Opening the black box of deep learning

Dian Lei , Xiaoxiao Chen , Jianfei Zhao
School of Mechatronics Engineering and Automation,
Shanghai University, Shanghai 200072, China

Abstract

The great success of deep learning shows that its technology contains profound truth, and understanding its internal mechanism not only has important implications for the development of its technology and effective application in various fields, but also provides meaningful insights into the understanding of human brain mechanism. At present, most of the theoretical research on deep learning is based on mathematics. This dissertation proposes that the neural network of deep learning is a physical system, examines deep learning from three different perspectives: microscopic, macroscopic, and physical world views, answers multiple theoretical puzzles in deep learning by using physics principles. For example, from the perspective of quantum mechanics and statistical physics, this dissertation presents the calculation methods for convolution calculation, pooling, normalization, and Restricted Boltzmann Machine, as well as the selection of cost functions, explains why deep learning must be deep, what characteristics are learned in deep learning, why Convolutional Neural Networks do not have to be trained layer by layer, and the limitations of deep learning, etc., and proposes the theoretical direction and basis for the further development of deep learning now and in the future. The brilliance of physics flashes in deep learning, we try to establish the deep learning technology based on the scientific theory of physics.

1 Introduction

Deep learning is the main representative of the breakthrough in artificial intelligence today, it has reached nearly human level in image classification [1], speech recognition [2], natural language processing [3] and so on. The method of deep learning is developing rapidly, which almost subverts all branches of computer vision field. However, the fundamental problem of deep learning at present is the lack of theoretical research on its internal principles, and there is no accepted theoretical explanation, namely, the so-called black box problem: Why use such a deep model in deep learning? Why is deep learning successful? What's the key inside? The lack of theoretical basis has led to the academic community being unable to explain the fundamental reason for the success of deep learning. The theoretical basis is not clear, and we simply do not know from what angle to look at it. The black box model is purely based on data without considering the physical laws of the model, it lacks the ability to adhere to mechanistic understandings of the underlying physical processes. Hence, even if the model achieves high accuracy but it lacks of theoretical support, it cannot be used as a basis for subsequent scientific developments [4]. We must not rely solely on intuition designed algorithmic structures and several empirically tried examples to prove the general validity of an algorithm. This research method has the potential to learn false modes from non-generic representations of data, the explanatory nature of the model is very low, and the resulting research results are difficult to pass on in the long term. As people's new ideas have been replaced by more and more complex model architectures, which are almost invisible under the weight of layers of models, calls for attention to the explanatory nature of machine learning are also getting higher. Therefore, we need to thoroughly understand the entire system operation of deep learning. We need to explain what the most fundamental problems are in the field of deep learning and whether these fundamental issues are mature enough to be accurately described in mathematical and physical languages.

The great success of deep learning shows that its technology contains profound truth, but the most widely understood way is mathematical analysis, so far, very little attention has been paid to its scientific issues. However, purely mathematical explanations may lead to misdirection. For example, the neural network is mathematically trying to approximate any function. In mathematics, it has been proved that a single-layer neural network can approximate any function if it is long enough, this viewpoint has greatly hindered the development of neural networks, this is why most people in the past neglected multi-layer networks for a long time and without studying in depth. Only a small number of people such as Yann LeCun, Geoffrey Hinton, and Yoshua Bengio still insist on research in multi-layer neural networks [5]. Therefore, from the great successes achieved in deep learning, it is far from enough to explain deep learning in mathematics, and the technology of deep learning needs to be based on scientific theory.

As deep learning has made breakthroughs in many aspects such as images, phonetics, and text, methods based on deep learning are increasingly being applied in various other fields, for example, recently effective in solving many-body quantum physics problems has also been proved. Therefore, the theory of deep learning methods must reflect some objective laws of the real world. obviously the most basic and universal theory is quantum physics and statistical physics. What is science? Physics is the most perfect science that has been developed so far. Just as most engineering disciplines are based on physics, the engineering foundation for deep learning now and in the future will be physics. We need to describe the deep learning concept model in the language of physics, so that we can scientifically guide the development and design of deep learning. From this we say that the key to the current and future success of artificial intelligence depends not only on the mathematical calculation, but also on the laws of physics. The theory of deep learning requires physics.

