



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

MNM: Multi-level Neuroimaging Meta-analysis with Hyperbolic Brain-Text Representations

Seunghun Baek, Jaejin Lee, Jaeyoon Sim, Minjae Jeong, and Won Hwa Kim

Pohang University of Science and Technology, Pohang, South Korea
{habaek4, jlee4923, simjy98, minjaetidtid, wonhwa}@postech.ac.kr

Abstract. Various neuroimaging studies suffer from small sample size problem which often limit their reliability. Meta-analysis addresses this challenge by aggregating findings from different studies to identify consistent patterns of brain activity. However, traditional approaches based on keyword retrieval or linear mappings often overlook the rich hierarchical structure in the brain. In this work, we propose a novel framework that leverages hyperbolic geometry to bridge the gap between neuroscience literature and brain activation maps. By embedding text from research articles and corresponding brain images into a shared hyperbolic space via the Lorentz model, our method captures both semantic similarity and hierarchical organization inherent in neuroimaging data. In the hyperbolic space, our method performs multi-level neuroimaging meta-analysis (MNM) by 1) aligning brain and text embeddings for semantic correspondence, 2) guiding hierarchy between text and brain activations, and 3) preserving the hierarchical relationships within brain activation patterns. Experimental results demonstrate that our model outperforms baselines, offering a robust and interpretable paradigm of multi-level neuroimaging meta-analysis via hyperbolic brain-text representation.

Keywords: Neuroimaging Meta-analysis · Hyperbolic Representation.

1 Introduction

Recent research in neuroscience is rapidly expanding our understanding of brain function and structure, uncovering diverse aspects of neural organization across various regions [18,?]. However, neuroimaging studies are often underestimated due to small sample sizes of their experiments [3,19]. Brain meta-analysis tackles these challenges by integrating neuroscientific descriptions (i.e., body text) with corresponding brain activation maps from multiple studies to identify consistent patterns in the brain. Traditional approaches [10,20,7] have relied on predefined keywords and regression models to discover these patterns by predicting brain activation maps from textual descriptions. More recent approaches utilize large language models (LLMs) to handle longer sequences [14], and employ contrastive learning to map textual descriptions and activation coordinates into a joint representation space [12] for text-to-brain activation prediction.

Notice that the human brain exhibits a naturally nested organization, where broad regions are gradually subdivided into more specialized areas as in Fig. 1(a).

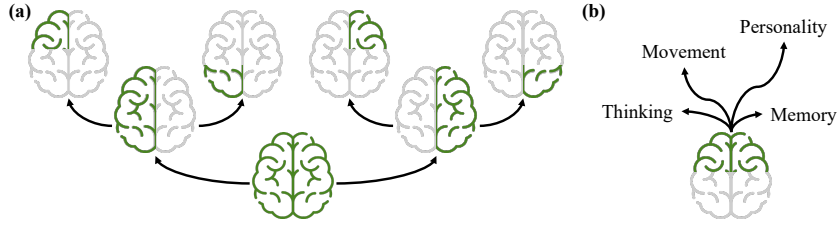


Fig. 1. Motivation of MNM. (a) The brain exhibits a spatial hierarchy, where broad regions can be further subdivided into more specific regions. (b) Each brain region can correspond to various articles covering diverse topics, such as functionalities.

Capturing these *multi-level relationships* is essential for robustly linking textual descriptions from a coarse activation (i.e., involved hemisphere) to fine details (i.e., the most relevant region of interests (ROIs)). Also, as depicted in Fig. 1(b), each brain region is defined by its unique functionality and can be linked to numerous articles that explore diverse topics related to these functions. Such insights on hierarchy are not included in previous approaches, where they map brain activation and text pairs to the identical point in the Euclidean space [12].

