Electromagnetic Source Imaging via a Data-Synthesis-Based Convolutional Encoder–Decoder Network

Gexin Huang, Ke Liu, Jiawen Liang, Chang Cai, Zheng Hui Gu, Member, IEEE, Feifei Qi, Yuanqing Li, Fellow, IEEE, Zhu Liang Yu, Member, IEEE, and Wei Wu, Senior Member, IEEE

Abstract—Electromagnetic source imaging (ESI) requires solving a highly ill-posed inverse problem. To seek a unique solution, traditional ESI methods impose various forms of priors that may not accurately reflect the actual source properties, which may hinder their broad applications. To overcome this limitation, in this article, a novel data-synthesized spatiotemporally convolutional encoder–decoder network (DST-CedNet) method is proposed for ESI. The DST-CedNet recasts ESI as a machine learning problem, where discriminative learning and latent-space representations are introduced in CedNet to learn a robust mapping from the measured electroencephalography/magnetoencephalography (E/MEG) signals to the brain activity. In particular, by incorporating prior knowledge regarding dynamical brain activities, a novel data synthesis strategy is devised to generate large-scale samples for effectively training CedNet. This stands in contrast to traditional ESI methods where the prior information is often enforced via constraints primarily aimed for mathematical convenience. Extensive numerical experiments as well as analysis of a real MEG and epilepsy EEG dataset demonstrate that the DST-CedNet outperforms several state-of-the-art ESI methods in robustly estimating source signals under a variety of source configurations.

Index Terms—Convolutional encoder–decoder network (CedNet), data synthesis, deep learning, electromagnetic source imaging (ESI).

I. INTRODUCTION

THE forward problem of electroencephalography/magnetoencephalography (E/MEG) involves determining the external surface potentials or magnetic fields from a given configuration of neuronal current activities. In contrast, reconstructing sources of brain activities from the external measurements of electromagnetic signals is defined as the E/MEG inverse problem, also known as the electromagnetic source imaging (ESI) problem. The solution to the ESI problem provides comprehensive spatial–temporal patterns of brain activities associated with a multitude of functions, leading to broad applications in the diagnosis and prediction of brain diseases [1], [2], [3]. However, due to the limited number of channels for the external measurements compared with the number of unidentified sources, the ESI problem is highly ill-posed, i.e., there are infinite possible solutions for a given measured E/MEG signal, which poses a long-standing challenge for ESI.

To enable a unique solution, previous studies were centered on the optimization or probabilistic framework where regularization techniques or pieces of prior information are exploited to constrain the solution space. Minimum norm estimate (MNE) [4] employs the $L_2$-norm regularization to estimate spatially smooth sources. Its variant, weighted MNE (wMNE) [5], assigns additional weights to the sources to mitigate the bias toward superficial sources. Low-resolution electromagnetic tomography (LORETA) [6] is an alternative $L_2$-norm-based approach that utilizes the Laplace matrix to spatially regularize the solution. Although these methods may be suitable for imaging spatially extended sources, the $L_2$-norm regularization may lead to overly diffuse estimates for focal sources [7], [8]. To overcome this limitation, focal underdetermined system solution (FOCUSS) [9] and sparse source imaging (SSI) [10] introduce $L_p$-norm regularizers that encourage sparsity of the solution. Along this line, the $L_p$-norm iterative sparse solution (LPiSS) [11] is an iterative sparse learning algorithm based on an $L_p$ norm. Sparse Bayesian learning (SBL) approaches [12], [13], [14] cast the inverse problem under a empirical Bayesian framework where hyperparameters can be automatically determined, and sparse solutions can be obtained. In particular, Champagne [15], [16] is an SBL approach that estimates the number,
location, and time course of the sources in a principled fashion. Nevertheless, these sparse approaches are fundamentally limited in recovering the spatial extents for extended sources and may suffer from discontinuity in the reconstructed time courses under low signal-to-noise settings [17]. In addition to the smoothness and sparseness constraints, prior studies also explored the use of other constraints for source imaging. A spatiotemporal tomographic nonnegative-independent component analysis (STTONNICA) [18] decomposes source signals into temporal and spatial components via nonnegative matrix factorization. In contrast, methods, such as [19], [20], and [21], exploit the physical properties of neural activity, information about the cortical tissue structures, or other modalities with high spatial resolutions, such as magnetic resonance imaging (MRI), to obtain more accurate source imaging results.

The aforementioned approaches attempt to solve the inverse problem in a direct manner that use varying forms of prior assumptions to constrain the solution space of the optimization problem. From a machine learning perspective, since there is no labeled data involved to train the underlying models, these approaches can also be viewed as unsupervised learning methods. However, prior information regarding source signals is enforced via constraints often adopted to enable mathematical tractability yet may not accurately reflect the actual source configurations. Moreover, computational complexity may be increased considerably with a more accurate modeling analysis [22]. For the purpose of learning the mapping from the sensor to the source space, some prior studies employed artificial neural networks to solve the dipole fitting problem [23], [24], [25], [26], which assumes a small number of dipole sources are activated. Nonetheless, the problem is tractable only for one or two dipoles due to the difficulty of optimizing the neural network with more dipoles. Because of the rapid development of deep learning in recent years [27], deep neural networks are now capable of learning sophisticated mappings through large-scale training sets for distributed ESI, in which the mappings are represented by stacked nonlinear layers. Thus, deep neural networks may serve as a potential new solution for the ESI problem. Recently, the source imaging framework network (SIFNet) [28] was proposed in an attempt to convert the ESI inverse problem to a multiclassification problem, which falls in the realm of supervised learning. Experimental results show that the network can recover the spatial extents and the time courses of the sources under even low signal-to-noise settings, outperforming the existing ESI algorithms.

In this article, we propose a novel data-synthesized spatiotemporally convolutional encoder–decoder network (DST-CedNet) method that reformulates the ESI problem as a machine learning problem, where discriminative learning and latent-space representations are integrated in a convolutional encoder–decoder to learn a robust mapping from E/MEG signals to sources. Notably, by incorporating the spatiotemporal properties of dynamical brain activities, a data synthesis strategy is devised to generate the large-scale training samples for effectively training the CedNet, where no complicated mathematical modeling of prior information is required as in the traditional framework. In addition, the structure of the CedNet is tailored to E/MEG signals in that there are dedicated layers performing the spatiotemporal filtering of E/MEG signals, ensuring that the spatiotemporal patterns of E/MEG signals are sufficiently leveraged to enhance the performance.

Compared with the SIFNet, which focuses on classifying the activation state of each brain region, the DST-CedNet is able to not only locate the region of brain activity but also effectively reconstruct their time courses. Furthermore, the latent-representation learning task introduced in CedNet renders DST-CedNet robust to the noise in the E/MEG signals. Experimental results show that the network can recover the spatial extents and the time courses of the sources under even low signal-to-noise settings, outperforming the existing ESI algorithms.

The remainder of this article is organized as follows. The proposed DST-CedNet approach is introduced in Section II. Section III presents the experimental results on simulated and real datasets. Discussions about the merits and limitations of DST-CedNet are presented in Section IV. Finally, the conclusions are provided in Section V.

