

# Electromagnetic Source Imaging With a Combination of Sparse Bayesian Learning and Deep Neural Network

Jiawen Liang<sup>1b</sup>, Zhu Liang Yu<sup>1b</sup>, *Member, IEEE*, Zhenghui Gu<sup>1b</sup>, *Member, IEEE*,  
and Yuanqing Li<sup>1b</sup>, *Fellow, IEEE*

**Abstract**—Accurate reconstruction of the brain activities from electroencephalography and magnetoencephalography (E/MEG) remains a long-standing challenge for the intrinsic ill-posedness in the inverse problem. In this study, to address this issue, we propose a novel data-driven source imaging framework based on sparse Bayesian learning and deep neural network (SI-SBLNN). Within this framework, the variational inference in conventional algorithm, which is built upon sparse Bayesian learning, is compressed via constructing a straightforward mapping from measurements to latent sparseness encoding parameters using deep neural network. The network is trained with synthesized data derived from the probabilistic graphical model embedded in the conventional algorithm. We achieved a realization of this framework with the algorithm, source imaging based on spatio-temporal basis function (SI-STBF), as backbone. In numerical simulations, the proposed algorithm validated its availability for different head models and robustness against distinct intensities of the noise. Meanwhile, it acquired superior performance compared to SI-STBF and several benchmarks in a variety of source configurations. Additionally, in real data experiments, it obtained the concordant results with the prior studies.

**Index Terms**—E/MEG source imaging, sparse Bayesian learning, deep neural network, variational inference.

## I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG) and magnetoencephalography (MEG) are widely used noninvasive techniques to explore human brain function with millisecond temporal resolution. Reconstruction of the cortical activities from the E/MEG recordings plays an important role in the field of neuroscience research and disease diagnose [1], [2]. The aforementioned problem can be addressed by solving the forward and inverse models sequentially. In forward model, the lead field matrix, which represents the linear relation between sources and measurements, is obtained depending on the configuration of sensors as well as the geometric and electric attributes of individual brain anatomy [3]. The inverse problem aims at estimating brain current sources from E/MEG recordings based on the known lead field matrix. In distributed source imaging models, current sources are located on voxels discretized over the entire cortical surface. Unfortunately, it involves solving an inherently ill-posed linear inverse problem since the source space dimensionality ( $\sim \mathcal{O}(10^4)$ ) always enormously exceeds the sensor space's ( $\sim \mathcal{O}(10^2)$ ).

In order to acquire a unique meaningful estimate in the neurophysiology context, prior assumptions are used to narrow the solution space. In the minimum-norm estimate (MNE) [4],  $L_2$ -norm of sources was introduced and interpreted as the underlying energy should be minimized. As a variant of MNE, low resolution brain electromagnetic tomography (LORETA) [5] and standard LORETA [6] employed the discrete Laplacian operator to reflect the spatial correlation between discretized voxels. Notably, the aforementioned algorithms can be framed into the Bayesian schema with  $L_2$ -norm interpreted as the ingredient of negative log of likelihood function. Under empirical Bayesian framework [7], also referred to as sparse Bayesian learning (SBL), the hyper parameters introduced in regularization terms can be released and optimized with other variables simultaneously, which induce the estimates with high sparsities. Champagne [8] was a representative SBL algorithm and outperformed several benchmarks in various

Manuscript received 11 October 2022; revised 1 February 2023; accepted 16 February 2023. Date of publication 15 March 2023; date of current version 22 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61836003, Grant 62276102, and Grant 61906211; and in part by the Technology Innovation 2030 under Grant 2022ZD0211700. (*Corresponding author: Zhu Liang Yu.*)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Jiawen Liang is with the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou 510641, China, and also with the Pazhou Laboratory, Guangzhou 510330, China.

Zhu Liang Yu is with the School of Automation Science and Engineering and the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou 510641, China, and also with the Pazhou Laboratory, Guangzhou 510330, China (e-mail: zlyu@scut.edu.cn).

Zhenghui Gu and Yuanqing Li are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China, and also with the Pazhou Laboratory, Guangzhou 510330, China.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3251420>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3251420

source configurations. In addition, the usage of constraints in temporal domain also benefits to source localization. Temporal basis functions (TBFs) are widely used in the inverse problem, with which we are allowed to formulate the model in latent space with dimensionality tremendously smaller than the number of time points. TBFs can be predefined [9] or extracted in data-driven manner, such as singular value decomposition (SVD) [10] and stimulus-evoked factor analysis (SEFA) [11]. Furthermore, in source imaging based on spatio-temporal basis function (SI-STBF) [12], TBFs were alternately updated with other coefficients in the variational inference.

With the rapid development of deep learning in last decade, more and more attempts have been made to solve the inverse problem using deep neural networks (DNNs), attracted by whose effectiveness in building up the mapping between variables via extracting the underlying correlation from large-scale training dataset [13]. For example, a DNN model composed of long short-term memory (LSTM) layers was presented in [14] to estimate the location and time-course of single-dipole source. It formulated a sequence-to-sequence (seq2seq) network with the same length between spatio-temporal input and output. Hecker et al. proposed a tailored convolutional neural network (CNN) called ConvDip [15], which converted an interpolated 2D image of EEG data on single time point into the corresponding estimated source vector. In source imaging framework network (SIFNet) [16], the inverse problem was reinterpreted as a supervised multi-class classification task and the discriminative model was constructed based on residual convolutional layers [17]. In data-synthesized spatio-temporal denoising autoencoder (DST-DAE) [18], a denoising autoencoder [19] based model was trained with samples generated via a data synthesis strategy with the usage of TBFs.

Different from conventional imaging algorithms, where the prior knowledge is imposed through constructing the probabilistic generative model with explicit mathematical expressions, the emerging DNN-based methods implicitly integrate the prior assumptions in the derivation of training data. Nevertheless, most of the presented approaches generate the data in heuristic fashion with little mathematical foundation. On the contrary, probabilistic generative model has the advantage in deriving the explicit and hierarchical correlation between variables, but it often remains intractable to acquire the true posterior distributions of latent variables. Even though we consider the approximation using mean field assumption, it takes much time on the alternately update for groups of parameters in every complete variational inference process.

In this paper, to solve the inverse problem, we propose a novel source imaging framework with a combination of sparse Bayesian learning and deep neural network (SI-SBLNN). Specifically, the network is employed to construct a mapping from observed measurements to latent sparseness encoding parameters in the probabilistic graphical model of specific SBL based algorithm. The network is trained with the synthesized data derived from the above probabilistic model. Once the training is accomplished, the variational inference in original algorithm can be significantly compressed because the critical parameters are directly output from the trained network and the convergence can be readily fulfilled. This proposed framework

is flexible to achieve different realizations based on distinct SBL algorithms or probabilistic graphical models. In this work, we selected SI-STBF as the backbone to illustrate the implementation details and corresponding imaging results. The main contributions of this paper are presented as follows:

- 1) A novel framework is proposed to solve the inverse problem in source imaging, where sparse Bayesian learning is incorporated with deep neural network.
- 2) Gamma distribution is utilized for sampling the sparseness encoding parameters in training data.
- 3) Adaptive weighted mean square error (MSE) is presented as the loss function to measure the discrepancy between the estimated parameters and ground truth.

The rest of this paper is organized as follows. In Section II, we derive the novel framework for inverse problem and illustrate the implementation in detail. Meanwhile, we also expound the experimental protocol and performance metrics. Section III demonstrates the source imaging results on both simulated and real E/MEG data. Some discussion and conclusions are illustrated in Section IV and V respectively.

For notation, upper and lower case bold face letters denote matrices and vectors respectively. Scalars are represented by roman lowercase variables.  $\|\cdot\|_{\mathcal{F}}$  denotes the Frobenius norm of a matrix and  $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes a Gaussian distribution of  $\mathbf{x}$  with the mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .

## II. METHODS

### A. Probabilistic Generative Model

With the usage of Maxwell's equation [3], E/MEG recordings can be approximately expressed as the linear mapping of brain current sources as:

$$\mathbf{B} = \mathbf{L}\mathbf{S} + \boldsymbol{\varepsilon}, \quad (1)$$

where  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_T] \in \mathbb{R}^{P \times T}$  is the E/MEG measurements with  $P$  sensors and  $T$  time points,  $\mathbf{b}_t$  represents a snapshot of measurements at time point  $t$ .  $\mathbf{S} \in \mathbb{R}^{D \times T}$  indicates the sources and  $D$  is the number of discretized voxels over the cortical mesh.  $\mathbf{L} \in \mathbb{R}^{P \times D}$  is the lead field matrix whose columns reflect the sensors response induced by the unit current sources. The orientation of each source is restricted perpendicular to the cortical surface.  $\boldsymbol{\varepsilon} \in \mathbb{R}^{P \times T}$  denotes the measurement noise with independent and identical Gaussian distribution  $\mathcal{N}(\boldsymbol{\varepsilon}_t | \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$  at each time point. Because  $D$  always enormously exceeds  $P$ , it is inherently ill-posed to estimate brain sources from E/MEG measurements with the known lead field matrix. Therefore, prior knowledge is of great importance to acquire unique meaningful estimate in the context of neurophysiology.

Because of the convenience to integrate prior assumptions, probabilistic graphical model is widely used for the inverse problem. For instance, in SI-STBF, hierarchical structure is constructed between multiple latent variables, which are respectively imposed different prior distributions. As shown in Fig. 1, sources  $\mathbf{S}$  are decomposed into the product of three ingredients including spatial basis functions  $\mathbf{A}$ , temporal basis functions  $\boldsymbol{\Phi}$ , and coefficients  $\boldsymbol{\Theta}$ :

$$\mathbf{S} = \mathbf{A}\boldsymbol{\Theta}\boldsymbol{\Phi}, \quad (2)$$

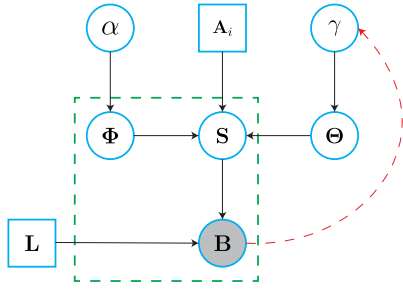


Fig. 1. Probabilistic graphical model of SI-STBF. The variables inside green dashed square are time series while the ones outside the square are time-invariant. The shaded node indicates the observed variable. The square-node variables are known and the circle-node ones are unknown. Directed arrows between nodes indicate probabilistic dependencies. The red dashed arrow represents the mapping constructed by deep neural network beyond the scope of original SI-STBF.

where  $\Phi = [\phi_1, \phi_2, \dots, \phi_K]^T$ ,  $A = [A_{[1]}, \dots, A_{[d_y]}]$ ,  $\Theta = [\Theta_{[1]}^T, \dots, \Theta_{[d_y]}^T]^T$ .  $\phi_k \in \mathbb{R}^T$  indicates the  $k$ th TBF with  $T$  time points. Spatial basis functions  $A$  are constructed based on a data driven parcelization [20] of the cortical mesh with each  $A_{[i]} \in \mathbb{R}^{D \times r_i}$  derived from one parcel. The parcelization is implemented based on Multivariate Source Pre-localization (MSP) [21], which estimates the contribution of each dipole to the measurements. Coefficients  $\Theta$  build up the connection between spatial and temporal basis functions. Additionally,  $\Theta_{[i]} \in \mathbb{R}^{r_i \times K}$  is imposed the prior distribution with precision parameter  $\gamma_i$  as:

$$p(\Theta_{[i]}|\gamma_i) \propto \exp\left(-\frac{\gamma_i}{2} \|\Theta_{[i]}\|_{\mathcal{F}}^2\right). \quad (3)$$

Hence, the prior probability of entire  $\Theta$  can be expressed as:

$$p(\Theta|\boldsymbol{\gamma}) = \prod_{k=1}^K \mathcal{N}(\theta_k|\mathbf{0}, \Gamma^{-1}) \quad (4)$$

$$\text{diag}(\Gamma) = \begin{bmatrix} \underbrace{\gamma_1, \dots, \gamma_1}_{r_1}, \dots, \underbrace{\gamma_{d_y}, \dots, \gamma_{d_y}}_{r_{d_y}} \end{bmatrix}^T, \quad (5)$$

where  $\theta_k$  represents the  $k$ th column in  $\Theta$ . Meanwhile,  $\Phi$  is imposed the prior distribution with parameters  $\alpha$  as:

$$p(\Phi|\alpha) = \prod_{t=1}^T \mathcal{N}(\phi_t|\mathbf{0}, \text{diag}(\alpha^{-1})). \quad (6)$$

With the application of mean field assumption, the variational posterior distribution is decomposed as:

$$q(\Theta, \Phi) = q(\Phi) \prod_{k=1}^K q(\theta_k). \quad (7)$$

In VB-EM process,  $q(\theta_k)$ ,  $q(\Phi)$ ,  $\boldsymbol{\gamma}$ , and  $\alpha$  are alternately updated until the convergence of variational lower bound  $\mathcal{F}$ , also called free energy, is fulfilled. Specifically, free energy has the definition as:

$$\mathcal{F} = \langle \log p(\mathbf{B}, \Theta, \Phi) - \log q(\Theta, \Phi) \rangle_{q(\Theta, \Phi)}, \quad (8)$$

where  $\langle f(x) \rangle_{q(x)}$  denotes the expectation of  $f(x)$  under the distribution  $q(x)$ . The specific expression of free energy in

SI-STBF is presented in [12]. Additionally,  $q(\theta_k)$  is updated as follows:

$$\begin{aligned} q(\theta_k) &= \mathcal{N}(\theta_k|\bar{\theta}_k, \Sigma_{\theta_k}) \\ \bar{\theta}_k &= \Sigma_{\theta_k} \mathbf{F}^T \left( \mathbf{B}(\phi_k)_{q(\Phi)} - \mathbf{F} \sum_{j \neq k} \langle \phi_j^T \phi_k \rangle_{q(\Phi)} \bar{\theta}_j \right) \\ \Sigma_{\theta_k}^{-1} &= \mathbf{F}^T \mathbf{F} \langle \phi_k^T \phi_k \rangle_{q(\Phi)} + \Gamma, \end{aligned} \quad (9)$$

where  $\mathbf{F} = \mathbf{L}\mathbf{A}$ .

## B. A Combination of Sparse Bayesian Learning and Neural Network

It is worth noting that precision parameters  $\boldsymbol{\gamma}$  have the dominant influence on the reconstructed sources. Optimized values of  $\boldsymbol{\gamma}$  always achieve enormous difference in magnitude between ingredients and hence it leads to sparseness in coefficients  $\Theta$  owing to the formulation in (3). Moreover, if  $\boldsymbol{\gamma}$  are settled down during the optimization in VB-EM, iterative update process is tremendously compressed and the computational cost can be greatly reduced.

In this study, we introduce a mapping based on deep neural network represented as the red dashed arrow in Fig. 1, which directly reconstructs precision parameters  $\boldsymbol{\gamma}$  from measurements  $\mathbf{B}$ . Using this mapping, we are allowed to acquire the estimated sources with much fewer epochs of update in variational inference. Based on the hierarchical structure of the graphical model, through repeatedly conducting the top-down sampling procedure, we can obtain the pairs  $\{\mathbf{B}^{(n)}, \boldsymbol{\gamma}^{(n)}\}_{n=1}^N$  for network training. Specifically, above all, spatial and temporal basis functions  $\mathbf{A}$  and  $\Phi$  are predefined before the sampling procedure. Precision parameters  $\boldsymbol{\gamma}$  are drawn from a specific distribution with the same dimensions as the number of components in  $\mathbf{A}$ . Afterwards,  $\Theta$ ,  $\mathbf{S}$ , and  $\mathbf{B}$  are sequentially generated based on (4), (2), and (1).

### 1) Gamma Distribution for Precision Parameters Sampling:

As mentioned above, precision parameters  $\boldsymbol{\gamma}$  need to be sampled before other variables. Nevertheless, how to derive the training samples  $\{\boldsymbol{\gamma}^{(n)}\}_{n=1}^N$ , which maintain the innate attributes in SI-STBF, is not straightforward. Intuitively, we have the preference for  $\boldsymbol{\gamma}$  in which outliers exist with small values, because it implies the sparseness in sources  $\mathbf{S}$ .

