



AI meets physics: a comprehensive survey

Licheng Jiao¹ · Xue Song¹ · Chao You¹ · Xu Liu¹ · Lingling Li¹ · Puhua Chen¹ ·
Xu Tang¹ · Zhixi Feng¹ · Fang Liu¹ · Yuwei Guo¹ · Shuyuan Yang¹ · Yangyang Li¹ ·
Xiangrong Zhang¹ · Wenping Ma¹ · Shuang Wang¹ · Jing Bai¹ · Biao Hou¹

Accepted: 22 July 2024 / Published online: 16 August 2024
© The Author(s) 2024

Abstract

Uncovering the mechanisms of physics is driving a new paradigm in artificial intelligence (AI) discovery. Today, physics has enabled us to understand the AI paradigm in a wide range of matter, energy, and space-time scales through data, knowledge, priors, and laws. At the same time, the AI paradigm also draws on and introduces the knowledge and laws of physics to promote its own development. Then this new paradigm of using physical science to inspire AI is the physical science of artificial intelligence (PhysicsScience4AI, PS4AI). Although AI has become the driving force for development in various fields, there is still a “black box” phenomenon that is difficult to explain in the field of AI deep learning. This article will briefly review the connection between relevant physics disciplines (classical mechanics, electromagnetism, statistical physics, quantum mechanics) and AI. It will focus on discussing the mechanisms of physics disciplines and how they inspire the AI deep learning paradigm, and briefly introduce some related work on how AI solves physics problems. PS4AI is a new research field. At the end of the article, we summarize the challenges facing the new physics-inspired AI paradigm and look forward to the next generation of artificial intelligence technology. This article aims to provide a brief review of research related to physics-inspired AI deep algorithms and to stimulate future research and exploration by elucidating recent advances in physics.

Keywords Artificial intelligence · Physics inspiration · Classical mechanics · Statistical physics · Electromagnetism · Quantum mechanics

1 Introduction

Artificial intelligence contains a wide range of algorithms (Yang et al. 2023; LeCun et al. 1998; Krizhevsky et al. 2012; He et al. 2016) and modeling tools (Sutskever et al. 2014) for large-scale data processing tasks. The emergence of massive data and deep neural networks provides elegant solutions in various fields. The academic community has also begun to

✉ Licheng Jiao
lchjiao@mail.xidian.edu.cn

¹ The Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China

explore the application of AI to various traditional disciplines. The objective is to promote the development of AI while further improving the possibilities of traditional analytical modeling (Hsieh 2009; Ivezić et al. 2019; Karpatne et al. 2017, 2018; Kutz 2017; Reichstein et al. 2019). Realizing general artificial intelligence is the goal that human beings have been pursuing. Although AI has made considerable progress in the past few decades, it is still difficult to achieve general machine intelligence and brain-like intelligence Jiao et al. (2016).

At present, researchers are beginning to explore the field of “AI + Physics” (Muther et al. 2023; Mehta et al. 2019). The objectives of current research are: (1) Utilise the findings of physical science and artificial intelligence to investigate the principles governing brain learning; (2) Utilise AI to facilitate the advancement of physics; (3) Apply physical science to inform the development of novel AI paradigms. We review relevant research on the intersection between AI and physical sciences in a selective manner. This includes the development of AI conceptual and algorithmic driven by physical insights, the application of artificial intelligence technology in multiple fields of physics, and the intersection between these two fields (Zdeborová 2020; Meng et al. 2022).

Physics As we all know, physics is a natural science that plays a heuristic role in the cognition of the objective world, focusing on the study of matter, energy, space, and time, especially their respective properties and the relationship between them. Broadly speaking, physics explores and analyzes the phenomena that occur in nature to understand its rules. Statistical Mechanics describes the theoretical progress made by neural networks in statistical physics Engel (2001). In the long history, physical knowledge (a priori) has been collected, verified, and integrated into practical theories. It is a simplified induction of the laws of nature and human behavior in many important disciplines and engineering applications. If the prior knowledge and AI are properly combined, more abundant and effective feature information can be extracted from the scarce data set, which helps to improve the generalization ability and interpretability of the network model Meng et al. (2022).

Artificial intelligence Artificial intelligence is a discipline that researches and develops theories and application systems for simulating and extending human brain intelligence. The purpose of artificial intelligence is to enable machines to simulate human intelligent behavior (such as learning, reasoning, thinking, planning, etc.) Widrow and Lehr (1990), so that machines have intelligence and complete “complex work”. Today, artificial intelligence is widely valued in the computer field, involving machine vision (Krizhevsky et al. 2012; Heisele et al. 2002), natural language processing Devlin et al. (2018), psychology (Rogers and McClelland 2004; Saxe et al. 2018) and pedagogy Piech et al. (2015) and other disciplines Khammash (2022) to model, is an interdisciplinary subject. The convergence of physical sciences and deep learning offers exciting prospects for theoretical science, providing valuable insights into the learning and computational capabilities of deep networks.

Relationship The development of physics is a simplified induction of nature, which promotes the research of brain-like science in artificial intelligence. And the brain that perceives any “experience” technology is close to the so-called “physical sense”, and physics opens new avenues and provides new tools for current artificial intelligence research McCormick (2022). To some extent, both artificial intelligence models and physical models can share information and predict the behavior of complex systems Tiberi et al. (2022), that is, they share certain methods and goals, but the implementation methods are different. Thus, physics should understand natural mechanisms, using prior knowledge Niyogi et al. (1998), regularity, and inductive reasoning to inform models, while model-agnostic AI should provide “intelligence” Werner (2013) through data extraction.

Main contributions Based on these analyses, this study aims to provide a comprehensive review and classification of the field of physics-inspired AI deep learning (Fig. 1) and summarize potential research directions and open questions that need to be addressed urgently in the future. The main contributions of this paper are summarized as follows:

1. Comprehensiveness and readability. This article comprehensively reviews over 400 physical science ideas in progress and physics-inspired deep learning AI algorithms. It also summarizes existing physics-inspired learning and modeling research from four aspects: classical mechanics, electromagnetism, statistical physics, and quantum mechanics.
2. Inspirational. The latest progress in artificial intelligence technology to solve physical science problems is summarized in the article. Finally, in the new generation of deep learning artificial intelligence algorithms, we analyzed the outlooks and implications between AI and physics.

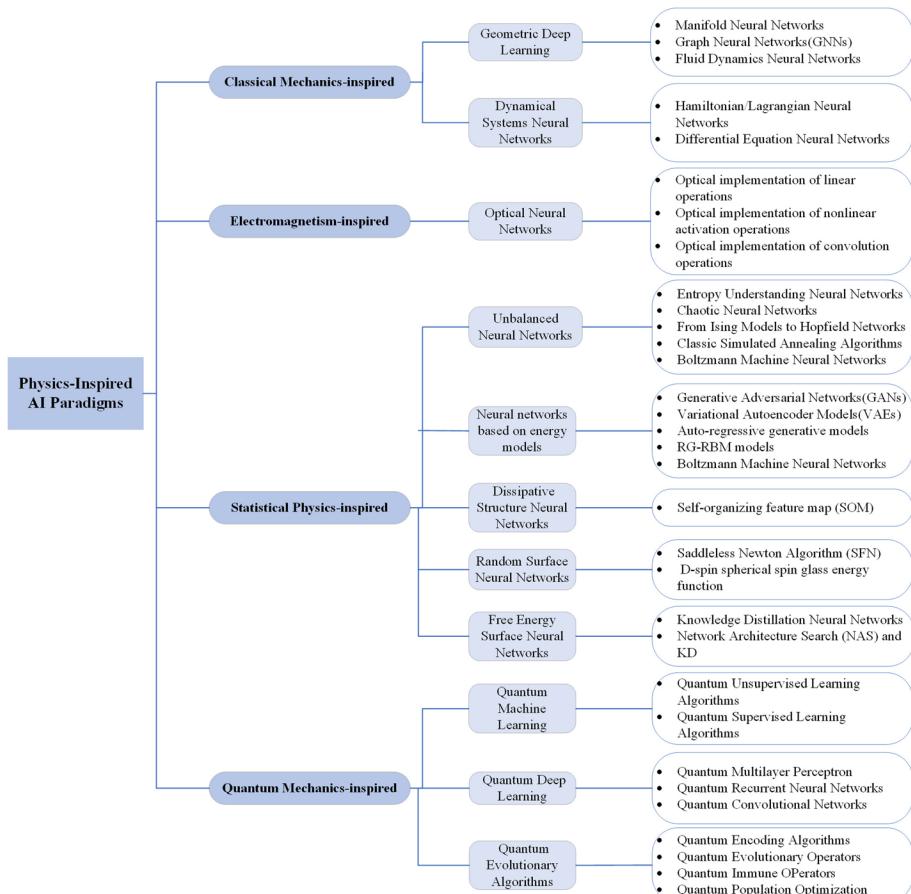


Fig. 1 This article presents an overall taxonomic structure for physics-inspired AI paradigms. We mainly outline the main contents of AI deep algorithms inspired by classical mechanics, electromagnetism, statistical physics and quantum mechanics

3. In-depth analysis. This article reviews open questions that need to be addressed to facilitate future research and exploration.

In this review, we attempt to provide a coherent review of the different intersections of deep learning artificial intelligence and physics. The rest of the paper is organized as follows: Chapter 2 presents artificial intelligence algorithms inspired by the perspective of classical mechanics and how AI can solve physical problems. Chapter 3 briefly reviews the electromagnetics-inspired artificial intelligence algorithms and the applications of AI in electromagnetics. Chapters 4 and 5 provide an overview of AI algorithms and applications inspired by statistical physics and quantum mechanics, respectively. Chapter 6 explores potential applications and challenges currently facing the intersection of AI and physics. Chapter 7 is the conclusion of this paper.

2 Deep neural network paradigms inspired by classical mechanics

In this section, we briefly introduce manifolds, graphs and fluid dynamics in geometric deep learning, as well as the basics of Hamiltonian/Lagrangian and differential equation solvers in dynamic neural network systems. Then it explains the related work inspired by it, and finally introduces the deep learning method of graph neural networks to solve physical problems. We summarize the structure of this section and an overview of representative methods in Table 1.

2.1 Geometric deep learning

Deep learning simulates the symmetry of the physical world (meaning the invariance of the laws of physics under various transformations). From the invariance of physical laws, an invariable physical quantity can be obtained, which is called a conserved quantity or invariant, and the universe follows translational/rotational symmetry (conservation of momentum). Momentum conservation is the embodiment of space uniformity (distortion degree), which is explained by mathematical group theory: space has translational symmetry—after the spatial translational transformation of an object, the physical motion trend and related physical laws remain unchanged. In the 20th century, Noether proposed Noether's theorem that every continuous symmetry corresponds to a conservation law, relevant expressions see references (Torres 2003, 2004; Frederico and Torres 2007) and references therein, related applications are shown in Fig. 2.

The translation invariance, locality, and compositionality of Convolutional Neural Networks (CNNs) make them naturally suitable for tasks dealing with Euclidean-structured data like images. However, there are still complex non-Euclidean data in the world, and Geometric Deep Learning (GDL) Gerken et al. (2023) emerges from this. From the perspective of symmetry and invariance, the design of deep learning framework in the case of non-traditional plane data (non-Euclidean data) structure is studied Michael (2017). The term was first proposed by Michael Bronstein in 2016, and GDL attempts to generalize (structured) deep networks to non-Euclidean domains such as graphs and manifolds. The data structure is shown in Fig. 3 .

Table 1 An overview of methods for AI DNNs inspired by classical mechanics

Categories	Detail	Typical Methods	Keywords	Publication & Years
Geometric deep learning	Manifold neural networks	GCNN Masci et al. (2015)	Feature descriptors; Non-Euclidean manifolds; ShapeNet	ICCVW 2015
		MoNet Monti et al. (2017)	Non-Euclidean; Graphs and manifolds; CNN	CVPR 2017
		Garcia Satorras et al. (2021)	E(n) Equivariant Normalizing Flows; Particle systems	NeurIPS 2021
		Gerken et al. (2022)	Rotational equivariance; S2CNNs; Data augmentation	PMLR 2022
		MGCN Hanik et al. (2024)	Manifold-valued features; Diffusion; GNN	arXiv 2024
		GM-VAE Cho et al. (2024)	Gaussian distributions; Manifolds	NeurIPS 2024
		RRNN Katsman et al. (2024)	Riemannian manifolds; ResNet; Euclidean	NeurIPS 2024
	Graph neural networks	GNN Gori et al. (2005)	Topological information; Learning algorithm	IJCNN 2005
		GRNN You et al. (2018)	Graph modeling; Distribution of graphs; Benchmark suite	ICML 2018
		GWNN Xu et al. (2019)	Graph wavelet transform; Semi-supervised	arXiv 2019
		MolGAN Lee et al. (2021)	Neural generative models; RL; GANs	MI 2021
		FedGNN Wu et al. (2021)	GNN; Privacy preserving; Federated learning	arXiv 2021
		PI-GNN Schuetz et al. (2022)	Combinatorial optimization; Scalability	Nat.Mach.Intell. 2022
Skeleton	SkeletonMAE Yan et al. (2023)	Skeleton sequence learning; Spatial-Temporal	ICCV 2023	
	VIHGNN Han et al. (2023)	Hypergraph structure; Fuzzy C-Means method	ICCV 2023	
	HypDiff Fu et al. (2024)	Diffusion models; Hyperbolic geometry	arXiv 2024	
	FedGCN Yao et al. (2024)	GCN; Federated learning; Semi-supervised	NeurIPS 2024	
Federated	FedGL Chen et al. (2024)	GNN; Federated learning; Self-supervised	ISCI 2024	

Table 1 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
Dynamic neural network systems	Fluid dynamics neural networks	HFM Raissi et al. (2020) FluidsNet Zhang et al. (2020) NeuroFluid Guan et al. (2022)	Flow visualization; Navier–Stokes equations Lagrangian fluid simulation; physical-based animation Physical dynamics; Fluid particle systems	Science 2020 Exp.Syst.Appl. 2020 arXiv 2022
Hamiltonian neural networks	Neural SPH Toshev et al. (2024) HNN Greydanus et al. (2019) HGN Toth et al. (2019) AHNN Han et al. (2021) EHNN Dierkes and Flaßkamp (2021)	Smoothed particle hydrodynamics; GNN Inductive biases; Conservation laws; Reversibility Neural Hamiltonian Flow; Continuous time evolution Chaotic systems; Physical constraints; Adaptive System invariances; Energy conservation	NeurIPS 2019 arXiv 2019 arXiv 2019 Phys. Rev. Res. 2021	
PHNN Eidnes et al. (2023) DeepH-E3 Gong et al. (2023) Ma et al. (2023) HDNN Kaltas (2024)	Pseudo-Hamiltonian; Physics-informed; Hybrid DFT; Euclidean symmetry; Priori knowledge Adaptive estimation scheme; Lyapunov stability Hamilton–Dirac equations; PINNs	PAMM 2021 Phys. D. 2023 Nat. Commun. 2023 AI Rev. 2023 arXiv 2024		
Lagrangian neural networks	HANG Zhao et al. (2024) DeLaN Lutter et al. (2019) LNNs Cranmer et al. (2020) LGNN Bhattoo et al. (2023) GLNN Xiao et al. (2024) ResNet He et al. (2016)	Hamiltonian flows; Lyapunov stability Prior knowledge; Physics models; Model-based control Symmetries; Conservation laws; Energy conservation Graph neural network; Dynamical systems Lagrangian system; Non-conservative system Residual; Optimization; Deeper neural networks	NeurIPS 2024 ICLR 2019 arXiv 2020 MLST 2023 arXiv 2024 CVPR 2016	
Neural network differential equation solvers	PolyNet Zhang et al. (2017) LatSeqODE Shi and Morris (2021)	Structural diversity; PolyInception modules Hybrid systems; Latent ODEs; Reconstructive	CVPR 2017 ICML 2021	

Table 1 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
mmODE	Yi (2023)	mmODE	Inverse ODE	AI Rev. 2023
FANN	Joshi et al. (2023)	ANN	Fractional calculus	AI Rev. 2023
PINN	Raissi et al. (2019)	Data-driven	Predictive modeling; Nonlinear dynamics	J. Comput. Phys. 2019
DPINN	Dwivedi et al. (2019)	PDEs	Burgers' equation; Representation capability	arXiv 2019
PINN-HFM	Raissi et al. (2020)	Navier–Stokes equations	Quantifying	Science 2020
NCDES	Morrill et al. (2021)	Rough path theory	Rough differential equations	ICML 2021
SPINN	Zhang et al. (2023)	Invariant surface conditions	Lie symmetry	J. Comput. Phys. 2023
LPINNs	Mojgani et al. (2023)	Deep learning	Kolmogorov n-width	CMAE 2023
SHoP	Xiao et al. (2024)	High-order PDEs	Taylor series	AAAI 2024
OptPDE	Kantamneni et al. (2024)	Integrable systems	Machine learning	arXiv 2024

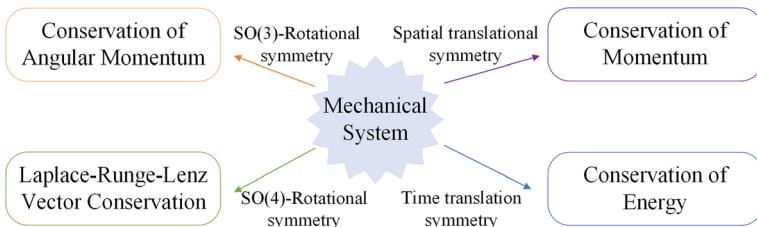


Fig. 2 Related applications of Noether's theorem

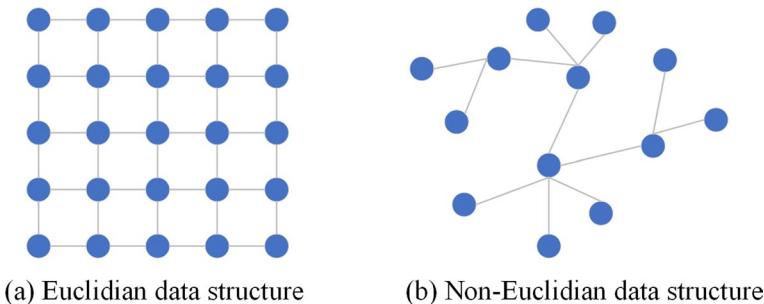


Fig. 3 Euclidian/Non-Euclidian data structures

2.1.1 Manifold neural networks

A manifold is a space with local Euclidean space properties and is used in mathematics to describe geometric shapes, such as the spatial coordinates of the surfaces of various objects returned by radar scans. A Riemann manifold is a differential manifold with a Riemannian metric, where the Riemannian metric is a concept in differential geometry. Simply put, a Riemannian manifold is a smooth manifold given a smooth, symmetric, positive definite second-order tensor field. For example, in physics, the phase space of classical mechanics is an instance of a manifold, and the four-dimensional pseudo-Riemannian manifold that constructs the space-time model of general relativity is also an instance of a manifold.

Often, manifold data have richer spatial information, such as magnetoencephalography on a sphere Defferrard et al. (2020) and human scan data (Armeni et al. 2017; Bogo et al. 2014), which contain local structures and spatial symmetries Meng et al. (2022). At present, a new type of manifold convolution has been introduced into the physics-informed manifold (Masci et al. 2015; Monti et al. 2017; Boscaini et al. 2016; Cohen et al. 2019; De Haan et al. 2020) to make up for the defect that convolutional neural networks cannot fully utilize spatial information.

Manifold learning is a large class of manifold-based frameworks, and recovering low-dimensional structures is often referred to as manifold learning or nonlinear dimensionality reduction, which is an instance of unsupervised learning. Examples of manifold learning include: (1) A multidimensional scaling (MDS) algorithm Tenenbaum et al. (2000) that focuses on preserving “similarity (usually Euclidean distance)” information in high-dimensional spaces, is another linear dimensionality reduction method; (2) Focus on the local linear embedding (LLE) algorithm that preserves the local linear features of the sample during dimensionality reduction Roweis and Saul (2000), abandoning the global optimal

dimensionality reduction of all samples; (3) The stochastic neighbor embedding (t-SNE) algorithm Maaten and Hinton (2008) uses the t distribution of heavy-tailed distribution to avoid the crowding problem and optimization problem, it is only suitable for visualization and cannot perform feature extraction; (4) The Uniform Manifold Approximation and Projection (UMAP) McInnes et al. (2018) algorithm is built on the theoretical framework of Riemannian geometry and algebraic topology. UMAP, like t-SNE, is only suitable for visualization, and the performance of UMAP and t-SNE is determined by different initialization choices (Kobak and Linderman 2019, 2021); (5) Spectral embedding, such as Laplacian feature map, is a graph-based dimensionality reduction algorithm to construct the relationship between data from a local perspective. It hopes that the points that are related to each other (the points connected in the graph) are as close as possible in the space after dimensionality reduction so that the original data structure can still be maintained after dimensionality reduction Belkin and Niyogi (2003); (6) The diffusion map method Wang (2012) uses the diffusion map to construct the data kernel, which is also a nonlinear dimensionality reduction algorithm; (7) The deep model Hadsell et al. (2006) learns a model that can evenly map the data to the output a globally consistent nonlinear function (invariant map) on a manifold for dimensionality reduction. Cho et al. (2024) proposed a Gaussian manifold variational autoencoder (GM-VAE) that addresses common limitations previously reported in hyperbolic VAEs. Katsman et al. (2024) studied ResNet and showed how to extend this structure to general Riemannian manifolds in a geometrically principled way.

2.1.2 Graph neural networks

Another type of non-Euclidean geometric data is graph. Graph refers to network structure data composed of nodes and edges, such as social networks. The concept of graph neural network (GNN) was first proposed by Gori et al. to extend existing neural networks to handle more types of graph data Gori et al. (2005), and then further inherited and developed by Scarselli et al. (2008). In 2020, Wu et al. (2020) proposed a new classification method to provide a comprehensive overview of graph neural networks (GNN) in the field of data mining and machine learning. Zhou et al. (2020) proposed a general design process for GNN models and systematically classified and reviewed applications. The network proposed for the first time in the context of spectral graph theory extends convolution and pooling operations in CNN to graph-structured data. The input is the graph and the signal on the graph, and the output is the node on each graph Defferrard et al. (2016).

The graph convolutional neural network (GCN) is the “first work” of GNN. It uses a semi-supervised learning method to approximate the convolution kernel in the original graph convolution operation and improves the original graph convolution algorithm Kipf and Welling (2016), as shown in the Fig. 4 . For the application of GCN in recommender systems, refer to Monti et al. (2017). Graph convolutional networks are the basis for many complex graph neural network models, including autoencoder-based models, generative models, and spatiotemporal networks. Inspired by physics, Martin et al. published an article using graph neural networks to solve combinatorial optimization problems in the journal Nature Machine Intelligence in 2022 Schuetz et al. (2022). In order to solve the limitation of the large amount of computation of GCN, Xu et al. proposed a graph wavelet neural network (GWNN) Xu et al. (2019) that uses graph wavelet transform to reduce the amount of computation.

Graph Attention Network is a space-based graph convolutional network, which combines the attention mechanism in natural language processing with the new graph geometry

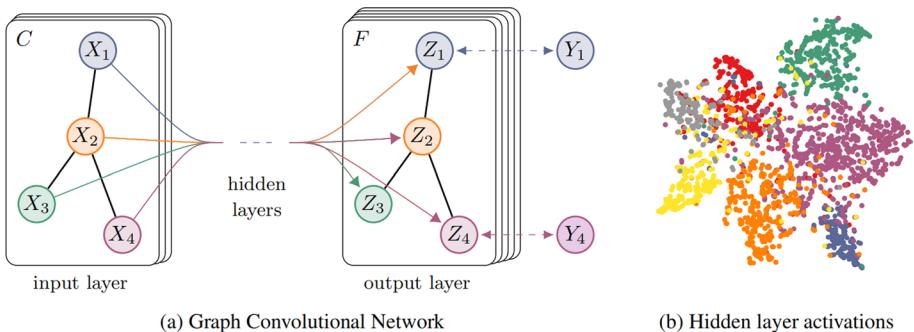


Fig. 4 **a** Schematic depiction of GCN for semi-supervised learning with C input channels and F feature maps in the output layer. The graph structure (edges shown as black lines) is shared over layers, labels are denoted by Y_i . **b** Visualization of hidden layer activations of a two-layer GCN. Colors denote document class Kipf and Welling (2016)

data learning of graph structure data. The attention mechanism is used to determine the weight of the node neighborhood, resulting in a more effective feature representation Velikovi et al. (2017), which is suitable for (graph-based) inductive learning problems and transductive learning problems. The graph attention model proposes a recurrent neural network model that can solve the problem of graph classification. It processes graph information by adaptively visiting the sequence of each important node.

Graph autoencoders are a class of graph embedding methods that aim to represent the vertices of a graph as low-dimensional vectors using a neural network structure. At present, GCN-based autoencoder methods mainly include: GAE Kipf and Welling (2016) and ARGA Pan et al. (2018), and other variants are NetRA Yu et al. (2018), DNGR Cao et al. (2016), DRNE Ke et al. (2018).

The purpose of a graph generation network is to generate new graphs given a set of observed graphs. MolGAN Lee et al. (2021) integrates relational GCNs, modified GANs, and reinforcement learning objectives to generate a graph of the properties required by the models. DGMG Li et al. (2018) utilizes graph convolutional networks with spatial properties to obtain hidden representations of existing graphs, which is suitable for expressive and flexible relational data structures (such as natural language generation, pharmaceutical fields, etc.). GRNN You et al. (2018) generates models through depth graphs of two layers of recurrent neural networks.

In addition to the above-mentioned classic models, researchers have conducted further studies on GCN. For example, GLCN Jiang et al. (2018), RGDCN Brockschmidt (2019), GIC Jiang et al. (2019), HA-GCN Zhou and Li (2017), HGCN Liu et al. (2019), BGCN Zhang et al. (2018), SAGNN Zhang et al. (2019), DVNE Zhu et al. (2018), SDNE Wang et al. (2016), GC-MC Berg et al. (2017), ARGA Pan et al. (2018), Graph2Gauss Bojchevski and Günnemann (2017) and GMNN Qu et al. (2019) and other network models. In fact, the DeepMind team has also begun to pay attention to deep learning on graphs. In 2019, the Megvii Research Institute proposed a GeoConv for modeling the geometric structure between points and a hierarchical feature extraction framework Geo-CNN Lan et al. (2019). Hernández et al. (2022) proposed a method to predict the time evolution of dissipative dynamical systems using graph neural networks. Yao et al. (2024) introduced the Federated Graph Convolutional Network (FedGCN) algorithm for semi-supervised node classification, which has the characteristics of fast convergence and low communication cost.

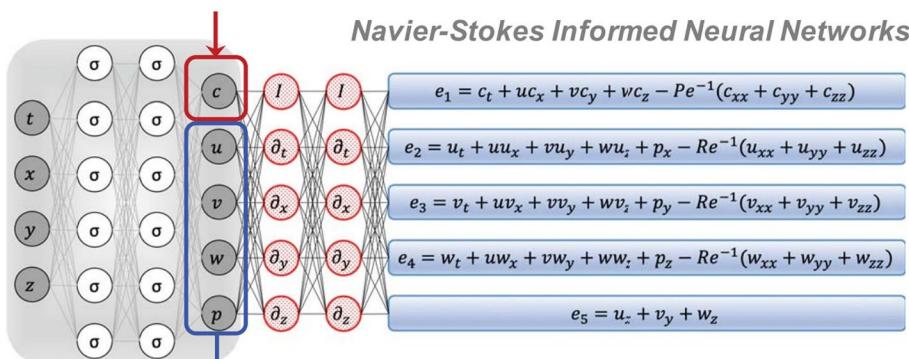


Fig. 5 A physics-uninformed neural network (left) takes the input variables t , x , and y and outputs c , u , v , and p . By applying automatic differentiation on the output variables, we encode the transport and NS equations in the physics-informed neural networks e_i , $i = 1, \dots, 4$ (right) Raissi et al. (2020)

2.1.3 Fluid dynamics neural networks

Computational fluid dynamics (CFD) is the product of the combination of modern fluid mechanics. The research content is to solve the governing equations of fluid mechanics through computer and numerical methods, and to simulate and analyze fluid mechanics problems.

Maziar et al. proposed a physical neural network in science - Hidden Fluid Mechanics Network Framework (HFM) to solve partial differential equations Raissi et al. (2020). The motion of the fluid in Raissi et al. (2020) is governed by the transport equation, the momentum equation, and the continuity equation, and these equations (knowledge of fluid mechanics) are encoded into the neural network, and the governing equations, and a feasible solution is obtained by combining the residuals of the neural network, as shown in Fig. 5. The HFM framework is not limited by boundary conditions and initial conditions. Realizing the prediction of fluid physics data has the advantages of strong versatility of machine learning and strong pertinence of computational fluid dynamics.

Wsewles et al. proposed the NPM–Neural Particle Method Wessels et al. (2020), a computational fluid dynamics using an updated Lagrangian physics-informed neural network, even with discrete point locations highly irregular, NPM is also stable and accurate. A new end-to-end learning deep learning neural network for automatic generation of fluid animations based on Lagrangian fluid simulation data Zhang et al. (2020). Guan et al. proposed the “NeuroFluid” model, which uses the artificial intelligence differentiable rendering technology based on neural implicit fields, regards fluid physics simulation as the inverse problem of solving the 3D rendering problem of fluid scenes and realizes fluid dynamic inversion Guan et al. (2022).

2.2 Dynamic neural network systems

The methods used to express nonlinear functions include dynamic systems and neural networks. At the same time, various nonlinear functions are actually information waves propagating between various layers. If physical systems in the real world are represented by neural networks, it will greatly improve the possibility of applying these physical systems to the field of artificial intelligence for analysis. Neural networks usually use a large

amount of data for training, and adjust the weight and bias of the data through a large amount of information obtained. Minimizing the difference between the actual output and the expected output value, approximating the ground truth. Thereby imitating the behavior of human brain neurons to make judgments. However, this training method has the disadvantage of “chaos blindness”, that is, the AI system cannot respond to the chaos (or mutation) in the system.

2.2.1 Hamiltonian/Lagrangian neural networks

The steepest descent curve problem proposed by the Swiss mathematician Johann Bernoulli makes the variational method an essential tool for solving extreme value problems in mathematical physics. The variational principle of physical problems (or problems in other disciplines) is transformed into the problem of finding a function's extreme value (or stationary value) by using the variational method. The variational principle is also called the principle of least action Feynman (2005). Karl Jacobbit called the principle of least action the mother of analytical mechanics. When applied to the action of a mechanical system, the equation of motion of the mechanical system can be obtained. The study of this principle led to the development of Lagrangian and Hamiltonian formulations of classical mechanics.

Hamiltonian neural networks Hamilton's principle is a variational principle proposed by Hamilton in 1834 for dynamic complete systems. The Hamiltonian (conservation of momentum) embodies complete information about a dynamic physical system, that is, the total amount of all energies, kinetic and potential energies that exist. The Hamiltonian principle is often used to establish dynamic models of systems with continuous mass distribution and continuous stiffness distribution (elastic systems). Hamilton is the “special seasoning” that gives neural networks the ability to learn order and chaos. Neural networks understand underlying dynamics in a way that conventional networks cannot. This is the first step toward neural networks in physics. The NAIL team incorporated the Hamiltonian structure into a neural network, applying it to the known Hénon-Heiles model of stellar and molecular dynamics models Choudhary et al. (2020), accurately predicting the system dynamics moving between order and chaos.