II. METHODOLOGY

A. Problem Formulation

The forward model of the E/MEG signals can be represented as follows:

\[ X = LS + \epsilon \]  \quad (1)

where \( X = [x_1, \ldots, x_t, \ldots, x_T] \in \mathbb{R}^{N_c \times T} \) denotes the measured E/MEG signals of \( N_c \) channels at \( T \) time points. Each column \( x_t = [x_{1t}, \ldots, x_{Nct}]^T \) represents a snapshot of the signal vector at time \( t \). \( S = [s_1, \ldots, s_t, \ldots, s_T] \in \mathbb{R}^{N_c \times T} \) denotes the source signals. Each column \( s_t = [s_{1t}, \ldots, s_{Nct}]^T \) represents the source signal vector at time \( t \). \( \epsilon \in \mathbb{R}^{N_c \times T} \) denotes the measurement noise. \( L \) is the lead field matrix, which encodes how each unit source at a certain location is related to the measured E/MEG signals. The lead field matrix can typically be obtained by modeling tissues of the human head via numerical methods, such as the finite-element model (FEM) [30] or boundary-element model (BEM) [31], based on structural MRI scans of individual subjects or an average template.

Assuming that the lead field matrix \( L \) has been obtained by solving the forward problem, the ESI problem aims to find a mapping \( P(\cdot) \) from E/MEG signals \( X \) to source signals \( S \):

\[ X \xrightarrow{P(\cdot)} S. \]  \quad (2)
due to the ill-posedness of the ESI problem when \( N_s \gg N_r \), traditional methods introduce regularization techniques to obtain a unique solution \( S^* \). This can be typically performed explicitly by adding a regularization term to the loss function under the optimization framework or implicitly priors in probabilistic modeling, both leading to the following solution:

\[
S^* = \arg \min_S D(S) + R(S) \tag{3}
\]

where \( D(S) \) measures the discrepancy between the actual and the reconstructed sensor signals based on the estimated source signals, e.g., \( D(S) = \|X - LS\|_F^2 \) \([4], [5]\), where \( \| \cdot \|_F \) denotes the Frobenius norm. \( R(S) \) is the regularization term based on the estimated source signals. For instance, \( R(S) \) can be employed to promote the sparsity (\( L_1 \) norm) or minimize the overall energy (\( L_2 \) norm) of the source signals \([19]\).

To overcome the limitation that \( R(S) \) is often devised for mathematical convenience yet may not accurately reflect the actual source configurations, in this article, we propose a novel deep learning approach based on a data synthesis strategy.

### B. Data-Synthesized Deep Learning Strategy

The DST-CedNet views the ill-posed inverse problem as a multitarget regression problem that aims to uncover the mapping \( P_\theta(\cdot) \) via a nonlinear neural network, where \( \theta \) represents the unknown weights of the neural network. To render the neural network approach feasible, sufficient source and E/MEG data samples should be provided for training the neural network. However, in vivo source data samples are hardly acquired, in practice, as direct recordings of electrical brain activities are invasive. In this article, to better reflect the spatiotemporal properties of the source signals, we inject prior information regarding the temporal realistic activities and the spatial compactness into data synthesis. These properties encoded in the synthesized data can then be effectively learned by training the neural network, of which the specific architecture is illustrated in Section II-C.

The proposed DST-CedNet is depicted in Fig. 1. Specifically, the data synthesis strategy is based on the realistic forward model derived from anatomical MRI scans and temporal basis functions (TBFs) derived from realistic scalp signals. First, the source data are decomposed into temporal and spatial components where various spatial and temporal states are generated by uniformly sampling from the respective prior space (details are provided in Section II-D). Second, the generated source signals, denoted as \( S_g \), are obtained via the product of the temporal signals and the source active states, and the corresponding E/MEG signals, denoted as \( X_g \), can be simulated by passing \( S_g \) through the forward model. As such, a large-scale dataset of source and E/MEG signals with varying states can be generated as the training set to learn \( P_\theta(\cdot) \). Under the loss function \( L_\theta(\cdot) \), the optimal parameters can be obtained as \( \theta^* = \arg \min_\theta L_\theta(X_g, S_g) \) based on the training set. Finally, the learned inverse mapping \( P_\theta(\cdot) \) can be used for ESI on unseen E/MEG signals. The implementation details are provided in Section II-D.

### C. Neural Network Design

As opposed to taking a pure discriminative approach to solving the multitarget regression problem, which may be prone to overfitting in the presence of noise in the E/MEG signals, we propose a more robust approach that integrates discriminative learning and latent-space representation under the framework of the CedNet. Inspired by the EEG decoding network \([32], [33], [34]\), through blocks with spatiotemporal stepwise convolution/deconvolution tailored to E/MEG signals, the latent-space representation learning in the CedNet is capable of extracting robust features from noisy E/MEG signals that capture the underlying dependencies and spatiotemporal structure, hence effectively regularizing the solution when combined with discriminative learning.

1) Architecture: The overall network structure is shown in Fig. 2(a). The generated E/MEG signals \( X_g \) and the corresponding generated source signals \( S_g \) are fed into the neural network as training samples, and \( \hat{X}_g \) is the noise corrupted version of \( X_g \). The encoder contains a noise corruption step and two feature extraction blocks: temporal and spatial encoding blocks. Specifically, through the temporal encoding block, the temporal feature \( \hat{X}_t \) is extracted from the noise corrupted E/MEG signals afforded by the noise layer. The spatial block then generates the spatiotemporal feature \( \hat{X}_{st} \).

In the decoder, which consists of spatial and temporal decoding blocks and a forward transformation layer, the spatial decoding block first decodes spatial information from \( \hat{X}_{st} \), yielding the source temporal features \( S_t \). The temporal block then computes the reconstructed source signal \( S_{re} \) by decoding the temporal information. Finally, the forward transformation layer calculates the reconstructed scalp signal \( \hat{X}_{re} \) as the output of the CedNet. Unlike in the traditional CedNet, here, the loss function is the weighted sum of the discrepancy between \( S_{re} \) and \( S_g \) and the discrepancy of \( \hat{X}_{re} \) and \( \hat{X}_g \). The former measures the error between the predicted and true response variables from the perspective of multitarget regression, hence falling under the umbrella of discriminative learning. The latter introduces an auxiliary learning task, which measures the adequacy of the CedNet for latent-space representation.
the size of the spatial dimension of the feature maps [35]. Collectively, the deconvolution layers upsample the size of the spatial dimension from 1 to \( N_s \) with a twofold increase per layer. Therefore, this block decodes spatial information from \( X_{st} \) to yield the source temporal features \( S_i = g^{-1}(X_{st}) \) that contain \( F_2 \) feature maps with a size of \( N_s \times K_t \).

**Temporal Decoding Block** \( h^{-1} : (\mathbb{R}^{N_s \times K_t} \times F_2 \rightarrow \mathbb{R}^{N_s \times T}) \): This block contains multiple 2-D deconvolution layers with a kernel size of \( 1 \times 6 \) and a step size of \( 1 \times 2 \) applied to the feature maps along the temporal dimension, upsampling its size from \( K_t \) to \( T \) with a twofold increase per layer. To enhance the capacity of the model, a deconvolution layer with the same kernel size but a \( 1 \times 1 \) step size is also added after each layer. We set the kernel size of the deconvolution layer to an even number (i.e., 6) to avoid checkerboard artifacts [36]. This block further decodes \( S_i \) to estimate the reconstructed source signals \( S_{re} = h^{-1}(S_i) \).

**Forward Transformation Block** \( f : (\mathbb{R}^{N_s \times T} \rightarrow \mathbb{R}^{N_s \times T}) \): This block calculates the reconstructed E/MG signals through the lead field matrix: \( X_{re} = L_{re} S_{re} \).

