IEEE International Conference on
Automatic Face and Gesture Recognition
FG 2021 Jodhpur, India (Virtual Event)
December 15 - 18, 2021

Spatially Constrained GAN for Face and Fashion Synthesis

Songyao Jiang[1], Hongfu Liu[2], Yue Wu[1] and Yun Fu[1]

[1] Northeastern University, Boston MA, USA

[2] Brandeis University, Waltham MA, USA

Presented by Songyao Jiang at the IEEE International Conference on Automatic Face and Gesture Recognition 2021

Synergetic Media Learning Lab

About Us

Songyao Jiang (songyaojiang.com)

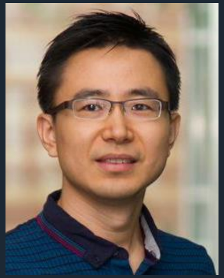
- Ph.D. candidate in Computer Engineering at Northeastern University, Boston MA, USA.
- Research interests: human detection, pose estimation, face recognition, skeleton-based action recognition, sign language recognition, and generative adversarial networks.
- *Email:* jiang.so@northeastern.edu

Dr. Hongfu Liu (hongfuliu.com)

- Faculty member affiliated with Michtom School of Computer Science at Brandeis University, Waltham MA, USA.
- Research interests: data mining and machine learning, with special interests in ensemble learning.
- *Email:* hongfuliu@brandeis.edu

Dr. Yue Wu(wuyuebupt.github.io)

- Ph.D. in Computer Engineering.
- Microsoft.
- Research interests: face recognition and object recognition.
- *Email:* yuewubupt@gmail.com



Dr. Yun Fu (www1.ece.neu.edu/~yunfu/)

- Principal Investigator and founding director of SmileLab.
- Faculty member affiliated with College of Engineering and Khoury College of Computer Science at Northeastern University, Boston MA, USA.
- Fellow of IEEE, OSA, SPIE, IAPR.
- Successful Serial Entrepreneur
- Research interests: machine learning, computational intelligence, big data mining, computer vision, pattern recognition, and cyber-physical systems.
- *Email:* yunfu@ece.neu.edu

Smile lab

Synergetic Media Learning Lab

web.northeastern.edu/smilelab/

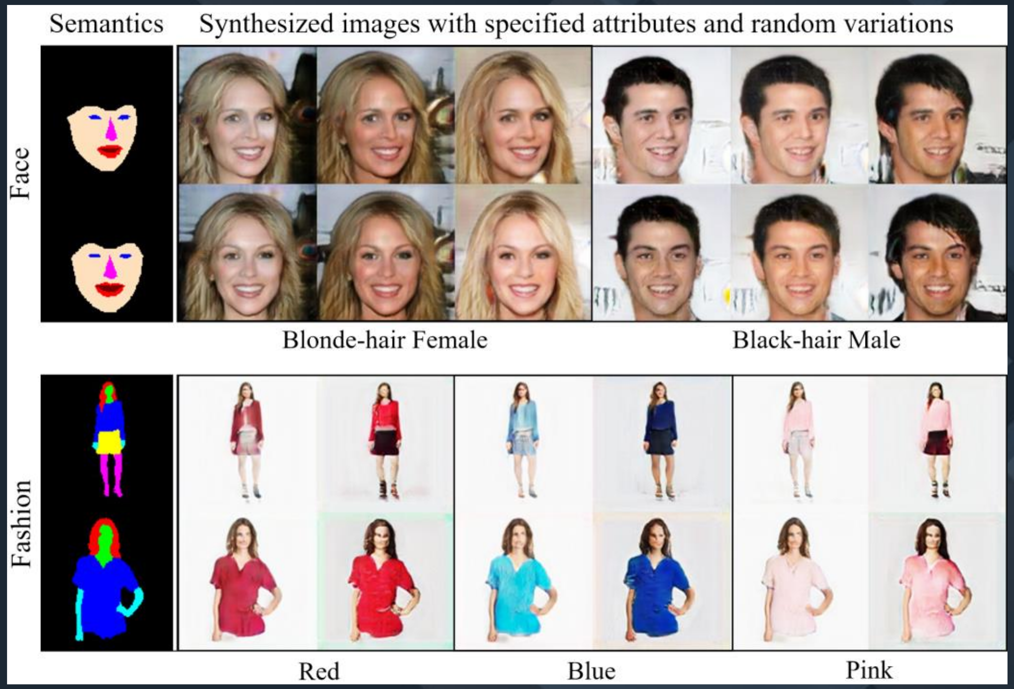
Problem Definition

Spatially Constrained Image Synthesis

- **Goal:**
  - Add spatial constraints to the image synthesis task.
  - Decouple the image synthesis task into three dimensions (i.e., spatial, attribute and latent dimensions), control the spatial and attribute-level contents, and randomize the other unregulated contents.
  - Train a neural network G to synthesize face and fashion images from semantic segmentations.
- **Motivation:**
  - Face and fashion synthesis are inherently one-to-many mappings from semantic segmentations to real images.
  - Existing GAN methods lack spatial constraints, thus not explicitly controllable in spatial configuration.
- **Mathematically:**
  - Our goal can be described as finding the mapping:

