RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

2021 ICCV OCTOBER 11-17 VIRTUAL

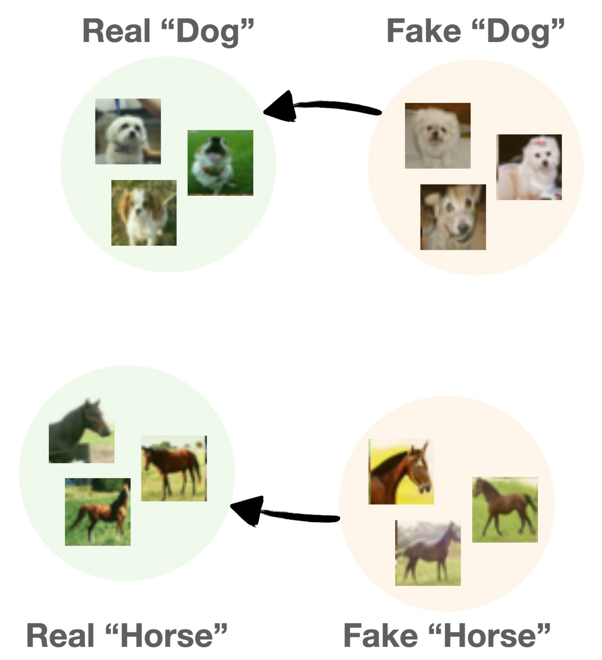
Dual Projection Generative Adversarial Networks for Conditional Image Generation

Ligong Han[1], Martin Renqiang Min[2], Anastasis Stathopoulos[1], Yu Tian[1], Ruijiang Gao[3], Asim Kadav[2], and Dimitris Metaxas[1]

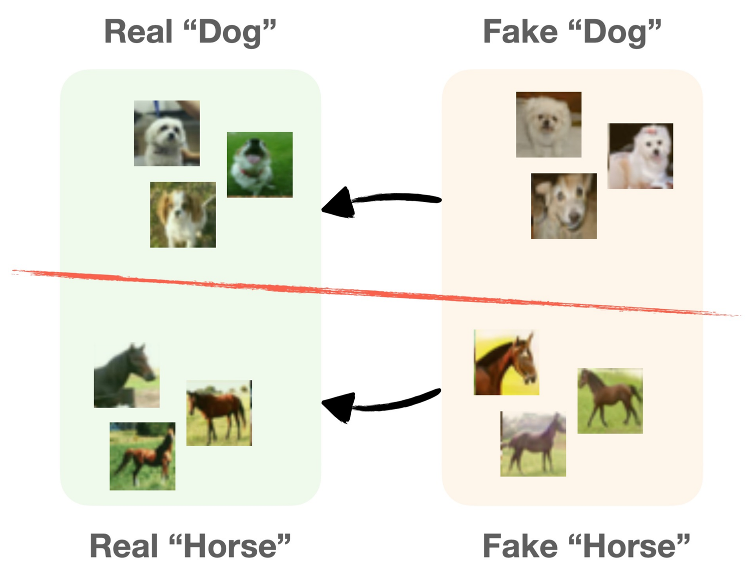
[1]Rutgers University [2]NEC Labs America [3]The University of Texas at Austin

Introduction

- Data Matching and Label Matching
  - Conditional Data Matching



Marginal Matching + Label Matching



- Conditional GAN
  - Objectives

$$L_D = \mathbb{E}_{x,y\sim P_{XY}}\mathcal{A}(-\tilde{D}(x,y)) + \mathbb{E}_{z\sim P_Z, y\sim Q_Y}\mathcal{A}(\tilde{D}(G(z,y),y))$$

$$L_G = \mathbb{E}_{z\sim P_Z, y\sim Q_Y}\mathcal{A}(-\tilde{D}(G(z,y),y))$$

$\mathcal{A}$ is *activation function*, $\tilde{D}$ is the *logit*. Vanilla GAN, $\mathcal{A}(t) = softplus(t) = \log(1+\exp(t))$.

