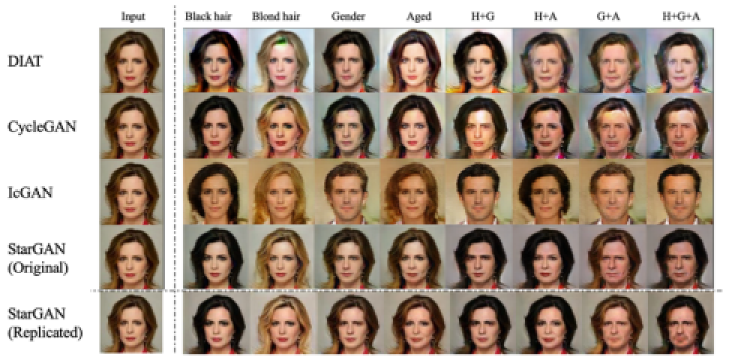
StarGAN

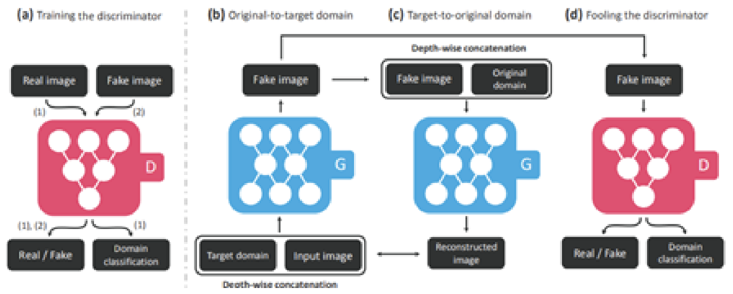
Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

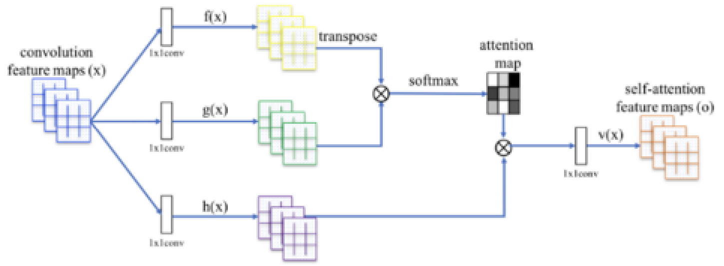
TEAM 41

20170264 박승민
20190006 강산아
20190545 임승재
20200673 최현민

Abstract







01

Multi-Domain Translation

- cGAN
- Limitation

02

StarGAN

- Loss function
- Multiple Dataset

03

Analysis

- CelebA
- AffectNet Dataset

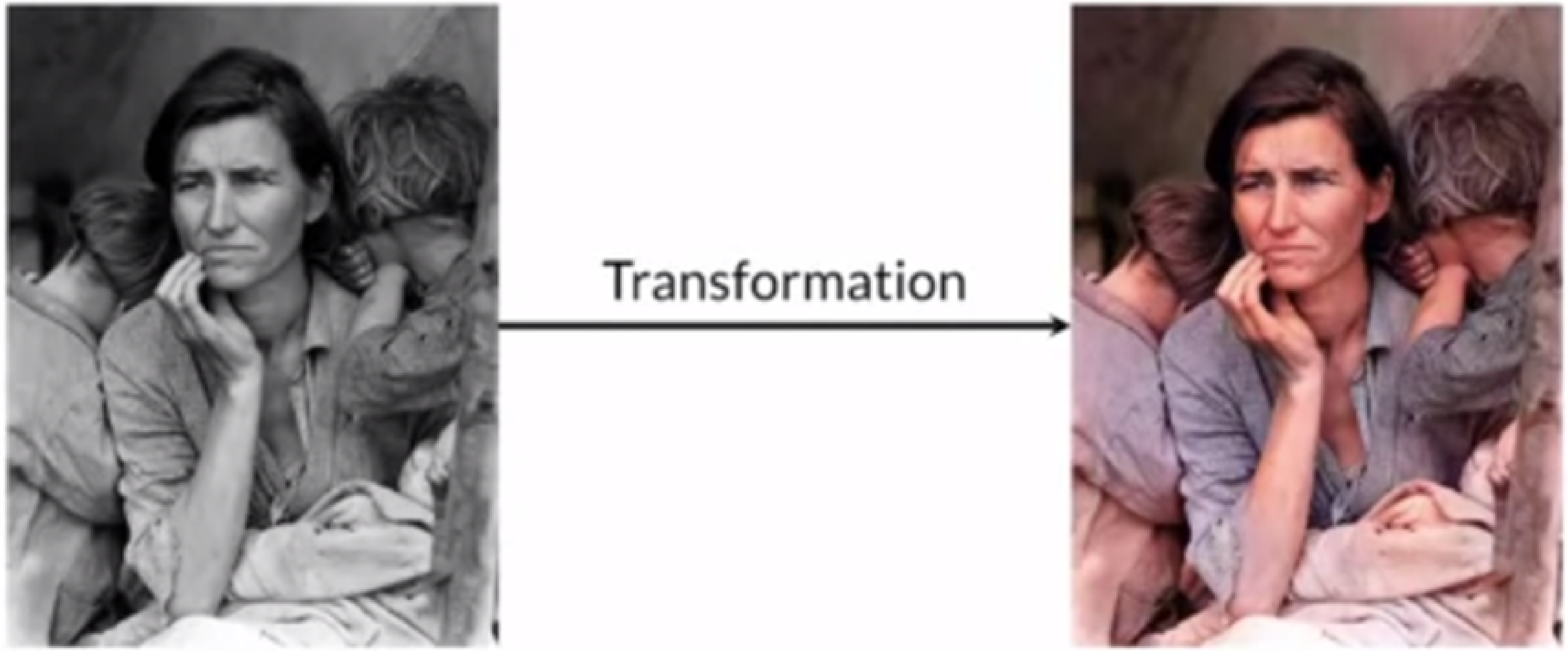
04

Self Attention

- Improvement
- Geometric, Structal pattern problem

Introduce

Image-to-Image Translation



Given training data from **two different domains,**
train the model to **translate images** from one domain to the other