To accelerate network training, the exponential linear unit (elu) function [37] is employed as the activation function for all convolution and deconvolution layers except the last deconvolution layer preceding \( S_i \), which uses the linear activation function instead. Furthermore, as suggested in [38], a batch normalization (BN) layer is added between the convolution/deconvolution layer and the activation layer. This overcomes the distribution shift and the covariate drift that may occur during feature transformation especially in training with a large-scale dataset, preventing loss oscillation, weight saturation, and decelerated learning. According to our experiments, with the BN layer introduced in the neural network, the solution is not only insensitive to the parameter initialization in each layer but also able to converge quickly and smoothly.

**2) Encoder and Decoder:** The details of the neural network are illustrated in Fig. 2(b). The network is comprised of the following blocks.

**Noise Corruption** \( \mathcal{R} : (\mathbb{R}^{N \times T} \rightarrow \mathbb{R}^{N \times T}) \): This step yields \( X_n \) by corrupting \( X \) with white Gaussian noise with a specific SNR. Note that this block is only used during the training stage.

**Temporal Encoding Block** \( h : (\mathbb{R}^{N \times T} \rightarrow \mathbb{R}^{N_s \times K_t \times F_1}) \): This block contains multiple stacks of 2-D convolution layers with a convolution kernel of \( 1 \times 5 \) and a step size of \( 1 \times 2 \), where a convolution layer with the same kernel size but a \( 1 \times 1 \) step size is added after each layer to enhance the modeling capability. The kernel size allows the block to perform temporal convolution separately for each channel of E/MG signals, thus extracting the temporal features \( X_t = h(X) \) that contains \( F_1 \) temporal feature maps with a temporal size of \( K_t \).

**Spatial Encoding Block** \( g : (\mathbb{R}^{N_s \times K_t \times F_1} \rightarrow \mathbb{R}^{1 \times K_s \times C}) \): This block contains a 2-D convolution layer with a convolution kernel of \( N_c \times 1 \) and a step size of \( 1 \times 1 \). Therefore, it performs a full-channel spatial convolution at each time point and extracts the spatiotemporal feature maps \( X_{st} = g(X_t) \) with a size of \( K_s \times K_t \).

**Spatial Decoding Block** \( g^{-1} : (\mathbb{R}^{1 \times K_s \times K_t} \rightarrow \mathbb{R}^{N_s \times K_t \times F_2}) \): This block contains multiple stacks of 2-D deconvolution (also known as transposed convolution) layers with a kernel size of \( 10 \times 1 \) and a step size of \( 2 \times 1 \) applied to the feature maps along the spatial dimension. Each deconvolution layer performs an inverse convolution operation that swaps the forward and backward passes of a convolution, increasing

---

Fig. 2. Overall architecture and design of the CedNet. (a) Overall architecture. The arrows indicate the input-to-output sequence of the CedNet, and the dashed lines indicate the connected tensors to calculate the corresponding loss function. (b) Detailed design of the neural network. \( N_c \) and \( N_s \) represent the number of sensors and sources, respectively. \( T \) is the total number of time points. In our experiments, we downsampled the time points to 40.

**3) Loss Function:** In this article, two error functions quantifying the reconstruction error for the source signals and that of the E/MG signals are proposed. The loss function is the weighted sum of the two error functions

\[
\mathcal{L}_d(X_g, S_g) = \lambda_1 \mathcal{L}_X(X_g, X_{re}) + \lambda_2 \mathcal{L}_S(S_g, S_{re})
\]

(4)

where \( \lambda_1 \geq 0 \) and \( \lambda_2 \geq 0 \) are the loss weights (see the Supplementary Material for further discussions), and

\[
\mathcal{L}_X(X_g, X_{re}) = \frac{1}{N} \sum_{i=1}^{N} \left\| X^{(i)}_g - X^{(i)}_{re} \right\|_F
\]

(5)

\[
\mathcal{L}_S(S_g, S_{re}) = \frac{1}{N} \sum_{i=1}^{N} \left[ \left\| S^{(i)}_g - S^{(i)}_{re} \right\|_2 + \delta \left\| S^{(i)}_g - S^{(i)}_{re} \right\|_1 \right]
\]

(6)

where \( \| \cdot \|_{1,1} \) is the \( L_{1,1} \) matrix norm defined as \( \| A \|_{1,1} = \sum_{j=1}^{N} \sum_{i=1}^{m} |a_{ij}| \); \( N \) is the total number of training samples, and \( \delta \geq 0 \) is a manually set parameter. To facilitate precise reconstruction of the sources by the network, we use the weighted sum of the mean square error (MSE) and the mean absolute error (MAE) to assess the reconstruction error for the source signals, since the MSE is more sensitive to high amplitude errors, while the MAE is more sensitive to subtle
differences in the amplitude. However, for the E/MEG signals, due to the presence of noise in the data, excessively pursuing the reconstruction accuracy may lead to overfitting. Therefore, we use the MSE alone to measure the reconstruction error. In contrast to the standard CedNet that only attempts to learn the source signals as in (6), which are what the ESI strives to estimate, the loss function for the DST-CedNet additionally incorporates the optimization of latent-space representation for E/MEG (5), which further enhance the denoising capability for the model. Moreover, since the generated source signals encode our prior knowledge regarding the true dynamics of brain activities, our approach provides a convenient framework for leveraging prior information regarding source signals as long as it is properly reflected in the data synthesis step. Thus, a notable difference of our approach from traditional ESI methods is that the prior information regarding source signals is not enforced via mathematical constraints typically presented as the $L_1$ norm or $L_2$ norm of the source signals for mathematical convenience, which, nonetheless, may not accurately reflect the actual source configurations.

Performance comparisons of different ESI algorithms are presented in Sections III-B–III-E to demonstrate the advantages of the DST-CedNet.

D. Implementation Details

The workflow of the DST-CedNet consists of three phases: data synthesis phase, training phase, and source estimation phase.

1) Data Synthesis Phase: To ensure that the training data faithfully reflect the spatiotemporal properties of realistic brain activity, the proposed data strategy incorporates spatial and temporal information of brain activity into the generation of source signals

$$S_g = W \Phi = \sum_{k=1}^{K} w_k \phi_k$$

(7)

where $\Phi \in \mathbb{R}^{K \times T}$ is the temporal state matrix, and $W \in \mathbb{R}^{N_s \times K}$ represents the activation state matrix of brain sources with $K$ spatial components. $\phi_k$ is the $k$th temporal component generated from a set of $m$ TBFs $\Theta \in \mathbb{R}^{m \times T}$. $w_k$ is the $k$th spatial component

$$w_k = \begin{cases} 1, & s_j \in \gamma \\ 0, & \text{else} \end{cases}$$

(8)

where $s_j$ represents the $j$th vertex of the source, and $\gamma$ represents the index set of activated sources.

For the purpose of large-scale training, we generate training samples with varying spatial activation states $W$ and temporal states $\Phi$ as follows. For the $i$th generated sample, first, we sample from a uniform distribution to obtain the seed source $s_{seed}$ in the source space. Subsequently, the spatial activation state is generated as a binary vector whose activated sources are indexed as 1, wherein the index set of activated sources is obtained by gradually expanding activated sources within the seed’s neighborhood $n$ until the total area is subjected to a given value $A: \gamma_k^{(i)} = \{s_j|\text{area}(\gamma) = A, s_j \in n(s_{seed})\}$, where $A$ is randomly sampled from a uniform distribution between 0 and $A_{max}$. $A_{max}$ is set to 20, as it has been shown the area of cortex activity is typically 10–20 cm$^2$ [39]. The generated spatial states can be spatially contiguous and locally homogeneous as suggested by previous studies [40], [41].

