Understanding and Designing Deep Neural Networks Through Theory-Guided Training

Karthika Nasir[1], Aradhana Reva[1], Jai Sekhar[1]

[1]Bharti Centre for Communication, Indian Institute of Technology Bombay, India
karthika.nasir@ee.iitb.ac.in , aradhana.reva@ee.iitb.ac.in,
jai.sekhar@ee.iitb.ac.in

Abstract. Deep Neural Networks (DNNs) have revolutionized numerous fields with their remarkable empirical success, yet their theoretical understanding remains incomplete. The emerging paradigm of *theory-trained deep neural networks* aims to bridge this gap by integrating rigorous theoretical principles into the design and training of deep models. This survey provides a comprehensive overview of the foundational theories, training methodologies, and practical applications that define this vibrant area of research. We categorize existing approaches based on the theoretical frameworks they leverage, including optimization landscapes, neural tangent kernels, implicit bias, and information-theoretic principles. Furthermore, we discuss empirical successes across diverse domains such as computer vision, natural language processing, scientific computing, and reinforcement learning. Finally, we outline key challenges and promising future directions, emphasizing the need for scalable, interpretable, and robust theory-guided learning algorithms. This survey serves as a resource for researchers and practitioners interested in the intersection of deep learning theory and practice.

Keywords: Deep Neural Networks, Theory-Guided Training, Optimization Theory, Neural Tangent Kernel, Implicit Bias, Information Theory, Generalization, Robustness, Interpretability

1 Introduction

Deep Neural Networks (DNNs) have achieved remarkable success across a wide spectrum of tasks, ranging from image classification and natural language processing to game playing and scientific discovery [1]. Despite their empirical performance, the theoretical understanding of why DNNs generalize well and how they should be trained remains incomplete [2]. In recent years, a new research direction has emerged that aims to bridge this gap: *Theory-Trained Neural Networks* [3]. This line of work leverages theoretical insights to guide the training process, architecture design, and optimization strategies of deep networks. The motivation for theory-trained approaches stems from the recognition that conventional training methods, often based on heuristics or trial-and-error, may not exploit the full potential of the underlying mathematical structure of deep models [4]. By contrast, theory-trained methods incorporate principles from optimization theory, statistical learning, dynamical systems, information theory, and approximation theory to derive algorithms with provable guarantees, interpretability, or improved generalization properties [5]. This survey aims to provide a structured overview of the emerging field of theory-trained DNNs [6]. We categorize existing approaches according to the theoretical principles they are based on, the ways in which these principles inform the training process, and the empirical outcomes they achieve [7]. Key topics include optimization landscape analysis, neural tangent kernels, implicit bias in gradient descent, overparameterization theory, spectral initialization, and structured inductive biases [8]. We also highlight connections to classical learning theory, recent advances in PAC-Bayes analysis, and information-theoretic frameworks [9]. The rest of the paper is organized as follows [10]. Section 2 reviews foundational concepts that underpin theory-guided training. Section 3 presents various theory-trained methods and their categorizations. Section 4 surveys empirical domains where these methods have shown promise [11]. Section 5 discusses open problems and future research directions. Finally, Section 6 summarizes key insights and outlines the trajectory of this growing field.

2 Foundations of Theory-Guided Deep Learning

In order to understand and evaluate theory-trained deep neural networks, it is essential to first establish the theoretical foundations upon which these approaches are built. This section provides a concise overview of key mathematical and conceptual tools that underpin the design and analysis of theory-guided training methods [12]. These foundations span several domains including statistical learning theory, optimization, dynamical systems, and information theory [13].

2.1 Statistical Learning Theory

Statistical learning theory provides a framework for understanding the generalization performance of machine learning models [14]. Central to this theory are notions such as empirical risk minimization, VC-dimension, Rademacher complexity, and uniform convergence bounds. These tools help quantify how well a model trained on finite data is expected to perform on unseen data [15]. In the context of deep learning, classical generalization bounds often prove too loose, motivating refined frameworks such as margin-based analyses, compression bounds, and PAC-Bayes theory.

