Generative AI and Minimalist Sculpture

Kavish Kondap[1], Andy Bass[1*], Elle O Hill[1*], Paloma Raffle[1*], Andrew Goldberg[1], Tianshuang Qiu[1], Zehan Ma[1], Letian Fu[1], Justin Kerr[1], Huang Huang[1], Kaiyuan Chen[1], Kuan Fang[2], Ken Goldberg[1]

Full Paper and Videos:
https://bloxnet.org/art/

***Abstract*— Art and sculpture convey human creativity; representational and abstract forms date back over 30,000 years. In this project, we explore how a large Vision Language Model (VLM) can be used to generate novel 3D sculptural compositions and specifically whether VLMs can design recognizable representational sculptures using a limited set of physical components. In a recent paper accepted to ICRA 2025 [1], we formalize Generative Design-for-Robot-Assembly (GDfRA): given a natural language prompt (e.g., "giraffe") and an image of available parts, the task is to generate a spatial assembly that visually resembles the prompt. In this paper, we apply Blox-Net to the task of generative sculptural assembly, exploring its ability to interpret and recreate iconic works of art. In particular, we examine three case studies: Frédéric Auguste Bartholdi's *Statue of Liberty*, Auguste Rodin's *The Thinker*, Tony Smith's *Cigarette*, and Christo and Jeanne-Claude's *The Gates*.**

I. Introduction

Art and technology have been linked throughout human history, from the earliest representational sculptures carved over 30,000 years ago to contemporary pieces. Sculpture transitioned from strict realism to impressionism, abstraction, symbolism, and conceptualism. Concurrently, recent advancements in artificial intelligence have shown that Vision Language Models (VLMs) can recognize images of art historical paintings and sculptures, describe complex scenes, and even generate new compositions that blur the boundaries between human and machine creativity.

In a paper accepted to ICRA 2025 [1], we introduced Blox-Net–a novel system capable of making creative decisions in the Design for Robot Assembly (DfRA) process. Unlike traditional DfRA systems that require human designers in the loop [2–4], Blox-Net autonomously designs recognizable sculptural forms using a limited set of physical building blocks, and assembles them with a 6-DoF robot arm equipped with a suction gripper. In this paper, we apply Blox-Net to the task of generative sculptural assembly, exploring its ability to interpret and recreate 4 iconic works of art.

Recent advances in Generative AI systems have demonstrated remarkable abilities to create novel texts, code, and images [5–7]. Researchers are actively exploring "text-to-video" [8–10] and "text-to-3D" [11–13] systems, where the latter generates 3D mesh structures from textual descriptions (and there are ongoing research efforts applying Gen AI for eCAD design of chips [14]). This suggests that Generative AI may have potential for DfRA, and that, if coupled with a physical robot, it may be possible in certain cases to fully automate the design cycle.

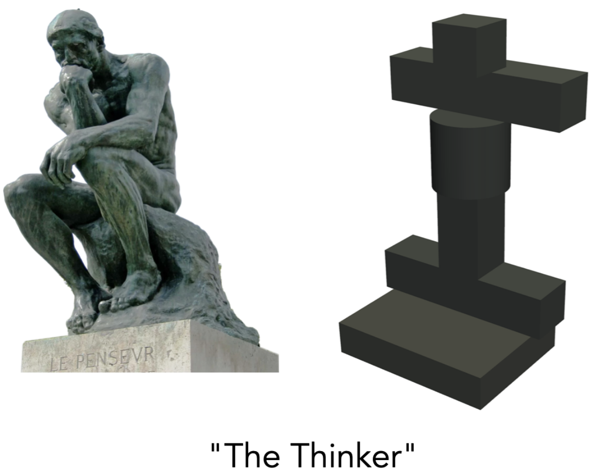

Fig. 1: We apply Blox-Net to the task of generating block-based renditions of famous sculptures, such as 'The Thinker'.

Blox-Net, a fully-implemented *generative* DfRA (GDfRA) system, utilizes the semantic planning and text generation capabilities of large language models (LLM) with physical analysis from a simulator. Blox-Net operates in three phases: 1) A customized iterative prompting process uses a vision language model (VLM) to design feasible 3D arrangements of available components—assemblies—that approximate the shape of a desired object (eg "giraffe"); 2) The proposed assemblies are tested in simulation to assess their physical constructability by a robot. Perturbation analysis identifies weaknesses, and revises a selected assembly as needed. 3) Using computer vision and motion planning a physical robot with a camera repeatedly builds a selected assembly, automatically resetting between trials. This process automatically evaluates the reliability of the physical construction.

We study how Blox-Net engages with artworks that embody complex cultural and emotional significance. In particular, we examine four case studies: Frédéric Auguste Bartholdi's *Statue of Liberty*, Auguste Rodin's *The Thinker*, Tony Smith's *Cigarette*, and Christo and Jeanne-Claude's *The Gates*. Each presents unique challenges for abstraction, embodiment, and narrative retention in a robotic system constrained by discrete, geometric building blocks.