Second, methods, such as the stimulus evoked factor analysis (SEFA) [42] or singular value decomposition (SVD) [43], can be used to estimate the TBFs from realistic E/MEG signals $X_{\text{real}}$ (see Sections III-A1 and III-G). Then, $\phi_k^{(i)}$ is generated from the subspace spanned by $\Theta$ as $\phi_k^{(i)} = \sum_{m=1}^{M} \gamma_k^{(i)} \Theta(m)$, where the weight $\gamma_k^{(i)}$ is obtained from uniform sampling.

Finally, the $i$th generated source signals $S_i^{(i)}$ can be obtained as the product of the spatial activation state $W^{(i)}$ and the temporal state $\Phi^{(i)}$, and the corresponding E/MEG signals $X_i^{(i)}$ can be generated by passing the generated source signals through the forward model with the lead field matrix $L$ derived from solving the forward problem.

2) Training Phase: The generated dataset $(S_i^{(i)}, X_i^{(i)})$ is used to train the neural network, where the SNR of the E/MEG signals determines the noise level in the noise corruption to corrupt $X_i^{(i)}$. Through minimizing the loss function until convergence, the trained neural network is capable of automatically learning the mapping from the E/MEG signals to the source signals.

3) Source Estimation Phase: In this phase, the trained neural network is applied to map unseen realistic E/MEG signals $X_{\text{real}}$ to source signals $S_i$.

III. EXPERIMENTS

In this section, we present experiments on both simulated and realistic data to demonstrate the efficacy of the DST-CedNet. Specifically, four simulated experiments were conducted to compare the performance of DST-CedNet to that of several existing ESI methods, including two $L_2$-norm methods, i.e., wMNE [5] and LORETA [6], which were proposed for the reconstruction of spatially extended source, three sparse-constrained methods, i.e., SBL [15], FOCUSS [9], and Champagne [15], [16], which were devised for estimating focal sources, and the spatiotemporally constrained method, STTONNICA [18], which decomposes source signals into spatial and temporal components that are optimized under the Bayesian information criterion (BIC). In addition, two recent deep learning algorithms are also included for comparison, namely, SIFNet [28] and ESBN [29]. Specifically, SIFNet reformulates the ESI problem as a multiclass classification problem. Although temporal information is utilized for data synthesis, the SIFNet does not reconstruct the temporal dynamics of the sources but rather focuses on the localization of sources. ESBN treats source activities as IID. Since the SIFNet is only able to determine the activation states of the sources as opposed to their temporal dynamics, we calculated all of the performance metrics (i.e., AUC, SD, and DLE) based on the peak values of the ground-truth source activities and those of the reconstructed source activities. As a result of hyperparameter optimization on the training set, the convolutional kernel size of SIFNet is $3 \times 3$, and the learning rate is $10^{-3}$. For ESBN, the learning rate is $10^{-6}$, and the loss weights are 1, $10^{-3}$, $10^{-6}$, and $10^{-6}$ for the MSE, cosine similarity, $L_1$
norm, and edge sparse norm, respectively. The other settings are identical to those used in [28] for SIFNET and [29] for ESBN. The performance of the DST-CedNet is assessed by evaluating the recovered source signals under a variety of experimental settings using multiple metrics. An additional experiment was conducted to study the impact of training sample size on the performance of the DAT-CedNet. Finally, real MEG and EEG datasets were analyzed to further illustrate the efficacy of the DST-CedNet on real-world data.

A. Experimental Description and Performance Metrics

1) Protocol of Simulation Experiments: Throughout the following experiments, the cortical surface was obtained by segmenting the white/gray matter interface of default MR images derived from MNI/ICBM152 using Brainvisa.\(^1\) The cortical surface was then tessellated with triangles, and the resulting cortical mesh was downsampling to a total of 1024 triangular elements to build the cortex space. The sensor template was configured as the 65-channel Neurosens Quicap system. The lead field matrix \(\mathbf{L}\) was computed using the BEM method. This forward modeling process was implemented using the OpenMEEG plug-in [44] in Brainstorm [45].

The time course of each patch source was simulated as a Gaussian damped sinusoidal time course: \(\sin(2\pi f) \exp^{-\left(2\pi f - \omega T\right)^2}\). Through the choice of different \(f, \omega\), and \(\tau\) values, event-related potential (ERP) signals with various temporal states could be simulated. A source active region was simulated by selecting a triangular element on the cortical mesh as a seed voxel in the regional center and enlarging the region by iteratively absorbing its neighbor elements until the required area condition was achieved. The source signals were constructed by placing the simulated ERP signals at the source active regions, and the corresponding EEG signals were calculated via the forward model. White Gaussian noise with a specific SNR was added to the simulated source and scalp signals. The SNR was defined as \(\text{SNR} = 10 \log_{10}\left(||\mathbf{S}||^2/||\varepsilon||^2\right)\). In our simulation experiments, we generated the ERP signal with 40 sampling points per trial, and the TBFs were obtained from the simulated E/MEG signals by SEFA. Without loss of generality, the number of TBFs was 4.

In simulation experiments, we considered four scenarios to assess the reconstruction performance of the DST-CedNet against aforementioned compared methods.

1) Single Active Patch Source With Varying SNRs: We fixed the source area at \(5 \pm 0.58 \text{ cm}^2\) and varied the SNR from \(-5, 0\), and 5 to 10 dB.
2) Single Active Patch Source With Varying Spatial Extents: We fixed the SNR at \(-5\) dB and varied the source area from 3, 6, and 10 to 15 cm\(^2\).
3) Two Patch Sources With Varying Spatial Patterns: We fixed the SNR at \(-5\) dB and varied the source pattern from 4 and 4, 4 and 10, and 10 and 4 to 10 and 10 cm\(^2\).
4) Two Patch Sources With Varying Correlated Sources: We fixed SNR at \(-5\) dB and the source area at 7 cm\(^2\). The coefficient correlation between the two patch source signals was varied from 0, 0.3, and 0.6 to 0.9.

In each experiment, the training set and the validation set were generated with an equal sample size of \(N = 27,000\) using our data synthesis approach. Each trial of 64-channel synthetic signals contained 40 time points and were corrupted by a randomly generated additive noise with an SNR of \(-5\) dB during the training of the CedNet. The source space consisted of 1024 vertices. Subsequently, the model was optimized on the training set with settings described in Section III-A2. The predictive performance of the trained models was then assessed on the validation set. A total of 100 Monte Carlo simulations were performed to statistically evaluate the reconstruction performance.

2) Parameter Settings: The proposed network was coded in Python with Keras library using TensorFlow Backend. Adam, which is a gradient-based optimization algorithm, was chosen as the optimizer. The impact of the hyperparameters on the predictive performance of the neural network was systematically studied on separately generated data (see Section IV-B for details). In short, the predictive performance of the neural network favors the following settings of the key hyperparameters: 1) small kernel sizes and small strides in the temporal encoding and decoding layers; 2) large kernel sizes and small strides in the spatial encoding and decoding layers; in particular, a kernel size equal to the channel number in the spatial convolution layer is preferred; and 3) increasing numbers of feature maps for the encoder and decreasing numbers of feature maps for the decoder. These observations, therefore, motivated us to set the kernel size and stride as in Fig. 2. Moreover, we set \(K_s = 5\) and \(K_t = 64\) for the spatiotemporal feature \(X_{st}\). The optimal ratio of loss weights is \(\lambda_2 : \lambda_1 = 15 : 1\). The learning rate of the Adam optimizer was set to \(10^{-4}\), and the other hyperparameters of the Adam optimizer were set to the values suggested in [46], i.e., \(\beta_1 = 0.9, \beta_2 = 0.999\), and \(\epsilon = 10^{-8}\). The batch size was set to 32 for the mini-batch stochastic gradient descent, and all the trainable parameters in the neural network were initialized with the Glorot Uniform method. The epoch number of the model training was set to 250. We used a weight decay strategy for every convolution/deconvolution layer by employing the \(L_1\) and \(L_2\) regularizers, for which the regularization penalty is \(10^{-3}\). \(\delta\) in the loss function was set to 0.1.

