bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ConvDip: A convolutional neural network for better M/EEG Source Imaging

Hecker, Lukas[1] (1,2,3); Rupprecht, Rebekka (4); Tebartz van Elst, Ludger (1,2); Kornmeier, Juergen (1,2,3)

1. Department of Psychiatry and Psychotherapy, Medical Center - University of Freiburg, Germany
2. Faculty of Medicine, University of Freiburg, Germany
3. Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany
4. Machine Learning Lab, Computer Science Department – University of Freiburg, Faculty of Engineering, University of Freiburg

Contents

Abstract **2**

Introduction **2**
The M/EEG and the inverse problem . . . 2

Methods **4**
Training Data – simulations of sources and EEG . . . 4
EEG simulation . . . 4
Convolutional neural network . . . 5
Implementation of eLORETA & LCMV . . . 7
Performance metrics . . . 8

Results **8**
Variance explained . . . 8
Mean localization error . . . 9
Normalized mean squared error . . . 9
Numbers of missed sources and ghost sources . . . 9
Covariates of source reconstruction success . . . 9

Discussion **12**
Overview . . . 12
General discussion of results . . . 12
Limitations . . . 14
Outlook . . . 15

Appendix **15**
A: Simulated sample . . . 15
B: Real sample . . . 15

Acknowledgements **15**

Competing interests **15**

[1]Corresponding author; lukas_hecker@web.de

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

EEG and MEG are well-established non-invasive methods in neuroscientific research and clinical diagnostics. Both methods provide a high temporal but low spatial resolution of brain activity. In order to gain insight about the spatial dynamics of the EEG one has to solve the inverse problem, which means that more than one configuration of neural sources can evoke one and the same distribution of EEG activity on the scalp. A large number of approaches have been developed in the past to handle the inverse problem by creating more accurate and reliable solutions. Artificial neural networks have been previously used successfully to find either one or two dipoles sources. These approaches, however, have never solved the inverse problem in a distributed dipole model with more than two dipole sources. We present ConvDip, a novel convolutional neural network (CNN) architecture that solves the EEG inverse problem in a distributed dipole model based on simulated EEG data in a semi-supervised approach. We show that (1) ConvDip learned to produce inverse solutions from a single time point of EEG data and (2) outperforms state-of-the-art methods (eLORETA and LCMV beamforming) on all focused performance measures. (3) It is more flexible when dealing with varying number of sources, produces less ghost sources and misses less real sources than the comparison methods. (4) It produces plausible inverse solutions for real-world EEG recordings and needs less than 40 ms for a single forward pass. Our results qualify ConvDip as an efficient and easy-to-apply novel method for source localization in EEG an MEG data, with high relevance for clinical applications, e.g. in Epileptology and real time applications.

Introduction

The M/EEG and the inverse problem

Magneto- and Electroencephalography (M/EEG) are among the most used imaging techniques for noninvasive measurements of electromagnetic brain activity. Their main advantage over other methods (e.g. functional magnetic resonance imaging; fMRI) is the high temporal resolution, which comes at the cost of a considerably low spatial resolution (Luck (2014)). As a consequence, M/EEG was mainly used to study temporal brain dynamics at fine time scales. In the past decades, however, there has been a steady growth of interest in the neural sources underlying the EEG signal (Koles (1998), Pascual-Marqui (1999), Grech et al. (2008)). The non-invasive estimations of neural generators, based on their projections to the scalp electrodes/sensors, constitutes an inverse problem. Without further constraints it is ill posed because it lacks an unique solution, since multiple configurations of neural sources can produce identical topographies of electrical activity at the scalp (see e.g. Nunez and Srinivasan (2006)).

Invasive multimodal methods (e.g. using combined EEG and electrocorticogram) help to bridge the gap between scalp recordings and neural generators and thus in handling the inverse problem in this constellation. However, access to these methods is limited and the conclusions that can be drawn are constrained by various factors such as placement of the cortical electrodes or coverage of restricted brain areas that project to the scalp electrodes. Combined EEG-fMRI has also been shown as a useful tool in providing insight into the spatiotemporal dynamics of the EEG (Ritter and Villringer (2006)). However, the costs of this technique are considerably high and the relation between electromagnetic and metabolic dynamics is yet not fully understood.

Classical approaches to handle the inverse problem in the EEG

Introducing some constraints on the solution, one can solve the inverse problem or at least reduce the number of possible solutions. One approach is the single dipole model, which is based on the assumption that the source of a signal, measured with the EEG, can be modeled by a single (or sometimes few) dipole(s). A more physiologically realistic approach is the distributed dipole model in which activity is expected to extend over larger areas of the brain (as opposed to tiny dipoles). Distributed dipole models aim to find the 3D distribution of neural activity underlying the EEG measurement Pascual-Marqui et al. (1994). A distributed dipole model proposes that sources of the EEG are better modeled using hundreds to many thousand dipoles and therefore aim to find a distributed activity that can explain the EEG data. This model can be viewed opposed to single- to few-dipole models, which assume that EEG data can be sufficiently modeled using point sources. It was e.g. shown that independent component analysis can decompose the EEG data into

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

independent sources that are 'dipolar', i.e. they can be easily modeled using single or few dipoles (Delorme et al. (2012).

A popular family of distributed dipole solutions is the *Minimum Norm Estimates* (MNE, e.g. Ioannides et al. (1990), Hämäläinen and Ilmoniemi (1994)), which aim to find the source configuration that minimizes the required power (i.e. neural energy) to generate a given potential at the scalp electrodes. Low Resolution Electromagnetic Tomography (LORETA) is a famous proponent of the MNE-family that assumes sources to be smoothly distributed (Pascual-Marqui et al. (1994)). In the most sophisticated version, exact LORETA (eLORETA, Pascual-Marqui (2007)) showed zero localization error when localizing single sources in simulated data, even in the presence of noise. Another family of inverse solutions is the beamforming approach. The linear constrained minimum variance (LCMV) beamforming approach is a spatial filter that assumes that neural sources are uncorrelated and in which portions of the data that don't belong to the signal are suppressed (Van Veen et al. (1997)).

Artificial neural networks (ANN) inverse solutions

ANN based inverse solutions follow a data-driven approach to predict source locations from EEG data. A large number of simulated EEG data samples is used to train an ANN to correctly map electrode/sensor-space signals to source-space locations. Long training periods are usually required for an ANN to generalize beyond the training set. After successful training it is capable of predicting the coordinates and orientations of a dipole correctly, given only the measurements at the scalp electrodes without further priors (Zhang et al. (1998), Abeyratne et al. (2001)).

Robert et al. (2002) reviewed the literature on ANN-based methods to solve the inverse problem of single dipoles and found that all reports achieved localization errors of less than 5%. The low computation time and robustness to measurement noise was highlighted when compared to iterative[2] approaches. However, iterative approaches were capable of achieving zero localization error in single-dipole simulations under zero-noise conditions (Hoey et al. (2000)).

In an effort to predict source locations of *two* dipoles, Yuasa et al. (1998) presented a feed-forward network with two hidden layers and achieved localization errors in a range between 3 – 9%. They found that if multiple simulated dipoles have a sufficient distance among each other the localization success of ANNs even for multiple simulated sources is equal to iterative approaches. Although ANNs as such gained popularity in the past two decades on other areas, the idea of ANN-based solutions to the inverse problem received little further attention.

Only very recently two studies about ANN-based solutions to the inverse problem have been published (to our best knowledge). Cui et al. (2019) showed that a neural network can be trained to reconstruct the position and time course of a single source using a long-short term memory recurrent neural network (LSTM) architecture (Hochreiter and Schmidhuber (1997)). LSTMs allow to not only use single time instances of (e.g. EEG-) data but instead learn from temporally lagged information.

