

Avoy Datta, Michael Cai | Stanford University

Motivation

- Utilize a Generative Adversarial Network (GAN) to generate artwork conditioned on class labels.
- Most conditional GANs that generate artwork do so with resolutions of 128x128 or less.
- Obtaining high-resolution samples requires considerable training time and results in poor differentiability.
- We propose **GANtor-v2**, a 2-stage Stacked GAN that generates 256x256 images using a 2-stage training process while still enforcing differentiability.

Dataset

- Our model is trained on the publicly available **WikiArt** dataset.
- 80,000 images labeled on style; 65,000 images labeled on genre.
- GANtor-v2** trained on genre labels, **GANtor-v1** trained on style labels.
- Stage-I** takes in 64x64 copies of art during training; **Stage-II** takes in 256x256 copies.
- Both stages take in labels as one-hot vectors.

Distribution of Samples per Genre

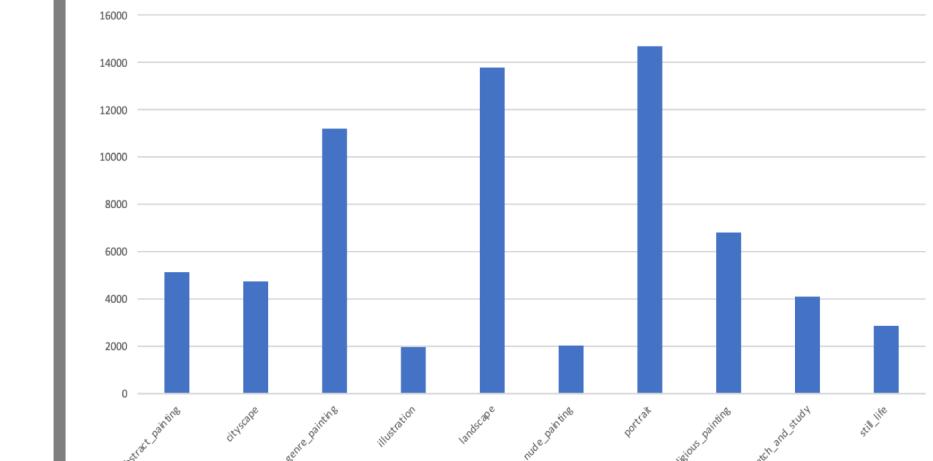
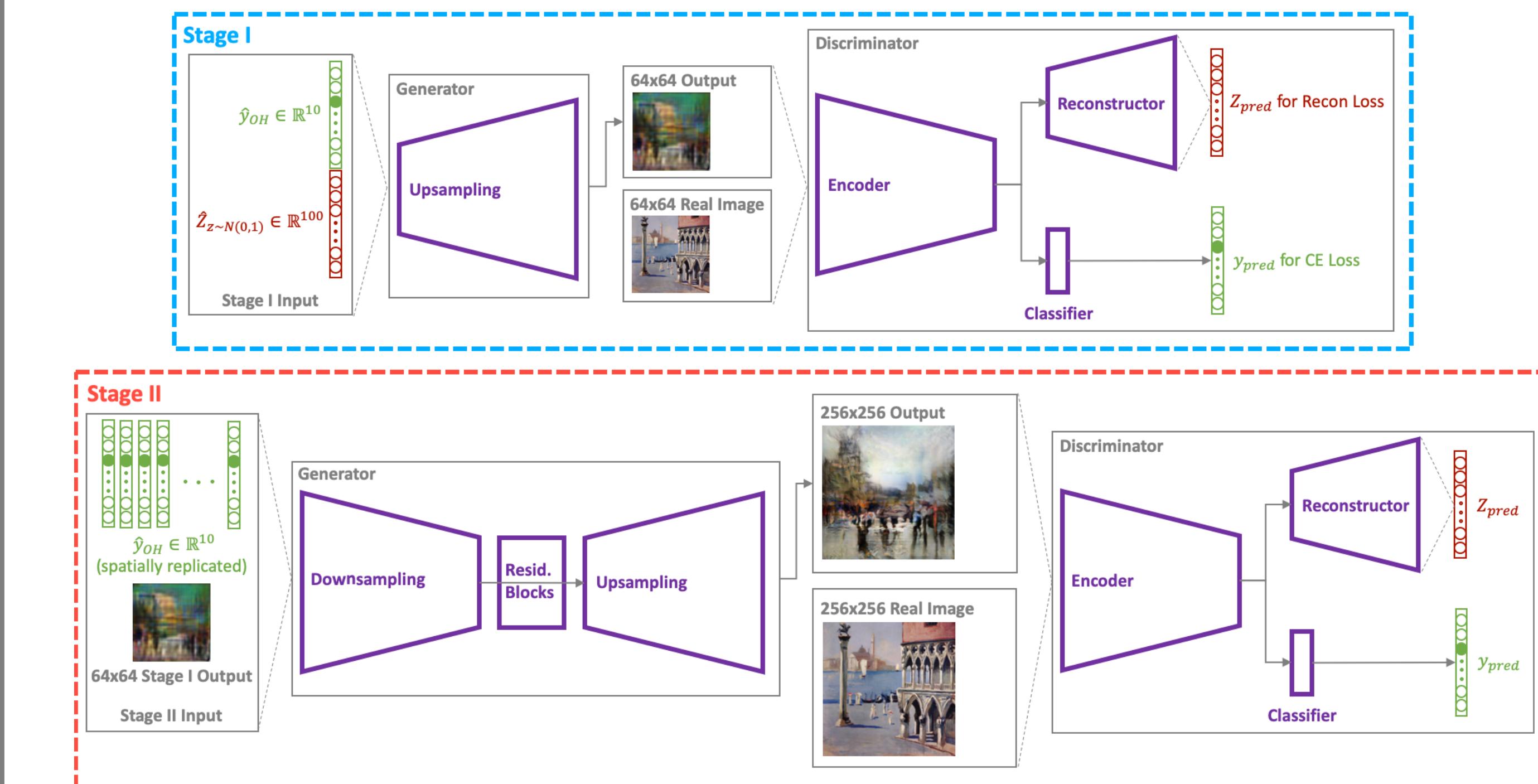


