Preprints.org

Review

Theory-Trained Deep Neural Networks: Insights, Techniques, and Future Directions

Aradhana Reva , Jai Sekhar , Meera Sharma , Karthika Nasir [*]

Review

Theory-Trained Deep Neural Networks: Insights, Techniques, and Future Directions

Aradhana Reva, Jai Sekhar, Meera Sharma and Karthika Nasir *

Bharti Centre for Communication, Indian Institute of Technology Bombay, India

* Correspondence: karthika.nasir@ee.iitb.ac.in

Abstract

Deep neural networks have achieved remarkable success across numerous domains, yet a comprehensive theoretical understanding of their training dynamics and generalization remains a fundamental challenge. Theory-Trained Deep Neural Networks (TT-DNNs) represent an emerging paradigm that integrates rigorous theoretical insights into the design and training of neural architectures. This survey provides a systematic overview of TT-DNNs, categorizing key approaches based on optimization theory, statistical learning theory, approximation theory, and information theory. We discuss theory-informed training paradigms that improve convergence, robustness, and interpretability, and highlight notable applications across computer vision, natural language processing, scientific computing, and healthcare. Furthermore, we identify open challenges and future directions to bridge the gap between theory and practice. Our aim is to offer a comprehensive resource that fosters deeper understanding and innovation at the intersection of theory and deep learning practice.

Keywords: deep neural networks; theory-trained models; optimization theory; statistical learning theory; approximation theory; information theory; deep learning; generalization; robustness; interpretability

---

1. Introduction

Deep neural networks (DNNs) have revolutionized a wide array of fields, from computer vision and natural language processing to reinforcement learning and scientific computing. Despite their empirical success, a theoretical understanding of how and why these models work remains elusive [1]. In recent years, the growing interest in bridging the gap between theory and practice has given rise to the concept of *Theory-Trained Deep Neural Networks* (TT-DNNs)—models designed or informed explicitly by theoretical principles. TT-DNNs represent a promising paradigm shift in the design and training of neural networks [2]. Instead of relying solely on heuristics or massive empirical exploration, this approach incorporates insights from optimization theory, statistical learning, information theory, and approximation theory into the architecture, initialization, and training procedures of DNNs [3]. These theoretically grounded strategies aim to improve generalization, robustness, interpretability, and training efficiency. The goal of this survey is to provide a comprehensive overview of the field of Theory-Trained Deep Neural Networks. We categorize the literature along several dimensions, including theoretical frameworks that guide network training, algorithmic developments that exploit these insights, and practical implementations that demonstrate the effectiveness of theory-informed models [4]. We also highlight open problems and future research directions, emphasizing the interdisciplinary nature of this evolving area. By synthesizing recent advances and organizing them within a unifying framework, this survey aims to serve both as a foundational reference for researchers entering the field and as a roadmap for future theoretical and practical developments in deep learning.

2. Background and Preliminaries

To provide context for Theory-Trained Deep Neural Networks (TT-DNNs), we first review foundational concepts in deep learning and the theoretical frameworks that inform them. This section covers the standard components of deep neural networks, outlines common training procedures, and introduces the mathematical tools often used to analyze and guide neural network design [5].

2.1. Deep Neural Networks

A deep neural network is a parameterized function $f_\theta : \mathbb{R}^d \to \mathbb{R}^k$ composed of multiple layers of affine transformations and nonlinear activations [6]. Formally, a feedforward network with L layers can be expressed as:

$$f_\theta(x) = f_L \circ f_{L-1} \circ \cdots \circ f_1(x),$$

where each layer $f_l(x) = \sigma(W_l x + b_l)$ for weight matrix W_l, bias b_l, and activation function σ [7]. Training typically involves minimizing a loss function $\mathcal{L}(f_\theta(x), y)$ over a dataset $\{(x_i, y_i)\}_{i=1}^n$ using stochastic gradient descent (SGD) or its variants [8].
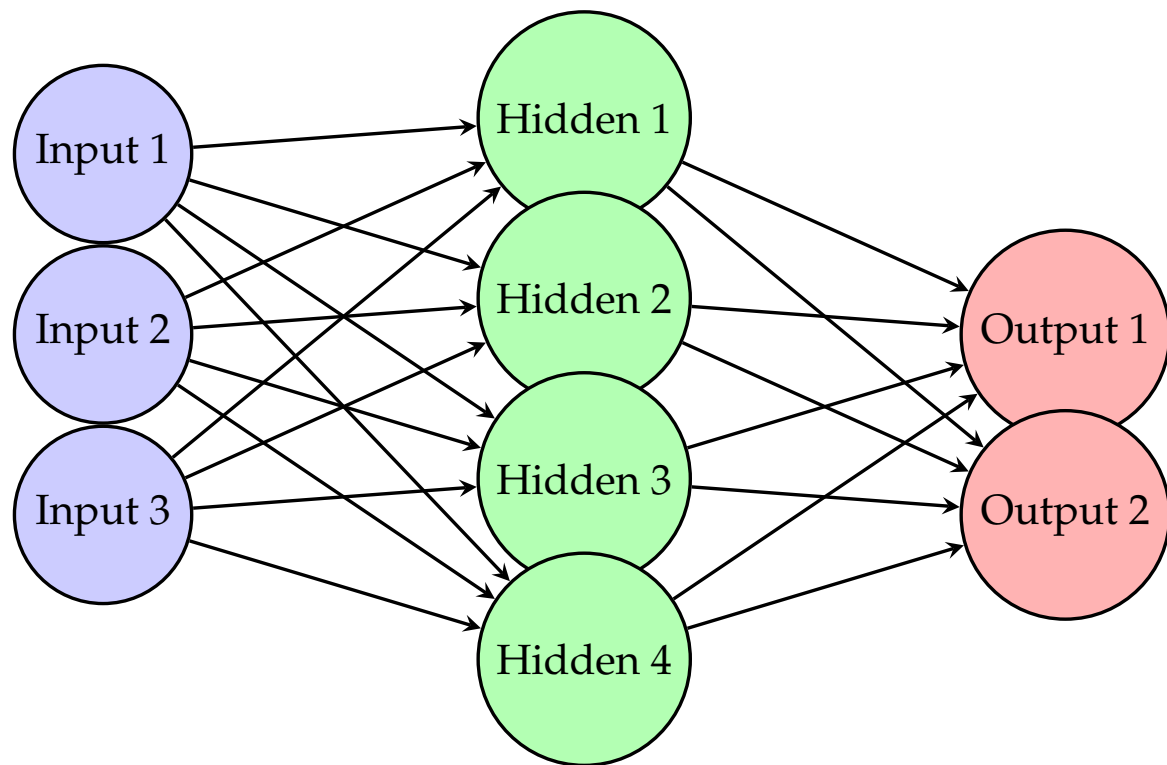
Figure 1. A simple feedforward neural network with one hidden layer [9]. Nodes represent neurons, and arrows indicate weighted connections.

2.2. Theoretical Foundations

Several branches of theory underpin recent advances in TT-DNNs:

- **Optimization Theory:** Analyzes the convergence properties of gradient-based methods, particularly in non-convex settings typical of deep networks.
- **Statistical Learning Theory:** Provides generalization bounds and capacity measures such as VC-dimension, Rademacher complexity, and uniform stability.
- **Approximation Theory:** Studies the expressivity of neural networks, including universal approximation theorems and the role of depth and width [10].
- **Information Theory:** Offers tools for understanding compression, mutual information, and the information bottleneck principle in learning dynamics.

2.3. Motivating the Need for Theory-Guided Training

Empirical deep learning often relies on trial-and-error heuristics, large-scale computation, and overparameterization [11]. While successful, this approach lacks principled guarantees and can lead to brittle models. TT-DNNs aim to reduce reliance on brute-force methods by leveraging the above theoretical insights to guide architecture design, training dynamics, and regularization strategies [12]. This can result in models that are more data-efficient, interpretable, and robust to perturbations [13]. The remainder of this survey builds on these foundations to explore how theory can be used not just to analyze but to actively shape the training of deep neural networks.

3. Taxonomy of Theory-Trained Deep Neural Networks

Theory-Trained Deep Neural Networks (TT-DNNs) encompass a diverse set of approaches that integrate theoretical principles into network design and training [14]. To provide clarity and structure, we categorize existing work into four primary classes based on the nature of the theoretical guidance and its role in the training process:

3.1. Optimization-Guided Networks

This category focuses on incorporating optimization theory to design architectures and training algorithms that improve convergence properties and stability. Examples include:

- Networks trained with provably convergent optimization schemes tailored for non-convex landscapes [15].
- Methods that exploit gradient flow dynamics or adaptive step sizes inspired by theoretical insights.

3.2. Regularization and Generalization-Driven Networks

These approaches employ statistical learning theory to derive principled regularization techniques and complexity control mechanisms:

- Architectures designed to enforce low-complexity function classes.
- Training methods that minimize generalization bounds, such as PAC-Bayes inspired algorithms or margin-based objectives.

3.3. Expressivity-Optimized Networks

Based on approximation theory, these methods design network structures (depth, width, activation choices) to maximize expressive power while maintaining trainability:

- Networks with theoretically grounded layer designs enabling efficient approximation of target functions [16].
- Architectures that leverage sparse or structured representations for compactness and interpretability [17].