An unstructured neural network, such as a multi-layer perceptron (MLP), can be utilized to parameterize the Hamiltonian. In 2019, Greydanus et al. proposed Hamiltonian Neural Networks (HNN) Greydanus et al. (2019) that learn the basic laws of physics (Hamiltonian of mass-spring systems) and accurately preserve a quantity similar to the total energy (energy conservation). In the same year, Toth et al. used the Hamiltonian principle (variational method) to transform the optimization problem into a functional extreme value problem (or stationary value) and proposed Hamiltonian Generative Networks (HGN) Toth et al. (2019). Due to the physical limitations defined by the Hamiltonian equations of motion, the research Han et al. (2021) introduces a class of HNNs that can adapt to non-linear physical systems. By training a time-series-based neural network, from a small number of bifurcation parameter values of the target Hamiltonian system, the dynamic state of other parameter values can be predicted. The work Dierkes and Flaßkamp (2021) introduced the Hamiltonian Neural Network (HNN) to explicitly learn the total energy of the system, training the neural network to learn the equations of motion to overcome the lack of physical rules.

In the field of neural networks applied to chaotic dynamic systems, the work by Haber and Ruthotto (2017) introduces a neural network model called “Stable Neural Networks,”

which is inspired by the differential equations of the Hamiltonian dynamical system. This model aims to address the issue of susceptibility to input data disturbance or noise that can affect the performance of neural networks obtained through the discretization of chaotic dynamic systems.

Another relevant research paper by Massaroli et al. (2019) offers a novel perspective on neural network optimization, specifically tackling the problem of escaping saddle points. The non-convexity and high dimensionality of the optimization problem in neural network training make it challenging to converge to a minimum loss function. The proposed framework guarantees convergence to a minimum loss function and avoids the saddle point problem. It also demonstrates applicability to neural networks based on physical systems and pH control, improving learning efficiency and increasing the likelihood of finding the global minimum of the objective function.

Additionally, there are other methods available for identifying Hamiltonian dynamical systems (HDS) using neural networks, as discussed in the referenced paper by Lin et al. (2017). These methods contribute to the exploration of neural network architectures and techniques for modeling and understanding HDS. Zhao et al. (2024) used conservative Hamiltonian neural flow to construct a GNN that is robust to adversarial attacks, greatly improving the robustness to adversarial perturbations.

Overall, these research works highlight important approaches and perspectives in applying neural networks to chaotic dynamic systems, addressing challenges such as input data disturbance, saddle point problems, and optimization difficulties.

Lagrangian neural networks The Lagrangian function of analytical mechanics is a function that describes the dynamical state of the entire physical system. The Lagrangian function of a system represents the properties of the system itself. If the world is symmetric (such as spatial symmetry), then after the system is translated, the Lagrangian function remains unchanged, and momentum conservation can be obtained using the variational principle.

Even if the training data satisfies all physical laws, it is still possible for a trained artificial neural network to make non-physical predictions (there are some scenarios where rigid body kinematics is not applicable, and it is even difficult to calculate with physical formulas). Therefore, in 2019, the object mass matrix in the Euler-Lagrangian equation is represented by a neural network, so that the relationship between the mass distribution and the robot pose can be estimated Lutter et al. (2019). Deep Lagrangian networks learn the equations of motion for mechanical systems, train faster than traditional feedforward neural networks, predict results more physically, and are more robust to new track predictions.

In order to enhance the sparsity and stability of the algorithm, the work Cranmer et al. (2020) proposes a new sparse penalty function based on the dimension reduction algorithm SCAD Fan and Li (2001), and adds it to the Lagrangian Constrained Neural Network to overcome the traditional blind source separation. The defects of the method and the independent component analysis method can effectively avoid the ill-conditioned problem of the equation and improve the sparsity, stability, and accuracy of blind image restoration. Since neural networks cannot conserve energy, it is difficult to model dynamics over a long period of time. In 2020, Cranmer et al. The research Cranmer et al. (2020) used neural networks to learn arbitrary Lagrangian quantities, inducing strong physical priors, as shown in Fig. 6. Xiao et al. (2024) introduce a breakthrough extension of the Lagrangian neural network (LNN) (generalized Lagrangian neural network), which is innovatively tailored for non-conservative systems.

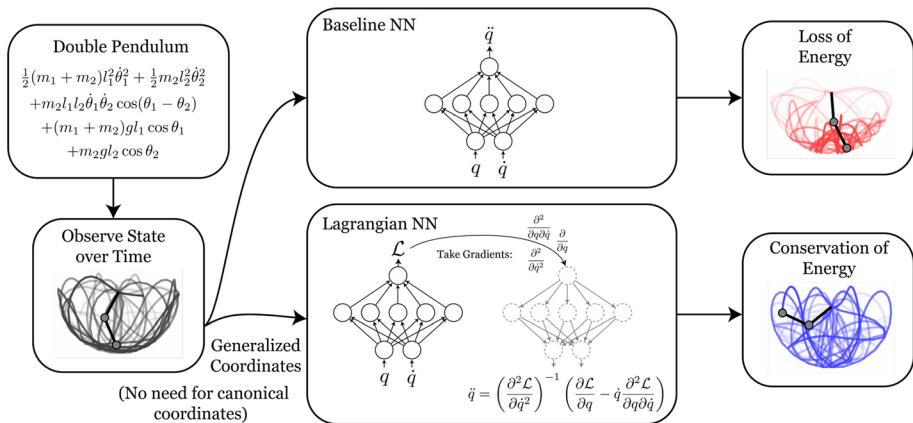


Fig. 6 Cranmer et al. (2020) propose a method to address the challenge of modeling the dynamics of physical systems using neural networks. They demonstrate that neural networks struggle to accurately represent these dynamics over long time periods due to their inability to conserve energy. To overcome this limitation, the authors introduce a technique for learning arbitrary Lagrangians with neural networks, which incorporates a strong physical prior on the learned dynamics. By leveraging the principles of Lagrangian mechanics, the neural networks are able to better capture the underlying physics of the system. This approach improves the accuracy of the neural network model (shown in blue) compared to traditional neural networks (shown in red), providing a promising avenue for enhancing the modeling of complex dynamical systems

2.2.2 Neural network differential equation solvers

In physics, due to the concepts of locality and causality equations, differential equations are basic equations, so it is a cutting-edge trend to treat neural networks as dynamic differential equations and to use numerical solution algorithms to design network structures.

Ordinary differential equation neural networks The general neural ODE is as follows:

$$\begin{aligned} y(0) &= y_0 \\ \frac{dy}{dt}(t) &= f_\theta(t, y(t)) \end{aligned} \tag{1}$$

where y_0 can be any dimension tensor, θ indicates some vector of learned parameters, f_θ indicates a neural network.

Neural networks offer powerful function approximation capabilities, while penalty terms help bridge the gap between theory and practice. One application is in turbulence modeling, as demonstrated in Ling et al. (2016), where a carefully designed neural network approximates closed relations (Reynolds stresses) while adhering to specific physical invariances. This approach enables the modeling of residuals between theoretical and observed data.

Latent ODEs emerge from this framework when incorporating time-varying components. Rubanova et al. (2019) utilize latent ODEs to simulate the dynamics of a small frog entering the air in a simulated environment. Additionally, Du et al. (2020) explore the applications of latent ODEs in reinforcement learning.

Another study by Shi and Morris (2021) combines latent ODEs with change-point detection algorithms to model switching dynamical systems. This approach provides a powerful tool for segmenting and understanding complex dynamics with abrupt changes.

In summary, neural networks coupled with penalty terms and latent ODEs offer valuable methods for modeling and simulating various dynamic systems, including turbulence, reinforcement learning, and switching dynamical systems. These approaches bridge the gap between theoretical principles and practical applications, opening up new possibilities in understanding and predicting complex phenomena.

Euler's method: The main idea of Euler's method is to use the first derivative of a point to linearly approximate the final value. Due to the different positions of the points where the first derivative is used, it is divided into the forward Euler method (also known as explicit Euler method) and backward Euler's method (Implicit Euler's method). The general form of deep residual network (ResNet) He et al. (2016) can be regarded as a discrete dynamical system, because each step of it is composed of the simplest nonlinear discrete dynamical system-linear transformation and non-linear linear activation function is formed. It can be said that the residual network is an explicit Euler discretization of a neural ODE. Now, the RevNet neural network Behrmann et al. (2019), as a further generalization of ResNet, is a residual learning with a symmetric form. The backward Euler algorithm corresponds to PolyNet Zhang et al. (2017), PolyNet can reduce the depth by increasing the width of each residual block, thereby achieving the most advanced classification accuracy. In addition, from the perspective of ordinary differential equations, the reverse Euler method has better stability than the forward Euler method. For more methods of using ordinary differential equations themselves as neural networks, see Chen et al. (2018).

Partial differential equation neural networks The general form of a second-order PDE:

$$\frac{\delta^2 \psi(x, y)}{\delta x^2} + \frac{\delta^2 \psi(x, y)}{\delta y^2} = f(x, y) \quad (2)$$

The design of FractalNet is based on self-similarity, by repeatedly applying a simple extension rule to generate deep networks whose structure is laid out as a truncated fractal Larsson et al. (2016), whose structure can be explained as the famous Runge- Kuta form. The activation and weight dynamics of neural networks in Ramacher (1993) are derived from partial differential equations and incorporate weights as parameters or variables. Results obtained using a combination of time-varying patterns of parameters and dynamics show that learning rules can be replaced by learning laws under equal performance.

Physics Informed Neural Network (PINN) Raissi et al. (2019) is a method of applying scientific machines in traditional numerical fields, especially for solving various problems related to PDE. The principle of PINN is to approximate the solution of PDE by training the neural network to minimize the loss function. The essence is to integrate the equation (physical knowledge) into the network and use the residual term from the governing equation to construct a loss function, which is used as a penalty term to limit the space of feasible solutions.

The PINN-HFM Raissi et al. (2020) algorithm fused with physical knowledge reconstructs the overall velocity field of resolution from sparse velocity information. That is, the loss term of the NS equation is minimized, and the velocity field and the pressure field are obtained at the same time, so that the result conforms to the “laws of physics”. Compared to traditional CFD solvers, PINN is better at integrating data (observations of flow) and physical knowledge (essentially the governing equations describing the physical phenomenon).

Considering that PINN is not robust enough for extreme gradient decline, and the depth increases with the PDE order, resulting in vanishing gradients and slower learning rates, Dwivedi et al. (2019) propose DPINN. In 2020, Meng et al. (2020) used the traditional parareal time domain segmentation method for parallelization to reduce the complexity and learning difficulty of the model. Unlike PINN and its variants, Fang (2021) proposed using the approximation of differential operators instead of automatic differentiation to solve hybrid physical information networks of PDEs. The research Moseley et al. (2021) presents a parallel approach to spatially partitioned regions. As a meshless method, PINN does not require a mesh. Therefore, an algorithm using the fusion differential format to accelerate information dissemination has also emerged Chen et al. (2021). Then the work Schiassi et al. (2022) utilizes PINN to solve the equation paradigm, which is used to “learn” the optimal control of the plane orbit transfer problem. Since the global outbreak of the Covid-19 virus, Treibert et al. used PINN to evaluate model parameters, built an SVIHDR differential dynamical system model Treibert and Ehrhardt (2021), extended Susceptible-Infected-Recovered (SIR) model Trejo and Hengartner (2022).

Although AI using PDE to simulate physical problems has been widely used, there are still limitations in solving high-dimensional PDE problems. This work Karniadakis et al. (2021) discusses the diverse applications of physical knowledge (discipline) learning integrating noisy data and mathematical models, under the condition of satisfying the physical invariance, improving the accuracy, and solving the hidden physical inverse problems and high-dimensional problems. Xiao et al. (2024) proposed a deep learning framework for solving high-order partial differential equations, named SHoP. At the same time, the network was expanded to the Taylor series, providing explicit solutions to partial differential equations.

Controlled differential equations neural networks Neural controlled differential equations (CDEs) rely on two concepts: Bounded paths and Riemann CStieltjes integrals, which are formulated as follows:

$$\begin{aligned} y(0) &= y_0 \\ \int_0^t f(y(s))dx(s) &= \int_0^t f(y(s)) \frac{dx}{ds}(s)ds \end{aligned} \quad (3)$$

Modeling the dynamics of time series using neural differential equations is a promising option, however, the performance of current methods is often limited by the choice of initial conditions. The neural CDEs model generated by Kidger et al. (2020) can handle irregularly sampled and partially observed input data (i.e., time series), and has higher performance than ODE or RNN-based models. Additional terms in the numerical solver are introduced in Morrill et al. (2021) to incorporate substep information to obtain neural rough differential equations. When dealing with data with missing information, it is standard practice to add observation masks Che et al. (2018), which is the appropriate continuous-time analogy.

Stochastic differential equation neural networks Stochastic Differential Equations (SDE) have been widely used to model real-world stochastic phenomena such as particle systems (Coffey and Kalmykov 2012; Pavlitos 2014), financial markets Black and Scholes (2019), population dynamics Arató (2003) and genetics Huillet (2007). Latent ODEs serve as a natural extension of ordinary differential equations (ODEs) for modeling systems that evolve in continuous time while accounting for uncertainty Kidger (2022).

The dynamics of a stochastic differential equation (SDE) encompass both a deterministic term and a stochastic term:

$$dy(t) = \mu(t, y(t))dt + \sigma(t, y(t))dw(t) \quad (4)$$

where μ, σ is a regular function, w is a d dimensional Brownian motion, and y is the resulting d dimensional continuous random process.

The inherent randomness in stochastic differential equations (SDEs) can be viewed as a generative model within the context of modern machine learning. Analogous to recurrent neural networks (RNNs), SDEs can be seen as an RNN with random noise, specifically Brownian motion, as input, and the generated sample as the output. Time series models are classic interest models. Predictive models such as Holt-Winters Holt (2004), ARCH Engle (1982), ARMA Hannan and Rissanen (1982), GARCH Bollerslev (1986), etc.

More deep learning libraries for solving differential equations combined with physical knowledge and machine learning such as the literature Lu et al. (2021).

2.3 Graph neural networks to solve physical problems

Molecular Design: The most critical problem in the fields of materials and pharmaceuticals is to predict the physical, chemical, and biological properties of new molecules from their structures. Recent work from Harvard University Duvenaud et al. (2015) proposes to model molecules as graphs and use graph convolutional neural networks to learn the desired molecular properties. Their method significantly outperforms the handcrafted capabilities of Morgan (1965), Rogers and Hahn (2010), a work that opens up opportunities for molecular design in a new way.

Medical Physics: The field of medical physics Manco et al. (2021) is one of the most important areas of artificial intelligence application, which can be roughly divided into radiotherapy and medical imaging. With the success of AI in imaging tasks, AI research in radiotherapy (Hrinivich and Lee 2020; Maffei et al. 2021) and medical imaging (such as x-ray, MRI, and nuclear medicine) Barragán-Montero et al. (2021) has grown rapidly. Among them, magnetic resonance imaging (MRI) technology in medical image analysis Castiglioni et al. (2021) plays a vital role in the diagnosis, management, and monitoring of many diseases Li et al. (2022). A recent study from Imperial College Ktena et al. (2017) uses graph CNNs on non-Euclidean brain imaging data to detect disruptions in autism-related brain functional networks. Zegers et al. outlined the current state-of-the-art applications of deep learning in neuro-oncology MRI Zegers et al. (2021), which has broad potential applications. Rizk et al. introduced deep learning models for meniscal tear detection after external validation Rizk et al. (2021). The discussion and summary of MRI image reconstruction work Montalt-Tordera et al. (2021) provides great potential for the acquisition of future clinical data pairs.

High-energy physics experiments: Introducing graph neural networks to predict the dynamics of N-body systems (Battaglia et al. 2016; Chang et al. 2016) with remarkable results.

Power System Solver: The research Donon et al. (2019) combines graph neural networks to propose a neural network architecture for solving power differential equations to calculate power flow (so-called “load flow”) in the grid. The work Park and Park (2019) proposes a physics-inspired data-driven model for wind farm power estimation tasks.

Structure prediction of glass systems (glass phase transitions): DeepMind published a paper in Nature Physics Bapst et al. (2020) to model glass dynamics with a graph neural network model, linking network predictions to physics. The long-term evolution of glassy systems can be predicted using only the structures hidden around the particles. The model

works well across different temperature, pressure, and density ranges, demonstrating the power of graph networks.

3 Deep neural network paradigms inspired by electromagnetics

3.1 Optical design neural networks

Optical neural networks(ONNs) are novel types of neural networks designed with optical technology such as optical connection technology, optical device technology, and so on. The idea of optical neural networks is to imitate neural networks by attaching information to optical features utilizing modulation. At the same time, taking advantage of the optical propagation principle of light such as interference, diffraction, transmission, and reflection to realize neural networks and their operators. The first implementation of ONNs was optical Hopfield networks, proposed by Demetri Psaltis and Farhat (1985) in 1985. There are three main operators involved in traditional neural networks: linear operations, nonlinear activation operations, and convolution operations, and in this subsection, the optical implementation of the above operators is presented in that order. We summarize the structure of this section and an overview of representative methods in Table 2.

3.1.1 Optical implementation of linear operations

The main linear operators of neural networks are matrix multiplication operators and weighted summation operators. The weighted summation operators are easy to implement due to the property of optical coherence and incoherence, so the challenge of optical implementation of linear operations lies in the optical implementation of matrix multiplication. As early as 1978, J. W. Goodman et al. (1978) first implemented an optical vector-matrix multiplier with a lens set according to the principle of optical transmission; And the optical implementation of matrix-matrix multiplier was first implemented using a 4f-type system consisting of a lens set by Chen (1993).

Optical implementation of vector-matrix multiplications The vector \mathbf{p} is obtained by multiplying the matrix \mathbf{A} with the vector \mathbf{b} . The mathematical essence is to use each row of the matrix \mathbf{A} to make an inner product with the vector \mathbf{b} to obtain the value of the corresponding position of the vector \mathbf{p} . The mathematical expression is:

$$\mathbf{p}(i) = \sum_j \mathbf{A}(i,j)\mathbf{b}(j) \quad (5)$$

The optical vector-matrix multiplier is mainly composed of two parts: the light source such as light-emitting diode light source arrays, and the optical path system composed of a spherical lens, a cylindrical lens, a spatial light modulator, and an optical detector. Its mathematical idea is to transform the vector-matrix multiplication into the matrix-matrix point-wise multiplication.

As shown in Fig. 7, the vector \mathbf{b} is modulated into optical features of the incoherent light source (LS) such as the amplitude, intensity, phase, and polarization, then the incident light passes the first spherical lens L1. Since the LS array is located in the front focal plane of the spherical lens L1, the light through L1 is emitted in parallel. Next, the light passes the cylindrical lens CL1, which is located in the post-focal plane of the L1. Due to the vertical placement of the cylindrical lens CL1, the light through CL1 is only converged on the

Table 2 An overview of methods for AI DNNs inspired by electromagnetism

Categories	Detail	Typical Methods	Keywords	Publication & Years
Optical design neural networks	Optical implementation of linear operations	ONN Psaltis and Farhat (1985) Chen (1993)	Hopfield model; Collective computational properties Matrix multiplication; Holographic mask; Fourier lenses	Opt. Lett. 1985 Opt. Eng. 1993
MOINN	Francis et al. (1991)	Mirror-array interconnection; High-light-efficiency		Opt. Lett. 1991
Nitta et al. (1993)		Optical neurochip; Light-emitting diode array		Appl. Opt. 1993
POINN	Wang et al. (1997)	Prism-array interconnection; New architecture		Opt. Eng. 1997
AOML	Lin et al. (2018)	Diffractive optical elements; All-optical		Science 2018
F-D ² NN	Yan et al. (2019)	Fourier-space diffractive; Diffractive modulation		Phys. Rev. Lett. 2019
D ² NNs	Mengu et al. (2019)	Optoelectronic neural networks;	J. Sel. Top. Quan. Elec. 2019	
Hamerly et al. (2019)		Optical computing		Phys. Rev. X 2019
Du et al. (2023)		Photonic accelerator; Coherent detection; Spatial multiplexing Mach-Zehnder interferometer; Micro-ring resonator	IET. Opt. 2023	
Xbar	Giamougiannis et al. (2023)	SVD; Lower insertion losses	J. Lightw. Technol. 2023	
Li et al. (2023)		Massively parallel; Wavelength-multiplexed	Adv. Photonics 2023	
Huang et al. (2024)		Optical encoder; Metasurface; ONN	Nanophotonics 2024	
Optical implementation of nonlinear activation	Photonic RC Antonik et al. (2019)	Human action recognition; Reservoir computing; RNN	Nat. Mach. Intell. 2019	

Table 2 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
	PhRC Katumba et al. (2019)	Neuromorphic computing; Silicon photonics	J. Lightw. Technol. 2019	
	Feldmann et al. (2019)	Neurosynaptic system; Brain-inspired computing; All-optical Graphene; Carrier dynamics; Ultrafast photonics	Nature 2019	
	Bao et al. (2011)	ANN; ONN; Power efficiency	Nano Res. 2011	
	FONN Shen et al. (2017)	Neuromorphic compute paradigms; All-optical perception	Nat. Photonics 2017	
	Miscuglio et al. (2018)	ANNS; Nonlinear optical activation functions	Opt. Mater. Express 2018	
	AONN Zuo et al. (2019)	Optical computing; Diffractive optical element	Optica 2019	
	HoCNN Chang et al. (2018)	Tensor core; Photonic in-memory	Sci. Rep. 2018	
	IPHA Feldmann et al. (2021)	Photonic neural network; Micro-disk modulator	Nature 2021	
	E-O NAFs Wang et al. (2023)	Optical imaging; Nonlinear ONN encoder	Phys. Opt. 2023	
	MNONNN Wang et al. (2023)	Multimode fibers; Nonlinear optical computation	Nat. Photonics 2023	
	Oguz et al. (2024)	Complex-valued residual; Nonlinear inverse scattering	AP 2024	
Optical implementation of convolutional neural networks	DeepNIS Li et al. (2018)	Inverse scattering problems; ICLM	IEEE Trans. Ante Propag. 2018	
	CEE-CNN Wei and Chen (2019)	Contrast Source Inversion; Micro-wave imaging	IEEE Trans. Ante Propag. 2019	
	CV-Pix2pix Guo et al. (2021)		Electronics 2021	

Table 2 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
	Bernstein et al. (2023)	Single-shot; CMOS-compatible DNN processor	Sci. Adv. 2023	
	OCNN Yang et al. (2023)	CNNs; Optical nonlinearities	Opt. Express 2023	
	OP-FCNN Huang et al. (2024)	End-to-end; Spatial light modulators	Opt. Express 2024	

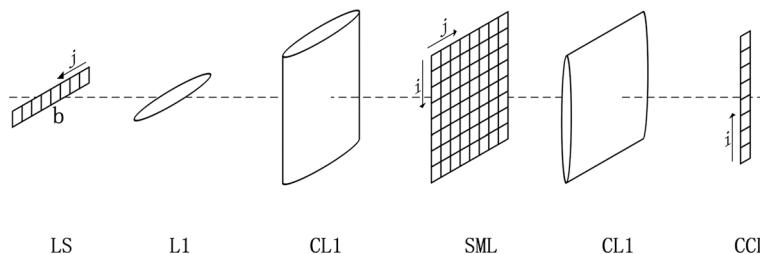


Fig. 7 Optical implementation of vector–matrix from Goodman et al. (1978)

post-focal plane in the horizontal direction, and the light is emitted in parallel in the vertical direction. At this time the light field carries the information:

$$\mathbf{B} = \begin{bmatrix} \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix} \in R^{J \times J} \quad (6)$$

There is a spatial light modulator(SLM) being placed on the back post-focal plane of CL1, which contains the information of matrix \mathbf{A} . The process of passing through the SLM can be seen as the process of dot multiplication of matrix \mathbf{A}, \mathbf{B} . At this time, the light field carries the information as:

$$\mathbf{P}(i,j) = \mathbf{A}(i,j)\mathbf{B}(i,j) \quad (7)$$

Then, the light through SLM passes the cylindrical lens CL2, between which and SLM the distance is the focal length f of CL2. Due to the horizontal placement of the cylindrical lens CL2, the light through SLM is only converged on the post-focal plane in the vertical direction, and the light is emitted in parallel in the horizontal direction. At this time, the light field carries the information of the multiplication result of the vector \mathbf{p} :

$$\mathbf{p}(i) = \sum_j \mathbf{P}(i,j) = \sum_j \mathbf{A}(i,j)\mathbf{B}(i,j) \quad (8)$$

Finally, the light through CL2 is demodulated and the vector \mathbf{p} can be obtained with a charge-coupled device(CCD).

Optical implementation of matrix–matrix multiplications Compared to vector–matrix multiplication, matrix–matrix multiplication is more complicated. The multiplication of matrix \mathbf{A} and matrix \mathbf{B} is the inner product for each row of matrix \mathbf{A} and each column of matrix \mathbf{B} . Assuming that the result matrix is \mathbf{P} , the expression is as follows:

$$\mathbf{P}(x,y) = \sum_l \mathbf{A}(x,l)\mathbf{B}(l,y) \quad (9)$$

The matrix–matrix multiplication is implemented with the help of an optical 4f-type system, which consists of Fourier lenses, holographic masks(HM), and charge-coupled devices. Taking advantage of the discrete Fourier transform(DFT), the matrix \mathbf{B} can be constructed with the discrete Fourier transform matrix to implement the multiplication.

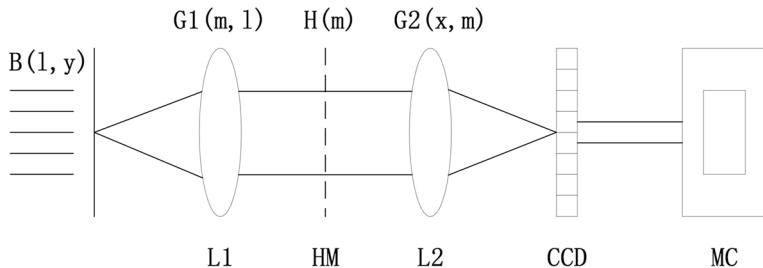


Fig. 8 Optical implementation of matrix-matrix from Chen (1993)

As shown in Fig. 8, the matrix \mathbf{B} is modulated in the complex amplitude of the input light, and the result matrix \mathbf{P} is obtained in the output plane. The multiplication operation of matrix \mathbf{A} and matrix \mathbf{B} is completed during the light propagation from the input plane to the output plane. Let the matrix \mathbf{B} and the function F be the input light field, the Fourier transform function at the front-focal plane of the Fourier lens, respectively. According to the principle of Fresnel diffraction, the complex amplitude distribution of the light field at the post-focal plane of the lens is the Fourier transform of the complex amplitude distribution of the light field at the front-focal plane, and the expression is as follows:

$$\mathbf{P}(x, y) = \frac{1}{i\lambda f} F\left(\mathbf{B}\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right) \quad (10)$$

Since the DFT can be implemented with the DFT matrix, combining with the Equation (10), the discretized light field is expressed as:

$$\mathbf{P}(x, y) = \sum_l \mathbf{G}(x, l) \mathbf{B}(l, y) \quad (11)$$

In this case, the DFT matrix \mathbf{G} of the lens is only related to the focal length and the wavelength, so the matrix \mathbf{A} must be moderated with a holographic mask, which is used to adjust the complex amplitude distribution of the light field. The whole optical system is composed of two Fourier lenses and a holographic mask, so the output light field is:

$$\begin{aligned} \mathbf{P}(x, y) &= \sum_m \mathbf{G}_2(x, m) \mathbf{H}(m) \left(\sum_l \mathbf{G}_1(m, l) \mathbf{B}(l, y) \right) \\ &= \sum_m \sum_l (\mathbf{G}_2(x, m) \mathbf{H}(m) \mathbf{G}_1(m, l) \mathbf{B}(l, y)) \end{aligned} \quad (12)$$

where the matrices \mathbf{G}_1 and \mathbf{G}_2 denote the DTF matrices of the two lenses, respectively, and $\mathbf{H}(m)$ is the complex amplitude distribution function of the holographic mask. Comparing the Equation (9) and the Equation (12):

$$\mathbf{A}(x, l) = \sum_m \mathbf{G}_2(x, m) \mathbf{H}(m) \mathbf{G}_1(m, l) \quad (13)$$

The relationship between the sampling periods and the sampling numbers in the input plane, the output plane, and the holographic mask satisfies:

$$\begin{cases} \frac{\Delta x_1 \Delta x}{f\lambda} = \frac{1}{M} \\ \frac{\Delta x \Delta x_2}{f\lambda} = \frac{1}{X} \\ M = X \times L \end{cases} \quad (14)$$

where $\Delta x_1, \Delta x, \Delta x_2, L, M, X$ are the sampling periods and the sampling numbers in the input plane, the holographic mask, and the output plane, respectively. According to the equation (13) and the equation (14), $\mathbf{H}(m)$ can be obtained:

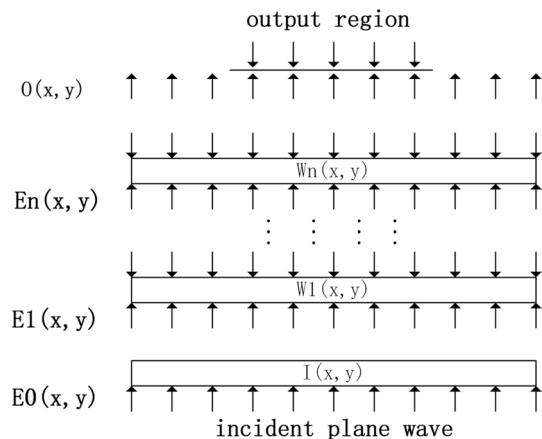
$$\mathbf{H}(m) = \sum_x \sum_l \exp\left(\frac{i2\pi mx}{X}\right) \mathbf{A}(x, l) \exp\left(\frac{i2\pi lm}{M}\right) \quad (15)$$

Optical matrix multipliers The vector-matrix multiplier was first proposed by J. W. Goodman et al. (1978) in 1978. With this multiplier, the DFT was implemented in an optical way. These works (Liu et al. 1986; Francis et al. 1990; Yang et al. 1990) proposed to construct a spatial light modulator with a miniature liquid crystal television (LCTV) to replace the matrix mask and lens to implement matrix multiplication. The research Francis et al. (1991) proposed to use a mirror array instead of the commonly used lens array to realize the optical neural network that uses a mirror-array interconnection; And the work Nitta et al. (1993) removed two cylindrical lenses from the matrix multiplier, improved light-emitting diode arrays and the variable-sensitivity photodetector arrays, and produced the first optical neural chip. The research Chen (1993) proposed to construct an optical 4f-type system, which used the optical Fourier transform and inverse transform of Fourier lenses to implement matrix-matrix multiplication. The research Wang et al. (1997) proposed a new optical neural network architecture that uses two perpendicular 1-D prism arrays for optical interconnection to implement matrix multiplication.

