

Scaffold-GAT: Grounding Brain Dynamics in Neural Anatomy

Beomjin Kim
superbdi@o.cnu.ac.kr

Chungnam National University
Daejeon, South Korea

Sungsu Lim*
sungsu@cnu.ac.kr

Chungnam National University
Daejeon, South Korea

Abstract

Reconstructing natural language from functional Magnetic Resonance Imaging (fMRI) is a fundamental objective in non-invasive Brain-Computer Interfaces (BCIs). Current fMRI encoding methods—mapping linguistic features to brain activity—typically follow two paradigms. Linear models are robust under extreme data scarcity but treat voxels as isolated units, failing to capture the brain’s complex neural topology. Conversely, graph-based models capture regional interactions but rely on unstable functional connectivity (FC) due to data-scarce conditions. These limitations highlight the necessity of modeling dynamic functional relationships and stable anatomical constraints together. To address this, we propose Scaffold-GAT, a multiplex graph attention network that explicitly anchors dynamic functional signals onto a stable anatomical scaffold. Our framework achieves a BERTScore of 0.8259, outperforming standard linear baselines. Furthermore, our scaling analysis reveals that the proposed model achieves its best performance when its representational capacity and spatial coverage align with biologically meaningful scales of the language network.

CCS Concepts

• **Computing methodologies** → **Neural networks**; Supervised learning; • **Information systems** → *Data mining*.

Keywords

Graph Neural Networks, Graph Attention Networks, Multi Relational Graphs, fMRI Encoding, Structural Regularization

ACM Reference Format:

Beomjin Kim and Sungsu Lim. 2026. Scaffold-GAT: Grounding Brain Dynamics in Neural Anatomy. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '26)*. ACM, New York, NY, USA, 8 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

In modern neurotechnology, decoding natural language from fMRI is an important step toward non-invasive brain-computer interfaces and a deeper understanding of human semantic processing [18, 23]. Recent advances in Large Language Models (LLMs) have greatly

* Advisor.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '26, Jeju, Korea

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06
<https://doi.org/XXXXXXX.XXXXXXX>

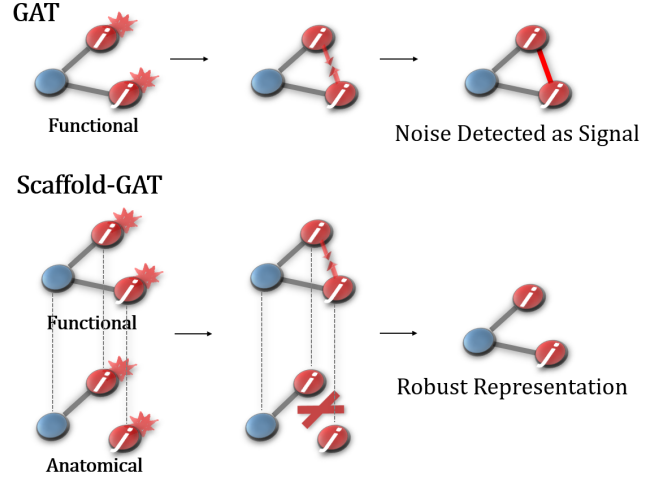


Figure 1: While GAT (top) is vulnerable to spurious noise, Scaffold-GAT (bottom) anchors functional dynamics to anatomical scaffold, ensuring robust fMRI encoding.

improved the decoding stage by enabling the reconstruction of continuous linguistic meaning from brain activity [26]. In contrast, the encoding stage—mapping linguistic representations to fMRI responses—relies largely on simple voxel-wise linear regression [11, 20]. This gap between powerful decoders and simple encoders has become a major limitation in fMRI-based language modeling.

Existing fMRI encoding methods mainly follow two directions. The first uses linear models, which are widely adopted because fMRI data are noisy and limited in size. In benchmark datasets, the number of temporal samples is often small compared with the dimensionality of features, so linear regression provides a practical way to reduce overfitting [19]. However, this simplicity comes with a clear limitation: it treats each voxel independently and ignores the fact that brain activity arises from interactions among distributed regions [3, 19]. As a result, linear models have limited ability to capture the structured dependencies.

The second direction uses graph-based models, especially Graph Attention Networks (GATs), to explicitly model relationships between voxels [1, 25]. By representing the brain as a graph, these methods can aggregate information from related regions and are therefore better suited to the networked nature of brain activity. In particular, FC-based GATs can selectively integrate signals from functionally related voxels, which helps overcome the limitations of isolated voxel-wise modeling. However, FC estimated from fMRI is often unstable because of the low signal-to-noise ratio of neural recordings. Temporary statistical fluctuations may be mistaken for meaningful interactions, making functional GAT sensitive to noisy or spurious connections.

These limitations suggest that robust fMRI encoding requires more than functional dependency alone. Brain activity is shaped not only by functional dependency, which reflects statistical co-activation, but also by anatomical dependency, which provides the structural scaffold that constrains neural communication. We therefore argue that both types of dependency should be modeled together [4, 10]. Anatomical information can provide a stable prior for physically plausible interactions, while functional information captures task-relevant dynamic relationships. Combining them is especially important for language decoding, where meaningful representations emerge from coordinated activity across distributed but anatomically grounded brain networks [9].

