

Few-shot adaptation of generative adversarial networks

Robb, Esther, et al. "Few-shot adaptation of generative adversarial networks." *arXiv preprint arXiv:2010.11943* (2020).

ABSTRACT

INTRODUCTION

gan을 학습하는데 있어 주요 문제점

1. 각 시각적 domain에 맞는 크고 다양한 dataset이 필요함
2. adversarial optimization은 loss function의 최적해가 아닌, saddle point를 표현하며 거짓말을 침.

데이터 수가 적을 시에 gan은 memorization이나 instability문제를 야기함

개선 방안

- transfer learning을 통한 gan의 sample efficiency를 향상
데이터 수가 적은 경우에 target domain과 상관 없는 영상 생성
- 파라미터를 작게 하도록 optimization을 제한하는 batchnorm 사용
low shot setting에선 gan base의 optimization보다 mle base의 optimization이 좋은 결과를 내었
다고 함.
하지만, 영상이 blur하다던가, 상세한 표현을 하지 못한다.

이 논문에서는 사전학습된 gan의 가중치들을 singular value decomposition(SVD)을 적용하여 행렬분
해를 한 후, 고정된 left/right singular vectors를 대상으로 gan optimization을 적용하여 singular
values를 조정함.

BACKGROUND

SVD



원본 이미지



50개 singular value로 근사한 이미지

FEW-SHOT GAN

ADAPTATION PROCEDURE

$$\begin{aligned}
 G^{(l)} &= \text{generator layer} \\
 D^{(l)} &= \text{discriminator layer} \\
 G^{(l)}, D^{(l)} &\in \mathbb{R}^{c_{in} \times c_{out}} \text{ or } \mathbb{R}^{k \times k \times c_{in} \times c_{out}} \\
 &\mathbb{R}^{k \times k \times c_{in} \times c_{out}} \rightarrow \mathbb{R}^{k^2 c_{in} \times c_{out}}
 \end{aligned}$$

g, d 의 layer들은 fully connected weights(c_{in}, c_{out})나 convolutional filter weights(k, k, c_{in}, c_{out})의 형태의 weight를 가진다.

SVD는 2D matrix에서 수행 가능하므로 fully connected weights는 바로 적용 가능하지만, convolutional filter weights는 4D이므로 $k^2(c_{in}, c_{out})$ 의 2D로 재구성한다.

$$W_{\Sigma} = U_0 \begin{bmatrix} \sigma_1 & 0 & \dots \\ 0 & \sigma_2 & \dots \\ \vdots & \ddots & \\ 0 & & \sigma_s \end{bmatrix} V_0^{\top}$$

$k^2 c_{in} \times s$ $s \times s$ $s \times c_{out}$

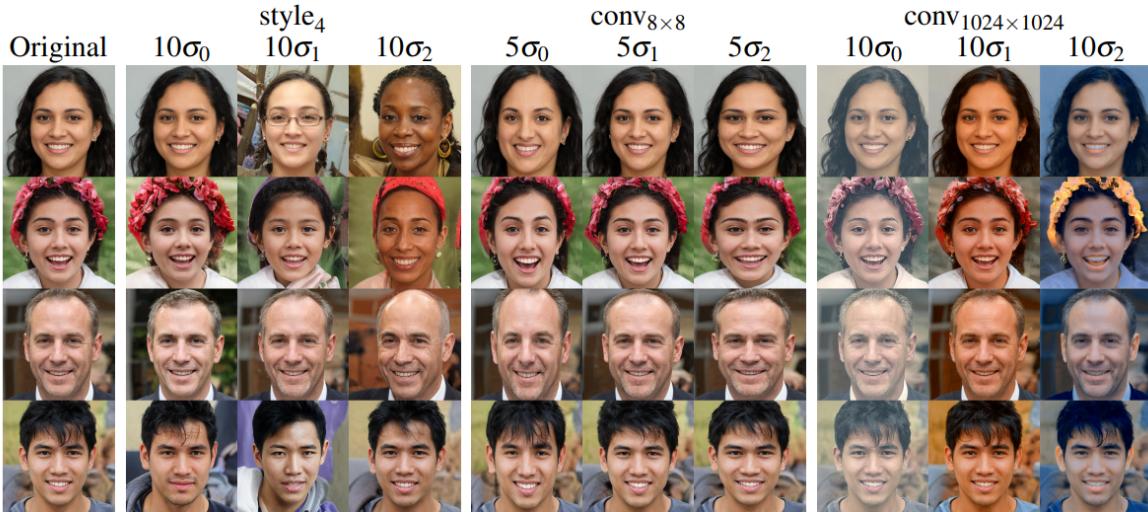
(b) FSGAN singular value adaptation.

