A Style-aware Discriminator for Controllable Image Translation

POSTECH Intelligent Media Lab.

Kunhee Kim Sanghun Park Eunyeong Jeon Taehun Kim Daijin Kim

Pohang University of Science and Technology

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation

Context

- Cannot be trained without domain (or attribute) labels.
- Users can only control the classes used for training.
- Editing the results via latent space exploration is limited [1].

Our Method:

- Does not require labels for training. It can be trained on unlabeled datasets (e.g., FFHQ or LSUN).
- Discovers **style prototypes**, which users can use like domain labels.
- Allows users to **control** the synthesis results. Users can interpolate between two prototypes (or references) or edit specific region only.

Key Observations

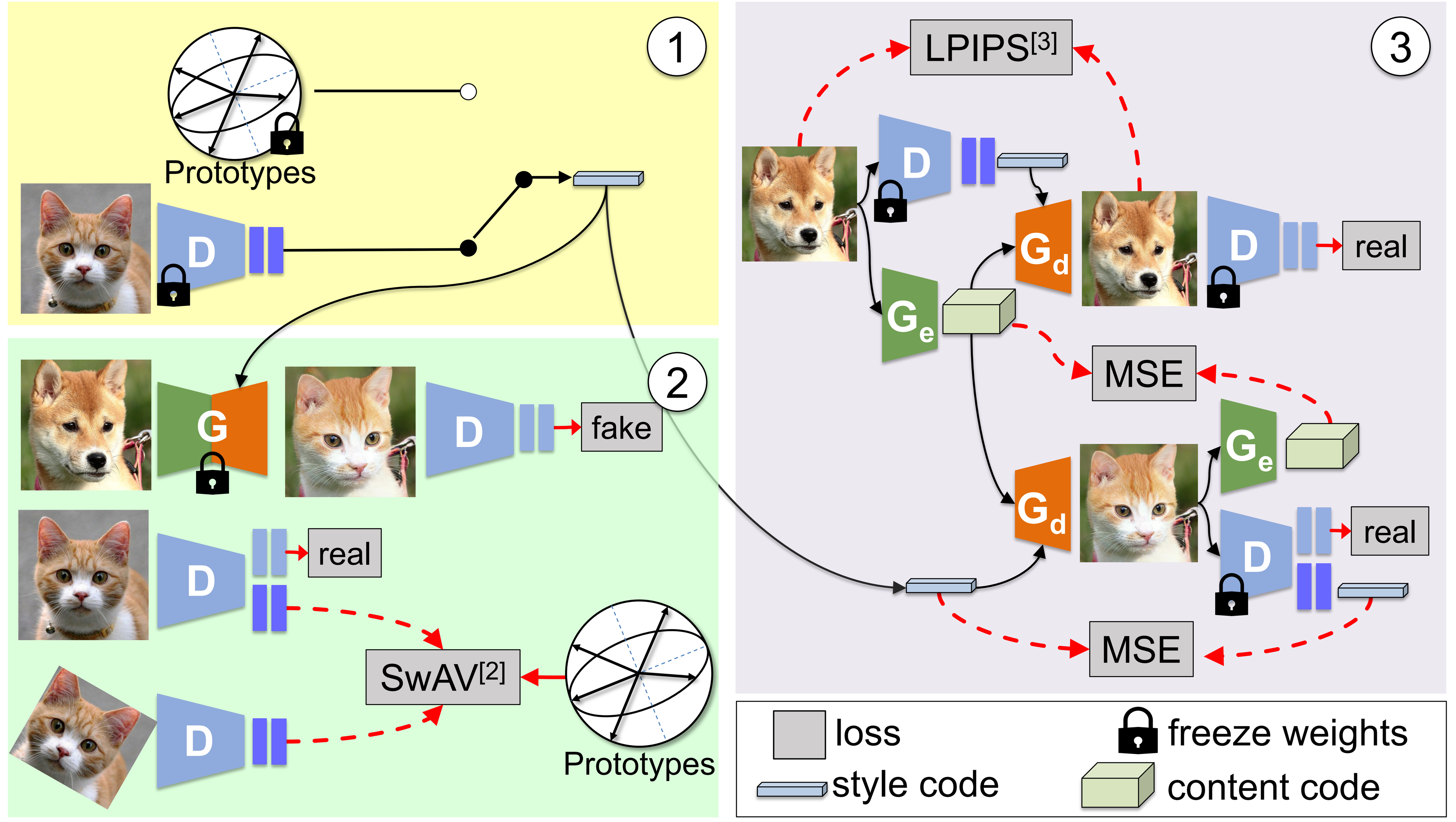
(1) We can control the feature space using augmentation

