Deep Learning Methods in Communication Systems: A Review

Feifan Liao, Shengyun Wei, Shun Zou
College of Information and Communication, National University of Defense Technology, Wuhan, China
liaofeifan@126.com; yumocloud@protonmail.ch; zoushun@nudt.edu.cn

Abstract—With the rapid development of modern communication systems, the amount of data has exploded, the system structure has become increasingly complex, and existing communication theories and technologies are facing huge challenges. The successful application of deep learning technology in the fields of images, speech, natural language processing, and games provides a possible solution for the theory and technology of communication systems that goes beyond traditional ideas and performance. This article mainly summarizes the application cases of deep learning methods in channel estimation, signal detection, and modulation recognition, and shows their outstanding performance compared to traditional communication theory and technology. Finally, we analyze the opportunities and challenges faced by deep learning-based communication technologies.

1. INTRODUCTION
Modern communication systems have performance requirements such as large data volumes, high transmission rates, and fast response speeds, which pose challenges to existing communication technologies. Especially in the processing and data mining of massive data, the existing theoretical framework has fundamental limitations. Therefore, many researchers have turned their attention to deep learning technology, and deep learning-based communication technologies have shown great potential in end-to-end communication systems, channel estimation, signal detection, modulation identification [1].

A general communication system framework is shown in the figure 1. It consists of three basic parts: transmitter, channel, and receiver. Starting from the source, the modules in the communication system appear in pairs, such as the source encoder and source decoder, channel encoder and signal decoder, modulator and demodulator. In order to optimize the performance of this modular design of the communication system, the general approach of researchers is to design a corresponding performance optimization algorithm for each module, but this method does not guarantee that the performance of the entire communication system is optimal. Therefore, the end-to-end communication system optimization scheme has become a research hotspot, and deep learning can provide a powerful solution for this end-to-end application demand. In addition, deep learning can effectively improve the utilization efficiency of massive data in communication systems through automated feature extraction.
In order to meet the new requirements in future wireless communication scenarios, this paper mainly studies the key technologies in the communication system and its process, summarizes the application of deep learning technology in communication systems, and provides references for the further development of deep learning technologies in communication systems.

2. DEEP LEARNING IN COMMUNICATION

Computer vision, speech, and natural language processing are the three main fields of artificial intelligence. Deep learning uses deep neural networks to simulate human pattern recognition process to perceive the outside world by feature extraction, recognition and classification on visual pattern. Compared with traditional computer vision methods and statistical machine learning methods, deep learning methods represented by convolutional neural network (CNN) show more powerful feature learning and representation capabilities, and are widely used in the field of computer vision. CNN is successfully applied to the entire end-to-end process of computer vision. The core mission of computer vision is the understanding of the environment, mainly including image classification, semantic segmentation, object detection, and pose estimation, etc.

Image classification is the most well-known computer vision task. Since AlexNet [2] got a breakthrough on ImageNet at 2012, the classification performance has been improving with network evolution, the state-of-the-art model EfficientNet [3] can achieve 88.4% top-1 accuracy on ImageNet. In addition, some networks are also applied to mobile and embedded vision applications by reducing network parameters without downgrading too much performance, such as MobileNets [4] and ShuffleNet [5]. Semantic segmentation is pixel-level prediction. FCN [5] could be marked as an initial breakthrough, DeepLabv3 [7] and HRNetV2 [8] achieve best performance on public dataset currently.

