Towards End-to-End Prompt-Vision-Physics Neural Network for Fast Design Discovery

Qingshan Xu[1*], Jiao Liu[1*], Melvin Wong[1*], Ge Jin[2*], Ryan Lau[1*],
Yew-Soon Ong[1,5], Stefan Menzel[3], Thiago Rios[3], Joo-Hwee Lim[4], Chin Chun Ooi[5]

[1]*College of Computing & Data Science (CCDS), Nanyang Technological University (NTU), Singapore*
[2]*School of Aerospace Engineering, Beijing Institute of Technology, Haidian, Beijing, China*
[3]*Honda Research Institute Europe (HRI-EU), Offenbach am Main, Germany*
[4]*Institute for Infocomm Research (I²R), Agency for Science, Technology and Research (A*STAR), Singapore*
[5]*Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore*

{qingshan.xu, jiao.liu, wong1357, RLAU014, asysong}@ntu.edu.sg, jinge52293@gmail.com,
{stefan.menzel, thiago.rios}@honda-ri.de, joohwee@i2r.a-star.edu.sg, ooicc@cfar.a-star.edu.sg

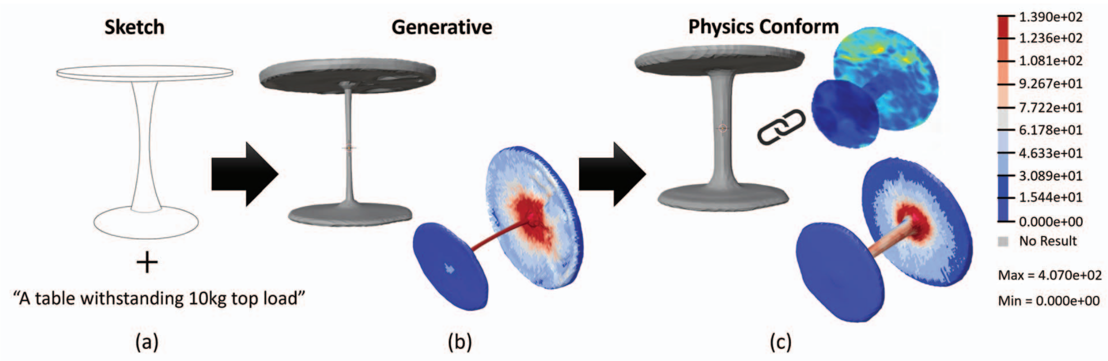


Fig. 1: We present the study of end-to-end prompt-vision-physics neural network (PVPNN) to discover novel physics-conform designs using a joint optimization strategy. Based on the user's visual and physics preferences (Fig. 1(a)), the design generated using a conventional computer vision technique has an unnatural shape (Fig. 1(b)). We observed uneven stress distribution and high concentration (deep red regions) at the table leg and joint when validated with finite element method simulation. In contrast, our proposed method learns the target object's physics and can discover a viable design (Fig. 1(c)) that has a realistic appearance and yet adheres to visual and physics preferences.

***Abstract*—In this paper, we present the study of an end-to-end prompt-vision-physics neural network (PVPNN) to speed up the design discovery process in science and engineering. PVPNN has the potential to transform design discovery by enabling researchers and engineers to specify design requirements seamlessly via prompts, analyze available large-scale relevant historical data sets to generate novel physics-conform designs and predict the efficacy of potential design candidates rapidly. Starting from a given image or prompt for an ideal design as the baseline, our proposed framework first generates potential designs from a vision foundation model. Based on the practicality requirements provided by the user through text prompt, the PVPNN initializes a resource-efficient physics-informed neural network and jointly optimizes both geometry and physics models in an end-to-end mode. An experimental study on a simple engineering structure validates that our proposed framework can seamlessly satisfy visual preferences and practicality for fast design discovery.**

***Index Terms*—AI for design discovery, prompt-vision-physics neural network, PVPNN, design discovery, computer vision, engineering design**

I. Introduction

Recent advances in using artificial intelligence (AI) to speed up scientific discovery have garnered significant interest to drive further progress across diverse domains [1]–[3]. A substantial body of research in the field of engineering has been dedicated to discovering novel structures or geometries that fulfill specific design criteria [4]. Both the demands of researchers and engineers and the wealth of accumulated data over the past decades are propelling and substantiating the exploration of AI for design discovery.