이를 통해 구성한 가중치 w 를 SVD를 사용하여 행렬분해를 한다.

$$W_0^{(l)} = (U_0 \sum_0 V_0^T)^{(l)}$$

$$\sum = \lambda \sum_0, \quad W_{\sum}^{(l)} = (U_0 \sum_0 V_0^T)^{(l)}$$

forward propagation 할 때, 기존 가중치를 위 식처럼 재구성하여 진행하게 된다.



기존 stylegan2에서 각 layer에 실험한 결과로, svd 수행시에 singular values 중 상위 3개를 뽑아서 수정한 결과이다. 각 layer는 좌측부터 나이나 피부톤, 얼굴의 크기, 전체적인 채도, 대비 같이 자연스럽게 이미지가 변화되는 것을 관찰할 수 있다.

TRAINING & INFERENCE

t	FID	Interpolation
0	121.21	
20	154.25	
40	134.22	
80	102.87	
120	93.65	
180	92.94	
Train set (10-shot):		

위 이미지는 10개의 train set과 train image를 보여준 횟수(t)에 관한 실험 결과로, t에 커짐에 따라 fid는 작아짐을 볼 수 있지만, train set을 거의 암기한 것과 같은 현상을 볼 수 있다. 그렇기 때문에 low-shot setting에서는 생성되는 영상의 품질과 다양성을 위해 학습 시간을 줄이는 것이 필요함을 알 수 있다.

본 논문의 실험은 stylegan2으로 진행하였고, singular value를 재학습하였으며, inference에는 input noise를 $N(0,1)$ 에서 sampling한 후 threshold 값 이상인 sample은 다시 resampling 하는 기법인 truncation trick을 적용하였다고 한다.

EVALUATION IN FEW-SHOT SYNTHESIS

기존의 gan은 training set에 대해 FID를 evaluation metrics로 사용하지만 low-shot setting에서는 고품질의 이미지 생성을 위해 training set을 암기해버렸을 경우가 있으므로 권장되지 않는다.

$$FID = d^2((m, C), (m_w, C_w)) = ||m - m_w||_2^2 + Tr(C + C_w - 2\sqrt{(CC_w)})$$

m=평균, C=특징벡터의 공분산 행렬

gan의 evaluation metrics는 생성되는 이미지의 다양성, training data를 암기했는지에 대한 패널티가 고려되어야 한다.

기존의 gan은 training set에 대해 FID를 evaluation metrics로 사용하지만 low-shot setting에서는 고품질의 이미지 생성을 위해 training set을 암기해버렸을 경우가 있으므로 권장되지 않는다.

그렇기 때문에 FID 대신, training step을 제한한다.

