



¹Snap Inc.



²UC Merced



³CMU



Omni-Attribute:

Open-vocabulary Attribute Encoder for Visual Concept Personalization

Tsai-Shien Chen^{1,2} Aliaksandr Siarohin¹ Gordon Guocheng Qian¹ Kuan-Chieh Jackson Wang¹ Egor Nemchinov¹
Moayed Haji-Ali¹ Riza Alp Guler¹ Willi Menapace¹ Ivan Skorokhodov¹ Anil Kag¹ Jun-Yan Zhu³ Sergey Tulyakov¹

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



Person Identity "A person riding a roller coaster."	Artistic Style "A person riding a roller coaster."	Identity "A man in a shirt sitting at a bar."	Clothing "A man with a hat looking at the camera."	Person Hat Lighting "A dog in a kimono walking on the sand."	Dog Clothing Tone "A close-up view of a woman making a face."	Expression Hairstyle Makeup "A rabbit lying in the snow."



Personal Webpage



Project Webpage

Motivation



Copy-Paste Artifacts

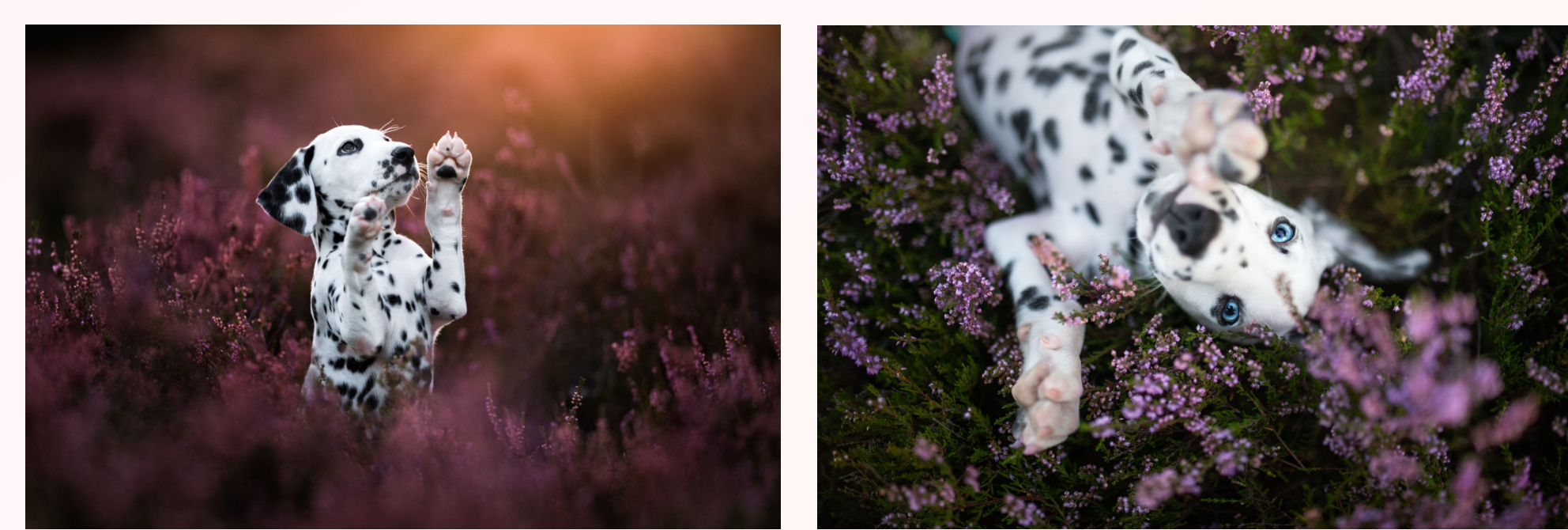


Video Alchemist [CVPR'25]

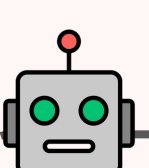


Phantom [ICCV'25]