Introduce

StarGAN

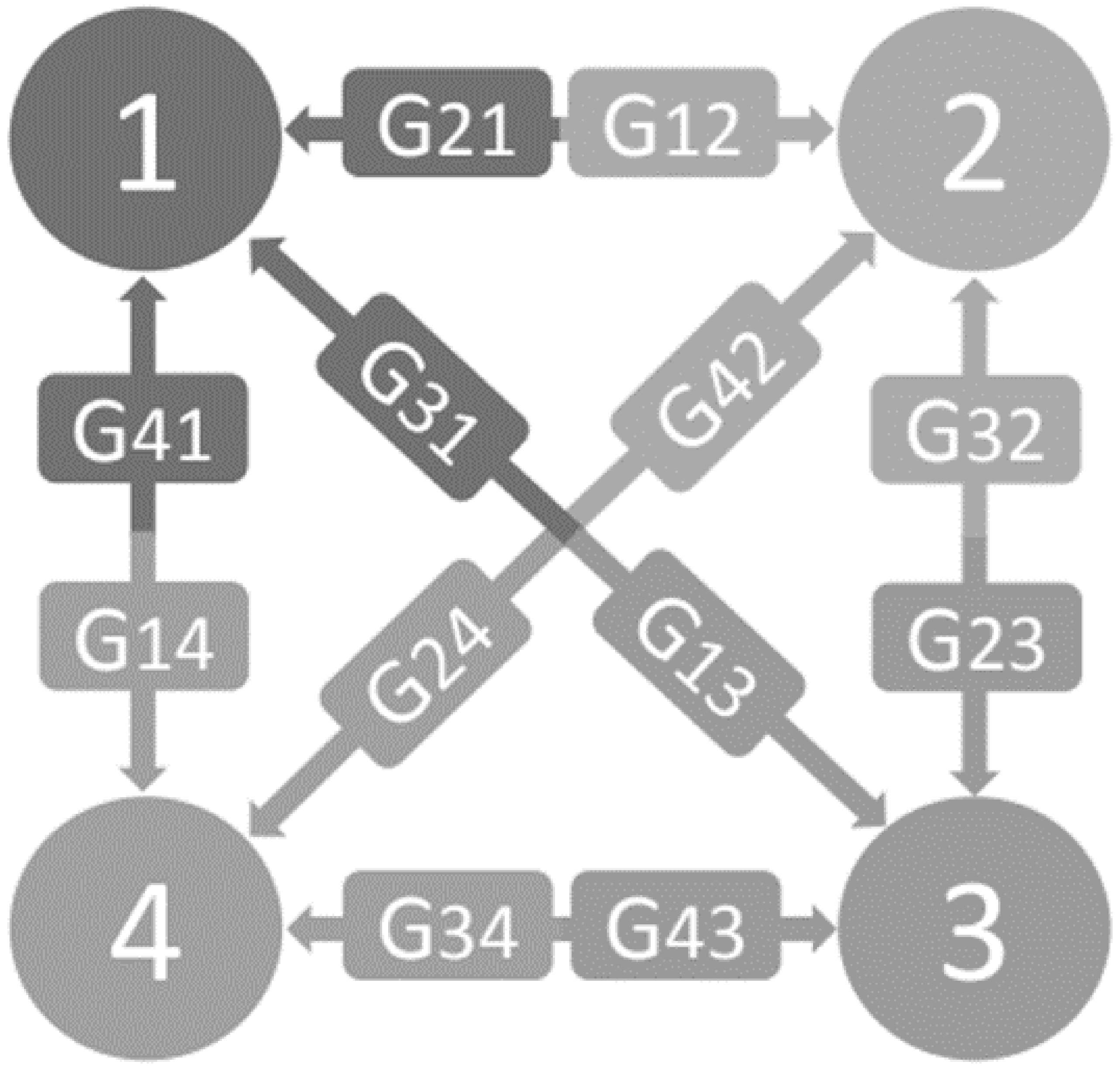
k(k-1) generators need for k domains, **expensive**!

Problems

Each generator **doesn’t share global features**
Can’t fully utilize the entire training data

Solution

Incapable of jointly training domains from different datasets



Introduce

Objective

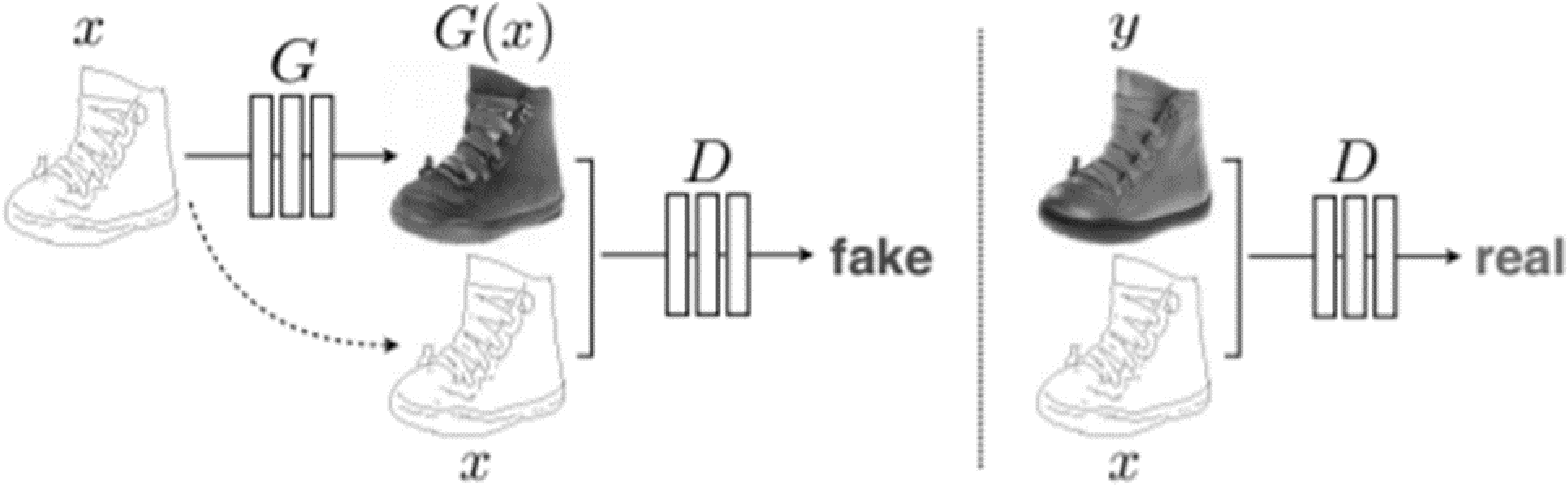
Using **single generator and discriminator**,
build more elaborate multi-domain translation model

Build a novel generative adversarial network that learns the mappings
among **multiple domains** using only a **single generator and a discriminator**

