A Dataset for Mechanical Mechanisms

Farshid Ghezelbash* Amir H Eskandari Amir J Bidhendi

August 2024

Abstract

This study introduces a dataset consisting of approximately 9,000 images and corresponding descriptions of mechanical mechanisms, aimed at supporting research in mechanism design. The dataset includes a diverse collection of 2D and 3D sketches, meticulously curated to ensure relevance and quality. We demonstrate the application of this dataset by fine-tuning two models: 1- Stable Diffusion (for generating new mechanical designs), and 2- BLIP-2 (for captioning these designs). While the results from Stable Diffusion show promise, particularly in generating coherent 3D sketches, the model struggles with 2D sketches and occasionally produces nonsensical outputs. These limitations underscore the need for further development, particularly in expanding the dataset and refining model architectures. Despite these challenges, this work serves as a step towards leveraging generative AI in mechanical design, highlighting both the potential and current limitations of these approaches.

1 Introduction

Mechanical mechanism design has traditionally followed a structured approach where engineers begin by identifying a problem, followed by exploring and analyzing existing mechanisms to inspire or directly apply to their designs. This method, while effective, can be time-consuming and limited by the scope of known mechanisms. One can use generative models to assist in the design process or at least to aid in the brainstorming step of designing a mechanism; although there are some text-to-image generation tools already available (such as DALL-E and Midjourney) these tools are generally geared towards creating generic images and have not been fine-tuned for mechanical systems, often resulting in outputs that are not suitable for engineering purposes. Therefore, there is a need for specialized datasets that can serve as the foundation for fine-tuning or training these generative models for the design of mechanical mechanisms. With the advent of these models, there is now a potential paradigm shift in how mechanical mechanisms are conceived, as these AI tools can assist in generating novel mechanism ideas or serve as a brainstorming aid, potentially streamlining the design process.

Therefore, this work presents a dataset of 9,000 images of mechanical mechanisms (3D and 2D sketches), each accompanied by a text description. This dataset is intended to serve as a resource for those exploring the application of generative AI in the design of mechanical mechanisms. To demonstrate the utility of this dataset, we have applied it in generating new mechanism designs as well as in developing image captioning models. This work provides a specialized data-set for the research community, and illustrates the potential of AI-driven approaches in advancing mechanical design processes.

2 Methods

The dataset of mechanical mechanisms was compiled through web scraping from various sources. The first source was a YouTube channel focused on mechanical design (Table 1), where mechanisms were primarily modeled using CAD software [6]. For each mechanism, we extracted a frame from the video along with its description. The second source was a digital library dedicated to mechanisms and gears [1], which included a video section featuring 3D reconstructions. These videos provided detailed visualizations of mechanisms,

*Department of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA. ghezelbash.far@gmail.com

from which we extracted images and descriptions (Table 1). The third source was a book that contains a vast collection of 2D sketches and comprehensive descriptions of mechanical mechanisms [2]. In total, these sources yielded 8,994 images and corresponding descriptions.

To ensure the quality of the dataset, we conducted a thorough manual review of all the images. During this process, we identified and removed any images that were blank, irrelevant, or did not make sense. To clean up descritions, we manually adjusted and corrected descriptions to ensure consistency and relevance. We used ChatGPT to remove references to patents/designs, creators' names, or other verbose and/or unrelated content by providing specific instructions to remove any such references, including mentions of detailed variants or referrals to other designs.

Table 1: Summary of datasets with the number of images, and references.

Number of Images	Sketch Type	Reference
3872	3D	[6]
980	3D	[1]
4142	2D	[2]
Total: 8994		

2.1 Textual Data Analysis and Visualization

To analyze the text descriptions of mechanisms, we processed the dataset by tokenizing the text and removing common stop words [3]. The frequency of the remaining words was calculated using a word counter to identify key terms. A word cloud was generated to visually represent the most frequent words, and the top words were further analyzed to identify potential synonyms using WordNet, a lexical database that groups English words into sets of synonyms.

2.2 Design Generation Using Stable Diffusion

To showcase the likely application of this dataset, we fine-tuned the version 1.6 using our complete dataset [5]. Stable Diffusion is a deep learning model known for its ability to generate high-quality images from textual descriptions. For our purpose, we utilized the entire dataset of 8,994 images and descriptions to fine-tune this model specifically for generating mechanical mechanism designs. This fine-tuning process allowed the model to learn the specific characteristics and visual features of mechanical systems.