To address this issue, we propose Scaffold-GAT, a multiplex graph attention network that models anatomical and functional dependencies as two distinct but related graph layers defined over the same set of voxels. Figure 1 demonstrates how Scaffold-GAT prevents the features of each network from getting blurred. Instead of merging all relations into a single graph, Scaffold-GAT keeps the two dependency types separate to preserve their different roles. It then uses a cross-layer attention mechanism in which the anatomical layer guides the flow of information in the functional layer. In this way, the model suppresses implausible or noisy functional interactions while still retaining the flexibility needed to capture higher-order structure-function relationships.

While Scaffold-GAT remains robust across various settings, its peak performance is reached when its representational capacity and spatial scale align with biological constants of the human brain. Notably, optimal configuration aligns with known neural dimensions rather than scaling with more data or voxels. This alignment provides strong empirical validation for the model’s biological plausibility, proving that our framework effectively mirrors the intrinsic scale of human semantic processing.

Our contributions are summarized as follows:

- We propose a multiplex fMRI encoding framework that explicitly separates anatomical and functional dependencies.
- We introduce a cross-layer structural gating mechanism that uses anatomical structure to regularize functional message passing under noisy and data-scarce conditions.
- We show that our model achieves best performance when its representational capacity and spatial coverage align with biologically meaningful scales of the language network.

2 Related Work

2.1 Evolution of Neural Language Encoding

The methodology for mapping stimulus features to neural responses has transitioned from voxel-wise classification [18] to the continuous reconstruction of naturalistic narratives [11, 23, 26]. Recent advancements have integrated LLMs to bridge the gap between high-order semantic embeddings and blood-oxygen-level-dependent (BOLD) signals. However, while decoding has benefited from generative architectures, the encoding stage remains reliant on voxel-wise linear models [11, 13]. While robust in data-scarce settings, these models ignore the brain’s networked topology, treating voxels as isolated units and thus reaching a representational bottleneck when modeling complex cognitive processes [15, 27].

2.2 Graph Neural Networks in Neuroanatomy

To capture the brain’s functional connectivity, researchers have increasingly adopted Graph Neural Networks (GNNs) [2]. Early efforts utilized isotropic filters [16], but subsequent developments proposed anisotropic mechanisms such as GAT to dynamically weight interactions [29]. Recent frameworks like BrainGNN have further integrated region-of-interest (ROI)-aware kernels to address the unique identity of brain regions, moving away from the translation invariance assumed in conventional GNNs [6, 17]. Despite these gains, many models still rely on FC in isolation. When integrating multimodal data, existing approaches often resort to simple fusion (e.g., concatenation or summation), which can lead to *relational blurring* and the loss of layer-specific features that define the hierarchical interaction between anatomy and function [30].

2.3 Structure-Function Synergy and Multiplex Systems

Anatomical pathways provide the structural foundation for brain function, guiding its activity without strictly determining it [9, 22]. This relationship varies across the brain. Sensory regions follow their anatomical wiring, whereas regions responsible for complex thought operate with greater flexibility [24, 28]. To model these complex interactions, multiplex graphs are highly effective; they represent different data types—such as structure and function—as distinct but connected layers over the same brain regions [8].

Recent research has begun to explore the role of the anatomical connectome as a stable anatomical scaffold that guides functional dynamics over constrained latent manifolds [5, 21]. Inspired by biological gating processes in neural channels, models such as Brain-Symphony attempt to use structural priors as an inductive bias to reduce the search space and improve representational capacity in data-scarce environments [14]. However, a remaining challenge involves designing a principled gating mechanism that can robustly ground transient functional signals within the fixed anatomical substrate, especially in high-noise contexts. Building upon these foundations, our work introduces a structural gating framework to facilitate more biologically grounded information flow in the encoding process.

3 Methodology

In this section, we present the framework of Scaffold-GAT. Our approach focus on stabilizing information flow in high-noise fMRI environments by modeling the brain as a multiplex graph. We first detail the construction of **multiplex representation** and then introduce the **structural gating mechanism**.

3.1 Problem Formulation and Notation

We define the neural encoding task as learning a mapping from stimulus space to brain response space. Let $\mathcal{S} = \{s_1, s_2, \dots, s_T\}$ be a sequence of natural language stimuli and $Y \in \mathbb{R}^{N \times T}$ be the observed fMRI BOLD signals across N voxel nodes $V = \{v_1, \dots, v_N\}$ over T time points. The brain is represented as a multiplex graph $G = (V, E_{func}, E_{anat})$, capturing both dynamic synchronization and structural constraints.

