# Tangible Diffusion: Exploring Artwork Generation via Tangible Elements and AI Generative Models in Arts and Design Education

**Kuntong Han**
Tsinghua University
Beijing, China
hkt21@mails.tsinghua.edu.cn

**Keyang Tang**
Tsinghua University
Beijing, China
tangkeyang@mail.tsinghua.edu.cn

**Meng Wang**
Beijing Institute for General Artificial Intelligence
Beijing, China
wangmeng@bigai.ai

## ABSTRACT

Generative models have revolutionized the field of art and design, providing an emerging and accessible approach to creating diverse artwork. However, effectively utilizing these models still requires significant expertise, making the system inaccessible to novice users, such as young children. This paper introduces a novel approach to artwork generation, combining tangible elements with AI generative models, resulting in a more engaging and immersive learning experience. With materials prepared by teachers, students can easily create digital artwork by manipulating tangible building blocks. The experiments demonstrate that the proposed pipeline can be applied to various scenarios, using either off-the-shelf or carefully designed tangible elements. This approach provides an interdisciplinary learning platform for arts and design education, fostering creativity and exploration of various art styles and design topics.

**Figure 1**. Artwork generated by building blocks illustrating a traditional Chinese poem.



## Author Keywords

Arts and Design Education; Generative art; Tangible Interaction; Interactive Learning

## CSS Concepts

**Applied computing ~ Education;** *Arts and humanities*

## INTRODUCTION

Generative art is an emerging field that utilizes artificial intelligence (AI) algorithms to automatically create digital artwork. These algorithms have the capability to learn from existing art and generate new pieces based on the style or characteristics of an artist. However, for novice users, especially young children, generating desired artwork using AI generative models can be challenging, as it requires significant computational power, a deep understanding of algorithms, and complex interactions. Moreover, even with the existence of inpaint methods, it remains difficult to conveniently adjust the content of the generated image.

On the other hand, tangible interaction relies on the utilization of physical objects to create unique and interactive experiences. In particular, modular tangible interfaces enable students to engage in hands-on construction, providing an easy way to create and manipulate digital information [1,2]. This offers new possibilities for embodied learning [3,4,5] and promoting individual's creativity [6,7,8].

Combining AI generative models and tangible elements in artwork generation has not been extensively explored in previous research. This paper proposes a pipeline for using tangible elements to generate artwork. Teachers are responsible for designing prompts and selecting or training LoRA models that are appropriate for the chosen theme. They also choose or design tangible elements that can be used by students to create their artwork. To ensure greater control over the final output, teachers can tune ControlNet parameters to modify the output. Students, on the other hand, are responsible for building with the tangible elements and taking pictures of their creations. The pictures are then used in combination with ControlNet to generate highly customized and expressive artwork automatically.

Through this paper, we aim to explore the potential of combining AI generative models and tangible elements in artwork generation, specifically in the context of arts and design education. This approach provides a more immersive and interactive learning experience, and can make education more accessible, inclusive and engaging.