3.4. Information-Theoretic Training Paradigms

This category leverages information theory to inform training objectives and representations:

- Models trained using the information bottleneck principle to balance compression and relevance [18].
- Approaches that quantify and optimize mutual information between layers or between inputs and outputs [19].

3.5. Hybrid Approaches

Many state-of-the-art TT-DNN frameworks combine elements from multiple theoretical domains, resulting in hybrid methods that simultaneously address optimization, generalization, expressivity, and information flow. This taxonomy provides a high-level map of the landscape, helping researchers

to identify where their work fits and how different theoretical insights complement each other in advancing the understanding and performance of deep neural networks [20].

4. Theoretical Foundations of Optimization in Deep Neural Networks

Optimization lies at the heart of training deep neural networks. Given a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, the goal is to find network parameters $\theta \in \mathbb{R}^p$ that minimize the empirical risk defined by a loss function $\ell : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}_+$:

$$\min_{\theta \in \mathbb{R}^p} \mathcal{L}(\theta) := \frac{1}{n} \sum_{i=1}^{n} \ell\big(f_\theta(x_i), y_i\big),$$

where $f_\theta : \mathbb{R}^d \to \mathbb{R}^k$ is the neural network parameterized by θ [21]. Due to the non-convex nature of $\mathcal{L}(\theta)$, classical convex optimization guarantees do not directly apply [22]. However, recent theoretical advances analyze the loss landscape through tools such as the Neural Tangent Kernel (NTK) [23]. The NTK at initialization, denoted Θ_0, characterizes the first-order behavior of the network output with respect to parameters:

$$\Theta_0(x, x') = \nabla_\theta f_{\theta_0}(x)^\top \nabla_\theta f_{\theta_0}(x'),$$

where θ_0 is the initial parameter vector [24]. In the infinite-width limit, training dynamics under gradient descent are approximated by a kernel regression with kernel Θ_0, yielding linear convergence rates:

$$\mathcal{L}(\theta_t) \leq \exp(-\lambda_{\min}(\Theta_0)t)\mathcal{L}(\theta_0),$$

where $\lambda_{\min}(\Theta_0)$ is the minimal eigenvalue of the NTK Gram matrix. Furthermore, the implicit bias of gradient-based optimization methods can favor solutions with desirable properties [25]. For example, gradient descent on linear models converges to the minimum norm solution. Extending this to deep nonlinear networks remains an active area of research. Understanding these theoretical foundations informs the design of optimization algorithms and initialization schemes that promote faster convergence and better generalization in deep learning.

5. Generalization Theory in Deep Neural Networks

A fundamental question in deep learning is understanding why highly overparameterized neural networks generalize well despite their capacity to perfectly fit training data. Classical learning theory typically bounds the generalization error $\mathcal{E}_{\mathrm{gen}}$ by controlling model complexity, but these bounds often fail to explain deep networks' empirical behavior [26]. Formally, the generalization error is defined as the difference between the expected risk and the empirical risk:

$$\mathcal{E}_{\mathrm{gen}}(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{P}}\big[\ell(f_\theta(x), y)\big] - \frac{1}{n} \sum_{i=1}^{n} \ell\big(f_\theta(x_i), y_i\big),$$

where $\mathcal{P}$ denotes the true data distribution [27]. Recent theoretical advances have introduced norm-based capacity measures that better capture generalization in deep networks. For example, the path norm $\|\theta\|_{\mathrm{path}}$ is defined as the sum of products of weights along all paths from input to output, serving as a complexity measure:

$$\|\theta\|_{\mathrm{path}} = \sum_{\pi \in \Pi} \prod_{l=1}^{L} \|W_l^{(\pi)}\|,$$

where Π denotes the set of all paths and $W_l^{(\pi)}$ are weights along path π at layer l [28]. PAC-Bayesian frameworks provide probabilistic bounds on generalization by considering distributions over parameters rather than point estimates. Given a prior distribution P and a learned posterior Q, the PAC-Bayes bound with probability at least $1 - \delta$ states:

$$\mathbb{E}_{\theta \sim Q}\big[\mathcal{E}_{\mathrm{gen}}(\theta)\big] \leq \sqrt{\frac{KL(Q\|P) + \log\frac{2\sqrt{n}}{\delta}}{2n}},$$

where $KL(Q\|P)$ is the Kullback–Leibler divergence between Q and P. Moreover, margin-based theories argue that the robustness of classification depends on the margin distribution of training samples. For a classifier f_θ, the margin of a sample (x_i, y_i) is defined as:

$$\gamma_i = y_i f_\theta(x_i),$$

and maximizing the minimum margin across samples can improve generalization [29]. These theoretical tools collectively contribute to a richer understanding of deep network generalization, guiding the development of regularization techniques and training algorithms that yield more robust models [30].

