31st Youth Academic Annual Conference of Chinese Association of Automation
Wuhan, China; November 11-13, 2016

Overview of Deep Learning

Xuedan Du, Yinghao Cai, Shuo Wang, and Leijie Zhang
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, CAS
Beijing, China
Emails: duxuedan2015@ia.ac.cn; yinghao.cai@ia.ac.cn; shuo.wang@ia.ac.cn; zhangleijie2014@ia.ac.cn

***Abstract*—In recent years, deep learning has achieved great success in many fields, such as computer vision and natural language processing. Compared to traditional machine learning methods, deep learning has a strong learning ability and can make better use of datasets for feature extraction. Because of its practicability, deep learning becomes more and more popular for many researchers to do research works. In this paper, we mainly introduce some advanced neural networks of deep learning and their applications. Besides, we also discuss the limitations and prospects of deep learning.**

Keywords—deep learning; machine learning; neural network

I. INTRODUCTION

Deep learning was developed from artificial neural network, and now it is a prevalent field of machine learning. The research of artificial neural network began from 1940s. McCulloch *et al*. [1] proposed the McCulloch-Pitts (MP) model by analyzing and summarizing the characteristics of neurons. Hebb *et al*. [2] proposed a cell assembly theory to explain the adaptation of cerebral neuron during the learning process. This theory had an important influence on the development of neural networks. Then Rosenblatt *et al*. [3] invented the perceptron algorithm. This algorithm is a kind of binary classifier which belongs to supervised learning. Widrow proposed the adaptive linear element, and it is a single layer artificial neural network based on the MP model. Unfortunately, Minsky and Papert pointed that the perceptron algorithm had great limitations in theory and made a negative evaluation on the prospects of neural networks, which led the development of neural networks to hit a nadir. However, Hopfield *et al*. [4] proposed the Hopfield network in the early 1980s. This made artificial neural network revived. Then Hinton *et al.* [5] proposed the Boltzmann machine by using simulated annealing algorithm. In the 1990s, various shallow machine learning methods were proposed one after another, such as support vector machine [6], Boosting [7]. Due to the advantages of these methods both in theory and in application, artificial neural network hit a nadir again. After Hinton et al. put forward the concept of deep learning in the journal Science in 2006, artificial neural network once again received much interest from the research community.

Deep learning models usually adopt hierarchical structures to connect their layers. The output of a lower layer can be regarded as the input of a higher layer via simple linear or nonlinear calculations. These models can transform low-level features of the data into high-level abstract features. Owning to this characteristic, deep learning models can be stronger than shallow machine learning models in feature representation. The performance of traditional machine learning methods usually rely on users' experiences, while deep learning approaches rely on the data. Therefore, we can find out that deep learning approaches have reduced the demands for users. With the progress of computer technology, computers' performance is rapidly improved. Meanwhile, information on the Internet is also spewing out. These factors provide a strong impetus for deep learning to develop and make deep learning become the prevalent method in machine learning.

In this paper, we make a systematic introduction for deep learning from many aspects by expatiating its research progresses, state-of-the-art models, frameworks, and applications respectively. First, we introduce the research progresses in Section II. Then, we introduce several typical deep learning models in Section III and several deep learning frameworks in Section IV. Next, we list some applications of deep learning in Section V. Finally we conclude this paper in Section VI.

II. RESEARCH PROGRESSES

The concept of deep learning was put forward in 2006 at first. After that, deep learning is still continually developing at abroad. At present, there are many outstanding figures, such as Geoffrey Hinton, Yoshua Bengio, Yann LeCun and Andrew Ng. They are leading the research direction of deep learning. Some companies, like Google and Facebook, have made lots of research achievements in deep learning and applied them to various fields. In this year, Google's AlphaGo program defeated Lee Sedol in Go competition, which showed that deep learning had a strong learning ability. What's more, Google's DeepDream is an excellent software which can not only classify images but generate strange and artificial paintings based on its own knowledge. Facebook announced a new artificial intelligence system named Deep Text. Deep Text is a deep learning-based text understanding engine which can classify massive amounts of data, provide corresponding services after identifying users' chatting messages and clean up spam messages.

