STARGAN V2: DIVER IMAGE SYNTHESIS FOR MULTIPLE DOMIANS

Kangyeol Kim

Vision Study 2020204

DAVIAN

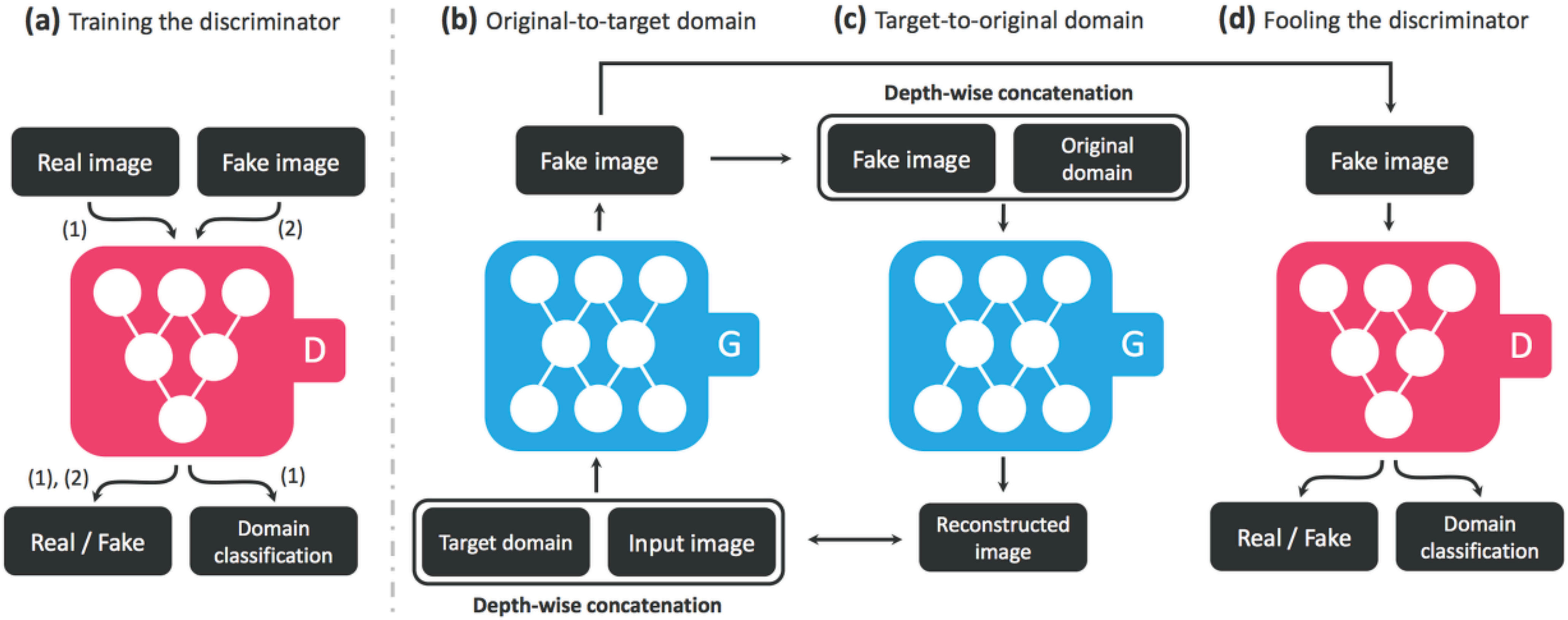
Data and Visual Analytics Lab

Contents Overview

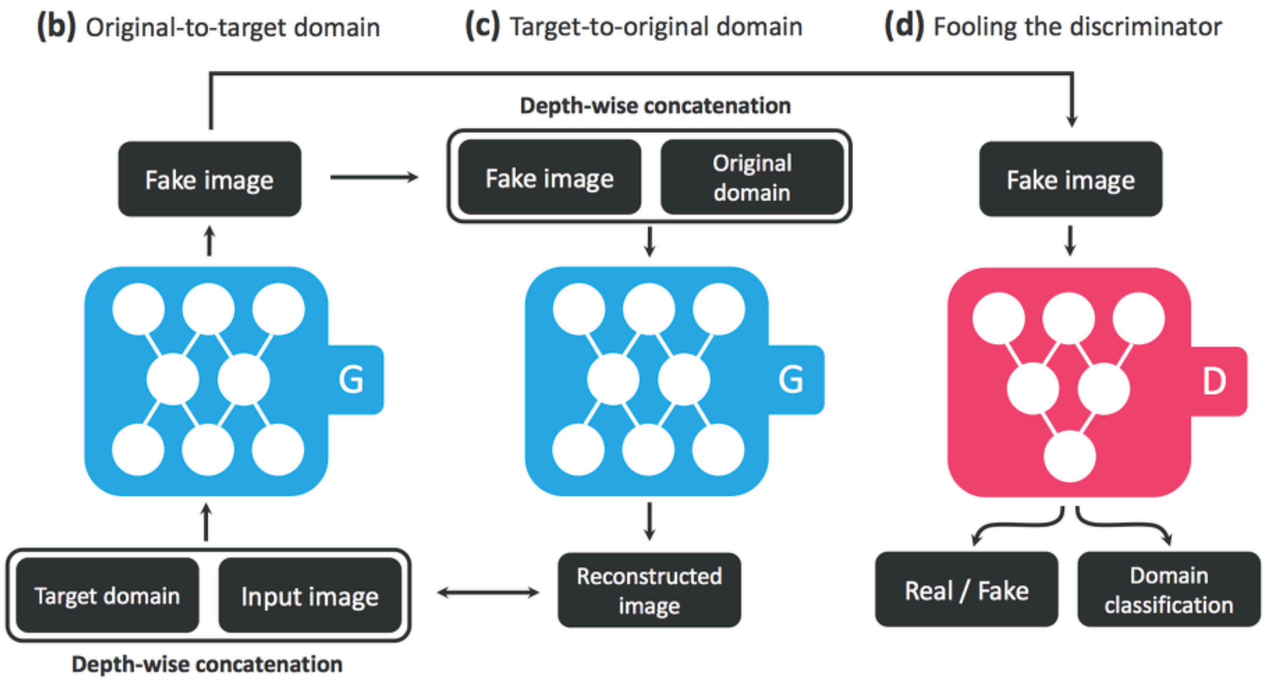
