APPENDIX

A. KADL for Various Machine Learning Models

Knowledge-augmented deep learning has emerged as a versatile approach that can be seamlessly integrated with a broad spectrum of machine learning (ML) algorithms, enhancing their capabilities and extending their applicability across various domains. A diverse range of ML algorithms have been paired with knowledge augmentation, spanning traditional methods to advanced deep neural architectures, such as the convolutional neural network (CNN) [1], recurrent neural network (RNN) [2], [3], variational autoencoder [4], [5], and generative adversarial network (GAN) [6], [7]. For instance, when combined with CNNs [1], knowledge might be incorporated as physics-constrained intermediate variables as a way of embedding physical principles. Besides, attention mechanisms [8], graph-based neural networks [9], and fusion layers [10] are common techniques used to effectively blend the acquired knowledge with the model's internal computations.

Beyond deep models, knowledge augmentation has also been harnessed to bolster other ML methods. For instance, prior knowledge on probabilistic dependencies has not only been considered for boosting deep models, but more widely considered for learning of probabilistic graphical models (PGMs). Abstract structural knowledge represented as block-structured priors is shown to help improve PGM structure learning from sparse data [11]. The domain knowledge on relatedness of tasks has been applied in multi-task BN structure learning to exploit the structural commonalities across tasks [12]. Basides, Markov logic network (MLN) [13] combines logic rules with probabilistic graphical models to model first-order logic (FOL) and its uncertainties. Besides PGMs, experts' rating as domain knowledge is used as an input of the data mining classification methods, such as naive Bayes, logistic regression, decision tree, decision table, k-nearest neighbor, and support vector machine (SVM), to reduce the misclassification cost on evaluating applications in the domain of indirect bank lending [14]. Similarly, features are ranked by domain experts and are incorporated into SVM classifier [15].

In this survey, we focus on discussing knowledge augmentation for deep learning models.

B. Evaluations of KADL

Knowledge-augmented deep learning (KADL) demonstrates its potential to enhance deep learning performance across diverse applications. Through numerous empirical studies, we observe consistent improvements achieved by various KADL algorithms. To comprehensively assess the impact of KADL, we focus on three critical performance aspects: accuracy, data efficiency, and generalization. Since the evaluation varies with respect to different applications and thus different benchmarks, we choose representative works as case studies to illustrate the effectiveness of KADL.

Accuracy

KADL has been demonstrated effective in improving the accuracy for various applications. In computer vision, KADL has been harnessed to advance human behavior understanding, such as 3D body pose estimation and reconstruction [16], [17], and facial action units (AU) detection [18], [19]. It has also been applied for motion understanding such as force prediction [1], [20], [21] and tracking [22]. Leveraging its integration with semantic knowledge, KADL has also advanced scene understanding [23], object and relation detection [24], [25], and visual question answering (VQA) [10] tasks. For example, in the object detection task, Fang et al. [24] leveraged semantic consistency constraints derived from knowledge graph to regularize the neural network. Their empirical evaluation on two benchmark datasets highlighted a substantial enhancement in recall by up to 6.3 points, without compromising mean average precision, when compared to the state-of-the-art model. In the context of VQA, the KRISP model proposed by Marino et al. [10] stands as an illustrative example. By harnessing both implicit and explicit symbolic knowledge, KRISP achieved remarkable accuracy gains. Its evaluation on the OK-VQA benchmark dataset [26] showcased an accuracy of 38.35%, notably surpassing the best ConceptBert performance of 33.66%. These outcome underscores the ability of KADL to fuse diverse forms of knowledge to elevate accuracy in various tasks.