2.2 Optimization Landscapes and Gradient Dynamics

The training of deep networks is typically carried out via variants of gradient descent [16]. Understanding the optimization landscape—the geometry of the loss function over the parameter space—has been crucial in explaining phenomena such as convergence to global or near-global minima. Recent advances reveal that overparameterization can induce benign landscapes with no spurious local minima, and that gradient descent exhibits implicit regularization behaviors, especially under constraints like small initialization or low-rank structure.

2.3 Neural Tangent Kernel and Linearization

The Neural Tangent Kernel (NTK) regime provides a powerful approximation of deep network training dynamics in the infinite-width limit [17]. In this setting, the network behaves approximately as a linear model in function space, and its training can be analyzed using kernel methods [18]. This linearization allows for closed-form expressions of training trajectories and generalization behavior, offering insight into how architectural choices and depth affect learning.

2.4 Implicit Bias and Inductive Priors

Empirical evidence suggests that even in the absence of explicit regularization, optimization algorithms like stochastic gradient descent (SGD) tend to find solutions that generalize well [19]. This phenomenon is studied through the lens of implicit bias: the tendency of learning algorithms to prefer certain types of solutions over others. Understanding the nature of these biases is key to developing theory-trained networks that exploit them effectively [20].

2.5 Information-Theoretic and Dynamical Systems Perspectives

Information theory offers alternative viewpoints for understanding representation learning and compression in deep networks [21]. Concepts such as mutual information, the Information Bottleneck principle, and entropy-based generalization bounds have been proposed to explain deep learning behavior. Additionally, the dynamical systems perspective models training as a continuous-time flow, revealing stability properties and convergence characteristics not evident in purely discrete analyses [22]. These foundational elements form the theoretical backbone of many modern approaches to theory-trained neural networks. In the next section, we categorize and analyze methods that directly leverage these principles to inform training procedures, architecture design, and regularization strategies.

3 Theory-Trained Methods

This section surveys prominent methods and frameworks that incorporate theoretical insights directly into the training of deep neural networks [23]. We organize these approaches according to the theoretical principles they exploit, the way theory guides their design, and their practical impact on training efficiency, generalization, and interpretability [24].

3.1 Optimization-Informed Training

Optimization-informed methods explicitly utilize knowledge of the loss landscape and gradient dynamics to improve training [25]. Techniques in this category include:

- **Landscape Smoothing and Regularization:** By designing loss functions or adding regularizers that smooth nonconvex landscapes, these methods facilitate convergence to favorable minima with better generalization [26].
- **Adaptive Gradient Methods with Theoretical Guarantees:** Algorithms such as AMSGrad and others are designed with provable convergence guarantees under certain assumptions, improving stability during training.
- **Overparameterization and Lazy Training Regimes:** Exploiting the benign geometry of overparameterized models, training schemes initialize networks in a regime where optimization behaves nearly convexly, simplifying convergence analysis [27].

3.2 Neural Tangent Kernel-Based Approaches

Building on the NTK framework, several methods train networks in a regime where the network function evolves approximately linearly:

- **Kernel Regression via Infinite-Width Networks:** Approximating training by kernel regression with NTK enables the derivation of explicit generalization bounds.
- **NTK-Guided Architecture Design:** Theoretical analysis of NTK spectra informs architecture choices that improve expressivity and trainability [28].
- **Hybrid Methods:** Combining kernel methods with finite-width network training to leverage the benefits of both linear and nonlinear regimes.

3.3 Implicit Bias Exploitation

These methods leverage the implicit regularization effect of gradient-based optimizers:

- **Norm-Based Implicit Regularization:** Training protocols are designed to encourage solutions with minimum norm, which often correlate with better generalization [29].
- **Margin Maximization Strategies:** By controlling the margin through optimization dynamics, these approaches improve robustness and generalization.
- **Algorithmic Modifications:** Variants of SGD and other optimizers are proposed to accentuate implicit bias towards desirable solutions [30].

