SEGA:
Instructing Text-to-Image Models using Semantic Guidance

TECHNISCHE UNIVERSITÄT DARMSTADT · dfki ai · hessian.AI


Manuel Brack


Felix Friedrich


Dominik Hintersdorf


Lukas Struppek


Patrick Schramowski


Kristian Kersting

NEURAL INFORMATION PROCESSING SYSTEMS

Semantic Control over Diffusion Models

- Producing **high-fidelity** images with text-to-image models is an **iterative, interactive** process
- Users want to make **changes** until **satisfied** with the outputs
- Fine-grained **semantic control** is needed which should be as **easy to use** and **versatile**

Method	Manipulation Success
Disentanglement (Wu, 2022)	41.35 %
Prompt2Prompt (Hertz, 2023)	43.25 %
Composable Diffusion (Liu, 2022)	60.50 %
SEGA (Ours)	**72.72 %**

User Preference

Semantic Guidance **83%** vs **13%** Composable Diffusion

Original

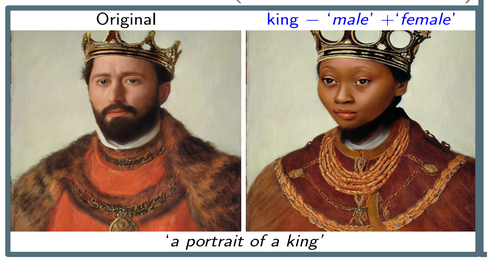
+ *'van Gogh's Starry Night'*
+ *'boat in the river'*

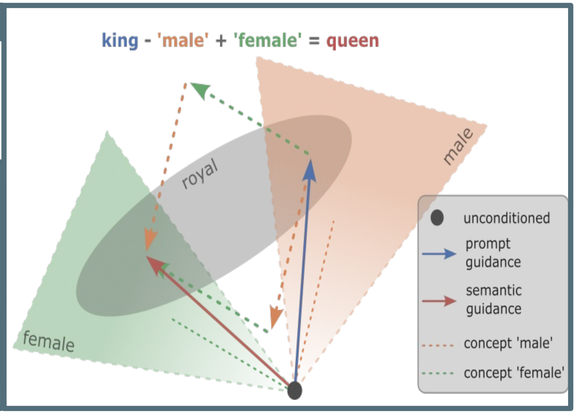
Semantic Guidance

Semantic Guidance

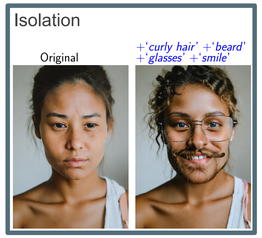
$$\overline{\epsilon_\theta}(z_t, c_p, c_e) = \epsilon_\theta(z_t) + s_g\left(\epsilon_\theta(z_t, c_p) - \epsilon_\theta(z_t)\right) + \gamma(z_t, c_e)$$

$$\gamma(z_t, c_e) = \mu(\psi; s_e, \lambda)\left(\pm\left(\epsilon_\theta(z_t, c_e) - \epsilon_\theta(z_t)\right)\right)$$

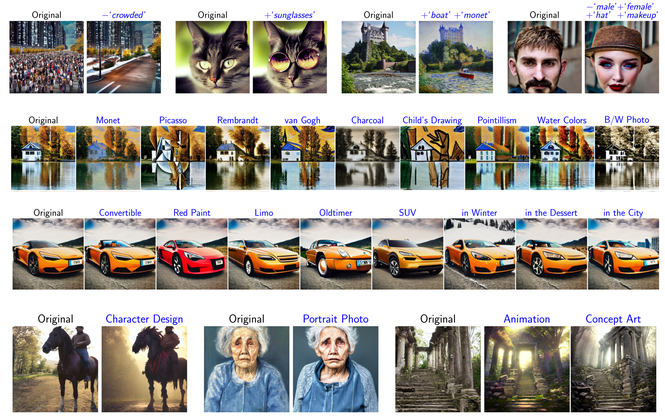




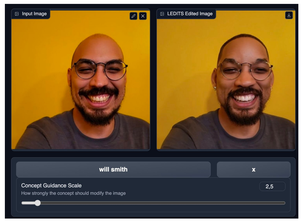
Properties



Applications



Follow up work: Real Image Editing with LEdits++



Workshop on ML for Creativity and Design

- Sat 16 Dez.
- Talk: 9:20 am
- Poster: 1:30 pm

Failure when changing input prompt & coherent manipulation

"A photo of a cat"

"A photo of a cat wearing sunglasses"

"A photo of a cat"

+SEGA: *"sunglasses"*

Code

Integrated in Diffusers

Demo

Conclusion

- SEGA enables **semantic control** over image generation
- at **inference**, with **no training**, and **architecture-agnostic**

Check out the Project Page