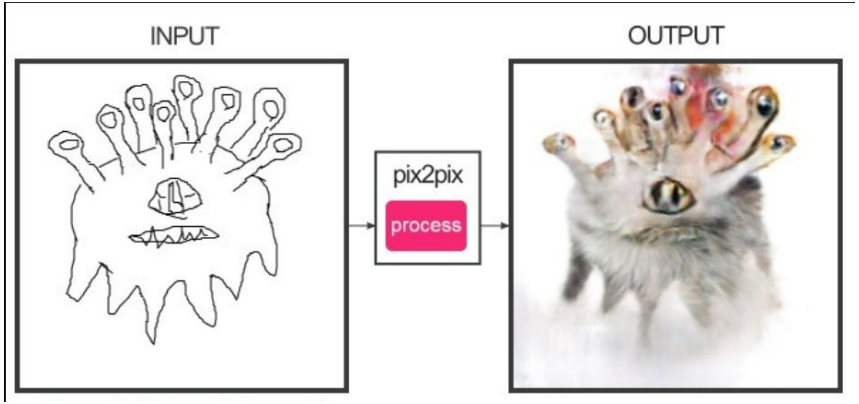
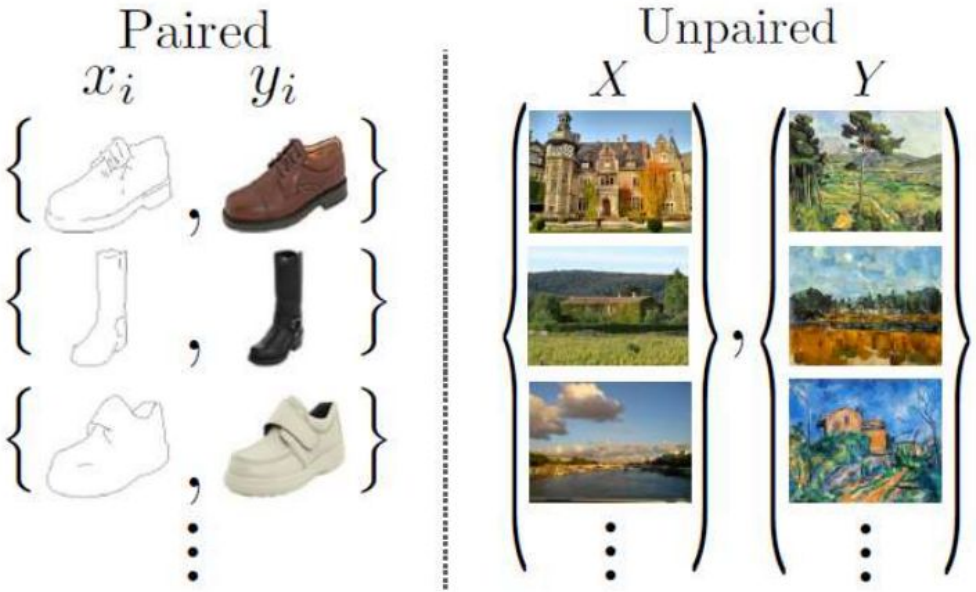
Подвиды генерации изображений

Image2Image translation



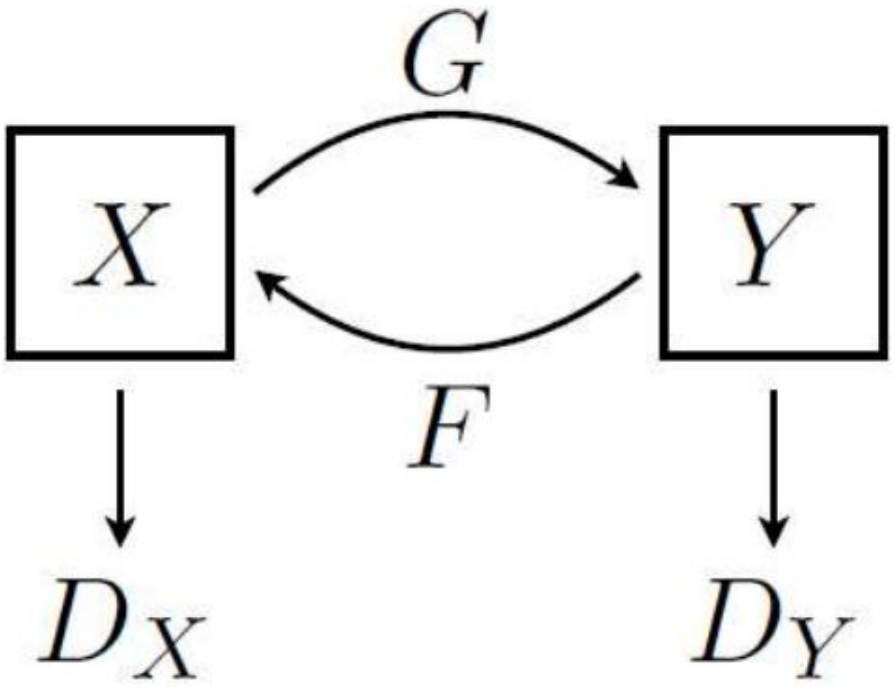
Paired and Unpaired Image to Image Translation

- **Paired** Training data- yi corresponding to each xi is given
- **Unpaired** data- No information provided as to which xi corresponds to which y



Cyclic Model

- G maps X to Y
- F is inverse mapping from Y to X
- DX and DY are discriminators

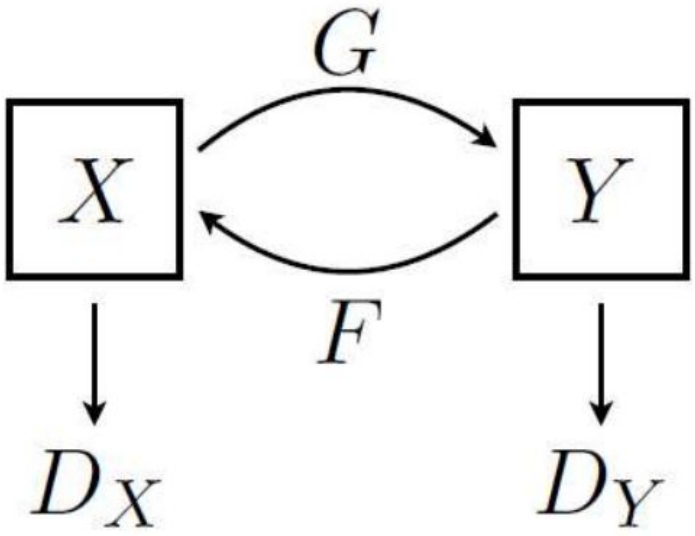


Adversarial Loss

$$G : X \rightarrow Y$$
$$G^* = \arg\min_G \max_{D_Y} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y)$$
$$F : Y \rightarrow X$$
$$F^* = \arg\min_F \max_{D_X} \mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$$

where,

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$



Cycle Loss

- Forward Cycle Consistency

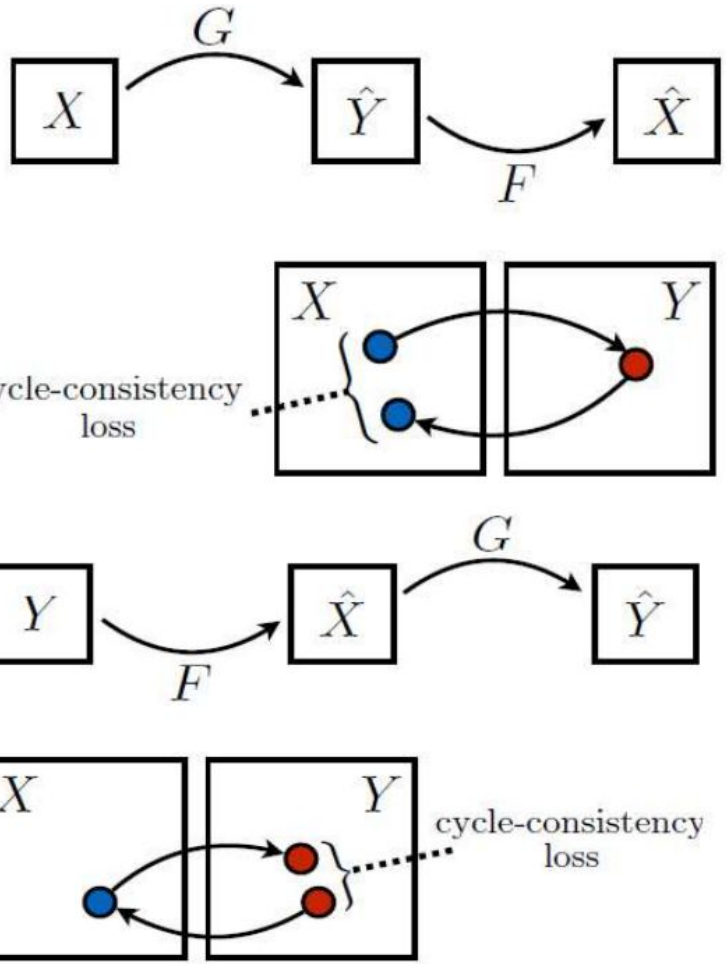
$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$$