3.4 Information-Theoretic and Compression-Based Training

This category encompasses methods grounded in information theory and model compression principles:

- **Information Bottleneck Regularization:** Incorporating objectives that trade off accuracy with compression of learned representations [31].
- **PAC-Bayes Inspired Training:** Training schemes that optimize bounds on generalization via PAC-Bayes theory, often incorporating stochasticity or noise injection.
- **Structured Sparsity and Pruning:** Guided pruning strategies informed by theoretical insights reduce model complexity while preserving performance [32].

3.5 Theory-Guided Architecture Design

Beyond training algorithms, theory-trained approaches inform architecture choices:

- **Spectral Initialization and Normalization:** Initializing and normalizing weights based on spectral theory improves stability and trainability.
- **Architectures with Provable Properties:** Designing networks with guaranteed expressivity, robustness, or optimization behavior.
- **Incorporation of Inductive Biases:** Embedding domain knowledge or structural priors into network design to reduce sample complexity [33].

Collectively, these theory-trained methods demonstrate how integrating theoretical insights into the core of deep learning training can enhance performance and deepen our understanding of neural networks [34]. The following section will explore practical applications and empirical evaluations of these methods across diverse domains [35].

4 Applications of Theory-Trained Deep Neural Networks

Theory-trained deep neural networks have found impactful applications across a variety of domains where leveraging theoretical insights improves performance, reliability, and interpretability [36]. This section highlights key application areas and summarizes empirical findings demonstrating the benefits of theory-guided training.

4.1 Computer Vision

In computer vision, theory-trained models have been used to improve robustness and generalization, particularly in settings with limited labeled data or adversarial perturbations [37]. For instance, optimization-informed regularization methods enhance convergence to flatter minima, leading to better resistance against adversarial attacks. Neural tangent kernel-inspired training has also facilitated more efficient transfer learning and fine-tuning by clarifying feature representation dynamics [38].

4.2 Natural Language Processing

Theory-trained approaches have contributed to the understanding and optimization of large language models [39]. Implicit bias exploitation helps guide training toward solutions that generalize well despite overparameterization. Information-theoretic methods have been applied to compress large transformer architectures, reducing inference costs without significant loss in accuracy [40]. Furthermore, theory-driven architecture design has influenced the development of more efficient attention mechanisms [41].

4.3 Scientific Computing and Physics-Informed Learning

Incorporating domain knowledge through inductive biases and theory-guided architectures enables neural networks to solve differential equations and model physical systems more accurately. Theory-trained methods guarantee stability and convergence in physics-informed neural networks (PINNs), ensuring consistency with underlying physical laws [42]. Optimization techniques grounded in theory improve the training of networks for simulating complex dynamical systems.

4.4 Reinforcement Learning

In reinforcement learning, theory-trained neural networks enhance policy optimization by exploiting implicit regularization and stability guarantees. Techniques such as spectral normalization and theory-informed initialization improve the sample efficiency and convergence properties of deep RL agents [43]. Additionally, NTK-based analyses provide theoretical explanations for the success of value function approximators in high-dimensional environments [44].

4.5 Healthcare and Bioinformatics

The interpretability and reliability afforded by theory-trained methods are critical in healthcare applications [45]. Approaches that incorporate explicit regularization and theoretical generalization guarantees have improved diagnostic accuracy in medical imaging and genomic data analysis [46]. Information-theoretic frameworks help manage noisy, high-dimensional data by encouraging robust feature extraction and compression [47].

4.6 Summary of Empirical Outcomes

Across these domains, theory-trained deep neural networks often demonstrate:

- Improved convergence speed and stability during training [48].
- Enhanced generalization on unseen data, especially in low-data regimes [49].
- Greater robustness to adversarial or distributional shifts.
- Reduced model complexity without sacrificing accuracy [50].

