The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

ニューラルネットワークを用いた ヒトの脳活動からの動的視覚刺激の再構成

Reconstructing Dynamic Visual Stimuli from Human Brain Activity using Deep Neural Networks

永野 雄大 *1 Yudai Nagano　小林 一郎 *2 Ichiro Kobayashi　西本 伸志 *3 Shinji Nishimoto　中山 英樹 *1 Hideki Nakayama

*1東京大学 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

*2お茶の水女子大学 基幹研究院 自然科学系
Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

*3情報通信研究機構　脳情報通信融合研究センター
Center for Information and Neural Networks, National Institute of Information and Communications Technology

Decoding is one of the important fields in Neuroscience, which is considered to be useful for analysis of brain function, clarification of disease and development of Brain Machine Interfaces. The purpose of this study is to decode visual stimuli from human brain activity. To reconstruct the visual stimuli, we used Neural Networks and GAN-based Neural Networks. We compared recent GAN-based models to confirm which one works the best. We also examined the difference in reconstruction quality when brain area was changed. To improve the quality of reconstruction, we combine multiple consecutive frames of human brain activity. Finally, we calculated the effect of multi-frame by quantative evaluation. The results show the effectiveness of decoding with multi-frame inputs.

1. はじめに

1.1 BMI とデコーディング

生物の脳活動の解析は科学の中でも長年注目を集めている分野であり，特にヒトの脳活動の解析は様々な側面から重視されている．その理由として挙げられるのが，ブレイン・マシン・インターフェース (BMI) への応用である．BMI の重要度は大きく，例えば運動機能に障碍を持つ人への運動機能の補助や，ロボットの遠隔操作の助けになる．なお，BMI には，侵襲性と非侵襲性の二種類がある．非侵襲性の BMI は，脳活動を直接取得できないため侵襲性の BMI より信号の取得精度が落ちるものの，装着した対象の安全性を確保できるため，今後の発展が期待されている．このような BMI を開発するために，非侵襲性の装置によってヒトの脳活動を解読 (デコーディング) することは重要である．

1.2 本研究の目的

本研究の目的は，ヒトの脳活動から，そのとき与えられていた視覚刺激を再構成することである．そのため，動的視覚刺激を再構成しやすいように複数フレームを入力とすることを提案，導入し，その効果の測定を行う．また，Generative Adversarial Network (GAN) [1] を用いたもの，用いていないものによる脳活動の再構成の実証的研究も示し，どのような方法が脳活動から視覚刺激を再構成するのにもっとも良いかを確かめる．また，学習・推論に使用する領野を選択し，どの領野が再構成に影響を与えているかを確認する．これにより，実際にヒトの脳活動においても同様の情報が用いられているとし，脳活動の一部を解読することを試みる．

連絡先: 永野 雄大, 東京大学, nagano@nlab.ci.i.u-tokyo.ac.jp

2. 関連研究

2.1 GAN: Generative Adversarial Nets

GAN [1] は，Generator, Discriminator と呼ばれる二つのネットワークからなる．Generator は Discriminator を騙すようなデータをノイズから生成し，真のデータとの KL 距離を最小化するように学習する．逆に Discriminator は Generator から生成されたデータが本物のデータかどうかを区別することで学習を進めていく．最終的には，Discriminator は入力データが Generator からのものか，真のデータからなのか区別できなくなる．一方，Generator は真の分布に近い分布を生成できるようになる．

2.2 デコーディング技術

脳活動からの視覚刺激の再構成の初期の研究は，Thirion らの研究であり [2] 脳活動から解像度 3×3 のパターン予測である．その後，Miyawaki らにより，解像度 10×10 のパターンを予測する研究 [3] が行われた．また，Naselaris らはベイズ的手法を用いた事前知識を必要とする，解像度 64×64 の自然画像を再構成する研究 [4] を行った．その後，Nishimoto らにより vim-2 [5] データセットの作成とこれを用いた動画の再構成 [6] が行われた．一方 Shen らは，事前学習モデルなどの外部知識を用いないモデルを使用し，fMRI から取得した脳活動から視覚刺激を再構成する研究 [7] を行った．この研究において，Shen らはニューラルネットワークに GAN の学習法を取り入れている．

2.3 vim-2 データセット

使用したデータは vim-2 データセット [5] であり，これには 3 名の実験対象者の脳活動が含まれている．この脳活動は実験対象者が視覚刺激を見ているときに fMRI で取得された信号となっている。刺激データは動画のフレームの連続となっており，フレームレートは 15 Hz，対する脳活動は 1 Hz で取得されている．よって，視覚刺激は合計 108,000 フレーム，

脳活動は 7,200 のデータ点からなる．実験対象者は 600 秒の動画を 12 セット，計 7,200 秒見る．これが学習用データとして保存されている．また，テスト用データとして，実験対象者は 60 秒の動画を 9 セット，これを 10 回見る．10 回分は平均が取られ，最終的なテストデータとなる．訓練用データとテスト用データの間に重複は存在しない．

