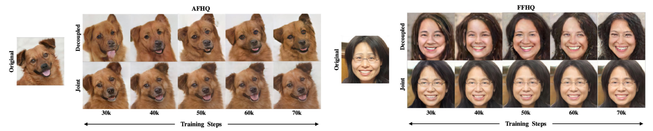
# AE-StyleGAN: Improved Training of Style-Based Auto-Encoders

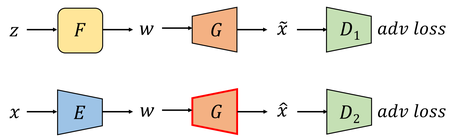
Ligong Han*[1], Sri Harsha Musunuri*[1], Martin Renqiang Min[2], Ruijiang Gao[3], Yu Tian[1], Dimitris Metaxas[1]
[1]Rutgers University [2]NEC Labs America [3]The University of Texas at Austin

WACV WAIKOLOA, HAWAII JANUARY 4-8, 2022



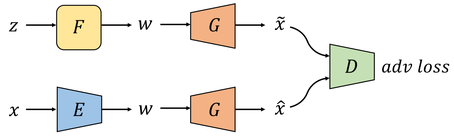

## Introduction


## Decoupled AE-StyleGAN




Decoupled AE-StyleGAN

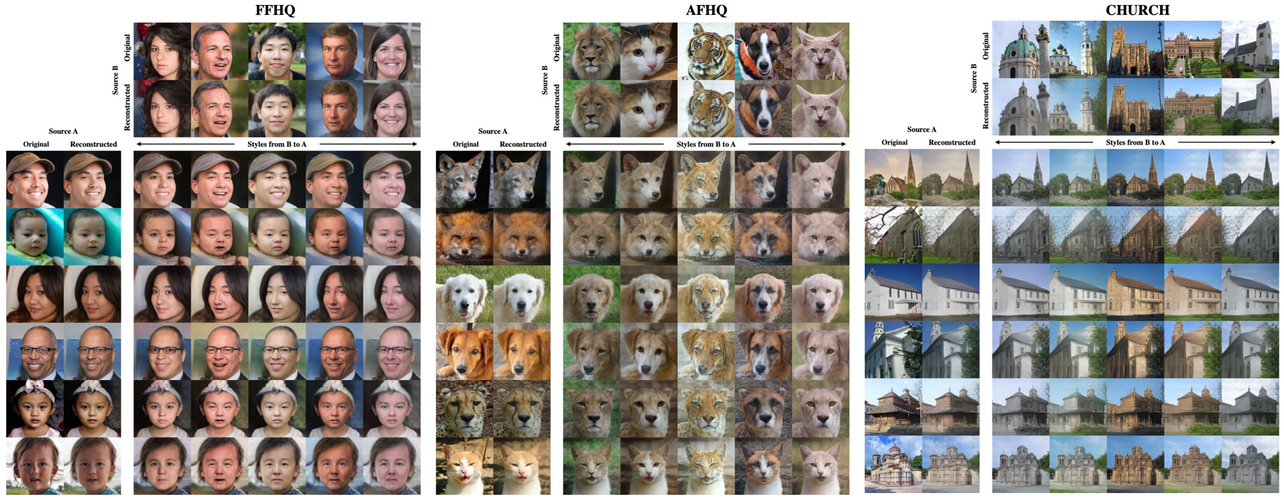
## Joint AE-StyleGAN




Joint AE-StyleGAN

## Experiments

- FFHQ, AFHQ, CHURCH, METFACES – Reconstructions & Style Transfers




## Method

- Loss Functions

$$V_{\text{GAN}}(G \circ F, D) = -\mathbb{E}_{x \sim P_X} \mathcal{A}(-\tilde{D}(x)) - \mathbb{E}_{z \sim P_Z} \mathcal{A}(\tilde{D}(G(F(z)))) \quad (1)$$

**Decoupled AE-StyleGAN.** One straightforward way to train encoder and generator end-to-end is to simultaneously train an encoder with GAN inversion algorithms along with the generator. Here we choose in-domain inversion. To keep the generator's generating ability intact, one can decouple GAN training and GAN inversion training by introducing separate discriminator models $D_1$ and $D_2$, and freezing $G$ in inversion step. Specifically, $D_1$ is involved in $V_{\text{GAN}}(G \circ F, D_1)$ in Equation 1 for GAN training steps, and $D_2$ is involved in $L_{\text{idinv}}(E, D_2, G)$ in Equation 2. Training of $D_2$ follows:

$$V_{\text{GAN}}(G \circ F, D_2) = -\mathbb{E}_{x \sim P_X} \mathcal{A}(-\tilde{D}_2(x)) - \mathbb{E}_{x \sim P_X} \mathcal{A}(\tilde{D}_2(G(E(x)))). \quad (4)$$