Deep learning started relatively late but developed very rapidly at home. There have achieved remarkable progress in colleges, universities, research institutes and companies. Baidu has established a deep learning institute to explore how to complete many a task with deep learning. Baidu's unmanned ground vehicle has accomplished road test under complicated road conditions. IFLYTEK started the research of speech recognition based on Deep Neural Network (DNN) in 2010. They launched the first online Chinese speech recognition system and an advanced technology to recognize different

This work is supported in part by the National Natural Science Foundation of China under Grants 61503381.

978-1-5090-4423-8/16/$31.00 ©2016 IEEE

languages. And now, they have published a high performance computing (HPC) platform in cooperation with Intel.

III. Deep Learning Models

From the beginning to the present, there are a lot of deep learning models. The typical models include Autoencoder (AE), Deep Belief Network (DBN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). In this section, we mainly introduce some state-of-the-art models.

A. Autoencoder

Autoencoder is mainly used to process complex high-dimensional data. Its aim is to learn how to represent a set of data via dimensionality reduction. When we process the input x by using a series of weighting and mapping methods, we can get the low-dimensional output y. Then we adopt the inverse weighting and mapping methods to make y transform to the output x' whose dimension is as the same as the input x. Now, all we have to do is to make the error function $L(x, x')$ be the smallest by training iteratively the network weights. The basic principle of AE is shown in Fig. 1.

AE also has many improved structures like Denoising Autoencoder [8], and Sparse Autoencoder [9]. For Denoising Autoencoder, it uses the original data with random noise to train network weights, which makes extracted features become more robust. For Sparse Autoencoder, besides increasing the number of hidden layers and neurons, Sparse Autoencoder limits the activation state of hidden nodes, which only a small number of hidden nodes are in the activated state and most of hidden nodes are in the unactivated state.

Xiong *et al*. [10] proposed a modified autoencoder network to recognize and separate anomalous ones from a set of geochemical samples. Continuous Restricted Boltzmann Machine (CRBM) is used as the part of the autoencoder network in [8]. The authors adopt three steps to train the model, which are pre-training CRBMs, unrolling CRBNs to construct the network and fine-tuning parameters via backpropagation. Finally, this approach achieves good results in recognizing multivariate geochemical anomalies.

Louizos *et al*. [9] proposed a variational fair autoencoder model, which could make latent representations maximally informative about observed random variables but minimally informative about sensitive or nuisance variables. In other words, the model could separate undesired factors from the variations while retaining as much information as possible from what remains. In order to remove sensitive or nuisance variables from latent representations, [9] added a penalty term based on Maximum Mean Discrepancy measure to the model. At last, they applied this model to some tasks and got great results.

B. Deep Belief Network

Deep Belief Network is a kind of neural network which is stacked by several restricted Boltzmann machines (RBMs). RBM is a kind of generative stochastic neural network models, which comes from the Boltzmann machine. Although RBM has inherited the two-layer neuron structure of the Boltzmann machine, there is no connection between neurons in the same layer with only the whole connection between the visual layer and the hidden layer. The basic structure of RBM is shown in

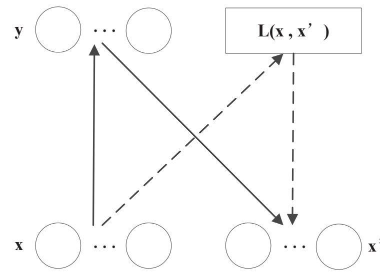


Fig. 1. The basic principle of Autoencoder.

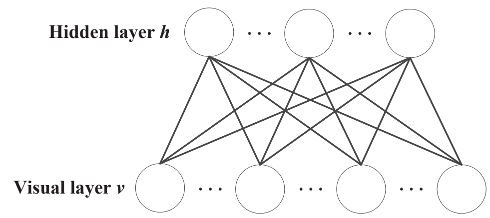


Fig. 2. The basic structure of restricted Boltzmann machine.