To further improve the performance of the DST-CedNet, we utilized a warm-up strategy [47] to dynamically fine-tune the learning rate during training. The learning rate was linearly increased to the maximum in the initial 20 epochs and linearly decreased to the minimum in the last 20 epochs.

3) Performance Metrics: In the simulation experiments, the performance of all the ESI methods was assessed with the following four metrics.

1) Area under the receiver operating characteristic curve (AUC), which evaluates the sensitivity and specificity of the source detection.
2) Relative MSE (RMSE), which evaluates the accuracy of the reconstructed time courses.
3) Distance of localization error (DLE), which measures the localization error through the average distance from

\(^1\)http://www.brainvisa.info
TABLE I
RMSE VALUES BETWEEN SIMULATED EEG SIGNALS AND RECONSTRUCTED SIGNALS. THE MEAN AND STANDARD DEVIATION ARE CALCULATED BY 100 MONTE CARLO SIMULATIONS PER SNR

| SNR (dB) | RMSE     |
|----------|----------|
| -5       | 0.781±0.036 |
| 0        | 0.519±0.029  |
| 5        | 0.269±0.015  |
| 10       | 0.132±0.005  |

Fig. 3. Denoising performance of CedNet for EEG signals with varying SNRs. (a) Noiseless EEG signals randomly sampled from each group of test samples. (b) Corresponding noisy signals. (c) Reconstructed signals that are denoised via the neural network.

4) Spatial dispersion (SD), which measures the extent to which the estimated sources are spatially blurred.

These metrics allow us to comprehensively evaluate each ESI method in terms of its ability to localize and estimate the spatial extent of the sources. In general, a desirable method is expected to achieve a relatively high AUC value and small RMSE, SD, and DLE values. Furthermore, as an indicator of the generalization performance of the trained network, we calculated the R-square ($R^2$) values between the estimated and true source signals in the validation set unused in the training stage. Source maps were thresholded using the Otsu’s method [49]. Details of each metric and the Otsu’s method are provided in the Supplementary Material.

B. Experiment I: Single Patch Source With Varying SNRs

The first experiment evaluates the reconstruction performance for source signals and EEG signals with varying SNRs when the spatial extent of the patch source is fixed at 5 cm$^2$. As shown in Fig. 3 and Table I, the DST-CedNet can accurately reconstruct the time courses of the EEG signals even with heavy noise in the data. Fig. 4 depicts the four performance metrics over 100 Monte Carlo simulations. For the AUC value, the performance of wMNE, LORETA, ESBN, and Champagne worsens as the noise level increases, while STTONNICA, SBL, FOCUSS, and SIFNet are more robust to noise. As can be seen, SIFNet achieves the highest AUCs when SNR is 5 or 10 dB, which is perhaps not surprising as it is designed to specifically solve ESI as a classification problem. Among all the methods, the DST-CedNet achieves the highest AUC value under low SNRs ($\leq 0$ dB), which remains stable across all the noise levels. For the RMSE value, all the algorithms’ performance declines with increasing noise. However, the DST-CedNet substantially outperforms the other algorithms under all the SNRs with at least five-fold improvement. Furthermore, the DST-CedNet achieves low DLE and SD values as SBL and FOCUSS do. However, SBL and FOCUSS are sparse algorithms that favor focal sources at the price of higher RMSE and lower AUC values.

To illustrate, a source imaging example is presented in Fig. 5. As can be seen, the $L_2$-norm based methods, namely, wMNE and LORETA, lead to diffuse source estimates, which aggravates as the noise increases. By contrast, the sparse methods, namely, SBL, FOCUSS, and Champagne, tend to split patch sources into focal sources, resulting in spurious sources with increasing noise. STTONNIC yields relatively compact sources. However, the amplitude of the sources is not accurate compared with the ground truth. Although ESBN and SIFNet can both accurately localize sources, errors are observable for the extent of the reconstructed sources with increasing noise. Since the DST-CedNet does not make biased assumptions regarding the sources, it can faithfully recover the spatial extent of the sources with accurate estimation of their amplitude. Therefore, both the quantitative and qualitative results demonstrate the robustness of DST-CedNet compared with the other algorithms.

C. Experiment II: Single Patch Source With Varying Spatial Extents

We next assess the performance of the DST-CedNet for a single patch source with varying spatial extents when the SNR is fixed at $-5$ dB, which is arguably a challenging scenario for any ESI algorithms to handle. The performance metrics for the compared algorithms are shown in Fig. 6. Again, the
Fig. 5. Source estimates with varying SNRs. Results are presented as the absolute values of the current activities at the peak of the simulated sources and thresholded using Otsu’s method. Areas of the sources are all 5 cm².

Fig. 6. Performance metrics with varying spatial extents. Results are assessed across 100 Monte Carlo simulations and shown as mean ± SEM. SNR is −5 dB.

Fig. 7. Source estimates with varying spatial extents. Results are presented as the absolute value of the current activities at the peak and thresholded using Otsu’s method.

DST-CedNet achieves the highest AUC value and the lowest RMSE and DLE values across all the spatial extents. Moreover, despite the low SD value for FOCUSS and SBL when the spatial extent is less than 6 cm², the performance degrades when the spatial extent increases. However, DST-CedNet’s SD remains low across all the spatial extents, demonstrating its ability to localize sources precisely with small spatial diffusion.

Fig. 7 provides a source imaging example. The observations align with those in Fig. 5 in that SBL, FOCUSS, Champagne, wMNE, LORETA, and STTONNICA are unable to accurately estimate the spatial extent due to their biased prior assumptions.

Moreover, while the performance of ESBN enhances as the spatial extent of the sources increases, the estimates by SIFNet become fragmented. By contrast, owing to the diversity of the training data comprising comprehensive spatiotemporal configurations, the DST-CedNet can recover sources with accurate estimation of their locations and spatial extent.

D. Experiment III: Two Patch Sources With Varying Spatial Patterns

In the third experiment, we further evaluate the performance of different algorithms in the case of multiple patch sources. To simplify settings, we consider two patch sources, each with a different ERP time source that is independently generated. In line with the results for the single patch source, the DST-CedNet yields the best values across AUC, RMSE, and DLE. FOCUSS and SBL achieve slightly lower SD values than the DST-CedNet for certain source areas (Fig. 8). However, these two algorithms favor focal sources at the price of sacrificing accuracy for estimating the spatial extents of the patch sources as reflected by their high DLE and RMSE values.
As illustrated in Fig. 9, the localization accuracy of all the algorithms declines when there are multiple patch sources. Notably, as the extent of the sources gets larger, SIFNet misses a large portion of the sources. Among all the compared algorithms, the DST-CedNet provides the best source recovery performance, demonstrating its capability of accurately reconstructing multiple patch sources.

E. Experiment IV: Two Patch Sources With Varying Correlations

Finally, we examine the performance of different algorithms when the activities of the two patch sources are correlated, where the areas of both sources are fixed to 7 cm², and the strength of the correlation is varied from 0 (uncorrelated) to 0.9 (highly correlated).