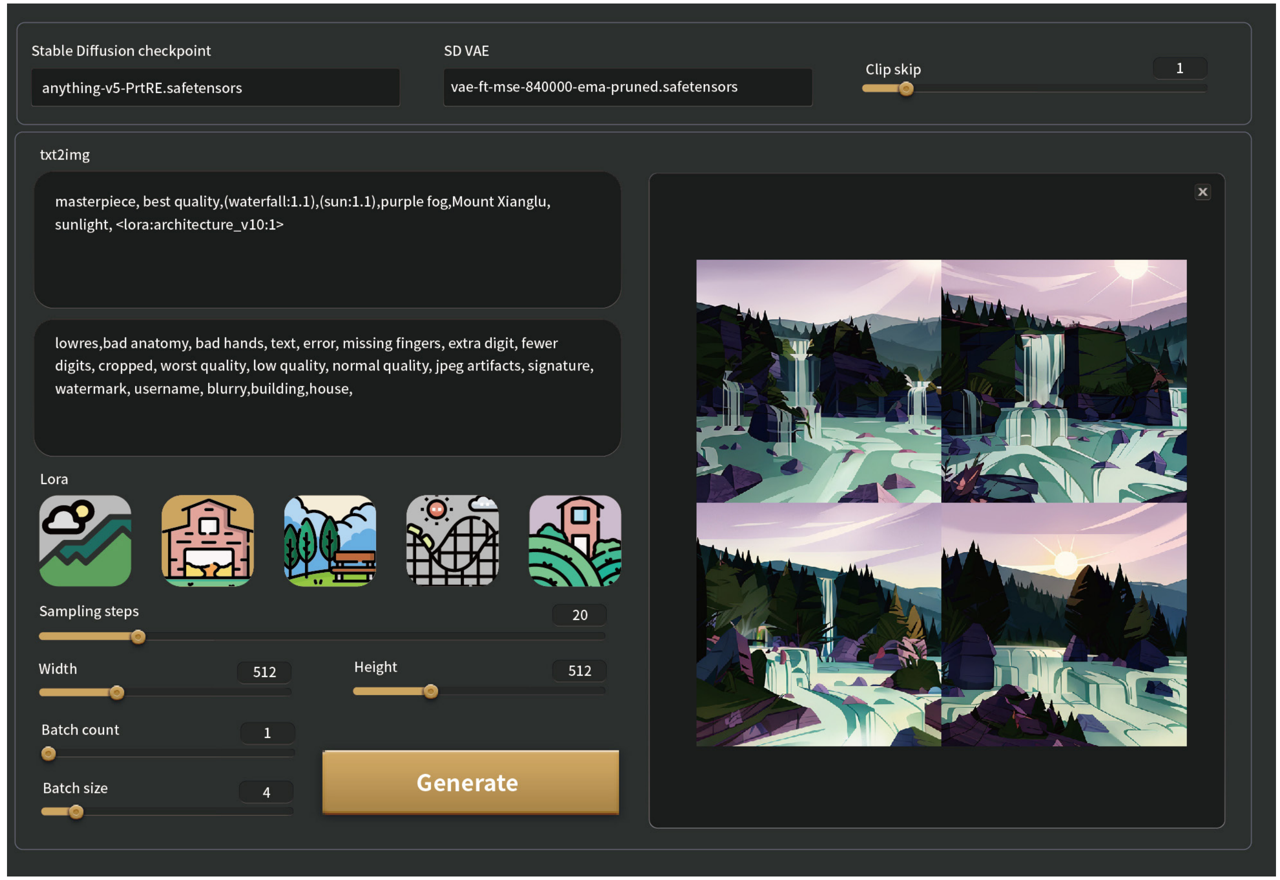



**Figure 2**. A simplified GUI for novice users to utilize the Stable Diffusion and LoRA models.

## GENERATIVE MODELS

### Stable Diffusion and LoRA

Stable Diffusion[9] and LoRA[10] are two powerful AI generative models utilized for generating high-quality artwork. While Stable Diffusion focuses on capturing the distribution of a set of images to generate consistent results in a specific style, LoRA's unique low-rank autoregressive approach enables it to produce diverse and detailed images with fewer training examples. Unlike traditional deep learning models that require large datasets, LoRA can effectively learn from a limited number of images while still generating artwork that aligns with the desired style and content. Through a Web-based Stable Diffusion UI[11] and a LoRA extension[12], non-expert users can easily create a diverse set of high-quality images with thematic coherence. This research leverages a general Stable Diffusion checkpoint across all setups while applying different LoRA models and prompts tailored to specific themes, enhancing the artistic exploration and generating artwork that reflects various artistic styles and subjects.

## CONTROLNET

ControlNet[13] is a recently developed algorithm that can be used to guide the generation of images by AI generative models. ControlNet works by providing a set of high-level controls that can be used to manipulate the generated images, such as changing the pose or orientation of an object, adding or removing objects, or adjusting the style or color of the image.

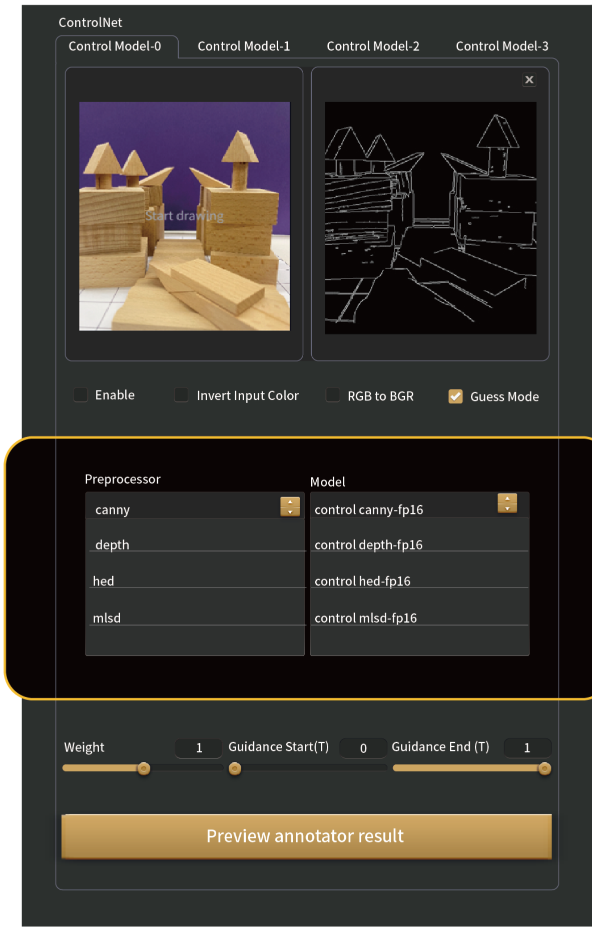



**Figure 3**. A simplified GUI for novice users to utilize the ControlNet preprocessors and models.

ControlNet can provide a powerful tool for generating highly customized and controllable artwork. By using ControlNet to manipulate the output of Stable Diffusion and LoRA[14], it is possible to generate images that are tailored to the specific needs and preferences of the user. This approach can be particularly useful in our educational settings, where students can use the tangible elements to implement diverse scenarios, and use a snapshot of a scenario to generate their artwork.

Different models can be used to preprocess and generate annotator for ControlNet. Popular models include Canny [15], Depth [16], HED [17], M-LSD [18], etc. They each have their own respective areas of application, and should be selected thematically:

*Canny*: A general-purpose edge detector that extracts outlines, preserving the composition of the original image and generating artwork following the outlines.

*Depth*: Estimation of depth information from a reference image, enabling the generation of artwork with a sense of depth and perspective.

*M-LSD*: Specialized in extracting outlines with straight edges, making it useful for generating artwork featuring interior designs, buildings, street scenes, picture frames, and paper edges.

*HED*: Leveraging deep learning, it automatically learns hierarchical representations for resolving ambiguity in edge and object boundary detection. Preserves fine details in input images, making it suitable for tasks like stylizing.

The parameters of *Weight, Guidance Start*, and *Guidance End* play a crucial role in guiding the generation of artwork. The *Weight* parameter allows users to adjust the influence of the ControlNet on the generated image. By fine-tuning this weight, users can emphasize or de-emphasize the impact of the ControlNet's guidance. The *Guidance Start* parameter defines the starting point at which the ControlNet's guidance is applied during the generation process, while the *Guidance End* parameter determines the point at which the guidance concludes.