Semantic Connections between Images

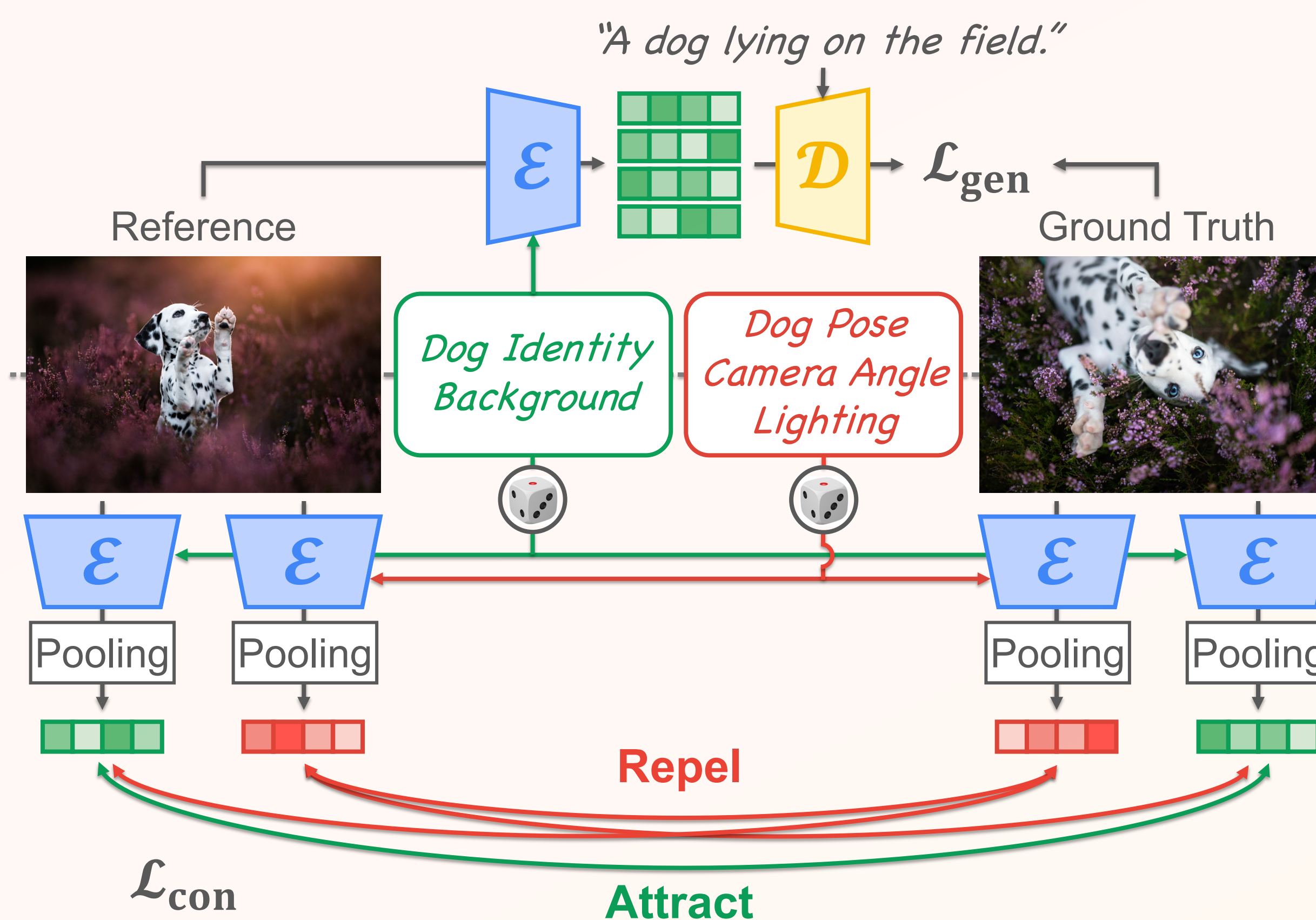


What **positive attributes** are shared by two images?
What **negative attributes** differentiate two images?

