Factorized Diffusion

Chad Yu Ryan Noonan Andrew Richmond Sahib Manjal Everett Lee

Motivation

Many visual aspects of the physical world can be understood through image decompositions. For example differences in lighting or viewing distance can impact our perception of objects. The goal of our project is to control various factors of an image decomposition, to produce hybrid images.

Our chosen paper, *Factorized Diffusion: Perceptual Illusions by Noise Decomposition* uses many different image factorizations to generate different types of hybrid images.

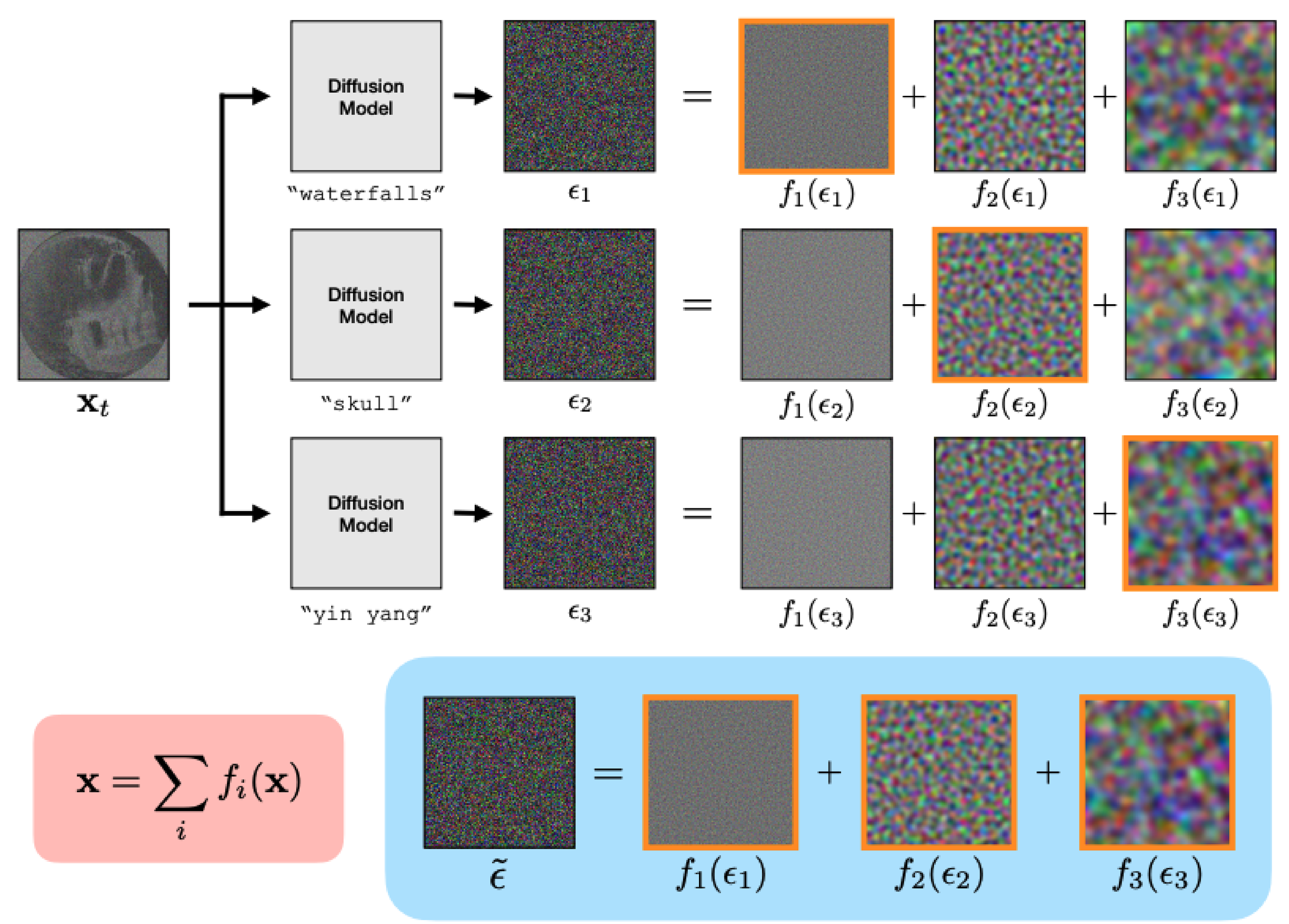
Methodology

Pixel space diffusion model used: DeepFloyd/IF-I-M-v1.0

Python **diffusers** library for integration

Sampling process

- Denoised image at step t: $x_t \rightarrow$ predict different noises for each prompt.
- Factorize each noise estimate, and keep only one of the factors.
- Add the noise factors together, and use this noise to denoise x_t



