

## **A Report of System Design (2<sup>nd</sup>)**

in Image & Auditory Reconstruction From EEG Brain Signals



Graduate Project 1 (3183) - Karpjoo Jeong

Team 7, EEmage – Advisor : Eun Yi Kim

Computer Science and Engineering 201811218 Hyun Woo Lee

Computer Science and Engineering 201811182 Wonjun Park

Computer Science and Engineering 201911193 Esan Woo

## Abstract

이 보고서는 EEG (Electroencephalography) Brain Signals로부터 인간의 시각 이미지와 청각 소리를 재현하는 연구의 시스템 설계서입니다. EEG data는 Brain-Computer Interface (BCI) 분야에서 각광받고 있는 비용 효율적인 데이터 중 하나로, 이를 통해 시각과 청각을 재현할 수 있다는 사실이 보여지고 있습니다. 이 연구는 다음의 두 가지 가설을 가지고 있습니다; 1) *이미지와 소리에 대한 인공지능 모델 학습이 상관성을 지닌다.* 2) *인간 체계를 모방하기 위해서는 Supervision이 아닌 다른 방법이 필요하다.* 이를 토대로, 연구는 궁극적으로 (Autoencoder 등의) Unsupervised Learning을 적용한 A Universal EEG Encoder의 제안을 목표로 합니다. 본 보고서의 본론에서는 이 Universal Encoder의 설계와, 현재 초점을 맞추고 있는 부분을 중심으로 한 설계를 소개합니다.

## I. Introduction

본 연구의 시발점은 “인간의 시각 체계가 형성되는 데에는 Supervision이 필요하지 않다”는 가설이었습니다. 이로부터 시각 이미지 재현 프레임워크 중 인간의 시각 체계에 해당하는 인코딩 방법에 Supervised Learning을 사용하는 것이 실제 인간과 맞지 않다는 결론을 도출하였고, 궁극적으로 이를 탈피하는 새로운 프레임워크를 제안하고자 합니다.

이미지와 소리는 서로 다른 도메인에 속하는 데이터입니다. 하지만 이미지와 소리는 데이터에 연속성이 있다는 점에서 그 상관성이 있습니다. 실제로 {1}, {2}와 같이 이 둘을 함께 분석하려는 시도들이 있었고, 이 연구 또한 이 가설을 채택하고자 합니다.

ICASSP 2024: A EEG Auditory Challenge {3}은 EEG와 소리 데이터 간의 상관관계를 찾는 공개 챌린지입니다. 이 챌린지의 Deadline이 2023년 12월 26일이기 때문에, 본 연구에서는 현재 청각 데이터의 재현에 초점을 맞추고 있습니다. 이 챌린지를 해결하기 위한 아이디어 또한 연구의 가설로부터 나옵니다. *가설 1*에 의거해, EEG-ChannelNet {4}를 챌린지의 인코더에 적용해보는 실험을 진행하고 있습니다.

결론적으로 본 연구에서는 EEG 데이터로부터 시각과 청각을 모두 재현할 수 있는 Universal Encoder를 제안하고자 합니다. 이 연구를 통해 뇌의 비밀을 풀고, 신체가 불편한 이들에게 도움이 되며, 범용적인 Brain-Computer Interface (BCI)에 한 단계 더 다가가기 기대합니다.

## II. System Overview

본 섹션에서는 *II.A. A Universal EEG Encoder*와 *II.B. An Architecture of Auditory Specialized Encoder*에 대해 소개하고자 합니다. 본 연구에서 궁극적으로 제안하고자 하는 범용 EEG 인코더를 먼저 설명하고, 현재 챌린지에서 집중하여 풀고자 하는 모델을 설명합니다. *II.B*는 이상적인 성능의 *II.A*를 제안하기 위해, EEG와 Audio의 상관성을 분석, 입증하고 효과적인 모델 구조를 찾기 위한 필수적인 단계입니다.

### A. A Universal EEG Encoder

현재까지 설계한 구조는 다음과 같습니다.

