

---

# Learning Input-Agnostic Manipulation Directions in StyleGAN with Text Guidance

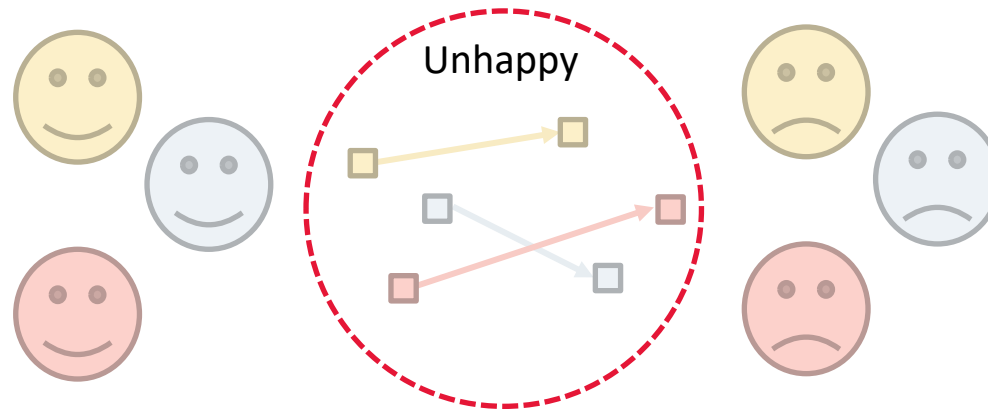
Yoonjeon Kim, Hyunsu Kim , Junho Kim , Yunjey Choi, Eunho Yang

ICLR 2023

# Recall: Text-Guided Manipulation with StyleGAN

---

- **Input-Dependent method requires...**
  - Different directions for every Image & Text pair
  - Time consuming

