🎨 *Omni-Attribute:*

Open-vocabulary Attribute Encoder for Visual Concept Personalization

Tsai-Shien Chen[1,2] Aliaksandr Siarohin[1] Gordon Guocheng Qian[1] Kuan-Chieh Jackson Wang[1] Egor Nemchinov[1]
Moayed Haji-Ali[1] Riza Alp Guler[1] Willi Menapace[1] Ivan Skorokhodov[1] Anil Kag[1] Jun-Yan Zhu[3] Sergey Tulyakov[1]

[1]Snap Inc. [2]UC Merced [3]CMU

CVPR JUNE 3-7, 2026 DENVER COLORADO

Copy-Paste Artifacts

Phantom [ICCV'25]

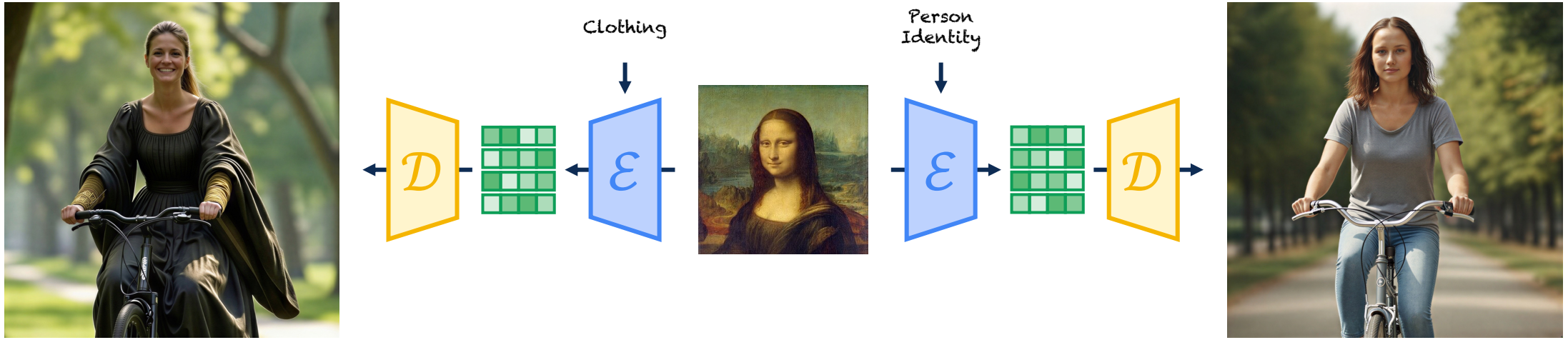
Movie Weaver [CVPR'25]

BindWeave [arXiv'25]

Can we prevent such information leakage at the encoder stage?

Can we learn an attribute encoder that only extracts the information of the user-specified attributes?