- StarGAN Review and Problems
- Motivations
- Method
- Experiments

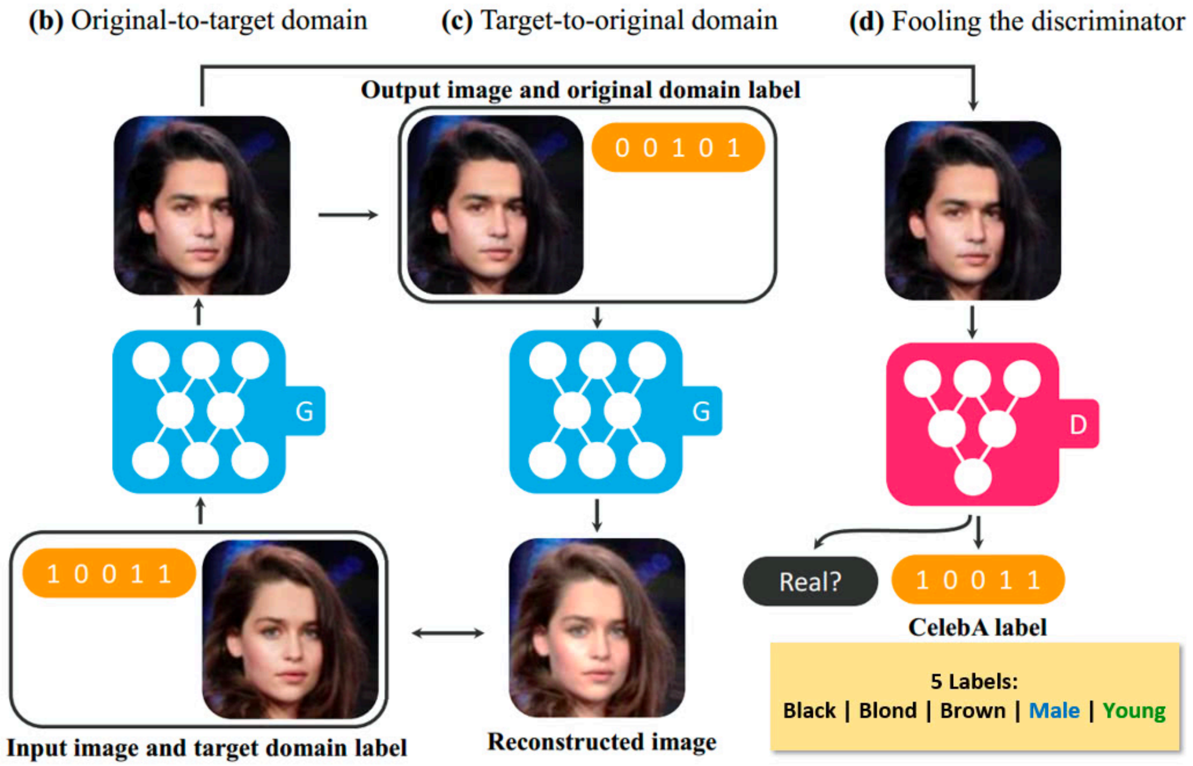
StarGAN Review

- StarGAN Motivations
  - 단일 모델을 가지고 여러 가지 domain에 대해 image-to-image translation
  - 동시에 다른 domain을 가진 데이터셋들을 동시에 학습 (one-hot vector)