$$G(z, c, s) \rightarrow y$$

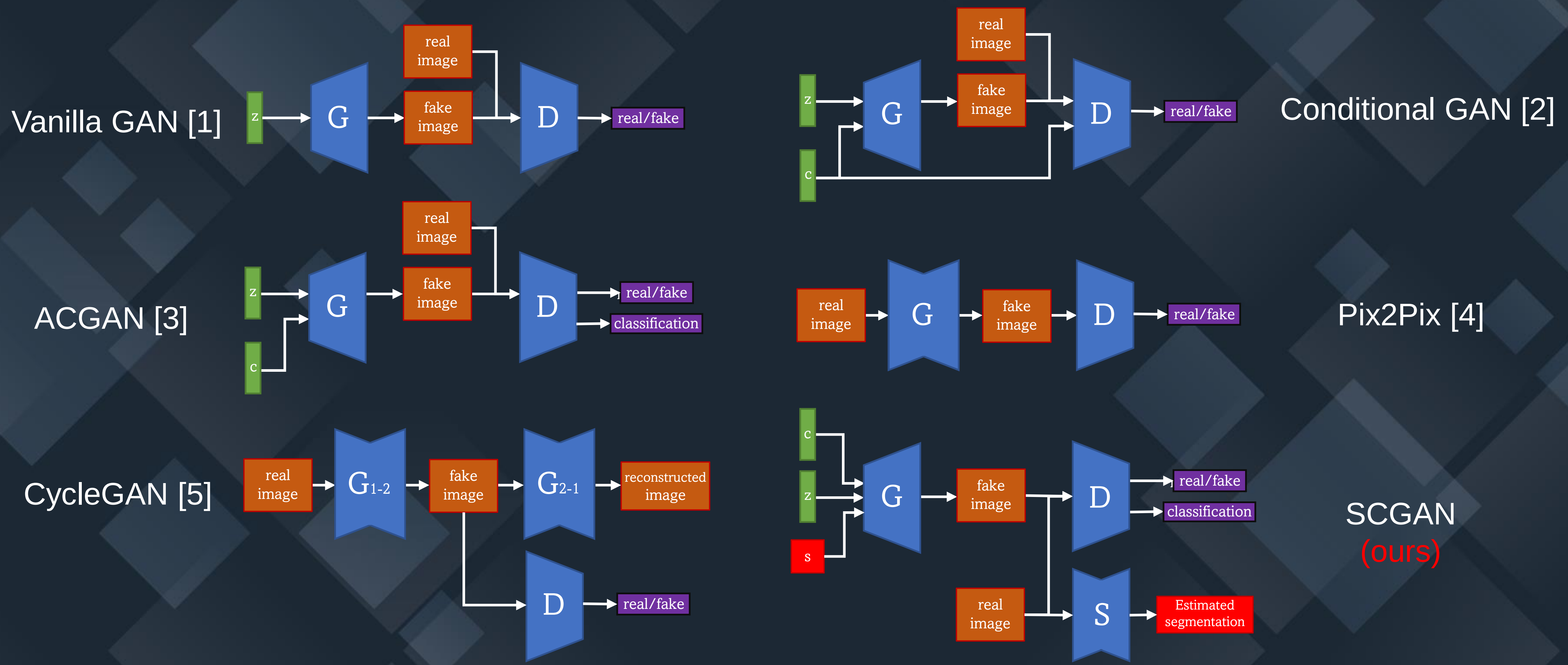
where *G* is the generative function, *z* is the latent vector and *y* is the conditionally generated image which complies with target attribute *c* and target semantic segmentation *s*.



Spatially Constrained Image Synthesis

Previous GAN Models

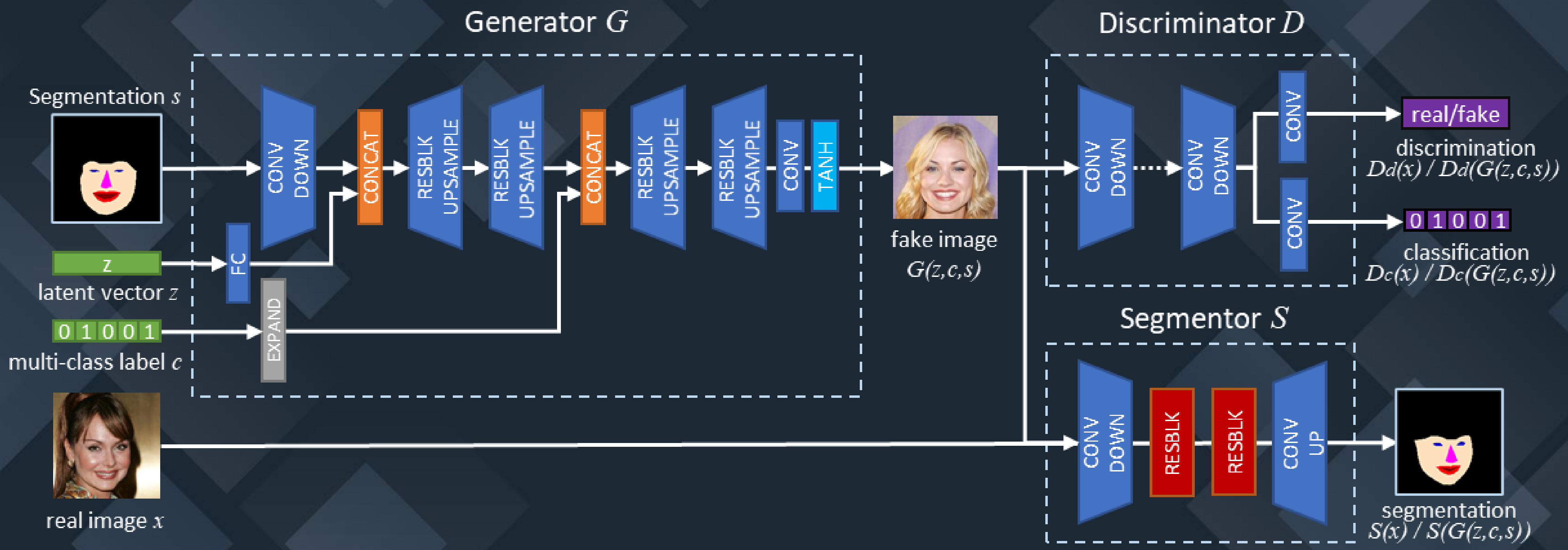
And Our Proposed Solution



[1] Goodfellow *et al.*, Generative adversarial nets. In *NeurIPS*, 2014.
[3] Odena *et al.*, Conditional image synthesis with auxiliary classifier gans. In *ICML*, 2017.
[5] Zhu *et al.*, Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.
[2] Mirza *et al.*, Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.
[3] Isola *et al.*, Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