Given this multiplex brain topology—represented by functional and refined anatomical adjacency matrices (A_{func}, A_{ref})—and the

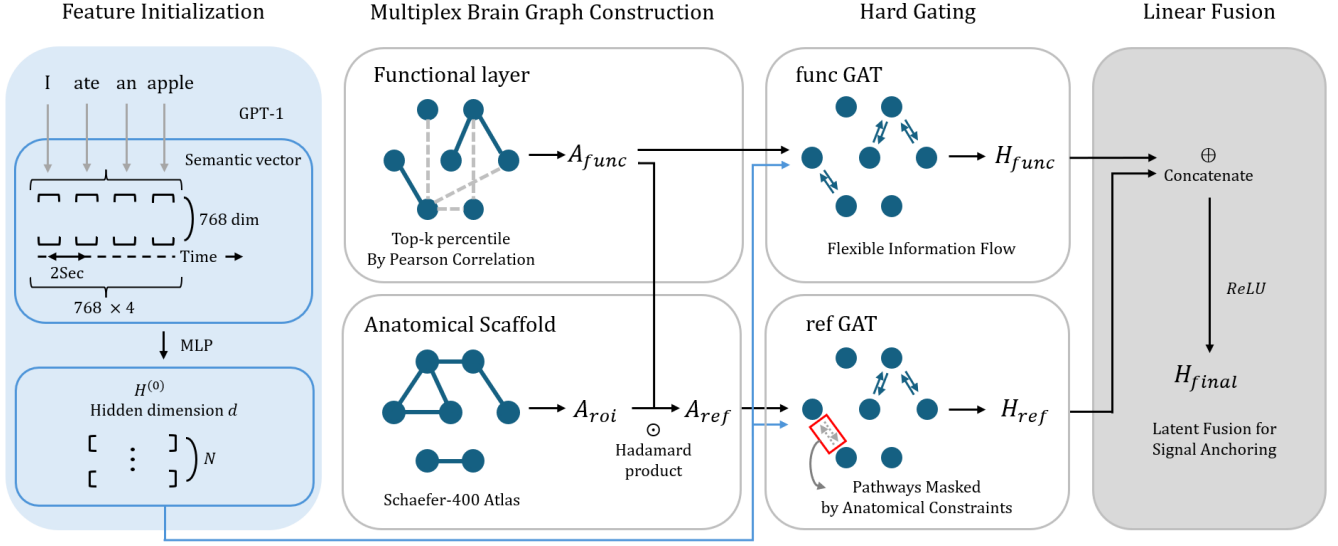


Figure 2: The framework integrates functional and anatomical brain connectivities through structural hard-gating and latent-level signal anchoring to discover robust neural representations from high-noise fMRI data.

corresponding linguistic stimulus sequence $\mathcal{S}$, the goal is to learn a mapping function $\mathcal{F}_\Theta$ to predict the fMRI BOLD signals $\hat{Y}$. Here, Θ represents all learnable parameters. Thus, the neural encoding task is formulated as follows:

$$[\{s_t\}_{t=1}^T; (A_{func}, A_{ref})] \xrightarrow{\mathcal{F}_\Theta} \{\hat{y}_t\}_{t=1}^T \quad (1)$$

3.1.1 Feature Initialization. Each node $v \in V$ is initialized with semantic features extracted from neural language representative model. To account for the $\sim 10s$ hemodynamic delay and match the fMRI repetition time ($TR = 2s$), we temporally concatenate embeddings across a four-volume window. These concatenated features, comprising 768×4 dimensions, are projected via an MLP into a d -dimensional latent space, yielding initial node representations $H^{(0)} \in \mathbb{R}^{N \times d}$, where $h_i^{(0)} \in \mathbb{R}^d$ for each node i .

3.2 Multiplex Brain Graph Construction

We construct a multiplex representation consisting of two distinct topological layers to capture multifaceted brain organization.

3.2.1 Functional Connectivity Layer. The functional layer captures the dynamic synchronization between voxels. For each node i , we extract the BOLD time-series $Y_i \in \mathbb{R}^T$. The edge weight between node i and j is defined by the Pearson correlation coefficient r_{ij} :

$$r_{ij} = \frac{\sum_{t=1}^T (Y_{i,t} - \bar{Y}_i)(Y_{j,t} - \bar{Y}_j)}{\sqrt{\sum_{t=1}^T (Y_{i,t} - \bar{Y}_i)^2 \sum_{t=1}^T (Y_{j,t} - \bar{Y}_j)^2}}. \quad (2)$$

To mitigate noise, we apply a global thresholding strategy, retaining the top- k percentile (e.g., $k = 30\%$) of correlations to construct a sparse functional adjacency matrix $A_{func} \in \mathbb{R}^{N \times N}$.

3.2.2 Anatomical Scaffold. While functional edges are dynamic, the anatomical connectome provides a stable physical substrate. We introduce a Hadamard-based refinement operator to ground functional activity within the anatomical substrate. Let $A_{roi} \in \{0, 1\}^{N \times N}$

be the anatomical adjacency matrix defined by shared boundaries in the Schaefer-400 atlas. The refined scaffold A_{ref} is computed as:

$$A_{ref} = A_{roi} \odot A_{func}, \quad (3)$$

where $\odot$ denotes the Hadamard product. This operation serves as a biological filter, ensuring information in this layer is propagated only through pathways that are both physically adjacent and functionally synchronized. By constraining the search space to intra-ROI interactions, we provide a strong inductive bias that filters out spurious long-range correlations in fMRI data.

