Conditional Image Synthesis by Generative Adversarial Modeling

Zhifei Zhang

EECS

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Conditional Image Synthesis by Generative Adversarial Modeling

Conditional Image Synthesis by **Generative Adversarial Modeling**

Conditions **→ GAN →** ***Synthesis***

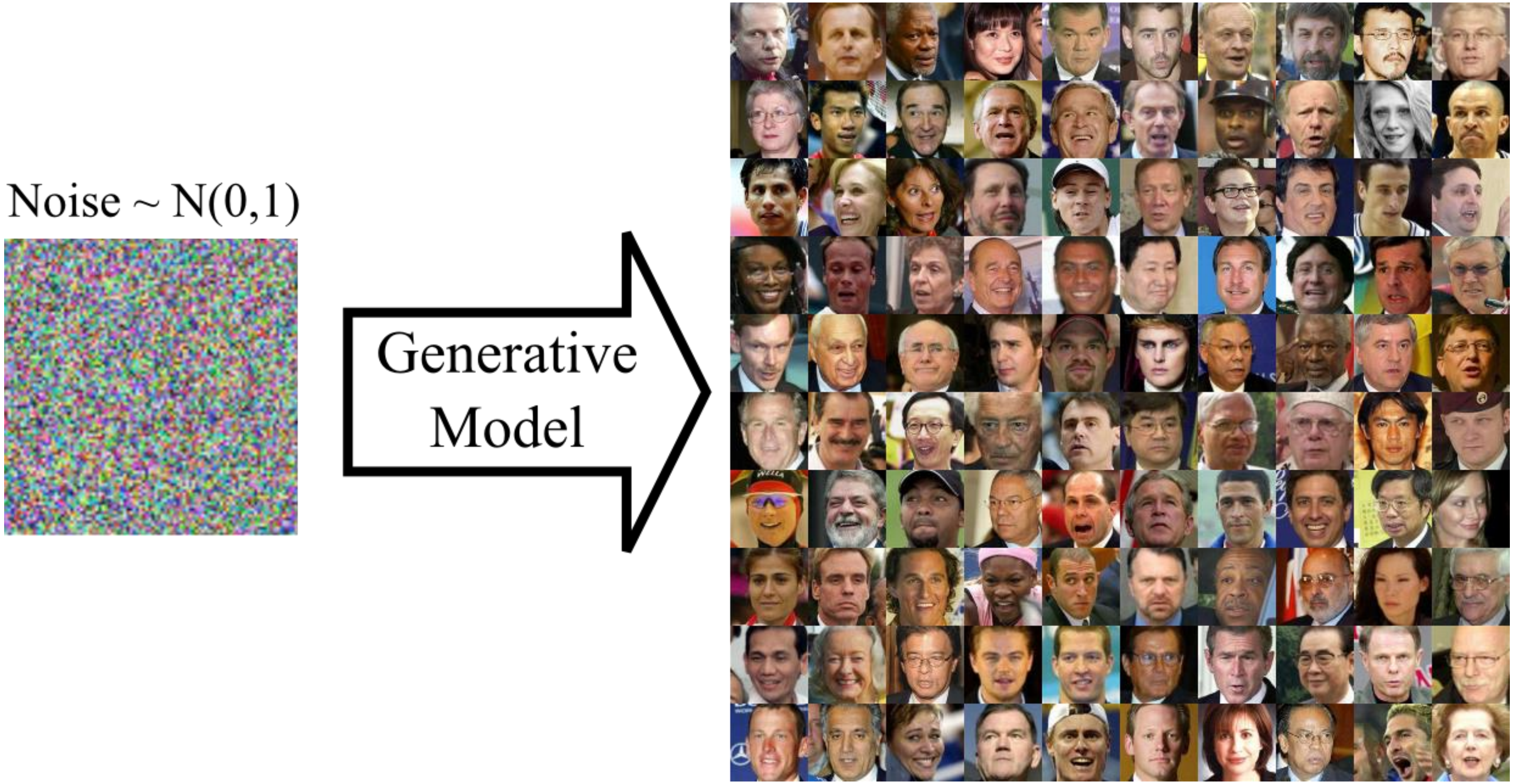
Instable and hard to train

Limited on empirical studies

Contents

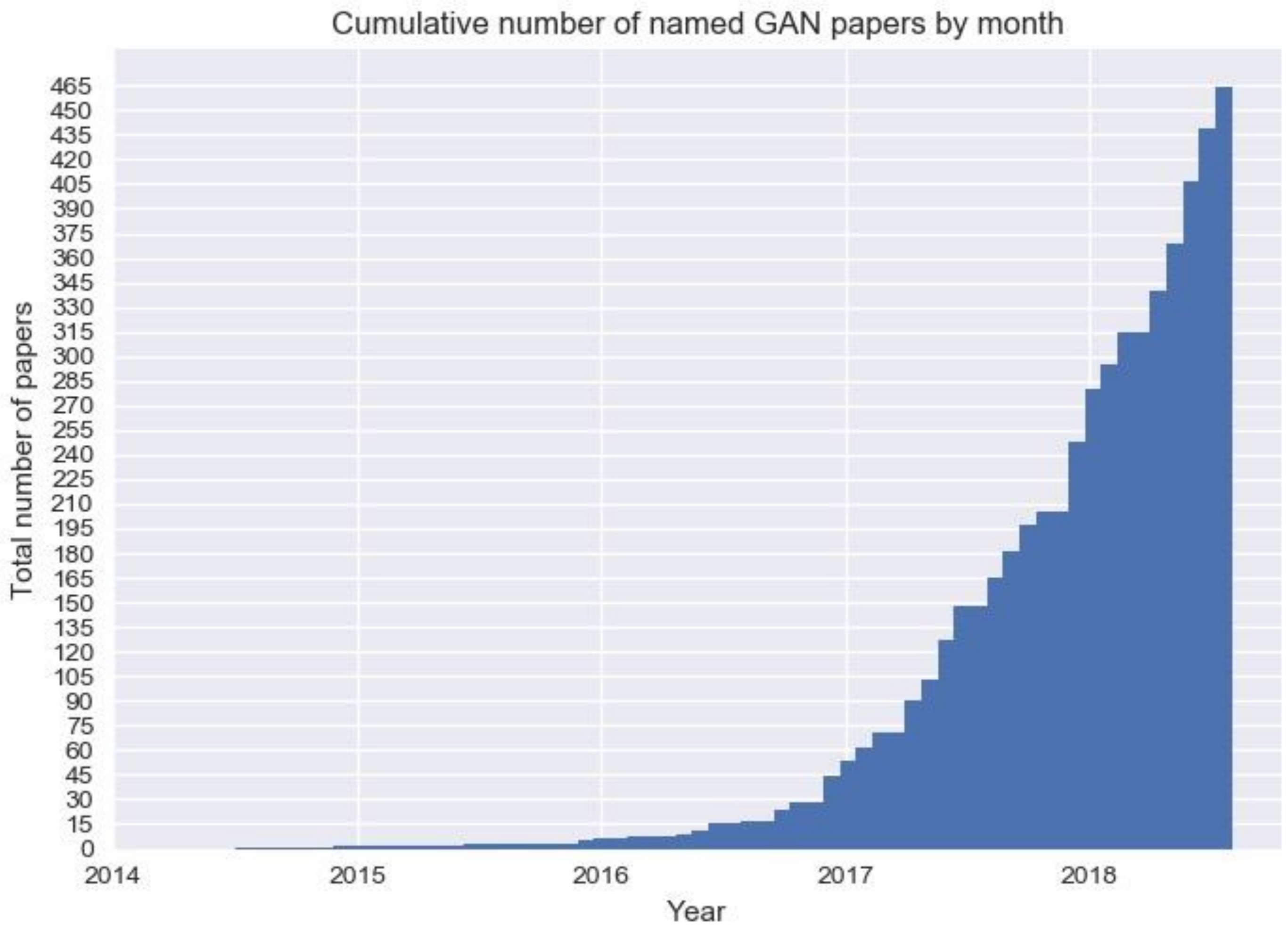
Contents

Generative Adversarial Network (GAN)



Generative Adversarial Network (GAN)

Advantage of GAN: Achieving more perceptually/visually realistic images.



https://github.com/hindupuravinash/the-gan-zoo

Generative Adversarial Network (GAN)

Advantage of GAN --- An example of super-resolution

Original Image

Down scaling

LR

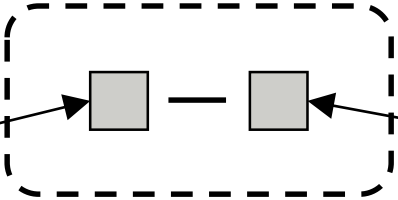
Super-resolution

L_2 loss

GAN loss

Generative Adversarial Network (GAN)

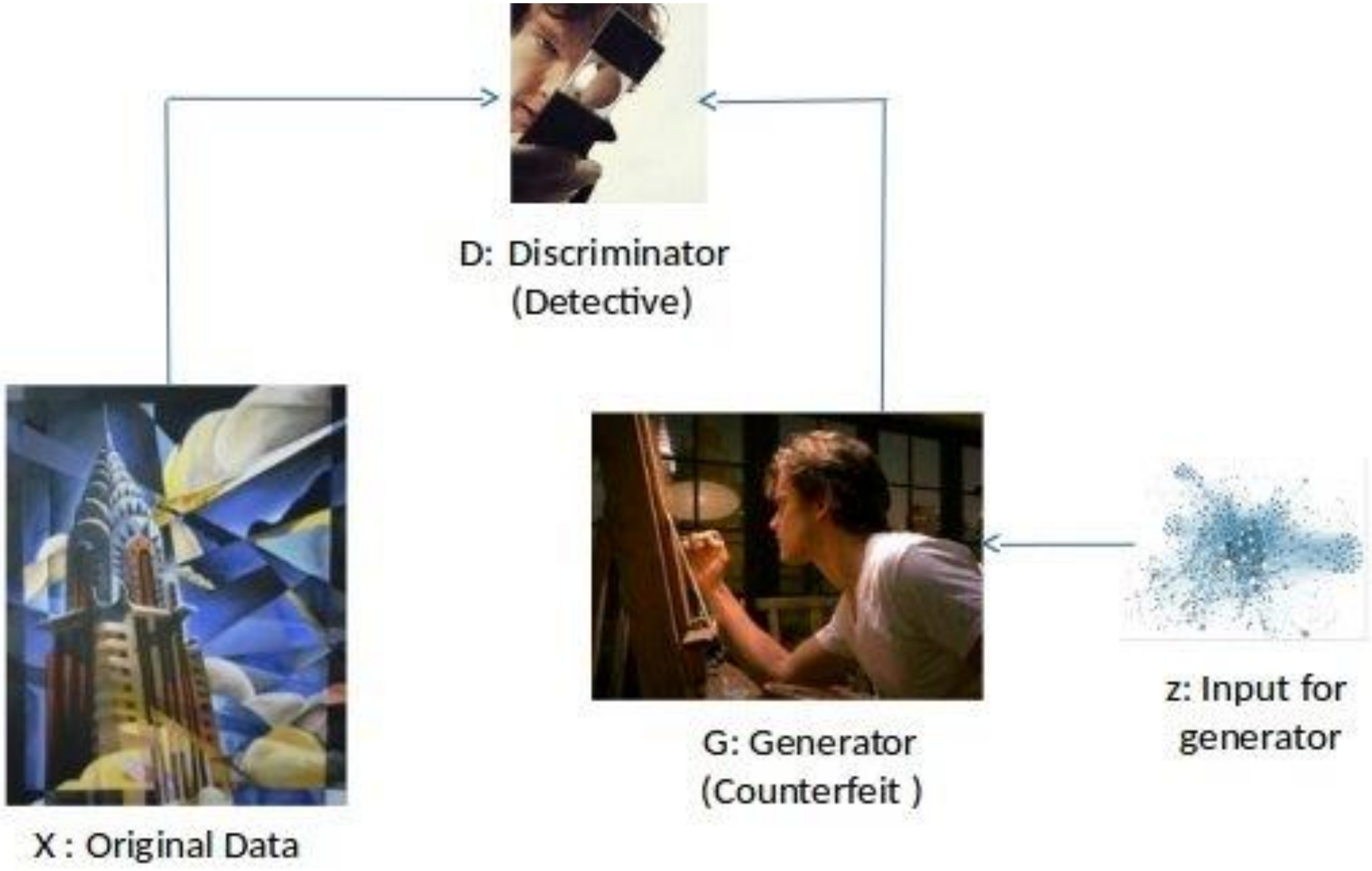
Traditional loss for image synthesis ---
pixel-wise distance



GAN loss --- perceptual distance

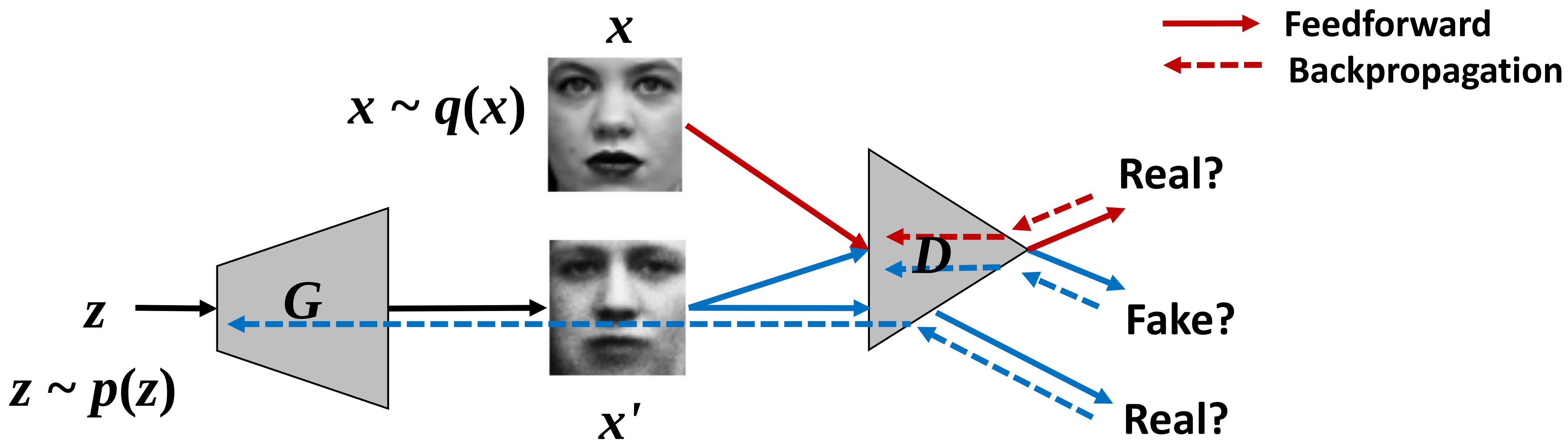
Perceptual distance is more sensitive to edges and texture

Generative Adversarial Network (GAN)



[G. Ramachandra, 2017]

Generative Adversarial Network (GAN)



The objective function:

$$\min_G \max_D \; \mathbb{E}_{x\sim q(x)}[\log(D(x))] + \mathbb{E}_{z\sim p(z)}[\log(1 - D(G(z)))]$$

