# StarGAN

CVPR

2018

## StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

Yunjey Choi[1,2] Minje Choi[1,2] Munyoung Kim[2,3] Jung-Woo Ha[2] Sunghun Kim[2,4] Jaegul Choo[1,2]
[1] Korea University [2] Clova AI Research, NAVER Corp.
[3] The College of New Jersey [4] Hong Kong University of Science & Technology

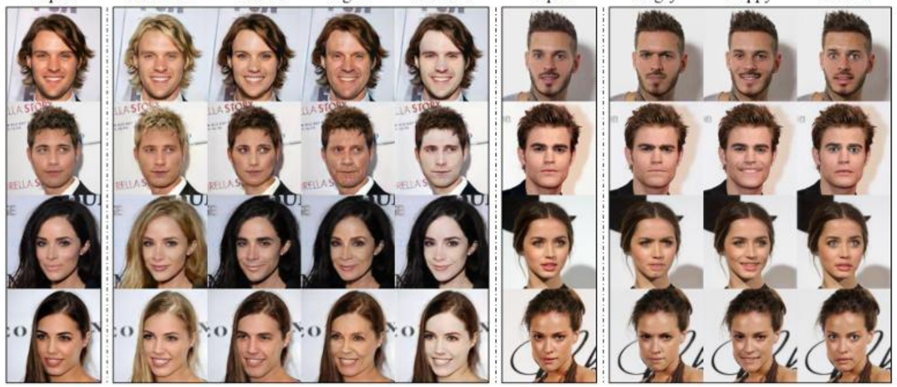



Figure 1. Multi-domain image-to-image translation results on the CelebA dataset via transferring knowledge learned from the RaFD dataset. The first and sixth columns show input images while the remaining columns are images generated by StarGAN. Note that the images are generated by a single generator network, and facial expression labels such as angry, happy, and fearful are from RaFD, not CelebA.

### Abstract

*Recent studies have shown remarkable success in image-to-image translation for two domains. However, existing approaches have limited scalability and robustness in handling more than two domains, since different models should be built independently for every pair of image domains. To address this limitation, we propose StarGAN, a novel and scalable approach that can perform image-to-image translations for multiple domains using only a single model. Such a unified model architecture of StarGAN allows simul-*

### 1. Introduction

The task of image-to-image translation is to change a particular aspect of a given image to another, e.g., changing the facial expression of a person from smiling to frowning (see Fig. 1). This task has experienced significant improvements following the introduction of generative adversarial networks (GANs), with results ranging from changing hair color [9], reconstructing photos from edge maps [7], and changing the seasons of scenery images [33].

Given training data from two different domains, these

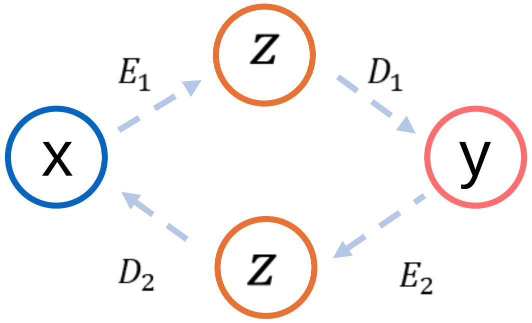
Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

Image-to-Image Translation

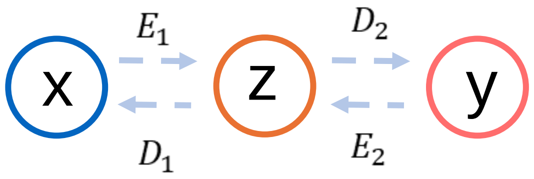
# StarGAN

## Comparison of Approaches

**CycleGAN**

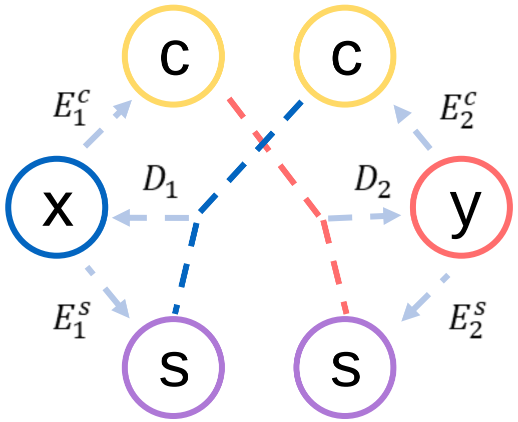



**UNIT**



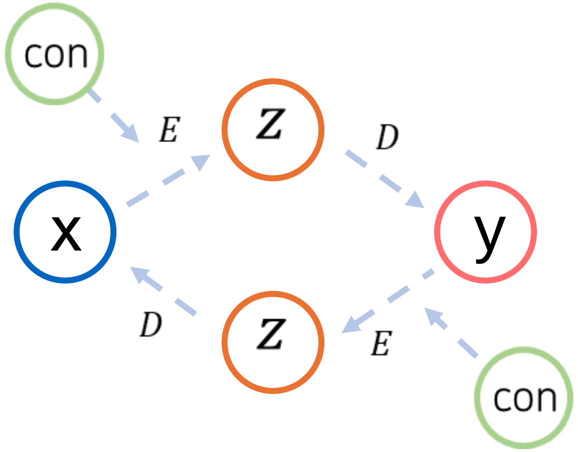

Shared Latent Space

**MUNIT**




Content & Style

**StarGAN**




Domain Condition

Image-to-Image Translation

# StarGAN

## Background & Goal

### Previous research limitation

- CycleGAN is inefficient in such multi-domain image translation tasks. (k×(k-1))

### Goal

- Multi-domain image translation using only one generator.