Open-vocabulary Image Attribute Encoder

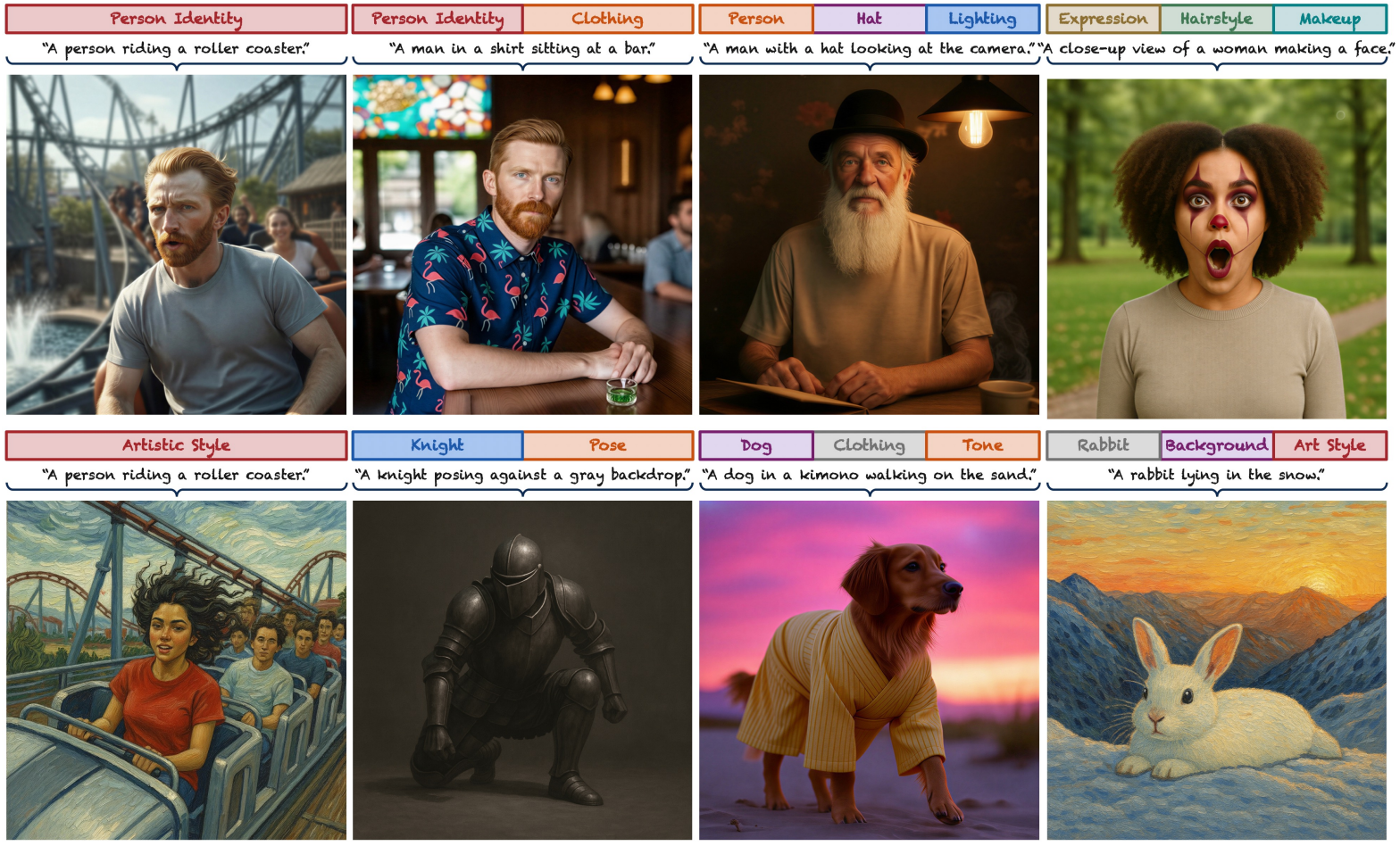


Interactive Demo

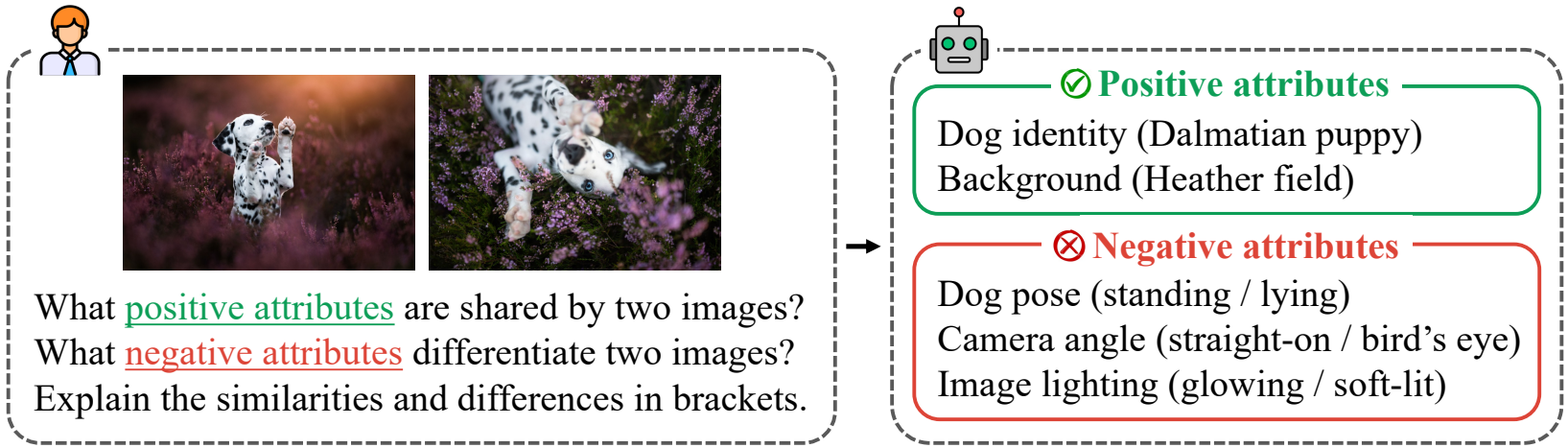
"A vase is standing against a plain background."

