Triple Generative Adversarial Nets

Chongxuan Li, Kun Xu, Jun Zhu, Bo Zhang Dept. of Comp. Sci. & Tech, Tsinghua University, Beijing, China

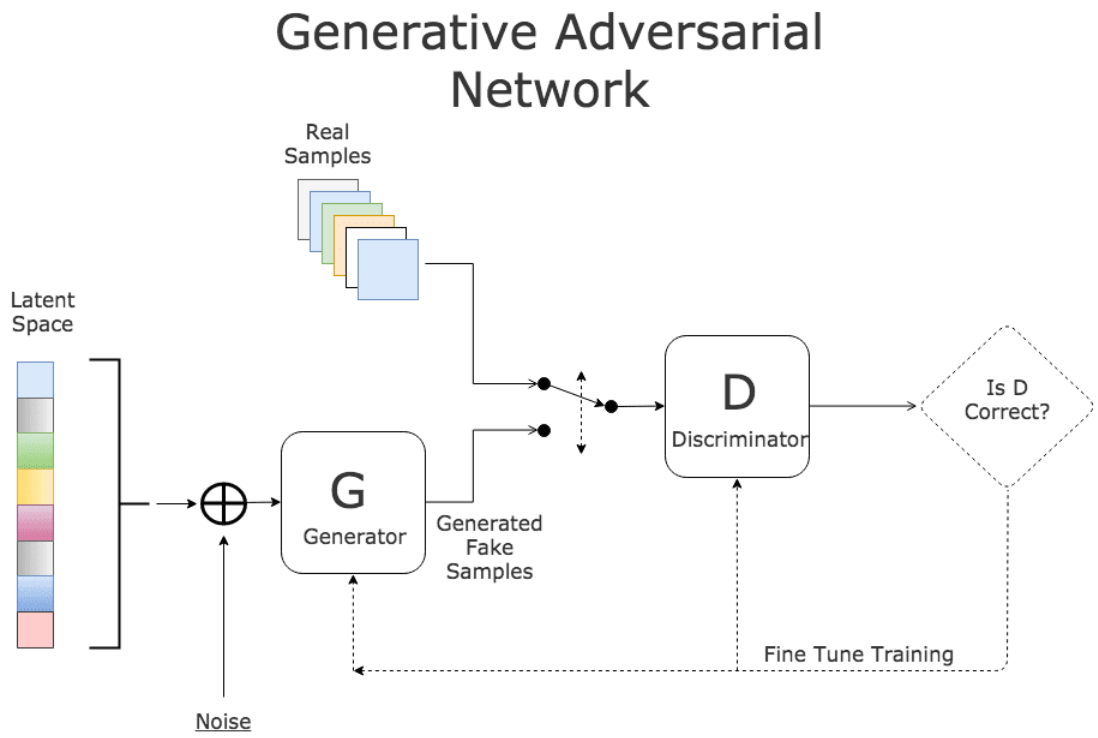
Advances in Neural Information Processing Systems (NeurIPS / NIPS 2017)

Introduction

Existing GANs in SSL have two problems:

(1) The Generator and the discriminator (i.e., the classifier) may not be optimal at the same time

(2) The Generator cannot control the semantics of the generated samples. The problems essentially arise from the two-player formulation, where a single discriminator shares incompatible roles of identifying fake samples and predicting labels and it only estimates the data without considering the labels

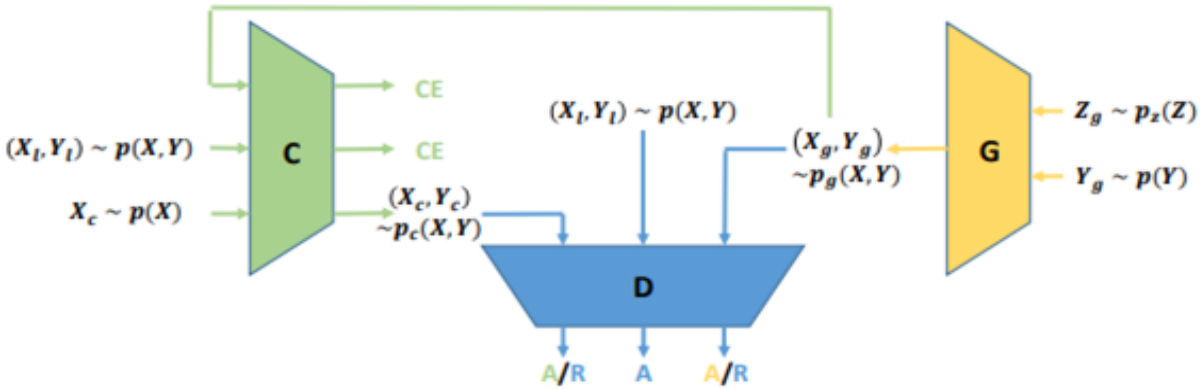


Methodology: Three Player Formulation

A classifier C that (approximately) characterizes the conditional distribution $p_c(y|x) \approx p(y|x)$

A class-conditional generator G that (approximately) characterizes the conditional distribution in the other direction $p_g(x|y) \approx p(x|y)$; and

A discriminator D that distinguishes whether a pair of data (x, y) comes from the true distribution p(x, y). All the components are parameterized as neural networks



Loss Function: $\min_{C,G} \max_D E_{p(x,y)} [\log D(x, y)] + \alpha E_{pc(x,y)} [\log(1 - D(x, y))] + (1 - \alpha) E_{pg(x,y)} [\log(1 - D(G(y, z), y))] + R_C + \alpha_P R_P$

where $\alpha \in (0, 1)$, pseudo discriminative loss - R_P

Results

- Triple-GAN ensures that both the classifier and the generator can achieve their own optima respectively in the perspective of game theory and enable the generator to sample data in a specific class

- Disentangle the classes and styles of the input and transfer smoothly in the data space via interpolation in the latent space class-conditionally.