SCGAN Framework

Spatially Constrained GAN Overview



Generator Network *G*:

- Synthesize the fake image from segmentation, latent vector and attribute label step-by-step.

Segmentor Network *S*:

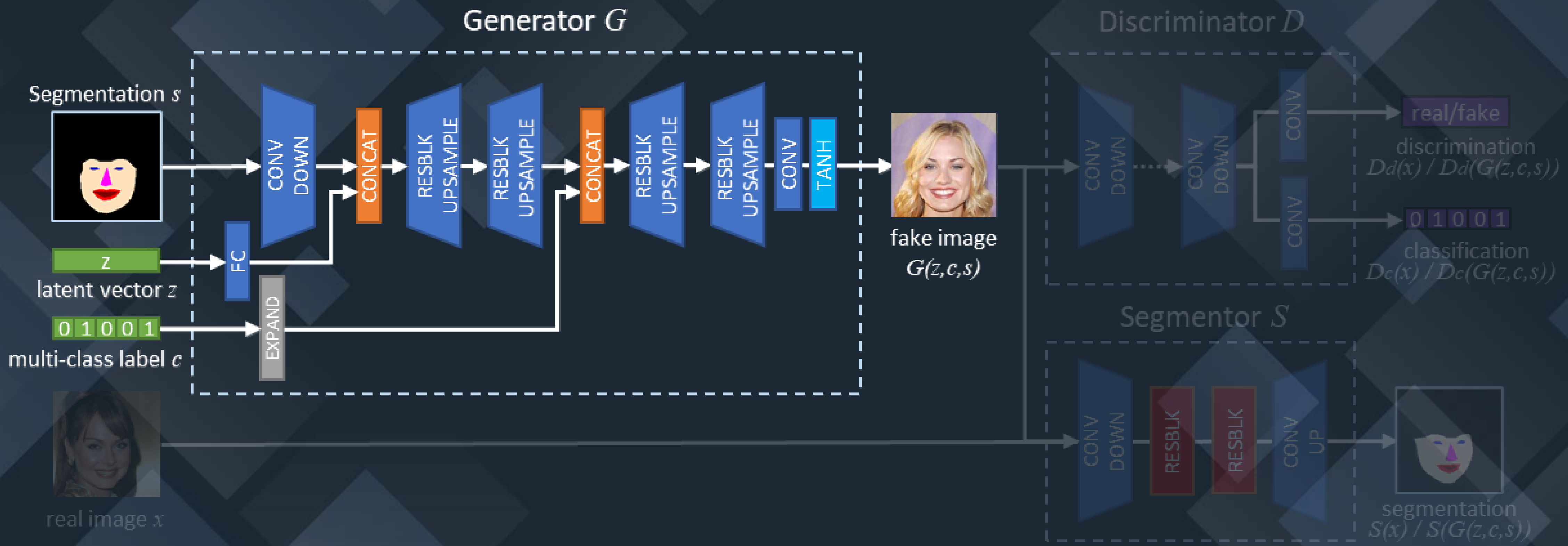
- Do semantic segmentation on both real and fake images.
- Provide G spatial constraints.

Discriminator Network *D*:

- Distinguish between real and fake images.
- Classify the images into attribute classes via an embedded auxiliary classifier.

SCGAN Framework

Generator Network



Goal:

- Learn the target mapping function:

$$G\left(z, c, s\right) \rightarrow y$$

Inputs:

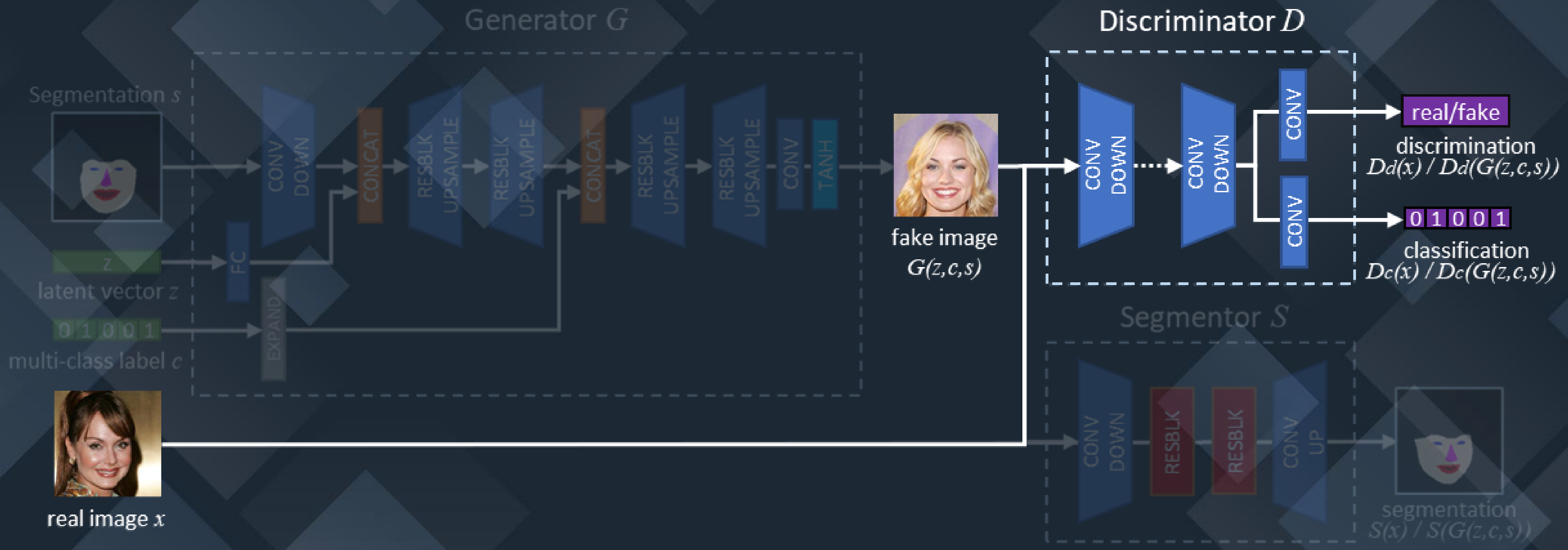
- Target segmentation *s*
- Random latent vector *z*
- Attribute-level class label *c*

Output:

- Synthesized image *G(z,c,s)*

SCGAN Framework

Discriminator Network



Adversarial Loss: $\mathcal{L}_{adv} = L_{adv}^{real} + L_{adv}^{fake} + L_{gp},$

$$\mathcal{L}_{adv} = \mathbb{E}_x\left[D_d(x)\right] + \mathbb{E}_{z,c,s}\left[D_d\left(G\left(z,c,s\right)\right)\right] + \lambda_{gp}\mathbb{E}_{\hat{x}}\left[\left(\|\nabla_{\hat{x}} D_d(\hat{x})\|_2 - 1\right)^2\right],$$

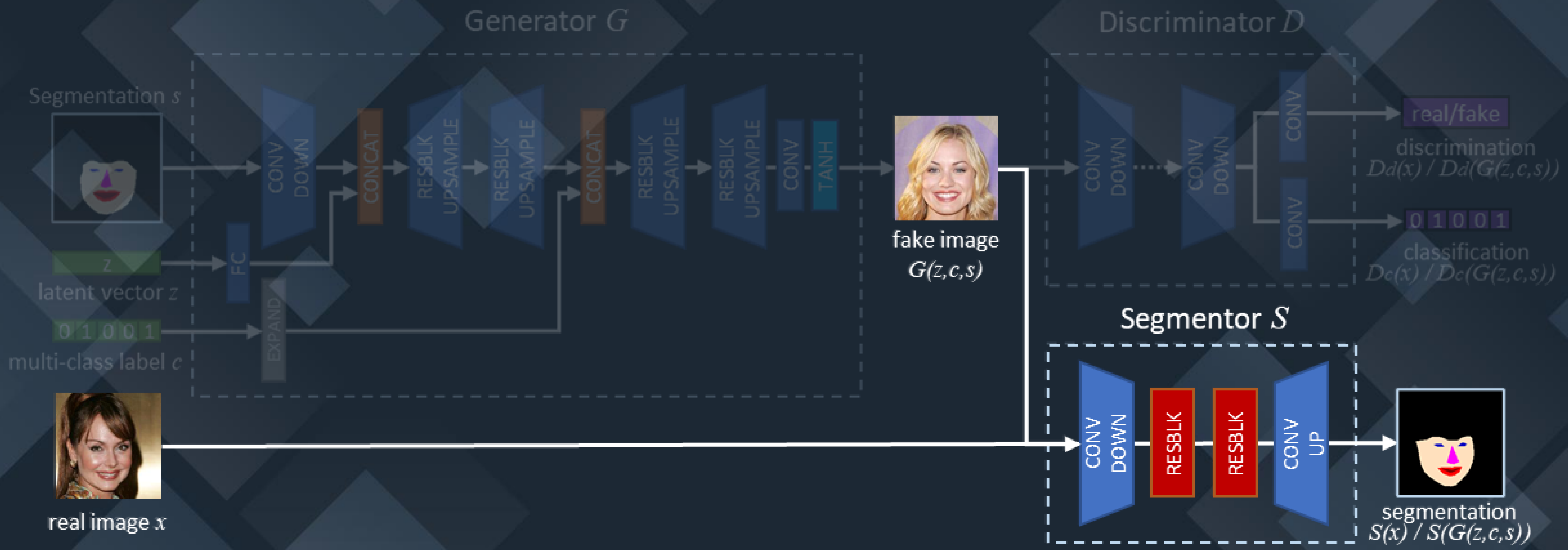
Classification Loss:

$$\mathcal{L}_{cls}^{real} = \mathbb{E}_{x,c}\left[A_c(c, D_c(x))\right],$$

$$\mathcal{L}_{cls}^{fake} = \mathbb{E}_{z,c,s}\left[A_c(c, D_c(G(z,c,s)))\right],$$

SCGAN Framework

Segmentor Network



Segmentation Loss:

$$\mathcal{L}_{seg}^{real} = \mathbb{E}_{x,s}[A_s(s, S(x)],$$

$$\mathcal{L}_{seg}^{fake} = \mathbb{E}_{z,c,s}\left[A_s(s, S(G(z,c,s)))\right],$$

Pixel-wise Cross Entropy:

$$A_s(a,b) = -\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{n_s} a_{i,j,k}\log b_{i,j,k},$$

Provide spatial constraints to the generator

Overall Objectives

Training SCGAN

Overall objectives to optimize SCGAN:

$$\mathcal{L}_S = \mathcal{L}_{seg}^{real},$$

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{real},$$

$$\mathcal{L}_G = \mathcal{L}_{adv}^{fake} + \lambda_{cls}\mathcal{L}_{cls}^{fake} + \lambda_{seg}\mathcal{L}_{seg}^{fake},$$

$\mathcal{L}_S$: Segmentor Loss.
$\mathcal{L}_D$: Discriminator Loss.
$\mathcal{L}_G$: Generator Loss.