Within the domain of weather and climate understanding, KADL is explored in lake temperature modeling [2], [27], [28] and climate modeling [29]–[31]. An illustrative example can be the physics-guided architecture (PGA) [27] for lake temperature modeling. PGA was evaluated on two benchmark lakes: the Lake Mendota in Wisconsin, USA (Table I) and the Falling Creek Reservoir (FCR) in Virginia, USA (Table II). The evaluation metrics are the rooted mean square error (RMSE) on test set and the physical inconsistency. In particular, the test RMSE measures if every sample is accurate (per sample) while the physical inconsistency measures if the proposed model makes sense in scientific applications on average (mean). In direct comparison with two baseline models (namely LSTM and PGL), the proposed PGA model, by leveraging the incorporated physics principles, achieved the smallest Test RMSE, always preserving physical consistency. In particular, the PGL is the baseline where the physics principles are incorporated using physics-based loss functions which measure the physical inconsistency. When comparing PGL with PGA, it becomes evident that integrating physics principles via architecture customization is more effective than utilizing a physics-guided training loss. This observation is substantiated by superior performance in two benchmark lack temperature modeling tasks.

TABLE I
ACCURACY EVALUATION OF LAKE TEMPERATURE MODELING ON LAKE MENDOTA IN WISCONSIN, USA. TABLE IS REPRODUCED FROM [27].

	Test RMSE (in ^oC; per sample)	Physical inconsistency (mean)
LSTM	2.25 ± 0.14	0.10 ± 0.02
PGL	2.30 ± 0.12	0.12 ± 0.02
PGA	$\mathbf{2.09 \pm 0.18}$	$\mathbf{0.00 \pm 0.00}$

TABLE II
ACCURACY EVALUATION OF LAKE TEMPERATURE MODELING ON FALLING CREEK RESERVOIR (FCR) IN VIRGINIA, USA. TABLE IS REPRODUCED FROM [27].

	Test RMSE (in ^oC; per sample)	Physical inconsistency (mean)
LSTM	2.96 ± 0.22	0.07 ± 0.03
PGL	2.84 ± 0.16	0.08 ± 0.03
PGA	**2.19 ± 0.21**	**0.00 ± 0.00**

Data efficiency

KADL has demonstrated a notable capacity to significantly enhance data efficiency. One effective approach to leverage domain knowledge involves training deep models with data synthesized from conventional mechanistic models. By incorporating synthesized data, the deep models' reliance on training data is notably diminished [21], [23], [32], [33]. A concrete illustration is evident in the field of robotics. The success achieved in grasping 36 diverse and previously unseen physical objects by GraspGAN [33] substantially improves with a reduction in training samples, facilitated by the utilization of simulated samples. As depicted in Table III, they showed the grasp success performance with different quantities of real-world samples (100%, 20%, 10%, 2%, and 1%) and 8 million simulated samples. Real-only refers to the model that is given only real data. Rand., DANN, and DANN-R are three different baselines with different combinations of simulated data generation and domain adaptation methods. As shown, the proposed model GraspGAN achieved competitive performance across different quantities of real-world samples. Notably, when evaluated with only 1% of real-world samples, GraspGAN achieved a remarkable success rate of 59.95%. This outcome significantly outperforms the model trained exclusively on real-world data, which achieves a success rate of only 31.13%.

TABLE III
DATA EFFICIENCY EVALUATION OF SUCCESSFUL GRASPS FOR 36 DIVERSE AND PREVIOUSLY UNSEEN PHYSICAL OBJECTS. TABLE IS REPRODUCED FROM [33].

Method	All	20%	10%	2%	1%
Real-only	67.65%	64.93%	62.75%	35.46%	31.13%
Rand.	75.58%	70.16%	73.31%	63.61%	50.99%
DANN	76.26%	68.12%	71.93%	61.93%	59.27%
DANN-R.	72.60%	66.46%	**74.88**%	63.73%	43.81%
GraspGAN	**76.67**%	**74.07**%	70.70%	**68.51**%	**59.95**%

Furthermore, the utilization of domain knowledge in improving data efficiency extends to label-free supervisions in training neural networks [22]. Rather than relying on annotations, the domain knowledge is represented as physical equations and used for training. The effectiveness of this label-free supervision was assessed across various computer vision tasks. For instance, in the context of tracking an object in free fall, the model trained with the proposed label-free supervision resulted in a correlation of 90.1% with respect to the ground truth. For comparison, a supervised network trained with the annotations achieved a correlation of 94.5%. In the context of tracking the position of a walking man, the proposed model exhibited predictions with a 95.4% correlation to the ground truth. Surprisingly, the same network trained using direct supervision encountered challenges in generalization and yielded a notably lower correlation of 80.5% on the test set. These experiments illuminate the potential of supervising a neural network by employing only the relevant physics equations governing the object's behavior, diminishing the reliance on extensive annotations for training purposes.