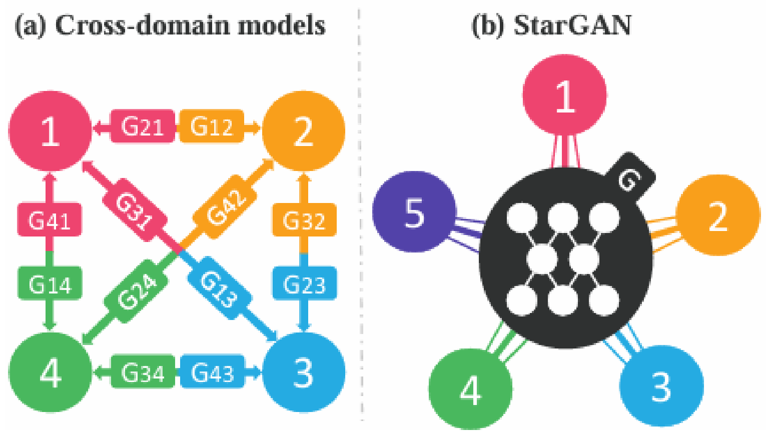


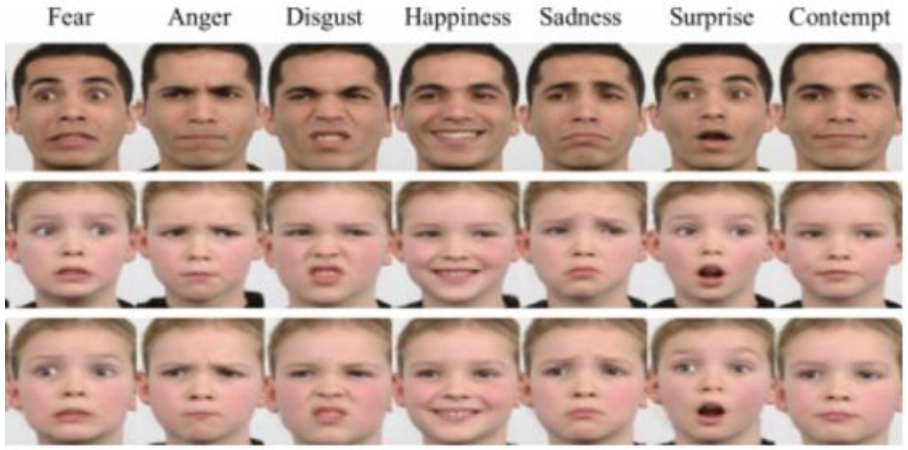

Fig 1. Comparison between cross-domain models and StarGAN

Image-to-Image Translation

# StarGAN

## What is domain?

- A set of images sharing the same attribute value.

