEICE). Tokyo, Japan.
[11] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4),
[12] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to
document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.
[13] Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural
networks. science, 313(5786), pp.504-507.
[14] Zhang, W., 1988. Shift-invariant pattern recognition neural network and its optical architecture.
In Proceedings of annual conference of the Japan Society of Applied Physics.
[15] Lek S and Park Y S 2008 Multilayer Perceptron Encyclopedia of Ecology, Five-Volume Set.
[16] Goodfellow, I., Bengio, Y., Courville, A.. Deep learning (Vol. 1). Cambridge: MIT press, 2016: 326-366.
[17] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang,
G. and Cai, J., 2015. Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108.
[18] Dumoulin, V. and Visin, F., 2016. A guide to convolution arithmetic for deep learning. arXiv
preprint arXiv:1603.07285.
[19] Estrach, J.B., Szlam A. and LeCun, Y., 2014. Signal recovery from pooling representations. In:
Proceedings of the International Conference on Machine Learning (ICML), pp. 307–315.
[20] Krizhevsky A, Sutskever I and Geoffrey E H 2012 Imagenet Adv. Neural Inf. Process. Syst. 25.
[21] Vamsi Madhav Kota, V Manoj Kumar, C Bharatiraja, Deep Learning-A Review. IOP Conf. Series:
Materials Science and Engineering 912 (2020) 032068.
[22] Zhang, W., 1988. Shift-invariant pattern recognition neural network and its optical architecture.
In Proceedings of annual conference of the Japan Society of Applied Physics.
[23] Chollet, F. (2017). Keras: The python deep learning library. https://keras.io/, Accessed 5
November 2017.
[24] ‘Albert Einstein World Award of Science 2005’. Archived from the original on October 23, 2013. 
Retrieved August 13, 2013.
[25] Jordan, Michael I. (1997-01-01). Serial Order: A Parallel Distributed Processing Approach.
Neural-Network Models of Cognition-Biobehavioral Foundations. Advances in Psychology. 
Neural-Network Models of Cognition. 121. pp. 471–495. doi:10.1016/s0166-4115(97)80111-2. 
ISBN 9780444819314.
[26] 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.
[27] Denny Britz, Recurrent Neural Networks Tutorial. September 17, 2015. 
http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-
rnns/
[28] Elman, Jeffrey L. (1990). Finding Structure in Time. Cognitive Science. 14 (2): 179–211. 
doi:10.1016/0364-0213(90)90002-E.
[29] Gers, F.A. and Schmidhuber, J., 2000. Recurrent nets that time and count. In International Joint
Conference on Neural Networks, IJCNN (p. 3189). IEEE.
[30] He Liquan, Zhan Yongping. Research on perceptron neural network model [J]. Journal of Jiujiang
University (NATURAL SCIENCE EDITION), 2014,29 (04): 37-39 + 43
[31] Liu Jianwei, Shen Fanglin, Luo xionglin. Research on perceptron learning algorithm [J]. 
Computer Engineering, 2010,36 (07): 190-192
[32] Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient
descent is difficult. IEEE transactions on neural networks, 5(2), 157-166.
[33] Kong Yueping, Dang Shuang, Zeng Jun, Gao Kai. Saddle recognition method based on
convolutional neural network and multilayer perceptron [J]. Mini computer system, 2021,42
(02): 409-413
[34] Zhou Feiyan, Jin Linpeng, Dong Jun. review of convolutional neural networks [J]. Acta computa
Sinica, 2017,40 (06): 1229-1251
[35] Yang Sibo, Bao Yuan, Luo Jinming, Wu Wei, Zhang Chao. Computational power of multi-layer
perceptron for multi class problems [J]. Journal of Computational Mathematics of colleges and universities, 2020,42 (03): 277-288

[36] Lim T Y Ratnam M M Khalid M A Automatic classification of weld defects using simulated data and an MLP neural network [J] INSIGHT,2007

[37] Khaled J. Assi; Kh Md Nahiduzzaman; Nedal T. Ratrout; Adel S. Aldosary Mode choice behavior of high school goers: Evaluating logistic regression and MLP neural networks [j] Case Studies on Transport Policy,2018

[38] Rui Feng, et al. "Analysis and accurate prediction of ambient PM 2.5 in China using Multi-layer Perceptron." Atmospheric Environment 232,(2020):. doi: 10.1016/j.atmosenv.2020.117534.

[39] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. ImageNet classification with deep convolutional neural networks. Nips'12. Curran Associates Inc.: 1097-1105. 3 December 2012.

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

[41] Translations Made Simple: The Usefulness of Translation Apps. April 8, 2020. Archived from the original on April 29, 2020. Retrieved April 29, 2020.

[42] Moubarak, Paul M.; Ben-Tzvi, Pinhas (2011). Adaptive manipulation of a Hybrid Mechanism Mobile Robot. 2011 IEEE International Symposium on Robotic and Sensors Environments (ROSE). pp. 113–118. doi:10.1109/ROSE.2011.6058520. ISBN 978-1-4577-0819-0. S2CID 8659998.

[43] 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. S2CID 393948.