[1]The AUTOLab at UC Berkeley, [2]Cornell University
Emails: {kavish_kondap, andybass, elleohill, palomaraffle, apgoldberg, ethantqiu, zehanma, max.fu.letian, justin_kerr, huangr, kych, goldberg}@berkeley.edu, {kuanfang}@cornell.edu
AUTOLab Website: https://autolab.berkeley.edu/

These case studies raise broader questions at the intersection of art, robotics, and AI: How do machines reinterpret human creativity under physical and algorithmic constraints? What is lost, transformed, or revealed in the process of translation?

II. Related Work

A. Design for Robot Assembly

The concept of Design for Assembly (DfA) was pioneered by Geoffrey Boothroyd and Peter Dewhurst in the early 1980s [15], with Hitachi developing its Assemblability Evaluation Method (AEM) in 1986 [16]. These seminal works laid the foundation for systematic approaches that follow product design guidelines [17] facilitate facilitate efficient assembly processes. As robotics automation in manufacturing became prevalent, Design for Robot Assembly (DfRA) emerged as an extension of DfA principles, specifically addressing the unique capabilities and limitations of robotic systems in assembly tasks [18, 19].

Design for Robot Assembly (DfRA) [18, 20–22] has evolved significantly with the advent of Computer-Aided Design (CAD) and Computer-Aided Manufacturing (CAM) software, which expedite design and evaluation of components and assemblies using Finite Element Methods and perturbation analysis [2–4, 23, 24]. While these tools facilitate visualization and analysis of tolerances, stresses, and forces, all existing DfRA systems require extensive human input [2–4]. A persistent challenge in DfRA is accurately modeling assembly reliability, given the inherent uncertainties in perception, control, and physics [25–30]. Simulation can partially address this, but struggles to capture 3D deformations and collisions crucial to robot grasping, necessitating iterative real-world testing and redesign [31–36]. Recent advancements leverage large language models (LLMs) [6, 37] for various aspects of design, including task planning, robot code generation [38, 39], engineering documentation understanding [40], and generating planar layouts or CAD models [14, 41–43]. However, these methods primarily focus on determining assembly sequences for fixed designs. In contrast, we tackle both the design and execution aspects of robot assembly, aiming to create physically feasible designs for robotic assembly with minimal human supervision.

B. Text-to-Shape Generation

Semantic generation of 3D shapes and structures is a long-standing problem in computer vision and computer graphics [44]. Deep generative models have enabled a wide range of approaches that learn to capture the distribution of realistic 3D shapes, in the format of voxel maps [45], meshes [46], point clouds [47], sign distance functions [48], CAD models [49], and implicit representations [50]. A growing number of approaches have also been proposed to generate objects and environments to create digital twins or curricula for robotic control and scale up learning [51–58]. With the advances of aligned text-image representations and vision-language models, an increasing number of works have aimed to generate semantically meaningful shapes specified by natural language instructions [11, 59–61]. Unlike these methods, Blox-Net generates 3D shapes using only the available physical building blocks by prompting an LLM (ChatGPT 4o [62]) to generate a plan to assemble the blocks into the desired shape.

C. Art and Generative AI

Advances in generative models have enabled artists to integrate AI into creative workflows. AI systems can assist traditional artistic practices and create new forms of generative, autonomous, and interactive art. Tools like DALL-E [63], Midjourney, and Sora [64] allow users to generate images, animations, and environments from natural language prompts. Several notable artworks have deployed robots as fabricators, performers, or autonomous agents in galleries and public spaces. Catie Cuan choreographs and performs alongside robotic systems, exploring the interaction between human movement and machine autonomy [65]. Sun Yuan and Peng Yu's *Can't Help Myself* at the Guggenheim used a large industrial robot tasked with a futile, repetitive cleaning task, to produce commentary on labor, control, and autonomy [66]. Patrick Tresset builds semi-autonomous robots that mimic human drawing behavior, prompting reflection on whether artistic value lies in creating these automatons or in the resulting drawings [67]. AI technologies have been employed by artists in a variety of contexts to create immersive and visual works that explore our evolving relationship between data and material form. Refik Anadol creates AI-generated data sculptures and immersive environments, processing massive datasets into machine-driven aesthetic experiences [68].

III. GDfRA Problem

Generative Design for Robot Assembly (GDfRA) considers the design of a 3D structure that can be assembled with an industrial robot arm (see Figure 1). The input is a word or phrase (e.g., *"bridge"*) and an image of available components for assembly. The objective of a GDfRA system is to design an assembly that is (1) "recognizable", meaning the assembly visually resembles the provided text input and (2) "constructible" from a given set of minimalist blocks.

IV. Method

We utilize Blox-Net, a GDfRA system that assumes (1) components are cuboids and cylinders and (2) components are lying in stable poses within a reachable planar area.

