2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10781876

Scepter: Weakly Supervised Framework for Spatiotemporal Dense Prediction of 4D Dynamic Brain Networks

Behnam Kazemivash*
dept. Computer Science
Georgia State University
Atlanta, USA
bkazemivash1@gsu.edu

Pranav Suresh
dept. Computer Science
Georgia State University
Atlanta, USA
pnadigapusuresh1@gsu.edu

Jingyu Liu
dept. Computer Science
Georgia State University
Atlanta, USA
jliu75@gsu.edu

Dong Hye Ye
dept. Computer Science
Georgia State University
Atlanta, USA
dongye@gsu.edu

Vince Calhoun
dept. Computer Science
Georgia State University
Atlanta, USA
vcalhoun@gsu.edu

***Abstract*—Spatiotemporal brain dynamism is a complex phenomenon, characterized by dynamic patterns of neural activity that unfold across both space and time. However, capturing these dynamic patterns poses a formidable challenge due to the sheer complexity of neural interactions and the demand for advanced computational models. In this context, we have harnessed advances in computer vision and formulated this challenging issue as the weakly supervised spatiotemporal dense prediction of dynamic brain networks. To accomplish this, we have developed a novel framework for encoding spatiotemporal characteristics of functional magnetic resonance imaging (fMRI) data to densely predict dynamic brain networks, each encompassing 4D maps that vary over time and between subjects. The backbone of our framework is an isotropic model architecture that contains a deep stack of pre-activated ConvMixer modules. Furthermore, we introduce a strategy for generating prior information, which serves as weak supervision for training the model, since no benchmark currently exists for addressing the dynamic brain network issue and annotating fMRI data proves to be an expensive and inaccurate process. We also address some of the significant drawbacks in popular brain parcellation methods. Finally, our experimental results indicate the method's ability to generate plausible brain network maps that are highly dynamic and consistent with previous findings in brain dynamics. The proposed advancement in generating brain dynamic maps transcends the boundaries of conventional neuroscience research, ushering in a paradigm shift which facilitates the discovery of new perspectives on the complexity of brain function.**

***Index Terms*—Dense Prediction, Weakly Supervised Learning, Spatiotemporal Encoding, Dynamic Brain Mapping, Brain Parcellation**

I. Introduction

The human brain is a spatiotemporal complex system with billions of interconnected neurons forming intricate computational networks and memory units that exhibit dynamic activity patterns across both space and time, playing vital role in shaping cognitive ability, memories, emotions, sensory perception, and eventually our identity. Thus, studying brain dynamism is essential for advancing our understanding of consciousness and cognition, as well as for finding reliable biomarkers and developing interventions for brain disorders [1]. However, this phenomenon has been comparatively less explored due to the inherent complexities of the brain organization.

Over the past decade, research on brain dynamics has grown, with a focus on temporal rather than spatial dynamism [2], [3]. Some methods analyze brief, spontaneous co-activation patterns in fMRI data [4], but they may encounter challenges in identifying meaningful patterns when activities temporally overlap [5]. Additionally, quasi-periodic patterns in spatially dynamic approaches reveal recurring spatiotemporal brain activity patterns and evolving spatial networks (spatial chronnectome) [6]. However, drawbacks include the significant impact of user-defined parameters on results. The spatial chronnectome approach allows for the estimation of overlapping temporally evolving networks within an ICA-based brain parcellation framework [7], providing an approach to jointly parcellate and estimate dynamics; however, initial approaches focus on linear decomposition methods.

