Learning with Limited Data

Guandao Yang

Artificial intelligence (AI) has achieved remarkable success in fields like computer vision and natural language processing, mostly enabled by large-scale training data. However, AI's transformative potential remains largely underexploited in areas without big data. This limitation is evident in fields involving 3D geometry, such as 3D computer vision, computer graphics, and physics simulation, because existing 3D datasets are still orders of magnitude smaller than their counterparts in vision and language domains. In contrast, these fields have developed rich mathematical models, such as differential geometry in graphics and governing equations in physics, without relying on large data. **My research philosophy centers on combining classical mathematical models with modern machine learning approaches to unlock AI's potential in data-constrained domains.**

Building on this philosophy, I study how to create AIs that can generate [1–8] and analyze [9–13] 3D spatial data. With backgrounds in computer vision, machine learning, and computer graphics, I have pioneered the paradigm shift toward using continuous neural fields to represent 3D geometry, transforming how 3D data is generated and processed. My works have garnered significant recognition and impacted other scientific disciplines such as chemistry [14] and physics [15].

Looking ahead, my long-term research goal is to empower different scientific disciplines to harness AI's potential through the synergy of modeling and learning approaches. I plan to build a lab to explore the fundamental principles of applied machine learning in data-constrained domains, from geometry-related areas to other scientific fields. My lab will study questions such as how to design data-efficient neural network architectures exploiting domain-specific properties [16] and how to apply machine learning to accomplish complex designs in graphics and engineering [6]. I look forward to initiating interdisciplinary collaborations to apply machine learning to empower different scientific and engineering applications like solving partial differential equations under limited observations [17].

1 Learning to Generate Spatial Data

Generating and editing 3D shapes is crucial in fields like computer vision, graphics, and mechanical engineering. Although traditional methods have provided tools for capturing real-world 3D shapes or creating digital ones, using these tools remains manual and time-consuming. Machine learning models have the potential to automate this process by directly predicting the desired 3D shape, yet their effectiveness is limited by the scarcity of readily usable 3D training data. My works address this issue by designing shape representations that can make use of the data and knowledge accumulated by the vision and graphic communities while remaining compatible with machine learning models.

3D point cloud generation. Point clouds are perhaps the most widely available source of 3D data, thanks to classical 3D reconstruction techniques. While generative fixed-dimensional tensors were commonly studied in machine learning, generating point clouds was still challenging due to the inherent irregularity, as 3D shapes are usually represented with different numbers of points. In PointFlow [1], we address this issue by modeling each shape as a continuous probability density field of 3D points. A point cloud can be viewed as a set of discrete samples from such a distribution, regardless of its cardinality. We further improve the efficiency of this representation in our ECCV 2020 paper, ShapeGF [2], by modeling only the gradient of the density fields, akin to the methods that later evolved into diffusion models. This representation can combine deep generative models with data created directly by traditional 3D vision and graphics methods. Using this framework, both PointFlow and ShapeGF achieved state-of-the-art performance in point cloud generation at their time. These two works are among the first successful instantiations of a new paradigm in which 3D generative models output shapes as continuous fields. My works on 3D shape generation have been cited over 1,000 times, and our code has received over 900 stars. Their impact has extended beyond generating 3D point clouds to scientific applications such as molecule generation [14].

Figure 1: *Left*: generation process of PointFlow [1]. *Middle*: complex shapes created by the efficient generation process of ShapeGF [2]. *Right*: discretization-free geometry processing using neural fields [3].

