

# Manipulating Embeddings of Stable Diffusion Prompts

---

IJCAI 2024

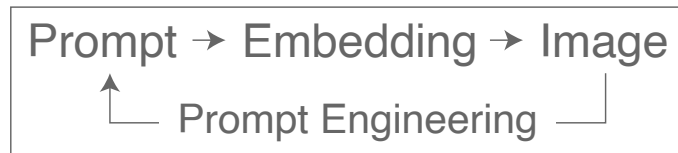
**Niklas Deckers**, Julia Peters, Martin Potthast

Leipzig University, ScaDS.AI, University of Kassel, hessian.AI



# Prompt Engineering Means Trial and Error

- ❑ Text-to-image models allow users to directedly control the output
- ❑ Finding the right prompt is difficult
- ❑ Trial and error
- ❑ Frustrating, inaccessible



# User Stories

1. Users want to achieve certain aesthetics: Prompt modifiers like  
4k high resolution award winning image
2. Users seek inspiration based on a given prompt:  
Expanding the prompt with keywords
3. Users try to verbalize aspects seen when using a certain seed:  
Expanding the prompt with details

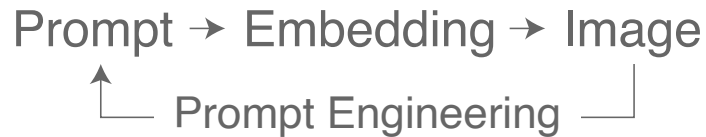
# Prompt Embedding Space

- Prompt embedding space is continuous
- Allowing manipulations like interpolation and gradient-based methods



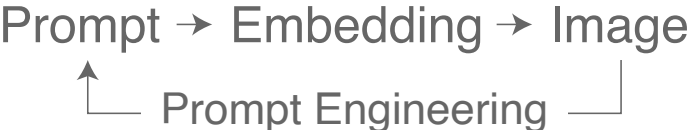
# Manipulating Prompt Embeddings

- Allow users to directly manipulate prompt embeddings instead of the prompt
- We introduce three methods for directed prompt embedding manipulation



# Manipulating Prompt Embeddings

- Allow users to directly manipulate prompt embeddings instead of the prompt
- We introduce three methods for directed prompt embedding manipulation



# (1) Optimization Direction by Image Quality Metric

Prompt:  
Autumn  
lake at  
dusk with  
sun ball.



Update by  
gradient descent or  
ascent (metric-dependent)



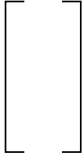
5.8

Metric  
(e.g., aesthetics)

7.0

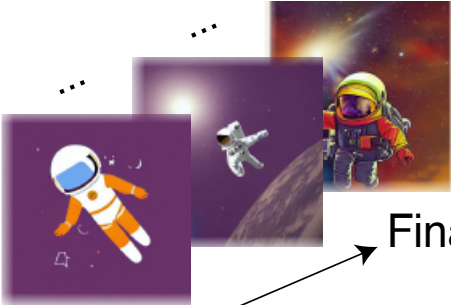
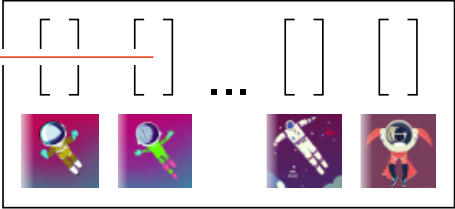
# (2) Optimization Direction by Human Feedback

Prompt:  
Flying  
comic  
astronaut.



Random variants

Update by  
interpolation



Initial






Final

Refinement through manual selection among variants.