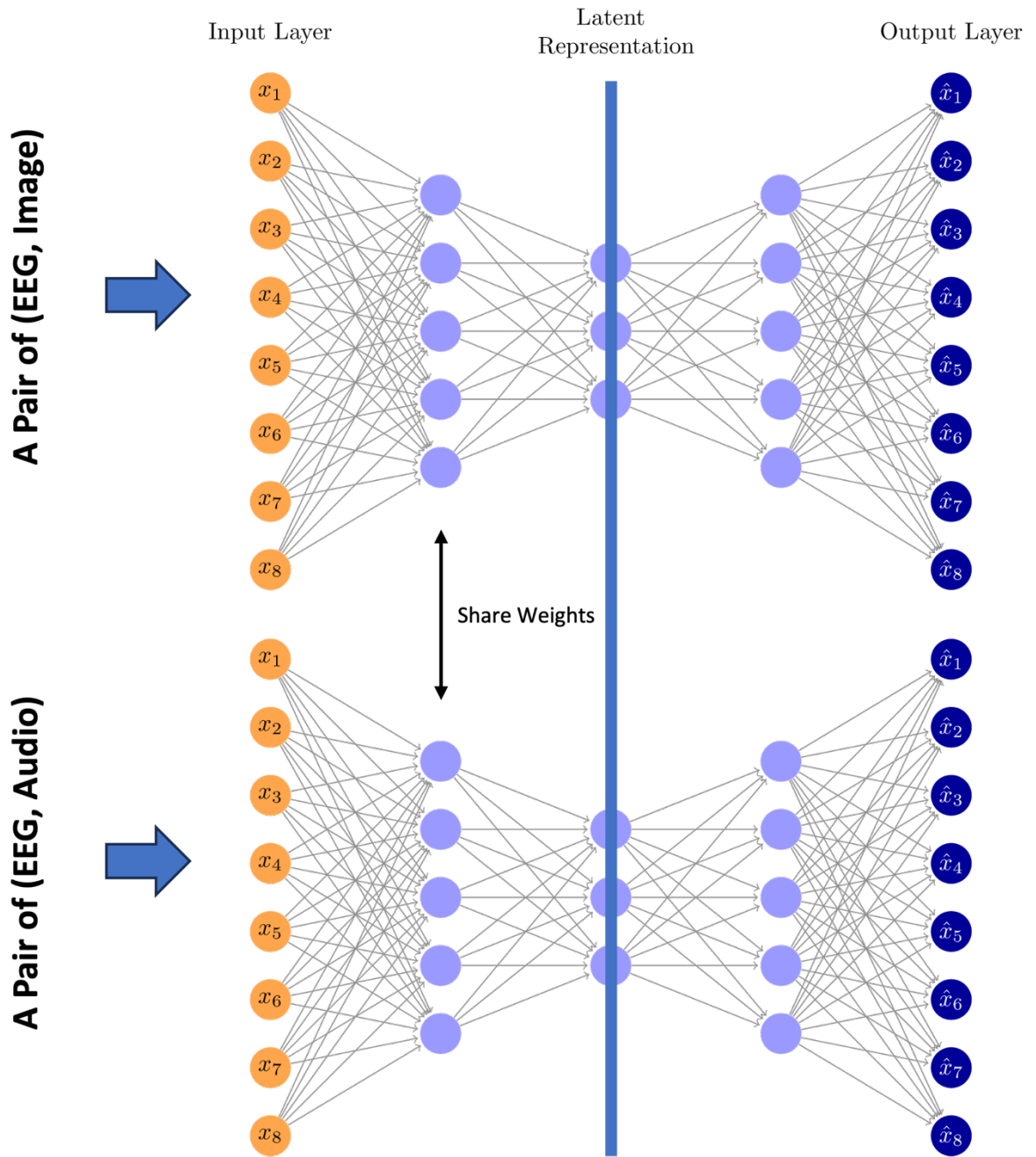


Figure 1. A Model of an Universal Encoder Based on Autoencoder

두 개의 Autoencoder를 이어붙여 (EEG, Image)와 (EEG, Audio)를 함께 Input으로 받습니다. Encoder 부분에서는 두 Autoencoder가 가중치를 공유하도록 합니다. Decoder에서는 각각 Image와 Audio에 대한 loss value를 계산하고, 둘의 가중평균 값을 최종 loss 값으로 합니다.

