Deep learning approaches for neural decoding across architectures and recording modalities*

Jesse A. Livezey[1,2,*] and Joshua I. Glaser[3,4,5,*]

jlivezey@lbl.gov, j.glaser@columbia.edu
*equal contribution
[1]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States
[2]Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, California, United States
[3]Department of Statistics, Columbia University, New York, United States
[4]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, United States
[5]Center for Theoretical Neuroscience, Columbia University, New York, United States

November 4, 2020

Abstract

Decoding behavior, perception, or cognitive state directly from neural signals is critical for brain-computer interface research and an import tool for systems neuroscience. In the last decade, deep learning has become the state-of-the-art method in many machine learning tasks ranging from speech recognition to image segmentation. The success of deep networks in other domains has led to a new wave of applications in neuroscience. In this article, we review deep learning approaches to neural decoding. We describe the architectures used for extracting useful features from neural recording modalities ranging from spikes to fMRI. Furthermore, we explore how deep learning has been leveraged to predict common outputs including movement, speech, and vision, with a focus on how pretrained deep networks can be incorporated as priors for complex decoding targets like acoustic speech or images. Deep learning has been shown to be a useful tool for improving the accuracy and flexibility of neural decoding across a wide range of tasks, and we point out areas for future scientific development.

Contact: jlivezey@lbl.gov, j.glaser@columbia.edu

Jesse A. Livezey is a post-doctoral research scientist in the Neural Systems and Data Science Lab at the Lawrence Berkeley National Laboratory. He obtained his PhD in Physics from the University of California, Berkeley. His research interests include applications of machine learning and information theory to neuroscience datasets.
Joshua I. Glaser is a post-doctoral research scientist in the Center for Theoretical Neuroscience and Department of Statistics at Columbia University. He obtained his PhD in Neuroscience from Northwestern University. His research interests include neuroscience, machine learning, and statistics.

1 Introduction

Using signals from the brain to make predictions about behavior, perception, or cognitive state, i.e., "neural decoding", is becoming increasingly important within neuroscience and engineering. One common goal of neural decoding is to create brain computer interfaces, where neural signals are used to control an output in real time [1, 2]. This could allow patients with neurological or motor diseases or injuries to, for example, control a robotic arm or cursor on a screen, or produce speech through a synthesizer. Another common goal of neural decoding is to gain

*This is a pre-copyedited, author-produced version of an article accepted for publication in Briefings in Bioinformatics following peer review.

a better scientific understanding of the link between neural activity and the outside world. To provide insight, decoding accuracy can be compared across brain regions, cell types, different types of subjects (e.g., with different diseases or genetics), and different experimental conditions [3–11]. Additionally, the representations learned by neural decoders can be probed to better understand the structure of neural computation [12–16]. These uses of neural decoding span many different neural recording modalities and span a wide range of behavioral outputs (Fig. 1A).

Within the last decade, many researchers have begun to successfully use deep learning approaches for neural decoding. A decoder can be thought of as a function approximator, doing either regression or classification depending on whether the output is a continuous or categorical variable. Given the great successes of deep learning at learning complex functions across many domains [17–26], it is unsurprising that deep learning has become a popular approach in neuroscience. Here, we review the many uses of deep learning for neural decoding. We emphasize how different deep learning architectures can induce biases that can be beneficial when decoding from different neural recording modalities and when decoding different behavioral outputs. We aim to provide a review that is both useful to deep learning researchers looking to understand current neural decoding problems and to neuroscience researchers looking to understand the state-of-the-art in neural decoding.

2 Deep learning architectures

At their core, deep learning models share a common structure across architectures: 1) simple components formed from linear operations (typically addition, matrix multiplication, or convolution) plus a nonlinear operation (for example, rectification or a sigmoid nonlinearity); and 2) composition of these simple components to form complex, layered architectures [27]. The simplest fully-connected neural networks combine matrix multiplication and nonlinearities. While more complex deep network layer types, e.g., graph neural networks [28] or networks that use attention mechanisms [29], have been developed, they have not seen much use in neuroscience. Additionally, given that datasets in neuroscience typically have limited numbers of trials, shallower neural networks are often used for neural decoding compared with the networks used in common machine learning tasks.

Recurrent neural networks (RNNs) act on a sequence of inputs of potentially varying length, which occurs in neuroscience data (e.g., trials of differing duration). This is unlike a fully-connected network, which requires a fixed dimensionality input. In an RNN, the inputs, X_t, are then projected (with weights w_X) into a hidden layer, H_t, which recurrently connects to itself (with weights w_H) across time (Fig. 1B)

$$\begin{aligned} H_{t+1} &= f(w_H \cdot H_t + w_X \cdot X_t) \\ Y_t &= g(w_Y \cdot H_t) \end{aligned} \tag{1}$$

where $f(\cdot)$ and $g(\cdot)$ are nonlinearities and Y_t is the RNN output. Finally, the hidden layer projects to an output, Y_t which can itself be a sequence (Fig. 1B), or just a single data point. Commonly used RNN architectures like LSTMs and GRUs [22, 27, 30] have multiplicative "gating" operations in addition to element-wise nonlinearities. Recurrent networks are commonly used for neural decoding since they can flexibly incorporate information across time.

Convolutional neural networks (CNNs) can be trained on input and output data in many different formats. For example, convolutional architectures can take in structured data (1d timeseries, 2d images, 3d volumes) of arbitrary size [23, 27, 31, 32]. Input neural data, X, may have one or more channels indexed by c (which may be combined in a filter as in Eq 2 or operated on individually) and temporal or spatial dimensions indexed by $(t, \ldots)$. A convolutional layer has weights with multiple filters (f), that combine across channels, that have temporal (or spatial) extent $(T, \ldots)$ followed by a nonlinearity, $g(\cdot)$. For example, for a 1d convolution the activations in a layer are calculated as

$$h_{t,f} = g\Big(\sum_{\tau=0,c}^{T-1} w_{\tau,f,c} X_{t+\tau,c}\Big). \tag{2}$$

The 2d and 3d extensions have more output indices in addition to t for the additional dimensions, and more dimensions are summed over for each filter. The convolutional layers will then learn filters of the corresponding dimensions, in order to extract meaningful local structure (Fig. 1C). The convolutional layers are commonly used if there are important features that are translation invariant, as in images. This is done hierarchically, in order to learn filters of varying scales (i.e., varying temporal or spatial frequency content), which is a useful prior for multi-scale data, such as images. Next, depending on the output that is being predicted, the convolutional layers are fed into other types of layers to produce the final output (e.g., into fully connected layers to classify an image).

Weight-sharing, where the values of some parameters are constrained to be the same, is often used for neural decoding. For instance, the parameters of a convolutional (in time) layer can be made the same for differing input channels or neurons, so that these different inputs are filtered in the same way. For neural decoding, this can be beneficial for learning a shared set of data-driven features for different recording channels (e.g., a relevant frequency pattern for ECoG datasets) as an alternative to human-engineered features.

Training a neural decoder uses supervised learning, where the network's parameters are trained to predict target outputs based on the inputs. Recent work has combined supervised deep networks with unsupervised learning techniques, which learn lower dimensional representations that reproduce one data source. One common unsupervised method, generative adversarial networks (GANs) [33, 34], generate an output, e.g., an image, given a vector of noise as input. GANs are trained to produce images that fool a classifier deep network about whether they are real versus generated images. Another method is convolutional autoencoders, which are trained to encode an image into a latent state, and then reconstruct a high fidelity version [35]. These unsupervised methods can produce representations of the decoding input or output that are sometimes more conducive for decoding and can potentially leverage larger datasets for training than are available for neural decoding.

3 Stages of neural decoding

In order to go from the raw neural signal to the final predicted output (e.g. speech), the neural decoding pipeline can be conceptually broken down into a few components, each of which can incorporate deep learning.

1. Preprocessing / Feature Engineering. First, the raw neural signals are processed to create features that are beneficial for neural decoding. Sometimes, these features are hand-engineered based on previous knowledge, traditionally with the goal of creating features that are most compatible with linear decoders. More recently, supervised feature engineering has been incorporated into deep learning architectures. That is, a more raw form of the input is provided into the neural decoder, and a first stage of the deep network decoder will automatically learn to extract relevant features. Specific neural network architectures can be beneficial for this automatic feature engineering (Fig. 2). It is also possible to generate features from the neural data with deep learning [42, 43] in an unsupervised manner and then use those features with simple linear decoders.

2. Mapping from features to final (or intermediate) output. This central part transforms the features to an output representation, and deep learning tools allow this mapping to be a flexible nonlinear function.

3. Mapping from intermediate to final output (optional). Neural decoding is used to predict many outputs, including movement, speech, vision, and more. Sometimes, the output variable will be directly predicted from the neural inputs, e.g., when predicting movement velocities (and thus this stage is not relevant). Other times, the neural decoder may be trained to predict some intermediate representation, which has a predetermined mapping to the output (Fig. 3). For example, a GAN can be trained to generate an image using a small number of latent variables. This mapping from the low-dimensional variables to images can be learned without having to simultaneously record neural activity. Then, to decode an image from neural activity, one can train the neural decoder to predict the latent variables to be fed into the GAN, rather than the entire high-dimensional image. This two-step approach can be especially beneficial when the output data is complex and high-dimensional, as is often the case in vision or speech. In effect, the generative model can act as a prior on the underconstrained decoding solution.

We expand on these stages below - first focusing on neural recordings and how they are transformed into features, and then focusing on the deep learning methods used for predicting the final outputs of neural decoding.

4 The inputs of decoding: neural recording modalities and feature engineering

To understand how varying neural network architectures can be preferable for processing different neural signals, it is important to understand the basics of neural recording modalities. These modalities differ in their invasiveness, and their spatial and temporal precision.

4.1 Spikes

The most invasive recordings involve inserting electrodes into the brain to record voltages. This allows experimentalists to record spikes, or action potentials, the basic unit of neural signaling. Action potentials are the fast electrical transients that individual neurons use to signal and are triggered when a neuron's membrane potential depolarizes past its threshold. To get binary spiking events, the recorded signals are high-pass filtered and thresholded. They are then often sorted into waveforms attributed to individual neurons, sometimes using deep learning tools [44, 45]. Datasets with spikes are thus binary time courses from all of the recording channels or neurons (Fig. 1A). Spikes are more commonly recorded from animal models than humans because of their invasive nature.

