Decoding EEG by Visual-guided Deep Neural Networks

Zhicheng Jiao[1] , **Haoxuan You**[2] , **Fan Yang**[1] , **Xin Li**[1] , **Han Zhang**[1] and **Dinggang Shen**[1*]
[1]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
[2]BNRist, KLISS, School of Software, Tsinghua University, China

Abstract

Decoding visual stimuli from brain activities is an interdisciplinary study of neuroscience and computer vision. With the emerging of Human-AI Collaboration, Human-Computer Interaction, and the development of advanced machine learning models, brain decoding based on deep learning attracts more attention. Electroencephalogram (EEG) is a widely used neurophysiology tool. Inspired by the success of deep learning on image representation and neural decoding, we proposed a visual-guided EEG decoding method that contains a decoding stage and a generation stage. In the classification stage, we designed a visual-guided convolutional neural network (CNN) to obtain more discriminative representations from EEG, which are applied to achieve the classification results. In the generation stage, the visual-guided EEG features are input to our improved deep generative model with a visual consistence module to generate corresponding visual stimuli. With the help of our visual-guided strategies, the proposed method outperforms traditional machine learning methods and deep learning models in the EEG decoding task.

1 Introduction

Vision is one of the most important components in the human perception system. When eyes receive visual stimulation, neural spikes are produced in brain [Jacobs *et al.*, 2009]. Decoding neural spikes produced in the brain due to visual stimuli is instrumental to the exploration of visual information processing mechanism of humans and to development of computer vision [Nestor *et al.*, 2016]. With the development of contemporary machine learning methods, it becomes possible to obtain high-quality decoding results from functional magnetic resonance imaging (fMRI) and EEG.

fMRI sequences provide information from several visual cerebral areas, and play an important role in visual decoding tasks varying from categorizing to reconstruction [Haxby *et al.*, 2001; Haynes and Rees, 2005; Miyawaki *et al.*, 2008; Naselaris *et al.*, 2009; Cowen *et al.*, 2014]. In previous works, machine learning models are proposed to capture the relationship between visual stimuli and fMRI. More recently, deep learning methods have also been proposed to reconstruct images from fMRI [Horikawa and Kamitani, 2017; Güçlütürk *et al.*, 2017; Du *et al.*, 2017].

The huge volume and high cost render fMRI based neural decoding impossible in daily life [Poldrack and Farah, 2015]. On the contrary, EEG has significant advantages in aspects of portability and price which promote the application of EEG decoding systems. For example, EEG devices with integrated machine learning algorithms are the key components in brain-controlled systems for disabled persons [Liu *et al.*, 2017; Park *et al.*, 2011]. In addition, deep neural networks have also been reported to perform competitively on decoding visual related EEG trials [Adamos *et al.*, 2016; Spampinato *et al.*, 2017; Palazzo *et al.*, 2017]. Assuming that EEG signal exists in a cognitive domain, the workflow of existing EEG decoding methods is summarized by Fig 1., i.e.s (1) Visual stimuli are shown to subjects; (2) EEG trails are recorded; (3) EEG trials are represented by handcrafted features or neural networks for decoding. Importantly, this decoding pipeline only utilizes information in cognitive domain. However, there is an inevitable fact that state-of-the-art deep learning models outperform humans on representing and classifying images [He *et al.*, 2016]. So, information in the related visual domain can assist to improve the performance of EEG decoding.

In this paper, we propose using visual representation to guide the decoding of EEG. Our decoding framework contains two stages: (1) Visual-guided EEG classification stage; (2) Visual-guided stimuli generation stage. In the first stage, a classification network is guided by visual representation to category EEG into the classes of related visual stimuli. Then, in the generation stage, an improved generative adversarial network (GAN) [Goodfellow *et al.*, 2014] produces the corresponding visual stimuli from visual-guided EEG representations. Compared with state-of-the-art neural decoding models, our method achieves superior performance.