Provide **qualitative** and **quantitative** result on I2I translation

Theoretical Background

Pix2pix

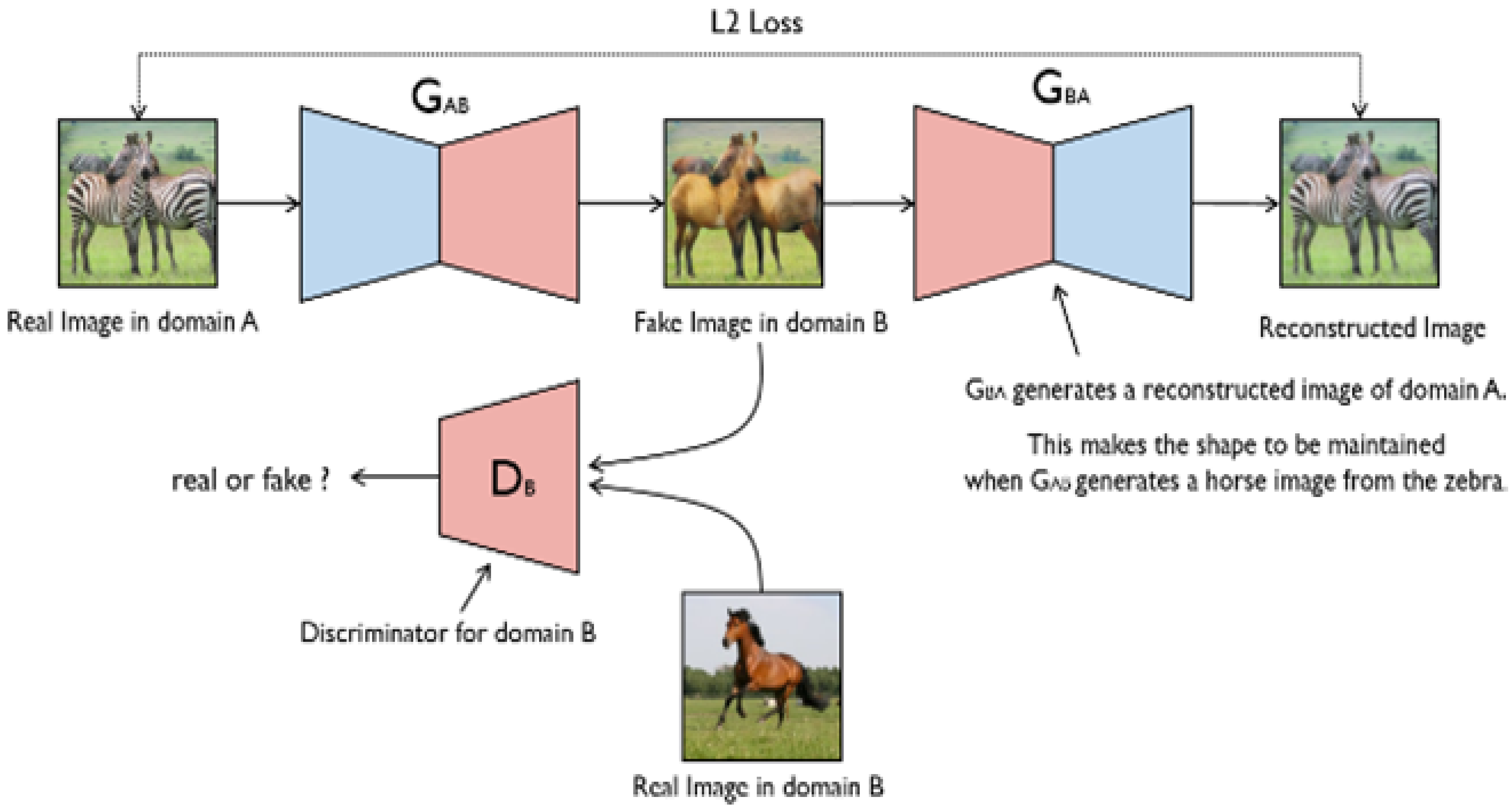


Proceeds the learning with **sketch data**

Using the cGAN, learns tasks in the supervised manner

Theoretical Background

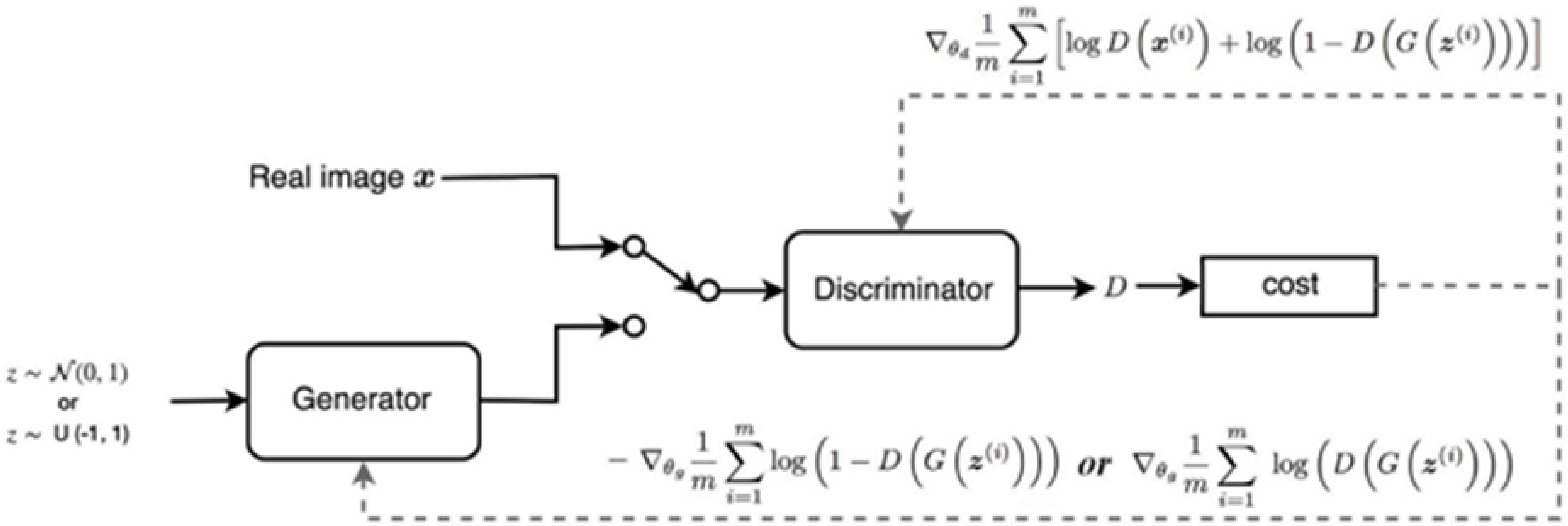
CycleGAN



Utilize the **cycle consistency loss**

Theoretical Background

WGAN



Uses the newly defined **critic** instead of the discriminator

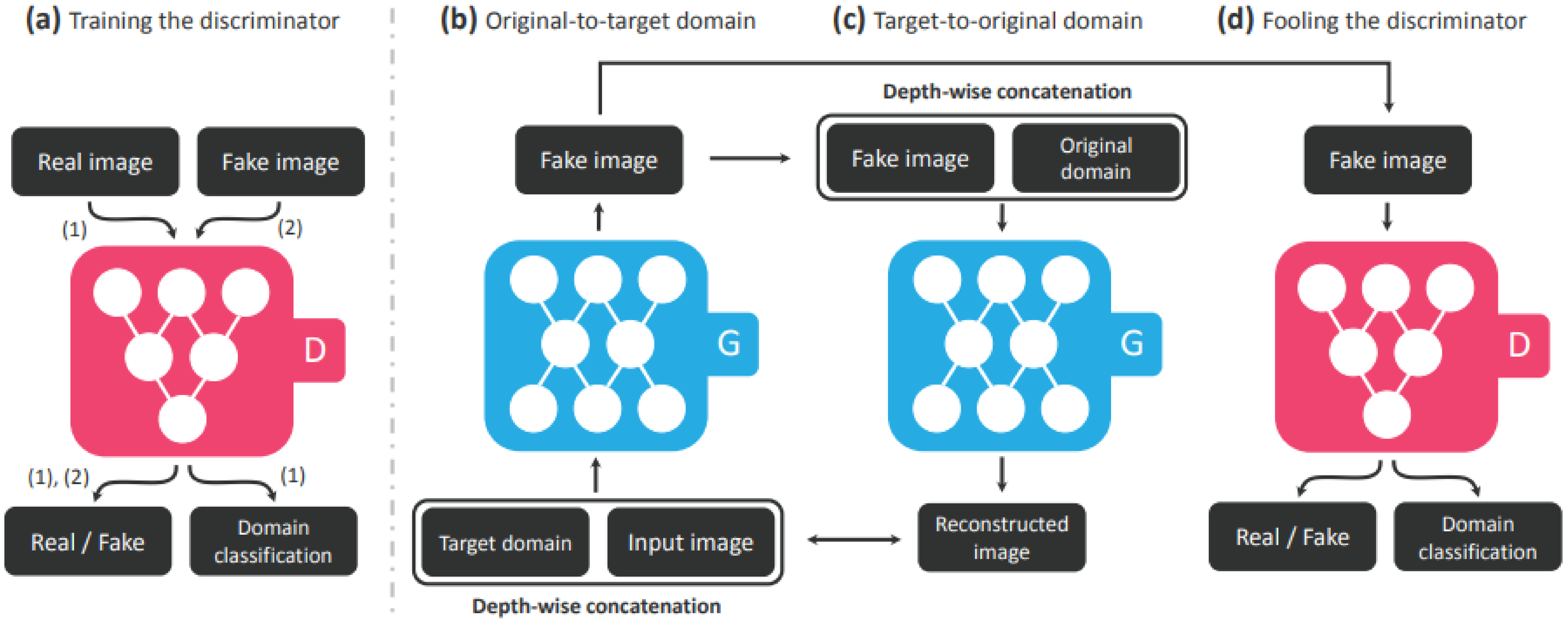
Challenges

Limitation

- difficult to **maintain the balance**
  between the discriminator and generator when training a model
- **no model** has proposed an efficient method
  for efficient image-to-image translation **in multiple domains**

StarGAN Design

Overview



StarGAN Design

Loss functions and Implementation

Adversarial Loss

$$L_{adv} = E_x[\log D_{src}(x)] + E_{x,c}[\log(1 - D_{src}(G(x,c)))]$$

Domain Classification Loss

$$L^r_{cls} = E_{x,c'}[-\log D_{cls}(c'|x)]$$
$$L^f_{cls} = E_{x,c}[-\log D_{cls}(c|G(x,c))]$$

Reconstruction Loss

$$L_{rec} = E_{x,c,c'}[\|x - G(G(x,c),c')\|_1]$$

Full Objective

$$L_D = -L_{adv} + \lambda_{cls} L^r_{cls}$$
$$L_G = L_{adv} + \lambda_{cls} L^f_{cls} + \lambda_{rec} L_{rec}$$

StarGAN Design

Loss functions and Implementation

Improved GAN Training (WGAN)

$$L_{adv} = E_x[D_{src}(x)] - E_{x,c}\left[D_{src}\left(G(x,c)\right)\right] - \lambda_{gp} E_{x'}[(|\nabla_{x'} D_{src}(x')|_2 - 1)^2]$$