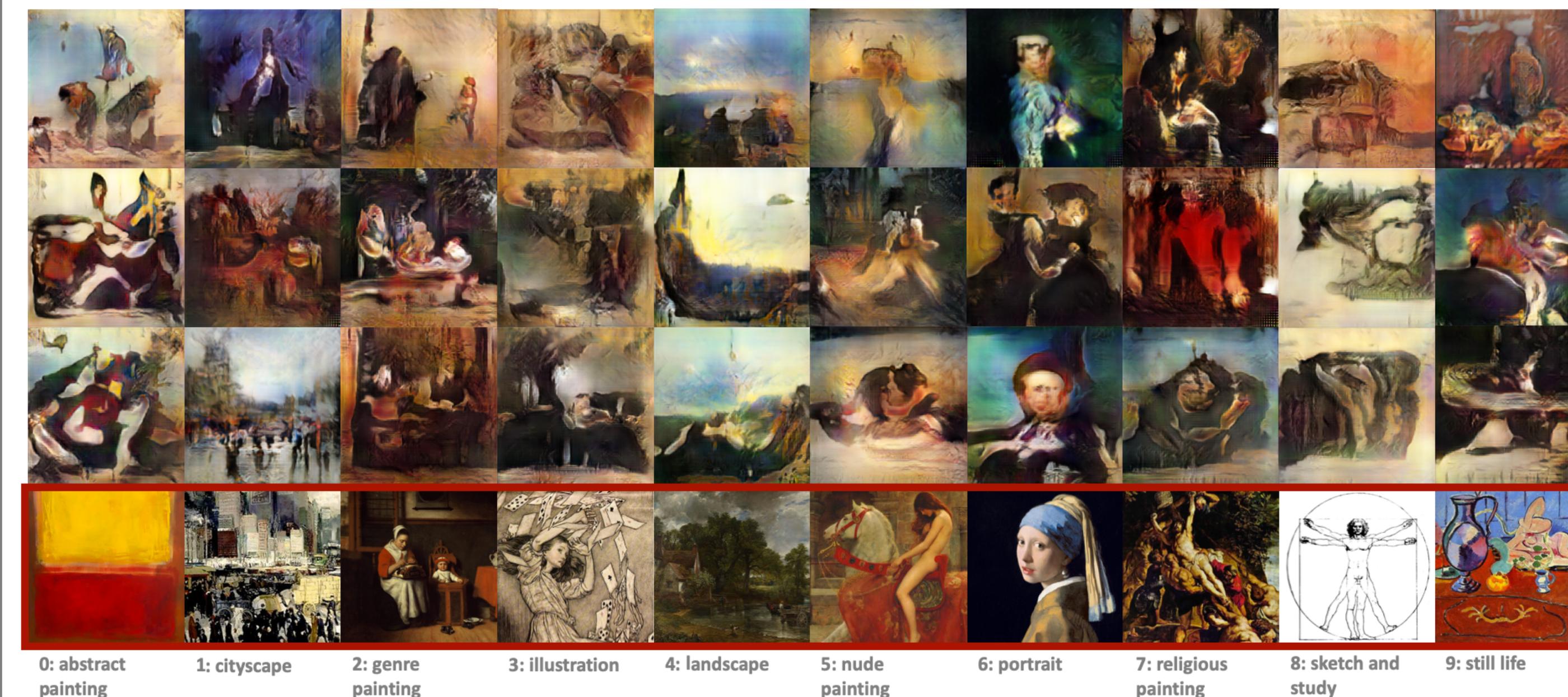
Figure 1: Distribution of images between each genre label. Note the class imbalance

