Analisi di Immagini e Video (Computer Vision)

Francesco Sergio Pisani

GAN Evaluation

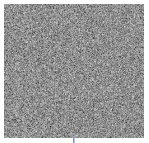
Evaluation of a classification model is simple, but how do we evaluate a generation model?

Classification Task

Classification Model

Car

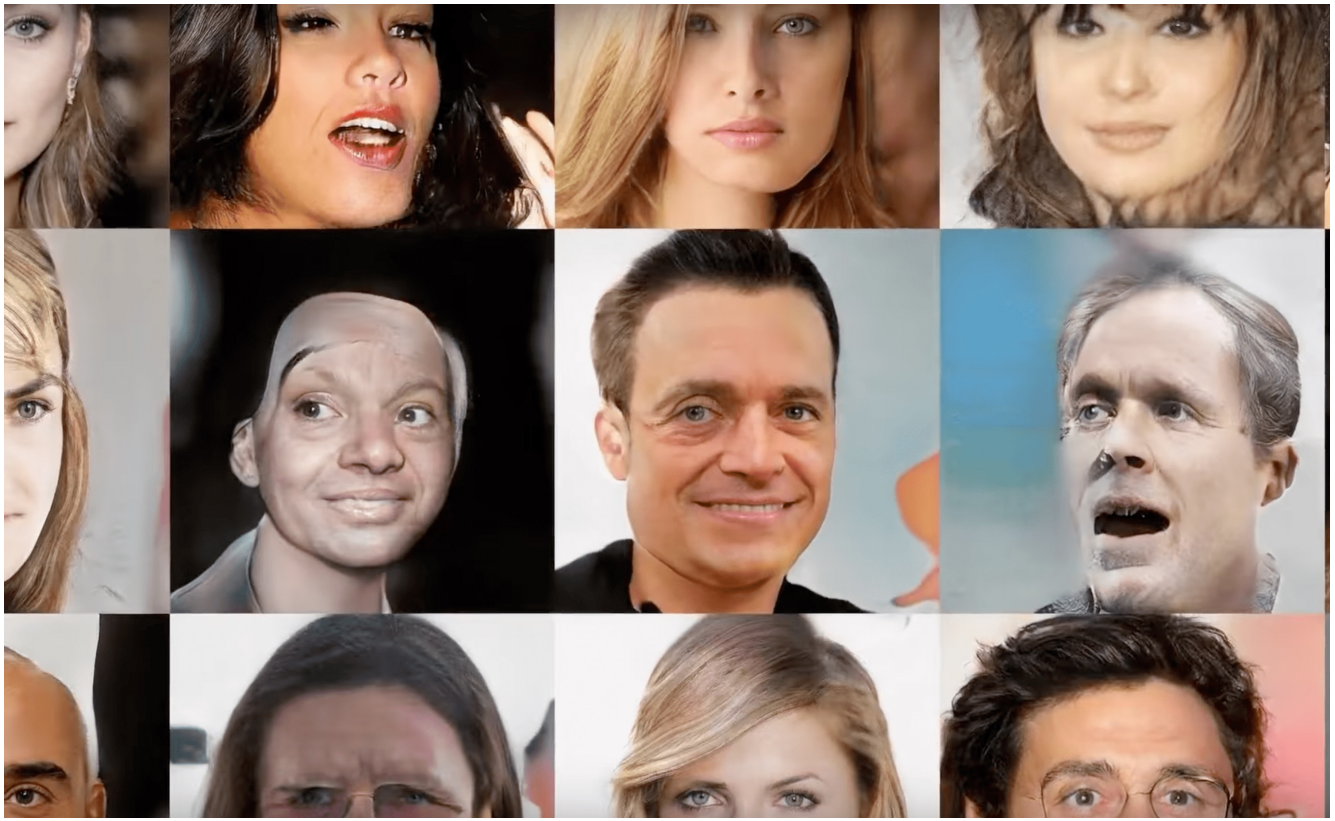
Generation Task



GAN

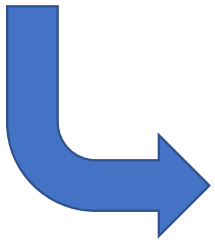
GAN Evaluation

Not all results are equals…



GAN Evaluation – Main properties

- **Fidelity**, how realistic each input is
- **Diversity**, how well fake samples capture the variations in real samples