Fig. 2.

After increasing the number of the hidden layers of RBM, we can get deep Boltzmann machine. Then, we adopt a top-down directed connection near the visual layer so that we can get DBN model. When training the network, the greedy unsupervised layer-wise pre-training method can be used to get the network weights. It only trains one layer at a time with the output of the lower layer being used as the input of the higher layer. Then, back-propagation algorithm is used to fine-tune the whole network.

Liu *et al*. [11] proposed a novel Boosted Deep Belief Network (BDBN), which consists of several DBNs. Each DBN is used to learn hierarchical feature representations and all DBNs which are regarded as weak learners are connected together through a boosted classifier. BDBN adopts a bottom-up unsupervised feature learning (BU-UFL) process and a boosted top-down supervised feature strengthen (BTD-SFS) process. It is used to recognize facial expressions. The network divides facial images into many partially overlapped patches. And then it uses BU-UFL process to learn feature representation from each patch with one DBN and BTD-SFS process to fine-tune the features by processing classification errors produced by boosted classifier and weak learners. Finally, the model gets better results than other related work. The structure of this model is shown in Fig. 3.

Kim *et al*. [12] proposed a fingerprint liveness detection method which could distinguish whether a scanned fingerprint is live or fake prior to the recognition. The model in [12] uses a deep belief network. The structure of DBN is as the same as a normal DBN except for the last layer which has two output nodes to make liveness decision. Before inputting data to DBN, fingerprint images should be processed. The authors use the two dimensional Harris corner detector to infer the average location which contained the region of interest. The model is pre-trained by unsupervised learning with and fine-tuned with a training set of labeled inputs. Lastly, the experiments show that the proposed method can provide efficient and effective fingerprint liveness detection to prevent spoofing by fake fingerprints.

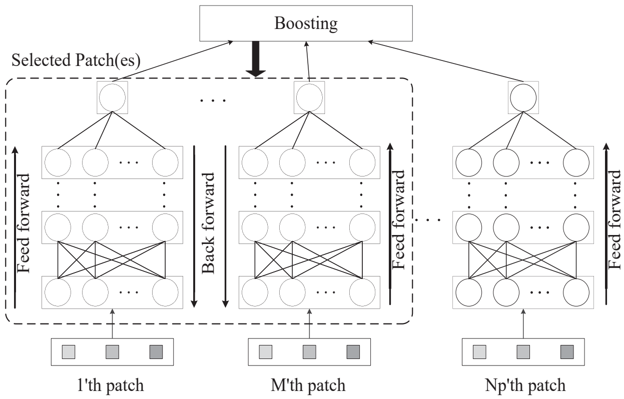


Fig. 3. The structure of the Boosted Deep Belief Network which consists of multiple deep belief networks. Only the ones in the dotted box will be fine-tuned jointly (Adapted from [11]).

C. Convolutional Neural Network

In 1960s, the concept of the receptive field was proposed. The neocognitron [13] based on the receptive field proposed in 1980s was considered as the predecessor of CNN. The remarkable characteristic of CNN is that the network uses the local receptive field and weight sharing. By using these two strategies, the number of training parameters is reduced significantly, which can make the network becomes less complicated. A typical CNN structure consists of some convolutional layers, pooling layers and fully-connected layers. The convolutional layer is used for feature extraction. Each input of the neuron in this layer is connected to a local receptive field of the previous one. The pooling layer is used for feature mapping. It can reduce the dimension of data and be able to maintain the invariance of the network structure.