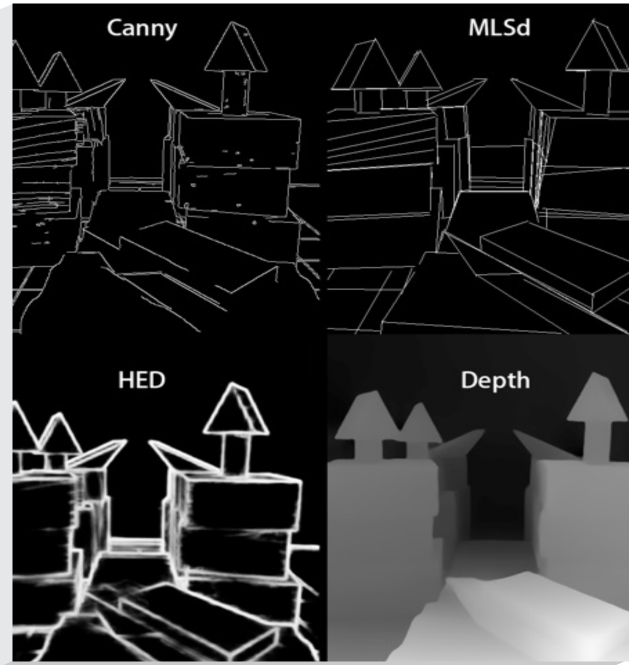



**Figure 4**. Annotators generated from the same input using different preprocessors and models.

These parameters enable users to precisely define the portion of the image generation process that is influenced by the ControlNet's guidance. In most of our cases:

- *Weight* is tuned between 0.5~1, providing a weak control to avoid unnatural patterns;
- *Guidance Start* is set to 0, to achieve better control over the overall layout of the artwork;
- *Guidance End* is more flexible between 0.5~1, and a smaller value allows the generative model to have more freedom and generate more unexpected details.

By using these different models in combination with Stable Diffusion and LoRA, it is possible to generate highly customized and controllable artwork with a variety of different styles, themes, and features.

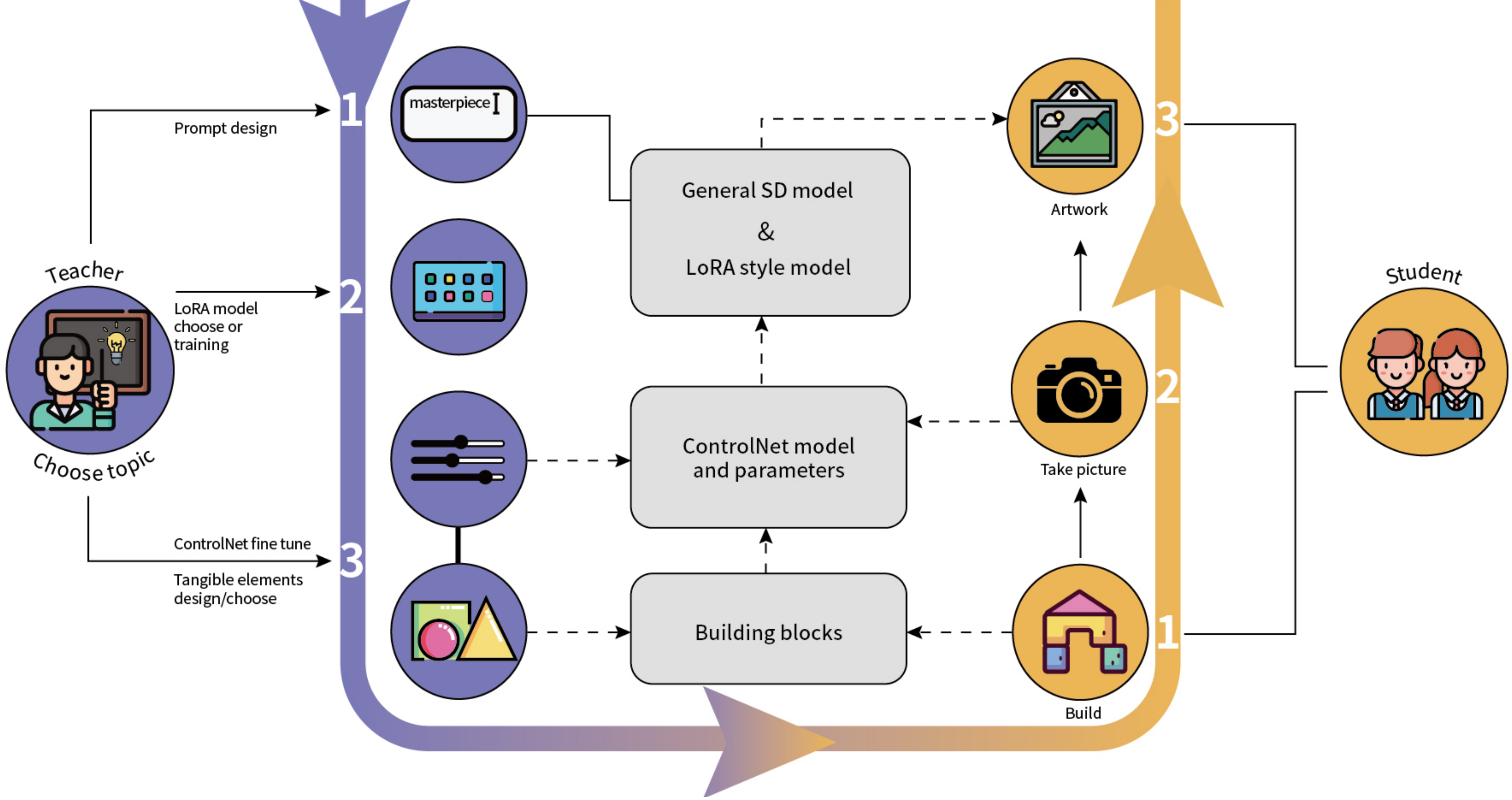



**Figure 5**. Our proposed pipeline, facilitating student participation in the AI generative process by assembling and capturing tangible building blocks.

**PIPELINE**

### 1 Designing Prompts and Choosing/Training LoRA Model:

Teachers initiate the pipeline by designing prompts that align with the educational objectives and the chosen theme, such as landscape or architecture. These prompts serve as a keynote for the students' artwork. Additionally, teachers select or train a LoRA model that is suitable for the specific prompt and desired artistic style. This step serves as the foundation for generating the initial set of images, providing a starting point for creative exploration. Other configurations, such as the sampling method, can either use default settings or be further adjusted to achieve better results.

### 2 Choosing/Designing Tangible Elements:

In this step, teachers carefully choose or design tangible elements that correspond to the theme and prompts. A good starting point is to utilize readily available building blocks such as LEGO or blocks from board games like Jenga. However, if these options do not fully satisfy the requirements, teachers can design and manufacture specific tangible elements through handcrafting or 3D printing techniques. This flexibility allows for customization and ensures that the tangible elements align perfectly with the desired educational objectives.