6. Information-Theoretic Perspectives on Deep Learning

Information theory provides a powerful framework to analyze and understand the learning dynamics of deep neural networks [31]. One of the most influential concepts in this context is the *Information Bottleneck* (IB) principle, which formalizes the trade-off between compression of the input data and preservation of relevant information for the output task. The IB principle views the learning process as finding a representation T of the input X that maximally compresses the information about X while retaining sufficient information to predict the output Y [32]. Mathematically, this trade-off is formulated as the optimization problem:

$$\min_{p(t|x)} \; I(X;T) - \beta I(T;Y),$$

where $I(\cdot;\cdot)$ denotes mutual information and $\beta > 0$ is a Lagrange multiplier balancing compression and prediction [33]. Here, $I(X;T)$ measures the amount of information T retains about the input, and $I(T;Y)$ quantifies how informative T is about the output. By minimizing this objective, the network is encouraged to discard irrelevant details from the input while preserving predictive features. This perspective leads to a conceptual interpretation of deep layers as successive representations $T_1, T_2, \ldots, T_L$ that progressively compress the input information, forming a Markov chain:

$$Y \leftrightarrow X \leftrightarrow T_1 \leftrightarrow T_2 \leftrightarrow \cdots \leftrightarrow T_L,$$

where each T_ℓ ideally contains less redundant information than the previous layer but retains essential characteristics for prediction [34]. Empirical studies have observed phases of fitting and compression during training, lending support to the IB hypothesis [35].
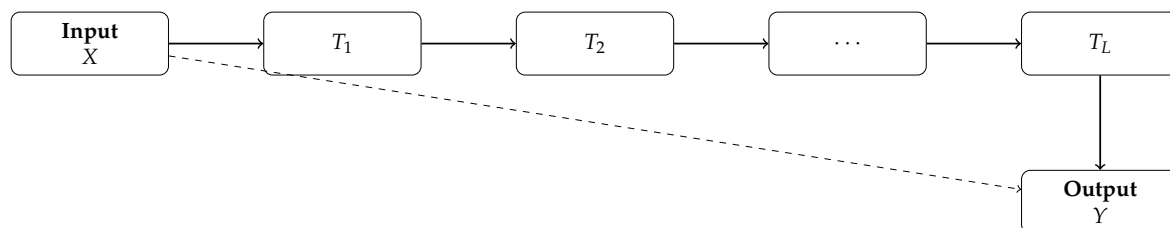
Figure 2. Schematic of the Information Bottleneck framework: input X is transformed through successive representations $T_1, \ldots, T_L$ which compress irrelevant information while preserving predictive content about output Y. The dashed arrow indicates the direct but complex relationship between input and output.

Despite its elegant formulation, computing mutual information terms in high-dimensional neural networks is notoriously challenging due to the intractability of the true distributions involved. This has motivated the development of variational approximations and surrogate objectives such as the Variational Information Bottleneck (VIB), which enables practical implementation by parameterizing distributions with neural networks and optimizing tractable bounds. The information-theoretic perspective also informs the design of regularization techniques, such as noise injection and dropout, which can be interpreted as mechanisms encouraging compression and robustness. Moreover, maximizing mutual information between intermediate representations and the input or output can improve

feature quality and interpretability. Overall, viewing deep learning through the lens of information theory provides profound insights into representation learning and generalization, while inspiring novel training strategies that balance expressivity and compression in neural networks.

7. Theory-Informed Training Paradigms

In this section, we review prominent training methodologies that explicitly incorporate theoretical principles to improve the learning process and final model performance [36]. These paradigms demonstrate how theory can guide practical algorithm design, enhancing convergence, generalization, and interpretability.

7.1. Optimization-Theoretic Approaches

Optimization theory has inspired several modifications to standard training algorithms, including:

- **Adaptive Gradient Methods with Theoretical Guarantees:** Algorithms such as Adam, AdaGrad, and their variants have been analyzed to better understand convergence rates and stability in non-convex settings [37].
- **Implicit Regularization through Gradient Descent:** Recent works show that gradient-based optimization implicitly biases solutions towards simpler models, leading to better generalization without explicit regularization.
- **Optimization in the Overparameterized Regime:** Theory-driven approaches leverage the Neural Tangent Kernel (NTK) and mean-field analysis to explain and improve training dynamics when networks are heavily overparameterized [38].