In recent years, CNN has got lots of attentions from many researchers. It is an excellent model that can accomplish tasks efficiently. There are many types of CNN structures, such as LeNet [14], AlexNet [15], ZFNet [16], VGGNet [17] and GoogleNet [18]. LeCun et al. proposed a convolutional neural network namely LeNet, and applied to handwriting recognition. AlexNet is mainly used to object detections. After that, ZFNet, VGGNet and GoogleNet were put forward based on AlexNet. At present, CNN is still an active topic with many directions to explore. Some researchers want to increase the complexity of CNN structures. Others want to combine CNN with other traditional machine learnings.

Liu *et al*. [19] proposed an SSD model which could detect objects efficiently with high accuracy. The model consists of a truncated base network structure and auxiliary structure. The truncated base network in [19] adopts VGG-16, and the auxiliary structure adopts some feature layers to the end of VGG-16. The network can produce a set of fixed-size bounding boxes from many feature maps. It can also give category scores if there is an object in the bounding boxes and corresponding offsets. When training SSD, the loss function which is the weighted sum of the localization loss and confidence loss will be produced on forward propagation. Lastly, the loss function can be used to fine-tune the model on back propagation. The structure of this model is shown in Fig. 4.

Kontschieder *et al*. [20] proposed Deep Neural Decision Forests which is a novel structure that unifies classification trees with CNN. In their paper, the network structure is very clear that they replace the softmax layer with a stochastic and differentiable decision tree model. The decision tree is a kind of tree-structured classifier which consists of decision nodes and prediction nodes. The decision nodes decide the routes that how samples pass along the tree. The prediction nodes are to calculate their predictions. Finally, all predictions will be averaged separately and the samples can be judged that which class they belong to. On back propagation, they adopt minimum empirical risk principle to fine-tune the network. The structure of the tree model is shown in Fig. 5.

Levine *et al*. [21] proposed an approach which coordinates the hand and eye of a robot to grasp things from monocular images. The approach unifies reinforcement learning with deep learning. In their paper, authors use convolutional neural networks to make a prediction that whether the motion of the gripper could result in a successful grasp. The current images I_t and the original image I_0 are regarded as inputs of the network. It also provides the command vector v_t as input to the network after the two images are processed by the first five layers. Then they are inputted to the following set of layer so that the probability of a successful grasp.

D. Recurrent Neural Network

Recurrent Neural Network is a kind of artificial neural network. Apart from having the structure of the feedforward neural network, there exists directed cycles in RNN. This structure allows the information to be circulated in the network, so the output of each time is not only related to the input at present, but related to the input at previous timestamps.

Although the traditional RNN is able to deal with time series data, there exists a serious problem about gradient vanishing in the process of back propagation. Therefore RNN can only be used for short-term memory in most cases. In order to solve this problem, many researchers began to put forward several kinds of improved structures, such as Long Short-Term Memory (LSTM). Different from the traditional RNN, LSTM has a memory cell and an input-output gate structure. The memory cell is used to record information and the input-output gate determines whether the information is capable of flowing into or out of the memory cell. Due to these characteristics, LSTM has a better performance than RNN in long-term memory tasks.

Byeon *et al*. [22] proposed a completely learning-based approach for scene labeling using a kind of 2D LSTM recurrent neural networks. The network is divided into three main layers: input layer, hidden layer and output layer. The input image is split into several non-overlapping windows to the input network. The hidden layer consists of a 2D LSTM layer and feedforward layer. The 2D LSTM layer is used to memorize context information in all directions and the feedback layer combines information together. The output layer normalizes the outputs from the last hidden layer with a softmax function and generates probabilities about which classes the targets belong to. Experimental results show the effectiveness of the proposed model. The network is shown in Fig. 6.