Albeit it was only tested on single dipoles, this approach could be the next big step towards a spatiotemporal reconstruction of neural activity from scalp potentials. Tankelevich (2019) showed that a deep feed forward network can find the correct set of source clusters that produced a given scalp signal. To our knowledge this was the first ANN approach to calculate distributed dipole solutions.

In the very recent years, convolutional neural networks (CNNs) have proven to be a useful tool in a steadily increasing number of domains, like image classification (Krizhevsky et al. (2012)), sentence classification (Kim (2014)) and decoding of single trial EEG (Schirrmeister et al. (2017)) etc. CNNs are capable of learning patterns in data with a preserved temporal (e.g. time sequences) or spatial (e.g. images) structure by optimizing filter kernels that are convolved with a given input. Two famous CNNs are AlexNet (Krizhevsky et al. (2012)) and VGG16 (Simonyan and Zisserman (2014)), that won the ImageNet classification competition in 2012 and 2014, respectively.

In the current study we explored the feasibility of CNNs to solve the EEG inverse problem. Specifically, we constructed a CNN, named *ConvDip*, that is capable of detecting multiple sources using training data with biologically plausible constraints. ConvDip solves the EEG inverse problem using a distributed dipole

[2]We use the expression *iterative* for methods that iteratively estimate an inverse solution, such as MNE, eLORETA or LCMV bemformin. Although, conceptually, ANNs are iterative (as opposed to analytical), they do not require iterative processing once they are trained.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

solution in a data-driven approach. ConvDip was trained to work on single time-instances of EEG data and predicts the position of sources from potentials measured with scalp electrodes.

A key factor for the development of ConvDip is the simulation of EEG data using a forward model. This data is used to train the network to predict neural sources in the brain. It turned out that the parameters chosen for the EEG simulation are critical constrain the solution space of the neural network. For example, a neural network that was trained to reconstruct single dipoles, can only find point sources even if it will be fed by electroencephalographic data of smoothly distributed sources. This shows that the assumptions about the origin of the M/EEG will be incorporated into the implicit knowledge of the ANN. For the simulation of our EEG training data we thus incorporated the properties of brain-electric activity underlying the EEG based on physiological findings from both in-vivo and in-vitro studies (Nunez and Srinivasan (2006)).

In the present study we simulated sources of brain activity, calculated EEG scalp distributions thereof and trained ConvDip to infer the originally designed sources from scalp activity. After training we simulated a new set of sources and used them to compare the performance of ConvDip with two state of the art inverse solutions: eLORETA and LCMV beamformer. We demonstrate how data-driven approaches provide unique opportunities regarding the choice of assumed brain-electric activity, increased accuracy on multiple measures and rapid computation times suitable for real-time applications. We further provide a perspective on future developments of ANN-based inverse solutions.

Methods

Training Data – simulations of sources and EEG

We simulated 100,000 independent samples of brain activity. These samples were projected forward through the leadfield K which produces EEG-like data, similar to those one would have measured given the brain-electric distribution of activity. In the following we describe the detailed settings for these simulations.

Number of sources

For each simulated data sample the number of active sources was chosen randomly between 1 and 5. During exploratory runs it was observed that the CNN can quickly learn to reconstruct a single source, whereas higher numbers of sources were more difficult to learn. Therefore, sampling weights for the number of source between 1 and 5 were 10, 15, 20, 25 and 30%. This means that only 10% of all source simulations contained a single source, whereas 30% of all simulations contained 5 sources. In other words: The more difficult the data situation (e.g. more sources), the more training samples were given. This strategy turned out to successfully aid the training process.

Size of sources

The origin of the EEG signal stems from large scale synchronized neuronal activity and is therefore highly correlated with local field potentials (LFPs, Cohen (2017)). In order to detect a cortical signal at the scalp it needs synchronized activity of neurons within a large cortical area of approximately 6cm2 (Nunez and Srinivasan (2006)). We therefore decided to simulate sources in the range of 25 mm to 35 mm diameter (i.e. between approximate square roots of 5 and 6). No constraint concerning overlapping sources was set in the simulation, wherefore simulations could contain larger sources of two or more overlapping source patches. The orientation of dipoles was set to be orthogonal to the cortical surface, which is in accordance to the orientation of the large pyramidal cells in the human cortex.

Noise

In order to generate realistic training data we added random Gaussian noise to the EEG simulations, targeting a SNR between 6 and 9 dB. These values are based on signal-to-noise estimations from own EEG data.

EEG simulation

We based our simulation on 32 scalp electrodes according to the 10-20 system (FP1, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO9,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

O1, Oz, O2, PO10). Source activity was projected through a leadfield K, which serves as a transformation matrix between source-space data and electrode-/sensor-space data. This leadfield was generated using a three shell boundary element method (BEM) head model, which describes different conductivities for the brain, skull and scalp tissue for realistic modeling of brain activity. Our template brain was based on an average cortical surface model (Fischl et al. (1999)) as provided by the Freesurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/). Source positioning was restricted to gray matter only to add a simple biological constraint that reduces the search space. Furthermore, orientation of the resulting dipole sheets was restricted to be orthogonal to the cortex surface. The forward-projected scalp potentials were re-referenced to common average.

Convolutional neural network

I/O of ConvDip

The spatial structure of the EEG electrodes data was preserved as a low-resolution scalp map. In order to obtain a simplistic, representation of the scalp map they were re-shaped into a 7x11 matrix that kept the spatial relation between electrodes. Missing values within this scalp matrix were inpainted (Damelin and Hoang (2018)) using the scikit-image library for Python. These low-resolution scalp maps were used as input for ConvDip (see Appendix 1 left column for an example).

ConvDip architecture

The design and training of the neural network was accomplished using the Tensorflow (Abadi et al. (2016)) and Keras (Chollet et al. (2015)) libraries, which are based on Python 3. Training was partially accomplished on a Nvidia Titan V graphical processing unit (GPU).

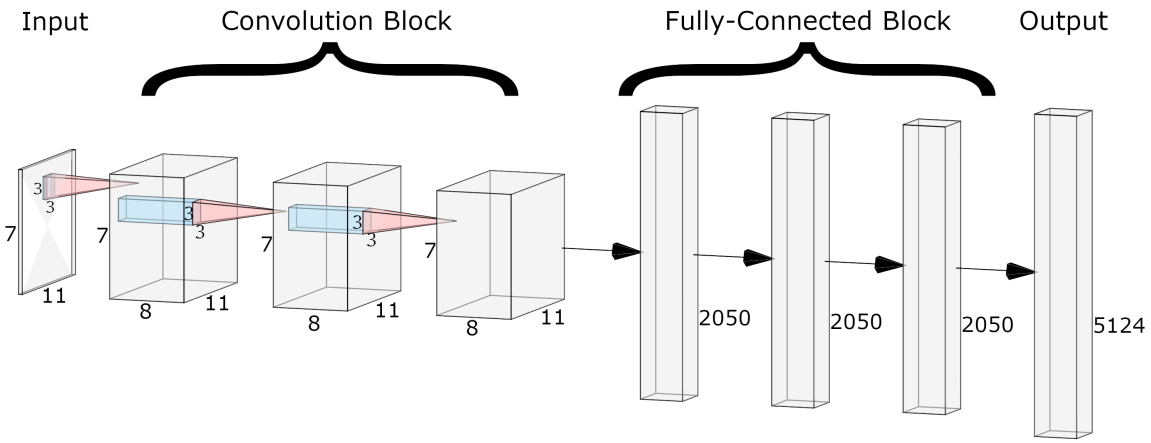

Figure 1: ConvDip architecture. The *input* layer is a 7x11 matrix, corresponding to a single low-resolution scalp map of EEG data (see Appendix A for an example). The *convolution block* consists of three convolutional layers, each of which uses 8 convolution kernels of size 3x3. The convolution block is followed by a fully-connected block consisting of three layers of 2050 neurons each. Finally, the output layer contains 5124 neurons, which correspond to the voxels in the brain. (This diagram was created using a web application at http://alexlenail.me/NN-SVG/.

