Revisiting ImprovedGAN with Metric Learning for Semi-Supervised Learning

Jaewoo Park Yoon Gyo Jung Andrew Beng Jin Teoh

Yonsei University

Objective and Contribution

Objective: The adversarial loss in ImprovedGAN is analyzed under a metric learning framework, General Pair Weighting.

Contributions:

- Its theoretical properties related to class-wise cluster separation are observed, and further verified experimentally.
- In particular, adversarial losses in ImprovedGAN is observed to induce class-wise cluster separation on the features of all samples (both labeled and unlabeled).
- Based on the finding, two techniques are provided to enhance the class-wise cluster separation characteristic.

Preliminary

ImprovedGAN: Given a labeled set $\mathcal{L} = \{(x_1, y_1), \dots, (x_{|\mathcal{L}|}, y_{|\mathcal{L}|})\}$ with K classes, ImprovedGAN is trained by minimizing

$$\min_D L_u + L_s \tag{1}$$

and

$$\min_G L_g = \|\mathop{\mathbb{E}}_{\mathbf{x}\sim p_x} \mathbf{f}(\mathbf{x}) - \mathop{\mathbb{E}}_{\widehat{\mathbf{x}}\sim p_{G(\mathbf{z})}} \mathbf{f}(\widehat{\mathbf{x}})\|_1 \tag{2}$$

in an alternating manner for the discriminator D and generator G where L_u is the unsupervised discriminator loss

$$L_u = -\mathop{\mathbb{E}}_{\mathbf{x}\sim p_x} \log q(y \le K|\mathbf{x}) - \mathop{\mathbb{E}}_{\widehat{\mathbf{x}}\sim p_{G(\mathbf{z})}} \log q(y = K+1|\widehat{\mathbf{x}}), \tag{3}$$

$L_s = -\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{L}} \log q(y|\mathbf{x}, y \le K)$ is the supervision loss. The class predictor q is modeled by $q(y = k|\mathbf{x}) = \frac{e^{s_k(\mathbf{x})}}{1+\sum_{j=1}^K e^{s_j(\mathbf{x})}}$ with $s_{K+1} = 0$ and the $K+1$-th class serving as a fake class.

Observations

The role of the adversarial losses, namely, L_u and L_g is analyzed.

As a Metric Learning Loss: L_u is written as

$$L_u = \frac{1}{N}\sum_{i=1}^{N}\left[\log\left(1 + \frac{1}{\sum_{j=1}^K e^{s_{ij}}}\right) + \log\left(1 + \sum_{j=1}^K e^{\widehat{s}_{ij}}\right)\right] \tag{4}$$

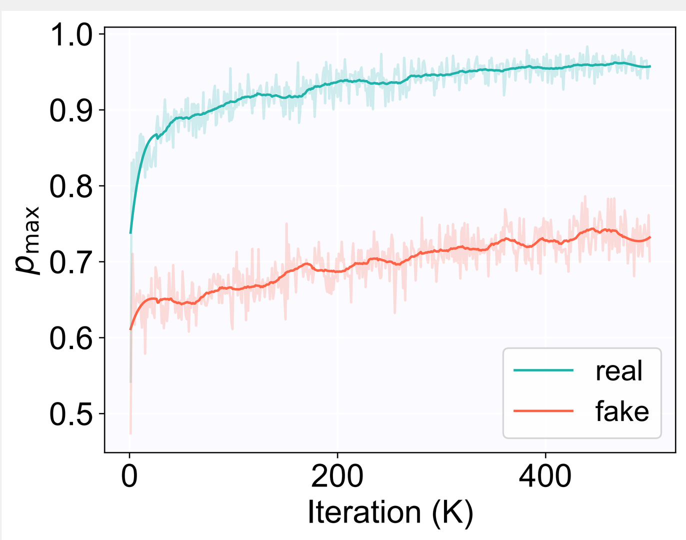
where the **similarity** s_{ij} is between the feature $\mathbf{f}_i = \mathbf{f}(\mathbf{x}_i)$ and class weight vector $\mathbf{w}_j$:

$$s_{ij} = \mathbf{f}_i \cdot \mathbf{w}_j = \|\mathbf{f}_i\|\|\mathbf{w}_j\| \cos\theta_{ij}. \tag{5}$$