Model Architecture



- Two-stage model:** Stage-I generates 64x64 inputs to Stage-II, Stage-II generates final 256x256 outputs.
- Discriminator Loss** = C.E. Loss (real vs. fake) + C.E. Loss (real labels) + Reconstruction Loss (real)
- Generator Loss** = C.E. Loss (fake labels) + Recon. Loss (fake)

Qualitative Analysis



Experimental Results

GANtor Model	Class category	Inception-v3 Metrics	
		I.S.	Top-1 (%)
v-1, Stage-I	genre	1.34 ± 0.08	21.47
v-2, Stage-I	genre	1.30 ± 0.04	10.16
v-1, Stage-II	style	1.35 ± 0.01	3.67
v-2, Stage-II	genre (epoch 50)	1.41 ± 0.03	17.46
v-2, Stage-II	genre (epoch 60)	1.44 ± 0.03	9.05

Figure 4: Inception Score and Top-1 Accuracy for each model.

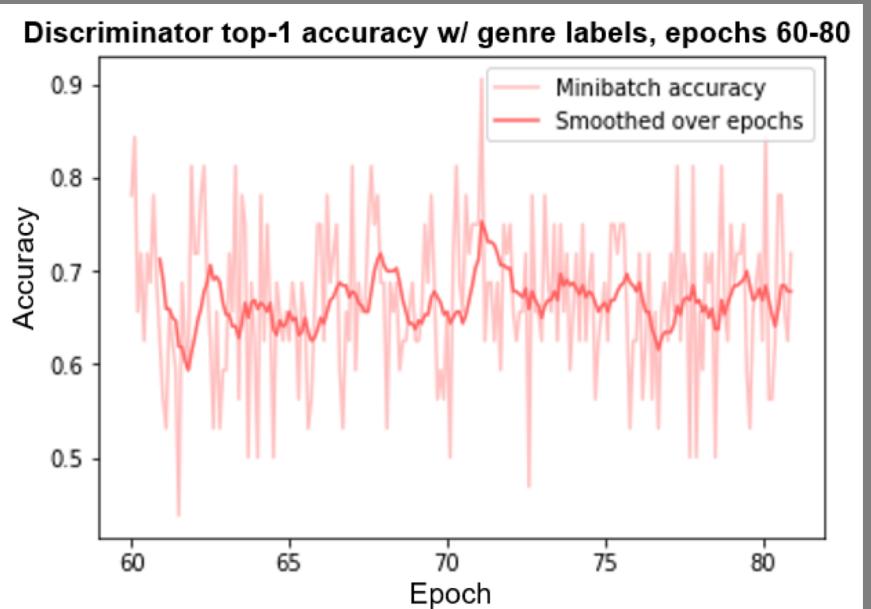


Figure 5: Discriminator performance over the final 20 epochs.