Here, we declare gamma distribution is an appropriate option to derive  $\{\boldsymbol{\gamma}^{(n)}\}_{n=1}^N$ . In SI-STBF, two update rules, including EM based manner and convex approximation, are tractable for  $\boldsymbol{\gamma}$  optimization. The latter one leads to much higher convergence rate but they have the same stationary points theoretically [22]. Suppose  $\boldsymbol{\gamma}$  are regarded as variables whose components are imposed gamma distributions as the prior. Simultaneously, mean field assumption is adopted and parameters in variational posterior are iteratively updated to promote the free energy. These update formulas have slight difference from the EM update rule in SI-STBF. Therefore, gamma distribution is satisfactory for  $\boldsymbol{\gamma}$  sampling.

Gamma distribution is defined in terms of two parameters, called shape  $a > 0$  and rate  $b > 0$ , which respectively control the sparsity and scale. To concretely illustrate how gamma distribution is influenced by the shape parameter,

in the supplementary material, we exhibit the histograms of samples drawn from gamma distributions with distinct  $a$  and fixed  $b$ . As demonstrated, shape  $a$  dominates the sparsity of gamma distribution and outliers with small scale can be readily sampled with tiny  $a$ . Hence, relatively small values of  $a$  were used to derive  $\{\boldsymbol{\gamma}^{(n)}\}_{n=1}^N$  in subsequent experiments.

**2) Network Architectures:** In this work, we chose a multi-layer perceptron (MLP) and two different CNN-type models for validation, including EEGNet [23] and residual convolutional network used in SIFNet [16], which are both widely used in temporal signal processing. EEGNet is a compact CNN composed of three convolution blocks and one fully-connected layer. The former one contains depthwise and separable convolution modules which extracts different patterns of information from the signal. The residual network used in SIFNet incorporates a sequential of residual blocks where each one consists of two CNN layers with batch normalization to alleviate the covariance shift [24]. For above models, softmax functions in last layers for classification were eliminated to fulfill the demand in our proposed framework. In the rest of this paper, EEGNet maintains its notation and residual network is denoted as ResNet.

**3) Loss Function:** Even though MSE is popular in the field of machine learning for its effectiveness, direct use of MSE between the estimates  $\hat{\boldsymbol{\gamma}}$  and ground truth cannot meet our requirement. Because of the inherent attribute of gamma distribution,  $\boldsymbol{\gamma}$  are readily drawn with components which cover a wide range of values. To avoid reconstructing extremely enormous values,  $\log \hat{\boldsymbol{\gamma}}$  is output from the neural network and the discrepancy between  $\log \boldsymbol{\gamma}$  and  $\log \hat{\boldsymbol{\gamma}}$  is adopted in the loss function. More critically, only the information of components in  $\boldsymbol{\gamma}$  with small scale can be completely propagated to the sensor space through the hierarchical structure in SI-STBF. From measurements  $\mathbf{B}$ , it is infeasible to explicitly estimate how small  $\Theta_{[i]}$ , is when whose information is covered by other components with huge scale. Owing to the coupling between  $\gamma_i$  and  $\Theta_{[i]}$  in (3), it remains unrealistic to build up a precise mapping from  $\mathbf{B}$  to great  $\gamma_i$ .

In this work, we proposed an adaptive weighted MSE to release the restriction of explicit reconstruction of  $\boldsymbol{\gamma}$  among all components, where the weighting coefficients are related to the scale of corresponding ingredients. Specifically, the large weighting coefficient is equipped for small  $\gamma_i$  because precise estimation is desired. In addition, we need to obviate small estimate for  $\gamma_i$  with great scale, otherwise it leads to redundant or interfered imaging result. Weighting coefficients are set to different values for distinct scenarios as shown in Table I, where the threshold  $\gamma^*$  is defined as the minimum of  $\boldsymbol{\gamma}$  multiplied by  $10^4$  and  $w \gg 1$ . Eventually, the loss function has the expression as:

$$\mathcal{L} = (\log \boldsymbol{\gamma} - \log \hat{\boldsymbol{\gamma}})^\top \mathbf{W} (\log \boldsymbol{\gamma} - \log \hat{\boldsymbol{\gamma}}), \quad (10)$$

where  $\mathbf{W}$  is a diagonal matrix with the weighting coefficients as the diagonal elements.

### C. Full Algorithm Implementation

The entire algorithm is composed of the data synthesis phase, model training phase, and source estimation phase.

TABLE I  
ADAPTIVE WEIGHTING COEFFICIENTS FOR LOSS FUNCTION

weighting coefficients	$\gamma_i \geq \gamma^*$	$\gamma_i < \gamma^*$
$\hat{\gamma}_i \geq \gamma^*$	0	$w$
$\hat{\gamma}_i < \gamma^*$	1	$w$

**1) Data Synthesis Phase:** First of all, spatial basis functions  $\mathbf{A}$  is constructed based on the partition set  $\mathcal{P}$  over cortical mesh. To achieve flexible source reconstruction,  $\mathcal{P}$  consists of parcelizations  $\mathcal{P}_s$  with different clustering scales  $s$ . Each  $\mathcal{P}_s$  is derived based on random MSP scores. Temporal basis functions  $\Phi^{(n)}$  are composed of specific number of Gaussian-damped sinusoidal time courses with the definition as  $\phi(t) = \sin(2\pi ft) \exp(-(\frac{t-\tau}{\omega})^2)$ , where  $f$ ,  $\tau$ , and  $\omega$  denote the frequency, phase, and damping rate respectively, which are randomly sampled from specific intervals to promote the diversity of training data. Afterwards, precision parameters  $\boldsymbol{\gamma}^{(n)}$ , with the same dimensionality as the number of clusters in  $\mathcal{P}$ , are drawn from gamma distribution with random shape  $a$ . According to (4) and (5),  $\theta_k^{(n)}$  is sampled from the Gaussian distribution with precision matrix  $\mathbf{\Gamma}^{(n)}$  derived from  $\boldsymbol{\gamma}^{(n)}$ . Negative values in  $\Theta^{(n)}$  are clamped and complemented with zeros. Sources  $\mathbf{S}^{(n)}$  are the product of aforementioned components as in (2). Finally, depending on (1), we obtain measurements  $\mathbf{B}^{(n)}$  corrupted by the noise  $\varepsilon$  with specific signal-to-noise ratio (SNR). Additionally,  $\mathbf{B}^{(n)}$  are re-scaled by division of their absolute maximum and  $\boldsymbol{\gamma}^{(n)}$  should be normalized compatible with  $\mathbf{B}^{(n)}$ . After repeatedly conducting the above top-down sampling procedure, we acquire abundant data  $\{\mathbf{B}^{(n)}, \boldsymbol{\gamma}^{(n)}\}_{n=1}^N$  for the model training phase. The detailed procedure is encapsulated in Procedure 1.

**2) Model Training Phase:** The neural network is trained through minimizing the loss function in (10) among the training dataset  $\{\mathbf{B}^{(n)}, \boldsymbol{\gamma}^{(n)}\}_{n=1}^N$ . Stochastic gradient descent is used for parameters update within the neural network.

**3) Source Estimation Phase:** Once the training phase is accomplished, using the trained neural network, the unseen realistic E/MEG signals  $\mathbf{B}$  are converted into the estimated precision parameters  $\hat{\boldsymbol{\gamma}}$ . With the fixed  $\hat{\boldsymbol{\gamma}}$ , variational inference is substantially compressed, where the convergence of the free energy is always fulfilled around five epochs of update with neglectable computational cost compared to SI-STBF. Estimated sources  $\hat{\mathbf{S}}$  can then be obtained based on  $\hat{\mathbf{S}} = \mathbf{A} \hat{\Theta} \Phi$ , where  $\hat{\Theta}$  is the variational mean and  $\Phi$  is extracted TBFs. Detailed procedure is exhibited in Procedure 2.

**4) Parameters Setting and Computation Requirements:** The partition set  $\mathcal{P}$  consisted of three parcels  $\mathcal{P}_s$  with  $s = 1, 2, 3$  unless otherwise specified. Shape parameters in gamma distributions were drawn from uniform distribution between 0.15 and 0.25. For Gaussian-damped sinusoidal time courses,  $f$  was sampled from log-uniformly distribution between 1 and 5, which implied the preference for low frequency. In addition,  $\tau$  and  $\omega$  were respectively drawn from uniform distribution over intervals (0.2, 0.5) and (0.05, 0.1). We set  $K = 3$ . In Monte-Carlo simulations, time courses were with sampling rate of 50 Hz and partitioned into pre- and

**Procedure 1** Data Synthesis Phase**Input:**

Lead field matrix  $L$ , spatial basis functions  $A$ , dataset size  $N$ , dimensionality of precision parameters  $d_\gamma$ , number of TBFs  $K$ , and specific SNR.