Liu *et al*. [23] introduced a new approach to jointly learn feature representations across multiple related tasks. The novelty

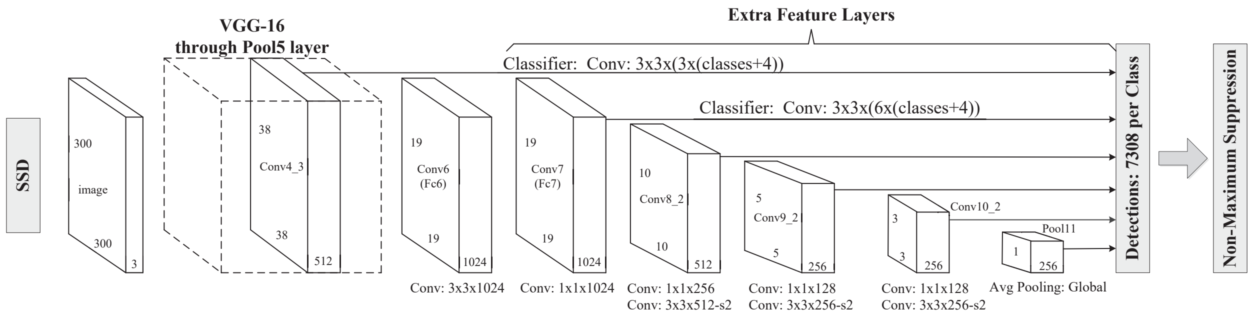


Fig. 4. The model in [19] adopts VGG-16 as the base network. Extra feature layers are added to the base network, so that [19] can use feature layers to produce detection predictions of every feature map (Adapted from [19]).

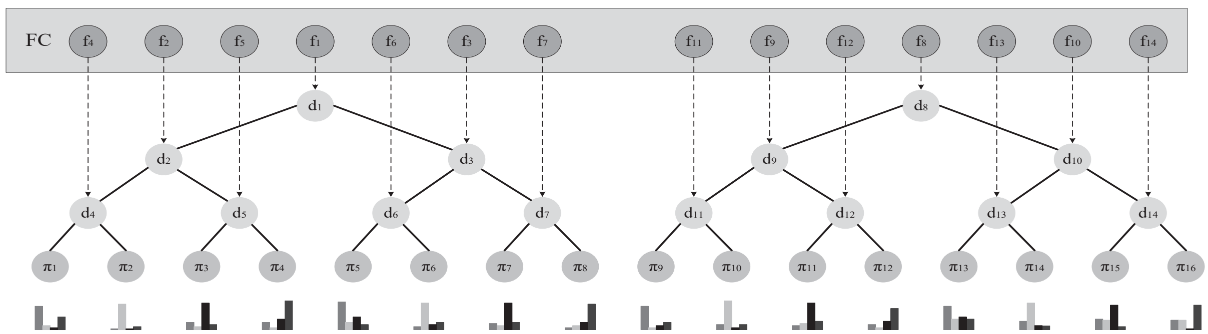


Fig. 5. The tree model of Deep Neural Decision Forests consists of several decision trees. The decision nodes decide how samples pass through the trees. The prediction nodes produce the decision probabilities after samples reach the corresponding final nodes (Adapted from [20]).

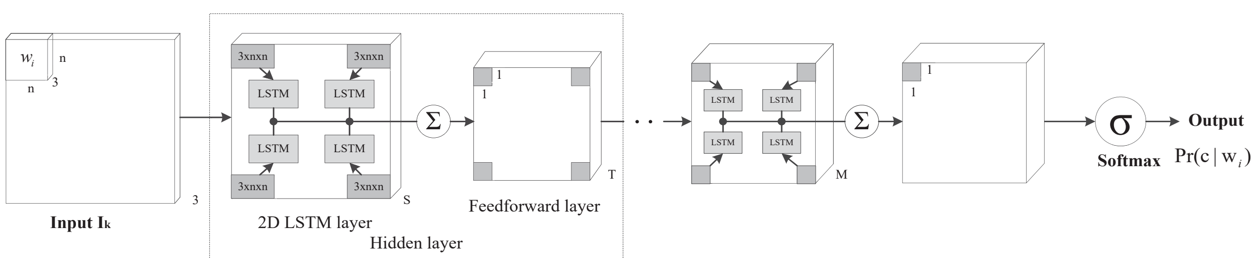


Fig. 6. The 2D long short-term memory network architecture. An input image Ik is divided into some windows. Each window is fed into four separate LSTM memory blocks. The output of each LSTM block is passed to the feedforward layer. At the last layer, the outputs of the final LSTM blocks are summed up and sent to the softmax layer. Finally, the networks output the class probabilities for each input window (Adapted from [22]).