In traditional engineering practices, the design discovery process typically entails a prolonged cycle and collaboration among people from diverse fields. Initially, designers draw upon past design experiences to create a design draft. Subsequently, engineers translate these drafts into 3D models and subject them to time-consuming computer simulations [5] to analyze the physical characteristics of the design. In response to simulation results, engineers iteratively refine the design through multiple rounds of adjustments and simulations [6], thus to better align with practical design requirements. This thorough design cycle demands a significant investment of time, ranging from days to months, and substantial resources, encompassing both human and material assets. Recently, with the wealth of historical design data and the rapid evolution of computer vision, generative AI models can now produce

[*]Equal Contribution

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00255

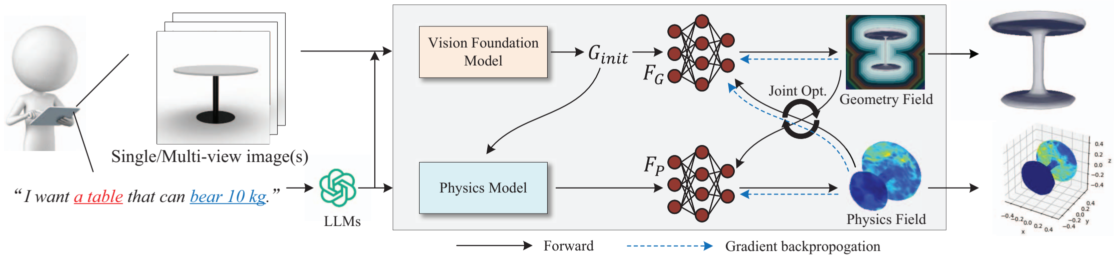


Fig. 2: Our proposed prompt-vision-physics neural network comprises a vision foundation model that synthesizes the initial design and a physics model that defines the precise physics of the design. We use a joint-optimization strategy, conditioned with the user's visual and physics preferences provided as images or prompts via LLMs, to co-train the geometry of the design and the physics-conform network, modifying the design towards a more physics-conform variant.

artistic designs with simple prompts [7]–[11]. Additionally, the evolution of physics-informed machine learning showcases its potential for enhanced physical simulation [12]. Together, these advancements lay the groundwork for an end-to-end design discovery framework, enabling the completion of complex design cycles with minimal prompts. Such a framework will provide ease of design specification and is anticipated to significantly reduce the traditional design cycle to hours or even minutes, requiring only minimal resources.

To realize the overarching vision described above, we have embarked on a preliminary endeavor towards an end-to-end design discovery framework, introducing a novel architecture known as the prompt-vision-physics neural network (PVPNN). Starting from a given image or prompt, PVPNN first synthesizes an initial neural geometry model using vision methods. Then, according to practicality requirements, our proposed method initializes a physics-conform network to estimate the physical information (e.g., stress fields) [12]. On this basis, we propose jointly optimizing the neural geometry and physics models end-to-end, considering visual preferences and practical design requirements. We leverage the differentiability capability of the physics-conform network to our framework, circumventing the need for employing black-box optimizers or the intricate design of a differentiable physical simulation process. PVPNN has the following strengths: (1) *Flexibility*. Our proposed technique can flexibly discover designs without relying on extensive expertise; (2) *Visual preferences*. PVPNN can synthesize realistic designs by inheriting the advantages of vision methods; (3) *Physics Conformance*. PVPNN can discover designs that align with physical principles and satisfy the physics preferences with a certain degree of flexibility through optimization, supporting real-world usage needs.

In summary, we proposed two contributions in this paper: (1) an end-to-end framework for design discovery, which can flexibly discover physics-conform designs from given images or prompts, and (2) a joint optimization framework of geometry and physics models integrates geometry and physics models in a fully differentiable way.

