2021 International Wireless Communications and Mobile Computing (IWCMC) | 978-1-7281-8616-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWCMC51323.2021.9498710

A Survey: Challenges and Future Research Directions of fMRI Based Brain Activity Decoding Model

Xiaomu Cheng, Huansheng Ning, Bing Du*
School of computer and communication engineering, University of science and technology Beijing
s20190669@xs.ustb.edu.cn, ninghuansheng@ustb.edu.cn, dubing@ustb.edu.cn

***Abstract*—Researchers have used deep neural networks (DNN) to reconstruct visual stimuli from human brain activities. In particular, convolution neural networks (CNN) processing functional magnetic resonance imaging (fMRI) signals can reconstruct visual stimulation in brain activity. Although visual stimulation has been successfully reconstructed in brain activity, the research on visual stimulation reconstruction is still in initial stage. The decoding model of brain activity based on fMRI is facing three challenges: the mapping ability of the decoding model; the limited paired data of visual stimulation and brain activity; and the presence of noise in fMRI signals. This paper reviews the potential solutions to the above difficulties and the future development directions.**

***Index Terms*—Brain Activity Decoding, Visual Reconstruction, functional Magnetic Resonance Imaging, Deep Learning**

I. INTRODUCTION

From the perspective of neuroscience and neuroimaging, functional magnetic resonance imaging can be used to decode human perception and semantic information of the cerebral cortex in a non-invasive manner [1]. And the researchers have decoded visual stimuli related to the activity of human brain neurons from fMRI data successfully [2], [3], [4], [5], [6].

In recent years, the application of deep learning methods in the field of computer vision, significant achievements have been made in image restoration and image super-resolution. And the structure of deep neural networks is similar to the feedforward of the human visual system [7], it is not surprising that DNN is used to decode the visual stimuli of brain activity [8], [9]. For deep neural networks, especially convolutional neural networks, researchers use pre-trained CNN to extract features from fMRI data in the convolutional layer [10], [11], [12]. Furthermore, the graph convolutional networks (GCN) can consider the topological structure of the functional areas of the brain to predict the cognitive state of the brain [13], [14]. In the case of limited amount of fMRI data and image paired data, the GCN-based decoding model can also provide an automated tool for marking the cognitive state of the brain [15]. Although previous studies on decoding visual stimuli in brain activity have made great achievements in classification and recognition, the performance of image reconstruction needs to be improved [16].

The challenges to brain decoding can be summarized in three dimensions: the limited ability of model mapping between brain activity and visual stimuli; there is not a large amount of matching data between visual stimuli and brain activity; fMRI signals are mixed with noise [16]. The following sections will propose potential solutions and future development directions for the challenges faced.

II. MODEL MAPPING CAPABILITY

A. Multi-Voxel Pattern Analysis and Deep Learning

In recent years, the combination of multi-voxel pattern analysis (MVPA) and deep learning is a popular method to identify brain states. By using deep neural networks to decode fMRI signals that record brain activity, the performance of brain decoding has been further improved [9], [15]. Although a decoding model based on MVPA has been proposed, the multi-voxel pattern analysis decoding model has poor interpretability, especially when the decoding model uses linear kernels [17]. And this technique is susceptible to image artifacts such as eye movement and cardiopulmonary artifacts [17]. Also, the speed of neuron vascular coupling, the sensitivity of BOLD activity and the signal-to-noise ratio of fMRI signals should also be considered. The efficiency of the algorithm and the processing speed of the hardware should not only be pursued, but should correspond to the blood coupling delay in the brain [18]. Although the existing deep learning-based decoding model has achieved satisfactory results [19]. In order to obtain a higher-precision decoding model, there are still many challenges in using deep learning to reconstruct the corresponding visual stimuli from fMRI data.