3.3 Structural Gating Mechanism

The structural gating mechanism modulates the flow of latent linguistic representations across the multiplex graph, ensuring that representations are biologically grounded.

3.3.1 Hard Gating. In our GAT layers, we implement "Hard Gating" to strictly enforce anatomical constraints. To capture node affinities within each multiplex layer, we utilize a scaled dot-product attention mechanism. For any node pair (i, j) , the raw attention score e_{ij} is calculated as:

$$e_{ij} = \frac{(Wh_i)^T(Wh_j)}{\sqrt{d_k}}, \quad (4)$$

where W is a learnable weight matrix that transforms the input features, and d_k represents the dimensionality of each attention head (where $d_k = d/n$ for n attention heads). The scaling factor $1/\sqrt{d_k}$ is applied to prevent the dot products from growing excessively large in high-dimensional spaces, which could otherwise lead to vanishing gradients during the softmax operation.

To block illegitimate information flow across the brain network, we apply a structural mask M_{ij} derived from the respective adjacency matrix (A_{func} or A_{ref}) corresponding to each multiplex layer before the final attention weights are computed:

$$\alpha_{ij} = \frac{\exp(e_{ij} + M_{ij})}{\sum_{k \in N_i} \exp(e_{ik} + M_{ik})}. \quad (5)$$

In this formulation, the mask is defined as $M_{ij} = 0$ if a valid edge exists in the corresponding adjacency matrix A ($A_{ij} > 0$), and $M_{ij} = -\infty$ otherwise. By assigning $-\infty$ to unvalidated edges, the model ensures that the attention weight α_{ij} becomes exactly zero for any interaction not supported by the anatomical or functional constraints defined in the scaffold. This mechanism effectively prevents information leakage from biologically implausible neighbors, stabilizing the model’s inference in high-noise fMRI environments.

3.3.2 Linear Fusion and Signal Anchoring. To preserve the distinct properties of each modality, Scaffold-GAT processes the functional (A_{func}) and refined anatomical (A_{ref}) structures through parallel message-passing paths. At the final stage of each multiplex block, representations from the functional and anatomical paths (H_{func}, H_{ref}) are integrated through a linear fusion layer:

$$H_{final} = \text{ReLU}(W_{fuse}[H_{func} \oplus H_{ref}] + b), \quad (6)$$

where $\oplus$ denotes concatenation and W_{fuse} is a learnable weight matrix. Crucially, Scaffold-GAT performs fusion at the latent level in addition to the adjacency-level refinement, specifically to avoid relational blurring that occurs when collapsing heterogeneous connectivity into a single matrix. This separation preserves modality-specific identities within their unique manifolds during independent message-passing. W_{fuse} then adaptively anchors transient functional dynamics onto the stable anatomical scaffold, capturing high-dimensional synergy while maintaining the relational integrity of the brain network.

3.4 Training and Optimization

To predict neural activity, integrated representations are projected into the temporal BOLD space. For each voxel i , the predicted time-series $\hat{Y}_i \in \mathbb{R}^T$ is generated via a linear output layer:

$$\hat{Y}_i = W_{out}h_{i,final} + b_{out}, \quad (7)$$

where $W_{out} \in \mathbb{R}^{T \times d}$ is a learnable weight matrix and $b_{out} \in \mathbb{R}^T$ is a bias term. This operation effectively decodes the d -dimensional semantic features into a T -dimensional temporal sequence, matching the acquisition length of the fMRI signal.

Scaffold-GAT is then optimized to minimize the discrepancy between these predicted signals and the observed BOLD responses. We define a multi-objective loss function $\mathcal{L}$ that combines Mean Squared Error (MSE) for amplitude reconstruction and a Pearson-based loss ($1 - r$) for temporal pattern alignment:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left(\text{MSE}(Y_i, \hat{Y}_i) + \lambda(1 - r(Y_i, \hat{Y}_i)) \right), \quad (8)$$

where λ balances the reconstruction of BOLD magnitude and the preservation of neural dynamics. This dual-loss approach ensures that the learned representations are not only numerically accurate but also biologically plausible in temporal behavior, capturing the fine-grained semantic fluctuations recorded during the experiment.

4 Experiments

4.1 Experimental Setup

4.1.1 Dataset and Preprocessing. We evaluated the performance of Scaffold-GAT using the natural language fMRI dataset provided by Tang et al. (2023) [26]. We specifically utilized 16 hours of BOLD responses from subject S3. This subject was selected because they exhibited the highest baseline decoding performance. Following the baseline protocol, raw signals were preprocessed with motion correction and z-scored within each scan to ensure unit variance.

4.1.2 Model Configuration. The top 2,000 cortical voxels were selected as the default node set. For the Scaffold-GAT architecture, we set the hidden dimension (d) at 128. For a comprehensive list of all hyperparameters, including weight decay and learning rate schedules, please refer to the Reproducibility Statement.

4.1.3 Graph Construction and Density. In constructing the functional layer A_{func} , we adaptively adjusted the Pearson correlation thresholds for each node count to maintain a consistent graph density within a range of 1%–2%. This baseline density preserves robust functional interactions; a detailed analysis of its theoretical justification regarding dimensional capacity is provided in Section 4.4.2.