- $V(C, G) = -\log 4 + 2JSD(p(x, y)||p_{\alpha}(x, y))$

Comments

- Increased complexity with three networks, potentially harder to train and tune.
- Risk of mode collapse or bias from pseudo labels in practice.
- Hyperparameter Sensitivity, α(=1/2, fixed in the paper), α_P, learning rates the paper's limited exploration of their impact suggests potential sensitivity that could hinder reproducibility or generalization across datasets.

- The mode collapse class specific examples can be used as negative samples for adversarial training

- Current SOTA are Diffusion – GAN hybrids

Three Player Formulation

In the game, after a sample x is drawn from p(x), C produces a fake label y given x following the conditional distribution $p_c(y|x)$.

The fake input-label pair is a sample from the joint distribution $p_c(x, y) = p(x)p_c(y|x)$.

Similarly, a fake input-label pair is sampled from G by first drawing $y \sim p(y)$ and then drawing $x|y \sim p_g(x|y)$,

Hence from the joint distribution $p_g(x, y) = p(y)p_g(x|y)$.For $p_g(x|y)$, we assume that x is transformed by the latent style variables z given the label y, namely, $x = G(y, z)$, $z \sim p_z(z)$, where $p_z(z)$ is a simple distribution (e.g., uniform or standard normal). The fake input-label pairs (x, y) generated by both C and G are sent to the discriminator D. D can also access the input-label pairs from the true data distribution as positive samples.

Our desired equilibrium is that the joint distributions defined by the classifier and the generator both converge to the true data distribution

- The objective functions in the process as adversarial losses :
- $\min_{C,G} \max_{D} E_{p(x,y)} [\log D(x, y)] + \alpha E_{pc(x,y)} [\log(1 - D(x, y))] + (1 - \alpha) E_{pg(x,y)} [\log(1 - D(G(y, z), y))]$
- where $\alpha \in (0, 1)$ is a constant that controls the relative importance of classification and generation
- for convenience $\alpha = 1/2$.

For any fixed C and G, the optimal D of the game defined by the utility function U(C, G, D) is:

$$D^*_{C,G}(x, y) = p(x, y)/ (p(x, y) + p_\alpha(x, y)) ,$$

where $p_\alpha(x, y) := (1 - \alpha)p_g(x, y) + \alpha p_c(x, y)$ is a mixture distribution for $\alpha \in (0, 1)$.

$$V (C, G) = \max_D U(C, G, D)$$
$$V (C, G) = - \log 4 + 2JSD(p(x, y)||p_\alpha(x, y))$$

- $p_\alpha(x, y) := (1 - \alpha)p_g(x, y) + \alpha p_c(x, y)$

The equilibrium indicates that if one of C and G tends to the data distribution, the other will also go towards the data distribution, which addresses the competing problem

- Given $p(x, y) = p_\alpha(x, y)$, the marginal distributions are the same for p, p_c and p_g, i.e. $p(x) = p_g(x) = p_c(x)$ and $p(y) = p_g(y) = p_c(y)$

However, it may not be unique, and we should minimize an additional objective to ensure the uniqueness

Final Loss Function

- The objective functions in the process as adversarial losses :
- $\min_{C,G} \max_{D} E_{p(x,y)} [\log D(x, y)] + \alpha E_{pc(x,y)} [\log(1 - D(x, y))] + (1 - \alpha) E_{pg(x,y)} [\log(1 - D(G(y, z), y))] + R_C + \alpha_P R_P$
- where $\alpha \in (0, 1)$
- Because label information is extremely insufficient in SSL, we propose pseudo discriminative loss
- $R_P = E_{pg} [- \log p_c(y|x)]$
  - The Cross Entropy Loss to C
- $R_C = E_{(x,y)\sim p(x,y)} [- \log p_c(y|x)]$

Disentangle the classes and styles of the input and transfer smoothly in the data space via interpolation in the latent space class-conditionally”.

in simpler terms:

- To **disentangle** the classes and styles of the input means to separate the different categories (such as dogs or cats) and visual features (such as colour or shape) of the images that are given to the model.
- To **transfer smoothly** in the data space means to create new images that look realistic and natural by changing some aspects of the original images gradually.
- To do this **via interpolation** in the latent space means to use a mathematical technique that finds intermediate values between two points in a hidden representation of the data that captures its essential characteristics.
- To do this **class-conditionally** means to do this only for images that belong to the same category (such as dogs or cats), and not across different categories.