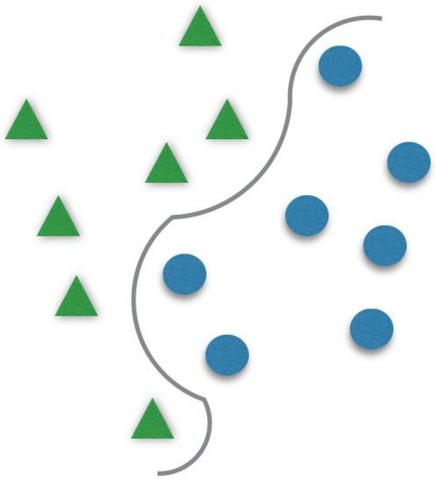
Contents

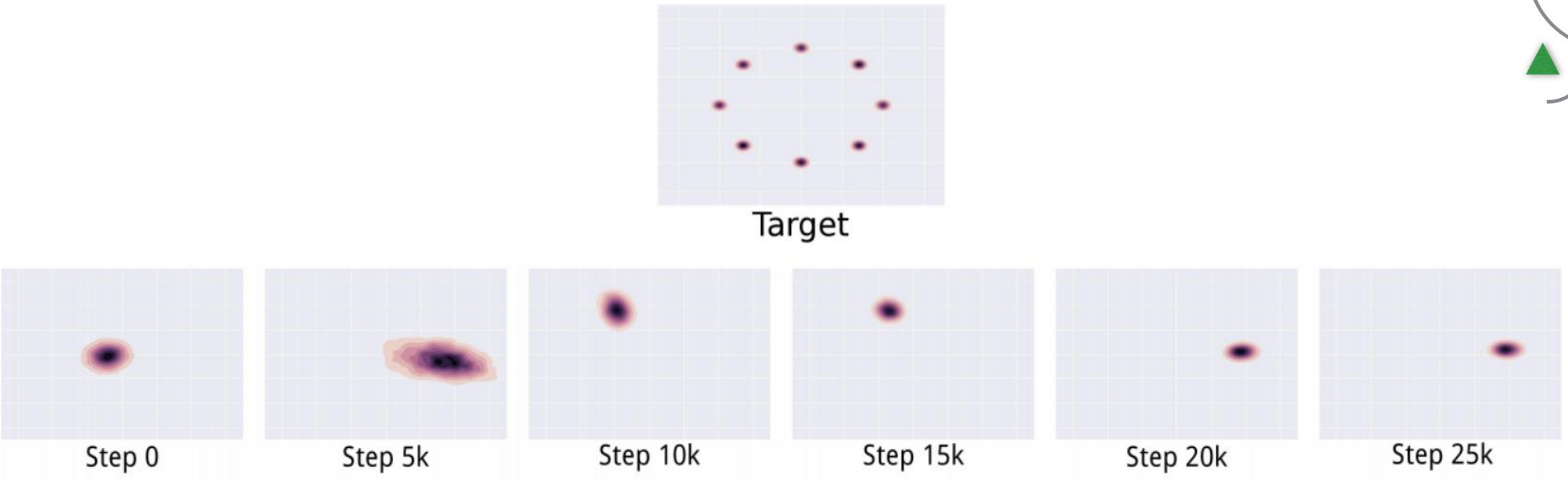
1. Introduction to generative adversarial network (GAN)
2. **Instability of GAN and Stabilization by Conditional GAN**
3. Image Synthesis by Conditional GAN --- Face Aging
4. Further Stabilize Conditional GAN --- Decoupled Learning
5. Reference-Conditioned Super-Resolution

Instability of GAN

- Mode missing
- Gradient vanishing

Perfect Discriminator causes gradient vanishing





[L. Metz et al., 2017]

Instability of GAN

The objective function

$$\min_G \max_D \ \mathbb{E}_{x\sim q(x)}[\log(D(x))] + \mathbb{E}_{z\sim p(z)}[\log(1 - D(G(z)))]$$

Since, $\mathbb{E}_{x\sim q(x)}[p(x)] = \int_x \ q(x)p(x)\, dx$

$$\min_G \max_D \ \int_x \ q(x)\log(D(x))\, dx + \int_z \ p(z)\log(1 - D(G(z)))\, dz$$

Instability of GAN

Alternative update of G and D

How does it come? (click to see)

Fix G,

$$\min_G \max_D \int_x q(x)\log(D(x))\,dx + \int_z p(z)\log(1-D(G(z)))\,dz$$

$$= \max_D \int_x q(x)\log\big(D(x)\big) + p_g(x)\log\big(1-D(x)\big)\,dx$$

$$\boldsymbol{D^*(x) = \frac{q(x)}{q(x)+p_g(x)}}$$

Fix D^*,

$$= \min_G \int_x p_g(x)\log\left(1-\frac{q(x)}{q(x)+p_g(x)}\right)dx$$

$$= \min_G \int_x p_g(x)\log\left(\frac{p_g(x)}{q(x)+p_g(x)}\right)dx$$

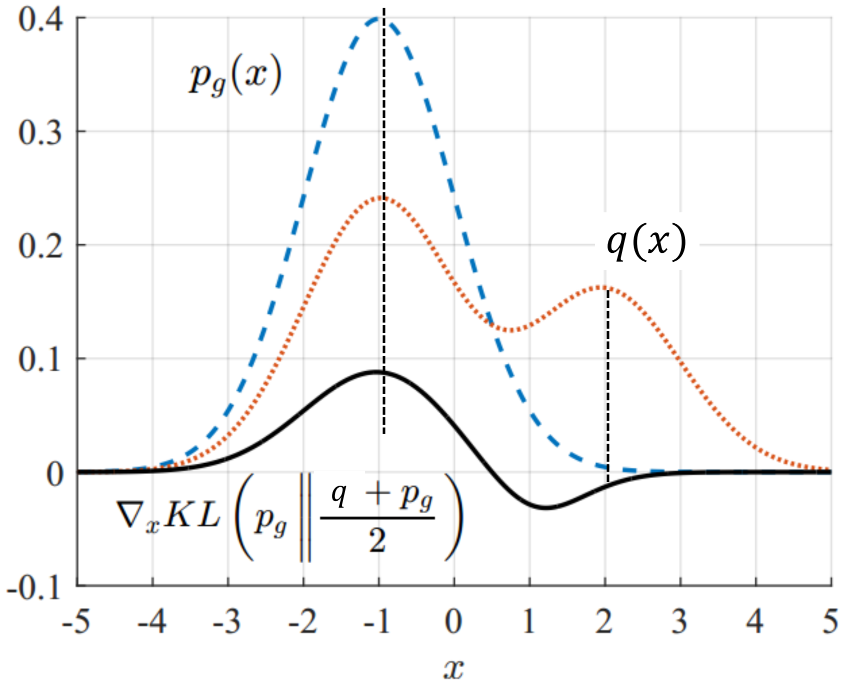
$$= \min_G D_{KL}(p_g||q+p_g)$$

$$= \min_G D_{KL}(p_g||\frac{q+p_g}{2}) - 2\log 2$$

Bothe mode missing and gradient vanishing are caused by KL-divergence

Instability of GAN

KL-divergence is an unsymmetrical measurement



$p_g(x) > q(x)$: Unrealistic samples
$q(x) > p_g(x)$: Mode missing

Reasons of mode missing:

- **Punish more on unrealistic samples**
- **Punish less on mode missing**

Reason of gradient vanishing:

- **KL-divergence is constant (i.e., zero) if two distribution are not overlapped. Then, the gradient will be zero.**

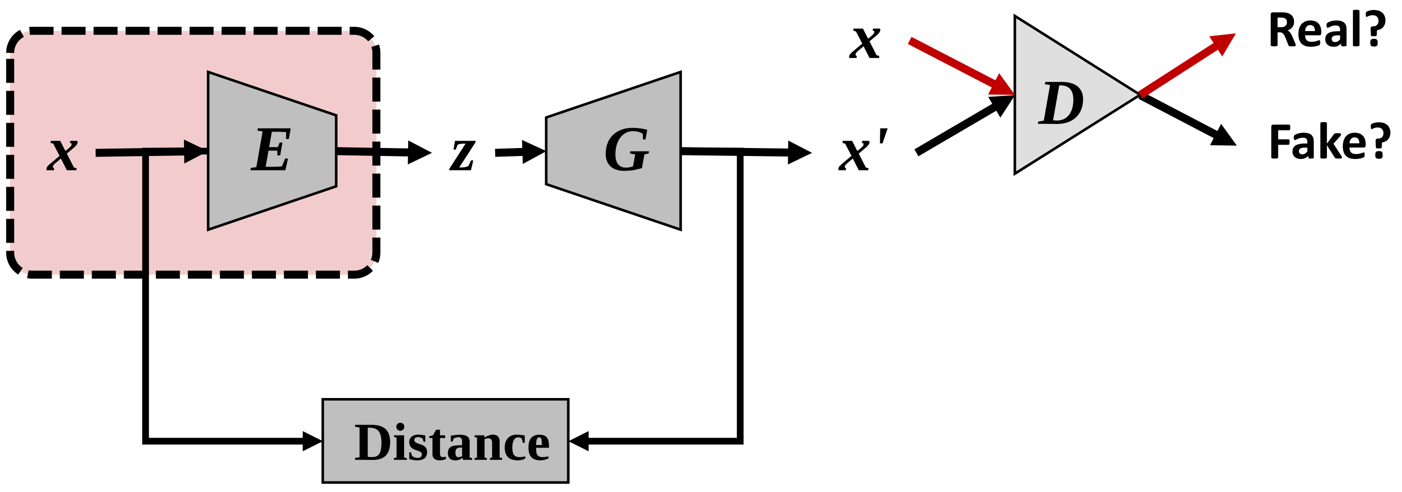
$$\min_G \; D_{KL}(p_g || \frac{q+p_g}{2}) - 2\log 2$$
$$= \min_G \int_x p_g(x) \log\left(\frac{p_g(x)}{q(x)+p_g(x)}\right) dx$$

Conditional GAN

Conditional GAN can relax the instabilities

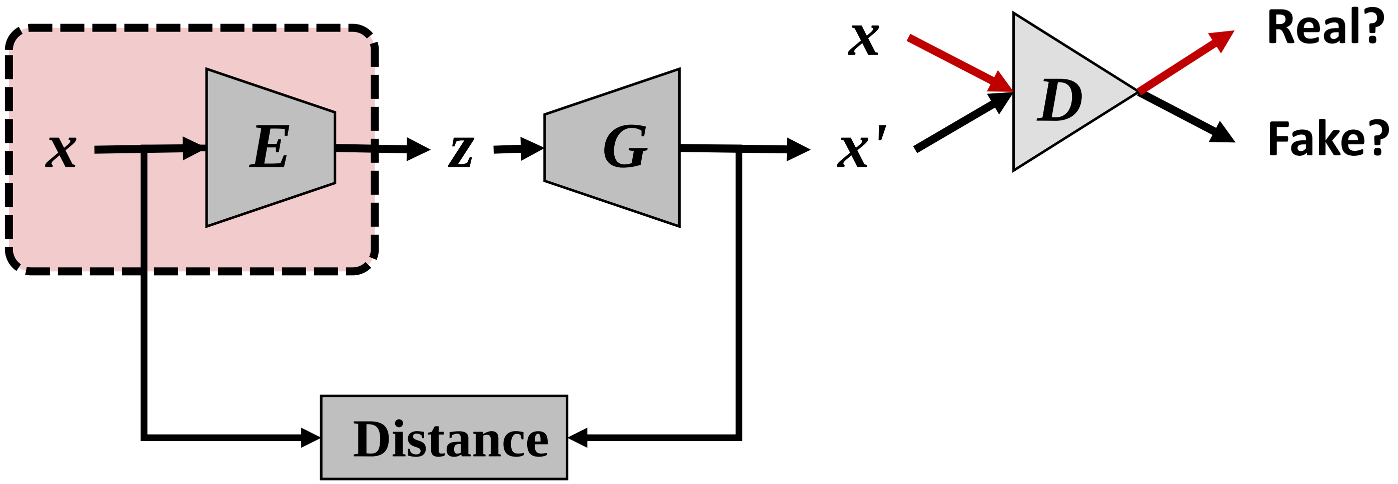
In GAN, $D(x|p_g) = D(x|q)$, which is indirect matching.

Ideally, $p_g(x) = q(x)$, which is direct matching, avoiding mode missing.



An extra loss based on Euclidian distance relaxes gradient vanishing

Conditional GAN



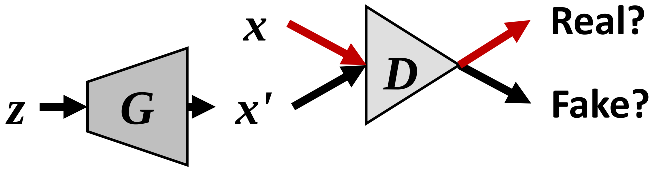
GAN: $\mathbb{E}_{x \sim p_x}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$

Conditional GAN: $\mathbb{E}_{x \sim p_x}[\log (D(x)(1 - D(\mathcal{H}(x))) + \lambda \mathcal{L}(x, \mathcal{H}(x))]$

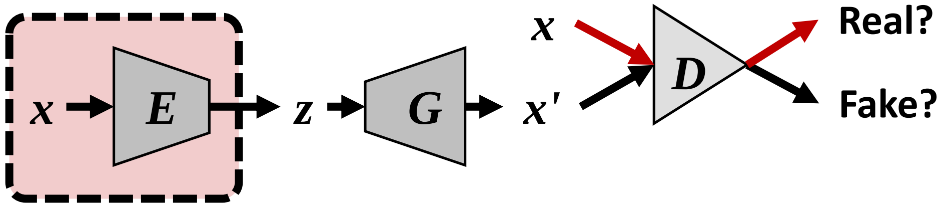
where λ balances the effect of reconstruction error, and $\mathcal{H}(x) = G(E(x))$.

