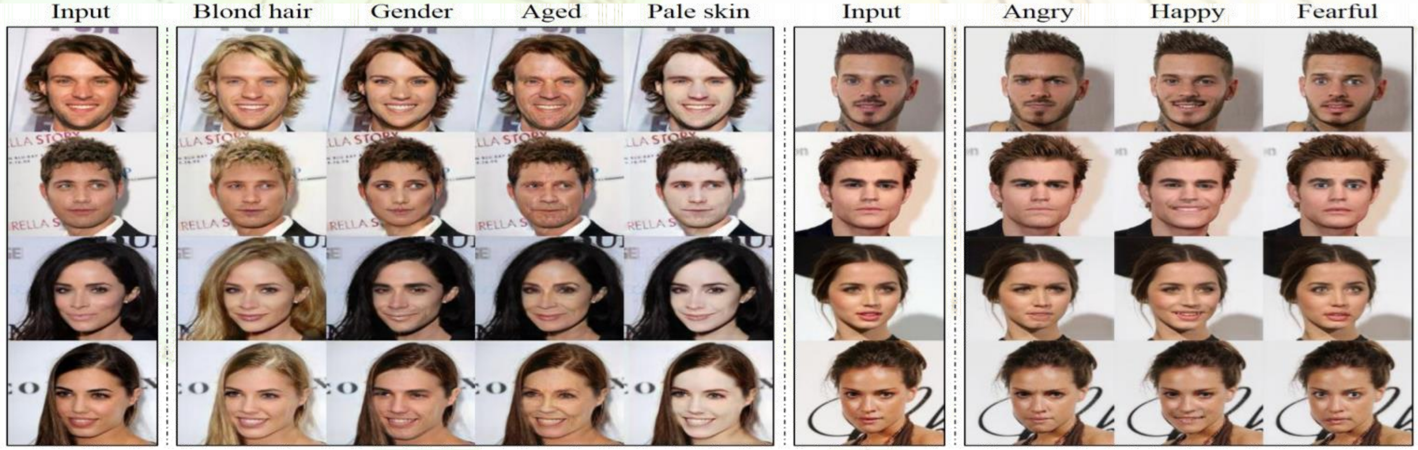
StarGAN

Pattern Recognition & Machine Learning Laboratory
Tae-jin Woo

Aug 17, 2021

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (1/7)

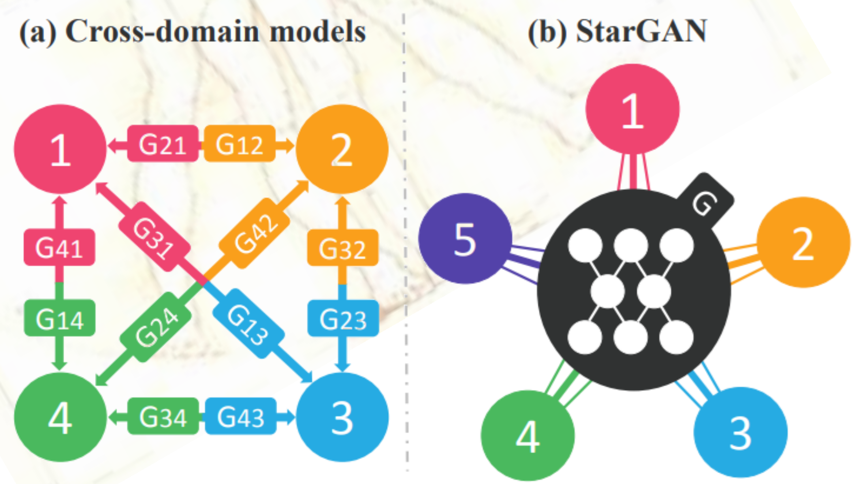
- **Goal**
  - Solving scalability and robustness in handling more than two domains
- **Motivation**
  - Improving network efficiency in multi-domain image-to-image translation
  - Learning general features in multi-domain image-to-image translation
- **Contribution**
  - Proposing a single network for multi-domain image-to-image translation
  - Proposing joint training method on multi-dataset
  - Achieving State-of-the-Art (SOTA) in image-to-image translation



Multi-domain image-to-image translation results

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (2/7)

