Bridging Brain Anatomy and Function: Diagnosing Underfitting in Contrastive T1-fMRI Learning

Dhanapanot Rakyutithamkul
bugattmark@gmail.com
github.com/bugattmark/Structure-Function-Brain
linkedin.com/in/bugatt

Abstract

Connecting brain structure to its dynamic function remains a long-standing challenge in neuroscience. The ultimate goal is to build a machine learning pipeline that can infer a participant's functional connectivity, and by extension, cognitive and clinical traits directly from a T1 scan. Such a tool would enable fast and cheap assessments in clinical and research applications.

In this first report, I present a contrastive learning framework to align 3D T1 volumes with ROI[1]-wise fMRI time-series and deploy rigorous diagnostic probes: single-batch memorization, gradient-norm tracking, module freezing, and plateau analysis to validate my implementation and find the core bottleneck: underfitting due to data scarcity vs. model capacity. Training loss plateaus at ≈ 2.773, not the intuitive value of –ln (1/32) because I used 40 samples with a final partial batch of 8, so the reported loss is the arithmetic mean of ln 32 and ln 8: (ln 32 + ln 8)/2$\approx$2.773. Validation loss similarly caps at ln 5$\approx$1.609 due to val batch_size=5. Low cosine similarity ($\leq$0.23) confirms the gap between chance and learned signal, not collapse. I propose a pivot toward more lightweight encoders, data augmentation, and self-supervised pretraining to overcome this barrier.

[1]Region of Interest: areas in the brain where analysis is focused on.

1 Introduction

Connecting structural MRI to functional dynamics is important for understanding how the brain works and developing biomarkers. High-capacity transformers have more expressive power but risk overfitting when applied to small neuroimaging datasets. My contributions:

- **Ultimate goal**: Map T1 scans to ROI-wise time-series.
- **Implementation diagnostics**: Overfit a single batch, track gradient norms, and freeze modules to isolate issues.
- **Key finding**: Loss plateaus at $-\ln(1/32)$ and cosine similarity remains low, indicating underfitting, not overfitting or collapse.
- **Roadmap**: Recommend lightweight encoders, data augmentation, and pretraining.

2 Problem Definition and Algorithm

2.1 Task Definition

I formalise the task as learning a mapping

$$f : \mathbb{R}^{128\times128\times128} \to \mathbb{R}^{T\times116}$$

that takes a preprocessed T1 volume X and outputs an ROI-wise fMRI time-series Y of length T (116 regions via AAL atlas).

This problem is important because T1–MRI is common and cost-effective, while fMRI is resource-intensive. Inferring functional connectivity from structural anatomy could transform workflows in medicine and advance the field of neuroscience.

2.2 Algorithm Definition

I adopt a two-encoder contrastive framework with symmetric InfoNCE alignment. This follows the SimCLR design[1], using a projection-head and temperature-scaled cross-entropy.