- Backward Cycle Consistency

$$y \rightarrow F(y) \rightarrow G(F(y)) \approx y$$

- Cycle Consistency Loss

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]$$



Objective Function

$$L(G, F, D_x, D_y) = L_{GAN}(G, D_y, X, Y) + L_{GAN}(F, D_x, X, Y) + \lambda L_{cyc}(G, F)$$

where, λ controls relative importance of the two objective functions

- The objective is to solve-

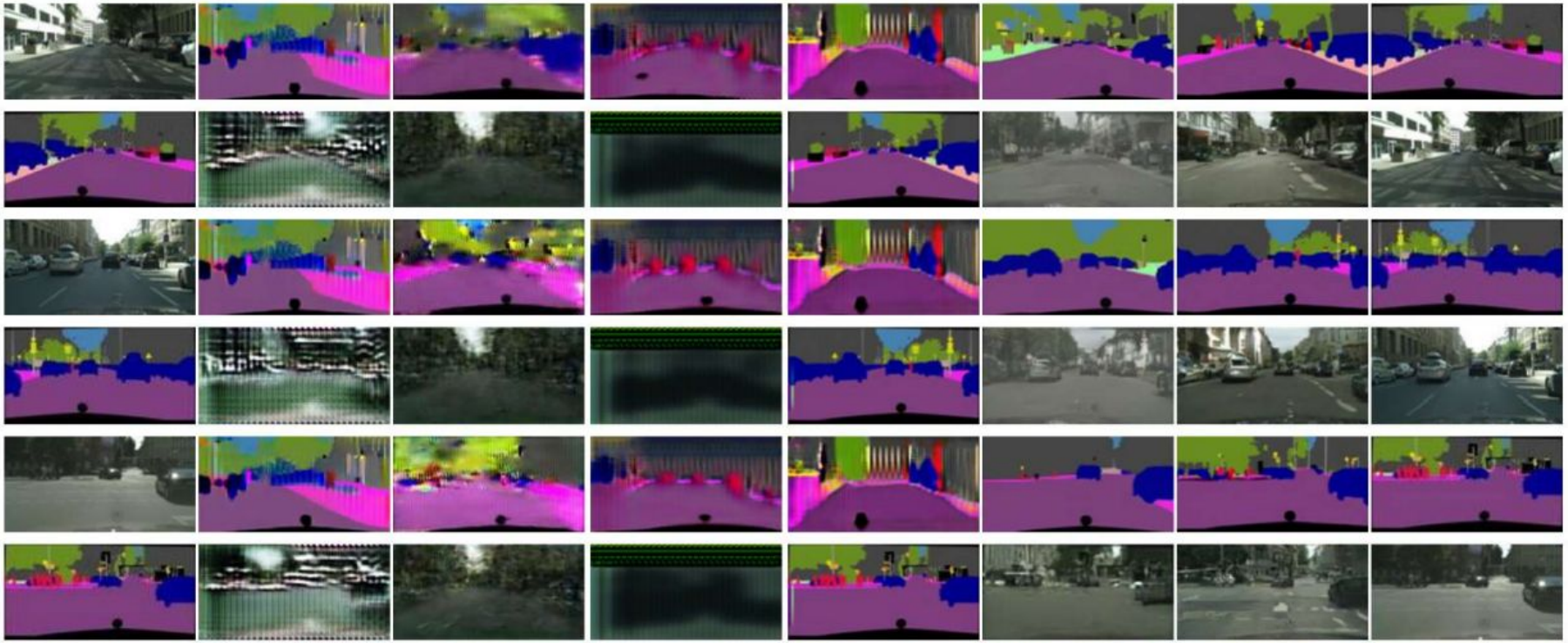
$$G^*, F^* = \arg\min_{F,G} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y)$$

- For implementation purpose we use-

$$\mathcal{L}_{\text{LSGAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[D_Y(G(x))^2]$$

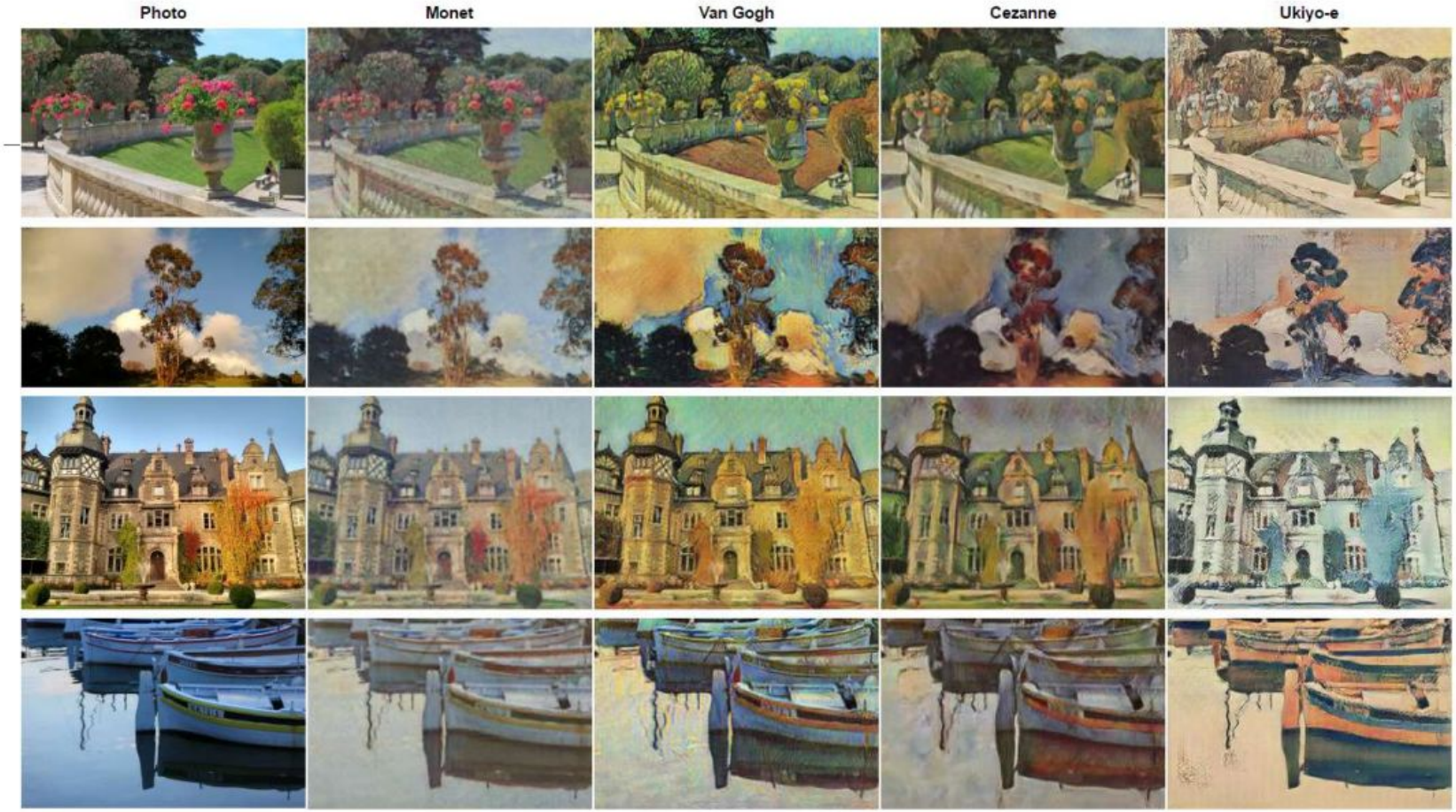
- Srivastava et al. - *Learning from simulated and unsupervised images through adversarial training.*