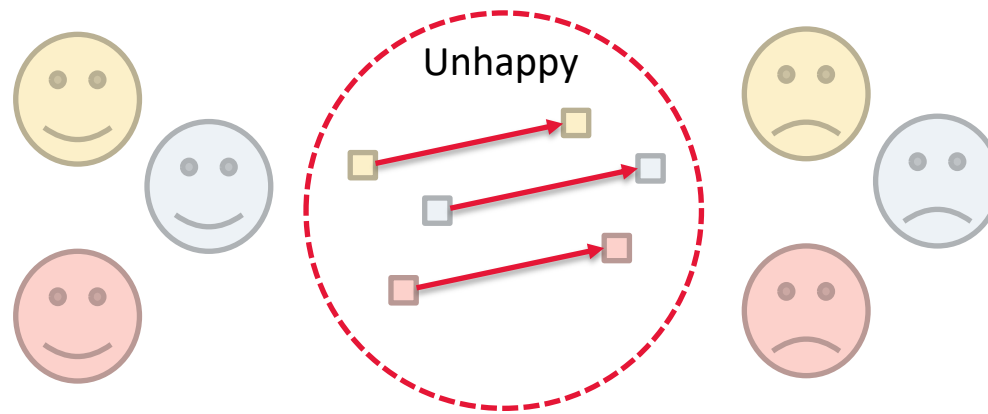


# Recall: Text-Guided Manipulation with StyleGAN

---

- **Input-Agnostic method requires...**

- Universal direction for one text
- Efficient



# Recall: Text-Guided Manipulation with StyleGAN

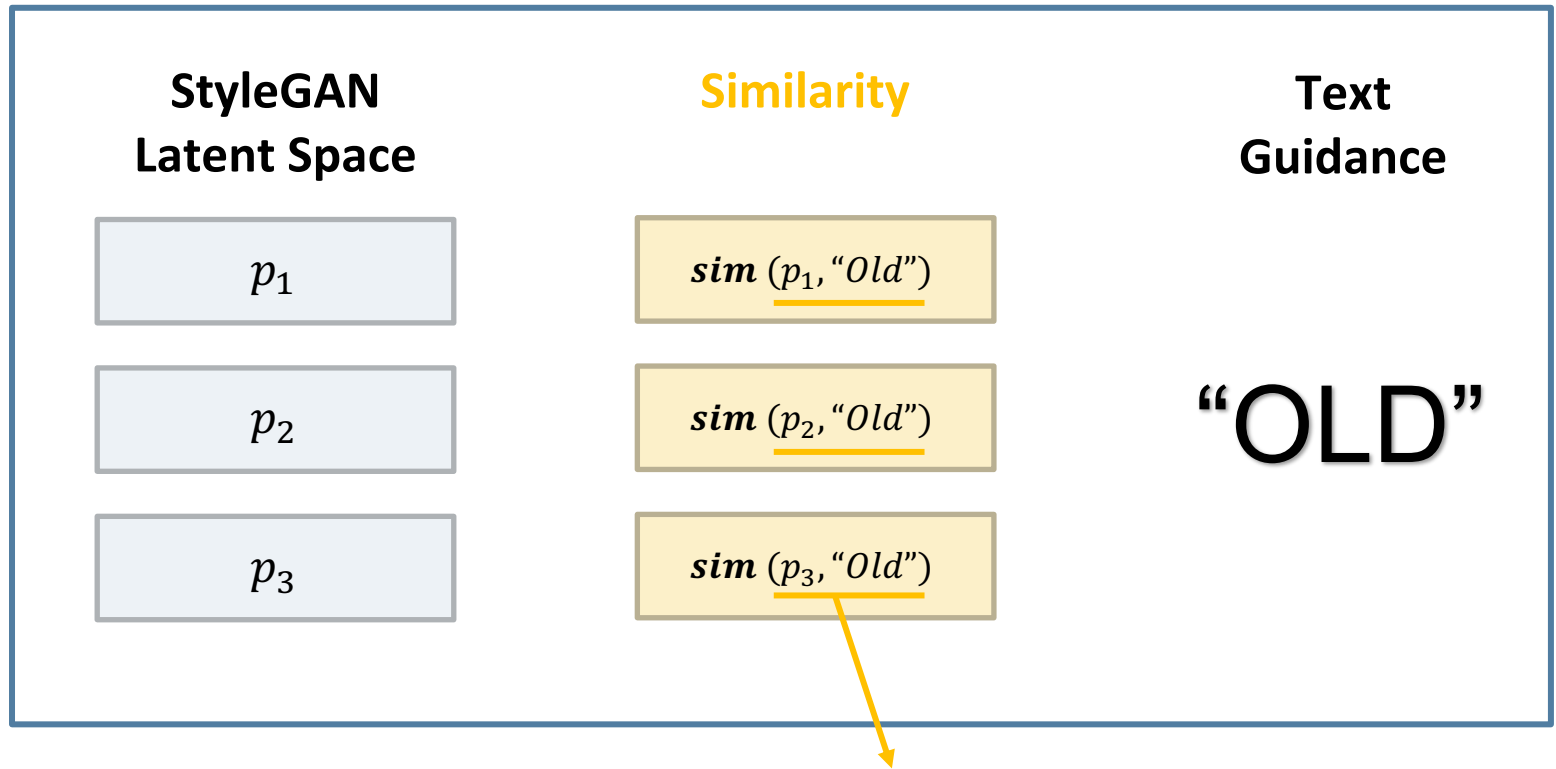
---

- **Two methods on Discovery of Input-Agnostic Direction**
  - Global Mapper<sup>1)</sup>:
    - needs optimization for every text prompt
  - Global Direction<sup>1)</sup>:
    - finds a universal direction of a given text instantly

1) StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery (ICCV 2021)

# Recall: StyleCLIP Global Direction

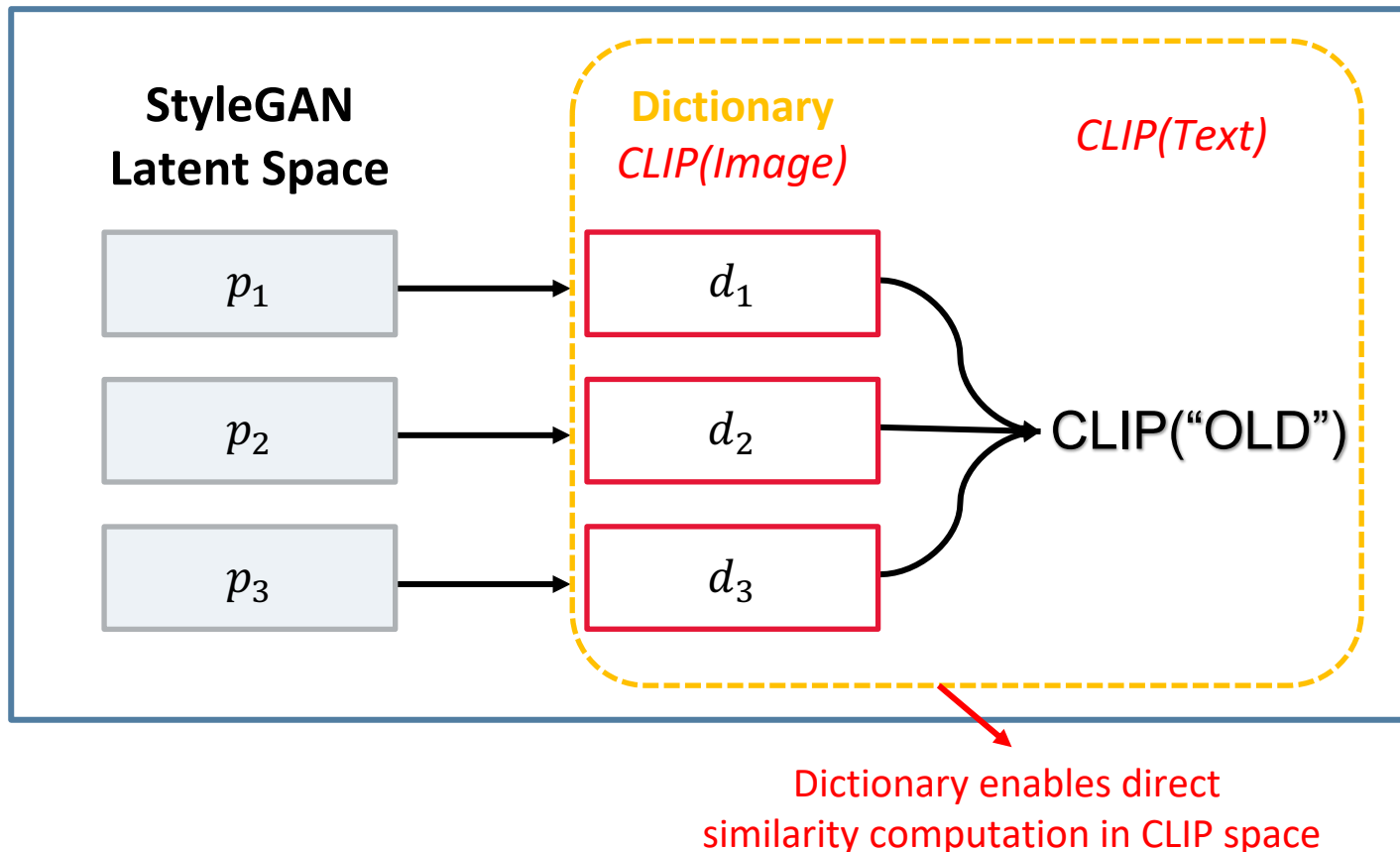
- Similarity between latent space parameter and text guidance