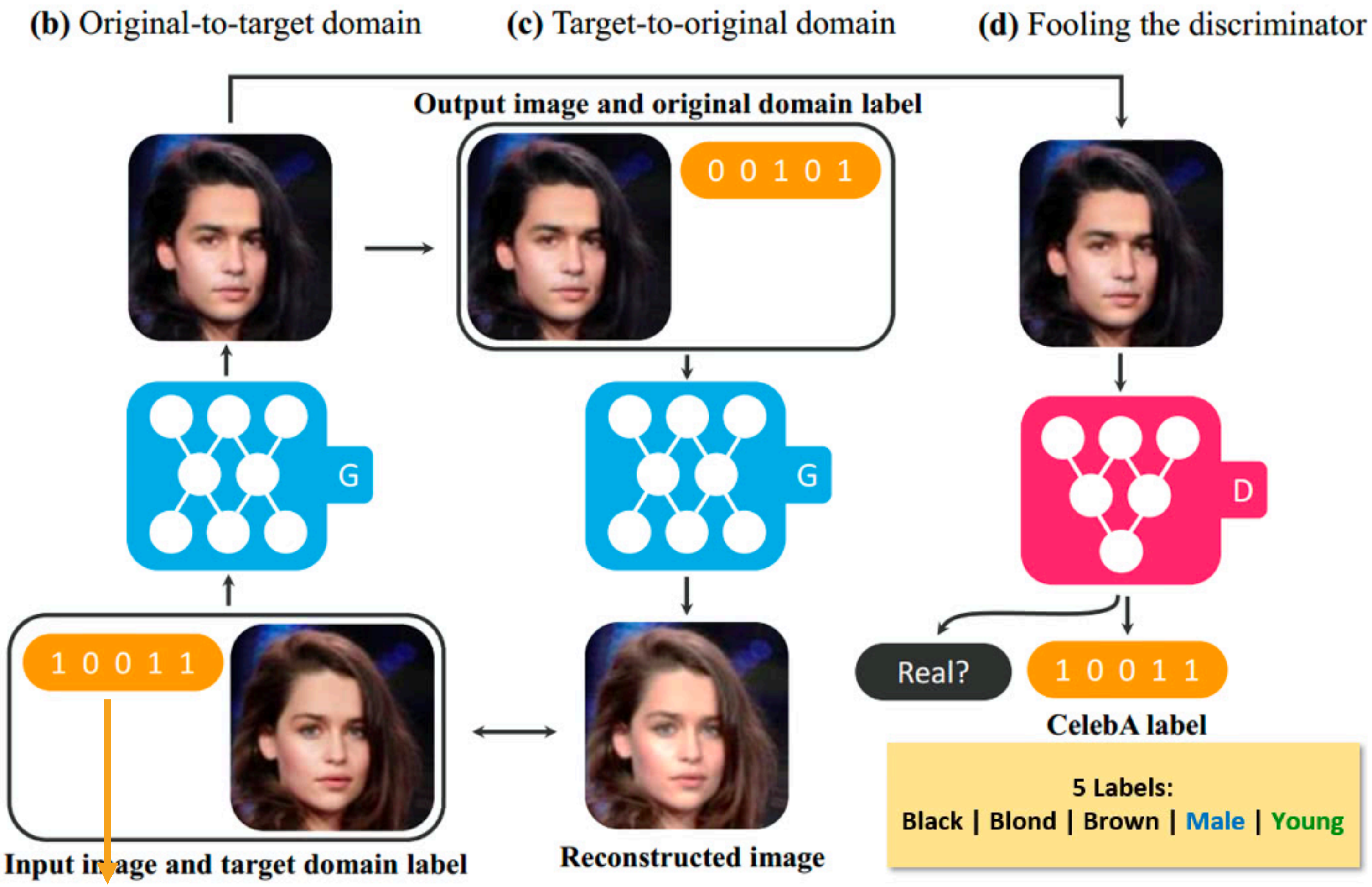


StarGAN Review





StarGAN Problems

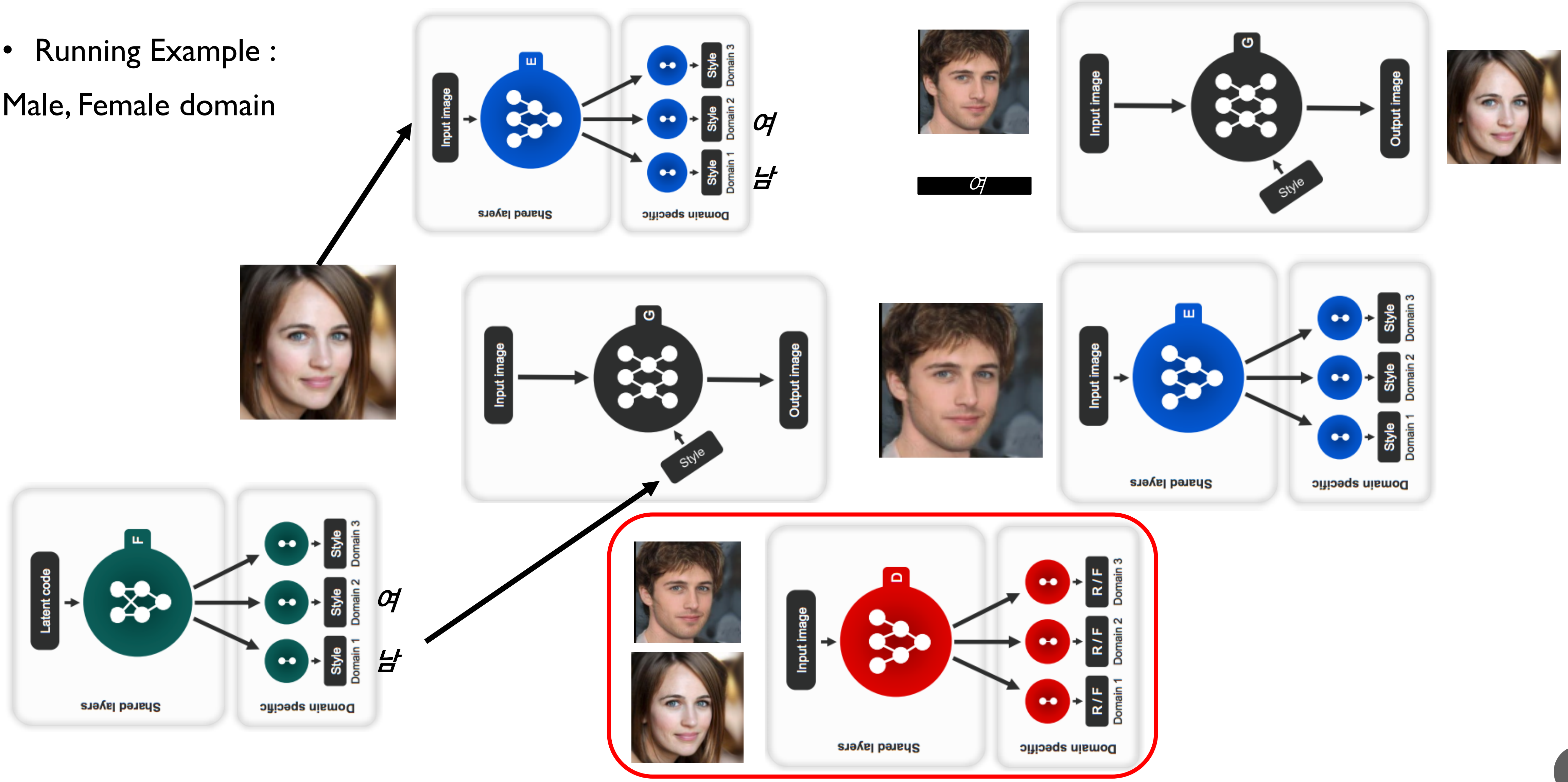


deterministic mapping per each domain,
cannot represent diverse *styles* of a specific *domain*

Method (1/3) : Training Forwarding

- Running Example :