For example the color factorization is

$$f_{\text{gray}}(\epsilon_{\mathbf{i}}) = \frac{1}{3} \sum_{c \in \{R,G,B\}} (\epsilon_{\mathbf{i}})_c$$

and $f_{\text{color}}(\mathbf{x}) = \mathbf{x} - f_{\text{gray}}(\mathbf{x})$. Formulas for other factorizations are in the original paper. [1]

Results

We implemented the factorizations described in the paper, and generated images using various prompts, sometimes using the prompts from the paper to compare quality, and exploring new prompts. Below is various images we generated labeled with the factorization used in their creation.

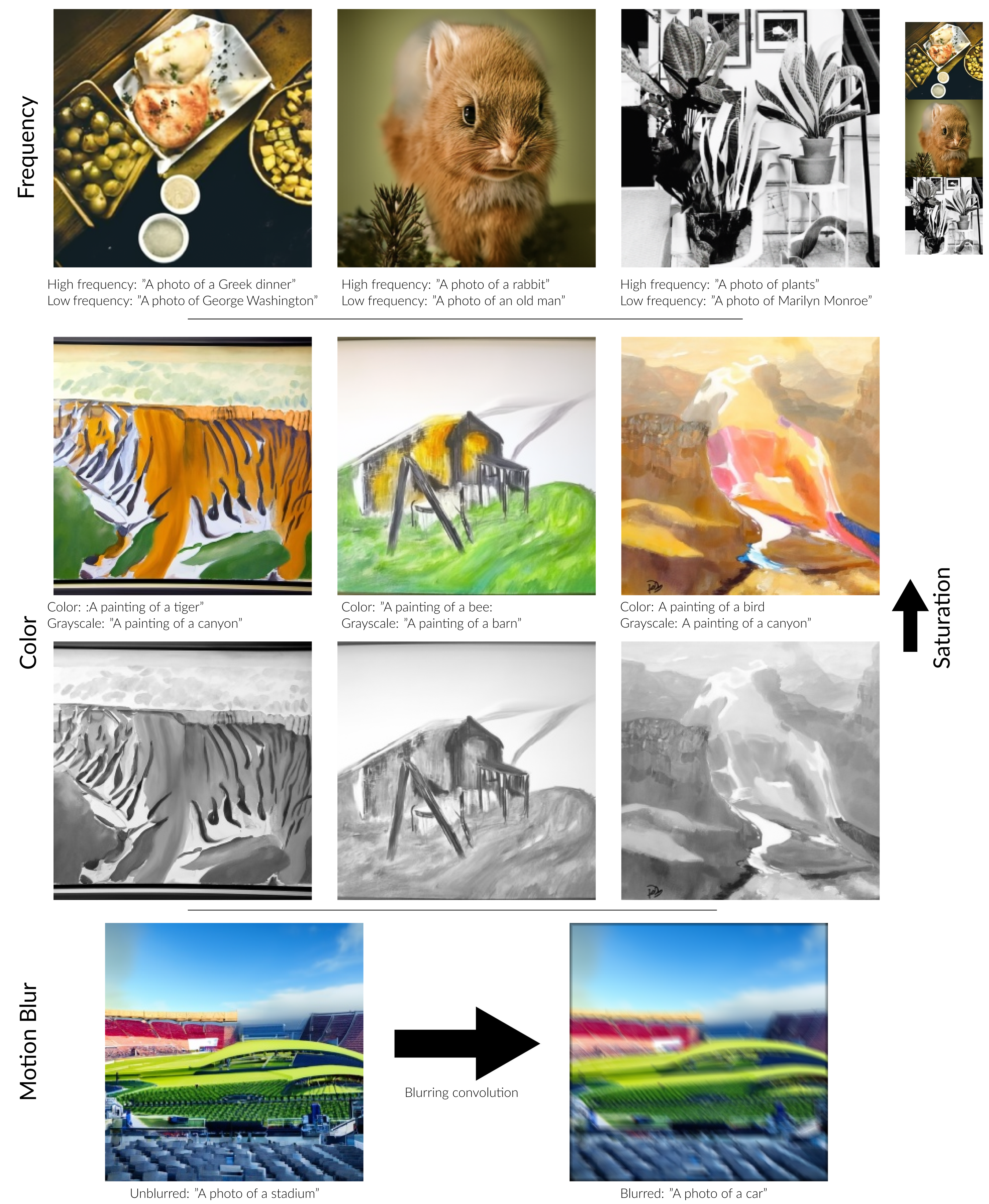
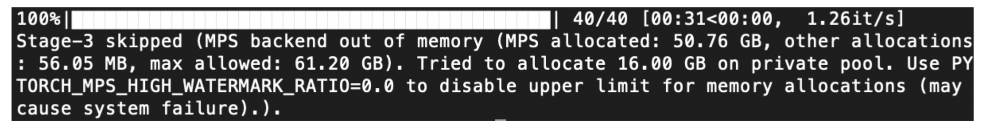