Blox-Net includes three phases. In phase I (Figure 2), Blox-Net prompts a VLM (GPT-4o [62]) to generate multiple assembly designs, from which the VLM selects the top candidate based on stability and visual fidelity. In phase II, the chosen assembly design undergoes an iterative refinement process in a customized physics simulator. This simulation-based approach applies controlled perturbations to enhance the design's constructability while maintaining its core characteristics. Phase II and phase III are not used in this work. In phase III, Blox-Net utilizes a robot arm equipped with a wrist-mounted stereo camera and suction gripper to construct

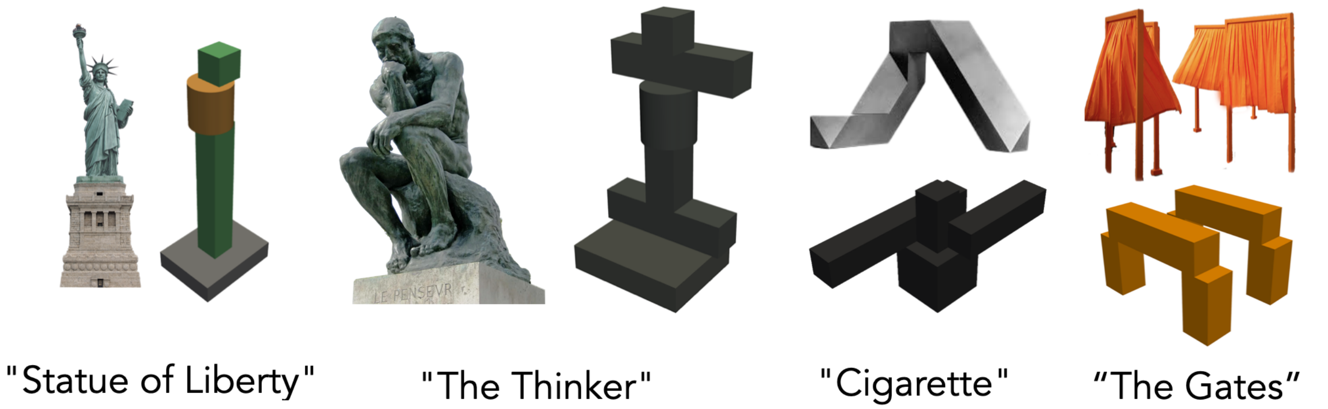

Fig. 2: **Case Study Generations:** From left to right: Statue of Liberty by Frédéric Auguste Bartholdi, The Thinker by Auguste Rodin, Cigarette by Tony Smith, and The Gates by Christo and Jeanne-Claude. For each artwork, the original sculpture is shown alongside the Blox-Net interpretation.

the optimized design using 3D printed blocks. The assembly is constructed on a tilt plate, which the robot actuates to automatically reset the blocks back into a tray.

V. Case Studies

Artists have continually redefined sculpture by embracing new tools, materials, and techniques. Today, AI and robotic systems introduces unprecedented possibilities into this ongoing evolution, challenging our understanding of creativity, interpretation, and authorship. By reconstructing renowned sculptures using Blox-Net, we explore how familiar forms can be reimagined through the lens of algorithmic design and robotic assembly. In the following section, we examine four iconic works—Frédéric Auguste Bartholdi's *Statue of Liberty*, Auguste Rodin's *The Thinker*, Tony Smith's *Cigarette*, and Christo and Jeanne-Claude's *The Gates*—to explore how Blox-Net reshapes both the aesthetic qualities and cultural meanings of historically significant sculptures in the context of emerging technologies.

A. Statue of Liberty - Frédéric Auguste Bartholdi, 1876

The Statue of Liberty, designed by Frédéric Auguste Bartholdi and gifted by France to the U.S. in 1886, has endured as one of the most charged symbols in American visual culture. From the torch she holds to the broken chains below her robe, the monument was built to both welcome immigrants and declare a national ideal of refuge and freedom [69]. Over time, artists have reimagined the statue to engage with and question that ideal. Andy Warhol's 1986 *Statue of Liberty* series turns her into a pop symbol, repeating its image in layered ghost tones [70], while JR's *Inside Out: Ellis Island* overlays the statue's pedestal with immigrant faces, grounding abstraction in lived experience [71]. Today, the statue is found in augmented reality filters, AI image generators, and viral media—its presence is preserved, yet its meaning is increasingly fractured. These reinterpretations mirror a broader cultural shift: the transformation of complex political symbols into consumable visuals, inviting reflection on what is remembered, what is erased, and what remains remixable [72].

Blox-Net's minimalist reconstruction of the Statue of Liberty continues this trend. In its rendering of the Statue of Liberty, small details fall away. The spiked crown becomes a flat green plane. The raised arm is reduced to a vertical column. A golden hemisphere, loosely referencing the torch, sits atop a minimalist frame. While the form signals the statue's identity, it also flattens its meaning. What remains is a stylized echo of the statue, stripped of the emotional and historical density that once made it monumental. Yet this simplification is the point. Blox-Net's version, assembled from data scraped from billions of inputs, becomes a mirror of society's own abstractions. When it rebuilds the Statue of Liberty, it compiles what we've culturally encoded: a structure recognized globally but often emptied of its original political meaning. Speculative renderings by models like DALL·E [63]—where Liberty is imagined underwater, made of candy, or transplanted into dystopian cityscapes—underscore this trend toward stylized, low-resolution symbolism. Like other digital representations, it reveals how contemporary society trivializes complexity, turning the ongoing story of immigration into a caricature of "welcome" while real borders harden and debates polarize. Blox-Net's Statue of Liberty becomes both object and critique, a minimalist artifact reflecting our current political upheaval.