**Output:**

Training dataset  $\{\mathbf{B}^{(n)}, \boldsymbol{\gamma}^{(n)}\}_{n=1}^N$ .

- 1: **for**  $n = 1$  **to**  $N$  **do**
- 2:   Sample  $a^{(n)}$  from uniform distribution;
- 3:   **for**  $i = 1$  **to**  $d_\gamma$  **do**
- 4:      $\gamma_i^{(n)} \sim \text{Gam}(a^{(n)}, 1)$ ;
- 5:   **end for**
- 6:   Sample  $f^{(n)}$  from log-uniformly distribution;
- 7:   **for**  $k = 1$  **to**  $K$  **do**
- 8:     Sample  $\tau_k^{(n)}$  and  $\omega_k^{(n)}$  from uniform distributions respectively;
- 9:      $\phi_k^{(n)(t)} = \sin(2\pi f^{(n)}t) \exp\left(-\left(\frac{t-\tau_k^{(n)}}{\omega_k^{(n)}}\right)^2\right)$ ;
- 10:    Sample  $\theta_k^{(n)}$  based on (4) and (5);
- 11:   **end for**
- 12:    $\mathbf{S}^{(n)} = \mathbf{A}\boldsymbol{\Theta}^{(n)}\boldsymbol{\Phi}^{(n)}$ ;
- 13:   Obtain  $\mathbf{B}^{(n)}$  based on (1) where Gaussian white noise is sampled with specific SNR
- 14:    $s^{(n)} = \max |\text{vec}(\mathbf{B}^{(n)})|$
- 15:    $\mathbf{B}^{(n)} = \mathbf{B}^{(n)}/s^{(n)}$ ,  $\boldsymbol{\gamma}^{(n)} = \boldsymbol{\gamma}^{(n)} \cdot (s^{(n)})^2$
- 16: **end for**

**Procedure 2** Source Estimation Phase**Input:**

Measurement  $\mathbf{B}$ , lead field matrix  $L$ , spatial basis functions  $A$ , extracted TBFs  $\Phi$ , trained network  $f(x; \theta)$ , and tolerance  $\delta$ .

**Output:**

Estimated sources  $\hat{\mathbf{S}}$ .

- 1:  $\hat{\boldsymbol{\gamma}} = \exp(f(\mathbf{B}; \boldsymbol{\theta}))$ ;
- 2: **repeat**
- 3:   update the variational posterior distribution of  $\boldsymbol{\theta}_k$  with fixed  $\hat{\boldsymbol{\gamma}}$  based on (9) for each  $k$  respectively
- 4:   compute the free energy  $\mathcal{F}$ ;
- 5:   **until** relative change of  $\mathcal{F} < \delta$
- 6:    $\hat{\mathbf{S}} = \mathbf{A}\hat{\boldsymbol{\Theta}}\boldsymbol{\Phi}$ , where  $\hat{\boldsymbol{\Theta}} = [\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \dots, \hat{\boldsymbol{\theta}}_K]$ .

post-stimulus periods with 50 time points per phase. Sources were regarded as active during the post-stimulus period. Noise was integrated with distinct values of SNR including  $-5$ ,  $0$ ,  $5$ ,  $10$  dB. The whole neural networks were implemented in Pytorch and trained on single 16 GB NVIDIA Tesla T4 GPU. Detailed configurations of the neural networks, including the architectures and parameters, were presented in the supplementary document. For the loss function in (10), the weighting coefficient was set  $w = 100$ . Adam [25] was chosen as the optimizer where learning rate was  $10^{-3}$  and other parameters for gradient rectification was set as recommended with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$ . Batch size was 256 and 100 epochs of training were adequate for convergence. Tolerance  $\delta$  was set to be  $10^{-6}$ .

**D. Evaluation**

1) *Simulation Protocol*: To validate the availability of SI-SBLNN across head models, three different cortical meshes distributed in Brainstorm [26] were used in the numerical simulations, which were derived from the default MR images of ICBM152, FsAverage, and Colin27. They were respectively downsampled with 1000, 1500, and 2000 voxels while the orientation of located dipoles was imposed perpendicular to the cortical surface. Lead field matrix was acquired via OpenMEEG [27] software package embedded in Brainstorm, based on the sensor configuration in 64-channel Neuroscan Quik-cap system with two reference electrodes removed and three layers head surface obtained from Boundary Element Method (BEM).

In order to obtain spatial extended sources for Monte-Carlo experiments, we randomly selected the seed dipoles on cortical mesh and then iteratively absorb neighbor elements into the patches until the desired area of activated region is fulfilled. To reduce the variance of performance metrics caused by diverse positions of the patches, 100 Monte-Carlo experiments were conducted for each experimental scenario such that the majority of cortical surface is covered by the generated extended patches. For the evaluation data, dipoles were activated with the time courses randomly generated with the same manner as illustrated in Procedure 1. Simulated sources were projected onto the sensor space with lead field matrix and Gaussian noise was integrated with specific SNR based on (1). Simulated measurements were normalized by division of their absolute maximum and the estimated sources required to be rescaled.

2) *Performance Metrics*: To acquire the quantitatively and comprehensively evaluation of the source imaging performance, we utilized four validation metrics for assessment. The first metric is the area under the receiver operating characteristic curve (AUC) [28], which is an effective metric for sensitivity and specificity of the detection. The large AUC value implies the algorithm is capable to detect sensitively. The next metric is the spatial dispersion (SD), which measures the spatial blurriness of the estimated sources [29], [30]. To evaluate the correctness of the localization, the distance of localization error (DLE) [29], [30] is introduced. Finally, a mean square error (MSE) between the simulated and estimated sources is utilized to evaluated the accuracy of time courses reconstruction over the whole brain [31]. For good source imaging performance, the small values of SD, DLE and MSE are desired. The detailed calculations of AUC, SD, DLE, and MSE are presented in the supplementary material.

3) *Comparison With Conventional Algorithms*: In this paper, we derive a realization of SI-SBLNN with SI-SBTF as the backbone, which plays the essential role in data synthesis and source estimation phases. Hence, it is necessary to have the proposed algorithm compared to SI-STBF in order to explore whether it maintains or even promotes the imaging performance. In addition, the algorithm was tested against the benchmark methods including SBL [7], LORETA [5], and VB-GLM [9]. In this comparison, both LORETA and VB-GLM were transformed within the empirical Bayesian framework

TABLE II

PERFORMANCE METRICS FOR SOURCE RECONSTRUCTION WITH DIFFERENT NETWORK ARCHITECTURES AND DIFFERENT SIZES OF TRAINING DATASETS. RESULTS ARE ASSESSED ACROSS 100 MONTE-CARLO SIMULATIONS AND SHOWN AS MEDIANS. THE BEST RESULTS ACROSS DIFFERENT ARCHITECTURES ARE HIGHLIGHTED IN BOLD FONT