Conditional GAN --- Summary

GAN
(unstable in training)



Conditional GAN
(more stable in practice)



- Theatrical analysis of the instability in the GAN
- How conditional GAN stabilizes the training

Contents

Image Synthesis by Conditional GAN

Becomes a common framework for image synthesis

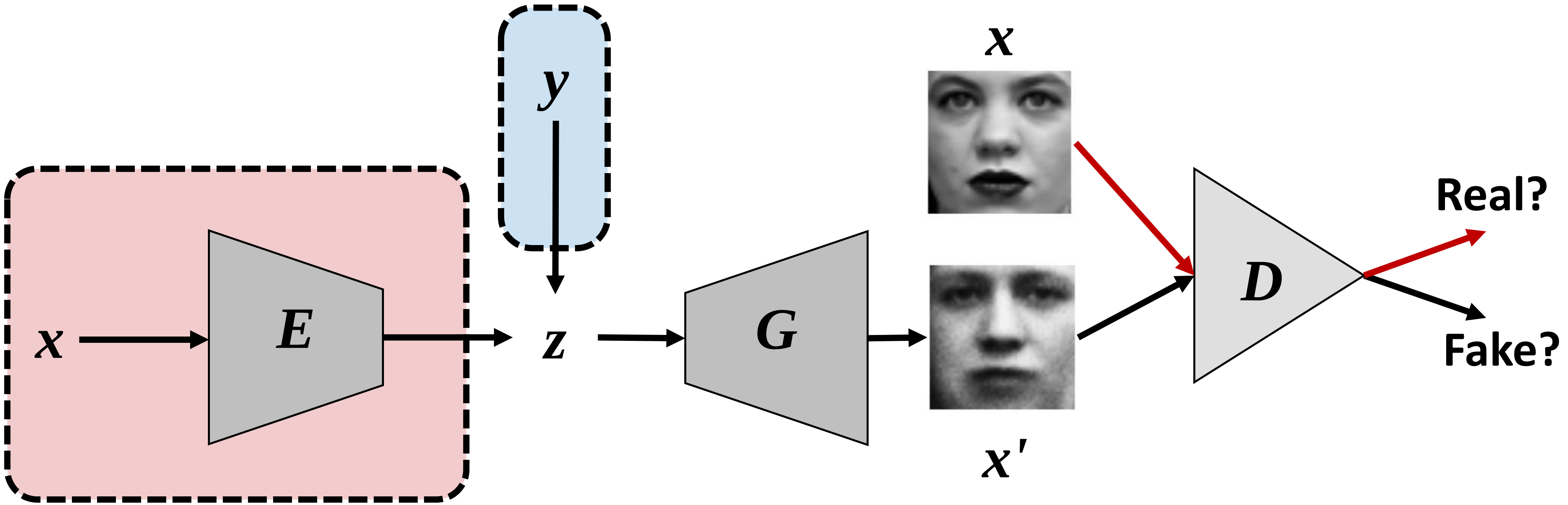
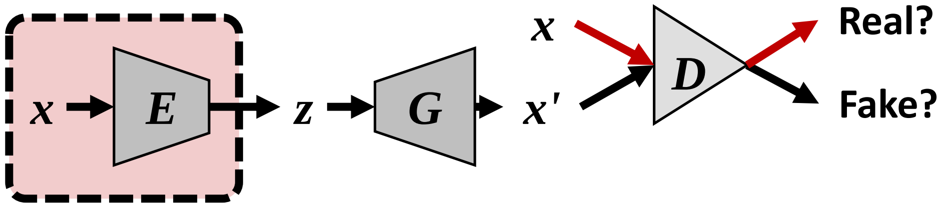


Image Synthesis by Conditional GAN

Image condition [Isola et al., CVPR 2017]



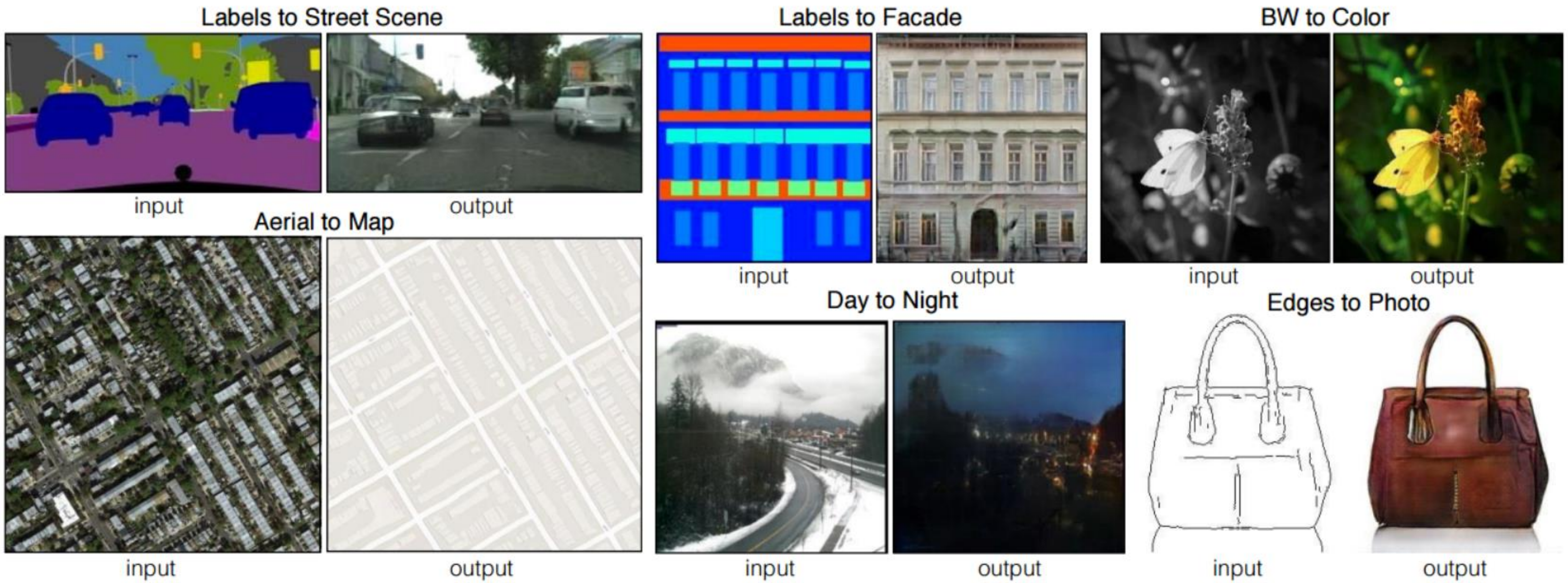
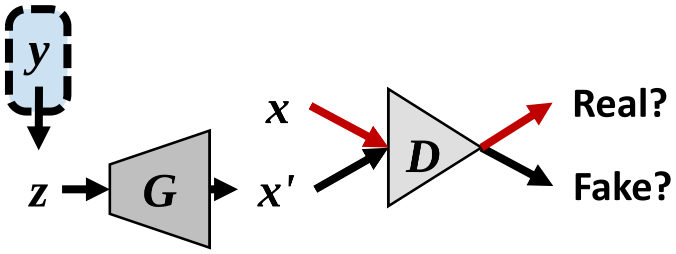


Image Synthesis by Conditional GAN

Label condition [Reed et al., CVPR 2016]



this small bird has a pink breast and crown, and black primaries and secondaries.

this magnificent fellow is almost all black with a red crest, and white cheek patch.

the flower has petals that are bright pinkish purple with white stigma

this white and yellow flower have thin white petals and a round yellow stamen

Face Aging by Conditional GAN

Regression/Rejuvenation Progression/Aging

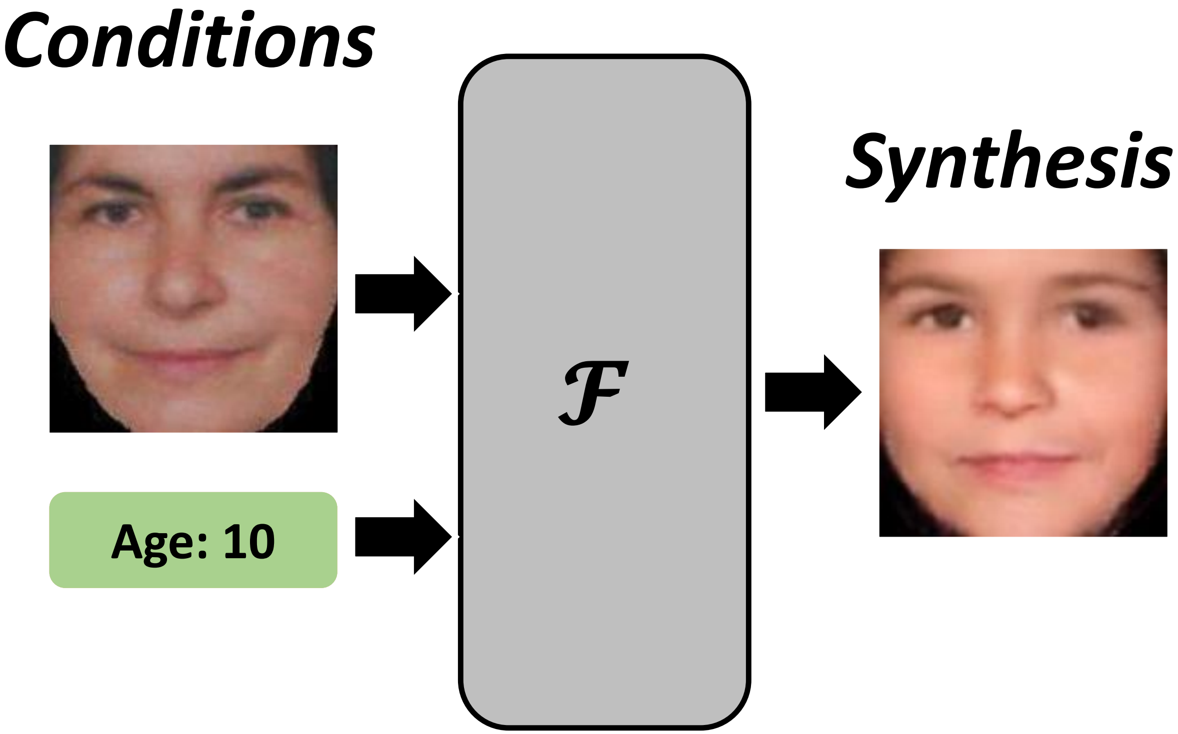
Given face

10 20 35 years old 40 60

Image synthesis conditioned on age and identity

Face Aging by Conditional GAN

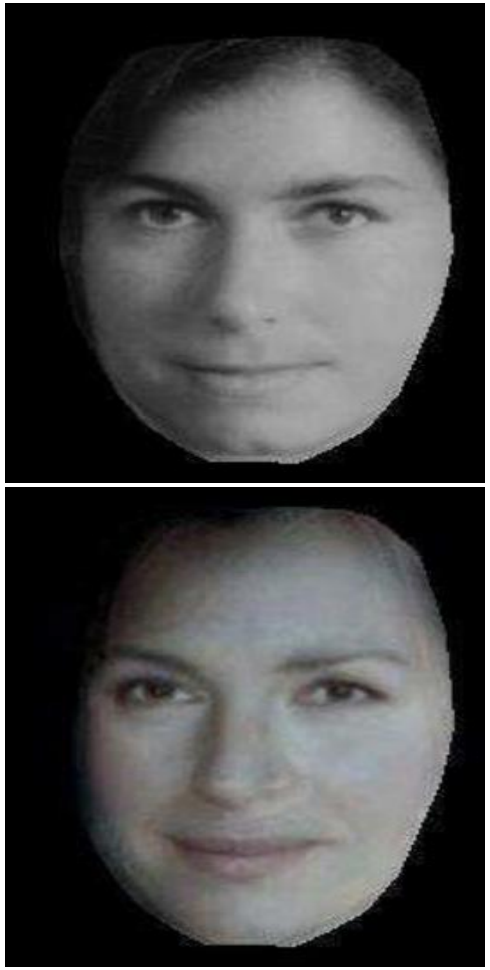
[Zhang et al., CVPR 2017]