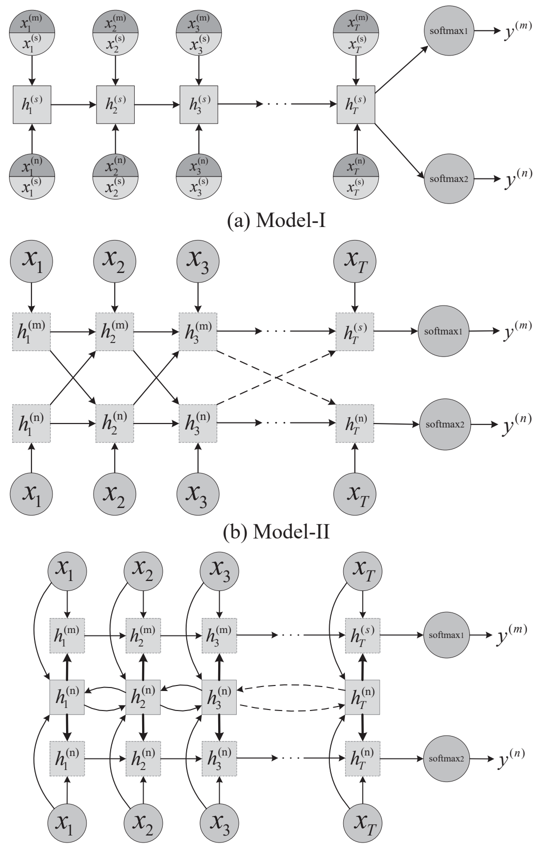
of this approach is that they integrate LSTM into the multi-task learning framework. In their paper, they proposed three architectures of sharing information to model text. The first one shares only a LSTM layer for all tasks. The second one assigns a LSTM layer for each task and every LSTM layer could use information from others. The last one not only has the characteristic of the second's, but also builds a bidirectional LSTM layer for all tasks. At the end of their paper, we can find that their models can improve the performances of the tasks they mentioned. The three architectures are shown in Fig. 7.

IV. Deep Learning Frameworks

Besides learning these models, we should also know about several deep learning frameworks and how to use them in different applications. The most commonly used deep learning frameworks include Caffe, TensorFlow, Torch, and Theano.

Caffe [24] is a kind of deep learning framework that suitable for CNN models based on several computing libraries like MKL, OpenBLAS and cuBLAS. Caffe provides a set of tools to be used for training, predicting, fine-tuning and so force. It also has many reference models and routines for learners to use. The configuration files of Caffe are simple to set up. And the Matlab and Python interfaces it provided are convenient to use. Compared to other frameworks, Caffe is easier to understand so that many beginners prefer to choose it.

TensorFlow [25] is a large-scale machine learning framework which provides an interface for machine learning algorithms to execute. It has been used in many fields, including speech recognition, computer vision, robotics, information retrieval, and natural language processing. Tensorflow is developed from DistBelief. It takes computations described using a dataflow-like model and maps them onto different hardware platforms such as such as Android and iOS. And it supports single-device, multi-device and distributed execution.



(a) Model-I

(b) Model-II

(c) Model-III

Fig. 7. Three architectures of the network for text classification with multi-task learning (Adapted from [23]).

Torch [26] can support most of the machine learning algorithms. It includes most popular algorithms and models such as multi-layer perceptrons, support vector machines, Gaussian mixture models, hidden Markov models, spatial and temporal convolutional neural networks, AdaBoost, Bayes classifiers and so on. Besides supporting CPU and GPU, Torch also can be embedded into iOS, Android, and FPGA.