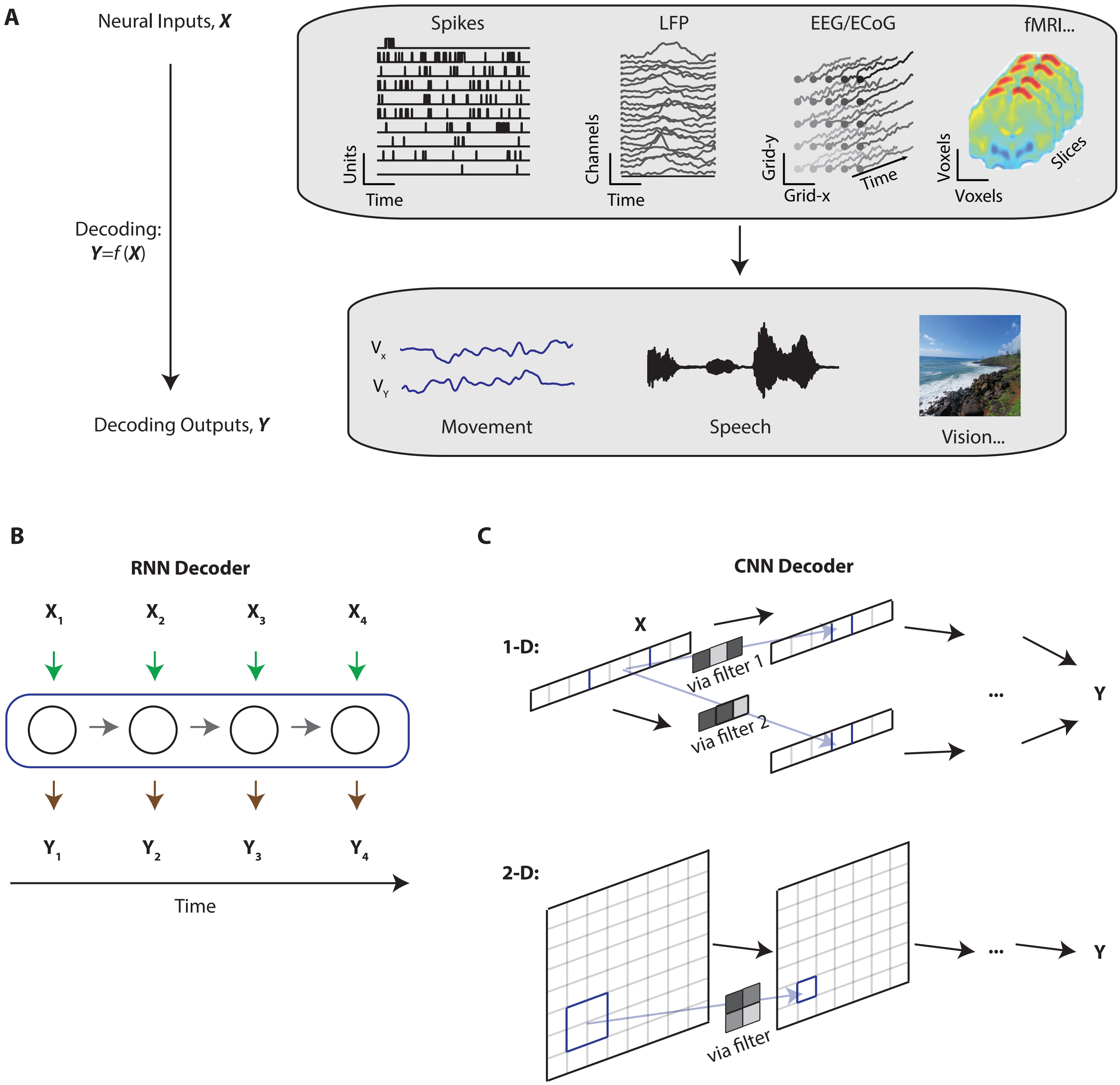

Figure 1: Schematics. **A:** Schematics of neural decoding, which can use many different neural modalities as input (top) and can predict many different outputs (bottom), such as movement velocities (left) [36], a waveform of speech (center) [37], or visual images (right) [38]. Embedded figures are adapted from [39–41]. **B:** A schematic of a standard recurrent neural network (RNN). Each arrow represents a linear transformation followed by a nonlinearity. Arrows of the same color represent the same transformations occurring. The circles representing the hidden layer typically contain many hidden units. More sophisticated versions of RNNs, which include gates that control information flow through various parts of the network, are commonly used. For example, see [27] for a schematic of an LSTM. **C:** A schematic of convolutional neural networks (CNNs). A convolutional transformation takes a learned filter and convolves it with the input, and then passes this through a nonlinearity. As an example of a 1-dimensional convolutional transformation (top), as may be the case a single time series, a filter of length 3 (for example) is multiplied element-wise with all input segments of length 3 to get the values of the next network layer. In CNNs, typically multiple filters (here, filter 1 and filter 2) are learned within each layer, and the outputs of all filters are combined in a subsequent layer. In our example of 2-dimensional convolutional transformation (bottom), as may be the case for spatial data, a $2{\times}2$ filter is multiplied pixel-wise with all $2{\times}2$ blocks to get the values of the next layer in the network. Convolutions can also occur in three dimensions for neural decoding.

For use in neural decoding, spikes are typically first converted into firing rates by determining the number of spikes in time bins, sometimes with additional temporal smoothing. Then, these firing rates are fed into the neural decoder. Commonly, these firing rates are considered the relevant features, and thus additional neural network architectures are not used to extract unknown features from the input.

One form of feature engineering that is used for spike trains, especially when many neurons are recorded, is dimensionality reduction. That is, a lower-dimensional representation of the firing rates is used to predict the outputs. This dimensionality reduction can use a variety of methods, from classical linear methods, e.g. PCA, to deep learning approaches, e.g. autoencoders [43]. This dimensionality reduction step is usually done prior to decoding the output, but it is also possible to incorporate this step into a single neural network decoder [46], so that the learned lower-dimensional representations are particularly relevant for predicting the output. We note that dimensionality reduction is not specific to decoding with spiking activity, but can also be applied to the neural recording modalities described below [46, 47].

Finally, in future research, it might be advantageous to provide a more raw form of spiking as input, rather than binned spike counts. Then, one could use deep learning architectures to do feature engineering. For example, with binary spiking events as input, the best size and temporal placement of time bins could be automatically determined, or even features related to the precise timing of spikes could be learned. It was also recently shown that using the envelope of spiking activity, a continuous signal, followed by feature extraction within a neural network, was able to improve decoding performance [48].

4.2 Calcium imaging

Another invasive technique for recording individual neurons' activities is calcium imaging, which uses microscopy to capture images of fluorescent calcium indicators that are sensitive to neurons' spiking activity [49]. These calcium indicators are genetically encoded within neurons in animal models, often within specific neuron types. The raw outputs of calcium imaging are videos: pixels measure fluorescence at the times when, and locations where, neurons are active. Calcium imaging is only used with animal models.

When analyzing calcium imaging data, the videos are typically preprocessed to extract time traces of fluorescences over time for each neuron [50]. Sometimes, additional processing will be done to estimate spiking events from the calcium traces [51]. Deep learning tools exist for both of these processing steps [52, 53]. For decoding, either the fluorescences, or the estimated firing rates (via the estimated spike trains), are then commonly used as input. While it could be possible to develop an end-to-end neural decoder that works with the videos as input, this may prove challenging given the potential for overfitting with high-dimensional input.

4.3 Wideband, LFPs, EEG, and ECoG

The electrode recordings for spikes simultaneously record local field potentials (LFPs), which are the low-pass filtered version (typically below ∼200Hz) of the same recorded voltage. LFPs are thought to be the sum of input activity of local neurons [54]. When all voltage is included across frequency bands, the voltage is generally referred to as wide-band activity. Datasets with LFP and wide-band are continuous time courses of voltages from all the recording channels (Fig. 1A). Note that traditionally, due to the distance between recording electrodes being greater than the spatial precision of recording, spatial relationships between electrodes are not utilized for neural decoding.

Electrical potentials measured from outside of the brain, that is electrocorticography (ECoG) and electroencephalography (EEG), are common neural recording modalities used in humans. ECoG recordings are from grids that record electrical potentials from the surface of the cortex, require surgical implantation, and often cover large functional areas of the cortex. EEG is a noninvasive method that records from the surface of the scalp from up to hundreds of spatially distributed channels. Like LFPs, datasets from ECoG and EEG recordings are continuous time courses of electrical potentials across recording channels (Fig. 1A), but here the spatial layout of the channels is also sometimes used in decoding. Note that as these electrical recording methods get less invasive, spatial precision decreases (from spikes to LFP to ECoG to EEG), which can lead to inferior decoding performance [55, 56]. Still, all these electrical signals can be recorded at high temporal resolution (100s-1000s of Hz) which make them good candidates for fast time-scale decoding.

When decoding from wide-band, LFP, EEG, and ECoG data, it is common to first extract spectrotemporal features from the data, for example the signals in specific frequency bands. Sometimes, only "task-relevant" frequencies will be used for decoding - for instance, using high gamma frequencies in ECoG to decode speech [57, 58] (Fig. 2A). More frequently, many frequencies will be included, to better understand which are contributing to decoding [15, 59]. In general, these extracted features can then be put into almost any type of neural decoder, such as linear (or logistic) regression or a deep neural network (e.g. [60]).

It is also possible to let a deep learning architecture do more of the feature extraction. One approach is to first convert each electrode's signal into a frequency domain representation over time (i.e., a spectrogram), often via a wavelet transform. Then, this 2-dimensional representation (like an image) is provided as input to a CNN [56, 61–63] (Fig. 2B). If multiple electrode channels are being used for decoding, each channel can be fed into an independent CNN, or alternatively, the CNN weights for each channel can be shared [56]. The CNN will then learn the relevant frequency domain representation for the decoding.

Another approach is to provide the raw input signals into a deep learning architecture (Fig. 2C). To learn temporal features, typically the signal is fed into a 1-dimensional CNN, where the convolutions occur in the time domain. This has been done with a standard CNN [64], in addition to variant architectures. Ahmadi et al. [65] used a temporal convolutional network, which is a more complex version of a 1-dimensional CNN that (among other things) allows for multiple timescales of inputs to affect the output. Li et al. [66] used parameterized versions of temporal filters that target synchrony between electrodes. These convolutional approaches will automatically learn temporal filters (like frequency bands) that are relevant for decoding.

In addition to temporal structure, there is often spatial structure of the electrode channels that can also be leveraged for neural decoding (Fig. 2A). Convolutional filters can be used in the spatial domain to learn spatial representations that are relevant for decoding, for example local functional correlation structure. It is common for the temporal filters and spatial filters to be learned in successive layers of the network, either temporal followed by spatial [67, 68] or vice-versa [69, 70]. Additionally, 3-dimensional convolutional filters can be learned that simultaneously incorporate both temporal and (2-dimensional) spatial dimensions [37] or 3 spatial dimensions [71]. Including spatial filters, which is most common in EEG and ECoG, can help learn spatial motifs that are most relevant for the task. Moreover, from a practical perspective, convolutional networks are an efficient way of processing high-dimensional spatial data.

