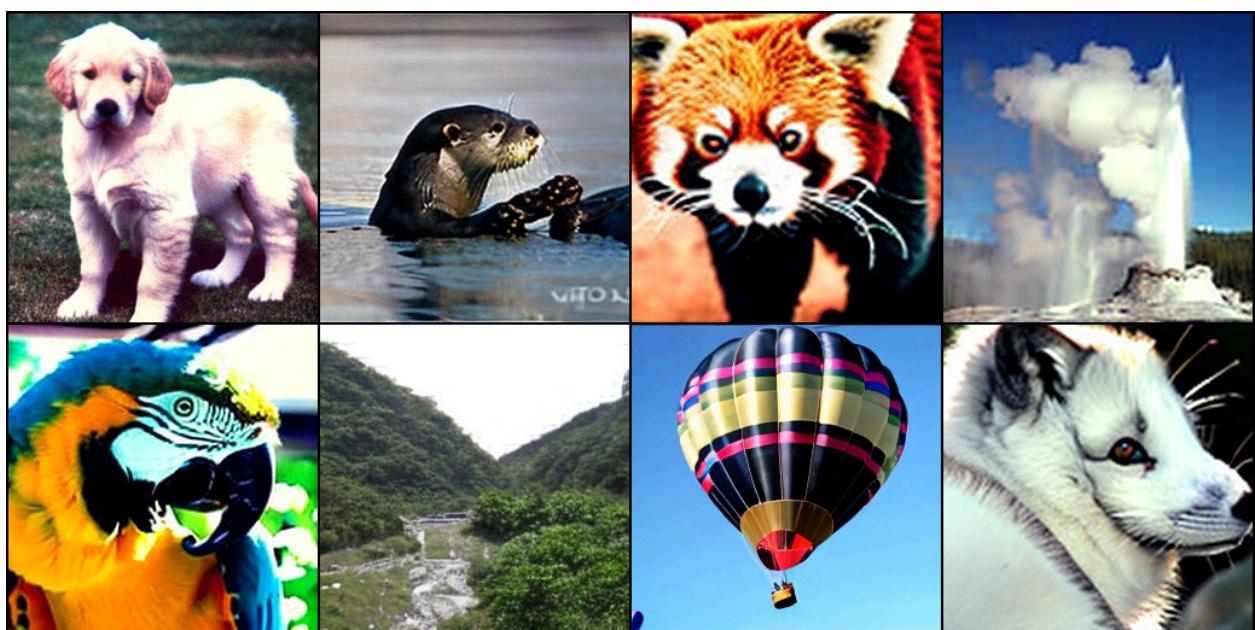




This is with cfg scale 0



This is with cfg scale 10

I also implemented with cfg scale 4 just for a comparison

CFG 4:



Now for sampling steps:

This is with 50 sampling steps:



250 steps:



500 steps:



Xformers task:

Avg time per 10 images: 0.05103s

Avg time per 10 images: 0.01535s

Testing batch size: 50

Avg time per 50 images: 0.26693s

Avg time per 50 images: 0.05673s

Testing batch size: 100

Avg time per 100 images: 0.45293s

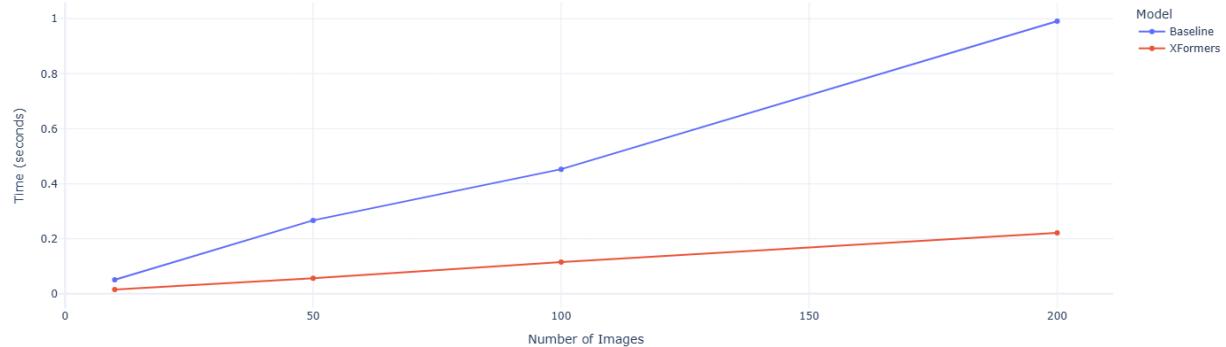
Avg time per 100 images: 0.11553s

Testing batch size: 200

Avg time per 200 images: 0.99126s

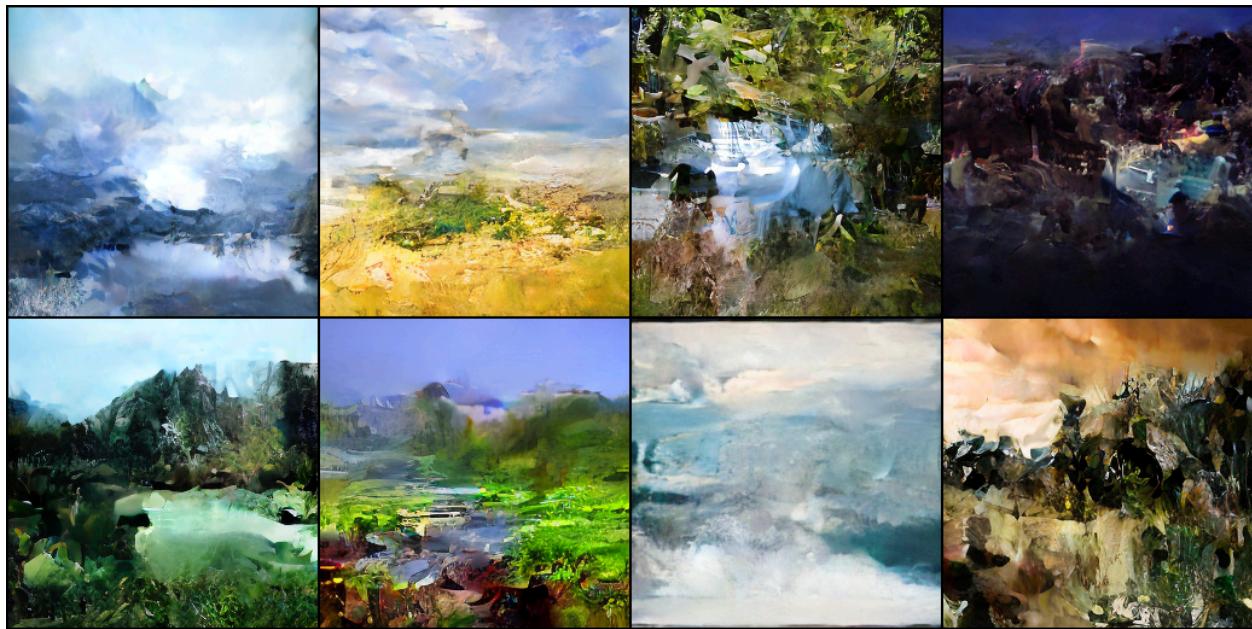
Avg time per 200 images: 0.22171s

Inference Time vs Batch Size



For the swa task:

This is the images generated by the Full attention



Sliding window implemented dit



## FID and CMMD calculations foCLIP/ SIGLIP and ALIGN MMD

```
{  
    "fid": {  
        "full": 246.8992661676288,  
        "swa": 388.87852498561443  
    },  
    "cmmd": {  
        "clip": {  
            "full": 0.2812194228172302,  
            "swa": 0.44345635175704956  
        },  
        "siglip": {  
            "full": 0.25273597240448,  
            "swa": 0.3985597491264343  
        },  
        "align": {  
            "full": 0.33260107040405273,  
            "swa": 0.47345298528671265  
        }  
    }  
}
```