Try it yourself on the project webpage!

Attribute-level Image Composition

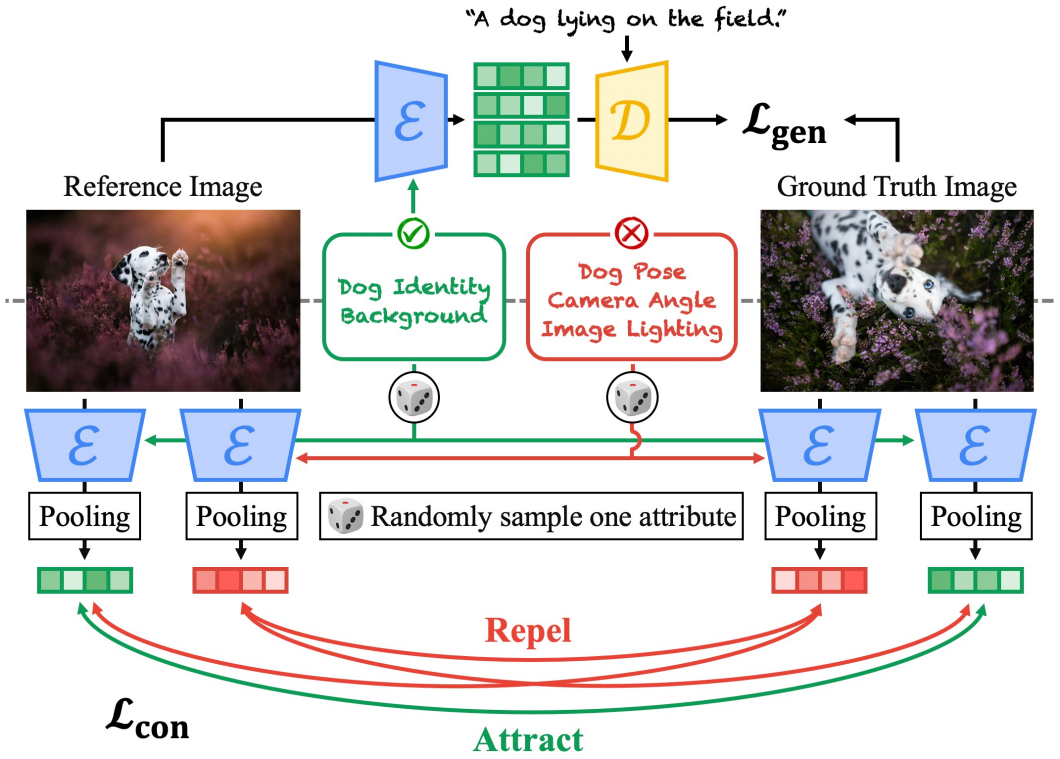


Semantic Connections between Image Pairs



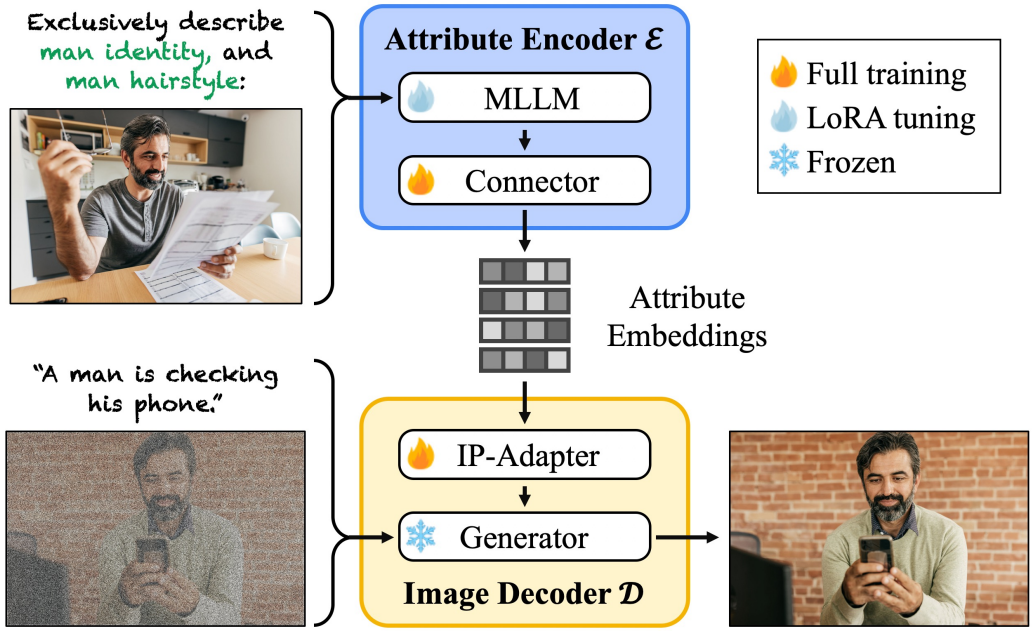
Attribute-level Representation Learning

[Goal 1] The embeddings need to capture sufficient information to ensure high-fidelity reconstruction of the target attributes.
[Goal 2] The embeddings need to remove irrelevant information from other attributes.

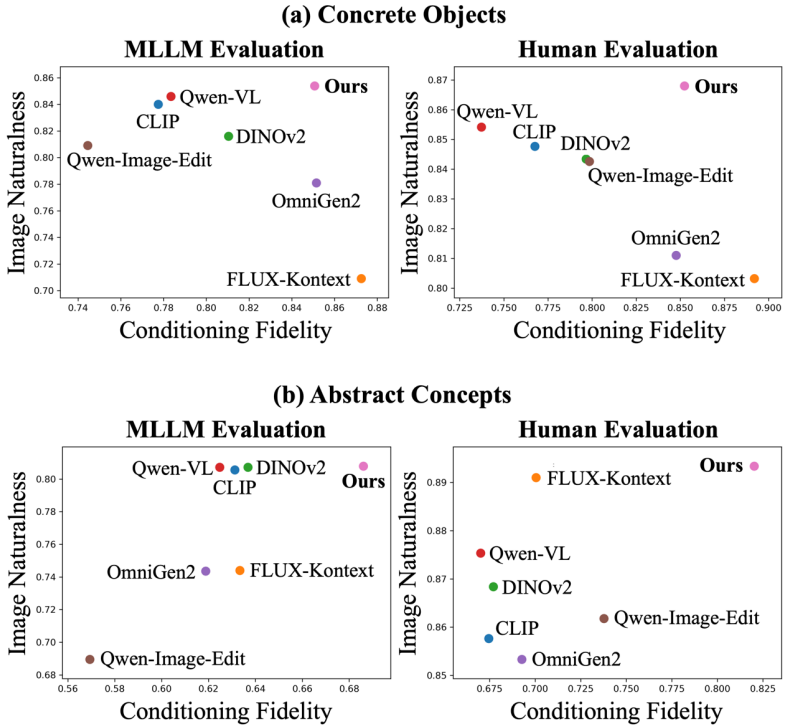


MLLM-based Attribute Encoder

[Goal 1] Capability to jointly process both image and text inputs.
[Goal 2] Strong vision-language prior to support attribute disentanglement.



Open-vocabulary Attribute Personalization



*For image encoders, we train IP-Adapter modules between each encoder and the same frozen image generation backbone.

*For image editing models, we reformulate the prompt as "*Preserve the <attribute> of the image and generate <prompt>*" to enable attribute personalization.

Composability of Attribute Embeddings

Motivated by Composable Diffusion [ECCV'22]:

$$\hat{\epsilon}(\boldsymbol{x}_t, t) = \underbrace{\epsilon_\theta(\boldsymbol{x}_t, t)}_{\text{Unconditional prediction}} + \sum_{i=1}^{n} w_i \underbrace{\big(\epsilon_\theta(\boldsymbol{x}_t, t|\boldsymbol{c}_i) - \epsilon_\theta(\boldsymbol{x}_t, t)\big)}_{\text{Conditional score direction (cond pred – uncond pred)}}$$