Previous Works

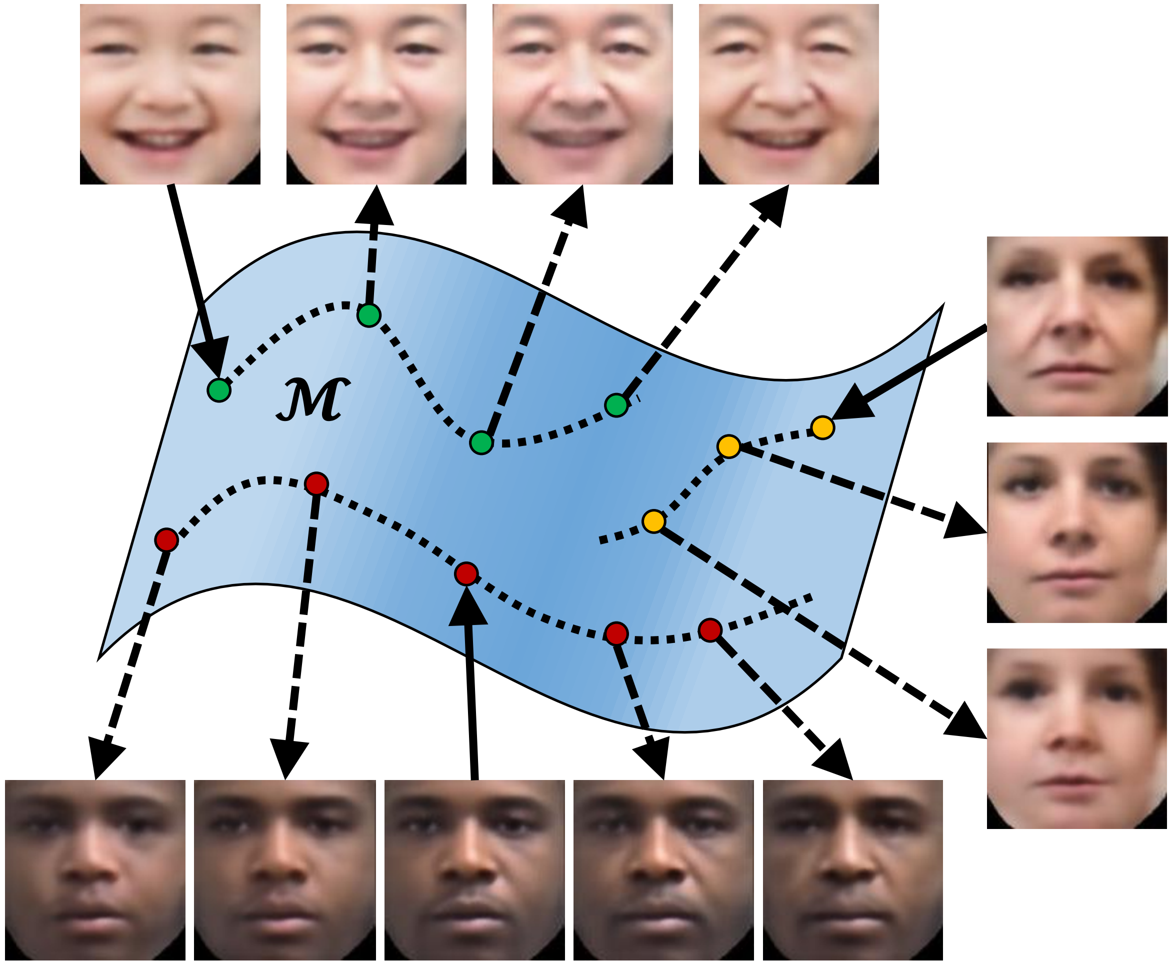
Kemelmacher, et al., CVPR2014

Shu, et al., ICCV2015

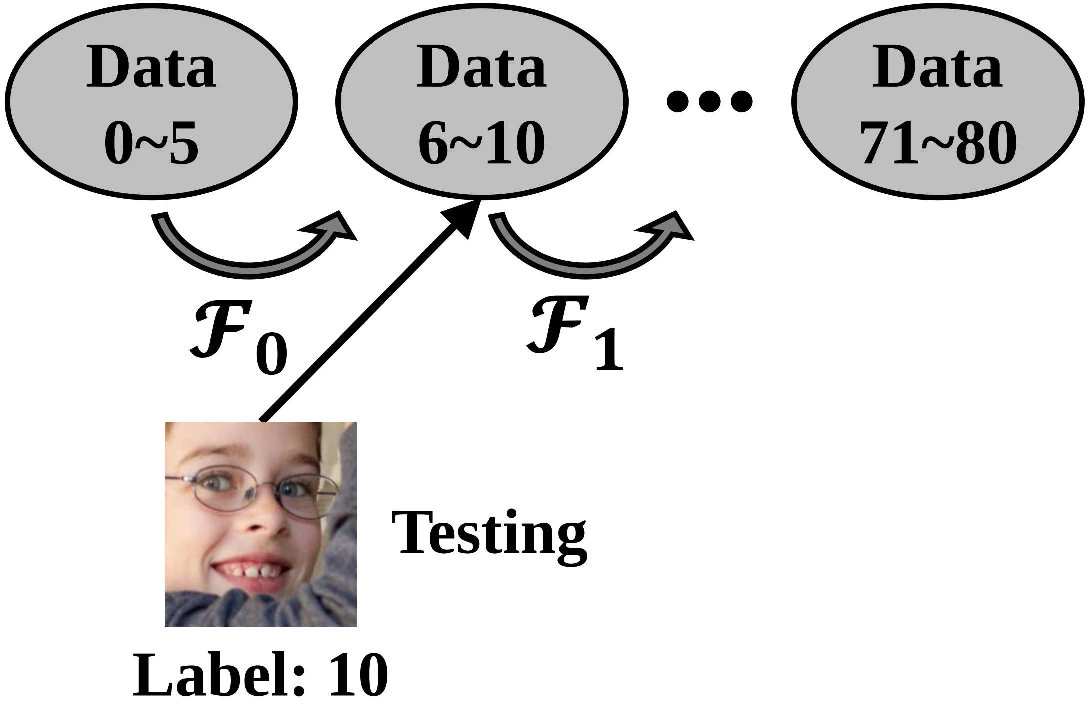


Wang, et al., CVPR2016

Main Idea --- Manifold Traversing

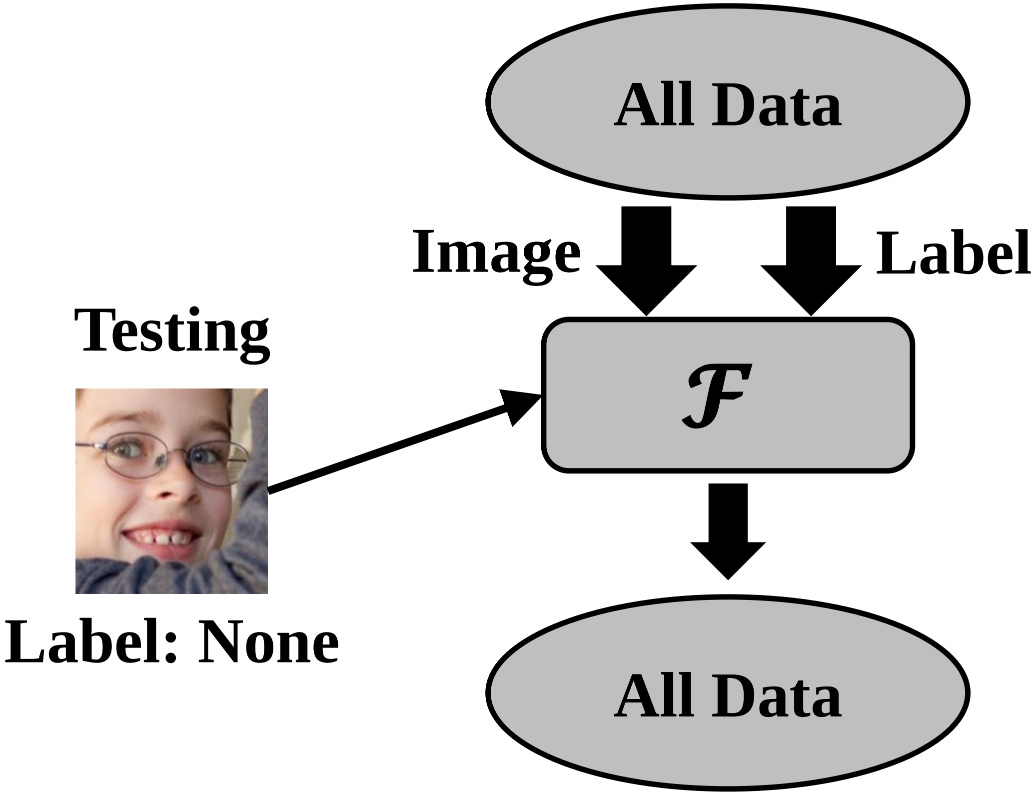


Previous Works



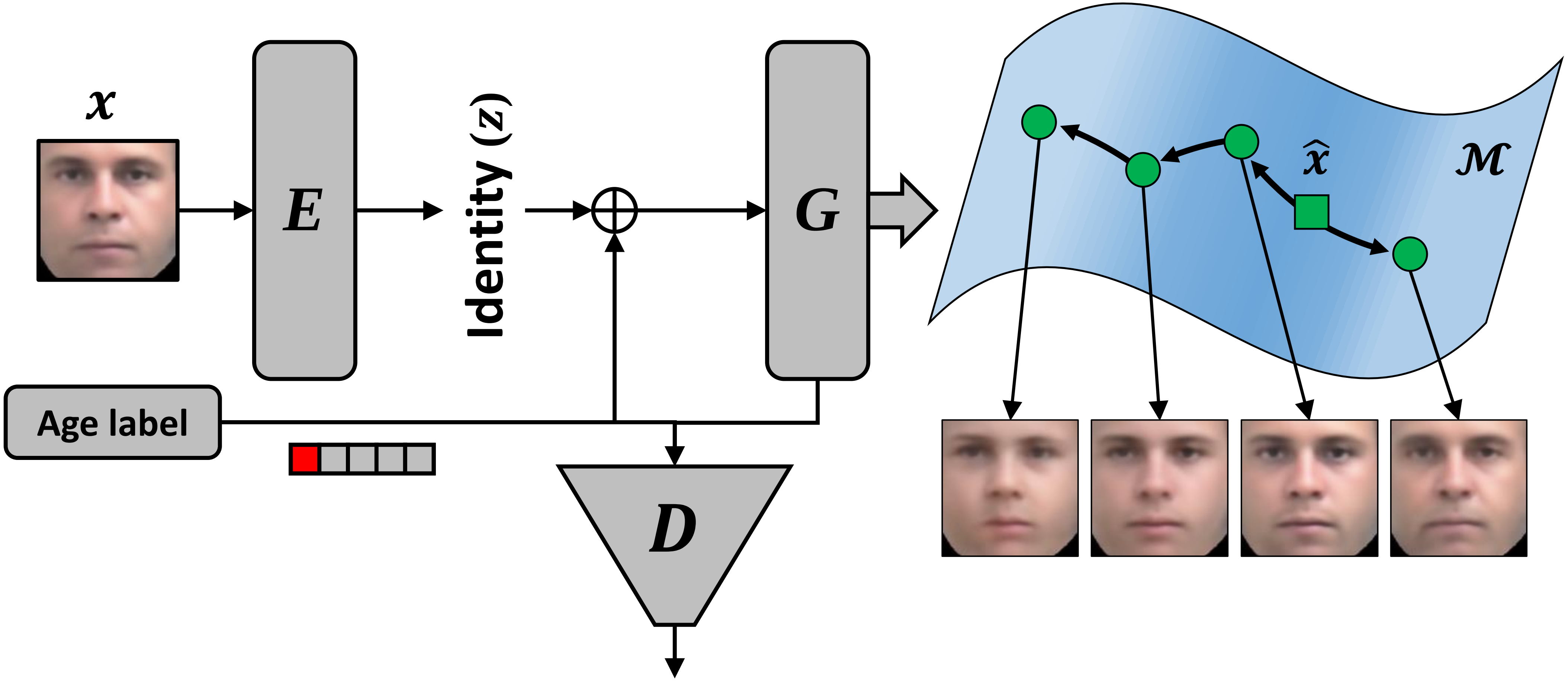
- Group-wise learning
- Unidirectional transition
- Label required in testing

Ours



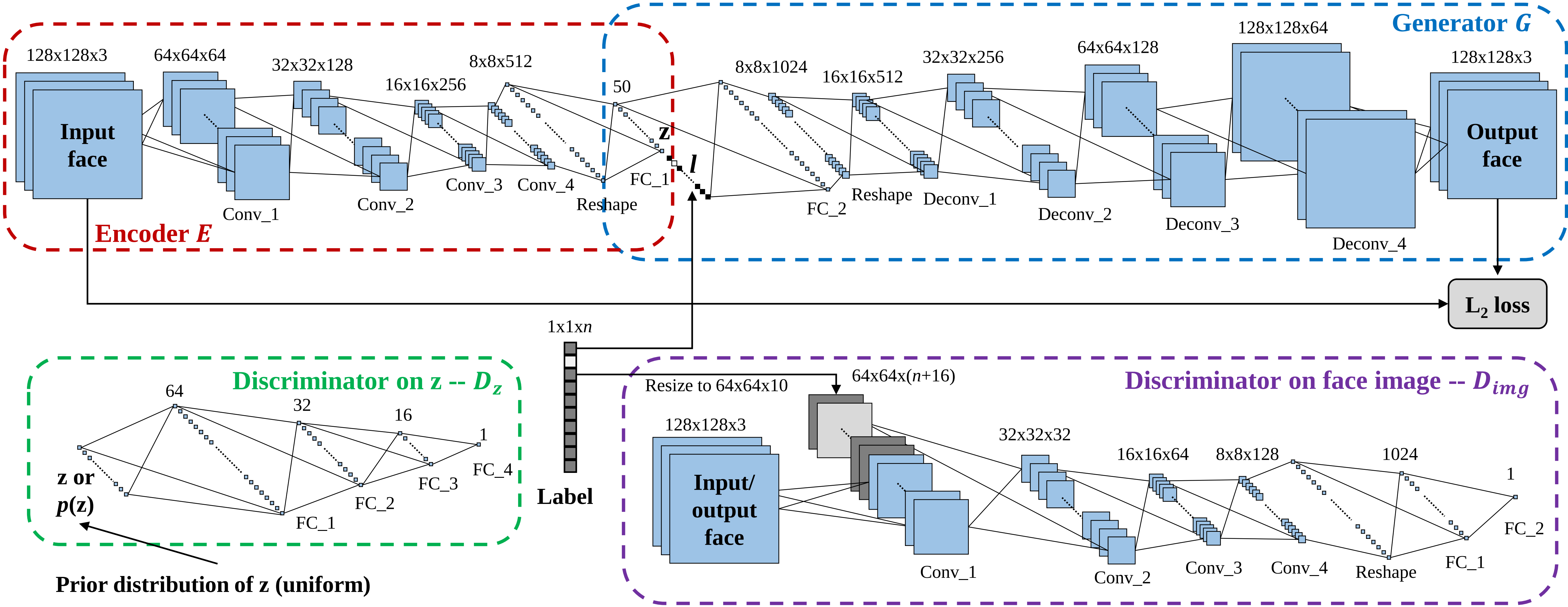
- Joint learning
- Bidirectional transition
- No label in testing

Conditional Adversarial Autoencoder



Real/fake conditioned on age

Conditional Adversarial Autoencoder (CAAE)



Conditional Adversarial Autoencoder (CAAE)

Objective function

D on z

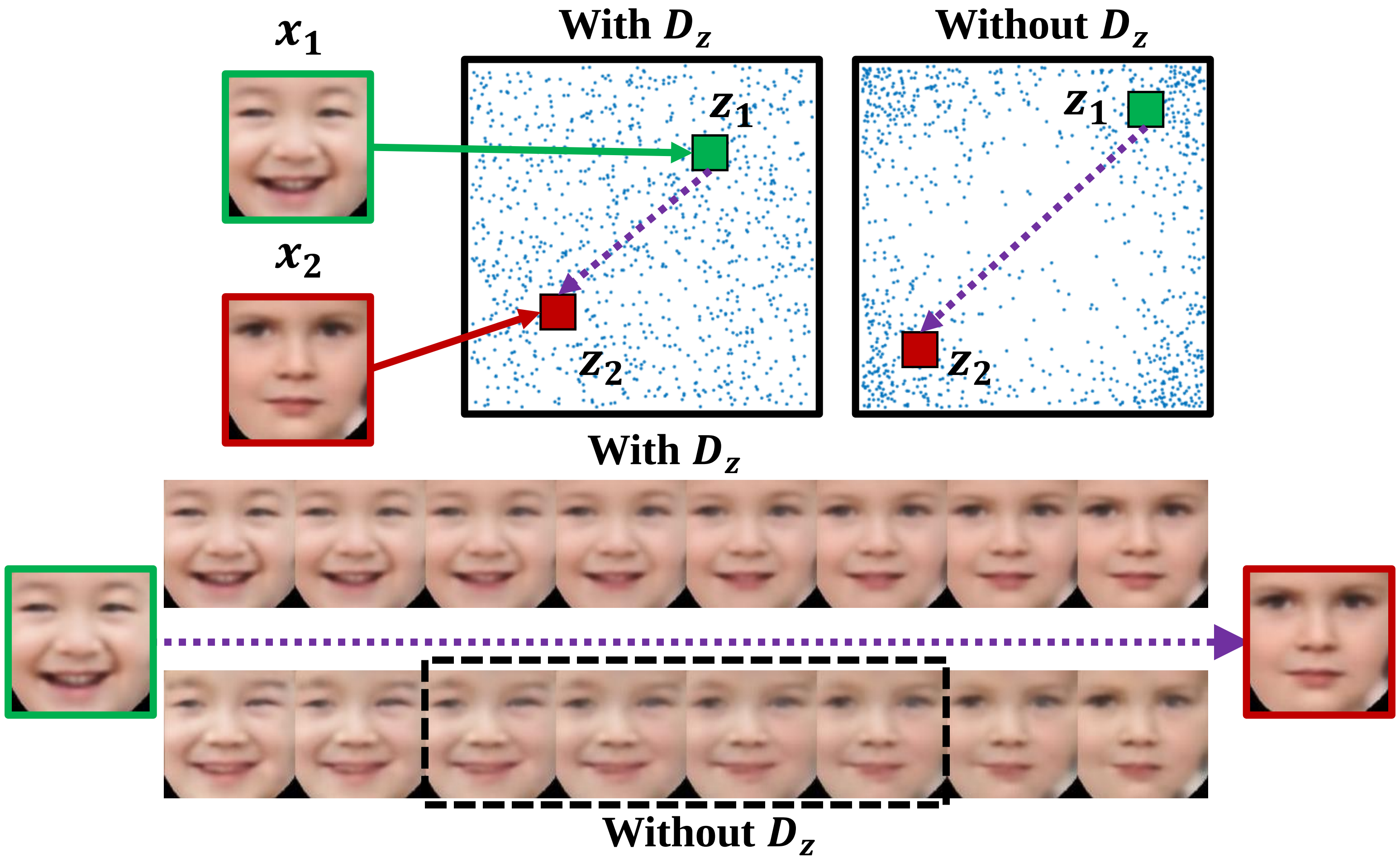
D on image

$$
\begin{aligned}
\min_{E,G} \max_{D_z, D_{img}} \; & \lambda \overbrace{\mathcal{L}\left(x, G(E(x), l)\right)}^{\text{Reconstruction error}} + \gamma \overbrace{TV\left(G(E(x), l)\right)}^{\text{Total variation}} \\
& + \mathbb{E}_{z^* \sim p(\mathbf{z})} \left[\log D_z(z^*)\right] \\
& + \mathbb{E}_{x \sim p_{data}(\mathbf{x})} \left[\log(1 - D_z(E(x)))\right] \\
& + \mathbb{E}_{x,l \sim p_{data}(\mathbf{x},\mathbf{l})} \left[\log D_{img}(x, l)\right] \\
& + \mathbb{E}_{x,l \sim p_{data}(\mathbf{x},\mathbf{l})} \left[\log(1 - D_{img}(G(E(x), l)))\right],
\end{aligned} \tag{5}
$$

where $TV(\cdot)$ denotes the total variation which is effective in removing the ghosting artifacts. The coefficients λ and γ balance the smoothness and high resolution.