Similarity between StyleGAN Latent Space & Text cannot be computed directly!!

# Recall: StyleCLIP Global Direction

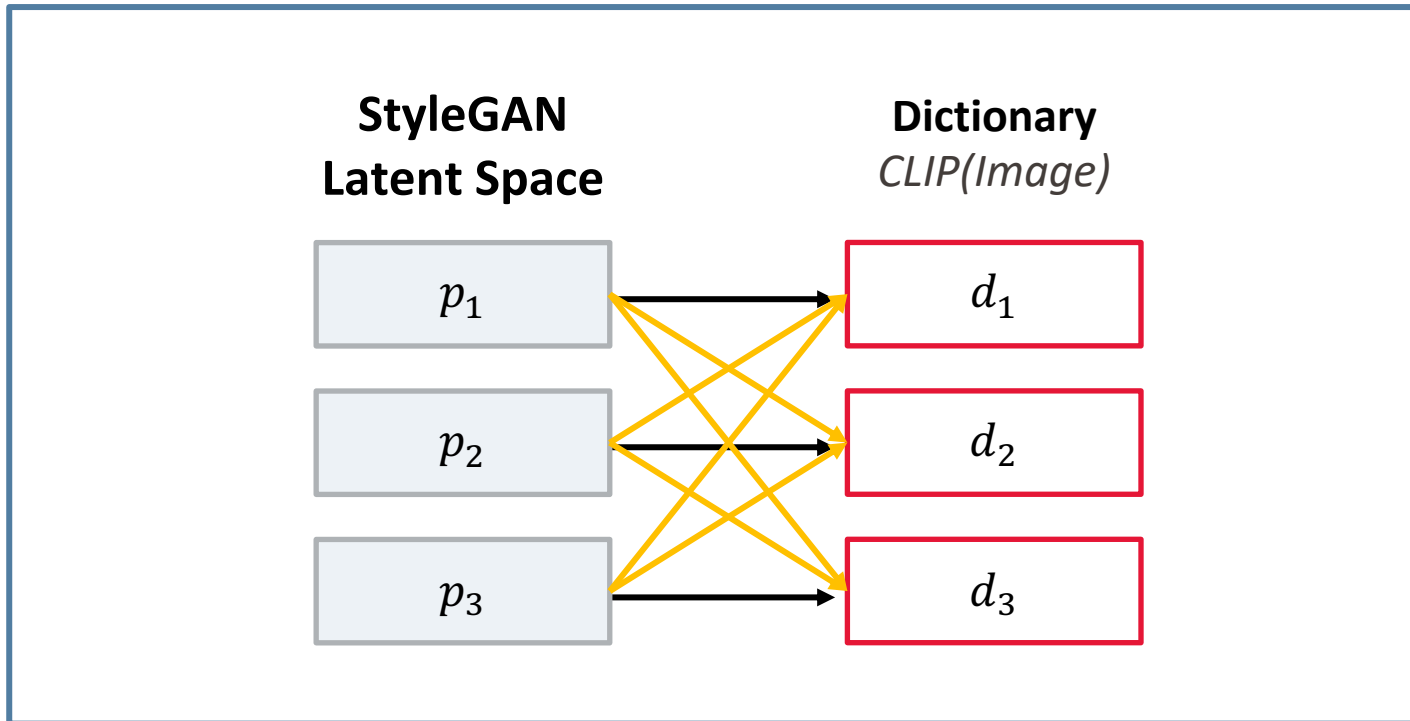
- Global Direction<sup>1)</sup> uses precomputed latent space **dictionary**



1) StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery (ICCV 2021)

# Problem of StyleCLIP Global Direction

















- **Global Direction<sup>1)</sup> computes single-channel CLIP representation**
  - Ignores **multi-channel interaction** in latent space!



1) StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery (ICCV 2021)

# Importance of Multi-Channel Interaction in Latent Space

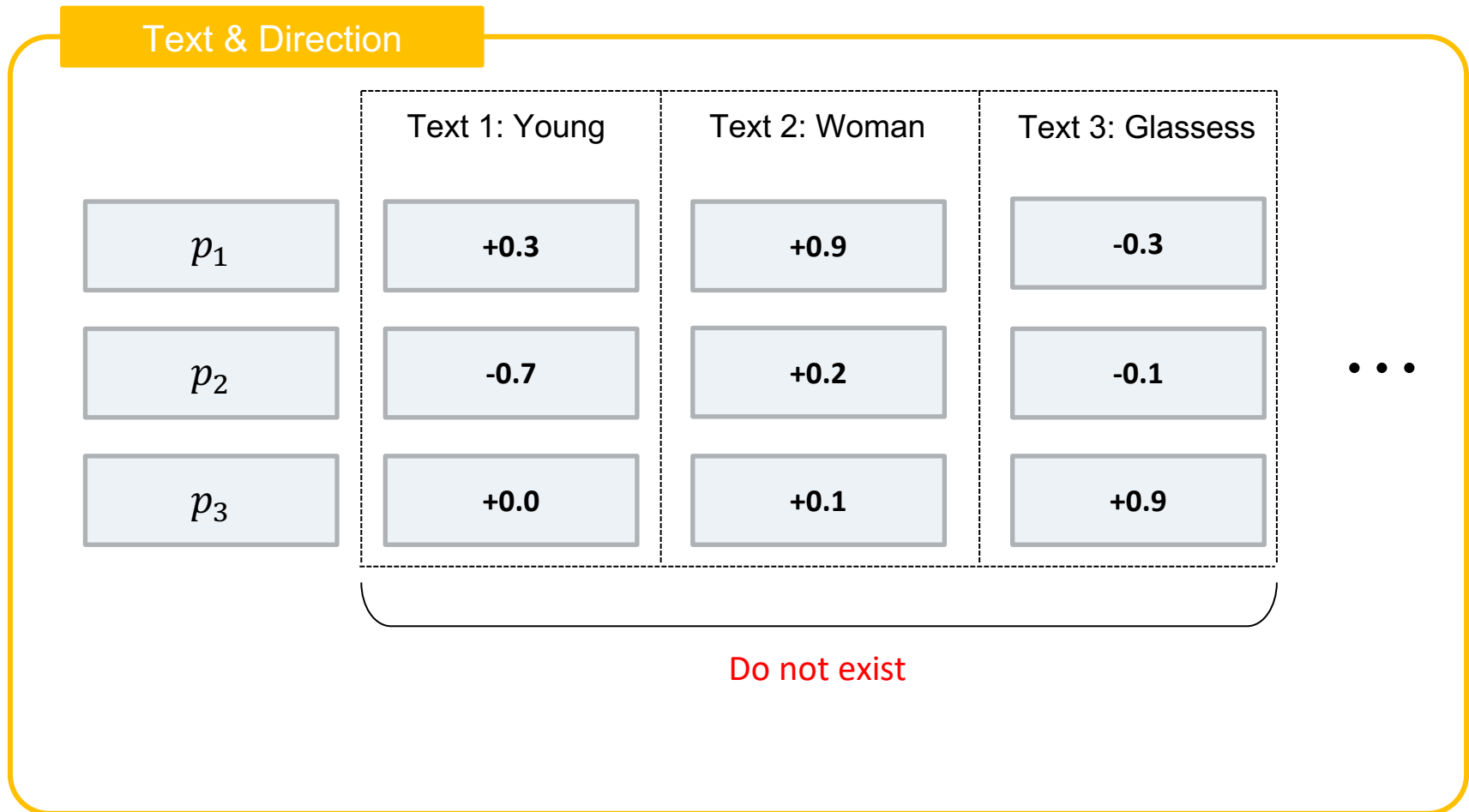
- Manipulation on channels in StyleGAN Latent Space

	Channel Index	Source	Manip.	Source	Manip.
No change	4520 (single)				
+					
Blondish hair	4426 (single)				
×					
↓					
White	4520+4426 (multiple)				
Guidance: "White" Cannot be found!	StyleCLIP GD				