EXPERIMENTS

NEAR-DOMAIN ADAPTATION

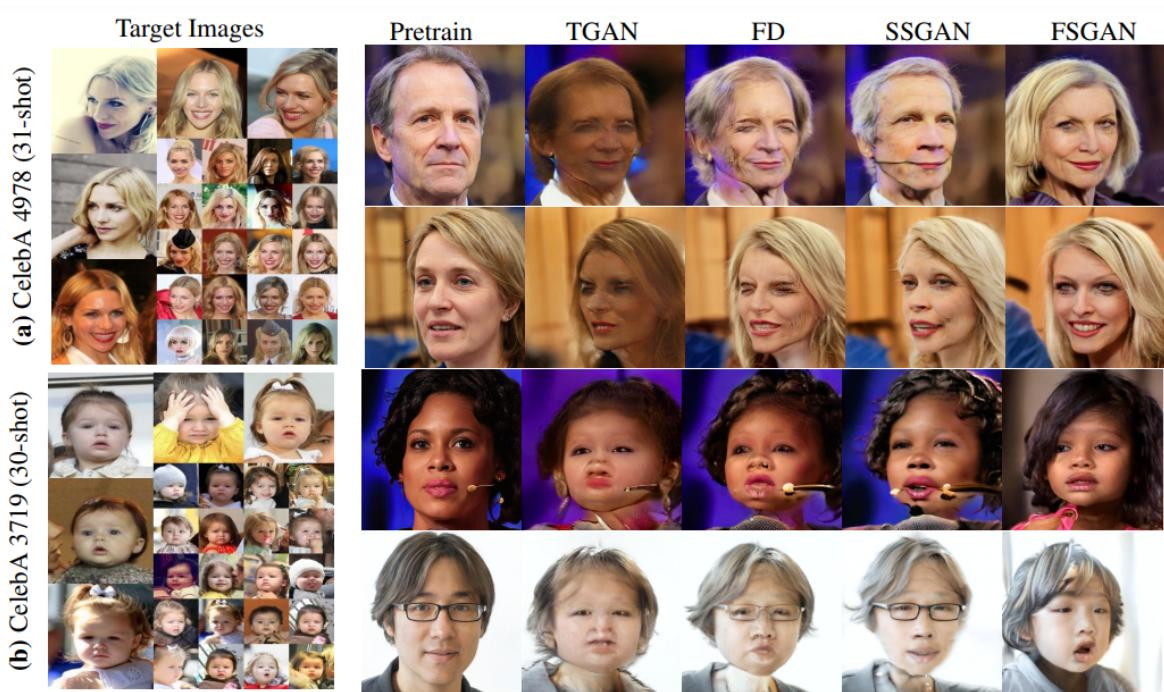


Figure 5: **Close-domain adaptation** (FFHQ→CelebA). Models adapted from a pretrained StyleGAN2 using ~ 30 target images (left-most column) of (a) CelebA ID 4978 and (b) CelebA ID 3719. The proposed FSGAN generates more natural face images without noticeable artifacts. Comparison methods include TGAN (Wang et al., 2018), FD (Mo et al., 2020), SSGAN (Noguchi & Harada, 2019), trained with a limited number of timesteps to prevent overfitting or degradation.

Table 1: **Quantitative comparisons** in three metrics: FID (Heusel et al., 2017b), Face Quality Index (FQI) (Hernandez-Ortega et al., 2019), and sharpness (Kumar et al., 2012). See Fig 5 for illustrations. FQI and Sharpness are evaluated on 1,000 images randomly generated with the same set of seeds. Bracketed/bold numbers indicated the best/second best results, respectively.

Method	CelebA 4978			CelebA 3719		
	FID	FQI	Sharpness	FID	FQI	Sharpness
Pretrain	—	0.40 ± 0.11	0.91 ± 0.06	—	0.37 ± 0.12	0.92 ± 0.06
TransferGAN	75.41	0.30 ± 0.07	0.61 ± 0.05	178.31	0.26 ± 0.09	0.61 ± 0.04
FreezeD	75.30	0.33 ± 0.09	0.58 ± 0.04	143.83	0.27 ± 0.09	0.56 ± 0.05
SSGAN	87.79	0.32 ± 0.08	$[0.67 \pm 0.05]$	147.14	0.27 ± 0.10	0.58 ± 0.05
FSGAN (ours)	78.90	$[0.36 \pm 0.07]$	0.65 ± 0.05	170.00	0.27 ± 0.08	$[0.68 \pm 0.07]$

위 사진은 near-domain transfer를 실험한 것으로, target domain과 source domain이 둘 다 실제 사람과 같이 domain이 유사한 것들을 의미한다.

위 사진과 아래의 테이블을 보면 fsgan이 전반적으로 더 자연스러운 얼굴을 생성하지만, FID가 제일 낮지 않음을 볼 수 있고, 이는 FID와 정성평가간의 상관관계가 적다는 것을 알 수 있다.