Psaltis et al. (1988) proposed the implementation of matrix multiplication using the dynamic holographic modification of photorefractive crystals, enabling the construction of most neuro networks. Slinger (1991) proposed a weighted N-to-N volume-holographic neural interconnect method and derived the coupled-wave solutions that describe the behavior of an idealized version of the interconnect. (Yang et al. 1994; Di Leonardo et al. 2007; Nogrette et al. 2014) proposed the use of the Gerchberg-Saxton algorithm to calculate holograms for each region. The research Lin et al. (2018) proposed the use of transmissive and reflective layers to form phase-only masks and construct all-optical neurons by optical diffraction. Yan et al. (2019) proposed a novel diffractive neural network implemented by placing diffraction modulation layers at the Fourier plane of the optical system. The research Qian et al. (2020) proposed to scatter or focus the plane wave at microwave frequencies in a diffractive manner on a compound Huygens metasurface to mimic the functionality of artificial neural networks.

Lin et al. Mengu et al. (2019) proposed to use five phase-only diffractive layers for complex-valued phase modulation and complex-valued amplitude modulation to implement an optical diffraction neural network. Shen et al. (2017), Bagherian et al. (2018) take advantage of the Mach-Zehnder interferometer array to implement matrix multiplication through the principle of singular value decomposition; Hamerly et al. (2019) proposed an optical interference-based zero-difference detection method to implement matrix multiplication and constructed a new type of photonic accelerator to implement optical neural networks. Zang et al. (2019) implemented the vector-matrix multiplications by stretching

Fig. 9 Optical neural network based on Kerr-type nonlinear materials from Skinner et al. (1994)



time-domain pulses. With the help of fiber loops, the multi-layer neural network can be implemented in optical.

3.1.2 Optical implementation of nonlinear activation

Nonlinear activation functions play an important role in neural networks, which enable them to approximate complex nonlinear mappings. However the lack of nonlinear response in optics and the limitations of the fabrication conditions of optical devices, the optical response of devices is often fixed, which limits the optical nonlinearity from being reprogrammed to achieve different forms of nonlinear activation functions. Therefore, previous nonlinearities in ONNs were generally achieved using optoelectronic hybrid methods Dunning et al. (1991). With the development of material fabrication conditions, the all-optical implementation of optical nonlinearity Skinner et al. (1994) has only emerged. This is presented below as an example (Fig. 9).

The all-optical neural network consists of linear layers and nonlinear layers, where the linear layers are composed of thick linear media, such as free space, and the nonlinear layers are composed of thin nonlinear media, such as Kerr-type nonlinear materials, whose refractive index satisfies the following relationship:

$$n(x, y, z) = n_0 + n_2 I_r(x, y, z) \quad (16)$$

where n_0 is the linear refractive index component, n_2 is the nonlinear refractive index coefficient, and $I_r(x, y, z)$ is the light field intensity. The material behaves as self-focusing if $n_2 > 0$, and the material behaves as self-scattering if $n_2 < 0$. Since its refractive index is dependent on the light intensity, the nonlinear layer can play the role of both weighted summation and nonlinear mapping.

When the input light is incident to the plane of the nonlinear layer, the refractive index will be different at various points of the nonlinear plane, which results in changes in the intensity and direction of the transmitted light and the appearance of interference phenomenon, so the nonlinear layer achieves the function of spatial light modulation. The final output light signal depends on the first layer input and the continuous weighting and nonlinear mapping of the nonlinear layer.

Photoelectric hybrid methods Dunning et al. (1991) processed video signals on a point-by-point basis by a frame grabber and image processor to implement programmable nonlinear activation functions. Larger et al. (2012) used an integrated telecom Mach-Zendel modulator to provide an electro-optical nonlinear modulation transfer function to achieve the construction of optical neural networks. Antonik et al. (2019) modulated the phase of spatially extended plane waves by means of a spatial light modulator to improve the parallelism of the optical system, which could significantly increase the scalability and processing speed of the network. Katumba et al. (2019) constructed nonlinear operators of networks with the nonlinearity of electro-optical detectors to achieve extremely high data modulation speed and large-scale network parameter update. Williamson et al. (2019), Fard et al. (2020) converted a small portion of the incident light into the electrical signal and modulated the original light signal with the help of an electro-optical modulator to realize the nonlinearity of the neural network, which increases the operating bandwidth and computational speed of the system.

All-optical methods Skinner et al. (1994) implemented weighted connectivity and nonlinear mapping using Kerr-type nonlinear optical materials as the thin layer separating the free space to improve the response speed of optical neural networks. Saxena and Fiesler (1995) used of liquid crystal light valve (LCLV) to achieve the threshold effect of nonlinear functions and constructed an optical neural network to avoid the energy loss problem of photoelectric conversion. Vandoorne et al. (2008), Vandoorne et al. (2014) used coupled semiconductor optical amplifiers (SOA) as the basic block to achieve nonlinearity in all-optical neural networks, making the networks with low power consumption, high speed, and high parallelism. Rosenbluth et al. (2009) used novel nonlinear optical fibers as thresholds to achieve nonlinear responses in networks, overcoming the scalar problem of digital optical calculations and the noise accumulation problem of analog optical calculations. Mesaritakis et al. (2013), Denis-Le Coarer et al. (2018), Feldmann et al. (2019) used the property of nonlinear refractive index variation of ring resonators to provide the nonlinear response of the network, enabling optical neural networks with high integration and low power consumption. Lin et al. (2018) proposed a method to build optical neural networks using only optical diffraction and passive optical components working in concert, avoiding the use of power layers and building an efficient and fast way to implement machine learning tasks. Bao et al. (2011); Shen et al. (2017); Schirmer and Gaeta (1997) exploited the saturable absorption properties of nanophotons to achieve nonlinearity in networks. Mis-cuglio et al. (2018) discussed two approaches to achieve nonlinearity in all-optical neural networks with the reverse saturable absorption property and electromagnetically induced transparency of nanophotonics; Zuo et al. (2019) used the spatial light modulator and Fourier lens to program for linear operation and electromagnetically induced transparency of laser-cooled atoms for nonlinear optical activation functions.

3.1.3 Optical implementation of convolutional neural networks

By imitating the information hierarchical processing mechanism of biological vision, the convolutional neural network(CNN) has the properties of local perception and weight sharing, which significantly reduces the computational complexity and makes networks with stronger fitting ability to fit more complex nonlinear functions.

A deep convolutional neural network is proposed in Shan et al. (2018) to accelerate electromagnetic simulations and predict the 3D Poisson equation for the electrostatic potential distribution through the powerful ability to approximate nonlinear functions. Li

et al. (2018) proposed a novel DNN architecture called DeepNIS for nonlinear inverse scattering problems (ISPs). DeepNIS consists of a cascade of multilayer complex-valued residual CNN to imitate the multi-scattering mechanism. This network takes the EM scattering data collected by the receiver as input and outputs a super-resolution image of EM inverse scattering, which maps the coarse images to the precise solutions to the ISPs. Wei and Chen (2019) proposed a physics-inspired induced current learning method (ICLM) to solve the full-wave nonlinear ISPs. In this method, a novel CEE-CNN convolutional network is designed, which feeds most of the induced currents directly to the output layer by jump connections and focuses on the other induced currents. The network defines the multi-label combination loss function to reduce the nonlinearity of the objective function to accelerate convergence. Guo et al. (2021) proposed a complex-valued Pix2pix generative adversarial network. This network consists of two parts: the generator and the discriminator. The generator consists of multilayer complex-valued CNNs, and the discriminator calculates the maximum likelihood between the original value and the reconstructed value. By adversarial training between the discriminator and the generator, the generator can capture more nonlinear features than the conventional CNN. The work Tsakyridis et al. (2024) provides an overview and discussion of the basics of photonic neural networks and optical deep learning. Matuszewski et al. (2024) discussed the role of all-optical neural networks.

4 Deep neural network paradigms inspired by statistical physics

The field of artificial intelligence contains a wide range of algorithms and modeling tools to handle tasks in various fields and has become the hottest subject in recent years. In the previous chapters, we reviewed recent research on the intersection of artificial intelligence with classical mechanics and electromagnetism. This includes the conceptual development of artificial intelligence powered by physical insights, the application of artificial intelligence techniques to multiple domains in physics, and the intersection between these two domains. Below we describe how statistical physics can be used to understand AI algorithms and how AI can be applied to the field of statistical physics. An overview of the representative methods is shown in Table 3.

4.1 Unbalanced neural networks

The most general problem in nonequilibrium statistical physics is the detailed description of the time evolution of physical (chemical or astronomical) systems. For example, different phenomena tending towards equilibrium states, considering the response of the system to external influences, metastability, and instability due to fluctuations, pattern formation and self-organization, the emergence of probabilities contrary to deterministic descriptions, and open systems, etc. Nonequilibrium statistical physics has created concepts and models that are not only relevant to physics, but also closely related to information, technology, biology, medicine, and social sciences, and even have a great impact on fundamental philosophical questions.

4.1.1 Neural networks understood from entropy

Entropy Proposed by German physicist Clausius in 1865, it was first a basic concept in the development of thermodynamics. Its essence is the "inherent degree of chaos" of a system,

Table 3 An overview of methods for AI DNNs inspired by statistical physics

Categories	Detail	Typical Methods	Keywords	Publication & Years
Unbalanced neural networks	Neural networks understood from entropy	InfoMap Rosvall et al. (2009) Munkres' algorithm Riesen and Bunke (2009) Word2Vec Mikolov et al. (2013a)	Modularity maximization; Map equation Graph edit distance; Bipartite graph matching Continuous vector representations; Word similarity	Eur.Phys.J.S.Topics 2009 IVC 2009 arXiv 2013
Chaotic neural networks		Goldfield et al. (2024) Pool et al. (2016)	Entropy measures; Quantum simulator Riemannian geometry; Mean field theory; DNN	Phys. Rev. A 2024 NeurIPS 2016
Ising models		Keup et al. (2021) DCNN Mohanrasu et al. (2023) Hopfield NN Hopfield (1982)	Spike communication; Rate units; Mean field theory Fractal interpolation functions; Vertical scaling factors	Phys. Rev. X 2021 Appl. Math. Model. 2023 PNAS 1982
Classic simulated annealing algorithms		HCNN Liu et al. (2019) MHNNs Lin et al. (2023) Ma et al. (2024) Laydevant et al. (2024)	Neurobiology; Collective properties; Ising Model Hopfield neural networks; Chaotic networks Variational autoregressive; Spin variables; Ising Model	IEEE Access 2019 Math. 2023 arXiv 2024 D-Wave Ising; Equilibrium propagation; Annealing
		OSA Kirkpatrick et al. (1983)	Combinatorial optimization; Annealing	Nat. Commun.2024
		DBN's Salakhutdinov and Murray (2008) Langevin-SA Bras and Pagès (2023)	RBM, AIS; Generative models Stochastic optimization; Langevin equation; Simulated annealing	ICML 2008 MCM.Appl. 2023

Table 3 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
Boltzmann machine neural networks	SAM/T Karacan et al. (2023)	Mixed-integer programming; Parallel simulated annealing		Processes 2023
	Milisav et al. (2024)	Morphospace representation; Randomized; SA		bioRxiv 2024
	CRBM Lee et al. (2009)	Hierarchical generative models; DBN		ICML 2009
	SRBM Lee et al. (2007)	Sparse coding; Visual area V2; DBN		NeurIPS 2007
	FRBM Feng and Chen (2016)	RBM; Symmetric triangular fuzzy numbers		ITrans. Fuzzy Syst. 2016
	NBM Lang et al. (2023)	Conditional generative models; highly flexible		arXiv 2023
Energy models design neural networks	GAN Goodfellow et al. (2014)	Minimax game; Multilayer perceptrons		NeurIPS 2014
	PixelRNN Van Oord et al. (2016)	Recurrent layers; Natural image modeling		ICML 2016
	CycleGAN Zhu et al. (2017)	Image2image; Adversarial loss; Cycle consistency loss		ICCV 2017
	StyleGAN Karras et al. (2019)	Style transfer; Latent factors of variation		CVPR 2019
	Wang et al. (2019)	High-dimensional; Deterministic process; SDE		NeurIPS 2019
	EGC Guo et al. (2023)	Joint distribution; Classifier and generator		ICCV 2023
Variational autoencoder models	FreeDoM Yu et al. (2023)	Conditional diffusion models; Energy functions		ICCV 2023
	VAEs Kingma and Welling (2013)	Latent variables; SGD; Stochastic variational inference		arXiv 2013

Table 3 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
Auto-regressive models	NVAE Vahdat and Kautz (2020) Cui et al. (2023) PixelCNN++ Salimans et al. (2017) XLNet Yang et al. (2019)	Auto-regressive; Deep energy-based models EBM prior; Variational joint learning Downsampling; Short-cut connections Bidirectional contexts; Autoregressive formulation; Transformer-XL	ICCV 2023 arXiv 2017 NeurIPS 2019	NeurIPS 2020
Dissipative structure neural networks	AR-Diffusion Wu et al. (2023) SOM Kohonen (1989) Budroni and De Wit (2017)	Diffusion models; Sequential dependency Weight Vector; Processing Unit; Minimal Span Tree Reaction-diffusion processes; Turing patterns; Chemo-hydrodynamic Non-convex error functions; Saddle points; Newton method	arXiv 2023 SOAM 1989 Chaos 2017	arXiv 2023
Random surface neural networks	SFN Dauphin et al. (2014) Choromanska et al. (2015)	Non-convex loss function; Spherical spin-glass model	NeurIPS 2014 AISTATS 2015	NeurIPS 2014
Free energy surface neural networks	Kawaguchi (2016) PoE Hinton (2002) Schneider et al. (2017)	Non-convex and non-concave loss function; Saddle points; Gap function; Latent-variable models; Rapid inference; Contrastive divergence Free energy landscapes; Ensemble averages; ANN	NeurIPS 2016 Neural Comp. 2002 Phys. Rev. Lett. 2017	NeurIPS 2016
KD to optimize neural networks	Sidky and Whitmer (2018) Noé et al. (2019) KDNN Hinton et al. (2015)	Adaptive bias techniques; ANN; Bayesian regularization Molecular dynamics; Monte Carlo methods; Invertible transformations Ensemble learning; Model averaging; Knowledge compression	J. Chem. Phys. 2018 Science 2019 arXiv 2015	Science 2019

Table 3 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
NST	Huang and Wang (2017)	Knowledge Transfer; Maximum Mean Discrepancy		arXiv 2017
BAN	Furlanello et al. (2018)	DenseNets; Distillation objectives; Born-Again Networks		ICML 2018
DML	Zhang et al. (2018)	Model distillation; Deep mutual learning		CVPR 2018
PCD	Huang and Guo (2023)	Dense prediction tasks; patialAdaptor		ICCV 2023
DaFKD	Wang et al. (2023)	Federated Distillation; Domain knowledge aware		CVPR 2023
DiffKD	Huang et al. (2024)	Diffusion; Adaptive noise matching		NeurIPS 2024
NEO-KD	Ham et al. (2024)	Neighbor knowledge distillation; Adversarial training		NeurIPS 2024
NAS-KD	Cream Peng et al. (2020)	NAS; Architecture distillation; Prioritized paths		NIPS 2020
DNA	Li et al. (2020)	Block-wise architecture search; ImageNet		CVPR 2020
DFA	Guan et al. (2020)	Feature aggregation; Differentiable architecture search		ECCV 2020
OKD	Kang et al. (2020)	Ensemble teacher networks; Oracle knowledge distillation		AAAI 2020
RNAS-CL	Nath et al. (2023)	Robust Neural Architecture Search; KD		arXiv 2023
MF-KD	Trofimov et al. (2023)	Bayesian optimization; Multi-fidelity optimization		IEEE Access 2023

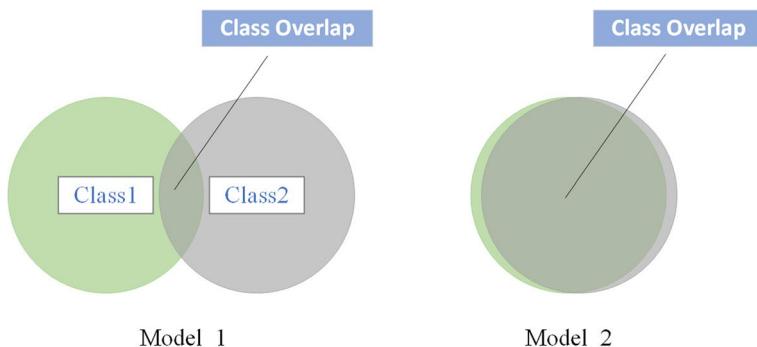


Fig. 10 De-redundancy

or the amount of information in a system (the more chaotic the system, the less the amount of information, the more difficult it is to predict, and the greater the information entropy), which is recorded as S in the formula. It summarizes the basic development law of the universe: things in the universe have a tendency to spontaneously become more chaotic, which means that entropy will continue to increase, which is the principle of entropy increase.

Boltzmann distribution In 1877, Boltzmann proposed the physical explanation of entropy: the macroscopic physical property of the system, which can be considered as the equal probability statistical average of all possible microstates.

Information entropy (learning cost) Until the development of statistical physics and information theory, Shannon extended the concept of entropy in statistical physics to the process of channel communication Shannon (1948) in 1948, and proposed information entropy and the universal significance of entropy became more obvious.

In deep learning, the speed at which the model receives information is fixed, so the only way to speed up the learning progress is to reduce the amount of redundant information in the learning target. The so-called “removing the rudiments and saving the essentials” is the principle of minimum entropy in the deep learning model, which can be understood as “removing unnecessary learning costs”(Fig. 10).

Application of algorithms inspired by the principle of minimum entropy, such as using information entropy to represent the shortest code length, InfoMap (Rosvall et al. 2009; Rosvall and Bergstrom 2008), cost minimization (Kuhn 1955; Riesen and Bunke 2009), Word2Vec (Mikolov et al. 2013a, b), t-SNE dimensionality reduction Maaten and Hinton (2008), etc.

4.1.2 Chaotic neural networks

Chaos refers to the unpredictable, random-like motion of a deterministic dynamic system because it is sensitive to initial values. Poole et al. (2016) published on NIPS in 2016 combines Riemannian geometry and dynamic mean field theory Sompolinsky et al. (1988) to analyze signals through the propagation of stochastic deep networks and form variance weights and biases in the phase plane. This work reveals the dynamic phase transition of signal propagation between ordered and chaotic states. Lin and Chen (2009) proposed a chaotic dynamic neural network based on a sinusoidal activation function, which is different from other models and has strong memory storage and retrieval capabilities. The 2020

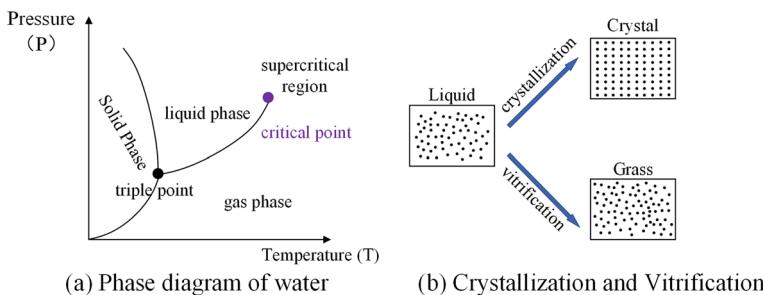


Fig. 11 Liquid, solid phase transition

edition of Keup et al. (2021) develops a statistical mean-field theory for random networks to solve transient chaos problems.

4.1.3 From Ising models to Hopfield networks

In everyday life, we see phase transitions everywhere changing from one phase to another. For example: liquid water is cooled to form ice, or heated and evaporated into water vapor (liquid phase to solid phase, liquid phase to gas phase). According to Landau's theory, the process of phase transition must be accompanied by some kind of "order" change. For example, liquid water molecules are haphazardly arranged, and once frozen, they are arranged in a regular and orderly lattice position (molecules vibrate near the lattice position, but not far away), so water freezes. The crystal order is created during the liquid–solid phase transition, as shown in Fig. 11.

Another important example of a phase transition is the ferromagnetic phase transition: a process in which a magnet (ferromagnetic phase) loses its magnetism and becomes a paramagnetic phase during heating. In the process of ferromagnetic phase transition (Fig. 12), the spin orientation of atoms changes from a random state in the paramagnetic phase to a specific direction, so the ferromagnetic phase transition is accompanied by the generation of spin orientation order, resulting in the macroscopic magnetism (spontaneous magnetization) of the material. According to Landau's theory, the order parameter changes continuously/discontinuously in the continuous/discontinuous phase transition, respectively.

Exactly 100 years ago, the mathematical key to solving the phase transition problem appeared, that is, the "primary version" of the spin glass model - the Ising model (the basic model of phase transition). The Ising model (also called the Lenz-Ising model) is one of the most important models in statistical physics. In 1920–1924, Wilhelm Lenz and Ernst Ising proposed a class of Ising describing the stochastic process of the phase transition of matter model. Taking the two-dimensional Ising lattice model as an example, the state of any point $p(s_i)$ can have two values ± 1 (spin up or down), and is only affected by the point adjacent to it (interaction strength J), the energy of the system can be obtained (Hamiltonian): For the Ising model, if all the spins are in the same direction, the Hamiltonian of the system is at a minimum, and the system is in the ferromagnetic phase. Likewise, the second law of thermodynamics tells us that, given a fixed temperature and entropy, the system seeks a configuration method that minimizes its energy, using the Gibbs-Bogoliubov-Feynman inequality to perform variational inference on the Ising model to obtain the optimal solution. In 1982, Hopfield, inspired by the Ising model, proposed a Hopfield neural

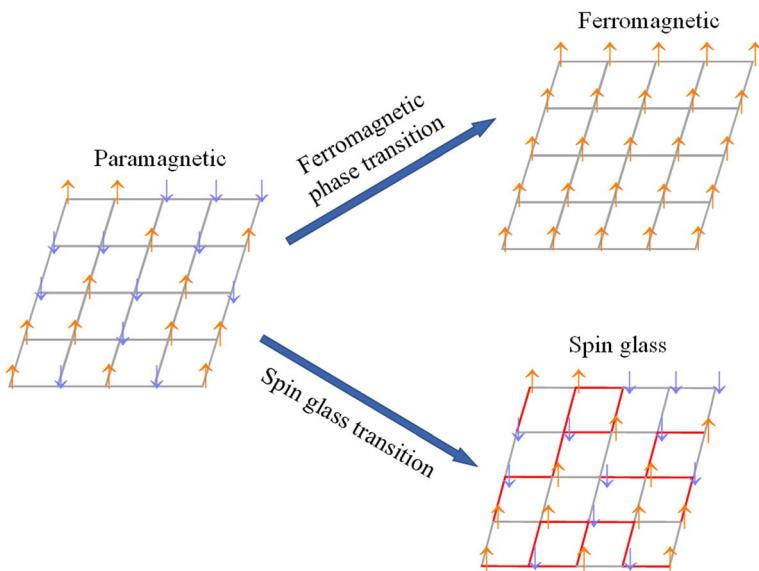


Fig. 12 Ferromagnetic phase transitions and spin-glass phase transitions. The grey edges represent ferromagnetic interactions, and the red edges represent antiferromagnetic interactions

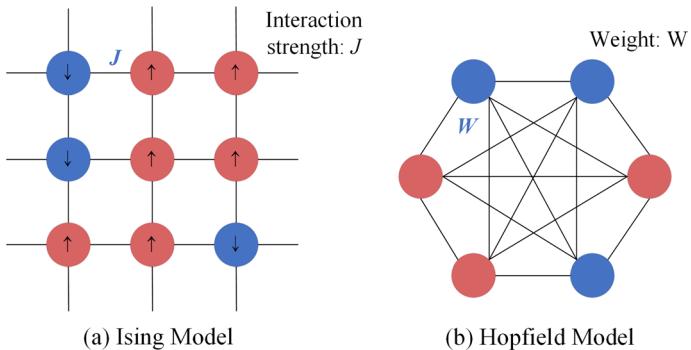


Fig. 13 Ising model and Hopfield network analogy diagram

network Hopfield (1982) that can solve a large class of pattern recognition problems and give approximate solutions to a class of combinatorial optimization problems. Its weight is to simulate the adjacent spin coherence of the Ising model; the neuron update is to simulate the Cell update in the Ising model. The unit of the Hopfield network (full connection) is binary, accepting a value of -1 or 1, or 0 or 1; it also provides a model that simulates human memory(Ising model and Hopfield network analogy diagram as shown in Fig. 13).

Hopfield formed a new calculation method with the idea of the energy function and clarified the relationship between neural networks and dynamics. He used the nonlinear dynamics method to study the characteristics of this neural network, and established the neural network stability criterion. At the same time, he pointed out that information is

stored on the connections between the various neurons of the network, forming the so-called Hopfield network. By comparing the feedback network with the Ising model in statistical physics, the upward and downward directions of the magnetic spin are regarded as two states of activation and inhibition of the neuron, and the interaction of the magnetic spin is regarded as the synaptic weight of the neuron value. This analogy paved the way for a large number of physical theories and many physicists to enter the field of neural networks. In 1984, Hopfield designed and developed the circuit of the Hopfield network model, pointing out that neurons can be implemented with operational amplifiers and the connection of all neurons can be simulated by electronic circuits, which is called a continuous Hopfield network. Using this circuit, Hopfield successfully solved the traveling salesman (TSP) computational puzzle (optimization problem).

Liu et al. (2019) discuss an image encryption algorithm based on the Hopfield chaotic neural network. This algorithm simultaneously scrambles and diffuses color images by utilizing the iterative process of a neural network to modify the pixel values. The encryption process results in highly randomized and complex encrypted images. During decryption, the original image is restored by reversing the iterative process of the Hopfield neural network.

In 2023, Lin et al. (2023) review the research on chaotic systems based on memory impedance Hopfield neural networks. It explores the construction method of chaotic systems using these neural networks, which incorporate memory impedance to preserve resistance changes. The article discusses the properties and applications of chaotic systems achieved through adjusting network parameters and connection weights. These studies offer new ideas and methods for understanding and applying image encryption and chaotic systems. Ma et al. (2024) proposed a variational autoregressive architecture with a message-passing mechanism, which can effectively exploit the interactions between spin variables. Laydevant et al. (2024) Train Ising machines in a supervised manner via a balanced propagation algorithm, which has the potential to enhance machine learning applications.

4.1.4 Classic simulated annealing algorithms

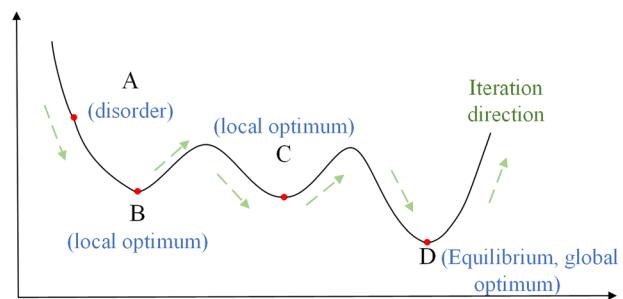
Physical annealing process: First the object is in an amorphous state, then the solid is heated to a sufficiently high level to be disordered, and then slowly cooled, annealing to a crystal (equilibrium state).

The simulated annealing algorithm was first proposed by Metropolis et al. In 1983, Kirkpatrick et al. applied it to combinatorial optimization to form a classical simulated annealing algorithm Kirkpatrick et al. (1983): Using the similarity between the annealing process of solid matter in physics and general optimization problems; Starting from a certain initial temperature, with the continuous decrease of temperature, combined with the probabilistic sudden jump characteristic of the Metropolis criterion (accepting a new state with probability), it searches in the solution space, and stays at the optimal solution with probability 1 (Fig. 14).

Importance Sampling (IS) is an effective variance reduction algorithm for rare events, as described in the seminal work by Marshall (1954). The fundamental concept of IS involves approximating the computation by taking a random weighted average of a simpler distribution function, representing the objective function's mathematical expectation.

Inspired by the idea of annealing, Radford proposed Annealed Importance Sampling (AIS) Salakhutdinov and Murray (2008) as a solution to address the high bias associated with IS. AIS, along with its extension known as Hamiltonian Annealed Importance

Fig. 14 Global optimal search process



Sampling (HAIS) Sohl-Dickstein and Culpepper (2012), represents generalizations of IS that enable the computation of unbiased expectations by reweighting samples from tractable distributions.

In AIS, a bridge is constructed between forward and reverse Markov chains, connecting the two distributions of interest. This bridge allows for the estimation of lower variance compared to what IS alone can provide. By leveraging the connections between the forward and reverse chains, AIS offers improved accuracy and efficiency in estimating expectations for rare events. In summary, Importance Sampling (IS) is a variance reduction algorithm for rare events, while Annealed Importance Sampling (AIS) and its extension HAIS provide solutions to overcome the bias issues associated with IS. AIS constructs a bridge between forward and reverse Markov chains, allowing for lower variance estimates than IS alone can achieve. These techniques offer improved accuracy and efficiency in estimating expectations for challenging problems involving rare events.

Later, Ranzato's MCRBM model (2010) Bengio et al. (2013), Dickstein's non-equilibrium diffusion model (2015) Sohl-Dickstein et al. (2015) and Menick's self-scaling pixel network autoregressive model (2016) Oord et al. (2016) followed Come. To adapt the network null model to weighted network inference, Milisav et al. (2024) proposed a simulated annealing process to generate random networks with strength sequence preservation. The simulated annealing algorithm is widely used and can efficiently solve NP-complete problems, such as the Travelling Salesman Problem, Max Cut Problem, Zero One Knapsack Problem, Graph Colouring Problem, and so on.

4.1.5 Boltzmann machine neural networks

Hinton proposed the Boltzmann Machine (BM) in 1985, BM is often referred to in physics as the inverse Ising model. BM is a special form of log-linear Markov random field (MRF), that is, the energy function is a linear function of the free variables. It introduces statistical probability in the state change of neurons, the equilibrium state of the network obeys Boltzmann distribution, and the network operation mechanism is based on a simulated annealing algorithm (Fig. 15), which is a good global optimal search method and is widely used in a certain range. See Nguyen et al. (2017) for the latest research on Boltzmann machines.

A Restricted Boltzmann Machine (RBM) is a type of Boltzmann Machine (BM) that exhibits a specific structure and interaction pattern between its neurons. In an RBM, the neurons in the visible layer and the neurons in the hidden layer are the two variables that interact through efficient coupling. Unlike a general BM, where all neurons can interact with each other, an RBM restricts the interactions to occur exclusively between the visible and hidden units.

