Psychometry: An Omnifit Model for Image Reconstruction from Human Brain Activity

Supplementary Material

This supplementary material provides more details of our implementations, the NSD dataset used, fMRI preprocessing techniques, the Omni MoE algorithm, and the Ecphory algorithm, as well as additional reconstruction results, quantitative results, broader impacts, and limitations. These topics are organized as follows:

- §A: Dataset and fMRI Data Preprocessing Details
- §B: More Reconstruction Results
- §C: More Implementation Details
- §D: Social Impacts
- §E: Limitations

A. Dataset and fMRI Preprocessing Details

The Natural Scenes Dataset[1] (NSD) [2] is currently the largest publicly available fMRI dataset. It contains densely sampled fMRI data from 8 participants (2 male and 6 female, age 19-32 years) viewing a total of 73,000 RGB images. Each participant saw a total of 10,000 unique images (repeated 3 times each) across 20 to 40 7T MRI sessions (whole-brain gradient-echo EPI, 1.8-mm iso-voxel and 1.6s TR). Each session consisted of 12 runs of 5 minutes each, where each image was seen for 3 s, with a 1-s blank interval between two successive image presentations. Among the 8 participants, only 4 (namely subject 1, subject 2, subject 5 and subject 7) completed all sessions. The images utilized in the NSD dataset were sourced from the Microsoft Common Objects in Context (COCO) database [33], square-cropped, and displayed at a size of $8.4° \times 8.4°$. A set of 982 among them was shared across all subjects while the remaining images for each individual were mutually exclusive across subjects. Existing fMRI-to-image reconstruction studies [21, 52, 62] that used NSD typically follow the same procedure: training subject-specific models for the four participants who finished all scanning sessions and employing a `test` set that corresponds to the common 982 images shown to each participant. However, in our experiments, *Psychometry* trains a single model on the amalgamated data for the four subjects. Specifically, the `train` set contains 35,436 images and 99,920 fMRI trials, and the `test` set consists of 982 images and 11,080 fMRI trials.

Pre-processing of the fMRI data involved performing temporal interpolation to correct slice time differences and spatial interpolation to account for head motion artifacts. Subsequently, a general linear model was employed to estimate single-trial beta weights. The NSD dataset also encompasses cortical surface reconstructions generated using FreeSurfer[2], with both volume- and surface-based versions of the beta weights being created for further analysis and interpretation. We masked preprocessed fMRI signals using the provided NSDGeneral ROI (Region-of-Interest) mask in 1.8 mm resolution. The ROI consists of 15,724, 14,278, 13,039, and 12,682 voxels for the 4 subjects respectively, and includes many visual areas from the early visual cortex to higher visual areas. To handle the different voxel numbers of different subjects, all fMRI data are first padded to the maximum length in a wrap-around manner. Additionally, training fMRI is normalized to have zero mean and unit standard deviation.

B. More Reconstruction Results

We provide more visual results that compare *Psychometry* to the state-of-the-art (MindEye [52]) in Figure 7. As can be seen, *Psychometry* is able to reconstruct high-quality and realistic images when utilizing a unified model (*UM*) trained on the amalgamated fMRI data from different subjects.

C. More Implementation Details

The framework of the fMRI representation learning network used in our *Psychometry* model is shown in Figure 2. Before training the network, we employ a RoI embedding layer [9, 46] to separately encode fMRI signals into patches from different RoI regions. The encoded features are utilized as input to the transformer blocks. We present the PyTorch implementation of both our *Omni MoE* layer and the *Ecphory* inference strategy in Algorithm 1 and Algorithm 2 respectively. During training, we train the model for 200 epochs with a batch size of 64. We update the parameters using AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.9999$, $\epsilon = 10^{-8}$, and $lr = 5 \times 10^{-4}$. During inference, α as the mix-up weight to enhance the output embedding is set as 0.5. When conducting fMRI-to-image generation, the forward and reverse diffusion blocks in the Versatile Diffusion model [67] are all pretrained and frozen. To ensure reproducibility and foster future research, our full implementation will be released after acceptance.

C.1. Omni MoE Algorithm

```python
def omni_moe(inputs, mix_experts, ssparams,
    subj_name):
  # Obtain the subject-specific parameter
```

[1] http://naturalscenesdataset.org/

[2] http://surfer.nmr.mgh.harvard.edu/

```
  subj_param = ssparams[subj_name]
  # Compute subject-specific features
  subj_feat  = torch.einsum("bmc,ce->bme", inputs
    , subj_param)
  # Compute splitting weights, Eq. 2
  sweights   = torch.nn.softmax(subj_feat, dim=1)
  # Obtain the splitted features, Eq.3
  m_feat     = torch.einsum("bmc,bme->bec",
    inputs, sweights)
  q_feat     = torch.stack([f_e(m_feat[:, i, :])
    for i, f_e in enumerate(mix_experts)], dim=1)
  # Compute lumping weights, Eq. 4
  lweights   = torch.nn.softmax(subj_feat, dim=2)
  # Compute output tokens, Eq. 5
  outputs    = torch.einsum("bec,bme->bmc",
    q_feat, lweights)
  return outputs
```