➢ Photo ↔ Label

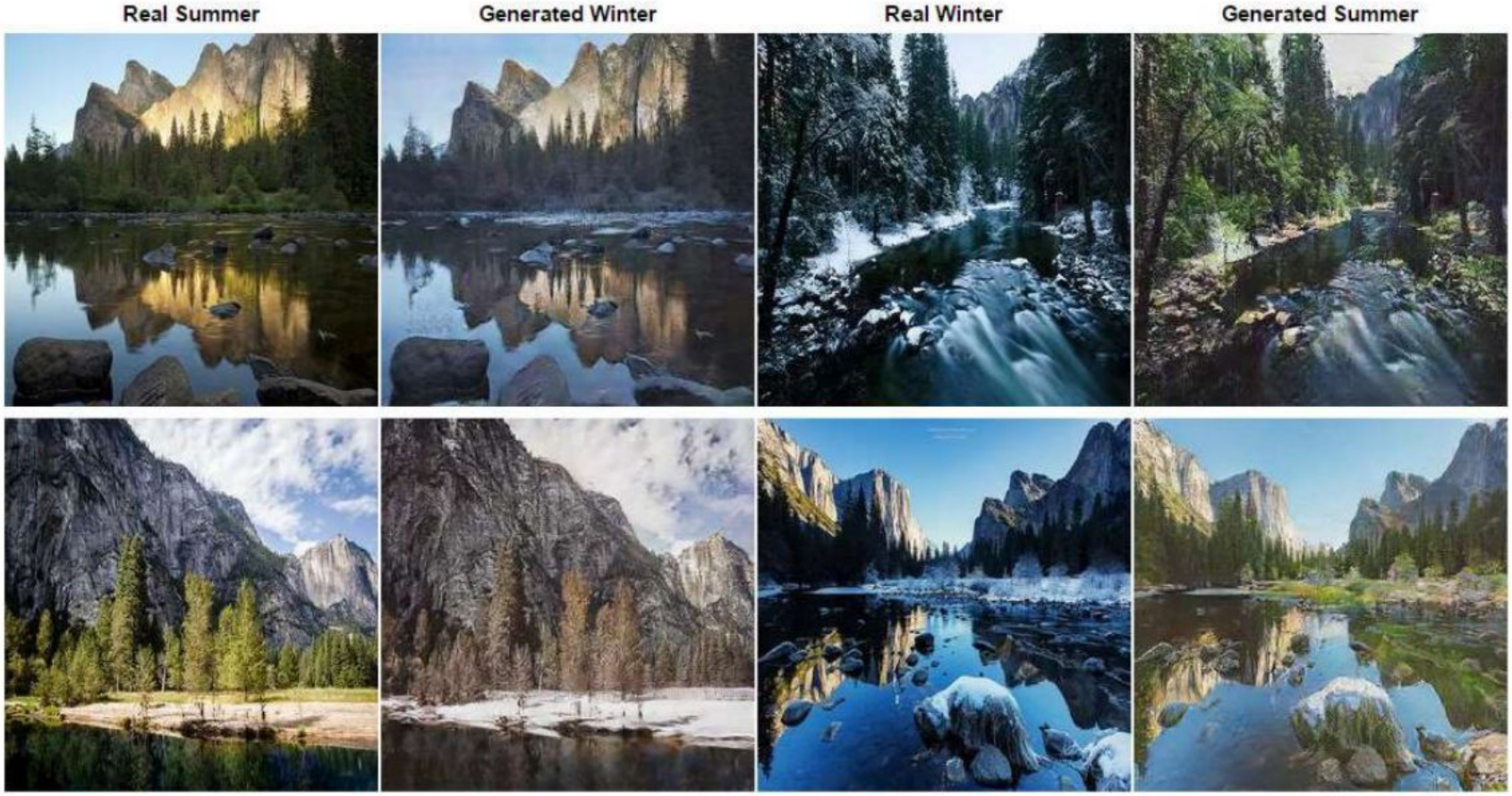

Input
BiGAN
CoGAN
featureGAN
pixelGAN
CycleGAN
pix2pixGAN
Gt

horse ↔ zebra

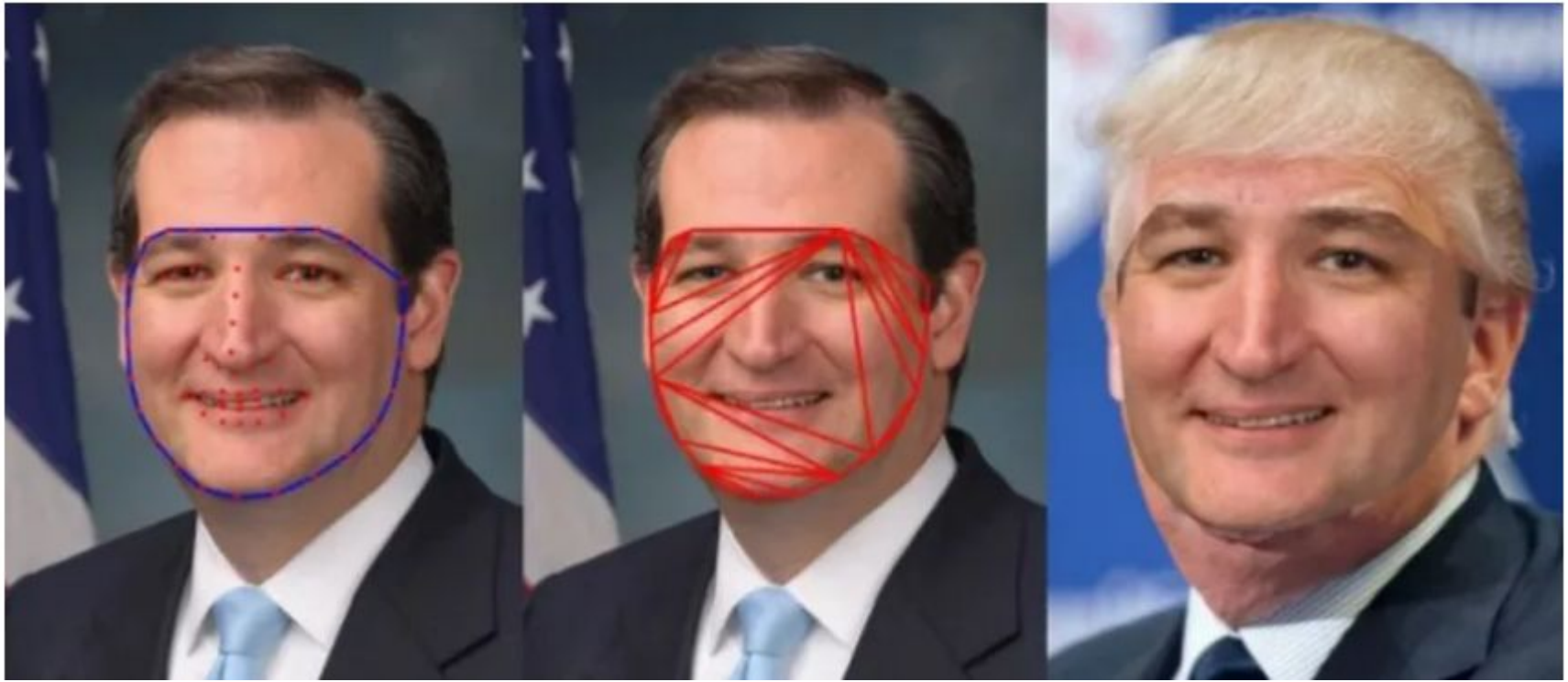
Style Transfer



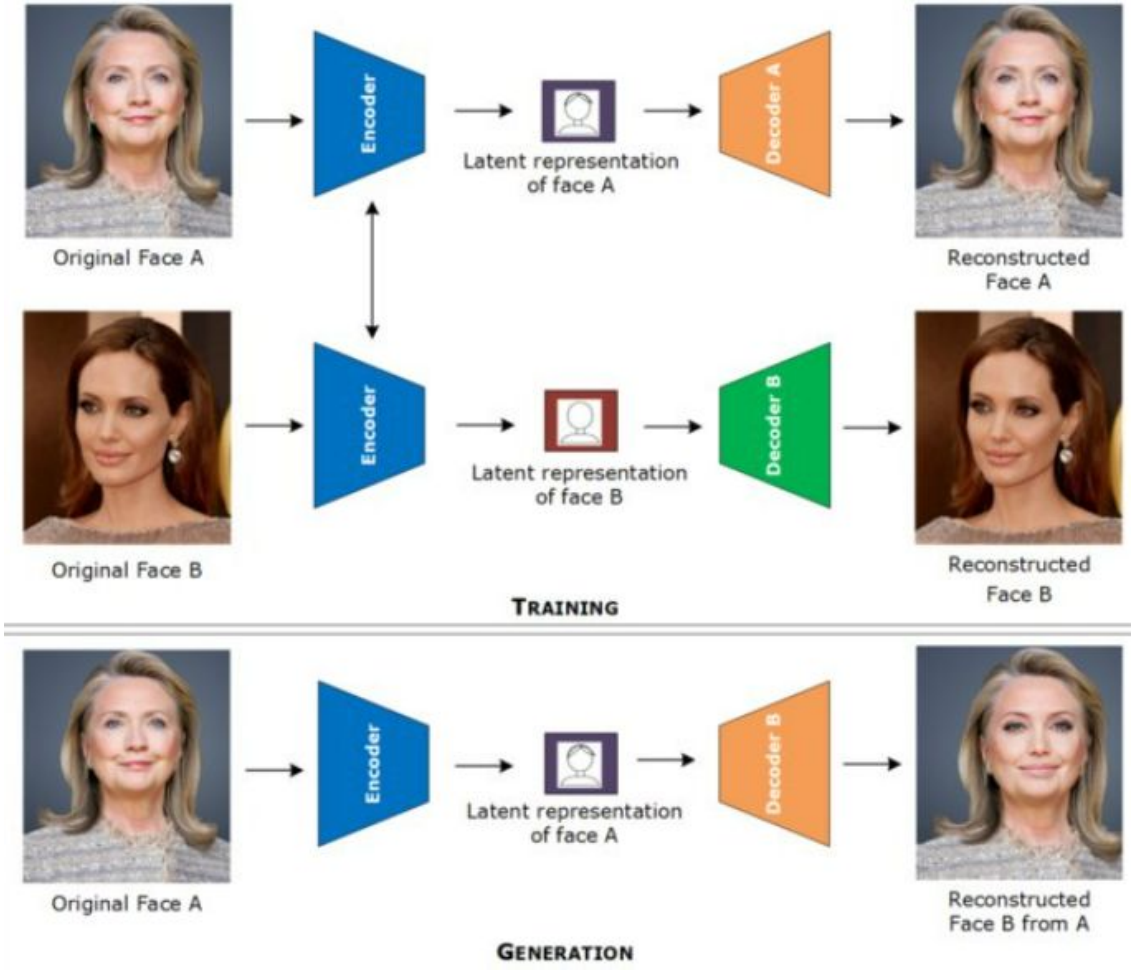
Season Transfer



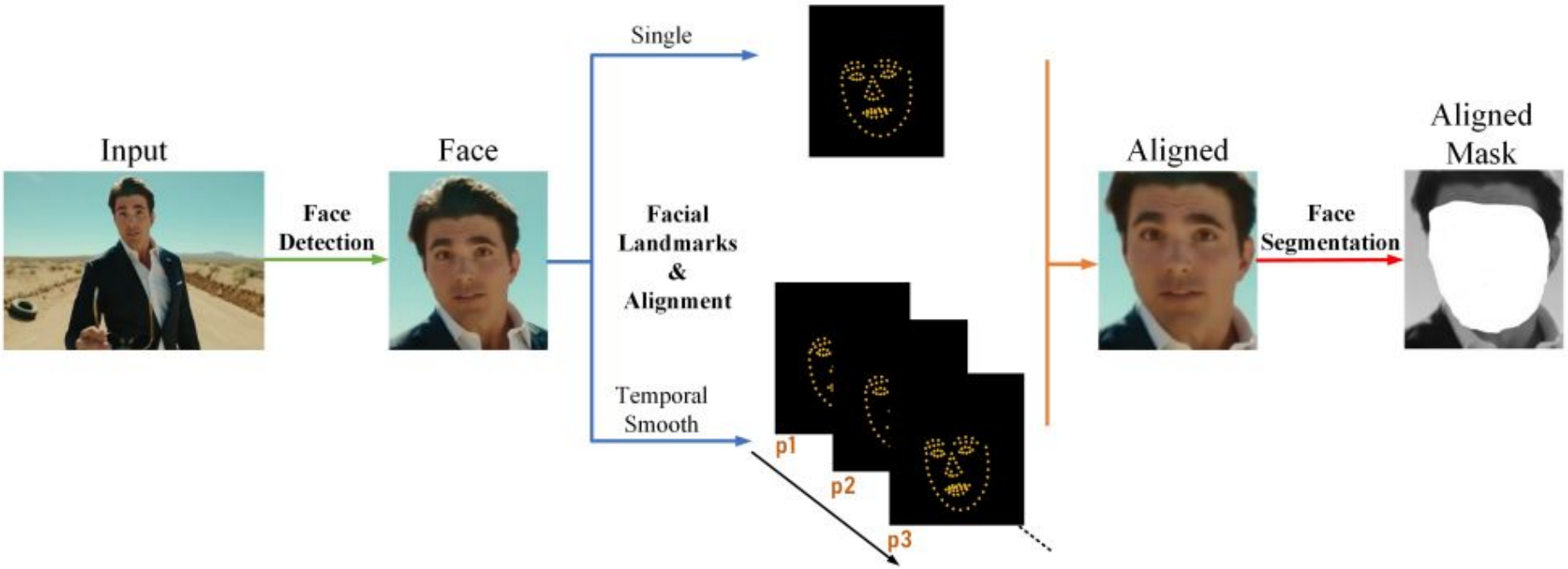
Face Swap