Recent advances in computer vision enable us to decipher the principles underlying brain dynamics, while moving beyond linear approaches. In this work, we contribute to the field by addressing a gap in neuroscience research related to dynamic brain maps. We introduce a novel, weakly supervised framework for dense prediction of dynamic brain networks. In computer vision, dense prediction entails the intricate task of generating pixel-wise (or voxel-wise) predictions, allowing for the capture of detailed spatial information essential for nuanced analysis [8]. Additionally, weakly supervised learning represents a paradigm in machine learning designed to train

This work was supported in part by the U.S. National Institutes of Health under grant R01MH123610 and the U.S. National Science Foundation under grant 2112455.

models using ambiguous, noisy, or incomplete data. This approach proves valuable in addressing challenges associated with the high expenses and complex nature of annotating procedures, offering a solution for gathering supervision information [9].

All in all, our goal is to encode the spatiotemporal characteristics of fMRI data, capturing brain dynamics and ultimately generating dynamic representations of input data that illustrate 4D brain maps varying over time for each voxel. This sets it apart from atlas-based methods, which provide fixed-size templates and ignore the natural variability of the brain in terms of size, shape, and gyrification, or ICA-based approaches (i.e., a soft parcellation method) that typically consider a fixed spatial map for each network over time, with variability only captured by a single time-course. Moreover, the dynamic nature of our method makes it feasible to study the contribution of each voxel in multiple networks and potential temporally overlapped patterns. Our method is computationally efficient during the deployment phase, surpassing other brain parcellation approaches, and has no user-defined parameters for end-users, enhancing its stability compared to other techniques. Last but not least, our method has the potential to expand the horizons of brain dynamics research and enrich relevant studies by providing 4D dynamic maps, which can be used as a fundamental tool for further studies.

II. Methodology

A. Scepter Framework

We formulate brain dynamism phenomena as a 4D soft brain parcellation task with weak supervision and then explore a strong backbone architecture for spatiotemporal dense prediction of dynamic brain networks. To do this, we design the Scepter framework including an isotropic encoder with pre-activated ConvMixer modules. We also leverage advances in weakly supervised learning paradigm to train the model with priors, generated by applying ICA on fMRI and merging extracted components and time-courses. The overview of our framework architecture is shown in Fig 1.

Let $S_i, P_i \in \mathbb{R}^{t\times x\times y\times z}$ denote a fMRI image and its prior for a specific brain network n, where t is the number of time-points and x, y, z are space dimension respectively. We want to learn a function like $\hat{f}_n$ for each brain network n, such a way that encodes spatiotemporal information of the data and generate a representation similar to the given prior while preserving temporal variability. To do this, the input image is split into a series of non-overlapping patches $S_i = \left[s_i^1, ..., s_i^k\right]$ where q is the patch size and k is the number of patches, using a patch embedding module that contains a 3D batch normalization, non-linearity function, and convolution layer respectively for implementing pre-activation scenario. We also encode index of timepoints using a temporal position embedding $e \in \mathbb{R}^{t\times q^3}$ with similar shape as patches to reduce asynchronization effects of brain activity patterns in different subjects.

$$A_0 = conv\left(\sigma\left(bn\left(S_i\right)\right)\right) + e \tag{1}$$

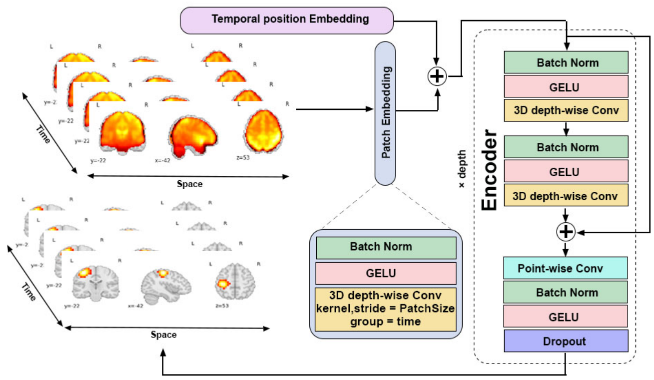


Fig. 1. Schematic of the proposed Scepter framework for encoding spatiotemporal information to densely predict 4D dynamic brain networks using weak supervision. Our framework comprises an isotropic architecture that considers the temporal dimension as an input channel and subdivides the fMRI input image into patches. These patches are subsequently fed into a series of k encoders to generate a spatiotemporal dynamic brain map.