Metrics extents (cm <sup>2</sup> )	AUC				SD(mm)				DLE(mm)				MSE				
	2	6	10	18	2	6	10	18	2	6	10	18	2	6	10	18	
D.1	ResNet-Shallow	0.9825	0.9803	<b>0.9742</b>	0.9813	20.22	<b>15.01</b>	13.52	10.06	12.87	<b>7.79</b>	<b>6.51</b>	<b>4.49</b>	0.9509	<b>0.8187</b>	0.8035	0.6753
	ResNet-Deep	0.9800	<b>0.9825</b>	0.9735	<b>0.9824</b>	21.37	16.15	13.37	<b>9.69</b>	13.16	10.00	9.26	5.09	0.9747	0.8833	0.8588	0.7672
	EEGNet-Narrow	<b>0.9850</b>	0.9797	0.9721	0.9743	23.21	19.20	15.13	11.86	13.91	10.76	7.11	4.85	0.9327	0.8574	<b>0.7576</b>	0.6788
	EEGNet-Broad	0.9800	0.9783	0.9718	0.9803	<b>20.02</b>	15.93	<b>12.74</b>	10.17	<b>11.90</b>	8.63	7.31	5.18	<b>0.9298</b>	0.8468	0.7987	<b>0.6577</b>
	MLP-90	0.9500	0.9596	0.9692	0.9489	50.40	44.89	43.43	40.59	20.33	15.22	11.98	9.81	0.9841	0.9497	0.9047	0.8829
	MLP-1000	0.9375	0.9358	0.9594	0.9330	49.74	42.37	39.42	36.84	22.67	18.44	10.78	11.45	0.9930	0.9762	0.8979	0.9008
D.3	ResNet-Shallow	<b>0.9900</b>	0.9830	0.9840	0.9862	17.66	<b>12.88</b>	10.17	8.58	9.87	<b>6.70</b>	5.27	4.29	0.9424	<b>0.7696</b>	0.7476	<b>0.6672</b>
	ResNet-Deep	<b>0.9900</b>	<b>0.9876</b>	<b>0.9854</b>	<b>0.9876</b>	17.30	13.62	<b>9.65</b>	<b>7.99</b>	<b>9.59</b>	6.81	<b>4.60</b>	<b>3.91</b>	<b>0.8874</b>	0.8120	<b>0.6141</b>	0.6957
	EEGNet-Narrow	0.9800	0.9756	0.9684	0.9733	23.10	17.56	15.29	12.20	15.21	11.22	7.71	6.46	0.9535	0.8893	0.8288	0.7154
	EEGNet-Broad	0.9863	0.9794	0.9748	0.9795	19.61	14.93	11.43	9.85	12.26	7.87	6.36	5.64	0.9211	0.8102	0.7701	0.6758
	MLP-90	0.9600	0.9653	0.9621	0.9597	50.90	45.10	43.41	40.09	20.30	16.16	13.76	8.60	0.9832	0.9492	0.9224	0.8782
	MLP-1000	0.9600	0.9688	0.9567	0.9568	37.67	31.78	29.77	23.32	19.35	15.10	12.91	10.40	0.9817	0.9501	0.9137	0.7974
D.6	ResNet-Shallow	<b>0.9900</b>	0.9844	0.9856	0.9858	<b>17.06</b>	12.85	10.17	8.15	10.22	<b>6.35</b>	4.98	3.93	0.9189	<b>0.7459</b>	<b>0.6871</b>	<b>0.6437</b>
	ResNet-Deep	<b>0.9900</b>	<b>0.9881</b>	<b>0.9874</b>	<b>0.9872</b>	17.25	<b>12.38</b>	<b>9.76</b>	<b>7.61</b>	<b>9.59</b>	7.08	<b>4.47</b>	<b>3.82</b>	<b>0.9088</b>	0.7572	0.7109	0.6473
	EEGNet-Narrow	0.9775	0.9742	0.9677	0.9751	23.65	17.90	13.86	11.31	15.12	10.47	7.63	5.36	0.9355	0.8727	0.8067	0.6939
	EEGNet-Broad	0.9825	0.9761	0.9734	0.9801	19.43	14.72	12.27	9.87	13.60	8.75	7.88	4.85	0.9232	0.8266	0.7420	0.6929
	MLP-90	0.9613	0.9644	0.9666	0.9544	49.89	44.54	41.39	40.12	21.07	15.64	12.59	9.24	0.9856	0.9521	0.9086	0.8823
	MLP-1000	0.9725	0.9731	0.9652	0.9641	33.43	28.41	24.31	20.49	16.68	13.28	9.17	7.63	0.9694	0.9229	0.8653	0.8080

with slight modification. The detailed derivation of benchmark approaches are presented in the supplementary document.

### E. Application to Real Data

To further evaluate the practical imaging performance of the proposed algorithm, two real datasets were used for validation. The first data was obtained from the OpenNeuro database. Its accession number is ds000117. The dataset is composed of simultaneous MEG/EEG recordings from 16 participants performing a simple visual recognition task from presentations of famous, unfamiliar and scrambled faces. The experimental design and multiple modalities data acquirement are expounded in [32]. Only the MEG measurements were used in our subsequent analysis, which were recorded at 1100Hz with an Elekta-Neuromag VectorView 306 system. The averaged response obtained from the subject 01 during famous face presentation epoch was downsampled at 100Hz and used for source estimation. The time window was chosen to be  $-500$  ms to  $500$  ms ( $0$  ms means stimulus onset). The noise covariance was estimated from the empty room recordings. The cortical surface was discretized with 1024 voxels. The second data is a public epilepsy dataset, which is available from the Brainstorm download page. The EEG data was recorded with sampling rate of 256 Hz from a patient suffering from focal fronto-parietal epilepsy, who was performed a left frontal tailored resection and was seizure-free with a follow-up of 5 years.<sup>1</sup> The average recordings, head model, and the lead field matrix can be acquired by following the procedures outlined in the Brainstorm tutorial. The time window was selected  $-200$  ms to  $200$  ms ( $0$  ms means the spike). Noise covariance was estimated from the recordings between 110 ms and 160 ms. The cortical surface was discretized with 1024 voxels.

## III. RESULTS

For clear visualization of the source imaging results at specific time point, the thresholded absolute value of the current

sources, which is normalized with unit maxima, is shown on the cortical surface in this paper. The thresholds are determined upon the level of background activity based on Otsu's method [20], [31], [33].

### A. Different Network Architectures and Different Sizes of Training Datasets

As a first step, we evaluated the performance of different network architectures and how they were influenced by the size of training dataset. Four networks were trained via three synthesized training datasets with different sizes. Concretely, we utilized two distinct configurations for ResNet and EEGNet respectively, which are indicated as ResNet-Shallow, ResNet-Deep, EEGNet-Narrow, and EEGNet-Broad. ResNet-Deep consists of more residual blocks than ResNet-Shallow. EEGNet-Broad shares the same architecture with EEGNet-Narrow, but the former one have more feature maps in each module. The detailed configurations including architectures and parameters are presented in supplementary document. Two MLP models with 90 and 1000 hidden units, respectively denoted as MLP-90 and MLP-1000, are also used for comparison. Three training datasets with distinct sizes are denoted as D.1, D.3, and D.6, which are composed of  $1 \times 10^4$ ,  $3 \times 10^4$ , and  $6 \times 10^4$  samples for each level of SNR. The training dataset with smaller scale is included in the one with larger scale and hence we have  $D.1 \subset D.3 \subset D.6$ . In this section, simulated sources were activated with single patch and different extents of 2, 6, 10, 18 cm<sup>2</sup>. Only the head model derived from ICBM152 was utilized. SNR was 5 dB.

First, we focus on the distinction between the MLP and CNN-type models. As demonstrated in Table II, for three training datasets, the CNN-type models consistently acquired the better results among all the validation metrics. Besides, it is easy to validate that there are enormously more learnable parameters in MLP-1000 than other compared models. Hence, compared to MLP, EEGNet and ResNet are more efficient in capturing the meaning information. Second, we investigate how the imaging performance of CNN-type models is

<sup>1</sup><http://neuroimage.usc.edu/brainstorm/Tutorials/Epilepsy>

TABLE III

PERFORMANCE METRICS FOR SOURCE RECONSTRUCTION WITH DIFFERENT CORTICES AND CORRESPONDING HEAD MODELS. RESULTS ARE ASSESSED ACROSS 100 MONTE-CARLO SIMULATIONS AND SHOWN AS MEDIANS. THE BEST RESULTS ACROSS DIFFERENT ALGORITHMS ARE HIGHLIGHTED IN BOLD FONT