As shown in Fig. 10, the DST-CedNet achieves the best performance under most settings. As other deep learning methods, the DST-CedNet’s performance remains stable as the source correlation varies. By contrast, the metrics of other traditional algorithms are more susceptible to the change of the source correlation. In particular, among all the algorithms, STTONNICA shows the most dramatic fluctuations across all the metrics, with the worst performance at \( r = 0.9 \).

To further assess the accuracy of estimating the source temporal dynamics, Fig. 11 shows the time courses of the reconstructed sources by different algorithms. The respective center source from each of the two correlated sources is selected as a reference for the comparison of different methods. Among the compared algorithms, wMNE and LORETA fail to capture the temporal dynamics of the sources entirely. SBL, FOCUSS, Champagne, and STTONNICA are able to identify the peaks and troughs of one of the sources but not of the other. ESBN can recover both peaks of correlated sources, but the temporal dynamics are noisy compared with the ground truth. Note that the reconstructed waveforms by STTONNICA are relatively smooth due to its model assumption regarding the spatiotemporal decomposition that has incorporated temporal constraints. As a consequence of the temporal smoothness of the training data and the powerful learning capability of the convolutional encoder–decoder, the DST-CedNet can precisely recover temporal dynamics of both sources.

F. Impact of Training Sample Size

To study the impact of the training sample size on the performance of the DST-CedNet, we train the neural network with data of different sample sizes under the single patch source condition. We progressively increase the sample size from 1500 to 36,000, with the batch size of the stochastic gradient descent fixed at 32 and the epoch number during optimization set to 250. As shown in Fig. 12, the training and validation loss functions as well as their gap decrease, as the training sample size increases from 1500 to 27,000 where a plateau is reached, suggesting the gradual improvement of both the representation and generalization performance of the neural network. This observation is corroborated by the trend of the \( R^2 \) score shown in Table II, where the increase is small when the training sample size reaches 27,000.

G. Experiment on Real MEG Data

Having validated DST-CedNet on the simulated data, we further investigate its performance in the analysis of a real MEG dataset. The MEG median nerve data of one subject were collected from a stimulation experiment on the left and right thumbs, which include 301 and 302 trials, respectively. The experiment was performed at the Massachusetts General Hospital using the Neuromag Vectorview 306 system. There are a total of 306 recording sensors, including 102 MEG magnetometers and 204 MEG planar gradiometers, for which the sampling rate is 1793 Hz. The length of each trial is 400 ms, including a 100-ms prestimulus baseline. Here, we use the 301 left-hand trials for analysis. Moreover, following the suggested data preprocessing settings, we select the 102 magnetometer channels and data within the poststimulus 250-ms time window (starting from 0 ms) as the test signal for the DST-CedNet. The lead field matrix is calculated based on the subject’s individual anatomy via the overlapping spheres model [50].

The TBFs of the MEG data are estimated via SEFA, which represent a total of 80% of the variance of the signal.

### Table II

| Training Sample Size | \( R^2 \) |
|-----------------------|----------|
| 1,500                 | 0.253    |
| 4,500                 | 0.448    |
| 9,000                 | 0.686    |
| 18,000                | 0.856    |
| 27,000                | 0.913    |
| 36,000                | 0.922    |

2http://neuroimage.usc.edu/bst/download.php
3https://neuroimage.usc.edu/brainstorm/Tutorials/TutMindNeuromag#Preprocessing
To reduce the number of training parameters, we ultimately obtain 48 sampling points after antialiasing filtering and downsampling. The cortex space is downsampled to 1024 to facilitate the calculation of the convolution kernel. The number of spatial components $K$ is set to 3. The SNR of the noise corruption is estimated using the prestimulus baseline data. The batch size for SGD is set at 32. The Adam optimizer is selected, in which the learning rate is set to $10^{-4}$, a hyperparameter $\beta_1$ is 0.9, $\beta_2$ is 0.999, and $\epsilon$ is $10^{-8}$. The number of training epochs is set to 250.

Fig. 13 depicts the source imaging results at 35 and 85 ms, as previous studies [51] reported that there are substantial brain activities in the contralateral primary somatosensory cortex (cSI) area at 35 ms, and in the cSI area and the secondary somatosensory cortex (SII) area at 85 ms. These prior findings are served as the “ground truth” for comparing different algorithms. As can be observed, the reconstructed sources by $w$MNE and LORETA are diffused in which the activation is smeared into the motor area. SBL, FOCCUS, and Champagne yield highly focal sources. However, ipsilateral SII (iSII) and contralateral SII (cSII) are not accurately localized by SBL and Champagne, respectively; spurious sources appear in the left ventral temporal cortex for FOCCUS and Champagne at 85 ms. As for STTONNICA, although it can locate the cSI area at 35 ms as well as cSI and cSII at 85 ms, it does not accurately localize the iSII area at 85 ms but instead yields a few spurious sources in the vicinity of cSI. ESBN is able to localize both the cSI and SII areas. However, spurious sources are noticeable at 85 ms. The cSII area is entirely missed by SIFNet, which is unable to resolve temporal information, since it only aims to classify source states. Interestingly, it is noted that the superior temporal gyrus (STG) region is also activated at 85 ms in the reconstructed source map by the DST-CedNet. This finding is consistent with the previous source imaging studies [43] where STG rather than cSII is typically activated in the reconstructed maps. This may be explained by the fact that STG and cSII are structurally closely connected.

H. Experiment on Real EEG Data

To further assess DST-CedNet’s performance on real-world data, a public epilepsy EEG dataset4 is used. There are a total of 41 recording electrodes, including 29 EEG electrodes and reference electrodes, for which the sampling

---

4https://neuroimage.usc.edu/brainstorm/DatasetEpilepsy
rate is 256 Hz. The length of each trial is 1000 ms, with a 300-ms prestimulus baseline. The average spikes, head model, and the lead field matrix are computed by following the procedures outlined in the Brainstorm tutorial. The EEG data were recorded from a patient suffering from focal frontal epilepsy, who underwent a left frontal tailored resection and was seizure-free with a follow-up of five years. Invasive EEG recordings on the same patient [52] suggested that source activities occurred in the left frontocentral region as reported by a previous study with invasive EEG recordings [52].

Fig. 13. Comparison of source imaging of MEG data at 35 and 85 ms. The MEG median nerve data were collected in a simulation experiment on the thumbs. The cSI area, the SIIs area, and the STG area are circled in the first row.

Fig. 14. Comparison of source imaging of epilepsy EEG data at 0 ms. The data were collected in a patient who suffered from left frontal epilepsy. Source activities are occurred in the left frontocentral region as reported by a previous study with invasive EEG recordings [52].

In this article, we propose a novel data-synthesis-based deep learning approach to solve the ESI problem. Compared with LORETA, wMNE, SBL, FOCCUSS, STTONNICA, Champagne, ESBN, and SIFNet, the DST-CedNet achieves higher detection sensitivity and smaller error in estimating the source extent and shape in simulation experiments. In addition, the DST-CedNet yields more neurophysiologically plausible source estimation on a realistic MEG and EEG dataset. The superiority of our method is owing to how the prior information is leveraged in the algorithm. Specifically, a notable difference between DST-CedNet and traditional ESI methods is that the prior information regarding source signals is not enforced via mathematical constraints but rather the data synthesis strategy that generates a variety of states from the spatial–temporal prior space, which further guarantees the generalizability of the neural network and the validity of the simulated data. Previous studies suggested that brain activities tend to be spatially contiguous and locally homogeneous [39], [40], [41]. Our method exploits these spatial properties to generate spatial states, which enables the network to yield estimates with compact and accurate extents. Unlike traditional methods [53], [54] where TBFs are directly utilized as the temporal components of source signals under probabilistic graphic models, our method instead uses TBFs as the basis to generate the temporal components of the source signals, which is more likely able to capture the diverse dynamics of brain activities. Moreover, the DST-CedNet can robustly extract spatiotemporal features from highly noisy E/MEG signals due to the denoising capability via the latent-space representation learning task introduced in the CedNet. The latent-space representation provided by the CedNet allows DST-CedNet to extract high-level features from noisy E/MEG signals, which effectively regularizes the solution obtained from discriminative learning.