7.2. Regularization and Generalization Techniques

These methods leverage statistical learning theory to enforce constraints that promote better generalization:

- **PAC-Bayesian Training:** Training frameworks that optimize PAC-Bayes bounds by incorporating prior distributions over parameters and minimizing expected risk [39].
- **Margin-Based Objectives:** Encouraging large margins in the feature space to improve robustness and reduce overfitting.
- **Norm-Based Regularization:** Using weight norms or path norms with provable generalization guarantees [40].

7.3. Architecture and Expressivity-Driven Training

Theory-informed choices about network architecture can substantially impact training outcomes:

- **Depth vs [41]. Width Trade-offs:** Architectures designed based on universal approximation theorems to optimize representational capacity.
- **Sparse and Structured Networks:** Leveraging theoretical sparsity results to build compact, efficient models.
- **Activation Function Design:** Theoretical insights guide the choice and modification of activation functions to improve gradient flow and expressivity [42].

7.4. Information-Theoretic Training Strategies

Information theory offers a principled framework to analyze and shape learning dynamics:

- **Information Bottleneck Principle:** Training objectives that compress intermediate representations while preserving relevant information about the output [43].
- **Mutual Information Maximization:** Methods that maximize mutual information between inputs and learned features to encourage informative representations [44].
- **Entropy-Regularized Training:** Introducing entropy constraints to promote robustness and diversity in learned features [45].

Each of these paradigms illustrates how theoretical insights can be directly translated into effective training practices, moving beyond purely empirical approaches towards more principled deep learning methodologies.

8. Applications and Case Studies

Theory-Trained Deep Neural Networks (TT-DNNs) have demonstrated significant benefits across a variety of domains, showcasing the practical value of integrating theoretical insights into model design and training. This section highlights notable applications and case studies that illustrate the impact of theory-informed approaches.

8.1. Computer Vision

In computer vision, theory-driven architectures and training algorithms have led to improved robustness and efficiency:

-   **Provably Robust Networks:** Models trained with adversarial robustness guarantees using optimization-based regularization and certification methods [46].
-   **Efficient Architectures:** Sparse and structured networks informed by approximation theory have reduced model size without sacrificing accuracy [47].

8.2. Natural Language Processing

Theoretical frameworks have guided training in NLP tasks to improve generalization and interpretability:

-   **Information Bottleneck in Language Models:** Using compression-based objectives to enhance representation quality and reduce overfitting [48].
-   **Optimization-Guided Fine-Tuning:** Adapting optimization schedules based on theoretical analyses to stabilize training of large transformer models [49].

8.3. Scientific Computing and Physics

TT-DNNs are increasingly employed in domains requiring reliable and interpretable models:

-   **Physics-Informed Neural Networks (PINNs):** Networks trained with PDE constraints that leverage approximation theory to solve complex physical systems.
-   **Uncertainty Quantification:** PAC-Bayes-based methods providing principled uncertainty estimates in scientific prediction tasks [50].

8.4. Healthcare and Medical Imaging

Theory-trained models contribute to trustworthy AI in sensitive areas:

-   **Robust Diagnostic Models:** Training with margin-based objectives to improve resistance to data shifts and adversarial attacks [51].
-   **Data-Efficient Learning:** Leveraging theory-inspired regularization to train with limited annotated medical data.

These case studies underscore how theoretical insights can translate into tangible improvements in real-world systems, driving advances that are not only empirically effective but also principled and reliable.

9. Open Challenges and Future Directions

Despite significant progress in integrating theoretical insights into the training of deep neural networks, several challenges remain unresolved [52]. Addressing these open problems is crucial for advancing the reliability, efficiency, and interpretability of theory-trained models. We highlight some key directions for future research:

9.1. Bridging Theory and Practice

Many theoretical results rely on assumptions or simplified models that do not fully capture the complexities of practical deep learning scenarios [53]. Developing theories that better reflect real-world architectures, datasets, and training dynamics remains an ongoing challenge.

9.2. Scalability of Theory-Informed Methods

While theory-guided algorithms often provide guarantees and interpretability, their computational overhead can be substantial. Designing scalable and efficient implementations that maintain theoretical benefits is a critical research area.

9.3. Understanding Generalization in Overparameterized Regimes

Overparameterized networks defy classical learning theory, yet they generalize remarkably well in practice [54]. Creating refined theoretical frameworks to explain and leverage this phenomenon will deepen our understanding of deep learning [55].

9.4. Robustness and Safety Guarantees

Providing provable robustness against adversarial attacks, distribution shifts, and noisy data is essential for deploying deep networks in safety-critical applications [56]. Developing theory-trained models with rigorous safety guarantees remains a priority [57].