- Pixel Distance
- Feature Distance

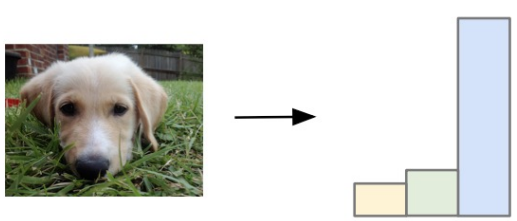
Cit. https://arxiv.org/pdf/2002.09797.pdf

Inception Score

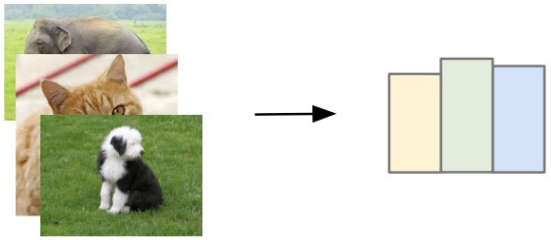
Main idea: compare label distribution of a pretrained classifier

Label probability $p_g(y|x) \approx f_\theta(x)$

Marginal Distribution $p_g(y) \approx E_{x\sim p_g}[f_\theta(x)]$

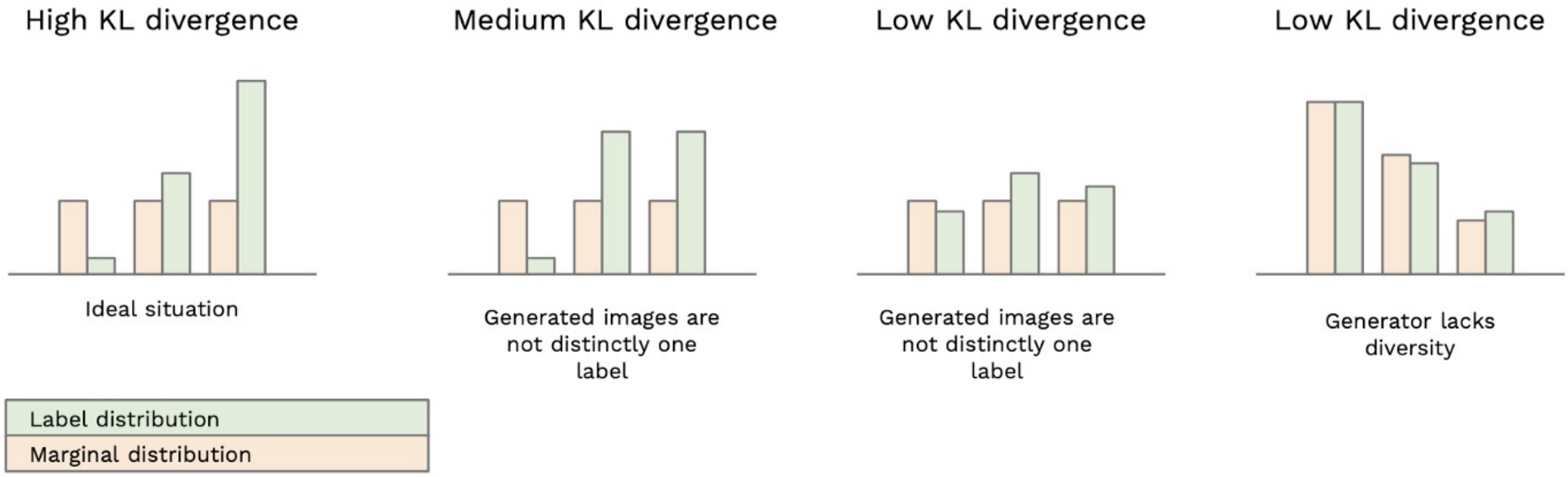


Ideal label distribution



Ideal marginal distribution

Inception Score



Inception Score - Problems

1. Highly correlated with training set data and its labels (how handle completely new data?)

2. If the classifier network cannot detect features relevant to your concept of image quality, then poor quality images may still get high scores

3. there is no measure of intra-class diversity

Frechet Inception Distance (FID)