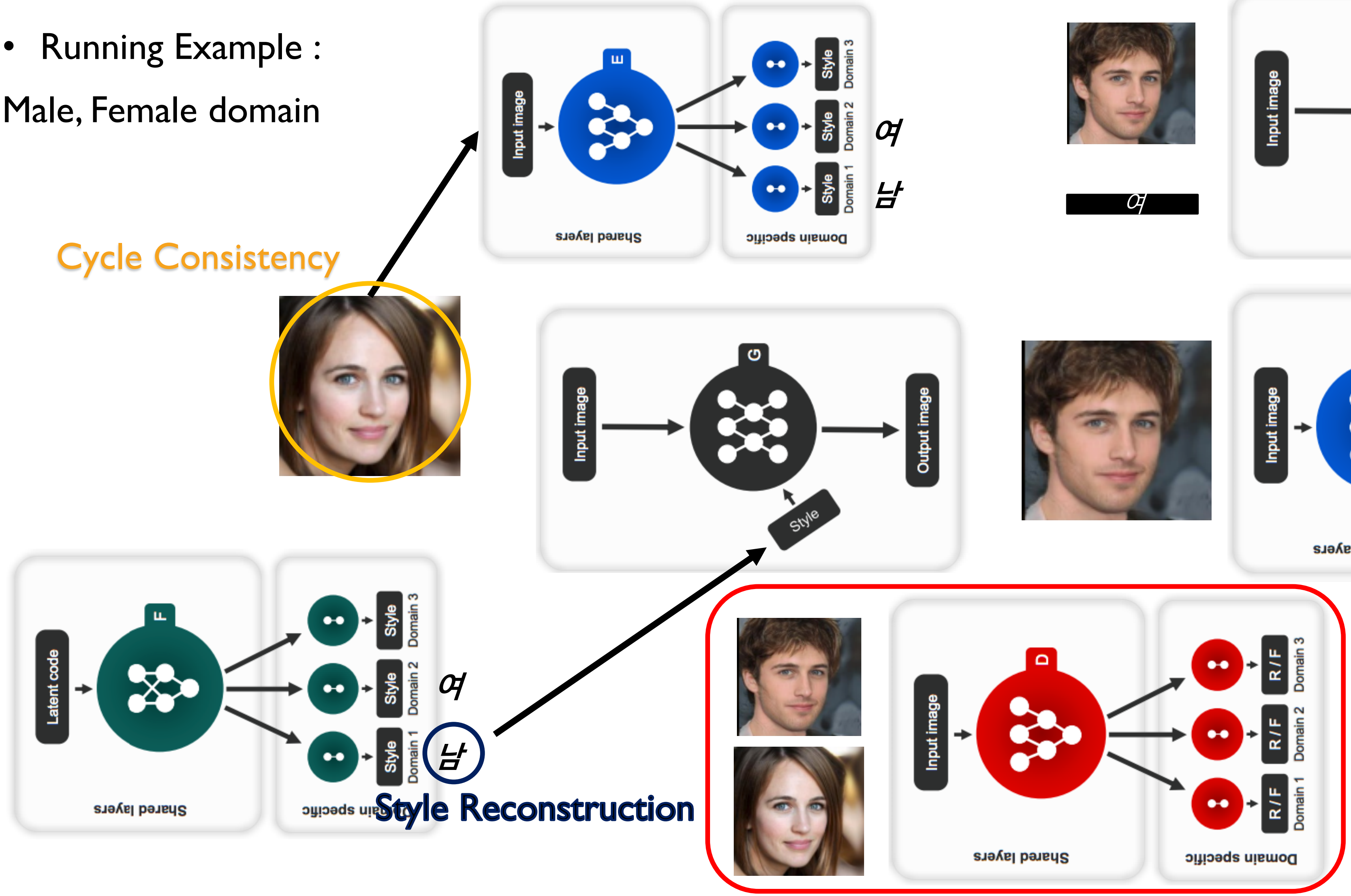
Male, Female domain

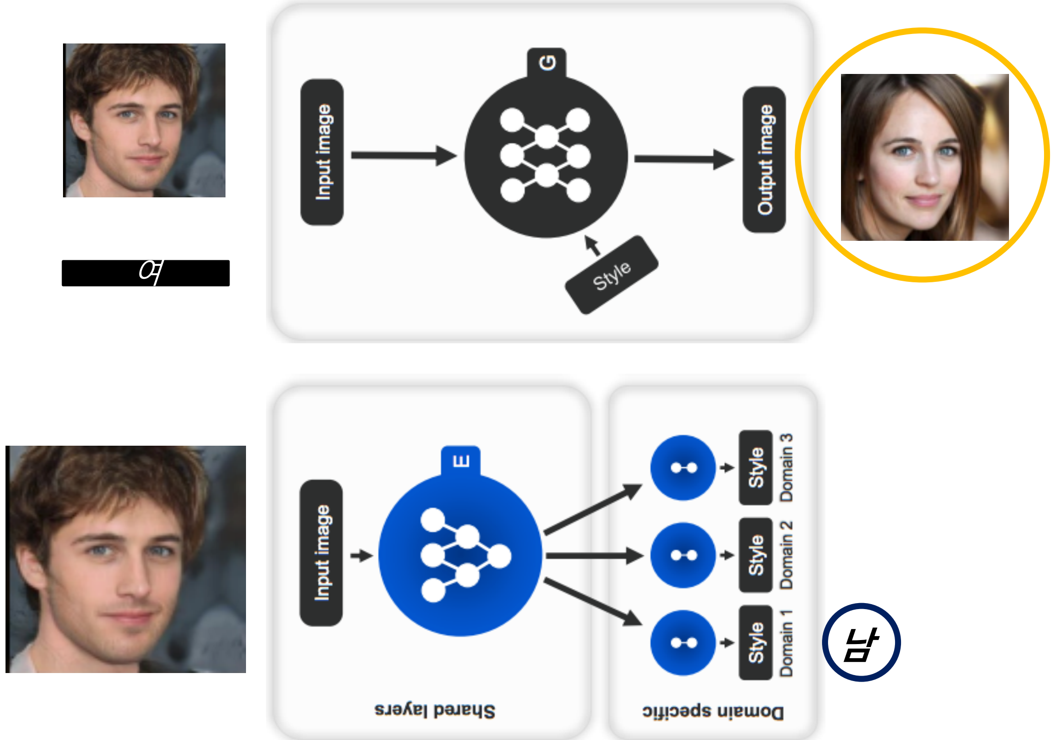


Method (1/3) : Losses

- Running Example :