Object detection is the task of detecting instances of objects of a certain class within an image. The state-of-the-art methods can be categorized into two main types: one-stage methods and two stage-methods. One-stage methods aim at detection efficiency, and example models include YOLO [9], SSD [10] and RetinaNet [11]. Two-stage methods prioritize detection accuracy, and example models include Faster R-CNN [12], Mask R-CNN [13] and Cascade R-CNN [14]. Post estimation is the task of detecting the position and orientation of an object. DeepCut [15], Stacked Hourglass Networks [16] and Cascade Feature Aggregation [17] are the representative networks in this field. Cascade Feature Aggregation is the SOTA with 93.9% PCKH-0.5 on MPII Human Pose dataset. Deep learning is very successful in speech processing. With Resnet and BiLSTM [18], the percentage error of speech recognition on Switchboard + Hub500 could be reduced into 5.5%. Speech processing has gone into current automatic speech recognition pipelines [19] from traditional hand-engineered components. Research on natural language processing originates from Turing test. Combined with other machine learning methods, deep learning can not only recognize the semantics of natural language, but also understand the topic and context of the conversation, and the user's intentions and emotions in the conversation. XLNet [20] outperforms BERT [21] on 20 tasks on SQuAD dataset including question answering, natural language inference, sentiment analysis, and document ranking. Transformer model [22] dominates lots of machine translation leaderboard. Deep learning makes breakthrough at every branch of the natural language processing.
As can be seen, the applications of deep learning have seen rapid growth in nearly every in computer vision, speech and natural language processing. Thanks to the development of DL libraries and specialized hardware, the application of deep learning also extend towards communications field now. Communication process can be defined as reliably transmitting a message from a transmitter to a receiver over a channel. There are typically two approaches to integrate deep learning with communications: holistic approach, which treats communication as an end-to-end process, and phase-oriented approach, which only investigates the application of DL in certain phase of communication process, such as channel modeling, equalization, decoding, compression, demodulation, and modulation recognition, etc. These investigations provide us a theoretical alternative to achieve performance bounds than traditional communication theory, although, they are still far from practical application.

2.1 Holistic Approach
Message is encoded by transmitter and then decoded by receiver during communication, which is essentially equivalent to the autoencoder DL model [23-24]. The holistic approach [1,25,26] represents an end-to-end communications system as an autoencoder, which is designed with an coupling encoding and decoding module composed of convolutional layers and fully connected layers to extract signal features, and additional regularizing layers are also integrated to modeling the channel effects, such as noise, dropout, channel delay, frequency and phrase offset, and etc. Then the network parameters will be optimized via stochastic gradient descent algorithm without much expert knowledge in the field of communications. Continuous data transmission through real channel could also be achieve by introducing a frame synchronization module based on neural network model [25] and a two-phases training process. The learned communications system outperforms some practical baselines without more expert knowledge for hyperparameter tuning. This holistic approach shows us a brand new way to optimize our communication systems than the traditional communication practices, but more investigations should be done before practical application such as training on arbitrary unknown channels.

2.2 Phase-Oriented Approach
Phase-oriented approach investigates the application of DL in certain phase or aspect of communication process. Channel estimation, modulation, signal detection, communication prototype classification will be summarized as follows.

1) Channel Estimation
Traditional channel estimation first estimate the channel state information(CSI) explicitly and then recover the transmitted information from the estimated CSI, DL-based approach formulate channel estimation as a signal recovery problem. Message transmission over channel could be defined as a nonlinear transformation of input data related to noise, spatial response, time response and frequency response along the transmitter/receiver antenna space, and etc, therefore it could be modeled with artificial neural network according to the well-known universal approximation theorem [27]. Some DL-based approaches have been investigated in OFDM [28-29] and MIMO [30-32] channels with the simulated communication dataset. In [28], the channel in OFDM system is modeled with several concatenated DNNs of 5 dense layers to include the channel noise and simple time response. In [30] and [31], the channel estimation problem is transformed into image super-resolution problem, and the beamspace channel vector $h$ is estimated with dense layers or convolutional neural network structure to cover the spatial response along the transmitter/receiver antenna space. In [29], the time-frequency response of OFDM channel is considered as a two-dimensional image, and estimated with DNN on pilot locations. The above works give us a new idea about modeling nonlinear channel with DNN, however, these ideas are only investigated in simplified channel models with simulated dataset, more works should be carried out to include as much conditions as possible on real communication channel.