Metrics extents (cm <sup>2</sup> )		AUC				SD(mm)				DLE(mm)				MSE			
		2	6	10	18	2	6	10	18	2	6	10	18	2	6	10	18
ICBM152 (1000)	SI-STBF	0.9783	0.9562	0.9578	0.9343	41.41	42.86	39.33	38.74	11.88	14.52	14.78	16.31	1.0668	1.2081	1.2513	1.5870
	SI-SBLNN	<b>0.9900</b>	<b>0.9876</b>	<b>0.9854</b>	<b>0.9876</b>	<b>17.30</b>	<b>13.62</b>	<b>9.65</b>	<b>7.99</b>	<b>9.59</b>	<b>6.81</b>	<b>4.60</b>	<b>3.91</b>	<b>0.8874</b>	<b>0.8120</b>	<b>0.6141</b>	<b>0.6957</b>
FSAverage (1500)	SI-STBF	0.9744	0.9583	0.9553	0.9387	43.69	39.99	39.17	41.66	9.39	14.50	15.91	14.33	1.0438	1.1731	1.3825	1.7911
	SI-SBLNN	<b>0.9933</b>	<b>0.9917</b>	<b>0.9906</b>	<b>0.9881</b>	<b>11.78</b>	<b>7.24</b>	<b>5.96</b>	<b>4.56</b>	<b>5.52</b>	<b>3.44</b>	<b>2.96</b>	<b>2.49</b>	<b>0.7930</b>	<b>0.6245</b>	<b>0.6370</b>	<b>0.7381</b>
Colin27 (2000)	SI-STBF	0.9700	0.9575	0.9575	0.9551	39.18	32.75	30.23	32.41	12.90	12.48	12.46	10.21	1.0310	1.0035	1.0033	1.1311
	SI-SBLNN	<b>0.9935</b>	<b>0.9929</b>	<b>0.9904</b>	<b>0.9896</b>	<b>12.38</b>	<b>6.76</b>	<b>6.15</b>	<b>4.19</b>	<b>5.90</b>	<b>2.56</b>	<b>2.92</b>	<b>2.02</b>	<b>0.7480</b>	<b>0.6424</b>	<b>0.6778</b>	<b>0.7037</b>

influenced by the size of training dataset. When the networks were trained with D.1, they acquired the comparable imaging results and no one had the superior values than others among entire metrics. As the training dataset enlarged, the difference between distinct models became clear. For D.3 and D.6, EEGNet-Broad had significantly larger AUCs and smaller SDs, DLEs, and MSEs than EEGNet-Narrow, while ResNet-Deep acquired slightly better performance than ResNet-Shallow among most metrics. Meanwhile, ResNets obtained remarkably superior imaging results than EEGNets when abundant synthesized data were used for training. It is worth noting that EEGNets were less sensitive than ResNets to the size of training dataset. Specifically, EEGNets obtained indistinguishable results with varying sizes, while ResNets acquired upgraded imaging performance as the dataset enlarged. Nevertheless, although using more training data led to better performance for ResNets, the improvement from D.3 to D.6 was empirically smaller than the one from D.1 to D.3. In the subsequent experiments, to achieve the balance between imaging results and computational cost, unless otherwise specified, we employed the network with architecture of ResNet-Deep and trained with  $3 \times 10^4$  samples for each SNR level.

### B. Different Head Models

To verify the availability of the proposed algorithm for different head models, in this section, we conducted the comparison between SI-SBLNN and SI-STBF across distinct head models derived from cortical surfaces including ICBM152, FsAverage, and Colin27. The experiments retained the settings in Section III-A. For Colin27, the parcelizations  $\mathcal{P}_s$  were generated with clustering scales of  $s = 2, 3, 4$  to avoid too many clusters over the cortical mesh, which lead to the difficulty in estimating precision parameters with huge dimensionality.

Table III depicts that, for all extents across different head models, SI-SBLNN acquired the superior performance compared to SI-STBF with larger AUCs and smaller SDs, DLEs, and MSEs. Moreover, as the spatial extents of activated region enlarged, SI-SBLNN obtained relatively stable values of AUC and decreased values in other metrics. It validated that the novel framework is available for different cortices and distinct numbers of discretized voxels over the cortical surface.

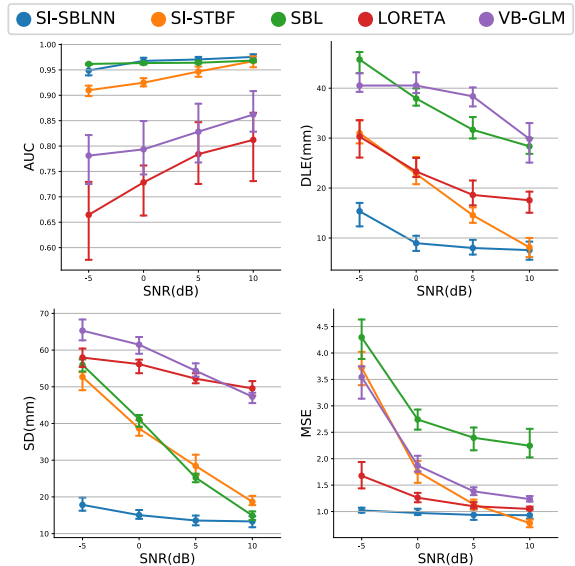


Fig. 2. Performance metrics for various levels of SNRs. The figures are results of 100 Monte-Carlo simulations and data are shown as medians surrounded by 95% confidence intervals.

### C. Influence of SNR

In this section, we tested the influence of SNR for distinct algorithms. We used the head model derived from FsAverage. Two patches of simulated sources were activated with spatial extents of approximately 6 cm<sup>2</sup> and white noise was integrated with different SNRs including  $-5$  dB,  $0$  dB,  $5$  dB, and  $10$  dB. Fig. 2 depicts the performance metrics for different algorithms with varying SNRs. As the SNR increased, all the algorithms produced increasing values of AUC and descending values of SD, DLE, and MSE implying the higher detection accuracy and smaller spatial dispersion, localization error, and time courses reconstruction error. SI-SBLNN was relatively robust to intensity of the Gaussian white noise and acquired more stable imaging result compared to conventional algorithms. Additionally, except AUCs, SI-SBLNN produced significantly better results than the benchmarks under small SNR. With increasing values of SNR, the discrepancy in multiple metrics between SI-SBLNN and benchmarks gradually descended. Specifically, with SNR of  $10$  dB, SI-STBF acquired the comparable performance with SI-SBLNN among AUC, DLE, and SD and even better results of MSE.

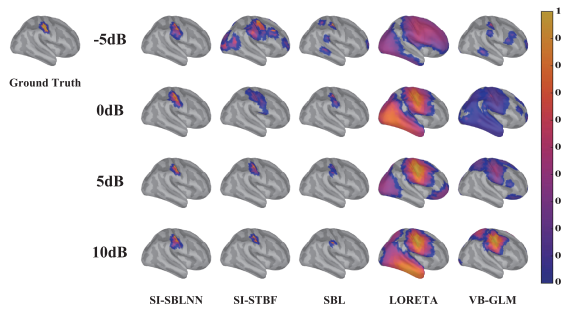


Fig. 3. Source imaging results of different methods for varying SNRs. Sources are activated with sizes of approximately  $6 \text{ cm}^2$ . The rescaled absolute values of current activities are presented at the peak of the time course.

To intuitively demonstrate the source imaging performance of different algorithms influenced by varying SNRs, one realization of randomized simulations is presented in Fig. 3. As depicted, under the Gaussian noise with SNR of  $-5 \text{ dB}$ , SI-SBLNN was capable to reconstruct the sources with spatial extents and locations basically in line with the ground truth. In the same circumstance, other algorithms obtained interfered imaging results in irrelevant regions over the cortical surface. As the noise intensity decreased, most methods acquired more plausible imaging results. With SNR = 5 and 10 dB, SI-SBLNN, SI-STBF, and SBL successfully localized the sources in correct region without spurious estimations. Notably, thanks to the usage of spatial basis functions, SI-SBLNN and SI-STBF reconstructed the sources with most accurate extents. In contrast, SBL acquired overly focal imaging results while LORETA and VB-GLM reconstructed diffuse sources covered many irrelevant regions.

#### D. Influence of the Number of Patches

In this section, we investigated how the imaging results of SI-SBLNN were influenced by the number of activated sources. We used the head model derived from FsAverage. We generated the simulated sources with various patches ranging from 1 to 4. The extent of each patch was fixed approximately  $6 \text{ cm}^2$  and SNR was 5 dB. The performance metrics for all compared methods with various numbers of patches are presented in Fig. 4. As the number of patches increased, SI-SBLNN produced degraded imaging performance with decreasing AUCs and growing DLEs, SDs and MSEs. However, it still outperformed the other algorithms among all the metrics except SBL in AUC. It is worth noting that, SI-SBLNN acquired the inverse trends in SD and DLE compared to most benchmarks, including its backbone SI-STBF. It is most probably caused by the relatively small values of shape  $a$  in the data synthesized phase. It induced the preference for source estimation with high sparsity which was empirically not appropriate for the cases where sources were activated simultaneously among so many cortical regions. To further exhibit the capability of different algorithms to reconstruct multiple sources, Fig. 5 presents the spatial maps for one realization of randomized simulations with three sources. The imaging results are depicted at the time points at which the peak of individual time courses were attained.