Besides the above two types of examples, KADL in general has consistently demonstrated its effectiveness in improving data efficiency. Through harnessing integrated domain knowledge, KADL enables deep models to maintain competitiveness even when confronted with a reduced pool of training samples.

Generalization

One representative category of works that improves generalization is through the incorporation of equivariance. For instance, Thomas et al. [34] introduced equivariances via tensor field networks to improve generalization of neural networks. The proposed model is applied to physics and chemistry. In the domain of physics, they addressed the prediction of vectors and tensors within the framework of classical mechanics. Through learning, the model is able to predict vectors and tensors that perfectly align with fundamental principles such as the Newtonian gravity inverse square law and the moment of inertia radial functions. In the domain of chemistry, they addressed the generalization to missing molecules. When the model is trained on molecules featuring 5 to 18 atoms, it exhibited the ability to predict missing molecules containing 19 atoms, 23 atoms, and even those with 25 to 29 atoms, achieving an impressive accuracy of 100% across the board.

Similarly, Wang et al. [35] studied the incorporation of symmetries for improved generalization. They performed the evaluation on Rayleigh-Bénard convection. Root Mean Square Error (RMSE) between the forward predictions and the ground truth over all pixels is used as a measurement of accuracy. Energy Spectrum Error (ESE) is used as a measurement of physical consistency. To simulate real-world conditions where each sample lacks a fixed reference frame, they consider random transformations from the relevant symmetry group for each sample during both training and testing. Four types of equivariance is considered: uniform motion equivariance (UM), magnitude equivariance (Mag), rotational equivariance (Rot), and scale equivariance ($Scal$), As outlined in Table IV, compared to the baseline model ResNet, the outcomes demonstrate that Equ-ResNet consistently outperforms ResNet, yielding a notable average improvement of 34% in terms of RMSE and 40% in terms of ESE.

Furthermore, to understand the benefits of incorporating the equivariance via architecture customization compared to conventional data augmentation, Equ-ResNets is compared to ResNets that are trained with data augmentation. Results show that ESE is uniformly worse for models trained with data augmentation compared to the equivariant models. It is believed by the authors that data augmentation presents a trade-off in learning. Though the model can be less sensitive to the various transformations considered, it necessitates larger model sizes and more extensive training on augmented data.

TABLE IV
GENERALIZATION EVALUATION ON RAYLEIGH-BÉNARD CONVECTION. ROOT MEAN SQUARE ERROR (RMSE) MEASURES THE ACCURACY AND ENERGY SPECTRUM ERROR (ESE) MEASURES THE PHYSICAL CONSISTENCY. TABLE IS REPRODUCED FROM [35].

	RMSE	ESE
ResNet	1.03 ± 0.05	0.96 ± 0.10
Equ_{UM}	$\mathbf{0.69 \pm 0.01}$	$\mathbf{0.35 \pm 0.13}$
ResNet	1.50 ± 0.02	0.55 ± 0.11
Equ_{Mag}	$\mathbf{0.75 \pm 0.04}$	$\mathbf{0.39 \pm 0.02}$
ResNet	1.18 ± 0.05	1.21 ± 0.04
Equ_{Rot}	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.68 \pm 0.01}$
ResNet	0.92 ± 0.01	1.34 ± 0.07
Equ_{Scal}	$\mathbf{0.74 \pm 0.03}$	$\mathbf{1.02 \pm 0.02}$

By comparison, customizing architectures to incorporate symmetries does not require additional training efforts associated with learning symmetries from augmented data. In the context of weather and climate modeling, the aspect of generalization has also been explored in evaluating the knowledge-augmented deep learning, as outlined in Table 2 of the survey from Kashinath et al. [31].