4.4 fMRI and other non-invasive modalities

Magnetoencephalography (MEG), functional near infrared spectroscopy (fNIRS), and functional magnetic resonance imaging (fMRI) are also noninvasive recording modalities which are most often used in human decoding experiments. In this paper, amongst these non-invasive modalities, we primarily consider examples of decoding from fMRI. fMRI measures blood oxygenation (a proxy for neural activity), through its absorption of light and with resonance imaging respectively, and its temporal resolution are temporally limited by its dynamics. fMRI datasets contain activity signals in different "voxels" (locations) of the brain over time. Due to the limited temporal resolution, sometimes the temporal continuity of this data is not used for decoding purposes (Fig. 1A).

In fMRI, feature engineering is often done by hand. Commonly, the fMRI voxels that are used for decoding are subselected by hand or with statistical tests. Additionally, other hand-engineered metrics like functional connectivity are sometimes used as decoder inputs [72, 73]. As in EEG and ECoG, CNNs can be used to automatically extract features. For instance, spatial features can be learned by inputting the entire set of voxels into a 3-dimensional CNN [71, 74].

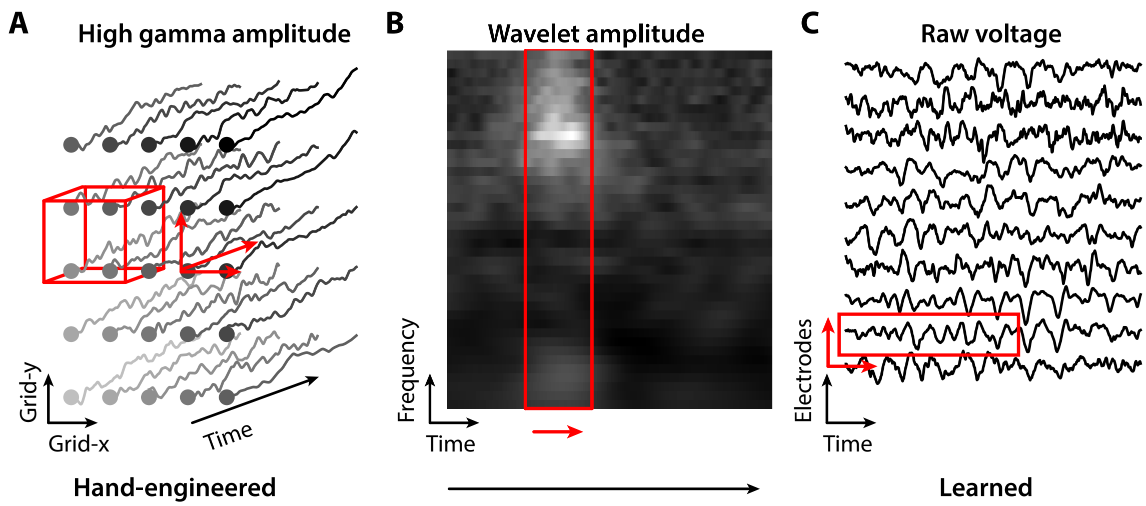

Figure 2: Feature engineering for neural decoding. Relevant features of neural data can be engineered completely by hand (left), automatically learned within a deep neural network (right), or somewhere in between. For all plots, the red box indicates a set of features across time, space, or frequency which will be filtered together by the first layer's convolutional or recurrent window. The red arrows indicate axes along which convolution or recurrence are performed. Sample data from [40]. **A:** High gamma amplitude, which is selected from a large filterbank of features from **B**, is shown spatially laid out in the ECoG grid locations. Deep network filters combine hand-engineered high gamma features across space and time. **B:** Spectrotemporal wavelet decomposition of one channel of the raw data, from **C**, may be used as the input to a deep network. The deep network filter shown combines features across frequency and time and can be shared across channels. **C:** Raw electrical potential recorded using ECoG across channels. The deep network filter shown combines features across time and can be shared across channels.

Table 1: Neural datasets and deep learning. Across the outputs of movement, speech, and vision, we overview a non-exhaustive list of deep learning applications to neural decoding. The column “Intermediate var.” refers to whether an intermediate variable was decoded, which was then used to predict the output. In papers where the goal was to compare many methods rather than focusing on a single method, we put a high performing method in the “Architecture” column. Lin. Reg., Linear Regression; Log. Reg., Logistic Regression; PCA, Principal Components Analysis; WF, Wiener Filter; WC, Wiener Cascade; FC, fully connected network; RNN, standard recurrent neural network; GRU, Gated recurrent unit network; LSTM, Long short term memory network; SVR, support vector regression; SVM, support vector machine; PLDS, Poisson linear dynamical system; TCN, Temporal Convolutional Network; LDA, linear discriminant anallysis; HMM, hidden markov model; (Bi)LSTM/GRU, (Bidirectional) Long Short-Term Memory/Gated Recurrent Unit; ASR, Automatic Speech Recognition

Paper	Decoding objective	Neural modality (subject)	Architecture	Methods compared against	Intermediate var.	Real-time
Movement						
Sussillo et al. [36]	Predict cursor movement on screen	Spikes (NHP)	Echostate network	KF	No	Yes
Sussillo et al. [75]	Predict cursor movement on screen, stitch across days	Spikes (NHP)	Multiplicative RNN	KF	No	Yes
Pandarinath et al. [43]	Predict reach kinematics, stitch across days	Spikes (NHP, Human)	Recurrent autoencoder + Lin. Reg.	Lin. Reg., GPFA + Lin. Reg.	No	No
Glaser et al. [39]	Predict reach kinematics	Spikes (NHP)	LSTM	WF, WC, KF, Naive Bayes, SVR, XGBoost, FC, RNN, Ensemble	No	No
Makin et al. [76]	Predict reach kinematics	Spikes (NHP)	Restricted boltzmann machine variant	WF, KF, Unscented KF	No	No
Ahmadi et al. [60]	Predict reach kinematics	LFP, Spikes (NHP)	LSTM	KF	No	No
Ahmadi et al. [65]	Predict reach kinematics	LFP (NHP)	TCN	LSTM	No	No
Ahmadi et al. [48]	Predict reach kinematics	Spikes (NHP)	QuasiRNN	WF, WC, KF, Unscented KF, RNN, GRU, LSTM	No	No
Park and Kim [77]	Predict reach kinematics	Spikes (NHP)	LSTM	KF	Speed, direction	No
Li et al. [78]	Predict forelimb reach location	Calcium imaging (Mouse)	CNN	None	No	No
Wang et al. [79]	Predict hindlimb kinematics	Spikes (NHP)	LSTM	WF, PLDS+WF, XGBoost, RNN,	No	No
Nakagome et al. [80]	Predict hindlimb kinematics	EEG (Human)	GRU	WF, Ridge Reg., Unscented KF, TCN, LSTM, Quasi RNN, CatBoost	No	No
Tseng et al. [81]	Predict reaching and hindlimb kinematics	Spikes (NHP)	Multilayer LSTM	WF, KF, Unscented KF, LSTM	No	No

Paper	Decoding objective	Neural modality (subject)	Architecture	Methods compared against	Intermediate var.	Real-time
Xie et al. [69]	Predict finger kinematics	ECoG (Human)	CNN+LSTM	Lin. Reg., Least Angle Reg., Random Forest, LSTM	No	No
Petrosuan et al. [70]	Predict finger kinematics	ECoG (Human)	CNN	WF	No	No
Naufel et al. [82]	Predict wrist EMG	Spikes (NHP)	LSTM	WF, WC	No	No
Farshchian et al. [83]	Predict wrist EMG, stitch across days	Spikes (NHP)	Recurrent autoencoder + adversarial domain adaptation network for alignment	CCA, KL Divergence minimization for alignment	No	No
Schwemmer et al. [84]	Classify wrist, index movements	Wide-band (Human)	LSTM+CNN	SVM	No	Yes
Skomrock et al. [85]	Classify hand, wrist, index movements	Wide-band (Human)	LSTM+CNN	SVM	No	Yes
Nurse et al. [86]	Classify hand squeeze	EEG (Human)	CNN	None	No	No
Pan et al. [87]	Classify hand gestures	ECoG (Human)	LSTM	Log. Reg., SVM, FC	No	No
Elango et al. [88]	Classify finger movements	ECoG (Human)	LSTM	LDA, HMM	No	No
Du et al. [89]	Classify finger movements	ECoG (Human)	FC + Multi-layer LSTM	SVM	No	No
Speech						
Livezey et al. [15]	Classify produced speech syllable	ECoG (Human)	FC	Log. Reg., Lin. SVM	Yes	
Sereshkeh et al. [90]	Classify yes/no/rest	EEG (Human)	FC	LDA, Lin. SVM, Poly. SVM, Naive Bayes, kNN	Yes	No
Wang et al. [91]	Classify produced phrase	MEG (Human)	FC	GMM	Yes	No
Dash et al. [92]	Classify imagined and produced phrase	MEG (Human)	CNN	FC	None	No
Wilson et al. [93]	Classify produced phonemes	LFP (Human), Unsorted spikes	BiGRU	Log. Reg.	None	No
Yang et al. [58]	Reconstruct perceived speech spectrogram	ECoG (Human)	FC	Lin. Reg.	Yes	No
Heelan et al. [94]	Reconstruct perceived speech/call spectrogram	Spikes (NHP)	LSTM	FC, RNN, GRU, KF, Weiner Filter, Weiner Cascade	Yes	No
Angrick et al. [37]	Speech reconstruction	ECoG (Human)	3dCNN + Wavenet	None	Spectrogram	No
Anumanchipalli et al. [95]	Produced speech synthesis	ECoG (Human)	BiLSTM + BiLSTM + Synthesizer	Ablation	Articulator kinematics	No
Sun et al. [96]	Speech recognition	ECoG (Human)	BiLSTM + CNN	LSTM + ASR, Ablation	Yes	No
Makin et al. [97]	Speech recognition	ECoG (Human)	CNN + BiLSTM	HMM, Ablation	Yes	No
Krishna et al. [98]	Speech recognition, Speech reconstruction	EEG (Human)	RNN, GAN	Ablation	None	No

