Measures of Generated Image Quality

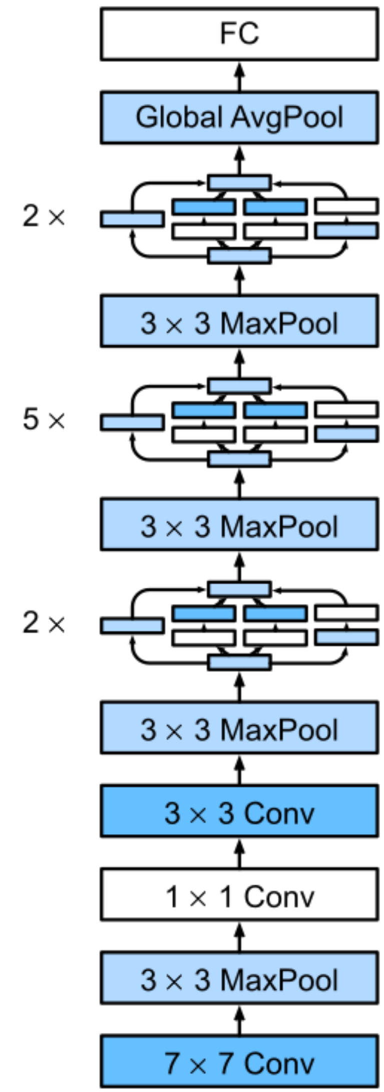
Bryan Pardo Fall 2024

What is Inception Score?

Inception V3

“We need to go deeper”

- Based on GoogLeNet
- Was the 3rd version of this net
- A famous image classifier released in 2016
- Trained on ImageNet (1000 class image data)
- Now mostly known for being used in calculating IS and FID



Inception Score

- Generate a bunch of images, conditioned on an ImageNet class (e.g. “dog”)
- Run each image through InceptionV3
- Gather their class probability distributions.
- Compare the distributions of things with similar labels to the distribution of things with different labels
- **HIGHER IS BETTER!**

https://proceedings.neurips.cc/paper_files/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf

Our ideal

Image classes are distinct

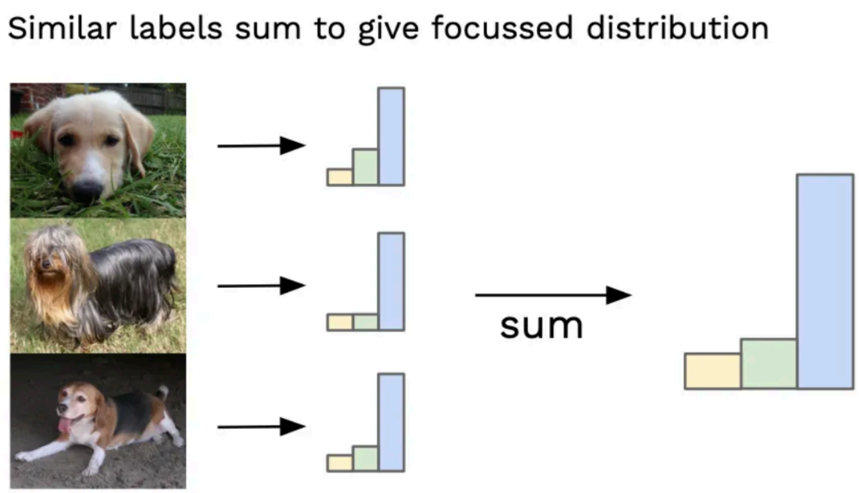
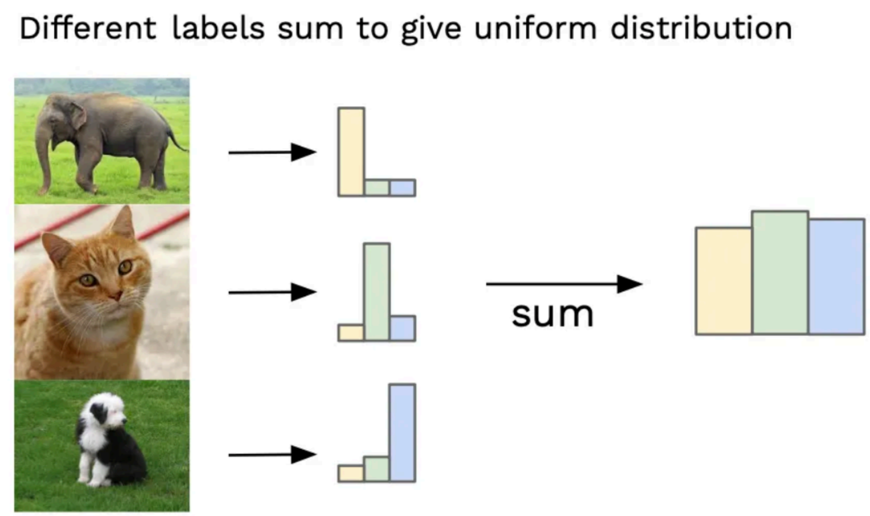
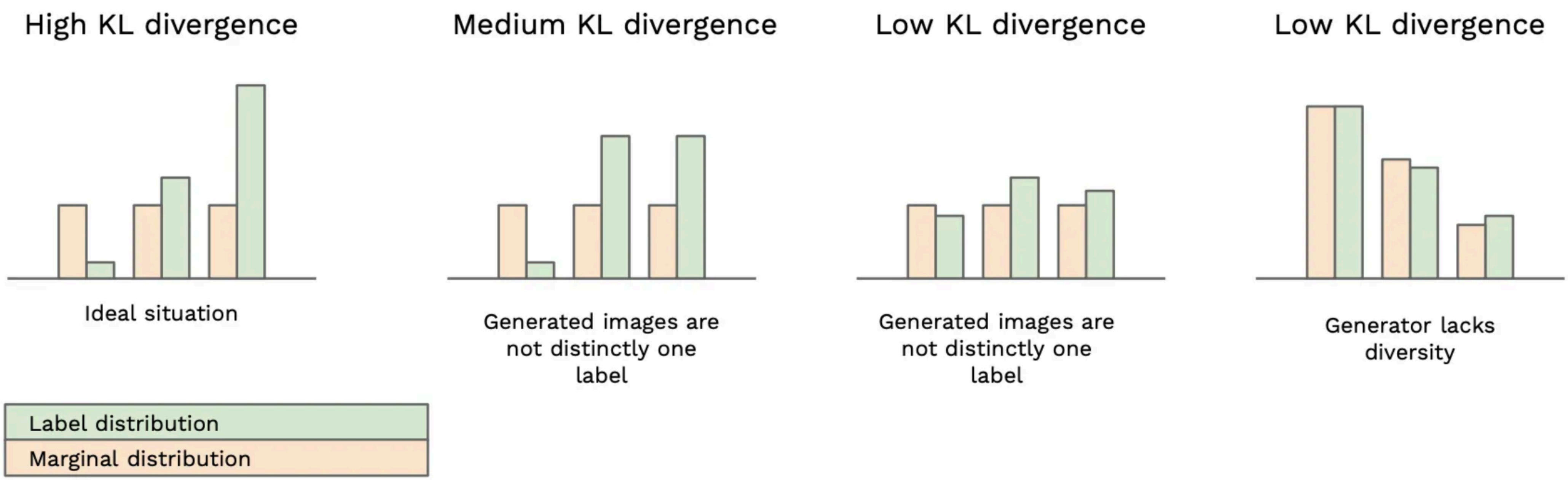


Image classes are diverse



https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a

KL divergence between distributions



https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a

In the author's words

As an alternative to human annotators, we propose an automatic method to evaluate samples, which we find to correlate well with human evaluation: We apply the Inception model[1] [19] to every generated image to get the conditional label distribution $p(y|\boldsymbol{x})$. Images that contain meaningful objects should have a conditional label distribution $p(y|\boldsymbol{x})$ with low entropy. Moreover, we expect the model to generate varied images, so the marginal $\int p(y|\boldsymbol{x} = G(z))dz$ should have high entropy. Combining these two requirements, the metric that we propose is: $\exp(\mathbb{E}_{\boldsymbol{x}} \mathrm{KL}(p(y|\boldsymbol{x}) || p(y)))$, where we exponentiate results so the values are easier to compare.

Inception Score = $\exp(\mathbb{E}_{\boldsymbol{x}} \mathrm{KL}(p(y|\boldsymbol{x}) || p(y)))$

https://proceedings.neurips.cc/paper_files/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf

How to interpret IS

Higher = better

What is Frechet Inception Distance?

Make 2 sets of image embeddings

- Create a set of generated images: G
- Collect a set of real images: R
- Run every image through InceptionV3 to get its embedding
- Fit a single Gaussian to the distribution of the embeddings for G
- Fit a single Gaussian to the distribution of the embeddings for R
- Measure the KL divergence between the 2 Gaussians.

https://proceedings.neurips.cc/paper_files/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf

Frechet Distance

A measure of difference between distributions

the FID compares the mean and standard deviation of two image sets, as represented by the deepest layer in Inception v3 (the 2048-dimensional activation vector of its last pooling layer.)