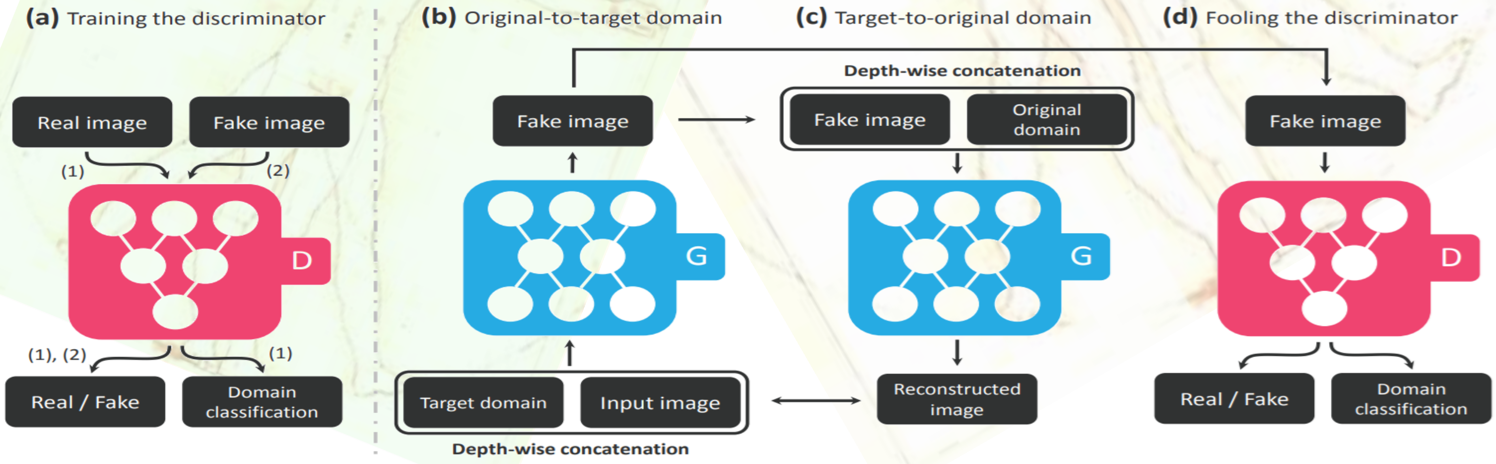
- **Introduction**
  - **Background**
    - **Required prior knowledge in image-to-image translation**
      - Attribute: Meaningful feature inherent in an image ex) color, gender
      - Attribute value: Particular value of an attribute ex) black, female
      - Domain: A set of images sharing the same attribute value
    - **Problems of previous cross-domain models**
      - Separate network required for each domain in multi-domain image translation
      - Unable to fully utilize the existing global features from entire training data
  - **Proposals**
    - **Discriminator**
      - Applying domain classification
      - Enable to handle multi-domain task
    - **Generator**
      - Applying domain classification
      - Applying reconstruction
      - Enable to learn general features
    - **Others**
      - Applying mask vector
      - Enable to handle multi-domain task



Comparison between cross-domain models and StarGAN

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (3/7)

- **Architecture**
  - **Modules**
    - **Discriminator**
      - Learning to distinguish between real and fake images
      - Classifying the real images to its corresponding domain
    - **Generator**
      - Taking in as input both the image and target domain and generates an fake image
      - Learning to reconstruct the original image from the fake and original domain
      - Learning the reconstructed image to be similar to the original image
      - Generating images indistinguishable from real images and classifiable