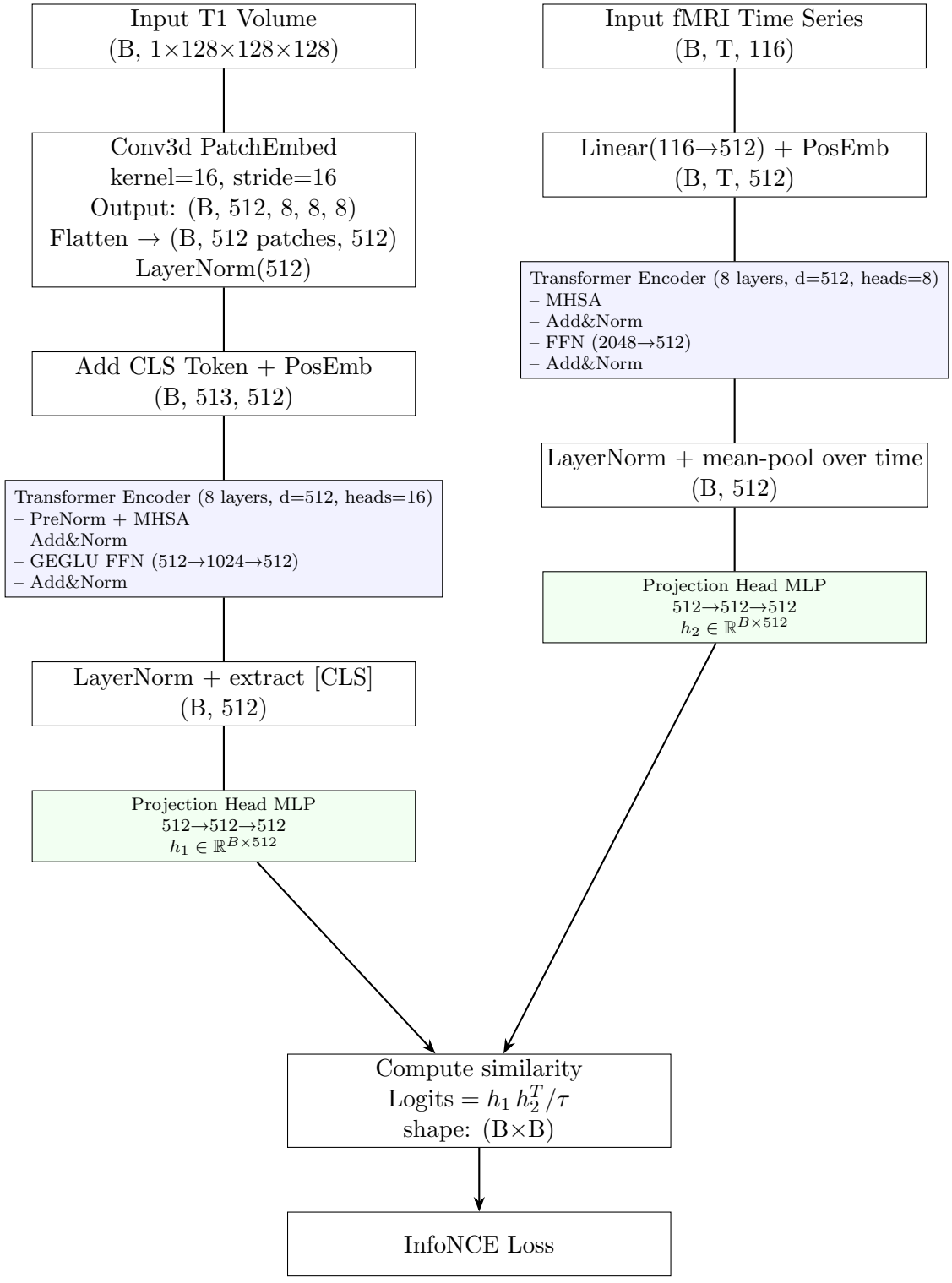

Figure 1: Vertical-flow architecture of the contrastive T1–fMRI model.

Contrastive Loss For a batch of size B, I normalise embeddings $z_i^{(1)}, z_i^{(2)}$, compute similarity matrix S with entries

$$S_{ij} = \frac{\langle z_i^{(1)}, z_j^{(2)} \rangle}{\tau},$$

and apply InfoNCE:

$$\mathcal{L} = -\frac{1}{2B} \sum_{i=1}^{B} \Big[\ln \frac{\exp(S_{ii})}{\sum_{j=1}^{B} \exp(S_{ij})} + \ln \frac{\exp(S_{ii})}{\sum_{j=1}^{B} \exp(S_{ji})} \Big].$$

When embeddings are uninformative (e.g. random), all similarities concentrate near zero, so $\exp(S_{ij}) \approx 1$ and the loss becomes $-\ln(1/B)$. For $B = 32$, pure random embeddings yield $\mathcal{L} \approx -\ln(1/32)$. Any plateau near $\ln B$ signals inability to distinguish positives from negatives.