To better jointly represent brain activation and text pairs, we rely on hyperbolic spaces with exponentially expanding geometry, which are suited for representing tree-like structures. Therefore, we propose to embed the corresponding pairs into a shared hyperbolic space, together with an angle-based contrastive learning to capture mutual context between brain activation and text. Our framework, **Multi-level Neuroimaging Meta-analysis with Hyperbolic Brain-Text Representations (MNM)**, preserves these inherent hierarchical relationships and overcomes the limitations of the Euclidean method, which lead to the following **major contributions**: **1)** We introduce a hyperbolic embedding framework for neuroimaging meta-analysis that naturally captures semantic associations between brain and text and brain’s hierarchical structure. **2)** We propose a brain structural hierarchy guidance such that brain structural hierarchy is preserved in the hyperbolic space. **3)** MNM is validated through extensive experiments on neuroimaging meta-analysis, such as cross-modal retrieval and brain activation prediction, demonstrating enhanced semantic alignment and interpretability in the same setting in [12]. Cross-modal retrieval and brain activation map prediction demonstrate that our approach not only improves semantic alignment between brain and text modalities but also preserves the multi-level organization of brain regions. Hierarchical structure analyses and ablation studies further confirm that incorporating both brain structural and brain-text hierarchies leads to more interpretable analyses and enhanced retrieval performance.

Related Works on Neuroimaging Meta-Analysis. Previous approaches for brain meta-analysis have transitioned from manual to automated methods. BrainMap [10] manually curated neuroimaging articles, a strategy that struggles to scale with the growing size of literature. NeuroSynth [20] and NeuroQuery [7] automate the extraction of brain activation coordinates using fixed keywords and

statistical techniques, yet they are limited by bag-of-words representations that fail to capture complex semantics. Text2Brain [14] addresses this by using SciBERT [2] embeddings and convolutional neural networks to generate 3D activation maps, though it faces challenges with longer texts. Recently, NeuroConText [12] employs a contrastive framework that leverages advanced language models such as Mistral-7B [9] alongside the Dictionary of Functional Mode (DiFuMo) atlas [5] to enhance the alignment between paired modalities. While we adopt identically processed dataset from [12], MNM outperforms it by leveraging hyperbolic geometry to capture the brain’s intrinsic hierarchical structure, thereby achieving superior semantic alignment and more flexible multi-level meta-analysis.

2 Preliminary: Hyperbolic Spaces and Lorentz Model

Unlike Euclidean space, hyperbolic space expands exponentially under constant negative curvature, making it highly effective for representing hierarchical structures. This enables capturing multi-level traits in both brain structure and text. **Lorentz model.** Lorentz model \mathbb{L}^n expresses the n -dimensional hyperbolic space with curvature c . It is defined in a $(n + 1)$ -dimensional Euclidean space \mathbb{R}^{n+1} as $\mathbb{L}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \langle \mathbf{x}, \mathbf{x} \rangle_{\mathbb{L}} = -1/c, x_0 > 0\}$, where Lorentzian inner product is defined as $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{L}} = -x_0 y_0 + \sum_{i=1}^d x_i y_i$. Following [13], $\mathbf{x} \in \mathbb{L}^n$ is expressed as $[x_{\text{time}} \in \mathbb{R}, \mathbf{x}_{\text{space}} \in \mathbb{R}^n]$. Then, x_{time} is determined as:

$$x_{\text{time}} = \sqrt{1/c + \sum_{i=1}^d x_i^2}. \quad (1)$$

Geodesics. The shortest distance between two points \mathbf{x}, \mathbf{y} on the hyperboloid is called Lorentzian distance $d_{\mathbb{L}}(\mathbf{x}, \mathbf{y})$ and given by:

$$d_{\mathbb{L}}(\mathbf{x}, \mathbf{y}) = \sqrt{1/c} \cosh^{-1}(-c\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{L}}). \quad (2)$$