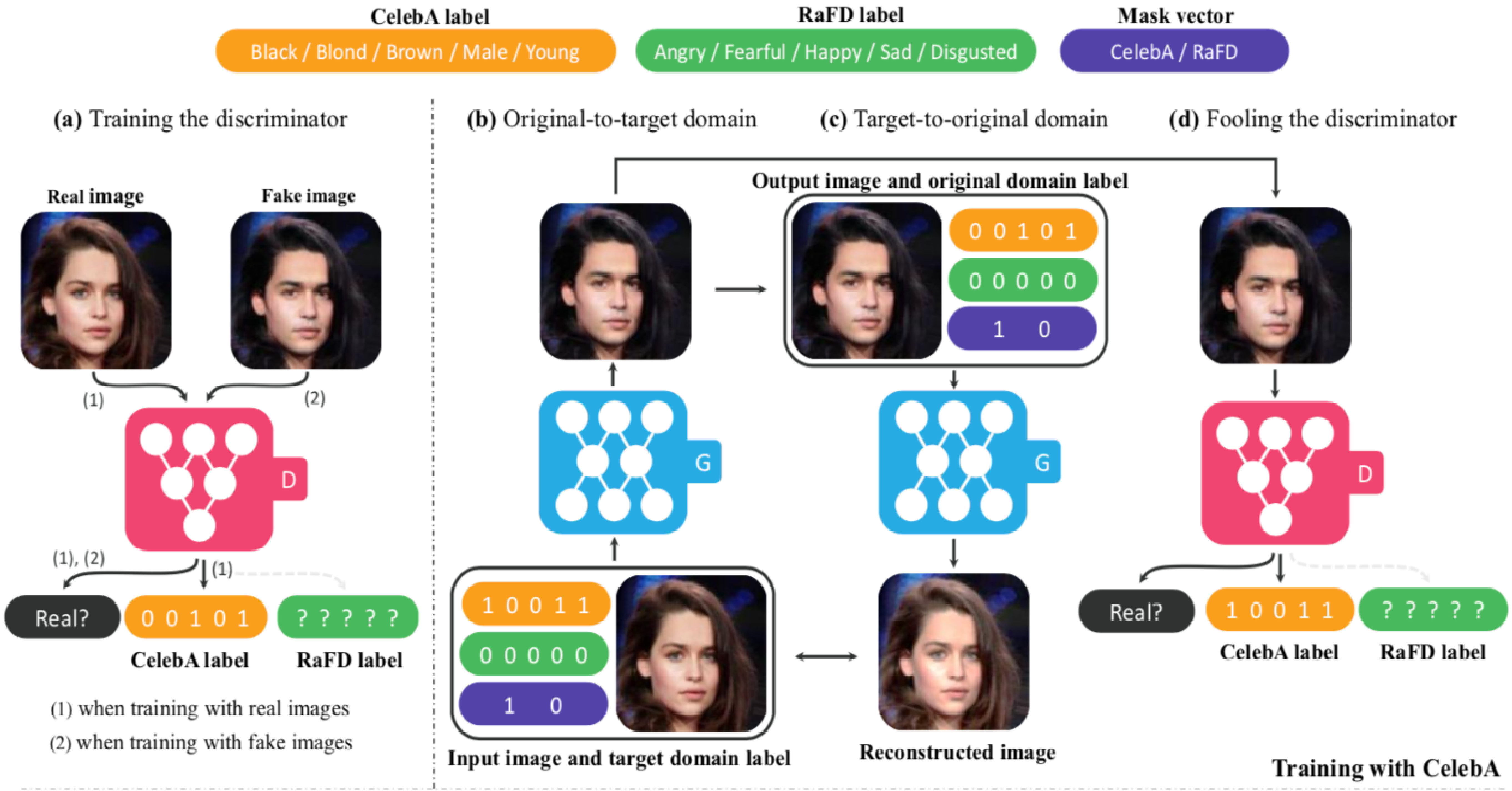
Network Architecture

Adapted from **CycleGAN**

two **convolutional** layers with stride size of two for down-sampling,
six **residual** blocks,
two **transposed convolutional** layers