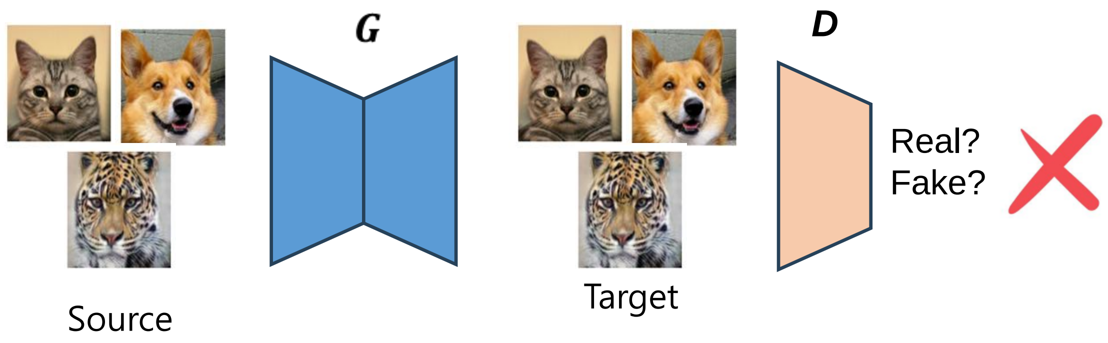

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

## Image-to-Image Translation

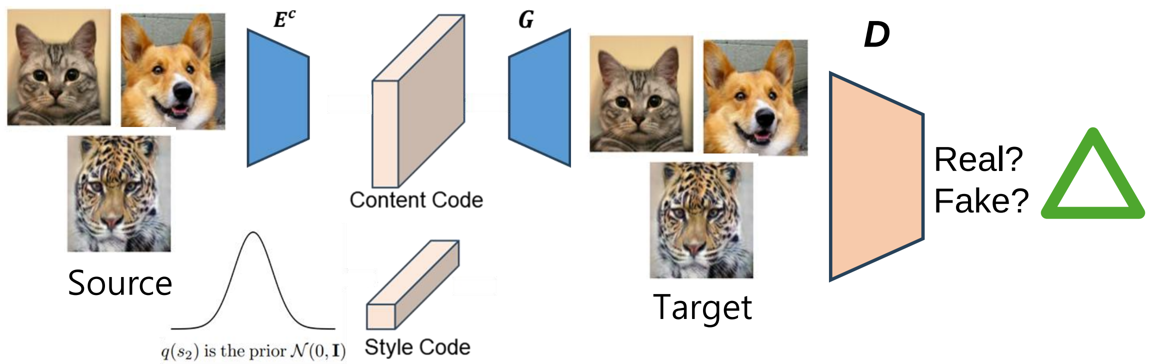
# StarGAN

### Source = Target Scenario

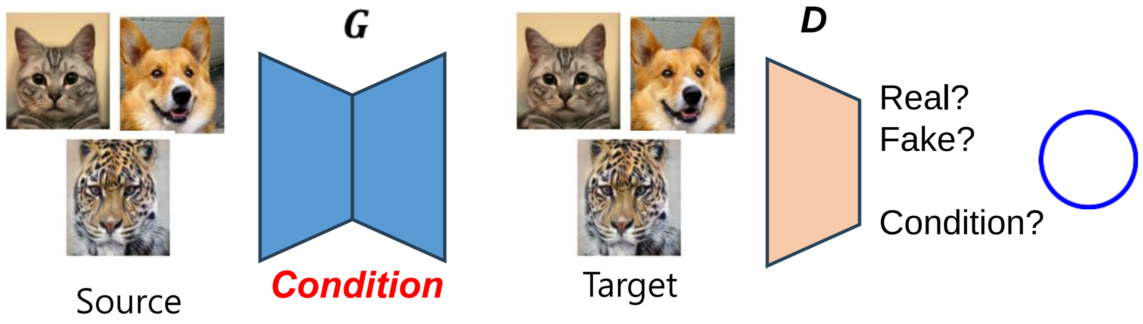
**CycleGAN & UNIT**




**MUNIT**




**StarGAN**




Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN



## Network Architecture

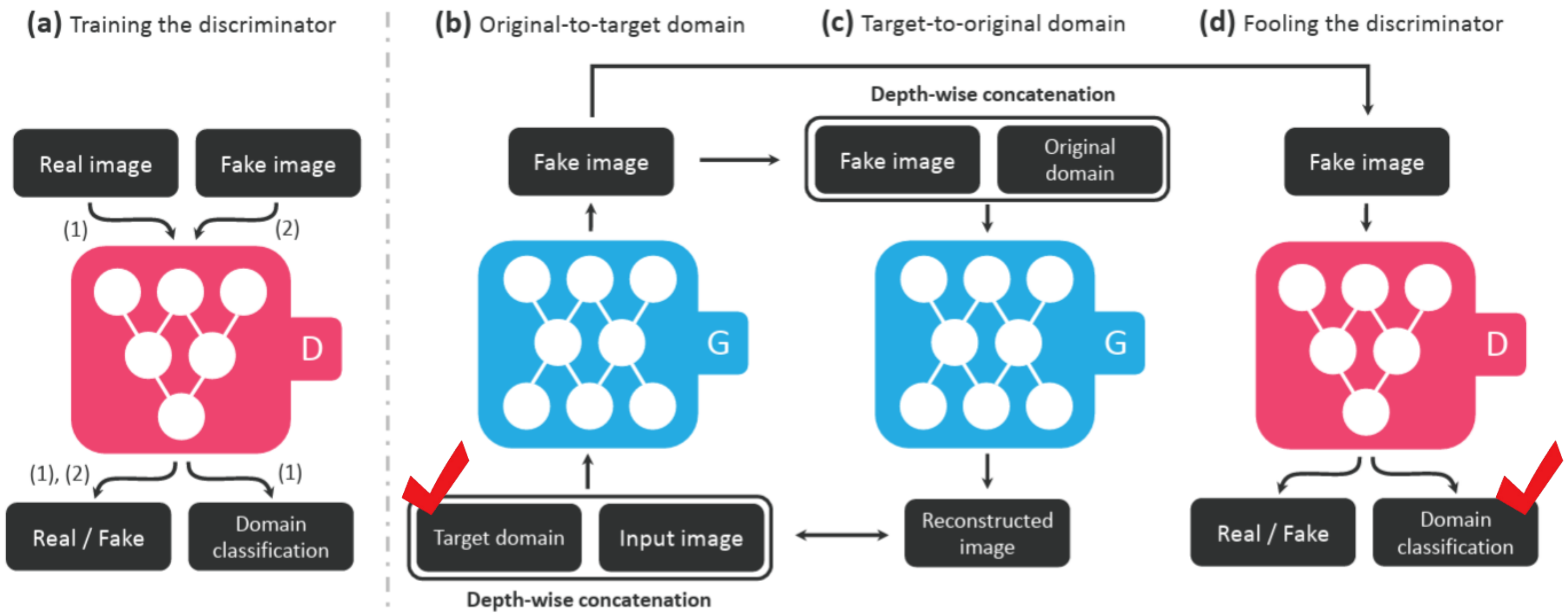



Fig 2. Overview of StarGAN

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN

## Loss Functions

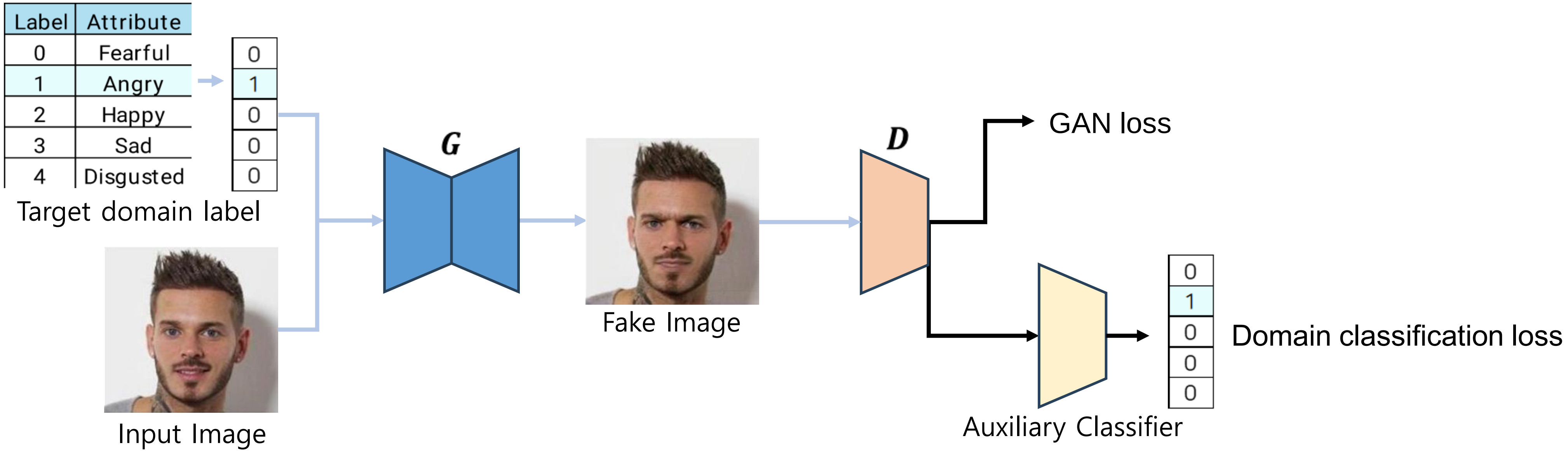



Fig 4. StarGAN Architecture

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN



## Loss Functions

### Why not use a separate classification network?

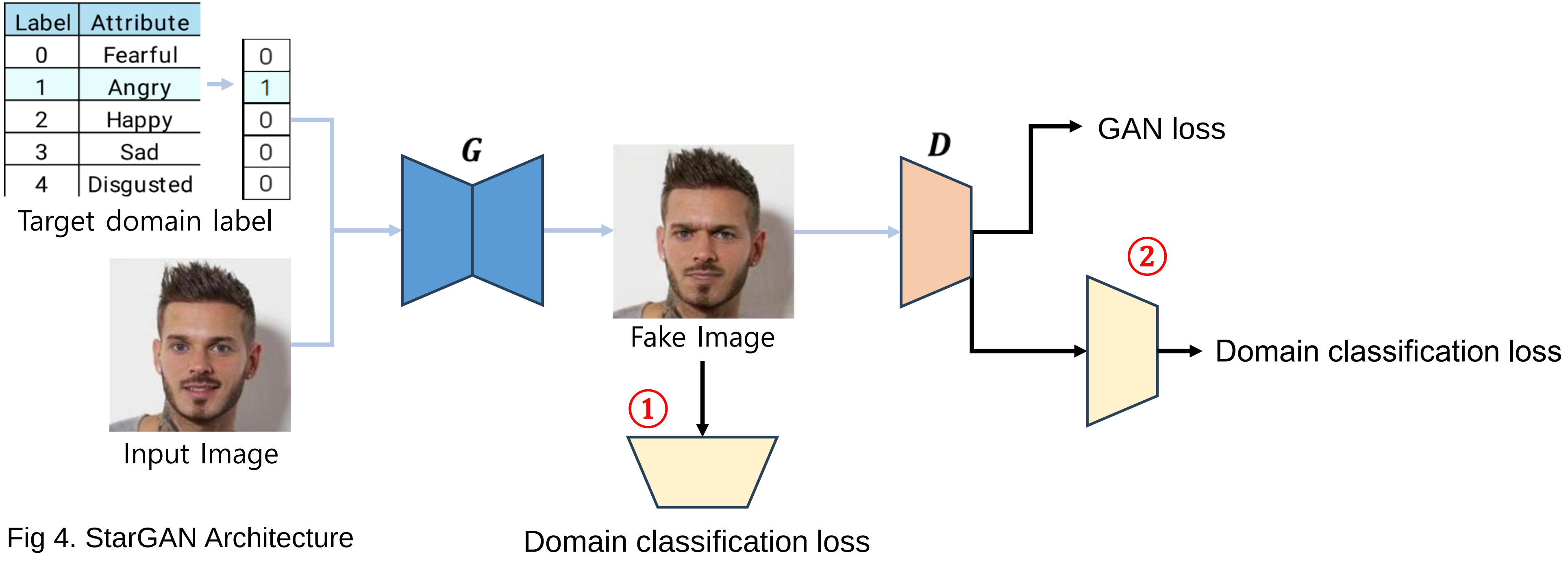



Fig 4. StarGAN Architecture

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN

## Loss Functions

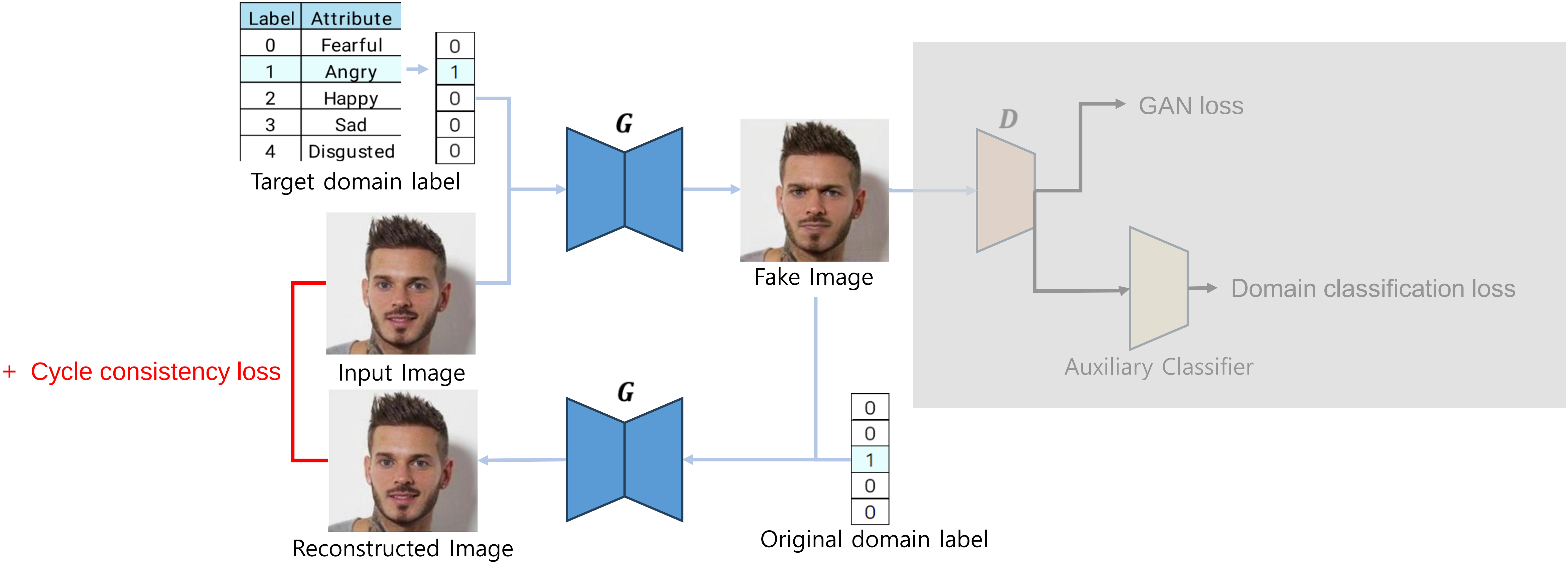



Fig 5. Cycle Consistency Architecture