Tangent space. Tangent space at a point $\mathbf{u} \in \mathbb{L}^n$ is an Euclidean space spanned by a set of vectors orthogonal to \mathbf{u} . For $\mathbf{z} \in \mathbb{R}^{n+1}$ in the tangent space, *exponential map* $\mathbf{Exp}_{\mathbf{u}}(\mathbf{z})$ transforms \mathbf{z} into $\mathbf{x} \in \mathbb{L}^n$ in the hyperbolic space as

$$\mathbf{x} = \mathbf{Exp}_{\mathbf{u}}(\mathbf{z}) = \cosh(\sqrt{c}\sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{L}}})\mathbf{u} + \frac{\sinh(\sqrt{c}\sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{L}}})}{\sqrt{c}\sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{L}}}}\mathbf{z}. \quad (3)$$

With the hyperboloid origin $\mathbf{O} = [1/c, \mathbf{0}]$ introduced in [6], $\mathbf{x}_{\text{space}}$ is defined as

$$\mathbf{x}_{\text{space}} = \mathbf{Exp}_{[1/c, \mathbf{0}]}(\mathbf{z}) = \frac{\sinh(\sqrt{c}\sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{L}}})}{\sqrt{c}\sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{L}}}}\mathbf{z} \quad (4)$$

Centroid. In the hyperbolic space, the *centroid* of N points in Lorentz space can be computed in the Klein space \mathbb{K}^n . Conversion between two spaces are

$$T_{\mathbb{L} \rightarrow \mathbb{K}}(\mathbf{x}) = \frac{\mathbf{x}_{\text{space}}}{x_{\text{time}}}, \quad T_{\mathbb{K} \rightarrow \mathbb{L}}(\mathbf{k}) = \frac{[1, \mathbf{k}]}{\sqrt{c(1 - \|\mathbf{k}\|^2)}}, \quad (5)$$

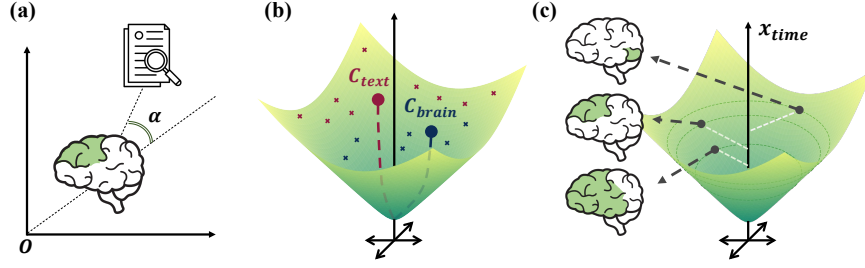


Fig. 2. Concept of Our Objective Functions. (a) The angle-based contrastive loss minimizes the exterior angle between a text embedding and its corresponding brain activation embedding. (b) The centroid loss encourages the centroid of brain embeddings $\mathbf{C}_{\text{brain}}$ to be positioned closer to the hyperboloid origin than the centroid of text embeddings \mathbf{C}_{text} . (c) The hierarchical loss aligns brain embeddings along the hyperbolic time axis by the number of activated regions for brain structural hierarchy.

where $\mathbf{k} \in \mathbb{R}^n$ is a point in \mathbb{K}^n . Lorentzian centroid is then defined as

$$\text{centroid}_{\mathbb{L}}(\{\mathbf{x}_j\}_{j=1}^N) = T_{\mathbb{K} \rightarrow \mathbb{L}} \left(\frac{\sum_{j=1}^N \gamma_j T_{\mathbb{L} \rightarrow \mathbb{K}}(\mathbf{x}_j)}{\sum_{j=1}^N \gamma_j} \right), \quad (6)$$

where the Lorentz factors $\gamma_j \in \mathbb{R}$ are defined as $1/\sqrt{1 - c\|T_{\mathbb{L} \rightarrow \mathbb{K}}(\mathbf{x}_j)\|^2}$.

3 Hyperbolic Brain-Text Representation

Given a neuroscience article indexed by i , let (b_i, t_i) denotes a pair of brain activation data and LLM-representation of its corresponding neuroscientific text. For neuroimaging meta-analysis, which seeks to identify the most relevant brain regions from a given textual description or retrieve corresponding brain activations, our goal is to embed b_i and t_i into a shared space while preserving both the brain-text hierarchy as well as the structural hierarchy of the brain.