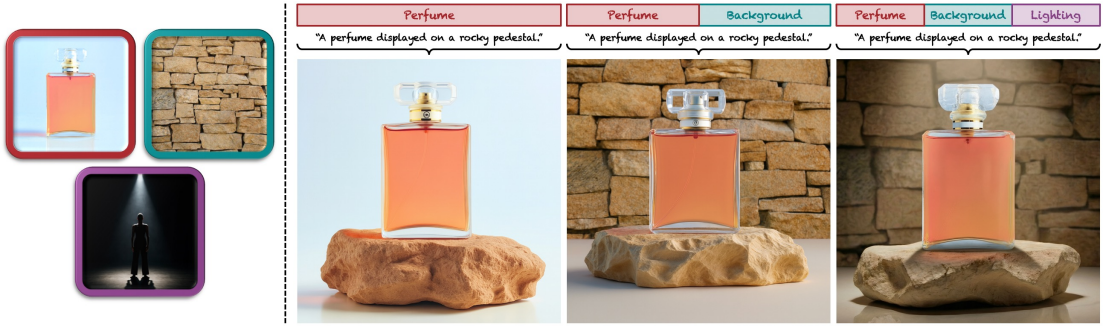
Adaption to Flow-matching Model:

Conditional flow field of **single** image-attribute condition: $\Delta_{(I_i, a_i)} = \mathcal{D}(\mathcal{E}(I_i, a_i), \varnothing) - \mathcal{D}(\varnothing, \varnothing)$

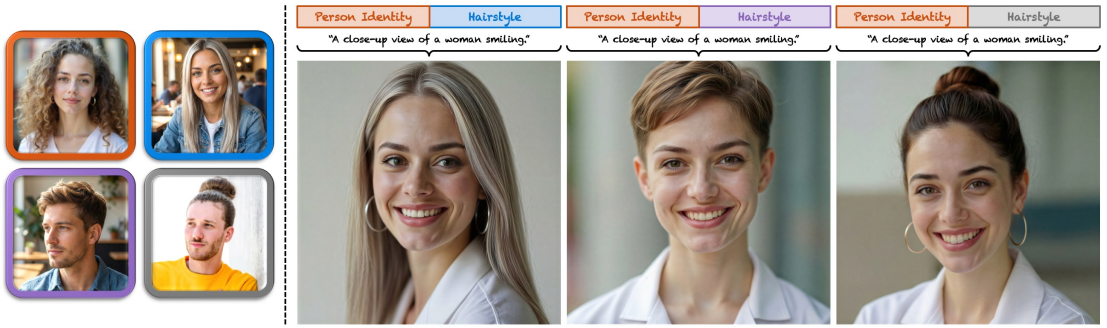
Composition of **multiple** image-attribute conditions: $v^* = \mathcal{D}(\varnothing, c) + \sum_{i=1}^{N} w_i \cdot \Delta_{(I_i, a_i)}$

Practical Applications

Advertisement Image Synthesis

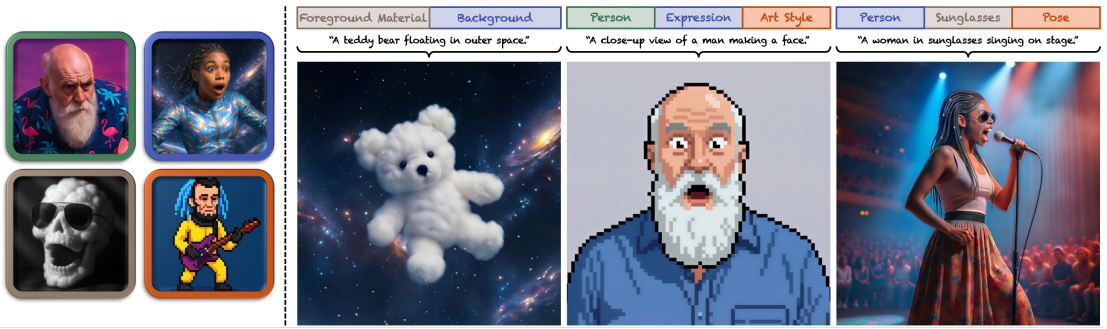


Hairstyle Customization



Storytelling Visualization

Creative Content Generation



🎨 *Omni-Attribute:*

Open-vocabulary Attribute Encoder for Visual Concept Personalization

Project Page

CVPR JUNE 3-7, 2026 DENVER COLORADO