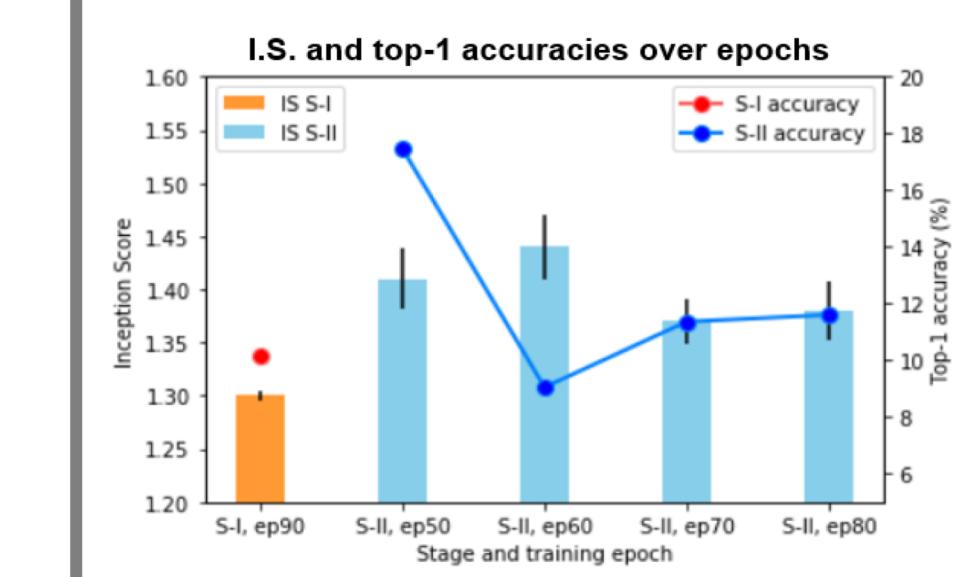


Figure 6: Inception Scores and Top-1 accuracies for different epochs of training (**GANtor-v2**).

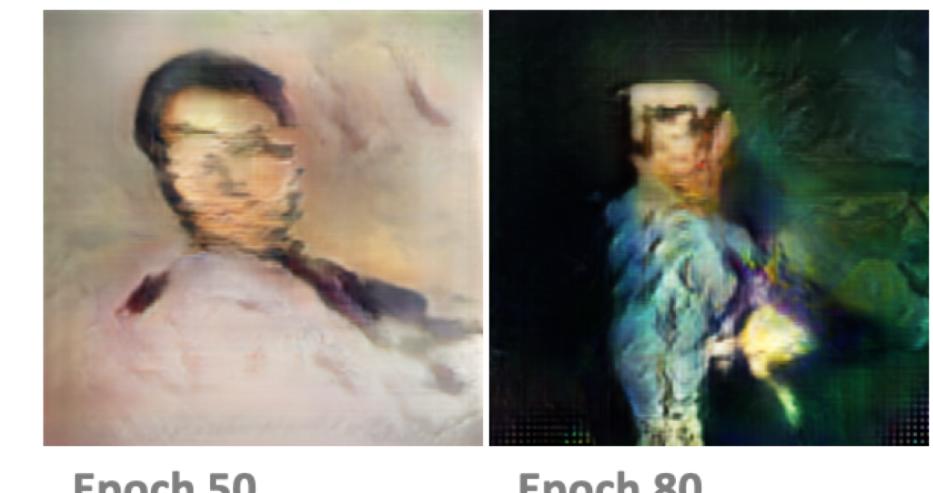


Figure 7: A comparison between portrait class outputs from epochs 50 and 80 (**GANtor-v2**)