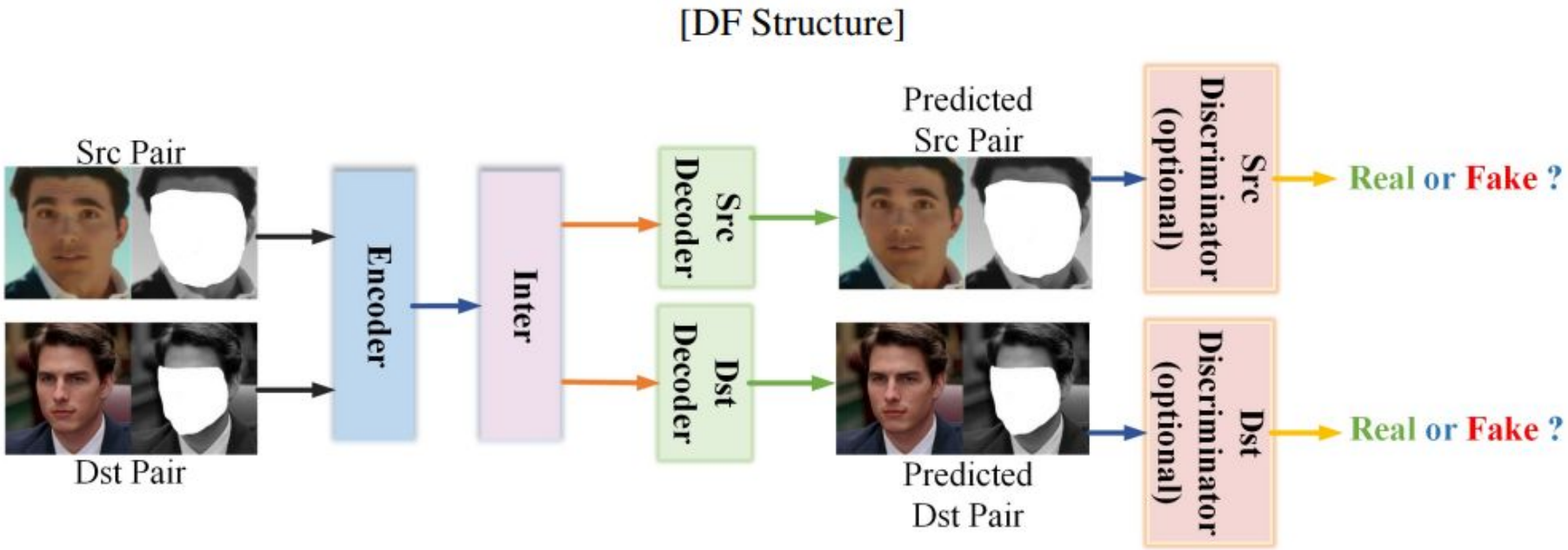
Autoencoders



Deep Face Lab: face extraction

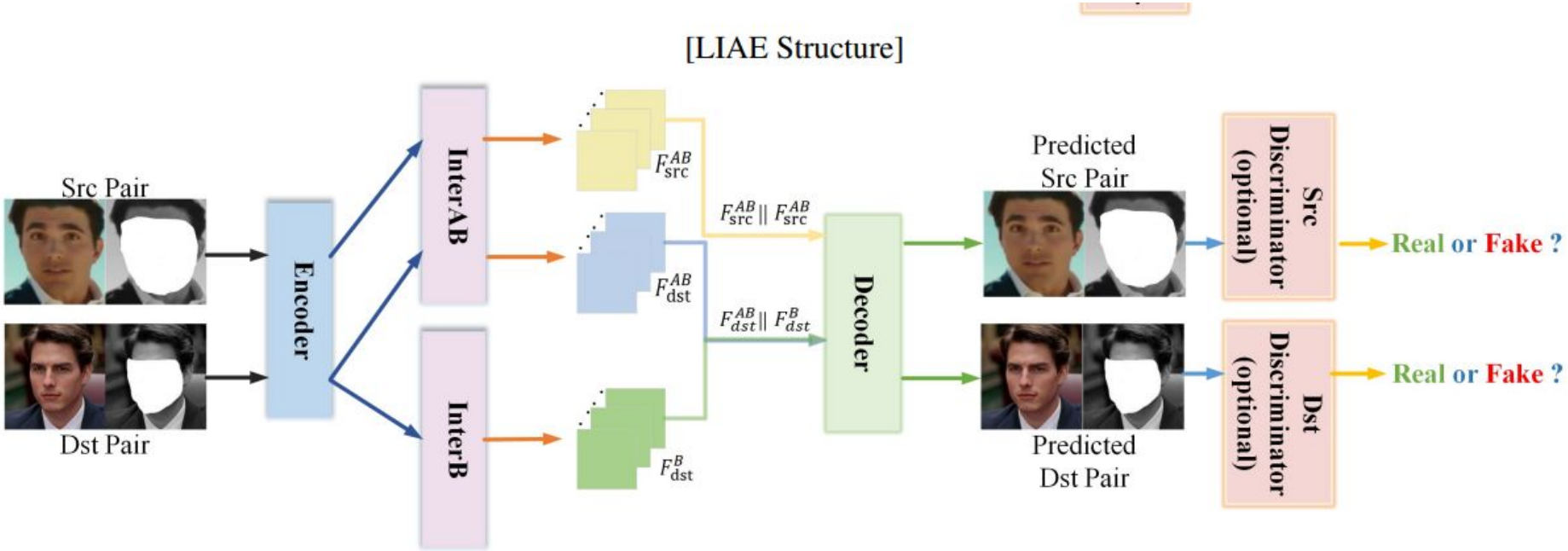


Deep Face Lab: training



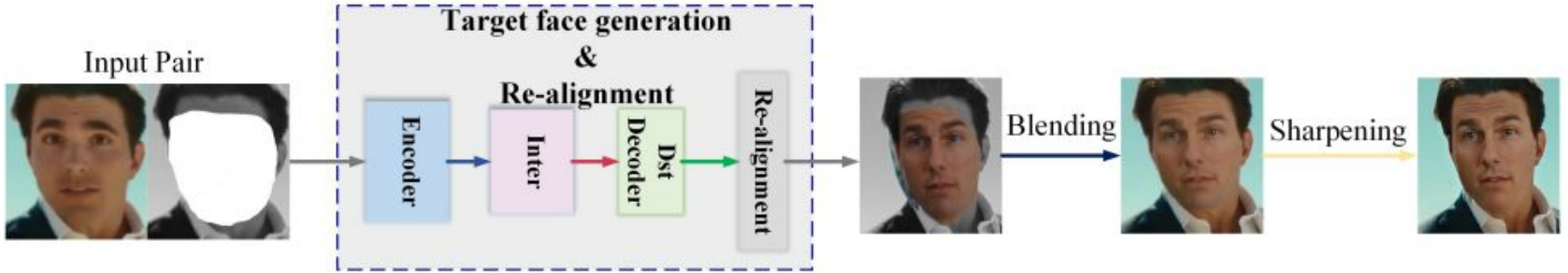
- учим общий Encoder и Inter, два Decoder’а работают в разных доменах
- есть проблема с переносом атрибутов (цветопередачи куда делаем пересадку)