Table 1: A_{func} thresholds and densities by node count.

Nodes (N)	Threshold	Density (%)
1,000	0.60	1.35%
2,000 (Default)	0.60	1.13%
5,000	0.45	1.66%
10,000	0.45	1.24%

4.2 Evaluation Protocol

4.2.1 Inference Procedure. Once the encoding models are trained, we evaluate their performance through a language reconstruction procedure. Following the protocol in [26], we utilize a beam search decoder ($k = 200$), employing the GPT-1 model as a drafter and the log-likelihood as a verifier.

4.2.2 Evaluation Metrics. We assess our model from two different angles. How well it captures the meaning and how accurately it predicts exact words.

In the context of brain decoding, literal metrics like WER and METEOR often reflect limited performance because they require near-perfect word matching, which is difficult given the noise in neural signals. Therefore, we treat BERTScore as our primary benchmark, as it captures the underlying semantic intent even if the specific vocabulary differs from the stimulus.

Table 2: Evaluation metrics overview.

Metric	Ideal	Primary Focus
BERTScore	↑	Semantic essence and context recovery
WER	↓	Strict word-for-word accuracy
METEOR	↑	Flexible alignment (synonyms/stems)

4.3 Ablation Study

To clarify the contribution of each module, we compare five architectures with varying degrees of structural constraints:

- **Linear (Baseline)**: The voxel-wise linear model from [26].
- **funcGNN & funcGAT**: Functional graph models using BOLD correlations, without anatomical priors.
- **refGAT**: A structurally-informed GAT integrating anatomical priors via adjacency-level fusion.
- **Scaffold-GAT**: Multiplex GAT anchoring functional signals to anatomy via latent fusion and structural gating.

The results, summarized in Table 3, provide key insights into the interaction between brain topology and data scarcity on subject S3.

Table 3: Ablation results for Scaffold-GAT components.

Model	BERTScore (↑)	WER (↓)	METEOR (↑)
Linear (Baseline) [26]	0.8116	0.9243	0.1703
funcGNN	0.7982	0.9549	0.1423
funcGAT	0.7997	0.9593	0.1390
refGAT	0.8085	0.9413	0.1455
Scaffold-GAT	0.8259	0.9183	0.1894

Interestingly, funcGNN and funcGAT exhibited lower BERTScore than the Linear baseline. This performance degradation is primarily due to the extreme data scarcity in fMRI datasets. Without structural constraints, the high degree of freedom in unconstrained GNNs leads to overfitting on localized noise, whereas the linear model remains robust by treating voxels independently.

Furthermore, while refGAT introduced anatomical priors and outperformed purely functional GNNs, its reliance on adjacency-level fusion ($A_{roi} \odot A_{func}$) still resulted in a performance bottleneck below the Linear baseline. This confirms our hypothesis that collapsing heterogeneous connectivities into a single matrix prematurely causes relational blurring, where the distinct topological identities of anatomy and function are lost.

In contrast, **Scaffold-GAT** achieves superior results by additionally performing fusion at the *latent level*. This separation preserves modality-specific identities during message-passing, providing a strong *structural inductive bias* that prevents overfitting. W_{fuse} then adaptively anchors transient functional signals onto the stable anatomical scaffold, capturing high-dimensional synergy without losing relational integrity.

4.3.1 Complexity-Data Imbalance in Linear vs. Graph. The performance gap between linear(Baseline) and simple graph models(funcGNN, funcGAT) stems from the ratio of learnable parameters to available data samples. The dataset provides $T = 28,800$ temporal samples, derived from 16 hours of fMRI BOLD responses at a 2-second repetition time.

The linear baseline maps stimulus features to voxels independently using a concatenated semantic vector of 3,072 dimensions (768×4). Because the number of parameters per voxel remains significantly lower than the sample size T , the linear mapping maintains robustness against noise even in data-scarce conditions.

In contrast, simple graph models jointly optimize non-linear interactions across $N = 2,000$ voxels. With a hidden dimension

of $d = 128$ and a graph density of 1.13%, these models must capture representations across approximately 45,000 edges. This high degree of freedom, combined with multi-head attention creates a complex parameter space that lacks sufficient constraints from only 28,800 samples.

Consequently, simple graph models overfit to spurious statistical fluctuations in the fMRI signal. Scaffold-GAT addresses this imbalance via structural gating, which anchors functional message-passing to the anatomical scaffold to reduce the search space to a biologically plausible range.

4.4 Scaling Analysis: Information Density and Dimensional Capacity

While Scaffold-GAT leverages network synergy to outperform linear models, its success depends on the balance between node count (N) and representational capacity (d). In this section, we analyze these scaling trends to understand the underlying principles of neural information integration. Specifically, we explore how the observed performance peak reflects a structural alignment between the model’s dimensional capacity and the intrinsic properties of neural semantic representations.