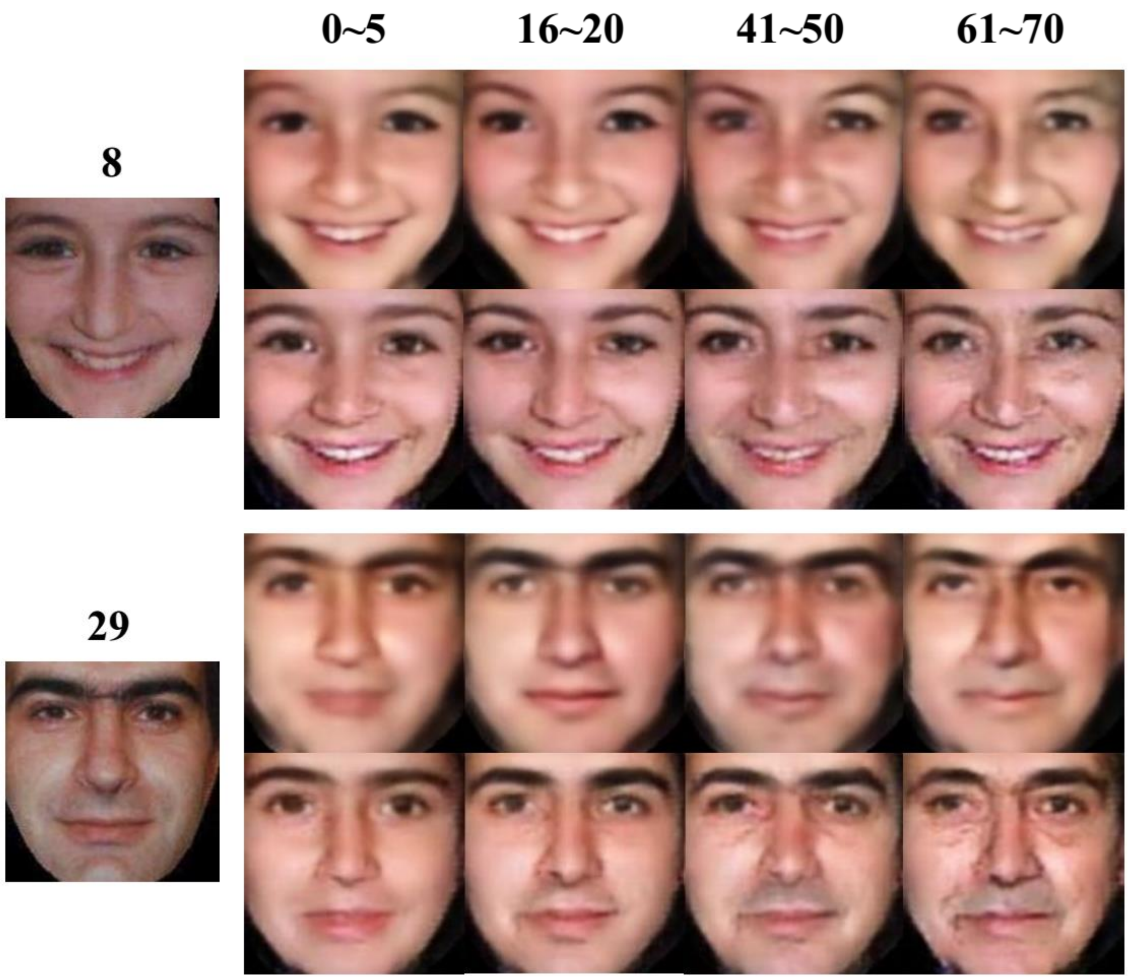
Conditional Adversarial Autoencoder (CAAE)

Effect of the Discriminator on z



Conditional Adversarial Autoencoder (CAAE)

Effect of the Discriminator on image

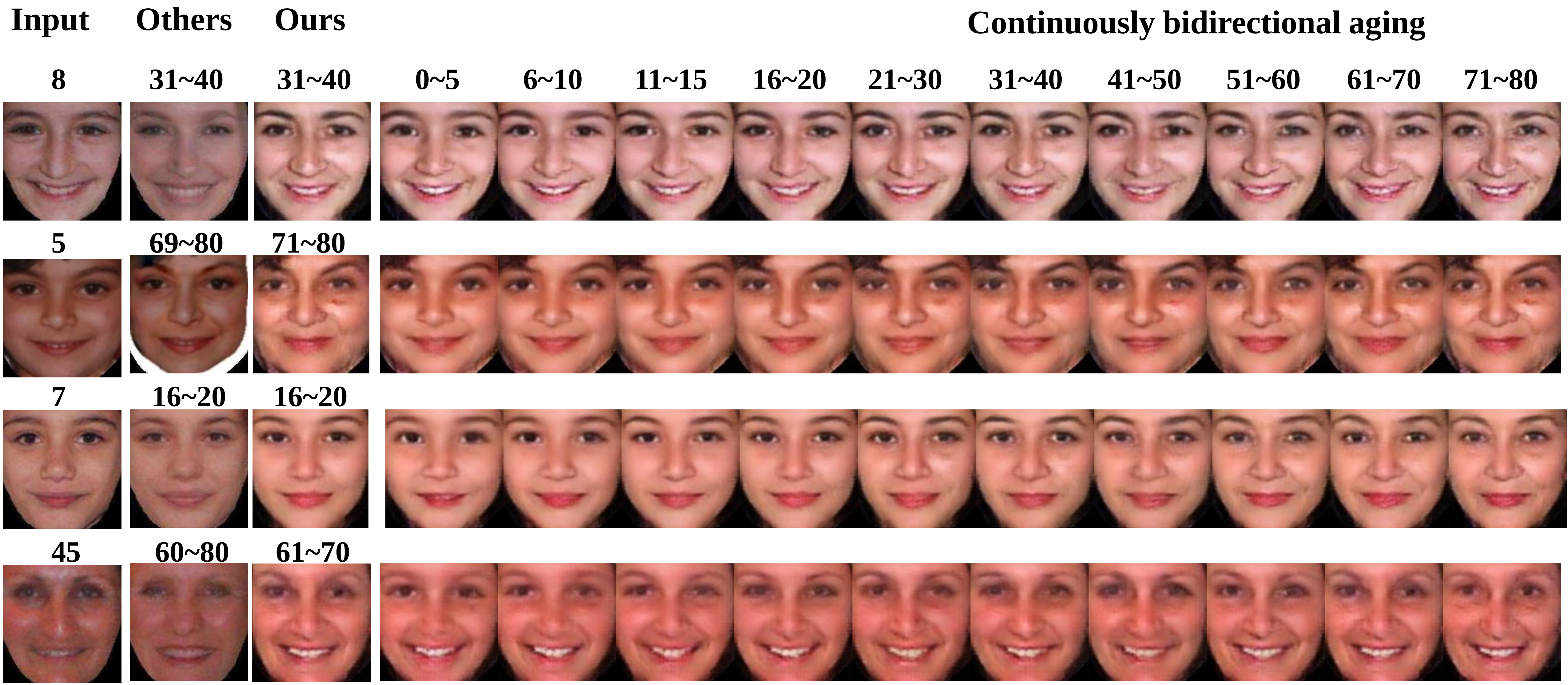


Without D_{img}

With D_{img}

GAN boosts the image resolution.

Experimental Evaluation



Project page: https://zzutk.github.io/Face-Aging-CAAE

Contents

Drawback of Conditional GAN

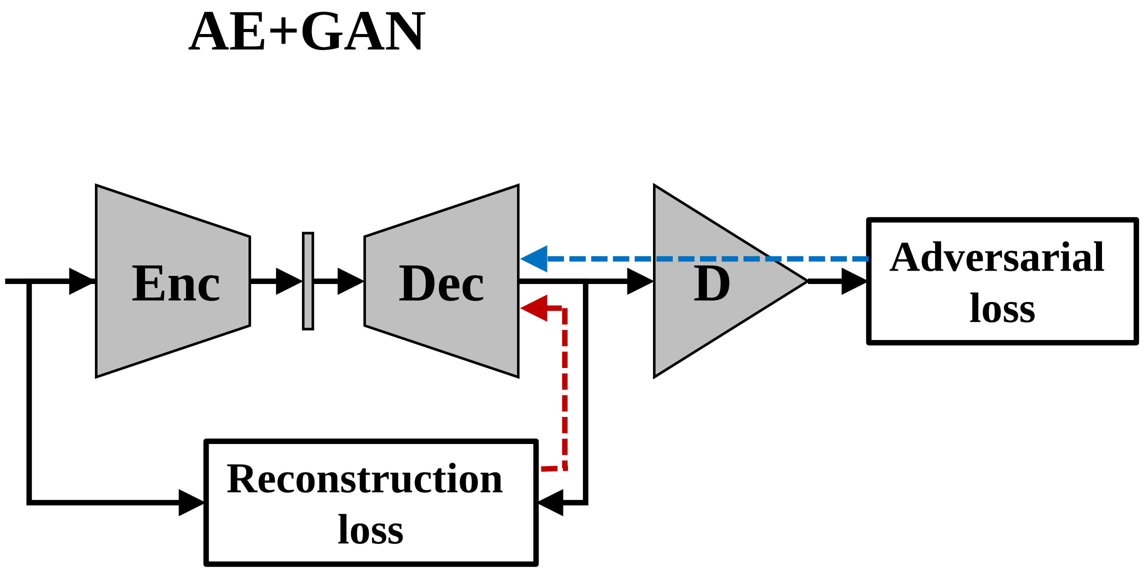
Need to carefully balance GAN loss and reconstruction loss

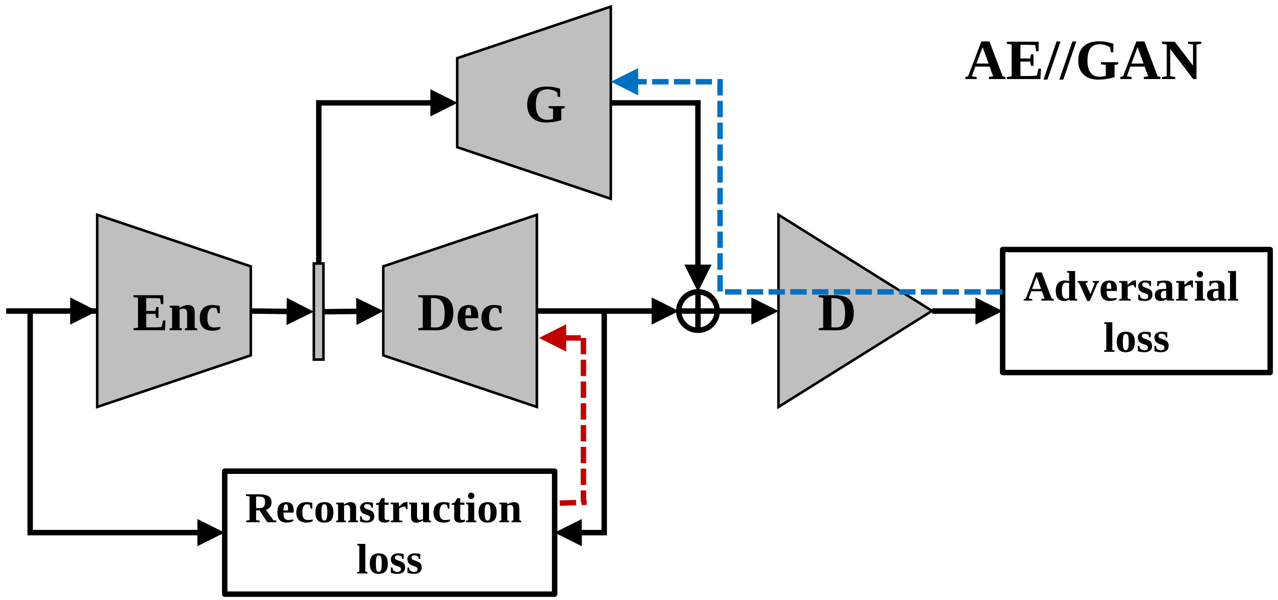
Assume the weight of AE is 1, and the weight of GAN varies from 0 to 0.1

