PROMPT PROGRAMMING FOR THE VISUAL DOMAIN

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work, we ask how text-to-image synthesis via large language models can effectively probe imagery that embodies fidelity and imagination. We investigate this question in the context of prompts (writing to language models) in a novel probing mechanism known as prompt programming, or programming in natural language. We start by refining existing techniques to characterize the effect of templates on visual fidelity then hone in on approaches to capture holistic nuances within the visual domain.

1 INTRODUCTION

AI models for text-prompted visual image generation have become very prominent in recent months PromptBase (2022) and we find it important to determine how to use these models so they may become useful tools for expressing creative intentionality. As such, we consider prompt programming to probe a model's capabilities in light of their specific benefit to language models, as quantified in recent work to often worth 100s of data points in enhancing a model's ability Scao & Rush (2021). We find prompts to be a valuable training method for invoking the development of language models' novel capabilities and transfigure their teaching benefits towards advancements in the visual domain, which in other domains, falls in line with research in the direction of prompt-based teaching Wu et al. (2022); Sanh et al. (2022).

2 NATURAL LANGUAGE TEMPLATES

We begin our exploration by referencing six templates from already established natural language prompt programming templates Reynolds & McDonell (2021). From there, we find that Metaprompt techniques along with a Memetic proxy and Constraining behavior yield promising results. A description of the templates is provided below:

1. **Memetic proxy:** task specification by prompting alongside a proxy to signify intention using memetic reinforcement techniques such as an archetypal situation.
2. **Constraining behavior:** specified prompting to enforce desired behavior by incorporating emphasis such as punctuation marks.
3. **Metaprompt programming:** prompting seeds encapsulating a more general intention that will unfold into a specific prompt along with additional information such as a prompt wrapper for the task to be solved.

3 LEARNED METAPROMPT TECHNIQUES

To generalize a framework specific to the visual domain, we begin by defining four Metaprompts that encapsulate a prompt as a separate entity. These are **L-Meta:** (Let's ...), **T-Meta:** (This prompt asks us to...), **I-Meta:** (In this prompt, we...), and **S-Meta:** (Shall we draw...). We then conceptualize our technique on a variety of concepts by devising inhabitants of a modern, botanical farm as modes of exploration. These include tomatoes, wheat, a robot, and fungi. Next, we set an exclamation mark as our constraining behavior and Unreal Engine (a popular gaming software known for its pixelated computer graphics) as our memetic proxy. As an inanimate proxy, Unreal Engine overpasses the model's inclination to infer a specified entity within the domain of interest and instead infers to the stylistic features of the methodology. Finally, we present our images generated via the CLIP + VGQAN model Crowson et al. (2022) in Figure 1 below:

Figure 1: | **L-Meta:** *Let's draw tomatoes in the style of an unreal engine!* | **T-Meta** *This prompt asks us to draw wheat in the style of an unreal engine!* | **I-Meta:** *In this prompt, we draw a robot in the style of an Unreal Engine 5!* | **S-Meta:** *Shall we draw fungi in the style of Unreal Engine 5!*

4 Beyond Templates

Metaprompts yield qualitative results yet signify the need for an alternate behavior. For example, while each image in Figure 1 was roughly trained in 1000 iterations, some were chosen interactively during various time stamps in part of the imagery that arose during training. Likewise, given that the resultant imagery is also subject to creativity via the prompt's intention or the prompter's interpretation, we present holistic assessments to garner visual fidelity with the factors outlined below. We also prompt an image as an amalgamation of these factors in Figure 2.

- **The Importance of a Model:** Generated images are subjective to the model responsible for their synthesis and model architectures play a significant role in the resultant fidelity Rombach et al. (2021).
- **Creative Prompt Programming:** This approach encourages a descriptive prompt that can in itself probe the capability of a model towards stark fidelity — with caution on "overfitting" the descriptive limits of details.
- **The Trade-off with Time:** The ability of a model to work at the speed of an artist's thoughts gives greater control to image generation however this tradeoff impacts the quality of the result and a balance must be held.

Figure 2: A Latent Diffusion model Dhariwal & Nichol (2021) with medium speed and mild description prompted beyond templates via *A beautiful VR 3D sci-fi painting of electrical activity in Fungi by Greg Rutkowski and Thomas Kinkade, trending on ArtStation.*

5 Conclusion

We present this study for the scholarly and artistic grounding of prompt programming within the visual domain. As such, we encourage practitioners to emphasize the holistic notion of text-to-image synthesis and explore creative prompting techniques with consideration of descriptive limits. As practicing artists, we also put forth these findings with the underlining ethical care to safeguard the curation and deployment of such generated artifacts Fuchs et al. (2020); Issak & Varshney (2022).

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. Vqgan-clip: Open domain image generation and editing with natural language guidance, 2022. URL https://arxiv.org/abs/2204.08583.

Prafulla Dhariwal and Alex Nichol. Diffusion Models Beat GANs on Image Synthesis. *arXiv:2105.05233 [cs, stat]*, June 2021. URL http://arxiv.org/abs/2105.05233. arXiv: 2105.05233.

Natalia Fuchs, Peter Kirn, Chris Salter, Emilio Vavarella, and Moises H. Valenzula. AI ART MANIFESTO, September 2020. URL https://www.transart.it/wp-content/uploads/2020/09/AI-Art-Manifesto.pdf.

Alayt Issak and Lav R Varshney. Artistic Autonomy in AI Art. In *Proc. of the Thirteenth International Conference on Computational Creativity (ICCC)*, Association of Computational Creativity, 2022.

PromptBase. PromptBase | DALL·E, Midjourney & GPT-3 Prompt Marketplace, 2022. URL https://promptbase.com.

Laria Reynolds and Kyle McDonell. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *arXiv:2102.07350 [cs]*, February 2021. URL http://arxiv.org/abs/2102.07350. arXiv: 2102.07350.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv:2112.10752 [cs]*, December 2021. URL http://arxiv.org/abs/2112.10752. arXiv: 2112.10752.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M. Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Tali Bers, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. Multitask Prompted Training Enables Zero-Shot Task Generalization. *arXiv:2110.08207 [cs]*, March 2022. URL http://arxiv.org/abs/2110.08207. arXiv: 2110.08207.

Teven Le Scao and Alexander M. Rush. How Many Data Points is a Prompt Worth? *arXiv:2103.08493 [cs]*, April 2021. URL http://arxiv.org/abs/2103.08493. arXiv: 2103.08493.

Tongshuang Wu, Michael Terry, and Carrie J. Cai. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. *arXiv:2110.01691 [cs]*, March 2022. URL http://arxiv.org/abs/2110.01691. arXiv: 2110.01691.