### 3 Tuning ControlNet Parameters:

To alleviate the challenges associated with generating high-quality artworks, teachers pre-adjust the ControlNet parameters. With the model, prompts, and tangible elements at hand, a predetermined set of parameters can effectively adapt to a wide range of distinct tangible scenes created by students. Nevertheless, teachers can also provide guidance to students in modifying specific aspects of the generation process afterward. This empowers students to exercise their creativity while enhancing their comprehension of how various controls influence the ultimate result.


**Figure 6**. (a) Assembling the given blocks. (b) A completed scene. (c) Taking a picture of the scene. (d) Generating artwork.

**1**

**Building with Tangible Elements:**
In this step, students engage in hands-on exploration and artistic expression using the tangible elements provided by the teachers. Students construct physical structures, arrange objects, and experiment with different compositions. This process encourages tactile engagement and enables students to express their ideas in a tangible and visually captivating manner.

**2**

**Taking Pictures of the Creations:**
Once students have built their artwork using the tangible elements, they capture photographs of their creations using digital cameras or smartphones. Teachers prepare a backdrop and advise students to produce a clean image. The image then serves as the input for the subsequent stages of the pipeline.

**3**

**Generating Artwork Automatically:**
With the captured photographs as inputs, students utilize prepared system on teacher's computer to automatically generate artwork. By feeding the photographs into the generative model, students witness their physical creations transformed into new digital art pieces. The ControlNet parameters further refine and guide the generative process, allowing students to experiment with different styles and interpretations. This automated generation process provides a bridge between the physical and digital realms, offering students a unique perspective on the intersection of traditional artistic techniques and AI-driven creativity.

By following this step-by-step pipeline, students' art creation experience becomes more immersive and interactive. Through the use of tangible elements, the digital results can be conveniently modified, allowing students to explore their creativity and develop a deeper understanding of the intersection of arts, design, and artificial intelligence.

**Figure 7**. Generated artwork.

**Figure 8**. Building blocks from Jenga.

## EXPERIMENT: ARCHITECTURE DESIGN

Incorporating architecture design in earlier education offers valuable benefits for young learners. By engaging students in this discipline at an early stage, they develop spatial awareness, design thinking, and interdisciplinary skills such as problem-solving and collaboration. Architectural design education fosters an understanding of the built environment, promoting cultural appreciation, sustainable practices, and social responsibility.

Many young children have an innate inclination to build structures using building blocks, which can be considered as prototypes for architectural design. However, professional architectural design requires advanced skills and knowledge, making it challenging to engage children in early education. Our approach facilitates architectural design to become more accessible and engaging by incorporating tangible elements and automated artwork generation. It expands the creative possibilities, enabling students to explore various styles, forms, and interpretations effortlessly. As students witness their tangible creations come to life through the automated generation process, it creates a more fulfilling and rewarding educational experience.

**Figure 9**. Children design museum with building blocks.

### Tangible Elements

The building blocks used in this experiment are sourced from the popular board game called Jenga. In Jenga, players take turns removing rectangular building blocks from a tower and placing them on top, gradually creating a taller and more unstable structure. These blocks are carefully crafted and typically made of wood or plastic, to be smooth and easy to handle, ensuring that players can remove and place them with precision.

In Jenga, each block has a uniform shape, with identical length and width dimensions, enabling them to stack together and form a sturdy yet precarious tower. The length-to-width-to-height ratio is approximately 3:1:0.5, which facilitates easy alignment and stacking, making them suitable for architecture prototypes.

### Details and Results

Five children, aged 7 to 8, participated in the experiment. They had experience with building blocks but no specific architectural knowledge. They were instructed to use the building blocks to create a museum. Once they completed their creations, the children were then directed to take pictures of their museum designs. These pictures were utilized as input to generate annotators via the *Depth* model. Alongside predefined prompts and the LoRA model, the annotators guided the generative process, resulting in digital images of the museums. After this step, we assisted the children in replacing the prompts from "modern museum" to "shopping mall" to observe how easily the creations could be transformed into another theme with the aid of AI generative models. Some selected results are presented in Figure 10.

Positive prompts:

*masterpiece, best quality, (modern museum:1.1), building, sky*

LoRA model:

*Fair-faced concrete architecture* [19]