Evaluating 3D generative models. In addition to developing 3D generative models, I contributed to establishing efficient and reliable evaluation metrics for them. Perhaps the most reliable evaluation method is through humans, such as Turing tests. However, these tests can be costly or impractical to perform as they require users to review and process many 3D shapes. To address this issue, I propose using an "Automatic Turing Test" (ATT) [18], which replaces human evaluators with machine learning models. In PointFlow [1], we propose *1NN accuracy*, a metric using nearest-neighbor classifiers to perform ATT. *1NN accuracy* has become *the facto* metric to measure the geometry quality of 3D generative models. Other than geometry quality, downstream applications often consider many other criteria when evaluating a generative model. To scale ATT to different criteria, we leverage the commonsense reasoning capability of Large Multimodal Models like GPT-4V. In our CVPR 2024 paper [5], we successfully prompt GPT-4V to perform ATT and achieve evaluation results that align with human judgments. Our work provides the first automatic quantitative metric to evaluate text-to-3D generative models holistically using multiple criteria. In just half a year since its publication, our paper has been cited more than 50 times, both for its effectiveness in evaluating 3D generative models and for pioneering the idea of using foundation models to perform ATT in different modalities. ATT is also adapted to other tasks, including image captioning [19] and language-driven 3D editing [20].

Geometry processing. Processing and editing 3D geometries are as crucial as generating them. Applying learning-based algorithms to geometry processing tasks is not trivial due to the lack of large data for editing operations. While classical methods have established rich mathematical models for manipulating well-discretized polygon meshes, their performance can be compromised by discretization errors caused by the meshing and remeshing steps in the pipeline. My NeurIPS 2021 paper [3] circumvents this issue by proposing a geometry processing pipeline based on neural fields, a novel representation encoding shapes as continuous fields parameterized by neural networks. My work formulates each geometry processing operation as finding a neural field that optimizes objectives derived from traditional mathematical models, such as thin-shell energy [21]. My proposed pipeline avoids discretization errors and supports many editing operations where data is expensive to collect, such as elastic deformation. I continued to work on extending the impact of neural field representation to other modalities, such as images [16], radiance fields [22], and fields that store physics quantities [11, 12]. My works in geometry processing have also inspired methods in applications like physics simulation [15, 23] and mechanical design [24].

2 Learning to Analyze Spatial Data

In addition to creating and editing shapes, it is also important for AI systems to understand the properties of 3D shapes. Directly applying learning-based methods here is challenging because it is difficult to obtain large-scale data with ground truth annotations for many properties, such as Young's modulus of a material. On the contrary, traditional methods predict these properties based on mathematical models without requiring large training data and labels. However, these classical models can be slow and not robust to noisy inputs like real-world images or 3D scans. My research integrates the strengths of classical approaches to create learning algorithms that can accurately and efficiently predict different shape properties.

Enchancing physics solvers with learning. Solving partial differential equations (PDEs) efficiently and accurately conditioned on a shape is an essential way to understand the physical properties of that shape. While Monte Carlo PDE solvers [25] have shown great promise in being robust to complex geometries, they can be expensive due to the large number of random samples needed to achieve the desired accuracy. We propose a novel PDE solver that combines the robustness of Monte Carlo methods with the

Figure 2: *Left:* fast and accurate PDE solution achieved by combining neural fields with Monte Carlo methods [11, 12]; *Middle:* physics-in-the-loop optimization method allows us to render realistic dressed avatars in novel poses and lighting [10]; *Right:* symmetry detection robust to noise [9].

speed of neural network [11, 12]. Our proposed solvers can produce accurate and high-resolution PDE solutions up to 20 times faster than existing methods.

Reconstructing material properties from visual observations. Preparing a 3D shape for physics simulation and rendering necessitates an understanding of its material properties. However, recovering these properties from real-world data, such as videos, presents a significant challenge, as it involves simultaneously solving for geometry, lighting, and tracking. Developing learning-based algorithms for this purpose is also difficult due to the lack of annotated data for material properties. In PhysAvatar [10], we address these challenges using a multi-stage optimization leveraging inductive biases from the physics simulator of loose garments and lighting. Our approach can recover plausible material properties directly from videos and achieve realistic renderings of the 3D avatar in unseen poses and under novel lighting conditions.

Symmetry detection. In addition to physical properties, it is also important to understand geometric properties, such as symmetry. Traditional modeling-based methods struggle when noise is present in the input 3D shapes, while learning-based approaches are hindered by the difficulty of annotating partial symmetries in scale. Our SIGGRAPH Asia 2024 paper [9] derives a theoretical connection between the traditional symmetry detection algorithms and the learning-based diffusion models from the machine learning community. We leverage this connection to develop a symmetry detector that is generalizable to unseen shapes and robust to different noise patterns.