The architecture of ConvDip (Fig. 1) is inspired by CNNs for image classification, in which the input layer is forwarded to a series of convolutional layers followed by a series of fully-connected layers followed by the output layer (e.g. AlexNet; Krizhevsky et al. (2012)). Convolution layers from typical CNNs are followed by pooling layers to reduce dimensionality and to promote invariance to an object's position. In ConvDip (Fig. 1), however, no pooling layers were used since spatial information would get lost or at least blurred.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

While invariance to the position of an object is desirable in image classification tasks (*what*-task), it would be detrimental in the case of source localization problems (*where*-task). Finally, we decided for a moderately deep architecture with a *convolution block* of three convolution layers followed by a *fully-connected* block with three fully-connected layers (Fig. 1).

We decided to use Rectified Linear Units (ReLUs) as activation functions after each layer (Glorot et al. (2011), Nair and Hinton (2010)). ReLUs have shown to exhibit the best performance in our preliminary tests compared to alternatives (e.g. sigmoid function). After each of the inner layers we added a batch normalization layer (Ioffe and Szegedy (2015)) in order to speed up training and to reduce internal covariance shifts. In an effort to reduce overfitting we used the method 'dropout' (Srivastava et al. (2014)) with a rate of 25 % after each fully connected layer. In the current version of ConvDip the output layer consists of 5124 neurons, which equals the number of voxels in the modeled brain. ReLU activation functions were used in the output layer. Typical CNNs use linear activation functions for regression tasks. However, since predictions are by definition non-negative in our application, ReLUs appeared to us as an appropriate alternative.

As a solver (i.e. an algorithm that backpropagates the error after each batch into the network) we used adaptive momentum estimation (ADAM, Kingma and Ba (2014)) with settings as suggested by the authors ($learningrate = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^8$).

As the loss function $\mathcal{L}_{total}$ we decided for a hybrid of *source-space loss* $\mathcal{L}_{source}$ and *sensor-space loss* $\mathcal{L}_{sensor}$. The source-space loss penalizes errors in the prediction of the sources. Sensor-space loss penalizes unplausible sources that, when projected forward to the scalp electrodes, should closely resemble the initial input. The total loss was calculated as follows:

$$\mathcal{L}_{total} = w_1 \cdot \mathcal{L}_{source} + \mathcal{L}_{sensor}, \tag{1}$$

where w_1 denotes a weighting factor to balance the two losses.

Source-space loss was implemented as the normalized and weighted mean squared error (nwMSE) between prediction and target sources.

First, target source y and predicted source $\hat{y}$ are normalized:

$$\begin{aligned} z &= \frac{y}{max(|y|)} \\ \hat{z} &= \frac{\hat{y}}{max(|\hat{y}|)} \end{aligned} \tag{2}$$

The rationale for this is to yield predictions of correct source locations while neglecting the absolute strength of sources. Then, the source-space loss was calculated as the weighted mean squared error between these normalized vectors:

$$\mathcal{L}_{source} = \frac{1}{n} \sum_{i=1}^{n} \begin{cases} (z_i - \hat{z}_i)^2, & \text{if } z_i \neq 0 \\ w_2 \cdot (z_i - \hat{z}_i)^2, & \text{if } z_i = 0, \end{cases} \tag{3}$$

where w_2 denotes a weighting factor that penalizes false-positives and n is the number of voxels in the head model.

Sensor-space loss was implemented to assure that the predictions of ConvDip represent brain-electric activity that, when projected to the scalp electrodes, resembles the spatial distribution of the original input. Therefore, this loss incentivizes plausibility of the inverse solution. Sensor-space loss was calculated as follows:

$$\mathcal{L}_{sensor} = 1 - \rho(x, \hat{x}), \tag{4}$$

where ρ denotes the Pearson correlation, x is a vector containing the input EEG data and $\hat{x}$ is the forward

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

projection of the prediction $\hat{y}$ of the sources through the leadfield K:

$$\begin{aligned} \hat{x} &= (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m), \\ \hat{x}_j &= \sum_{j=1}^{m} D_{i,j} \\ D &= \hat{y} \cdot K \end{aligned} \tag{5}$$

where m denotes the number of electrodes in the data.

We chose $w_1 = 50$ since this yielded good results in prior experiments and $w_2 = 10$ to reduce false-positive predictions (ghost sources).

Training

The simulated data was split into a training (80%) and validation (20%) set. The validation set was only used to monitor the networks' ability to generalize during training. For further evaluation of the trained ConvDip we simulated 1,000 additional samples (the 'evaluation set') for which no sample weighting for the number of sources was used. Notice, that all results presented here are based on the evaluation set only.

As mentioned above, ConvDip predicts source locations without correct global scaling (see also 2). In order to obtain the true amplitude of the sources we used Brent's method (Brent (1971)) to find a scalar $\hat{s}$ that minimizes the mean squared error between the forward-projected prediction $\hat{x}$ and the unscaled input vector x:

$$\hat{s} = \arg\min_{s} \frac{1}{m} \sum_{j=1}^{m} (\hat{x}_{scaled,j} \cdot s) - x_j \tag{6}$$

This scalar $\hat{s}$ can then be used to scale the prediction.

Implementation of eLORETA & LCMV

In order to compare our approach to eLORETA and LCMV beamformer, sources were calculated from the EEG samples in the evaluation set. Inverse solutions based on eLORETA and LCMV were carried out by the implementations in the Python library MNE (v 19). Each inverse solution was subjected to each sample of the evaluation set.

However, eLORETA and LCMV beamformer can't work properly with single time instances of EEG data. Both inverse solutions rely strongly on an estimation of the noise covariance matrix, which requires a noise interval and a signal interval. Thus, we simulated 100 trials of neural activity; each spanning 400 data points. The single trials were created by adding random Gaussian noise to the pure signal of the sample such that the averaged event-related potential contained exactly the amount of noise that was subjected to ConvDip. Each trial therefore contained a section (200 data points) of just Gaussian noise and 200 subsequent data points containing noise-contaminated signal. The first section was used for the estimation of the noise covariance matrix and followed by a signal section that was used for the actual estimation of the sources. The whole simulated epoch was re-referenced to common average and baseline corrected.

It was impossible to refrain from this noise covariance estimation since the resulting inverse solutions would have underperformed heavily if only a single data point would have been available. Noise estimation would have been impossible in this case.

For eLORETA we chose the regularization parameter λ_2 as the squared relative noise level, whereas for LCMV we chose the default regularization for the whitened data covariance at $reg = 0.05$. In both inverse algorithms we restricted solutions at fixed dipole orientations, which is in accordance with the forward model used for the simulations. For LCMV we chose to calculate the neural activity index which weighs the predicted sources by the estimated SNR (Van Veen et al. (1997)). Note, that eLORETA and LCMV beamformer can make use of substantially more information about the noise floor in the EEG than current version of ConvDip.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Performance metrics

We calculated a number of performance metrics to assess the quality of an inverse solution

1) **Variance explained**: The variance explained in the input EEG data by the forward-projected EEG of the respective inverse solution. We calculated the squared Pearson correlation coefficient between the original scalp EEG data and the forward projected prediction. This metric reflects how well the measured signal at the scalp is reflected by the forward projection of the predicted sources.

