Notes:

FID

We have changed the evaluation metric codes, which was the cause of the unrealistic results.

The following is the repository of the official pytorch FID metric implementation we used in this experiment:

https://github.com/mseitzer/pytorch-fid

The number of samples affects this metric as shown in this discussion:

https://github.com/mseitzer/pytorch-fid/issues/78

for stable results the authors in the original FID paper suggested using a minimum of 10,000 images and it must be more than 2048.

> IMPORTANT: The number of samples to calculate the Gaussian statistics (mean and covariance) should be greater than the dimension of the coding layer, here 2048 for the Inception pool 3 layer. Otherwise the covariance is not full rank resulting in complex numbers and nans by calculating the square root.
>
> We recommend using a minimum sample size of 10,000 to calculate the FID otherwise the true FID of the generator is underestimated.

but this implementation allows for multiple dimensions for smaller datasets:

> This might be useful if the datasets you want to compare have less than the otherwise required 2048 images. Note that this changes the magnitude of the FID score and you can not compare them against scores calculated on another dimensionality. The resulting scores might also no longer correlate with visual quality.
>
> You can select the dimensionality of features to use with the flag `--dims N`, where N is the dimensionality of features. The choices are:
>
> - 64: first max pooling features
> - 192: second max pooling features
> - 768: pre-aux classifier features
> - 2048: final average pooling features (this is the default)

IS

Measures the following:

1. The images generated should contain clear objects (i.e. the images are sharp rather than blurry), or p(y|x) should be low entropy. In other words, the Inception Network should be highly confident there is a single object in the image.
2. The generative algorithm should output a high diversity of images from all the different classes in ImageNet, or p(y) should be high entropy.

Pytorch implementation:

https://github.com/sbarratt/inception-score-pytorch?tab=readme-ov-file

The only function is inception_score. It takes a list of numpy images normalized to the range [0,1] and a set of arguments and then calculates the inception score. Please assure your images are 3x299x299 and if not (e.g. your GAN was trained on CIFAR), pass resize=True to the function to have it automatically resize using bilinear interpolation before passing the images to the inception network.

MS-COCO (64 sample)		
	IS ↑	FID ↓
Steps = 100	3.58859973117884	4.5111905052973
Steps = 150	3.61568802476438	3.2820010318482886
Steps = 200	3.472260540514136	3.32373158663426

To be continued…We will choose the best steps size to find the FID, IS scores for larger sample.

MS-COCO (192 sample)		
	IS	FID
Steps = X	…	…

Note: Because we are using a version that was trained on the ImageNet dataset, the generated images for the captions that included people descriptions were poor (there is another version specialized for generating people images that is trained on the FFHQ dataset).

```
Model = 'Imagenet-1024'

network_url = {
    "Imagenet-1024": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/imagenet1024.pkl",
    "Imagenet-512": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/imagenet512.pkl",
    "Imagenet-256": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/imagenet256.pkl",
    "Imagenet-128": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/imagenet128.pkl",
    "Pokemon-1024": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/pokemon1024.pkl",
    "Pokemon-512": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/pokemon512.pkl",
    "Pokemon-256": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/pokemon256.pkl",
    "FFHQ-256": "https://s3.eu-central-1.amazonaws.com/avg-projects/stylegan_xl/models/ffhq256.pkl"
}
```

'Giraffes and a bird behind a chain link fence at a zoo'

'A woman walking while carrying bananas on her hat.'

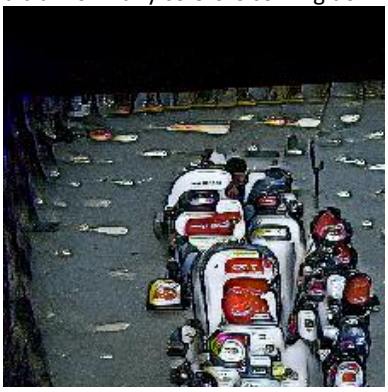
'A train with headlights traveling on a track past pedestrians.'

'Two zebras graze on grass inside of an enclosure. '

A shower curtain sits open in an empty and clean bathroom.

Many people standing in a field flying kites

a train of many colors is coming down the track



'a cat laying on a luggage bag on the ground'