2) Signal Detection
Signal detection is a task that detects the presence of signals in a specific bandwidth. Because the frequency information of the carrier signal is unknown, and signals of different modulation types
usually appear in a certain frequency band and are transmitted at the same time, this poses a challenge for signal detection in a communication system with unstable noise environments. In order to realize the blind detection of Morse signals from broadband wireless spectrum data, a DeepMorse [33] system based on deep neural networks was proposed. The system consists of three parts: data collection, data processing, and Morse signal detection. The first two parts process the signals collected from real-world scenes into broadband wireless spectrum data, so that the problem is converted from processing one-dimensional time-series signals to two-dimensional spectrum diagrams, which is convenient for efficient processing using convolutional neural networks. Because these data usually involve different types of modulation without prior information to estimate the distribution coefficient of the signal frequency, DeepMorse uses a two-stage method based on an energy-based multi-signal sensing module plus a deep convolutional neural network. The multi-signal sensing module automatically locates the Morse signal from the broadband spectral data, and uses the localized data as the input of the deep convolutional neural network. The deep convolutional neural network uses a 2-unit CNN architecture to automatically learn spatial features from the spectrum map, so that the texture of the Morse signal is distinguished from other types of modulation. Compared with typical feature learning methods such as Support Vector Machine (SVM) and Deep Neural Networks (DNN), DeepMorse has obtained more robust and excellent detection performance. Unlike DeepMorse, the end-to-end signal detection method does not need to manually locate the signal [34], but directly uses deep learning-based target detection technology in computer vision. [34] uses a single-stage detection network Single shot multibox detector (SSD) to detect the position of the signal. This method is a single-stage detection method, which is faster than the two-stage method (such as RCNN, Fast RCNN). Detection speed, but detection performance is relatively low. This type of method is a method of supervised learning. Therefore, in the data set construction process, the algorithm performs STFT on the broadband, records the start and end time, the carrier frequency, and the bandwidth, and then converts them to the time-frequency image as the detection algorithm. Data labels.

3) Modulation Recognition

Modulation recognition is also called modulation classification. It refers to identifying the modulation mode of a signal after receiving the signal. Its essence is a type of pattern classification problem. The basic architecture of a traditional modulation recognition system consists of three parts: signal preprocessing, feature extraction, and signal classification. The signal preprocessing part includes carrier synchronization, abnormal signal removal, noise suppression, and parameter estimation. The feature extraction part mainly extracts the features that can characterize the modulation type of the signal through expert knowledge, mainly including the instantaneous amplitude, phase and frequency, high-order statistics [35-36], and cyclostationary characteristics [37] of the signal. The classification and recognition part is based on the extraction of feature parameters, selecting and determining appropriate decision rules and recognition classifiers.

Figure 2 A typical deep learning-based modulation recognition framework
Compared with traditional modulation recognition methods, the main differences between modulation recognition methods based on deep learning are feature extraction and classifier construction. Figure 2 is a typical deep learning-based modulation recognition framework. The core of the traditional method is how to extract the features that can characterize the type of signal modulation and build the corresponding classifier. The deep learning-based method pays more attention to the automatic extraction of features using deep neural networks, especially in end-to-end systems. Expert features need to be extracted from the input signal, and the feature extraction process is completely implemented by the network. When the receiver receives the modulated signal, the modulation recognition system based on deep learning usually completes the identification of the modulated signal through the following workflow. The first step is to preprocess the signal, which usually includes data normalization, denoising, and fixed-length sampling. Then complete the feature engineering of the input terminal. If it is an end-to-end system, directly drop the original IQ data as input, and in order to improve the performance of the algorithm, researchers usually further process the IQ data, such as extracting the high order cumulants of the signal. In the latest research, in order to make full use of the efficient feature extraction performance of the convolutional neural network in the image, the raw IQ data is replaced with the form of pictures and input into the convolutional neural network. These picture forms include constellation diagrams, eye diagrams, and vector diagrams, polar features, etc. After the input is determined, a more important step than deep learning is to build a deep neural network that matches the input data. These networks usually include convolutional neural networks, recurrent neural networks, and a combination of the two. The output of the deep neural network is the classification information of the modulated signal.