4.4.1 Observation: Performance Peak and the Scaling Bottleneck. To investigate the scalability of Scaffold-GAT, we first evaluated the model by varying the voxel count ($N \in \{1000, 2000, 5000, 10000\}$) while fixing the hidden dimension at $d = 128$. For consistent comparison, we maintained a constant graph density of 1–2% across all settings. This constraint ensures that the average degree (k) of each node scales linearly with N (e.g., $k \approx 20$ for $N = 2,000$, and $k \approx 100$ for $N = 10,000$), allowing us to isolate the effect of neighbor density on latent representation.

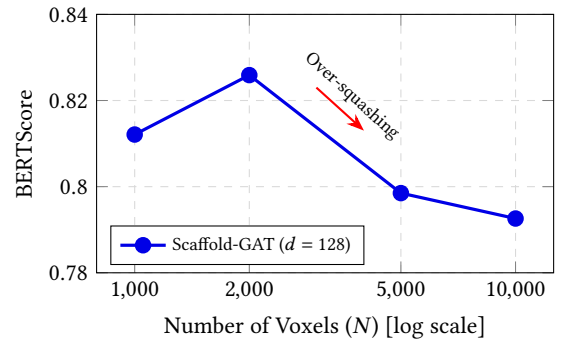


Figure 3: Performance scaling across N for $d = 128$.

As shown in Figure 3, the BERTScore followed an inverted U-shape curve, reaching its peak at $N = 2,000$. Interestingly, performance declined as N scaled toward 10,000, even though the model had access to a broader cortical area. This trend suggests that simply increasing the number of neural nodes does not guarantee better decoding; rather, it introduces a representational bottleneck within the fixed latent space.

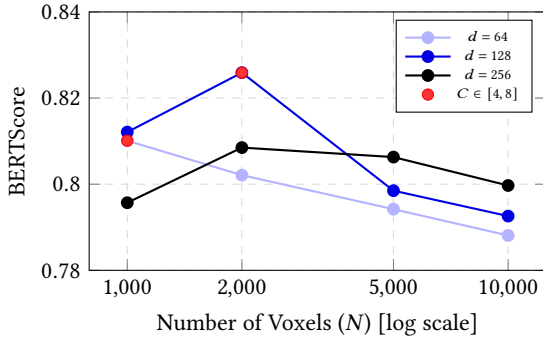
Table 4: Performance analysis across scaling configurations. (C) closest to the optimal biological range [4, 8] are highlighted in bold, and the second closest are underlined.

Voxels (N)	$d = 64$				$d = 128$				$d = 256$			
	$C = d/k$	BERTScore($\uparrow$)	WER($\downarrow$)	METEOR($\uparrow$)	$C = d/k$	BERTScore($\uparrow$)	WER($\downarrow$)	METEOR($\uparrow$)	$C = d/k$	BERTScore($\uparrow$)	WER($\downarrow$)	METEOR($\uparrow$)
1,000	4.74	0.8101	0.9275	0.1782	<u>9.48</u>	<u>0.8121</u>	<u>0.9241</u>	<u>0.1815</u>	18.96	0.7957	0.9482	0.1652
2,000	<u>2.83</u>	<u>0.8021</u>	<u>0.9388</u>	<u>0.1703</u>	5.66	0.8259	0.9183	0.1894	11.33	0.8085	<u>0.9314</u>	<u>0.1768</u>
5,000	0.77	0.7942	0.9428	0.1628	1.54	0.7985	0.9401	0.1685	3.08	<u>0.8063</u>	0.9310	0.1780
10,000	0.52	0.7881	0.9521	0.1564	1.03	0.7926	0.9489	0.1601	<u>2.06</u>	0.7997	0.9412	0.1672

4.4.2 Hypothesis: Dimensional Capacity per Edge. The performance drop at higher N led us to hypothesize a hidden constraint: Dimensional Capacity per Edge (C). We define C as the ratio of hidden dimensions to the number of neighbors (d/k), representing the bandwidth allocated to each connection during message passing.

To conceptualize this, assuming a baseline graph density of 1%, when $d = 128$ and $N = 2,000$ ($k = 20$), each neighbor is assigned $C = 6.4$ dimensions on average. This capacity appears to align with the intrinsic dimensionality of neural semantic representations. For instance, Huth et al. (2012) [12] demonstrated that the universal semantic space in the human brain across diverse stimuli can be effectively summarized by the **first 4 to 8 principal components**. The alignment between our optimized C on a distinct dataset and this known biological constant suggests that our model optimally integrates semantic signals while filtering out noise. In contrast, at $N = 10,000$ ($C = 1.28$ on average), the capacity becomes too narrow to distinguish individual neighbor messages, leading to over-squashing.

4.4.3 Validation: Joint Modulation of N and d . To validate this hypothesis, we conducted a second experiment where both N and d were modulated to maintain the capacity within a controlled biological range ($C \in [4, 8]$, indicated by red markers in Figure 4). Specifically, we compared optimal configurations such as ($N = 1,000, d = 64$) and ($N = 2,000, d = 128$), which provide similar per-edge bandwidths hovering securely within this target regime ($C \approx 4.74$ and $C \approx 5.66$, respectively).