Male, Female domain





\+ GAN Loss

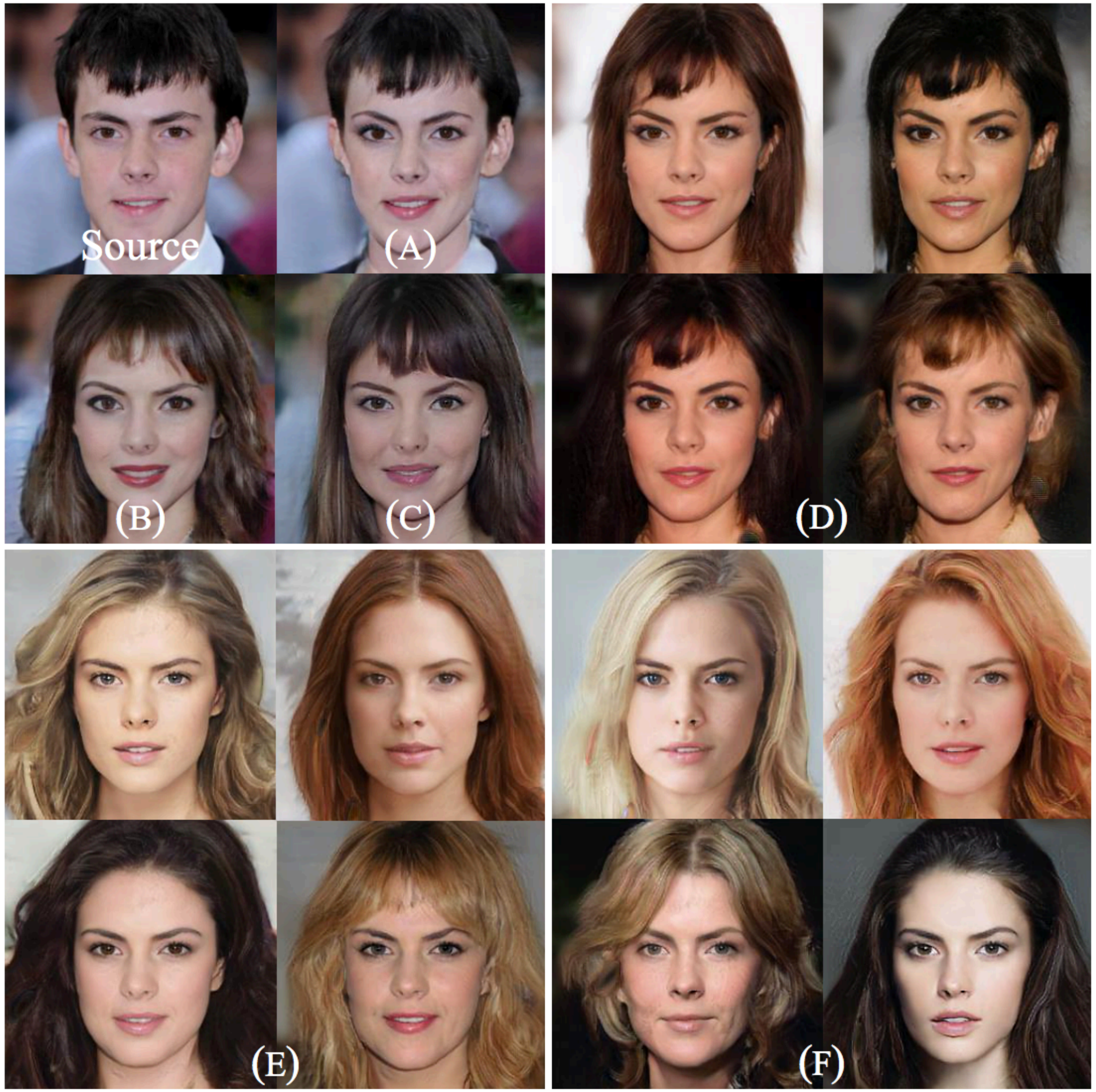
\+ Style diversification

$$\mathcal{L}_{ds} = \mathbb{E}_{\mathbf{x},\widetilde{y},\mathbf{z}_1,\mathbf{z}_2}\left[\|G(\mathbf{x},\widetilde{\mathbf{s}}_1) - G(\mathbf{x},\widetilde{\mathbf{s}}_2)\|_1\right],$$