2 Related Work

Spatial-temporal features of EEG are used to decode stimulus-related signals into two event-related potentials (ERPs) for capturing the distribution of signal with a Gaussian mixture model (GMM) [Tzovara *et al.*, 2012]. In a recent study [Kaneshiro *et al.*, 2015], principal component analysis

[*] Corresponding Author

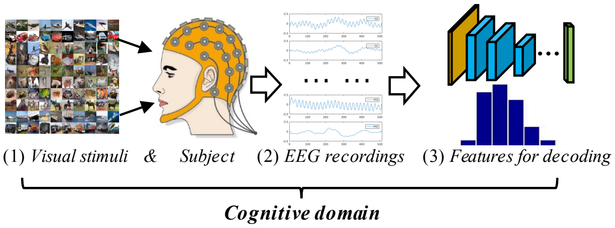

Figure 1: Traditional EEG decoding strategy.

(PCA) and linear discriminant analysis (LDA) are combined to recover fine-grained object category from EEG recordings.

Deep learning methods have become dominant in various traditional and bio-inspired computer vision tasks and bio [LeCun *et al.*, 2015; Zhang *et al.*, 2019]. Variations of CNN models are applied to represent EEG for detecting particular signal components [Cecotti and Graser, 2011; Plis *et al.*, 2014]. More recently, face image sequences and EEG data are analyzed jointly for emotion detection [Soleymani *et al.*, 2016]. Frameworks combing CNN and other deep learning structures are proposed [Bashivan *et al.*, 2015; Jiao *et al.*, 2018] to decode mental loads from EEG. Novel EEG-driven automated visual classification and generation methods have also been proposed [Spampinato *et al.*, 2017; Palazzo *et al.*, 2017]. Generally, these studies are based on the same assumption that EEG can guide deep learning models to achieve better results on image classification and generation. To the best of our knowledge, this work is the first to use visual representation to guide the decoding of visual stimulated EEG. Our study is based on the proven fact that state-of-the-art deep neural networks perform better than humans on visual representation tasks [He *et al.*, 2016].

3 Our Method

Since our decoding framework contains an EEG classification stage and a stimuli generation stage, they will be described respectively in the following subsections.

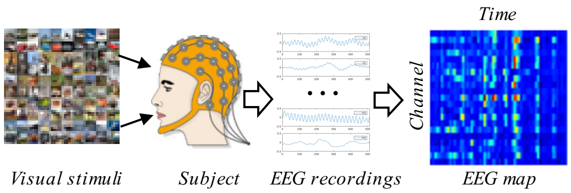

Figure 2: EEG map evoked by visual stimuli.

3.1 Visual-guided EEG Classification

In the cognitive domain, EEG signals are first organized to EEG maps X_{cog} as [Jiao *et al.*, 2018], axes of which stand for EEG channels and recordings of each channel (Fig 2).

As shown in Fig 3., X_{cog} and X_{vis} stand for input to representation layers in these two domains. Our classification model contains: (1) Feature representation layers (Cognitive CNN and Visual CNN) in both cognitive domain and visual domain; (2) Classification layer (Class) in cognitive domain. The loss function of our classification model is a modified softmax loss as shown in Equation (1).

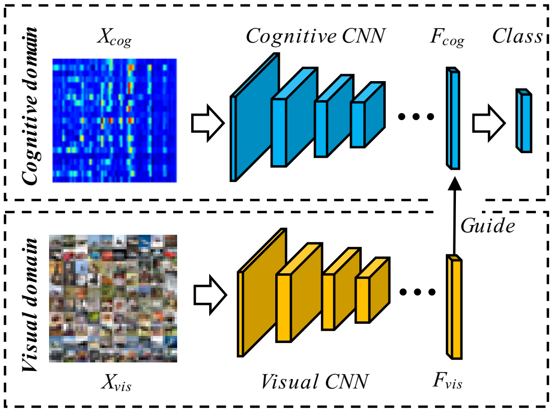

Figure 3: Visual-guided EEG classification.

$$L = -\sum_{j=1}^{N} y_j \log S_j + (F_{cog} - F_{vis})^2 \qquad (1)$$

In this equation, the first term is a standard softmax loss: S_j is the softmax value which stands for probability of one instance belonging to category j, y_j is the label of input, and N is the number of classes. The second term is a squared error to make the normalized cognitive representation F_{cog} (output of feature representation layers in the cognitive domain) be close to the normalized visual representation F_{vis} in the feature space. In the training process, parameters of pretrained visual CNN are fixed to render the cognitive representations more discriminative, assisting to obtain superior performance in the classification task.