The data in the information world is divided into two different types of data: one is symbolic data, which is designated by our humans; the other is physical data, which objectively reflects the data of the real world, any actual data set we care about (whether it is a natural image or a voice signal, etc.) is physical data. Reference [6] shows that the reason why neural networks can perform classification tasks well is that these physical data x follow some simple physical laws and models can be generated with relatively few free parameters: for example, they may exhibit symmetry, locality, or a simple form as an exponent of a low-order polynomial; and symbolic data, such as ”variable y=cat” is specified by humans, in this case the symmetry or polynomial is meaningless, and they are not related to physics. However, the probability distribution of non-physical data y can be obtained by Bayes' theorem using the physical characteristics of x. In the reference [4], a Physics-guided Neural Networks (PGNN) is proposed, which combines the scientific knowledge of physics-based models with the deep learning. The PGNN leverages the output of physics-based model simulations along with observational features to generate predictions using a neural network architecture. Reference [7] shows that deep learning is intimately related to one of the most important and successful techniques in theoretical physics, the renormalization group (RG). Reference [8] using DBM and RBM to represent quantum many-body states illustrates why the depth of neural networks in the quantum world is so important, revealing the close relationship between deep neural networks and quantum many-body problems. Reference [9] establishes a mapping of tensor network (TN) based on quantum mechanics and neural network in deep learning. Reference [10] mentions that people have found more and more connections between basic physics and artificial intelligence, such as Restricted Boltzmann Machine and spin systems, deep neural networks and renormalization groups; the effectiveness of machine learning allows people to think about the deeper connection between physics and machine learning, and perhaps it can help us gain insights into intelligence and the nature of the universe.

The research of the above reference mainly takes the neural network as a computational tool, or as a method to solve the quantum many-body problem. This dissertation studies the artificial deep neural network as a real physical system, considers that the neural network model is a real physical model. The goal of deep learning training is to obtain the neural network system model which accords with the physical laws by the interaction or response between the neural network system and the input physical information. Because the deep neural network is a physical system, its trained model and its evolution in training must meet the laws of physics.

This dissertation analyzes the principles of physics embodied in deep learning from three different perspectives: microscopic, macroscopic, and world view, and describes deep learning with physics language, aiming to provide theoretical guidance and basis for further study and future development direction, and tries to establish the technology of deep learning based on the scientific theory of physics.

2 A microscopic view of deep learning

The biggest rule of the universe is that the world is made up of microscopic particles such as atoms, electrons and photons, which obey quantum mechanics. Quantum mechanics is the science of studying the motion law of the microscopic particles in the material world, so the neural network model of deep learning as a physical system requires that the model must be governed by quantum mechanics. The following explains deep learning from the basic principles of quantum mechanics.

The human brain neural network is composed of atoms, the number of billions of neurons is the same, and the computational methods of the human brain should be similar. The neural network, as an interactive quantum many-body system, determines the deep learning system to be described by the wave function. The coordinate operator, momentum operator (corresponding translation operator), angular momentum operator (corresponding rotation operator), energy operator, and spin operator in the neural network are the most basic and important physical quantity or mechanical quantity operator.

2.1 The physical meaning of neurons

Information has both physical and symbolic meanings, so neurons also have two meanings: 1) physical, 2) symbolic mappings. Now discuss the meaning of its physics. In this dissertation, the first hypothesis is that the neuron is the scattering source of the quasi-particle wave and the superposition of receiving the quasi-particle wave. First look at a classic physics experiment—Young's double slit experiment.

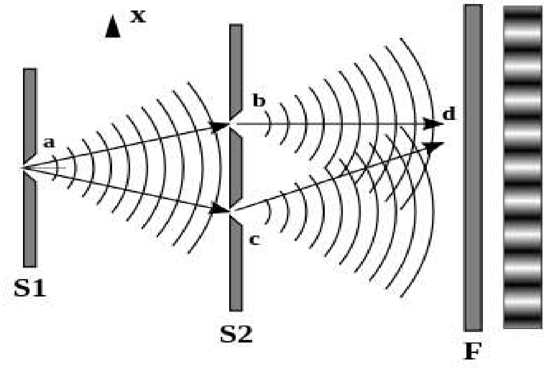

Figure 1: Young's double slit experiment.