2.3 Captioning Using BLIP-2

To present an application of this dataset, we fine-tuned the BLIP-2 (Bootstrapped Language-Image Pre-training) model using our dataset [4]. BLIP-2 is a vision-language model designed to generate descriptive captions for images by understanding the relationship between visual content and textual information. We fine-tuned BLIP-2 on our dataset to improve its ability to produce accurate and concise descriptions of mechanical mechanisms. By doing so, we ensured that the captions generated were not only relevant but also technically appropriate, enhancing the usability of the dataset for further research and application in mechanical design.

3 Results and Discussion

After processing and cleaning the datasets (Table 1), we finalized a collection of 8,994 images with their corresponding descriptions. This dataset, while relatively modest in size, provides a foundational resource for training/fine-tuning models in the context of mechanical mechanism design. The word cloud in Figure 1 visually represents the frequency of terms used in the descriptions, highlighting key concepts and components that are prevalent in mechanical mechanism design. Figure 2 showcases a sample of nine randomly selected mechanisms along with their descriptions, which illustrate the diversity of the mechanical mechanisms included in the dataset. Additionally, Table 2 summarizes the most frequent key terms and their synonyms, offering an insight into the linguistic patterns in the dataset.

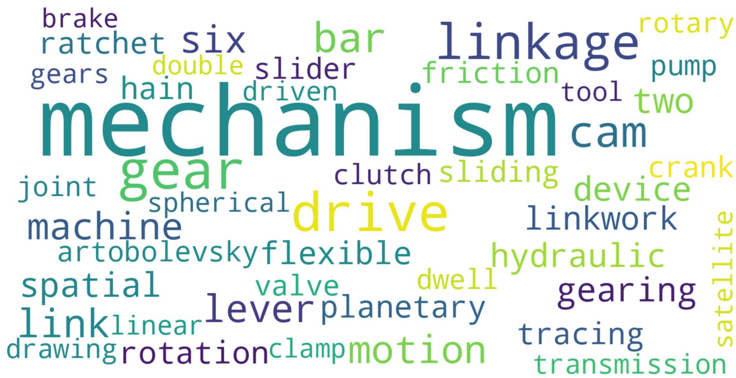


Figure 1: World cloud of text description of mechanisms.

Table 2: Frequency of key terms and their synonyms

Index	Word	Frequency	Synonyms
1	mechanism	4850	mechanism; chemical mechanism; mechanics
2	gear	541	paraphernalia; gearing; gear mechanism; appurtenance; ...
3	drive	533	motor; repulse; beat back; parkway; driving; ride; force; repel; ...
4	linkage	451	gene linkage; linkage
5	cam	362	Cam River; cam; River Cam; Cam
6	bar	318	saloon; streak; barricade; Browning automatic rifle; debar; bar; ...
7	link	283	radio link; contact; colligate; tie in; tie; unite; connect; relate; ...
8	motion	279	motion; apparent motion; motility; gesture; apparent movement; ...
9	lever	272	jimmy; prise; pry; prize; lever; lever tumbler
10	six	268	half dozen; half dozen; Captain Hicks; 6; vi; six spot; half a dozen; ...