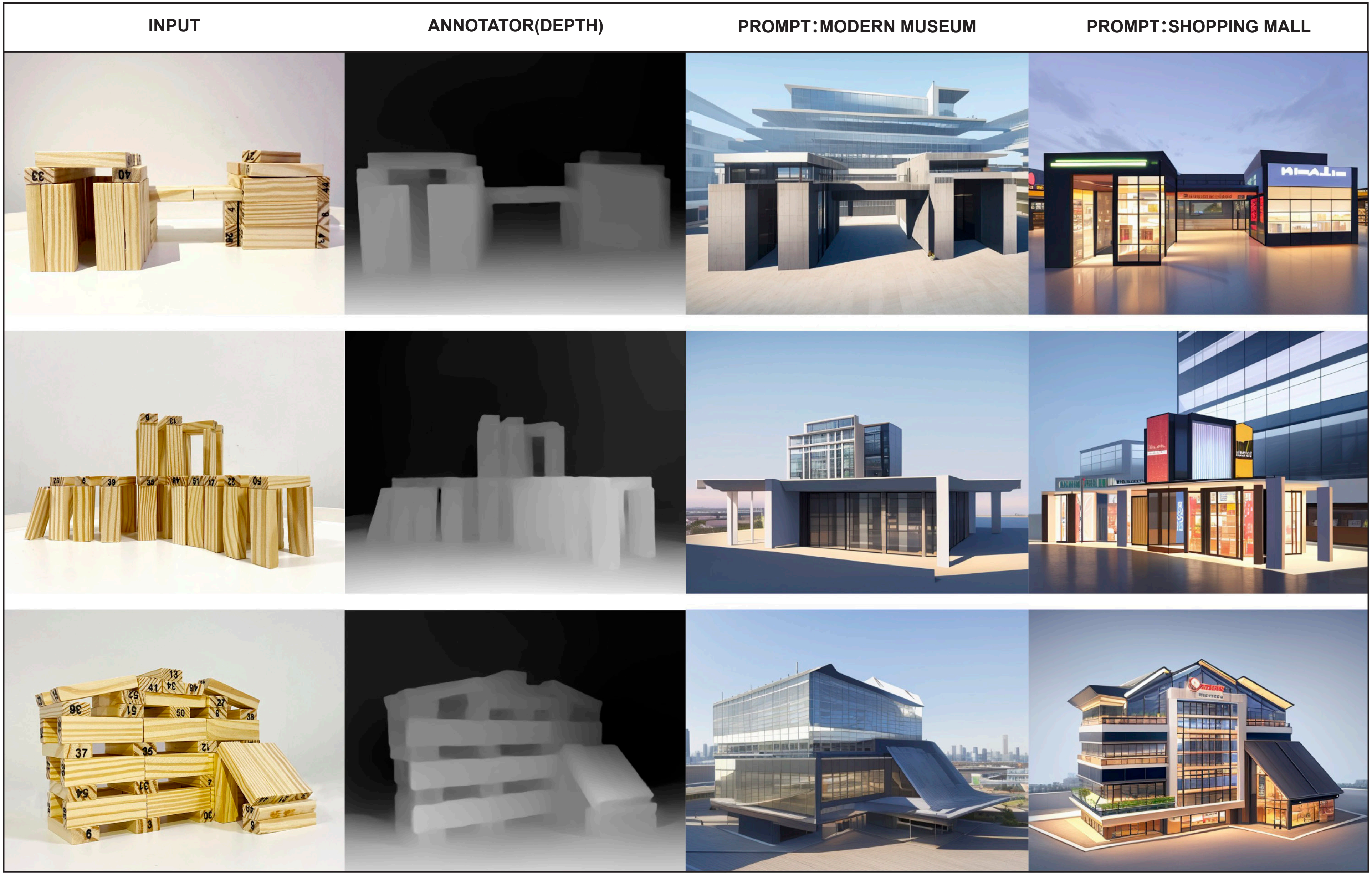



**Figure 10**. Results of the architecture design experiment.

**EXPERIMENT: POEM UNDERSTANDING**

Chinese classical poetry holds a significant place in language teaching in China, fostering a deep appreciation for the cultural richness and linguistic intricacies of Chinese literature. These poetic works provide avenues for exploring geographical, historical, philosophical, and moral themes. The study of classical poetry in early education ignites a lifelong passion for literature, nurtures cultural identity and pride in heritage, and enhances reading comprehension skills.

However, comprehending classical poetry can be challenging due to its distinct grammatical structures and linguistic conventions that differ from modern Chinese. Existing teaching methods often rely heavily on rote memorization, which can be monotonous and fail to engage young children and adolescents effectively. Our approach introduces an immersive and exploratory learning method specifically tailored for the abundant landscape poems found in classical poetry, fostering a tactile and interactive connection to the artistic conception evoked in the poems.

Unlike other forms of poetry like lyrical poems, landscape poems typically have distinct scene and spatial relationships in their content, making them suitable for further materializing through physical construction of scenes. Additionally, Chinese classical landscape poems are a typical form of artistic work completed collectively by the author and readers. With concise language, they leave ample room for readers' imagination, making them more likely to resonate. The construction of tangible elements also offers a unique way for each reader to build their own unique landscape poem according to their hearts.

Chinese classical poetry emphasizes the creation of artistic conception, with imagery being its fundamental building unit. In order to achieve a universal construction effect, we seek to use simple geometric shapes to construct the common imagery found in classical poetry. Based on this foundation, we then are able to compose the scenes depicted in the poems.

