

Project Proposal: Reproducibility Challenge

EECS 6322: Neural Networks and Deep Learning
Winter 2024

Team Member Names:
Yiqian Liu - 220616470
SeyedBahman Rouhani - 219509736

February 12, 2024

Summary

The chosen paper is Rethinking Visual Reconstruction: Experience-Based Content Completion Guided by Visual Cues by Chen et. al [1].

Reconstructing images from fMRI signals is a useful ability, however, extracting the seen image from recorded fMRI signals is usually hindered by limitations in different parts of the procedure: 1. The human eye has a limited field of vision and different sensitivity of receptors inside that field of vision, 2. there are far less ganglion cells in the brain with respect to photoreceptor cells, which means the visual information will need to be compressed in the brain, also 3. the fMRI recording techniques are not perfect and may not preserve all brain activity information. Due to these limitations, learning to reconstruct images from fMRI signals is not as easy as learning a mapping from signals to images.

This paper introduces VQ-fMRI. This method uses a VQ-VAE to learn a shared discrete representation between the images and the fMRI signals, which the decoder then learns to reconstruct the images from. The paper also introduces a self-supervised token-to-token inpainting network for visual content completion. This network leverages context and prior knowledge from the discrete latent space to infer uncertain or missing content.

The papers results are significantly better than those of competing methods based on qualitative assessment and the two quantitative metrics used in the paper (One to one evaluation of reconstructed images based on *SSIM*, *PSNR* and *PCC* and also a Pairwise Evaluation to check if the reconstructed image resembles the original image significantly more than a random image in dataset).

1 Deep Learning Framework

To reproduce this method, we will use PyTorch framework.

2 Datasets

The paper conducts experiments and reports its results on Generic Object Decoding (GoD) dataset introduced by Horikawa and Kamitani [2]. The dataset is fMRI recordings of 5 subjects while looking at ImageNet dataset.

3 List of Experiments

There are two general experiments and an ablation study part in this paper. The main claim of the paper is denoted in the first experiment where they demonstrate the superiority of their model compared to other state of the art models. We will conduct this part of the experiment.

4 Compute Resources

The proposed network in the paper uses VQ-VAEs and U-NETs. The implementation details in the paper indicate that none of these networks are specifically large (most network have about 5 or 6 layers, and the initial image size is 500×500). What's more, the network is not trained end to end and training process contains several phases. Considering these, a NVIDIA GeForce GTX 1080 GPU should be enough for training the network.

5 Anticipated Separation of Work

We divide expect the implementation to mainly included the following modules and subtasks, in front of each sub-task, the name of the team member expected to finish that part is written.

1. Encoder-Decoders: Yiqian Liu
2. U-Nets: Bahman Rouhani
3. Vector Quantization: Bahman Rouhani
4. Classifier: Yiqian Liu
5. Codebook module: Bahman Rouhani
6. Data-loader: Yiqian Liu
7. Overall pipeline:
 - (a) phase 1: Yiqian Liu
 - (b) phase 2: Yiqian Liu & Bahman Rouhani
 - (c) phase 3: Bahman Rouhani

References

References

- [1] Jiaxuan Chen, Yu Qi, and Gang Pan. Rethinking visual reconstruction: Experience-based content completion guided by visual cues. 2023.
- [2] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8(1):15037, 2017.