As shown in Figure 1, the electrons are diffracted through the aperture a, and then diffracted and interfered by b and c. The bright diffraction fringes and patches at F indicate that there is a greater probability of electrons appearing there, and the dark part is there is little or no chance of the appearance of electrons. This dissertation holds that neurons act as electron diffraction interference. When we look at a neuron as a physical unit, the neuron is a scattering potential well that causes scattering of quasiparticles (perhaps this scattering originates from the quantum effect in the microtubules of the neurons, perhaps the electron-phonon coherence coupling in the biological system, perhaps some other kind of elementary excitation). Therefore, the output after the input of the neuron calculation is like the scattering output of the electrons through the circular hole, and the law is determined by the quantum physics theorem. The physical meaning of neurons indicates that, as white light can be scattered as red, orange, yellow, green, cyan, blue, violet, it is a natural classifier, calculator.

smoking and cancer, or to analyze the risks of a construction project, and so on. Can these mathematical models be extended to the microscopic world dominated by quantum mechanics? Can it be incorporated into the deep learning quantum physics model? Since quantum mechanics itself has many strange features, for example, if two or more quantum systems are entangled with each other, it is difficult to deduce whether the statistical correlation between them is causal.

The concept of causal information actually exceeds statistical relevance. For example, we can compare these two sentences: ”The number of cars is related to the amount of air pollution” and ”The car causes air pollution.” The first sentence is a statistical statement, and the latter sentence is a causal statement. Causality differs from relevance because it tells us how the system will change under intervention. In statistics, causal models can be used to extract causality from the empirical data of complex systems. However, there is only one component in a system of quantum physicsłthe wave function ψ , so mathematical models that use causality derived from statistical data cannot be applied, including Bayesian inference. John-Mark Allen of Oxford University in the United Kingdom proposed a generalized quantum causal model based on Reichenbach's principle of common reason [25], successfully combining causal intervention and Bayesian inference into a model.

5 Conclusions

At present, the research on the internal theory of deep learning is very scarce, and the successful application of deep learning and the limitations of its existing technology further illustrate the importance of studying its internal technical mechanism from a scientific perspective. Only knowing why to do it can transform existing methods or means from a deeper level. This is the scientific way of thinking. Based on the principles of physics, this paper interprets the deep learning techniques from three different perspectives: microscopic, macroscopic, and physical world perspectives. Inspired by the biological neural network, a new neuron physics model was proposed. Based on this, it explains the success of deep learning well, and fully reveals the internal mechanism of deep learning by scientific methods. A good theory can not only explain existing experiments, but also predict new phenomena and technologies. Therefore, this dissertation also proposes the direction of further research in deep learning. Some of the main conclusions of this paper are as follows:

(1) The deep neural network is a physical system, and its architecture and algorithm should conform to the principles of physics. The technical foundation for deep learning is physics, especially quantum physics and quantum statistical physics.

(2) In this dissertation, the physical meaning of neurons in deep neural network is proposed: its output value is the distribution of quasi-particles.

(3) Two physical models of deep neural networks are proposed in this paper, one is pure ensemble deep neural network and the other is hybrid ensemble deep neural network. The former learning model corresponds to a quantum measurement of a microscopic state, such as CNN; the latter corresponds to a microscopically statistically averaged macroscopic state, such as RBM.

(4) The physical model of neurons in CNN is a quantum superposition of a quasi-particle incident wave (Figure 1) and is excited by the output. This excitation may be the elastic scattering caused by the incident wave (exit only includes the incident wave), or inelastic scattering (exit also includes internal new excited states), or various possible actions such as chemical reactions (exit only includes new quasiparticles). It obeys quantum mechanics. According to the superposition principle of quantum mechanics, the excitation output of a neuron is related not only to the intensity of incident quasi-particles in other neurons, but also to their coherence, and to their polarization direction or spin.

