

Supplementary Material for MarkovGen: Structured Prediction for Efficient Text-to-Image Generation

Sadeep Jayasumana Daniel Glasner Srikumar Ramalingam Andreas Veit
Ayan Chakrabarti Sanjiv Kumar

Google Research, New York

{sadeep, dglasner, rsrikumar, aveit, ayanchakrab, sanjivk}@google.com

Additional Qualitative Results: In Figure 1 we compare the results of the Early Exit Muse model to the results of the MarkovGen model. This direct comparison helps to illustrate the impact of the MRF model to image quality. We can see that the MarkovGen model with shows clear improvements to complex object structures as well as visual artifacts and textures. In Figure 2 we show generations of the MarkovGen model for a broad variety of text prompts demonstrating the ability of the proposed MarkovGen model to produce high quality images from a broad set of text prompts.

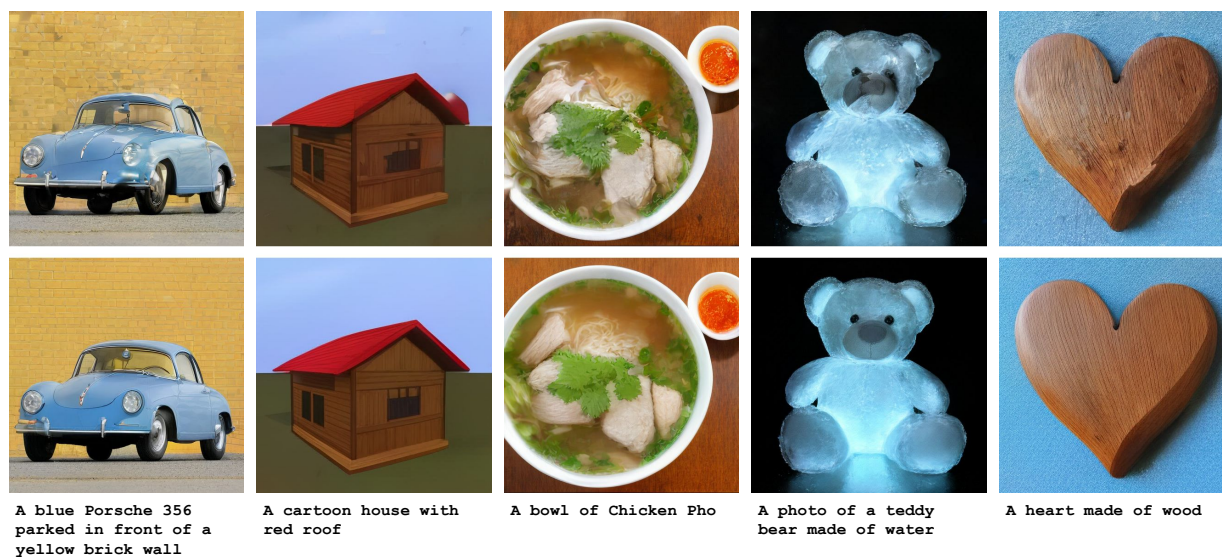


Figure 1. Example generations of the Early Exit Muse super-resolution model running for 3 (out of 8) steps (top) and the MarkovGen model after the application of the MRF model (bottom). We observe a significant reduction in visual artifacts, e.g., in the brick wall behind the car. We further see key improvements to complex object structures such as the blue car and the teddy bear’s face.



A solitary figure shrouded in mists peers up from the cobble stone street at the imposing and dark gothic buildings surrounding it. an old-fashioned lamp shines nearby. oil painting.



A realistic landscape painting, with an incomparable sense of reality. The painting features a super-wide view of a stormy sea with a sailing boat and a wooden boat in the foreground. Huge waves crash against the shore. The sky is filled with stars, and there are



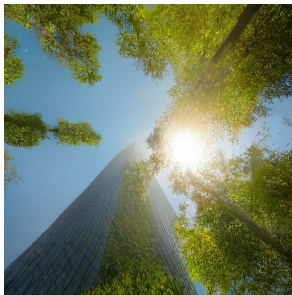
A hyper-realistic, cinematic, and detailed concept art of an alien pyramid landscape, inspired by the works of artists on ArtStation



A photorealistic, and atmospheric art of a ruined temple in a forest. The temple is surrounded by stairs, columns, and mist. The scene is lit by volumetric light, and the composition follows the rule of thirds. The art is rendered in 8K resolution and is trending



A snow-covered arctic landscape with a tent and warm LED lighting. The scene is backlit by the aurora borealis.



A photo of a modern skyscraper covered in trees. sun shining from the top.



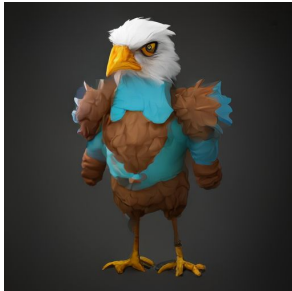
Three-quarters front view of a blue 1977 Corvette coming around a curve in a mountain road and looking over a green valley on a cloudy day.



A photorealistic painting of a blue whale swimming in the ocean, with a school of fish swimming below it!



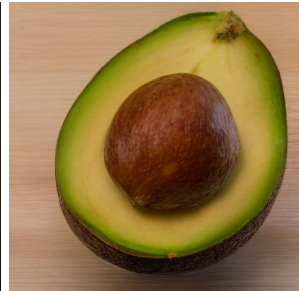
Red apples on a tree with green leaves



A trendy, anthropomorphic bird character for a multiplayer online battle arena (MOBA) game! Rendering.



A raccoon wearing formal clothes, wearing a top hat. The raccoon is holding a garbage bag. Oil painting.



An avocado on a table



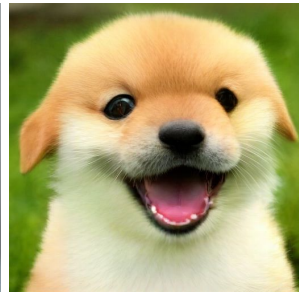
Cute baby lion



A squirrel in a forest



Cute panda



Cute happy puppy

Figure 2. Images generated by MarkovGen for a wide variety of prompts ranging from artistic to natural images.