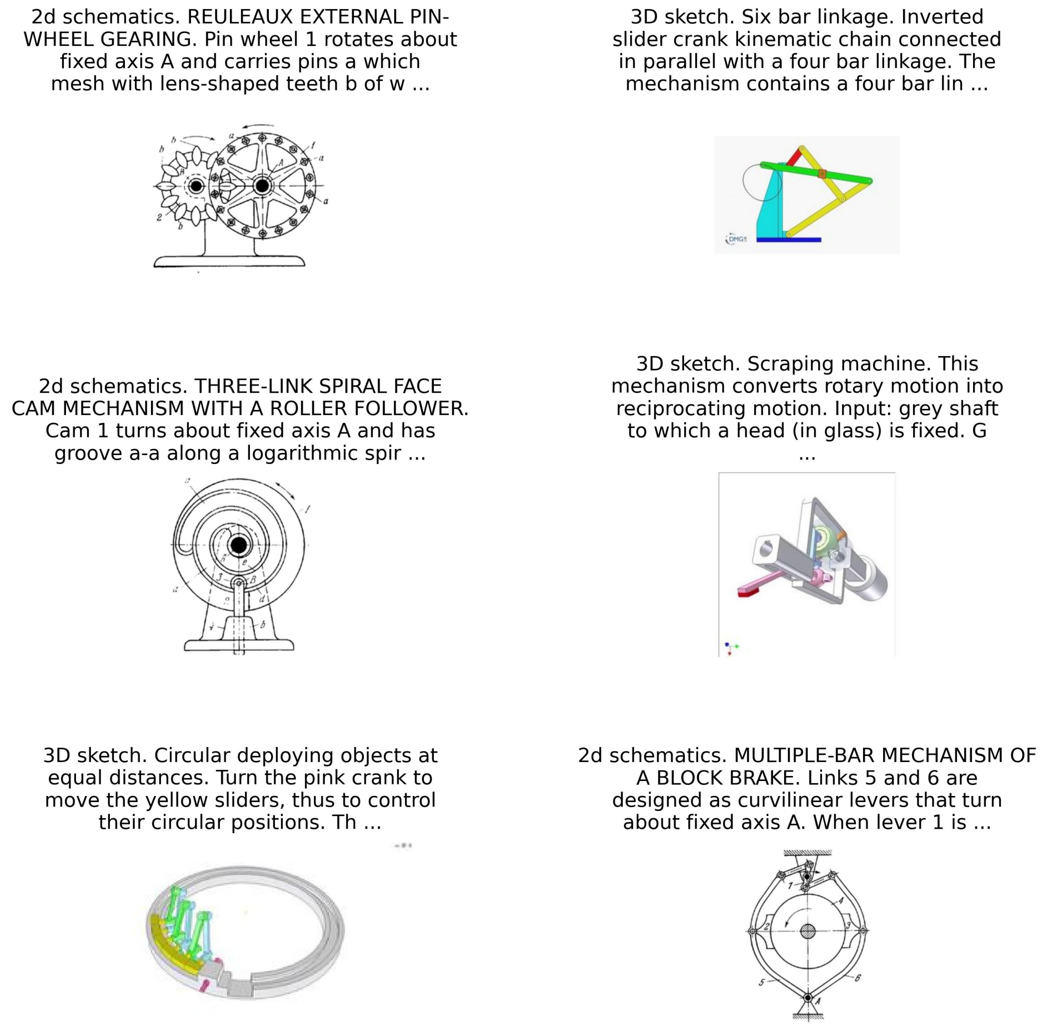


Figure 2: Nine randomly selected mechanisms with their descriptions (limited to 150 characters).

3.1 Design Generation Using Stable Diffusion

After fine-tuning the model (1500 epochs) for Stable Diffusion model on the presented dataset, we generated both 2D and 3D sketches (by adding ”2D schematic” or ”3D sketch” at the beginning of each prompt; Figure 3 and Figure 4). We noted that some of the generated mechanisms made relatively good sense (Figure 3), particularly the 3D sketches, which often included essential components of the provided descriptions. However, the 2D sketches in most examples were relatively meaningless and lacked coherence (Figure 3 and Figure 4). Despite these results, the model still has a lot of room for improvement. During experimentation, we also noticed that the model occasionally produced nonsensical outputs, especially for certain prompts where it tended to hallucinate and generate figures that were not only inaccurate but also entirely unrelated to the intended design (Figure 4). This tendency indicates the need for further refinement of the model, particularly in handling more complex or specific prompts to reduce instances of such errors and improve the overall reliability of the generated sketches.