II. End-to-End Prompt-Vision-Physics Neural Network

We envision a potential synergy with the power of large language models, computer vision, and engineering design techniques to discover novel designs that simultaneously adhere to visual and physics preferences. To this end, we conceptualize the idea that we define as follows:

Definition 1 (Prompt-Vision-Physics Neural Network (PVPNN))**.** *Prompt-vision-physics neural network combines visual information and physical concepts to discover novel designs that best satisfy the visual and physics preferences implied by the prompt.*

A. Overall Framework

Fig. 2 presents the overview of our proposed end-to-end framework. Given single/multi-view image(s) or a prompt for an ideal design, we aim to discover novel designs that satisfy visual preferences and practicality simultaneously. This can be decomposed into the following three objectives:

- Geometry conforms to visual preference.
- Ensuring the physics model can provide accurate predictions for the physics involved.
- Geometry aligns with specific design specifications of engineering implications.

To achieve the above goal, our proposed framework in Fig. 2 includes the following modules:

- Large Language Model (LLM) for text prompt: LLM represent users' specifications that are compatible with our framework, facilitating the subsequent vision and physics processing.
- Vision Foundation Model (VFM) for 3D design discovery: VFM synthesizes a 3D model that satisfies the images or text prompt to provide a rough design prototype (G_{init}). This prototype is used to initialize the 3D geometry model (F_G) and the differentiable physics model (F_P).
- Physics Model: Using the physics specification and the rough design prototype (G_{init}), the physics model computes the necessary physics information for initializing the differentiable physics model (F_P).

Once F_G and F_P are initialized, our proposed technique jointly optimizes the geometry and physics fields in a fully differentiable way and outputs the final novel design.

The representations of geometry and physics should be continuous and differentiable to allow end-to-end joint optimization. Therefore, two multi-layer perceptrons (MLPs) are employed to approximate the signed distance function (SDF) of the geometry and the physical fields, respectively.

B. Joint Optimization of Geometry and Physics

After the geometry and physics initialization, we can obtain the realistic initial geometry field. However, in the physics initialization, the computed force distribution of the initial geometry is usually not optimal according to the physics requirements. To satisfy the design visual preferences and practicality at the same time, we relax the geometry constraint and strictly impose the physics constraint to optimize the geometry and physics fields jointly. Specifically, to relax the geometry constraint, based on the initial geometry field G_{init}, we generate an additional subspace S_G, which is located at a distance from the initial geometry surface. This subspace prevents the geometry field from deforming too much. Therefore, we define a relaxed geometry constraint as:

$$\mathcal{L}_G = \frac{1}{|S_G|} \sum_{p \in S_G} ||F_G(p) - G_{init}(p)||^2. \tag{1}$$

This constrains the shape deformation to a specific subspace, guaranteeing the visual preference requirements. In terms of the physics constraint, we should make sure that the solution provided by the physics-conform network is correct. Thus the following loss should be minimized:

$$\mathcal{L}_P = ||\mathcal{N}[F_P(p)]||^2_{p\in\Omega} + ||\mathcal{B}[F_P(p)]||^2_{p\in\partial\Omega}, \tag{2}$$

where $F_P(p)$ is the approximation of the physics-conform network, and $\mathcal{N}[\cdot]$ and $\mathcal{B}[\cdot]$ are the operators used to form the following governing functions:

$$\begin{aligned} \mathcal{N}[f(p)] &= 0, p \in \Omega, \\ \mathcal{B}[f(p)] &= 0, p \in \partial\Omega. \end{aligned} \tag{3}$$

Note that Eq. (3) records a set of governing functions (e.g., linear elastic solid mechanics), Ω contains all of the points inside the geometry, $\partial\Omega$ includes all of the points on the boundary of the geometry, and $f(p)$ is the solution of the governing functions. Moreover, to ensure that the geometry satisfies the required engineering design objective(s), the following objective function should be optimized (a minimization problem is considered here), i.e.,

$$\mathcal{L}_D = \mathcal{D}[F_P(p)], \tag{4}$$

where $\mathcal{D}[\cdot]$ is an operator corresponding to the engineering requirement.

In minimizing the maximum stress of a geometry under a specific force (to withstand loads better), $\mathcal{D}[\cdot]$ denotes the maximum stress operator. Other objectives, such as the average stress in the geometry, can also be considered.