  - Decomposition

$$\tilde{D}^*(x,y) = \underbrace{\log\frac{P(x)}{Q(x)}}_{\text{Marginal Matching}} + \underbrace{\log\frac{P(y|x)}{Q(y|x)}}_{\text{Label Matching}}$$

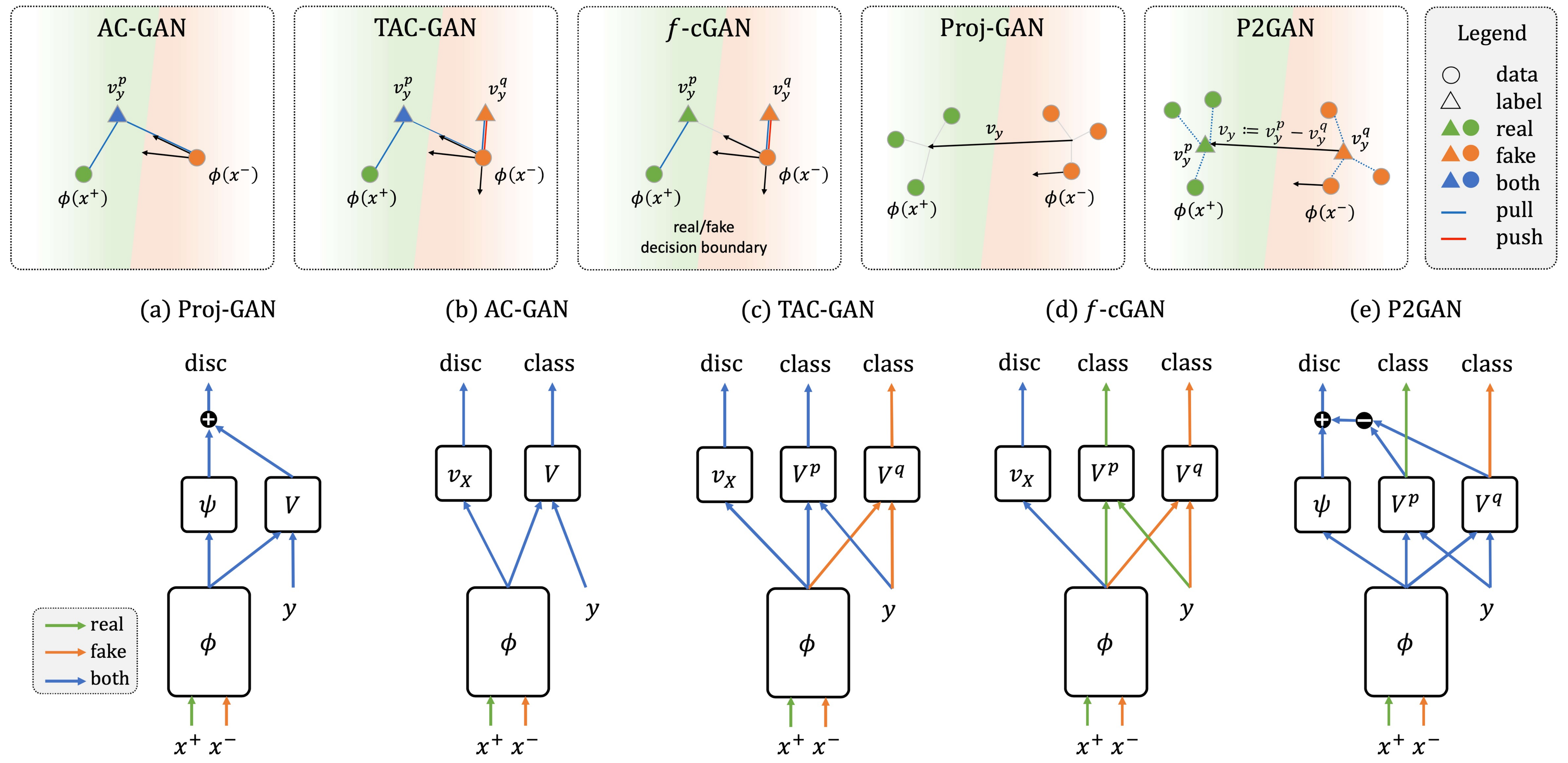
$$= \underbrace{\log\frac{P(x|y)}{Q(x|y)}}_{\text{Data Matching}} + \log\frac{P(y)}{Q(y)}$$

- Projection GAN (*Miyato 2018*)

$$\tilde{D}(x,y) = v_y^T\phi(x) + \psi(\phi(x))$$

$\phi(\cdot)$ is the image embedding function, v_y is embedding of class. $v_y := v_y^p - v_y^q$ is the difference of real and fake class embeddings.