(5) The input of the deep learning network is treated as a wave function, and the image is also a wave. The state of the neuron is also expressed by a wave function. If the measured neuron is the number or probability of excited quasi-particles, the value of a layer of neurons in the neural network is a probability distribution. The deep learning operator should be performed on the complex number field, but because the activation function is ReLU, the computational difference in the real number domain may not be large.

(6) Under such a physical model, the convolutional neural network algorithm is exactly the same as the quantum calculation method for measuring the number of quasi-particles ex-

cited by neurons, so that it can perfectly explain the technology of each important components of a convolutional neural network algorithm (convolution, rectification, activation, pooling, etc.) The purpose of convolution is to measure the number or probability of quasiparticles excited by each neuron. The convolution kernel is related to the Hamiltonian interaction potential of neurons. All neurons present an interference diffraction pattern - stripes and patches. That is, the deep convolutional neural network can measure the input wave vector or momentum, and its computational model can decompose white light into monochromatic light and decompose random-direction vibration into single-direction polarization. Therefore, the physical model of this paper can explain deep learning technology and success. This physical model shows that the deep convolutional neural network has natural learning ability and cognitive ability, and the model learns the ability to characterize the input micromechanical quantities, so it is reusable and can be applied across fields.

(7) The basis for parameter adjustment and optimization in the deep learning classification training process is the entropy in statistical physics, which is the number of microscopic states corresponding to the corresponding macroscopic state. Different types of training models should choose different cost functions according to the meaning of entropy, for example, the convolutional neural network model should use cross entropy as the objective function (cost function).

(8) A large number of operators, techniques, and methods in deep learning are related to the principles of physics such as energy, entropy, renormalization techniques, and translation operations; they are also related to physical world views such as symmetry, conjugacy, locality, hierarchy, etc.

The research in this paper shows that there are physics glimmers everywhere in deep learning. Deep learning techniques can be based on scientific theories. From the principles of physics, this dissertation presents the calculation methods for convolution calculation, pooling, normalization, and RBM, as well as the selection of cost functions, explains why deep learning must be deep, what characteristics are learned in deep learning, why convolutional neural networks do not have to be trained layer by layer, and the limitations of deep learning, etc. The physical model proposed in this paper can not only explain the successful technology of existing deep learning, but also predict many researchable directions and topics, such as positional neurons (these are in the research stage, and still need to be experimentally verified).

There is a striking homogeneity in the appearance and structure of the human cerebral cortex. In other words, the cerebral cortex uses the same calculations to accomplish all its functions. All the intelligence that humans exhibit (vision, auditory, physical movement...) are based on a unified set of algorithms. The deep learning technology is also based on a unified algorithm and is supported by physical theories. It will have a broad prospect for development.

References

[1] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898, 2014.

[2] David Imseng, Petr Motlicek, Philip N. Garner, and Herve Bourlard. Impact of deep mlp architecture on different acoustic modeling techniques for under-resourced speech recognition. In *Automatic Speech Recognition and Understanding*, pages 332–337, 2013.

[3] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *arXiv preprint arXiv:1708.02709*, 2017.

[4] Anuj Karpatne, William Watkins, Jordan Read, and Vipin Kumar. Physics-guided neural networks (pgnn): An application in lake temperature modeling. *arXiv preprint arXiv:1710.11431*, 2017.

[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[6] Henry W Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, 168(6):1223–1247, 2017.

[7] Pankaj Mehta and David J Schwab. An exact mapping between the variational renormalization group and deep learning. *arXiv preprint arXiv:1410.3831*, 2014.

[8] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature communications*, 8(1):662, 2017.

[9] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Bridging many-body quantum physics and deep learning via tensor networks. *arXiv preprint arXiv:1803.09780*, 2018.

[10] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[11] Ajit Narayanan and Tammy Menneer. Quantum artificial neural network architectures and components. *Information Sciences*, 128(3-4):231–255, 2000.