2) **Mean localization error (MLE)**: The Euclidean distance between the locations of the predicted source maximum and the target source maximum is a common measure of inverse solution accuracy. Centroids of source clusters were identified via mean shift cluster analysis Comaniciu and Meer (2002)) as implemented in the sciki-learn library for Python. Since the extent of sources is neglected with the MLE metric it is an estimate of *local similarity*. Localization errors larger than 30 mm were considered as missed sources.

3) **Normalized mean squared error (nMSE)**: Both, target and prediction, are normalized to have a maximum value of 1 before calculating the mean squared error. nMSE is a measure of *global similarity* and represents how well the prediction matches the target voxel-wise.

4) **Proportion of ghost sources**: A ghost source is a predicted source that is at least 30 mm misplaced to any true source. We calculated the proportion of samples that contain at least one ghost source. Ghost sources can be interpreted as false positives.

5) **Number of missed sources**: A source in the target volume that has no predicted source in the vicinity of 30 mm. The proportion of in which at least one source was missed was calculated. Missed sources can be interpreted as false-negatives.

In an effort to investigate the factors that lead to bad predictions, we calculated the Pearson correlation coefficient r between performance measures (nMSE and MLE) of all inverse solutions and a) the number, b) average depth of target sources, and c) the signal to noise ratio.

Results

In this section we evaluate the performance of ConvDip and compare it to two state-of-the-art inverse solutions, namely eLORETA and LCMV beamformer. Note that the evaluation set used for the performance measures was not used in the training phase of ConvDip; hence it is unknown to the model.

Variance explained

The following metric depicts the variance explained at sensor level by the predicted sources. Since the SNR at the simulated EEG samples was variable between 6 and 9 dB, the variance explained by a given inverse solution is optimally between 75 and 87 % (1 – relative noise level), depending on the actual noise level of a given sample. In Fig. 2 the explained variance is shown for each inverse solution approach and number of present sources, separately. The beamforming approach and eLORETA explain systematically less variance of the EEG signal than required, which has been discussed already (Friston et al. (2008)). ConvDip, on the other extreme, produces inverse solutions that explain more variance of the data than the signal had produced, i.e. ConvDip falsely interprets parts of the added Gaussian noise as signal.

ConvDip found inverse solutions that explained significantly more variance in the data than LCMV beamformer ($M = 0.42$, $t = 31.92$, $p < 0.001$) and eLORETA ($M = 0.39$, $t = 30.99$, $p < 0.001$). Further, eLORETA produced inverse solutions that explained slightly more variance than LCMV ($M = 0.03$, $t = 2.47$, $p < 0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

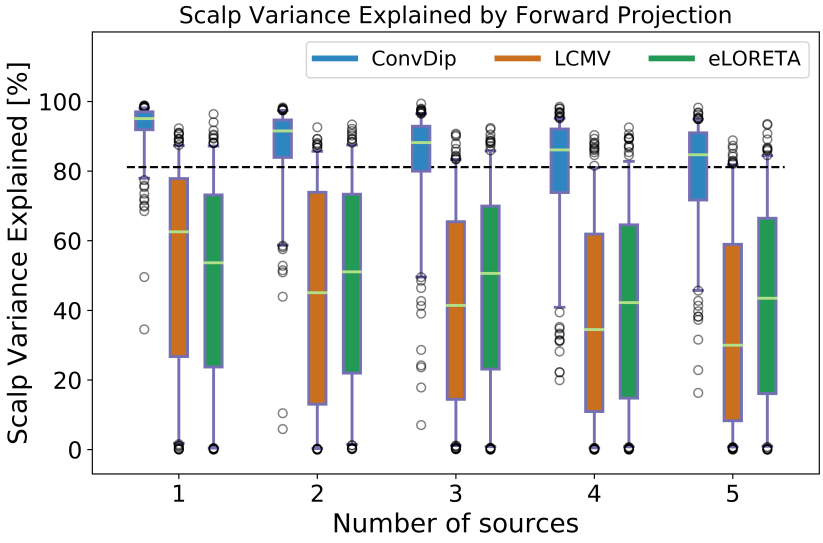

Figure 2: Variance explained by the predicted source projected to the scalp. Horizontal line (at 81 %) denotes proportion of signal as modeled in the artificial EEG samples. Central white lines show medians, boxes span the inter-quartile ranges and whiskers span over the central 95% of the data. ConvDip explains significantly more variance than the other two methods, regardless of the number of sources.

Mean localization error

Fig 3 shows the MLE as a function of the number of present sources. ConvDip shows substantially lower localization errors than both eLORETA ($M = 10.41$, $t = 28.53$, $p < 0.001$) and LCMV beamformer ($M = 3.67$, $t = 11.79$, $p < 0.001$), regardless of how many sources are present in the target (Fig. 3). Notably, LCMV beamformer yields lower localization errors compared to eLORETA ($M = 6.75$, $t = 19.00$, $p < 0.001$).

Normalized mean squared error

The normalized mean squared error (nMSE) provides a measure of global dissimilarity. ConvDip achieved lowest nMSE regardless of the number of sources present. This result indicates, that the ConvDip neural network learned to be flexible concerning the number and extent of the target sources during the training phase (Fig. 4). In particular, ConvDip achieved significantly lower nMSE compared to LCMV ($M = 0.15$, $t = 38.50$, $p < 0.001$) and eLORETA ($M = 0.07$, $t = 15.20$, $p < 0.001$).

Numbers of missed sources and ghost sources

Figs. 5 and 6 show the proportion of samples that contain missed and ghost sources. On average, ConvDip is capable of reconstructing more sources than LCMV ($M = 0.76$, $t = 10.05$, $p < 0.001$) and eLORETA ($M = 1.13$, $t = 15.13$, $p < 0.001$). LCMV, in turn, is capable of reconstructing slightly more sources than eLORETA ($M = 0.37$, $t = 4.77$, $p < 0.001$). Both LCMV and ConvDip produce considerably less ghost sources than eLORETA (ConvDip: $M = 1.34$, $t = 28.73$, $p < 0.001$; LCMV: $M = 1.23$, $t = 26.84$, $p < 0.001$). ConvDip yields slightly fewer ghost sources compared to LCMV beamforming ($M = 0.11$, $t = 3.27$, $p < 0.05$). These results are remarkable since eLORETA and LCMV inverse solutions use a noise covariance matrix based on 100 trials of neural activity to ensure a proper regularization of the inverse solution, but nevertheless ConvDip makes a more accurate prediction on a single time instance of EEG already.

