DreamDiffusion: High-Quality EEG-to-Image Generation with Temporal Masked Signal Modeling and CLIP Alignment –Supplementary Materials–

Yunpeng Bai[1,5], Xintao Wang[3], Yan-Pei Cao[4], Yixiao Ge[2], Chun Yuan[1](✉), and Ying Shan[2]

[1] Tsinghua Shenzhen International Graduate School, China
[2] ARC Lab, Tencent PCG, [3] Kuaishou Technology, [4] VAST
[5] The University of Texas at Austin
https://github.com/bbaaii/DreamDiffusion

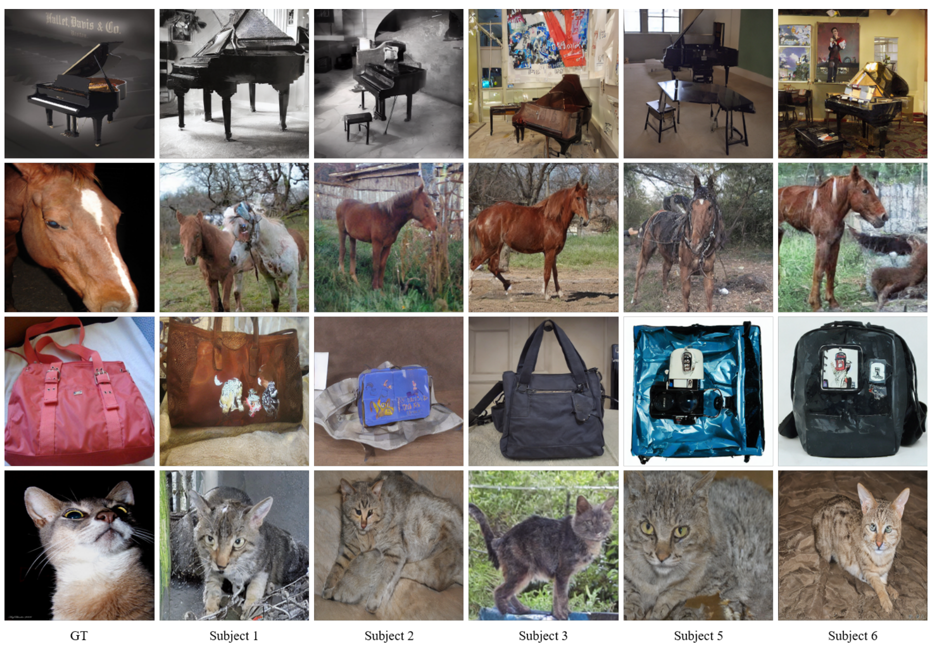

Fig. 1: The images generated from the signals of remaining subjects.

1 More experimental results

The results presented in the paper are generated from the signals of subject 4. Here, we provide the results of the remaining subjects, with the generated images shown in Figure 1. The quantitative evaluation results are shown in Table 1.

Table 1: The quantitative evaluation result.

Subject	1	2	3	5	6
Acc (%)	33.2	37.1	39.2	36.5	28.3
FID	3.96	3.87	3.12	3.78	4.11
IS	26.65	27.48	29.85	28.20	25.97

1.1 Comparison with previous method

We have added comparisons with several similar methods here, which are EEG2-IMAGE [3], Kumari et al. [2], and DCAE [4]. It's important to note that EEG2I-MAGE and the other two works did not utilize the ImagenetEEG dataset. Brain2Image is one of the few works that have been validated on the Imagene-tEEG dataset over the past six years and can be used for comparison. We also conducted experiments with other works on this dataset, but their performance was not satisfactory. The comparison results are displayed in Figure 2 and Table 2, respectively.

Fig. 2: Comparisons with other methods.

Table 2: More comparisons.

Methods	FID ↓	IS ↑	PSNR ↑	SSIM ↑	LPIPS ↓
Kumari et al.	48.62	4.21	9.8	0.142	0.893
DCAE	27.65	4.38	10.6	0.181	0.831
EEG2IMAGE	21.53	5.28	11.1	0.224	0.715
Ours	3.61	28.54	14.6	0.267	0.644

Fig. 3: Results on other dataset.