– Better interpretability and alignment with domain knowledge [51].

These empirical successes highlight the practical value of integrating theoretical principles into deep learning [52]. Nonetheless, challenges remain in scaling these methods and extending them to broader, more complex real-world problems, as discussed in the next section [53].

5 Challenges and Future Directions

Despite significant progress in theory-trained deep neural networks, several challenges remain that limit the widespread adoption and further advancement of these methods. Addressing these issues will be critical for bridging the gap between theory and practice and for unlocking the full potential of theory-guided training.

5.1 Scalability and Computational Efficiency

Many theoretical analyses rely on simplifying assumptions such as infinite width or convexity approximations, which do not fully capture practical large-scale networks. Extending theory-trained methods to scale efficiently on real-world, high-dimensional data and architectures remains an open problem[54]. Developing algorithms that retain theoretical guarantees without sacrificing computational tractability is a key research direction.

5.2 Bridging the Gap Between Theory and Practice

While theory-trained approaches offer rigorous insights, there is often a disconnect between theoretical models and the complexity of real neural networks used in practice [55]. Improving the fidelity of theoretical models, accounting for finite width, non-i.i.d [56]. data, and dynamic architectures will enhance their practical relevance. Furthermore, designing theory-informed heuristics that generalize across diverse tasks and architectures is essential [57].

5.3 Robustness and Generalization in Complex Settings

Achieving reliable generalization in the presence of distributional shifts, adversarial attacks, and noisy or scarce data remains challenging [58]. Theory-trained methods need to incorporate more sophisticated models of uncertainty and robustness [59]. Combining insights from robust optimization, causal inference, and information theory may provide promising avenues for enhanced resilience [60].

5.4 Interpretable and Explainable Models

As deep learning systems are deployed in safety-critical and high-stakes environments, interpretability becomes increasingly important [61]. Theoretical frameworks that facilitate transparent decision-making and provide guarantees on model behavior are still in early stages [3]. Future work should focus on integrating interpretability directly into theory-trained network design and training [62].

5.5 Integration with Emerging Paradigms

The rapid evolution of deep learning introduces new paradigms such as self-supervised learning, continual learning, and neural architecture search [63]. Adapting theory-trained methods to these emerging settings poses both opportunities and challenges. Leveraging theoretical insights to guide automated architecture design, optimize data efficiency, and manage catastrophic forgetting are promising directions [64].

5.6 Towards Unified Theoretical Frameworks

Current theoretical approaches often focus on isolated aspects of neural network behavior [65]. A comprehensive, unified framework that integrates optimization theory, statistical learning, information theory, and dynamical systems would provide deeper understanding and more powerful training methodologies. Developing such frameworks will require interdisciplinary collaboration and novel mathematical tools [66]. Addressing these challenges will pave the way for theory-trained deep neural networks to become a foundational component of future artificial intelligence systems, combining rigorous guarantees with empirical effectiveness [67].

6 Conclusion

Theory-trained deep neural networks represent a promising and rapidly evolving direction in deep learning research. By integrating rigorous theoretical principles from optimization, statistical learning, information theory, and dynamical systems into the training process, these methods provide valuable insights into the behavior of deep models and enhance their practical performance.

Throughout this survey, we have reviewed foundational concepts that underpin theory-guided training, categorized prominent methods that leverage theoretical insights, and examined diverse applications where these approaches have demonstrated tangible benefits. We also discussed the key challenges that remain, including scalability, bridging the gap between theory and practice, robustness, interpretability, and integration with emerging paradigms.

Looking forward, advancing theory-trained deep neural networks will require continued interdisciplinary efforts to develop scalable algorithms with strong theoretical guarantees, unified frameworks that connect disparate theoretical perspectives, and practical tools that empower practitioners across domains. By addressing these goals, theory-trained networks hold the potential to deepen our understanding of deep learning and drive innovations that are both principled and impactful.