## (2) Optimization Direction by Human Feedback

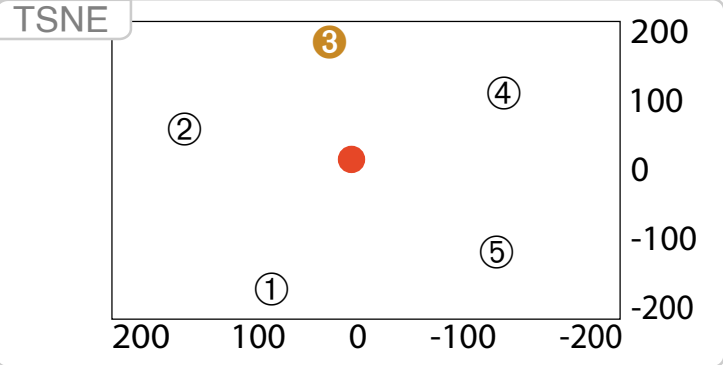
1. Initialization    2. Image Selection    3. History

1  2  3  4  5 

Interpolation Value

**Generate**

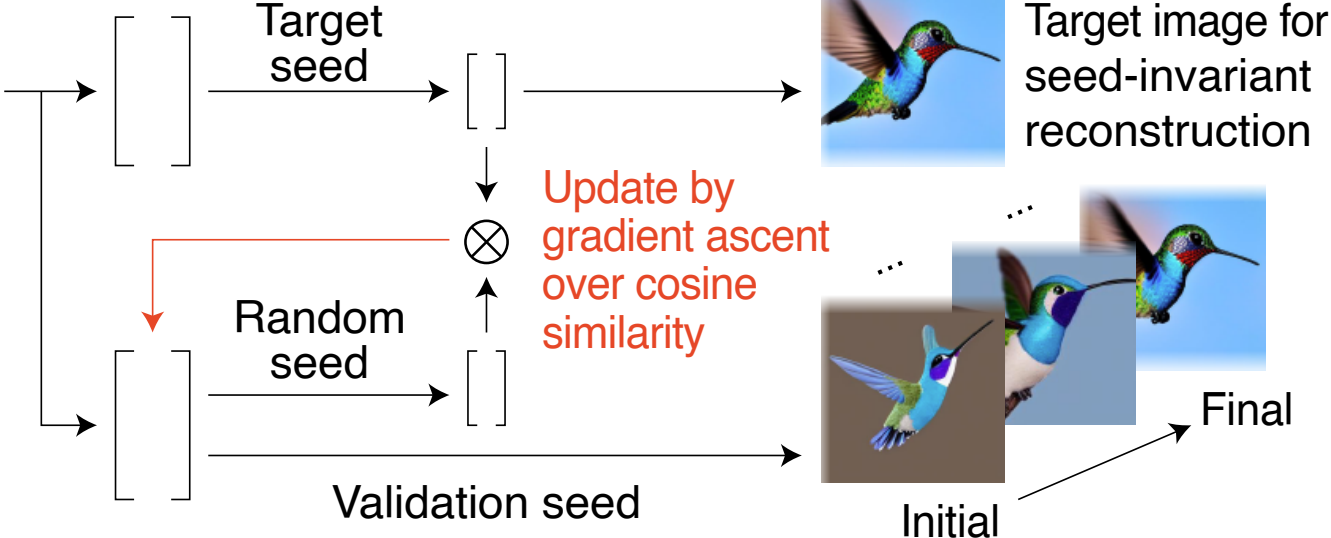
Current 

TSNE 

The TSNE plot shows the relationship between the five input images. The x-axis ranges from 200 to -200, and the y-axis ranges from -200 to 200. Points 1, 2, 4, and 5 are clustered together, while point 3 is positioned significantly higher on the y-axis. A red dot, representing the current state, is located near the center of the plot, between points 2 and 3.

# (3) Optimization Direction by Target Image

Prompt:  
Colorful  
Humming-  
bird.