9.5. Interpretable and Explainable Models

Theory-informed design can enhance model interpretability, but creating frameworks that offer clear, actionable explanations without sacrificing performance is still an open challenge [58].

9.6. Unified Theoretical Frameworks

Currently, theoretical insights come from diverse fields such as optimization, information theory, and approximation theory, often in isolation [59]. Integrating these perspectives into unified frameworks could lead to more holistic and powerful theory-trained neural networks.

9.7. Data Efficiency and Transfer Learning

Improving sample efficiency and enabling effective transfer of theory-trained models across tasks and domains are important practical goals that require further theoretical development [60,61]. Addressing these challenges will require interdisciplinary collaboration and continued innovation at the intersection of theory and empirical deep learning. We anticipate that future work in these directions will significantly enhance the capabilities and reliability of deep neural networks [62].

10. Conclusion

Theory-Trained Deep Neural Networks represent a promising frontier in the quest to unify the empirical success of deep learning with rigorous theoretical understanding. By integrating insights from optimization theory, statistical learning, approximation theory, and information theory, TT-DNNs aim to improve model robustness, generalization, interpretability, and training efficiency.

This survey has provided a comprehensive overview of the field, categorizing existing approaches, detailing theory-informed training paradigms, and highlighting key applications where these methods have made significant impact. We also identified open challenges and future directions that will shape the continued evolution of theory-trained models.

As deep learning continues to expand into critical and complex domains, the need for principled, theory-guided methodologies becomes ever more pressing. We believe that ongoing research in theory-trained deep neural networks will not only deepen our fundamental understanding but also drive the development of more reliable, efficient, and transparent AI systems.

We hope this survey serves as a useful resource for researchers and practitioners interested in bridging theory and practice in deep learning, inspiring future innovations at this vibrant intersection.

References

1. Gupta, A.; Mishra, B. Neuroevolving monotonic PINNs for particle breakage analysis. In Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI). IEEE, 2024, pp. 993–996.

2. Zhang, J.; Du, J.; Zhang, M.; Chen, Z.; Zhang, H.; Nie, C. Aerodynamic Performance Improvement of a Highly Loaded Compressor Airfoil with Coanda Jet Flap. *Journal of Thermal Science* **2022**, *31*, 151–162.

3. Zou, A.M.; Fan, Z. Fixed-time attitude tracking control for rigid spacecraft without angular velocity measurements. *IEEE Transactions on Industrial Electronics* **2019**, *67*, 6795–6805.

4. Gunning, D.; Aha, D. DARPA's explainable artificial intelligence (XAI) program. *AI magazine* **2019**, *40*, 44–58.

5. Lau, G.K.R.; Hemachandra, A.; Ng, S.K.; Low, B.K.H. PINNACLE: PINN Adaptive ColLocation and Experimental points selection. *arXiv preprint arXiv:2404.07662* **2024**.

6. Li, Z. Neural operator: Learning maps between function spaces. In Proceedings of the 2021 Fall Western Sectional Meeting. AMS, 2021.

7. Gupta, A.; Mishra, B.K. Globally optimized dynamic mode decomposition: A first study in particulate systems modelling. *Theoretical and Applied Mechanics Letters* **2024**, p. 100563.

8. Bengio, Y.; Bengio, S.; Cloutier, J. *Learning a synaptic learning rule*; Citeseer, 1990.

9. Tartakovsky, A.M.; Marrero, C.O.; Perdikaris, P.; Tartakovsky, G.D.; Barajas-Solano, D. Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research* **2020**, *56*, e2019WR026731.

10. Wang, C.; Li, S.; He, D.; Wang, L. Is $\mathcal{L}^2$ Physics-Informed Loss Always Suitable for Training Physics-Informed Neural Network? *arXiv preprint arXiv:2206.02016* **2022**.

11. Schmidhuber, J. Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook. PhD thesis, Technische Universität München, 1987.

12. Nichol, A.; Achiam, J.; Schulman, J. On First-Order Meta-Learning Algorithms, 2018, [arXiv:cs.LG/1803.02999].

13. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *nature* **2021**, *596*, 583–589.

14. Chen, X.; Duan, J.; Karniadakis, G.E. Learning and meta-learning of stochastic advection–diffusion–reaction systems from sparse measurements. *European Journal of Applied Mathematics* **2021**, *32*, 397–420.

15. Kharazmi, E.; Cai, M.; Zheng, X.; Zhang, Z.; Lin, G.; Karniadakis, G.E. Identifiability and predictability of integer-and fractional-order epidemiological models using physics-informed neural networks. *Nature Computational Science* **2021**, *1*, 744–753.

16. Psaros, A.F.; Kawaguchi, K.; Karniadakis, G.E. Meta-learning PINN loss functions. *Journal of computational physics* **2022**, *458*, 111121.