A. Generalization Capability

To investigate the generalization capability of our approaches, we examine the performance of the model trained with a specific SNR (−5 or 10 dB) on validation sets with different SNRs. Performance metrics, namely, $R^2$, AUC, RMSE, SD, and DLE, are then calculated on the validation set with an equal size as the training set. As shown in Table III, the model trained with a −5-dB SNR generalizes well to validation sets with varying SNRs in terms of $R^2$, with moderate performance degradation as the SNR in the validate sets increases. By contrast, the model trained with 10-dB SNR generalizes poorly to validation sets with a −5-dB SNR. This result demonstrates that good generalization performance can be achieved with our model trained with low SNR datasets.

B. Hyperparameter Tuning

In the proposed model, key hyperparameters are the kernel size, the stride (step size), and the number of feature maps. Here, we assess the impact of these hyperparameters on the predictive performance of the model. We first generate a training set and a validation set with equal sample size ($N = 13,500$ trials) using our data synthesis approach. Each trial of 64-channel synthetic signals contains 40 time points and is corrupted by an additive noise with an SNR of −5 dB. The source space consists of 1024 vertices. Subsequently,
we train multiple models by varying the hyperparameters using the training set, where the training batch size is 32, and the number of epochs is 250. The predictive performance of the trained models is then compared with the validation set. Since hyperparameter search is time-consuming, we only search within a confined set of hyperparameter values to save computational cost. Moreover, each time we only vary one hyperparameter with respect to a base model. The kernel size of the base model is $1 \times 1$, for the temporal encoding blocks, temporal decoding blocks, spatial encoding blocks, and spatial decoding blocks, respectively. These kernel sizes are smaller than the ones in Section II for the purpose of reducing the computational cost. Moreover, each time we only vary one hyperparameter with respect to a base model. The kernel size of the base model is $1 \times 3, 1 \times 4, 8 \times 1$, and $10 \times 1$, for the temporal encoding blocks, temporal decoding blocks, spatial encoding blocks, and spatial decoding blocks, respectively. These kernel sizes are smaller than the ones in Section II for the purpose of reducing the computational cost. Moreover, each time we only vary one hyperparameter with respect to a base model. The step size (i.e., stride) of the kernel refers to the rate of downsampling in the decoder layers and upsampling in the encoder layers, which determines the scale of the local information each feature map can capture and affects the number of the neural network layers. For instance, an input with 40 dimensions needs three layers with the stride of 2 to be reduced to an output with $40/2^3 = 5$ dimensions. We separately test three different strides, namely, two, four, and eight, for the temporal encoder and decoder and the spatial decoder. As shown in Table V, the predictive performance of the neural network deteriorates, as the step size of the temporal encoder/decoder increases. Notably, a drastic decrease of the performance occurs when the step size is 8, since a step size significantly larger than the kernel size may lead to adverse effects on training the neural network. Besides, the overfitting issue aggravates, as the step size of the spatial decoder increases.

3) Feature Map: We compare three modes for the number of feature maps: 1) the number of feature maps increases for the encoder and decreases for the decoder by twofold between adjacent layers; 2) the number of feature maps is fixed to 32 for the encoder while decreases by twofold between adjacent decoder layers; and 3) the number of feature maps decreases for both the encoder and the decoder by twofold between adjacent layers. As shown in Table V, the first mode attains the best predictive performance, arguing for an hourglass-shaped structure for the proposed neural network.

C. Time Complexity

For the real MEG dataset, the total number of trainable parameters is 509,846, the batch size of the stochastic gradient descent is 32, and the number of training epochs is 250. With the network architecture and hyperparameters fixed, the time complexity depends on the number of training samples, which is $O(n)$. The mean run times per epoch for different training sample sizes on a single 12-GB TITAN-Xp GPU are shown in Table VI. Furthermore, to measure the time it takes to make predictions using the trained neural network, we made 100 predictions on the real MEG data that contains 48 sampling points across 102 channels. The average prediction runtime for each dataset is $3.598 \pm 0.152$ s.

In the DST-CedNet, although the training process is time-consuming, once trained the model can be applied to new data to make predictions without the need of optimization as traditional methods do.

D. Limitations and Future Directions

Reconstructing a high-resolution source space entails the estimation of a large number of parameters and hyperparameters prohibitive to the available computing resource. To ease the computational burden, we reduce the dimensionality of the parameter space by sampling the source space at a relatively low spatial resolution. This limitation is expected to become less problematic with the rapid growth in computing power.
In this article, we have focused on assessing the predictive performance of DST-CedNet for single patch and two patch sources, where the number of the patch source is assumed to be known in generating the training data. It remains to be tested whether the predictive performance persists when the number of patch sources exceeds two. Since more source configurations may lead to highly similar E/MEG patterns in the case of multiple patch sources, a potential challenge will be designing an appropriate loss function sensitive to the subtle difference between these source configurations. In addition, in this study, we were unable to perform the quantitative evaluation of our method on real data, since the precise locations of the true sources are unavailable for both public datasets considered in this article. Future studies will seek opportunities to validate our method using concurrent intracranial and scalp EEG recordings with precise coordinate information.

Moreover, to the best of our knowledge, the current deep learning-based ESI methods restrict the possible numbers of patch sources in a single training set, as mixing different numbers of patch sources in the training set may lead to a decrease in the reconstruction performance. Therefore, our future work will focus on the design of better architectures and loss functions for dealing with multisources.

As a final note, although the encoding and decoding blocks in our network structure are tailored to E/MEG signals, it remains a challenge to interpret the features in the intermediate layers in terms of their neurophysiological meaning and how they impact source reconstruction, as is often the case for deep neural networks due to their black-box nature. Such model interpretability is crucial for clinical and health care practice and will be studied in our future work.

V. CONCLUSION

In conclusion, we have proposed a novel neural network method termed DST-CedNet to solve the ESI problem. Compared with traditional methods that rely on optimization constraints to estimate sources, the DST-CedNet generates large-scale training sets via a data synthesis strategy, allowing the neural network to automatically learn the sensor-to-source mapping. By leveraging prior knowledge regarding the spatiotemporal characteristics of brain activities during data generation, prior information can be readily incorporated without the need of sophisticated mathematical modeling and can conveniently be applied to new data without further optimization. Furthermore, the neural network integrates discriminative learning and latent-space representation under the framework of the convolutional encoder–decoder, where the auxiliary latent-space representation task in the CedNet designed with specific blocks is able to extract a robust spatiotemporal feature from highly noisy E/MEG signals, effectively regularizing the solution obtained from discriminative learning. Both simulation and realistic data analyses demonstrate that the DST-CedNet outperforms traditional and cutting edge deep learning ESI approaches in a range of source settings. Therefore, the method provides a useful tool for source localization in both basic science and translational studies.

ACKNOWLEDGMENT

The authors would like to thank Chen Wei and Kexin Lou at the Shenzhen Key Laboratory of Smart Healthcare Engineering, Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen, China, for assisting with the implementation of the edge sparse basis network (ESBN) algorithm.
REFERENCES

[1] B. He, L. Yang, C. Wilke, and H. Yuan, “Electrophysiological imaging of brain activity and connectivity—Challenges and opportunities,” IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1918–1931, Jul. 2011.