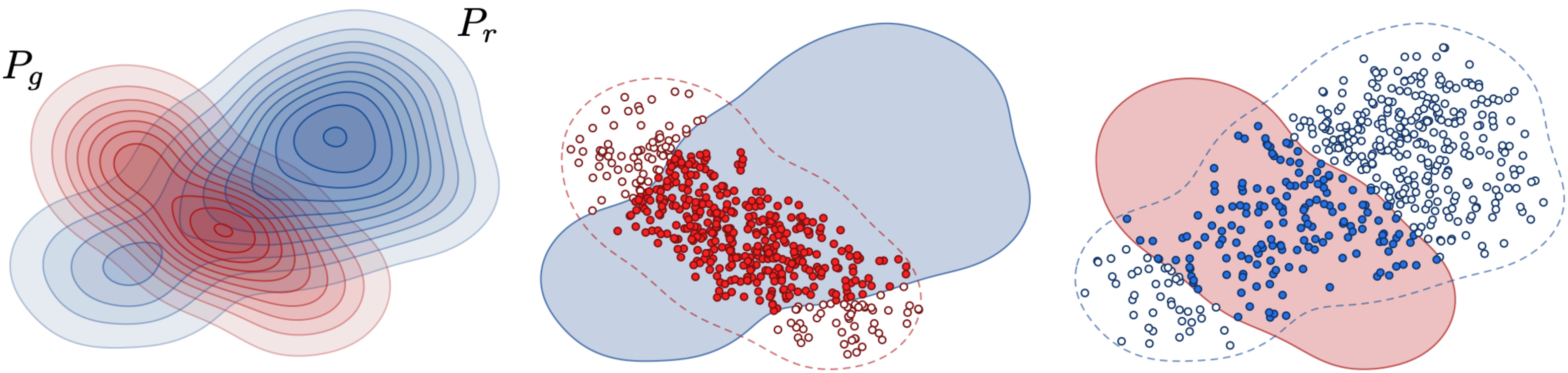
$$d_F(\mathcal{N}(\mu, \Sigma), \mathcal{N}(\mu', \Sigma'))^2 = \|\mu - \mu'\|_2^2 + \operatorname{tr}\left(\Sigma + \Sigma' - 2(\Sigma\Sigma')^{\frac{1}{2}}\right)$$

How to interpret FID

Lower = better

What is Precision/Recall?

https://arxiv.org/pdf/1904.06991



(a) Example distributions (b) Precision (c) Recall

Figure 1: Definition of precision and recall for distributions [25]. (a) Denote the distribution of real images with P_r (blue) and the distribution of generated images with P_g (red). (b) Precision is the probability that a random image from P_g falls within the support of P_r. (c) Recall is the probability that a random image from P_r falls within the support of P_g.

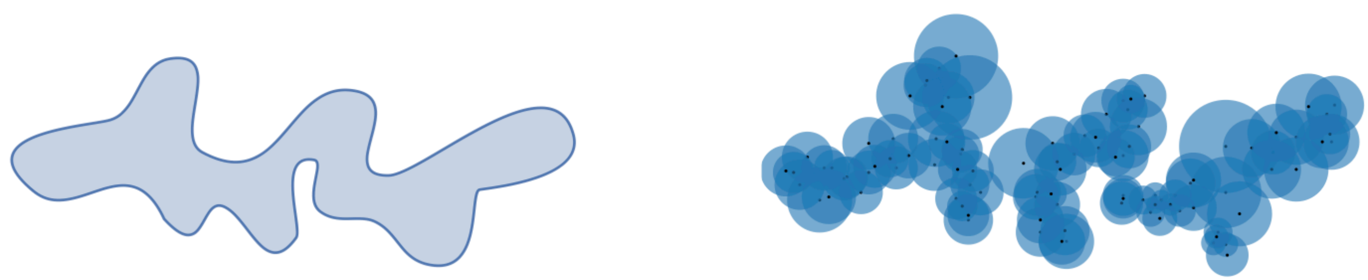
The nearest neighbor to ϕ'

image from the comparison set

An image

$$f(\phi, \mathbf{\Phi}) = \begin{cases} 1, & \text{if } \left\|\phi - \phi'\right\|_2 \le \left\|\phi' - \mathrm{NN}_k\left(\phi', \mathbf{\Phi}\right)\right\|_2 \text{ for at least one } \phi' \in \mathbf{\Phi} \\ 0, & \text{otherwise,} \end{cases}$$

The set of images we compare to ϕ



(a) True manifold (b) Approx. manifold

Figure 2: (a) An example manifold in a feature space. (b) Estimate of the manifold obtained b
sampling a set of points and surrounding each with a hypersphere that reaches its kth nearest neighbo

https://arxiv.org/pdf/1904.06991

How to interpret Precision & Recall

Higher = better