Angle-based Contrastive Learning. Unlike contrastive learning [6] that minimizes the spatial distances in a hyperbolic space, we adopt an angle-based similarity measure [16] since regularizing geodesic distance weakens the hierarchical dependency. We employ two parallel encoders: a brain activation encoder $\mathcal{E}_{\text{brain}}$ and a text encoder $\mathcal{E}_{\text{text}}$. After mapping the b_i and t_i into their respective Euclidean latent spaces, we project each of these onto d -dimensional hyperbolic space \mathbb{L}^d via Eq. (4) as $\mathbf{z}_i^{\text{text}} = \mathcal{E}_{\text{text}}(t_i)$ and $\mathbf{z}_i^{\text{brain}} = \mathcal{E}_{\text{brain}}(b_i)$. Since each brain region can be characterized by different functionalities as depicted in Fig. 1(b), we assume that brain activations have higher hierarchy (i.e., broader context) than textual descriptions. Denoting $z_{i,\text{time}}$ and $\mathbf{z}_{i,\text{space}}$ as the time and space component of \mathbf{z}_i , exterior angle of $\mathbf{z}_i^{\text{text}}$ from $\mathbf{z}_i^{\text{brain}}$ is then defined as

$$\text{ext}(\mathbf{z}_{\text{brain}}, \mathbf{z}_{\text{text}}) = \cos^{-1} \left(\frac{z_{i,\text{time}}^{\text{text}} + z_{i,\text{time}}^{\text{brain}} \cdot c \langle \mathbf{z}_{i,\text{space}}^{\text{brain}}, \mathbf{z}_{i,\text{space}}^{\text{text}} \rangle_{\mathbb{L}}}{\|\mathbf{z}_{i,\text{space}}^{\text{brain}}\| \sqrt{(c \langle \mathbf{z}_{i,\text{space}}^{\text{brain}}, \mathbf{z}_{i,\text{space}}^{\text{text}} \rangle_{\mathbb{L}})^2 - 1}} \right). \quad (7)$$

As exterior angle is defined in $[-\pi, \pi]$, smaller absolute angle indicates higher correspondence. To increase similarity between corresponding pair and discrepancies between non-corresponding pairs, we utilize InfoNCE loss [15] to define our angle-based loss by minimizing the absolute angles of positive pairs as

$$\mathcal{L}_{\text{angle}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(-|\text{ext}(\mathbf{z}_i^{\text{img}}, \mathbf{z}_i^{\text{text}})|/\tau)}{\sum_{j=1}^N \exp(-|\text{ext}(\mathbf{z}_i^{\text{img}}, \mathbf{z}_j^{\text{text}})|/\tau)}. \quad (8)$$

Centroid Regularization. To reflect the hierarchy between the brain activations and text, $\mathbf{z}^{\text{brain}}$ needs to be positioned closer to the hyperboloid origin \mathbf{O} than \mathbf{z}^{text} as a smaller distance from \mathbf{O} corresponds to a higher hierarchy. However, $\mathcal{L}_{\text{angle}}$ alone does not have explicit control over the spatial position of the embeddings. To explicitly enforce this ordering, we first compute the centroids of the brain activation embeddings and the article text embeddings, denoted by $\mathbf{c}_{\text{brain}}$ and \mathbf{c}_{text} each, as described in Eq. (6). We then impose a constraint that guides these centroids toward predetermined target distances respectively as

$$\mathcal{L}_{\text{cent}} = \left\| d_L(\mathbf{O}, \mathbf{c}_{\text{text}}) - \frac{1}{\sqrt{c}} \cosh^{-1}(cp) \right\| + \left\| d_L(\mathbf{O}, \mathbf{c}_{\text{img}}) - \frac{1}{\sqrt{c}} \cosh^{-1}(cq) \right\|, \quad (9)$$

where $p, q \in \mathbb{R}^+$ are hyperparameters satisfying $p > q$ and $\|\cdot\|$ is the Euclidean norm. This condition ensures that the ‘broader’ brain activation embeddings are positioned closer to the hyperboloid origin than the ‘detailed’ text embeddings.