Paper	Decoding objective	Neural modality (subject)	Architecture	Methods compared against	Intermediate var.	Real-time
Willett et al. [99]	Reconstruct text	Spikes (Human)	GRU	KF+HMM on moving cursor to letters	Imagined handwriting	Yes
Vision						
Qiao et al. [100]	Classify visual stimuli	fMRI (Human)	CNN feature selection + BiLSTM	Decision Tree, RF, AdaBoost, Lin. SVM, Kernel SVM, FC	Yes	No
Ellis and Michaelides [101]	Classify visual stimuli	Calcium imaging (Mouse)	CNN	Lin. SVM, FC	Yes	No
Güçlütürk et al. [38]	Reconstruct perceived faces	fMRI (Human)	Bayesian CNN + GAN	Bayesian Linear + GAN	No	No
Parthasarathy et al. [35]	Reconstruct images	Spikes (NHP)	Lin. Reg. + CNN Autoencoder	Low-fidelity image	No	
St-Yves and Naselaris [102]	Reconstruct images	fMRI (Human)	CNN+dAE+GAN	None	No	No
Wen et al. [103]	Reconstruct and classify images	fMRI (Human)	Lin. Reg. + CNN	None	CNN activations	No
Seeliger et al. [104]	Reconstruct images	fMRI (Human)	Lin. Reg. + GAN + CNN	None	GAN inputs	No
Shen et al. [105]	Reconstruct images	fMRI (Human)	Lin. Reg. + CNN	Ablation	CNN activations	No
Shen et al. [106]	Reconstruct images	fMRI (Human)	GAN + CNN	Ablation	None	No
VanRullen and Reddy [107]	Reconstructing perceived faces	fMRI (Human)	Lin. Reg. + VAE + GAN	Lin. Reg. + PCA	VAE inputs	No
Dado et al. [108]	Reconstruct perceived faces	fMRI (Human)	GAN	VAE + GAN, Eigenfaces	GAN inputs	No
Kim et al. [109]	Reconstruct images	Spikes (NHP)	Nonlin. Reg. + CNN Autoencoder	Low-fidelity image	No	

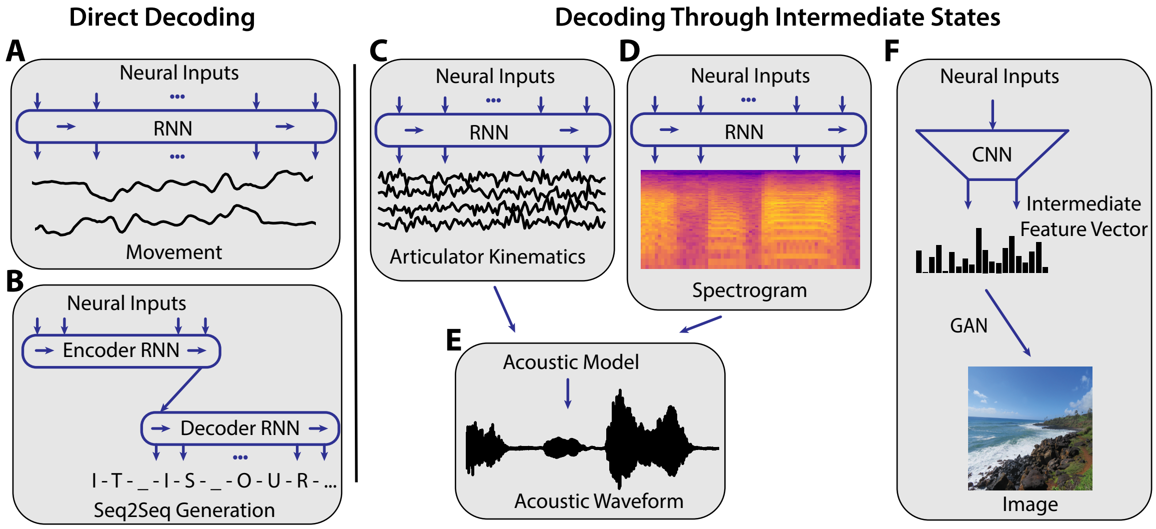

Figure 3: Architectures and outputs of neural decoding. We contrast examples of "direct decoding," in which the neural network decoder outputs the final desired variable, versus "decoding through intermediate states", in which the neural network predicts an intermediate variable that subsequently predicts the final desired variable. **A:** Sequential neural data is processed by RNNs that use past context to generate their output (or past and future in bidirectional RNNs). RNN outputs at each timestep can be mapped to behaviors, e.g., movements, measured concurrently, e.g., [36]. **B:** In a Seq2Seq-style RNN, as in Makin et al. [76], the final output of an encoder RNN is used as the input to a decoding RNN which produces a second sequence of potentially different length, such as the text representation of speech. **C:** RNNs can produce an intermediate state to be used in a second decoding step, such as articulator kinematics (movement of lips, tongue, etc.) [95]. **D:** Intermediate states can often be structured, such as a spectrogram [37]. **E:** Intermediate states can then be fed into an acoustic model such as Wavenet [122] or a speech synthesizer which produces acoustic waveforms [37, 95]. **F:** Temporal snapshot neural data, e.g., fMRI, can be processed by fully-connected networks or CNNs to produce intermediate feature vectors [38]. These feature vectors can be fed into generative image models, e.g., a GAN, to produce a more realistic looking image [38]

5 The outputs of decoding: behavior and perception

5.1 Movement

Some of the earliest uses of neural decoding were in the motor system [110]. Researchers have used neural activity from motor cortex to predict many different motor outputs, such as movement kinematics (e.g., position and velocity), muscle activity (EMG), and broad type of movement. Traditionally, this decoding has used methods (e.g., Kalman Filter or Wiener Filter) that assumed a linear mapping from neural activity to the motor output, which has led to many successes [111–115]. To improve the decoders, these methods were extended to allow specific nonlinearities (e.g., Unscented Kalman Filter, Point Process Filter, and Wiener Cascade [116–120]). Within the last decade, deep learning methods have become more common, frequently outperforming linear methods and their direct nonlinear extensions when compared (e.g., [39, 69, 81, 121]). Deep learning has shown to be a flexible tool for movement decoding, having been used to predict a wide range of movement variables from several different neural recording modalities (as catalogued in Table 1).

RNNs are by far the most common deep learning architecture for movement decoding. When predicting a continuous movement variable, there is generally a linear mapping from the RNN's output to the movement variable. When classifying movements, there is an additional softmax nonlinearity that determines the movement with the highest probability. From a deep learning perspective, given that this is a problem of converting one sequence (a temporal trace of neural activities) into another sequence (motor outputs), it would be expected that an RNN would be an appropriate architecture. Recurrent architectures also make sense from a scientific perspective: motor cortical activity has dynamics that are important for producing movements [123], plus movements themselves have dynamics.

LSTMs have generally been the most common and successful type of RNN for movement decoding [39, 60, 69, 77, 79, 81, 82, 87–89], although other standard types of RNN architectures (e.g., GRUs [80] and echostate networks [36]) have also proven successful. Additionally, researchers have found that stacking multiple layers of LSTMs [81, 89] can improve performance beyond a single LSTM [81]. LSTMs are likely successful because they are able to learn long-term dependencies better than a standard “vanilla” RNN [27].

A common goal of neural decoding of movement is to be able to create a usable brain computer interface for patients. While the majority of deep learning uses have been in offline scenarios (decoding after the neural recording), there are several successful examples of real-time uses of deep learning for movement decoding [36, 84, 85, 121]. The first use of deep learning for real-time movement decoding was in Sussillo et al. [36]. Monkeys with implanted electrode arrays were able to control the velocity of a cursor on a screen in real time via the use of an echostate network, which outperformed a Kalman filter. In a more recent example, in human patients with tetraplegia who had implanted electrode arrays, Schwemmer et al. [84] were able to classify planned movements of wrist extension, wrist flexion, index extension, and index flexion. This was done by inputting wide-band activity into an LSTM, followed by a CNN, followed by a fully connected layer for classification. After classifying the movement type, the authors then applied functional electrical stimulation to activate muscles according to this neural decoder, so that the patient was able to make these movements in real time.

While there has been great initial success, there are several challenges associated with using deep learning for real-time decoding for brain computer interfaces. One challenge is that the source of the recorded neural activity can change across days, for example due to slight movement of implanted electrodes. One approach that has dealt with this is the multiplicative RNN, an architecture that allows mappings from the neural input to the motor output to partially change across days [121]. Another approach that helps to utilize data across multiple days is that of Pandarinath et al. [43], which uses recurrent autoencoders to find a consistent low-dimensional dynamical model of the data that is shared across days. Incorporating this dynamical model leads to more accurate low-dimensional representations that are more predictive of movement kinematics. One other approach is that of [83], which uses adversarial domain adaptation networks in order to align neural recordings across days.

Another challenge of using deep learning for real-time neural decoding is computation time, as there is the need to make predictions through the deep learning architecture at very high temporal resolution. When using a less complicated echostate network, Sussillo et al. [36] were able to decode with less than 25 ms temporal resolution. However, when using a more complex architecture of LSTMs followed by CNNs, Schwemmer et al. [84] decoded at 100 ms resolution, slower than our perception. Relatedly, for linear methods that can be fit rapidly, researchers are able to adapt the neural decoder in real time to better match the subject’s intention (trying to get to a target) to improve performance [112, 115, 117, 120]. Developing similar approaches for deep learning based decoders is an exciting, unexplored area.

5.2 Speech

Vocal articulation of speech is a complex behavior that engages a large functional area of the brain to produce movements that have a high degree of articulatory temporal and spatial precision [124]. Its production is also a uniquely human ability which limits the recording modalities and neuroscientific interventions that can be used to study it. Due to the functional and temporal requirements of decoding speech, cortical surface electrical potentials recorded using ECoG is the typical recording modality used, although penetrating electrodes, MEG, EEG, and fNIRS are also used [90, 91, 125, 126]. When decoding from ECoG or EEG, researchers commonly use the signals’ high gamma amplitude [57], although some use more broad spectrotemporal features as well [57, 59, 127].

Many approaches to decoding speech from neural signals have used some combination of linear methods and shallow probabilistic models. Clustering, SVMs, LDA, linear regression, and probabilistic models have been used with spectrotemporal features of electrical potentials to decode vowel acoustics, speech articulator movements, phonemes, whole words, and semantic categories [57, 59, 125, 128–131].

Deep learning approaches to decoding speech from neural signals have emerged that can potentially learn nonlinear mappings (see Table 1). Some of these approaches have operated on temporally segmented neural data and have thus used fully connected neural network architectures. For example, spectrotemporal features derived from ECoG or EEG have been used to reconstruct perceived spectrograms, classify words or syllables, or classify entire phrases [15, 58, 90–93, 127]. These examples with temporally segmented neural data are useful for increasing understanding about neural representations, and as a step towards decoding natural speech.

Mapping directly from continuous, time-varying neural signals to speech is the goal of speech brain-computer interfaces [1, 132]. Both convolutional and recurrent networks are able to flexibly decode timeseries data and are often used for decoding naturalistic speech. Heelan et al. [94] reconstructed perceived speech audio from multi-unit spike counts from a non-human primate and found that LSTM-based networks outperformed other traditional and deep models. Speech represented as text does not have a simple one-to-one temporal alignment to regularly sampled neural signals. For this reason, speech-to-text decoding networks often use architectures and methods like sequence-to-sequence models or the connectionist temporal classification loss [24, 133], which are commonly used in machine translation or automated speech recognition applications. As such, several groups have decoded directly from neural signals to text during speech production or imagined handwriting using recurrent networks such as sequence-to-sequence models [96–99] (Fig. 3C).