Deep Face Lab: training



- вводим 2 разных Inter’a, InterAB на вход получает оба латента
- общий декодер

Deep Face Lab: conversion



Deep Face Lab

Плюсы

- хорошее качество переноса

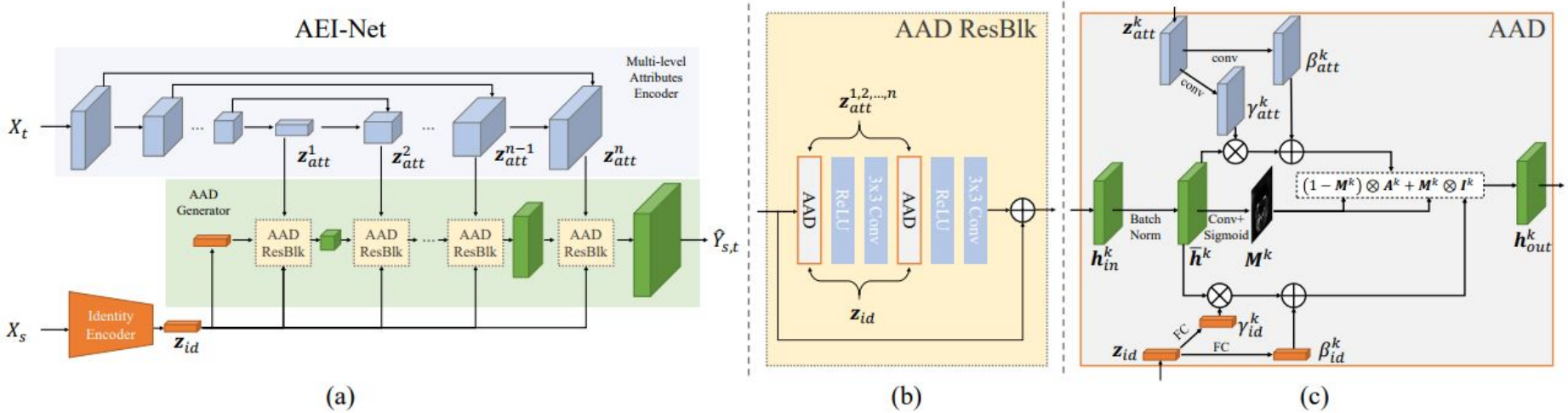
Минусы

- нужно учить под конкретную модель
- нужно много фото модели (от 100 и больше)