17. de Wolff, T.; Lincopi, H.C.; Martí, L.; Sanchez-Pi, N. Mopinns: an evolutionary multi-objective approach to physics-informed neural networks. In Proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference Companion, 2022, pp. 228–231.

18. Ataei, M.; Salehipour, H. XLB: A differentiable massively parallel lattice Boltzmann library in Python. *Computer Physics Communications* **2024**, *300*, 109187.

19. Ruiz Herrera, C.; Grandits, T.; Plank, G.; Perdikaris, P.; Sahli Costabal, F.; Pezzuto, S. Physics-informed neural networks to learn cardiac fiber orientation from multiple electroanatomical maps. *Engineering with Computers* **2022**, *38*, 3957–3973.

20. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nature Reviews Physics* **2021**, *3*, 422–440.

21. Ma, W.; Liu, Z.; Kudyshev, Z.A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep learning for the design of photonic structures. *Nature Photonics* **2021**, *15*, 77–90.

22. Tang, K.; Wan, X.; Yang, C. DAS-PINNs: A deep adaptive sampling method for solving high-dimensional partial differential equations. *Journal of Computational Physics* **2023**, *476*, 111868.

23. Baxter, J.; Caruana, R.; Mitchell, T.; Pratt, L.Y.; Silver, D.L.; Thrun, S. Learning to learn: Knowledge consolidation and transfer in inductive systems. In Proceedings of the NIPS Workshop, http://plato. acadiau. ca/courses/comp/dsilver/NIPS95_LTL/transfer. workshop, 1995.

24. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys* **2022**, *55*, 1–37.

25. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International conference on machine learning. PMLR, 2017, pp. 1126–1135.

26. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **2009**, *22*, 1345–1359.

27. Cuomo, S.; Di Cola, V.S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing* **2022**, *92*, 88.

28. Yu, G.; Ma, L.; Jin, Y.; Du, W.; Liu, Q.; Zhang, H. A survey on knee-oriented multiobjective evolutionary optimization. *IEEE transactions on evolutionary computation* **2022**, *26*, 1452–1472.

29. Zou, Z.; Karniadakis, G.E. L-HYDRA: Multi-head physics-informed neural networks. *arXiv preprint arXiv:2301.02152* **2023**.

30. Viana, F.A.; Subramaniyan, A.K. A survey of Bayesian calibration and physics-informed neural networks in scientific modeling. *Archives of Computational Methods in Engineering* **2021**, *28*, 3801–3830.

31. Pellegrin, R.; Bullwinkel, B.; Mattheakis, M.; Protopapas, P. Transfer learning with physics-informed neural networks for efficient simulation of branched flows. *arXiv preprint arXiv:2211.00214* **2022**.

32. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895* **2020**.

33. Kho, J.; Koh, W.; Wong, J.C.; Chiu, P.H.; Ooi, C.C. Design of turing systems with physics-informed neural networks. In Proceedings of the 2022 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2022, pp. 1180–1186.

34. Négiar, G.; Mahoney, M.W.; Krishnapriyan, A.S. Learning differentiable solvers for systems with hard constraints. *arXiv preprint arXiv:2207.08675* **2022**.

35. Yang, S.; Tian, Y.; He, C.; Zhang, X.; Tan, K.C.; Jin, Y. A gradient-guided evolutionary approach to training deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *33*, 4861–4875.

36. Gao, Y.; Cheung, K.C.; Ng, M.K. Svd-pinns: Transfer learning of physics-informed neural networks via singular value decomposition. In Proceedings of the 2022 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2022, pp. 1443–1450.

37. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **2019**, *378*, 686–707.

38. Cai, S.; Mao, Z.; Wang, Z.; Yin, M.; Karniadakis, G.E. Physics-informed neural networks (PINNs) for fluid mechanics: A review. *Acta Mechanica Sinica* **2021**, *37*, 1727–1738.

39. Qin, T.; Beatson, A.; Oktay, D.; McGreivy, N.; Adams, R.P. Meta-pde: Learning to solve pdes quickly without a mesh. *arXiv preprint arXiv:2211.01604* **2022**.

40. Pratt, L.Y.; Mostow, J.; Kamm, C.A. Direct transfer of learned information among neural networks. In Proceedings of the Proceedings of the ninth National conference on Artificial intelligence-Volume 2, 1991, pp. 584–589.

41. Bonfanti, A.; Bruno, G.; Cipriani, C. The Challenges of the Nonlinear Regime for Physics-Informed Neural Networks. *arXiv preprint arXiv:2402.03864* **2024**.