Brain Structural Hierarchy Guidance. Brain activation map exhibits varying levels of specificity from global regions (e.g., left hemisphere) to local regions (e.g., superior temporal gyrus) as illustrated in Fig. 1(a). To refine this structural hierarchy among brain embeddings, we introduce a brain structural hierarchy guidance based on the number of the activation regions. Specifically, we set R_i as how many ROIs in b_i have values greater than predefined threshold δ . The brain structural hierarchy guidance between two brain embeddings is defined as

$$\mathcal{L}_{\text{hier}} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N 1_{ij}(R_i - R_j) \cdot \max(\log(z_{i,\text{time}}^{\text{brain}}/z_{j,\text{time}}^{\text{brain}}), 0), \quad (10)$$

where 1_{ij} is an indicator function that equals 1 if $R_i > R_j$. This penalizes when a broader brain activation (i.e., a more general pattern) is wrongly placed at a lower hierarchical level (i.e., with a larger time value) than a localized activation.

Joint Optimization. The final training loss is a weighted combination as:

$$\mathcal{L} = \mathcal{L}_{\text{angle}} + \lambda_1 \mathcal{L}_{\text{cent}} + \lambda_2 \mathcal{L}_{\text{hier}}, \quad (11)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}^+$ are hyperparameters that balance the contributions of centroid regularization and brain hierarchical loss. By jointly minimizing this objective, the model learns to capture both the semantic alignment and the hierarchical structure necessary for robust neuroimaging meta-analysis.

4 Experimental Settings

Data Preparation. Our work follows the setting established by NeuroConText [12], which processed 20,674 neuroscientific articles from PubMed along with their corresponding brain activation peak coordinates. With a pretrained LLM (i.e., Mistral-7B), text representation t_i is derived from textual descriptions (e.g., abstract or body). A corresponding discrete brain activation peak coordinate is converted into continuous brain activation voxel maps via Kernel Density Estimation (KDE), and subsequently projected onto a 512-dimensional coefficients using the Dictionary of Functional Modes (DiFuMo) atlas as b_i .

Implementation Details. For $\mathcal{E}_{\text{text}}$ and $\mathcal{E}_{\text{brain}}$, we use a two-layer and three-layer residual MLP, respectively, followed by layer normalization. We train the model for 200 epochs using the AdamW optimizer [11] with a learning rate of 1×10^{-4} , a batch size of 4096, and a weight decay of 0.05. We set $\lambda_1 = 0.5$, $\lambda_2 = 30$, $\delta = 5$, $p = 2.0$, and $q = 0.5$. We use pretrained weights for Text2Brain [14] and NeuroQuery [7], while NeuroConText [12] is trained under the same settings as our model. *The code and dataset will be publicly available.*

5 Results and Analysis

Brain-Text Cross Modal Retrieval. To evaluate our semantic alignment performance, we perform text-to-brain and brain-to-text retrievals. In the text-to-brain setting, the model ranks candidate brain activations by their relevance to given textual descriptions (e.g., abstracts or full-body text). The brain-to-text setting reverses it, from a brain activation to the most relevant article. We adopt a 10-fold cross-validation to rigorously evaluate recall@K, $K \in \{5, 10, 100\}$. Recall@K evaluates whether the true pair appears among the top-K similar candidates, making it suitable for ranking tasks. For comparison, the following baselines are adopted. Since Text2Brain [14] and NeuroQuery [7] directly generate 3D brain activations from text, mean squared error (MSE) is used as similarity

Table 1. Comparison of Brain-Text Cross-Modal Retrieval Performance using two types of textual descriptions. The best performance is indicated in **bold**.