B. Region of Interest and Feature Selection

The sample size of fMRI signals and image pairing is small, and the dimensionality of fMRI signals is higher. When the model is trained with limited high-dimensional data samples, it is easy to produce the curse of dimensionality [17]. And traditional methods are easy to overfit on small datasets [19]. The efficiency of deep learning-based models depends on the number and reliability of training samples. A large number of neural activities and corresponding types of images are recorded. The quality and types of image reconstruction may be improved [20], [21]. However, the running time of the experiment should be proportional to the efficiency. It is particularly important to select the key features that contribute the most to the image reconstruction, so it is necessary to further improve the feature extraction ability of the decoding model for neuroimaging data [9], [21]. It can learn from the visual attention [22] and axiomatic attribution [23] methods proposed in computer vision, used to determine which voxels of neurons contribute the most to decoding visual stimuli.

In addition, the connection structure of the brain network of human cognition has become one of the important goals of neuroscience research [15]. The current decoding of brain activity is usually limited to specific cognitive areas that humans understand, and it takes a relatively long time to collect and record fMRI signals of brain activity [15]. Moreover, most of the current deep learning-based research cannot simultaneously consider the functional dependence and time dynamics between different regions of the brain [13]. In order to use the dependency between the regions of the brain to decode the brain, [13], [14], [15] have explored the use of GCN to predict or annotate the cognitive state of the brain. Especially based on the Spatio-Temporal Graph Convolution Networks (ST-GCN) model, the representation extracted from the fMRI signal can not only represent the temporal dynamic information of brain activity but also the functional dependence between brain regions, and has achieved success in the field of computer vision [13]. This method of integrating the importance of the edge of the map in the context of the spatio-temporal map may have potential effects on the development of neuroscience [13].

C. *Unsupervised Learning and Prior Knowledge*

To fully learn abstract representations of brain activity in an unsupervised way should be fully studied in the future [24]. Researchers' exploration of unsupervised learning methods led to the emergence of bidirectional generative models. For example, variational auto-encoder (VAE) is an unsupervised learning model. However, in the design process of the corresponding computing components of the VAE, the encoder and the decoder are not related, but in the activities of the cerebral cortex, the feedforward and feedback processes are related [24]. In addition, VAE does not have the ability to process dynamics and loops, but video information can be transmitted in time and space [24]. Moreover, human brain activity is dynamic, and reconstructing dynamic features from brain activity is a huge challenge [16].

In addition, some researchers use a large number of image priors to reconstruct visual stimuli [25]. But when there is a prior condition to decode the brain activity, the decoder output is a function of brain neuron activity and prior knowledge, so it is impossible to specifically determine which information of the brain is decoded [9]. Exploring the alternative prior knowledge of brain decoding problems still requires constant exploration by researchers [26]. In recent years, an encoding model based on deep learning has also emerged, which trains deep neural networks to perform some tasks to learn representations that can predict neuronal activity [27]. Specifically, the encoding model uses visual stimuli to predict the neural response of the brain and serves as a prior for the decoding model. The advancement of encode technology has many important practical applications for brain enhancement communication, machine and computer direct brain control, and disease state monitoring and diagnosis [26]. This method of complementing encoding and decoding models is a meaningful research direction [9].

III. LIMITED PAIRED FMRI AND IMAGE DATE

A. *Few-shot Learning*

Due to the high cost of fMRI research and the complicated research process, the collection of paired fMRI signals and image samples is also a difficult problem, so the number of pairs of fMRI signals and images is small [17], [20], [28]. Also inspired by the field of computer vision, [20] proposed few-shot learning for the decoding of brain activity. The experimental report proves that this kind of few-shot method is promising in solving the data problem of neural influence [20]. There are currently three main ways to learn few-shot: representation-based paradigm, initialization-based paradigm, and illusion-based paradigm [20].

Representation-based paradigm: This method aims to learn the representation of fMRI signals. This method regards the first layer of the neural network as a feature extractor and the last layer as a classifier. A large amount of training data is used to train the neural network to extract relevant hidden representations and complete the training of the classifier. Later, when processing small sample data, the classifier extracts a small amount of data characterization according to the feature extractor to complete the classification of the new data [20].