For decoding intelligible acoustic speech, it is also common to split neural decoding into a more constrained neural-to-intermediate mapping, followed by a second stage that maps this intermediate format into an acoustic waveform using acoustic priors for speech based on deep learning or hand-engineered methods. For instance, high gamma features recorded using ECoG have been used to decode spectrograms which were fed into a WaveNet [122] deep network to produce an acoustic waveform [37]. As a specific example of a split decoding setup, Anumanchipalli et al. [95] trained a bidirectional LSTM to decode articulator kinematic features (movement of lips, tongue, etc.) from a combination of high gamma amplitude and a low frequency component. The second stage was a separate bidirectional LSTM which decodes acoustic features (mel-frequency cepstral coefficients, voicing, etc.) from the decoded articulatory features. Finally, these acoustic features were passed into a speech synthesizer. Compared to a RNN that skips the intermediate articulator kinematics stage, their two-stage method decoded perceptually improved speech acoustics. The second stages do not require invasive neural data for training and were trained on a larger second corpus.

Deep learning models have improved the accuracy of primarily offline speech decoding tasks. Many of the preprocessing and decoding methods reviewed here are done offline using acausal or high-latency deep learning models. Developing deep learning methods, software, and hardware for real-time speech decoding is important for clinical applications of brain computer interfaces [131, 134].

5.3 Vision

Similar to decoding acoustic speech, decoding visual stimuli from neural signals requires strong image priors due to the large variability of natural scenes and the relatively small bit-rate of neural recordings. Early attempts to reconstruct the full visual experience restricted decoding to simple images [135] or relied on a filterbank encoding model and a large set of natural images as a sampled prior [136]. Qiao et al. [100] solved the simpler task of classifying perceived object category using one CNN to select a small set of fMRI voxels which were fed into a second RNN for classification. Similarly, Ellis and Michaelides [101] classified among many visual scenes from calcium imaging data using feedforward or convolutional neural networks.

As mentioned in Section 2, deep generative image models, such as GANs, can produce realistic images. In addition, CNNs trained to classify large naturalistic image databases [137] (discriminative models) have been shown to encode a large amount of textural and semantic meaning in their activations [138], which can be used as an image prior. Due to the variety of ways that natural image priors can be created with deep networks, there exist neural decoding methods that combine different aspects of both generative and discriminative networks.

Given a deep generative model of images, a simpler neural decoder can be trained to map from neural data to the latent space of the model [38, 104, 107, 108], and the generative model can be used for image reconstruction. As an example, Seeliger et al. [104] trained a Convolutional Generative Adversarial Network (GAN) [33, 34] to generate grayscale images of objects given a noise vector as input. The parameters of the network were then frozen. Then, a linear regression model was trained to generate GAN input vectors from fMRI to optimize both the reconstruction of an image's individual pixels and higher order image features. Similarly, a linear stage or combined linear and deep learning reconstruction followed by a deep network that cleans-up the image has been used with retinal ganglion cell output [35, 109]. Generative models can also be trained to reconstruct images directly from fMRI responses on real data with data augmentation from a simulated encoding model [102].

Alternatively, generative and discriminative models can be used together. By leveraging a pretrained CNN, a simple neural decoder can be trained to map neural data to CNN activations that can then be passed into a convolutional image reconstruction model [103]. Additionally, the input image in a pretrained CNN can be optimized so that the CNN activations match predictions given by the fMRI responses [105]. Researchers have also used an end-to-end approach in which they train the generative part directly on neural data with both an adversarial loss and a pretrained CNN feature loss [106]. Along with acoustic speech, decoding naturalistic visual stimuli presents one of the best cases to study the use of data-driven priors derived from deep networks.

5.4 Other

While we have chosen to focus on a few decoding outputs that are prevalent in the literature, deep learning has been used for a myriad of neural decoding applications. For instance, RNNs such as LSTMs have been used to decode an animal's location [39, 56, 139, 140] and direction [141] from spiking activity in the hippocampus and head-direction cells, respectively. Deep networks have been used to decode what is being remembered in a working memory task from [142] and to predict illness [71–74, 143, 144] from human fMRI. Researchers have used LSTMs [145] and feedforward neural networks [146] to classify different classes of behaviors, using spiking activity in animals [146] and fNIRS or fMRI measurements in humans [16, 145]. LSTMs [147, 148] and CNNs [149] have been used to classify emotions from EEG signals. Feedforward neural networks have been used to determine the source of a subject's attention, using EEG in humans [150, 151] and spiking activity in monkeys

[152]. CNNs [62–64], along with LSTMs [64] have been used to predict a subject's stage of sleep from their EEG. For almost any behavioral signal that can be decoded, examples exists of applying deep learning.

6 Discussion

Deep learning is an attractive method for use in neural decoding because of its ability to learn complex, nonlinear transformations from data. In many of the examples above, deep networks can outperform linear or shallow methods even on relatively small datasets; however, examples exist where this is not the case, especially when using fMRI [153, 154] or fNIRS data [155]. Relatedly, there are many times in which using hand-engineered features can outperform an end-to-end neural network that will learn the features. This is more likely with limited amounts of data, and also when there is strong prior knowledge about the relevant features. One general machine learning approach to efficiently use limited data is transfer learning, in which a neural network trained in one scenario (typically with more data) is used a separate scenario. This has been used in neural decoding to more effectively train decoders for new subjects [88, 97] and for new predicted outputs [84]. As the capability to generate ever larger datasets develops with automated, long-term experimental setups for single animals [156] and large scale recordings across multiple animals [157], deep learning is well poised to take advantage of this flood of data. As dataset sizes increase, this will also allow more features to be learned through data-driven network training rather than being selected by-hand.

Although deep learning will inevitably improve decoding accuracy as neuroscientists collect larger datasets, extracting scientific knowledge from trained networks is still an area of active research. That is, can we understand the transformations deep networks are learning? In computer vision, layers that include spatial attention [158] and methods for performing feature attribution [159] have been developed to understand what parts of the input are important for prediction, although the latter are an active area of research [160]. These methods could be used to attribute what channels, neurons (e.g. of different genetically-defined cell types), or time-points are most salient for neural decoding [159]. Additionally, there are methods for understanding deep network representations in computer vision that examine the representations networks have learned across layers [161, 162]. Using these methods may help to understand the transformations that occur within neural decoders, however results may be sensitive to the decoder's architecture and not purely the data's structure. While deep learning interpretability methods are not commonly used on decoders trained on neural data, there are a few examples of networks that were built with interpretability in mind or were investigated after training [15, 16, 66, 67, 70, 142].

When interpreting neural decoders, it is often assumed that the decoder reveals the information contained in the brain about the decoded variable. It is important to note that this is only partially true when priors are being used for decoding [163], which is often the case when decoding a full image or acoustic speech. In these scenarios, the decoded outputs will be a function of both neural activity and the prior, so one cannot simply determine what information the brain has about the output.

The software used to create, train, and evaluate deep networks has been steadily developed and is now almost as easy to use as other standard machine learning methods. A wide range of cost functions, layer types, and parameter optimization algorithms are implemented and accessible in deep learning libraries such as PyTorch or TensorFlow [164, 165] and libraries in other programming languages. Like other machine learning methods, care must be taken to carefully cross-validate results as deep networks can easily overfit to the training data.

In addition to their use in neural decoding, deep learning has other prominent uses within neuroscience [166, 167]. Neural networks have a long history in neuroscience as models of neural processing [168, 169]. More recently, there has also been a surge of papers using deep networks as encoding models [12, 14, 75]. There has been a specific focus on using the representations learned by deep networks trained to perform behavioral tasks (e.g., image recognition) to predict neural responses in corresponding brain areas (e.g., across the visual hierarchy [170]). Combining these multiple complementary approaches is one promising approach to understanding neural computation.

Future directions

There use of deep learning for neural decoding has greatly increased within the last few years. Here, we highlight several open challenges and potential future directions for research.

- Increased use of deep learning for online decoding. This will be benefited by speed-ups in computing performance and reduced latency hardware [171, 172], online adaptation of neural network decoders to subjects' intentions [112, 115, 117], and robustness of decoders to signal changes across days [43, 75, 83].
- Keeping pace with the state-of-the-art in deep learning methods. Machine learning datasets are typically much larger than neural datasets. Do the architectural improvements in deep learning that are beneficial in other domains translate to neural datasets?

- Standardized benchmark datasets for better comparisons with strong baseline models and clear cross validation and evaluation standards [173–175].
- Increased interpretability of the inner-workings of neural decoders, either by specifically creating more interpretable architectures [66, 70], or by post-hoc analyses of the neural networks [15, 56].

Key Points

- We review many deep learning approaches which have been used to create more accurate and flexible neural decoders.
- Traditionally, many decoders have used hand-engineered features as inputs; deep learning tools can help to automatically learn relevant features from neural inputs.
- Pretrained deep learning models can be used as priors for complex decoding targets such as images or acoustic speech.
- We discuss directions for future research.

Acknowledgements

We would like to thank Ella Batty and Charles Frye for very helpful comments on this manuscript.

Funding

This work was supported by National Science Foundation NeuroNex Award DBI-1707398 [J.I.G.]; The Gatsby Foundation AT3708 [J.I.G.]; and LBNL Laboratory Directed Research and Development [J.A.L.].

References

[1] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[2] Xiang Zhang, Lina Yao, Xianzhi Wang, Jessica Monaghan, David Mcalpine, and Yu Zhang. A survey on deep learning based brain computer interface: Recent advances and new frontiers. *arXiv preprint arXiv:1905.04149*, 2019.

[3] Rodrigo Quian Quiroga, Lawrence H Snyder, Aaron P Batista, He Cui, and Richard A Andersen. Movement intention is better predicted than attention in the posterior parietal cortex. *Journal of Neuroscience*, 26 (13):3615–3620, 2006.

[4] Stephenie A Harrison and Frank Tong. Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635, 2009.

[5] Soumyadipta Acharya, Matthew S Fifer, Heather L Benz, Nathan E Crone, and Nitish V Thakor. Electrocorticographic amplitude predicts finger positions during slow grasping motions of the hand. *Journal of Neural Engineering*, 7(4):046002, 2010.

[6] Martin Weygandt, Carlo R Blecker, Axel Schäfer, Kerstin Hackmack, John-Dylan Haynes, Dieter Vaitl, Rudolf Stark, and Anne Schienle. fmri pattern recognition in obsessive–compulsive disorder. *Neuroimage*, 60(2):1186–1193, 2012.