FID is a performance metric that calculates the distance between the feature vectors of real images and the feature vectors of fake images

$$x_{data} \sim N(\mu_1, \Sigma_1) \qquad x_{fake} \sim N(\mu_2, \Sigma_2)$$

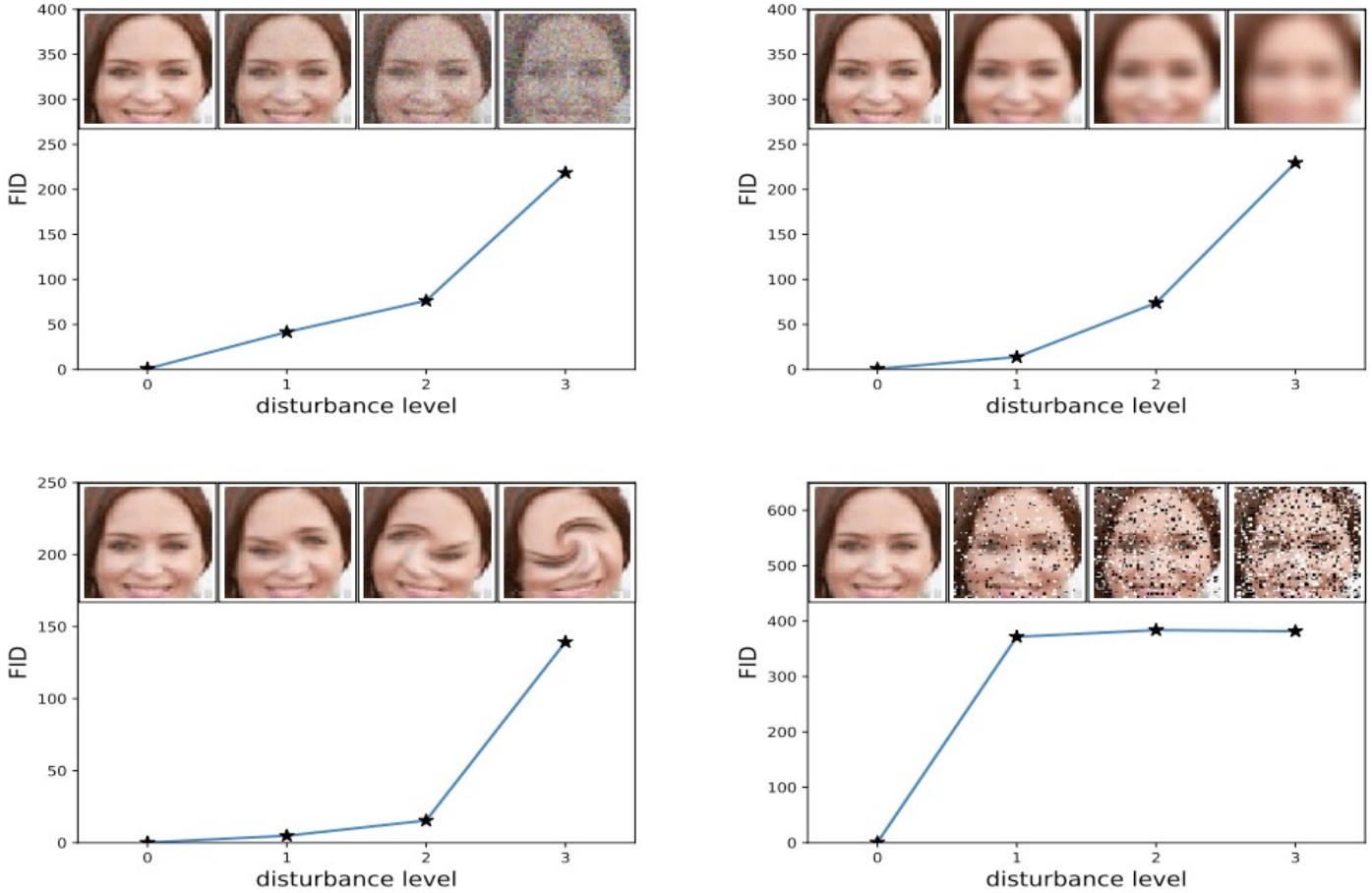
$$d(p_{data}, p_g) = \|\mu_1 - \mu_2\|_2^2 + Tr(\Sigma_1 + \Sigma_2 - 2\sqrt{\Sigma_1 \Sigma_2})$$

Mean and covariance in the feature space!!

Frechet Inception Distance (FID)

1. Robust to noise, image distortions, and perturbations.
2. A good metric for diverse datasets
3. Computationally efficient
4. Can detect intra-class mode dropping

Frechet Inception Distance (FID)



Frechet Inception Distance (FID) - Problems

- Still impossible to detect overfitting with it
- Not really a distance (only a distance for Gaussians distributions)