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

Medical Artificial Intelligence Laboratory
At Yonsei University

# StarGAN

## Depth-wise Concatenation

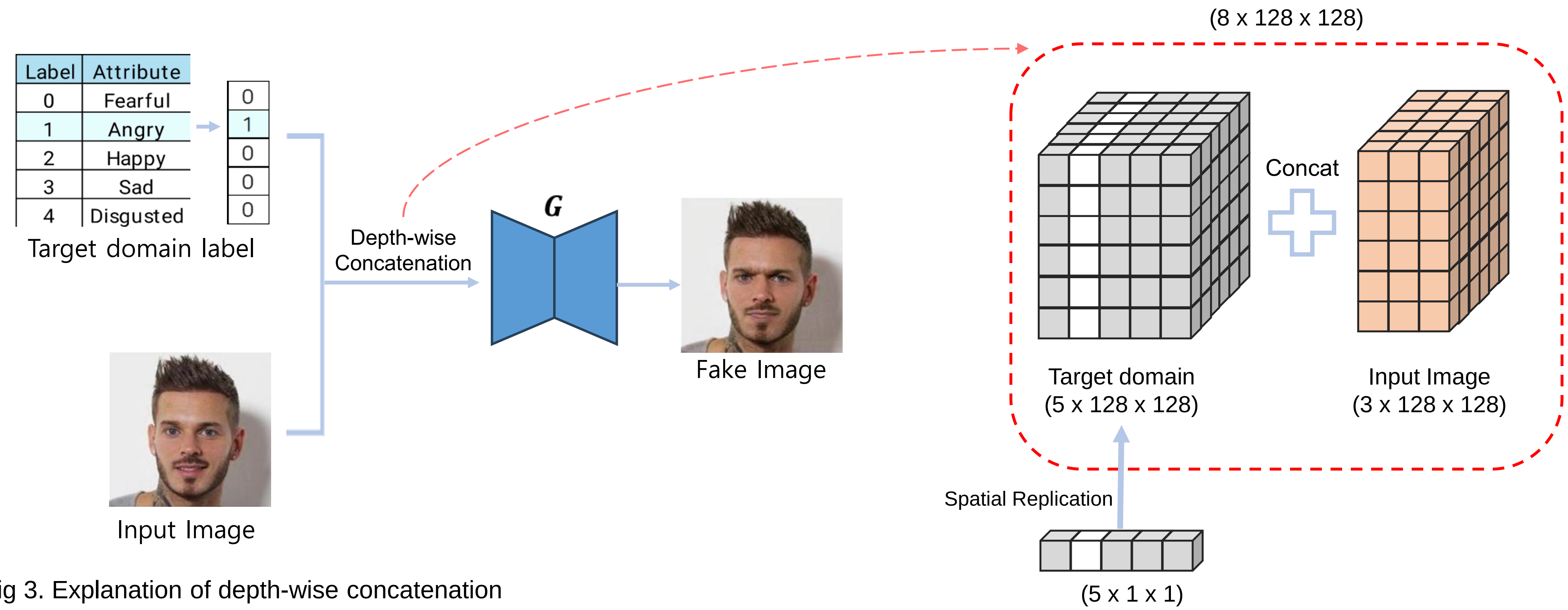



Fig 3. Explanation of depth-wise concatenation

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN

Medical Artificial Intelligence Laboratory
At Yonsei University

## Loss Functions

### GAN loss

- $G$ generates an image $G(x, c)$ conditioned on both the input image $x$ and the target domain $c$.