**Joint AE-StyleGAN.** The generator of a decoupled AE-StyleGAN would be exactly equivalent to a standard StyleGAN generator, however, we often find the encoder not capable of faithfully reconstruct real images. This phenomenon is illustrated in Figure 5. We hypothesize that with $G$ frozen at inversion step, $E$ cannot catch up with $G$'s update, thus lags behind $G$. To cope with this issue, we propose to train $G$ jointly with $E$ in the inversion step. We also use a single discriminator for both pathways. For the GAN pathway, the value function is written as:

$$V_{\text{AEGAN}}(G \circ F, D) = -\mathbb{E}_{x \sim P_X} \mathcal{A}(-\tilde{D}(x)) - \lambda_{\text{adv}} \mathbb{E}_{x \sim P_X} \mathcal{A}(\tilde{D}(G(E(x)))) - (1 - \lambda_{\text{adv}}) \mathbb{E}_{z \sim P_Z} \mathcal{A}(\tilde{D}(G(F(z)))). \quad (5)$$

## Results

| | StyleGAN | | ALAE | | AE-StyleGAN ($\mathcal{W}$) | | AE-StyleGAN ($\mathcal{W}^+$) | |
|---|---|---|---|---|---|---|---|---|
| | FID ↓ | LPIPS ↑ | FID ↓ | LPIPS ↑ | FID ↓ | LPIPS ↑ | FID ↓ | LPIPS ↑ |
| FFHQ | **7.359** | 0.432 | 12.574 | 0.438 | 8.176 | **0.448** | **7.941** | **0.451** |
| AFHQ | **7.992** | 0.496 | 21.557 | 0.508 | 15.655 | **0.522** | **10.282** | **0.518** |
| MetFaces | **29.318** | 0.465 | 41.693 | 0.462 | 29.710 | **0.469** | **29.041** | **0.471** |
| LSUN Church | **27.780** | 0.520 | 29.999 | 0.552 | 29.387 | **0.603** | **29.358** | **0.592** |

| | StyleGAN | | ALAE | | AE-StyleGAN ($\mathcal{W}$) | | AE-StyleGAN ($\mathcal{W}^+$) | |
|---|---|---|---|---|---|---|---|---|
| | Full | End | Full | End | Full | End | Full | End |
| FFHQ | **173.09** | **173.68** | 192.60 | 193.94 | 181.03 | 180.23 | **166.70** | **165.85** |
| AFHQ | 244.83 | 240.68 | **229.54** | **232.53** | 247.75 | 248.75 | **233.86** | **231.91** |
| MetFaces | **231.40** | **232.77** | **235.39** | 237.63 | 238.84 | 238.60 | 240.01 | **235.41** |
| LSUN Church | 245.22 | **239.62** | 298.06 | 295.01 | **241.46** | 231.73 | **240.01** | **231.49** |

## Algorithm

**Algorithm 1** Decoupled AE-StyleGAN Training

$\theta_E, \theta_{D_1}, \theta_{D_2}, \theta_F, \theta_G \leftarrow$ Initialize network parameters
**while** not converged **do**
  $x \leftarrow$ Random mini-batch from dataset
  $z \leftarrow$ Samples from $\mathcal{N}(0, I)$
  Step I. Update $D_1$, $D_2$
  $L_D \leftarrow -V_{\text{GAN}}(G \circ F, D_1) - V_{\text{GAN}}(G \circ E, D_2)$ in 1
  $\theta_{D_1}, \theta_{D_2} \leftarrow \text{ADAM}(\nabla_{\theta_{D_1}, \theta_{D_2}} L_D, \theta_{D_1}, \theta_{D_2})$
  Step II. Update $E$
  $L_E \leftarrow L_{\text{idinv}}(E, D_2, G)$ in 2 or 6
  $\theta_E \leftarrow \text{ADAM}(\nabla_{\theta_E} L_E, \theta_E)$
  Step III. Update $F$, $G$
  $L_G \leftarrow V_{\text{GAN}}(G \circ F, D_1)$ in 1
  $\theta_F, \theta_G \leftarrow \text{ADAM}(\nabla_{\theta_F, \theta_G} L_G, \theta_F, \theta_G)$
**end while**

**Algorithm 2** Joint AE-StyleGAN Training

$\theta_E, \theta_D, \theta_F, \theta_G \leftarrow$ Initialize network parameters
**while** not converged **do**
  $x \leftarrow$ Random mini-batch from dataset
  $z \leftarrow$ Samples from $\mathcal{N}(0, I)$
  Step I. Update $D$
  $L_D \leftarrow -V_{\text{AEGAN}}(G \circ F, D)$ in 5
  $\theta_D \leftarrow \text{ADAM}(\nabla_{\theta_D} L_D, \theta_D)$
  Step II. Update $E$, $G$
  $L_E \leftarrow L_{\text{idinv}}(E, D, G)$ in 2 or 6
  $\theta_E, \theta_G \leftarrow \text{ADAM}(\nabla_{\theta_E, \theta_G} L_E, \theta_E, \theta_G)$
  Step III. Update $F$, $G$
  $L_G \leftarrow V_{\text{AEGAN}}(G \circ F, D)$ in 5
  $\theta_F, \theta_G \leftarrow \text{ADAM}(\nabla_{\theta_F, \theta_G} L_G, \theta_F, \theta_G)$
**end while**

## Summary

In this paper, we proposed AE-StyleGAN, a novel algorithm that jointly trains an encoder with a style-based generator. With empirical analysis, we confirmed that this methodology provides an easy-to-invert encoder for real image editing. Extensive results showed that our model has superior image generation and reconstruction capability than baselines. We have explored the problem of training an end-to-end autoencoder. With improved generation fidelity and reconstruction quality, the proposed AE-StyleGAN model can serve as a building-block for further development and applications.