$\mathcal{L}_{adv}$: Adversarial Loss Term.
$\mathcal{L}_{cls}$: Classification Loss Term.
$\mathcal{L}_{seg}$: Segmentation Loss Term.

λ_{cls} and λ_{seg} are hyper-parameters that control the relative importance of loss terms.

Datasets

Face attribute dataset:
- 10,177 identities,
- 202,599 number of face images, and
- 5 landmark locations,
- 40 binary attributes annotations.

https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

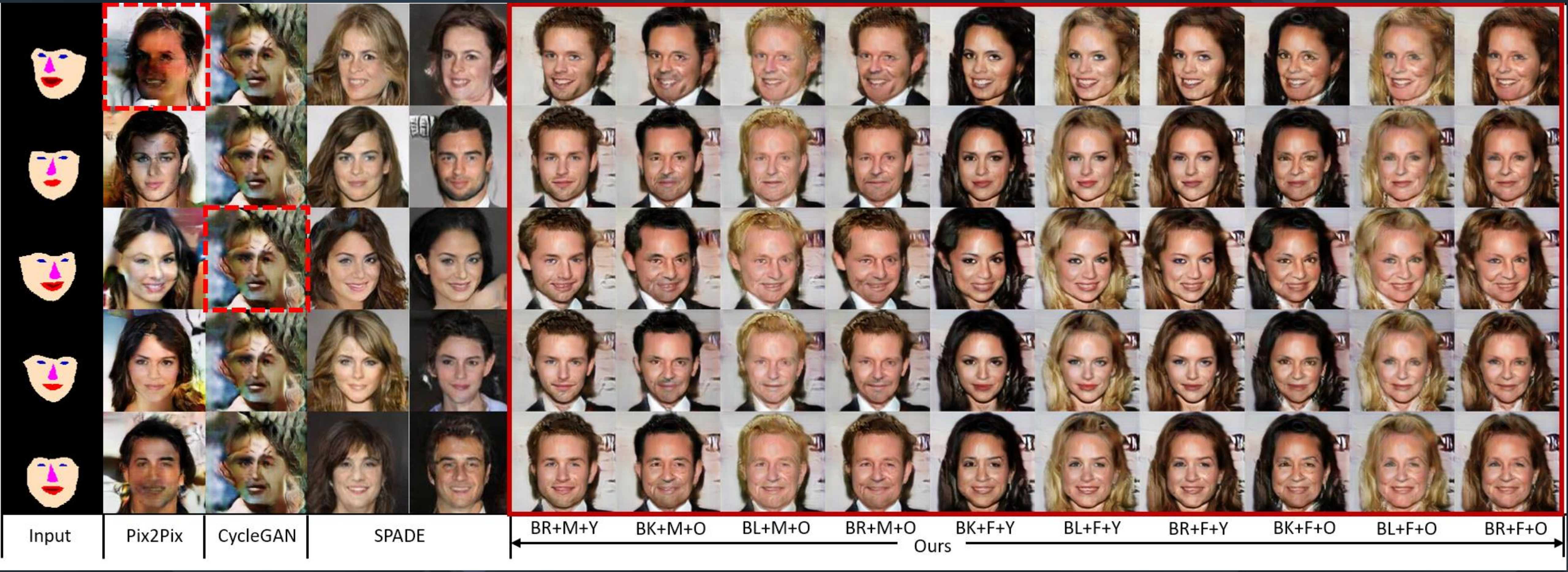
A large-scale clothes database
- 50 categories, 1,000 descriptive attributes