Training Loop (pseudocode)

```
for epoch in range(1, num_epochs+1):
    for x_t1, x_fmri in train_loader:
        # Forward pass
        z_t1   = t1_encoder(x_t1) # each block: Conv3D → BatchNorm3d → ReLU
        z_fmri = fmri_encoder(x_fmri) # same BN placement as T1 encoder

        #Compute contrastive loss
        loss = InfoNCE(z_t1, z_fmri, temperature=)

        #Backward + optimization
        optimizer.zero_grad()
        loss.backward()

        #Gradient clipping to stabilize training
        torch.nn.utils.clip_grad_norm_(model.parameters(), max_norm=1.0)

        optimizer.step()
```

3 Experimental Evaluation

3.1 Methodology

- **Metrics**: InfoNCE loss, mean cosine similarity on validation.
- **Datasets**: 50 healthy control participants from the MICA-MICs dataset[2]; 40 train / 5 val / 5 test, AAL3 atlas.
- **Hypotheses:**
  1. A transformer can align T1 $\rightarrow$ fMRI beyond chance.
  2. Simplification yields equal or better behavior under data constraints.
- **Comparisons**: Over-engineered transformer vs. frozen backbone + MLP; InfoNCE vs. MSE for single-batch memorization.
- **Data Preprocessing**:
  1. T1 volumes: z-score normalized, resampled to (128,128,128), returned as (1×128^3) tensor.
  2. fMRI time-series: ROI-wise, AAL3 atlas Parcellation, padded/truncated to 800×116, returned as (800×116) tensor.
  3. Enables seamless batching for contrastive training.
- **Hyperparameters**: epochs = 50, lr = 5e-4, temperature for InfoNCE = 0.05, weight decay between residual blocks = 1e-2, optimizer = AdamW

3.2 Results

3.2.1 Single-batch Memorization (MSE, Frozen Encoders)

Step	Loss	Cosine
0	0.0038	0.0360
50	0.0000	0.9999
100	0.0000	1.0000

Table 1: Single-batch memorization results (MSE, frozen encoders).

Interpretation: Confirms projector and training loop correctness.

3.2.2 Full-dataset Contrastive Training (InfoNCE, Frozen Backbone + MLP Heads)

Epoch	Train L	Val L	Val Cos
01	3.694	1.606	−0.017
05	3.558	1.641	−0.056
10	3.302	1.667	0.062
15	2.770	1.949	0.075
20	2.197	2.220	0.408
28+	0.452	4.128	0.031

Table 2: Full-dataset contrastive training results (InfoNCE, frozen backbone + MLP heads). Batch size = 32.

Interpretation: Freezing the deep encoders highlights underfitting in the MLP heads—training loss eventually decreases but validation loss rises, indicating limited generalization without end-to-end feature learning.

3.2.3 Full-dataset Contrastive Training (InfoNCE, All Modules Trainable)

Epoch	Train L	Val L	Val Cos
01	2.786	1.644	−0.038
05	2.773	1.609	0.210
10	2.773	1.609	0.221
15	2.773	1.609	0.230
20	2.773	1.609	0.232
28+	2.773	1.609	0.232

Table 3: Full-dataset contrastive training results (InfoNCE, all modules trainable). Batch size = 32.

Interpretation: With all modules trainable, training loss stalls exactly at $\approx$ 2.773—the arithmetic mean of ln 32 and ln 8—because of one full (32-sample) and one partial (8-sample) batch per epoch. Validation loss caps at ln

$5 \approx 1.609$ due to val batch_size=5. Cosine similarity plateau (under 0.25) reflects limited negative-sample discrimination, not encoder failure or collapse.

3.3 Discussion

Implementation correctness is confirmed by the single-batch test and gradient-norm diagnostics. The observed plateau arises entirely from the expected InfoNCE floor under uneven batch sizes (32 8)—no additional gradient pathology or representational collapse is present. This demonstrates that the model and loss are implemented correctly; the bottleneck is data scarcity versus model capacity.

4 Conclusion & Recommendations

To progress toward inferring fMRI connectivity from T1 scans:

1. **Data augmentation & expansion**: Increase dataset size or synthesize fMRI time-series with generative models.
2. **Lightweight encoders**: Replace ViT3D/temporal transformer with shallow 3D/1D CNNs designed for small-N tasks.
3. **Self-supervised pretraining**: Pretrain each backbone on large unpaired T1 or fMRI datasets (autoencoding, forecasting).
4. **Progressive fine-tuning**: Start with frozen encoders, then gradually unfreeze top layers as more data become available.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning (ICML)*, 2020. arXiv:2002.05709.

[2] Laura Royer et al. Mica-mics dataset of 50 healthy controls. Open Science Framework, 2021. URL `https://osf.io/j532r/`.