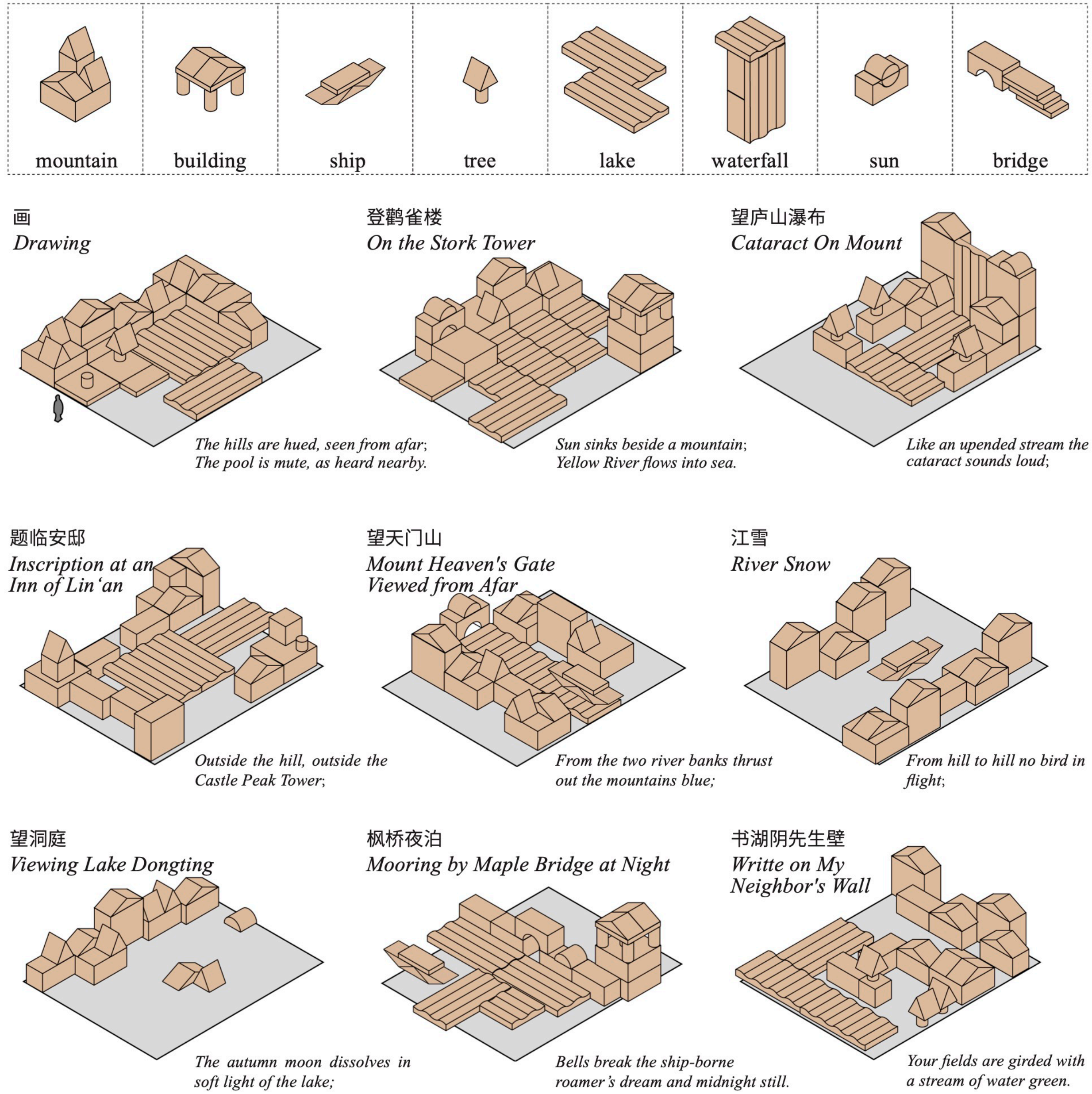
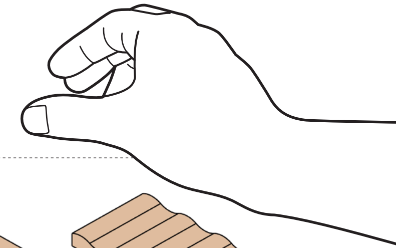
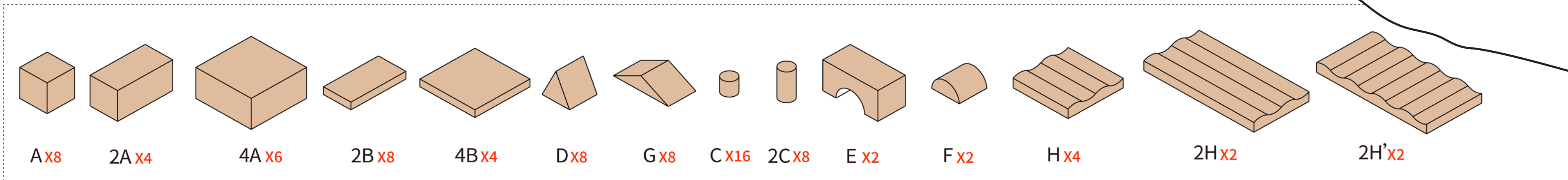



**Figure 11**. Some common imagery and poem scene constructed by simple geometric shapes.

**Figure 12**. The shapes and quantities of the building blocks in a set.






**Tangible Elements**

In this experiment, a set of building blocks is specifically designed for creating landscape scenes. The primary focus is to provide flexible and modular combinations, allowing the blocks to be freely assembled with stability. As a result, the blocks mainly consist of hexahedrons with integer multiple relations in size. Additionally, landscape poems often include imagery such as *mountains, trees, rivers, waterfalls, pavilions*, and others. Therefore, the blocks should also be capable of constructing as many imagery elements as possible. For example, a portion of the blocks is designed as cylinders and triangular prisms to represent *mountains and trees*, respectively, while hexahedrons with a single wavy face are used to represent *water*.

These building blocks are primarily designed for school and home education, with an overall size of 250mm x 200mm x 25mm, making them suitable for use on desks and portable. The smallest module measures 12mm x 12mm x 12mm, ensuring that children can conveniently handle them.

As the primary users of these building blocks are children, eco-friendly beech wood is chosen as the material. Beech wood is known for its relatively high density and weight compared to other types of wood, adding to the stability and durability of the blocks. The wood is milled using woodworking machines to achieve the desired shapes, and then the blocks undergo a polishing stage to smooth the surface, round the edges, and remove any roughness, ensuring that they are safe for children's use.

**Figure 13**. A set of wooden building blocks which can be neatly packed in a portable box.

**Details and Results**

Five children, aged 7 to 8, participated in the experiment. They had experience with building blocks, but it was their first time using this particular set. They were instructed to use the building blocks to construct a scene representing the famous poem '望庐山瀑布' (Cataract on Mount Lu), which they were familiar with from the classroom. Once the children completed their creations, they were directed to take pictures of their scenes. These pictures were utilized as input to generate annotators via the HED model. Alongside predefined prompts and the LoRA models, the annotators guided the generative process, resulting in digital images of the poem. Two different LoRA models were utilized to showcase the ease of generating artwork of different styles through AI generative models. In addition, children are also encouraged to take picture of the physical scene from a different perspective, to observe corresponding changes in the generated images.

LoRA models:

*Illustrate style* [20]; *Landscape style* [21]

Positive prompts:

*masterpiece, best quality, waterfall, no building, purple fog, Mount Xianglu, sunlight.*

Negative prompts:

*building.*

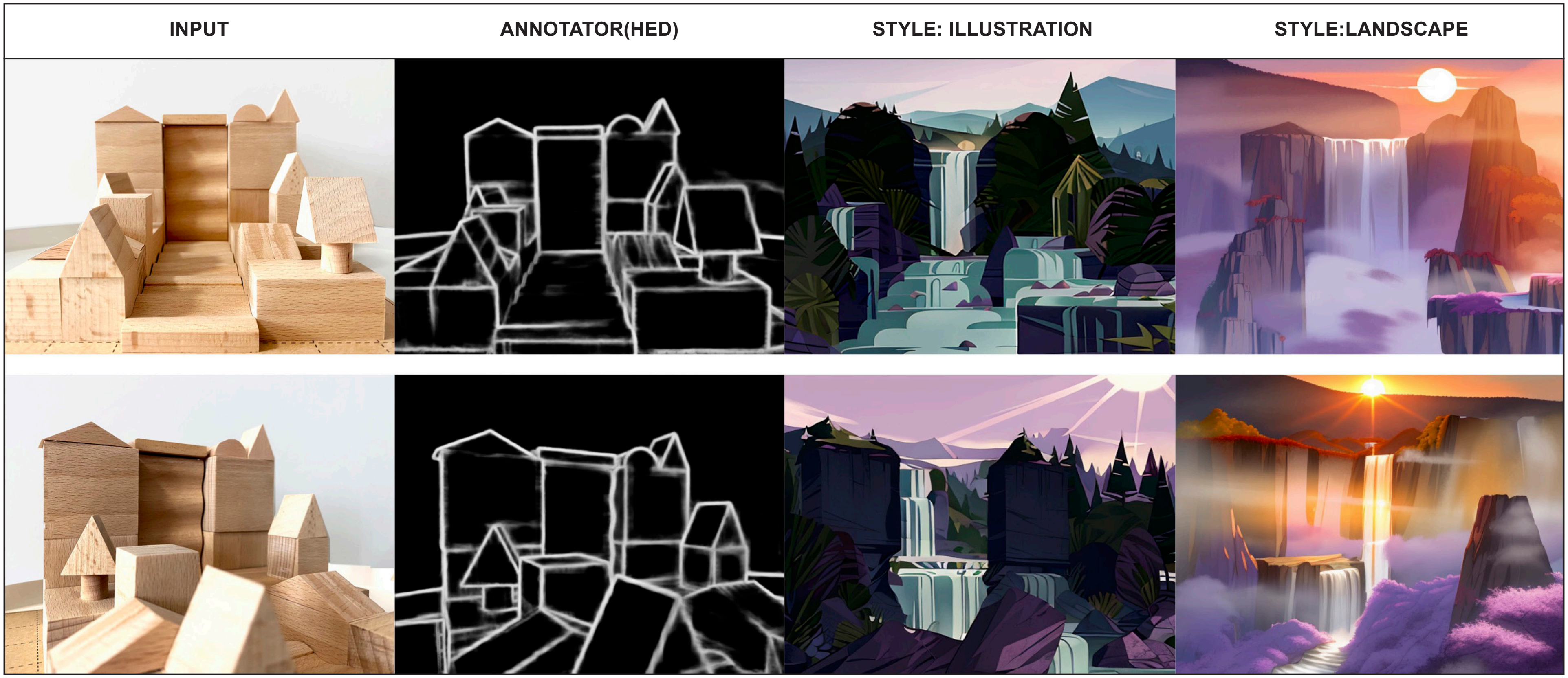



**Figure 14**. Results of the poem understanding experiment. These two annotators are generated from same physical scene but with different perspectives.

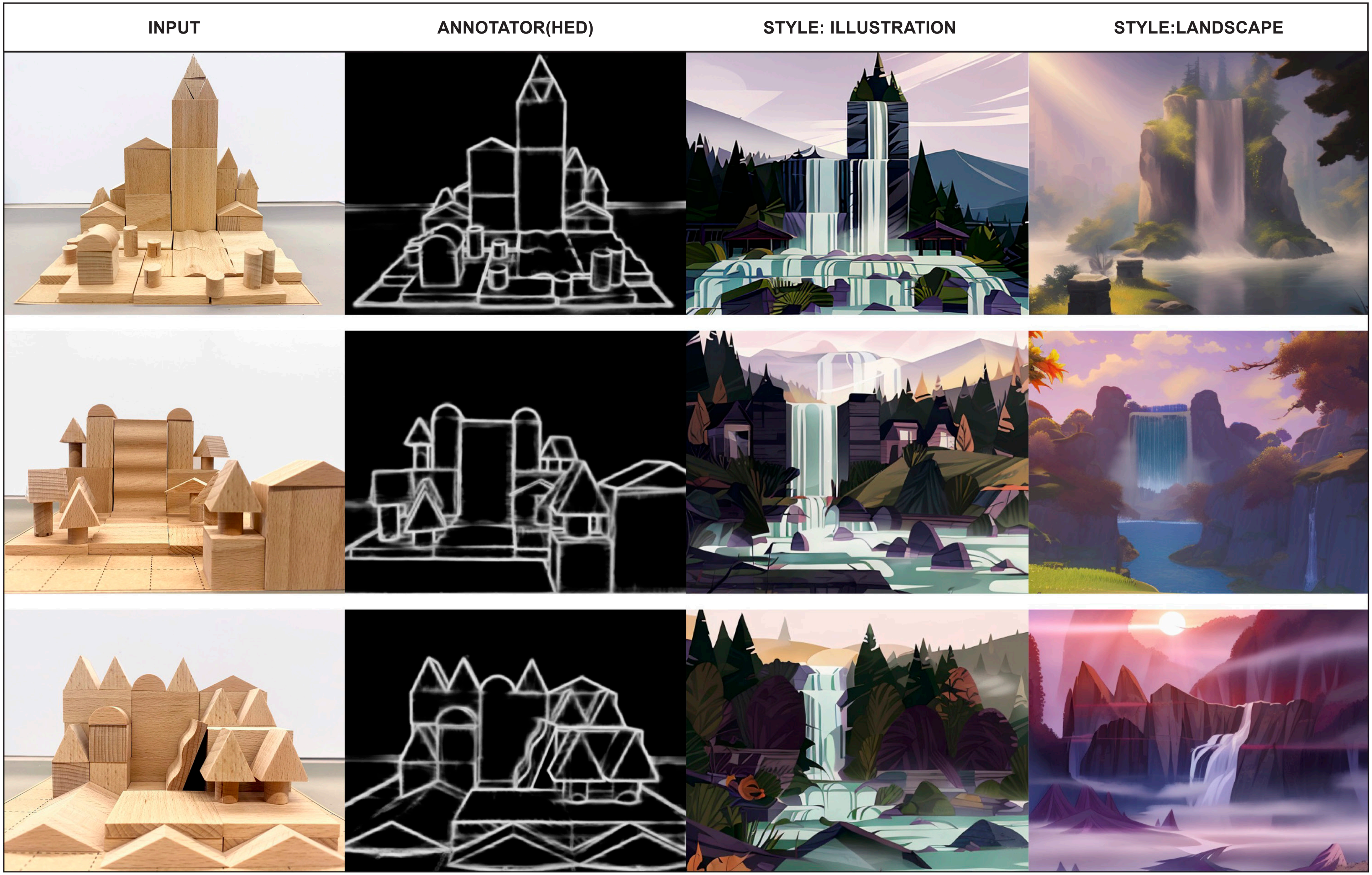



**Figure 15**. More results of the poem understanding experiment.

## LIMITATION AND FEEDBACK

As this research is in its early stages, we only prepared one set of building blocks for each experiment, and the experiments were conducted separately, one child at a time. We invited two colleagues with backgrounds in pedagogy to act as teachers. They have less experience in art or AI. We provided them with a detailed, hands-on walk-through and offered minimal assistance during the experiment. In addition, we designed the building blocks for the poems ourselves, rather than leaving it to the teachers.

