

Title: Inception score

URL: [https://en.wikipedia.org/wiki/Inception\\_score](https://en.wikipedia.org/wiki/Inception_score)

PageID: 71300141

Categories: Category:Computer graphics, Category:Machine learning

Source: Wikipedia (CC BY-SA 4.0).

-----

The Inception Score (IS) is an algorithm used to assess the quality of images created by a generative image model such as a generative adversarial network (GAN). [ 1 ] The score is calculated based on the output of a separate, pretrained Inception v3 image classification model applied to a sample of (typically around 30,000) images generated by the generative model. The Inception Score is maximized when the following conditions are true:

The entropy of the distribution of labels predicted by the Inceptionv3 model for the generated images is minimized. In other words, the classification model confidently predicts a single label for each image. Intuitively, this corresponds to the desideratum of generated images being "sharp" or "distinct".

The predictions of the classification model are evenly distributed across all possible labels. This corresponds to the desideratum that the output of the generative model is "diverse". [ 2 ]

It has been somewhat superseded by the related Fréchet inception distance . [ 3 ] While the Inception Score only evaluates the distribution of generated images, the FID compares the distribution of generated images with the distribution of a set of real images ("ground truth").

#### Definition

Let there be two spaces, the space of images  $\Omega_X$  and the space of labels  $\Omega_Y$ . The space of labels is finite.

Let  $p_{gen}$  be a probability distribution over  $\Omega_X$  that we wish to judge.

Let a discriminator be a function of type  $p_{dis} : \Omega_X \rightarrow M(\Omega_Y)$  where  $M(\Omega_Y)$  is the set of all probability distributions on  $\Omega_Y$ . For any image  $x$ , and any label  $y$ , let  $p_{dis}(y|x)$  be the probability that image  $x$  has label  $y$ , according to the discriminator. It is usually implemented as an Inception-v3 network trained on ImageNet.

The Inception Score of  $p_{gen}$  relative to  $p_{dis}$  is  $IS(p_{gen}, p_{dis}) := \exp \left( \mathbb{E}_{x \sim p_{gen}} [D_{KL}(p_{dis}(\cdot|x) \parallel p_{dis}(\cdot|x) p_{gen}(x) dx)] \right)$ . Equivalent rewrites include  $\ln IS(p_{gen}, p_{dis}) := \mathbb{E}_{x \sim p_{gen}} [D_{KL}(p_{dis}(\cdot|x) \parallel \mathbb{E}_{x \sim p_{gen}} [p_{dis}(\cdot|x)])]$  and  $\ln IS(p_{gen}, p_{dis}) := H[\mathbb{E}_{x \sim p_{gen}} [p_{dis}(\cdot|x)]] - \mathbb{E}_{x \sim p_{gen}} [H[p_{dis}(\cdot|x)]]$ .  $IS$  is nonnegative by Jensen's inequality .

Pseudocode:

INPUT discriminator  $p_{dis}$  .

INPUT generator  $g$  .

Sample images  $x_i$  from generator.

Compute  $p_{dis}(\cdot | x_i)$ , the probability distribution over labels conditional on image  $x_i$ .

Sum up the results to obtain  $\hat{p}$ , an empirical estimate of  $\int p_{dis}(\cdot | x) p_{gen}(x) dx$ .

Sample more images  $x_i$  from generator, and for each, compute  $D_{KL}(p_{dis}(\cdot | x_i) \parallel \hat{p})$ .

Average the results, and take its exponential.

RETURN the result.

#### Interpretation

A higher inception score is interpreted as "better", as it means that  $p_{gen}$  is a "sharp and distinct" collection of pictures.

$\ln IS(p_{gen}, p_{dis}) \in [0, \ln N]$ , where  $N$  is the total number of possible labels.

$\ln IS(p_{gen}, p_{dis}) = 0$  iff for almost all  $x \sim p_{gen}$   $p_{dis}(\cdot | x) = \int p_{dis}(\cdot | x) p_{gen}(x) dx$ . That means  $p_{gen}$  is completely "indistinct". That is, for any image  $x$  sampled from  $p_{gen}$ , discriminator returns exactly the same label predictions  $p_{dis}(\cdot | x)$ .

The highest inception score  $N$  is achieved if and only if the two conditions are both true:

For almost all  $x \sim p_{gen}$ , the distribution  $p_{dis}(y | x)$  is concentrated on one label. That is,  $H_y[p_{dis}(y | x)] = 0$ . That is, every image sampled from  $p_{gen}$  is exactly classified by the discriminator.

For every label  $y$ , the proportion of generated images labelled as  $y$  is exactly  $E_{x \sim p_{gen}}[p_{dis}(y | x)] = \frac{1}{N}$ . That is, the generated images are equally distributed over all labels.

#### References

v

t

e

MSE

MAE

sMAPE

MAPE

MASE

MSPE

RMS

RMSE/RMSD

R<sup>2</sup>

MDA

MAD

F-score  
P4  
Accuracy  
Precision  
Recall  
Kappa  
MCC  
AUC  
ROC  
Sensitivity and specificity  
Logarithmic loss  
Silhouette  
Calinski–Harabasz index  
Davies–Bouldin index  
Dunn index  
Hopkins statistic  
Jaccard index  
Rand index  
Similarity measure  
SMC  
DBCV index  
MRR  
NDCG  
AP  
PSNR  
SSIM  
IoU  
Perplexity  
BLEU  
Inception score  
FID  
Coverage  
Intra-list similarity  
Cosine similarity  
Euclidean distance  
Pearson correlation coefficient  
Confusion matrix