Representation based on initialization: This method is also called meta-learning, and the idea of meta-learning is to learn how to learn. This method aims to learn good initialization parameters so that the model can cope with various new datasets. In the process of meta-learning, the previous neural network can be understood as a low-level neural network, and the meta-learner is used to optimize the weights of the low-level neural network. The meta-learner inputs a list of samples and their corresponding labels. When training the meta-learner, the meta-loss (the error value of the prediction and the target label) can be used to measure the performance of the meta-learner on the target task. Then, another meta-learner is needed to update the weight of the current meta-learner [20].

Paradigm based on illusion: This method is to perform a series of deformation operations such as rotation or combination of samples in the original datasets to increase training examples [20].

B. *Transfer Learning*

When the amount of data is limited and the prior knowledge is sufficient, sometimes the functions designed by hand are better than the neural networks model learned from the data [9]. At present, the model based on deep learning is ready. As the amount of data in the fMRI datasets continues to increase, the future direction is not to manually design functions, but to learn more functions based on data driving [9]. In the field of neuroimaging, there is always a lack of datasets with large enough samples for specific experiments [29]. Most of the current transfer learning is to learn the data representation of the image in ImageNet, and then build a model to adjust the medical image [30]. Although transfer learning can effectively make up for the shortcomings of insufficient training data, natural images and medical images are still different in nature [29]. For example, Gabor filters are often used for edge detection of natural images, but have never been used in medical images [29].

More and more studies have shown that the human cognitive system is a function of multiple functional areas of the brain [31]. [29] used a large number of different experimental tasks and medical imaging datasets of experimental environments for training based on graph convolutional network models. The experimental report shows that the dynamics of the brain are transferable between different brain regions and different cognitive domains, and even between different scanning sequences. Through fine-tuning, on the basis of preserving the low-level representation of brain dynamics, learn more about the high-level representation of brain functional areas [15]. [15] proved that transfer learning can not only improve decoding performance, but also shows a potential role in neuroimaging.

C. Graph Convolutional Networks

If CNN is not trained, no effective features can be obtained at all. Even if GCN is not trained, it completely uses randomly initialized parameters, and the features extracted by GCN are very effective [32]. If labeling information is given, the effect of GCN will be even better [32]. Compared with other classifiers, the graph convolutional networks has better performance on a limited dataset [20]. The fMRI signals can represent the spatial structure of brain activity, and the graph neural networks (GNN) can take the connectivity of the brain into account to decode brain activity, which has the potential to solve the problem of limited data [15], [20], [33], [34].

IV. THE EFFECT OF FMRI NOISE

A. Hemodynamic Delay

The spatial resolution of functional magnetic resonance is very high, but its time resolution is relatively limited. It can only collect the average activity level in about two seconds, and there is a certain delay in the detection of neural activity [35]. The fMRI signals contains the position information in the brain voxels, but due to its limited time resolution, sometimes the time series can not be used to decode brain activity [9]. Because of the neurovascular coupling, the fMRI response is after the neurological response [24]. Therefore, at the stage of decoding fMRI signals into latent variables of visual stimulation, the delay of neurovascular dynamics should also be considered [24].

B. Brain Cognitive Limitation

Due to the high cost of fMRI research and the complex research process, fMRI-based brain-computer interface (BCI) learns the self-regulation ability of brain region in the way of neural feedback (NF), and then it can be transferred to the more flexible and lower cost electroencephalogram brain-computer interface [36], [37]. Combine variational autoencoder and generative adversarial networks (GAN), use fMRI data to supplement electroencephalogram (EEG) data, and encode condition vectors with less noise [38]. In addition to decoding low-level visual stimuli, researchers also decode brain activity into low-level pixel space and high-level semantic space at the same time [39], [40], [41]. Due to the inadequacy of human research on visual mechanism, the current reconstruction field is exploratory. In the reconstruction process, the decoded noise may be the true prediction of the cerebral visual cortex's response to the outside world, and the clear image reconstructed by the reconstruction algorithm may also be noise [42].