B. The Thinker - Rodin, 1880

First conceived in 1880 and recast numerous times in bronze due to its popularity, Auguste Rodin's *The Thinker* quickly became a universal symbol of human cognition. During the time of *The Thinker*'s creation, Western philosophy and art had long separated the mind from the body [73]. Intelligence was seen as elite, exclusive, and disembodied. Moreover, the late 1800s were the height of industrialization. Rodin prioritized hand labor and outwardly criticized works "created by means of industry rather than by art," claiming they "do not withstand the march of time" [74]. As human labor was increasingly mechanized, the human body was both glorified in nationalist propaganda and dehumanized by factory work.

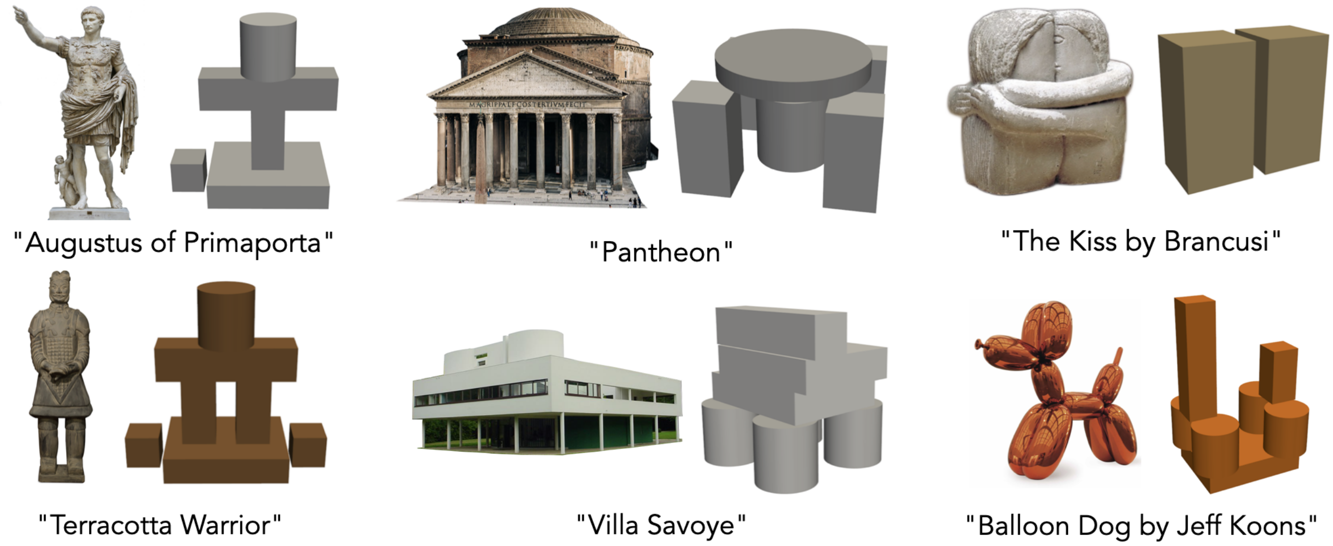

Fig. 3: **Additional Blox-Net Sculptures:** Blox-Net's interpretations of recognizable artworks and architecture.

Influenced by Romanticism and an emerging focus on the individual psyche, Rodin's innovation lay in making the body a language of its own. Whereas classical sculpture idealized the human form, Rodin was interested in its raw specificity [73]. Rodin himself explained, "what makes my *Thinker* think is that he thinks not only with his brain, with his knitted brow, his distended nostrils and compressed lips, but with every muscle of his arms, back, and legs, with his clenched fist and gripping toes" [75]. By emphasizing the physicality of thought, Rodin reclaims the human body as a site of meaning, emotion, and intelligence. One newspaper of the time wrote that the statue represented "the ordinary workman, anonymous, unknown," elevated to the level of "the egalitarian society" [76]. Rodin's *The Thinker* positioned intellectual depth not as the privilege of elites, but as a universal human trait.

Given only 21 blocks, Blox-Net reconstructs *The Thinker* and strips away the immense detail that Rodin used to democratize the labor of intelligence. It reduces the intricate form to a minimalist homage, acknowledging how AI reduces thinking to a disembodied and abstract process.

This version of *The Thinker* also shows how AI allows anyone to create sculptural forms. Where traditional sculpture demands immense skill, time, and physical labor, new technologies allow broader participation in both art and intellectual expression. This tension between embodiment and abstraction defines our new era of art, AI, and robotics. Ultimately, Blox-Net's *Thinker* reflects what we gain and what we sacrifice in separating thinking from the human body. As the nature of thought evolves with technology, Rodin's universal pose of thinking endures as a powerful symbol of human intelligence and reflection.