3.2 Visual-guided Stimuli Generation

Since our visual generation stage is in the form of an improved GAN model, we briefly introduce the basic GAN method first. Then, we show how our improved GAN model for decoding EEG data is constructed.

GAN models use a minimax game which guides two networks (a generative net and a discriminative net) to be trained in opposite directions. In the adversarial training process, the generative net tries to map a latent space to particular data distribution, while the discriminative net discriminates between real data and generated instances. In our EEG decoding task, the main role of GAN is learning to generate stimuli from corresponding EEG representations.

Our improved GAN contains a generative net G , a discriminative net D , and two representative networks R, which are shown as Fig 4. Both G and D are convolutional neural networks which share the same structures as ones proposed in [Palazzo *et al.*, 2017]. R are the fully convolutional networks (FCN) [Long *et al.*, 2015] that share the same structures and parameters. Goals of each network (G, D, and R) can be categorized as follow:

(1) G takes F_{cog} as input for generating plausible images X_{fake};

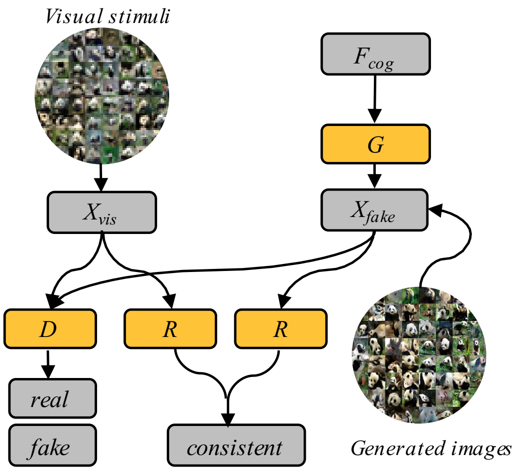

Figure 4: Illustration of generation stage. F_{cog} is visual-guided EEG representation, G is generative net, D is discriminative net, and two R are networks which are used to calculate consistence between generated images and real visual stimuli.

(2) D tries to distinguish X_{fake} from visual stimuli X_{vis};

(3) R assists to maintain consistency of visual components between X_{fake} and X_{vis} via lower-level visual perception and higher-level visual semantics.

The discriminative loss L_D is computed as Equation (2), where F_{cog-w} is wrongly corresponding EEG representation, randomly chosen from different classes of X_{vis}. The generator loss L_G is shown as Equation (3). Where $\lambda\,(L_{per}+L_{sem}\,)$ is the visual-consistent term which is achieved by visual representation networks R. In this visual-consistent term, L_{per} is in the form of $\parallel f\,(X_{vis}) - f\,(X_{fake})\parallel_2^2$, which stands for perceptual loss [Johnson *et al.*, 2016] between FCN feature maps (f in the first FCN layer for representing more low-level visual details) of X_{fake} and X_{vis}. L_{sem} (containing more semantic information) is the softmax cross entropy between semantic segmentation results of X_{fake} and X_{vis} after going through R. This visual-consistent term acts as auxiliary for generating higher-quality images.

$$\begin{aligned} L_D = -\log D\,(X_{vis}|F_{cog}) - \log(1 - D\,(X_{vis}|F_{cog-w})) \\ - \log(1 - D\,(X_{fake}|F_{cog})) \end{aligned} \tag{2}$$

$$L_G = -\log D\,(X_{fake}|F_{cog}) + \lambda\,(L_{per} + L_{sem}) \tag{3}$$

Methods which are the most similar with our decoding framework are [Spampinato *et al.*, 2017; Palazzo *et al.*, 2017]. The main differences between our method and theirs: (1) The representation F_{cog} is visual-guided, whereas previous works only utilize cognitive domain information in their generation stage. Our visual-guided F_{cog} is more discriminative in classification stage and more diverse in generation stage, contributing to better decoding performance; (2) The visual-consistent term in our generative loss helps to maintain X_{fake} with more qualitative consistence when observed X_{vis} .