[38] treats the IQ waveform of the communication signal as a two-dimensional image-like data, and then inputs the data into a narrow two-dimensional CNN. The input of the network is 2 × 128 IQ data, and the network consists of two Convolutional layer and two fully connected layers. The data set used by the research team was generated using the GNU radio channel model and published at http://radioml.com. This algorithm is compared with existing classifiers based on expert features. The accuracy of modulation recognition is significantly higher than SVM, KNN, DT and other methods. In order to further analyze the performance of different deep learning models on the modulation recognition data set, [39] comparatively analyzed the classification performance on CNN, Convolutional Long short-term Deep Neural Networks (CLDNN), Inception and Resnet. The results show that CLDNN The structure performs better when the SNRs are higher than -8dB. However, this type of model that uses only raw data as input has a poor classification effect on more complex modulation methods (such as 16QAM, 64QAM). Therefore, [34] considers that traditional eye diagram and vector diagram are binary images. The original image is enhanced by exponential and logarithm operations, and the enhanced image is combined into three channels of the input image and sent to the CNN network. Compared with a model that uses IQ data, constellation diagrams, and cumulants as inputs alone, a multi-input network has better recognition performance. Similarly, reference [40] uses a cascaded CNN structure. The first CNN takes IQ samples as input, and the second CNN is used to distinguish between 16QAM and 64QAM. The experimental results show that QAM signals can be effectively classified even when the signal-to-noise ratio is relatively low. In [41-42], IQ data is converted into polar coordinate features, and then fed into the convolutional neural network. The results on multiple data sets show the efficiency and lightness of the algorithm.

3. OPPORTUNITIES AND CHALLENGES

With the successful application of deep learning technology in the fields of vision, speech, games and other fields, combined with the development of communication systems, we have summarized recent cases in which researchers have successfully applied deep learning technology to communication systems and their key technical links. Deep learning intelligent communication systems, channel estimation, signal detection, and modulation recognition, and show their outstanding performance compared to traditional communication theory and technology. This article does not attempt to comprehensively summarize the research results in this field, but starts from several typical key links of
communication systems, and demonstrates the research direction of the deepest development potential of this technology in communication systems.

With the rapid development of fifth-generation communication technology (5G), the explosive growth of data volume, and the increasing demand for intelligence, as a data-driven cutting-edge technology, deep learning can be widely applied to all aspects of 5G system design and optimization. Including combinatorial optimization, detection, identification, estimation and other issues. Moreover, from the existing research, deep learning-based methods provide a feasible solution that surpasses traditional ideas and performance.

At the same time, we must also clearly understand that the existing research is still in the initial exploration stage, and the use of deep learning to solve problems in communication systems is a long and challenging road. For example, the theoretical framework and general engineering system of a communication system based on deep learning have not been fully established, and a general-purpose, open-source, representative data set has not yet been constructed. However, based on the successful application cases listed in this article, we believe that deep learning-based communication technology will lead modern communication technology and theory into a new stage of development.

REFERENCES

[1] T. J. O’Shea, J. Hoydis, An introduction to machine learning communications systems, arXiv preprint arXiv:1702.00832.
[2] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097-1105.
[3] Q. Xie, E. Hovy, M.-T. Luong, Q. V. Le, Self-training with noisy student improves imagenet classification, arXiv preprint arXiv:1911.04252.
[4] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilnets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861.
[5] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6848-6856.
[6] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.
[7] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801-818.
[8] Y. Yuan, X. Chen, J. Wang, Object-contextual representations for semantic segmentation, arXiv preprint arXiv:1909.11065.
[9] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.
[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21-37.
[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Doll’ar, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.
[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[13] K. He, G. Gkioxari, P. Doll’ar, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.
[14] Y. Liu, Y. Wang, S. Wang, T. Liang, Q. Zhao, Z. Tang, H. Ling, Cbnet: A novel composite backbone network architecture for object detection, arXiv preprint arXiv:1909.03625.
[15] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, B. Schiele, Deepcut: Joint subset partition and labeling for multi person pose estimation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4929-4937.

[16] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: European conference on computer vision, Springer, 2016, pp. 483-499.

[17] Z. Su, M. Ye, G. Zhang, L. Dai, J. Sheng, Cascade feature aggregation for human pose estimation (2019). arXiv:1902.07837.

[18] G. Saon, G. Kurata, T. Sercu, K. Audhkhasi, S. Thomas, D. Dimitriadiis, X. Cui, B. Ramabhadran, M. Picheny, L.-L. Lim, et al., English conversational telephone speech recognition by humans and machines, arXiv preprint arXiv:1703.02136.