C. Cigarette - Tony Smith, 1961

Tony Smith began his creative life as an architect, studying under Frank Lloyd Wright and designing over twenty private homes before turning to sculpture in the early 1960s [77–79]. His approach to space and structure was shaped not only by formal training but also by personal experience. As a child, he lived in isolation due to tuberculosis in a prefabricated hut built by his father [77, 79]. These early experiences with confinement, modular construction, and solitude later echoed through his sculptural vocabulary. When he began working in three dimensions, Smith rejected the traditional language of sculpture and described his works as "presences," forms that asserted themselves in space with quiet, psychological intensity [77].

A pivotal moment in Smith's transformation came in 1951 during a night drive down the unfinished New Jersey Turnpike. The road, unmarked and unlit, was surrounded by industrial structures and open darkness. Smith later described this environment as "a reality not yet aestheticized" [77]. This experience revealed the possibility of form without interpretation, of physical presence stripped of symbolic narrative. From that point forward, Smith sought to create works that embodied structure, volume, and void without relying on ornament or traditional meaning. Using steel, plywood, and automotive paint, Smith helped define postwar Minimalism while infusing it with interiority and unresolved mystery [78, 79].

Tony Smith's *Cigarette* sculpture (1961) is a memento-mori that captures this cultural shift not through narrative but through form. It is a bent triangular black steel tube, monumental in scale yet conceptually abject [80]. Smith described the piece as "a Cigarette from which one puff had been taken before it was ground in the ashtray" [77].

Blox-Net's reconstruction of *Cigarette* invites reflection on how AI may reframe cultural beliefs about mortality. Rebuilding *Cigarette* with uniform blocks strips away the quiet decay and physical subtlety that Tony Smith intended in the original form. It reduces the nuanced gesture of monumental discard into a series of simplified blocks, assembled based on algorithmic priorities of stability and recognizability.

Smith's original sculpture captured the aftermath of smoking, a consumption, a physical presence laden with traces of human use and abandonment. Like the skulls and extinguished candles of vanitas paintings, *Cigarette* serves as a formal reminder of mortality, presenting death not through narrative but through lingering presence. Algorithms that cannot die recreate forms to remind us that we do. Artificial intelligence cannot grieve, but it can emulate the structural logic of grief as envisioned by Tony Smith. As Blox-Net reconstructs *Cigarette*, it points to the hopes of machine immortality.

D. The Gates – Christo and Jeanne-Claude, 2005

The Gates, created by Christo and Jeanne-Claude in February 2005, remains one of the most ambitious ephemeral art installations ever constructed [81]. Spanning 23 miles of walkways in Central Park, it consisted of 7,503 free-hanging saffron fabric panels attached to vinyl gates. Though visually monumental, the installation lasted only sixteen days before all materials were dismantled and recycled. This impermanence was a hallmark of Christo and Jeanne-Claude's work, as they use the temporary nature of the projects "to endow the works of art with a feeling of urgency to be seen, and the love and tenderness brought by the fact that they will not last" [82].

In a cultural moment obsessed with permanence and archives, *The Gates* asked people to show up, move through, and let go. This focus on presence, temporality, and environment made the experience deeply human. But in a digital age, these qualities are being reinterpreted through new tools. Projects like The Shed's AR overlays of *The Gates* [83] and open-source 3D-printed miniatures on platforms like Thingiverse allow for re-engagement with the piece far beyond its lifespan. Blox-Net continues this trajectory—rendering *The Gates* as bright orange modular blocks assembled into arch-like structures. These reinterpretations cannot replicate the tactile, participatory richness of the original, but they enable new access points. This marks the "democratization of creativity"—transforming large-scale, permission-based public works into something reproducible, scalable, and globally shareable.

Recreating *The Gates* through Blox-Net raises new questions about how advanced machine learning models interpret ephemeral, fluid works. While *The Thinker* by Rodin grapples with anatomical specificity and emotional intensity, *The Gates* tests whether a generative AI system can convey "transience" or "flow" using static, rigid blocks. In the Blox-Net renditions, the saffron sheets become flat rectangular blocks or open gaps, and the sense of windblown fabric is reduced to simplified geometric frames. This formal reduction recalls the work of Minimalist sculptors like Donald Judd, whose *Untitled (Stack)* (1967) [84] repeats industrial units to emphasize structure over narrative, or Carl Andre's *Lever* (1966) [85], which lays 137 firebricks in a line to challenge the boundary between art and floor.

In this context, minimalism becomes metaphor: Blox-Net's geometric translation reflects how our culture tends to distill complexity into structure, flattening participatory experiences into consumable forms. However, the real critique is not of the AI. Blox-Net does not erase meaning—it reveals what society has already streamlined. Its outputs mirror the ways we already reduce civic discourse to slogans, protest to hashtags, and migration to policy soundbites. As D. Maria-Reina notes, the original *Gates* relied on impermanence to create urgency and tenderness; what happens when we recreate it endlessly and easily?

