CS5370 Deep Learning for Vision – Assignment 6

Submitted by:
Vishal Singh Yadav
CS20MTECH01001

(1) Objective:

$$\min_D \left[\frac{1}{2}\mathbb{E}_{x\sim p_{data}(x)}[(D(x)-b)^2] + \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}[(D(G(z))-a)^2]\right]$$

$$\Rightarrow \min_D \int_x \left(\frac{1}{2} p_{data}(x)(D(x)-b)^2 + \frac{1}{2} p_G(x)(D(x)-a)^2\right)dx$$

$$\Rightarrow \frac{1}{2}\min_D \int_x \left(p_{data}(x)(D(x)-b)^2 + p_G(x)(D(x)-a)^2\right)dx$$

Let $y = D(x)$

$a = p_G$

$b = p_{data}$.

$$\Rightarrow f(y) = b(y-b)^2 + a(y-a)^2$$

$$f'(y) = 2a(y-a) + 2b(y-b).$$

$f'(y) = 0$.

$$\Rightarrow 2a(y-a) + 2b(y-b) = 0.$$

$$= 2ay - 2a^2 + 2by - 2b^2 = 0.$$

$$2ay + 2by = 2a^2 + 2b^2$$

$$2y(a+b) = 2(a^2+b^2)$$

$$y = \frac{a^2+b^2}{a+b}$$

$\therefore \cancel{D^*_G(x) =}$ $\cancel{p_G^2 + p_{data}}$

$$\therefore D^*_G(x) = \frac{p_G(x)^2 + p_{data}(x)^2}{p_G(x) + p_{data}(x)}$$

<2> Inception score doesn't use stats of real world samples to compare with stats of synthetic examples. This results in IS failing in evaluating GAN.

To overcome this, we use Frechet Inception Distance (FID) which enables us to capture more subtle differences and measure diversity better.

<3> The subcategory of attacks which gets the image classified as specific target class, which is different from true class. are called targeted attacks.

They work by adding a noise to original image in a way that model misclassifies the image.

Some examples include:

- adding stickers or ~~specs~~ spectacles to images.

<4>. Since we need to classify as both bird & horse, the true class should consist of both.

Therefore, PGD eqn,

$$x^0_{adv} = x$$

$$x^{t+1}_{adv} = Proj\{x^t_{adv} + \epsilon \cdot sign(\nabla_x L(x^t_{adv}, (bird, horse))\}$$

<5> Contrastive loss keeps similiar samples together while classifying away from dissimiliar images.
We can use triplet loss in self supervised learning to reduce the minimum distance between positive and negative samples.

~~Since contrastive loss minimizes A-N di:~~
Since contrastive loss maximizes A-N distance & minimizes A-P distance. Whereas ~~triplet~~ triplet loss minimizes difference b/w A-N & A-P distance.
We can use triplet loss to control minimum distance b/w similiar & dissimiliar samples.

<6> $$\mathcal{L} = -\mathbb{E}\left[\log \frac{\exp(\text{score}(f(x), f(x^+))/\tau)}{\exp(\text{score}(f(x), f(x^+))/\tau) + \sum_{j=1}^{N-1} \exp(\text{score}(f(x), f(x_j^-))/\tau)}\right]$$

the temperature parameter is used to determine how spiky the probability distribution is.

when $\tau < 1$, i.e. temperature is low, we get a spiky distribution with a single or few prominent peaks.

when $\tau > 1$, i.e. temperature is high, we get a flat distribution without any prominent peaks.

<7><a> Even though regression seems a better option due to continous nature of images, a classification model works better.

Since an image is formed with pixels, each pixel can take various tones of Red, Green & Blue. A regression model will minimize l_2 loss for each pixel resulting in mean pixel value.

This ends ~~being~~ generating images that are desaturated and impure in colour tonality

<6> RGB has 3 color channels namely Red, Green and Blue. LAB colorspace contains lightness, A & B color channel. This contains the same information as RGB ~~[illegible]~~ but it makes processing easier.

<8> 50-way-6-shot learning.

Novel classes = 50.

Base classes = all other classes not included in novel classes

samples from each class = 6