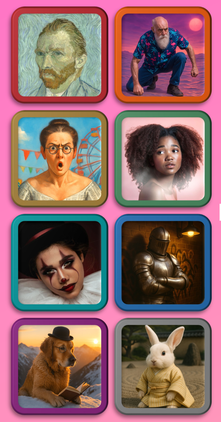
# 🎨 Omni-Attribute:
## Open-vocabulary Attribute Encoder for Visual Concept Personalization

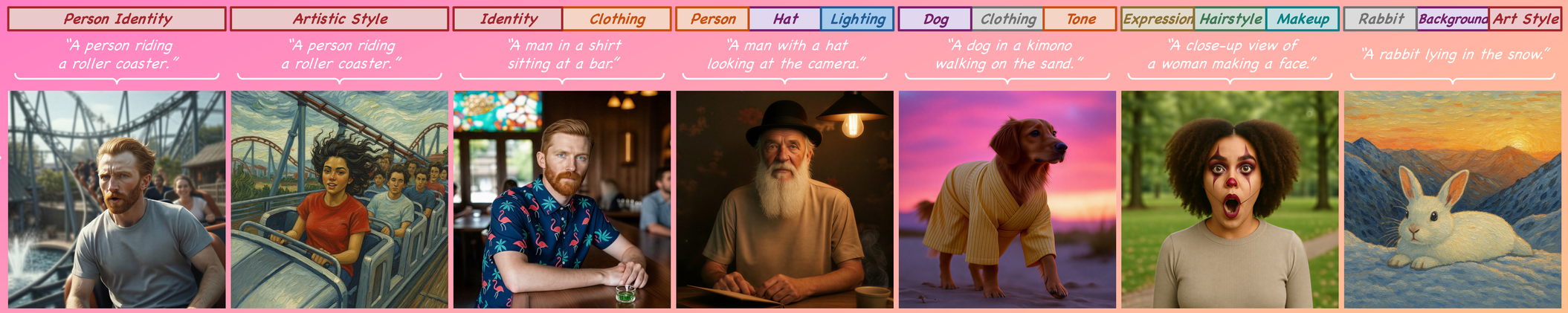

Tsai-Shien Chen[1,2] Aliaksandr Siarohin[1] Gordon Guocheng Qian[1] Kuan-Chieh Jackson Wang[1] Egor Nemchinov[1]
Moayed Haji-Ali[1] Riza Alp Guler[1] Willi Menapace[1] Ivan Skorokhodov[1] Anil Kag[1] Jun-Yan Zhu[3] Sergey Tulyakov[1]

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


CVPR JUNE 3-7, 2026 DENVER COLORADO

🔥 **Tsai-Shien Chen** is seeking **Industry Research** roles 🔥 **2025 Google PhD Fellowship** | **Image/Video/World Models** | **First author of Panda-70M & Video Alchemist**






Personal Webpage

Project Webpage

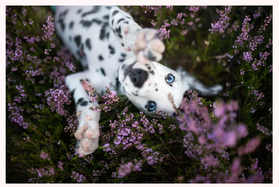
## Motivation

☹️ *Copy-Paste Artifacts*

Video Alchemist [CVPR'25]

Phantom [ICCV'25]

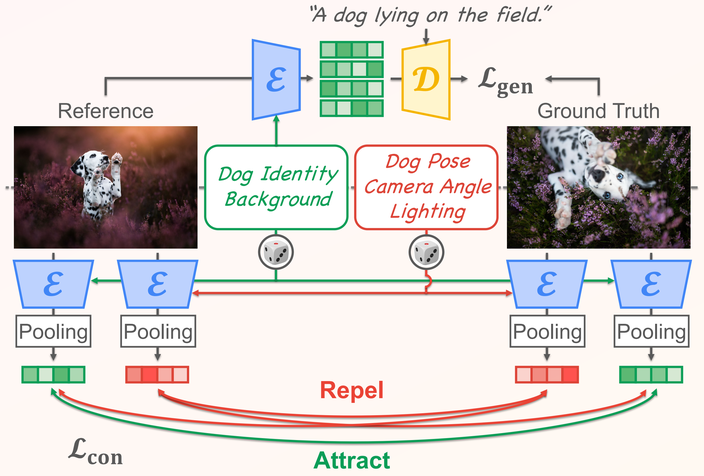
## Semantic Connections between Images



What positive attributes are shared by two images?
What negative attributes differentiate two images?

✅ Positives: dog identity / background
❌ Negatives: dog pose / camera angle / lighting

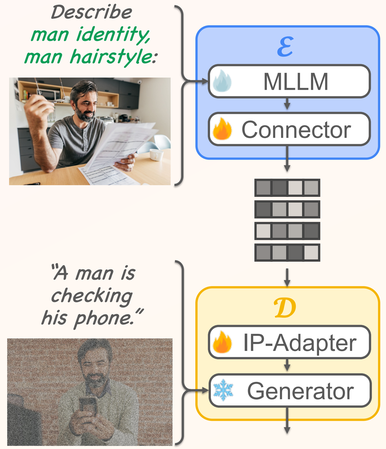
## Attribute-level Representation Learning




$\mathcal{L}_{gen}$

$\mathcal{L}_{con}$

## Model Architecture




## Composition of Attributes

Conditional flow field of **single** image-attribute condition:

$$\Delta_{(I_i,a_i)} = \mathcal{D}(\mathcal{E}(I_i, a_i), \emptyset) - \mathcal{D}(\emptyset, \emptyset)$$

Composition of **multiple** image-attribute conditions:

$$v^* = \mathcal{D}(\emptyset, c) + \sum_{i=1}^{N} \omega_i \cdot \Delta_{(I_i,a_i)}$$

$I_i$ : Reference image
$a_i$ : Reference attribute
$c$ : Text prompt