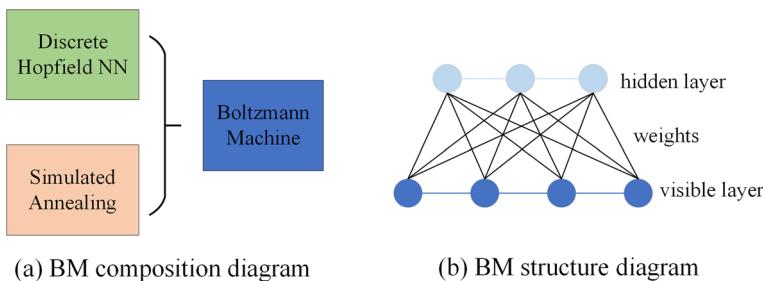


Fig. 15 BM composition and structure diagrams

The RBM's goal is to adjust its parameters in a way that maximizes the likelihood of the observed data. By learning the weights and biases of the connections between the visible and hidden units, the RBM aims to capture and represent the underlying patterns and dependencies present in the data. Through an iterative learning process, the RBM adjusts its parameters to improve the likelihood of generating the observed data and, consequently, enhance its ability to model and generate similar data instances.

Regarding RBMs, there are many studies in physics that shed light on how they work and what structures can be learned. Since Professor Hinton proposed RBM's fast learning algorithm contrast divergence, in order to enhance the expressive ability of RBM and take into account the specific structure of the data, many variant models of RBM have been proposed (Bengio 2009; Ranzato et al. 2010; Ranzato and Hinton 2010). Convolutional Restricted Boltzmann Machine (CRBM) Lee et al. (2009) is a new breakthrough in the RBM model. It uses filters and image convolution operations to share weight features to reduce the parameters of the model. Since most of the hidden unit states learned by RBM are not activated (non-sparse), researchers combined the idea of sparse coding to add a sparse penalty term to the log-likelihood function of the original RBM and proposed a sparse RBM model Lee et al. (2007), a sparse group restricted Boltzmann machine (SGRBM) model Salakhutdinov et al. (2007) and LogSumRBM model Ji et al. (2014), etc. In the articles (Cocco et al. 2018; Tubiana and Monasson 2017), the authors investigate a stochastic Restricted Boltzmann Machine (RBM) model with random, sparse, and unlearned weights. Surprisingly, they find that even a single-layer RBM can capture the compositional structure using hidden layers. This highlights the expressive power of RBMs in representing complex data.

Additionally, the relationship between RBMs with random weights and the Hopfield model is explored in Barra et al. (2018), Mézard (2017). These studies demonstrate the connections and similarities between RBMs and the Hopfield model, shedding light on the underlying mechanisms and properties of both models.

Overall, these works provide insights into the capabilities of RBMs with random weights in capturing compositional structures and their connections to the Hopfield model. Such research enhances our understanding of RBMs and their potential applications in various domains.

4.2 Energy models design neural networks

According to physical knowledge, the steady state of a thing actually represents its corresponding state with the lowest potential energy. Therefore, the steady state of a thing

corresponds to the lowest state of a certain energy and is transplanted into the network, thus constructing the definition of an energy function when the network is in a steady state.

In 2006, Lecun et al. reviewed the energy model-based neural network and its application. When the model reaches the optimal solution, it is in the lowest energy state (that is, it seeks to minimize positive data versus energy and maximize negative data versus energy) LeCun et al. (2006). The task is to find the configuration of those hidden variables that minimize the energy value given the observed variables (inference); and to find an appropriate energy function such that the energy of the observed variables is lower than that of the hidden variables (learning).

Normalized probability distributions are difficult to implement in high-dimensional spaces, leading to an interesting approach to generative modeling of data Pernkopf et al. (2014). Normalization can still be done analytically when normalizing (Dinh et al. 2014, 2016; Rezende et al. 2016), these interesting methods can be found in the reference Wang (2018).

4.2.1 Generative adversarial networks (GANs)

In 2014, Goodfellow et al. proposed a GAN Goodfellow et al. (2014) that aims to generate samples of the same type as the training set, which essentially uses learned discriminator judgments to replace explicit evaluation of probabilities, Unsupervised learning can be performed using the knowledge acquired during the supervised learning process. Physics-inspired GAN research is beginning to emerge, such as Wang et al. (2019) generalizing perceptrons in interpretable models of GANs using early online-learned statistical physics work.

Both the discriminator and generator of Deep Convolutional Generative Adversarial Networks (DCGAN) Radford et al. (2015) use CNN to replace the multilayer perceptron in GAN, which can connect supervised and unsupervised learning together. CycleGAN Zhu et al. (2017) can achieve mode conversion between the source domain and the target domain without establishing a one-to-one mapping between training data. GCGAN Fu et al. (2019) is to add convolution constraints to the original GAN, which can stabilize the learning configuration. WGAN Arjovsky et al. (2017) has improved the loss function based on GAN, and can also get good performance results on the full link layer.

4.2.2 Variational autoencoder models (VAEs)

Autoencoder (AE) is a feedforward neural network that aims to find a concise representation of data that still maintains the salient features of each sample, and an autoencoder with linear activation is closely related to PCA. VAE Kingma and Welling (2013) combines variational reasoning and autoencoders to simulate the transformation between energy distribution functions - building a generative adversarial network provides a deep generative model for the data, generating target data X from latent variables Z, which can be trained in an unsupervised manner. The VAE model is closer to a variant of the physicist's mindset, in which the autoencoder is represented by a graphical model and uses latent variables and variational priors for training inference (Cinelli et al. 2021; Vahdat and Kautz 2020). Rezende et al. (2014) is a fundamental version of understanding VAE.

An interesting approach to generative modeling involves decomposing the probability distribution into a product of one-dimensional conditional distributions in the autoregressive model, as discussed in the work by Van Oord et al. (2016). This decomposition allows

for efficient modeling of complex high-dimensional data, such as images, by sequentially generating each dimension conditioned on the previous dimensions.

In the context of variational autoencoders (VAEs), another intriguing approach is to replace the posterior distribution with a tractable variational approximation. This idea was introduced in the seminal works by Kingma and Welling (2013), Gregor et al. (2014), and Rezende et al. Ozair and Bengio (2014). By introducing an encoder network that maps the input data to a latent space and a decoder network that reconstructs the data from the latent space, VAEs enable efficient and scalable generative modeling.

These techniques, namely decomposing probability distributions in autoregressive models and using tractable variational approximations in VAEs, offer interesting and effective strategies for generative modeling. They provide insights into modeling complex data distributions and have found applications in various domains, including image generation and data synthesis.

4.2.3 Auto-regressive generative models

Auto-regressive generative model (Van Oord et al. 2016; Salimans et al. 2017) is a controllable method for modeling distributions that allow maximum likelihood training without latent random variables, where the conditional probability distribution is represented by a neural network. Since this model is a family of displayed probabilities, direct and unbiased sampling is possible. The application of these models has been realized in statistics Wu et al. (2019) and quantum physics problems Sharir et al. (2020).

Neural Autoregressive Distribution Estimation (NADE) is an unsupervised neural network built on top of autoregressive models and feedforward neural networks Zhang et al. (2019), which is a tractable and efficient estimator for modeling data distribution and density.

4.2.4 RG-RBM models

In a 2014 paper by Mehta and Schwab (2014), the concept of renormalization is applied to explain the performance of deep learning models. Renormalization is a technique used to study physical systems when detailed information about their microscopic components is unavailable, providing a coarse-grained understanding of the system's behavior across different length scales.

The authors propose that deep neural networks (DNNs) can be viewed as iterative coarse-graining schemes, similar to the renormalization group (RG) theory. In this context, each new high-level layer of the neural network learns increasingly abstract and high-level features from the input data. They argue that the process of extracting relevant features in deep learning is fundamentally the same as the coarse-graining process in statistical physics, as DNNs effectively mimic this process.

The paper highlights the close connection between RG and Restricted Boltzmann Machines (RBM) and suggests a possible integration of the physical conceptual framework with neural networks. This mapping between RG and RBM provides insights into the relationship between statistical physics and deep learning.

Overall, Mehta and Schwab's work demonstrates how renormalization can be applied to understand the performance of deep learning models. It emphasizes the similarity between feature extraction in deep learning and the coarse-graining process in statistical physics.

The mapping between RG and RBM offers a potential explanation for the combination of physical concepts and neural networks.

4.3 Dissipative structure neural networks

The theory of self-organization is that when an open system reaches a nonlinear region far away from the equilibrium state, once a certain parameter of the system reaches a certain threshold, the system can undergo a mutation through fluctuations, from disorder to order, and produce self-organization phenomena such as chemical oscillations. It consists of dissipative structure (disorder to order), synergy (synergy of various elements of the system), and mutation theory (threshold mutation).

Self-organizing feature map (SOM) (Kohonen 1989, 1990) was proposed by Professor Kohonen, when the neural network accepts external input, SOM will be divided into different regions, and each region has different response characteristics to the input mode. It self-organizes and adaptively changes the network parameters and structure by automatically finding the inherent laws and essential attributes in the samples. The self-organizing (competitive) neural network is an artificial neural network that simulates the functions of the above-mentioned biological nervous system. That is, in terms of the learning algorithm, it simulates the dynamic principle of information processing of excitation, coordination and inhibition, and competition between biological neurons to guide the study and work of the network. Since SOM is a tool that can visualize high-dimensional data and can effectively compress the transmission of information, Kohonen et al. (1996) summarizes some engineering applications of SOM.

A dissipative structure is when the system is far away from thermodynamic equilibrium, under certain external conditions. Due to the nonlinear interaction within the system, a new ordered structure can be formed through mutation, which is an important new aspect of non-equilibrium statistics in the physics branch. In 2017, Amemiya et al. discovered and outlined the role of glycolytic oscillations in cell rhythms and cancer cells Amemiya et al. (2017). In 2017, Kondepudi, et al. discussed the relevance of dissipative structures in understanding organisms and proposed a voltage-driven system Kondepudi et al. (2017) that can exhibit behaviors that are surprisingly similar to those we see in organisms. In the same year Burdoni and De Wit discussed how the interplay between reaction and diffusion produces localized spatiotemporal patterns Budroni and De Wit (2017) when different reactants come into contact with each other.

4.4 Random surface neural networks

In the field of artificial intelligence, early research was heavily influenced by the theoretical guarantees offered by optimization over convex landscapes, where each local minimum is also a global minimum Boyd et al. (2004). However, when dealing with non-convex surfaces, the presence of high-error local minima can impact the dynamics of gradient descent and affect the overall performance of optimization algorithms.

The statistical physics of smooth random Gaussian surfaces in high-dimensional spaces has been extensively studied, yielding various surface models that connect spatial information to probability distributions (Bray and Dean 2007; Fyodorov and Williams 2007). These models provide insights into the behavior and properties of non-convex surfaces, shedding light on the challenges posed by high-dimensional optimization problems.

In 2014, Dauphin et al. studied the connection between the neural network error surface model and statistical physics, that is, the connection between the energy functions of spherical spin glasses Choromanska et al. (2015).

In 2014, Pascanu proposed the Saddleless Newton Algorithm (SFN) in Dauphin et al. (2014) for the problem that high-dimensional non-convex optimization has a large number of saddle points instead of local extremums. It can quickly escape the saddle point where gradient descent is slowed down. Furthermore Kawaguchi (2016) introduces random surfaces into deeper networks.

By examining the statistical physics of random surfaces, researchers have gained a better understanding of the complex landscapes encountered in non-convex optimization. This knowledge has implications for improving optimization algorithms and enhancing the performance of artificial intelligence systems operating in high-dimensional spaces.

To summarize, research in statistical physics has explored different surface models to analyze the behavior of non-convex optimization landscapes. Understanding the properties of these surfaces is important not only for solving the challenges associated with high-dimensional optimization problems, but also for improving the performance of artificial intelligence algorithms.

4.5 Free energy surface (FES) neural networks

Free energy refers to the part of the reduced internal energy of the system that can be converted into external work during a certain thermodynamic process. It measures the “useful energy” that the system can output to the outside during a specific thermodynamic process. It can be divided into Helmholtz-free energy and Gibbs-free energy. The partition function is equivalent to free energy.

In the context of energy-based models, researchers have proposed a number of approaches to overcome the difficulty of calculating with free energy. These methods include exhaustive Monte Carlo, contrastive divergence heuristics Hinton (2002) and its variants Tieleman and Hinton (2009), fractional matching Hyvärinen and Dayan (2005), pseudo-likelihood Besag (1975), and minimum probability flow learning (MPF) (Battaglino 2014; Sohl-Dickstein et al. 2011) (where MPF itself is based on non-equilibrium statistical mechanics). Despite these advances, training expressive energy-based models on high-dimensional datasets remains an open challenge.

In the domain of energy-based models, several approaches have been proposed to address the challenge of computing with free energy. These methods aim to train models effectively despite the computational difficulties associated with estimating free energy. Some notable approaches include:

Exhaustive Monte Carlo: This method involves sampling from the model’s distribution using Monte Carlo techniques, which can be computationally expensive for high-dimensional datasets.

Contrastive Divergence (CD) and its variants: CD is a popular heuristic proposed by Hinton (2002) for training energy-based models. It approximates the gradient of the model’s parameters by performing a few steps of Gibbs sampling. Variants of CD, such as Persistent Contrastive Divergence (PCD) Tieleman and Hinton (2009), aim to improve the training process by maintaining a persistent chain of samples.

Fractional Matching: This approach, introduced by Hyvärinen and Dayan (2005), involves estimating the model’s parameters by matching the moments of the model’s distribution with the moments of the data distribution.

Pseudo-Likelihood: Proposed by Besag (1975), this method approximates the likelihood of the model by considering the conditional probabilities of each variable given the others.

Minimum Probability Flow Learning (MPF): MPF, based on non-equilibrium statistical mechanics, is a technique for training energy-based models introduced by Battaglino (2014) and Sohl-Dickstein et al. (2011). It minimizes the difference between the model's distribution and the data distribution using flow-based dynamics.

Machine learning methods learn the FES of a system as a function of collective variables to optimize AI algorithms. Using a functional representation of the FES of a neural network, the sampling of high-dimensional spaces can be improved. For example, Schneider et al. proposed a learnable FES to predict the NMR spin-spin coupling model of solid xenon under pressure Schneider et al. (2017). In 2018, Sidky et al. proposed a small neural network for FES, which can use data points generated by dynamic (real-time) adaptive sampling for iterative training Sidky and Whitmer (2018). This model verifies that when new data is generated, a smooth representation of the full configuration space can be obtained. Wehmeyer and Noé (2018) proposes a time-lagged autoencoder approach to identify slowly changing collective variables in the example of peptide folding. In 2018, Mardt et al. proposed a variational neural network-based approach to identify important dynamical processes during protein folding simulations and provided a framework for unified coordinate transformation and FES exploration Mardt et al. (2018), providing insights into the underlying dynamics of the system. In 2019, Noé et al. proposed to use the Boltzmann generator to sample the equilibrium distribution of the collective space to represent the state distribution on FES Noé et al. (2019).

Despite these advances, training expressive energy-based models on high-dimensional datasets remains a challenging task. Ongoing research aims to develop more efficient and effective training methods to tackle this open challenge in the field.

4.6 Knowledge distillation to optimize neural networks

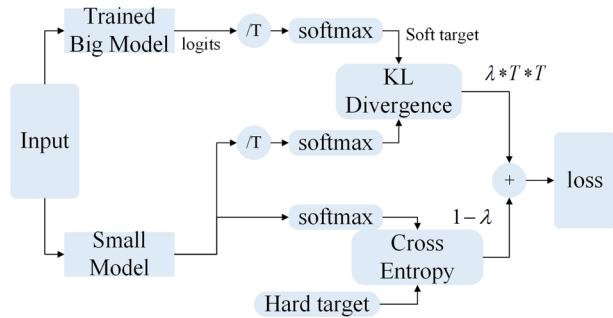
For neural networks: the larger the model, the deeper the layers, and the stronger the learning ability. In order to extract features from a large amount of redundant data, CNNs often require excessive parameters and larger models for training. However, the design of the model structure is difficult to design, so model optimization has become an important factor in solving this problem.

Knowledge distillation In 2015, Hinton's pioneering work, Knowledge Distillation (KD), promoted the development of model optimization Hinton et al. (2015). Knowledge distillation simulates the heating distillation in physics to extract effective substances and transfers the knowledge of the large model (teacher network) to the small model (student network), which makes it easy to deploy the model. In the process of distillation, the small model learns the generalization ability of the large model, speeds up the inference speed, and retains the performance close to the large model (Fig. 16).

4.6.1 Knowledge distillation neural networks

In 2017, TuSimple and Huang et al. proposed a distillation algorithm that uses the knowledge selection feature of neurons to transfer new knowledge (aligned selection style distribution), named Neuron Selectivity Transfer (NST) Huang and Wang (2017). NST models can be combined with other models to learn better features and improve performance. To enable the student network to automatically learn a good loss from the teacher network to

Fig. 16 Knowledge distillation process



preserve the relationship between classes and maintain polymorphism, Zheng et al. used conditional adversarial networks (CAN) in 2018 to build a teacher-student architecture Xu et al. (2017). The deep mutual learning (DML) model Zhang et al. (2018) and the Born Again Neural Networks (BAN) model Furlanello et al. (2018), which apply KD and do not aim to compress the model, were proposed in 2018. Huang et al. (2024) proposed a novel KD model that uses a diffusion model to explicitly denoise and match features, reducing computational costs. Ham et al. (2024) proposed a novel network based on a knowledge distillation adversarial training strategy, named NEO-KD, which improves robustness against adversarial attacks.

4.6.2 Network Architecture Search (NAS) and KD

KD transfers the knowledge in the teacher network to the student network, and there are a large number of networks in NAS, and the use of KD helps to improve the overall performance of the supernet. In 2020, Peng et al. proposed a network distillation algorithm based on priority paths to solve the inherent defects of weight sharing between models, that is, the problem of insufficient subnet training in HyperNetworks Peng et al. (2020), which improves the convergence of individual models. In the same year, Li et al. used the Distill the Neural Architecture (DNA) algorithm Li et al. (2020) to supervise the search for the internal structure of the network using knowledge distillation, which significantly improved the effectiveness of NAS. Wang et al. (2021) improves KL divergence by adaptively choosing alpha divergence, effectively preventing overestimation or estimating uncertainty in teacher models. Gu and Tresp (2020) combines network pruning and distilled learning to search for the most suitable student network. Kang et al. proposed the Oracle Knowledge Distillation (OKD) method in Kang et al. (2020), which distilled from the integrated teacher network and used NAS to adjust the capacity of the student network model, thereby improving the learning of the student network ability and learning efficiency. Inspired by BAN, Macko et al. (2019) proposes the Adaptive Knowledge Distillation (AKD) method to assist the training of the sub-network. To improve the efficiency and effectiveness of knowledge distillation, Guan et al. (2020) used differentiable feature aggregation (DFA) to guide the learning of the teacher network and the student network (network architecture search), and uses a method similar to differentiable architecture search (DARTS) Liu et al. (2018) to adaptively adjust the scaling factor.

4.7 DNNs to solve statistical physics classical problems

4.7.1 Rubik's cube problem

Professor Rubik invented the Rubik's Cube in 1974, initially called it "Magic Cube", and later the Rubik's Cube was featured in the Seven Towns toy business, issued by Ideal Toy Co and renamed "Rubik's Cube" European, Plastics, News, group (2015).

In 2018, DeepCube, a new algorithm without human assistance, solved the Rubik's cube by self-learning reasoning McAleer et al. (2018), which is a milestone in how to solve complex problems with minimal help. Agostinelli et al. proposed on Nature Machine Intelligence in 2019 to use the DL method DeepCubeA and search to solve the Rubik's cube problem Agostinelli et al. (2019), DeepCubeA can learn how to solve the Rubik's cube without any specific domain knowledge. Solve increasingly difficult Rubik's Cube problems in reverse from the target state. In 2021, Corli et al. introduced a deep reinforcement learning algorithm based on Hamiltonian reward and introduced quantum mechanics to solve the Rubik's cube problem in the combinatorial problem Corli et al. (2021). Colin's team, an associate professor at the University of Nottingham, published a paper on Expert Systems using a stepwise deep learning method to learn a "fitness function" to solve the Rubik's cube problem Johnson (2021) while highlighting the advantages of stepwise processing.

4.7.2 Neural networks to detect phase transition

Since each new high-level layer of a DNN learns more and more abstract high-level features from the data and previous layers can learn finer scales to represent the input data, researchers introduce renormalization into physics theory and extract macroscopic rules from microscopic rules. In 2017 Bradde and Bialek (2017) discussed the analogy between renormalization groups and principal component analysis. In 2018, Li and Wang et al., used neural networks to learn a new renormalization scheme (Koch-Janusz and Ringel 2018; Kamath et al. 2018).

Phase transitions are boundaries between different phases of matter, typically characterized by order parameters. However, neural networks have shown the ability to learn appropriate order parameters and detect phase transitions without prior knowledge of the underlying physics.

In a 2018 study by Morningstar and Melko (2017), unsupervised generative graphs were used to understand the probability distribution of two-dimensional Ising systems. This work demonstrated that neural networks can capture the essential features of the phase transitions in the Ising model.

The literature also provides positive evidence that neural networks can discriminate phase transitions in the Ising model. Carrasquilla and Melko (2017) and Wang (2016) utilized principal component analysis to detect phase transitions without prior knowledge of the system's physical properties.

Tanaka and Tomiya (2017) proposed a method for estimating specific phase boundary values from heatmaps, further demonstrating the possibility of discovering phase transition phenomena without prior knowledge of the physical system.

For a deeper understanding of these topics, interested readers can refer to the papers by Kashiwa et al. (2019) and Arai et al. (2018).

Overall, these studies highlight the potential of neural networks to identify and characterize phase transitions even without explicit knowledge of the underlying physics, opening up new avenues for studying complex systems and discovering emergent phenomena.

4.7.3 Protein sequence prediction and structural modeling

Protein sequence prediction and structural modeling are of great significance for providing valuable information in the fields of “AI + Big Health” such as precision medicine and drug research and development. In 2003 Bakk and Høye (2003) studied protein folding by introducing a simplified one-dimensional analogy of proteins composed of N-contacts (that is, using the one-dimensional Ising model). Stochastic RBM models Cocco et al. (2018) have recently been used to model protein families from their sequence information Tubiana et al. (2019). Analytical studies of the RBM learning process are extremely challenging, and this is usually done using a Gibbs sampling-based contrastive divergence algorithm Hinton (2002).

Wang et al. (2018) utilizes convolutional neural networks combined with extreme learning machine (ELM) classifiers to predict RNA-protein interactions. In 2019, Brian Kuhlman et al. reviewed the deep learning methods Kuhlman and Bradley (2019) that have been used for protein sequence prediction and 3D structure modeling problems. In Nature Communication, Ju et al. introduced a new neural network architecture, CopulaNet, which can extract features from multiple sequence alignments of target proteins and infer residue co-evolution, overcoming the defect of “information loss” in traditional statistical methods Ju et al. (2021).

4.7.4 Orderly glass-like structure design

Mehta’s experiments with the Ising model in Bukov et al. (2018) provide some initial ideas in this direction, highlighting the potential usefulness of reinforcement learning for applications of equilibrium quantities beyond quantum physics. In 2019, Greitemann and Liu et al. introduced and studied a kernel-based learning method in Greitemann et al. (2019), Liu et al. (2019), which is used to learn phases in frustrated magnetic materials, is easier to interpret and able to identify complex order parameters.

In 2016, Nussinov et al. also studied ordered glass-like solids, using multi-scale network clustering methods to identify the spatial and spatiotemporal structure of glasses Cubuk et al. (2015), learn to identify structural flow defects. It is also possible to discern subtle structural features responsible for the heterogeneous dynamics observed in broadly disordered materials. In 2017, Wetzel et al. used unsupervised learning for Ising and XY models Wetzel (2017), and in 2018 Wang and Zhai et al. in frustrated spin systems unsupervised learning also introduced in Wang and Zhai (2017), Wang and Zhai (2018), beyond the limitations of supervised learning, more be classified.

4.7.5 Prediction of nonlinear dynamical systems

AI also provides robust systems for studying, predicting, and controlling nonlinear dynamical systems. In 2016, Reddy et al. used reinforcement learning to teach autonomous gliders to use the heat in the atmosphere to make them fly like birds (Reddy et al. 2016, 2018). In 2017, Pathak et al. used a recurrent neural network or reservoir computer called an echo state network Jaeger and Haas (2004) to predict trajectories of chaotic dynamical systems

and a model for weather forecasting Pathak et al. (2018). Graafland et al. (2020) uses BNS to build data-driven complex networks to solve climate problems. The network topology of the correlated network (CNS) contains redundant information. Bayesian Networks (BNS), on the other hand, only include non-redundant information (from a probabilistic perspective) and thus can extract informative physical features from them using sparse topologies. Boers et al. (2014) used the extreme event synchronization method to study the global pattern of extreme precipitation and attempted to predict rainfall in South America. Ying et al. (2021) used the same method to study the carbon cycle and carbon emissions and formulated strategies and countermeasures for carbon emissions and carbon reduction. Chen et al. (2021) applies the method of Eigen Microstates to the distribution and evolution of ozone on different structures. Zhang et al. (2021) changes the traditional ETAS model for earthquake prediction by considering the memory effect of earthquakes through the long-term memory model. Uncertainty in ocean mixing parameters is a major source of bias in ocean and climate modeling, and traditional physics-driven parameterizations that lack process understanding perform poorly in the tropics. Zhu et al. (2022) exploring data-driven approaches to parameterize ocean vertical mixing processes using deep learning methods and long-term turbulence measurements, demonstrated good performance using limited observations good physical constraints generalization capabilities, and improved physical information for climate simulations.

5 Deep neural network paradigms inspired by quantum mechanics

Quantum algorithms are a class of algorithms that operate on a quantum computing model. By drawing on fundamental features of quantum mechanics, such as quantum superposition or quantum entanglement, quantum algorithms. Compared to traditional algorithms, quantum mechanics have a dramatic reduction in computational complexity, which can even reach exponential reductions. Back in 1992, David Deutsch and Richard Jozsa proposed the first quantum algorithm, the Deutsch-Jozsa algorithm Deutsch and Jozsa (1992). The algorithm requires only one measurement to determine the class to which the unknown function in the Deutsch-Jozsa problem belongs. Although this algorithm lacked practicality, it led to a series of subsequent traditional quantum algorithms. In 1994, Peter W. Shor (1994) proposed the famous quantum large number prime factorization algorithm, called the Shor algorithm. The computational complexity of traditional factorization algorithms varies exponentially with the size of the problem, however, the Shor algorithm can solve the prime factorization problem in polynomial time. In 1996, Lov K. Grover (1996) proposed the classical quantum search algorithm, also known as the Grover algorithm, which has a complexity of $O\sqrt{N}$ with a quadratic level of efficiency improvement compared to traditional search algorithms. Nature-inspired stochastic optimization algorithms have long been a hot topic of research. Recent work Sood (2024) provides a comprehensive overview of quantum-inspired metaheuristic algorithms, while work Kou et al. (2024) summarizes quantum dynamic optimization algorithms. An overview of the representative methods is shown in Table 4.

5.1 Quantum machine learning

Quantum machine learning(QML) is a combination of the speed of quantum computing and the learning and adaptation capabilities provided by machine learning. By simulating

Table 4 An overview of methods for AI DNNs inspired by quantum mechanics

Categories	Detail	Typical Methods	Keywords	Publication & Years
Machine learning	Unsupervised learning	QLA Lloyd et al. (2013) QPCA Lloyd et al. (2014)	Quantum machine learning; Cluster assignment; Exponential speed-up Quantum state; Density matrix; Quantum principal component analysis	arXiv 2013 Nat. Phys. 2014
	Supervised learning	QLDA Cong and Duan (2016)	Discriminant analysis; Quantum computation; Dimensionality reduction	NJP 2016
Deep learning	Multilayer perceptrons	QKNN Wiebe et al. (2014)	Nearest-neighbor learning; Distance metrics; Query complexity	arXiv 2014
		QDT Lu and Braunstein (2014)	Decision tree; Quantum entropy impurity criterion; Quantum states	QIP 2014
		QSVM Rebenrost et al. (2014)	Support vector machine; Matrix inversion; Quantum computer	PRL 2014
		QUINN Menneer and Narayanan (1995)	Neural network training; Many universes interpretation; Efficiency	TR 1995
		QCN Toth et al. (1996)	Coupled quantum dots; Coulomb forces; Analog cellular neural network	Super. Micr. 1996
		QCiNN Matsui et al. (2000)	quantum circuits; Qubitlike neural network; Neural state	Electron. Commun. Jpn. 2000
		QNN Kouda et al. (2005)	Qubit neuron model; Quantum computing; Parity check problems	Neural Comput. 2005
		QPNN Zhou et al. (2006)	Classical perceptron network; XOR function; Artificial intelligence	ICANN 2006

Table 4 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
Recurrent neural networks	QRNN Schuld et al. (2014)	Quantum walks; Quantum neural networks; Associative memory		Phys. Rev. A 2014
Convolutional neural networks	QRNN Bausch (2020)	Recurrent neural networks; Variational quantum eigensolvers; Seq2Seq		NeurIPS 2020
	QLSTM Chen et al. (2020)	Variational quantum circuits; LSTM; ICASSP 2022 RNN		
	QCNN Cong et al. (2019)	Quantum physics; Many-body systems; Quantum error correction		Nat. Phys. 2019
	MQCNN Kerenidis et al. (2019)	Quantum tomography; Computational paradigm; Quantum computing		arXiv 2019
Evolutionary algorithms	QCCNN Liu et al. (2021)	Wave functions; Gate operations; Quantum computing		Sci. China Phys. Mech. Astron. 2021
Encoding algorithms	RCQGA Chen et al. (2005)	Fuzzy neural; Networks chaotic; Global optimization		Contr. Decis. 2005
	ARQEAs Joshi et al. (2016)	Non-linear Optimization; Metaheuristic; Quantum Entanglement		CIECH 2016
Evolutionary operators	GAQPR Li and Zhuang (2002)	Quantum probability representation; Crossover operator; Mutation operator		IDEAL 2002
	IQPSO Jin and Jin (2015)	Optimization algorithm; Improvement operation; Ensemble strategem		Signal Process. 2015

Table 4 (continued)

Categories	Detail	Typical Methods	Keywords	Publication & Years
Population optimization	Immune operators	QICA Jiao et al. (2008) ICCoA Shang et al. (2014)	Global optimization; Immune clonal algorithm; Multiuser detection Dynamic multiobjective optimization; Pareto-front; NSGA-II	ITrans. Syst. 2008 Nat. Comput. 2014
	QICA-CARP Shang et al. (2018)	Qubit; Quantum rotation gate; Convergence rate	Quantum evolutionary algorithm; L5-based neighbors; Cellular structure	Memet. Comput. 2018
	LSCQEA Qi and Xu (2015)	RP-QPSO Mei and Zhao (2018)	Economic dispatch; Random perturbation; QPSO	ITME 2015 DCABES 2018
	VQAs Bonet-Monroig et al. (2023)	Variational algorithms; Classical optimization algorithms	Phys. Rev. A 2023	
	QIRO Finžgar et al. (2024)	Quantum-informed; Recursive; Combinatorial optimization	PRX Quantum 2024	

the characteristics of superposition, entanglement, coherence, and parallelism possessed by microscopic particles, traditional machine learning algorithms are quantized to enhance their ability to represent, reason, learn, and associate with data.