범용 EEG 인코더가 될 부분은 Autoencoder의 앞부분, 가중치를 공유하는 부분입니다. Image와 Audio에 대해 적절한 Feature Vector를 뽑아내는 것을 목표로 합니다. 이는 EEG로부터 뇌파의 일정한 규칙을 찾아낼 수 있을 것이라는 기대를 수반합니다.

## B. An Architecture of Auditory Specialized Encoder

현재 챌린지를 풀려고 시도하는 인코더의 설계는 다음과 같습니다.

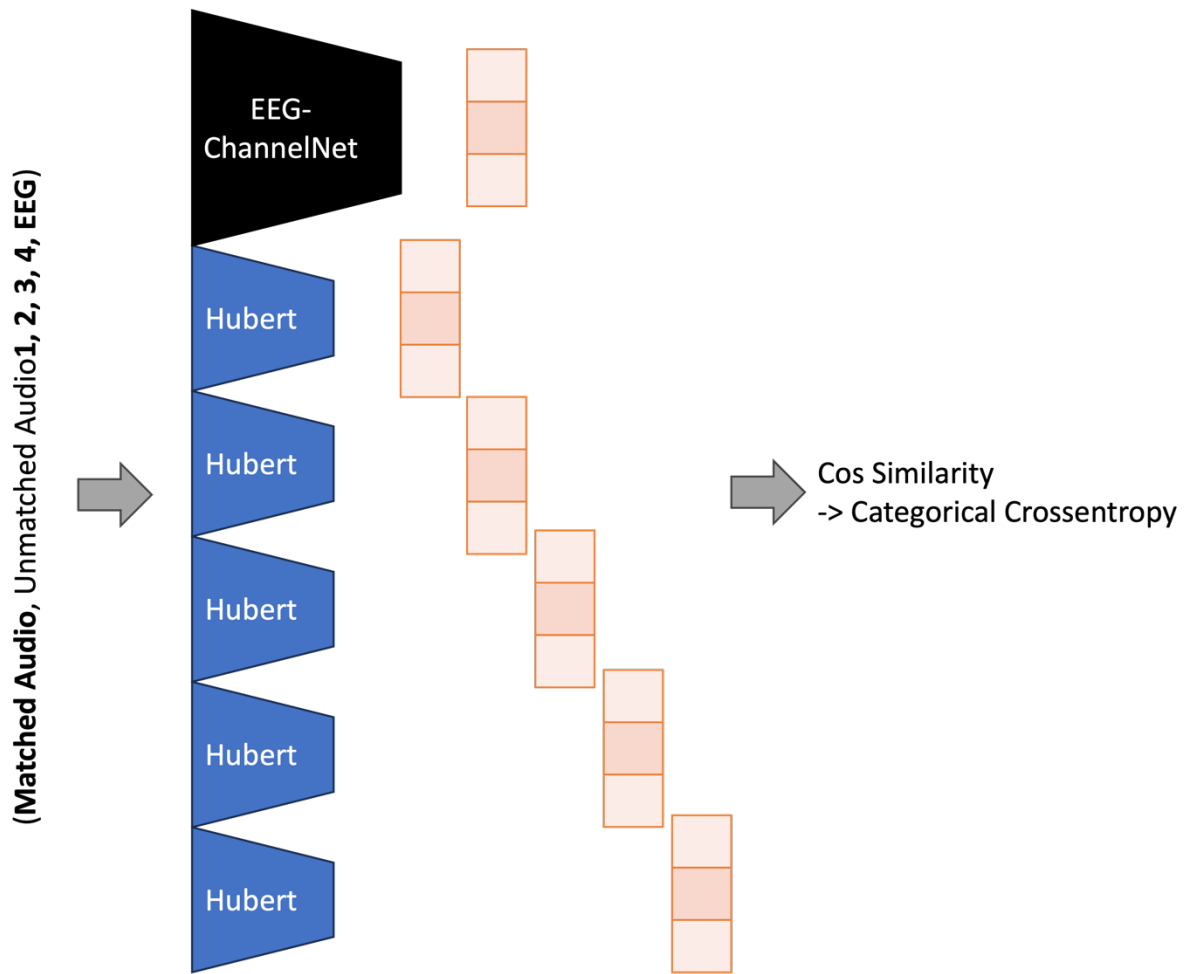


Figure 2. An Architecture of Auditory Specialized Encoder

Input은 EEG 데이터, EEG 데이터와 매치되는 Audio 1개, 불일치 하는 Audio 4개입니다. 현재 본 연구에서 사용하는 Encoder의 종류는 EEG-ChannelNet (4)와 Hubert (5) 입니다. EEG-ChannelNet은 EEG로부터 시각 이미지 재현을 효과적으로 수행하는 것이 입증된 인코더입니다. Hubert는 Self-Supervised 기반의 모델로, 본 설계에서는 Encoder 부분만 가져와 사용하고 있습니다. 각 인코더로부터 Input Data의 Vector를 뽑아내면, 이들의 Cosine Similarity 값을 구하여 유사도를 측정합니다. Loss Function으로는 Categorical Crossentropy를 사용하여 Multi-label 데이터 중 매칭되는 것에 학습되도록 했습니다.

이 챌린지를 통해 EEG-Image 인코더가 EEG-Audio 인코더에도 효과적으로 사용될 수 있음을 입증하고자 합니다. 이후에는 그 반대의 경우도 입증한 후, 두 데이터를 병합해 범용 EEG Encoder를 제작할 계획입니다.

### III. Conclusion

//A.는 현재 가상 설계 단계이고, //B.는 실험 도중에 있는 단계입니다. 아직 많은 epochs를 돌려보진 못했지만, epoch가 증가함에 따라 training accuracy가 90% 이상으로 증가하고 training loss value가 계속적으로 감소함을 확인하였습니다. 이를 토대로 모델이 데이터의 특징