**Figure 4: Performance scaling across N and d .**

As shown in Figure 4, while maintaining C mitigates severe over-squashing at larger scales, the $N = 2,000$ configuration ($d = 128$) consistently outperformed the others. Interestingly, even under $d = 256$, the $N = 2,000$ model slightly outperforms the $N = 5,000$

model despite possessing an excessive, sub-optimal capacity ($C = 11.33$ vs $C = 3.08$). This occurs because scaling to 5,000 voxels far exceeds the anatomical scale of the core language network, typically estimated at 1,100 to 1,800 voxels [26]. The massive influx of task-irrelevant noise degrades the semantic signal quality, overriding the theoretical benefits of a balanced dimensional capacity.

Rather, $N = 2,000$ represents a biologically grounded sweet spot. By focusing on this core, Scaffold-GAT effectively covers high-fidelity semantic regions [7] while avoiding spatial noise. While the architectural innovations provide the foundation, the final success of Scaffold-GAT stems from a dual alignment: matching the model's architectural capacity with the brain's intrinsic dimensionality, and aligning its spatial scale with functional anatomy.

5 Discussion

establishes a robust foundation for biologically grounded neural encoding, several avenues in graph learning remain. First, extending this framework to incorporate Dynamic Graph Neural Networks could capture non-stationary temporal shifts in brain connectivity, moving beyond static correlations. Second, building upon our findings of optimal spatial alignment ($N = 2,000$), future work will explore task-adaptive subgraph mining. This would allow the model to automatically discover the biological sweet spot for diverse cognitive tasks through differentiable subgraph extraction. Finally, prioritizing the stable anatomical scaffold opens a promising pathway for cross-subject transfer learning. By treating the structural network as a domain-invariant anchor, we aim to facilitate few-shot encoding across individuals, directly addressing neuroimaging data collection costs.

6 Conclusion

This research demonstrates that the trade-off between model capacity and data scarcity in fMRI encoding can be resolved by grounding neural dynamics in anatomical reality. While high-capacity GNNs typically suffer from overfitting in sparse data environments, our structural gating mechanism reveals that anatomical priors serve as a vital inductive bias, transforming functional noise into structured linguistic meaning. Beyond mere performance gains, the critical insight of this study lies in the convergence between a model's dimensional capacity and the brain's intrinsic spatial scale. The observation that peak encoding accuracy occurs at the alignment of C and N suggests that the language network is a structured manifold governed by specific topological constraints. This shifts the decoding paradigm from simple signal mapping to structural alignment, enabling robust semantic extraction from sparse, non-invasive recordings.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2025-25435830).