In general, quantum machine learning algorithms have the following three steps: (1) Quantum state preparation. Take advantage of the high parallelism of quantum computing, the original data must be converted to the form of a quantum bit so that the data has quantum characteristics; (2) Quantum algorithm processing. Quantum computers are no longer part of the von Neumann machine and its operating units are completely different from the traditional computer, so it is necessary to quantumize and transplant the traditional algorithm to the quantum computer. The transplantation of the algorithm should be combined with both the data structure of traditional algorithms and the characteristics of quantum theory to effectively accelerate the traditional algorithm, which makes the usage of quantum algorithms meaningful; (3) Quantum measurement operations. The result is output in a quantum state, which itself exists in the form of probability. By quantum measurement, quantum superposition wave packets collapse to classical states to extract the information contained in quantum states for subsequent information processing.

The history of QML can be traced back to 1995, and Subhash C. Kak (1995) first introduced the concept of "quantum neural computing". Kak considered quantum computers as a collection of conventional computers that can respond to stimuli and reorganize themselves to perform efficient computations. In the same way as traditional machine learning algorithms, quantum machine learning algorithms can be classified according to the data format: quantum unsupervised learning, and quantum supervised learning.

5.1.1 Quantum unsupervised learning algorithms

Quantum k-means algorithm The clustering algorithm is one of the most important classes of unsupervised learning algorithms. Clustering means partitioning some samples without labels into different classes or clusters according to some specific criteria (e.g., distance criterion), so that the difference between samples in the same cluster is as small as possible, and the difference between samples in different clusters is as large as possible.

For unsupervised clustering algorithms, the K-Means algorithm is the most common one. Its core idea is that for given a dataset consisting of U samples without labels and the number of clusters C ($C < U$), according to the distance between the sample and the centers of clusters, each sample is assigned to the nearest cluster:

$$\operatorname{argmin}_c |\mathbf{u} - \frac{1}{M} \sum_{j=1}^M \mathbf{v}_j^c| \quad (17)$$

where \mathbf{u} denotes the sample to be clustered and \mathbf{v}_j^c denotes the j th sample of class c . Then the centers of all clusters are iteratively updated until the position of centers converges. Since it is necessary to measure the distance between each sample and the center of every cluster and update the centers of all clusters when the K-means algorithm is performed, the time cost of the K-means algorithm will be very high when the number of clusters and samples is large.

In 2013, Lloyd et al. (2013) proposed the quantum version of Lloyd's algorithm for performing the K-Means algorithm. The main idea of this algorithm is the same as the traditional K-means algorithm, which compares the distances between quantum states, but the quantum states in Hilbert space have both entanglement and superposition and can be

processed in parallel to obtain the clusters samples belong to. First, the algorithm needs to transform the samples into quantum states $|u\rangle = \frac{\mathbf{u}}{|\mathbf{u}|}$. And the entangled states $|\varphi\rangle, |\phi\rangle$ are defined as

$$\begin{cases} |\varphi\rangle = \frac{1}{\sqrt{2}}(|u\rangle|0\rangle + \frac{1}{\sqrt{M}} \sum_{j=1}^M |v_j^c\rangle|j\rangle) \\ |\phi\rangle = \frac{1}{\sqrt{Z}}(|u\rangle|0\rangle - \frac{1}{M} \sum_{j=1}^M |v_j^c\rangle|j\rangle) \end{cases} \quad (18)$$

where $Z = |\mathbf{u}|^2 + \frac{1}{M} \sum_j |\mathbf{v}_j^c|^2$ is the normalization factor. It can be shown that the square of the expected distance $D_c^2 = |\mathbf{u} - \frac{1}{M} \mathbf{v}_j^c|^2$ between the sample to be measured and the cluster center is equal to Z times the probability of success of the measurement:

$$D_c^2 = 2|\langle\varphi|\phi\rangle|^2 Z \quad (19)$$

$|\langle\phi|\varphi\rangle|^2$ can be considered as the square of the modulus of the projection of $|\varphi\rangle$ in the direction of $|\phi\rangle$, which can be obtained from the probability of successfully performing the Swap quantum operation Nielsen and Chuang (2002). This algorithm's steps are only executed once for each sample in the sample space to find the cluster with the closest distance from the sample and assign the sample to that cluster.

The selection of the initial centers of clusters is important and the improper selection may lead to convergence to a local optimum. The general principle is to distribute the initial centers of clusters as sparsely as possible throughout the sample space. So Lloyd et al. proposed the method called quantum adiabatic computation to solve the optimization problem of finding the initial centers of clusters. Quantum adiabatic computation is based on quantum operations to evolve between states, and this method can be applied to quantum machine learning.

Quantum principal component analysis The dimensionality reduction algorithm is one of the most important unsupervised learning algorithms. Those algorithms map the features of samples in high-dimensional space to lower-dimensional space. The high-dimensional representation of samples contains noise information, which will cause errors and the reduction of accuracy. Through the dimensionality reduction algorithm, the noise information can be reduced which is beneficial to obtain the essential features of samples.

In dimensionality reduction algorithms, Principal Component Analysis (PCA) is one of the most common algorithms. The idea of PCA is to map the high-dimensional features of sample \mathbf{X} to the low-dimensional representing \mathbf{Y} by linear projection \mathbf{P} , so that the variance of the features in the projection space is maximized and the covariance between each dimension is minimum, i.e. the covariance matrix \mathbf{D} of \mathbf{Y} is diagonal. It can be shown that the matrix \mathbf{P} is the eigenmatrix of the covariance matrix \mathbf{C} of the sample matrix \mathbf{X} . In this way, fewer dimensions of features are used to preserve more properties of original samples, but the computational cost of PCA is prohibitive facing a large number of high-dimensional vectors.

In 2014, Lloyd et al. (2014) proposed the quantum principal component analysis (QPCA) algorithm. QPCA can be used for the discrimination and assignment of quantum states. Suppose there exist two sets consisting of m states and sample from them, the density matrix $\rho = \frac{1}{m} \sum_i |\phi_i\rangle\langle\phi_i|$ is obtained from the first set $\{|\phi_i\rangle\}$, and the density matrix $\sigma = \frac{1}{m} \sum_i |\psi_i\rangle\langle\psi_i|$ is obtained from the second set $\{|\psi_i\rangle\}$. Assuming that the quantum state

to be assigned is $|\chi\rangle$, the density matrix concatenation as well as the quantum phase estimation can be performed for $|\chi\rangle$ to obtain the eigenvectors and eigenvalues of $\rho - \sigma$:

$$|\chi\rangle|0\rangle \rightarrow \sum_j \chi_j |\xi_j\rangle|x_j\rangle \quad (20)$$

where $|\xi_j\rangle, x_j$ are the eigenvalues and eigenvectors of $\rho - \sigma$, respectively. By measuring the eigenvalue $|\xi_j\rangle$, $|\chi\rangle$ belongs to the first class $\{|\phi_i\rangle\}$ if the eigenvalue is positive, and it belongs to the second class $\{|\psi_i\rangle\}$ if the eigenvalue is negative. The above procedure is minimum error state discrimination, which has exponential speedup. The QPCA assumes the preparation of quantum states using QRAM, but QRAM is only a theoretical model and no reliable physical implementation has emerged.

5.1.2 Quantum supervised learning algorithms

Quantum linear discriminant analysis Similar to PCA, Linear Discriminant Analysis (LDA) is also a kind of dimensionality reduction algorithm. But unlike the PCA algorithm, the PCA algorithm reduces the dimensionality of unlabeled sample data, while LDA, reduces the dimensionality of labeled sample data. The idea of the LDA algorithm is to project the data to the low-dimensional space, and projected data of the same cluster is as close as possible, i.e. minimizing the intra-class scatter:

$$S_w = \sum_{i=1}^N \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T \quad (21)$$

and the distance between the centers of clusters to be as large as possible, i.e. maximizing inter-class scatter:

$$S_b = \sum_{i=1}^N (\mu_i - \bar{x})(\mu_i - \bar{x})^T \quad (22)$$

To satisfy these two conditions simultaneously, it is necessary to maximize the generalized Rayleigh quotient:

$$J = \frac{w^T S_b w}{w^T S_w w} \quad (23)$$

where w is the normal vector of the projected hyperplane. This optimization problem can be solved by the Lagrange multiplier method.

In 2016, Cong and Duan (2016) proposed the Quantum Linear Discriminant Analysis(QLDA) algorithm. Compared with the classical LDA algorithm, QLDA achieves exponential acceleration, which greatly reduces the computational difficulty and space utilization. Firstly, QPCA used the Oracles operator to obtain the density matrix:

$$\left\{ \begin{array}{l} |\Psi_1\rangle = O_2 \left(\frac{1}{\sqrt{k}} \sum_{c=1}^k |c\rangle |0\rangle |0\rangle \right) \\ \quad = \frac{1}{\sqrt{k}} \sum_{c=1}^k |c\rangle |||\mu_c - \bar{x}||| \langle \mu_c - \bar{x} \rangle \\ |\Phi_1\rangle = O_1 \left(\frac{1}{\sqrt{M}} \sum_{j=1}^M |j\rangle |0\rangle |0\rangle \right) \\ \quad = \frac{1}{\sqrt{M}} \sum_{j=1}^M |j\rangle |||x_j - \mu_{cj}||| \langle x_j - \mu_{cj} \rangle |c\rangle \end{array} \right. \quad (24)$$

If the norm of the vector forms an effectively productive distribution, the state can be obtained:

$$\left\{ \begin{array}{l} |\Psi_2\rangle = \frac{1}{\sqrt{A}} \sum_{c=1}^k ||\mu_c - \bar{x}|| |c\rangle ||\mu_c - \bar{x}|| \langle \mu_c - \bar{x} \rangle \\ |\Phi_2\rangle = \frac{1}{\sqrt{B}} \sum_{j=1}^M ||x_j - \mu_{cj}|| |j\rangle ||x_j - \mu_{cj}|| \langle x_j - \mu_{cj} \rangle |c_j\rangle \end{array} \right. \quad (25)$$

where $A = \sum_{c=1}^k ||\mu_c - \bar{x}||^2$, $B = \sum_{j=1}^M ||x_j - \mu_{cj}||^2$. After implementing the bias trace operation on the density matrix composed of the two states $|\Psi_2\rangle, |\Phi_2\rangle$, the Equation (26) can be obtained:

$$\left\{ \begin{array}{l} S_B = \frac{1}{A} \sum_{c=1}^k ||\mu_c - \bar{x}||^2 | \mu_c - \bar{x} \rangle \langle \mu_c - \bar{x} | \\ S_W = \frac{1}{B} \sum_{c=1}^k \sum_{i \in c} ||x_i - \mu_c||^2 | x_i - \mu_c \rangle \langle x_i - \mu_c | \end{array} \right. \quad (26)$$

The Equation (26) can be solved by the Lagrange multiplier method and obtain the solution:

$$(S_B^{-1/2} S_W^{-1} S_B^{-1/2})v = \lambda v \quad (27)$$

where $w = S_B^{-1/2}v$. The eigenvalues and eigenvectors of v can be obtained by the quantum phase estimation. Finally, the optimal projection direction w is obtained by the Ermey matrix concatenation solution.

The algorithm is similar to the QPCA algorithm. They both relate the covariance matrix of samples in original problems to the density matrix of the quantum system, and the eigenvalues and eigenvectors of the density matrix are investigated to obtain the optimal projection direction or the main eigenvectors.

Quantum k-nearest neighbors The K-Nearest Neighbors (KNN) is a very classical classification algorithm. The algorithm finds the k nearest samples with the label for the sample x to be classified, and then uses a classification decision rule, such as majority voting, to decide the cluster x belongs to according to the labels of these k samples. The advantages of the KNN algorithm are that the accuracy of the algorithm will be extremely high

Table 5 The comparison of time complexity. The Q-version and C-version denote the Quantum version and Classical version of the algorithm

ML Algorithms	Q-version	C-version
Deutsch-Jozsa	$O(\sqrt{n})$	$O(2^n)$
Shor	$O(\log(n))$	$O(\sqrt{n})$
Grover	$O(\sqrt{n})$	$O(n)$
QK-Means	$O(k\log(kmn))$	$O(mn)$
QPCA	$O(\log(d))$	$O(d)$
QLDA	$O(\log(mn)/\epsilon^3)$	$O(mn)$
QKNN	$O(\log(mn))$	$O(mn)$
QDT	$\frac{(\sqrt{n}-1)\sqrt{kd}}{\sqrt{n}(\sqrt{k}-1)}(T_1 + T_2)$	$O(mn\log(n))$
QSVM	$O(\log(mn))$	$O(mn)$

when the dataset is large enough, but the computational cost will be very high when the database is large or the dimensionality of the sample is large.

In 2014, Wiebe et al. proposed the Quantum K-Nearest Neighbor algorithm (QKNN) Wiebe et al. (2014). This algorithm can obtain the Euclidean distance as the inner product, which has polynomial reductions in query complexity compared with Monte Carlo algorithm. The QKNN algorithm first encodes the nonzero data of the vectors u, v to the probability magnitudes of the quantum states $|v\rangle, |u\rangle$:

$$\begin{cases} |v\rangle = d^{-\frac{1}{2}} \sum_{i: v_{ji} \neq 0} |i\rangle \left(\sqrt{1 - \frac{r_{ji}^2}{r_{max}^2}} e^{-i\phi_j i} |0\rangle + \frac{v_{ji}}{r_{max}} |1\rangle \right) |1\rangle \\ |u\rangle = d^{-\frac{1}{2}} \sum_{i: v_{0i} \neq 0} |i\rangle \left(\sqrt{1 - \frac{r_{0i}^2}{r_{max}^2}} e^{-i\phi_0 i} |0\rangle + \frac{v_{0i}}{r_{max}} |1\rangle \right) |1\rangle \end{cases} \quad (28)$$

where $v_0 = u$, $v_{ji} = r_{ji} e^{i\phi_j}$, r_{max} is the upper bound of the eigenvalue, and the inner product can be obtained by performing the SWAP operation on $|v\rangle, |u\rangle$:

$$|\langle u | v_j \rangle|^2 = |\langle v_0 | v_j \rangle|^2 = (2P(0) - 1)d^2 r_{max}^{-2} \quad (29)$$

where $P(0)$ denotes the probability when the measure is zero. Wiebe et al. also tried to use the Euclidean distance of two quantum states to determine their classification rules, but the experimental results showed that this method has more iterations and lower accuracy than the inner-product method. Therefore, the quantum Euclidean distance classification algorithm has not been promoted.

Quantum decision tree classifier Decision Tree (DT) algorithm is a classical supervised learning model. The DT algorithm represents a mapping relationship between the attributes and categories of objects. The node in the tree represents an attribute value, which determines the direction of the classification. Each path from the root node to the leaf node represents attribute value requirements, according to which the object can be identified as a category. The algorithm takes advantage of samples to learn the structure of decision trees and discriminant rules for the classification of samples. To improve the learning efficiency of DT, information gain is often used to select key features (Table 5).

In 2014, Lu and Braunstein (2014) proposed a Quantum Decision Tree(QDT) classifier, which clusters samples into subclasses by the quantum fidelity measure between two quantum states so that the QDT can control the quantum states. In addition, they proposed a quantum entropy impurity criterion to prune the decision tree. The QDT classifier first converts the sample features $\{x_i, y_i\}_{i=1}^n$ into a quantum state $\{|x_i\rangle, |y_i\rangle\}_{i=1}^n$, where $|x_i\rangle$ denotes the i quantum state corresponding to the i th sample. $|y_i\rangle$ denotes the quantum state of the known class corresponding to sample $|x_i\rangle$. The quantum entropy impurity criterion is defined as:

$$S(\rho) = -tr(\rho \log \rho) \quad (30)$$

where $\rho = \sum_{i=1}^{n_{dt}} p_i^{dt} |y_i^{(dt)}\rangle\langle y_i^{(dt)}|$ is the density matrix of the quantum states of the corresponding class at the node t . tr represents the trace (or trace number) of the matrix, which is the sum of the elements on the main diagonal. The expectation of the criterion is calculated:

$$S_e(\rho_i^{(t)}) = \sum_{j=1}^{t_i} p_j S(\rho_{ij}^{(t)}) \quad (31)$$

Finally, the Grover algorithm is used to solve the minimum of the expectation in the Equation (31), and the class the expectation corresponds to is the one the sample belongs to. The Shannon entropy in classical information theory is replaced by the quantum entropy impurity criterion, and the eigenvalue can be obtained by calculating the expectation of the criterion, which is the difference between this algorithm and the traditional decision tree algorithm.

Quantum support vector machine Support Vector Machine (SVM) is an important supervised linear classification algorithm. The idea of SVM is to classify by finding the classification hyperplane with maximum interval:

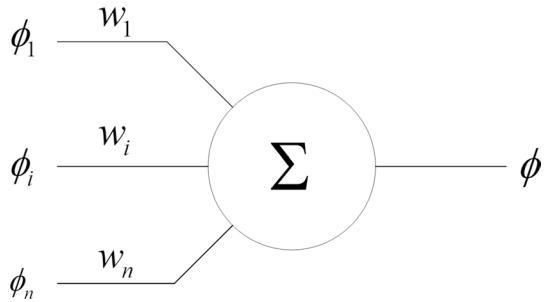
$$\arg \max_{w,b} \left(\min_i \frac{y_i(w^T \phi(x_i) + b)}{\|w\|} \right) \quad (32)$$

where w is the normal vector of the hyperplane, b is the bias, $y_i \in \{-1, 1\}$ is the label of the sample x_i . The solution x_i^* of x_i to the Equation (32) is called the support vector, which is closest to the classification hyperplane. and $d = \frac{1}{\|w\|} y_i^*(w^T \phi(x_i^*) + b)$ is the maximum interval from the sample to the hyperplane. The Equation (33) can be obtained by the Equation (32) with the scale transformation:

$$\begin{cases} \arg \min \left(\frac{1}{2} \|w\|^2 \right), \\ s.t. y_i(w^T x_i + b) \geq 1 \end{cases} \quad (33)$$

The Equation (33) is a conditional constraint problem which can be solved by the Lagrange multiplier method.

In 2014, Rebentrost et al. (2014) proposed the Quantum Support Vector Machine (QSVM). The QSVM uses a non-sparse matrix exponentiation technique for efficiently performing a matrix inversion and obtaining exponential acceleration. The QSVM firstly encodes eigenvectors to quantum state probability magnitudes using Oracles operators:

Fig. 17 Quantum Neurons Model

$$|x_i\rangle = \frac{1}{|x_i|} \sum_{k=1}^N (x_i)_k |k\rangle \quad (34)$$

where $(x_i)_k$ denotes the k th feature of the i th eigenvector. In order to obtain the normalized kernel matrix, it is necessary to obtain the quantum state:

$$|\chi\rangle = \frac{1}{\sqrt{N_\chi}} \sum_{i=1}^M |x_i||i\rangle|x_i\rangle \quad (35)$$

where $N_\chi = \sum_{i=1}^M |x_i|^2$. And the normalized kernel matrix can be solved by the bias trace of the density matrix $|\chi\rangle\langle\chi|$:

$$\text{tr}_2\{|x\rangle\langle x|\} = \frac{1}{N_\chi} \sum_{i,j=1}^M \langle x_j|x_i\rangle|x_i||x_j||i\rangle\langle j| = K/\text{tr}K \quad (36)$$

By this method, the quantum system is associated with the kernel matrix in traditional ML. Due to the high parallelism of evolutionary operations between quantum states, the computation of the kernel matrix in traditional ML can be accelerated.

5.2 Quantum deep learning

Similar to the QML, quantum deep learning(QDL) allows deep learning algorithms to take advantage of the basic properties of quantum mechanics. QDL uses quantum computing instead of the traditional von Neumann machine computation, making deep learning algorithms quantum, achieving the purpose of significantly improving the parallelism of algorithms and reducing the computational complexity (Fig. 17).

The basic principle of neurons is to simulate the signal of excitation or inhibition with weight parameters and simulate the information processing with connection weighting to obtain the output. So the neuron can be modeled as $Y = \sum_i w_i x_i$. In QDL, all neurons need to convert inputs into quantum states ϕ_j :

$$Y = \sum_i w_i \phi_j = \sum_i \sum_j w_{ij} |x_1, \dots, x_{2^n}\rangle, i = 1, 2, \dots, 2^n \quad (37)$$

where, 2^n denotes the number of nodes of the input.

If the quantum state ϕ_j is orthogonal, the output of neurons can be expressed by the quantum unitary transformation:

$$Y = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{12^n} \\ w_{12} & w_{22} & \cdots & w_{22^n} \\ \cdots & \cdots & \cdots & \cdots \\ w_{2^n 1} & w_{2^n 2} & \cdots & w_{2^n 2^n} \end{pmatrix} \times \begin{pmatrix} |0, 0, \dots, 0\rangle \\ |0, 0, \dots, 1\rangle \\ \vdots \\ |1, 1, \dots, 1\rangle \end{pmatrix} \quad (38)$$

In general, the process of training quantum neurons model involves five steps: First, initializing the weight matrix W^0 ; Second, constructing the training set $\{|\phi\rangle, |O\rangle\}$ according to the problem, Third, calculating the neuron output $|\Theta\rangle = W^t|\phi\rangle$, where t is the number of iterations. Fourth, updating the weight parameter $W_{ij}^{t+1} = W_{ij}^t + \alpha(|O\rangle_i - |\Theta\rangle_i)|\phi\rangle_j$, where α is the learning rate; Finally, repeating the third and fourth steps until the network converges.

The concept of quantum neural computation was introduced by Kak (1995) in 1995, but the concept of quantum deep learning was first proposed by Wiebe et al. (2014) in 2014. In the same year, Schuld et al. (2014) proposed three requirements satisfied by quantum neural networks: First, the input and output of the quantum system are encoded as quantum states; Second, the quantum neural network reflects one or more fundamental neural computational mechanisms; Third, the evolution based on quantum effects Must be fully compatible with quantum theory. This section will also present quantum multilayer perceptron, quantum recurrent networks, and quantum convolutional networks in that order.

5.2.1 Quantum multilayer perceptrons

In 1995, Menneer and Narayanan (1995) proposed the quantum-inspired neural network (QUINN). Traditional neural networks train neural networks to find parameters which make networks enable to obtain the correct results for each pattern. Inspired by quantum superpositionality, QUINNs train multiple isomorphic neural networks that process only a single pattern for each pattern, and the isomorphic networks corresponding to the different patterns are superimposed in a quantum way to produce the QUINNs. The weight vectors of the QUINNs are called the quantum-inspired wave-function (QUIWF), which will collapse and generate the classification results during measuring.

In 1996, Behrman et al. (1996) proposed the Quantum Dot Neural Networks (QDNN). The QDNN imitates a quantum dot molecule coupled to the substrate lattice in the time-varying field and uses discrete nodes of the time dimension as hidden layer neurons. It is shown that the QDNN can perform any desired classical logic gate in some regions of the phase space.

In 1996, Tóth et al. (1996) proposed Quantum Cellular Neural Networks (QCN). The QCNs use cells to form interacting quantum dots that communicate between cells with Coulomb forces, each cell encodes a continuous degree of freedom, and its state equation can be represented by the time-dependent Schrodinger equation to describe the cellular network.

In 2000, Matsui et al. (2000) proposed a quantum neural network based on quantum circuits. The basic unit of the network is a quantum logic gate consisting of a 1-bit rotating gate and a 2-bit controlled-NOT gate, which can implement all the basic logic operations. It controls the connection between neurons through the rotation gate and the computation within the neuron through the controlled-NOT gate. Since the construction of neurons

depends on the quantum logic gate, the number of logic gates will increase exponentially when the network structure is complex.

In 2005, Kouda et al. (2005) constructed a qubit neural network with quantum logic gates, and proposed the structure of a quantum perceptron. The state z of the neuron receiving inputs from K other neurons is denoted as:

$$\begin{cases} u = \sum_{k=1}^K f(\theta_k) \cdot x_k - f(\lambda) = \sum_{k=1}^K f(\theta_k) \cdot f(y_k) - f(\lambda) \\ y = \frac{\pi}{2} g(\delta) - \arg(u) \\ z = f(y) \end{cases} \quad (39)$$

where the quantum state $f(\varphi) = e^{i\varphi} = \cos \varphi + i \sin \varphi$, $g(\cdot)$ denotes the sigmoid function, and $\arg(u)$ denotes the phase angle of u .

In 2006, Zhou et al. (2006) proposed the Quantum Perceptron Network (QPN). Through experiment simulation, the quantum perceptron containing only one neuron can still realize the dissimilarity operation, which cannot be achieved by the conventional perceptron containing only one neuron. The structure of QPN is as follows:

$$\begin{cases} t = f(y) \\ \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \\ y = \frac{\pi}{2} \text{sigmoid}(\sigma) - \arctan(\text{Im}(\varphi)/\text{Re}(\varphi)) \\ \varphi = \sum_{n=1}^N f\left(\frac{\pi}{2} P_n\right) f(\theta_n) - f\left(\frac{\pi}{2}\right) f(\lambda) \end{cases} \quad (40)$$

where θ_n and λ are the weighting parameters and the phase parameters, respectively, σ is the phase control factor, and P_n is the input data.

In 2014, Schuld et al. (2014) proposed a quantum neural network based on quantum walking. The network uses the location of the quantum walking to represent the firing patterns of binary neurons, resting and activated states, which are encoded with a set of binary strings. To simulate the dissipative dynamics of the neural network, the network performs quantum walking in a decoherent manner to achieve retrieval of memorized patterns on non-fully initialized patterns.

5.2.2 Quantum recurrent neural networks

In 2014, Wiebe et al. (2014) first introduced the concept of “quantum deep learning”. They argued that quantum algorithms can effectively solve some problems that cannot be solved by traditional computers. Quantum algorithms provide a more efficient and comprehensive framework for deep learning. In addition, they also proposed an optimization algorithm for quantum Boltzmann machines, which reduced the training time of Boltzmann machines and provided significant improvements in the objective function.

Boltzmann machine(BM) is a kind of undirected recurrent neural network. From a physical perspective, the BM is modeled according to the Ising model of thermal equilibrium and uses the Gibbs distribution to model the probability of each hidden node. They proposed two quantum methods to solve the optimization problem of the BM: Gradient

Estimation via Quantum Sampling (GEQS) and Gradient Estimation via Quantum Amplitude Estimation (GEQAE). The GEQS algorithm uses mean-field theory to approximate the nonuniform prior distribution for each configuration and extracts the Gibbs states from the mean-field states, allowing the Gibbs distribution to be prepared accurately when the two states are close enough. Unlike the GEQS algorithm, the GEQAE uses the Oracle operators to quantize the training data, the idea of the GEQAE algorithm is to encode samples with amplitude estimation in quantum, which greatly reduces the computational complexity of gradient estimation.

In 2020, Bausch proposed Bausch (2020) quantum recurrent neural networks, which are mainly constructed by a novel quantum neuron. The nonlinear activation function of the neuron is implemented with the nonlinearity of the cosine function generated by the amplitude change when the basis vector of a quantum bit is rotated. These neurons are combined to form a structured QRNN cell which is iterated to obtain a recurrent model similar to the traditional RNN.

In 2020, Chen et al. (2020) proposed the Quantum Long short-term memory network (QLSTM). QLSTM utilizes Variational Quantum Circuits (VQCs) with tunable parameters to replace the LSTM cells in traditional Neural Networks. VQCs have the ability of feature extraction and data compression, which consist of data encoding layers, variational layers, and quantum measurement layers. Through numerical simulations, it can be demonstrated that the QLSTM learns faster and converges more robustly than the LSTM, and the typical spikes of the loss function don't appear in QLSTM, which appears in the traditional LSTM.

In 2021, Ceschini et al. (2021) proposed the method to implement LSTM cells in a quantum framework. The method uses quantum circuits to replicate the internal structure of the cell for inferences. In this method, an encoding method was proposed to quantize the operators of the LSTM cell, such as quantum addition, quantum multiplication, and quantum activation functions. Finally, the quantum architecture was verified by numerical simulations on the IBM Quantum Experience™ platform and classical devices.

5.2.3 Quantum convolutional networks

In 2019, Cong et al. (2019) first proposed a quantum convolution neural network(QCNN). QCNN is a variational quantum circuit model whose input is an unknown quantum state. The convolution layer consists of parametrized two-qubit gates applying a single quasilocal unitary in a translationally invariant manner. The pooling operation is implemented by applying unitary rotations to nearby qubits according to the measurement of a fraction of qubits. The convolution and pooling layers are performed until the system size is small enough to obtain qubits as the output. Similar to traditional convolutional neural networks, hyperparameters, such as the number of convolutional and pooling layers, are fixed in QCNN, while the parameters in the convolution and pooling layers of QCNN are learnable.

In 2020, Iordanis Kerenidis et al. (2019) proposed a modular quantum convolution neural network algorithm, which implements all modules with simple quantum circuits. The network achieves any number of layers and any number and size of convolution kernels. During the forward propagation, QCNN has exponential speedup compared with the traditional CNN.

In 2021, Liu et al. (2021) proposed the hybrid quantum-classical convolutional neural network (QCCNN). QCCNN utilizes interleaved 1-qubit layers and 2-qubit layers to form a quantum convolution layer. The 1-qubit layer consists of R_y gates, which contain tunable parameters. The 2-qubit layer consists of CNOT gates on the nearest-neighbor pairs of qubits. QCCNN converts the input into a separable quantum feature with the quantum

convolution layer, utilizes the pooling layers to reduce the dimensionality of the data, and finally measures the quantum feature to obtain the output scalar.

5.3 Quantum evolutionary algorithms

An evolutionary algorithm is a stochastic search algorithm constructed based on Darwin's theory of natural selection and Mendel's theory of genetic variation, which simulates reproductions, mutation, competition, and selection in biological evolution. And quantum evolutionary algorithm uses qubits to encode individuals and updates the individuals with rotation gates and NOT gates so that individuals can contain the information of multiple states at the same time and get more abundant populations, which greatly improves the parallelism and convergence speed of the algorithm.

In the quantum evolutionary algorithm, each individual is encoded with qubits in the population. After encoding, each gene of the individual contains all information in the superposition state:

$$|\phi\rangle = \alpha|0\rangle\beta|1\rangle \quad (41)$$

where α, β denotes the probability amplitude of the quantum state satisfying $|\alpha|^2 + |\beta|^2 = 1$, so the individuals encoded with qubits can be expressed as:

$$q_j^t = \begin{pmatrix} \alpha_{j1}^t & \alpha_{j2}^t & \dots & \alpha_{jm}^t \\ \beta_{j1}^t & \beta_{j2}^t & \dots & \beta_{jm}^t \end{pmatrix} \quad (42)$$

Where q_j^t denotes the j th individual in the population after the t th iteration and m denotes the number of genes in the individual. The individual encoded with qubits can express the superposition of multiple quantum states at the same time, making the individual more diverse. As the algorithm converges, $|\alpha|, |\beta|$ will also converge to 0 or 1, making the encoded individual converge to a single state.

In general, quantum evolutionary algorithm has the following steps: First, initialize the population to ensure that $\alpha = \beta = \frac{1}{\sqrt{2}}$; Second, generate random number $r \in [0, 1]$, and compare r with the probability amplitude of the quantum state α , the measurement value of the quantum state takes 1 if $r < \alpha^2$, otherwise the value takes 0. In this way each individual in the population is measured once to get a set of solutions for the population; Third, evaluate the fitness of each state; Fourth, compare the current best state of the population with the recorded historical best state, and then record the best state and the fitness; Fifth, update the population with quantum rotation gates and quantum NOT gates according to a certain strategy; Finally, loop the above steps until the convergence condition is reached.