And yet, this very accessibility is also where AI's potential lies. Systems like Blox-Net contribute to the "democratization of creativity"—lowering barriers for participation, making once-temporary installations available to students, educators, and independent creators. Projects like 3D-printable mini-*Gates* on Thingiverse or speculative AI-generated versions using DALL·E show how open tools can carry forward the form, rhythm, and visual logic of the original work, even if the wind and crowd are absent [86]. Blox-Net's rendition of the *Gates* reassemble a site-specific moment into something portable and shareable, not to diminish its meaning, but to invite us to examine what we choose to replicate, what we let go, and what our abstractions say about us.

VI. Discussion and Conclusion

The results from Blox-Net prompt broader questions about the nature of creativity, authorship, and artistic expression in an age increasingly shaped by AI and robotics. What does it mean for a machine to "create" art? Traditional conceptions of art emphasize intention, emotion, and individual perspective. In contrast, Blox-Net generates sculptures by sampling from distributions, optimizing for stability and recognizability under physical constraints. While it lacks consciousness or subjective experience, Blox-Net's outputs meaningful interpretations, can be the subject of case studies. One intriguing direction is the evolution of

Fig. 4: **GPT Image 1 Generations:** Images generated using OpenAI's image generation model. The model was prompted with a modified version of the Blox-Net input. Note that these generations do not adhere to the provided blockset and don't account for gravitational stability.

style. Human artists often develop distinct personal styles shaped by influences, experiences, and evolving sensibilities over time. Could future VLM-driven systems like Blox-Net similarly evolve stylistic signatures? Current models primarily optimize for task performance, but incorporating objectives related to aesthetic variation, historical styles, or novel formal experimentation could enable systems to develop novel "machine styles".

Furthermore, today's Blox-Net assemblies prioritize structural recognizability. Yet art is not solely about form; it often carries emotional weight, narrative layers, or sociopolitical critique. Could future VLM-robot systems generate conceptual sculptures that intentionally convey emotions or tell stories? Embedding narrative logic or affective resonance into generative design processes represents a significant but exciting challenge, demanding richer models of human perception, symbolism, and context.

The broader implications of this work touch on fundamental questions of authorship and collaboration. Who is the true "creator" of a Blox-Net sculpture: the human engineers who built the system, the vision-language model (VLM) that synthesized the structure, the robot that assembled it, or the cultural corpus that trained the model? Notably, because the Blox-Net sculptures were generated through a GDfRA pipeline, their forms were inherently constrained by the limitations of the target robotic construction technologies. Given a similar prompt, GPT Image 1 [87] produces a more detailed and recognizable version of the figure (Fig. 4), albeit deviating from the provided set of blocks and including overhanging components. These results suggest that as robotic construction capabilities continue to evolve alongside AI, it will become possible to realize increasingly complex and expressive forms. Future work will investigate how to ground image generation models with physical constraints, enabling new, imaginative AI-driven interpretations of art that remain feasible for real-world, robotic construction. Such human-machine collaboration could catalyze new forms of artistic production, where humans articulate high-level intentions or emotional objectives and AI-robotic systems traverse vast design spaces to propose novel and unexpected realizations.

Ultimately, Blox-Net suggests that creativity is not confined to human minds but can emerge from the interplay of algorithms, materials, and embodied action. As AI and robotics continue to evolve, new hybrid modes of creation will arise—challenging us to expand our definitions of art, authorship, and imagination.

References

[1] A. Goldberg *et al.*, "Blox-Net: Generative design-for-robot-assembly using VLM supervision, physics simulation, and a robot with reset," *arXiv preprint arXiv:2409.17126*, 2024.

[2] R. W. Kennard and L. A. Stone, "Computer aided design of experiments," *Technometrics*, vol. 11, no. 1, pp. 137–148, 1969.

[3] T.-C. Chang, R. A. Wysk, and H.-P. Wang, *Computer-aided manufacturing*. Prentice-Hall, Inc., 1991.

[4] Z. Bi and X. Wang, *Computer aided design and manufacturing*. John Wiley & Sons, 2020.

[5] A. Ramesh *et al.*, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*, PMLR, 2021, pp. 8821–8831.

[6] T. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[7] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Advances in neural information processing systems*, vol. 35, pp. 27 730–27 744, 2022.

[8] J. Ho *et al.*, "Imagen video: High definition video generation with diffusion models," *arXiv preprint arXiv:2210.02303*, 2022.

[9] T. Brooks *et al.*, "Generating long videos of dynamic scenes," *Advances in Neural Information Processing Systems*, vol. 35, pp. 31 769–31 781, 2022.

[10] L. Castrejon, N. Ballas, and A. Courville, "Improved conditional vrnns for video prediction," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7608–7617.

[11] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," *arXiv*, 2022.

[12] C.-H. Lin *et al.*, "Magic3d: High-resolution text-to-3d content creation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[13] H. Wang, X. Du, J. Li, R. A. Yeh, and G. Shakhnarovich, "Score jacobian chaining: Lifting pretrained 2d diffusion models for 3d generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 619–12 629.

[14] M. Liu *et al.*, "Chipnemo: Domain-adapted llms for chip design," *arXiv preprint arXiv:2311.00176*, 2023.

[15] G. Boothroyd and P. Dewhurst, *Design for Assembly - A Designer's Handbook*. Amherst, MA: University of Massachusetts, 1983.