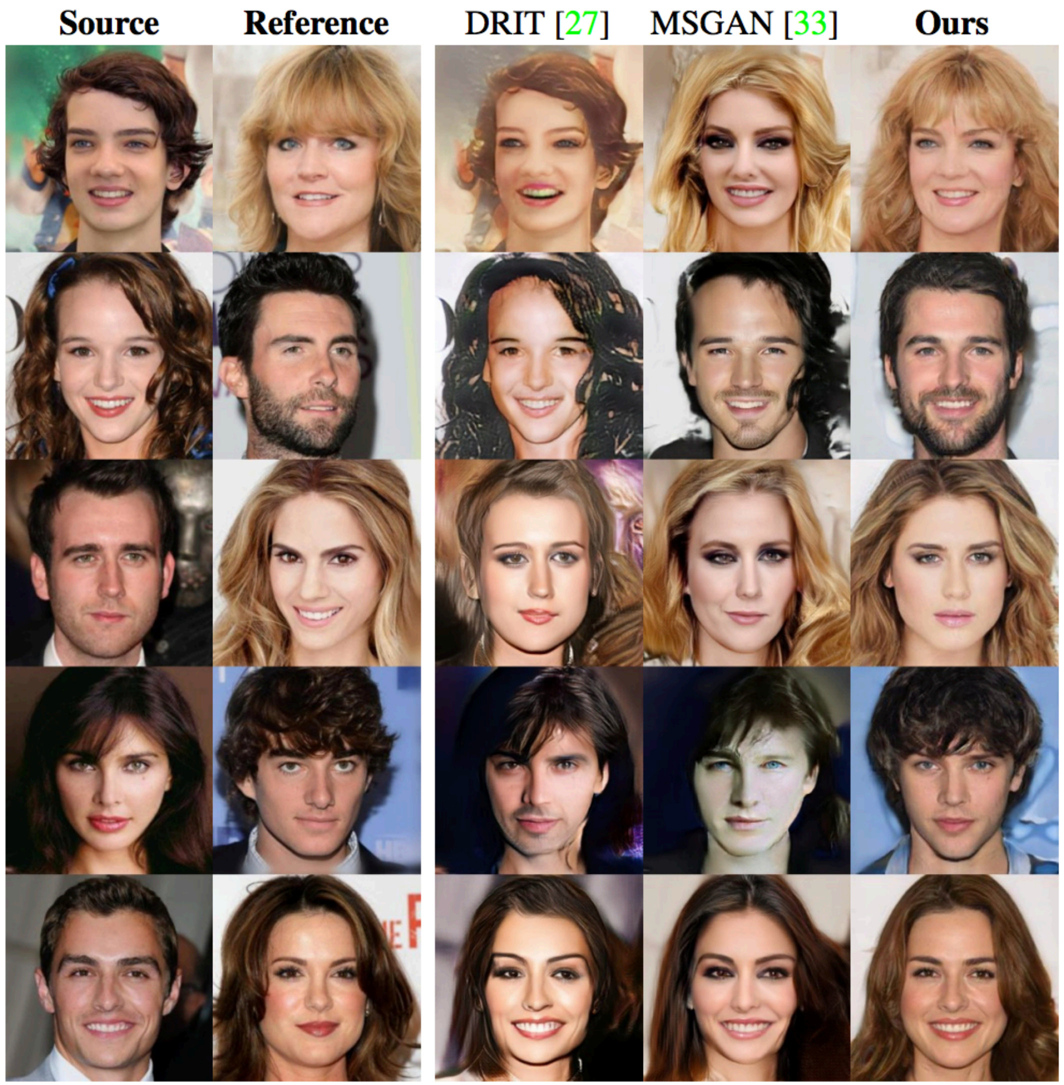
Experiment, Ablation Study

Concat=>AdaIN

	Method	FID	LPIPS
A	Baseline StarGAN [7]	98.4	-
B	+ Multi-task discriminator	91.4	-
C	+ Tuning (*e.g.*, R_1 regularization)	80.5	-
D	+ Latent code injection	32.3	0.312
E	+ Replace (D) with style code	21.2	0.406
F	+ Diversity regularization	**18.0**	**0.428**



