EEGMirror: Leveraging EEG data in the wild via Montage-Agnostic Self-Supervision for EEG to Video Decoding

Supplementary Material

7. More Dataset Information

We used the **S**JTU **E**E**G** **D**ataset for **D**ynamic **V**ision (SEED-DV) [31] for all experiments. This dataset consists of high-temporal-resolution EEG records collected from 20 healthy adult subject (10 males and 10 females). Each Subject watched 7 video blocks and each block contains 5 video clips for each 40 concept, resulting in 1400 EEG-video pairs in total. The 40 concepts investigated in SEED-DV cover 9 coarser classes: {*land animal, water animal, plant, exercise, human, natural scene, food, musical instrument, transportation*}. Consequently, for each concept, there is 7×5 EEG-video pairs, we use the first 6×5 pairs for training and the last 5 pairs for testing.

8. Implementation Details

8.1. EEG Signals Preprocessing

SEED-DV dataset. The raw data was recorded with the 62-channel EEG cap with a sample rate of 1000 Hz and stored in the continuous EEG data file format (.cnt), a single file for the experiment of each subject. We applied the 0.1-100 Hz band-pass filter to filter out the DC interference and very high-frequency interference and down-sampled the EEG data to 200Hz to accelerate computations.

Other EEG dataset. We use the same processing method as LaBraM [18]. We apply minimal yet essential preprocessing steps to the EEG signals. First, the signals are filtered between 0.1 Hz and 75 Hz to eliminate low-frequency noise. Next, a 50 Hz notch filter is applied to mitigate power-line interference. The EEG signals are then resampled to 200 Hz. Since EEG values typically range between -0.1 mV and 0.1 mV, we normalize the data by scaling the unit to 0.1 mV, ensuring the values predominantly fall within the range of -1 to 1.

8.2. Implementation Details

We reconstruct a two-second video clip from the corresponding two-second EEG segment. For efficient training and testing, we down-sampled the 24 FPS 1080p original videos to a small video of resolution of 512×288 (16:9) with 3 FPS, resulting in 6 frames for each video. We use the first 6 blocks from all the sessions as the training set and the last blocks as the testing set in our experiment. The inflated video generation model is fine-tuned using the same setting in [59] on the training set with learning rate of 0.00003 and cosine scheduler for 200 epochs, which takes about 5 hours. The inference is performed with 100 DDIM [44] steps.

9. Ablation Study: Visual Resutls

Visual results of the ablation studies are shown in Fig. 8. The Full model is pretrained on all datasets and finetune with contrastive loss. In contrastive learning ablation, we finetuned our model without contrastive loss. Similar to the numeric evaluations, using incomplete loss function gave an suboptimal results compared to full model, because the color of the cat cannot be recovered satisfactorily. However, without brain encoder, our method generated visually meaningless results.

10. Fail Cases

Some failure samples are displayed in Figure 9. These failures are typically caused by the inability of the model to infer either the semantic information or the low-level visual information correctly, resulting the irrelevantly generated videos. However, we can still see from these failed examples that the model reconstructs some features of the real video, like scenes and colors. For example, a turtle in the sea was recovered as a shark in the sea.

11. Broader Impacts

Reconstructing dynamic visual perception from brain activity advances our understanding of the brain's visual system. EEG, a physiological signal widely used in clinical practice and brain-computer interfaces, offers a portable and cost-effective alternative to non-portable and expensive neuroimaging techniques like fMRI and MEG. Our work provides a convenient and affordable solution for decoding visual information from brain activity, enabling the visualization of mental processes. This technique offers a novel approach to exploring the inner world of patients with mental illnesses such as autism and depression. However, this technology also raises concerns. Personal privacy could be compromised if brain activity data is accessed and exploited by malicious actors to "read minds" from EEG signals without consent. To address this, governments and medical institutions must establish stricter regulations to safeguard the privacy of individuals' biological data.

12. Limitations

Our framework has been evaluated under subject-dependent settings, but its cross-subject applicability remains unexplored due to individual variations. Future work could focus on enhancing the transferability of the framework.

Figure 8. Reconstruction samples for ablation studies. The Full model uses full modality contrastive learning and pretrain on all datasets.

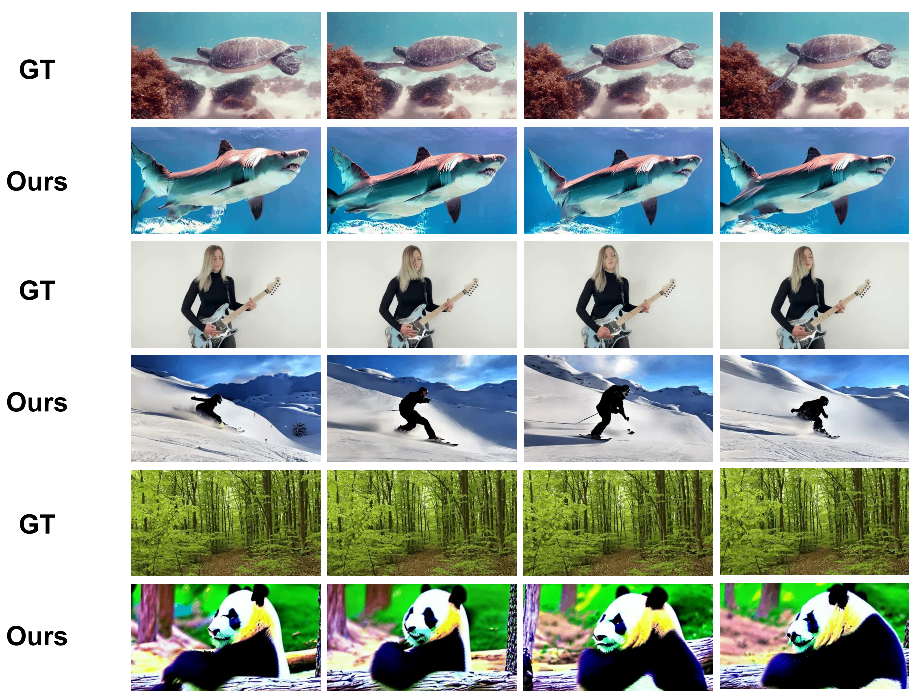

Figure 9. Fail cases.