Theano [27] is a framework based on Python. It can support some unsupervised and semi-supervised learning approaches as well as supervised learning approaches, such as logistic regression, multi-layer perceptron, deep CNN, AE, RBM, and DBN. Thanks to these functions, Theano is usually be used for teaching at aboard. However, Theano has a weakness that its speed is too slow.

V. Applications of Deep Learning

After we discuss these models and frameworks, we can find that deep learning approaches could help us to achieve performances in various applications. In this section, we introduce some applications of deep learning in computer vision and natural language processing.

Deep learning has had a wild development in computer vision, such as object detection, object tracking, and image segmentation. Object detection aims to recognize a class of objects from a large number of images. The traditional object detection methods mainly include candidate region selection, feature extraction and classification. This manual feature extraction method needs users to design what features they should extract. And these processes are often high-cost and time-consuming. Deep learning has the ability of unsupervised feature learning, and it can extract the features of images without any human intervention. Thus, it is gradually attracted more and more attention by researchers. After Krizhevsky *et al*. [15] got a breakthrough by using CNN in ImageNet LSVRC 2012, deep learning becomes more and more popular in computer vision and have made an excellent breakthrough up to now. So far, the method which conjuncts three residual Inception networks with one Inception-v4 [28] makes the image recognition task achieve 3.08% top-5 error. Learned-Miller *et al*. [29] proposed a deep learning method which made the accuracy of face recognition rise to about 87%. At present, the researchers in the Chinese University of Hong Kong have increased the face recognition accuracy above 99% [30].

Deep learning has been in continuous development in natural language processing and got many achievements in many applications, including speech recognition, speech synthesis and Question-Answering. The traditional speech recognition systems were mostly based on Gaussian Mixture Model and Hidden Markov Model in the past for a long time. However, these methods could not deal with deep characteristics well and are sensitive to disturbances from the outside environment. After adopting deep learning in speech recognition, the performances of the systems have improved dramatically. Now, the speech recognition system Deep Speech 2 which is designed by Baidu has reduced the error rate to 3.7% in Chinese speech test. At present, Google DeepMind published a new speech synthesis system which was named WaveNet [31]. WaveNet is a kind of deep neural network and can generate raw audio waveforms. Compared to other text-to-speech systems, WaveNet can generate more realistic sounds as well as music. From DeepMind, it showed that WaveNet reduced the gap between human and synthesized voices by over 50% in English and Chniese. Question-Answering (QA) is a hot research direction of natural language processing, which can give a correct and concise answer with the natural language form for natural language problems. The victory of Watson [32] on jeopardy has shown that QA based on deep learning has its own unique superiority.

VI. Conclusion

Deep learning approaches are practical for us to solve many problems. In this paper, we introduce deep learning models and frameworks in detail. Deep learning different kinds of models and frameworks, and it has had many applications in many aspects. From these, we can see that deep learning has a great development potential.

In future, it is foreseeable that deep learning could establish perfect theories to explain its performances. Meanwhile, its abilities of unsupervised learning will be enhanced since there are millions of data in the world but it is not applicable to add labels to all of them. It is also predicted that neural network structures will become more complex so that they can extract more semantically meaningful features. What's more, deep learning will combine with reinforcement learning better and we can use this advantages to accomplish more tasks.

REFERENCES

[1] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, pp. 115-133, 1943.

[2] D. O. Hebb, “The organization of behavior,” *J. Appl. Behav. Anal.*, vol. 25, pp. 575–577, 1949.

[3] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychol. Rev.*, vol. 65, no. 6, pp. 386-408, 1958.

[4] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *P. Natl. Acad. Sci. USA*, vol. 79, pp. 2554-8, 1982.

[5] D. H. Ackley, G. E. Hinton, and T.J. Sejnowski, “A learning algorithm for boltzmann machines,” *Cognitive Sci.*, vol. 9, pp. 147-169, 1985.

[6] C. Cortes and V. Vapnik , “Support-Vector Networks,” *Mach. Learn.*, vol. 20, pp. 273-297, 1995.