Retrieval		Text→Brain			Brain→Text		
Metric [%]		Recall@5	Recall@10	Recall@100	Recall@5	Recall@10	Recall@100
Abstract	Text2Brain [14]	0.247±0.026	0.469±0.022	4.837±0.075	-	-	-
	NeuroQuery [7]	4.266±0.262	7.560±0.389	36.055±0.761	-	-	-
	NeuroConText [12]	9.964±0.663	15.033±0.808	45.120±1.098	10.046±0.683	15.183±0.790	45.168±1.077
	MNM-Reverse	8.852±1.224	13.486±2.585	42.585±3.000	10.332±0.562	15.096±0.964	44.636±1.270
	MNM <i>w/o</i> $\mathcal{L}_{\text{hier}}$	10.699±0.650	15.807±0.653	45.734±1.286	10.458±0.590	16.218±0.764	46.406±1.057
	MNM	10.728±0.673	15.860±0.668	45.864±1.417	10.593±0.627	16.320±0.822	46.590±1.217
Body	Text2Brain [14]	0.237±0.014	0.479±0.015	4.832±0.079	-	-	-
	NeuroQuery [7]	4.421±0.374	7.696±0.458	36.055±0.562	-	-	-
	NeuroConText [12]	13.660±0.520	20.228±0.742	51.945±0.961	13.747±0.579	19.517±0.449	51.572±0.771
	MNM-Reverse	13.756±0.575	20.136±0.485	51.969±0.736	14.087±0.556	21.176±0.575	53.134±0.845
	MNM <i>w/o</i> $\mathcal{L}_{\text{hier}}$	14.288±0.747	20.649±0.850	52.670±0.735	14.569±0.648	21.302±0.396	53.342±0.797
	MNM	14.434±0.618	20.741±0.574	52.709±0.928	14.806±0.449	21.360±0.414	53.623±0.698

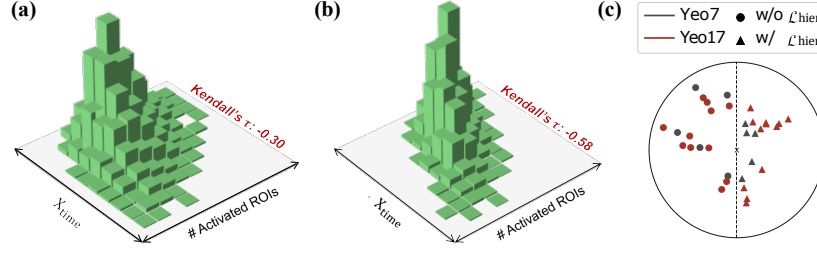


Fig. 3. Effect of Brain Hierarchical Loss. (a) Histogram between x_{time} and the number of activated ROIs without $\mathcal{L}_{\text{hier}}$, showing -0.30 of Kendall’s τ [17] (b) Same histogram with $\mathcal{L}_{\text{hier}}$, decreasing Kendall’s τ to -0.58 . (c) 2D Poincaré disk projection of Yeo7 and Yeo17 network [21] embeddings (left: w/o $\mathcal{L}_{\text{hier}}$, right: w/ $\mathcal{L}_{\text{hier}}$).

measure between the predicted brain activation and each candidate. Unlike [12] using cosine similarity, MNM adopts the negation of the absolute exterior angle.

For MNM, we employ two additional settings: (1) MNM-Reverse that reverses the hierarchy of MNM by flipping the brain and text relationship in \mathcal{L}_{ext} and $\mathcal{L}_{\text{cent}}$, and (2) MNM w/o $\mathcal{L}_{\text{hier}}$. As shown in Table 1, our method consistently outperforms all baselines in both text-to-brain and brain-to-text retrieval, with the largest improvement observed in brain-to-text retrieval when using full-body text. Degraded performances in two additional settings demonstrate the rationality of brain-text hierarchy and the need for explicit hierarchy guidance.