Algorithm 1. Pytorch implementation of our Omni MoE layer

C.2. Ecphory Algorithm

```
def ecphory(predictions, memories, K, alpha):
  # Normalize predictions and memories
  prediction_norm = F.normalize(predictions[:,0],
    dim=-1)
  memory_norm = F.normalize(memories[:,0],dim=-1)
  # Compute the similarities
  similarity    = prediction_norm @ memory_norm.T
  # Select memories with top K similarity scores
  topK_indexes  = torch.topk(similarity.flatten()
    , K).indices
  # In practice, we set K=1
  en_prediction = alpha * predictions + (1-alpha)
     * memories[topK_indexes]
  return en_prediction
```

Algorithm 2. Pytorch implementation of our Ecphory strategy

D. Social Impacts

This paper introduces *Psychometry*, a unified model for reconstructing images from human brain activity via fMRI data. The model is designed to be omnifit, capable of handling data from various individuals, and represents a significant advancement in the field of Brain-Computer Interface. It also has the potential to greatly enhance our understanding of brain function and could lead to breakthroughs in medical diagnostics and treatment, particularly in neurological disorders. Its ability to capture both inter-subject commonality and individual specificity can contribute to personalized medicine approaches. Ultimately, technology capable of interpreting human brain activity is expected to revolutionize how we interact with technology, leading to new interfaces that directly connect with human thought.

E. Limitations

Currently, *Psychometry* is specifically designed for image reconstruction via fMRI data. It is interesting to extend our method to handle more complicated human brain activity signals, for example, magnetoencephalography (MEG) and electroencephalography (EEG) signals. Moreover, while *Psychometry* offers substantial benefits, it is crucial to ensure that appropriate safeguards are in place to protect individual privacy, especially when amalgamating sensitive fMRI data from different subjects. We leave these as a part of our future work.

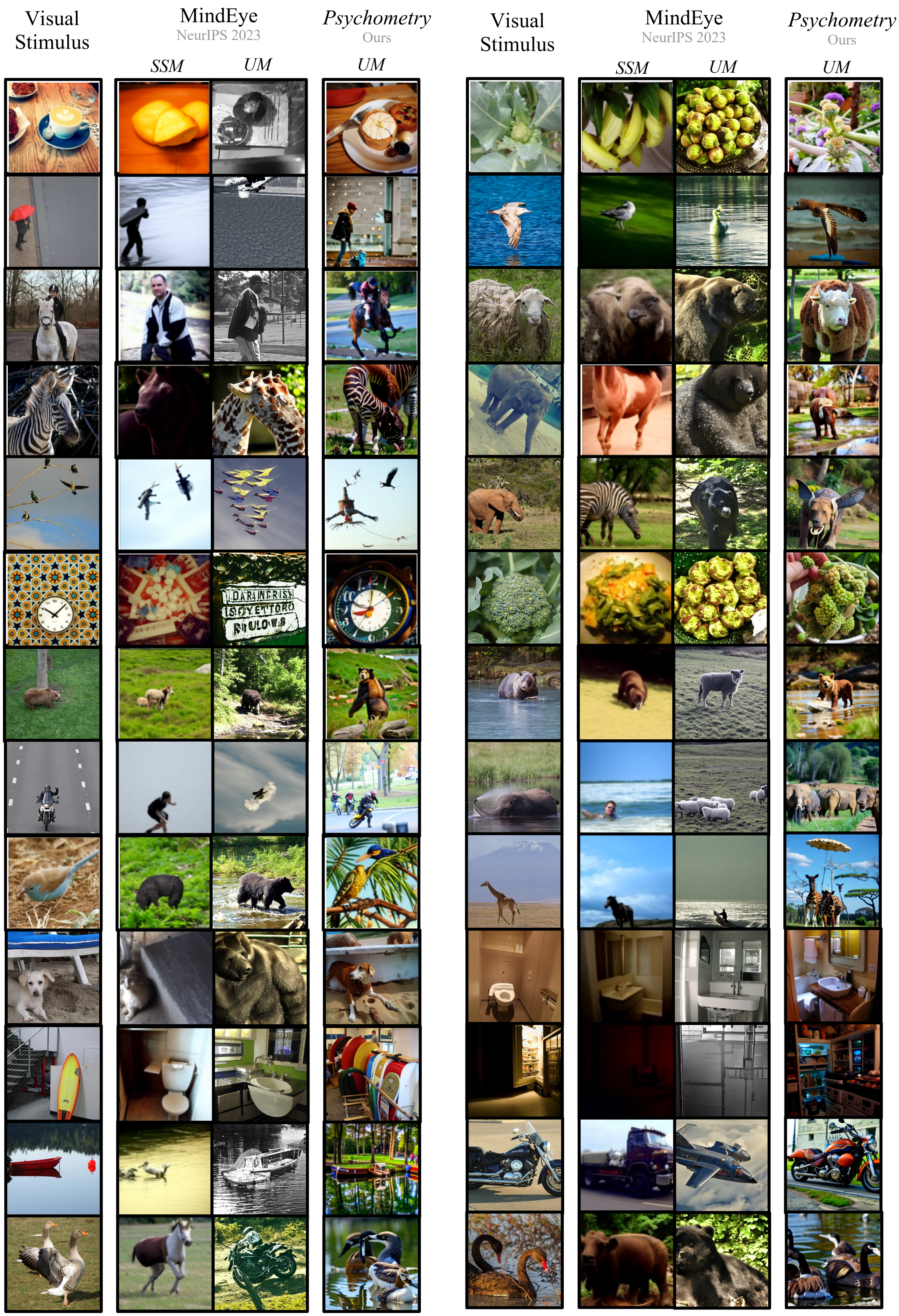

Figure 7. Additional visual comparison results on NSD `test`.