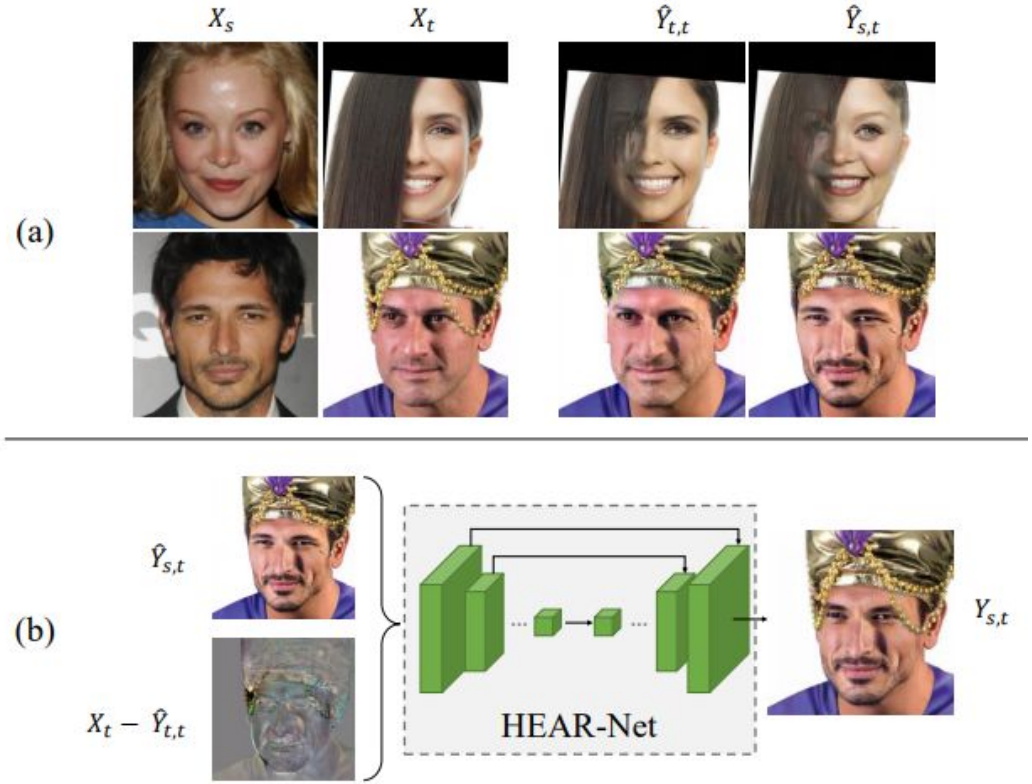
FaceShifter

1. Двухстадийный pipeline:
   a. AEI-Net - заменяет id у target на source, сохраняет atr target
   b. HEAR-Net - шлифует результат пересадки
2. Придумали *Adaptive Attentional Denormalization Generator*, с использованием attention

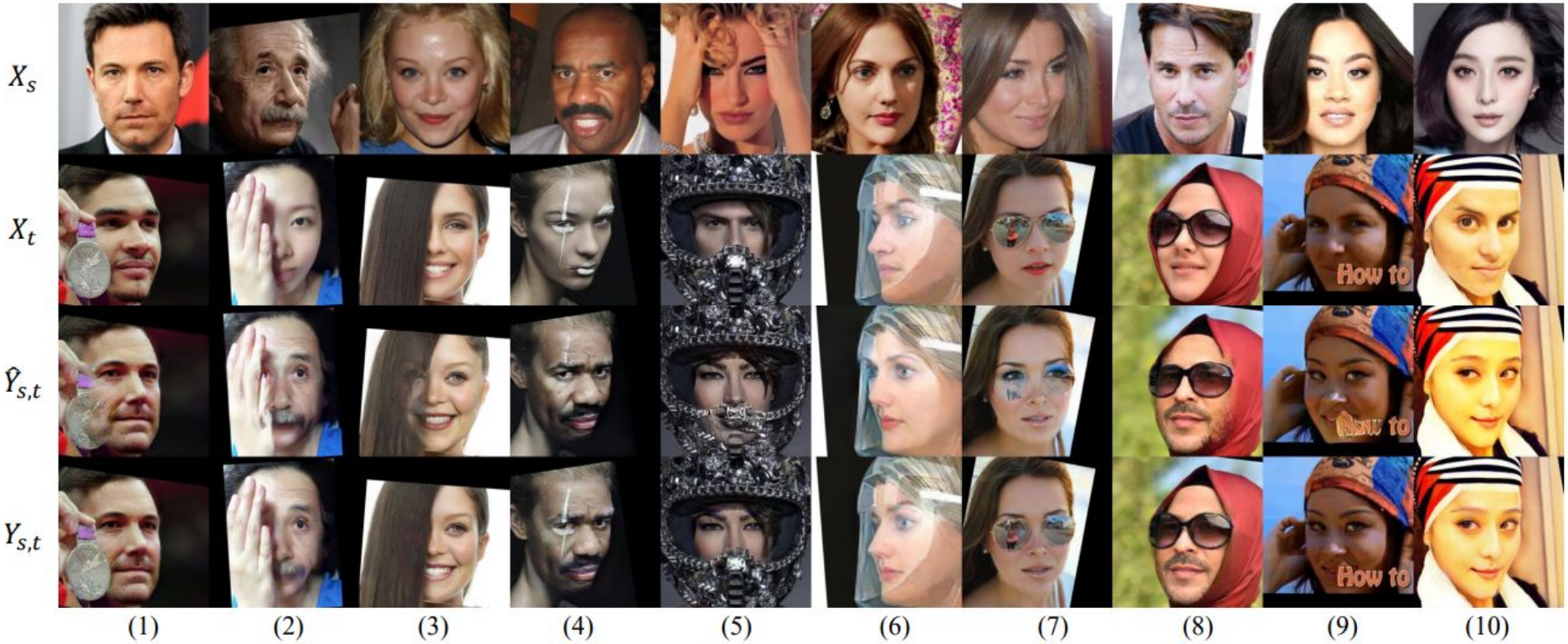
FaceShifter: Adaptive Embedding Integration Network



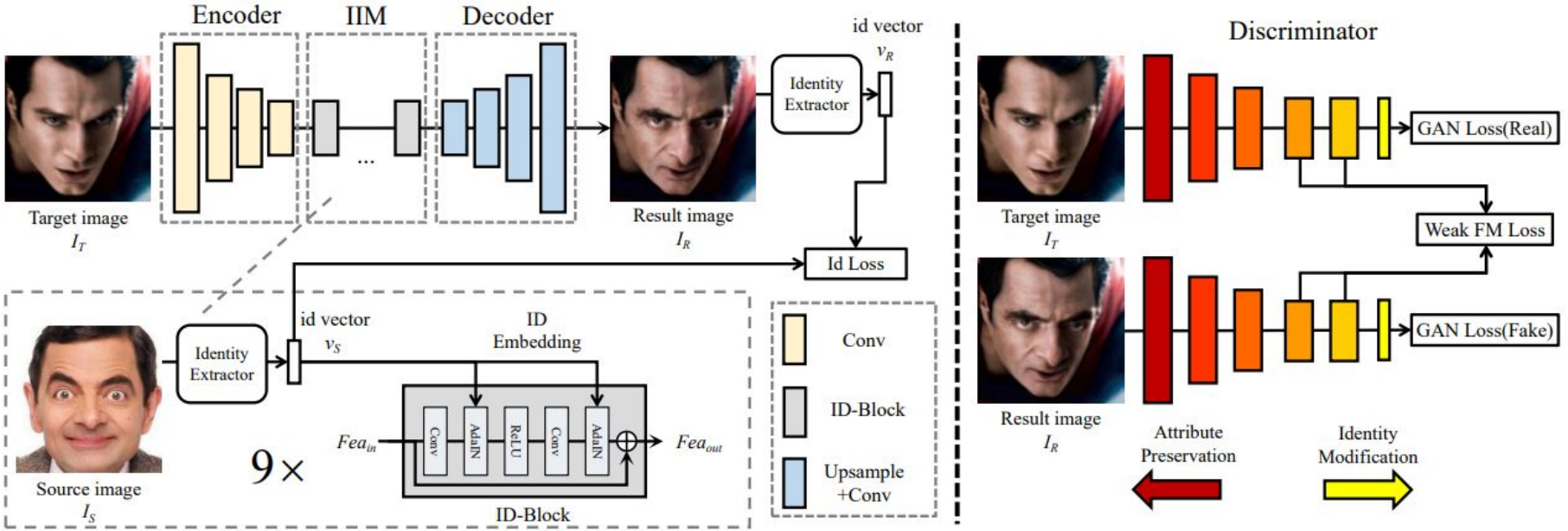
FaceShifter: Heuristic Error Acknowledging Refinement Network



FaceShifter: Heuristic Error Acknowledging Refinement Network



SimSwap


Encoder
IIM
Decoder
id vector
v_R
Identity Extractor
Target image
I_T
Result image
I_R
Id Loss
id vector
v_S
ID Embedding
Identity Extractor
Source image
I_S
9×
Fea_{in}
Conv
AdaIN
ReLU
Conv
AdaIN
Fea_{out}
ID-Block
Conv
ID-Block
Upsample +Conv
Discriminator
GAN Loss(Real)
Weak FM Loss
GAN Loss(Fake)
Target image
I_T
Result image
I_R
Attribute Preservation
Identity Modification