The earliest quantum evolutionary algorithm was proposed by Narayanan and Moore (1996). In 1996, they first combined quantum theory with genetic algorithms and proposed quantum genetic algorithms, which opened up the field of quantum evolutionary computation. The quantum evolutionary algorithm was proposed by Han et al. Based on the PGQA (Parallel Quantum-inspired Genetic Algorithm) Han et al. (2001) in 2001, they extended quantum genetic algorithms to quantum evolutionary algorithms (QEA) in 2002 Han and Kim (2000).

5.3.1 Quantum encoding algorithms

In 2008, Li and Li (2008) proposed a quantum evolutionary algorithm encoded by Bloch sphere coordinates. In this algorithm, individuals are encoded in Bloch coordinates of qubits, updated by a quantum rotation gate, and mutated by a quantum NOT gate. Compared with a simple genetic algorithm (SGA), the algorithm has higher effectiveness and feasibility.

In the same year, Cruz et al. (2007) proposed a quantum evolutionary algorithm based on real number encoding. The algorithm uses an interval in the search space to represent the genes in quantum individuals and calculates the pulse height by the pulse width value and the total number of quantum individuals in the population, which ensures that the total area of the probability density function used to generate classical individuals is equal to 1. Compared with similar algorithms, this algorithm can obtain better solutions with less computation, which greatly reduces the convergence time.

In 2009, Zhao et al. Chen et al. (2005) proposed a Real-coded Chaotic Quantum-inspired genetic Algorithm (RCQGA). The RCQGA maps individuals to qubits in the solution space and applies the crossover and mutation operations to search real individuals. The individuals representing the weights of networks are encoded as tunable vectors, which can be obtained by RCQGA. Compared with similar algorithms, this algorithm converges faster when searching for the best weights of fuzzy neural networks.

In 2016, Joshi et al. (2016) proposed an adaptive quantum evolutionary algorithm (ARQEAs) encoding with real numbers. The algorithm utilizes a parameter-free quantum crossover operator inspired by a rotation gate to generate new populations and amplifies the amplitude by a quantum phase rotation gate to search the desired element. The algorithm can avoid the tuning of evolutionary parameters.

5.3.2 Quantum evolutionary operators

In 2002, Li and Zhuang (2002) proposed a genetic algorithm based on the quantum probability representation (GAQPR), where a novel crossover operator and mutation operator are designed. The crossover operator makes individuals contain the best evolutionary information by exchanging the current evolutionary target and updating individuals, and the mutation operator is implemented by randomly selecting one quantum bit of each individual to exchange the probability amplitude. The GAQPR algorithm is more effective for multi-peaked optimization problems, which is demonstrated by two typical function optimization problems.

In 2004, Yang et al. (2004) proposed a novel discrete particle swarm optimization algorithm based on quantum individuals. The algorithm defines each particle as one qubit and uses random observation instead of a sigmoid function to approximate the optimal result step by step. The algorithm has also proved its effectiveness in simulation experiments and applications in CDMA.

In 2015, Jin and Jin (2015) proposed an improved quantum particle swarm algorithm (IQPSO) for visual feature selection(VFS). The algorithm obtains the reverse solution based on the reverse operation of the solution and selects the individual optimal solution and the global optimal solution by calculating the fitness function for all solutions and inverse solutions.

In 2019, Rehman et al. (2019) proposed an improved approach to the quantum particle swarm algorithm. The method uses a mutation strategy to change the mean best position

by randomly selecting the best particle to take part in the current search domain and then adds an enhancement factor to improve the global search capability to find the global best solution.

5.3.3 Quantum immune operators

In 2008, Li et al. Jiao et al. (2008) a quantum-inspired immune clonal algorithm (QICA), where antibody proliferation is divided into a set of subpopulations. The antibodies in the subpopulations are represented by multistate gene qubits. Antibody updates are implemented with the quantum rotation gate strategy and the dynamic angle adjustment mechanism to accelerate convergence, quantum mutations are implemented with the quantum NOT gate to avoid premature convergence, and a quantum recombination operator is designed for information communication between subpopulations to improve the search efficiency.

In the same year, Li et al. Yangyang and Licheng (2008) proposed a quantum-inspired immune clonal multiobjective optimization algorithm (QICMOA). The algorithm encodes the dominant population antibodies with qubits and designs quantum recombination operators and quantum NOT gates to clone, recombine, and update the dominant antibodies with less crowded density.

In 2013, Liu et al. (2013) proposed the cultural immune quantum evolutionary algorithm(CIQEA) which consists of a population space based on the QEA and a belief space based on immune vaccination. The population space periodically provides vaccines to the confident population. The confidence space continuously evolves these vaccines and optimizes the evolutionary direction for the population space, which greatly improves the global optimization capability and convergence speed.

In 2014, Shang et al. (2014) proposed an immune clonal coevolutionary algorithm (ICCoA) for dynamic multi-objective optimization(DMO). The algorithm solves the DMO problem based on the basic principle of an artificial immune system with an immune clonal selection method and designs coevolutionary competition and cooperation operators to improve the consistency and diversity of solutions.

In 2018, Shang et al. (2018) proposed a quantum-inspired immune clonal algorithm (QICA-CARP). The algorithm encodes antibodies in the population as qubits and controls the population evolution to a good schema with the current optimal antibody information. The quantum mutation strategy and quantum crossover operator speed up the convergence of the algorithm as well as the exchange of individual information.

5.3.4 Quantum population optimization

In 2005, Alba and Dorronsoro (2005) subdivided the grid population structure into squares, rectangles, and bars, and designed a quantum evolutionary algorithm that introduces a preprogrammed change of the relationship between individual fitness and population entropy to dynamically adjust the structure of the population and construct the first adaptive dynamic cellular model.

In 2008, Li et al. (2008) used a novel distance measurement method to maintain performance. The algorithm evolves the solutions population by a non-dominated sorting method and uses Pareto max-min distance to preserve population diversity, allowing a good balance between global and local search.

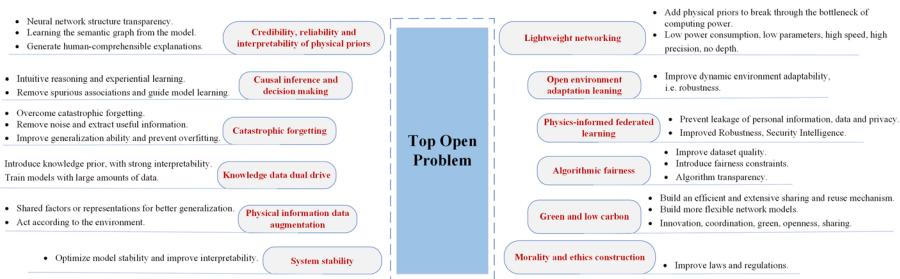


Fig. 18 Top open problems combining physics with AI

In 2009, Mohammad and Reza (2009) proposed a dynamic structured interaction algorithm among population members in the quantum evolutionary algorithm(QEA). The algorithm classified the population structure of QEA into ring structure, cellular structure, binary tree structure, cluster structure, lattice structure, star structure, and random structure, and proved that the best structure of QEA is cellular structure by comparing several structures.

In 2015, Qi and Xu (2015) proposed an L5-based simultaneous cellular quantum evolution algorithm (LSCQEA). In the LSCQEA algorithm, each individual is located in a lattice, and each individual in the lattice and its four neighboring individuals go through an iteration of QEA. In every iteration of QEA, different individuals exchange information with others by overlapping neighboring individuals, which makes the population evolve.

In 2018, Mei and Zhao (2018) proposed a random perturbation QPSO algorithm (RP-QPSO). By introducing a random perturbation strategy to the iterative optimization, the algorithm can dynamically and adaptively adjust, which improves the local search ability and global search ability.

6 Top open problems

The laws of physical knowledge are diverse and powerful, and the AI model simulates the brain composed of millions of neurons connected by weights to realize human behavior. Through the combination of physical knowledge and AI, mutual influence, and evolution, people's understanding of the deep neural network model is promoted, and then the development of a new generation of artificial intelligence is promoted. However, there are also huge challenges in combining the two, which we will discuss around the following issues (see Fig. 18).

6.1 Open problem 1: credibility, reliability, and interpretability of physical priors

Neural networks in AI are becoming more and more popular in physics as a general model in various fields (Redmon et al. 2016; He et al. 2017; Bahdanau et al. 2014). However, the intrinsic properties of neural networks (parameters and model inference results, etc.) are difficult to explain. Neural networks are therefore often labeled as a black box. Interpretability aims to describe the internal structure and inferences of the system in a way that humans can understand, which is closely related to the cognition, perception, and bias of the human brain. Today, the emerging and active intersection of physical neural networks

attempts to make the black box transparent by designing deep neural networks based on physical knowledge. By using this prior knowledge, deeper and more complex neural networks are made feasible. However, the reasoning and interpretation of the internal structure of neural networks is still a mystery, and physical information methods as a supplement to prior knowledge have become a major challenge in explaining artificial intelligence neural networks.

6.2 Open problem 2: causal inference and decision making

The purpose of AI is to let machines learn to “think” and “decide” like the brain, and the brain’s understanding of the real world, processing of incomplete information, and task processing capabilities in complex scenarios are unmatched by current AI technologies, especially in time series problems (Rubin 1974; Pearl 2009; Imbens and Rubin 2015). Since most of the existing AI models are driven by association, just like the decision output of a physical machine will be affected by the change of the mechanism or the intervention of other factors, these models usually only know the “how” (correlation) but not the “why” (causality), recent groundbreaking works (Runge 2018; Runge et al. 2019a, b; Nauta et al. 2019) on time series causality lays the foundation for AI. Introducing causal reasoning, statistical physics thinking, and multi-perspective cognitive activities of the brain into the AI field, removing false associations, and using causal reasoning and prior knowledge to guide model learning is a major challenge for AI to improve generalization capabilities in unknown environments.

6.3 Open problem 3: catastrophic forgetting

The brain memory storage system is an information filter, just like a computer clearing disk space, it can delete useless information in the data to receive new information. “Catastrophic forgetting” in neurobiological terms, when learning a new task, the connection weights between neurons will weaken or even disappear as the network deepens. That is, the appearance of new neurons will cause the weights to be reset, and the brain neurons in the hippocampus rewire and overwrite memories Abraham and Robins (2005). For humans, the occurrence of forgetting can improve decision-making flexibility by reducing the impact of outdated information on people, and it can also make people forget negative events and improve adaptability.

Achieving artificial general intelligence today requires agents to be able to learn and remember many different tasks, and the most important part of the learning process is forgetting (McCloskey and Cohen 1989; Goodfellow et al. 2013). Through the purification of selective forgetting (Kirkpatrick et al. 2017; Zhang et al. 2023), AI can better understand human commands, improve the generalization ability of the algorithm, prevent overfitting of the model, and solve more practical problems. Therefore, learning to forget is one of the major challenges facing artificial intelligence.

6.4 Open problem 4: optimization and collaboration driven by knowledge and data

When solving many practical optimization problems, it is difficult to solve due to their characteristics of non-convex or multi-modal, large scale, high constraints, multi-objectives, and large uncertainty of constraints, and most evolutionary optimization algorithms

evaluate the potential of candidate solutions. The objective function and constraint function are too simple and may not exist. In contrast, solving evolutionary optimization problems through the evaluation of objectives and/or constraints through numerical simulations, physical experiments, production processes, or data collected in everyday life is called data-driven evolutionary optimization. However, data-driven optimization algorithms also pose different challenges depending on the nature of the data (distributed, noisy, heterogeneous, or dynamic). Inspired by AI algorithms, the Physics-informed model not only reduces the cost of implementation and computation Belbute-Peres et al. (2020), but also has a stronger generalization ability Sanchez-Gonzalez et al. (2020). AI is mainly based on knowledge bases and inference engines to simulate human behavior, and knowledge, as a highly condensed embodiment of data and information, often means higher algorithm execution efficiency. Inspired by physics, knowledge-driven AI has a lot of experience and strong interpretability, so the knowledge-data dual-driven optimization synergy provides a new method and paradigm for general AI, combining the two will be a very challenging subject.

6.5 Open problem 5: physical information data augmentation

In real life, there are differences between the real data and the predicted data distribution, and it is crucial to obtain high-quality labeled data, so transfer learning (Tremblay et al. 2018; Bousmalis et al. 2018), multi-task learning, and reinforcement learning are indispensable tools for introducing physical prior knowledge.

In reality, many problems cannot be decomposed into sub-problems independently, even if they can be decomposed, each sub-problem is connected by some shared factors or shared representations. So the problem is decomposed into multiple independent single-task processing, it ignores the rich correlation information in the problems. Multi-task learning is to put multiple related tasks together to learn and share the information they have learned between tasks, which is not available in single-task learning. Associative multi-task learning Thanasutives et al. (2021) can achieve better generalization than single-task learning. However, the interference between tasks, the different learning rates and loss functions between different tasks, and the limited expressivity of the model make multi-task learning challenging in the AI field.

Reinforcement learning is a field in AI that emphasizes how to act based on the environment to maximize the intended benefit. The reasoning ability it brings is a key feature measurement of AI, and it gives machines the ability to learn and think by themselves. The laws of physics are a priori, and how to combine reinforcement learning with physics is a challenging topic.

6.6 Open problem 6: system stability

In physics, stability is a performance index that all automatic control systems must meet. It is a performance in which the motion of the system can return to the original equilibrium state after being disturbed. In the field of AI, the study of system stability refers to whether the output value of the system can keep up with the expected value, that is, the stability of the system is analyzed for the output value Chen et al. (2023). But, since the AI system has a dynamic system, the output value also has dynamic characteristics. The neural network model is a highly simplified approximation of the biological nervous system, that is, the neural network can approximate any function. From the perspective of the system, the

neural network is equivalent to the output function of the system, that is, the dynamic system of the system. It simulates the functions of the human brain's nervous system structure, machine information processing, storage, and retrieval at different degrees and levels. From the perspective of causality, there is a certain internal relationship between interpretability and stability, that is, by optimizing the stability of the model, its interpretability can be improved, thereby solving the current difficulties faced by artificial intelligence technology in the implementation.

As a new learning paradigm, stable learning attempts to combine the consensus basis between these two directions. How to reasonably relax strict assumptions to match more challenging real-world application scenarios and make machine learning more credible without sacrificing predictive ability is a key problem to be solved for stable learning in the future.

6.7 Open problem 7: lightweight networking

Deep learning is now playing a big role in the field of AI, but limited by the traditional computer architecture, data storage, and computing need to be completed by memory chips and central processing units, resulting in problems such as long time consumption and high power consumption for computers to process data. The physical prior knowledge is introduced into the search space of NAS to obtain the optimal knowledge so that the network structure and the prediction result can be balanced Skomski et al. (2021). Meanwhile, modularity also plays a key role in NAS based on physical knowledge (Xu et al. 2019; Chen et al. 2020; Goyal et al. 2019). At the same time, the deep neural network also has a complex structure and involves a large number of hyperparameters, which is extremely time-consuming and energy-consuming in the training process and is difficult to parallelize. Therefore, we should combine the physical structure and thinking behavior of the brain, add physical priors, break through the bottleneck of computing power, realize low-power, low-parameter, high-speed, high-precision, non-depth AI models, and develop more efficient artificial intelligence technology.

6.8 Open problem 8: physics-informed federated learning

Privacy protection: The wide application of artificial intelligence algorithms not only provides convenience for people but also brings great risks of privacy leakage. Mass data is the foundation of artificial intelligence. It is precisely because of the use of big data, the improvement of computing power, and breakthroughs in algorithms that AI can develop rapidly and be widely used. Acquiring and processing massive amounts of information data inevitably involves the important issue of personal privacy protection Wang and Yang (2024). Therefore, artificial intelligence needs to find a balance between privacy protection and AI capabilities.

Security Intelligence: With the widespread application of AI in all walks of life, the abuse or malicious destruction of AI systems will have a huge negative impact on society. In recent years, algorithm attacks, adversarial sample attacks, model stealing attacks, and other attack technologies targeting artificial intelligence algorithms have continued to develop, which has brought greater algorithm security risks to AI. Therefore, realizing the security intelligence of AI is a big challenge in the future.

6.9 Open problem 9: algorithmic fairness

While the rapid development of the AI field has brought benefits to people, there are also some fairness issues. Such as statistical (sampling) bias, the sensitivity of the algorithm itself, and discriminatory behavior introduced by human bias Pfeiffer et al. (2023). As an important tool to assist people in decision-making, improving the fairness of AI algorithms is an issue of great concern for artificial intelligence Xivuri and Twinomurinzi (2021). Given the physical distance and the large scale of data, improving dataset quality, improving the algorithm's dependence on sensitive attributes (introducing fairness constraints), defining index quantification, and fairness measures, and improving the algorithm's generalization ability are important solutions Chen et al. (2023). In addition, human-machine symbiosis and algorithm transparency are also important ways to achieve fairness.

The human-machine symbiosis of machine intelligence and human brain cognition, thinking, and decision-making plus human inductive reasoning of the laws of the real world (physical knowledge) will be the future development direction, and algorithm transparency (understandability and interpretability) is to achieve fairness important tool. The problem with algorithmic fairness is not to solve some complex statistical Rubik's Cube puzzle, but to try to embody Platonic perfection of fairness on the walls of a cave that can only capture shadows. Therefore, the continuous deepening of algorithmic fairness research is a key issue in AI governance.

6.10 Open problem 10: open environment adaptation learning

Today, the AI field is all about assumptions about closed environments, such as the iid. and distribution constancy assumptions for data. In reality, it is an open dynamic environment and there may be changes. The learning environment of the neural network is a necessary condition for the learning process. The open environment, as a mechanism for learning, needs to exchange information, which requires the future AI to have the ability to adapt to the environment, or the robustness of the AI. For example, in the field of autonomous driving Müller et al. (2018), there are always emergent situations in the real world that cannot be simulated by training samples, especially in rare scenarios. Therefore, the future development of AI must be able to overcome the "open environment" problem for data analysis and modeling, which poses a huge challenge to the adaptability or robustness of AI systems.

6.11 Open problem 11: green and low carbon

With the development of the AI field, the AI-enabled industry gradually requires a greener and lower-carbon environment Liu and Zhou (2024). At present, the three cornerstones of AI algorithm, data, and computing power are developing on a large scale, resulting in higher and higher consumption of resources. Therefore, to achieve green and low-carbon intelligence, it is necessary to do "subtraction" Yang et al. (2024). At the same time, the deep integration of new energy vehicles, smart energy, and artificial intelligence has also brought great challenges to green and low-carbon intelligence. On the one hand, it builds a more flexible network model; on the other hand, it builds a more efficient and extensive sharing and reuse mechanism to realize green and low-carbon from a macro perspective. In short, the five development concepts of "innovation, coordination, green, openness, and

sharing” point out the direction for the future development of AI, and propose fundamental compliance.

6.12 Open problem 12: morality and ethics construction

At present, artificial intelligence has created considerable economic benefits for human beings, but the negative impact and ethical issues of its application have become increasingly prominent Huang et al. (2022). Predictable, constrained, and behavior-oriented artificial intelligence governance has become the priority in the era of artificial intelligence proposition. For example, the privacy protection of user data and information; the protection of knowledge achievements and algorithms, the excessive demand for portrait rights by AI face-changing, and the accountability of autonomous driving safety accidents, etc. AI technology may also be abused by criminals, for example, to engage in cybercrime, produce and disseminate fake news, and synthesize fake images that are enough to disrupt audiovisuals. Artificial intelligence should protect user privacy as the principle of AI development. Only in this way can the development of artificial intelligence give back to human beings and provide new hope for the new ethics between people and AI Akinrinola et al. (2024).

7 Conclusion and outlooks

After a long period of evolution in physics, the laws of knowledge are diverse and powerful. Inevitably, our current understanding of theory is only the tip of the iceberg. With the development of the field of artificial intelligence, there is a close connection between the field of deep learning and the field of physics. Combining physical knowledge with AI is not only the driving force for the progress of physics concepts but also promotes the development of a new generation of artificial intelligence. This paper first introduces the mechanism of physics and artificial intelligence, and then gives a corresponding overview of deep learning inspired by physics, mainly including classical mechanics, electromagnetism, statistical physics, and quantum mechanics for the inspiration deep learning, and expounds how deep learning solve physical problems. Finally, the challenges of physics-inspired artificial intelligence and thinking about the future are discussed. Through the interdisciplinary analysis and design of artificial intelligence and physics, more powerful and robust algorithms are explored to develop a new generation of artificial intelligence.

Author contributions J.J.C. was responsible for the conceptualization, methodology, and formal analysis of the article. J.J.C., S.X., and Y.C. wrote the main manuscript (writing - original draft) and Latex compilation. L.X. and L.L.L. search for relevant literature and perform statistics and analysis. C.P.H., T.X., and F.Z.X. were responsible for visualization (drawing charts), writing - reviewed, and editing. L.F., J.J.C., and L.X. provided funding and resources for the project and provided supervision. Y.S.Y., L.Y.Y., and G.Y.W. researched, wrote - reviewed, and edited the project. Z.X.R., W.S., M.W.P., H.B., and B.J. checked the paper's formula, grammar, and format (figures, references, etc.). All authors reviewed and approved the final draft of the manuscript. All authors are accountable for all aspects of the work.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Yang Y, Lv H, Chen N (2023) A survey on ensemble learning under the era of deep learning. *Artif Intell Rev* 56(6):5545–5589
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25(2)
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778
- Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. *Adv Neural Inf Process Syst* 27
- Hsieh WW (2009) Machine learning methods in the environmental sciences: neural networks and Kernels. Cambridge University Press, Cambridge
- Ivezić Ž, Connolly AJ, VanderPlas JT, Gray A (2019) Statistics, data mining, and machine learning in astronomy: a practical python guide for the analysis of survey data. Princeton University Press, Princeton
- Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N, Kumar V (2017) Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Trans Knowl Data Eng* 29(10):2318–2331
- Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V (2018) Machine learning for the geosciences: challenges and opportunities. *IEEE Trans Knowl Data Eng* 31(8):1544–1554
- Kutz JN (2017) Deep learning in fluid dynamics. *J Fluid Mech* 814:1–4
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N et al (2019) Deep learning and process understanding for data-driven earth system science. *Nature* 566(7743):195–204
- Jiao L-C, Yang S-Y, Liu F, Wang S-G, Feng Z-X (2016) Seventy years beyond neural networks: retrospect and prospect. *Chin J Comput* 39(8):1697–1716
- Muther T, Dahaghi AK, Syed FI, Van Pham V (2023) Physical laws meet machine intelligence: current developments and future directions. *Artif Intell Rev* 56(7):6947–7013
- Mehta P, Bukov M, Wang C-H, Day AG, Richardson C, Fisher CK, Schwab DJ (2019) A high-bias, low-variance introduction to machine learning for physicists. *Phys Rep* 810:1–124
- Zdeborová L (2020) Understanding deep learning is also a job for physicists. *Nat Phys* 16(6):602–604
- Meng C, Seo S, Cao D, Griesemer S, Liu Y (2022) When physics meets machine learning: a survey of physics-informed machine learning. arXiv preprint [arXiv:2203.16797](https://arxiv.org/abs/2203.16797)
- Engel A (2001) Statistical mechanics of learning. Cambridge University Press, Cambridge
- Widrow B, Lehr MA (1990) 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proc IEEE* 78(9):1415–1442
- Heisele B, Verri A, Poggio T (2002) Learning and vision machines. *Proc IEEE* 90(7):1164–1177
- Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding
- Rogers TT, McClelland JL (2004) Semantic cognition: a parallel distributed processing approach. The MIT Press, Cambridge
- Saxe AM, McClelland JL, Ganguli S (2018) A mathematical theory of semantic development in deep neural networks. *Appl Math.* <https://doi.org/10.1073/pnas.1820226116>

- Piech C, Bassen J, Huang J, Ganguli S, Sahami M, Guibas LJ, Sohl-Dickstein J (2015) Deep knowledge tracing. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 28. Curran Associates, Inc., NY. <https://proceedings.neurips.cc/paper/2015/file/bac9162b47c56fc8a4d2a519803d51b3-Paper.pdf>
- Khammash MH (2022) Cybergenetics: theory and applications of genetic control systems. Proc IEEE 110(5):631–658
- McCormick K (2022) Quantum field theory boosts brain model. Physics 15:50
- Tiberi L, Stapmanns J, Kühn T, Luu T, Dahmen D, Helias M (2022) Gell-Mann-low criticality in neural networks. Phys Rev Lett 128(16):168301
- Niyogi P, Girosi F, Poggio T (1998) Incorporating prior information in machine learning by creating virtual examples. Proc IEEE 86(11):2196–2209
- Werner G (2013) Consciousness viewed in the framework of brain phase space dynamics, criticality, and the renormalization group. Chaos, Solitons Fractals 55:3–12
- Masci J, Boscaini D, Bronstein M, Vandergheynst P (2015) Geodesic convolutional neural networks on riemannian manifolds. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 37–45
- Monti F, Boscaini D, Masci J, Rodola E, Svoboda J, Bronstein MM (2017) Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5115–5124
- Garcia Satorras V, Hoogeboom E, Fuchs F, Posner I, Welling M (2021) E (n) equivariant normalizing flows. Adv Neural Inf Process Syst 34:4181–4192
- Gerken J, Carlsson O, Linander H, Ohlsson F, Petersson C, Persson D (2022) Equivariance versus augmentation for spherical images. In: International Conference on Machine Learning, pp. 7404–7421. PMLR
- Hanik M, Steidl G, Tylicowicz C (2024) Manifold gcn: Diffusion-based convolutional neural network for manifold-valued graphs. arXiv preprint [arXiv:2401.14381](https://arxiv.org/abs/2401.14381)
- Cho S, Lee J, Kim D (2024) Hyperbolic vae via latent gaussian distributions. Adv Neural Inf Process Syst 36
- Katsman I, Chen E, Holalkere S, Asch A, Lou A, Lim SN, De Sa CM (2024) Riemannian residual neural networks. Adv Neural Inf Process Syst 36
- Gori M, Monfardini G, Scarselli F (2005) A new model for learning in graph domains. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, vol. 2, pp. 729–734
- You J, Ying R, Ren X, Hamilton WL, Leskovec J (2018) Graphrnn: Generating realistic graphs with deep auto-regressive models
- Xu B, Shen H, Cao Q, Qiu Y, Cheng X (2019) Graph wavelet neural network. arXiv preprint [arXiv:1904.07785](https://arxiv.org/abs/1904.07785)
- Lee YJ, Kahng H, Kim SB (2021) Generative adversarial networks for de novo molecular design. Molecular Informatics
- Wu C, Wu F, Cao Y, Huang Y, Xie X (2021) Fedgnn: Federated graph neural network for privacy-preserving recommendation. arXiv preprint [arXiv:2102.04925](https://arxiv.org/abs/2102.04925)
- Schuetz MJ, Brubaker JK, Katzgraber HG (2022) Combinatorial optimization with physics-inspired graph neural networks. Nat Mach Intell 4(4):367–377
- Yan H, Liu Y, Wei Y, Li Z, Li G, Lin L (2023) Skeletonmae: Graph-based masked autoencoder for skeleton sequence pre-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5606–5618
- Han Y, Wang P, Kundu S, Ding Y, Wang Z (2023) Vision hggn: An image is more than a graph of nodes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 19878–19888
- Fu X, Gao Y, Wei Y, Sun Q, Peng H, Li J, Li X (2024) Hyperbolic geometric latent diffusion model for graph generation. arXiv preprint [arXiv:2405.03188](https://arxiv.org/abs/2405.03188)
- Yao Y, Jin W, Ravi S, Joe-Wong C (2024) Fedgcn: Convergence-communication tradeoffs in federated training of graph convolutional networks. Adv Neural Inf Process Syst 36
- Chen C, Xu Z, Hu W, Zheng Z, Zhang J (2024) Fedgl: federated graph learning framework with global self-supervision. Inf Sci 657:119976
- Raissi M, Yazdani A, Karniadakis GE (2020) Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations. Science 367(6481):1026–1030
- Zhang Y, Ban X, Du F, Di W (2020) Fluidsnet: end-to-end learning for Lagrangian fluid simulation. Expert Syst Appl 152:113410
- Guan S, Deng H, Wang Y, Yang X (2022) Neurofluid: Fluid dynamics grounding with particle-driven neural radiance fields. arXiv preprint [arXiv:2203.01762](https://arxiv.org/abs/2203.01762)