[7] Erin L Rich and Jonathan D Wallis. Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience*, 19(7):973, 2016.

[8] Joshua I Glaser, Matthew G Perich, Pavan Ramkumar, Lee E Miller, and Konrad P Kording. Population coding of conditional probability distributions in dorsal premotor cortex. *Nature Communications*, 9(1): 1–14, 2018.

[9] Liberty S Hamilton, Erik Edwards, and Edward F Chang. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Current Biology*, 28(12):1860–1871, 2018.

[10] Nora Brackbill, Colleen Rhoades, Alexandra Kling, Nishal P Shah, Alexander Sher, Alan M Litke, and EJ Chichilnisky. Reconstruction of natural images from responses of primate retinal ganglion cells. *bioRxiv*, 2020.

[11] Barnaly Rashid and Vince Calhoun. Towards a brain-based predictome of mental illness. *Human Brain Mapping*, 2020.

[12] Lane McIntosh, Niru Maheswaranathan, Aran Nayebi, Surya Ganguli, and Stephen Baccus. Deep learning models of the retinal response to natural scenes. In *Advances in Neural Information Processing Systems*, pages 1369–1377, 2016.

[13] Tasha Nagamine and Nima Mesgarani. Understanding the representation and computation of multilayer perceptrons: A case study in speech recognition. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2564–2573. JMLR. org, 2017.

[14] Alexander JE Kell, Daniel LK Yamins, Erica N Shook, Sam V Norman-Haignere, and Josh H McDermott. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3):630–644, 2018.

[15] Jesse A Livezey, Kristofer E Bouchard, and Edward F Chang. Deep learning as a tool for neural data analysis: speech classification and cross-frequency coupling in human sensorimotor cortex. *PLoS Computational Biology*, 15(9):e1007091, 2019.

[16] Hojin Jang, Sergey M Plis, Vince D Calhoun, and Jong-Hwan Lee. Task-specific feature extraction and classification of fmri volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *NeuroImage*, 145:314–328, 2017.

[17] Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. *Nature Biotechnology*, 33(8):831–838, 2015.

[18] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. In *Advances in Neural Information Processing Systems*, pages 505–513, 2015.

[19] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Calogan: Simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Physical Review D*, 97(1):014021, 2018.

[20] Thorsten Kurth, Sean Treichler, Joshua Romero, Mayur Mudigonda, Nathan Luehr, Everett Phillips, Ankur Mahesh, Michael Matheson, Jack Deslippe, Massimiliano Fatica, et al. Exascale deep learning for climate analytics. In *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 649–660. IEEE, 2018.

[21] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

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

[23] 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.

[24] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112, 2014.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[26] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, Guoliang Chen, et al. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International Conference on Machine Learning*, pages 173–182, 2016.

[27] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[28] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[30] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[31] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10):1995, 1995.

[32] Dan Claudiu Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Twenty-second International Joint Conference on Artificial Intelligence*, 2011.

[33] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[34] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[35] Nikhil Parthasarathy, Eleanor Batty, William Falcon, Thomas Rutten, Mohit Rajpal, EJ Chichilnisky, and Liam Paninski. Neural networks for efficient bayesian decoding of natural images from retinal neurons. In *Advances in Neural Information Processing Systems*, pages 6434–6445, 2017.

[36] David Sussillo, Paul Nuyujukian, Joline M Fan, Jonathan C Kao, Sergey D Stavisky, Stephen Ryu, and Krishna Shenoy. A recurrent neural network for closed-loop intracortical brain–machine interface decoders. *Journal of Neural Engineering*, 9(2):026027, 2012.

[37] Miguel Angrick, Christian Herff, Emily Mugler, Matthew C Tate, Marc W Slutzky, Dean J Krusienski, and Tanja Schultz. Speech synthesis from ecog using densely connected 3d convolutional neural networks. *Journal of Neural Engineering*, 16(3):036019, 2019.

[38] Yağmur Güçlütürk, Umut Güçlü, Katja Seeliger, Sander Bosch, Rob van Lier, and Marcel AJ van Gerven. Reconstructing perceived faces from brain activations with deep adversarial neural decoding. In *Advances in Neural Information Processing Systems*, pages 4246–4257, 2017.

[39] Joshua I Glaser, Ari S Benjamin, Raeed H Chowdhury, Matthew G Perich, Lee E Miller, and Konrad P Kording. Machine learning for neural decoding. *Eneuro*, 7(4), 2020.

[40] Kristofer E. Bouchard and Edward F Chang. Human ecog speaking consonant-vowel syllables, 2019. URL `doi.org/10.6084/m9.figshare.c.4617263.v4`.

[41] Samaneh Kazemifar, Kathryn Y Manning, Nagalingam Rajakumar, Francisco A Gomez, Andrea Soddu, Michael J Borrie, Ravi S Menon, Robert Bartha, Alzheimer's Disease Neuroimaging Initiative, et al. Spontaneous low frequency bold signal variations from resting-state fmri are decreased in alzheimer disease. *PloS One*, 12(6), 2017.

[42] David Sussillo, Rafal Jozefowicz, LF Abbott, and Chethan Pandarinath. Lfads-latent factor analysis via dynamical systems. *arXiv preprint arXiv:1608.06315*, 2016.

[43] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15(10):805–815, 2018.

[44] JinHyung Lee, Catalin Mitelut, Hooshmand Shokri, Ian Kinsella, Nishchal Dethe, Shenghao Wu, Kevin Li, Eduardo B Reyes, Denis Turcu, Eleanor Batty, et al. Yass: Yet another spike sorter applied to large-scale multi-electrode array recordings in primate retina. *bioRxiv*, 2020.

[45] Melinda Rácz, Csaba Liber, Erik Németh, Richárd Fiáth, János Rokai, István Harmati, István Ulbert, and Gergely Márton. Spike detection and sorting with deep learning. *Journal of Neural Engineering*, 17 (1):016038, 2020.

[46] Melikasadat Emami, Mojtaba Sahraee-Ardakan, Parthe Pandit, Alyson K Fletcher, Sundeep Rangan, Michael Trumpis, Brinnae Bent, Chia-Han Chiang, and Jonathan Viventi. Low-rank nonlinear decoding of μ-ecog from the primary auditory cortex. *arXiv preprint arXiv:2005.05053*, 2020.

[47] Luke Y Prince and Blake A Richards. Inferring hierarchies of latent features in calcium imaging data. *NeurIPS 2019 Workshop Neuro AI*, 2019.

[48] Nur Ahmadi, Timothy Constandinou, and Christos-Savvas Bouganis. Robust and accurate decoding of hand kinematics from entire spiking activity using deep learning. *BioRxiv*, 2020.

[49] Tsai-Wen Chen, Trevor J Wardill, Yi Sun, Stefan R Pulver, Sabine L Renninger, Amy Baohan, Eric R Schreiter, Rex A Kerr, Michael B Orger, Vivek Jayaraman, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, 2013.

[50] Andrea Giovannucci, Johannes Friedrich, Pat Gunn, Jeremie Kalfon, Brandon L Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L Gauthier, Pengcheng Zhou, et al. Caiman an open source tool for scalable calcium imaging data analysis. *Elife*, 8:e38173, 2019.

[51] Joshua T Vogelstein, Adam M Packer, Timothy A Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–3704, 2010.

[52] Somayyeh Soltanian-Zadeh, Kaan Sahingur, Sarah Blau, Yiyang Gong, and Sina Farsiu. Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. *Proceedings of the National Academy of Sciences*, 116(17):8554–8563, 2019.

[53] Artur Speiser, Jinyao Yan, Evan W Archer, Lars Buesing, Srinivas C Turaga, and Jakob H Macke. Fast amortized inference of neural activity from calcium imaging data with variational autoencoders. In *Advances in Neural Information Processing Systems*, pages 4024–4034, 2017.

[54] György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13(6):407–420, 2012.

[55] Robert D Flint, Christian Ethier, Emily R Oby, Lee E Miller, and Marc W Slutzky. Local field potentials allow accurate decoding of muscle activity. *Journal of Neurophysiology*, 108(1):18–24, 2012.

[56] Markus Frey, Sander Tanni, Catherine Perrodin, Alice O'Leary, Matthias Nau, Jack Kelly, Andrea Banino, Christian F Doeller, and Caswell Barry. Deepinsight: a general framework for interpreting wide-band neural activity. *bioRxiv*, page 871848, 2019.

[57] Kristofer E Bouchard and Edward F Chang. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6782–6785. IEEE, 2014.

[58] Minda Yang, Sameer A Sheth, Catherine A Schevon, Guy M Mckhann Ii, and Nima Mesgarani. Speech reconstruction from human auditory cortex with deep neural networks. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[59] Emily M Mugler, James L Patton, Robert D Flint, Zachary A Wright, Stephan U Schuele, Joshua Rosenow, Jerry J Shih, Dean J Krusienski, and Marc W Slutzky. Direct classification of all american english phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering*, 11(3):035015, 2014.

[60] Nur Ahmadi, Timothy G Constandinou, and Christos-Savvas Bouganis. Decoding hand kinematics from local field potentials using long short-term memory (lstm) network. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 415–419. IEEE, 2019.

[61] Hosein M Golshan, Adam O Hebb, and Mohammad H Mahoor. Lfp-net: A deep learning framework to recognize human behavioral activities using brain stn-lfp signals. *Journal of Neuroscience Methods*, 335: 108621, 2020.

[62] Jialin Wang, Yanchun Zhang, Qinying Ma, Huihui Huang, and Xiaoyuan Hong. Deep learning for single-channel eeg signals sleep stage scoring based on frequency domain representation. In *International Conference on Health Information Science*, pages 121–133. Springer, 2019.

[63] Zeke Barger, Charles G Frye, Danqian Liu, Yang Dan, and Kristofer E Bouchard. Robust, automated sleep scoring by a compact neural network with distributional shift correction. *PloS One*, 14(12), 2019.

[64] Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, 2017.

[65] Nur Ahmadi, Timothy G Constandinou, and Christos-Savvas Bouganis. End-to-end hand kinematic decoding from lfps using temporal convolutional network. In *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2019.

[66] Yitong Li, Kafui Dzirasa, Lawrence Carin, David E Carlson, et al. Targeting eeg/lfp synchrony with neural nets. In *Advances in Neural Information Processing Systems*, pages 4620–4630, 2017.

[67] 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.

[68] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

[69] Ziqian Xie, Odelia Schwartz, and Abhishek Prasad. Decoding of finger trajectory from ecog using deep learning. *Journal of Neural Engineering*, 15(3):036009, 2018.

[70] Artur Petrosuan, Mikhail Lebedev, and Alexei Ossadtchi. Decoding neural signals with a compact and interpretable convolutional neural network. *bioRxiv*, 2020.