StarGAN Design

Training with Multiple Datasets

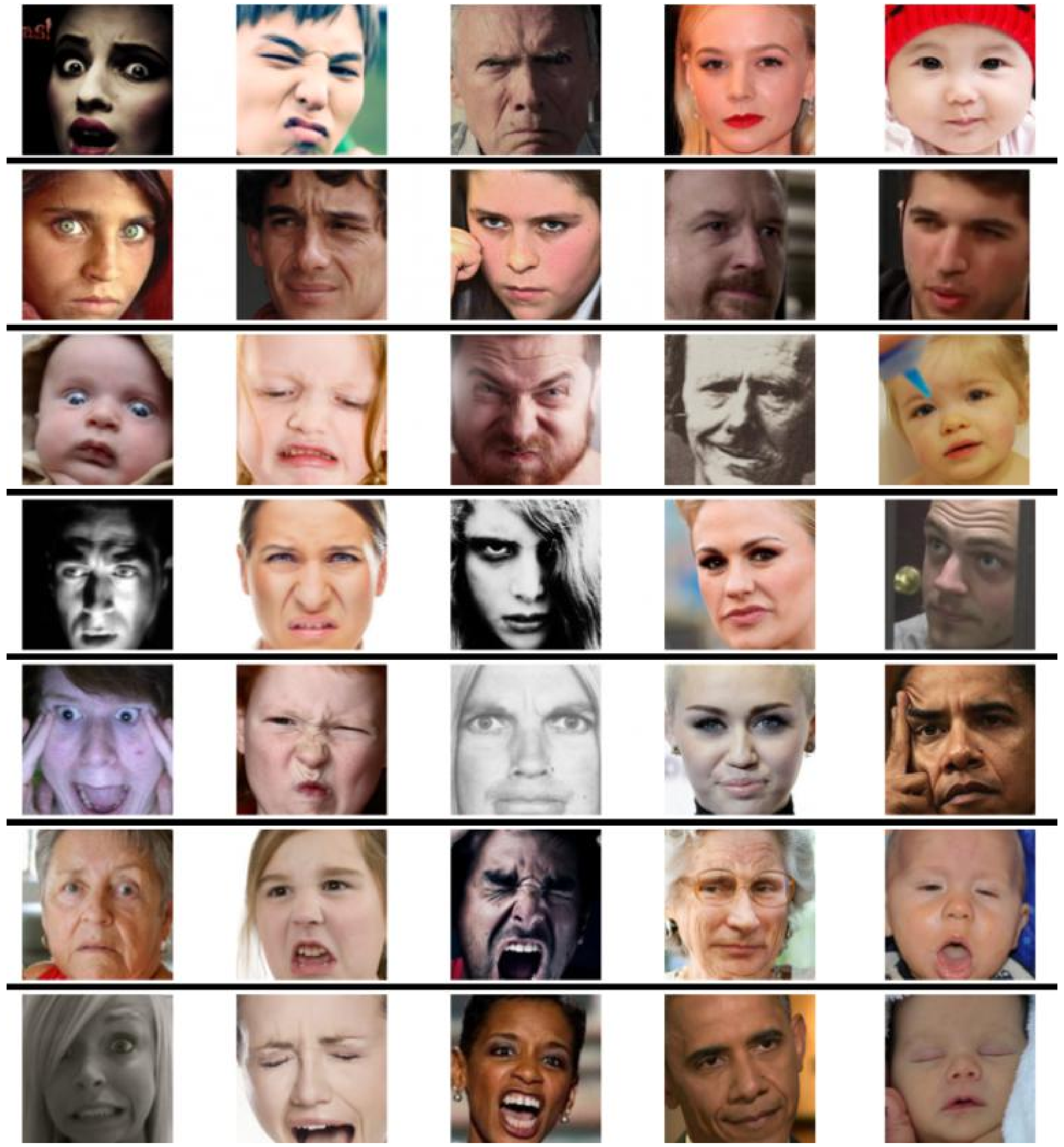


Mask Vector ignore unspecified labels, focus on the explicitly known label

Experimental Results

Datasets

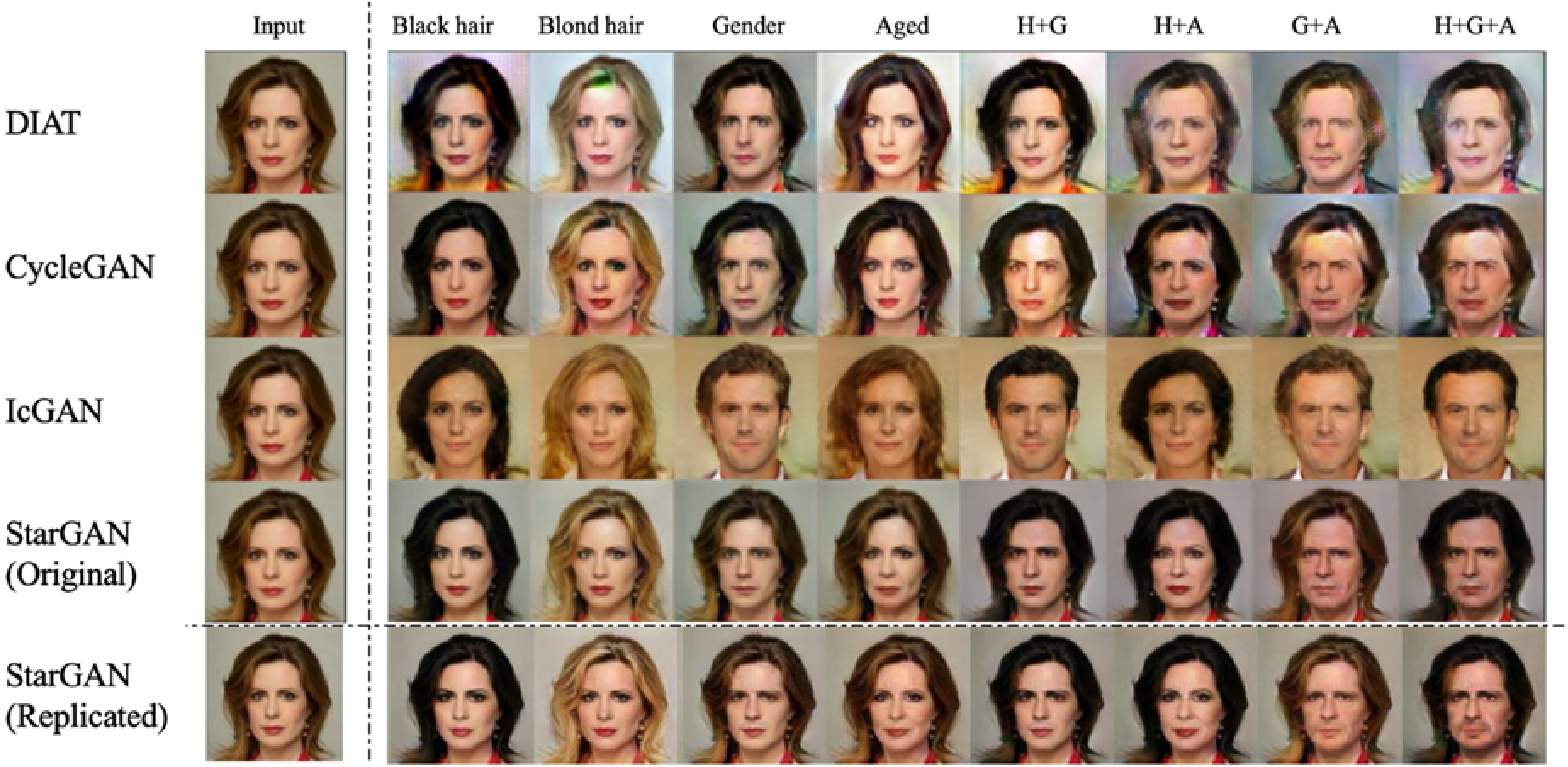
CelebA



AffectNet Dataset

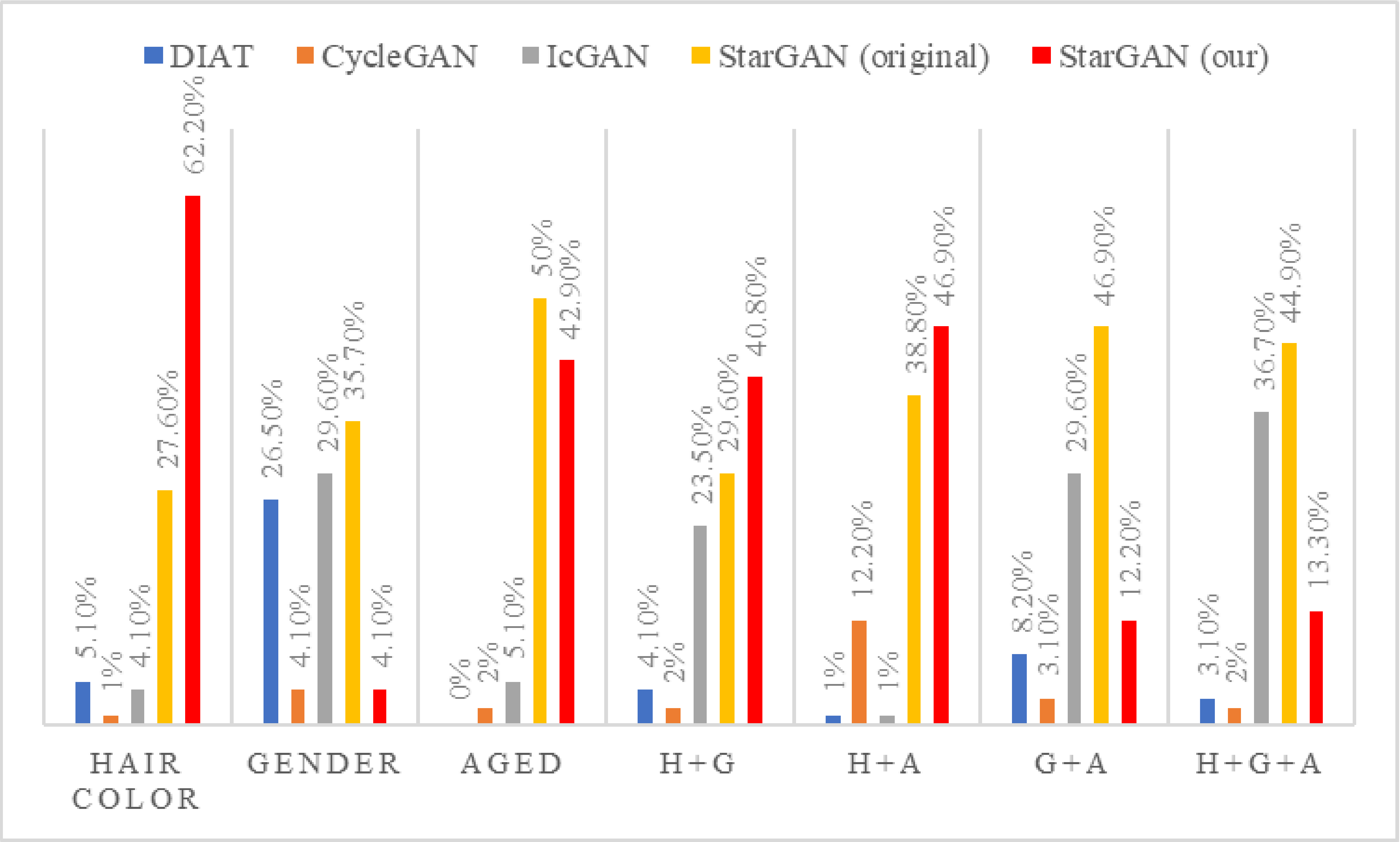
Experimental Results

Experimental Results on CelebA



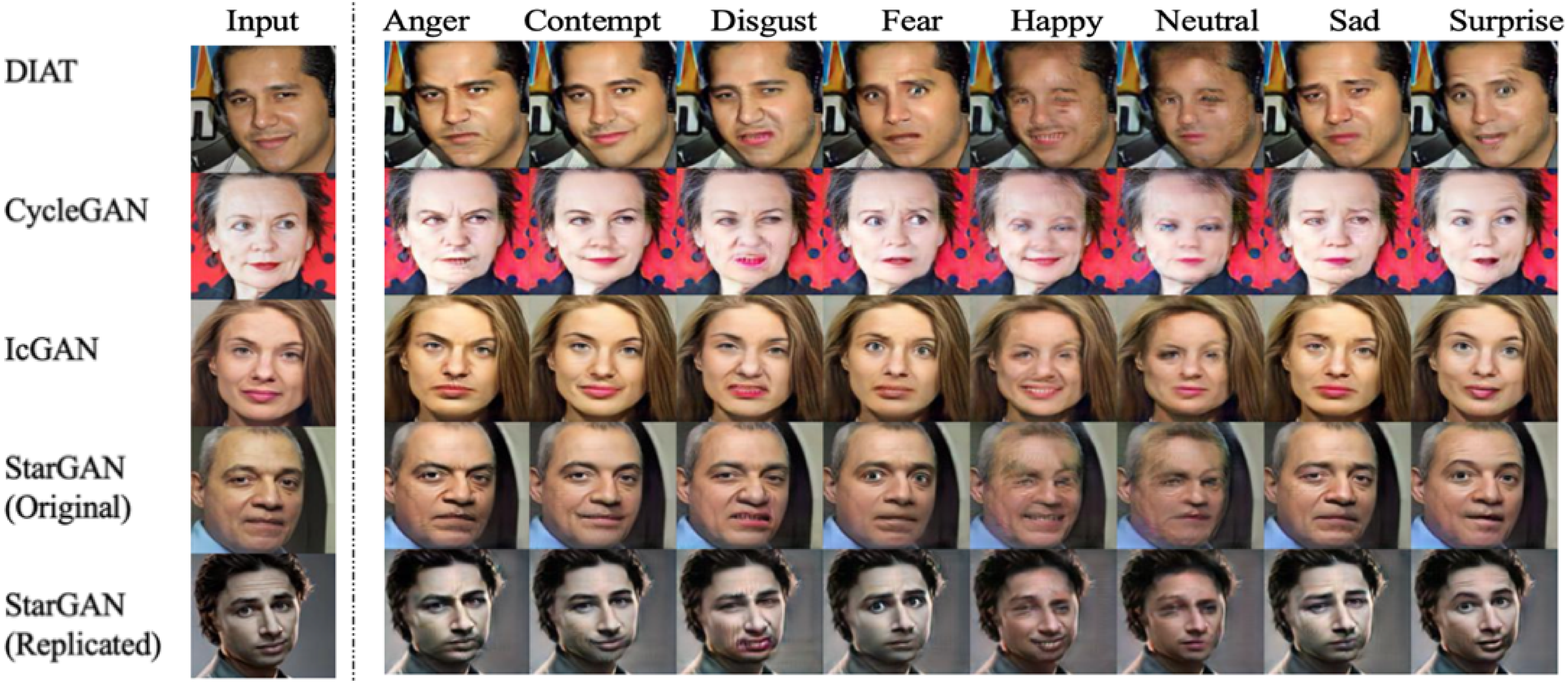
Experimental Results

Experimental Results on CelebA



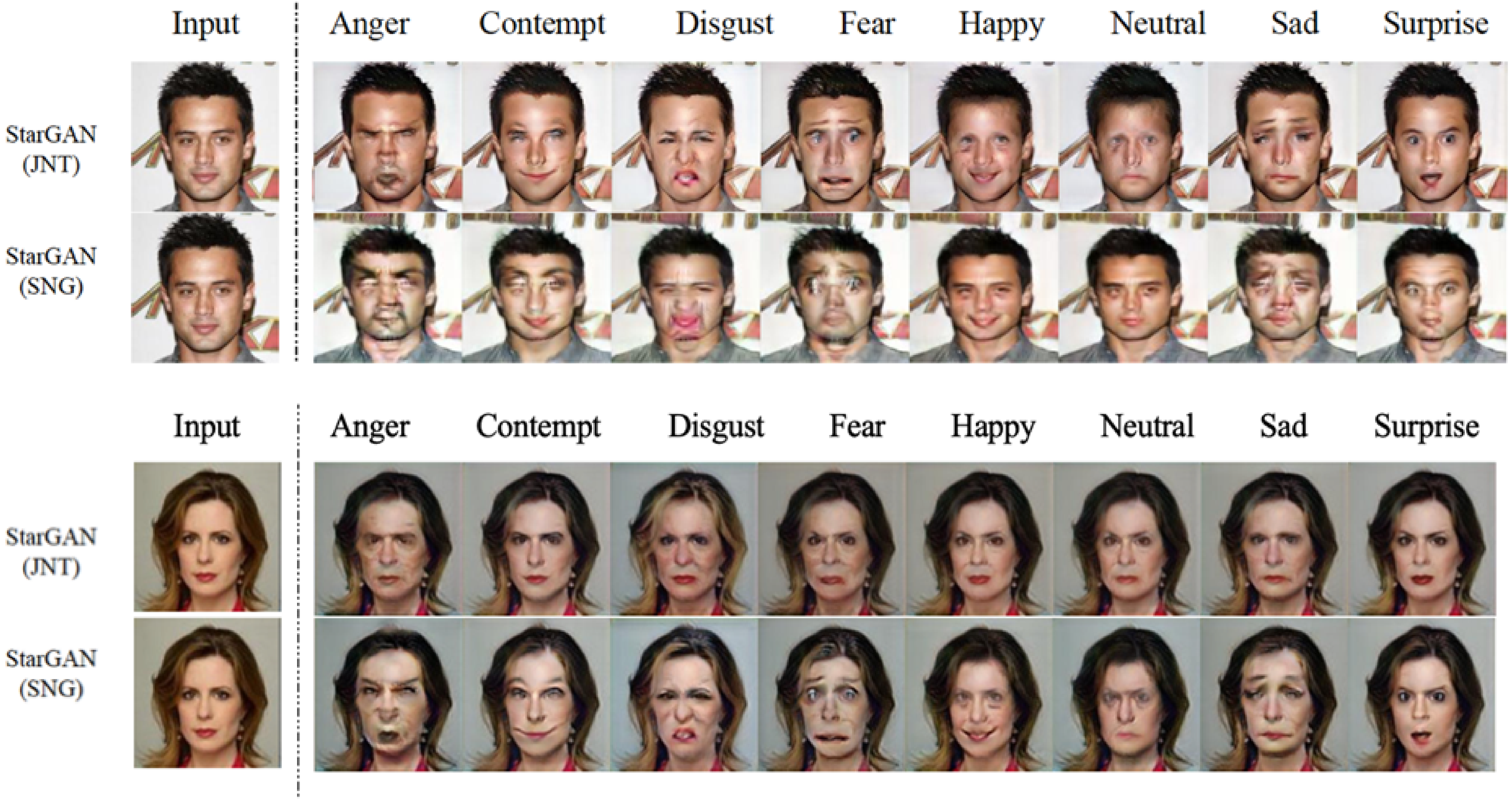
Experimental Results

Experimental Results on AffectNet



Experimental Results

Experimental Results on CelebA + AffectNet



Model Improvement Suggestions

Problems

current model has low performance at
recognizing **geometric or structural patterns** in face

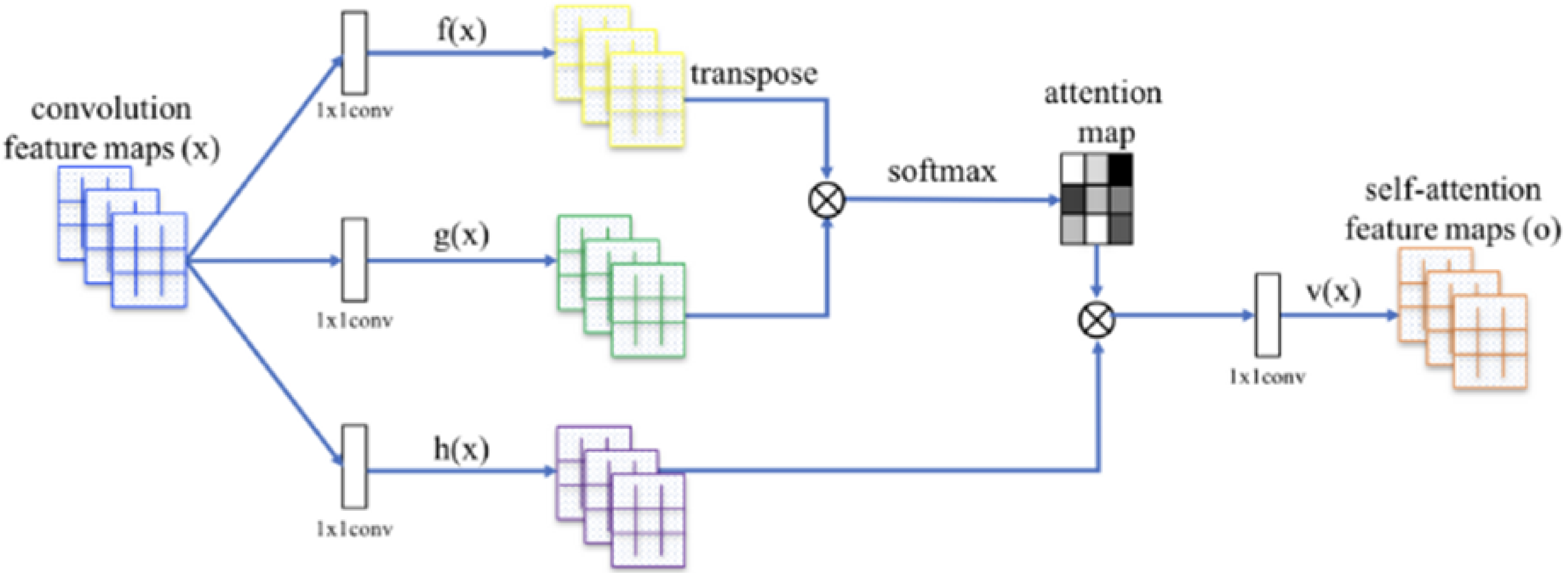