Without loss of generality, our joint optimization loss can be defined as:

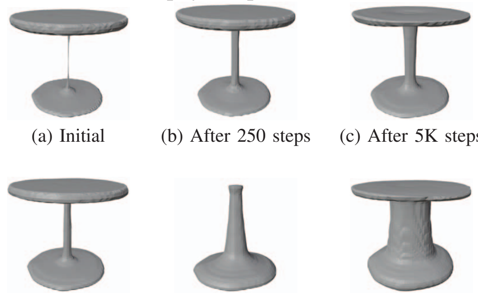
$$\mathcal{L} = \lambda_G \cdot \mathcal{L}_G + \lambda_P \cdot \mathcal{L}_P + \lambda_D \cdot \mathcal{L}_D, \tag{5}$$

where λ_G, λ_P and λ_D are weights to balance different terms.

III. Experiments

We assume the geometry of the design follows the linear elastic material and conducted a two-phase training process using Nvidia V100 GPU (32GB). In the first phase, we obtain the boundary conditions from an initial design, discovered using a vision foundation model [8]. We then pre-train the geometry and physics-conform neural network, using Eq. (1) and Eq. (2), respectively, to extract the 3D representation and initialize the physics model. In the final phase, the design is optimized by simultaneously training both the geometry network and the physics-conform neural network, where the loss function in Eq. (5) is used.

To evaluate our proposed method, we provide PVPNN an image and a prompt "A table withstanding 10kg top load," to discover viable designs for a table, as illustrated in Fig. 1(a). Fig. 1(b) showcases the 3D design discovered by the vision foundation model. We validated the design using FEM simulation to observe its stress distribution, as shown in the color shades plots in Fig. 1. Note that light-color shaded regions indicate areas with high-stress concentrations on the geometry. Hence, these hotspots are more susceptible to integrity failure. Notice that while the discovered design appears viable, its corresponding stress distribution indicates that this design is infeasible. Specifically, the tabletop and joint are more fragile than other table parts, making it less durable. On the other hand, after optimizing with our proposed framework, as shown in Fig. 1(c), the thickness of the table joints and leg has increased, resulting in a much lower stress distribution. This demonstrates that PVPNN can discover viable designs that conform to visual and physics preferences.



(a) Initial (b) After 250 steps (c) After 5K steps

(d) Without $\mathcal{L}_D$ (e) Without $\mathcal{L}_P$ (f) Without $\mathcal{L}_G$

Fig. 3: PVPNN is efficient in optimizing the design as shown in Fig. 3 (a)-(c). Futhermore, the effect of excluding design $\mathcal{L}_D$, physics $\mathcal{L}_P$, and geometry $\mathcal{L}_G$ components in our training objective (see Eq. (5)) has a profound effect on the design Fig. 3 (d)-(f).

Our investigation of the geometric evolution during optimization has shown in Fig. 3(a)-(c) that our proposed framework is efficient in strengthening the weak spots of the design and still conforms to the visual preference of a table.

We also assess the efficacy of the three proposed losses by selectively excluding each loss term, namely $\mathcal{L}_G$, $\mathcal{L}_P$, and $\mathcal{L}_D$. In Fig. 3(d), the absence of $\mathcal{L}_D$ resulted in a slight thinning of the table leg compared to the PVPNN with all of the losses. This deviation may compromise the load-bearing capacity of the table. The rationale behind this outcome lies

(a) Malformed base (b) Malformed top

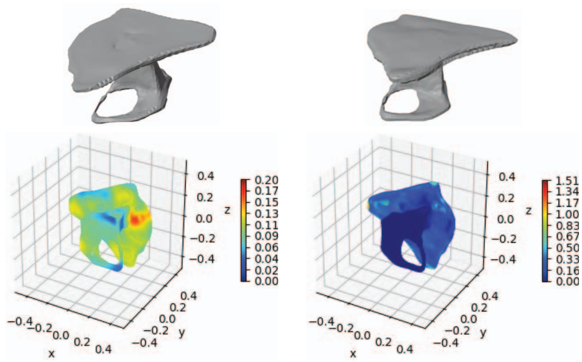
Fig. 4: A vision foundation model [9] with the prompt "A table with a teardrop shape" discovered a malformed design with tabletop holes and a fragile table base, and using this model alone is not enough to discover a feasible design.