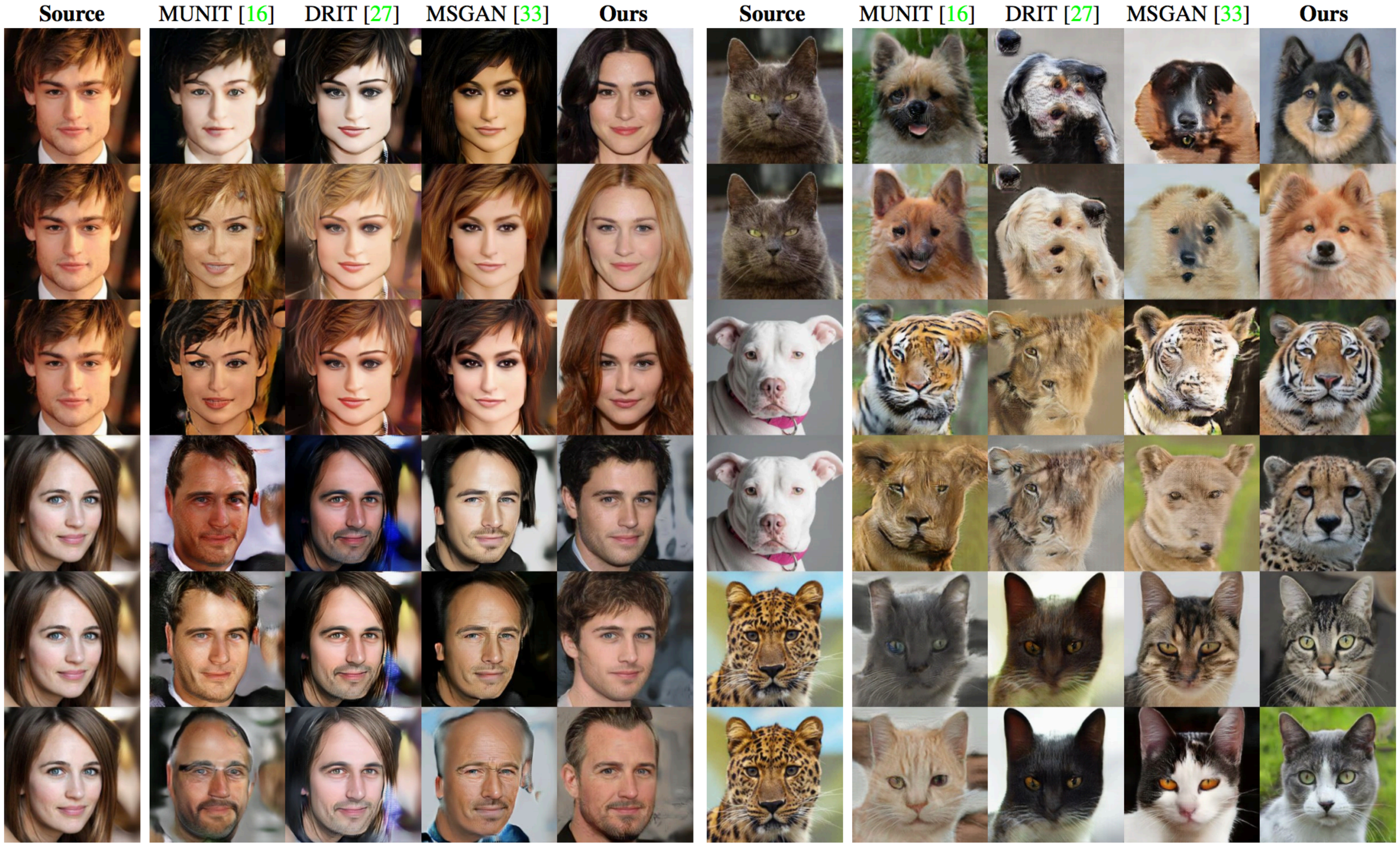
Experiment, Exemplar guided translation



(a) Reference-guided synthesis on CelebA-HQ

(b) Reference-guided synthesis on AFHQ

Experiment, Latent guided translation



(a) Latent-guided synthesis on CelebA-HQ

(b) Latent-guided synthesis on AFHQ

Experiment, Latent guided translation

First, following the insight of StyleGAN, our style space is produced by non-linear transformations from a Gaussian distribution. **This provides more flexibility to our model compared to assuming a fixed prior distribution.**

Second, our **style code is separately generated per each domain by the multi-branch encoder and mapping networks.** By doing so, **our generator can only focus on using the style code**, whose domain-specific information is already taken care of by the encoder or the mapping network.

Third, our modules benefit from fully exploiting training data from multiple domains. By design, the shared part of each module should learn domain-invariant features which induces the regularization effect, encouraging better generalization to unseen samples.

Qualitative results of third point.

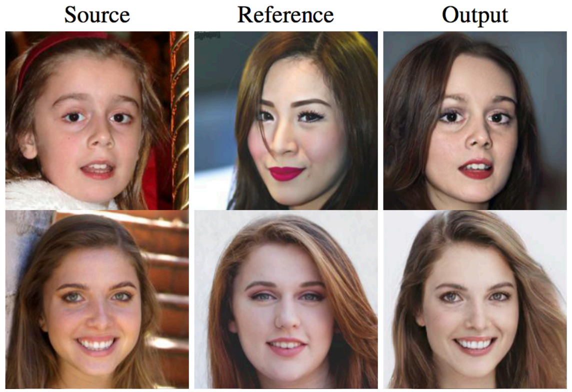


Figure 7. Reference-guided synthesis results on FFHQ with the model trained on CelebA-HQ. Despite the distribution gap between the two datasets, StarGAN v2 successfully extracts the style codes of the references and synthesizes faithful images.