Fashion synthesis subset:
- 78,979 images,
- Captions, and segmentations

https://mmlab.ie.cuhk.edu.hk/projects/DeepFashion/FashionSynthesis.html

Experiment

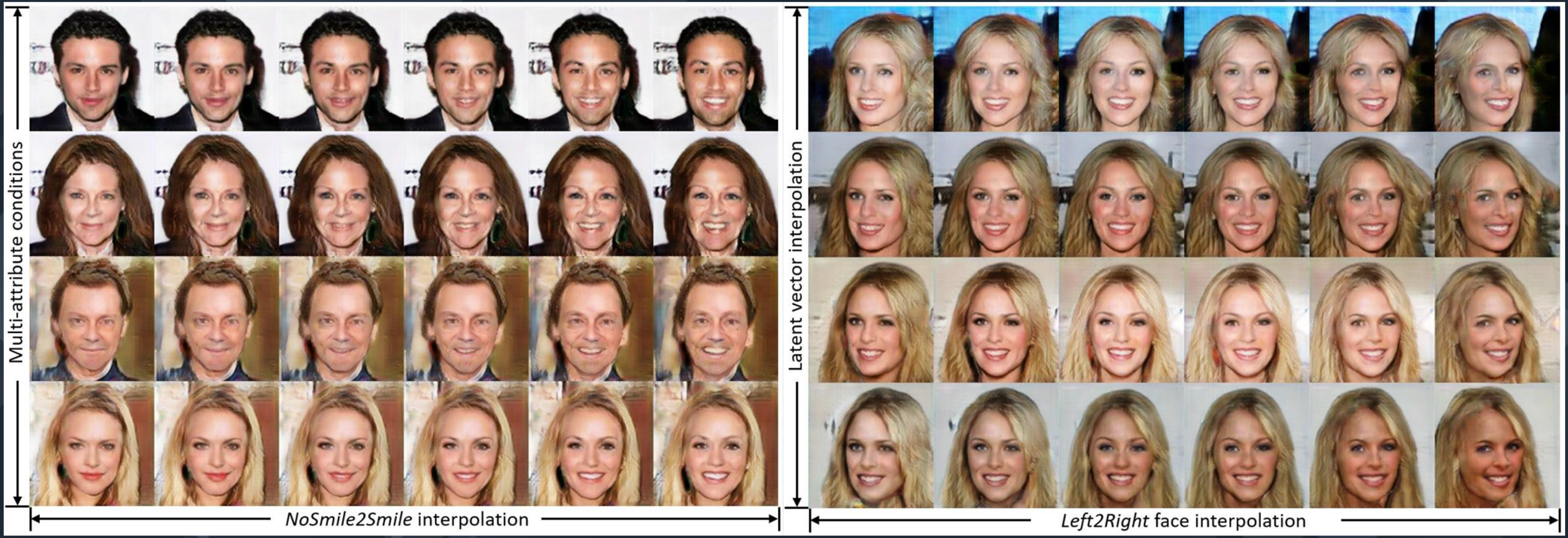
Comparison on CelebA Dataset



Low-resolution Side-view Faces

Experiment

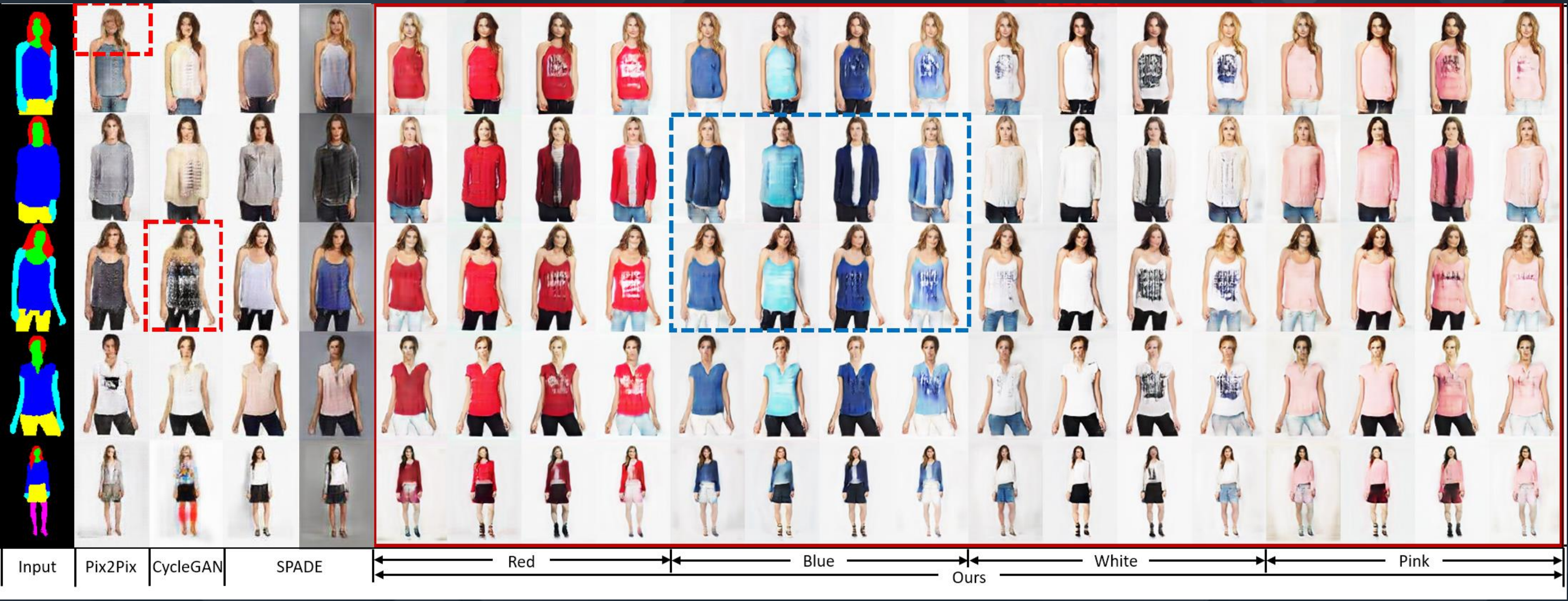
Face Interpolation



***NoSmile2Smile* Interpolation** ***Left2Right* Interpolation**

Experiment

Comparison on DeepFashion Dataset



Experiment

Quantitative Evaluation

- **Evaluation:**
  - **Visual quality**
  - **Spatial correctness**
- **Metrics:**
  - Frechet Inception Distance (FID) [1]
  - Pixel Accuracy
  - Mean IoU (intersection over union)