[2] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. G. de Peralta, “EEG source imaging,” Clin. Neurophysiol., vol. 115, no. 10, pp. 2195–2222, Oct. 2004.

[3] 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.

[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] W. Wu, S. Nagarajan, and Z. Chen, “Bayesian machine learning: A review,” NeuroImage, vol. 44, no. 3, pp. 947–966, Feb. 2009.

[6] K. Liu, Z. L. Yu, W. Wu, Z. Gu, and Y. Li, “STRAPS: A fully data-driven algorithm for simulated and real MEG data,” NeuroImage, vol. 60, no. 1, pp. 305–323, Mar. 2012.

[7] W. Ou, M. S. Hämäläinen, and P. Golland, “A distributed spatio-temporal method and its accuracy,” NeuroImage, vol. 11, no. 4, pp. 349–359, Aug. 2001.

[8] S. Sakhavi, C. Guan, and S. Yan, “Learning temporal information for brain-computer interface using convolutional neural networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[9] B. He, L. Yang, C. Wilke, and H. Yuan, “Electrophysiological imaging of brain activity and connectivity—Challenges and opportunities,” IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1918–1931, Jul. 2011.

[10] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. G. de Peralta, “EEG source imaging,” Clin. Neurophysiol., vol. 115, no. 10, pp. 2195–2222, Oct. 2004.

[11] 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.

[12] M. S. Hämäläinen and J. Sarvas, “Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data,” IEEE Trans. Biomed. Eng., vol. 36, no. 2, pp. 165–171, Feb. 1989.

[13] R. T. Schirmeister et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” 2017, arXiv:1703.05051.

[14] 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.

[15] M. Zervakis, “Review on solving the inverse problem in EEG source localization,” IEEE Trans. Biomed. Eng., vol. 58, no. 4, pp. 769–778, Mar. 2010.

[16] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, “A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation,” arXiv:1703.05051.

[17] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” 2016, arXiv:1603.07285.

[18] D. A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” 2015, arXiv:1511.07289.

[19] A. M. Dale and M. I. Sereno, “Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach,” J. Cognit. Neurosci., vol. 19, no. 11, pp. 1595–1606, Jun. 1999.

[20] B. He, L. Yang, C. Wilke, and H. Yuan, “Electrophysiological imaging of brain activity and connectivity—Challenges and opportunities,” IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1918–1931, Jul. 2011.

[21] F. A. Rodriguez, J. S. George, and B. D. Rao, “Neuroradiographic source imaging with FOCUSS: A recursive weighted minimum norm algorithm,” IEEE Trans. Biomed. Eng., vol. 45, no. 5, pp. 605–615, Sep. 1998.

[22] M. S. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Onnasmaa, “Magnetoecephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain,” Rev. Modern Phys., vol. 65, no. 2, p. 413, 1993.

[23] M. Zervakis, “Review on solving the inverse problem in EEG source localization,” IEEE Trans. Biomed. Eng., vol. 58, no. 4, pp. 769–778, Mar. 2010.

[24] F. A. Rodriguez, J. S. George, and B. D. Rao, “Neuroradiographic source imaging with FOCUSS: A recursive weighted minimum norm algorithm,” IEEE Trans. Biomed. Eng., vol. 45, no. 5, pp. 605–615, Sep. 1998.

[25] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, “A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation,” arXiv:1703.05051.

[26] D. A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” 2015, arXiv:1511.07289.

[27] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, “A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation,” arXiv:1703.05051.

[28] D. A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” 2015, arXiv:1511.07289.
Gexin Huang received the B.S. degree in automatic control from the Civil Aviation University of China, Tianjin, China, in 2016, and the M.S. degree in pattern recognition and intelligent systems from the South China University of Technology, Guangzhou, China, in 2021.

His research interests include pattern recognition, graph representation learning, and Bayesian deep learning.

Ke Liu received the B.S. degree in automatic control from Southwest University, Chongqing, China, in 2011, and the Ph.D. degree in pattern recognition and intelligent systems from the South China University of Technology, Guangzhou, China, in 2016.

He is currently an Associate Professor with the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing. His research interests include pattern recognition, Bayesian inference, and their applications in electroencephalography (EEG) data analysis.

Jiawen Liang received the B.S. degree in mathematics and applied mathematics from the South China University of Technology, Guangzhou, China, in 2017, where he is currently pursuing the Ph.D. degree in control science and engineering.

His research interests include brain signal processing, pattern recognition, and machine learning.

Chang Cai received the B.Sc. and Ph.D. degrees in computer science from the China University of Geosciences, Wuhan, China, in 2011 and 2017, respectively.

He is currently an Associate Professor with the Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan. His research interests include electromagnetic/function brain mapping, machine learning, magnetoencephalography (MEG)/electroencephalography (EEG) signal processing, high-performance computing, and neuroinformatics.

Zheng Hui Gu (Member, IEEE) received the Ph.D. degree from Nanyang Technological University, Singapore, in 2003.

From 2002 to 2008, she was with the Institute for Infocomm Research, Singapore. Since 2008, she has been with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where she is currently a Professor. Her research interests include the fields of signal processing and pattern recognition for brain–computer interfaces.

Feifei Qi received the B.S. degree in applied mathematics from Qufu Normal University, Qufu, China, in 2010, and the Ph.D. degree in pattern recognition and intelligent systems from the South China University of Technology, Guangzhou, China, in 2017.

Since 2017, she has been with the School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou. Her research interests include pattern recognition, brain signal analysis, and brain–computer interfaces.

Yuanqing Li (Fellow, IEEE) was born in Hunan, China, in 1966. He received the B.S. degree in applied mathematics from Wuhan University, Wuhan, China, in 1988, the M.S. degree in applied mathematics from South China Normal University, Guangzhou, China, in 1994, and the Ph.D. degree in control theory and applications from the South China University of Technology, Guangzhou, in 1997.

From 2002 to 2004, he was a Researcher with the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, Japan.

From 2004 to 2008, he was a Research Scientist with the Laboratory for Neural Signal Processing, Institute for Infocomm Research, Singapore. Since 1997, he has been with the South China University of Technology, where he became a Full Professor in 2004. He has authored or coauthored more than 150 scientific papers in journals and conference proceedings. His research interests include blind signal processing, sparse representation, machine learning, brain–computer interface, electroencephalography (EEG), and functional magnetic resonance imaging (fMRI) data analysis.

Zhu Liang Yu (Member, IEEE) received the B.S.E.E. and M.S.E.E. degrees in electronic engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1995 and 1998, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2003.

He joined the Center for Signal Processing, Nanyang Technological University, in 2000, as a Research Engineer and then as a Group Leader in 2001. In 2008, he joined the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where he was promoted to be a Full Professor in 2010. His research interests include signal processing, machine learning, and their applications in biomedical engineering and intelligent robotics.

Wei Wu (Senior Member, IEEE) received the Ph.D. degree in biomedical engineering from Tsinghua University, Beijing, China, in 2012.

He was a Research Assistant with the Neuroscience Statistics Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. He was a Post-Doctoral Research Fellow and an Instructor with the Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA.

He was a Full Professor with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China. He is currently a Co-Founder and the Chief Data Science Officer of Alto Neurosciences, Inc., Los Altos, CA, USA. His current research interests include computational psychiatry, neural signal processing, neural engineering, and brain stimulation.