C. Expressive power of knowledge-augmented deep learning

The depth or width of deep neural networks (DNNs) enable them to approximate a wide range of functions, making them highly flexible and adaptable for various tasks. This capacity is referred to as their "expressive power" or "representation power". Recently, people seek to gain deeper and more theoretical understanding of how and why neural network architectures achieves their practical successes. This includes measuring how the architectural properties of a neural network (depth, width, layer type) affect its performance [36], [37]. Knowledge augmentation is effective in enhancing the expressive power of deep neural networks. For example, the architecture design of LSTM is bio-inspired. In LSTM, memory cells with gating are introduced to store related data, while discarding unrelated data. Both scientific and experiential knowledge are used in design of DNN architecture to enhance its expressive power, as discussed in the main manuscript. As a result, DNNs become capable of capturing even more intricate and domain-specific patterns, leading to improved generalization and predictive accuracy.

However, characterizing the expressiveness of neural networks, and understanding how expressiveness varies with parameters of the architecture, has been a challenging problem due to the difficulty in identifying meaningful notions of expressivity and in linking their analysis to implications for these networks in practice. Furthermore, deep theoretical insights on how knowledge augmentation helps improve the expressivity of a neural network remains an unanswered question in the literature. Future efforts are required to establish deep theoretical understanding on the effectiveness of incorporating domain knowledge in enhancing the deep models' performance via quantifying the expressive power.

REFERENCES

[1] N. Muralidhar, J. Bu, Z. Cao, L. He, N. Ramakrishnan, D. Tafti, and A. Karpatne, "Phynet: Physics guided neural networks for particle drag force prediction in assembly," in *Proceedings of the 2020 SIAM International Conference on Data Mining*. SIAM, 2020, pp. 559–567.

[2] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, "Physics guided rnns for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 2019, pp. 558–566.

[3] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar, "Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles," *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–26, 2021.

[4] M. Jaques, M. Burke, and T. M. Hospedales, "Newtonianvae: Proportional control and goal identification from pixels via physical latent spaces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4454–4463.

[5] P. Liu, L. Tian, Y. Zhang, S. Aylward, Y. Lee, and M. Niethammer, "Discovering hidden physics behind transport dynamics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 082–10 092.

[6] J.-L. Wu, K. Kashinath, A. Albert, D. Chirila, H. Xiao *et al.*, "Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems," *Journal of Computational Physics*, vol. 406, p. 109209, 2020.

[7] V. Shah, A. Joshi, S. Ghosal, B. Pokuri, S. Sarkar, B. Ganapathysubramanian, and C. Hegde, "Encoding invariances in deep generative models," *arXiv preprint arXiv:1906.01626*, 2019.

[8] F. Liu, C. You, X. Wu, S. Ge, X. Sun *et al.*, "Auto-encoding knowledge graph for unsupervised medical report generation," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[9] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," *Advances in Neural Information Processing Systems*, vol. 31, pp. 1853–1863, 2018.

[10] K. Marino, X. Chen, D. Parikh, A. Gupta, and M. Rohrbach, "Krisp: Integrating implicit and symbolic knowledge for open-domain knowledge-based vqa," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 111–14 121.

[11] V. Mansinghka, C. Kemp, T. Griffiths, and J. Tenenbaum, "Structured priors for structure learning," *arXiv preprint arXiv:1206.6852*, 2012.

[12] D. Oyen and T. Lane, "Leveraging domain knowledge in multitask bayesian network structure learning," in *Twenty-Sixth AAAI conference on artificial intelligence*, 2012.

[13] M. Richardson and P. Domingos, "Markov logic networks," *Machine learning*, vol. 62, no. 1-2, pp. 107–136, 2006.