Using **only convolution layers** leads to lack of recognition of structural patterns

Model Improvement Suggestions

Our solution: Self-Attention Mechanism

Self-Attention

Fully utilize the **long-range dependency** in internal representation of image



Model Improvement Suggestions

Our solution: Self-Attention Mechanism

Amount of relation between x_i and x_j

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})}, \qquad where\ \boldsymbol{s_{ij}} = \boldsymbol{f}(\boldsymbol{x_i})^T \boldsymbol{g}(\boldsymbol{x_j})$$

Output feature map

$$o = v\left(\sum_{i=1}^{N} \beta_{j,i}\, h(x_i)\right), h(x_i) = W_h x_i, v(x_i) = W_v x_i$$

Final output

$$y_i = \gamma o_i + x_i$$

Loss function of attention applied

$$L_D = -E_{x,y\sim P_{data}}\left[\min\left(0, -1 + D(x,y)\right)\right] - E_{z\sim P_z, y\sim P_{data}}\left[\min\left(0, -1 - D(G(z),y)\right)\right]$$

$$L_G = -E_{z\sim P_z, y\sim P_{data}}[D(G(z),y)]$$

Discussion

StarGAN Design

Solve the **limited scalability** and **robustness issue**

Modified the adversarial loss function to the objective function of **WGAN** (Wasserstein GAN)

Increase the performance of StarGAN at recognizing geometric or structural patterns in face by **Self-Attention**

Discussion

Qualitative Evaluation

CelebA

Our StarGAN model **outperforms** the baseline model

AffectNet

Observed a **low-quality** conversion

Gave a **logical reason** for this phenomenon

StarGAN-JNT

Joint training with CelebA and AffectNet

Conclusion

StarGAN

High
Generalization
Capability

Multiple
Dataset

Mask Vector

Improvement

Self Attention

Future
Investigation

Balance
Discriminator
Generator