V. CONCLUSION

Although researchers have successfully reconstructed the input visual stimuli from fMRI data, they are still in the initial stage of reconstruction. This paper reviews the current decoding methods of brain activity based on fMRI, which mainly face three challenges: mapping ability of decoding model; limited paired fMRI and image data; fMRI signals are mixed with noise. This paper also reviews the potential solutions to the above challenges in computer science and neuroscience. With the advancement of brain signal measurement technology, the development of more complex encoding and decoding models, and a better understanding of brain structure, "mind reading" will become a reality.

VI. ACKNOWLEDGMENT

This research was financially supported by the Civil Aviation Joint Funds of the National Natural Science Foundation of China (GrantNo.U1633121).

References

[1] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nature Reviews Neuroscience*, vol. 7, no. 7, pp. 523–534, 2006.

[2] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[3] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature neuroscience*, vol. 8, no. 5, pp. 679–685, 2005.

[4] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[5] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, "Encoding and decoding in fmri," *Neuroimage*, vol. 56, no. 2, pp. 400–410, 2011.

[6] M. A. Van Gerven, B. Cseke, F. P. De Lange, and T. Heskes, "Efficient bayesian multivariate fmri analysis using a sparsifying spatio-temporal prior," *NeuroImage*, vol. 50, no. 1, pp. 150–161, 2010.

[7] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[8] Y. Güçlütürk, U. Güçlü, K. Seeliger, S. Bosch, R. van Lier, and M. A. van Gerven, "Reconstructing perceived faces from brain activations with deep adversarial neural decoding," in *Advances in Neural Information Processing Systems*, 2017, pp. 4246–4257.

[9] J. A. Livezey and J. I. Glaser, "Deep learning approaches for neural decoding: from cnns to lstms and spikes to fmri," *arXiv preprint arXiv:2005.09687*, 2020.

[10] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature communications*, vol. 8, no. 1, pp. 1–15, 2017.

[11] G. Shen, T. Horikawa, K. Majima, and Y. Kamitani, "Deep image reconstruction from human brain activity," *PLoS computational biology*, vol. 15, no. 1, p. e1006633, 2019.

[12] H. Wen, J. Shi, Y. Zhang, K.-H. Lu, J. Cao, and Z. Liu, "Neural encoding and decoding with deep learning for dynamic natural vision," *Cerebral Cortex*, vol. 28, no. 12, pp. 4136–4160, 2018.

[13] S. Gadgil, Q. Zhao, E. Adeli, A. Pfefferbaum, E. V. Sullivan, and K. M. Pohl, "Spatio-temporal graph convolution for functional mri analysis," *arXiv preprint arXiv:2003.10613*, 2020.

[14] A. Grigis, J. Tasserie, V. Frouin, B. Jarraya, and L. Uhrig, "Predicting cortical signatures of consciousness using dynamic functional connectivity graph-convolutional neural networks," *bioRxiv*, 2020.

[15] Y. Zhang, L. Tetrel, B. Thirion, and P. Bellec, "Functional annotation of human cognitive states using deep graph convolution," *bioRxiv*, 2020.

[16] C. Du, C. Du, L. Huang, and H. He, "Reconstructing perceived images from human brain activities with bayesian deep multiview learning," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 8, pp. 2310–2323, 2018.

[17] S. M. LaConte, "Decoding fmri brain states in real-time," *Neuroimage*, vol. 56, no. 2, pp. 440–454, 2011.

[18] R. Sitaram, A. Caria, and N. Birbaumer, "Hemodynamic brain–computer interfaces for communication and rehabilitation," *Neural networks*, vol. 22, no. 9, pp. 1320–1328, 2009.

[19] D. Changde, L. Jinpeng, H. Lijie, H. Huiguang *et al.*, "Brain encoding and decoding in fmri with bidirectional deep generative models," 2019.