3 Future Directions

In addition to advancing the abovementioned thrusts in 3D vision and graphics, I am eager to expand the scope of my research to other data-constrained fields in science and engineering. My interests include:

Data-efficient Network Architectures. Domain-specific data often possesses characteristics unique to its domain. For instance, 3D point clouds are typically sparse, and images usually exhibit self-similarities across scales. General network architectures without corresponding inductive biases, such as self-attention, can be inefficient when dealing with limited domain-specific data. This raises an important question: can we design more efficient network architectures by leveraging the data's known structure? In our NeurIPS 2022 paper [16], we demonstrated that leveraging the scale-space structure can lead to a novel class of efficient neural field architectures, highlighting the potential of integrating domain knowledge as inductive biases in neural networks. Moving forward, exploring how to incorporate insights from geometry and physics into network design remains an exciting direction.

AI for Engineering Design. Designing high-quality shapes that are both manufacturable and functional in engineering applications requires multi-step reasoning to achieve optimal results. Existing research usually emphasizes efficiency, using neural networks to directly output the final design [1, 2]. However, accomplishing design tasks with low error tolerance demands a "slow thinking" approach [26] – one that systematically explores multiple hypotheses before arriving at a conclusion. With BlenderAlchemy [6], we have taken the first step toward building such a system and showcased the advantages of "slow thinking" for graphical editing. Expanding this concept to other engineering design tasks, which may involve thousands of reasoning steps, can be a fruitful direction that requires insights from reinforcement learning, physical simulation, and foundation models. Breakthroughs in this direction can streamline and accelerate workflows in industries such as aerospace and mechanical engineering.

AI for Scientific Modeling. Building on the exploration of "slow thinking" systems for engineering design tasks, a similar paradigm can be applied to create mathematical models to facilitate scientific discovery, where reasoning over multiple hypotheses via simulations and experiments is critical. In my NeurIPS 2024 paper, I have begun such exploration by developing non-linear PDE solvers under limited observations [17]. I am excited to take a dive deep into scientific applications by initiating interdisciplinary projects collaborating with domain experts, aiming to develop AI systems tailored to these needs. I believe the knowledge gained through these collaborations has the potential to advance scientific discovery and address key challenges in the development of artificial general intelligence.

References

[1] **Guandao Yang***, Xun Huang*, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *ICCV*, 2019.

[2] Ruojin Cai*, **Guandao Yang***, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning gradient fields for shape generation. In *ECCV*, 2020.

[3] **Guandao Yang**, Serge Belongie, Bharath Hariharan, and Vladlen Koltun. Geometry processing with neural fields. In *NeurIPS*, 2021.

[4] **Guandao Yang**, Yin Cui, Serge Belongie, and Bharath Hariharan. Learning single-view 3d reconstruction with limited pose supervision. In *ECCV*, 2018.

[5] Tong Wu*, **Guandao Yang***, Zhibing Li*, Kai Zhang, Ziwei Liu, Leonidas Guibas, Dahua Lin, and Gordon Wetzstein. Gpt-4v(ision) is a human-aligned evaluator for text-to-3d generation. In *CVPR*, 2024.

[6] Ian Huang, **Guandao Yang**, and Leonidas Guibas. Blenderalchemy: Editing 3d graphics with vision-language models. In *ECCV*, 2024.

[7] Youming Deng*, Wenqi Xian*, **Guandao Yang**, Leonidas Guibas, Gordon Wetzstein, Steve Marschner, and Paul Debevec. Self-calibrating gaussian splatting. In Submission, 2024.

[8] Kiyohiro Nakayama, Jan Ackermann, Timur Kesdogan, Yang Zheng, Maria Korosteleva, Olga Sorkine-hornung, Leonidas Guibas, **Guandao Yang**, and Gordon Wetzstein. Aipparel: A large multimodal generative model for digital garments. In Submission, 2024.