[16] S. Miyakawa, "The hitachi assemblability evaluation mrthod (aem)," in *Proceedings of 1st Int. Conf. on Product Design for Assembly, 1986*, 1986.

[17] General Electric Co., *Manufacturing Producibility Handbook*. Schenectady, NY: Manufacturing Services, General Electric Co., 1960.

[18] S. Y. Nof, W. E. Wilhelm, and H.-J. Warnecke, "Design for assembly," in *Industrial Assembly*. Boston, MA: Springer US, 1997, pp. 84–134.

[19] G. Boothroyd, "Design for assembly—the key to design for manufacture," *The International Journal of Advanced Manufacturing Technology*, vol. 2, pp. 3–11, 1987.

[20] T. Ohashi, M. Iwata, S. Arimoto, and S. Miyakawa, "Extended assemblability evaluation method (aem)(extended quantitative assembly producibility evaluation for assembled parts and products)," *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, vol. 45, no. 2, pp. 567–574, 2002.

[21] G. Boothroyd, P. Dewhurst, and W. A. Knight, *Product design for manufacture and assembly*. CRC press, 2010.

[22] H. K. Rampersad, *Integrated and Simultaneous Design for Robotic Assembly: Product Development, Planning...* John Wiley & Sons, Inc., 1994.

[23] K. Millheim, S. Jordan, and C. Ritter, "Bottom-hole assembly analysis using the finite-element method," *Journal of Petroleum Technology*, vol. 30, no. 02, pp. 265–274, 1978.

[24] S. Lee and H. H. Asada, "A perturbation/correlation method for force guided robot assembly," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 4, pp. 764–773, 1999.

[25] S. Y. Nof, *Handbook of industrial robotics*. John Wiley & Sons, 1999.

[26] K. Y. Goldberg, "Orienting polygonal parts without sensors," *Algorithmica*, vol. 10, no. 2, pp. 201–225, 1993.

[27] A. A. Apolinarska *et al.*, "Robotic assembly of timber joints using reinforcement learning," *Automation in Construction*, vol. 125, p. 103 569, 2021.

[28] Y. Tian *et al.*, "Assemble them all: Physics-based planning for generalizable assembly by disassembly," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 1–11, 2022.

[29] J. Luo and H. Li, "A learning approach to robot-agnostic force-guided high precision assembly," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2021, pp. 2151–2157.

[30] L. Fu, H. Huang, L. Berscheid, H. Li, K. Goldberg, and S. Chitta, "Safe self-supervised learning in real of visuo-tactile feedback policies for industrial insertion," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 10 380–10 386.

[31] Y. Tian *et al.*, *Asap: Automated sequence planning for complex robotic assembly with physical feasibility*, 2023. arXiv: 2309.16909 [cs.RO].

[32] J. Xu, M. Danielczuk, E. Steinbach, and K. Goldberg, "6dfc: Efficiently planning soft non-planar area contact grasps using 6d friction cones," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 7891–7897.

[33] C. M. Kim, M. Danielczuk, I. Huang, and K. Goldberg, "Ipc-graspsim: Reducing the sim2real gap for parallel-jaw grasping with the incremental potential contact model," in *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 6180–6187.

[34] R. M. Murray, Z. Li, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 2017.

[35] H. Huang *et al.*, "Mechanical search on shelves with efficient stacking and destacking of objects," in *The International Symposium of Robotics Research*, Springer, 2022, pp. 205–221.

[36] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning," in *2018 IEEE International Conference on robotics and automation (ICRA)*, IEEE, 2018, pp. 5620–5627.

[37] H. Touvron *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288*, 2023.

[38] A. Macaluso, N. Cote, and S. Chitta, "Toward automated programming for robotic assembly using chatgpt," *arXiv preprint arXiv:2405.08216*, 2024.

[39] H. You, Y. Ye, T. Zhou, Q. Zhu, and J. Du, "Robot-enabled construction assembly with automated sequence planning based on chatgpt: Robogpt," *Buildings*, vol. 13, no. 7, p. 1772, 2023.

[40] A. C. Doris, D. Grandi, R. Tomich, M. F. Alam, H. Cheong, and F. Ahmed, "Designqa: A multimodal benchmark for evaluating large language models' understanding of engineering documentation," *arXiv preprint arXiv:2404.07917*, 2024.

[41] A. Gaier, J. Stoddart, L. Villaggi, and S. Sudhakaran, "Generative design through quality-diversity data synthesis and language models," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2024, pp. 823–831.

[42] A. Badagabettu, S. S. Yarlagadda, and A. B. Farimani, "Query2cad: Generating cad models using natural language queries," *arXiv preprint arXiv:2406.00144*, 2024.

[43] S. Wu *et al.*, "Cad-llm: Large language model for cad generation," in *Proceedings of the neural information processing systems conference. neurIPS*, 2023.

[44] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 28, no. 3, Aug. 2009.

[45] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum, *Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling*, 2017. arXiv: 1610.07584 [cs.CV].

[46] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," in *ECCV*, 2018.

[47] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, Jul. 2017, pp. 2463–2471.