4 Datasets

Performance of our framework is evaluated and compared with state-of-the-art methods on two public datasets: (1) ImageNet subset [Spampinato *et al.*, 2017; Kavasidis *et al.*, 2017]; (2) Face and object [Kaneshiro *et al.*, 2015]. Details of datasets are described in the subsections.

4.1 ImageNet Subset

This dataset is a 40-class subset of ImageNet [Deng *et al.*, 2009]. The related classes: dog, cat, butterfly, sorrel, capuchin, elephant, panda, fish, airliner, broom, canoe, phone, mug, convertible, computer, watch, guitar, locomotive, espresso, chair, golf, piano, iron, jack, mailbag, missile, mitten, bike, tent, pajama, parachute, pool, radio, camera, gun, shoe, banana, pizza, daisy, and bolete (fungus). During the EEG acquisition experiments, 2,000 images (50 images in each class) are shown to 6 subjects. A 128-channel Brainvision EEG system is applied to record related neural signal during the process described above, and totally 12,000 visual-evoked EEG sequences are acquired. More details can be found in the related reference [Spampinato *et al.*, 2017].

4.2 Face and Object

Visual stimuli applied in this dataset are categorized into two classes: face and object (12 images in each class). EEG trials are recorded by a 128-channel EGI system when 10 subjects viewed a sequence of images. Details of stimulating experiments and subjects are described in [Kaneshiro *et al.*, 2015]. Meanwhile, data acquisition and preprocessing are also detailed in the related paper. Finally, a dataset of 17,281 sequenced EEG recordings is formed, of which 8,641 samples are evoked by stimuli of faces and 8,640 samples are evoked by visual stimuli of objects. Formulation of this dataset is similar to that of ImageNet subset as mentioned above.

5 Experiments and Analysis

Our decoding method can not only classify but also reconstruct visual stimuli from single-trial EEG data. So, the first goal of our experiments is to investigate whether EEG data which represent visual stimuli of different categories can be classified accurately. The other goal is to evaluate whether our method can reconstruct plausible images from related EEG recordings. The performance of classification and generation are detailed in the following subsections respectively. Our models are based on the deep learning toolkit of TensorFlow [Abadi *et al.*, 2016].

5.1 Performance of Classification

Both objective and subjective evaluations are applied to compare classification performance of our method with state-of-the-art ones. The objective evaluation criterion is classification accuracy, while the subjective one is t-distributed stochastic neighbor embedding (t-SNE) [Maaten and Hinton, 2008]. t-SNE can project high-dimensional data into a 2-D scatter plot which is widely used for evaluating discriminative ability of feature vectors. For t-SNE visualization, the evaluation criterion is: the more instances of one class can

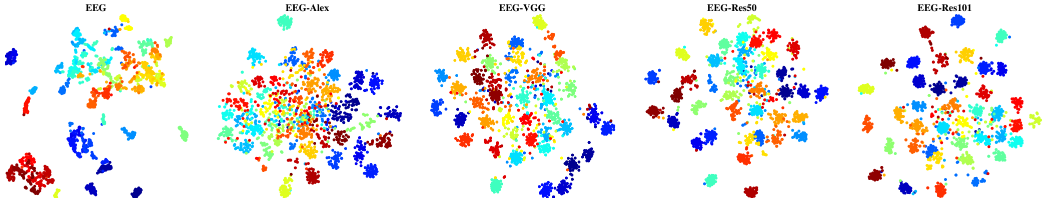

Figure 5: t-SNE maps on *ImageNet subset* dataset.

Method	Classification Accuracy(%)
LDA	80.27
LSTM	83.09
CNN	85.60
CNN-Alex	88.80
CNN-VGG16	90.97
CNN-ResNet50	92.65
CNN-ResNet101	**92.99**

Table 1: Mean classification accuracy on *ImageNet subset* dataset.

Method	Classification Accuracy(%)
LDA	81.06
LSTM	80.67
CNN	83.10
CNN-ResNet101	**85.50**

Table 2: Mean classification accuracy on *face and object* dataset.

be separated from ones of other classes, the better the related features perform.