[9] Jihyeon Je*, Jiayi Liu*, **Guandao Yang***, Boyang Deng*, Shengqu Cai, Gordon Wetzstein, Or Litany, and Leonidas Guibas. Robust symmetry detection via riemannian langevin dynamics. In *SIGGRAPH Asia*, 2024.

[10] Yang Zheng*, Qingqing Zhao*, **Guandao Yang**, Wang Yifan, Donglai Xiang, Florian Dubost, Dmitry Lagun, Thabo Beeler, Federico Tombari, Leonidas Guibas, et al. Physavatar: Learning the physics of dressed 3d avatars from visual observations. In *ECCV*, 2024.

[11] Zilu Li*, **Guandao Yang***, Xi Deng, Christopher De Sa, Bharath Hariharan, and Steve Marschner. Neural caches for monte carlo partial differential equation solvers. In *SIGGRAPH Asia*, 2023.

[12] Zilu Li*, **Guandao Yang***, Qingqing Zhao, Xi Deng, Leonidas Guibas, Bharath Hariharan, and Gordon Wetzstein. Neural control variates with automatic integration. In *SIGGRAPH*, 2024.

[13] Aditya Chetan, **Guandao Yang**, Zichen Wang, Steve Marschner, and Bharath Hariharan. Accurate differential operators for hybrid neural fields. *arXiv preprint*, 2023.

[14] Lemeng Wu, Chengyue Gong, Xingchao Liu, Mao Ye, and Qiang Liu. Diffusion-based molecule generation with informative prior bridges. In *NeurIPS*, 2022.

[15] Honglin Chen, Rundi Wu, Eitan Grinspun, Changxi Zheng, and Peter Yichen Chen. Implicit neural spatial representations for time-dependent pdes. In *ICML*, 2023.

[16] **Guandao Yang***, Sagie Benaim*, Varun Jampani, Kyle Genova, Jonathan Barron, Thomas Funkhouser, Bharath Hariharan, and Serge Belongie. Polynomial neural fields for subband decomposition and manipulation. In *NeurIPS*, 2022.

[17] Jiahe Huang, **Guandao Yang**, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. In *NeurIPS*, 2024.

[18] Ryan Lowe, Michael Noseworthy, Iulian Vlad Serban, Nicolas Angelard-Gontier, Yoshua Bengio, and Joelle Pineau. Towards an automatic Turing test: Learning to evaluate dialogue responses. In *ACL*, 2017.

[19] Yin Cui, **Guandao Yang**, Andreas Veit, Xun Huang, and Serge Belongie. Learning to evaluate image captioning. In *CVPR*, 2018.

[20] Hansheng Chen, Bokui Shen, Yulin Liu, Ruoxi Shi, Linqi Zhou, Connor Z Lin, Jiayuan Gu, Hao Su, Gordon Wetzstein, and Leonidas Guibas. 3d-adapter: Geometry-consistent multi-view diffusion for high-quality 3d generation. *arXiv preprint*, 2024.

[21] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4, pages 109–116. Citeseer, 2007.

[22] Mikaela Angelina Uy, Kiyohiro Nakayama, **Guandao Yang**, Rahul Thomas, Leonidas J Guibas, and Ke Li. Nerf revisited: Fixing quadrature instability in volume rendering. In *NeurIPS*, 2024.

[23] Zeshun Zong, Xuan Li, Minchen Li, Maurizio M Chiaramonte, Wojciech Matusik, Eitan Grinspun, Kevin Carlberg, Chenfanfu Jiang, and Peter Yichen Chen. Neural stress fields for reduced-order elastoplasticity and fracture. In *SIGGRAPH Asia*, 2023.

[24] Arturs Berzins, Andreas Radler, Sebastian Sanokowski, Sepp Hochreiter, and Johannes Brandstetter. Geometry-informed neural networks. *arXiv preprint*, 2024.

[25] Rohan Sawhney and Keenan Crane. Monte carlo geometry processing: A grid-free approach to pde-based methods on volumetric domains. *ACM ToG*, 2020.

[26] Daniel Kahneman. *Thinking, Fast and Slow*. 2017.

* denotes equal contribution.