The 4 samples closest to the reference in the features space
(left) Flip, RandomResizedCrop, ColorDistortion
(right) (left) − ColorDistortion + RandomAffine

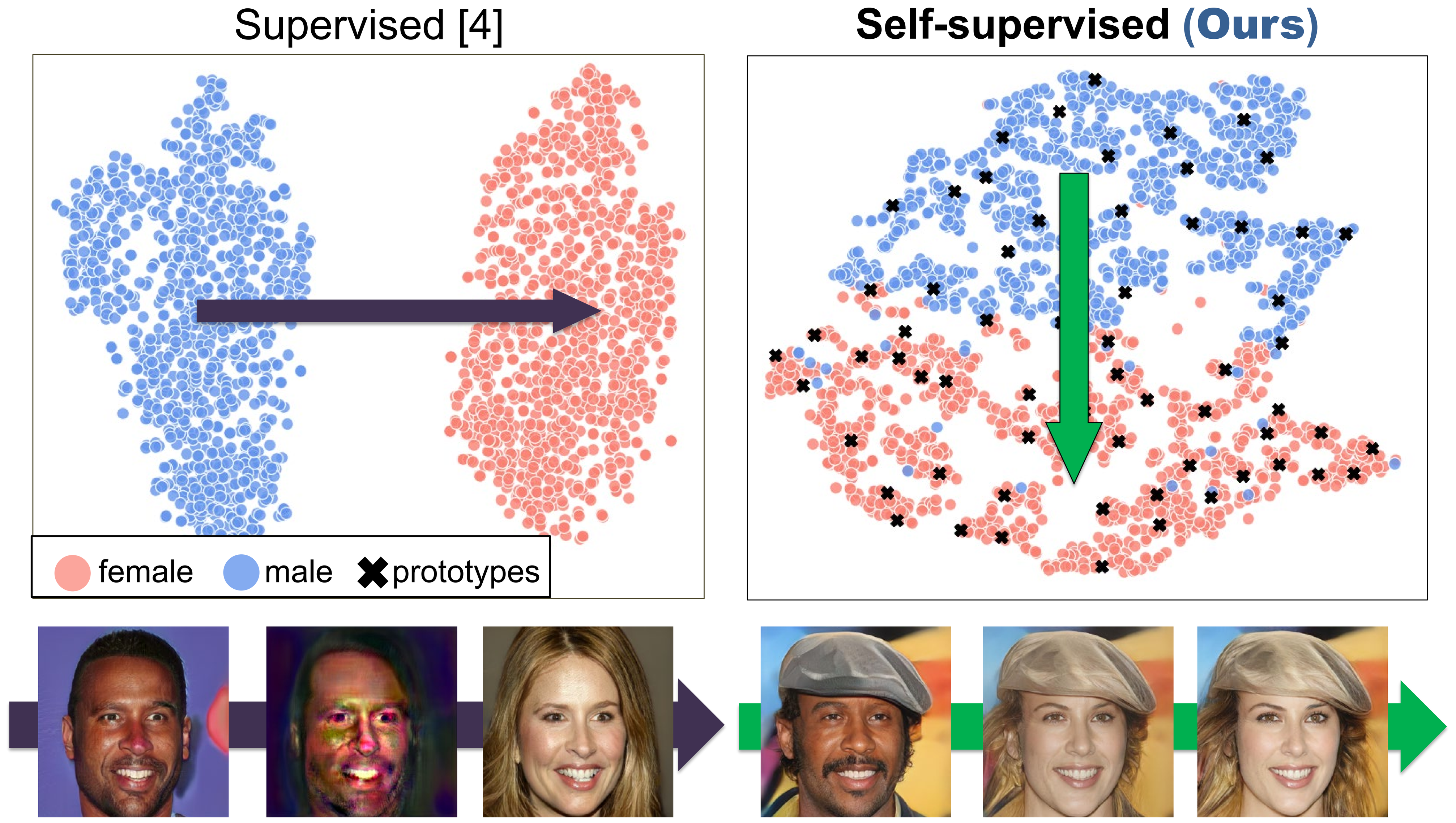
(2) Discriminators can be reused as style encoders