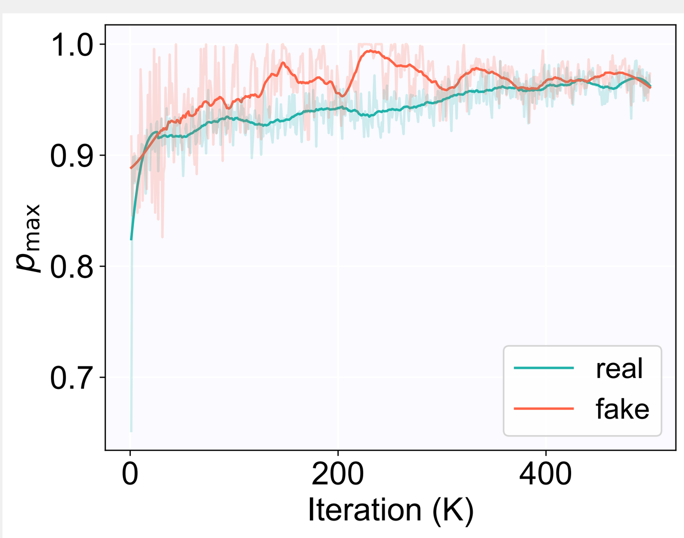
Under GPW, the followings can be proved:

Prediction Confidence:

Prop 1. Minimizing L_u maximizes $\max_j s_{ij}$ and thus the prediction confidence $p_{\max}(\mathbf{x}) = \max_y q(y|y < K, \mathbf{x})$ for real $\mathbf{x}$.