We hope this survey serves as a useful resource and catalyst for further research at the intersection of theory and practice in deep learning.

References

1. Shengze Cai, He Li, Fuyin Zheng, Fang Kong, Ming Dao, George Em Karniadakis, and Subra Suresh. Artificial intelligence velocimetry and microaneurysm-on-a-chip for three-dimensional analysis of blood flow in physiology and disease. *Proceedings of the National Academy of Sciences*, 118(13):e2100697118, 2021.
2. Sifan Wang, Hanwen Wang, and Paris Perdikaris. On the eigenvector bias of fourier feature networks: From regression to solving multi-scale pdes with physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 384:113938, 2021.
3. Maryam Toloubidokhti, Yubo Ye, Ryan Missel, Xiajun Jiang, Nilesh Kumar, Ruby Shrestha, and Linwei Wang. Dats: Difficulty-aware task sampler for meta-learning physics-informed neural networks. In *The Twelfth International Conference on Learning Representations*, 2023.
4. Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
5. Shijun Cheng and Tariq Alkhalifah. Meta-pinn: Meta learning for improved neural network wavefield solutions. *arXiv preprint arXiv:2401.11502*, 2024.
6. Hesheng Tang, Yangyang Liao, Hu Yang, and Liyu Xie. A transfer learning-physics informed neural network (tl-pinn) for vortex-induced vibration. *Ocean Engineering*, 266:113101, 2022.
7. Chen Xu, Ba Trung Cao, Yong Yuan, and Günther Meschke. Transfer learning based physics-informed neural networks for solving inverse problems in engineering structures under different loading scenarios. *Computer Methods in Applied Mechanics and Engineering*, 405:115852, 2023.
8. Hua-Ming Huang, Elena Raponi, Fabian Duddeck, Stefan Menzel, and Mariusz Bujny. Topology optimization of periodic structures for crash and static load cases using the evolutionary level set method. *Optimization and Engineering*, 25(3):1597–1630, 2024.
9. Kalyanmoy Deb and Matthias Ehrgott. On generalized dominance structures for multi-objective optimization. *Mathematical and Computational Applications*, 28(5):100, 2023.
10. Jaehyeok Doh, Nandhini Raju, Nagarajan Raghavan, David W Rosen, and Samyeon Kim. Bayesian inference-based decision of fatigue life model for metal additive manufacturing considering effects of build orientation and post-processing. *International Journal of Fatigue*, 155:106535, 2022.
11. Rongye Shi, Zhaobin Mo, and Xuan Di. Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 540–547, 2021.

12. Carlos A Coello Coello. *Evolutionary algorithms for solving multi-objective problems*. Springer, 2007.
13. Zhenkun Wang, Qingfu Zhang, Yew-Soon Ong, Shunyu Yao, Haitao Liu, and Jianping Luo. Choose appropriate subproblems for collaborative modeling in expensive multiobjective optimization. *IEEE Transactions on Cybernetics*, 53(1):483–496, 2021.
14. David Gunning and David Aha. Darpa's explainable artificial intelligence (xai) program. *AI magazine*, 40(2):44–58, 2019.
15. Lulu Cao, Haokai Hong, and Min Jiang. Fast solving partial differential equations via imitative fourier neural operator. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024.
16. Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.
17. Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019.
18. Jian Cheng Wong, Abhishek Gupta, and Yew-Soon Ong. Can transfer neuroevolution tractably solve your differential equations? *IEEE Computational Intelligence Magazine*, 16(2):14–30, 2021.
19. S Chandra Mouli, Muhammad Alam, and Bruno Ribeiro. Metaphysica: Improving ood robustness in physics-informed machine learning. In *The Twelfth International Conference on Learning Representations*, 2023.
20. John Eason and Selen Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers & Chemical Engineering*, 68:220–232, 2014.
21. Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.
22. Raphaël Pellegrin, Blake Bullwinkel, Marios Mattheakis, and Pavlos Protopapas. Transfer learning with physics-informed neural networks for efficient simulation of branched flows. *arXiv preprint arXiv:2211.00214*, 2022.
23. Ehsan Haghighat, Sahar Abouali, and Reza Vaziri. Constitutive model characterization and discovery using physics-informed deep learning. *Engineering Applications of Artificial Intelligence*, 120:105828, 2023.
24. Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
25. Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.
26. Lulu Cao, Zimo Zheng, Chenwen Ding, Jinkai Cai, and Min Jiang. Genetic programming symbolic regression with simplification-pruning operator for solving differential equations. In *International Conference on Neural Information Processing*, pages 287–298. Springer, 2023.
27. Jian Cheng Wong, Pao-Hsiung Chiu, Chinchun Ooi, My Ha Dao, and Yew-Soon Ong. Lsa-pinn: Linear boundary connectivity loss for solving pdes on complex geometry. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10. IEEE, 2023.

28. Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry P Vetrov, and Andrew G Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns. *Advances in neural information processing systems*, 31, 2018.
29. Mohammadamin Mahmoudabadbozchelou and Safa Jamali. Rheology-informed neural networks (rhinns) for forward and inverse metamodelling of complex fluids. *Scientific reports*, 11(1):1–13, 2021.
30. Linlin Zhong, Bingyu Wu, and Yifan Wang. Accelerating physics-informed neural network based 1d arc simulation by meta learning. *Journal of Physics D: Applied Physics*, 56(7):074006, 2023.
31. Franz M Rohrhofer, Stefan Posch, Clemens Gößnitzer, and Bernhard C Geiger. On the apparent pareto front of physics-informed neural networks. *IEEE Access*, 2023.
32. Guangyong Sun, Fengxiang Xu, Guangyao Li, and Qing Li. Crashing analysis and multiobjective optimization for thin-walled structures with functionally graded thickness. *International Journal of Impact Engineering*, 64:62–74, 2014.
33. Alex Nichol and John Schulman. Reptile: a scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999*, 2(3):4, 2018.
34. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.
35. Guo Yuan, Fu Zhuojia, Min Jian, Liu Xiaoting, and Zhao Haitao. Curriculum-transfer-learning based physics-informed neural networks for long-time simulation of nonlinear wave. , 56:1–11, 2023.
36. Wei Ma, Zhaocheng Liu, Zhaxylyk A Kudyshev, Alexandra Boltasseva, Wenshan Cai, and Yongmin Liu. Deep learning for the design of photonic structures. *Nature Photonics*, 15(2):77–90, 2021.
37. Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.
38. Xu Liu, Wei Peng, Zhiqiang Gong, Weien Zhou, and Wen Yao. Temperature field inversion of heat-source systems via physics-informed neural networks. *Engineering Applications of Artificial Intelligence*, 113:104902, 2022.
39. Ariel Norambuena, Marios Mattheakis, Francisco J González, and Raúl Coto. Physics-informed neural networks for quantum control. *arXiv preprint arXiv:2206.06287*, 2022.
40. Jürgen Schmidhuber. *Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook*. PhD thesis, Technische Universität München, 1987.
41. Yaochu Jin, Handing Wang, Tinkle Chugh, Dan Guo, and Kaisa Miettinen. Data-driven evolutionary optimization: An overview and case studies. *IEEE Transactions on Evolutionary Computation*, 23(3):442–458, 2018.
42. Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738, 2021.
43. Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11:341–359, 1997.
44. Abhishek Gupta and BK Mishra. Neuroevolving monotonic pinns for particle breakage analysis. In *2024 IEEE Conference on Artificial Intelligence (CAI)*, pages 993–996. IEEE, 2024.