We performed classification experiments under the same condition as described in related references [Spampinato *et al.*, 2017; Kaneshiro *et al.*, 2015]. Training strategy of our classification network is Adam. The independent separations of EEG signal datasets are 80% for training, 10% for validation, and 10 % for testing. Structure of our EEG classification net in cognitive domain is the same form as that of AlexNet [Krizhevsky *et al.*, 2012], except for numbers of neurons in the classification layer (40 for ImageNet subset dataset, and 2 for face and object dataset). Data augmentation methods and network training strategies are drawn from [Jiao *et al.*, 2018]. Pretrained networks in the visual domain are AlexNet [Krizhevsky *et al.*, 2012], VGG16 [Simonyan and Zisserman, 2014], ResNet50, and ResNet101 [He *et al.*, 2016], which are used to obtain visual representation to guide EEG decoding.

For ImageNet subset dataset, classification accuracy of different methods are listed in Table 1. LDA and long short term memory (LSTM) are state-of-the-art methods [Kaneshiro *et al.*, 2015; Spampinato *et al.*, 2017]. CNN is our EEG classification net which is not guided by visual representation. CNN-Alex, CNN-VGG16, CNN-ResNet50, and CNN-RestNet101 stand for classification performance guided by different pretrained visual networks. Results listed in this table show that our visual-guided frameworks outperform LDA and LSTM, among which the ResNet101 guided classification method achieves a new state-of-the-art result, with our method improving the performance of the EEG classification stage. EEG features extracted by CNN in the cognitive domain and EEG features guided by visual net are visualized by t-SNE in Fig 5., data points with different colors representing categories of EEG features. It is obvious that the visual-guided EEG representations of different visual classes are more separable in the corresponding feature space, while instances in the same class are more uniformly distributed. These properties demonstrate superior performance of our method in handling diversity (more separable data points in feature space assist to generate diverse visual stimuli) in the generation stage. Besides, performance of visual guided model (CNN-ResNet101) and no visual guided CNN are compared by category in Fig 6., a positive number or negative number (superiority of classification accuracy (%)) representing respectively superior or inferior performance of our strategy compared to traditional strategies in a given category. Our visual guided strategy achieves superior classification performance among most of the categories, except for *Airliner, Folding chair, Mailbag, Radio telescope, Revolver, and Running shoe*, which are attributable to the same synsets (categories of some image classes) of *Artifact* in *ImageNet* dataset. We hypothesize that this illustrates that human knowledge in cognitive domain keeps the advantage on classifying visual stimuli of man-made objects.

We also compare the performance of our CNN and ResNet101 guided CNN with LDA and LSTM on face and object dataset. The related objective results are shown in Table 2, and the subjective results are illustrated in Fig 7. It is obvious to see that our method also performs better than the non-visual guided models on this dataset.

5.2 Performance of Generation

The subjective visual quality of generated images is usually chosen as an evaluation criterion for generative models. In addition, inception score and inception classification accuracy [Palazzo *et al.*, 2017; Kavasidis *et al.*, 2017] are widely used quantitative evaluations for GANs (higher values of inception score and inception classification accuracy stand for a superior performance of generation). We also use these methods to evaluate performance of our generation stage. Since the number of images in face and object dataset is very small, it does not meet the requirement of diversity for training an effective GAN model [Goodfellow *et al.*, 2014]. So, generation performance of both previous works [Spampinato *et al.*, 2017; Palazzo *et al.*, 2017; Kavasidis *et al.*, 2017] and our method are discussed just on ImageNet subset dataset.