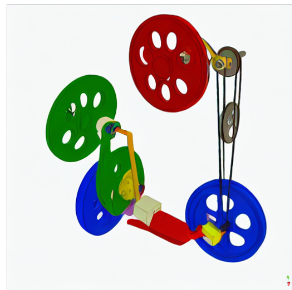
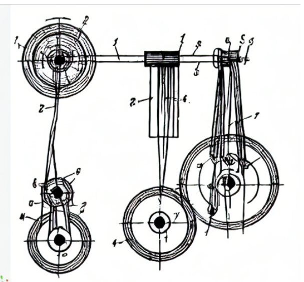
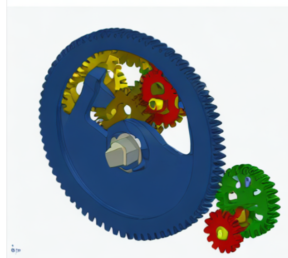
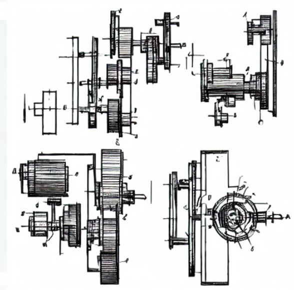
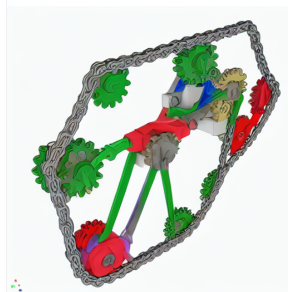
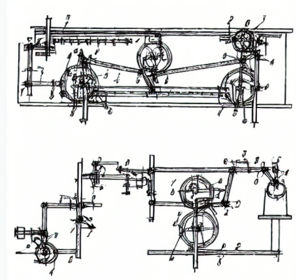
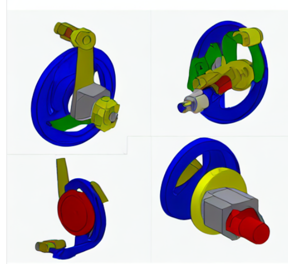
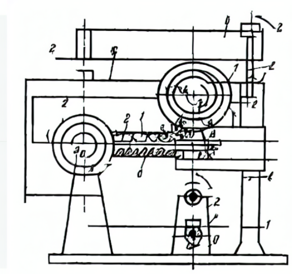
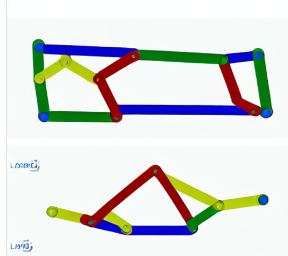
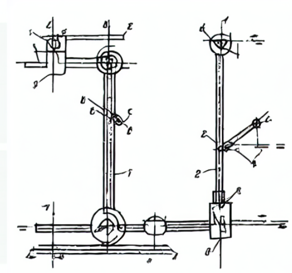
Description	3D	2D
Pulley. A wheel with a grooved rim around which a cord passes; used to lift loads.		
Gear. A rotating machine part with cut teeth that engage with another gear to transmit torque.		
Chain Drive. A method of transmitting mechanical power between two rotating shafts using a chain that meshes with sprockets on the shafts.		
Ratchet. A mechanical device that allows continuous linear or rotary motion in only one direction while preventing motion in the opposite direction.		
Linkage. A series of rigid links connected by joints to transfer motion and force.		

Figure 3: Examples of 2D (right) and 3D (middle) sketches generated by the fine-tuned Stable Diffusion model from a text input (left). The 3D sketches generally align better with the provided descriptions, capturing key components of mechanical mechanisms, while the 2D sketches often lack coherence and meaningful structure.