SimSwap

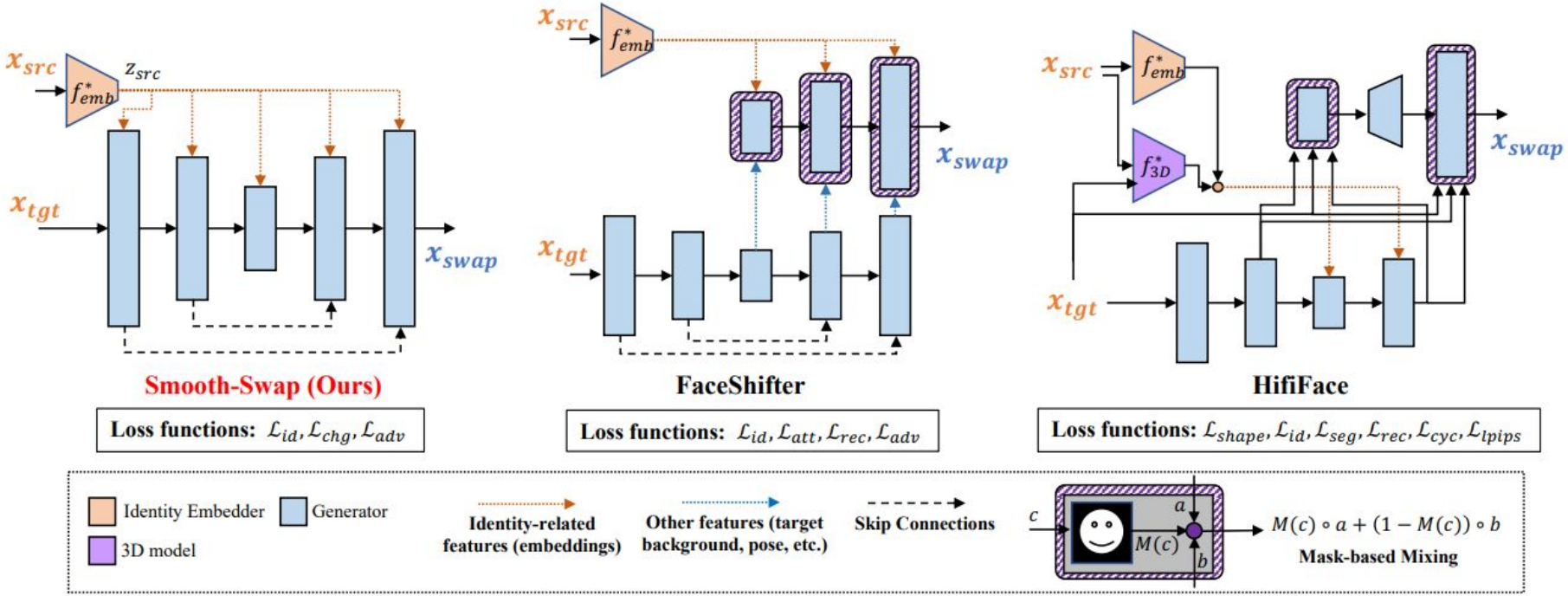
Плюсы

- одна модель на всех
- простая архитектура

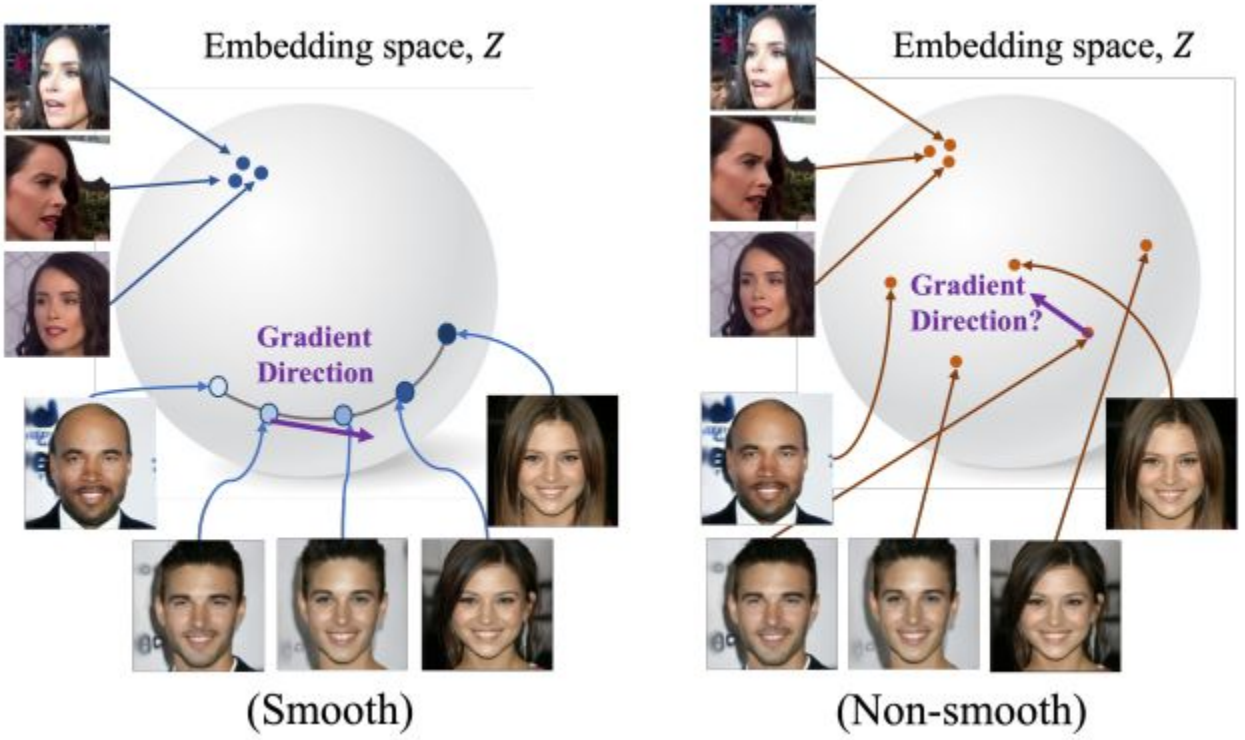
Минусы

- нестабильно переносит id

Smooth-Swap


x_{src}
f^*_{emb}
z_{src}
x_{tgt}
x_{swap}
Smooth-Swap (Ours)
Loss functions: $\mathcal{L}_{id}, \mathcal{L}_{chg}, \mathcal{L}_{adv}$
x_{src}
f^*_{emb}
x_{swap}
x_{tgt}
FaceShifter
Loss functions: $\mathcal{L}_{id}, \mathcal{L}_{att}, \mathcal{L}_{rec}, \mathcal{L}_{adv}$
x_{src}
f^*_{emb}
f^*_{3D}
x_{swap}
x_{tgt}
HifiFace
Loss functions: $\mathcal{L}_{shape}, \mathcal{L}_{id}, \mathcal{L}_{seg}, \mathcal{L}_{rec}, \mathcal{L}_{cyc}, \mathcal{L}_{lpips}$
Identity Embedder
Generator
3D model
Identity-related features (embeddings)
Other features (target background, pose, etc.)
Skip Connections
c
a
$M(c)$
b
$M(c) \circ a + (1 - M(c)) \circ b$
Mask-based Mixing