Training strategies for our improved GAN model mainly follow those in [Palazzo *et al.*, 2017]. Parameters of *D* and *G* are listed in Table 3 and Table 4, in which Conv and Deconv stand for convolution and deconvolution blocks with ReLU

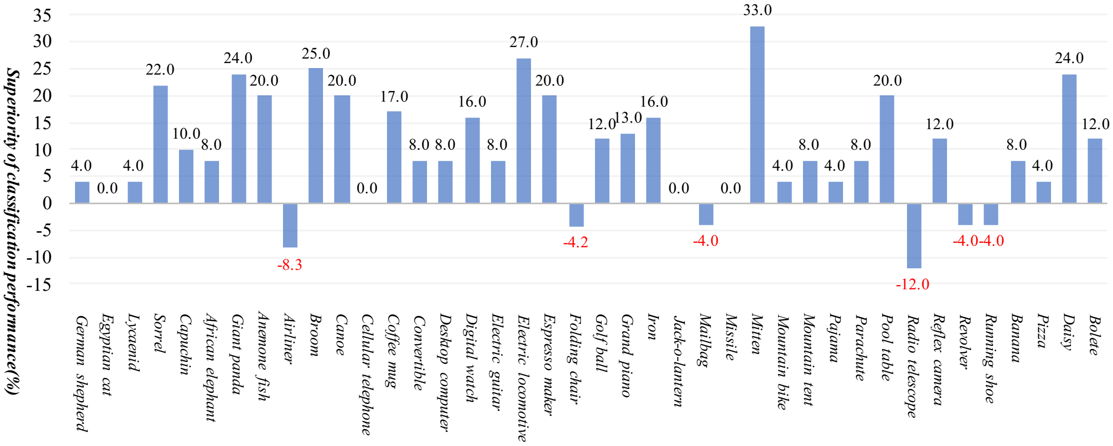

Figure 6: Visual guided classification ***VS*** Non-visual guided classification

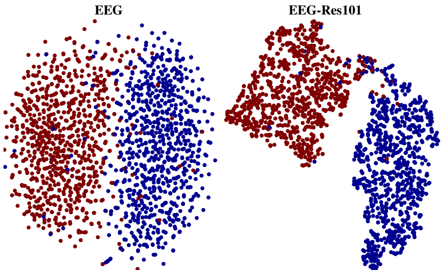

Figure 7: t-SNE maps on *face and object* dataset.

Layer	Filter size	Filter dimension
Conv1	4	64
Conv2	4	128
Conv3	4	256
Conv4	4	256
Fc5	1	1024

Table 3: Parameters of *D* net.

Layer	Filter size	Filter dimension
Deconv1	4	512
Deconv2	4	256
Deconv3	4	128
Deconv4	4	64
Deconv5	4	32

Table 4: Parameters of *G* net.

activation and batch normalization operation, and Fc is fully-connected layer. We choose the best-performed visual-guided representation (guided by ResNet101) as F_{cog} in generation stage. FCN for the visual-consistent term λ $(L_{per} + L_{sem})$ is pretrained on VOC2012 dataset [Everingham *et al.*, 2010], and λ is set to 0.5. Since only limited images in the dataset are labeled by EEG, previous works initialize training of GAN with random noise Z and unlabeled images (50,000 unlabeled images in ImageNet). In our work, visual features F_{vis} of 50,000 images can be obtained from visual networks, and they are applied in the pretrain stage of GAN model. Then, training of GAN is refined by the visual guided representation F_{cog}.

In previous papers, high-quality results of three classes (*Airliner*, *Jack-oa-Lantern*, and *Panda*) and low-quality results of other three classes (*Banana*, *Capuchin*, and *Bolete*) are listed for subjective evaluation. We follow this evaluation strategy and list generated instances of different methods in Fig 8 and Fig 9. In these figures, the top rows named EEG-GAN are the results generated by state-of-the-art method [Palazzo *et al.*, 2017; Kavasidis *et al.*, 2017], while the bottom ones are generated by our visual-guided GAN with the visual-consistent term. Importantly, images generated by our method contain more visual details and they correspond better with related classes. The subjective visual quality of our results on all these classes are superior to the state-of-the-art one. This phenomenon demonstrates the efficiency of our proposed decoding method in visual stimuli generation stage.