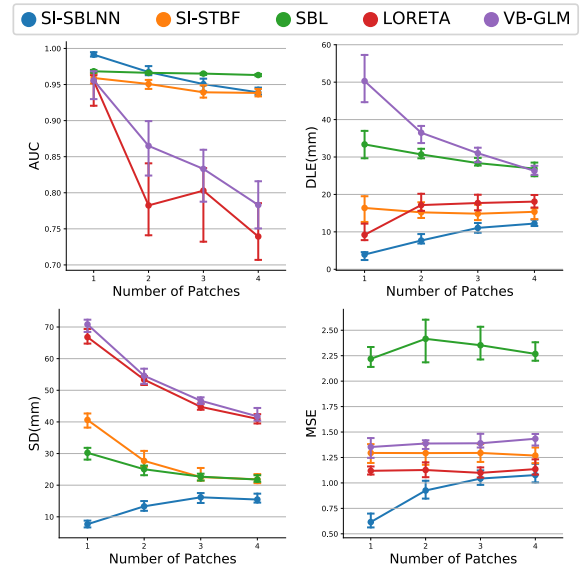


Fig. 4. Performance metrics for various numbers of patches. The figures are results of 100 Monte-Carlo simulations and data are shown as medians surrounded by 95% confidence intervals.

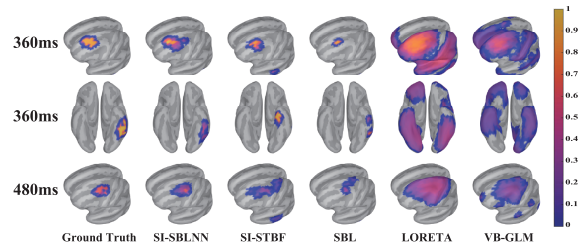


Fig. 5. Imaging results of different methods with three simultaneous sources. The sources are activated with sizes of approximately  $6 \text{ cm}^2$  and individual temporal patterns. The rescaled absolute values of current activities are presented at the peak of each source. The SNR is 5 dB.

SI-SBLNN produced the imaging results with consistent locations and extents with the ground truth at entire time points. SI-STBF and SBL were capable to localize the sources at correct regions. However, the former obtained slightly larger extents than the ground truth while the latter acquired smaller ones. Additionally, SI-STBF produced the irrelevant estimate at inferior temporal area. LORETA and VB-GLM acquired overly extended estimations with too many redundant sources.

#### E. Result of Real Data

1) *Results With Face Processing Experiment MEG Data:* In this section, we investigated the imaging result of SI-SBLNN for real MEG data, whose description is presented in Section II-E. Both the data synthesis and model training phases maintained the same configurations as expounded in Section II-C.

Fig. 6 presents the visual responses of distinct algorithms at 100 ms and 170 ms after the presentation of famous faces stimulus. As shown in the plots, SI-SBLNN successfully localized the bilateral neural activation at the occipital area at 100 ms, which is regarded for early visual processing. At 170 ms, SI-SBLNN reconstructed the sources in the vicinity of right



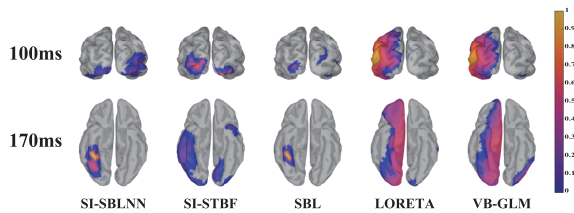


Fig. 6. Source imaging results for MEG recordings in face processing experiment. The results of different methods are demonstrated from the left to right columns. The estimates are presented as the rescaled absolute values of current activities at 100 ms and 170 ms after stimulus onset.

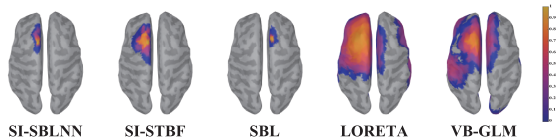


Fig. 7. Comparison of distinct algorithms on the Brainstorm epilepsy tutorial data. The rescaled absolute values of current activities are presented at the spike of recordings (i.e., 0 ms).

fusiform area concordant with the previous reports [9], [34]. SBL and SI-STBF estimated the neural activation at the similar regions over the cortex. Meanwhile, the former reconstructed sources with smaller extents than SI-SBLNN while the latter acquired spurious reconstruction at left temporal pole. Besides, LORETA and VB-GLM failed to locating the neural activation at right occipital area and obtained too diffuse estimates which covered many functionally irrelevant regions.

2) *Results With Public Epilepsy EEG Data*: In this section, we validated the imaging performance of proposed algorithm for real EEG recordings with its acquisition declaration presented in Section II-E. Both the data synthesis and model training phases maintained the same configurations as presented in Section II-C.

As shown in Fig. 7, all the methods except SBL can successfully localized the sources on left frontal region over the cortical surface. Nevertheless, both LORETA and VB-GLM reconstructed too widespread estimates including many spurious sources at functionally irrelevant areas. On the contrary, SI-SBLNN and SI-STBF produced the estimations with good spatial contiguity, which are consistent with the previous clinical report in [35].

#### IV. DISCUSSION

Because of the inherent ill-posedness of E/MEG source imaging problem, prior constraints are of great importance to acquire meaningful solution in the context of neurophysiology. It has been widely accepted by previous studies [36], [37] that cortical activations enjoy spatial coherence and local homogeneity. With the employment of Bayesian modeling, the sources are factorized into multiple ingredients which can be imposed prior knowledge with distinct attributes. Many conventional source imaging methods can be framed into this schema, such as MNE, LORETA, and VB-GLM. When hyper parameters introduced in the prior distribution are released to be updated, the algorithm is transformed

within empirical Bayesian framework, which has the preference for sparse estimations, like Champagne and SI-STBF. The Bayesian modeling based algorithm constructs the probabilistic model with explicit mathematical formulas. However, the more sophisticated model we derive, the more difficultly we can obtain the posterior distribution of latent variables. In most cases, it is even intractable to acquire the true posterior and we can merely achieve the approximate estimation via variational inference, which always has high computational cost and slight use for real-time applications. On the contrary, DNN-based methods, such as ConDvid, SIFNet, and DST-DAE, avoid the iteration and hence reduce the computation time in each single inference process. Additionally, prior assumptions are implicitly imposed through the synthesized training data, instead of constructing the probabilistic graphical model.

In this paper, we proposed a novel framework termed SI-SBLNN, which combines sparse Bayesian learning and deep neural network, and implemented an algorithm with SI-STBF as the backbone for training data synthesis and source estimation. As demonstrated in Table III, the proposed algorithm had the availability for different cortices and corresponding head models. Meanwhile, the plots in Fig. 2 depict that SI-SBLNN was robust against SNRs and acquired superior imaging performance than the benchmarks under low SNRs ( $\leq 5$  dB). Furthermore, as shown in Section III-D, SI-SBLNN was capable to duel with the cases where multiple sources were activated. On the contrary, the DNN-based method proposed in [14] was limited in single dipole reconstruction. In [18], individual models are required to be trained to acquired promising results for different scenarios. In Section III-E, we demonstrated that SI-SBLNN was capable to produce plausible imaging results for realistic E/MEG recordings, even though it was merely trained with synthesized data.

For the Bayesian learning algorithms with complicated graphical structures where latent variables are heavily correlated, variational inference is widely used to acquire the approximation of the posterior distribution, such as mean field assumption. However, the decomposition over variables can induce severe deviation from the true correlation. Besides, the error is readily accumulated and amplified through the alternate updates in VB-EM process. On the contrary, the proposed method directly extracts the relationship between measurements and essential latent variables from the synthesized training data, which in principle reduces the bias caused by the user-specified approximate posterior and the iterative update process. Therefore, compared to the backbone algorithm, it makes sense for SI-SBLNN to obtain the more accurate estimation of precision parameters and acquire the superior imaging performance in multiple scenarios as presented in Fig. 2 and 4. Moreover, in SI-SBLNN, the computational cost of variational inference are tremendously compressed using the trained neural network. As shown in Table S5 of the supplementary material, across Monte-Carlo simulations, the average inference time for SI-SBLNN was around one-thirtieth that of SI-STBF.