Too low → blurry

Too high → noisy

Decoupled Learning --- AE//GAN

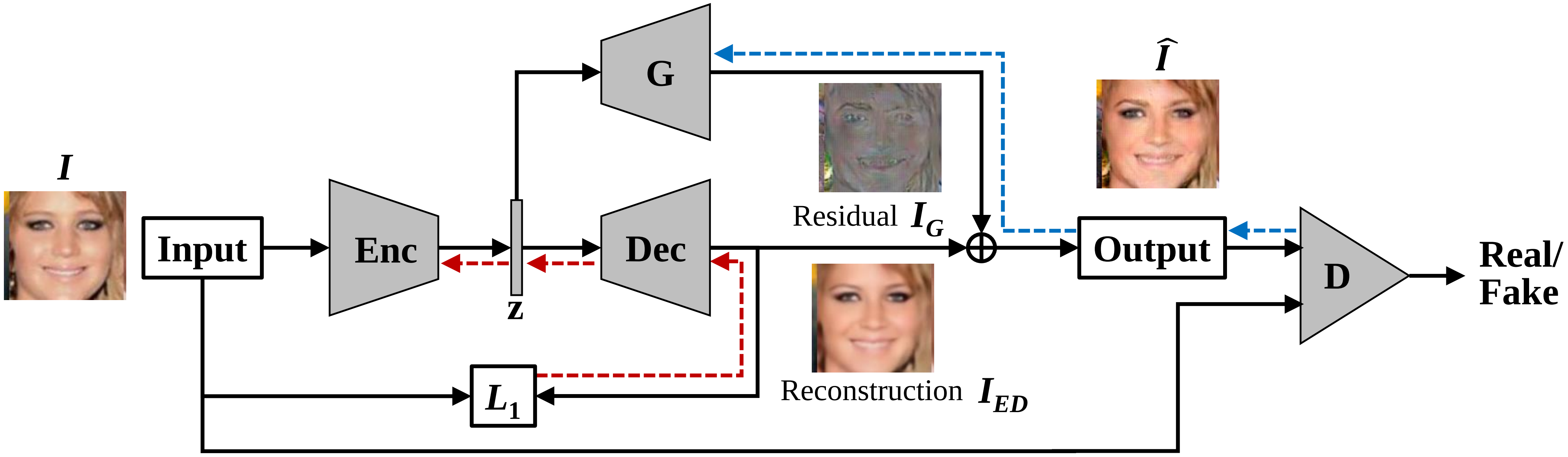




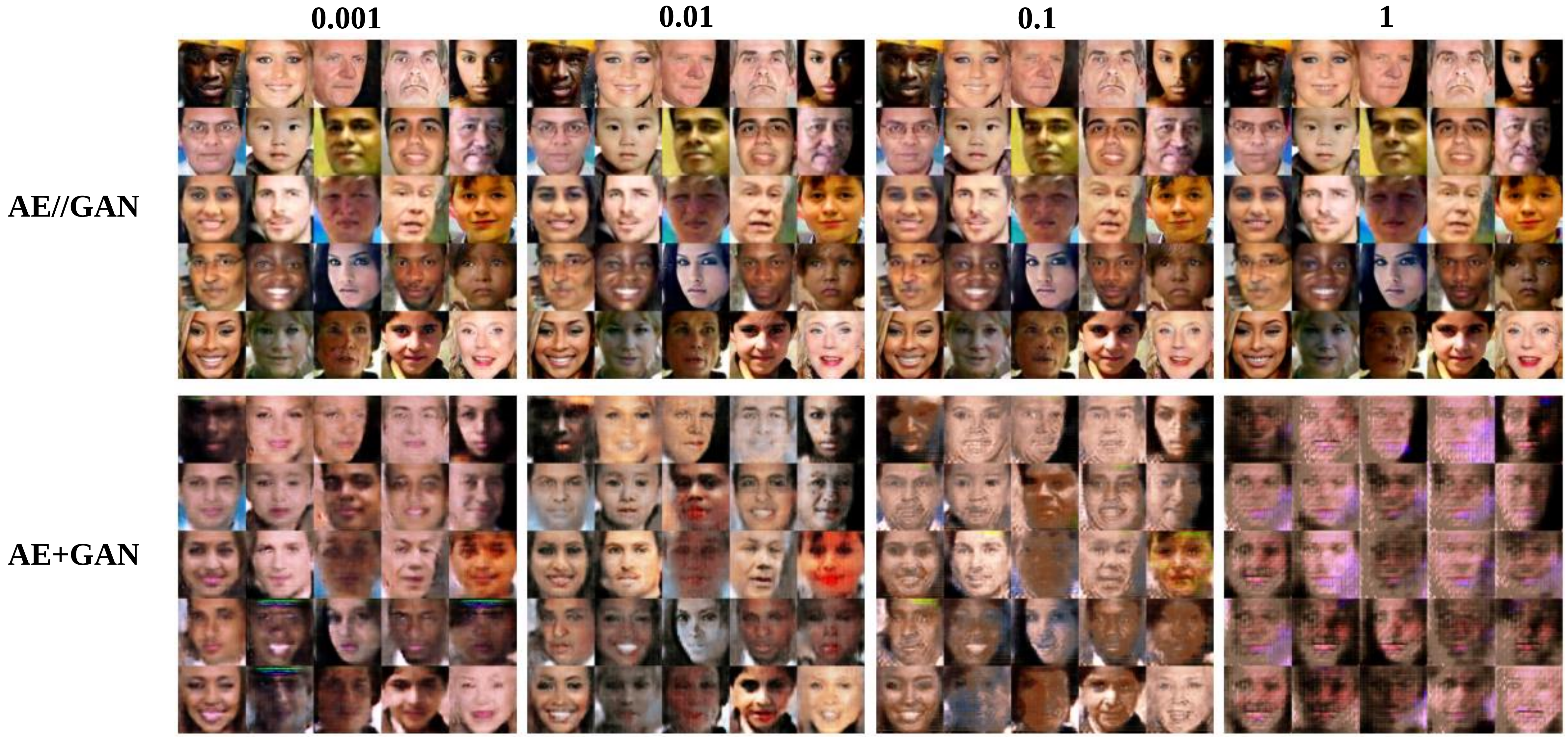
AE and GAN are tied in the parts of decoder/generator. Therefore, the reconstruction loss and adversarial loss interact/compete with each other, potentially causing unstable results.

The path of backpropagation is decoupled that avoids the interaction between the two losses, and thus relaxes the effort on balancing them.

Decoupled Learning --- AE//GAN



Experimental Evaluation



Evaluation Metric of GANs

- It is challenge to evaluate image quality/GAN performance
- There is no convincing metric for evaluating image quality

Inception score

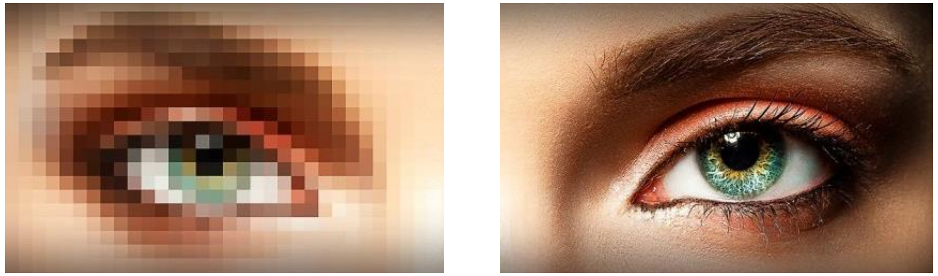
Downsides of inception score:

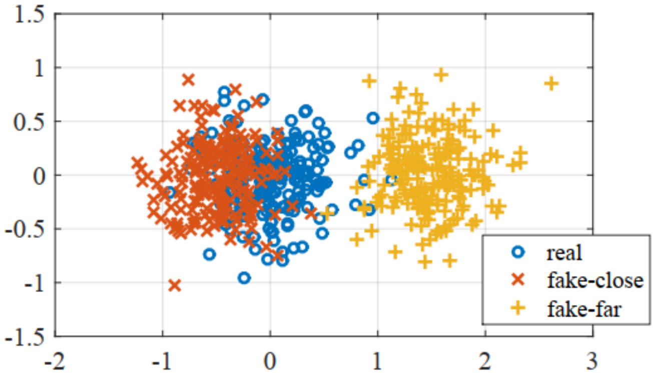
- Heavily depend on the pre-trained classifier
- The results may vary with different datasets
- Absolut score misaligns actual performance

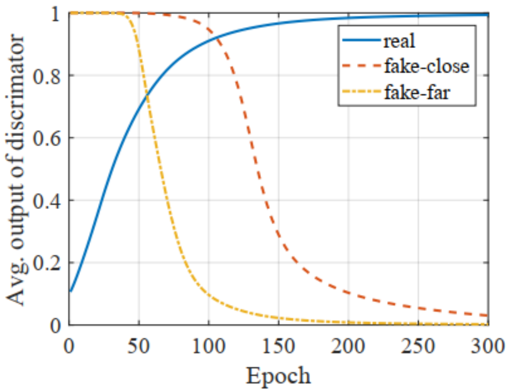
New Evaluation Metric of GANs

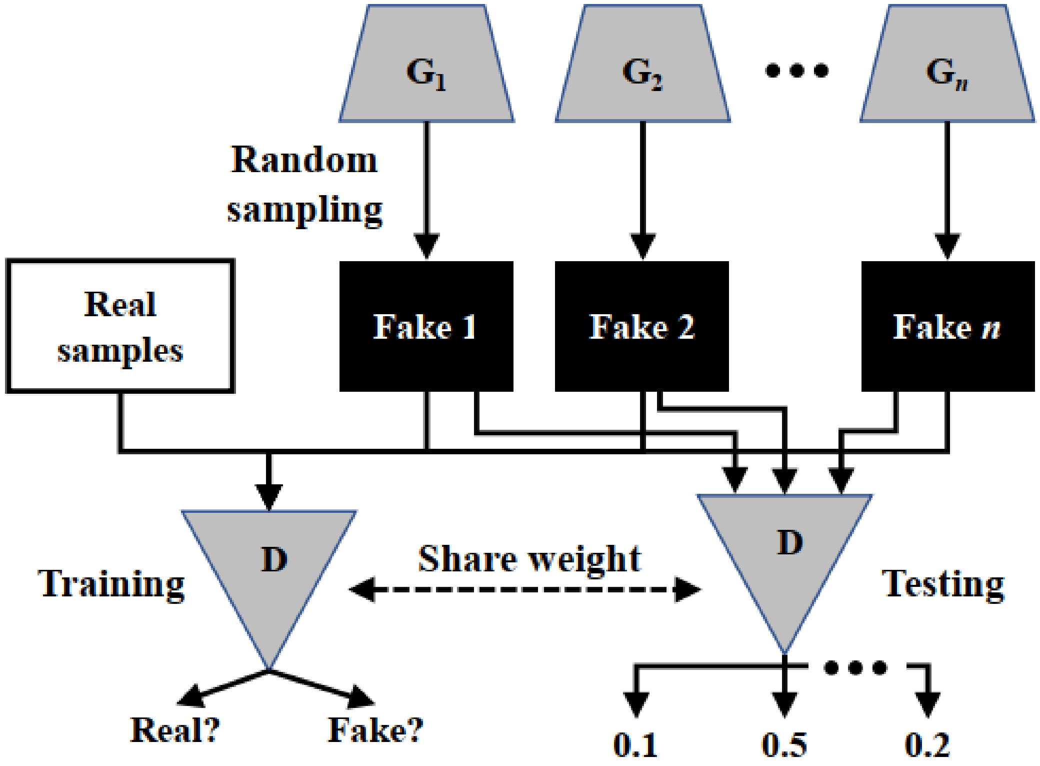
Idea: relative comparison of models

Which is better?









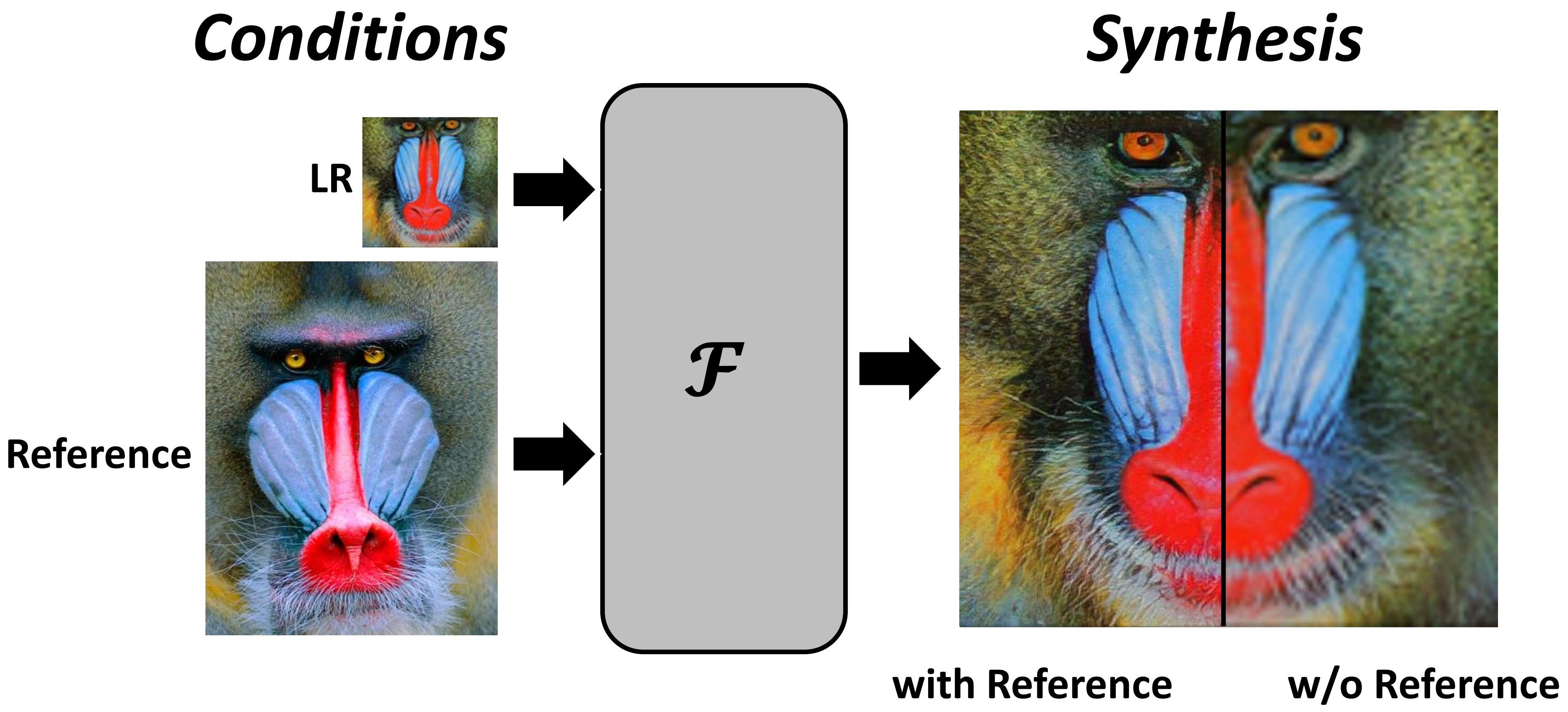
Project page: https://github.com/ZZUTK/Decoupled-Learning-Conditional-GAN

Conditional Image Synthesis --- Summary

- Conditional GAN outperforms previous works in image synthesis
- Decoupled learning could further stabilize the training of Conditional GAN
- The relative comparison is proposed for image quality evaluation.