Smooth-Swap: main idea



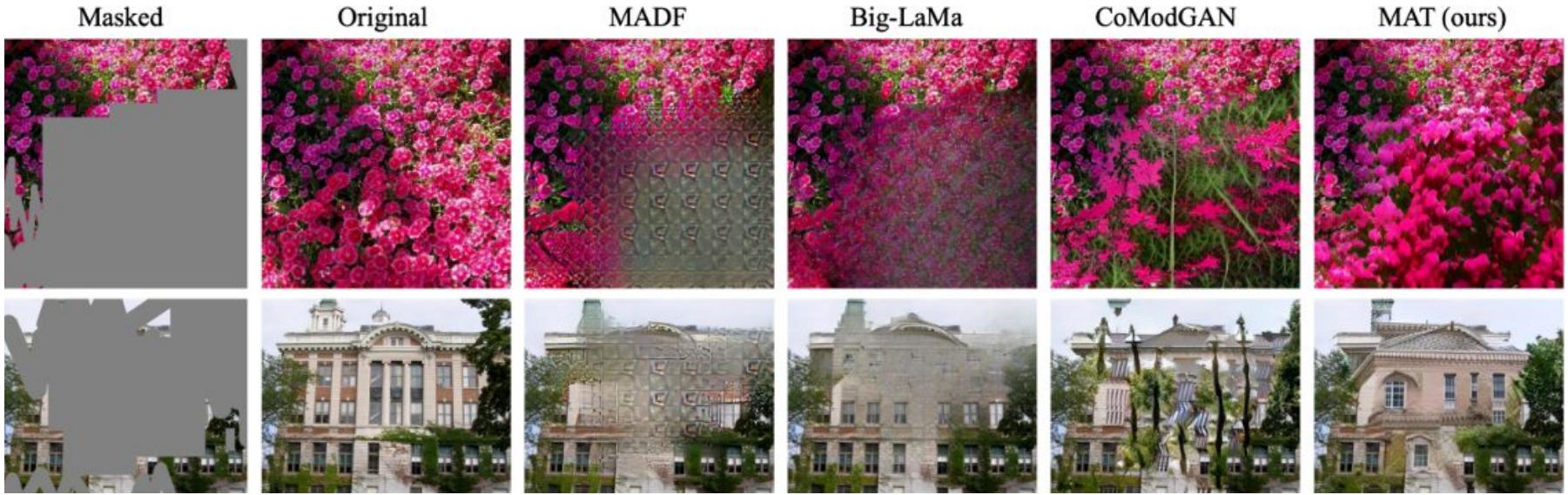
Inpainting and super-resolution

MAT

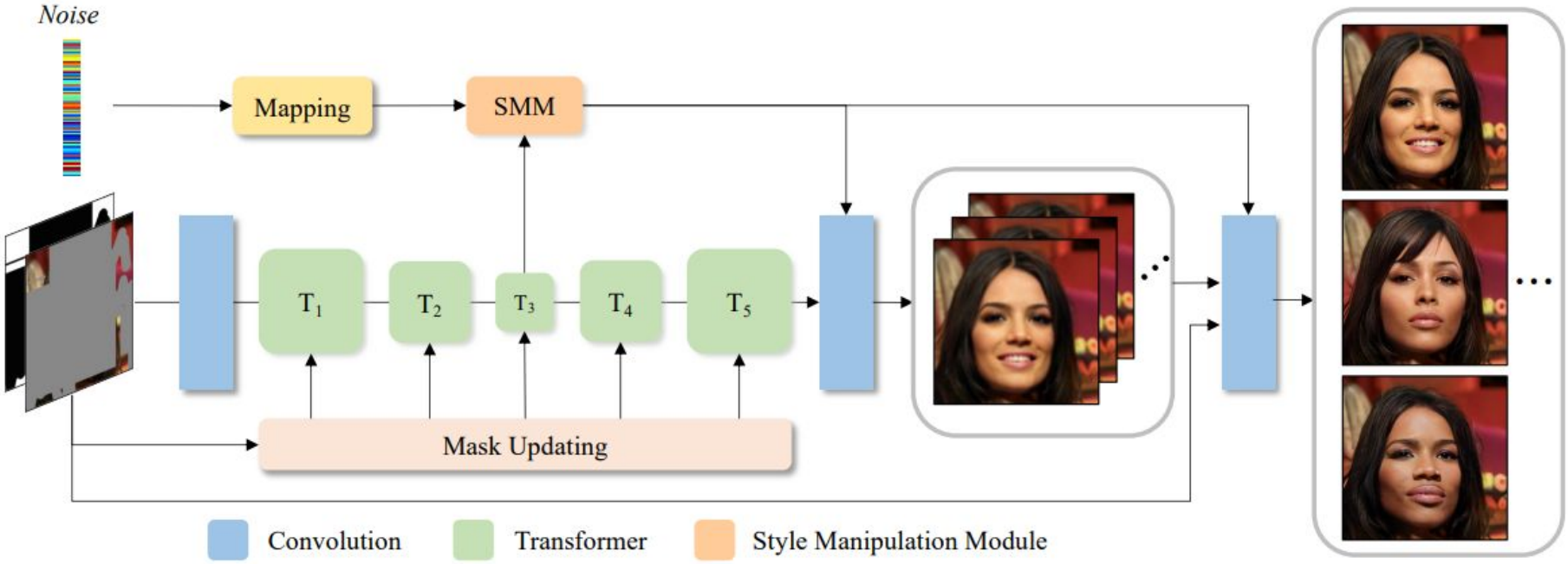
GPEN

SWIN

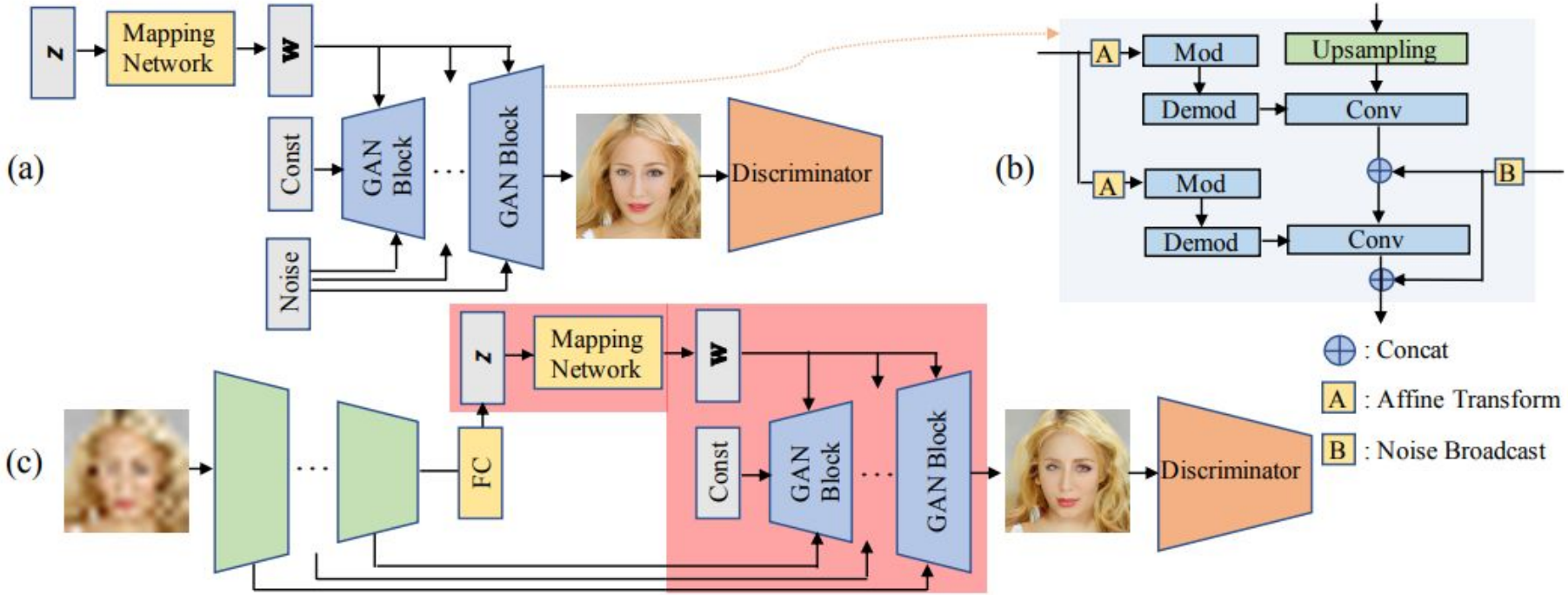
MAT: Mask-Aware Transformer for Large Hole Image Inpainting



MAT: архитектура



GAN Prior Embedded Network for Blind Face Restoration in the Wild



GPEN

- pre-train the GAN prior network using a dataset of HQ face images following the training strategies of StyleGAN
- fine-tune the whole network using a set of synthesized LQ-HQ face image pairs
- the adversarial loss, the content los, feature matching loss

$$I^d = ((I \otimes \mathbf{k}) \downarrow_s + \mathbf{n}_\sigma)_{JPEG_q}, \tag{4}$$

where I, $\mathbf{k}$, $\mathbf{n}_\sigma$, I^d are respectively the input face image, the blur kernel, the Gaussian noise with standard deviation σ and the degraded image. $\otimes$, $\downarrow_s$, $JPEG_q$ respectively denote the two-dimensional convolution, the standard s-fold downsampler and the JPEG compression operator with a quality factor q.

Восстановление видео

Короткий отрезок видео с измененной персоной, сниженным разрешением, и с наложенной маской.

Задача: восстановить закрашенные фрагменты, увеличить разрешение, получить указанную персону (Анжелину Длжоли)

Метрика: покадровый и средний по всем кадрам SSIM и PSNR

Inpainting: MAT, GPEN, SwinIr

Face Swap: SimSwap, Face Shifter