References

- [1] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Andrea De Maria, Greg Wayne, Peter Rezende, David Silver, and Oriol Vinyals. 2018. Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261*
- [2] Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. 2022. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 5 (2022), 5833–5848.
- [3] Jeffrey R Binder, Rutvik H Desai, William W Graves, and Lisa L Conant. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex* 19, 12 (2009), 2767–2796.
- [4] Ed Bullmore and Olaf Sporns. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience* 10, 3 (2009), 186–198.
- [5] Jeongwhan Choi, Seungjun Park, Sumin Park, Sung-Bae Cho, and Noseong Park. 2026. Are Graph Transformers Necessary? Efficient Long-Range Message Passing with Fractal Nodes in MPNNs. *Proceedings of the AAAI Conference on Artificial Intelligence* 40, 25 (Mar. 2026), 20553–20561. doi:10.1609/aaai.v40i25.39191
- [6] Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. 2022. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International conference on medical image computing and computer-assisted intervention*. Springer, 375–385.
- [7] Evelina Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, Susan Whitfield-Gabrieli, and Nancy Kanwisher. 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of neurophysiology* 104, 2 (2010), 1177–1194.
- [8] Sharmistha Guha, Jose Rodriguez-Acosta, and Ivo D Dinov. 2024. A bayesian multiplex graph classifier of functional brain connectivity across diverse tasks of cognitive control. *Neuroinformatics* 22, 4 (2024), 457–472.
- [9] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106, 6 (2009), 2035–2040. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.0811168106 doi:10.1073/pnas.0811168106
- [10] Jinghan Huang, Moo K. Chung, and Anqi Qiu. 2023. Heterogeneous Graph Convolutional Neural Network via Hodge-Laplacian for Brain Functional Data. In *Information Processing in Medical Imaging*, Alejandro Frangi, Marleen de Bruijne, Demian Wassermann, and Nassir Navab (Eds.). Springer Nature Switzerland, Cham, 278–290.
- [11] Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 7600 (2016), 453–458.
- [12] Alexander G Huth, Shinji Nishimoto, An T Vu, and Jack L Gallant. 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 6 (2012), 1210–1224.
- [13] Shailee Jain, Vy Vo, Shivangi Mahto, Amanda LeBel, Javier S Turek, and Alexander Huth. 2020. Interpretable multi-timescale models for predicting fMRI responses to continuous natural speech. *Advances in Neural Information Processing Systems* 33 (2020), 13738–13749.
- [14] Moein Khajehnejad, Forough Habibollahi, Devon Stoliker, and Adeel Razi. 2026. BrainSymphony: A parameter-efficient multimodal foundation model for brain dynamics with limited data. *arXiv:2506.18314 [q-bio.QM]* <https://arxiv.org/abs/2506.18314>
- [15] Tim C Kietzmann, Patrick McClure, and Nikolaus Kriegeskorte. 2017. Deep neural networks in computational neuroscience. *BioRxiv* (2017), 133504.
- [16] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [17] Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. 2021. Brainn: Interpretable brain graph neural network for fmri analysis. *Medical image analysis* 74 (2021), 102233.
- [18] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. 2008. Predicting human brain activity associated with the meanings of nouns. *science* 320, 5880 (2008), 1191–1195.
- [19] Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. 2011. Encoding and decoding in predictive cognitive neuroscience. *Neuron* 70, 3 (2011), 402–415.
- [20] Kenneth A Norman, Sean M Polyn, Greg J Detre, and James V Haxby. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences* 10, 9 (2006), 424–430.
- [21] James C Pang, Kevin M Aquino, Marianne Oldehinkel, Peter A Robinson, Ben D Fulcher, Michael Breakspear, and Alex Fornito. 2023. Geometric constraints on human brain function. *Nature* 618, 7965 (2023), 566–574.
- [22] Hae-Jeong Park and Karl Friston. 2013. Structural and functional brain networks: from connections to cognition. *Science* 342, 6158 (2013), 1238411.
- [23] Francisco Pereira, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. 2018. Toward a universal decoder of linguistic meaning from brain activation. *Nature communications* 9, 1 (2018), 963.
- [24] Maria Giulia Preti and Dimitri Van De Ville. 2019. Decoupling of brain function from structure reveals regional behavioral specialization in humans. *Nature communications* 10, 1 (2019), 4747.
- [25] Olaf Sporns. 2016. *Networks of the Brain*. MIT press, Cambridge, MA.
- [26] Jerry Tang, Amanda LeBel, Shailee Jain, and Alexander G Huth. 2023. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience* 26, 5 (2023), 858–866.
- [27] Gaël Varoquaux and R Cameron Craddock. 2013. Learning and comparing functional connectomes across subjects. *NeuroImage* 80 (2013), 405–415.
- [28] Bertha Vázquez-Rodríguez, Laura E Suárez, Ross D Markello, Golia Shafiei, Casey Paquola, Patric Hagmann, Martijn P Van Den Heuvel, Boris C Bernhardt, R Nathan Spreng, and Bratislav Misic. 2019. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences* 116, 42 (2019), 21219–21227.
- [29] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, Vancouver, Canada, 12 pages.
- [30] Yanteng Zhang, Xiaohai He, Yi Hao Chan, Qizhi Teng, and Jagath C Rajapakse. 2023. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Computers in biology and medicine* 164 (2023), 107328.

Reproducibility Statement

To ensure the reproducibility of our results, we provide comprehensive technical definitions for the data, hardware, and hyperparameter configurations used in the development of Scaffold-GAT.

Data and Code Accessibility

The dataset is based on the natural language fMRI recordings provided by Tang et al. (2023).

- **Training Data:** Stimulus and response data for subject S3 can be downloaded from OpenNeuro (ds003020, v3.1.1) at <https://openneuro.org/datasets/ds003020/versions/3.1.1>.
- **Test Data:** Data for perceived and imagined experiments can be downloaded from OpenNeuro (ds004510, v1.1.0) at <https://openneuro.org/datasets/ds004510/versions/1.1.0>.

Complete source code for our model Scaffold-GAT is available at <https://github.com/bermuchin/semantic-decoding>.

Hardware and Computational Complexity

All experiments were executed on a workstation equipped with an NVIDIA H100 GPU. The model was trained for 100 epochs for all voxel configurations. The training duration varied according to the number of voxels (N):

- $N = 1,000$: ~1–2 hours
- $N = 2,000$ (**Default**): ~3–4 hours
- $N = 5,000$: ~6–8 hours
- $N = 10,000$: ~30 hours

Implementation Details and Hyperparameters

The following parameters derived from our `config.py` specify the exact configuration used to achieve the reported performance:

Encoding Model Training.

- **Architecture:** Hidden dimension $d = 128$, attention heads $n = 4$, and dropout rate 0.3.
- **Optimization:** AdamW optimizer with a learning rate of 5×10^{-5} and weight decay of 1×10^{-4} .
- **Loss Function:** Hybrid loss combining MSE and Pearson correlation with a weighting coefficient of $\lambda = 0.5$.

Decoding and Beam Search.

- **Language Model:** Features were extracted from the 9th layer (GPT_LAYER) of GPT-1 using a 5-word context window (GPT_WORDS).
- **Beam Search:** Beam width was fixed at 200. Nucleus sampling utilized a cumulative probability mass threshold of 0.9 (LM_MASS) and a relative probability threshold of 0.1 (LM_RATIO).
- **Stimulus Modeling:** `STIM_DELAYS = [2, 4, 6, 8]` represents the discrete temporal offsets (in seconds) used to align stimulus features with the hemodynamic delay.
- **Noise Model:** `NM_ALPHA = 0.9` defines the shrinkage intensity parameter used for regularizing the noise covariance matrix in precision estimation.