



Tests on VQGAN

VQGAN의 일러스트에 대한 reconstruct 성능 확인

결과

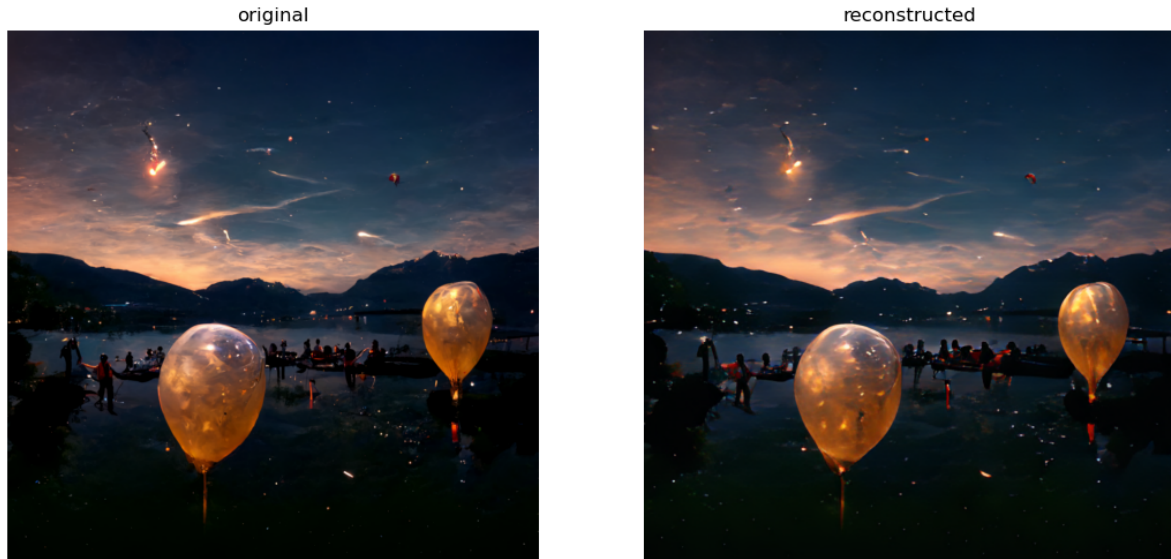
https://s3-us-west-2.amazonaws.com/secure.notion-static.com/8f70646d-b7ac-45c2-b416-01f31c422118/vqgan_performance_test.py

original



reconstructed





Test를 통해 VQGAN의 입력/출력 이미지 사이즈를 자유롭게 변경할 수 있음을 확인할 수 있었음. 이는 VQGAN이 CNN 기반 아키텍처를 사용하기 때문임으로 보임. 이에 따라 VQGAN token 또한 사이즈가 변할 것이므로 Dalle 내부 코드에서도 이가 반영가능한지 확인할 필요가 있어 보인다.

결론

VQGAN의 Decoder는 일러스트를 reconstruct하는데 그대로 사용해도 큰 무리는 없을 것 같음. Dalle 모델 결과의 성능은 VQGAN Token을 모방하여 생성하는 Transformer Decoder가 얼마나 잘 생성하는지에 달려 있는듯 하며, Dalle가 일러스트만 생성하는지 여부 또한 Transformer Decoder를 일러스트에 대해서만 훈련시키면 되는 것으로 보인다.

따라서 VQGAN을 직접 훈련시킬 필요는 없어보임.