Despite this limitation, this innovative activity has received positive feedback from both teachers and children. Teachers have witnessed a remarkable increase in student engagement and enthusiasm during the hands-on exploration with tangible elements. The use of tangible elements has ignited creativity, enabling students to experiment, construct, and express their ideas in a tangible and visually captivating manner. Moreover, children have eagerly embraced the opportunity to witness their physical creations come to life in digital form, allowing them to explore new artistic possibilities and experience a sense of empowerment. The combination of tangible elements and automated artwork generation has provided a unique and engaging learning experience that has left both teachers and children inspired, motivated, and eager to further explore the intersection of arts, design, and AI.

## CONCLUSION AND FUTURE WORK

In conclusion, this paper explores the innovative approach of artwork generation via tangible elements. By integrating tangible elements with AI generative models, this approach offers an immersive learning experience for students. The use of tangible elements enhances hands-on engagement, promotes creativity, and fosters a deeper understanding of spatial relationships and design principles. The automatic generation of artwork through AI algorithms expands artistic possibilities and encourages exploration of various art styles and interpretations. The two experiments demonstrated that the proposed pipeline can be applied to different scenarios, using either off-the-shelf or carefully designed tangible elements. The feedback from both teachers and children involved in these experiments has been overwhelmingly positive, highlighting increased engagement and a sense of accomplishment.

**Figure 16**. Teacher and student in the experiment.

This proposed innovative approach has the potential to revolutionize arts and design education, empowering students to explore their creativity, develop critical thinking skills, and shape the future of artistic expression. By bridging the physical and digital realms, artwork generation via tangible elements opens new avenues for interdisciplinary learning, cultural appreciation, and the cultivation of lifelong artistic passion and expression.

While the system has not been fully tested from the teacher's perspective, the experiments indicate that using this system at a basic level requires little knowledge of art or AI. Nevertheless, for teachers aiming to design specific high-quality courses, a fundamental understanding of design and generative art may be necessary. This requirement can be met through training or collaboration.

In future work, the proposed pipeline may integrate more algorithms to improve the results, for example, using reference control to maintain consistent digital image content when modifying the physical scene. This will provide students with more control over the generated outcomes and enable a seamless creative process.


## ACKNOWLEDGMENTS

We thank all the participants, who provided helpful comments on this project.

## REFERENCES


[1] Pruszko, L., Gu, H., Bourgeois, J., Laurillau, Y., & Coutrix, C. (2023, February). Modular Tangible User Interfaces: Impact of Module Shape and Bonding Strength on Interaction. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 1-15).

[2] Wang, M., Lei, K., Li, Z., Mi, H., & Xu, Y. (2018, March). Twistblocks: Pluggable and twistable modular tui for armature interaction in 3d design. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 19-26).

[3] Skulmowski, A., & Rey, G. D. (2018). Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications*, 3(1), 1-10.

[4] Stusak, S., Hobe, M., & Butz, A. (2016, February). If your mind can grasp it, your hands will help. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 92-99).

[5] Taylor, J. L., Aboriginal Shire Council, W. W., Soro, A., Esteban, M., Vallino, A., Roe, P., & Brereton, M. (2020, April). Crocodile language friend: Tangibles to foster children's language use. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

[6] Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction* (TOCHI), 20(1), 1-30.

[7] Frith, E., Miller, S., & Loprinzi, P. D. (2020). A review of experimental research on embodied creativity: Revisiting the mind–body connection. *The Journal of Creative Behavior*, 54(4), 767-798.

[8] Marichal, S., Rosales, A., Perilli, F. G., Pires, A. C., Bakala, E., Sansone, G., & Blat, J. (2017, September). Ceta: designing mixed-reality tangible interaction to enhance mathematical learning. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-13).

[9] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).

[10] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. *arXiv preprint* arXiv:2106.09685.

[11] AUTOMATIC1111. Stable Diffusion web UI. 2023. Retrieved August 1, 2023 from https://github.com/AUTOMATIC1111/stable-diffusion-webui

[12] kohya-ss. Additional Networks for generating images. 2023. Retrieved August 1, 2023 from https://github.com/kohya-ss/sd-webui-additional-networks

[13] Zhang, L., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. *arXiv preprint* arXiv:2302.05543.

[14] Mikubill. ControlNet for Stable Diffusion We bUI. 2023. Retrieved August 1, 2023 from https://github.com/Mikubill/sd-webui-controlnet

[15] J. Canny, "A Computational Approach to Edge Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[16] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2020). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3), 1623-1637.

[17] Gu, G., Ko, B., Go, S., Lee, S. H., Lee, J., & Shin, M. (2022, June). Towards light-weight and real-time line segment detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 1, pp. 726-734).

[18] Xie, S., & Tu, Z. (2015). Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 1395-1403).

[19] BYCJ. Fair-faced concrete architecture. 2023. Retrieved August 1, 2023 from https://civitai.com/models/34597

[20] JerryQAQ. Architecture_illustrate. 2023. Retrieved August 1, 2023 from https://civitai.com/models/31673/architectureillustrate

[21] nucleardiffusion. Samuri Jack s05 Landscape Style. 2023. Retrieved August 1, 2023 from https://civitai.com/models/14914/samuri-jack-s05-landscape-style