$$L_{\text{adv}} = \mathbb{E}_x[\log D_{\text{src}}(x)] + \mathbb{E}_{x,c}[\log(1 - D_{\text{src}}(G(x, c)))]$$

Eq 1. Adversarial Loss

### Domain classification loss

- A domain classification loss uses an auxiliary classifier. ($c'$ : original domain label, $c$ : target domain label)

$$L^r_{\text{cls}} = \mathbb{E}_{x,c'}[-\log D_{\text{cls}}(c'|x)] \qquad L^f_{\text{cls}} = \mathbb{E}_{x,c}[-\log D_{\text{cls}}(c|G(x, c))]$$

Eq 2. Domain Classification Loss

### Cycle consistency loss

- Apply Cycle consistency loss

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x, c), c')||_1]$$

Eq 3. Reconstruction Loss

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

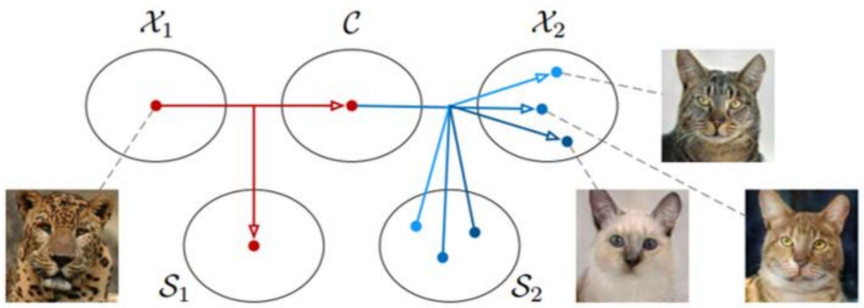
Image-to-Image Translation

# StarGAN

## Comparison with MUNIT

### MUNIT

- Content-style disentanglement
- Diverse outputs within the same domain
- If multiple domains are trained together, there is a risk of e ntanglement between different attributes.




### StarGAN

- No disentanglement; relies only on domain labels
- Single output per domain
- Handles multiple domains with a single network

Image-to-Image Translation

# StarGAN

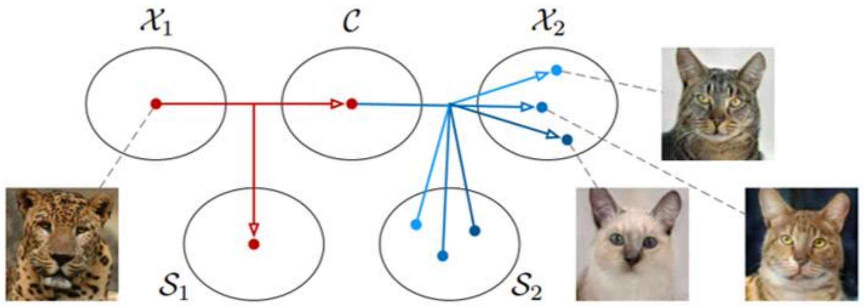
## Comparison with MUNIT

### MUNIT

- Content-style disentanglement
- Diverse outputs within the same domain
- If multiple domains are trained together, there is a risk of entanglement between different attributes.

### StarGAN

- No disentanglement; relies only on domain labels
- Single output per domain
- Handles multiple domains with a single network