Covariates of source reconstruction success

In this section we evaluate the determinants of source reconstruction errors. The influence of (1) the number of present sources, (2) depth of sources in the brain and (3) the relative noise level were evaluated separately

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

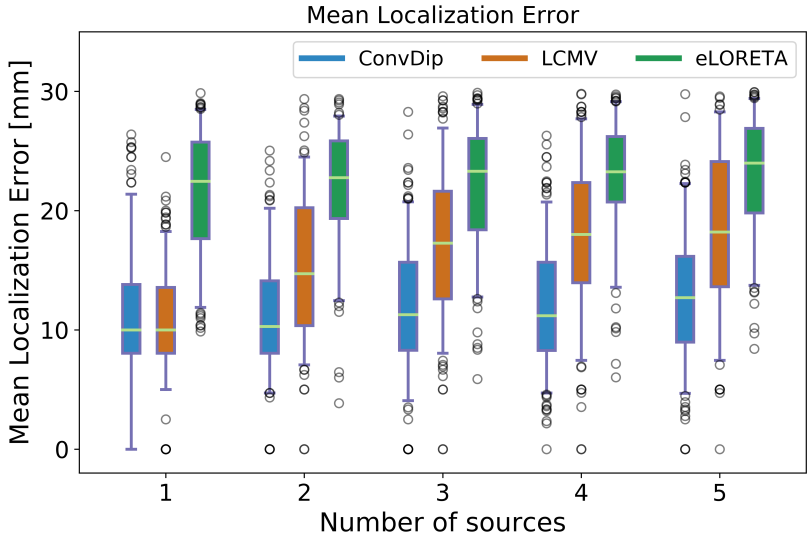

Figure 3: Mean localization error as a function of the number of target sources. Central white lines show medians, boxes span the inter-quartile ranges and whiskers span over the central 95% of the data. ConvDip shows substantially less localization errors than the other two methods, except when only a single source is present, where it is comparable to LCMV and more accurate than eLORETA.

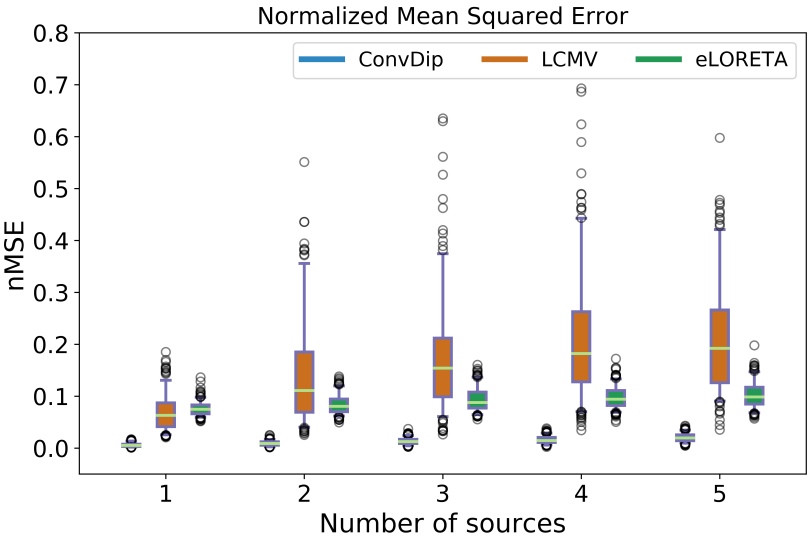

Figure 4: Normalized mean squared error for each inverse solution and for different numbers of sources. Central white lines show medians, boxes span the inter-quartile ranges and whiskers span over the central 95% of the data. ConvDip produces significantly less normalized mean squared errors than the other two methods.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

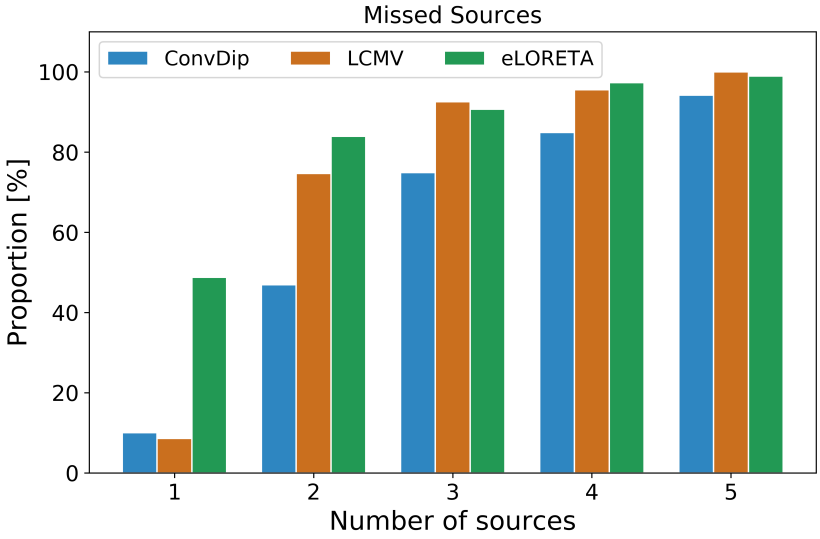

Figure 5: Proportion of samples containing missed sources.

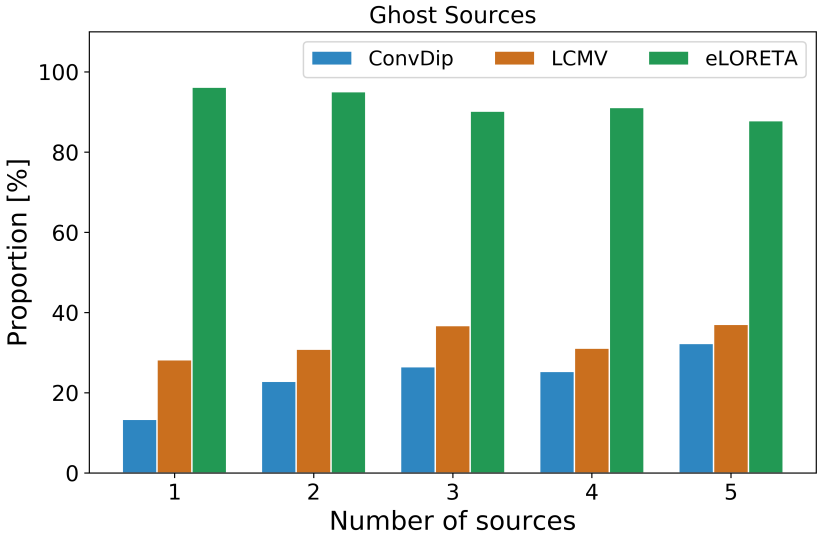

Figure 6: Proportion of samples containing ghost sources.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Number of sources		Source depth		Noise	
Inverse Solution	MLE	nMSE	MLE	nMSE	MLE	nMSE
ConvDip	0.12**	0.63***	-0.20**	-0.32***	0.10**	0.07*
LCMV	0.42***	0.46***	-0.14*	-0.24***	0.06	0.17***
eLORETA	0.10**	0.42***	-0.02	-0.46***	0.01	0.19***

Table 1: Covariates of source reconstruction success. Values express the Pearson correlation coefficient r. MLE: mean localization error, nMSE: normalized Mean Squared Error. $* < 0.05$, $** < 0.01$, $*** < 0.001$.

for each of the focused inverse solution methods in the following way: We calculated for each of the three factors separately Pearson correlation coefficients with the MLE (local similarity) and with nMSE (global similarity) across the samples in the evaluation set. Results are shown in Tab 1.

We found that the strongest predictors of EEG source localization errors for eLORETA and LCMV are the average depth and the number of sources in the target. Interestingly, noise does not play a significant role for these measures and only a weak role for ConvDip.

The strongest predictor of inverse solution errors when using ConvDip is the number of present sources. The correlation coefficient r^2 shows that 39.69 % of the nMSE in ConvDip can be explained by an elevated number of sources in the target, whereas only 10.24 % are explained by the average depth of target sources.

The inverse solution success of LCMV beamformer and eLORETA are also strongly influenced by the number of sources (21.16 - 17.64 %), albeit substantially less than ConvDip. These analyses show that ConvDip yields overall substantially lower nMSE. However, beyond the overall better performance, the errors produced by ConvDip are more directly influenced by physical boundaries (such as the attenuation of source signal strength over distance or the mixing of various sources) which in general renders any inverse solution calculation highly difficult.