Dual Projection GAN (P2GAN)



- Objective

$$\tilde{D} = \underbrace{(v_y^p - v_y^q)^T}_{\text{Untying Class Embeddings}}\phi(x) + \psi(\phi(x))$$

$$L_D^{P2} = L_D(\tilde{D}) + \underbrace{L_{mi}^p + L_{mi}^q}_{\text{Classification Losses}}$$

$$L_G^{P2} = L_G(\tilde{D})$$

- Weighted Dual Projection GAN (P2GAN-w)

$$L_D^{P2w} = L_D + \lambda\cdot(L_{mi}^p + L_{mi}^q)$$

Theoretical Analysis

Proposition 1. *When $\psi = 0$, a Proj-GAN reduces to K unconditional GANs, each of them minimizes the Jensen-Shannon divergence between $P_{X|y}$ and $Q_{X|y}$ with mixing ratio $\{\frac{P(y)}{P(y)+Q(y)}, \frac{Q(y)}{P(y)+Q(y)}\}$. Its value function can be written as,*

$$\mathbb{E}_{P_Y}\left\{\mathbb{E}_{P_{X|Y}}\log D(x|y) + \frac{Q_y}{P_y}\mathbb{E}_{Q_{X|Y}}\log(1-D(x|y))\right\}.$$

Proposition 2. *Given a generator G, if cross entropy losses L_{mi}^p and L_{mi}^q are minimized optimally, then the difference of two losses evaluated at fake data equals the reverse KL-divergence between $P_{Y|X}$ and $Q_{Y|X}$,*

$$L_{mi}^p(x^-) - L_{mi}^q(x^-) = \mathbb{E}_{Q_X}KL(Q_{Y|X}\|P_{Y|X}). \quad (10)$$

DM-GAN (explicit data matching)

- Set $\psi(\cdot) = 0$

***f*-cGAN (explicit label matching)**

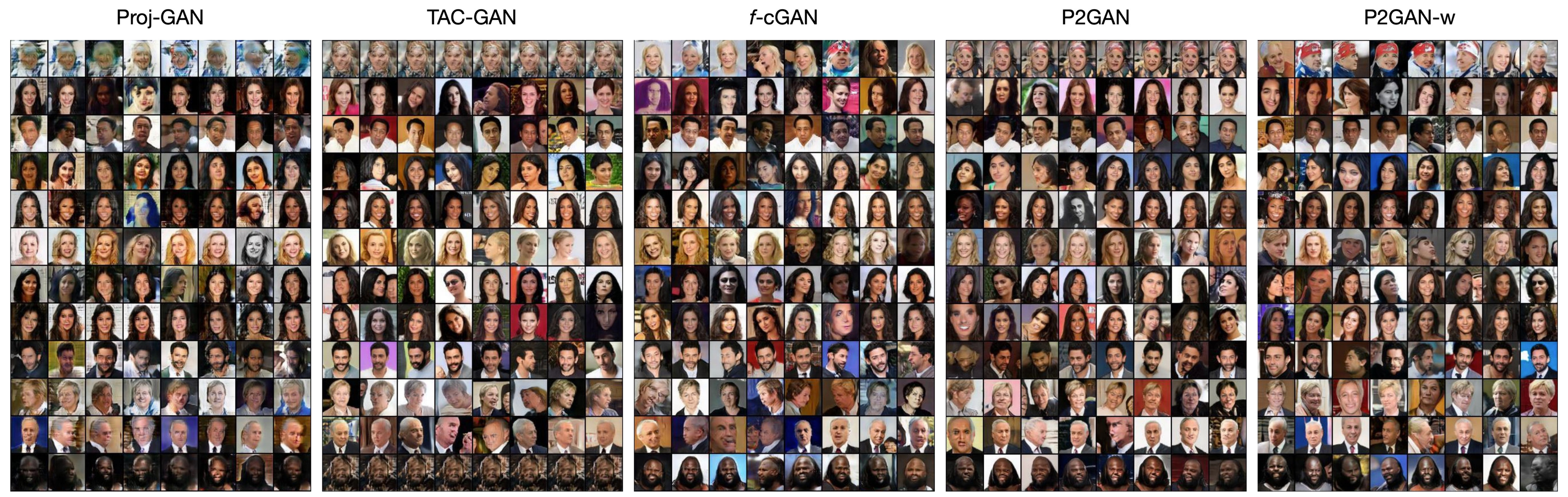
$$L_D^f = L_D(\tilde{D}) + L_{mi}^p(x^+) + L_{mi}^q(x^-)$$