Compared to the purely DNN-based methods, the proposed algorithm has the main distinctions in following aspects. First, rather than generating the training data in heuristic manner, we employed SI-STBF as the backbone, which consists of probabilistic graphical model with explicit mathematical expressions, to derive the training data by conducting the top-down sampling procedure. Second, rather than making estimations in the source space with huge dimensionality like SIFNet and DST-DAE, the proposed algorithm reconstructs the precision parameters with enormously less dimensions thanks to the matrix factorization in SI-STBF, which implies the spatial coherence of cortical activations. It substantially alleviates the difficulty in network training. In the supplementary document, we compared the performance for different head models between SI-SBLNN and SIFNet. As shown in Table S4, SIFNet acquired the superior results for ICBM152 and FsAverage, which consist of relatively small number of dipoles. For the head model with most discretized voxels over the cortex, the source detection sensitivity and localization accuracy of SIFNet significantly degraded. In contrast, SI-SBLNN was capable to obtain consistently satisfactory results among different head models with tolerable additional computational cost. As depicted in Table S5, the inference time of SI-SBLNN and SIFNet are essentially of comparable magnitude.

In this paper, rather than deriving a specific source imaging algorithm which outperforms the benchmarks, the main contribution is that we proposed a flexible framework, which combines the sparse Bayesian learning and deep neural network and is allowed to achieve different realizations with slight modification in many aspects. Specifically, the proposed strategy can be applied to other algorithms within empirical Bayesian framework in addition to SI-STBF. For instance, we can choose BESTIES [38] as backbone and the mappings are formulated from measurements to both precision parameters and the learnable scale in spatial smoothing kernel. Additionally, a variety of network architectures can be used to construct the mapping. Graphical convolutional neural network [39] is a reasonable option because of its inherent compatibility to face connectivity over the cortical mesh. Furthermore, we can build up the network by unrolling the iteration in variational inference [40]. The network is composed of stacked modules with the same architectures but different learnable parameters, where the architecture is elaborately designed to match the update procedure in single epoch. LSTM layer can be employed to get rid of the constraint of fixed-size temporal dimensions [14].

To the best of our knowledge, current DNN-based algorithms, including the proposed algorithm in this study, have to train individual neural networks for distinct head models. It restricts the clinical applications owing to the long training time for subject-specific models. In the future, we will concentrate on the design of network architecture and collaborative algorithm towards the generalizability across different head models, which eliminate the network training phase for new subjects and sensor configurations. It will significantly improve the practicality of DNN-based algorithm for source imaging.

## V. CONCLUSION

The current study presented a novel E/MEG source imaging framework termed SI-SBLNN, in which the neural network is incorporated into a source imaging algorithm based on sparse Bayesian learning through constructing the straightforward mapping from measurements to latent sparseness encoding parameters. We achieved an implementation with SI-STBF as the backbone and employed MLP, EEGNet and ResNet to formulate the mappings. In numerical simulations, MLP performed poorly among the comparison and ResNet obtained better imaging performance than EEGNet with abundant training data. The proposed algorithm validated its availability for different head models and robustness against distinct intensities of noise. Meanwhile, it acquired superior performance compared to SI-STBF and several benchmarks in a variety of source configurations. Additionally, in real data experiments, SI-SBLNN produced the concordant results with the prior studies when it was trained with synthesized data merely.

## REFERENCES

- [1] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 14–30, Nov. 2001.
- [2] W. Wu, S. Nagarajan, and Z. Chen, "Bayesian machine learning: EEG/MEG signal processing measurements," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 14–36, Jan. 2016.
- [3] H. Hallez et al., "Review on solving the forward problem in EEG source analysis," *J. Neuroeng. Rehabil.*, vol. 4, no. 1, p. 46, 2007.
- [4] M. S. Hämäläinen and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain: Minimum norm estimates," *Med. Biol. Eng. Comput.*, vol. 32, no. 1, pp. 35–42, Jan. 1994.
- [5] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, no. 1, pp. 49–65, Oct. 1994.
- [6] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details," *Methods Find Exp. Clin. Pharmacol.*, vol. 24, pp. 5–12, Jan. 2002.
- [7] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947–966, Feb. 2009.
- [8] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, no. 1, pp. 641–655, Jan. 2010.
- [9] N. J. Trujillo-Barreto, E. Aubert-Vázquez, and W. D. Penny, "Bayesian M/EEG source reconstruction with spatio-temporal priors," *NeuroImage*, vol. 39, no. 1, pp. 318–335, Jan. 2008.
- [10] W. Ou, M. S. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, Feb. 2009.
- [11] S. S. Nagarajan, H. T. Attias, K. E. Hild, and K. Sekihara, "A probabilistic algorithm for robust interference suppression in bioelectromagnetic sensor data," *Statist. Med.*, vol. 26, no. 21, pp. 3886–3910, Sep. 2007.
- [12] K. Liu et al., "Bayesian electromagnetic spatio-temporal imaging of extended sources based on matrix factorization," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2457–2469, Sep. 2019.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [14] S. Cui et al., "EEG source localization using spatio-temporal neural network," *China Commun.*, vol. 16, no. 7, pp. 131–143, Jul. 2019.
- [15] L. Hecker, R. Rupperecht, L. T. Van Elst, and J. Kornmeier, "ConvDip: A convolutional neural network for better EEG source imaging," *Frontiers Neurosci.*, vol. 15, Jun. 2021, Art. no. 569918.
- [16] R. Sun, A. Sohrabpour, S. Ye, and B. He, "SIFNet: Electromagnetic source imaging framework using deep neural networks," *bioRxiv*, pp. 1–29, May 2020. [Online]. Available: <https://www.biorxiv.org/content/early/2020/05/13/2020.05.11.089185>

- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [18] G. Huang et al., "Electromagnetic source imaging via a data-synthesis-based convolutional encoder–decoder network," 2020, *arXiv:2010.12876*.
- [19] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 1–38, 2010.
- [20] R. A. Chowdhury, J. M. Lina, E. Kobayashi, and C. Grova, "MEG source localization of spatially extended generators of epileptic activity: Comparing entropic and hierarchical Bayesian approaches," *PLoS ONE*, vol. 8, no. 2, Feb. 2013, Art. no. e55969.
- [21] J. Mattout, M. Pelegriniissac, L. Garnero, and H. Benali, "Multivariate source prelocalization (MSP): Use of functionally informed basis functions for better conditioning the MEG inverse problem," *NeuroImage*, vol. 26, no. 2, pp. 356–373, Jun. 2005.
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [23] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
- [24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [26] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–13, Jan. 2011.
- [27] A. Gramfort, T. Papadopoulos, E. Olivi, and M. Clerc, "OpenMEEG: Opensource software for quasistatic bioelectromagnetics," *BioMed. Eng. OnLine*, vol. 9, p. 45, Sep. 2010.
- [28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [29] W.-T. Chang, A. Nummenmaa, J.-C. Hsieh, and F.-H. Lin, "Spatially sparse source cluster modeling by compressive neuromagnetic tomography," *NeuroImage*, vol. 53, no. 1, pp. 146–160, Oct. 2010.
- [30] M. Zhu, W. Zhang, D. L. Dickens, and L. Ding, "Reconstructing spatially extended brain sources via enforcing multiple transform sparseness," *NeuroImage*, vol. 86, pp. 280–293, Feb. 2014.
- [31] C. Grova, J. Daunizeau, J.-M. Lina, C. Bénar, H. Benali, and J. Gotman, "Evaluation of EEG localization methods using realistic simulations of interictal spikes," *NeuroImage*, vol. 29, no. 3, pp. 734–753, Feb. 2006.
- [32] D. G. Wakeman and R. N. Henson, "A multi-subject, multi-modal human neuroimaging dataset," *Sci. Data*, vol. 2, no. 1, Jan. 2015, Art. no. 150001.
- [33] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [34] J. P. Owen, D. P. Wipf, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data," *NeuroImage*, vol. 60, no. 1, pp. 305–323, Mar. 2012.
- [35] M. Dümpelmann, T. Ball, and A. Schulze-Bonhage, "SLORETA allows reliable distributed source reconstruction based on subdural strip and grid recordings," *Hum. Brain Mapping*, vol. 33, no. 5, pp. 1172–1188, May 2012.
- [36] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413–497, Apr. 1993.
- [37] A. Destexhe, D. Contreras, and M. Steriade, "Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states," *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 19, no. 11, pp. 4595–4608, 1999.
- [38] K. Liu, Z. L. Yu, W. Wu, Z. Gu, Y. Li, and S. Nagarajan, "Bayesian electromagnetic spatio-temporal imaging of extended sources with Markov Random field and temporal basis expansion," *NeuroImage*, vol. 139, pp. 385–404, Oct. 2016.
- [39] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [40] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 141–151, Jan. 2020.