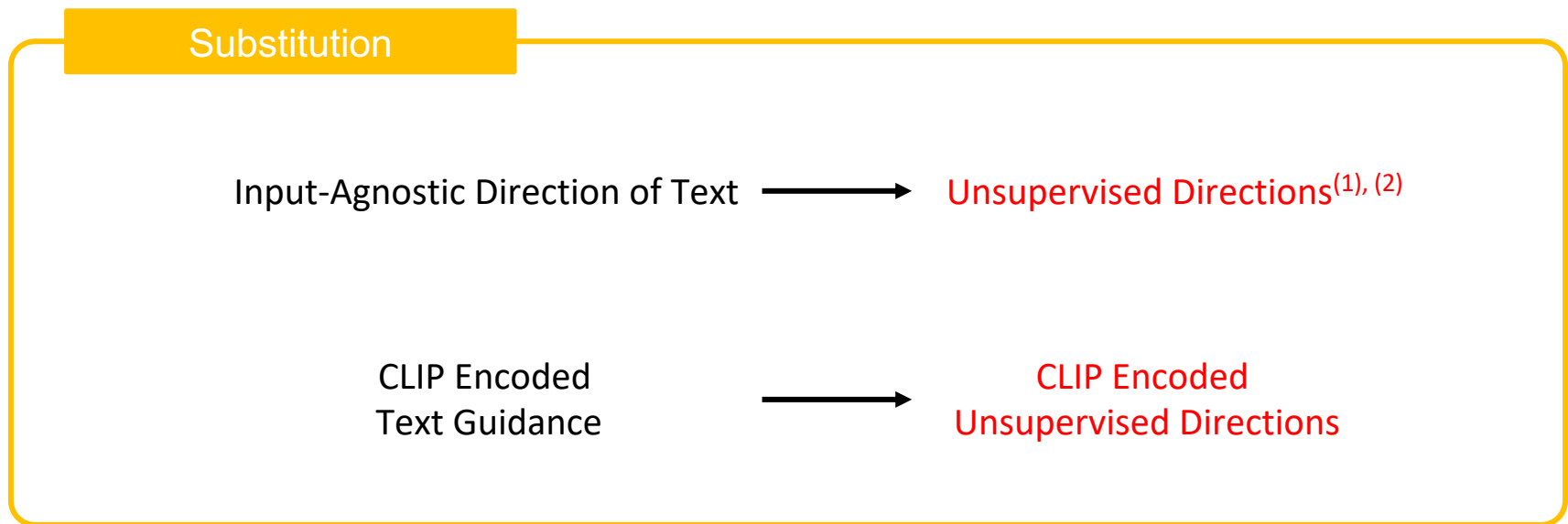
# Learning Multi-Channel Interaction into a Dictionary

- Optimal Case: Text and Direction pair exists



# Learning Multi-Channel Interaction into a Dictionary

- **Substitute** Text – Direction pairs!