Where $conv$ is a normal 3D convolution, $\sigma(.)$ is the non-linear function (Gaussian Error Linear Unit - GELU), and bn refers to 3D batch normalization layer. Next, our architecture is followed by k number of encoders, including two redundant set of residual pre-activated depth-wise convolution layers followed by a normal point-wise convolution and a dropout layer.

$$A_l = \widetilde{conv}\left(\sigma\left(bn\left(\widetilde{conv}\left(\sigma\left(bn\left(A_{l-1}\right)\right)\right)\right)\right)\right) \tag{2}$$

$$A_{l+1} = \sigma\left(bn\left(c\dot{o}nv\left(A_l\right)\right)\right) \tag{3}$$

Where $\widetilde{conv}$ is a depth-wise convolution with depth of t and $c\dot{o}nv$ refers to point-wise convolution. Our base architecture begins with a patch embedding layer and d encoders all stacked together to form our isotropic architecture.

B. Generating Prior

Generating dynamic brain maps faces a significant barrier due to the lack of annotated data, which is costly, complex, and prone to inaccuracies. These challenges drive the need for alternative approaches, motivating us to leverage ICA spatial maps and time-courses for generating priors as weak supervision for model training.

To extract spatial maps and time-courses, we employed the NeuroMark template [10] — a template consists of replicable independent components (ICs) that are computed by spatially aligning correlated group-level ICs from two extensive healthy control fMRI datasets with more than 800 subjects. Subsequently, we utilized the template for running a spatially constrained ICA algorithm on a per-subject basis, enabling us to detect 53 maximally independent resting state networks (RSNs), each comprising a spatial map $C_i^n \in \mathbb{R}^{53\times 63\times 52}$ and its corresponding time-course $\vec{TC}_i^n$ with size of timepoint count. Therefore, we compute the prior for each of networks by applying a brain mask on spatial maps and vectorizing it, then computing outer product of the time-course and vectorized spatial map as follows:

$$P_i^n = \vec{TC}_i^n \otimes \vec{C}_i^n \tag{4}$$

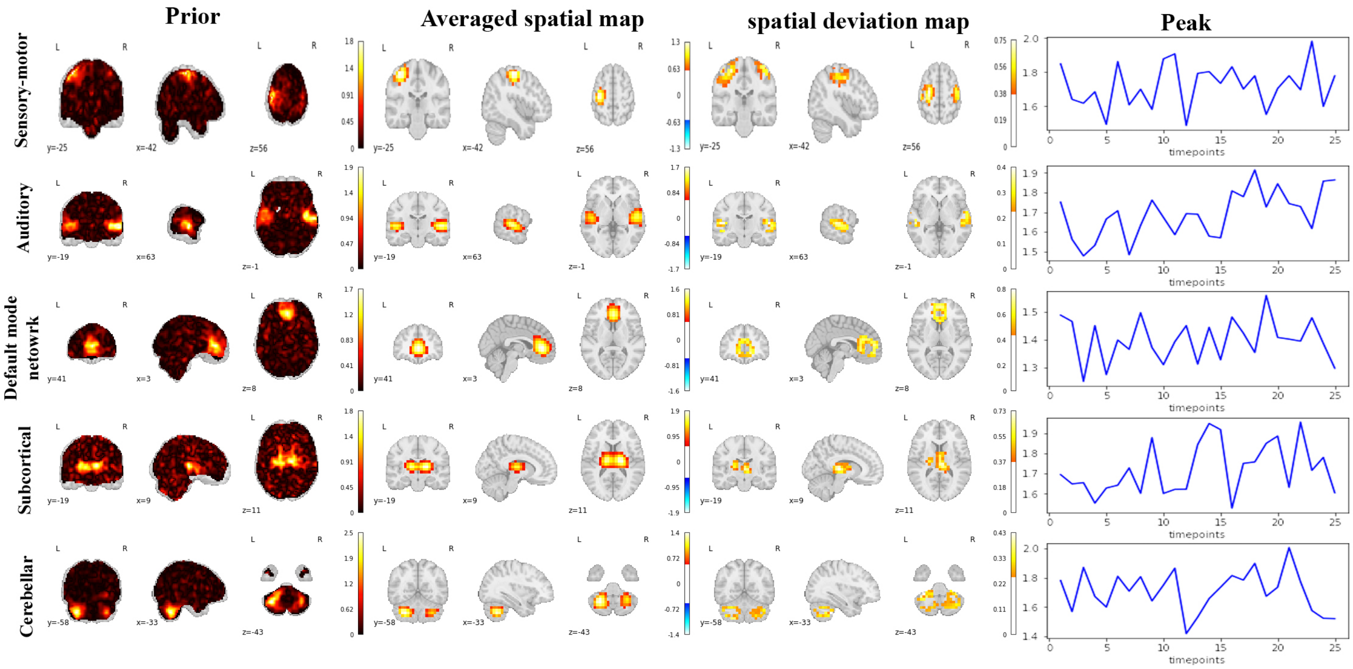
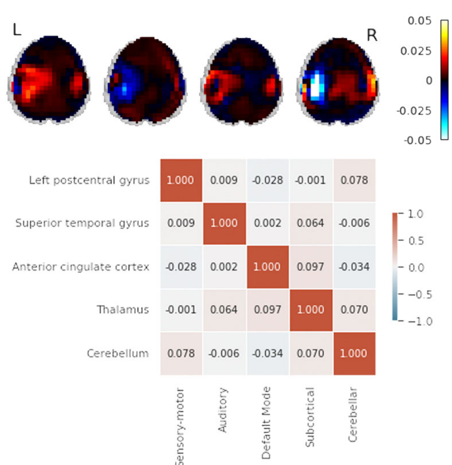


Fig. 2. Qualitative comparison with the prior was conducted for a randomly selected subject from the test set. Spatial variability in sample sensory-motor, auditory, default mode network, subcortical, and cerebellar regions was analyzed for the given subject. The first column displays the prior averaged over time, while the second column demonstrates the generated dynamic map averaged over time. In the third column, spatial deviation is presented, showcasing the remarkable ability of our framework to encode spatial dynamics. Interestingly, the right postcentral gyrus reveals distinct activities in the spatial deviation map within the sensory-motor network, not visible in the averaged map. The fourth and final column showcases the peak voxel value over time, providing further evidence of spatial variability over time. The model's output is thresholded to deliver more detailed information.



	Sensory-motor	Auditory	Default Mode	Subcortical	Cerebellar
Left postcentral gyrus	1.000	0.009	-0.028	-0.001	0.078
Superior temporal gyrus	0.009	1.000	0.002	0.064	-0.006
Anterior cingulate cortex	-0.028	0.002	1.000	0.097	-0.034
Thalamus	-0.001	0.064	0.097	1.000	0.070
Cerebellum	0.078	-0.006	-0.034	0.070	1.000

Fig. 3. The upper panel displays horizontal slice of 5 consecutive timepoints post mean removal for the Sensory-Motor network, emphasizing the spatial variation. In the lower panel, a heatmap illustrates the functional network connectivity across sensory-motor, auditory, default mode, subcortical, and cerebellar networks. The figure unveils a lower correlation between these diverse networks, consistent with previous research findings.

Finally, we unmask the computed matrix to get the 4D spatiotemporal tensor and use the generated prior $P_i \in \mathbb{R}^{t \times x \times y \times z}$ as weak supervision to train our model.