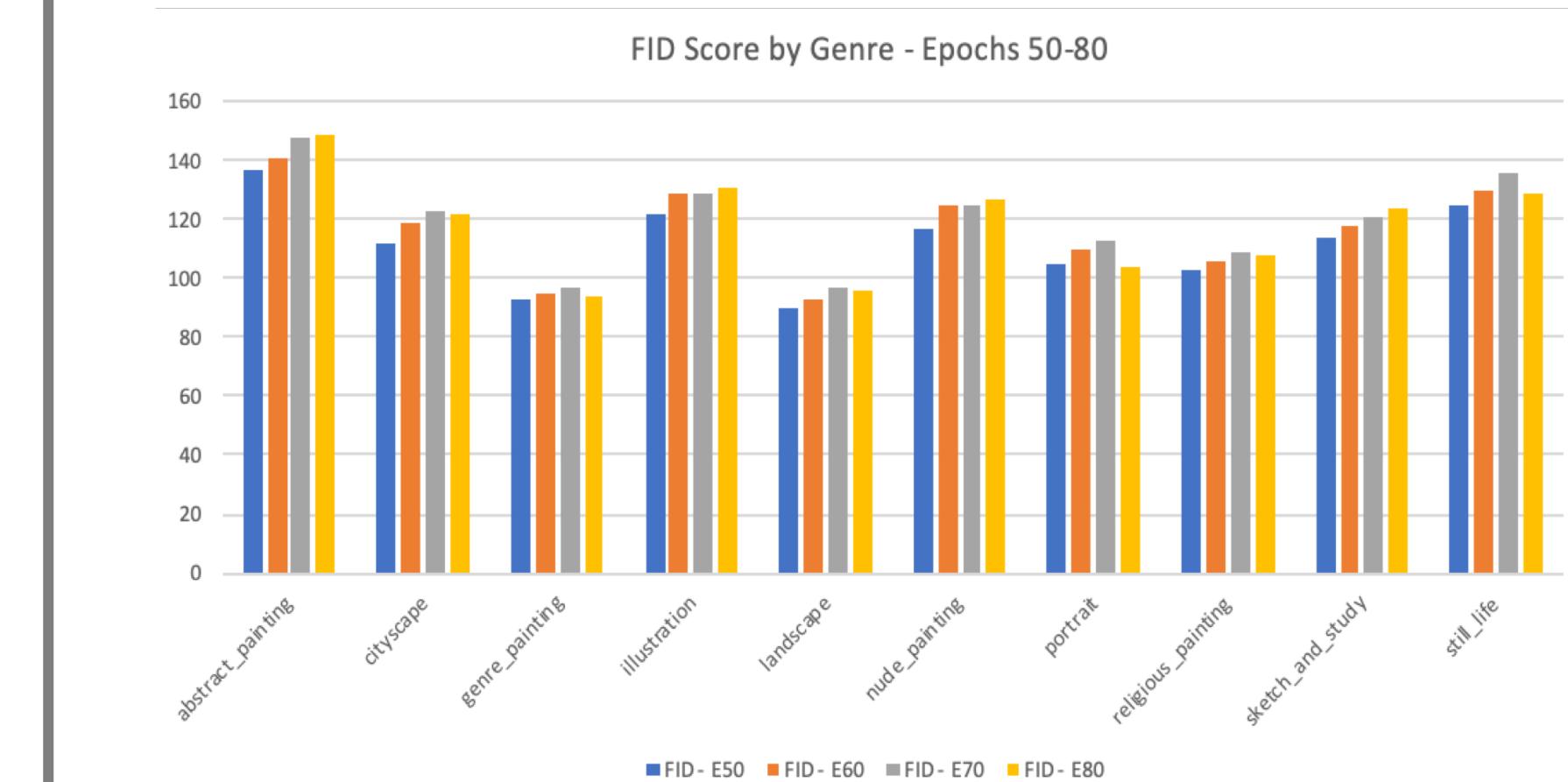


Figure 8: FID scores calculated across training epochs 50-80 on **GANtor-v2 Stage-II** for each genre.

References

- [1] Wei Ren Tan, Chee Seng Chan, Hernan E. Aguirre, and Kiyoshi Tanaka. Artgan: Artwork synthesis with conditional categorial GANs. *CoRR*, abs/1702.03410, 2017
- [2] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In 2017 IEEE International Conference on Computer Vision (ICCV)