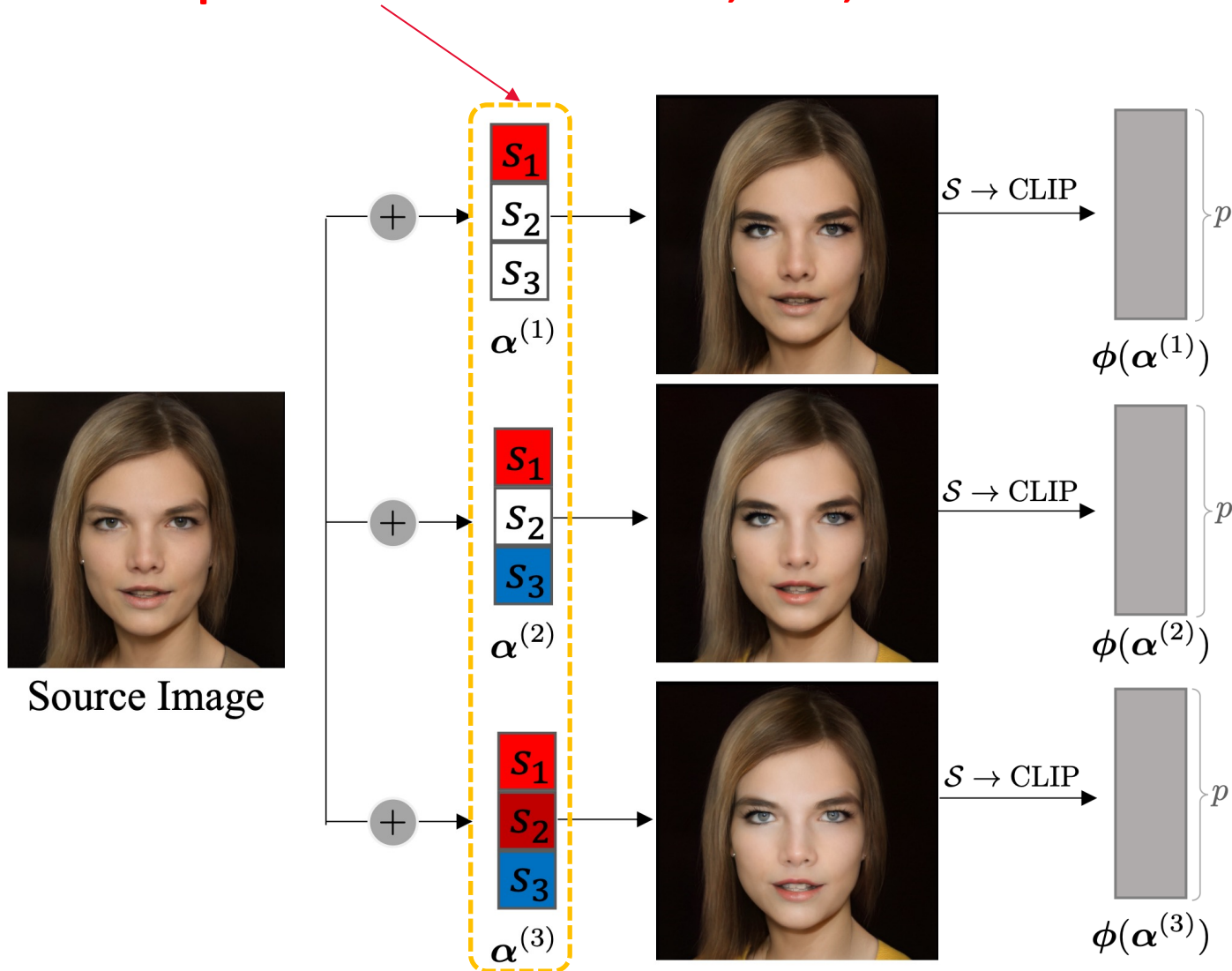


1) Closed-Form Factorization of Latent Semantics in GANs, CVPR 2021

2) GANSpace: Discovering Interpretable GAN Controls, NeurIPS 2020

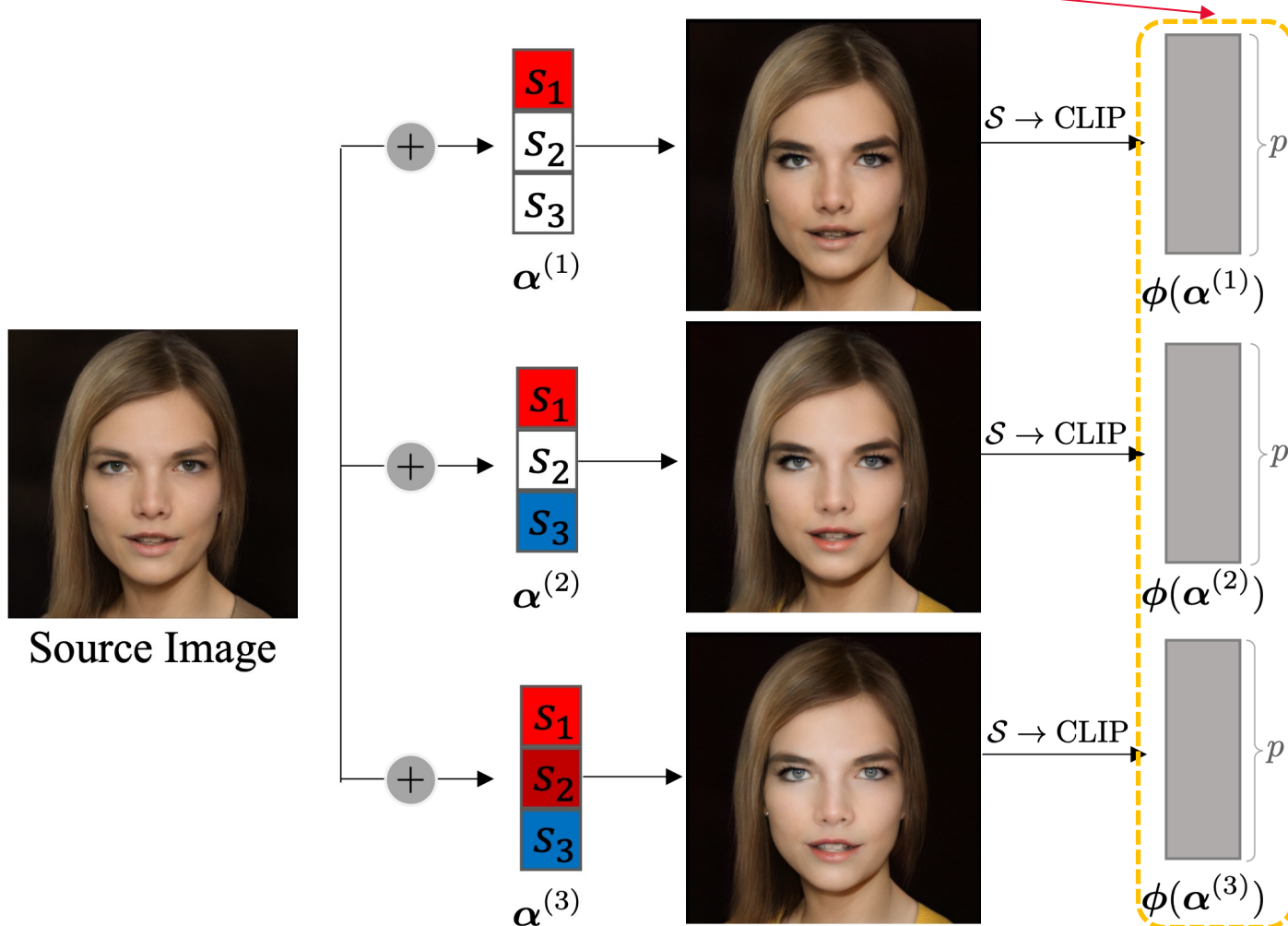
# Method: Unsupervised Directions and CLIP encodings