[12] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[13] Isma Hadji and Richard P Wildes. What do we understand about convolutional networks? *arXiv preprint arXiv:1803.08834*, 2018.

[14] Ramamurti Shankar. *Principles of quantum mechanics*. Springer Science & Business Media, 2012.

[15] Weiyang Liu, Zhen Liu, Zhiding Yu, Bo Dai, Rongmei Lin, Yisen Wang, James M Rehg, and Le Song. Decoupled networks. *arXiv preprint arXiv:1804.08071*, 2018.

[16] Avraham Ruderman, Neil Rabinowitz, Ari S Morcos, and Daniel Zoran. Learned deformation stability in convolutional neural networks. *arXiv preprint arXiv:1804.04438*, 2018.

[17] Francois Chollet. *Deep learning with Python*. Manning Publications Co., 2017.

[18] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted boltzmann machines and tensor network states. *Physical Review B*, 97(8):085104, 2018.

[19] Song Cheng, Jing Chen, and Lei Wang. Quantum entanglement: from quantum states of matter to deep learning. *Physics*, 2017.

[20] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[21] T Poggio, K Kawaguchi, Q Liao, B Miranda, L Rosasco, X Boix, J Hidary, and HN Mhaskar. Theory of deep learning iii: the non-overfitting puzzle. Technical report, CBMM memo 073, 2018.

[22] Shuo-Hui Li and Lei Wang. Neural network renormalization group. *arXiv preprint arXiv:1802.02840*, 2018.

[23] Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, page 1, 2018.

[24] Rohan Anil, Gabriel Pereyra, Alexandre Passos, Robert Ormandi, George E Dahl, and Geoffrey E Hinton. Large scale distributed neural network training through online distillation. *arXiv preprint arXiv:1804.03235*, 2018.

[25] John-Mark A Allen, Jonathan Barrett, Dominic C Horsman, Ciarán M Lee, and Robert W Spekkens. Quantum common causes and quantum causal models. *Physical Review X*, 7(3):031021, 2017.

[26] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep learning and quantum physics: A fundamental bridge. *arXiv preprint arXiv:1704.01552*, 2017.

[27] W Kinzel. Physics of neural networks. *Europhysics News*, 21(6):108–110, 1990.

[28] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[29] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep learning and quantum entanglement: Fundamental connections with implications to network design. *CoRR, abs/1704.01552*, 2017.

[30] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[31] IE Lagaris, A Likas, and DI Fotiadis. Artificial neural network methods in quantum mechanics. *Computer Physics Communications*, 104(1-3):1–14, 1997.

[32] Alaa Sagheer and Mohammed Zidan. Autonomous quantum perceptron neural network. *arXiv preprint arXiv:1312.4149*, 2013.

[33] Max Tegmark. Why the brain is probably not a quantum computer. *Information Sciences*, 128(3-4):155–179, 2000.

[34] G Perry, ET Rolls, and SM Stringer. Continuous transformation learning of translation invariant representations. *Experimental brain research*, 204(2):255–270, 2010.

[35] Giuseppe Carleo, Matthias Troyer, Giacomo Torlai, Roger Melko, Juan Carrasquilla, and Guglielmo Mazzola. Neural-network quantum states. *Bulletin of the American Physical Society*, 2018.

[36] Rongxin Xia and Sabre Kais. Quantum machine learning for electronic structure calculations. *arXiv preprint arXiv:1803.10296*, 2018.

[37] Rene Vidal, Joan Bruna, Raja Giryes, and Stefano Soatto. Mathematics of deep learning. *arXiv preprint arXiv:1712.04741*, 2017.

[38] IE Lagaris, A Likas, and DI Fotiadis. Artificial neural network methods in quantum mechanics. *Computer Physics Communications*, 104(1-3):1–14, 1997.

[39] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, page 1, 2018.

[40] MV Altaisky. Quantum neural network. *arXiv preprint quant-ph/0107012*, 2001.

[41] Judea Pearl. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*, 2018.

[42] Daniel J Buehrer. A mathematical framework for superintelligent machines. *arXiv preprint arXiv:1804.03301*, 2018.