Brain Hierarchy Analysis. To show the effect of $\mathcal{L}_{\text{hier}}$, we analyze how samples are distributed in the latent space. One of our goals is to model the inherent hierarchy of brain activation patterns, such that more general (i.e., widely activated) regions lie closer to the hyperbolic origin than more specific (i.e., smaller or functionally narrower) ones. To verify the assumption with the trained embeddings using $\mathcal{L}_{\text{hier}}$, we plot histograms of x_{time} against the number of activated ROIs, along with Kendall’s τ [17] to quantify the ordinal correlation in a range $[-1, 1]$. A negative τ indicates that embeddings with broader activations tend to reside at a higher “level” (i.e., smaller x_{time}) in the hyperbolic space. In Fig. 3, while embeddings are dispersed (a) without $\mathcal{L}_{\text{hier}}$ resulting in -0.30 of Kendall’s τ , they are more inversely aligned (b) with $\mathcal{L}_{\text{hier}}$ resulting in Kendall’s $\tau = -0.58$.

We further examine two well-known cortical parcellation schemes, Yeo7 and Yeo17 networks [21]. The Yeo7 network segments the brain into seven broad functional systems, while the Yeo17 network refines this into 17 detailed subdivisions. To visualize these divisions, we map them into the learned hyperbolic space and project them onto a 2D Poincaré disk [4]. Due to page limitations, we visualize only half of the Poincaré disk, varying the presence of $\mathcal{L}_{\text{hier}}$. In Fig. 3(c), the embeddings from both networks exhibit no discernible order in the left half when $\mathcal{L}_{\text{hier}}$ is absent. However, with $\mathcal{L}_{\text{hier}}$, Yeo7 embeddings cluster near the origin in the right half, enhancing interpretability. This organization supports a coarse-to-fine analysis by selecting x_{time} from brain activation map candidates.

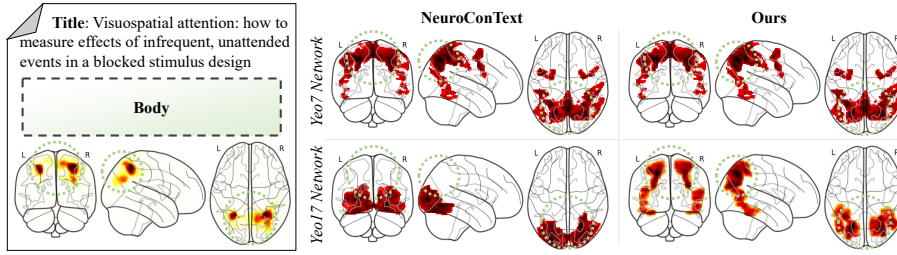


Fig. 4. Reconstructed Brain Activation Map from a Neuroscience Article [8]. While NeuroConText [12] fails to capture activation in the intraparietal sulcus using Yeo17 networks, MNM assigns high similarity scores to them as in the ground truth.

Brain Activation Map Prediction. To evaluate text-brain alignment beyond the task-specific decoder, we use the embeddings of Yeo7 and Yeo17 networks [21] as our brain activation bases. With body text of given articles, we calculate the similarity of the text embedding and brain activation bases of two networks. NeuroConText [12] is a unique baseline, since it aligns both modalities in the shared latent space as we do. We employ the same similarity metric as in the brain-text cross-modal retrieval experiment. The resulting similarities are transformed into probabilities using a softmax function and assigned to the corresponding voxels in the 3D brain. For better visualization, we threshold the top 10% of voxels and visualize using brain glass schematics via the Nilearn [1] software.

Figure 4 illustrates how embeddings can be recovered using the Yeo7 and Yeo17 networks for a study [8] that investigates the neural mechanisms underlying visuospatial attention reorientation. The study demonstrates that both blocked and event-related analyses consistently activate the intraparietal sulcus and superior parietal cortex. In coarse prediction using the Yeo7 networks, both NeuroConText and MNM exhibit a high correlation (i.e., top 10% similarity) with the activated regions. However, with the finer-grained Yeo17 network, NeuroConText fails to capture the target regions, whereas our method successfully identifies them. Since most detected regions overlap across different parcellation levels, MNM demonstrates greater consistency compared to NeuroConText. These results indicate that our hyperbolic representation can flexibly adapt to various parcellation schemes when inferring activation maps from articles.