45. Craig Michoski, Miloš Milosavljević, Todd Oliver, and David R Hatch. Solving differential equations using deep neural networks. *Neurocomputing*, 399:193–212, 2020.
46. Levi McClenny and Ulisses Braga-Neto. Self-adaptive physics-informed neural networks using a soft attention mechanism. *arXiv preprint arXiv:2009.04544*, 2020.
47. Neil Savage. Tapping into the drug discovery potential of ai. *Nature. com*, 2020.
48. Abdul Hannan Mustajab, Hao Lyu, Zarghaam Rizvi, and Frank Wuttke. Physics-informed neural networks for high-frequency and multi-scale problems using transfer learning. *Applied Sciences*, 14(8):3204, 2024.
49. Yujin Tang, Yingtao Tian, and David Ha. Evojax: Hardware-accelerated neuroevolution. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 308–311, 2022.
50. Ehsan Haghighat and Ruben Juanes. Sciann: A keras/tensorflow wrapper for scientific computations and physics-informed deep learning using artificial neural networks. *Computer Methods in Applied Mechanics and Engineering*, 373:113552, 2021.
51. Sebastian Flennerhag, Pablo G Moreno, Neil D Lawrence, and Andreas Damianou. Transferring knowledge across learning processes. *arXiv preprint arXiv:1812.01054*, 2018.
52. Felipe AC Viana and Arun K Subramaniyan. A survey of bayesian calibration and physics-informed neural networks in scientific modeling. *Archives of Computational Methods in Engineering*, 28:3801–3830, 2021.
53. Ana Kaplarević-Mališić, Branka Andrijević, Filip Bojović, Srđan Nikolić, Lazar Krstić, Boban Stojanović, and Miloš Ivanović. Identifying optimal architectures of physics-informed neural networks by evolutionary strategy. *Applied Soft Computing*, 146:110646, 2023.
54. Yassine Zniyed, Thanh Phuong Nguyen, et al. Enhanced network compression through tensor decompositions and pruning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
55. Qiming Zhu, Zeliang Liu, and Jinhui Yan. Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks. *Computational Mechanics*, 67:619–635, 2021.
56. Nvidia modulus: An open-source framework for physics-based deep learning in science and engineering.
57. Bing-Chuan Wang, Zhen-Dong Ji, Yong Wang, Han-Xiong Li, and Zhongmei Li. A physics-informed composite network for modeling of electrochemical process of large-scale lithium-ion batteries. *IEEE Transactions on Industrial Informatics*, 2024.
58. Lorien Y Pratt, Jack Mostow, and Candace A Kamm. Direct transfer of learned information among neural networks. In *Proceedings of the ninth National conference on Artificial intelligence-Volume 2*, pages 584–589, 1991.
59. Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
60. Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.
61. Yifan Wang and Linlin Zhong. Nas-pinn: neural architecture search-guided physics-informed neural network for solving pdes. *Journal of Computational Physics*, 496:112603, 2024.

62. Roger Temam. *Navier–Stokes equations: theory and numerical analysis*, volume 343. American Mathematical Society, 2024.
63. An-Min Zou and Zhun Fan. Fixed-time attitude tracking control for rigid spacecraft without angular velocity measurements. *IEEE Transactions on Industrial Electronics*, 67(8):6795–6805, 2019.
64. Yew-Soon Ong, Meng-Hiot Lim, Ning Zhu, and Kok-Wai Wong. Classification of adaptive memetic algorithms: a comparative study. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(1):141–152, 2006.
65. Apostolos F Psaros, Kenji Kawaguchi, and George Em Karniadakis. Meta-learning pinn loss functions. *Journal of computational physics*, 458:111121, 2022.
66. Ye Tian, Xingyi Zhang, Chao Wang, and Yaochu Jin. An evolutionary algorithm for large-scale sparse multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 24(2):380–393, 2019.
67. Sifan Wang, Shyam Sankaran, and Paris Perdikaris. Respecting causality for training physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116813, 2024.