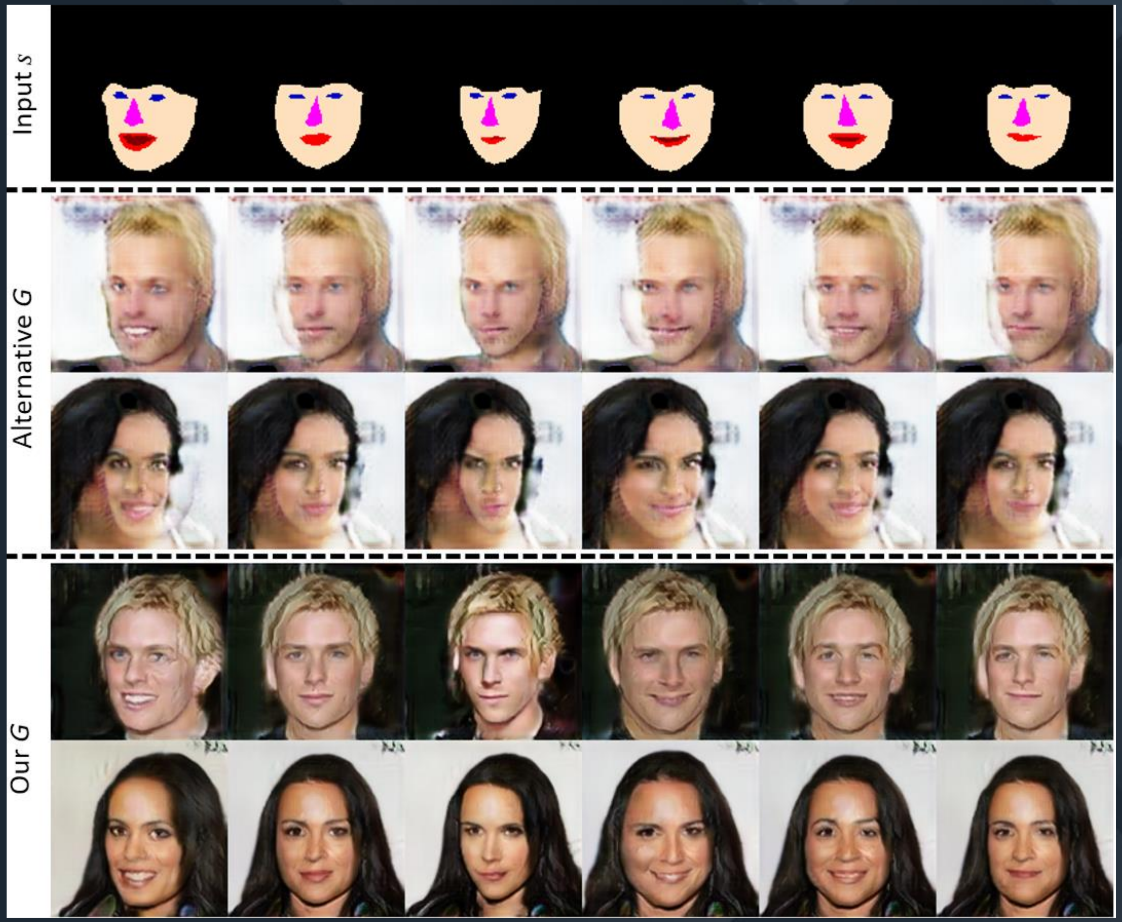
Methods	CelebA			DeepFashion		
	FID	mIoU	pAcc	FID	mIoU	pAcc
CycleGAN [2]	N/A	N/A	N/A	30.1	63.26	82.21
Pix2Pix [3]	20.4	78.71	98.05	24.4	65.41	82.91
SPADE [4]	18.5	74.76	97.82	20.2	75.80	83.10
SCGAN	**10.2**	**79.11**	**98.95**	**19.8**	**77.20**	**83.23**

[1] Heusel *et al*., Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.
[2] Zhu *et al*., Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.
[3] Isola *et al*., Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.
[4] Park *et al*., Semantic image synthesis with spatially-adaptive normalization. In *CVPR*, 2019.

Experiment

Ablation Study of Generator Architecture

- **Our proposed architecture:**
  - Step-by-step generator *G.*
  - From coarse to fine synthesis.
- **Alternative architecture:**
  - Input all at once generator *G.*
- **Comparison:**
  - Better visual quality.
  - Sharper details
  - No foreground-background mismatch.