### Recent Trend

- Hybrid : Single network with domain label + Content-style disentanglement

Image-to-Image Translation

# StarGAN

## Experiment Results

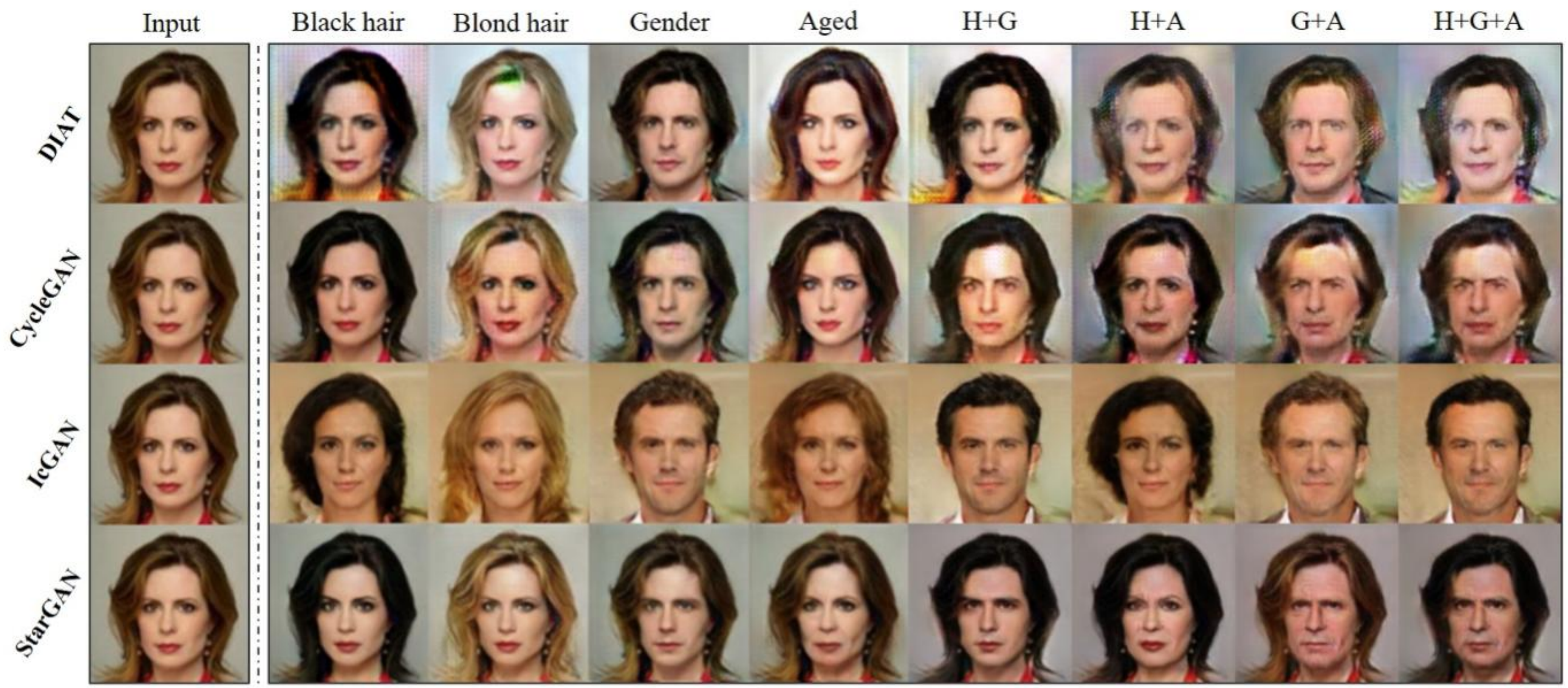



Fig 6. Results on the CelebA dataset

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

Image-to-Image Translation

# StarGAN

## Experiment Results

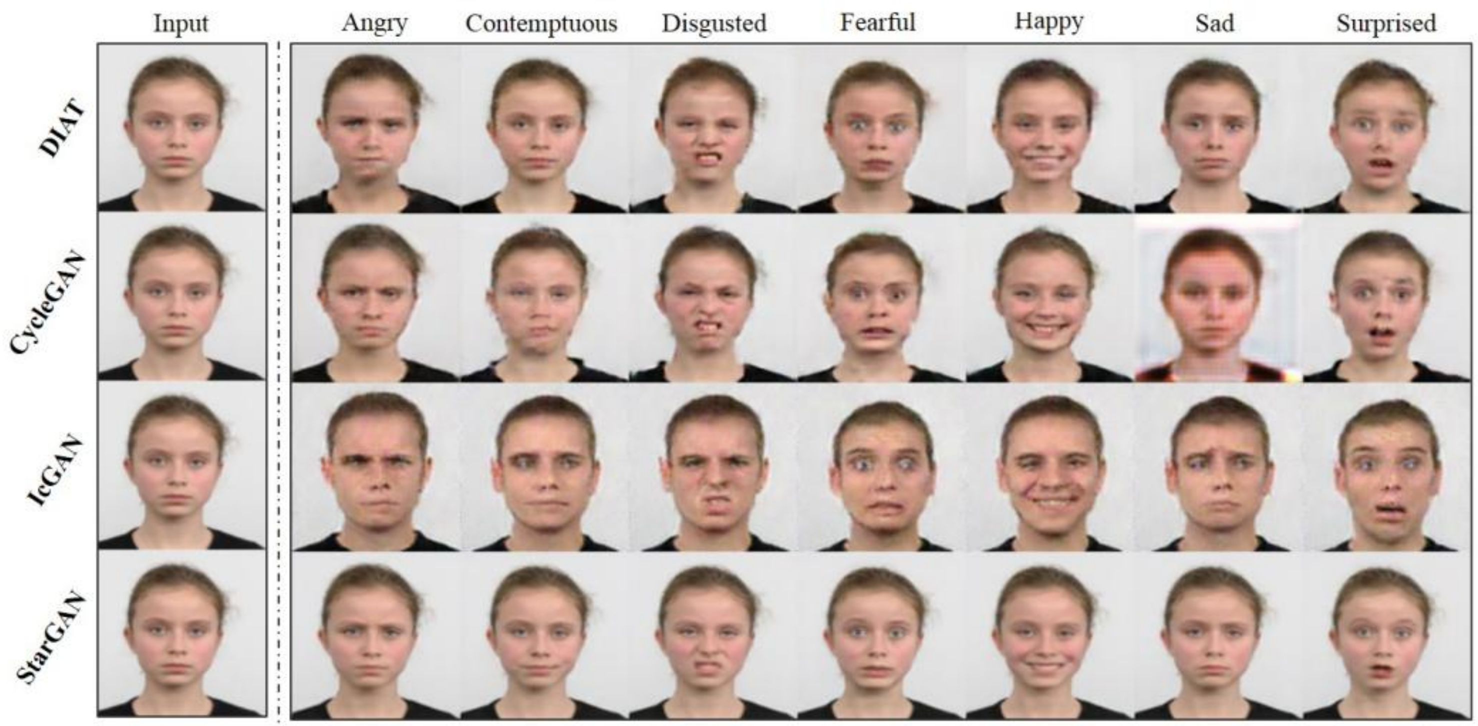



Fig 7. Results on the RaFD dataset

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN

## Comparison with Single-task Network

### Pros

- Feature sharing: A single network learns common features (e.g., facial contours, skin texture) across domains, improving data efficiency and generalization.
- Memory efficiency: As the number of domains increases, memory usage remains relatively stable since a single network is used.