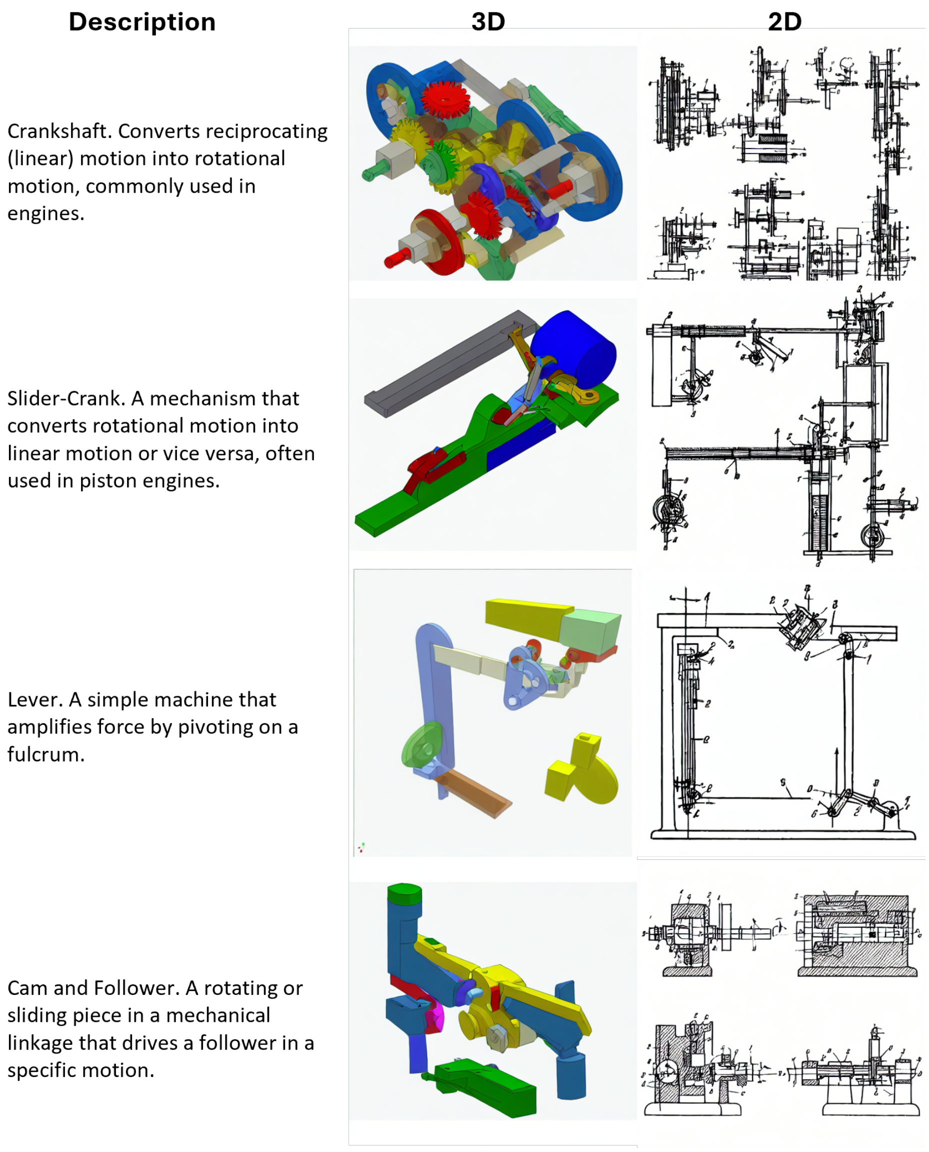
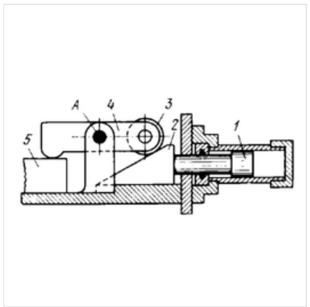


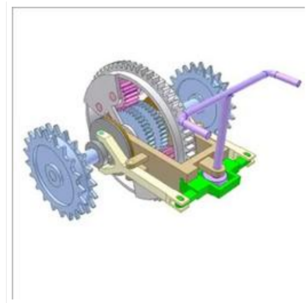
Figure 4: Examples of nonsensical/hallucinated outputs generated by the fine-tuned Stable Diffusion model from a text input (2D: right; 3D: middle; input prompt: left). These examples illustrate the model's difficulty in accurately interpreting prompts, leading to outputs that do not align with the intended mechanical designs and lack meaningful structure.

3.2 Captioning Using BLIP-2

The results of the BLIP-2 model fine-tuning for captioning mechanical mechanisms are mixed. Most of the generated captions are incorrect, with only a few containing elements of truth. This inconsistency is likely due to the limited number of training epochs (=10), which is far from sufficient for achieving accurate results. Our primary goal was to demonstrate the potential of this approach, despite significant limitations in training resources, particularly GPU access. The generated captions, along with their corresponding images, are presented in Figure 5.