Discussion

Overview

The M/EEG is a widely used technique to study in vivo neurophysiology at a temporal resolution on which neural processes occur. The spatial resolution of non invasive measures, however, is strongly limited due to the highly underdetermined problem when using scalp electrodes/ sensors. While the relationship between electromagnetic brain activity and its projection to scalp electrodes is unique, inferring the sources of activity from scalp signals is not. This missing injectivity constitutes an ill-posed inverse problem.

In this study we present ConvDip, a convolutional neural network that has been designed to solve the inverse problem of the M/EEG in a distributed dipole model. ConvDip was successfully trained to infer physiologically plausible inverse solutions based on single time-instances of (simulated) EEG data. Exemplary inverse solutions of a simulated sample and real data sample can be found in Appendix A and B, respectively.

General discussion of results

Computation speed

As already pointed out by Sclabassi et al. (2001), ANNs naturally yield faster inverse solutions than iterative methods. On our workstation (CPU: Intel i9 9900K, GPU: Nvidia Titan V, 64 Gb RAM) one forward pass of ConvDip took on average 26.07 ms. When all necessary preprocessing steps (such as rereferencing, scaling and interpolation to 2-D scalp maps) are taken into account we reached 38.37 ms of computation time on average. These short computation times are at a distinct advantage of using ANNs over iterative methods, e.g. in real-time applications as in neurofeedback experiments or in the development of brain-computer interfaces. When comparing our approach to the existing literature on ANN-based inverse solutions, we can now confirm that an ANN not only provides competitively accurate single dipole locations but also a distributed dipole model of the brain-electric activity, taking into account more than one dipole. This can be attributed mostly to the rapid developments in the machine learning domain such as the introduction

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of convolutional layers (LeCun and Bengio (1995)) and the improvements of graphical processing units that render the training of large ANNs possible in acceptable time.

Scalp variance explained

We used multiple measures to evaluate different aspects of the performance of ConvDip and to establish a fair comparison to alternative inverse solutions. The first measure of choice was the EEG variance if the forward projection explained by the inverse solution (Fig. 2). This measure aids as a plausibility check for the inverse solution in general since the input EEG signal should be reflected in the forward projection of the inverse solution. The maximal value of explained variance of the simulated data we used was calculated at 81 %, given our simulated signal to noise ratio.

It became evident that ConvDip explained on average more variance than produced by the source signals, whereas eLORETA and LCMV beamformer explained consistently less variance. Hence, ConvDip produces more liberal inverse solutions, whereas eLORETA and LCMV produce more conservative estimates

We assume that ConvDip's behavior can be attributed to the training procedure in which we introduced the sensor-space loss. By calculating the loss between forward projected predictions and the input EEG (including simulated noise), liberal source reconstructions were incentivized. This problem should be addressed in the future to build models, that may be more robust to noise. Nevertheless, ConvDip was significantly closer to the optimal explained variance than both alternative inverse solutions and therefore shows its capability to produce plausible inverse solutions.

Mean localization error

For single sources the localization error of ConvDip was comparable to the error produced by the beamforming approach, and clearly better than the eLORETA approach. However, increasing the number of sources up to five, ConvDip shows an overall lower MLE compared to eLORETA and LCMV beamformer (Fig. 3).

Normalized mean squared error

During our validation of the performance we found that ConvDip produced lower nMSE compared to the two state-of-the-art methods of comparison (Fig. 4). The nMSE reflects not only the capability to produce a plausible inverse solution but also that the general topology of sources in the simulations are well reflected in the inverse solutions yielded by ConvDip. However, given how we simulated our artificial EEG data, at this point ConvDip is put at an 'unfair advantage' over any other inverse solution where the nature of brain-electric signals is assumed to be different, e.g. in the smoothness assumption of LORETA-like inverse solutions or in sparse inverse solutions such as MxNE (Gramfort et al. (2012)). For example, in a world where brain signals exhibit exactly the shape and smoothness as LORETA-like solutions assume, LORETA would be the best choice to reconstruct them. However, the underlying assumptions have shown to be incomplete (Nunez and Srinivasan (2006)) - an aspect that we took into account when modeling the artificial EEG data. In summary, the advantage of ConvDip regarding nMSE partially reflects the different assumptions underlying the EEG signal. And it may be a matter of debate, which assumption is closest to the brain's reality.

Missed & ghost sources

ConvDip is capable to discover significantly more true active sources and at the same time shows an even lower false-positive rate than eLORETA and LCMV (Fig. 5, 6). This renders ConvDip a valuable tool for EEG source imaging in *in-vivo* neuroscientific research as it might help to recover a more accurate and nuanced picture of the neural networks underlying EEG recordings.

A special situation is the identification of a single source, which is highly important e.g. in pre-surgical epileptic diagnostic. The current version of ConvDip does not show higher capabilities than LCMV beamforming for this case with respect to missed and ghost sources (Fig. 3 and 5). For such specific cases it might be advantageous to train a specific ConvDip variant with a training set specifically adjusted towards this task. Specifically, one could design a training set for ConvDip that contains only single sources of certain spatial extent to more reliably locate epileptic foci.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

During development of ConvDip it became evident that the rate of ghost versus missed sources can be directly controlled by the loss function in the training phase of the artificial neural network (see 1 and 2). Further investigations into ANN-based inverse solutions should focus on tweaking these weights w_1 and w_2 depending on the task at hand. For clinically relevant data (e.g. recordings of epileptic seizures) it might be more important to minimize the false-positive rate by penalizing false non-zero predictions more.

Causes of EEG inverse solution errors

We identified the driving forces of source reconstruction success and demonstrated how the number of present sources, noise and depth of sources in the brain are critical determinants of localization errors and global similarity between predicted and target sources. It was evident that the global similarity of the inverse solutions yielded by ConvDip were more severely influenced by the overall difficulty of the samples (SNR, depth of sources in the brain, number of sources) compared to the two classical approaches. In total, 50.42 % of the nMSE in ConvDip inverse solutions was explained by these factors, as presented in Table , compared to 41.41 % in eLORETA and 29.82 % in LCMV beamforming.

We conclude that ConvDip supersedes the other tested inverse solutions in all measures we focused on. Further, ConvDip seems to cover a larger range of possible neural source configurations and seems to become only limited by physical and related mathematical limitations given by the nature of EEG data.

Perspectives for improvement

While increasing the number of sources reasonably increases the inverse problem, the susceptibility to elevated noise levels should be tackled in the future development of the neural network approach to the EEG inverse problem. Notably, the present version of ConvDip is trained on single time instances of EEG data. From a single EEG data point it is, however, not possible to estimate a baseline of activity. As a consequence, such a baseline has not yet been taken into account in ConvDip, which leaves space for improvement regarding a more regularized solution. In contrast, LCMV and eLORETA explicitly make use of a noise covariance matrix which is used to regularize the inverse solution. Data augmentation that involves varying noise levels for the same samples might be a first step for more robustness to noise levels. Further, progress may be made by extending the input of ConvDip to an EEG time series to allow the neural network to learn noise and signal patterns. Possible further advances could be made by preserving the 3D structure of the output space by, instead of flattening the voxel data into a 1D vector. This can be realized using a deconvolutional neural network for 2D-3D projections (Lin et al. (2018)) or by using spatiotemporal information with a LSTM network (Cui et al. (2019)).