### Cons

- Task interference: Handling diverse domain transformations simultaneously can degrade performance.
- Limited specialization : Handling all domains uniformly within a single network may result in suboptimal transformation performance for each domain.

# StarGAN

## Comparison with Single-task Network

### Pros

- Feature sharing: A single network learns common features (e.g., facial contours, skin texture) across domains, impr oving data efficiency and generalization.
- Memory efficiency: As the number of domains increases, memory usage remains relatively stable since a single net work is used.

### Cons

- Task interference: Handling diverse domain transformations simultaneously can degrade performance.
- Limited specialization : Handling all domains uniformly within a single network may result in suboptimal transformati on performance for each domain.

### Recent Trend: MoE (Mixture of Experts)

- Combining shared feature learning with domain-specific expert sub-networks to overcome these limitations.

# StarGAN



## Training Strategy

### Mask vector

- To ignore unspecified labels and focus on the explicitly known label provided by a dataset

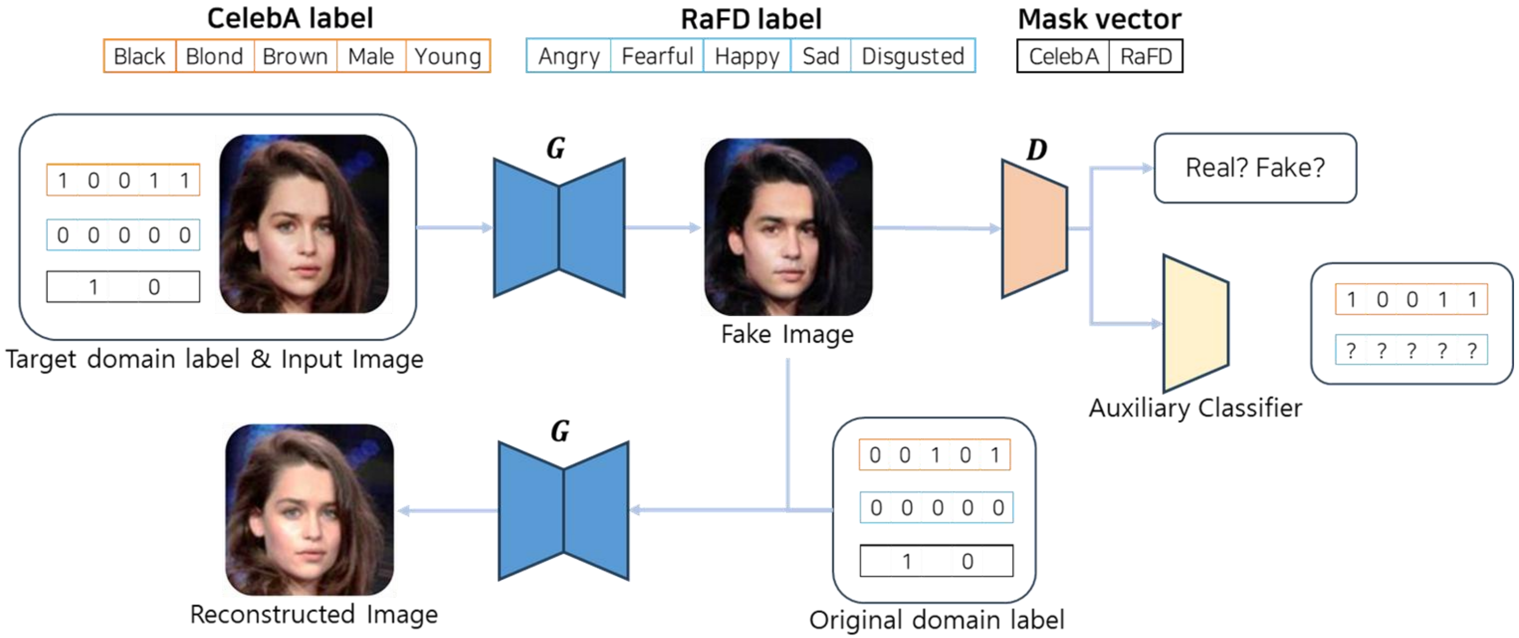



Fig 8. StarGAN with mask vector

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

Image-to-Image Translation

# StarGAN

## Training Strategy


Fig 9. Facial expression synthesis results on the CelebA dataset

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

# StarGAN

## Implications & Limitations

### Implications

- A scalable image-to-image translation model among multiple domains using a single generator & discriminator
- It can handle multiple datasets with different domain label sets.

### Limitations

- Accurately categorized domain labels are essential.
- A discrete model that generates a single deterministic output for a given domain label.