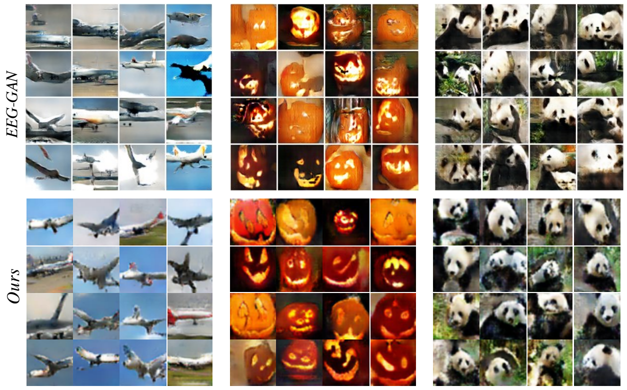

Figure 8: Good results of three classes on *ImageNet subset* dataset. From left to right: *Airliner*, *Jack-oa-Lantern*, and *Panda*.

Inception scores and inception classification accuracy of different methods are listed in Table 5, where EEG-GAN stands for the method (state-of-the-art EEG decoding model) proposed in [Palazzo *et al.*, 2017; Kavasidis *et al.*, 2017],

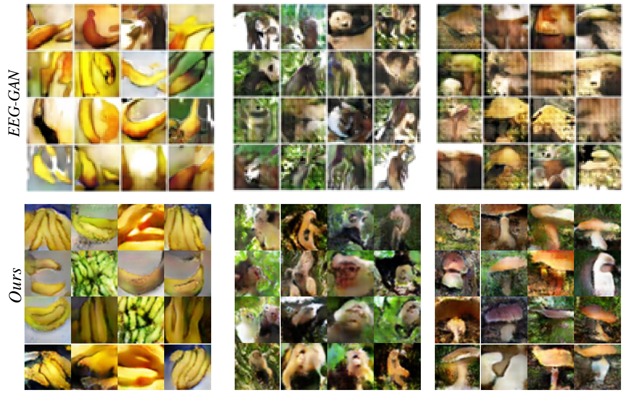

Figure 9: Bad results of three classes on *ImageNet subset* dataset. From left to right: *Banana*, *Capuchin*, and *Bolete*.

Method	IS	IC
EEG-GAN	5.07	0.43
VG-GAN	5.54	0.51
VG-GAN-VC	**6.33**	**0.53**

Table 5: Inception scores (IS) and Inception classification accuracy (IC) on *ImageNet subset* dataset.

VG-GAN stands for our visual-guided GAN without visual-consistent term, VG-GAN-VC represents our visual-guided GAN with visual-consistent term. Results in this table demonstrate that our visual-guided method achieves better results in both these two aspects. In addition, the visual-consistent term is efficient to further improve the objective visual quality of generated images for obtaining better decoding performance.

6 Conclusion

Deep learning methods have achieved significant improvements on neural decoding tasks. Inspired by the representative studies of other researchers, we propose a visual-guided decoding framework for EEG data in this paper. To obtain superior decoding results, our work takes full advantage of visual representations which are obtained from state-of-the-art deep learning models on computer vision tasks. The proposed framework contains a visual-guided EEG classification stage and a visual-guided generation stage. In the classification stage, visual-guided EEG representations bridge the gap between cognitive domain and visual domain for categorizing EEG recordings evoked by different visual stimuli more accurately. In the visual-guided visual stimuli generation stage, the visual-guide EEG representation can also improve the performance of generation. Besides, our improved GAN model can improve consistency between visual representation of real stimuli and generated instances to further improve the subjective and objective quality of generated images. However, as the per-class classification results shown in Fig 6., human can achieve higher classification accuracy in certain classes. This phenomenon shows opportunities to combine the advantages of both human brain and deep learning models to improve the performance. In the future, we will extend the simultaneous and synergistic utilization of both visual information and cognitive knowledge to further improving neural decoding tasks, thus augmenting the areas such as hybrid intelligence and human-AI collaboration.

Acknowledgments

In this research, we perform the deep learning experiments on a Titan V graphics card provided by the NVIDIA Academic Program of *GPU Grant Program*.

References

[Abadi *et al.*, 2016] 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 *OSDI*, volume 16, pages 265–283, 2016.

[Adamos *et al.*, 2016] Dimitrios A Adamos, Stavros I Dimitriadis, and Nikolaos A Laskaris. Towards the biopersonalization of music recommendation systems: A single-sensor eeg biomarker of subjective music preference. *Information Sciences*, 343:94–108, 2016.