[71] Liang Zou, Jiannan Zheng, Chunyan Miao, Martin J Mckeown, and Z Jane Wang. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access*, 5: 23626–23636, 2017.

[72] Junghoe Kim, Vince D Calhoun, Eunsoo Shim, and Jong-Hwan Lee. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*, 124:127–146, 2016.

[73] Tetsuya Iidaka. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, 63:55–67, 2015.

[74] Dong Nie, Han Zhang, Ehsan Adeli, Luyan Liu, and Dinggang Shen. 3d deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 212–220. Springer, 2016.

[75] David Sussillo, Mark M Churchland, Matthew T Kaufman, and Krishna V Shenoy. A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, 18(7):1025–1033, 2015.

[76] Joseph G Makin, Joseph E O'Doherty, Mariana MB Cardoso, and Philip N Sabes. Superior arm-movement decoding from cortex with a new, unsupervised-learning algorithm. *Journal of Neural Engineering*, 15(2): 026010, 2018.

[77] Jisung Park and Sung-Phil Kim. Estimation of speed and direction of arm movements from m1 activity using a nonlinear neural decoder. In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–4. IEEE, 2019.

[78] Chunyue Li, Danny CW Chan, Xiaofeng Yang, Ya Ke, and Wing-Ho Yung. Prediction of forelimb reach results from motor cortex activities based on calcium imaging and deep learning. *Frontiers in Cellular Neuroscience*, 13:88, 2019.

[79] Yinong Wang, Wilson Truccolo, and David A Borton. Decoding hindlimb kinematics from primate motor cortex using long short-term memory recurrent neural networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1944–1947. IEEE, 2018.

[80] Sho Nakagome, Trieu Phat Luu, Yongtian He, Akshay Sujatha Ravindran, and Jose L Contreras-Vidal. An empirical comparison of neural networks and machine learning algorithms for eeg gait decoding. *Scientific Reports*, 10(1):1–17, 2020.

[81] Po-He Tseng, Núria Armengol Urpi, Mikhail Lebedev, and Miguel Nicolelis. Decoding movements from cortical ensemble activity using a long short-term memory recurrent network. *Neural Computation*, 31(6): 1085–1113, 2019.

[82] Stephanie Naufel, Joshua I Glaser, Konrad P Kording, Eric J Perreault, and Lee E Miller. A muscle-activity-dependent gain between motor cortex and emg. *Journal of Neurophysiology*, 121(1):61–73, 2019.

[83] Ali Farshchian, Juan A Gallego, Joseph P Cohen, Yoshua Bengio, Lee E Miller, and Sara A Solla. Adversarial domain adaptation for stable brain-machine interfaces. *arXiv preprint arXiv:1810.00045*, 2018.

[84] Michael A Schwemmer, Nicholas D Skomrock, Per B Sederberg, Jordyn E Ting, Gaurav Sharma, Marcia A Bockbrader, and David A Friedenberg. Meeting brain–computer interface user performance expectations using a deep neural network decoding framework. *Nature Medicine*, 24(11):1669–1676, 2018.

[85] Nicholas D Skomrock, Michael A Schwemmer, Jordyn E Ting, Hemang R Trivedi, Gaurav Sharma, Marcia A Bockbrader, and David A Friedenberg. A characterization of brain-computer interface performance trade-offs using support vector machines and deep neural networks to decode movement intent. *Frontiers in Neuroscience*, 12:763, 2018.

[86] Ewan Nurse, Benjamin S Mashford, Antonio Jimeno Yepes, Isabell Kiral-Kornek, Stefan Harrer, and Dean R Freestone. Decoding eeg and lfp signals using deep learning: heading truenorth. In *Proceedings of the ACM International Conference on Computing Frontiers*, pages 259–266, 2016.

[87] Gang Pan, Jia-Jun Li, Yu Qi, Hang Yu, Jun-Ming Zhu, Xiao-Xiang Zheng, Yue-Ming Wang, and Shao-Min Zhang. Rapid decoding of hand gestures in electrocorticography using recurrent neural networks. *Frontiers in Neuroscience*, 12:555, 2018.

[88] Venkatesh Elango, Aashish N Patel, Kai J Miller, and Vikash Gilja. Sequence transfer learning for neural decoding. *bioRxiv*, page 210732, 2017.

[89] Anming Du, Shuqin Yang, Weijia Liu, and Haiping Huang. Decoding ecog signal with deep learning model based on lstm. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 0430–0435. IEEE, 2018.

[90] Alborz Rezazadeh Sereshkeh, Robert Trott, Aurélien Bricout, and Tom Chau. Eeg classification of covert speech using regularized neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(12):2292–2300, 2017.

[91] Jun Wang, Myungjong Kim, Angel W Hernandez-Mulero, Daragh Heitzman, and Paul Ferrari. Towards decoding speech production from single-trial magnetoencephalography (meg) signals. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3036–3040. IEEE, 2017.

[92] Debadatta Dash, Paul Ferrari, and Jun Wang. Decoding imagined and spoken phrases from non-invasive neural (meg) signals. *Frontiers in Neuroscience*, 2020.

[93] Guy H Wilson, Sergey D Stavisky, Francis R Willett, Donald T Avansino, Jessica N Kelemen, Leigh R Hochberg, Jaimie M Henderson, Shaul Druckmann, and Krishna V Shenoy. Decoding spoken english phonemes from intracortical electrode arrays in dorsal precentral gyrus. *bioRxiv*, 2020.

[94] Christopher Heelan, Jihun Lee, Ronan O'Shea, Laurie Lynch, David M Brandman, Wilson Truccolo, and Arto V Nurmikko. Decoding speech from spike-based neural population recordings in secondary auditory cortex of non-human primates. *Communications Biology*, 2(1):1–12, 2019.

[95] Gopala K Anumanchipalli, Josh Chartier, and Edward F Chang. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493, 2019.

[96] Pengfei Sun, Gopala K Anumanchipalli, and Edward F Chang. Brain2char: A deep architecture for decoding text from brain recordings. *arXiv preprint arXiv:1909.01401*, 2019.

[97] Joseph G Makin, David A Moses, and Edward F Chang. Machine translation of cortical activity to text with an encoder–decoder framework. Technical report, Nature Publishing Group, 2020.

[98] Gautam Krishna, Yan Han, Co Tran, Mason Carnahan, and Ahmed H Tewfik. State-of-the-art speech recognition using eeg and towards decoding of speech spectrum from eeg. *arXiv preprint arXiv:1908.05743*, 2019.

[99] Francis R Willett, Donald T Avansino, Leigh R Hochberg, Jaimie M Henderson, and Krishna V Shenoy. High-performance brain-to-text communication via imagined handwriting. *bioRxiv*, 2020.

[100] Kai Qiao, Jian Chen, Linyuan Wang, Chi Zhang, Lei Zeng, Li Tong, and Bin Yan. Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Frontiers in Neuroscience*, 13, 2019.

[101] Randall Jordan Ellis and Michael Michaelides. High-accuracy decoding of complex visual scenes from neuronal calcium responses. *BioRxiv*, page 271296, 2018.

[102] Ghislain St-Yves and Thomas Naselaris. Generative adversarial networks conditioned on brain activity reconstruct seen images. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1054–1061. IEEE, 2018.

[103] Haiguang Wen, Junxing Shi, Yizhen Zhang, Kun-Han Lu, Jiayue Cao, and Zhongming Liu. Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral Cortex*, 28(12):4136–4160, 2018.

[104] Katja Seeliger, Umut Güçlü, Luca Ambrogioni, Yagmur Güçlütürk, and Marcel AJ van Gerven. Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage*, 181:775–785, 2018.

[105] Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani. Deep image reconstruction from human brain activity. *PLoS Computational Biology*, 15(1):e1006633, 2019.

[106] Guohua Shen, Kshitij Dwivedi, Kei Majima, Tomoyasu Horikawa, and Yukiyasu Kamitani. End-to-end deep image reconstruction from human brain activity. *Frontiers in Computational Neuroscience*, 13, 2019.

[107] Rufin VanRullen and Leila Reddy. Reconstructing faces from fmri patterns using deep generative neural networks. *Communications Biology*, 2(1):1–10, 2019.

[108] Thirza Dado, Yagmur Gucluturk, Luca Ambrogioni, Gabrielle Ras, Sander Erik Bosch, Marcel van Gerven, and Umut Guclu. Hyperrealistic neural decoding: Linear reconstruction of face stimuli from fmri measurements via the gan latent space. *bioRxiv*, 2020.

[109] Young Joon Kim, Nora Brackbill, Ella Batty, JinHyung Lee, Catalin Mitelut, William Tong, EJ Chichilnisky, and Liam Paninski. Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. *bioRxiv*, 2020.

[110] Apostolos P Georgopoulos, Roberto Caminiti, John F Kalaska, and Joseph T Massey. Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations. *Experimental Brain Research*, 49(Suppl. 7):327–336, 1983.

[111] Wei Wu, Michael J Black, Yun Gao, M Serruya, A Shaikhouni, JP Donoghue, and Elie Bienenstock. Neural decoding of cursor motion using a kalman filter. In *Advances in Neural Information Processing Systems*, pages 133–140, 2003.

[112] Vikash Gilja, Paul Nuyujukian, Cindy A Chestek, John P Cunningham, M Yu Byron, Joline M Fan, Mark M Churchland, Matthew T Kaufman, Jonathan C Kao, Stephen I Ryu, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, 15(12):1752, 2012.

[113] Mijail D Serruya, Nicholas G Hatsopoulos, Liam Paninski, Matthew R Fellows, and John P Donoghue. Instant neural control of a movement signal. *Nature*, 416(6877):141–142, 2002.

[114] Jose M Carmena, Mikhail A Lebedev, Roy E Crist, Joseph E O'Doherty, David M Santucci, Dragan F Dimitrov, Parag G Patil, Craig S Henriquez, and Miguel AL Nicolelis. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology*, 1(2), 2003.

[115] Amy L Orsborn, Helene G Moorman, Simon A Overduin, Maryam M Shanechi, Dragan F Dimitrov, and Jose M Carmena. Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. *Neuron*, 82(6):1380–1393, 2014.

[116] Zheng Li, Joseph E O'Doherty, Timothy L Hanson, Mikhail A Lebedev, Craig S Henriquez, and Miguel AL Nicolelis. Unscented kalman filter for brain-machine interfaces. *PloS One*, 4(7), 2009.

[117] Trieu Phat Luu, Yongtian He, Samuel Brown, Sho Nakagome, and Jose L Contreras-Vidal. Gait adaptation to visual kinematic perturbations using a real-time closed-loop brain–computer interface to a virtual reality avatar. *Journal of Neural Engineering*, 13(3):036006, 2016.