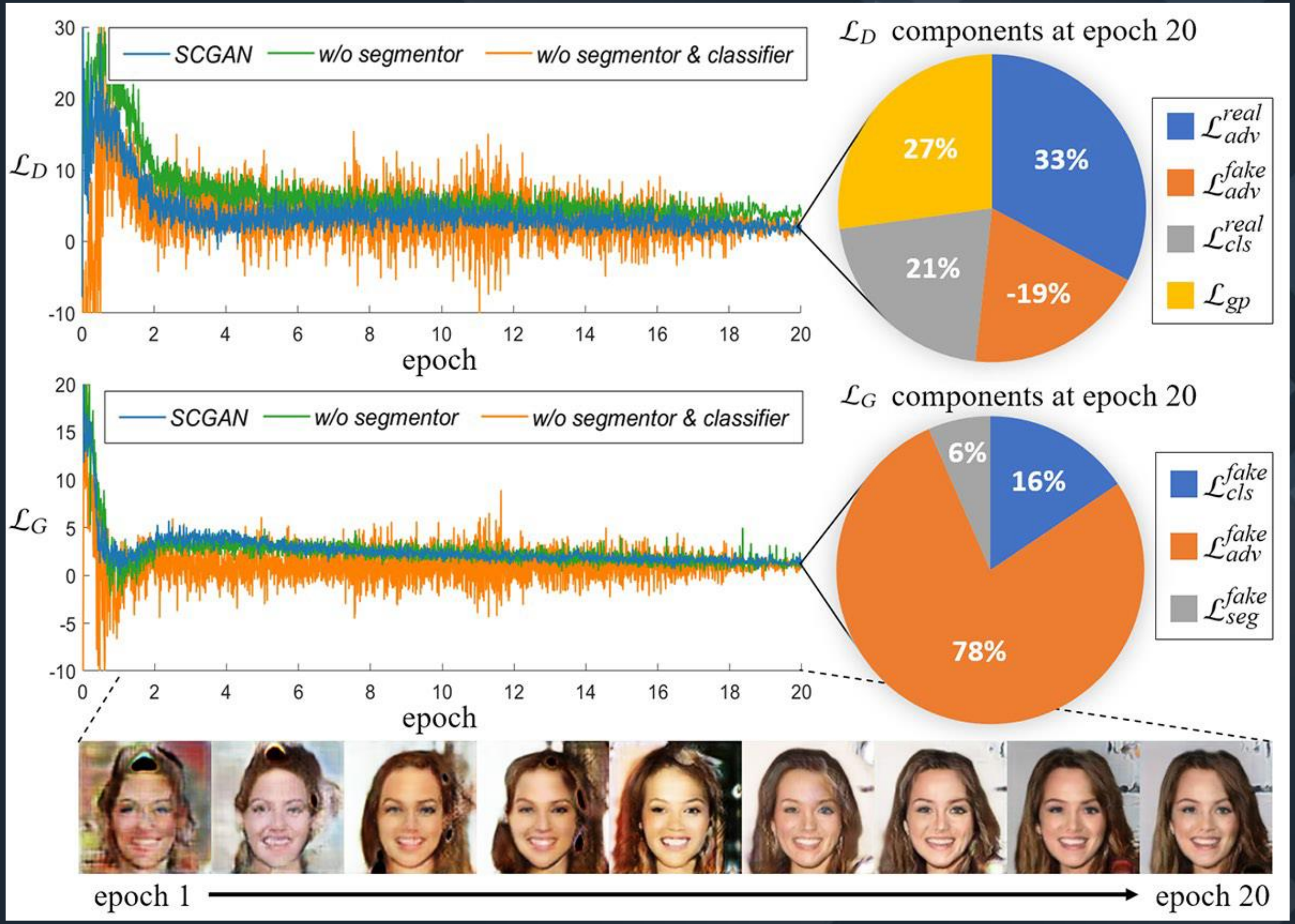


Experiment

Ablation Study of Model Convergence

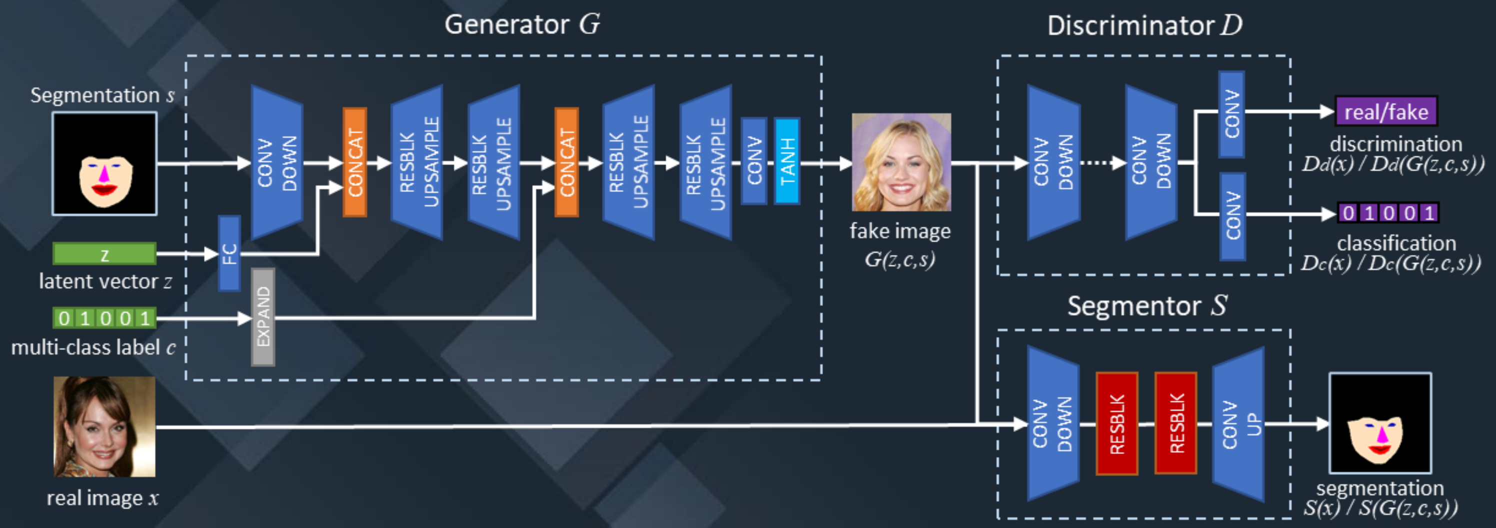
- Study of model convergence:
  - SCGAN.
  - w/o Segmentor.
  - w/o Segmentor & Classifier.

- Benefits of Segmentor *S:*
  - Stabilize training.
  - Faster convergence.
  - Lower loss when converged.
  - Better image quality.



The End.
Thank You!

Spatially Constrained GAN for Face and Fashion Synthesis



Scan QR Code

Code and more details available on our project website.

https://jackyjsy.github.io/SCGAN/