FAR-DOMAIN ADAPTATION

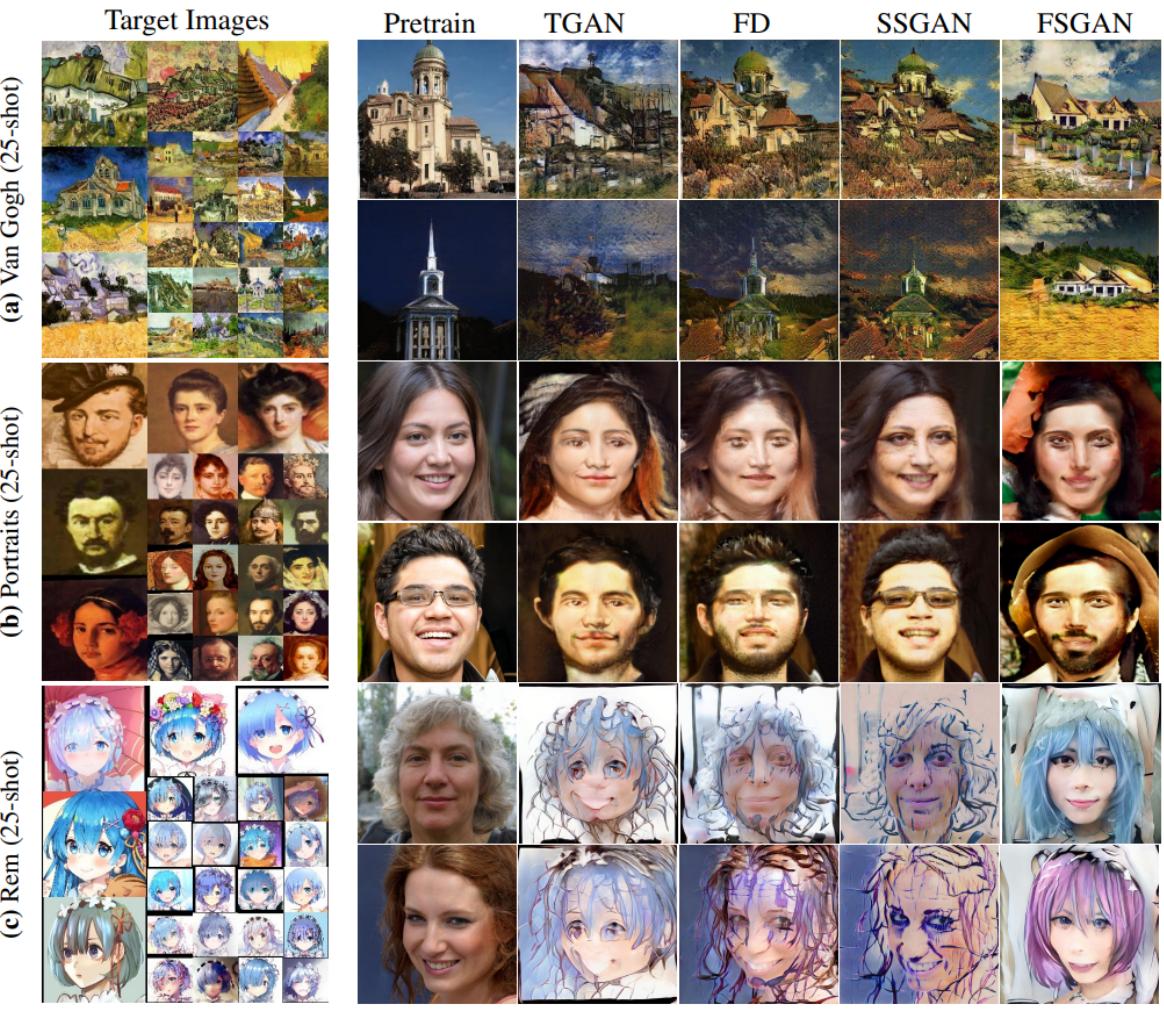


Figure 6: **Far-domain adaptation** (Photo→Art). Comparing FSGAN with alternative GAN adaptation methods in the photo-to-art setting. **(a)** FSGAN more effectively alters building layouts and adds landscape in the foreground to match the Van Gogh paintings, maintaining better spatial coherency. **(b)** FSGAN adopts features from the Portraits dataset (hats, beards, artistic backgrounds), while other methods primarily alter image textures. **(c)** FSGAN transforms natural hair and facial features to imitate the anime target while retaining spatial consistency. Note the occurrence of pink hair in our generated images, which does not exist in the few-shot target but is visually consistent.

위 사진은 far-domain에 관한 실험으로 건물 모양이 다르다던지, 실제 사진에서 그림으로 변환하는 것을 보여준다.