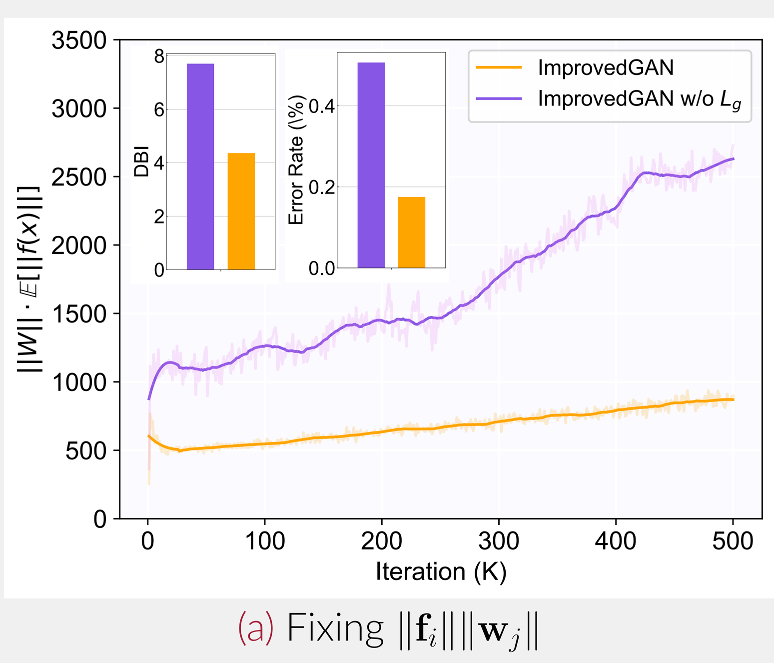
(a) w/o L_g



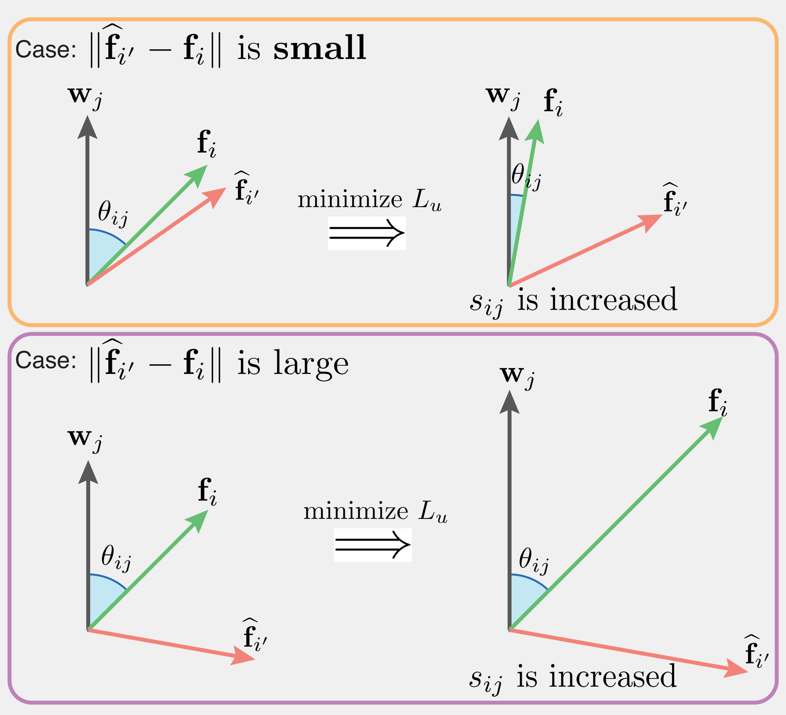
(b) w L_g

Angle Minimization (i.e., cosine similarity maximization):

Prop 2. If $\mathbf{f}_i$ and $\widehat{\mathbf{f}}_{i'} = (\widehat{\mathbf{x}}_{i'})$ with a generated sample $\widehat{\mathbf{x}}_{i'}$ are sufficiently near to each other, then minimizing L_u decreases the angle θ_{ij} while constraining $\|\mathbf{f}_i\|\|\mathbf{w}_j\|$ to be fixed.



(a) Fixing $\|\mathbf{f}_i\|\|\mathbf{w}_j\|$

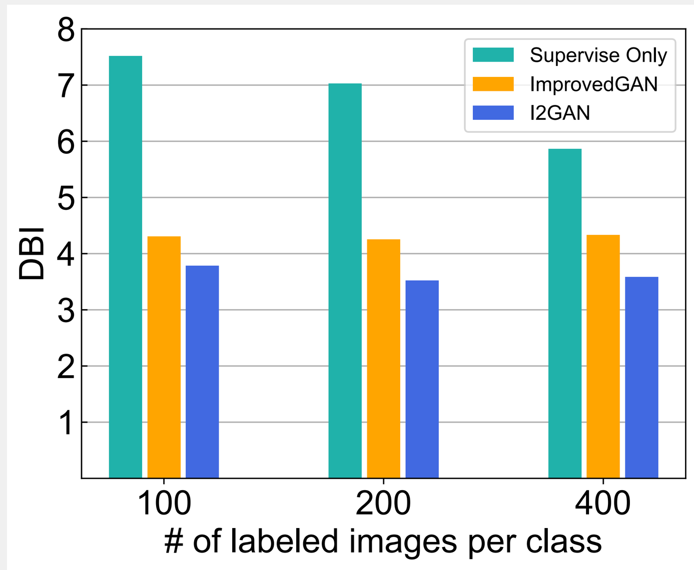


(b) On angle minimization

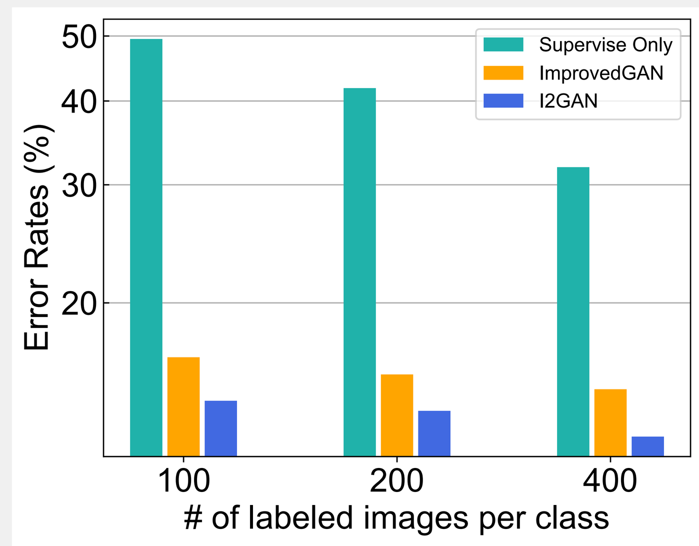
Figure: The class-wise cluster separation is measured by Davies-Bouldin Index (DBI).