Contents

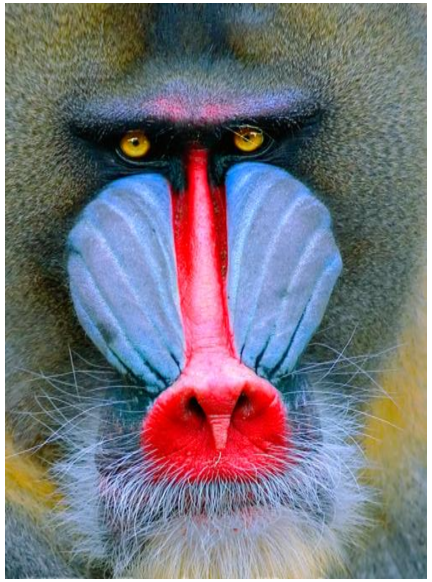
Reference-Conditioned Super-Resolution



Super-Resolution

Input

Reference



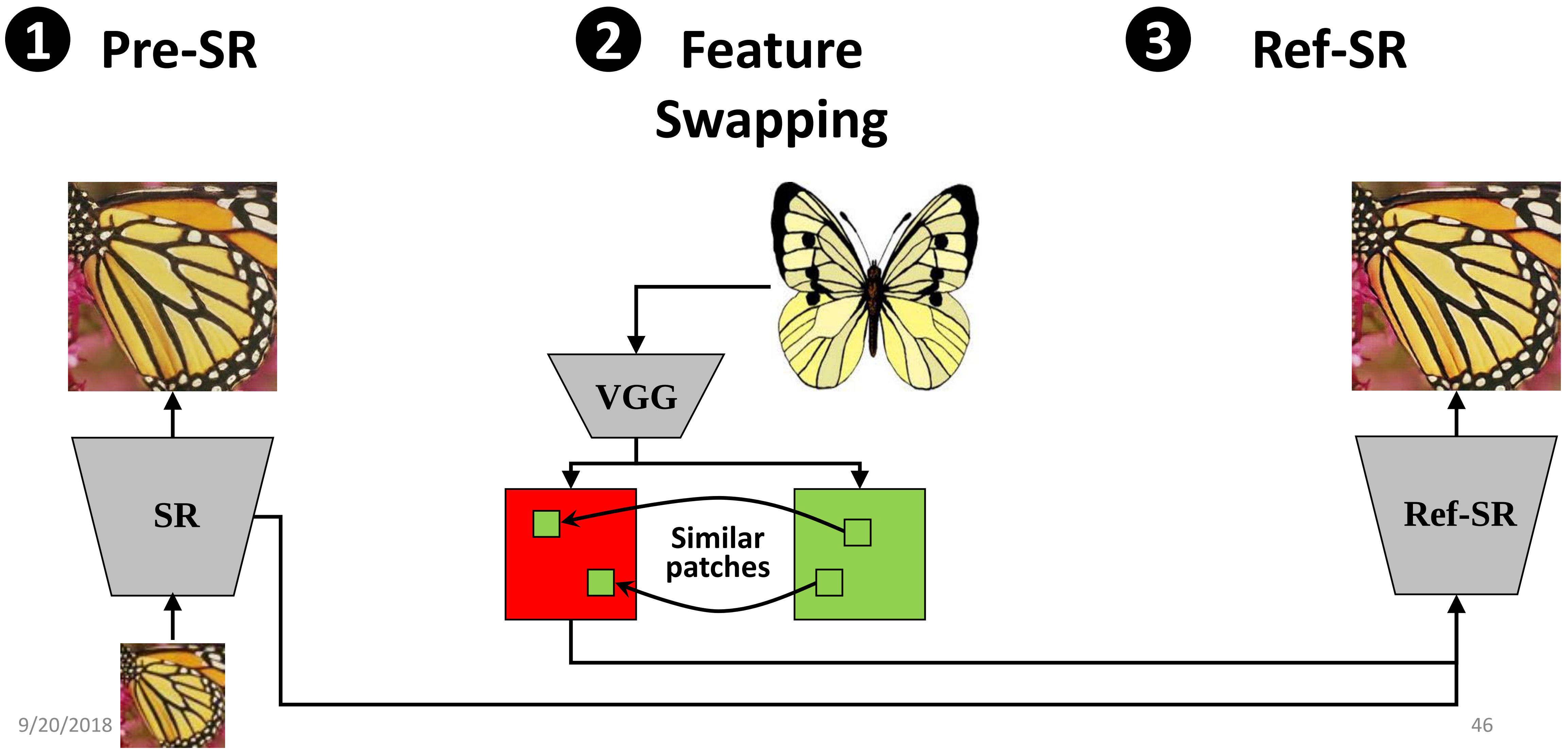
Original

Bicubic

Ours

SRResNet [Ledig et al., CVPR2017]

Our Approach



Reference-Conditioned Super-Resolution


Bicubic (4x)


SRCNN


SRGAN


SRNTT (ours)

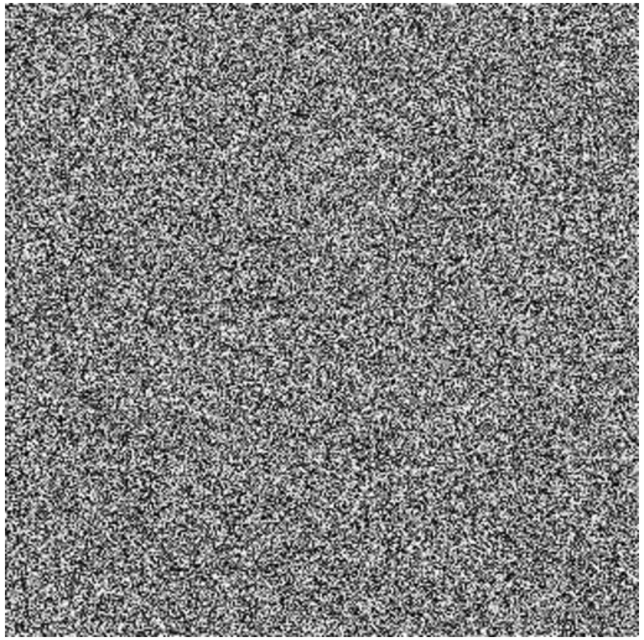

Reference

Project page: http://web.eecs.utk.edu/~zzhang61/project_page/SRNTT/SRNTT.html

What If A Bad Reference

LR

Reference

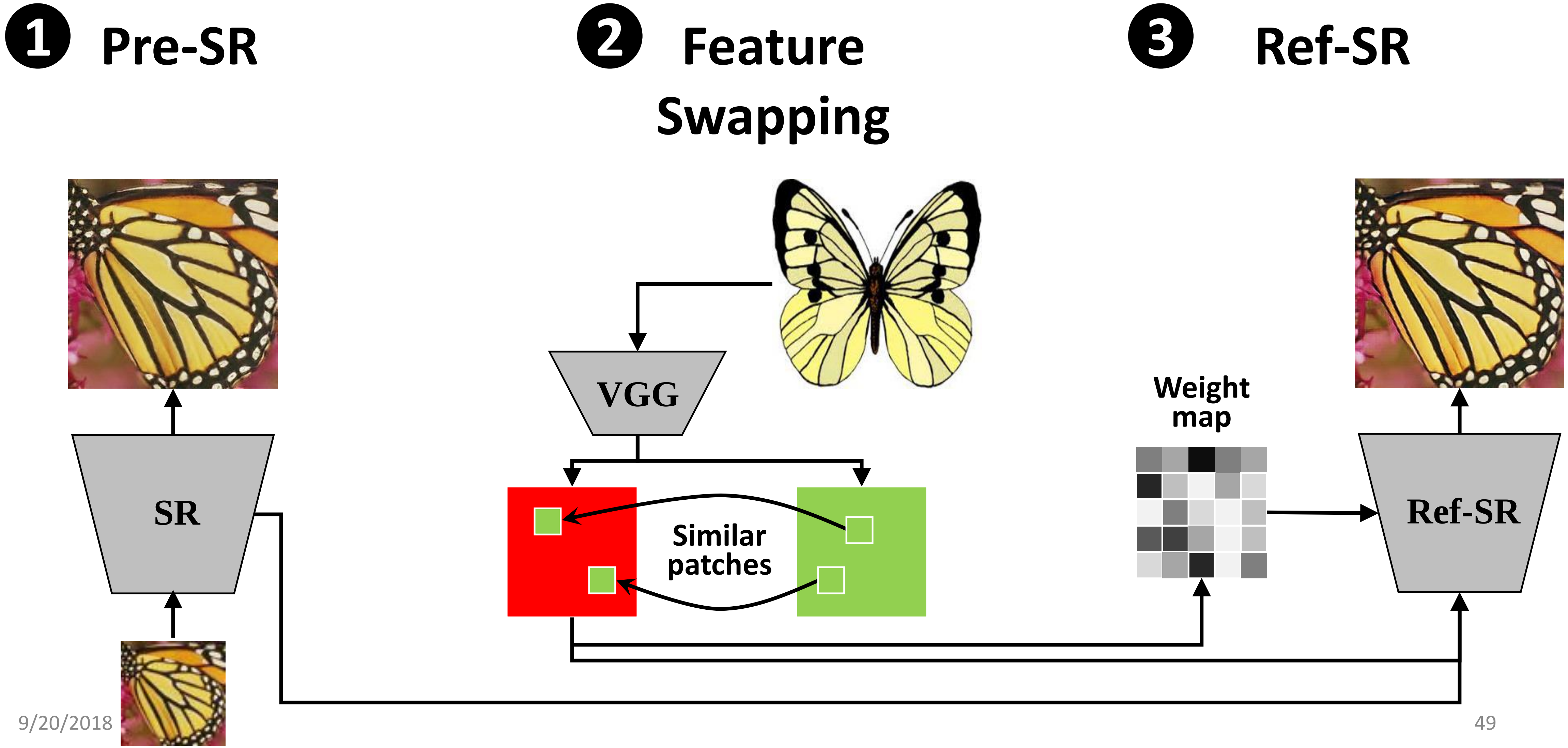


Original

SR 4x

Negative effect from the reference is introduced to the output

Our Approach



Effect of the Weight Map

LR

Reference

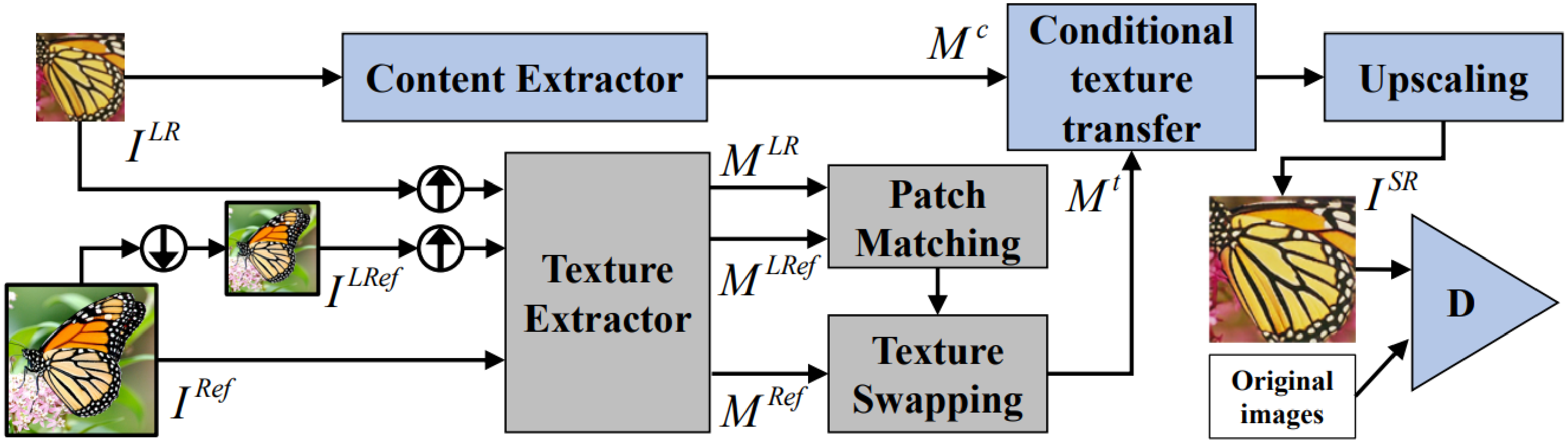
Original

w/o weight

Weighted

The weight reduce negative effect from the reference

Reference-Conditioned Super-Resolution

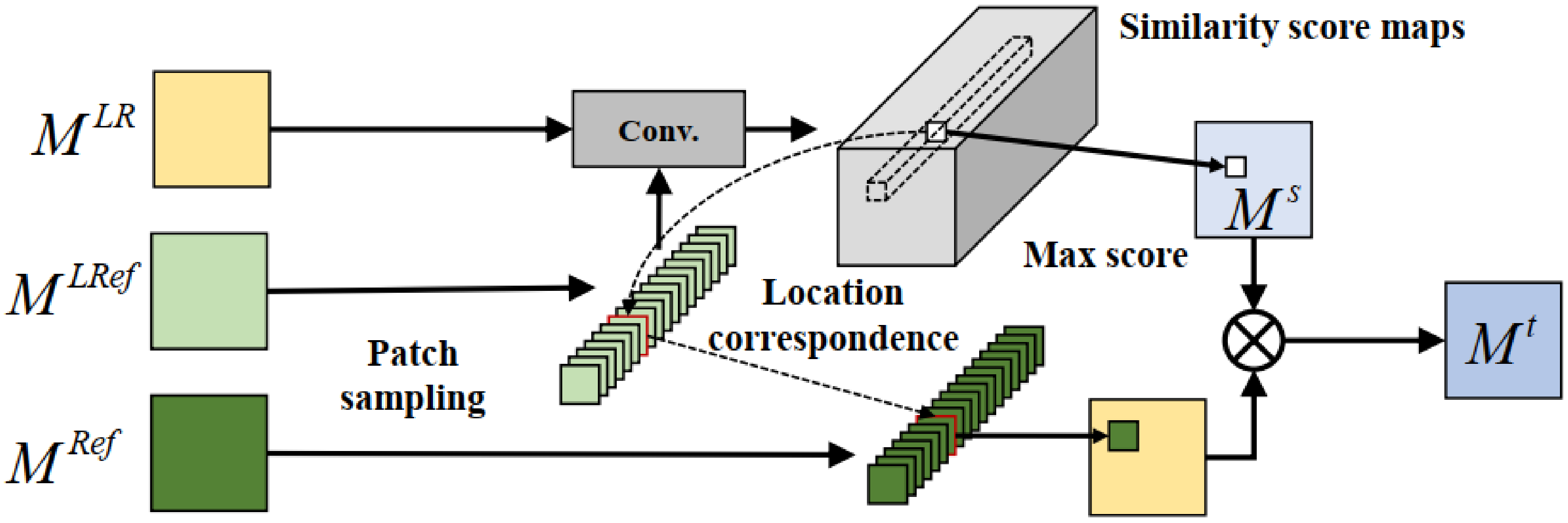


Texture loss

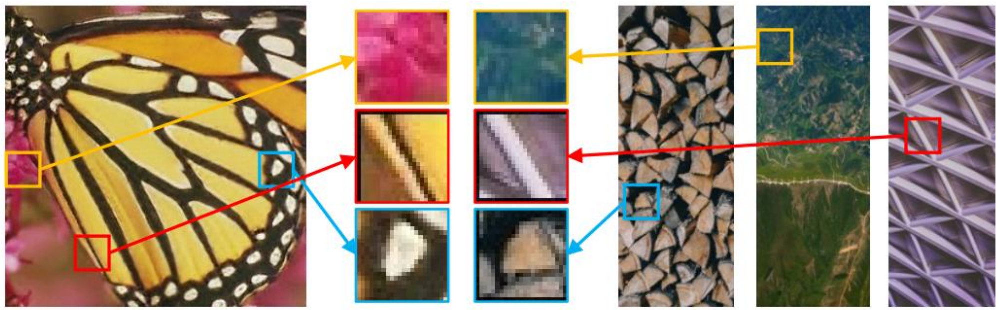
$$\mathcal{L}_t = \frac{1}{4V^2} \left\| Gr(\phi(I^{SR}) \otimes M^s) - Gr(M^t) \right\|_F$$

Reference-Conditioned Super-Resolution

Texture loss: $$\mathcal{L}_t = \frac{1}{4V^2} \left\| Gr(\phi(I^{SR}) \otimes M^s) - Gr(M^t) \right\|_F$$