[20] M. Bontonou, N. Farrugia, and V. Gripon, "Few-shot learning for decoding brain signals," *arXiv preprint arXiv:2010.12500*, 2020.

[21] R. Hayashi and H. Kawata, "Image reconstruction from neural activity recorded from monkey inferior temporal cortex using generative adversarial networks," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 105–109.

[22] A. Show and K. Xu, "Tell: Neural image caption generation with visual attention," *Kelvin Xu et. al.. arXiv Pre-Print*, vol. 23, 2015.

[23] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," *arXiv preprint arXiv:1703.01365*, 2017.

[24] K. Han, H. Wen, J. Shi, K.-H. Lu, Y. Zhang, D. Fu, and Z. Liu, "Variational autoencoder: An unsupervised model for encoding and decoding fmri activity in visual cortex," *NeuroImage*, vol. 198, pp. 125–136, 2019.

[25] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current Biology*, vol. 21, no. 19, pp. 1641–1646, 2011.

[26] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, no. 6, pp. 902–915, 2009.

[27] L. McIntosh, N. Maheswaranathan, A. Nayebi, S. Ganguli, and S. Baccus, "Deep learning models of the retinal response to natural scenes," in *Advances in neural information processing systems*, 2016, pp. 1369–1377.

[28] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "Fmri brain-computer interface: a tool for neuroscientific research and treatment," *Computational intelligence and neuroscience*, vol. 2007, 2007.

[29] Y. Zhang and P. Bellec, "Transferability of brain decoding using graph convolutional networks," *bioRxiv*, 2020.

[30] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[31] J. M. Shine, P. G. Bissett, P. T. Bell, O. Koyejo, J. H. Balsters, K. J. Gorgolewski, C. A. Moodie, and R. A. Poldrack, "The dynamics of functional brain networks: integrated network states during cognitive task performance," *Neuron*, vol. 92, no. 2, pp. 544–554, 2016.

[32] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[33] A. Kazeminejad and R. C. Sotero, "Topological properties of resting-state fmri functional networks improve machine learning-based autism classification," *Frontiers in neuroscience*, vol. 12, p. 1018, 2019.

[34] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Distance metric learning using graph convolutional networks: Application to functional brain networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 469–477.

[35] R. W. Cox, A. Jesmanowicz, and J. S. Hyde, "Real-time functional magnetic resonance imaging," *Magnetic resonance in medicine*, vol. 33, no. 2, pp. 230–236, 1995.

[36] T. Hinterberger, N. Weiskopf, R. Veit, B. Wilhelm, E. Betta, and N. Birbaumer, "An eeg-driven brain-computer interface combined with functional magnetic resonance imaging (fmri)," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 971–974, 2004.

[37] T. Hinterberger, R. Veit, U. Strehl, T. Trevorrow, M. Erb, B. Kotchoubey, H. Flor, and N. Birbaumer, "Brain areas activated in fmri during self-regulation of slow cortical potentials (scps)," *Experimental brain research*, vol. 152, no. 1, pp. 113–122, 2003.

[38] I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano, and M. Shah, "Brain2image: Converting brain signals into images," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1809–1817.

[39] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain," *Neuron*, vol. 76, no. 6, pp. 1210–1224, 2012.

[40] A. G. Huth, T. Lee, S. Nishimoto, N. Y. Bilenko, A. T. Vu, and J. L. Gallant, "Decoding the semantic content of natural movies from human brain activity," *Frontiers in systems neuroscience*, vol. 10, p. 81, 2016.

[41] D. B. Walther, E. Caddigan, L. Fei-Fei, and D. M. Beck, "Natural scene categories revealed in distributed patterns of activity in the human brain," *Journal of neuroscience*, vol. 29, no. 34, pp. 10 573–10 581, 2009.

[42] K. Seeliger, U. Güçlü, L. Ambrogioni, Y. Güçlütürk, and M. A. van Gerven, "Generative adversarial networks for reconstructing natural images from brain activity," *NeuroImage*, vol. 181, pp. 775–785, 2018.