Limitations

Realistic simulations

The simulation of EEG data in this study was based on basic assumptions on the number and distribution of electromagnetic brain signals as described by Nunez and Srinivasan (2006). The validity of predictions of any inverse solution is only granted if the assumptions on the origin of the EEG signal are correct. This is indeed one of the critical challenges towards realistic inverse solutions in general, both for data-driven inverse solutions (such as ConvDip) as well as iterative inverse solutions with physiological constraints (e.g. eLORETA). Therefore, when interpreting the applicability of ConvDip to real data, it is important to specify and justify these assumptions. We show that the knowledge about brain-electric activity underlying EEG signals can be seamlessly integrated into the training data for an ANN to calculate inverse solutions. However, it is evident from the present work, how critical the particular choice of parameters is with respect to the performance, as ConvDip produces solutions that closely resemble the training samples (see supplementary Material, Fig. 7).

An important observation in real EEG data is the inter-individual variability, which is in part a direct result of individual brain anatomies. It is thus advised to collect anatomical MRI brain scans of each individual subject. Individual neural networks then need to be trained based on each subjects' individual anatomy provided by the MRI data. This may help to achieve accurate group-study source estimations. A next step of ANN based inverse solutions could include anatomical information in the training process,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

such that individual anatomical properties (e.g. conductivity, folding of gyri and sulci) can be used to guide inference. Another option is an inter-individual transfer learning approach, where ConvDip is trained on one subjects' anatomy and fine-tuned for each additional subject with new training data of the individual anatomies. Fine tuning could be achieved by replacing the output-layer, lowering the learning rate and retraining for only few epochs.

Computational costs

Although the availability of high-performing hardware resources (e.g. high performing graphic cards) is growing rapidly, the training of the presented neural network architecture requires a considerable amount of processing power and processing time. The version of ConvDip shown in this work required ~1.5 hours of training using a NVIDIA Titan V graphical processing unit. Improving the architecture of ConvDip for shorter training time is an important task for future research, especially when individualized models are required. Limiting the solution space to fewer voxels may be one way to decrease the computation time since most trainable parameters are located at the final fully-connected layers.

Another computationally expensive processing step is the generation of realistic artificial training data. In this study, we projected neural activity using a three-layer BEM head model. Generating 100,000 simulations of neural activity took 12 minutes using a PC with CPU @ 4x4.3GHz, 16 GB of RAM and a NVIDIA Titan V. Software that enables EEG researchers to perform physiologically realistic simulations of neural activity for their own ConvDip application is available from different resources, e.g. functions provided by the MNE library in Python (Gramfort et al. (2014)).

Outlook

We showed that ConvDip, a deep CNN, is capable of solving the inverse problem using a distributed dipole solution. The association between single time points of EEG measurements and underlying source distributions can be learned and used to predict plausible inverse solutions. These inverse solutions were furthermore shown to globally reflect the topology of the brain-electric activity given by the training data.

Predictions yielded from ConvDip exceed the performance of the LCMV beamformer and eLORETA on all focused measures. Further, it the application of ConvDip to real EEG data yields reasonable results. The fully trained ConvDip requires only little computational costs of 38 ms, which makes it a promising tool for real-time applications where only single time points of EEG data are available.

We want to emphasize the importance of realistic simulations of *real* neural activity that can be measured by M/EEG. Although some knowledge about the generation of scalp potentials was implemented in our EEG simulations, including more physiological constraints can further reduce the complexity of the inverse problem. One example is the estimation of cortical column orientation (Bonaiuto et al. (2019)) or the incorporation of empirical knowledge about region-specific activity from in-vivo (e.g. electrocorticogram) studies. Taking into account these additional aspects may further improve ConvDip's performance.

The current ConvDip version was based on single time points of artificial EEG data. Exploiting temporal aspects of brain dynamics may provide additional valuable information and predictive power that may further refine future ConvDip versions.

Appendix

A: Simulated sample

B: Real sample

Acknowledgements

The authors want to thank NVIDIA for supporting this work as part of the academic seeding program with a Nvidia Titan V GPU.

Competing interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

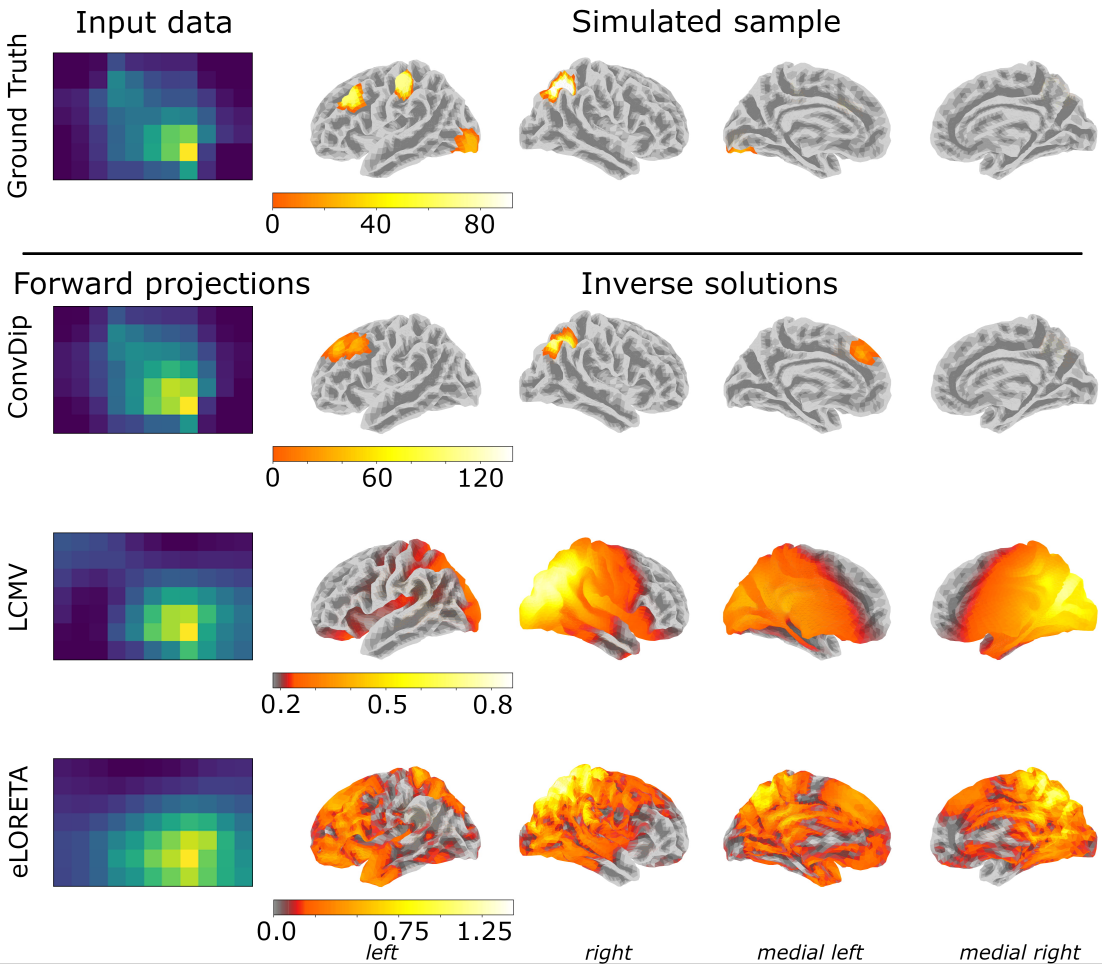

Figure 7: Inverse solutions of one exemplary sample of simulated data. Top Row: Simulated brain activity (right) and corresponding scalp map + added gaussian noise (left). Below top row: Brain plots of inverse solutions and respective scalp maps (forward projection) are shown for ConvDip, LCMV beamformer and eLORETA.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