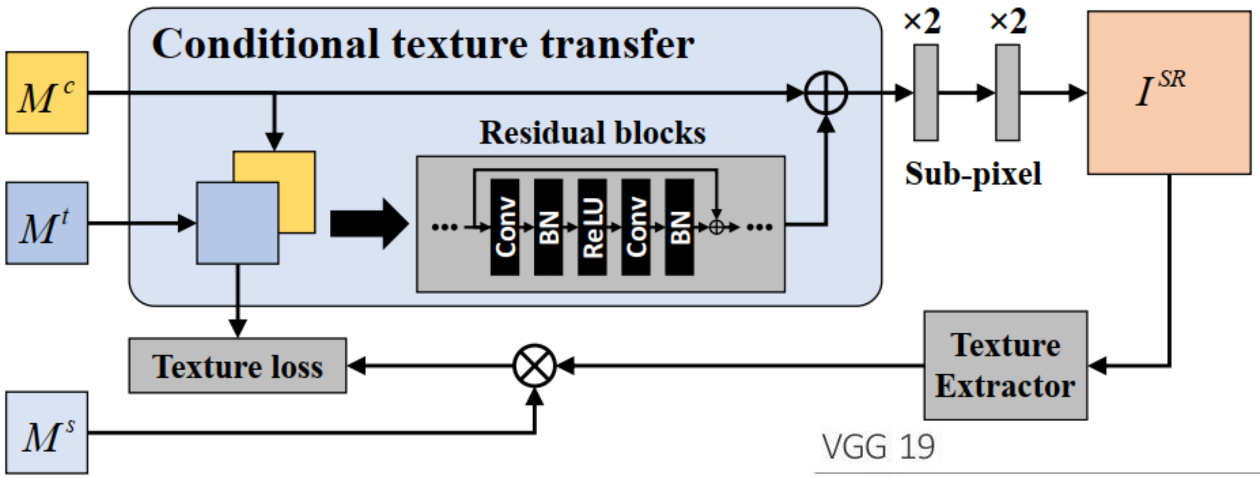


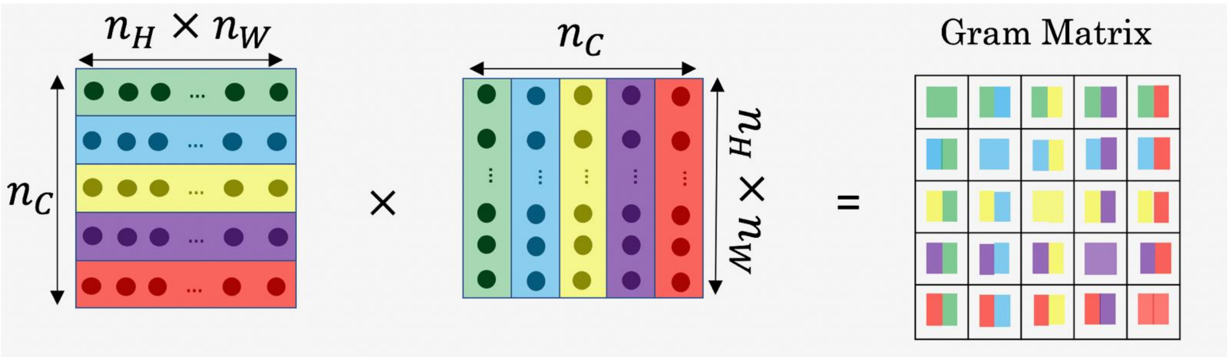
$$s_{i,j} = \left\langle \frac{p_i^{LR}}{\|p_i^{LR}\|}, \frac{p_j^{LRef}}{\|p_j^{LRef}\|} \right\rangle$$

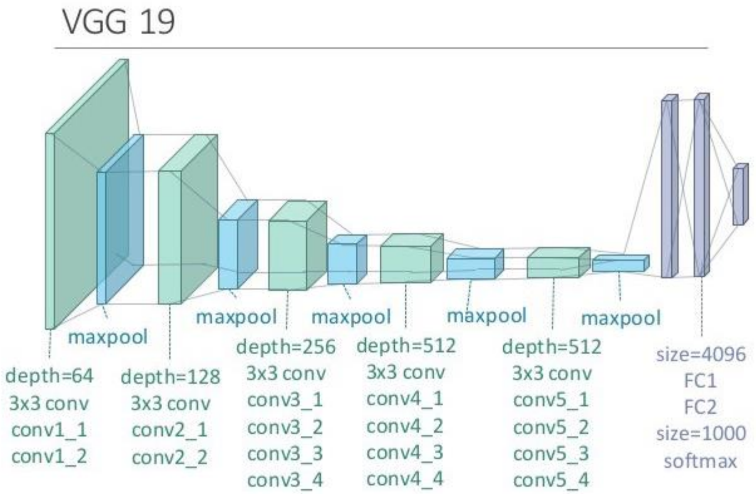


Reference-Conditioned Super-Resolution

Texture loss: $$\mathcal{L}_t = \frac{1}{4V^2} \left\| Gr(\phi(I^{SR}) \otimes M^s) - Gr(M^t) \right\|_F$$

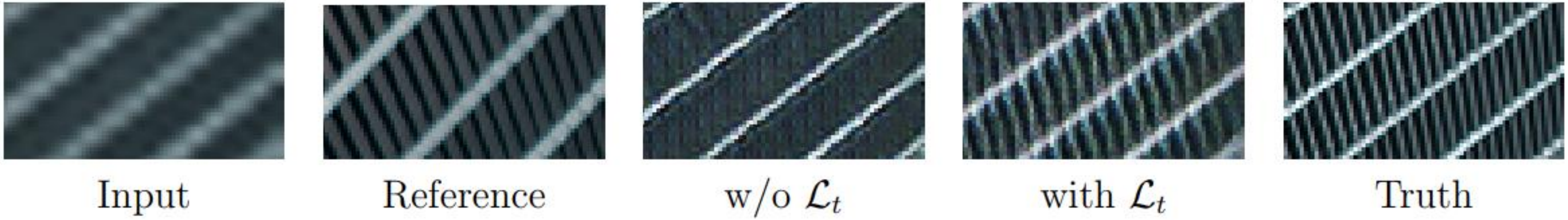






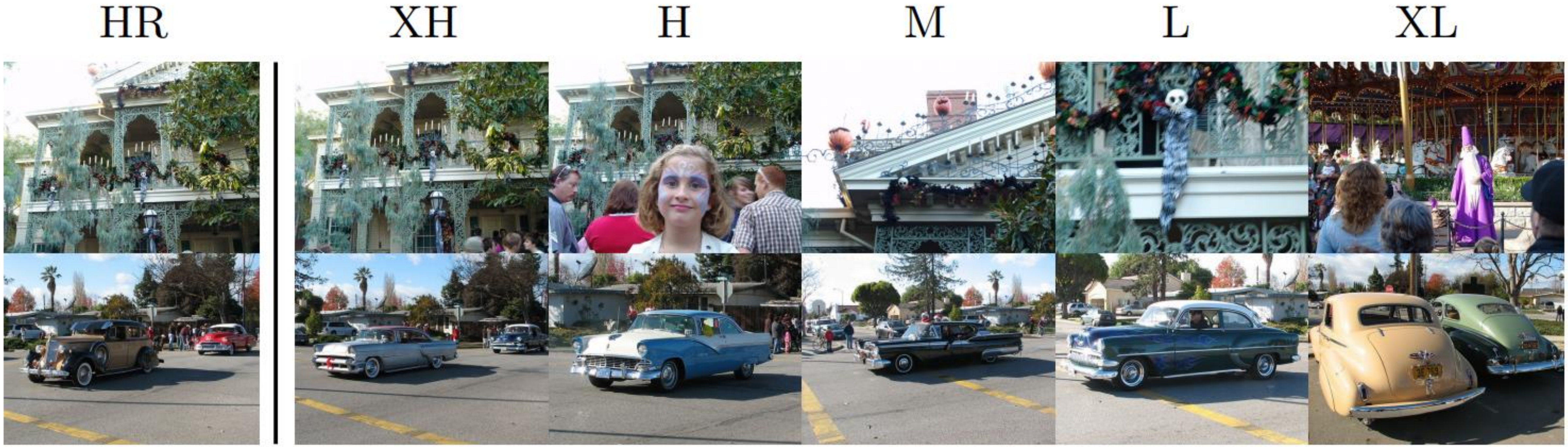
Reference-Conditioned Super-Resolution

Texture loss: $$\mathcal{L}_t = \frac{1}{4V^2} \left\| Gr(\phi(I^{SR}) \otimes M^s) - Gr(M^t) \right\|_F$$



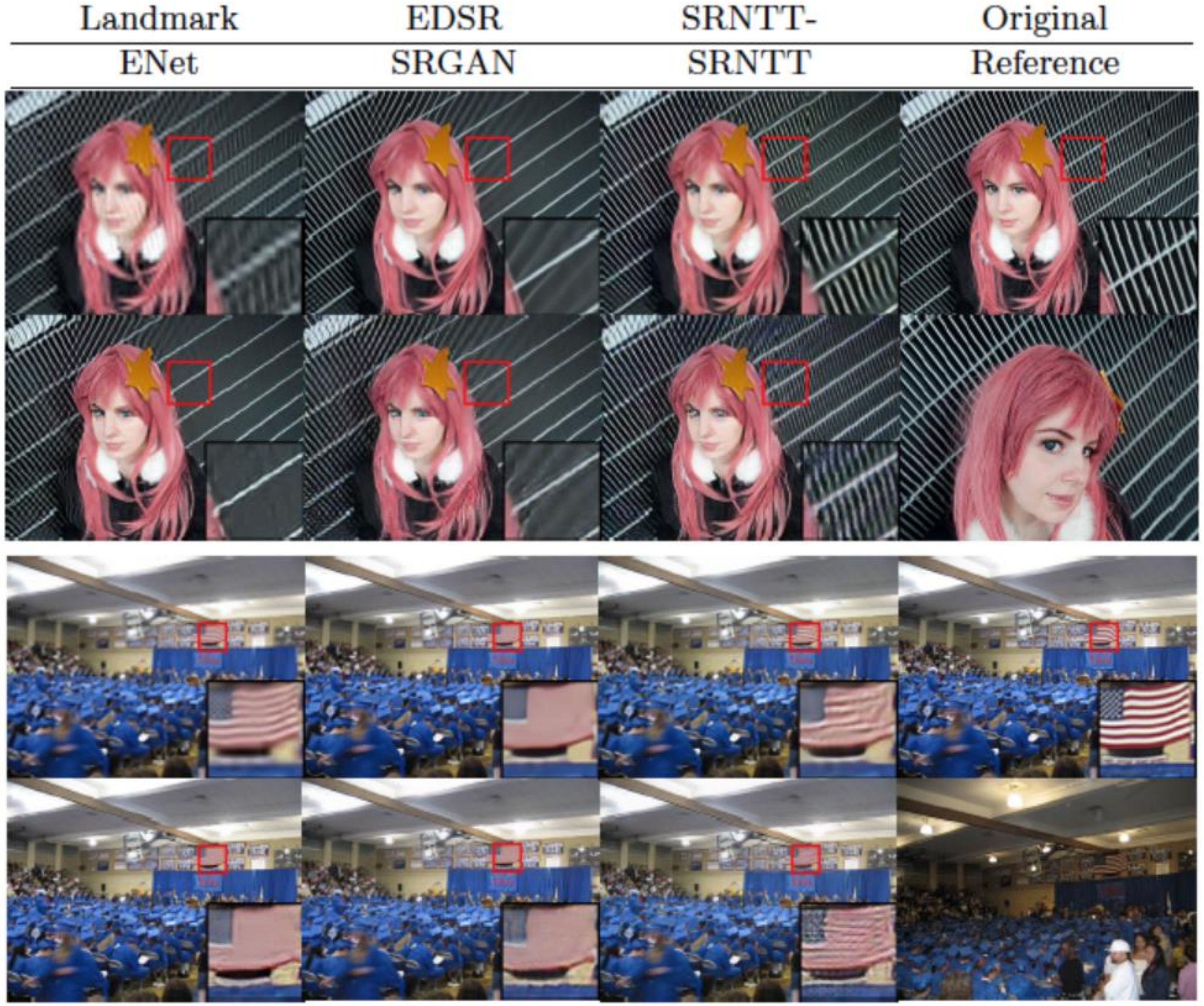
Data Collection --- CUFED5

There is no benchmark dataset for evaluating the reference-based SR methods

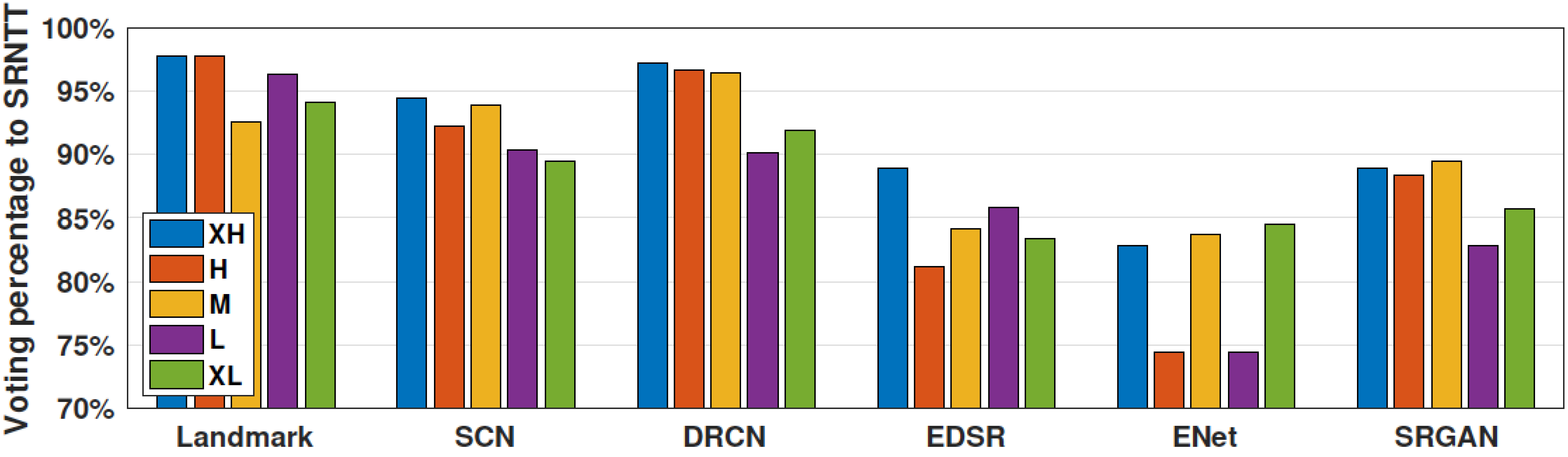


Project page: http://web.eecs.utk.edu/~zzhang61/project_page/SRNTT/SRNTT.html

Experimental Results



Experimental Results



	XH	H	M	L	XL
PSNR	24.57	24.58	24.56	24.63	**24.69**

 Project page: http://web.eecs.utk.edu/~zzhang61/project_page/SRNTT/SRNTT.html

Summary of Contributions

1. **Theoretically analyze GAN, i.e., drawbacks and improvements. GAN → conditional GAN → decoupled learning**
2. **Demonstrate the advantages of conditional GAN in image synthesis through the face aging task.**
3. **Extend conditional image synthesis to a traditional area, i.e., super-resolution, significantly boosting the visual quality.**
4. **Exploit relative comparison of GAN-based models, providing an alternative for image evaluation.**
5. **Extensive experimental evaluation is conducted to support the proposed designs.**


Weisheng Tang
Razieh Kaviani Baghbaderani
Chengcheng Li
Dr. Zhibo Wang
Dr. Husheng Li
Yang Song
Quan Zhou
Dr. Jens Gregor
Dr. Wei Wang
Dr. Austin P. Albright
Dr. Hairong Qi
Dr. Ying Qu
Elliot Davis Greenlee
Steven Patrick
Dr. Russell Zaretzki
Waofeng Tang
Dr. Jiajia Luo
Fangi Wang
Dr. Ali Taalimi
Dr. Rafael C. Gonzalez
Dr. Li He
Alireza Rahimpour
Dr. Liu Liu
Dr. Shuangjiang Li
Jia (Jason) Liang
Mohamad Ramin Nabati
Dr. Rui Guo

Thank you