[48] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019.

[49] H. Izadinia, Q. Shan, and S. M. Seitz, *Im2cad*, 2017. arXiv: 1608.05137 [cs.CV].

[50] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[51] A. Kar *et al.*, "Meta-sim: Learning to generate synthetic datasets," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[52] K. Fang, Y. Zhu, S. Savarese, and L. Fei-Fei, "Adaptive procedural task generation for hard-exploration problems," *International Conference on Representation Learning (ICLR)*, 2021.

[53] K. Fang, Y. Zhu, S. Savarese, and L. Fei-Fei, "Learning generalizable skills via automated generation of diverse tasks," *Robotics: Science and Systems (RSS)*, 2021.

[54] K. Fang, T. Migimatsu, A. Mandlekar, L. Fei-Fei, and J. Bohg, "Active task randomization: Learning robust skills via unsupervised generation of diverse and feasible tasks," *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–8, 2022.

[55] Z. Jiang, C.-C. Hsu, and Y. Zhu, "Ditto: Building digital twins of articulated objects from interaction," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[56] C.-C. Hsu, Z. Jiang, and Y. Zhu, "Ditto in the house: Building articulation models of indoor scenes through interactive perception," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[57] L. Wang *et al.*, *Gensim: Generating robotic simulation tasks via large language models*, 2024. arXiv: 2310.01361 [cs.LG].

[58] H. Ha, P. Florence, and S. Song, "Scaling up and distilling down: Language-guided robot skill acquisition," in *Proceedings of The 7th Conference on Robot Learning*, J. Tan, M. Toussaint, and K. Darvish, Eds., ser. Proceedings of Machine Learning Research, vol. 229, PMLR, Jun. 2023, pp. 3766–3777.

[59] K. Chen, C. B. Choy, M. Savva, A. X. Chang, T. Funkhouser, and S. Savarese, "Text2shape: Generating shapes from natural language by learning joint embeddings," *arXiv preprint arXiv:1803.08495*, 2018.

[60] A. Jain, B. Mildenhall, J. T. Barron, P. Abbeel, and B. Poole, "Zero-shot text-guided object generation with dream fields," *CVPR*, 2022.

[61] A. Haque, M. Tancik, A. A. Efros, A. Holynski, and A. Kanazawa, "Instruct-nerf2nerf: Editing 3d scenes with instructions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 740–19 750.

[62] OpenAI, *Gpt-4o system card*, https://cdn.openai.com/gpt-4o-system-card.pdf, Accessed: 2024-12-22, 2024.

[63] M. Chen, R. Child, A. Ramesh, P. Mishkin, *et al.*, *Dall-e: Creating images from text*, 2021.

[64] B. Peebles *et al.*, *Sora*, 2024.

[65] C. Cuan, T. Qiu, S. Ganti, and K. Goldberg, "Breathless: An 8-hour performance contrasting human and robot expressiveness," in *Proceedings of the International Symposium on Robotics Research (ISRR)*, 2024.

[66] S. Yuan and P. Yu, *Can't help myself*, Guggenheim Museum, 2016.

[67] P. Tresset, *Patrick tresset: Semi-autonomous drawing robots*, 2024.

[68] R. Anadol, T. Art, and MYND, *Sense of healing*, 2022.

[69] N. P. Service, *Statue of liberty history*, 2024.

[70] W. Foundation, *Statue of liberty series*, 2024.

[71] J. Art, *Inside out project: Ellis island*, 2024.

[72] B. Zelizer, *About to Die: How News Images Move the Public*. Oxford University Press, 2019.

[73] C. R. Morey, "The art of auguste rodin," *The Art Bulletin*, 1917.

[74] P. Griener, "Rodin's family: The intersection between myth and reality," *Nineteenth-Century Art Worldwide*, 2021.

[75] *European sculpture of the 19th century*, n.d.

[76] *Rodin museum website*.

[77] T. Magazine, *Art: Monumental minimalist*, Vol. 90, No. 15, 1967.

[78] E. Britannica, *Tony smith: American artist and architect*, 2024.

[79] T. S. Estate, *About tony smith*, 2024.

[80] P. A. Fund, *Cigarette by tony smith*, 2024.

[81] Christo and Jeanne-Claude, *The gates*, Accessed: 2025-04-28, 2025.

[82] D. Maria-Reina, *Smarthistory – christo and jeanne-claude, the gates*, Accessed: 2025-04-28, 2025.

[83] The Shed, *Christo and jeanne-claude: The gates and unrealized projects for new york city*, Accessed: 2025-04-28, 2025.

[84] D. Judd, *Untitled. 1967*, The Museum of Modern Art; Accessed: 2025-04-28, 2025.

[85] D. T. Folland, *Smarthistory – carl andre, lever*, Accessed: 2025-04-28, 2025.

[86] M. J. Jacob, *Culture in Action: A Public Art Program of Sculpture Chicago*. Bay Press, 2012.

[87] OpenAI, *Gpt-image-1 model documentation*, https://platform.openai.com/docs/models/gpt-image-1, Accessed: 2025-04-28, 2025.