# Results: Metric-Based Optimization

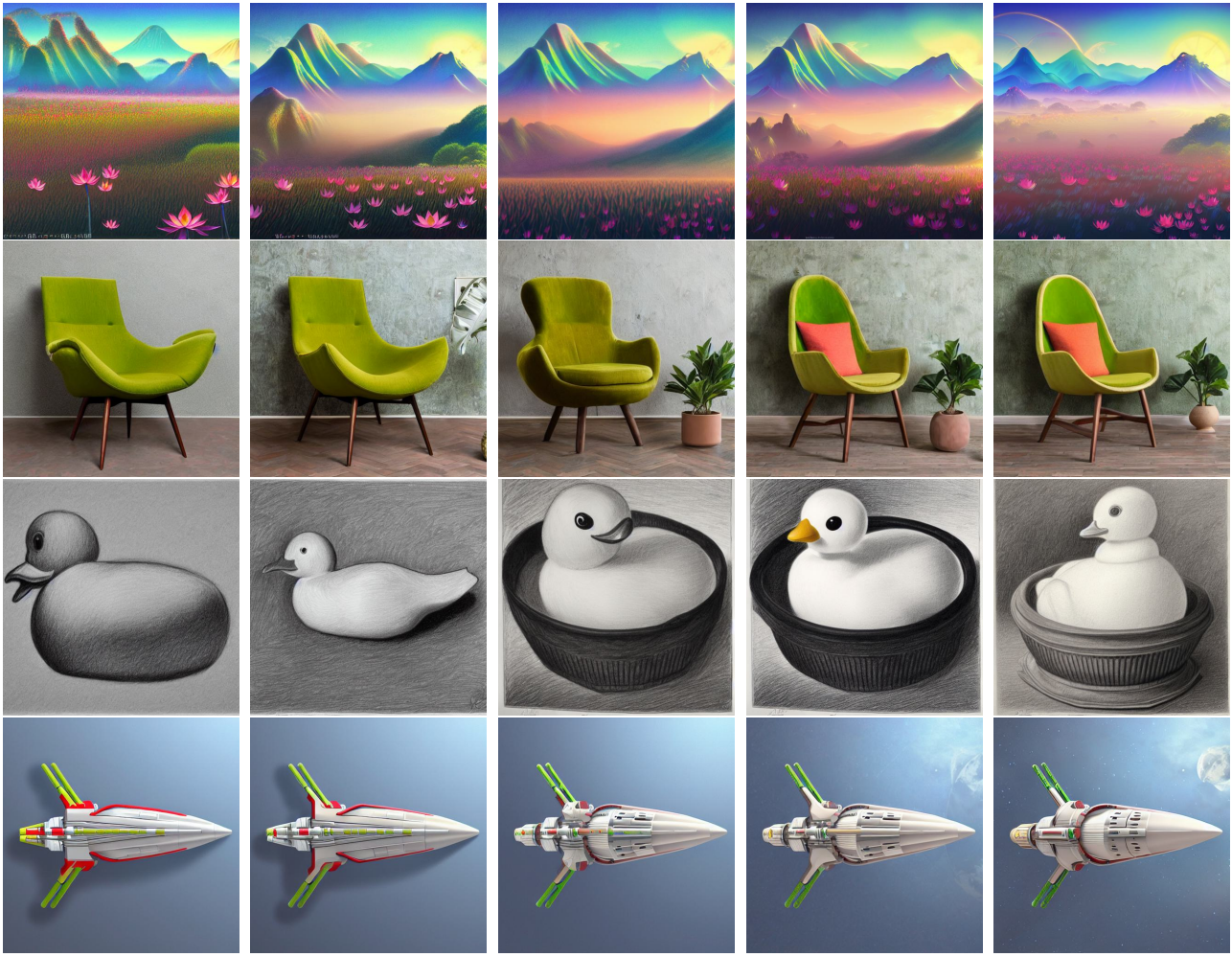


Prompt

—— Metric: ▲ blurriness ▼ sharpness ——▷



# Results: Metric-Based Optimization



Prompts 1-4

————— Metric: aesthetics —————>



# Results: Optimization Based on Human Feedback



Prompts 1-3

———— Prompt engineering ———>

# Results: Optimization Based on Human Feedback



Prompts 1-3

————— Our method —————>