Transformer Architecture Details for Brain Decoder

fMRI Encoder

Input Projection (3092 → 512)

Multi-Head Attention 1

Layer Norm + Residual

Feed Forward

Multi-Head Attention 2

Layer Norm + Residual

Feed Forward

Multi-Head Attention 3

Layer Norm + Residual

Feed Forward

Multi-Head Attention Details:

- Number of heads: 8
- Head dimension: 64 (512/8)
- Query, Key, Value projections
- Scaled dot-product attention
- Dropout: 0.1
- Residual connections
- Layer normalization

Cross-Attention:
- Queries from decoder
- Keys & Values from encoder
- Enables fMRI-stimulus alignment

Mathematical Formulation:

$$\text{Attention}(Q,K,V) = \text{softmax}(QK^T/\sqrt{d_k})V$$

$$\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1,...,\text{head}_h)W^O$$
where $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

$$\text{LayerNorm}(x) = \gamma(x-\mu)/\sigma + \beta$$

Encoded Representation

Transformer Decoder

Self-Attention 1

Cross-Attention

Feed Forward

Self-Attention 2

Cross-Attention

Feed Forward

Self-Attention 3

Cross-Attention

Feed Forward

Self-Attention 4

Cross-Attention

Feed Forward

Self-Attention 5

Cross-Attention

Feed Forward

Self-Attention 6

Cross-Attention

Feed Forward