in the absence of an objective, as $\mathcal{L}_D$ serves to guide the geometry toward satisfying the design requirements. Consequently, the geometry fails to optimize effectively in alignment with the design specifications. Similarly, Fig. 3(e) highlighted the consequences of omitting $\mathcal{L}_P$, wherein the absence of physics-related information hinders the optimization process. This lack of guidance can cause the geometry to deviate towards configurations that do not align with the intended design requirements, resulting in an unconventional geometry. Moreover, as demonstrated in Fig. 3(f), the exclusion of $\mathcal{L}_G$ markedly influences the geometry, restricting its ability to adhere to visual preferences. This deviation results in an unnatural shape, emphasizing the role of $\mathcal{L}_G$ in driving the geometry towards user-preferred designs.

Furthermore, we employed our proposed method to discover a novel teardrop-shaped table design. The design from a vision foundation model [9] as presented in Fig. 4 is infeasible. After initializing the design using our proposed technique, the initial design filled the small holes as shown in Fig. 5(a). However, moderate stress distribution is observed throughout the entire design in the physics field, and depending on the material, the narrow strip of the table base may suffer breakage. After optimization, while the peak stress is higher than the initial design, we have seen a significant reduction of stress at the highlighted potential breakage area (Fig. 5(b)). This result reiterates PVPNN's conformity to physics preferences to a certain degree. Nevertheless, this optimized design has a smoother surface with some uneven regions leveled and better fits the visual and physics preferences.

IV. Conclusion and Future Direction

We propose PVPNN for fast design discovery that optimizes geometry and physics models, seamlessly integrating the representation of computer vision and engineering design. PVPNN efficiently produces realistic and durable novel designs, catering to diverse user preferences and practical considerations. However, in our current framework, there is a notable absence of the incorporation of historical physical data in the training process. The data-driven differentiable physics model may be a promising technique to alleviate the limitation. This avenue represents a focal point for our future research efforts. We envision the limitless potential of our proposed framework, extending beyond the design discovery to scientific discovery [2], [3].



(a) Before Optimization (b) Optimized Design

Fig. 5: Before optimization, the initial design suffers moderate stress. Depending on its material, it has the potential to disintegrate at its base. After optimization, this area has a stress reduction and a much smoother appearance, rendering the design a better fit to user preferences. Note that the color bars are not to scale.

Acknowledgment

This research is partly supported by the Honda Research Institute Europe (HRIEU) and the College of Computing & Data Science (CCDS), Nanyang Technological University (NTU).

References

[1] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac *et al.*, "Scientific discovery in the age of artificial intelligence," *Nature*, vol. 620, no. 7972, pp. 47–60, 2023.

[2] H. Kitano, "Nobel turing challenge: creating the engine for scientific discovery," *npj Systems Biology and Applications*, vol. 7, no. 1, p. 29, 2021.

[3] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, "Autonomous chemical research with large language models," *Nature*, vol. 624, no. 7992, pp. 570–578, 2023.

[4] H. A. Eschenauer and N. Olhoff, "Topology optimization of continuum structures: a review," *Appl. Mech. Rev.*, vol. 54, no. 4, pp. 331–390, 2001.

[5] G. Dhatt, E. Lefrançois, and G. Touzot, *Finite element method.* John Wiley & Sons, 2012.

[6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, pp. 455–492, 1998.

[7] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," *ICLR*, 2023.

[8] R. Liu, R. Wu, B. Van Hoorick, P. Tokmakov, S. Zakharov, and C. Vondrick, "Zero-1-to-3: Zero-shot one image to 3d object," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9298–9309.

[9] M. Li, Y. Duan, J. Zhou, and J. Lu, "Diffusion-sdf: Text-to-shape via voxelized diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 642–12 651.

[10] M. Wong, Y. Ong, A. Gupta, K. Bali, and C. Chen, "Prompt evolution for generative ai: A classifier-guided approach," in *2023 IEEE Conference on Artificial Intelligence (CAI).* IEEE Computer Society, 2023, pp. 226–229. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CAI54212.2023.00105

[11] T. Rios, S. Menzel, and B. Sendhoff, "Large language and text-to-3d models for engineering design optimization," in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2023, pp. 1704–1711.

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