Figure 1. Hybrid images generated using various factor decompositions and prompts

Challenges

Memory: The forward pass in the diffusion step uses very large models (UNets), and our personal GPUs could not handle it, hence we had to switch to CPU, slowing down inference.



Stable Diffusion: We initially used a Stable Diffusion model, which produced blurry non-sensical results. The paper details that operating on the latent space leaves "artifacts".

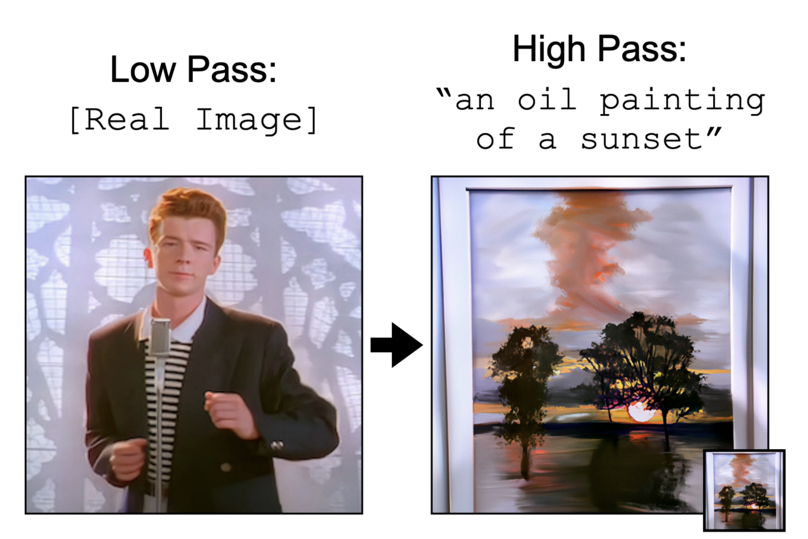
Prompting: The order of the prompts matters: It is easy to make Marilyn Monroe out of plants but hard to make plants out of Marilyn Monroe. Here is a high frequency Marilyn Monroe and low frequency plants

Conclusion

It has been known that diffusion models have high image quality output, even on complex prompts. However, these methods apply existing diffusion models to hybrid tasks, something individual diffusion models are unable to do directly. This shows the versatility of the use cases of diffusion models.

Future Work

- **Addressing Current Limitations:** Diffusion models are currently quite limited. Techniques like this can expand upon them to access generation 'spaces' we can't currently reach.
- **Inverse Hybrids:** The paper [1] includes Inverse Hybrids - a method to generate hybrid images based off of some fixed reference image.



References

[1] Daniel Geng, Inbum Park, and Andrew Owens. Factorized diffusion: Perceptual illusions by noise decomposition, 2025.