[7] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *J. comput. Syst. Sci.*, vol. 55, no. 1, pp. 119-139, 1997.

[8] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” *Proceedings of the 25th international conference on Machine learning*, *New York*, *pp. 1096-1103*, 2008.

[9] A. Ng, “Sparse autoencoder”, *CS294A Lecture Notes*, *Stanford Univ. California*, pp. 72, 2011.

[10] Y. Xiong and R. Zuo, “Recognition of geochemical anomalies using a deep autoencoder network,” *Comput. Geosci.-UK*, vol. 86, pp. 75-82, 2016.

[11] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial Expression Recognition via a Boosted Deep Belief Network,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *Columbus*, pp. 1805-1812, 2014.

[12] S. Kim, B. Park, B.S. Song, and S. Yang, “Deep belief network based statistical feature learning for fingerprint liveness detection,” *Pattern Recogn. Lett*., vol. 77, pp. 58-65, 2016.

[13] K. Fukushima, S. Miyake, and T. Ito, “Neocognitron: A neural network model for a mechanism of visual pattern recognition,” *IEEE Trans. Syst.*, *Man*, *Cybern.*, vol. 1, pp. 826-834, 1983

[14] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, pp. 2278-2324, 1998

[15] A. Krizhevsky, Sutskever, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Proc. Neural Information and Processing Systems*, pp. 1097-1105, 2012.

[16] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” *European Conference on Computer Vision*, *Zurich*, pp. 818-833, 2014.

[17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv*: 1409.1556, 2014

[18] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, “Going deeper with convolutions,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *Boston*, pp. 1-9, 2015.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed, “SSD: Single Shot MultiBox Detector,” *arXiv preprint arXiv: 1512.02325*, 2015

[20] P. Kontschieder, M. Fiterau, A. Criminisi, and S.R. Bulo, “Deep Neural Decision Forests,” *IEEE International Conference on Computer Vision*, *Santiago*, pp. 1467-1475, 2015.

[21] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, “Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection,” *arXiv preprint arXiv: 1603.02199*, 2016

[22] W. Byeon, T.M. Breuel, F. Raue, and M. Liwicki, “Scene labeling with LSTM recurrent neural networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *Boston*, pp. 3547-3555, 2015.

[23] P. Liu, X. Qiu, and X. Huang, “Recurrent Neural Network for Text Classification with Multi-Task Learning,” *arXiv preprint arXiv: 1605.05101*, 2016.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al*., “Caffe: Convolutional architecture for fast feature embedding,” *Proceedings of the 22nd ACM international conference on Multimedia*, *Orlando*, pp. 675-678, 2014.

[25] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. F. Chen, C. Citro, *et al*., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” *arXiv preprint arXiv: 1603.04467*, 2016.

[26] R. Collobert, S. Bengio, and J. Mariéthoz, “Torch: a modular machine learning software library,” *Idiap*, *No. EPFL-REPORT-82802*, 2002.

[27] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, *et al*., “Theano: A Python framework for fast computation of mathematical expressions,” *arXiv preprint arXiv: 1605.02688*, 2016.

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” *arXiv preprint arXiv: 1602.07261*, 2016.

[29] G. B. Huang, H. Lee, and E. Learned-Miller, “Learning hierarchical representations for face verification with convolutional deep belief networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Rhode Island*, pp. 2*518-2525, 2012.

[30] Y. Sun, X. Wang, and X. Tang, “Deeply learned face representations are sparse, selective, and robust,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *Boston*, pp. 2892-2900, 2015.

[31] A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A Graves, *et al*., “WaveNet: A Generative Model for Raw Audio,” *arXiv preprint arXiv: 1609.03499*, 2016.

[32] D. Ferrucci, A. Levas, S. Bagchi, D. Gondek, and E.T. Mueller, “Watson: Beyond Jeopardy!” *Artif. Intell.*, vol. 199-200, pp. 93-155, 2013.