- Unsupervised Directions:  $\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}$



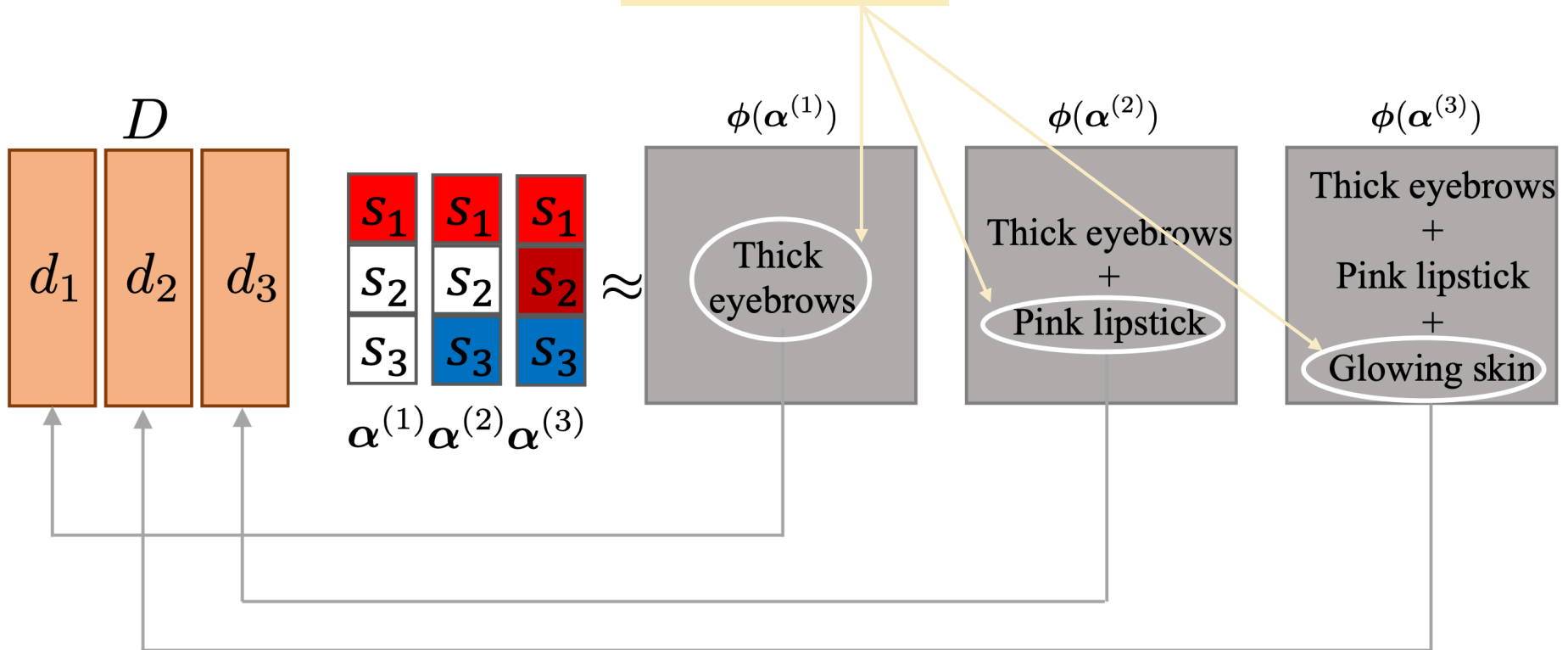
# Method: Unsupervised Directions and CLIP encodings

- CLIP encoded directions:  $\phi(\alpha^{(1)})$ ,  $\phi(\alpha^{(2)})$ ,  $\phi(\alpha^{(3)})$



# Learning a Dictionary with Pairs

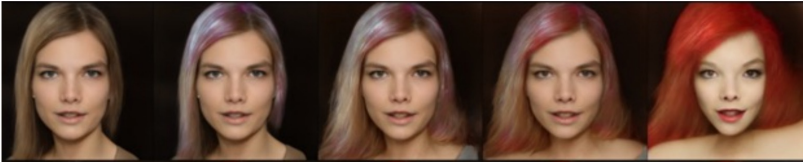
- Manipulation effect of unsupervised directions is decomposed
  - The parameter relevant to decomposed effect is identified



# Qualitative Results

Little Mermaid

StyleCLIP  
(G.D.)