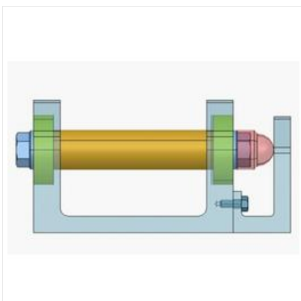


Generated caption: 2d schematics. LEVER-TYPE SCREW-ACTUATED CLAMP.
Real caption: 2d schematics. WEDGE-LEVER MECHANISM OF A HYDRAULIC CLAMPING DEVICE.

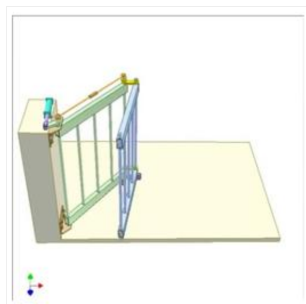


Generated caption: 3D sketch. Transmission with teeth-uncompleted gears.
Real caption: 3D sketch. Braked differential steering. It is for steering vehicles of continuous tracks.

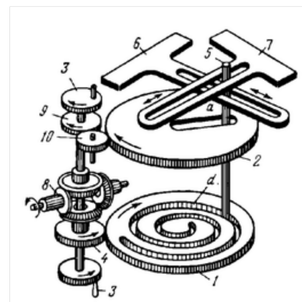
Generated caption: 3D sketch. Machine tool fixture. The pink gear is fixed to the green shaft.
Real caption: 3D sketch. Friction roller drive. The arm and the red roller create normal forces on contact surfaces.



Generated caption: 3D sketch. Machine tool fixture.
Real caption: 3D sketch. Kinematic rotation torque freely engaged with a bearing.



Generated caption: 3D sketch. Gate to open inward uphill.
Real caption: 3D sketch. Bi-folding gate. Orange conrod, green and upper yellow cranks create a parallelogram mechanism.



Generated caption: 2d schematics. LEVER-GEAR MECHANISM WITH A DWELL.
Real caption: 2d schematics. SLOTTED-LEVER-GEAR SPIRAL-TYPE RECTANGULAR COORDINATE MECHANISM.

Figure 5: Randomly selected 6 mechanisms with their generated captions and real captions.

3.3 Limitations

One of the primary limitations of this study is the relatively small size of the dataset, which includes only 8,994 images and descriptions; however, we believe this is a foundational step to move toward using generative models for mechanism design. This modest size have limited the model's ability to generalize across a broader range of mechanical designs, leading to less reliable outputs, especially for more complex or novel inputs.

3.4 Future Directions

- Expand the dataset by incorporating a wider variety of mechanical mechanisms, including more complex and diverse designs.
- Refine the model's architecture and training procedures to reduce the occurrence of nonsensical outputs and improve the coherence of 2D sketches.
- Explore alternative generative models or integrate multiple models to enhance the quality of the generated designs.
- Apply the model to real-world design challenges and iteratively improve its performance based on feedback from engineering professionals, transitioning the research from theoretical to practical application.

4 Code and Dataset Availability

The code used for fine-tuning the models and generating the results presented in this paper is available on GitHub: https://github.com/farghea/database_for_mechanical_mechanism. The dataset, including 256x256 versions (https://drive.google.com/file/d/1yC6nKih8HcAAoKCVM-Lo6bxGQ2O8T5-_/view?usp=sharing) and higher resolution (https://drive.google.com/file/d/1jqSKDypbN3vfGBA2SnUuQLuSnZC3BPYh/view?usp=sharing), can be accessed from Google Drive.

5 Responsibility of the Use of the Dataset

All data included in this dataset was collected from publicly available sources on the internet. We have ensured that the data was freely accessible at the time of collection. To respect and acknowledge the contributions of the original creators, users of this dataset are strongly encouraged to cite the corresponding references provided in the dataset documentation. It is the responsibility of the users to ensure that the dataset is used ethically and in accordance with any applicable laws or regulations.

References

[1] Dmg-lib: Digital mechanism and gear library - video collection. https://www.dmg-lib.org/dmglib/main/portal.jsp?mainNaviState=browsen.video, 2024. Accessed: 2024.

[2] Ivan I Artobolevsky. Mechanisms in modern engineering design. *Mir publishers*, 1, 2, 3, 4, 1975-1980.

[3] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[4] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023.

[5] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[6] Thang. thang010146 - youtube channel. https://www.youtube.com/@thang010146/featured, 2024. Accessed: 2024-08-11.