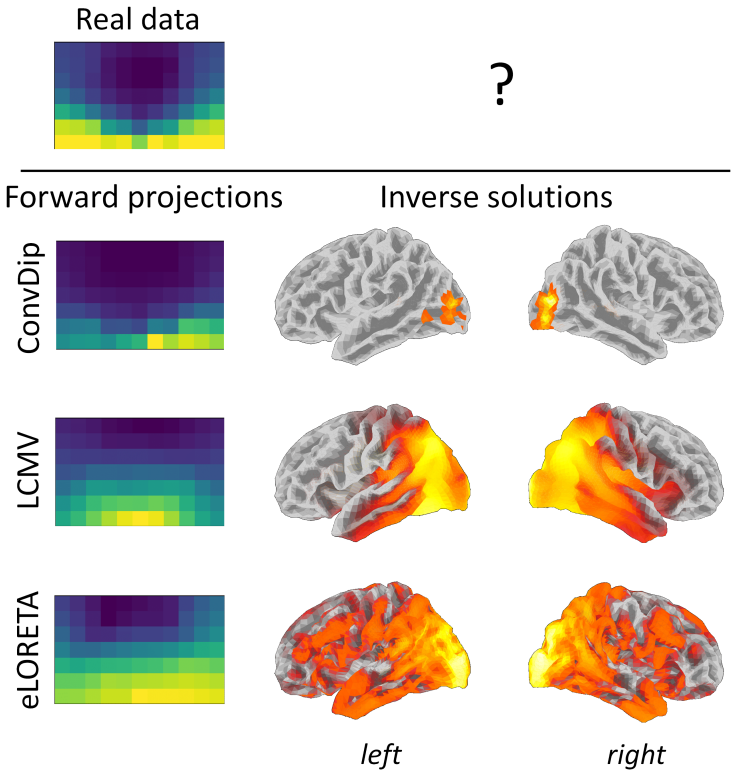

Figure 8: Exemplary inverse solution of real data. Visually evoked potential of face stimulus at 150 ms post-stimulus of one person. Left: Simplistic low-resolution matrix representations of scalp map of real ERP data. Top row: Scalp map of real ERP data as input image for the different inverse solutions (ConvDip, eLORETA and LCMV beamformer). Second to fourth row: Estimated sources in the brain, calculated by the different inverse solution methods (middle and right column) and the forward projections of these solutions to the scalp (left column).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, and Matthieu Devin. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

Udantha R. Abeyratne, G. Zhang, and P. Saratchandran. EEG source localization: A comparative study of classical and neural network methods. *International journal of neural systems*, 11(04):349–359, 2001.

James J. Bonaiuto, Fardin Afdideh, Maxime Ferez, Konrad Wagstyl, Jérémie Mattout, Mathilde Bonnefond, Gareth R. Barnes, and Sven Bestmann. Estimates of cortical column orientation improve MEG source inversion. *bioRxiv*, page 810267, 2019.

Richard P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 14(4):422–425, 1971.

François Chollet et al. Keras. 2015.

Michael X. Cohen. Where does EEG come from and what does it mean? *Trends in neurosciences*, 40(4): 208–218, 2017.

Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

Song Cui, Lijuan Duan, Bei Gong, Yuanhua Qiao, Fan Xu, Juncheng Chen, and Changming Wang. EEG source localization using spatio-temporal neural network. *China Communications*, 16(7):131–143, 2019.

S. B. Damelin and N. S. Hoang. On surface completion and image inpainting by biharmonic functions: Numerical aspects. *International Journal of Mathematics and Mathematical Sciences*, 2018, 2018.

Arnaud Delorme, Jason Palmer, Julie Onton, Robert Oostenveld, and Scott Makeig. Independent EEG sources are dipolar. *PloS one*, 7(2), 2012.

B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284, 1999. ISSN 1065-9471. doi: 10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4.

Karl Friston, Lee Harrison, Jean Daunizeau, Stefan Kiebel, Christophe Phillips, Nelson Trujillo-Barreto, Richard Henson, Guillaume Flandin, and Jérémie Mattout. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3):1104–1120, 2008.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

Alexandre Gramfort, Matthieu Kowalski, and Matti Hämäläinen. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine & Biology*, 57(7):1937, 2012.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S. Hämäläinen. MNE software for processing MEG and EEG data. *Neuroimage*, 86:446–460, 2014.

Roberta Grech, Tracey Cassar, Joseph Muscat, Kenneth P. Camilleri, Simon G. Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste. Review on solving the inverse problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, 5(1):25, 2008.

Matti S. Hämäläinen and Risto J. Ilmoniemi. Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & biological engineering & computing*, 32(1):35–42, 1994.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gert Van Hoey, Jeremy De Clercq, Bart Vanrumste, Rik Van de Walle, Ignace Lemahieu, Michel D'Havé, and Paul Boon. EEG dipole source localization using artificial neural networks. *Physics in Medicine and Biology*, 45(4):997–1011, April 2000. ISSN 0031-9155, 1361-6560. doi: 10.1088/0031-9155/45/4/314.

A. A. Ioannides, J. P. R. Bolton, and C. J. S. Clarke. Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Problems*, 6(4):523, 1990.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Zoltan J. Koles. Trends in EEG source localization. *Electroencephalography and clinical Neurophysiology*, 106(2):127–137, 1998.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

Chen-Hsuan Lin, Chen Kong, and Simon Lucey. Learning efficient point cloud generation for dense 3d object reconstruction. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Steven J. Luck. *An Introduction to the Event-Related Potential Technique*. MIT press, 2014.

Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

Paul L. Nunez and Ramesh Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, USA, 2006.

RD Pascual-Marqui, Christoph M. Michel, and Dietrich Lehmann. Low-resolution electromagnetic tomography–a new method for localizing electrical activity in the brain. *International Journal of psychophysiology*, 18:49–65, 1994.

Roberto D. Pascual-Marqui. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization. *arXiv preprint arXiv:0710.3341*, 2007.

Roberto Domingo Pascual-Marqui. Review of methods for solving the EEG inverse problem. *International journal of bioelectromagnetism*, 1(1):75–86, 1999.

Petra Ritter and Arno Villringer. Simultaneous EEG–fMRI. *Neuroscience & Biobehavioral Reviews*, 30(6): 823–838, 2006.

Claude Robert, Jean-François Gaudy, and Aimé Limoge. Electroencephalogram processing using neural networks. *Clinical Neurophysiology*, 113(5):694–701, 2002.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

Robert J. Sclabassi, Murat Sonmez, and Mingui Sun. EEG source localization: A neural network approach. *Neurological research*, 23(5):457–464, 2001.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.09.033506; this version posted April 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15 (1):1929–1958, 2014.

Roman Tankelevich. Inverse problem's solution using deep learning: An EEG-based study of brain activity. Part 1 - rel. 1.0. February 2019.

B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE transactions on bio-medical engineering*, 44(9):867–880, September 1997. ISSN 0018-9294. doi: 10.1109/10.623056.

Motohiro Yuasa, Qinyu Zhang, Hirofumi Nagashino, and Yohsuke Kinouchi. EEG source localization for two dipoles by neural networks. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, volume 4, pages 2190–2192. IEEE, 1998.

Q. Zhang, M. Yuasa, H. Nagashino, and Y. Kinouchi. Single dipole source localization from conventional EEG using BP neural networks. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, volume 4, pages 2163–2166, Hong Kong, China, 1998. IEEE. ISBN 978-0-7803-5164-6. doi: 10.1109/IEMBS.1998.747038.