- Toshev AP, Erbesdöbler JA, Adams NA, Brandstetter J (2024) Neural sph: Improved neural modeling of lagrangian fluid dynamics. arXiv preprint [arXiv:2402.06275](https://arxiv.org/abs/2402.06275)
- Greydanus S, Dzamba M, Yosinski J (2019) Hamiltonian neural networks. *Adv Neural Inf Process Syst* 32
- Toth P, Rezende DJ, Jaegle A, Racanière S, Botev A, Higgins I (2019) Hamiltonian generative networks. arXiv preprint [arXiv:1909.13789](https://arxiv.org/abs/1909.13789)
- Han C-D, Glaz B, Haile M, Lai Y-C (2021) Adaptable Hamiltonian neural networks. *Phys Rev Res* 3(2):023156
- Dierkes E, Fläskamp K (2021) Learning mechanical systems by Hamiltonian neural networks. *PAMM* 21(1):202100116
- Eidnes S, Stasik AJ, Sterud C, Bøhn E, Riemer-Sørensen S (2023) Pseudo-hamiltonian neural networks with state-dependent external forces. *Physica D* 446:133673
- Gong X, Li H, Zou N, Xu R, Duan W, Xu Y (2023) General framework for e (3)-equivariant neural network representation of density functional theory Hamiltonian. *Nat Commun* 14(1):2848
- Ma B, Yao X, An T, Dong B, Li Y (2023) Model free position-force control of environmental constrained reconfigurable manipulators based on adaptive dynamic programming. *Artif Intell Rev*, 1–29
- Kaltsas DA (2024) Constrained hamiltonian systems and physics informed neural networks: Hamilton-dirac neural nets. arXiv preprint [arXiv:2401.15485](https://arxiv.org/abs/2401.15485)
- Zhao K, Kang Q, Song Y, She R, Wang S, Tay WP (2024) Adversarial robustness in graph neural networks: a Hamiltonian approach. *Adv Neural Inf Process Syst* 36
- Lutter M, Ritter C, Peters J (2019) Deep Lagrangian networks: using physics as model prior for deep learning. arXiv preprint [arXiv:1907.04490](https://arxiv.org/abs/1907.04490)
- Cranmer M, Greydanus S, Hoyer S, Battaglia P, Spergel D, Ho S (2020) Lagrangian neural networks. arXiv preprint [arXiv:2003.04630](https://arxiv.org/abs/2003.04630)
- Bhattoo R, Ranu S, Krishnan NA (2023) Learning the dynamics of particle-based systems with Lagrangian graph neural networks. *Mach Learn: Sci Technol* 4(1):015003
- Xiao S, Zhang J, Tang Y (2024) Generalized Lagrangian neural networks. arXiv preprint [arXiv:2401.03728](https://arxiv.org/abs/2401.03728)
- Zhang X, Li Z, Change Loy C, Lin D (2017) Polynet: a pursuit of structural diversity in very deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 718–726
- Shi R, Morris Q (2021) Segmenting hybrid trajectories using latent odes. In: International Conference on Machine Learning, pp. 9569–9579. PMLR
- Yi Z (2023) nmode: neural memory ordinary differential equation. *Artif Intell Rev*, pp. 1–36
- Joshi M, Bhosale S, Vyawahare VA (2023) A survey of fractional calculus applications in artificial neural networks. *Artif Intell Rev*, pp. 1–54
- Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707
- Dwivedi V, Parashar N, Srinivasan B (2019) Distributed physics informed neural network for data-efficient solution to partial differential equations. arXiv preprint [arXiv:1907.08967](https://arxiv.org/abs/1907.08967)
- Morrill J, Salvi C, Kidger P, Foster J (2021) Neural rough differential equations for long time series. In: International Conference on Machine Learning, pp. 7829–7838. PMLR
- Zhang Z-Y, Zhang H, Zhang L-S, Guo L-L (2023) Enforcing continuous symmetries in physics-informed neural network for solving forward and inverse problems of partial differential equations. *J Comput Phys* 492:112415
- Mojgani R, Balajewicz M, Hassanzadeh P (2023) Kolmogorov n-width and Lagrangian physics-informed neural networks: a causality-conforming manifold for convection-dominated pdes. *Comput Methods Appl Mech Eng* 404:115810
- Xiao T, Yang R, Cheng Y, Suo J (2024) Shop: a deep learning framework for solving high-order partial differential equations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 16032–16039
- Kantamneni S, Liu Z, Tegmark M (2024) Optpde: Discovering novel integrable systems via AI-human collaboration. arXiv preprint [arXiv:2405.04484](https://arxiv.org/abs/2405.04484)
- Torres DF (2003) Quasi-invariant optimal control problems. arXiv preprint [arXiv:math/0302264](https://arxiv.org/abs/math/0302264)
- Torres DF (2004) Proper extensions of Noether's symmetry theorem for nonsmooth extremals of the calculus of variations. *Commun Pure Appl Anal* 3(3):491
- Frederico GS, Torres DF (2007) A formulation of Noether's theorem for fractional problems of the calculus of variations. *J Math Anal Appl* 334(2):834–846
- Gerken JE, Aronsson J, Carlsson O, Linander H, Ohlsson F, Petersson C, Persson D (2023) Geometric deep learning and equivariant neural networks. *Artif Intell Rev*, pp. 1–58
- Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P (2017) Geometric deep learning: going beyond Euclidean data. *IEEE Sig Process Magaz* 34(4):18–42

- Defferrard M, Milani M, Gusset F, Perrauidin N (2020) Deepsphere: a graph-based spherical cnn. arXiv preprint [arXiv:2012.15000](https://arxiv.org/abs/2012.15000)
- Armeni I, Sax S, Zamir AR, Savarese S (2017) Joint 2d-3d-semantic data for indoor scene understanding. arXiv preprint [arXiv:1702.01105](https://arxiv.org/abs/1702.01105)
- Bogo F, Romero J, Loper M, Black MJ (2014) Faust: Dataset and evaluation for 3d mesh registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3794–3801
- Boscaini D, Masci J, Rodolà E, Bronstein MM, Cremers D (2016) Anisotropic diffusion descriptors. In: Computer Graphics Forum, vol. 35, pp. 431–441. Wiley Online Library
- Cohen TS, Weiler M, Kicanaoglu B, Welling M (2019) Gauge equivariant convolutional networks and the icosahedral CNN
- De Haan P, Weiler M, Cohen T, Welling M (2020) Gauge equivariant mesh cnns: anisotropic convolutions on geometric graphs. arXiv preprint [arXiv:2003.05425](https://arxiv.org/abs/2003.05425)
- Tenenbaum JB, Silva VD, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319–2323
- Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326
- Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9(86):2579–2605
- McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint [arXiv:1802.03426](https://arxiv.org/abs/1802.03426)
- Kobak D, Linderman GC (2019) Umap does not preserve global structure any better than t-sne when using the same initialization. *BioRxiv*, 2019–12
- Kobak D, Linderman GC (2021) Initialization is critical for preserving global data structure in both t-sne and umap. *Nat Biotechnol* 39(2):156–157
- Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. MIT Press, Cambridge
- Wang J (2012) Diffusion maps. Springer, Berlin
- Hadsell R, Chopra S, Lecun Y (2006) Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
- Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks
- Scarselli F, Gorji M, Tsai AC, Hagenbuchner M, Monfardini G (2008) The graph neural network model. *IEEE Trans Neural Netw* 20(1):61–80
- Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY (2020) A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst* 32(1):4–24
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) Graph neural networks: a review of methods and applications. *AI Open* 1:57–81
- Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering
- Monti F, Bronstein MM, Bresson X (2017) Geometric matrix completion with recurrent multi-graph neural networks
- Veliković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks
- Kipf TN, Welling M (2016) Variational graph auto-encoders
- Pan S, Hu R, Long G, Jiang J, Yao L, Zhang C (2018) Adversarially regularized graph autoencoder for graph embedding. arXiv preprint [arXiv:1802.04407](https://arxiv.org/abs/1802.04407)
- Yu W, Cheng Z, Wei C, Aggarwal CC, Wei W (2018) Learning deep network representations with adversarially regularized autoencoders. In: the 24th ACM SIGKDD International Conference
- Cao S (2016) Deep neural networks for learning graph representations. In: Thirtieth AAAI Conference on Artificial Intelligence
- Ke T, Peng C, Xiao W, Yu PS, Zhu W (2018) Deep recursive network embedding with regular equivalence. In: the 24th ACM SIGKDD International Conference
- Li Y, Vinyals O, Dyer C, Pascanu R, Battaglia P (2018) Learning deep generative models of graphs
- Jiang B, Zhang Z, Lin D, Tang J (2018) Graph learning-convolutional networks
- Brockschmidt M (2019) Gnn-film: Graph neural networks with feature-wise linear modulation
- Jiang J, Cui Z, Xu C, Yang J (2019) Gaussian-induced convolution for graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 4007–4014
- Zhou Z, Li X (2017) Graph convolution: a high-order and adaptive approach
- Liu Q, Nickel M, Kiela D (2019) Hyperbolic graph neural networks
- Zhang Y, Pal S, Coates M, Stebey D (2018) Bayesian graph convolutional neural networks for semi-supervised classification
- Zhang R, Zou Y, Ma J (2019) Hyper-sagnn: a self-attention based graph neural network for hypergraphs. arXiv preprint [arXiv:1911.02613](https://arxiv.org/abs/1911.02613)

- Zhu D, Cui P, Wang D, Zhu W (2018) Deep variational network embedding in Wasserstein space. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2827–2836
- Wang D, Cui P, Zhu W (2016) Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1225–1234
- Berg Rvd, Kipf TN, Welling M (2017) Graph convolutional matrix completion. arXiv preprint [arXiv:1706.02263](https://arxiv.org/abs/1706.02263)
- Bojchevski A, Günnemann S (2017) Deep gaussian embedding of graphs: unsupervised inductive learning via ranking. arXiv preprint [arXiv:1707.03815](https://arxiv.org/abs/1707.03815)
- Qu M, Bengio Y, Tang J (2019) Gmnn: Graph markov neural networks. In: International Conference on Machine Learning, pp. 5241–5250. PMLR
- Lan S, Yu R, Yu G, Davis LS (2019) Modeling local geometric structure of 3d point clouds using geo-cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 998–1008
- Hernández Q, Badías A, Chinesta F, Cueto E (2022) Thermodynamics-informed graph neural networks. arXiv preprint [arXiv:2203.01874](https://arxiv.org/abs/2203.01874)
- Wessels H, Weißenfels C, Wriggers P (2020) The neural particle method—an updated Lagrangian physics informed neural network for computational fluid dynamics. Comput Methods Appl Mech Eng 368:113127
- Feynman RP (2005) The principle of least action in quantum mechanics. In: Feynman's Thesis: A New Approach To Quantum Theory, pp. 1–69. World Scientific, USA
- Choudhary A, Lindner JF, Holliday EG, Miller ST, Sinha S, Ditto WL (2020) Physics-enhanced neural networks learn order and chaos. Phys Rev E 101(6):062207
- Haber E, Ruthotto L (2017) Stable architectures for deep neural networks. Inverse Prob 34(1):014004
- Massaroli S, Poli M, Califano F, Faragasso A, Park J, Yamashita A, Asama H (2019) Port-Hamiltonian approach to neural network training. In: 2019 IEEE 58th Conference on Decision and Control (CDC), pp. 6799–6806. IEEE
- Lin HW, Tegmark M, Rolnick D (2017) Why does deep and cheap learning work so well? J Stat Phys 168(6):1223–1247
- Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96(456):1348–1360
- Ling J, Kurzawski A, Templeton J (2016) Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. J Fluid Mech 807:155–166
- Rubanova Y, Chen RT, Duvenaud DK (2019) Latent ordinary differential equations for irregularly-sampled time series. Adv Neural Inf Process Syst 32
- Du J, Futoma J, Doshi-Velez F (2020) Model-based reinforcement learning for semi-Markov decision processes with neural odes. Adv Neural Inf Process Syst 33:19805–19816
- Behrmann J, Grathwohl W, Chen RT, Duvenaud D, Jacobsen J-H (2019) Invertible residual networks. In: International Conference on Machine Learning, pp. 573–582. PMLR
- Chen RT, Rubanova Y, Bettencourt J, Duvenaud DK (2018) Neural ordinary differential equations. Adv Neural Inf Process Syst 31
- Larsson G, Maire M, Shakhnarovich G (2016) Fractalnet: Ultra-deep neural networks without residuals. arXiv preprint [arXiv:1605.07648](https://arxiv.org/abs/1605.07648)
- Ramacher U (1993) Hamiltonian dynamics of neural networks. In: Neurobionics, pp. 61–85. Elsevier, Neuherberg, Germany
- Meng X, Li Z, Zhang D, Karniadakis GE (2020) Ppinn: parareal physics-informed neural network for time-dependent pdes. Comput Methods Appl Mech Eng 370:113250
- Fang Z (2021) A high-efficient hybrid physics-informed neural networks based on convolutional neural network. IEEE Transactions on Neural Networks and Learning Systems
- Moseley B, Markham A, Nissen-Meyer T (2021) Finite basis physics-informed neural networks (fbpinns): a scalable domain decomposition approach for solving differential equations. arXiv preprint [arXiv:2107.07871](https://arxiv.org/abs/2107.07871)
- Chen Y, Huang D, Zhang D, Zeng J, Wang N, Zhang H, Yan J (2021) Theory-guided hard constraint projection (hcp): a knowledge-based data-driven scientific machine learning method. J Comput Phys 445:110624
- Schiassi E, D'Ambrosio A, Drozd K, Curti F, Furfaro R (2022) Physics-informed neural networks for optimal planar orbit transfers. J Spacecr Rockets 10(2514/1):A35138
- Treibert S, Ehrhardt M (2021) An unsupervised physics-informed neural network to model covid-19 infection and hospitalization scenarios

- Trejo I, Hengartner NW (2022) A modified susceptible-infected-recovered model for observed under-reported incidence data. *PLoS ONE* 17(2):0263047
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L (2021) Physics-informed machine learning. *Nat Rev Phys* 3(6):422–440
- Kidger P, Morrill J, Foster J, Lyons T (2020) Neural controlled differential equations for irregular time series. *Adv Neural Inf Process Syst* 33:6696–6707
- Che Z, Purushotham S, Cho K, Sontag D, Liu Y (2018) Recurrent neural networks for multivariate time series with missing values. *Sci Rep* 8(1):1–12
- Coffey W, Kalmykov YP (2012) The Langevin equation: with applications to stochastic problems in physics, chemistry and electrical engineering, vol 27. World Scientific, USA
- Pavliotis GA (2014) Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations, vol 60. Springer, Berlin
- Black F, Scholes M (2019) The pricing of options and corporate liabilities. In: World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 1: Foundations of CCA and Equity Valuation, pp. 3–21. World Scientific, Singapore
- Arató M (2003) A famous nonlinear stochastic equation (Lotka-Volterra model with diffusion). *Math Comput Model* 38(7–9):709–726
- Huillet T (2007) On wright-fisher diffusion and its relatives. *J Stat Mech: Theory Exp* 2007(11):11006
- Kidger P (2022) On neural differential equations. *CoRR* [arXiv:abs/2202.02435](https://arxiv.org/abs/2202.02435)
- Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast* 20(1):5–10
- Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: J Econ Soc* 50:987–1007
- Hannan EJ, Rissanen J (1982) Recursive estimation of mixed autoregressive-moving average order. *Biométrika* 69(1):81–94
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econ* 31(3):307–327
- Lu L, Meng X, Mao Z, Karniadakis GE (2021) Deepxde: a deep learning library for solving differential equations. *SIAM Rev* 63(1):208–228
- Duvenaud DK, Maclaurin D, Iparragirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP (2015) Convolutional networks on graphs for learning molecular fingerprints. *Adv Neural Inf Process Syst* 28
- Morgan HL (1965) The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *J Chem Doc* 5(2):107–113
- Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50(5):742–754
- Manco L, Maffei N, Strolin S, Vichi S, Bottazzi L, Strigari L (2021) Basic of machine learning and deep learning in imaging for medical physicists. *Physica Med* 83:194–205
- Hrinivich WT, Lee J (2020) Artificial intelligence-based radiotherapy machine parameter optimization using reinforcement learning. *Med Phys* 47(12):6140–6150
- Maffei N, Manco L, Aluisio G, D'Angelo E, Ferrazza P, Vanoni V, Meduri B, Lohr F, Guidi G (2021) Radiomics classifier to quantify automatic segmentation quality of cardiac sub-structures for radiotherapy treatment planning. *Physica Med* 83:278–286
- Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, Willems S, Vandewincke L, Holmström M, Löfman F et al (2021) Artificial intelligence and machine learning for medical imaging: a technology review. *Physica Med* 83:242–256
- Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, D'Amico NC, Sardanelli F (2021) Ai applications to medical images: from machine learning to deep learning. *Physica Med* 83:9–24
- Li C, Li W, Liu C, Zheng H, Cai J, Wang S (2022) Artificial intelligence in multiparametric magnetic resonance imaging: a review. *Med Phys* 49(10):1024–1054
- Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, Rueckert D (2017) Distance metric learning using graph convolutional networks: Application to functional brain networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 469–477. Springer
- Zegers C, Posch J, Traverso A, Eekers D, Postma A, Backes W, Dekker A, Elmpet W (2021) Current applications of deep-learning in neuro-oncological MRI. *Physica Med* 83:161–173
- Rizk B, Brat H, Zille P, Guillain R, Pouchy C, Adam C, Ardon R, d'Assignies G (2021) Meniscal lesion detection and characterization in adult knee MRI: a deep learning model approach with external validation. *Physica Med* 83:64–71
- Montalt-Tordera J, Muthurangu V, Hauptmann A, Steeden JA (2021) Machine learning in magnetic resonance imaging: image reconstruction. *Physica Med* 83:79–87

- Battaglia P, Pascanu R, Lai M, Jimenez Rezende D et al (2016) Interaction networks for learning about objects, relations and physics. *Adv Neural Inf Process Syst* 29
- Chang MB, Ullman T, Torralba A, Tenenbaum JB (2016) A compositional object-based approach to learning physical dynamics. arXiv preprint [arXiv:1612.00341](https://arxiv.org/abs/1612.00341)
- Donon B, Donnot B, Guyon I, Marot A (2019) Graph neural solver for power systems. In: 2019 International Joint Conference on Neural Networks (ijcnn), pp. 1–8. IEEE
- Park J, Park J (2019) Physics-induced graph neural network: an application to wind-farm power estimation. *Energy* 187:115883
- Bapst V, Keck T, Grabska-Barwińska A, Donner C, Cubuk ED, Schoenholz SS, Obika A, Nelson AW, Back T, Hassabis D et al (2020) Unveiling the predictive power of static structure in glassy systems. *Nat Phys* 16(4):448–454
- Psaltis D, Farhat N (1985) Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Opt Lett* 10(2):98–100
- Chen Y (1993) 4f-type optical system for matrix multiplication. *Opt Eng* 32(1):77–79
- Francis T, Yang X, Yin S, Gregory DA (1991) Mirror-array optical interconnected neural network. *Opt Lett* 16(20):1602–1604
- Nitta Y, Ohta J, Tai S, Kyuma K (1993) Optical learning neurochip with internal analog memory. *Appl Opt* 32(8):1264–1274
- Wang Y-J, Zhang Y, Guo Z (1997) Optically interconnected neural networks using prism arrays. *Opt Eng* 36:2249–2253
- Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, Ozcan A (2018) All-optical machine learning using diffractive deep neural networks. *Science* 361(6406):1004–1008
- Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, Fang L, Lin X, Dai Q (2019) Fourier-space diffractive deep neural network. *Phys Rev Lett* 123(2):023901
- Mengu D, Luo Y, Rivenson Y, Ozcan A (2019) Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J Sel Top Quantum Electron* 26(1):1–14
- Hamerly R, Bernstein L, Sludds A, Soljačić M, Englund D (2019) Large-scale optical neural networks based on photoelectric multiplication. *Phys Rev X* 9(2):021032
- Du Y, Su K, Yuan X, Li T, Liu K, Man H, Zou X (2023) Implementation of optical neural network based on mach-zehnder interferometer array. *IET Optoelectron* 17(1):1–11
- Giamougiannis G, Tsakyridis A, Ma Y, Totović A, Moralis-Pegios M, Lazovsky D, Pleros N (2023) A coherent photonic crossbar for scalable universal linear optics. *J Lightwave Technol* 41(8):2425–2442
- Li J, Gan T, Bai B, Luo Y, Jarrahi M, Ozcan A (2023) Massively parallel universal linear transformations using a wavelength-multiplexed diffractive optical network. *Adv Photonics* 5(1):016003–016003
- Huang L, Tanguy QA, Fröhlich JE, Mukherjee S, Böhringer KF, Majumdar A (2024) Photonic advantage of optical encoders. *Nanophotonics* 13(7):1191–1196
- Antonik P, Marsal N, Brunner D, Rontani D (2019) Human action recognition with a large-scale brain-inspired photonic computer. *Nat Mach Intell* 1(11):530–537
- Katumba A, Yin X, Dambre J, Bienstman P (2019) A neuromorphic silicon photonics nonlinear equalizer for optical communications with intensity modulation and direct detection. *J Lightwave Technol* 37(10):2232–2239
- Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WH (2019) All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* 569(7755):208–214
- Bao Q, Zhang H, Ni Z, Wang Y, Polavarapu L, Shen Z, Xu Q-H, Tang D, Loh KP (2011) Monolayer graphene as a saturable absorber in a mode-locked laser. *Nano Res* 4(3):297–307
- Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Laroche H, Englund D et al (2017) Deep learning with coherent nanophotonic circuits. *Nat Photonics* 11(7):441–446
- Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George J, Kildishev AV, Pelton M, Sorger VJ (2018) All-optical nonlinear activation function for photonic neural networks. *Opt Mater Express* 8(12):3851–3863
- Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y-C, Chen P, Jo G-B, Liu J, Du S (2019) All-optical neural network with nonlinear activation functions. *Optica* 6(9):1132–1137
- Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G (2018) Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci Rep* 8(1):1–10
- Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Le Gallo M, Fu X, Lukashchuk A, Raja AS et al (2021) Parallel convolutional processing using an integrated photonic tensor core. *Nature* 589(7840):52–58
- Wang B, Yu W, Duan J, Yang S, Zhao Z, Zheng S, Zhang W (2023) Microdisk modulator-assisted optical nonlinear activation functions for photonic neural networks. arXiv preprint [arXiv:2306.04361](https://arxiv.org/abs/2306.04361)
- Wang T, Sohoni MM, Wright LG, Stein MM, Ma S-Y, Onodera T, Anderson MG, McMahon PL (2023) Image sensing with multilayer nonlinear optical neural networks. *Nat Photonics* 17(5):408–415

- Oguz I, Hsieh J-L, Dinc NU, Teğin U, Yıldırım M, Gigli C, Moser C, Psaltis D (2024) Programming nonlinear propagation for efficient optical learning machines. *Adv Photon* 6(1):016002–016002
- Li L, Wang LG, Teixeira FL, Liu C, Nehorai A, Cui TJ (2018) Deepnis: deep neural network for nonlinear electromagnetic inverse scattering. *IEEE Trans Antennas Propag* 67(3):1819–1825
- Wei Z, Chen X (2019) Physics-inspired convolutional neural network for solving full-wave inverse scattering problems. *IEEE Trans Antennas Propag* 67(9):6138–6148
- Guo L, Song G, Wu H (2021) Complex-valued pix2pix—deep neural network for nonlinear electromagnetic inverse scattering. *Electronics* 10(6):752
- Bernstein L, Sludds A, Panuski C, Trajtenberg-Mills S, Hamerly R, Englund D (2023) Single-shot optical neural network. *Sci Adv* 9(25):7904
- Yang M, Robertson E, Esguerra L, Busch K, Wolters J (2023) Optical convolutional neural network with atomic nonlinearity. *Opt Express* 31(10):16451–16459
- Huang Z, Gu Z, Shi M, Gao Y, Liu X (2024) Op-fcnn: an optronic fully convolutional neural network for imaging through scattering media. *Opt Express* 32(1):444–456
- Goodman JW, Dias A, Woody L (1978) Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Opt Lett* 2(1):1–3
- Liu H-K, Kung S, Davis JA (1986) Real-time optical associative retrieval technique. *Opt Eng* 25(7):853–856
- Francis T, Lu T, Yang X, Gregory DA (1990) Optical neural network with pocket-sized liquid-crystal televisions. *Opt Lett* 15(15):863–865
- Yang X, Lu T, Francis T (1990) Compact optical neural network using cascaded liquid crystal television. *Appl Opt* 29(35):5223–5225
- Psaltis D, Brady D, Wagner K (1988) Adaptive optical networks using photorefractive crystals. *Appl Opt* 27(9):1752–1759
- Slinger C (1991) Analysis of the n-to-n volume-holographic neural interconnect. *JOSA A* 8(7):1074–1081
- Yang G-Z, Dong B-Z, Gu B-Y, Zhuang J-Y, Ersoy OK (1994) Gerchberg-saxton and yang-gu algorithms for phase retrieval in a nonunitary transform system: a comparison. *Appl Opt* 33(2):209–218
- Di Leonardo R, Ianni F, Ruocco G (2007) Computer generation of optimal holograms for optical trap arrays. *Opt Express* 15(4):1913–1922
- Nogrette F, Labuhn H, Ravets S, Barredo D, Béguin L, Vernier A, Lahaye T, Browaeys A (2014) Single-atom trapping in holographic 2d arrays of microtraps with arbitrary geometries. *Phys Rev X* 4(2):021034
- Qian C, Lin X, Lin X, Xu J, Sun Y, Li E, Zhang B, Chen H (2020) Performing optical logic operations by a diffractive neural network. *Light: Sci Appl* 9(1):1–7
- Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D et al (2017) Deep learning with coherent nanophotonic circuits. *Nat Photonics* 11(7):441–446
- Bagherian H, Skirlo S, Shen Y, Meng H, Ceperic V, Soljacic M (2018) On-chip optical convolutional neural networks. arXiv preprint [arXiv:1808.03303](https://arxiv.org/abs/1808.03303)
- Zang Y, Chen M, Yang S, Chen H (2019) Electro-optical neural networks based on time-stretch method. *IEEE J Sel Top Quantum Electron* 26(1):1–10
- Dunning G, Oweckyo K, Soffer B (1991) Hybrid optoelectronic neural networks using a mutually pumped phase-conjugate mirror. *Opt Lett* 16(12):928–930
- Skinner SR, Steck JE, Behrman EC (1994) Optical neural network using Kerr-type nonlinear materials. In: Proceedings of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems, pp. 12–15. IEEE
- Larger L, Soriano MC, Brunner D, Appeltant L, Gutiérrez JM, Pesquera L, Mirasso CR, Fischer I (2012) Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Opt Express* 20(3):3241–3249
- Williamson IA, Hughes TW, Minkov M, Bartlett B, Pai S, Fan S (2019) Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J Sel Top Quantum Electron* 26(1):1–12
- Fard MMP, Williamson IA, Edwards M, Liu K, Pai S, Bartlett B, Minkov M, Hughes TW, Fan S, Nguyen T-A (2020) Experimental realization of arbitrary activation functions for optical neural networks. *Opt Express* 28(8):12138–12148
- Saxena I, Fiesler E (1995) Adaptive multilayer optical neural network with optical thresholding. *Opt Eng* 34(8):2435–2440
- Vandoorne K, Dierckx W, Schrauwen B, Verstraeten D, Baets R, Bienstman P, Van Campenhout J (2008) Toward optical signal processing using photonic reservoir computing. *Opt Express* 16(15):11182–11192
- Vandoorne K, Meche P, Van Vaerenbergh T, Fiers M, Morthier G, Verstraeten D, Schrauwen B, Dambre J, Bienstman P (2014) Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat Commun* 5(1):1–6

- Rosenbluth D, Kravtsov K, Fok MP, Prucnal PR (2009) A high performance photonic pulse processing device. *Opt Express* 17(25):22767–22772
- Mesaritakis C, Papataxiaris V, Syvridis D (2013) Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system. *JOSA B* 30(11):3048–3055
- Denis-Le Coarer F, Sciamanna M, Katumba A, Freiberger M, Dambre J, Bienstman P, Rontani D (2018) All-optical reservoir computing on a photonic chip using silicon-based ring resonators. *IEEE J Sel Top Quantum Electron* 24(6):1–8
- Schirmer RW, Gaeta AL (1997) Nonlinear mirror based on two-photon absorption. *JOSA B* 14(11):2865–2868
- Shan T, Dang X, Li M, Yang F, Xu S, Wu J (2018) Study on a 3d possion's equation slover based on deep learning technique. In: 2018 IEEE International Conference on Computational Electromagnetics (ICCEM), pp. 1–3. IEEE
- Tsakyridis A, Moralis-Pegios M, Giampougiannis G, Kirtas M, Passalis N, Tefas A, Pleros N (2024) Photonic neural networks and optics-informed deep learning fundamentals. *APL Photonics* 9(1)
- Matuszewski M, Prystupiuk A, Opala A (2024) Role of all-optical neural networks. *Phys Rev Appl* 21(1):014028
- Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. *The Euro Phys J Special Topics* 178(1):13–23
- Riesen K, Bunke H (2009) Approximate graph edit distance computation by means of bipartite graph matching. *Image Vis Comput* 27(7):950–959
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint [arXiv:1301.3781](https://arxiv.org/abs/1301.3781)
- Goldfeld Z, Patel D, Sreekumar S, Wilde MM (2024) Quantum neural estimation of entropies. *Phys Rev A* 109(3):032431
- Poole B, Lahiri S, Raghu M, Sohl-Dickstein J, Ganguli S (2016) Exponential expressivity in deep neural networks through transient chaos. *Adv Neural Inf Process Syst* 29
- Keup C, Kühn T, Dahmen D, Helias M (2021) Transient chaotic dimensionality expansion by recurrent networks. *Phys Rev X* 11(2):021064
- Mohanrasu S, Udhayakumar K, Priyanka T, Gowrisankar A, Banerjee S, Rakkiyappan R (2023) Event-triggered impulsive controller design for synchronization of delayed chaotic neural networks and its fractal reconstruction: an application to image encryption. *Appl Math Model* 115:490–512
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79(8):2554–2558
- Liu L, Zhang L, Jiang D, Guan Y, Zhang Z (2019) A simultaneous scrambling and diffusion color image encryption algorithm based on hopfield chaotic neural network. *IEEE Access* 7:185796–185810
- Lin H, Wang C, Yu F, Sun J, Du S, Deng Z, Deng Q (2023) A review of chaotic systems based on memristive hopfield neural networks. *Mathematics* 11(6):1369
- Ma Q, Ma Z, Xu J, Zhang H, Gao M (2024) Message passing variational autoregressive network for solving intractable ising models. arXiv preprint [arXiv:2404.06225](https://arxiv.org/abs/2404.06225)
- Laydevant J, Marković D, Grollier J (2024) Training an Ising machine with equilibrium propagation. *Nat Commun* 15(1):3671
- Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680
- Salakhutdinov R, Murray I (2008) On the quantitative analysis of deep belief networks. In: Proceedings of the 25th International Conference on Machine Learning, pp. 872–879
- Bras P, Pagès G (2023) Convergence of Langevin-simulated annealing algorithms with multiplicative noise II: total variation. *Monte Carlo Methods Appl* 29(3):203–219
- Karacan I, Senvar O, Bulkan S (2023) A novel parallel simulated annealing methodology to solve the no-wait flow shop scheduling problem with earliness and tardiness objectives. *Processes* 11(2):454
- Milisav F, Bazinet V, Betzel R, Misic B (2024) A simulated annealing algorithm for randomizing weighted networks. bioRxiv, 2024–02
- Lee H, Grosse R, Ranganath R, Ng AY (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 609–616
- Lee H, Ekanadham C, Ng A (2007) Sparse deep belief net model for visual area v2. *Adv Neural Inf Process Syst* 20
- Feng S, Chen CP (2016) A fuzzy restricted Boltzmann machine: Novel learning algorithms based on the crisp possibilistic mean value of fuzzy numbers. *IEEE Trans Fuzzy Syst* 26(1):117–130
- Lang AH, Loukianov AD, Fisher CK (2023) Neural Boltzmann machines. arXiv preprint [arXiv:2305.08337](https://arxiv.org/abs/2305.08337)

- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. *Adv Neural Inf Process Syst* 27
- Van Oord A, Kalchbrenner N, Kavukcuoglu K (2016) Pixel recurrent neural networks. In: International Conference on Machine Learning, pp. 1747–1756. PMLR
- Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232
- Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410
- Wang C, Hu H, Lu Y (2019) A solvable high-dimensional model of gan. *Adv Neural Inf Process Syst* 32
- Guo Q, Ma C, Jiang Y, Yuan Z, Yu Y, Luo P (2023) Egc: Image generation and classification via a diffusion energy-based model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 22952–22962
- Yu J, Wang Y, Zhao C, Ghanem B, Zhang J (2023) Freedom: Training-free energy-guided conditional diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 23174–23184
- Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114)
- Vahdat A, Kautz J (2020) Nvae: a deep hierarchical variational autoencoder. *Adv Neural Inf Process Syst* 33:19667–19679
- Cui J, Wu YN, Han T (2023) Learning hierarchical features with joint latent space energy-based prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2218–2227
- Salimans T, Karpathy A, Chen X, Kingma DP (2017) Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. arXiv preprint [arXiv:1701.05517](https://arxiv.org/abs/1701.05517)
- Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV (2019) Xlnet: generalized autoregressive pretraining for language understanding. *Adv Neural Inf Process Syst* 32
- Wu T, Fan Z, Liu X, Gong Y, Shen Y, Jiao J, Zheng H-T, Li J, Wei Z, Guo J et al (2023) Ar-diffusion: autoregressive diffusion model for text generation. arXiv preprint [arXiv:2305.09515](https://arxiv.org/abs/2305.09515)
- Kohonen T (1989) Self-organizing feature maps. In: Self-organization and Associative Memory, pp. 119–157. Springer, USA
- Budroni M, De Wit A (2017) Dissipative structures: from reaction-diffusion to chemo-hydrodynamic patterns. *Chaos: Interdiscip J Nonlinear Sci* 27(10):104617
- Dauphin YN, Pascanu R, Gulcehre C, Cho K, Ganguli S, Bengio Y (2014) Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Adv Neural Inf Process Syst* 27
- Choromanska A, Henaff M, Mathieu M, Arous GB, LeCun Y (2015) The loss surfaces of multilayer networks. In: Artificial Intelligence and Statistics, pp. 192–204. PMLR
- Kawaguchi K (2016) Deep learning without poor local minima. *Adv Neural Inf Process Syst* 29
- Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comput* 14(8):1771–1800
- Schneider E, Dai L, Topper RQ, Drechsler-Grau C, Tuckerman ME (2017) Stochastic neural network approach for learning high-dimensional free energy surfaces. *Phys Rev Lett* 119(15):150601
- Sidky H, Whitmer JK (2018) Learning free energy landscapes using artificial neural networks. *J Chem Phys* 148(10):104111
- Noé F, Olsson S, Köhler J, Wu H (2019) Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* 365(6457):1147
- Hinton G, Vinyals O, Dean J et al (2015) Distilling the knowledge in a neural network. arXiv preprint [arXiv:1503.02531](https://arxiv.org/abs/1503.02531) 2(7)
- Huang Z, Wang N (2017) Like what you like: knowledge distill via neuron selectivity transfer. arXiv preprint [arXiv:1707.01219](https://arxiv.org/abs/1707.01219)
- Furlanello T, Lipton Z, Tschanen M, Itti L, Anandkumar A (2018) Born again neural networks. In: International Conference on Machine Learning, pp. 1607–1616. PMLR
- Zhang Y, Xiang T, Hospedales TM, Lu H (2018) Deep mutual learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4320–4328
- Huang J, Guo Z (2023) Pixel-wise contrastive distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 16359–16369
- Wang H, Li Y, Xu W, Li R, Zhan Y, Zeng Z (2023) Dafkd: domain-aware federated knowledge distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20412–20421
- Huang T, Zhang Y, Zheng M, You S, Wang F, Qian C, Xu C (2024) Knowledge diffusion for distillation. *Adv Neural Inf Process Syst* 36

- Ham S, Park J, Han D-J, Moon J (2024) Neo-kd: Knowledge-distillation-based adversarial training for robust multi-exit neural networks. *Adv Neural Inf Process Syst* 36
- Peng H, Du H, Yu H, Li Q, Liao J, Fu J (2020) Cream of the crop: distilling prioritized paths for one-shot neural architecture search. *Adv Neural Inf Process Syst* 33:17955–17964
- Li C, Peng J, Yuan L, Wang G, Liang X, Lin L, Chang X (2020) Block-wisely supervised neural architecture search with knowledge distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1989–1998
- Guan Y, Zhao P, Wang B, Zhang Y, Yao C, Bian K, Tang J (2020) Differentiable feature aggregation search for knowledge distillation. In: European Conference on Computer Vision, pp. 469–484. Springer
- Kang M, Mun J, Han B (2020) Towards oracle knowledge distillation with neural architecture search. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 4404–4411
- Nath U, Wang Y, Yang Y (2023) Rnas-cl: Robust neural architecture search by cross-layer knowledge distillation. arXiv preprint [arXiv:2301.08092](https://arxiv.org/abs/2301.08092)
- Trofimov I, Klyuchnikov N, Sahnikov M, Filippov A, Burnaev E (2023) Multi-fidelity neural architecture search with knowledge distillation. *IEEE Access*
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA* 105(4):1118–1123
- Kuhn HW (1955) The Hungarian method for the assignment problem. *Naval Res Log Q* 2(1–2):83–97
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst* 26
- Sompolinsky H, Crisanti A, Sommers H-J (1988) Chaos in random neural networks. *Phys Rev Lett* 61(3):259
- Lin W, Chen G (2009) Large memory capacity in chaotic artificial neural networks: a view of the anti-integrable limit. *IEEE Trans Neural Networks* 20(8):1340–1351
- Marshall AW (1954) The use of multi-stage sampling schemes in monte Carlo computations. Technical report, RAND CORP SANTA MONICA CALIF
- Sohl-Dickstein J, Culpepper BJ (2012) Hamiltonian annealed importance sampling for partition function estimation. arXiv preprint [arXiv:1205.1925](https://arxiv.org/abs/1205.1925)
- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828
- Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S (2015) Deep unsupervised learning using non-equilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265. PMLR
- Oord A, Kalchbrenner N, Espeholt L, Vinyals O, Graves A et al (2016) Conditional image generation with pixelcnn decoders. *Adv Neural Inf Process Syst* 29
- Nguyen HC, Zecchina R, Berg J (2017) Inverse statistical problems: from the inverse ising problem to data science. *Adv Phys* 66(3):197–261
- Bengio Y (2009) Learning deep architectures for AI. Now Publishers Inc, USA
- Ranzato M, Krizhevsky A, Hinton G (2010) Factored 3-way restricted boltzmann machines for modeling natural images. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 621–628. JMLR Workshop and Conference Proceedings
- Ranzato M, Hinton GE (2010) Modeling pixel means and covariances using factorized third-order boltzmann machines. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2551–2558. IEEE
- Salakhutdinov R, Mnih A, Hinton G (2007) Restricted boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning, pp. 791–798
- Ji N, Zhang J, Zhang C, Yin Q (2014) Enhancing performance of restricted Boltzmann machines via log-sum regularization. *Knowl-Based Syst* 63:82–96
- Cocco S, Monasson R, Posani L, Rosay S, Tubiana J (2018) Statistical physics and representations in real and artificial neural networks. *Physica A* 504:45–76
- Tubiana J, Monasson R (2017) Emergence of compositional representations in restricted Boltzmann machines. *Phys Rev Lett* 118(13):138301
- Barra A, Genovese G, Sollich P, Tantari D (2018) Phase diagram of restricted Boltzmann machines and generalized hopfield networks with arbitrary priors. *Phys Rev E* 97(2):022310
- Mézard M (2017) Mean-field message-passing equations in the hopfield model and its generalizations. *Phys Rev E* 95(2):022117
- LeCun Y, Chopra S, Hadsell R, Ranzato M, Huang F (2006) A tutorial on energy-based learning. *Predicting Struct Data* 1(0)

- Pernkopf F, Peharz R, Tschiatschek S (2014) Introduction to probabilistic graphical models. In: Academic Press Library in Signal Processing vol. 1, pp. 989–1064. Elsevier, Academic Press
- Dinh L, Krueger D, Bengio Y (2014) Nice: Non-linear independent components estimation. arXiv preprint [arXiv:1410.8516](https://arxiv.org/abs/1410.8516)
- Dinh L, Sohl-Dickstein J, Bengio S (2016) Density estimation using real nvp. arXiv preprint [arXiv:1605.08803](https://arxiv.org/abs/1605.08803)
- Rezende D, Danihelka I, Gregor K, Wierstra D et al (2016) One-shot generalization in deep generative models. In: International Conference on Machine Learning, pp. 1521–1529. PMLR
- Wang L (2018) Generative models for physicists
- Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Fu H, Gong M, Wang C, Batmanghelich K, Zhang K, Tao D (2019) Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2427–2436
- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv. <https://doi.org/10.48550/ARXIV.1701.07875>. <https://arxiv.org/abs/1701.07875>
- Cinelli LP, Marins MA, Da Silva EAB, Netto SL (2021) Variational methods for machine learning with applications to deep networks. Springer, Berlin
- Rezende DJ, Mohamed S, Wierstra D (2014) Stochastic backpropagation and variational inference in deep latent gaussian models. In: International Conference on Machine Learning, vol. 2, p. 2. Citeseer
- Gregor K, Danihelka I, Mnih A, Blundell C, Wierstra D (2014) Deep autoregressive networks. In: International Conference on Machine Learning, pp. 1242–1250. PMLR
- Ozair S, Bengio Y (2014) Deep directed generative autoencoders. arXiv preprint [arXiv:1410.0630](https://arxiv.org/abs/1410.0630)
- Wu D, Wang L, Zhang P (2019) Solving statistical mechanics using variational autoregressive networks. Phys Rev Lett 122(8):080602
- Sharir O, Levine Y, Wies N, Carleo G, Shashua A (2020) Deep autoregressive models for the efficient variational simulation of many-body quantum systems. Phys Rev Lett 124(2):020503
- Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning based recommender system: a survey and new perspectives. ACM Comput Surv (CSUR) 52(1):1–38
- Mehta P, Schwab DJ (2014) An exact mapping between the variational renormalization group and deep learning. arXiv preprint [arXiv:1410.3831](https://arxiv.org/abs/1410.3831)
- Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480
- Kohonen T, Oja E, Simula O, Visa A, Kangas J (1996) Engineering applications of the self-organizing map. Proc IEEE 84(10):1358–1384
- Amemiya T, Shibata K, Itoh Y, Itoh K, Watanabe M, Yamaguchi T (2017) Primordial oscillations in life: direct observation of glycolytic oscillations in individual hela cervical cancer cells. Chaos: Interdiscip J Nonlinear Sci 27(10):104602
- Kondopudi D, Kay B, Dixon J (2017) Dissipative structures, machines, and organisms: a perspective. Chaos: Interdiscip J Nonlinear Sci 27(10):104607
- Boyd S, Boyd SP, Vandenberghe L (2004) Convex Optimiz. Cambridge University Press, UK
- Bray AJ, Dean DS (2007) Statistics of critical points of gaussian fields on large-dimensional spaces. Phys Rev Lett 98(15):150201
- Fyodorov YV, Williams I (2007) Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity. J Stat Phys 129(5):1081–1116
- Tieleman T, Hinton G (2009) Using fast weights to improve persistent contrastive divergence. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 1033–1040
- Hyvärinen A, Dayan P (2005) Estimation of non-normalized statistical models by score matching. J Mach Learn Res 6(4)
- Besag J (1975) Statistical analysis of non-lattice data. J R Stat Soc: Series D (The Statistician) 24(3):179–195
- Battaglino PB (2014) Minimum probability flow learning: a new method for fitting probabilistic models. University of California, Berkeley
- Sohl-Dickstein J, Battaglino PB, DeWeese MR (2011) New method for parameter estimation in probabilistic models: minimum probability flow. Phys Rev Lett 107(22):220601
- Wehmeyer C, Noé F (2018) Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. J Chem Phys 148(24):241703
- Mardt A, Pasquali L, Wu H, Noé F (2018) Vampnets for deep learning of molecular kinetics. Nat Commun 9(1):1–11
- Xu Z, Hsu Y-C, Huang J (2017) Training shallow and thin networks for acceleration via knowledge distillation with conditional adversarial networks. arXiv preprint [arXiv:1709.00513](https://arxiv.org/abs/1709.00513)

- Wang D, Gong C, Li M, Liu Q, Chandra V (2021) Alphanet: Improved training of supernets with alpha-divergence. In: International Conference on Machine Learning, pp. 10760–10771. PMLR
- Gu J, Tresp V (2020) Search for better students to learn distilled knowledge. arXiv preprint [arXiv:2001.11612](https://arxiv.org/abs/2001.11612)
- Macko V, Weill C, Mazzawi H, Gonzalvo J (2019) Improving neural architecture search image classifiers via ensemble learning. arXiv preprint [arXiv:1903.06236](https://arxiv.org/abs/1903.06236)
- Liu H, Simonyan K, Yang Y (2018) Darts: Differentiable architecture search. arXiv preprint [arXiv:1806.09055](https://arxiv.org/abs/1806.09055)
- European, Plastics, News, group: Rubik's cube (1974). European Plastics News (2015)
- McAleer S, Agostinelli F, Shmakov A, Baldi P (2018) Solving the rubik's cube without human knowledge. arXiv preprint [arXiv:1805.07470](https://arxiv.org/abs/1805.07470)
- Agostinelli F, McAleer S, Shmakov A, Baldi P (2019) Solving the Rubik's cube with deep reinforcement learning and search. Nat Mach Intell 1(8):356–363
- Corli S, Moro L, Galli DE, Prati E (2021) Solving Rubik's cube via quantum mechanics and deep reinforcement learning. J Phys A: Math Theor 54(42):425302
- Johnson CG (2021) Solving the Rubik's cube with stepwise deep learning. Expert Syst 38(3):12665
- Bradde S, Bialek W (2017) Pca meets RG. J Stat Phys 167(3):462–475
- Koch-Janusz M, Ringel Z (2018) Mutual information, neural networks and the renormalization group. Nat Phys 14(6):578–582
- Kamath A, Vargas-Hernández RA, Krems RV, Carrington T Jr, Manzhos S (2018) Neural networks vs gaussian process regression for representing potential energy surfaces: a comparative study of fit quality and vibrational spectrum accuracy. J Chem Phys 148(24):241702
- Morningstar A, Melko RG (2017) Deep learning the ising model near criticality. arXiv preprint [arXiv:1708.04622](https://arxiv.org/abs/1708.04622)
- Carrasquilla J, Melko RG (2017) Machine learning phases of matter. Nat Phys 13(5):431–434
- Wang L (2016) Discovering phase transitions with unsupervised learning. Phys Rev B 94(19):195105
- Tanaka A, Tomiya A (2017) Detection of phase transition via convolutional neural networks. J Phys Soc Jpn 86(6):063001
- Kashiwa K, Kikuchi Y, Tomiya A (2019) Phase transition encoded in neural network. Prog Theor Exp Phys 2019(8):83–84
- Arai S, Ohzeki M, Tanaka K (2018) Deep neural network detects quantum phase transition. J Phys Soc Jpn 87(3):033001
- Bakk A, Høye JS (2003) One-dimensional ising model applied to protein folding. Physica A 323:504–518
- Tubiana J, Cocco S, Monasson R (2019) Learning protein constitutive motifs from sequence data. Elife 8:39397
- Wang L, You Z-H, Huang D-S, Zhou F (2018) Combining high speed elm learning with a deep convolutional neural network feature encoding for predicting protein-rna interactions. IEEE/ACM Trans Comput Biol Bioinf 17(3):972–980
- Kuhlman B, Bradley P (2019) Advances in protein structure prediction and design. Nat Rev Mol Cell Biol 20(11):681–697
- Ju F, Zhu J, Shao B, Kong L, Liu T-Y, Zheng W-M, Bu D (2021) Copulanet: learning residue co-evolution directly from multiple sequence alignment for protein structure prediction. Nat Commun 12(1):1–9
- Bukov M, Day AG, Sels D, Weinberg P, Polkovnikov A, Mehta P (2018) Reinforcement learning in different phases of quantum control. Phys Rev X 8(3):031086
- Greitemann J, Liu K, Pollet L et al (2019) Probing hidden spin order with interpretable machine learning. Phys Rev B 99(6):060404
- Liu K, Greitemann J, Pollet L et al (2019) Learning multiple order parameters with interpretable machines. Phys Rev B 99(10):104410
- Cubuk ED, Schoenholz SS, Rieser JM, Malone BD, Rottler J, Durian DJ, Kaxiras E, Liu AJ (2015) Identifying structural flow defects in disordered solids using machine-learning methods. Phys Rev Lett 114(10):108001
- Wetzel SJ (2017) Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders. Phys Rev E 96(2):022140
- Wang C, Zhai H (2017) Machine learning of frustrated classical spin models. I. Principal component analysis. Phys Rev B 96(14):144432
- Wang C, Zhai H (2018) Machine learning of frustrated classical spin models (II): Kernel principal component analysis. Front Phys 13(5):1–7
- Reddy G, Celani A, Sejnowski TJ, Vergassola M (2016) Learning to soar in turbulent environments. Proc Natl Acad Sci USA 113(33):4877–4884

- Reddy G, Wong-Ng J, Celani A, Sejnowski TJ, Vergassola M (2018) Glider soaring via reinforcement learning in the field. *Nature* 562(7726):236–239
- Jaeger H, Haas H (2004) Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*. <https://doi.org/10.1126/science.1091277>
- Pathak J, Hunt B, Girvan M, Lu Z, Ott E (2018) Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys Rev Lett* 120(2):024102
- Graafland CE, Gutiérrez JM, López JM, Pazó D, Rodríguez MA (2020) The probabilistic backbone of data-driven complex networks: an example in climate. *Sci Rep* 10(1):1–15
- Boers N, Bookhagen B, Barbosa HM, Marwan N, Kurths J, Marengo J (2014) Prediction of extreme floods in the eastern central andes based on a complex networks approach. *Nat Commun* 5(1):1–7
- Ying N, Wang W, Fan J, Zhou D, Han Z, Chen Q, Ye Q, Xue Z (2021) Climate network approach reveals the modes of co₂ concentration to surface air temperature. *Chaos: Interdiscip J Nonlinear Sci* 31(3):031104
- Chen X, Ying N, Chen D, Zhang Y, Lu B, Fan J, Chen X (2021) Eigen microstates and their evolution of global ozone at different geopotential heights. *Chaos: Interdiscip J Nonlinear Sci* 31(7):071102
- Zhang Y, Zhou D, Fan J, Marzocchi W, Ashkenazy Y, Havlin S (2021) Improved earthquake aftershocks forecasting model based on long-term memory. *New J Phys* 23(4):042001
- Zhu Y, Zhang R-H, Moum JN, Wang F, Li X, Li D (2022) Physics-informed deep learning parameterization of ocean vertical mixing improves climate simulations. *Nat Sci Rev*. <https://doi.org/10.1093/nsr/nwac044>
- Deutsch D, Jozsa R (1992) Rapid solution of problems by quantum computation. *Proc R Soc Lond A* 439(1907):553–558
- Shor PW (1994) Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th Annual Symposium on Foundations of Computer Science, pp. 124–134. IEEE
- Grover LK (1996) A fast quantum mechanical algorithm for database search. In: Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, pp. 212–219
- Sood SK et al (2024) Scientometric analysis of quantum-inspired metaheuristic algorithms. *Artif Intell Rev* 57(2):1–30
- Kou H, Zhang Y, Lee HP (2024) Dynamic optimization based on quantum computation-a comprehensive review. *Comput Struct* 292:107255
- Lloyd S, Mohseni M, Rebentrost P (2013) Quantum algorithms for supervised and unsupervised machine learning. arXiv preprint <arXiv:1307.0411>
- Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. *Nat Phys* 10(9):631–633
- Cong I, Duan L (2016) Quantum discriminant analysis for dimensionality reduction and classification. *New J Phys* 18(7):073011
- Wiebe N, Kapoor A, Svore K (2014) Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning. arXiv preprint <arXiv:1401.2142>
- Lu S, Braunstein SL (2014) Quantum decision tree classifier. *Quantum Inf Process* 13(3):757–770
- Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big data classification. *Phys Rev Lett* 113(13):130503
- Menneer T, Narayanan A (1995) Quantum-inspired neural networks. *Tech. Rep.* R329
- Tóth G, Lent CS, Tougaw PD, Brazhnik Y, Weng W, Porod W, Liu R-W, Huang Y-F (1996) Quantum cellular neural networks. *Superlattices Microstruct* 20(4):473–478
- Matsui N, Takai M, Nishimura H (2000) A network model based on qubitlike neuron corresponding to quantum circuit. *Electron Commun Jpn (Part III: Fundamental Electronic Science)* 83(10):67–73
- Kouda N, Matsui N, Nishimura H, Peper F (2005) Qubit neural network and its learning efficiency. *Neural Comput Appl* 14(2):114–121
- Zhou R, Qin L, Jiang N (2006) Quantum perceptron network. In: International Conference on Artificial Neural Networks, pp. 651–657. Springer
- Schuld M, Sinayskiy I, Petruccione F (2014) Quantum walks on graphs representing the firing patterns of a quantum neural network. *Phys Rev A* 89(3):032333
- Bausch J (2020) Recurrent quantum neural networks. *Adv Neural Inf Process Syst* 33:1368–1379
- Chen SY-C, Yoo S, Fang Y-LL (2020) Quantum long short-term memory. arXiv preprint <arXiv:2009.01783>
- Cong I, Choi S, Lukin MD (2019) Quantum convolutional neural networks. *Nat Phys* 15(12):1273–1278
- Kerenidis I, Landman J, Prakash A (2019) Quantum algorithms for deep convolutional neural networks. arXiv preprint <arXiv:1911.01117>
- Liu J, Lim KH, Wood KL, Huang W, Guo C, Huang H-L (2021) Hybrid quantum-classical convolutional neural networks. *Sci China Phys Mech Astron* 64(9):1–8
- Chen H, Zhang J-S, Zhang C (2005) Real-coded chaotic quantum-inspired genetic algorithm. *Control Decis* 20(11):1300

- Joshi D, Jain A, Mani A (2016) Solving economic load dispatch problem with valve loading effect using adaptive real coded quantum-inspired evolutionary algorithm. In: 2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity (CIPECH), pp. 123–128. IEEE
- Li B, Zhuang Z-q (2002) Genetic algorithm based-on the quantum probability representation. In: International Conference on Intelligent Data Engineering and Automated Learning, pp. 500–505. Springer
- Jin C, Jin S-W (2015) Automatic image annotation using feature selection based on improving quantum particle swarm optimization. *Signal Process* 109:172–181
- Jiao L, Li Y, Gong M, Zhang X (2008) Quantum-inspired immune clonal algorithm for global optimization. *IEEE Trans Syst Man Cybern Part B (Cybernetics)* 38(5):1234–1253
- Shang R, Jiao L, Ren Y, Wang J, Li Y (2014) Immune clonal coevolutionary algorithm for dynamic multi-objective optimization. *Nat Comput* 13(3):421–445
- Shang R, Du B, Dai K, Jiao L, Esfahani AMG, Stolk R (2018) Quantum-inspired immune clonal algorithm for solving large-scale capacitated arc routing problems. *Memetic Computing* 10(1):81–102
- Qi F, Xu L (2015) A 15-based synchronous cellular quantum evolutionary algorithm. In: 2015 7th International Conference on Information Technology in Medicine and Education (ITME), pp. 321–324. IEEE
- Mei J, Zhao J (2018) An enhanced quantum-behaved particle swarm optimization for security constrained economic dispatch. In: 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), pp. 221–224. IEEE
- Bonet-Monroig X, Wang H, Vermetten D, Senjean B, Moussa C, Bäck T, Dunjko V, O'Brien TE (2023) Performance comparison of optimization methods on variational quantum algorithms. *Phys Rev A* 107(3):032407
- Finžgar JR, Kerschbaumer A, Schuetz MJ, Mendl CB, Katzgraber HG (2024) Quantum-informed recursive optimization algorithms. *PRX Quantum* 5(2):020327
- Kak SC (1995) Quantum neural computing. *Adv Imaging Electron Phys* 94:259–313
- Nielsen MA, Chuang I (2002) Quantum computation and quantum information. Am Assoc Phys Teachers Wiebe N, Kapoor A, Svore KM (2014) Quantum deep learning. arXiv preprint [arXiv:1412.3489](https://arxiv.org/abs/1412.3489)
- Schuld M, Sinayskiy I, Petruccione F (2014) The quest for a quantum neural network. *Quantum Inf Process* 13(11):2567–2586
- Behrman EC, Niemel J, Steck JE, Skinner SR (1996) A quantum dot neural network. In: Proceedings of the 4th Workshop on Physics of Computation, pp. 22–24
- Ceschini A, Rosato A, Panella M (2021) Design of an lstm cell on a quantum hardware. Express Briefs, *IEEE Transactions on Circuits and Systems II*
- Narayanan A, Moore M (1996) Quantum-inspired genetic algorithms. In: Proceedings of IEEE International Conference on Evolutionary Computation, pp. 61–66. IEEE
- Han K-H, Park K-H, Lee C-H, Kim J-H (2001) Parallel quantum-inspired genetic algorithm for combinatorial optimization problem. In: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546), vol. 2, pp. 1422–1429. IEEE
- Han K-H, Kim J-H (2000) Genetic quantum algorithm and its application to combinatorial optimization problem. In: Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512), vol. 2, pp. 1354–1360. IEEE
- Li P, Li S (2008) Quantum-inspired evolutionary algorithm for continuous space optimization based on bloch coordinates of qubits. *Neurocomputing* 72(1–3):581–591
- Cruz A, Velasco MMBR, Pacheco MAC (2007) Quantum-inspired evolutionary algorithm for numerical optimization. In: Hybrid Evolutionary Algorithms, pp. 19–37. Springer, USA
- Yang S, Wang M et al (2004) A quantum particle swarm optimization. In: Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753), vol. 1, pp. 320–324. IEEE
- Rehman OU, Yang S, Khan S, Rehman SU (2019) A quantum particle swarm optimizer with enhanced strategy for global optimization of electromagnetic devices. *IEEE Trans Magn* 55(8):1–4
- Yangyang L, Licheng J (2008) Quantum immune cloning multi-objective optimization algorithm. *J Electron Inf* 30(6):1367–1371
- Liu S, You X, Wu Z (2013) A cultural immune quantum evolutionary algorithm and its application. *J Comput* 8(1):163–169
- Alba E, Dorronsoro B (2005) The exploration/exploitation tradeoff in dynamic cellular genetic algorithms. *IEEE Trans Evol Comput* 9(2):126–142
- Li Z, Xu K, Liu S, Li K (2008) Quantum multi-objective evolutionary algorithm with particle swarm optimization method. In: 2008 Fourth International Conference on Natural Computation, vol. 3, pp. 672–676. IEEE

- Mohammad T, Reza ATM (2009) Improvement of quantum evolutionary algorithm with a functional sized population. In: Applications of Soft Computing, pp. 389–398. Springer, USA
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969
- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473)
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66(5):688
- Pearl J (2009) Causality. Cambridge University Press, Cambridge
- Imbens GW, Rubin DB (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, Cambridge
- Runge J (2018) Causal network reconstruction from time series: from theoretical assumptions to practical estimation. *Chaos: Interdiscip J Nonlinear Sci* 28(7):075310
- Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D (2019) Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci Adv* 5(11):4996
- Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J et al (2019) Inferring causation from time series in earth system sciences. *Nat Commun* 10(1):1–13
- Nauta M, Bucur D, Seifert C (2019) Causal discovery with attention-based convolutional neural networks. *Mach Learn Knowl Extr* 1(1):312–340
- Abraham WC, Robins A (2005) Memory retention—the synaptic stability versus plasticity dilemma. *Trends Neurosci* 28(2):73–78
- McCloskey M, Cohen NJ (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In: Psychology of Learning and Motivation vol. 24, pp. 109–165. Elsevier, Amsterdam
- Goodfellow IJ, Mirza M, Xiao D, Courville A, Bengio Y (2013) An empirical investigation of catastrophic forgetting in gradient-based neural networks. arXiv preprint [arXiv:1312.6211](https://arxiv.org/abs/1312.6211)
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A et al (2017) Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci USA* 114(13):3521–3526
- Zhang T, Cheng X, Jia S, Li CT, Poo M-M, Xu B (2023) A brain-inspired algorithm that mitigates catastrophic forgetting of artificial and spiking neural networks with low computational cost. *Sci Adv* 9(34):2947
- Belbute-Peres FDA, Economou T, Kolter Z (2020) Combining differentiable pde solvers and graph neural networks for fluid flow prediction. In: International Conference on Machine Learning, pp. 2402–2411. PMLR
- Sanchez-Gonzalez A, Godwin J, Pfaff T, Ying R, Leskovec J, Battaglia P (2020) Learning to simulate complex physics with graph networks. In: International Conference on Machine Learning, pp. 8459–8468. PMLR
- Tremblay J, Prakash A, Acuna D, Brophy M, Jampani V, Anil C, To T, Cameracci E, Boochoon S, Birchfield S (2018) Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 969–977
- Bousmalis K, Irpan A, Wohlhart P, Bai Y, Kelcey M, Kalakrishnan M, Downs L, Ibarz J, Pastor P, Konolige K et al (2018) Using simulation and domain adaptation to improve efficiency of deep robotic grasping. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 4243–4250. IEEE
- Thanasutives P, Numao M, Fukui K-i (2021) Adversarial multi-task learning enhanced physics-informed neural networks for solving partial differential equations. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–9. IEEE
- Chen Y, Zhang N, Yang J (2023) A survey of recent advances on stability analysis, state estimation and synchronization control for neural networks. *Neurocomputing* 515:26–36
- Skomski E, Drgoňá J, Tuor A (2021) Automating discovery of physics-informed neural state space models via learning and evolution. In: Learning for Dynamics and Control, pp. 980–991. PMLR
- Xu K, Li J, Zhang M, Du SS, Kawarabayashi K-i, Jegelka S (2019) What can neural networks reason about? arXiv preprint [arXiv:1905.13211](https://arxiv.org/abs/1905.13211)
- Chen Y, Friesen AL, Behbahani F, Doucet A, Budden D, Hoffman M, Freitas N (2020) Modular meta-learning with shrinkage. *Adv Neural Inf Process Syst* 33:2858–2869

- Goyal A, Lamb A, Hoffmann J, Sodhani S, Levine S, Bengio Y, Schölkopf B (2019) Recurrent independent mechanisms. arXiv preprint [arXiv:1909.10893](https://arxiv.org/abs/1909.10893)
- Wang Q, Yang K (2024) Privacy-preserving data fusion for traffic state estimation: a vertical federated learning approach. arXiv preprint [arXiv:2401.11836](https://arxiv.org/abs/2401.11836)
- Pfeiffer J, Gutschow J, Haas C, Mösllein F, Maspfuhl O, Borgers F, Alpsancar S (2023) Algorithmic fairness in AI: an interdisciplinary view. *Bus Inf Syst Eng* 65(2):209–222
- Xivuri K, Twinomurinzi H (2021) A systematic review of fairness in artificial intelligence algorithms. In: Responsible AI and Analytics for an Ethical and Inclusive Digitized Society: 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, September 1–3, 2021, Proceedings 20, pp. 271–284. Springer
- Chen RJ, Wang JJ, Williamson DF, Chen TY, Lipkova J, Lu MY, Sahai S, Mahmood F (2023) Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng* 7(6):719–742
- Müller M, Dosovitskiy A, Ghanem B, Koltun V (2018) Driving policy transfer via modularity and abstraction. arXiv preprint [arXiv:1804.09364](https://arxiv.org/abs/1804.09364)
- Liu T, Zhou B (2024) The impact of artificial intelligence on the green and low-carbon transformation of Chinese enterprises. *Managerial and Decision Economics*
- Yang S, Wang J, Dong K, Dong X, Wang K, Fu X (2024) Is artificial intelligence technology innovation a recipe for low-carbon energy transition? a global perspective. *Energy*, 131539
- Huang C, Zhang Z, Mao B, Yao X (2022) An overview of artificial intelligence ethics. *IEEE Trans Artif Intell* 4(4):799–819
- Akinrinola O, Okoye CC, Ofodile OC, Ugochukwu CE (2024) Navigating and reviewing ethical dilemmas in ai development: strategies for transparency, fairness, and accountability. *GSC Adv Res Rev* 18(3):050–058

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.