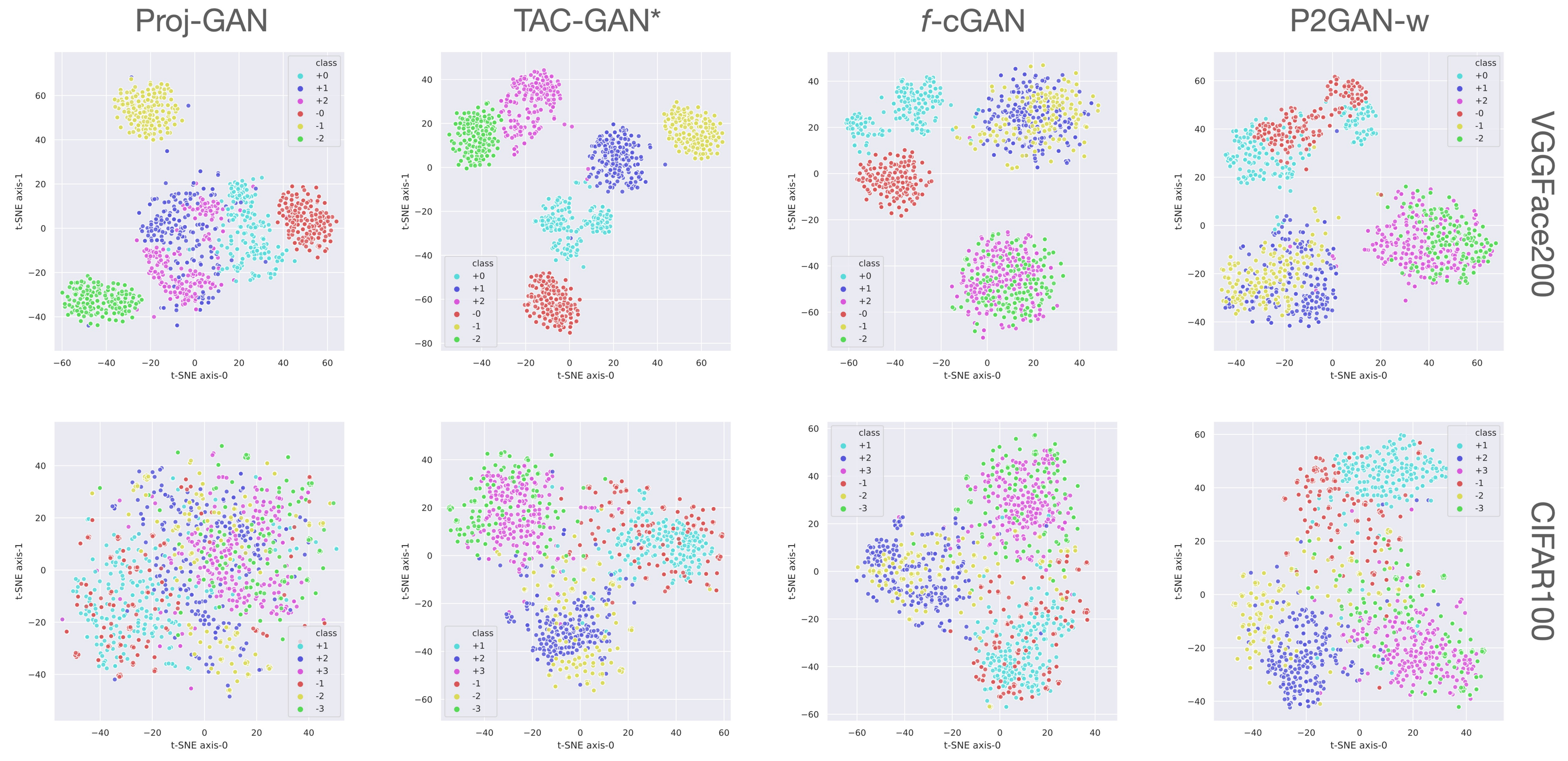
$$L_G^f = L_G(\tilde{D}) + L_{mi}^p(x^-) - L_{mi}^q(x^-)$$

$$\tilde{D} = v_X^T\phi(x) + b$$

Experiments

- VGGFace2 (200 identity subset)





- ImageNet and CIFAR10

	ImageNet 128 × 128		CIFAR10	
	IS ↑	FID ↓	IS ↑	FID ↓
Proj-GAN	30.73	23.07	*9.85	*8.03
P2GAN	**59.24**	**16.86**	9.76	8.00
P2GAN-w	42.69	19.19	**9.87**	**7.99**

Summary

We give insights on projection form of conditional discriminators and propose a new conditional generative adversarial network named Dual Projection GAN (P2GAN). We demonstrate its flexibility in modeling and balancing data matching and label matching. We further rigorously analyze the underlying connections between AC-GAN, TAC-GAN, and Proj-GAN. From the analysis, we first propose *f*-cGAN, a general framework for learning conditional GAN. We demonstrate the efficacy of our proposed models on various synthetic and real-world datasets.