Class-wise Cluster Separation:

The above two propositions suggests that **the adversarial interaction by L_u and L_g induces class-wise cluster separation of the real features $\mathbf{f}_i$.**



(a) DBI of the normalized features $\frac{\mathbf{f}(\mathbf{x})}{\|\mathbf{f}(\mathbf{x})\|}$ from real samples $\mathbf{x}$



(b) Semi-supervision error-rate in inference

Method

To **enhance class-wise cluster separation characteristic** of ImprovedGAN, we propose:

- **Scaling-up the unsupervised discriminator loss**: replace L_u by

$$L_u \leftarrow L_{u,\tau} := \tau L_u \tag{6}$$

  to make the model optimization end up with higher prediction confidence.
- **Excessive sampling on generated samples**: for the loss L_g, replace

$$\{\widehat{\mathbf{x}}_{i'}\}_{i=1}^{N} \leftarrow \{\widehat{\mathbf{x}}_{i'}\}_{i=1}^{N'} \quad \text{where } N' > N \tag{7}$$

  to better satisfy the sufficient condition of Prop 2.

The enhanced ImprovedGAN is termed as **I2GAN**.

Experiments

Table: The SSL performance in error rates (%) on CIFAR-10

# labels	100	200	400
Mean Teacher*	5.45 ± 0.14	5.21 ± 0.21	
LP* (CVPR'19)	16.93 ± 0.70	13.22 ± 0.29	10.61 ± 0.28
ICT* (NIPS'19)	15.48 ± 0.78	9.26 ± 0.09	7.29 ± 0.02
SWA* (ICLR'19)	15.58	11.02	9.05
ALI*	19.98 ± 0.89	19.09 ± 0.44	17.99 ± 1.62
TripleGAN*	81.08 ± 0.57	18.21 ± 0:37	16.99 ± 0.36
Local-GAN*	17.44 ± 0.25	-	14.23 ± 0.27
ImprovedGAN*	21.83 ± 2.01	19.61 ± 2.09	18.63 ± 2.32
BadGAN*	22.42 ± 0.17	18.64 ± 0.08	14.41 ± 0.30
ImprovedGAN w/ $\mathcal{M}$ Inv.*	19.52 ± 1.5	-	16.20 ± 1.6
ImprovedGAN w/ $\mathcal{M}$ Reg.*	16.37 ± 0.42	15.25 ± 0.35	14.34 ± 0.17
ImprovedGAN	16.80 ± 0.54	15.64 ± 0.12	14.86 ± 0.26
I2GAN	**14.29** ± **0.22**	13.80 ± 0.20	**12.63** ± **0.17**
e-I2GAN	14.93 ± 0.25	**13.77** ± **0.07**	13.29 ± 0.35

Table: The SSL performance in error rates (%) on CIFAR-100

# labels	40
Supervise Only	74.85 ± 0.55
BadGAN*	61.49 ± 0.73
ImprovedGAN (our implementation)	56.14 ± 0.64
I2GAN	**51.31** ± **0.32**
e-I2GAN	52.50 ± 1.25

Table: The SSL performance in error rates (%) on SVHN

# of labeled images for each class	50	100
Temporal Ensemble*	7.01 ± 0.29	5.73 ± 0.16
SPCTN*	-	7.73 ± 0.30
Pseudo-Labeling*		9.94 ± 0.61
Mean Teacher*	5.45 ± 0.14	5.21 ± 0.21
VAT*	-	5.77
ALI*	-	7.41 ± 0.65
TripleGAN*	5.33 ± 0.12	5.77 ± 0.17
LocalGAN*	5.48 ± 0.29	4.73 ± 0.29
ImprovedGAN*	18.44 ± 4.80	8.11 ± 1.3
BadGAN*	5.79 ± 0.45	**4.68** ± **0.07**
ImprovedGAN (our implementation)	5.79 ± 0.19	5.60 ± 0.09
I2GAN	**5.27** ± **0.13**	5.17 ± 0.16
e-I2GAN	5.43 ± 0.13	5.27 ± 0.10