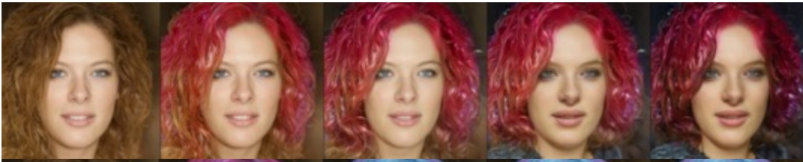


Ours



Blue Hair

StyleCLIP  
(G.D.)



Ours



Joker Smile

StyleCLIP  
(G.D.)

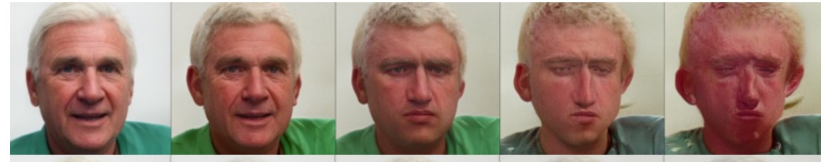


Ours

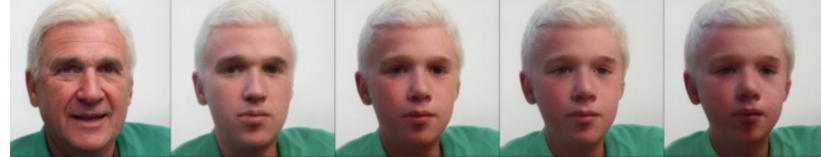


Young

StyleCLIP  
(G.D.)



Ours



# Qualitative Results



# Qualitative Results

Brick Church

StyleCLIP  
(G.D.)



Ours



Golden Church

StyleCLIP  
(G.D.)



Ours





# Qualitative Results

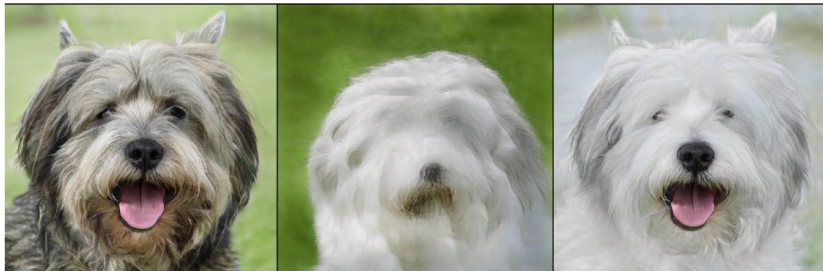
Source

GD

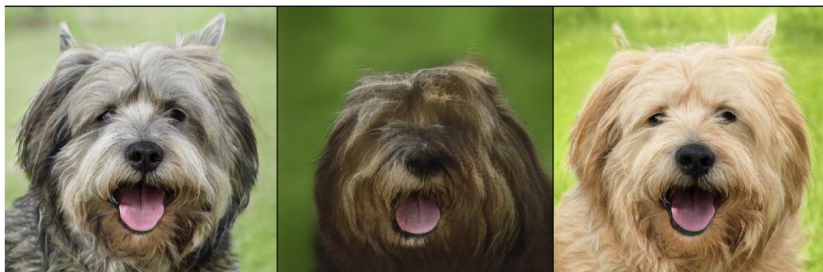
Ours



Black



White

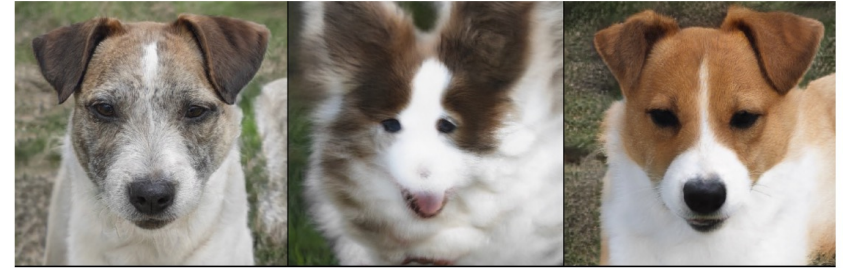


Yellow

Source

GD

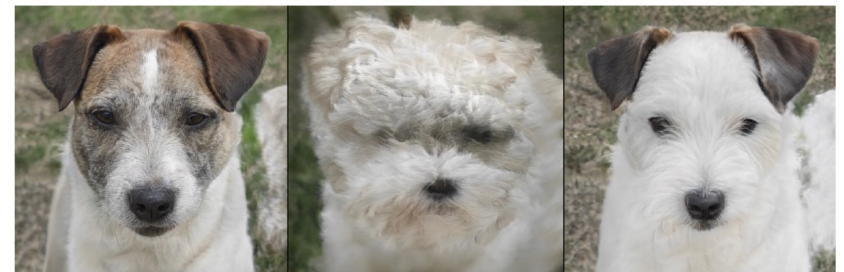
Ours



Corgi



Chihuahua



Maltese

---

**Thank you !**

Any Questions ?