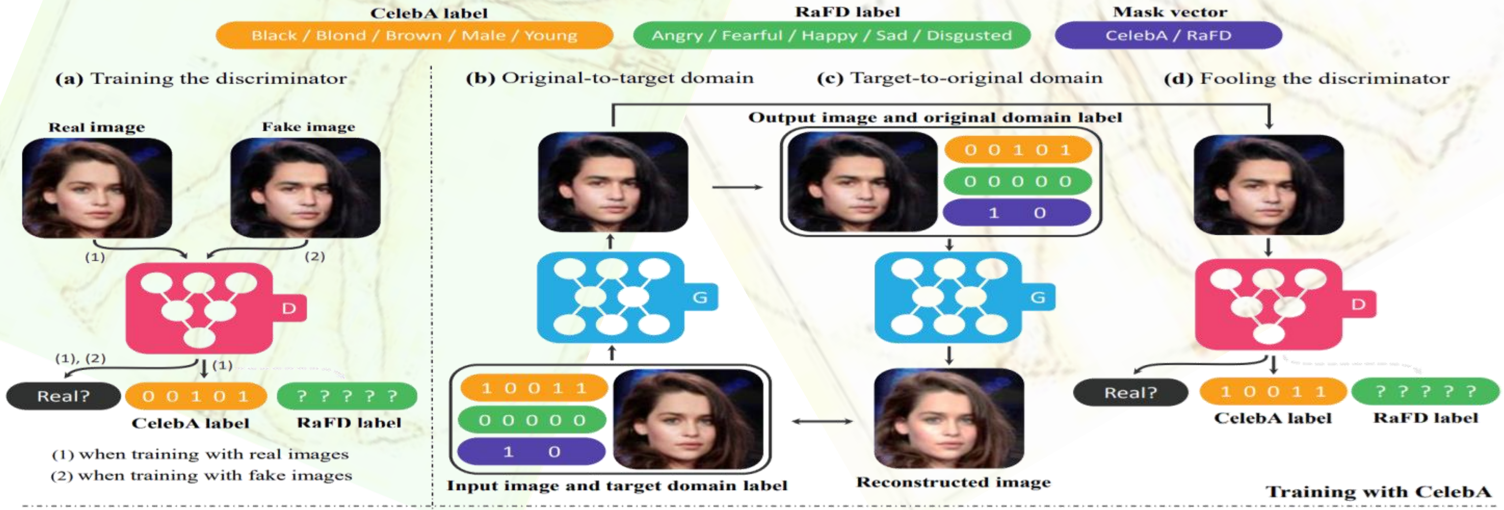
Overview of StarGAN architecture

- **Training for multi-domain image-to-image translation**
  - **Loss functions**
    - **Adversarial loss**
      - $\mathcal{L}_{adv} = E_x[\log D_{src}(x)] + E_{x,c}[\log(1 - D_{src}(G(x,c)))]$
      - x: Input real image
      - $D_{src}(x)$: Probability distribution over sources given by G
      - c: Target domain label
    - **Domain classification loss**
      - $\mathcal{L}_{cls}^{r} = E_{x,c'}[-\log D_{cls}(c'|x)]$ for real images
      - c': Original domain label
      - $D_{cls}(c'|x)$: Probability distribution over original domain labels
      - $\mathcal{L}_{cls}^{f} = E_{x,c}[-\log D_{cls}(c|G(x,c))]$ for fake images
    - **Reconstruction loss**
      - $\mathcal{L}_{rec} = E_{x,c,c'}[||x - G(G(x,c),c')||_1]$
      - Adopting the L1 norm as reconstruction loss
      - Translating into target domain first, and then reconstructing the original image
    - **Full objective loss**
      - $\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{r}$ for discriminator
      - $\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{f} + \lambda_{rec}\mathcal{L}_{rec}$ for generator
      - $\lambda_{cls}, \lambda_{rec}$: Hyper-parameters that control the relative importance among losses

Y. Choi et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," Computer Vision & Pattern Recognition, 2018.

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (5/7)

- Training with multiple dataset
  - Mask vector
    - Previous problems
      - Only partially known label information is given when learning multiple dataset
    - Solution
      - Introducing mask vector $\tilde{c} = [c_1, c_2, \ldots, c_n, m]$
  - Training strategy
    - Cross-over method
      - Ignoring and focusing on specified labels by taking mask vector as an input



Overview of StarGAN training with both CelebA and RaFD

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (6/7)