III. Experiments

Our backbone architecture contains a patch embedding unit that split and project the input image into set of patches by utilizing a pre-activated depth-wise 3D convolution with group size of t, and same patch size of 2 for both kernel and stride. We use a small patch size as the spatial dimension of input image is not large enough and using a bigger patch size will result in artifacts in generated map. Next to the patch embedding unit, there is a stack of 9 encoders to capture spatiotemporal information, each of which has two identical pre-activated depth-wise convolution layers with kernel size of 3 and padding of 1 to preserve shape of extracted features, followed by a normal point-wise convolution all with group ratio of t, and a dropout layer with ratio of 0.1. Moreover, we trained the model on two V100 GPU with 32 GB of dedicated memory by minimizing mean squared error between prior and generated 4D tensor along with Adam optimizer with learning ratio of 0.001, batch size of 3, weight decay ratio of 0.05 for 1000 epochs and eventually early stopping policy for preventing overfitting issue with $\epsilon = 10^{-8}$. Also, we train and evaluate our model on BSNIP dataset [11] using 182 healthy control samples (train set =150, test set = 32), collected from 4 different sites by 3.0 T Siemens Verio Scanner with an 8-channel SENSE head coil, 3.0 T Signa HDx GE Scanner, and 3.0 T Siemens Trio Trim Scanner. Additionally, we employed a straightforward technique for uniformly sampling timepoints with step size of 4 as utilizing all timepoints would be computationally intensive resulted in sample shape of $S_i \in \mathbb{R}^{25\times53\times63\times52}$, and smoothed the input image by gaussian filter with $\sigma = 0.5$ and then applied z-scoring to the input images to ensure proper normalization. Finally, we trained the model for different brain regions including subcortical, cerebellar, sensory-motor, auditory, default mode network.

A. Qualitative results

Our qualitative results demonstrate the capability of our framework to generate plausible maps with globally similar patterns to the prior, as depicted in Fig 2. Moreover, the spatial deviation maps —depicting absolute differences between consecutive timepoints— exhibit smooth changes around active regions in all the maps, with variable amplitudes for different subjects. This collective evidence highlights the model's proficiency in capturing spatiotemporal brain dynamism. More intriguingly, activities in the right postcentral gyrus of the spatial deviation map within the sensory-motor network become apparent, which remains unseen in the averaged map. This observation aligns with previous anatomical knowledge of the sensory-motor network. Furthermore, we explored functional network connectivity, as shown in Fig 3, revealing a significant low correlation among distinct networks, indicating their functional differentiation.

B. Quantitative results

Currently, as there is no similar work on dynamic brain maps, there are no direct comparisons to competing state-of-the-art methods. Moreover, there is no specific metric to evaluate generated maps directly. Thus, we study the quality of generated maps in terms of the global difference with prior maps using the mean Absolute Relative Error (mARE), localization of the active region with Intersection over Union (IOU), visual quality of maps with the structural similarity index measure (SSIM), and eventually homogeneity which is correlation of voxel time-series within the ROI. The model's ability to capture dynamic brain activities while preserving the global pattern of the brain map for the given networks is demonstrated by high IOU, SSIM, and Homogeneity values, along with low mARE, as shown in Table I.

TABLE I
RESULTS ON THE LOCALIZATION AND VISUAL QUALITY OF GENERATED DYNAMIC BRAIN NETWORKS.

Brain Networks	**Quantitative metrics**			
	mARE $\downarrow$	***IOU*** $\uparrow$	***SSIM*** $\uparrow$	**Homogeneity** $\uparrow$
Default Mode	0.18	0.64	0.82	0.98
Auditory	0.23	0.57	0.8	0.99
Sensory-Motor	0.24	0.55	0.8	0.94
Subcortical	0.21	0.58	0.81	0.98
Cerebellar	0.23	0.56	0.81	0.98