[19] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al., Deep speech 2: End-to-end speech recognition in english and mandarin, in: International conference on machine learning, 2016, pp. 173-182.

[20] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Advances in neural information processing systems, 2019, pp. 5754-5764.

[21] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805.

[22] S. Edunov, M. Ott, M. Auli, D. Grangier, Understanding back-translation at scale, arXiv preprint arXiv:1808.09381.

[23] A. Ng, et al., Sparse autoencoder, CS294A Lecture notes 72 (2011) (2011) 1-19.

[24] J. Masci, U. Meier, D. Cire¸san, J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in: International conference on artificial neural networks, Springer, 2011, pp. 52-59.

[25] S. Dörner, S. Cammerer, J. Hoydis, S. ten Brink,Deep learning based communication over the air, IEEE Journal of Selected Topics in Signal Processing 12 (1) (2017) 132-143.

[26] T. J. O’Shea, K. Karra, T. C. Clancy, Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention, in: 2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE, 2016, pp. 223-228.

[27] K. Hornik, M. Stinchcombe, H. White, et al., Multilayer feedforward networks are universal approximators., Neural networks 2 (5) (1989) 359-366.

[28] H. Ye, G. Y. Li, B.-H. Juang, Power of deep learning for channel estimation and signal detection in ofdm systems, IEEE Wireless Communications Letters 7 (1) (2017) 114-117.

[29] M. Soltani, V. Pourahmadi, A. Mirzaei, H. Sheikhzadeh, Deep learning-based channel estimation, IEEE Communications Letters 23 (4) (2019) 652-655.

[30] H. He, C.-K. Wen, S. Jin, G. Y. Li, Deep learning-based channel estimation for beamspace mmwave massive mimo systems, IEEE Wireless Communications Letters 7 (5) (2018) 852-855.

[31] H. Huang, J. Yang, H. Huang, Y. Song, G. Gui, Deep learning for super-resolution channel estimation and doa estimation based massive mimo system, IEEE Transactions on Vehicular Technology 67 (9) (2018) 8549-8560.

[32] D. Neumann, T. Wiese, W. Utschick, Learning the mmse channel estimator, IEEE Transactions on Signal Processing 66 (11) (2018) 2905-2917.

[33] Y. Yuan, Z. Sun, Z. Wei, K. Jia, Deepmorse: A deep convolutional learning method for blind morse signal detection in wideband wireless spectrum, IEEE Access 7 (2019) 80577-80587.

[34] X. Zha, H. Peng, X. Qin, G. Li, S. Yang, A deep learning framework for signal detection and modulation classification, Sensors 19 (18) (2019) 4042.

[35] A. Swami, B. M. Sadler, Hierarchical digital modulation classification using cumulants, IEEE Transactions on communications 48 (3) (2000) 416-429.
[36] A. Abdelmutalab, K. Assaleh, M. El-Tarhuni, Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers, Physical Communication 21 (2016) 10-18.

[37] L. Hong, Classification of bpsk and qpsk signals in fading environment using the ica technique, in: Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, 2005. SSST’05., IEEE, 2005, pp. 491-494.

[38] T. J. O Shea, J. Corgan, T. C. Clancy, Convolutional radio modulation recognition networks, in: International conference on engineering applications of neural networks, Springer, 2016, pp. 213-226.

[39] N. E. West, T. O’Shea, Deep architectures for modulation recognition, in: 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), IEEE, 2017, pp. 1-6.

[40] Y. Wang, M. Liu, J. Yang, G. Gui, Data-driven deep learning for automatic modulation recognition in cognitive radios, IEEE Transactions on Vehicular Technology 68 (4) (2019) 4074-4077.

[41] Teng C F, Liao C C, Chen C H, et al. Polar feature based deep architectures for automatic modulation classification considering channel fading[C]//2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2018: 554-558.

[42] Teng C F, Chou C Y, Chen C H, et al. Accumulated polar feature-based deep learning for efficient and lightweight automatic modulation classification with channel compensation mechanism[J]. arXiv, 2020: arXiv: 2001.01395.