[Bashivan *et al.*, 2015] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*, 2015.

[Cecotti and Graser, 2011] Hubert Cecotti and Axel Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):433–445, 2011.

[Cowen *et al.*, 2014] Alan S Cowen, Marvin M Chun, and Brice A Kuhl. Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage*, 94:12–22, 2014.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. IEEE, 2009.

[Du *et al.*, 2017] Changde Du, Changying Du, and Huiguang He. Sharing deep generative representation for perceived image reconstruction from human brain activity. *arXiv preprint arXiv:1704.07575*, 2017.

[Everingham *et al.*, 2010] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[Güçlütürk *et al.*, 2017] 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 *NIPS*, pages 4249–4260, 2017.

[Haxby *et al.*, 2001] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

[Haynes and Rees, 2005] John-Dylan Haynes and Geraint Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience*, 8(5):686–691, 2005.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[Horikawa and Kamitani, 2017] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8:15037, 2017.

[Jacobs *et al.*, 2009] Adam L Jacobs, Gene Fridman, Robert M Douglas, Nazia M Alam, Peter E Latham, Glen T Prusky, and Sheila Nirenberg. Ruling out and ruling in neural codes. *Proceedings of the National Academy of Sciences*, 106(14):5936–5941, 2009.

[Jiao *et al.*, 2018] Zhicheng Jiao, Xinbo Gao, Ying Wang, Jie Li, and Haojun Xu. Deep convolutional neural networks for mental load classification based on eeg data. *Pattern Recognition*, 76:582–595, 2018.

[Johnson *et al.*, 2016] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2016.

[Kaneshiro *et al.*, 2015] Blair Kaneshiro, Marcos Perreau Guimaraes, Hyung-Suk Kim, Anthony M Norcia, and Patrick Suppes. A representational similarity analysis of the dynamics of object processing using single-trial eeg classification. *PloS one*, 10(8):e0135697, 2015.

[Kavasidis *et al.*, 2017] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. Brain2image: Converting brain signals into images. In *ACM MM*, pages 1809–1817. ACM, 2017.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[LeCun *et al.*, 2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[Liu *et al.*, 2017] Feng Liu, Shouyi Wang, Jay Rosenberger, Jianzhong Su, and Hanli Liu. A sparse dictionary learning framework to discover discriminative source activations in eeg brain mapping. In *AAAI*, pages 1431–1437, 2017.

[Long *et al.*, 2015] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

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

[Naselaris *et al.*, 2009] Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6):902–915, 2009.

[Nestor *et al.*, 2016] Adrian Nestor, David C Plaut, and Marlene Behrmann. Feature-based face representations and image reconstruction from behavioral and neural data. *Proceedings of the National Academy of Sciences*, 113(2):416–421, 2016.

[Palazzo *et al.*, 2017] Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, and Daniela Giordano. Generative adversarial networks conditioned by brain signals. In *ICCV*, pages 3410–3418, 2017.

[Park *et al.*, 2011] Jaeyoung Park, Kee-Eung Kim, and Yoon-Kyu Song. A pomdp-based optimal control of p300-based brain-computer interfaces. In *AAAI*, 2011.

[Plis *et al.*, 2014] 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, 2014.

[Poldrack and Farah, 2015] Russell A Poldrack and Martha J Farah. Progress and challenges in probing the human brain. *Nature*, 526(7573):371–379, 2015.

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

[Soleymani *et al.*, 2016] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. Analysis of eeg signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, 7(1):17–28, 2016.

[Spampinato *et al.*, 2017] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *CVPR*, pages 6809–6817, 2017.

[Tzovara *et al.*, 2012] Athina Tzovara, Micah M Murray, Gijs Plomp, and Michael H Herzog. Decoding stimulus-related information from single-trial eeg responses based on voltage topographies. *Pattern Recognition*, 45(6):2109–2122, 2012.

[Zhang *et al.*, 2019] Xiaodan Zhang, Xinbo Gao, Wen Lu, and Lihuo He. A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction. *IEEE Transactions on Multimedia*, 2019.