6 Conclusion

In this work, we propose MNM, a novel framework that leverages hyperbolic geometry to capture the inherent hierarchical structure of brain activations while aligning them more effectively with corresponding textual descriptions. Through extensive experiments, we demonstrate that MNM not only surpasses baselines in cross-modal retrieval but also provides greater interpretability in activation map prediction. By providing robust performances across various parcellation schemes, MNM paves the way for multi-level neuroimaging meta-analysis.

Acknowledgments. This research was supported by NRF-2022R1A2C2092336 (50%), RS-2025-02216257 (20%), RS-2022-II220290 (20%), and RS-2019-III191906 (AI Graduate Program at POSTECH, 10%).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abraham, A., Pedregosa, F., Eickenberg, M., et al.: Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics* **8**, 71792 (2014)
2. Beltagy, I., Lo, K., Cohan, A.: Scibert: A pretrained language model for scientific text. In: *EMNLP-IJCNLP*. pp. 3615–3620. *ACL* (2019)
3. Button, K.S., Ioannidis, J.P., Mokrysz, C., et al.: Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience* **14**(5), 365–376 (2013)
4. Cvetkovski, A., Crovella, M.: Multidimensional scaling in the poincaré disk. *arXiv preprint arXiv:1105.5332* (2011)
5. Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., et al.: Fine-grain atlases of functional modes for fmri analysis. *NeuroImage* **221**, 117126 (2020)
6. Desai, K., Nickel, M., Rajpurohit, T., et al.: Hyperbolic image-text representations. In: *ICML*. pp. 7694–7731. *PMLR* (2023)
7. Dockès, J., Poldrack, R.A., Primet, R., et al.: Neuroquery, comprehensive meta-analysis of human brain mapping. *elife* **9**, e53385 (2020)
8. Giessing, C., Thiel, C.M., Stephan, K.E., et al.: Visuospatial attention: how to measure effects of infrequent, unattended events in a blocked stimulus design. *Neuroimage* **23**(4), 1370–1381 (2004)
9. Jiang, A.Q., Sablayrolles, A., Mensch, A., et al.: Mistral 7b (2023), <https://arxiv.org/abs/2310.06825>
10. Laird, A.R., Lancaster, J.J., Fox, P.T.: Brainmap: the social evolution of a human brain mapping database. *Neuroinformatics* **3**, 65–77 (2005)
11. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: *ICLR* (2019)
12. Meudec, R., Ghayem, F., Dockès, J., et al.: Neurocontext: Contrastive text-to-brain mapping for neuroscientific literature. In: *MICCAI*. pp. 325–335. *Springer* (2024)
13. Minkowski, H.: *Raum und zeit*. *Springer* (1988)
14. Ngo, G.H., Nguyen, M., Chen, N.F., et al.: Text2brain: Synthesis of brain activation maps from free-form text query. In: *MICCAI*. pp. 605–614. *Springer* (2021)
15. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018)
16. Ramasinghe, S., Shevchenko, V., Avraham, G., et al.: Accept the modality gap: An exploration in the hyperbolic space. In: *CVPR*. pp. 27263–27272 (2024)
17. Sen, P.K.: Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American statistical association* **63**(324), 1379–1389 (1968)
18. Sim, J., Lee, M., Wu, G., Kim, W.H.: Multi-modal graph neural network with transformer-guided adaptive diffusion for preclinical alzheimer classification. In: *MICCAI*. pp. 511–521. *Springer* (2024)
19. Szucs, D., Ioannidis, J.P.: Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS biology* **15**(3), e2000797 (2017)

20. Yarkoni, T., Poldrack, R.A., Nichols, T.E., et al.: Large-scale automated synthesis of human functional neuroimaging data. *Nature methods* **8**(8), 665–670 (2011)
21. Yeo, B.T., Krienen, F.M., Sepulcre, J., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* (2011)