The experiment of the baseline model

Testing StyleGAN against the dataset

StyleGAN+ CLIP text encoder

Configurations:

Steps = 40 , seed =-1 , optimizer = Adam , learning rate = 0.3

Caption	Dataset image	Generated image	
A plate with eggs, sausage, tomato and other meats.			
A herd of zebra standing next to each other.			
A box filled with some bacon topped donuts.			

A woman wearing sunglasses with a yellow oriental parasol opened over her head			
A man riding a skateboard up the side of a ramp			
A beach area with boats and people walking about it			
A man swinging a tennis racquet on a tennis court			

a woman riding a little pink motorcycle by a car			
a women combing her young daughters hair			
A man preparing to throw a frisbee while another man attempts to block him			
a man with a baseball bat in field			

a girl is next a flower garden on her phone		
an airplain sites outside the airport next to trucks		
snowboarders strike a pose with a variety of boards		

several diners are seated at a restaurant table		
A white bathroom has a striped shower curtain		

A group of people on a grass field together playing a game with a frisbee			
A baseball player running on a field during a game			

A home computer and coffee mug on a desk			
Female tennis player swinging racket for a high ball			

Evaluation using the following measures:

1-IS score which indicates the quality of images generated by the model. It measures two essential traits: diversity and distinctiveness.

2-FID score which indicates the similarity between two sets of images. A lower FID score indicates that the distributions of features extracted from the real and generated images are closer, suggesting better similarity between the two sets of images.

MS-COCO (20)		
	IS	FID
StyleGAN-XL model	0.00011784392700064927	384.1173296048412

SDXL diffusion model		