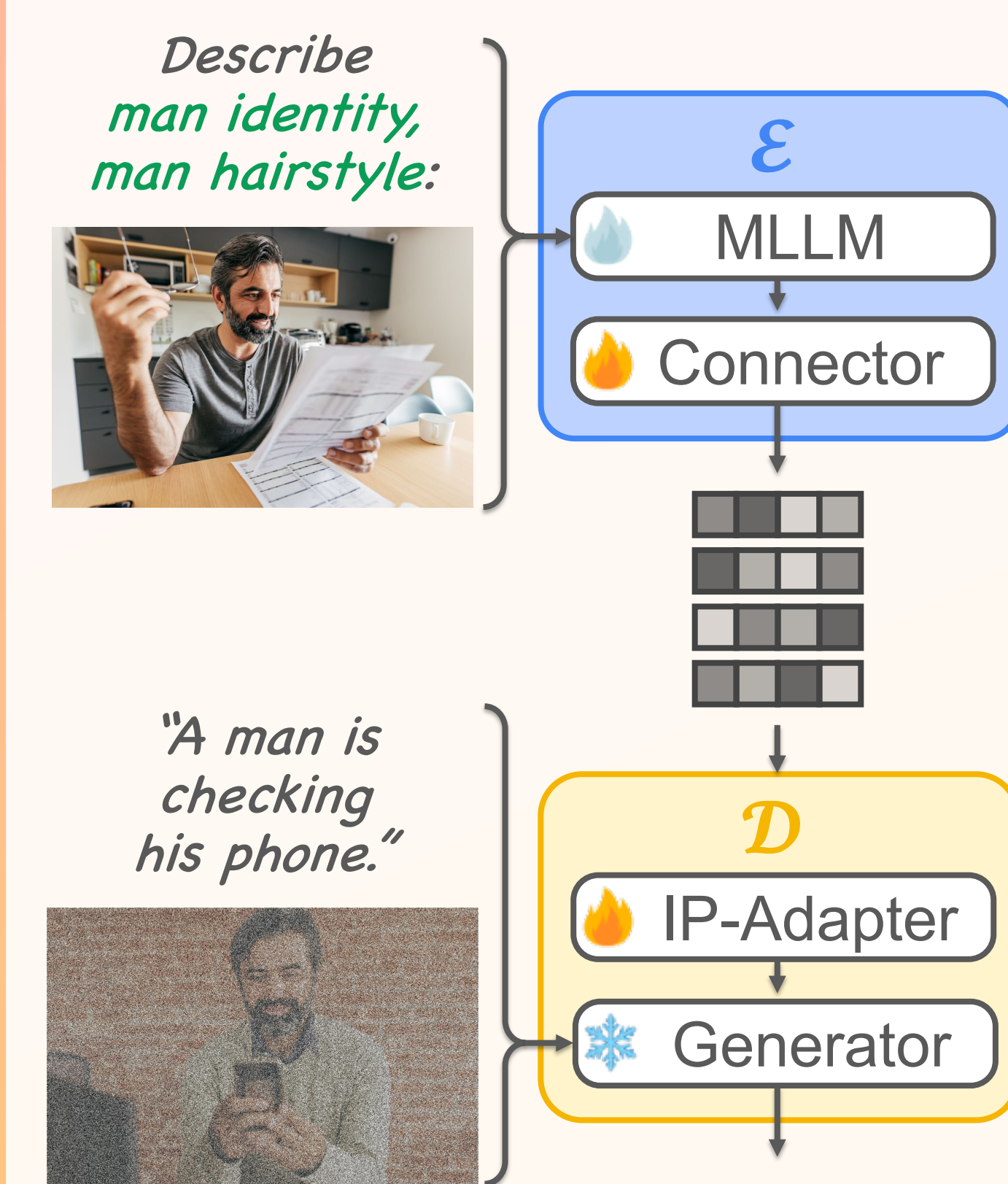


- ✔ Positives: dog identity / background
- ✘ Negatives: dog pose / camera angle / lighting

Attribute-level Representation Learning



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