- Significantly reduce the number of parameters (25~35 %▼).
- Can leverage the advantages of the self-supervised discriminators.
- Don't have to throw away the discriminator after training!

Methods Overview

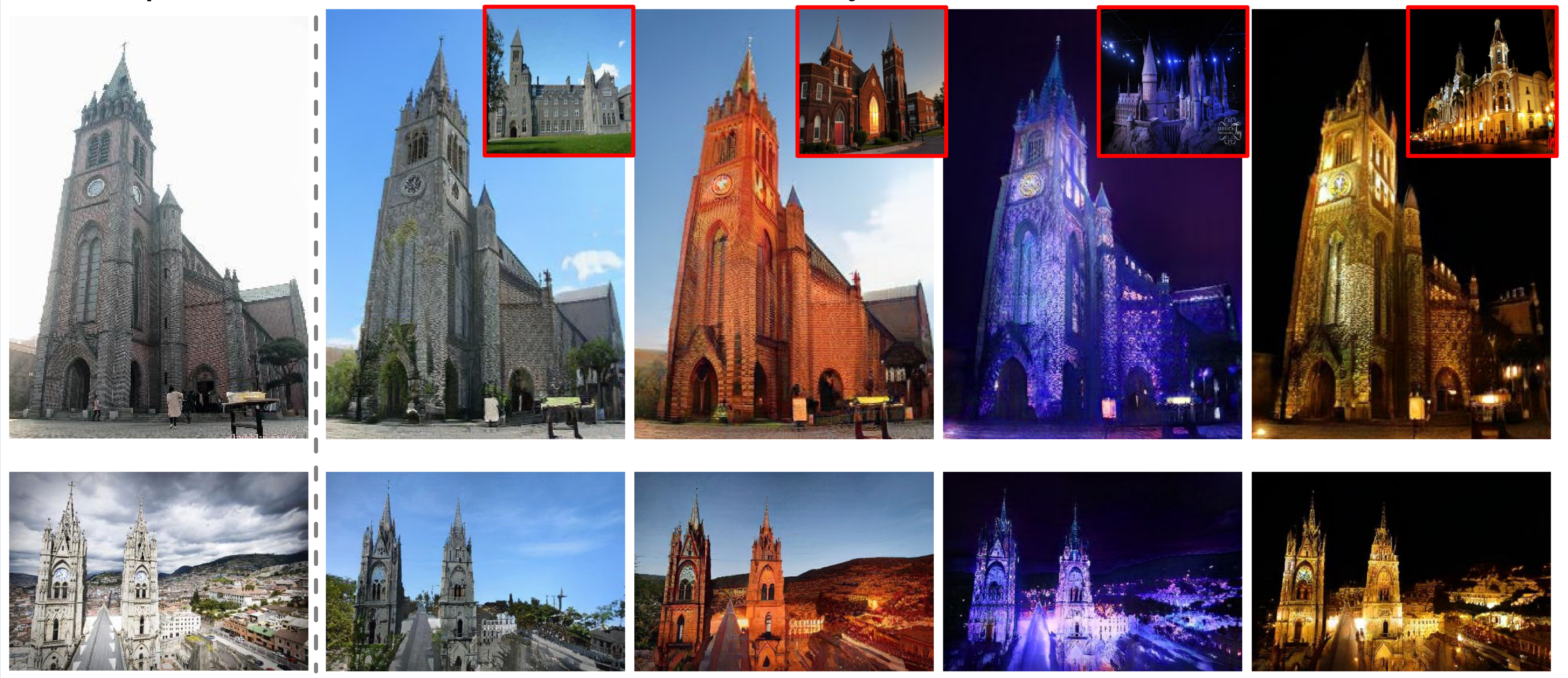


Latent Space Comparison

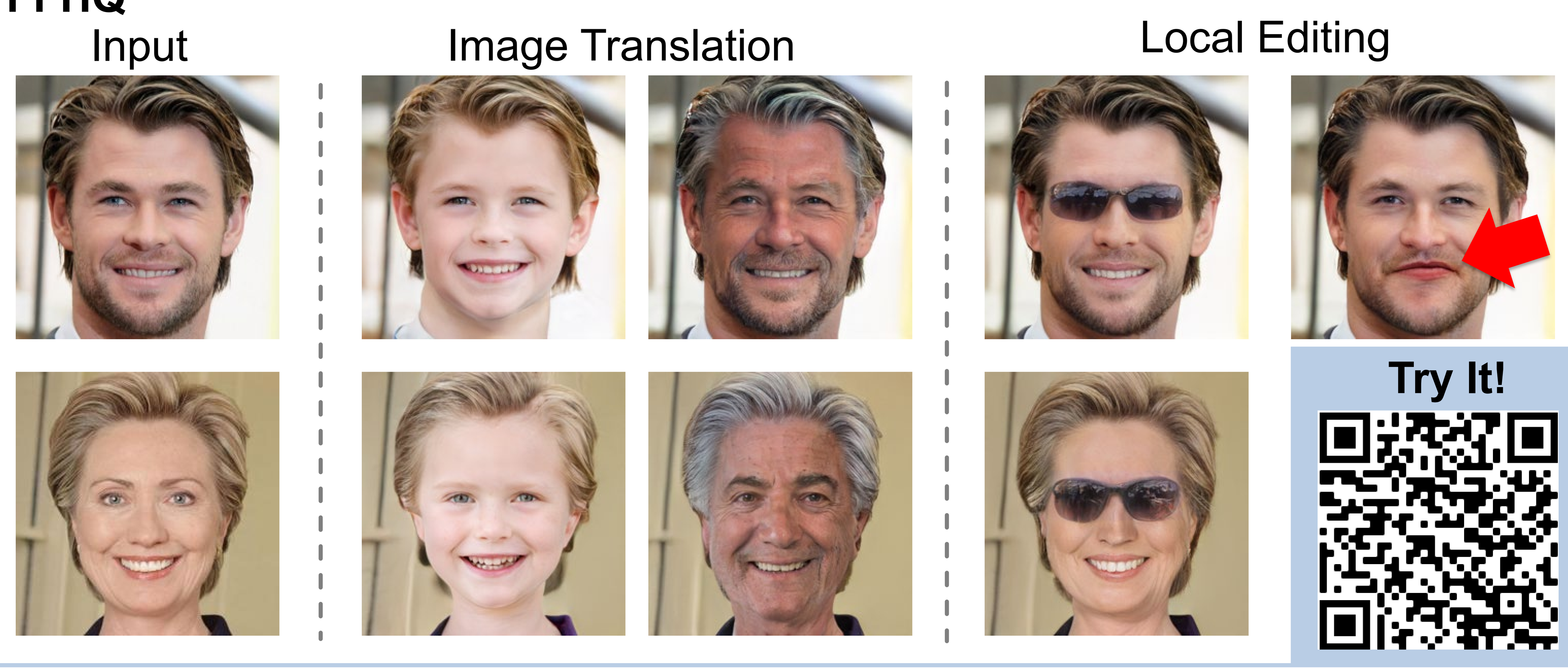


Qualitative Examples

LSUN Churches (□: reference)



FFHQ



Try It!

References

[1] Liu et al. "Smoothing the Disentangled Latent Style Space for Unsupervised Image-to-Image Translation." In CVPR, 2021.

[2] Caron et al. "Unsupervised learning of visual features by contrasting cluster assignments." In NIPS, 2020.

[3] Zhang et al. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." In CVPR, 2018.

[4] Choi et al. "StarGAN v2: Diverse Image Synthesis for Multiple Domains." In CVPR, 2020.