[14] A. P. Sinha and H. Zhao, "Incorporating domain knowledge into data mining classifiers: An application in indirect lending," *Decision Support Systems*, vol. 46, no. 1, pp. 287–299, 2008.

[15] K. Small, B. C. Wallace, C. E. Brodley, and T. A. Trikalinos, "The constrained weight space svm: learning with ranked features," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, pp. 865–872.

[16] J. Chen, S. Nie, and Q. Ji, "Data-free prior model for upper body pose estimation and tracking," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4627–4639, 2013.

[17] Y. Zhang, H. Wang, J. O. Kephart, and Q. Ji, "Body knowledge and uncertainty modeling for monocular 3d human body reconstruction," *arXiv preprint arXiv:2308.00799*, 2023.

[18] Z. Cui, T. Song, Y. Wang, and Q. Ji, "Knowledge augmented deep neural networks for joint facial expression and action unit recognition," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[19] T. Song, Z. Cui, W. Zheng, and Q. Ji, "Hybrid message passing with performance-driven structures for facial action unit detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6267–6276.

[20] Z. Li, J. Sedlar, J. Carpentier, I. Laptev, N. Mansard, and J. Sivic, "Estimating 3d motion and forces of person-object interactions from monocular video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8640–8649.

[21] K. Ehsani, S. Tulsiani, S. Gupta, A. Farhadi, and A. Gupta, "Use the force, luke! learning to predict physical forces by simulating effects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 224–233.

[22] R. Stewart and S. Ermon, "Label-free supervision of neural networks with physics and domain knowledge," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[23] R. Mottaghi, H. Bagherinezhad, M. Rastegari, and A. Farhadi, "Newtonian scene understanding: Unfolding the dynamics of objects in static

images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3521–3529.

[24] Y. Fang, K. Kuan, J. Lin, C. Tan, and V. Chandrasekhar, "Object detection meets knowledge graphs." International Joint Conferences on Artificial Intelligence, 2017.

[25] R. Yu, A. Li, V. I. Morariu, and L. S. Davis, "Visual relationship detection with internal and external linguistic knowledge distillation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1974–1982.

[26] K. Marino, M. Rastegari, A. Farhadi, and R. Mottaghi, "Ok-vqa: A visual question answering benchmark requiring external knowledge," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2019, pp. 3195–3204.

[27] A. Daw, R. Q. Thomas, C. C. Carey, J. S. Read, A. P. Appling, and A. Karpatne, "Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling," in *Proceedings of the 2020 siam international conference on data mining*. SIAM, 2020, pp. 532–540.

[28] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (pgnn): An application in lake temperature modeling," *arXiv preprint arXiv:1710.11431*, 2017.

[29] T. Beucler, S. Rasp, M. Pritchard, and P. Gentine, "Achieving conservation of energy in neural network emulators for climate modeling," *arXiv preprint arXiv:1906.06622*, 2019.

[30] P. Nowack, P. Braesicke, J. Haigh, N. L. Abraham, J. Pyle, and A. Voulgarakis, "Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations," *Environmental Research Letters*, vol. 13, no. 10, p. 104016, 2018.

[31] K. Kashinath, M. Mustafa, A. Albert, J. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh *et al.*, "Physics-informed machine learning: case studies for weather and climate modelling," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200093, 2021.

[32] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.

[33] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige *et al.*, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 4243–4250.

[34] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, "Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds," *arXiv preprint arXiv:1802.08219*, 2018.

[35] R. Wang, R. Walters, and R. Yu, "Incorporating symmetry into deep dynamics models for improved generalization," *arXiv preprint arXiv:2002.03061*, 2020.

[36] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein, "On the expressive power of deep neural networks," in *international conference on machine learning*. PMLR, 2017, pp. 2847–2854.

[37] S. Zhou and A. P. Schoellig, "An analysis of the expressiveness of deep neural network architectures based on their lipschitz constants," *arXiv preprint arXiv:1912.11511*, 2019.