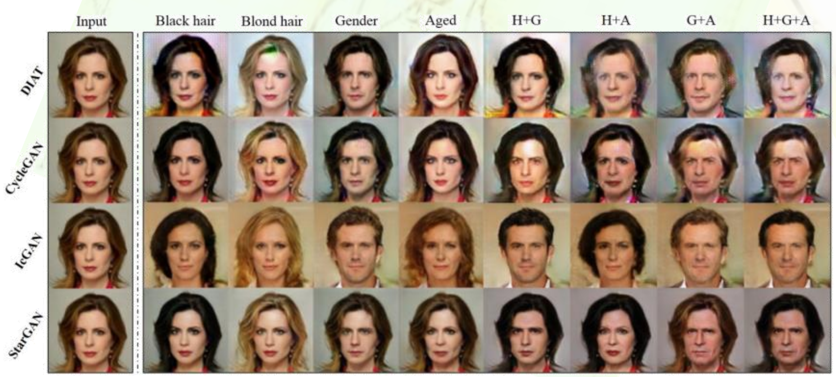
- **Experiments**
  - Qualitative evaluation
    - Reasons for SOTA
      - Regularization effect through multi-domain
      - Learning reliable features universally
      - Maintaining spatial information using ConvNet
  - Quantitative evaluation
    - AMT perceptual evaluation
      - Achieving 1st ranking on both single and multi-attribute transfer task
    - Efficiency
      - Having fewer parameters than others

Method	H+G	H+A	G+A	H+G+A
DIAT	20.4%	15.6%	18.7%	15.6%
CycleGAN	14.0%	12.0%	11.2%	11.9%
IcGAN	18.2%	10.9%	20.3%	20.3%
StarGAN	**47.4%**	**61.5%**	**49.8%**	**52.2%**

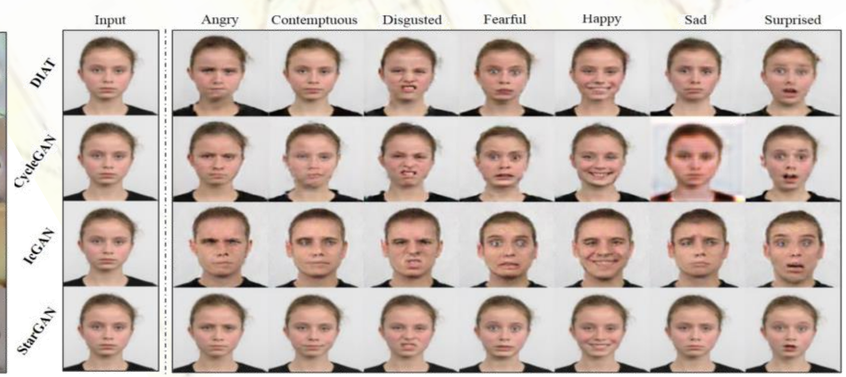
AMT perceptual evaluation results

Method	Classification error	# of parameters
DIAT	4.10	52.6M × 7
CycleGAN	5.99	52.6M × 14
IcGAN	8.07	67.8M × 1
StarGAN	**2.12**	**53.2M × 1**
Real images	0.45	-

Classification errors and the number of parameters



Facial attribute transfer results on the CelebA dataset



Facial attribute transfer results on the CelebA dataset

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [Y. Choi et al., 2018] (7/7)

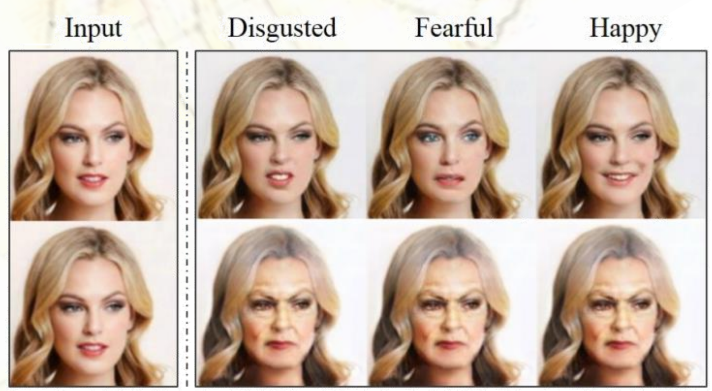
➢ Effect of joint training

- **Qualitative evaluation**
  – Showing higher visual quality in joint training, not blurry and gray backgrounds
- **Reasons for higher visual quality**
  – Enable to learn with both datasets in joint training, but not in single training
  – Enable to improve low-level tasks, which is beneficial to learning

➢ Effect of mask vector

- **Testing method**
  – Intentionally making to training G with wrong mask vector
- **Qualitative evaluation**
  – Showing fails to synthesis facial expressions
  – Enable to confirm that mask vector makes StarGAN better

Facial expression synthesis results of single and joint dataset



Learned role of the mask vector