[118] Eric A Pohlmeyer, Sara A Solla, Eric J Perreault, and Lee E Miller. Prediction of upper limb muscle activity from motor cortical discharge during reaching. *Journal of Neural Engineering*, 4(4):369, 2007.

[119] Christian Ethier, Emily R Oby, Matthew J Bauman, and Lee E Miller. Restoration of grasp following paralysis through brain-controlled stimulation of muscles. *Nature*, 485(7398):368–371, 2012.

[120] Maryam M Shanechi, Amy L Orsborn, and Jose M Carmena. Robust brain-machine interface design using optimal feedback control modeling and adaptive point process filtering. *PLoS Computational Biology*, 12 (4):e1004730, 2016.

[121] David Sussillo, Sergey D Stavisky, Jonathan C Kao, Stephen I Ryu, and Krishna V Shenoy. Making brain–machine interfaces robust to future neural variability. *Nature Communications*, 7:13749, 2016.

[122] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[123] Krishna V Shenoy, Maneesh Sahani, and Mark M Churchland. Cortical control of arm movements: a dynamical systems perspective. *Annual Review of Neuroscience*, 36:337–359, 2013.

[124] Kristofer E Bouchard, Nima Mesgarani, Keith Johnson, and Edward F Chang. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327, 2013.

[125] Alexander M Chan, Eric Halgren, Ksenija Marinkovic, and Sydney S Cash. Decoding word and category-specific spatiotemporal representations from meg and eeg. *Neuroimage*, 54(4):3028–3039, 2011.

[126] Christian Herff and Tanja Schultz. Automatic speech recognition from neural signals: a focused review. *Frontiers in Neuroscience*, 10:429, 2016.

[127] Hassan Akbari, Bahar Khalighinejad, Jose L Herrero, Ashesh D Mehta, and Nima Mesgarani. Towards reconstructing intelligible speech from the human auditory cortex. *Scientific Reports*, 9(1):1–12, 2019.

[128] David F Conant, Kristofer E Bouchard, Matthew K Leonard, and Edward F Chang. Human sensorimotor cortex control of directly measured vocal tract movements during vowel production. *Journal of Neuroscience*, 38(12):2955–2966, 2018.

[129] Spencer Kellis, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger. Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of Neural Engineering*, 7(5):056007, 2010.

[130] Christian Herff, Dominic Heger, Adriana De Pesters, Dominic Telaar, Peter Brunner, Gerwin Schalk, and Tanja Schultz. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience*, 9:217, 2015.

[131] Frank H Guenther, Jonathan S Brumberg, E Joseph Wright, Alfonso Nieto-Castanon, Jason A Tourville, Mikhail Panko, Robert Law, Steven A Siebert, Jess L Bartels, Dinal S Andreasen, et al. A wireless brain-machine interface for real-time speech synthesis. *PloS One*, 4(12), 2009.

[132] Tanja Schultz, Michael Wand, Thomas Hueber, Dean J Krusienski, Christian Herff, and Jonathan S Brumberg. Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(12):2257–2271, 2017.

[133] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine learning*, pages 369–376, 2006.

[134] David A Moses, Matthew K Leonard, Joseph G Makin, and Edward F Chang. Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, 10(1):1–14, 2019.

[135] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008.

[136] Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19):1641–1646, 2011.

[137] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009.

[138] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

[139] Ardi Tampuu, Tambet Matiisen, H Freyja Ólafsdóttir, Caswell Barry, and Raul Vicente. Efficient neural decoding of self-location with a deep recurrent network. *PLoS Computational Biology*, 15(2):e1006822, 2019.

[140] Mohammad R Rezaei, Anna K Gillespie, Jennifer A Guidera, Behzad Nazari, Saeid Sadri, Loren M Frank, Uri T Eden, and Ali Yousefi. A comparison study of point-process filter and deep learning performance in estimating rat position using an ensemble of place cells. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4732–4735. IEEE, 2018.

[141] Zishen Xu, Wei Wu, Shawn S Winter, Max L Mehlman, William N Butler, Christine M Simmons, Ryan E Harvey, Laura E Berkowitz, Yang Chen, Jeffrey S Taube, et al. A comparison of neural decoding methods and population coding across thalamo-cortical head direction cells. *Frontiers in Neural Circuits*, 13, 2019.

[142] Hongming Li and Yong Fan. Interpretable, highly accurate brain decoding of subtly distinct brain states from functional mri using intrinsic functional networks and long short-term memory recurrent neural networks. *NeuroImage*, 202:116059, 2019.

[143] Sergey M Plis, Devon R Hjelm, Ruslan Salakhutdinov, Elena A Allen, Henry J Bockholt, Jeffrey D Long, Hans J Johnson, Jane S Paulsen, Jessica A Turner, and Vince D Calhoun. Deep learning for neuroimaging: a validation study. *Frontiers in Neuroscience*, 8:229, 2014.

[144] Shaoqiang Han, Wei Huang, Yan Zhang, Jingping Zhao, and Huafu Chen. Recognition of early-onset schizophrenia using deep-learning method. In *Applied Informatics*, volume 4, pages 1–6. SpringerOpen, 2017.

[145] So-Hyeon Yoo, Seong-Woo Woo, and Zafar Amad. Classification of three categories from prefrontal cortex using lstm networks: fnirs study. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, pages 1141–1146. IEEE, 2018.

[146] Eleanor Batty, Matthew Whiteway, Shreya Saxena, Dan Biderman, Taiga Abe, Simon Musall, Winthrop Gillis, Jeffrey Markowitz, Anne Churchland, John P Cunningham, et al. Behavenet: nonlinear embedding and bayesian neural decoding of behavioral videos. In *Advances in Neural Information Processing Systems*, pages 15680–15691, 2019.

[147] Simon M Hofmann, Felix Klotzsche, Alberto Mariola, Vadim V Nikulin, Arno Villringer, and Michael Gaebler. Decoding subjective emotional arousal during a naturalistic vr experience from eeg using lstms. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 128–131. IEEE, 2018.

[148] Anumit Garg, Ashna Kapoor, Anterpreet Kaur Bedi, and Ramesh K Sunkaria. Merged lstm model for emotion classification using eeg signals. In *2019 International Conference on Data Science and Engineering (ICDSE)*, pages 139–143. IEEE, 2019.

[149] Samarth Tripathi, Shrinivas Acharya, Ranti Dev Sharma, Sudhanshu Mittal, and Samit Bhattacharya. Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In *Twenty-Ninth IAAI Conference*, 2017.

[150] Gregory Ciccarelli, Michael Nolan, Joseph Perricone, Paul T Calamia, Stephanie Haro, James O'Sullivan, Nima Mesgarani, Thomas F Quatieri, and Christopher J Smalt. Comparison of two-talker attention decoding from eeg with nonlinear neural networks and linear methods. *Scientific Reports*, 9(1):1–10, 2019.

[151] Tobias de Taillez, Birger Kollmeier, and Bernd T Meyer. Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech. *European Journal of Neuroscience*, 2017.

[152] Elaine Astrand, Pierre Enel, Guilhem Ibos, Peter Ford Dominey, Pierre Baraduc, and Suliann Ben Hamed. Comparison of classifiers for decoding sensory and cognitive information from prefrontal neuronal populations. *PloS One*, 9(1), 2014.

[153] Marc-Andre Schulz, Thomas Yeo, Joshua Vogelstein, Janaina Mourao-Miranada, Jakob Kather, Konrad Kording, Blake A Richards, and Danilo Bzdok. Deep learning for brains?: Different linear and nonlinear scaling in uk biobank brain images vs. machine-learning datasets. *bioRxiv*, page 757054, 2019.

[154] Rajat Mani Thomas, Selene Gallo, Leonardo Cerliani, Paul Zhutovsky, Ahmed El-Gazzar, and Guido van Wingen. Classifying autism spectrum disorder using the temporal statistics of resting-state functional mri data with 3d convolutional neural networks. *Frontiers in Psychiatry*, 11:440, 2020.

[155] Johannes Hennrich, Christian Herff, Dominic Heger, and Tanja Schultz. Investigating deep learning for fnirs based bci. In *2015 37th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2844–2847. IEEE, 2015.

[156] Ashesh K Dhawale, Rajesh Poddar, Steffen BE Wolff, Valentin A Normand, Evi Kopelowitz, and Bence P Ölveczky. Automated long-term recording and analysis of neural activity in behaving animals. *Elife*, 6: e27702, 2017.

[157] Allen Brain Observatory. Available at: http://observatory.brain-map.org/visualcoding, 2016.

[158] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*, pages 2048–2057, 2015.

[159] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3319–3328. JMLR. org, 2017.

[160] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In *Advances in Neural Information Processing Systems*, pages 9505–9515, 2018.

[161] Chris Olah, Arvind Satyanarayan, Ian Johnson, Shan Carter, Ludwig Schubert, Katherine Ye, and Alexander Mordvintsev. The building blocks of interpretability. *Distill*, 3(3):e10, 2018.

[162] The OpenAI microscope. `microscope.openai.com/models`, 2020. Accessed: 2020-05-12.

[163] Nikolaus Kriegeskorte and Pamela K Douglas. Interpreting encoding and decoding models. *Current Opinion in Neurobiology*, 55:167–179, 2019.

[164] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, 2019.

[165] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[166] Tim Christian Kietzmann, Patrick McClure, and Nikolaus Kriegeskorte. Deep neural networks in computational neuroscience. *BioRxiv*, page 133504, 2018.

[167] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11):1761–1770, 2019.

[168] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[169] David Zipser and Richard A Andersen. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158):679–684, 1988.

[170] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356, 2016.

[171] Javier Duarte, Song Han, Philip Harris, Sergo Jindariani, Edward Kreinar, Benjamin Kreis, Jennifer Ngadiuba, Maurizio Pierini, Ryan Rivera, Nhan Tran, et al. Fast inference of deep neural networks in fpgas for particle physics. *Journal of Instrumentation*, 13(07):P07027, 2018.

[172] Stylianos I Venieris and Christos-Savvas Bouganis. Latency-driven design for fpga-based convolutional neural networks. In *2017 27th International Conference on Field Programmable Logic and Applications (FPL)*, pages 1–8. IEEE, 2017.

[173] Jeffery L Teeters, Keith Godfrey, Rob Young, Chinh Dang, Claudia Friedsam, Barry Wark, Hiroki Asari, Simon Peron, Nuo Li, Adrien Peyrache, et al. Neurodata without borders: creating a common data format for neurophysiology. *Neuron*, 88(4):629–634, 2015.

[174] Ben Walker and Konrad Kording. The database for reaching experiments and models. *PloS One*, 8(11): e78747, 2013.

[175] Russell A Poldrack, Deanna M Barch, Jason Mitchell, Tor Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham. Toward open sharing of task-based fmri data: the openfmri project. *Frontiers in Neuroinformatics*, 7:12, 2013.