은 뽑아내고 있음을 알 수 있습니다. 하지만 현재까지 진행된 epoch가 약 6-7 정도인 점과 Validation Accuracy, Loss Value는 적절한 값을 보여주지 못하고 있는 문제가 발생하고 있습니다. //B.에서 적절한 수정이 가해질 수도 있음을 시사합니다.

본 프로젝트는 궁극적으로 인간의 시각과 청각을 재현할 수 있는 Universal EEG Encoder의 제안을 목표로 합니다. 그 과정에서 (EEG-Image), (EEG-Audio) 데이터의 상관관계를 파악하고자 (3)의 챌린지를 준비하고 있습니다. 챌린지의 결과와 상관없이, 본 연구의 궁극적인 목표에 도움이 될 것이라 기대합니다.

## Reference

다음의 목록은 MLA 포맷을 따랐습니다.

- [1] Likhoshesterov, Valerii, et al. "Polyvit: Co-training vision transformers on images, videos and audio." *arXiv preprint arXiv:2111.12993* (2021).
- [2] Gong, Yuan, et al. "Uavm: Towards unifying audio and visual models." *IEEE Signal Processing Letters* 29 (2022): 2437-2441.
- [3] "Auditory EEG Challenge 2024." *ICASSP 2024 SP Grand Challenges*, <https://exporl.github.io/auditory-eeeg-challenge-2024/>.
- [4] Palazzo, Simone, et al. "Decoding brain representations by multimodal learning of neural activity and visual features." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11 (2020): 3833-3849.
- [5] Hsu, Wei-Ning, et al. "Hubert: Self-supervised speech representation learning by masked prediction of hidden units." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021): 3451-3460.