IV. CONCLUSION

In this work, we propose a novel weakly supervised framework for spatiotemporal dense prediction of 4D dynamic brain networks. Our approach surpasses current brain parcellation methods by offering insights into complete spatiotemporal variations across space, time, and networks. We leverage advances in weakly supervised learning to generate a prior for training the model by merging spatial maps and timecourses extracted from input fMRI data through the application of the ICA algorithm. This choice is driven by the lack of a benchmark for creating 4D dynamic brain maps. The results from the test set reveal that our framework generates maps with sensitivity to individual variations. We additionally examined spatial variability by computing the summation of absolute spatial deviation. This method unveils regions that exist within the network boundaries but deviate from the overall mean activity. This unique aspect of our approach offers a further understanding of brain activity. Moreover, we assessed the temporal relationships between networks by calculating functional network connectivity, which exhibited the anticipated modular pattern. We believe that a pivotal aspect of Scepter, involving the encoding of spatiotemporal information to construct dynamic brain networks, offers extensive suitability for investigating brain disorders, potentially identifying biomarkers, and understanding cognitive processes. Several upcoming studies in this vein are part of our planned research agenda.

REFERENCES

[1] R. Liégeois, J. Li, R. Kong, C. Orban, D. Van De Ville, T. Ge, M. R. Sabuncu, and B. T. Yeo, "Resting brain dynamics at different timescales capture distinct aspects of human behavior," *Nature communications*, vol. 10, no. 1, p. 2317, 2019.

[2] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı, "The chronnectome: time-varying connectivity networks as the next frontier in fmri data discovery," *Neuron*, vol. 84, no. 2, pp. 262–274, 2014.

[3] A. Iraji, R. Miller, T. Adali, and V. D. Calhoun, "Space: a missing piece of the dynamic puzzle," *Trends in cognitive sciences*, vol. 24, no. 2, pp. 135–149, 2020.

[4] T. Matsui, T. Q. Pham, K. Jimura, and J. Chikazoe, "On co-activation pattern analysis and non-stationarity of resting brain activity," *NeuroImage*, vol. 249, p. 118904, 2022.

[5] A. Iraji, A. Faghiri, Z. Fu, P. Kochunov, B. M. Adhikari, A. Belger, J. M. Ford, S. McEwen, D. H. Mathalon, G. D. Pearlson *et al.*, "Moving beyond the 'cap'of the iceberg: Intrinsic connectivity networks in fmri are continuously engaging and overlapping," *Neuroimage*, vol. 251, p. 119013, 2022.

[6] A. Abbas, M. Belloy, A. Kashyap, J. Billings, M. Nezafati, E. H. Schumacher, and S. Keilholz, "Quasi-periodic patterns contribute to functional connectivity in the brain," *Neuroimage*, vol. 191, pp. 193–204, 2019.

[7] A. Iraji, T. P. Deramus, N. Lewis, M. Yaesoubi, J. M. Stephen, E. Erhardt, A. Belger, J. M. Ford, S. McEwen, D. H. Mathalon *et al.*, "The spatial chronnectome reveals a dynamic interplay between functional segregation and integration," *Human brain mapping*, vol. 40, no. 10, pp. 3058–3077, 2019.

[8] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3614–3633, 2021.

[9] K. P. Murphy, *Probabilistic machine learning: an introduction*. MIT press, 2022.

[10] Y. Du, Z. Fu, J. Sui, S. Gao, Y. Xing, D. Lin, M. Salman, A. Abrol, M. A. Rahaman, J. Chen *et al.*, "Neuromark: An automated and adaptive ica based pipeline to identify reproducible fmri markers of brain disorders," *NeuroImage: Clinical*, vol. 28, p. 102375, 2020.

[11] S. Liu, H. Wang, M. Song, L. Lv, Y. Cui, Y. Liu, L. Fan, N. Zuo, K. Xu, Y. Du *et al.*, "Linked 4-way multimodal brain differences in schizophrenia in a large chinese han population," *Schizophrenia bulletin*, vol. 45, no. 2, pp. 436–449, 2019.