42. Liu, Y.; Liu, W.; Yan, X.; Guo, S.; Zhang, C.a. Adaptive transfer learning for PINN. *Journal of Computational Physics* **2023**, *490*, 112291.

43. Wang, S.; Perdikaris, P. Long-time integration of parametric evolution equations with physics-informed deeponets. *Journal of Computational Physics* **2023**, *475*, 111855.

44. Sung, N.; Wong, J.C.; Ooi, C.C.; Gupta, A.; Chiu, P.H.; Ong, Y.S. Neuroevolution of physics-informed neural nets: benchmark problems and comparative results. In Proceedings of the Proceedings of the Companion Conference on Genetic and Evolutionary Computation, 2023, pp. 2144–2151.

45. Lu, L.; Meng, X.; Mao, Z.; Karniadakis, G.E. DeepXDE: A deep learning library for solving differential equations. *SIAM review* **2021**, *63*, 208–228.

46. Cho, W.; Lee, K.; Rim, D.; Park, N. Hypernetwork-based meta-learning for low-rank physics-informed neural networks. *Advances in Neural Information Processing Systems* **2024**, *36*.

47. Wang, S.; Yu, X.; Perdikaris, P. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics* **2022**, *449*, 110768.

48. Mahmoudabadbozchelou, M.; Jamali, S. Rheology-informed neural networks (RhINNs) for forward and inverse metamodelling of complex fluids. *Scientific reports* **2021**, *11*, 1–13.

49. de Avila Belbute-Peres, F.; Chen, Y.f.; Sha, F. HyperPINN: Learning parameterized differential equations with physics-informed hypernetworks. In Proceedings of the The symbiosis of deep learning and differential equations, 2021.

50. Huang, X.; Ye, Z.; Liu, H.; Ji, S.; Wang, Z.; Yang, K.; Li, Y.; Wang, M.; Chu, H.; Yu, F.; et al. Meta-auto-decoder for solving parametric partial differential equations. *Advances in Neural Information Processing Systems* **2022**, *35*, 23426–23438.

51. Wang, Z.; Zhang, Q.; Ong, Y.S.; Yao, S.; Liu, H.; Luo, J. Choose appropriate subproblems for collaborative modeling in expensive multiobjective optimization. *IEEE Transactions on Cybernetics* **2021**, *53*, 483–496.

52. Battaglia, P.W.; Hamrick, J.B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261* **2018**.

53. Garbet, X.; Idomura, Y.; Villard, L.; Watanabe, T. Gyrokinetic simulations of turbulent transport. *Nuclear Fusion* **2010**, *50*, 043002.

54. Von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Ramamurthy, R.; et al. Informed Machine Learning–A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering* **2021**, *35*, 614–633.

55. Kapoor, T.; Wang, H.; Núñez, A.; Dollevoet, R. Transfer learning for improved generalizability in causal physics-informed neural networks for beam simulations. *Engineering Applications of Artificial Intelligence* **2024**, *133*, 108085.

56. Cai, S.; Wang, Z.; Fuest, F.; Jeon, Y.J.; Gray, C.; Karniadakis, G.E. Flow over an espresso cup: inferring 3-D velocity and pressure fields from tomographic background oriented Schlieren via physics-informed neural networks. *Journal of Fluid Mechanics* **2021**, *915*, A102.

57. Wandel, N.; Weinmann, M.; Neidlin, M.; Klein, R. Spline-pinn: Approaching pdes without data using fast, physics-informed hermite-spline cnns. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2022, Vol. 36, pp. 8529–8538.

58. Chiu, P.H.; Wong, J.C.; Ooi, C.; Dao, M.H.; Ong, Y.S. CAN-PINN: A fast physics-informed neural network based on coupled-automatic–numerical differentiation method. *Computer Methods in Applied Mechanics and Engineering* **2022**, *395*, 114909.

59. Jin, Y.; Wang, H.; Chugh, T.; Guo, D.; Miettinen, K. Data-driven evolutionary optimization: An overview and case studies. *IEEE Transactions on Evolutionary Computation* **2018**, *23*, 442–458.

60. Prantikos, K.; Chatzidakis, S.; Tsoukalas, L.H.; Heifetz, A. Physics-informed neural network with transfer learning (TL-PINN) based on domain similarity measure for prediction of nuclear reactor transients. *Scientific Reports* **2023**, *13*, 16840.

61. Zniyed, Y.; Nguyen, T.P.; et al. Enhanced network compression through tensor decompositions and pruning. *IEEE Transactions on Neural Networks and Learning Systems* **2024**.

62. Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explainable machine learning for scientific insights and discoveries. *Ieee Access* **2020**, *8*, 42200–42216.