Table 3: Result on other dataset.

Metrics	FID ↓	IS ↑	PSNR ↑	SSIM ↑	LPIPS ↓
Ours	2.14	29.33	15.8	0.283	0.615

1.2 Results on other dataset

We also validate our method on another dataset [1]. The results on this dataset are shown on Figure 3 and Table 3. These results are sufficient to demonstrate the robustness of our method.

2 Pretraining visualization

We visualize the reconstruction results of one channel from the EEG data in Figure 4. We can observe that the overall trend is accurate, but the details are influenced by the dataset, as the EEG signals in these datasets are relatively noisy.

References

1. Gifford, A.T., Dwivedi, K., Roig, G., Cichy, R.M.: A large and rich eeg dataset for modeling human visual object recognition. NeuroImage **264**, 119754 (2022)
2. Kumari, N., Anwar, S., Bhattacharjee, V., Sahana, S.K.: Visually evoked brain signals guided image regeneration using gan variants. Multimedia Tools and Applications **82**(21), 32259–32279 (2023)
3. Singh, P., Pandey, P., Miyapuram, K., Raman, S.: Eeg2image: image reconstruction from eeg brain signals. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
4. Zeng, H., Xia, N., Tao, M., Pan, D., Zheng, H., Wang, C., Xu, F., Zakaria, W., Dai, G.: Dcae: A dual conditional autoencoder framework for the reconstruction from eeg into image. Biomedical Signal Processing and Control **81**, 104440 (2023)

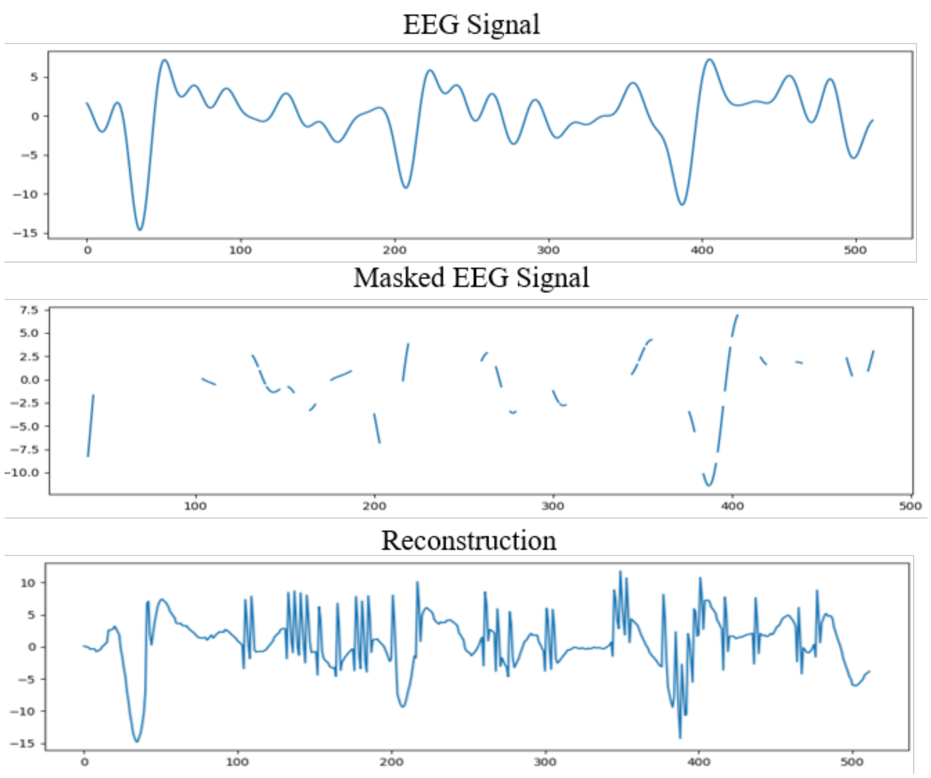

Fig. 4: Masked signals modeling with large-scale noisy EEG data.