3. 提案手法

図 1 に本研究における脳活動からの動的視覚刺激再構成の概要を示す．時間 t[s] から時間 $t+d-1$[s] までの間に生じた脳活動をまとめて Generator への入力とする．また，今回 d は 3 を基準としている．これにより Generator は視覚刺激の再構成を行い 1 フレーム の刺激画像を出力する．本研究の方法では，これを時間方向に繰り返し，動的視覚刺激を再構成する．

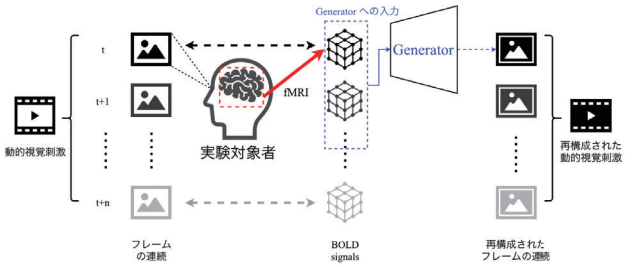

図 1: 脳活動からの動的視覚刺激再構成の概要図．実験対象者から取得された BOLD 信号は連続するフレームとして，ニューラルネットワークへの入力となる．出力は再構成された映像の 1 フレームとする．これを，動画の各フレームに対して行い，フレームを連続的に再構成することで動画の再構成とする．

4. 実験

4.1 評価指標

定量評価のためには，テスト用データをどの程度再構成できたのかを計算する指標が必要である．ここでは，その評価を行うため，一般的な画質評価指標である Peak Signal-to-Noise Ratio (PSNR) と Structural Similarity (SSIM) を用いる．

4.1.1 評価指標: Peak Signal-to-Noise Ratio (PSNR)

PSNR はピーク信号対雑音比と呼ばれ，2 枚の画像間の画質を評価する指標である．今回は再構成を行う視覚刺激に対し，脳活動からそのフレームをどの程度画像として復元できるのかを確かめるために PSNR を用いる．

PSNR は式 2 で表せる．ここで，$\hat{\boldsymbol{y}}$ は正解データ，$\boldsymbol{y}$ はモデルの出力，MAX_I は画像の中の画素の最大値 (今回は 1.0) である．この式を見ると，MSE の値によって PSNR が変動し，0 から ∞ の値をとることがわかる．

$$\mathrm{MSE} = |\hat{\boldsymbol{y}} - \boldsymbol{y}|_2 \tag{1}$$

$$\mathrm{PSNR} = 10 \cdot \log \frac{\mathrm{MAX}_I^2}{\mathrm{MSE}} \tag{2}$$

4.1.2 評価指標: Structural Similarity (SSIM)

SSIM [8] は，PSNR 同様，2 枚の画像間の画質を評価する指標である．SSIM は PSNR とは異なる画質評価指標であり，画素の微小な変化に強く，画像の構造を重視しており，PSNR よりも人間の知覚に近いと考えられている．

2 画像 x と y の SSIM は式 3 で表せる．ここで，SSIM は少領域のウィンドウの評価値の平均を取る．μ_x, μ_y は，画像 x と画像 y のウィンドウ内の画素の平均値．σ_x, σ_y は，画像 x と画像 y ウィンドウ内の画素の分散．σ_{xy} は，画像 x と画像 y のウィンドウ内の共分散である．ここで，c_1, c_2 はゼロ除算を防ぐための微小な定数である．

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

4.2 結果

各モデルによって，画質評価指標である PSNR と SSIM を比較した結果を表 1 に示す．ここで，ランダムサンプリングでは訓練用データからランダムに選択してきた視覚刺激を出力に代えたものである．また，最近傍法は視覚刺激に対する脳活動にもっとも近い脳活動を訓練用データから検索し、対応する視覚刺激を予測として利用する．NN は GAN による学習法を行わない畳み込みニューラルネットワークによる出力である．SNGAN + PD は SNGAN[9] に projection discriminator[10] を追加したモデルである．ここでは，全て時間長は $d = 3$ として実験を行っている．

表 1: 各モデルの PSNR / SSIM の比較.

モデル	実験対象者 1	実験対象者 2	実験対象者 3
ランダムサンプリング	9.43 / 0.0716		
最近傍法	9.93 / 0.089	9.84 / 0.096	9.71 / 0.090
NN	**12.6** / 0.098	**12.7** / 0.092	**12.6** / 0.098
GAN	12.1 / 0.119	12.0 / 0.115	11.9 / 0.116
WGAN-gp[11]	12.0 / 0.109	11.7 / 0.110	11.7 / 0.110
SNGAN + PD	11.7 / **0.126**	11.6 / **0.131**	11.6 / **0.124**

次に，使用する脳領野を限定し，ニューラルネットワークの学習を行う．これにより，どの脳の部位の情報を削ると，再構成の再現力が低下するかが分かり，視覚野の部位における役割の予測ができる．この結果を表 2 に示す．

表 2: 扱う部位の変化による PSNR/SSIM の比較

脳領野	実験対象者 1	実験対象者 2	実験対象者 3
V1∼V2	11.6 / 0.106	11.5 / 0.111	11.7 / 0.116
V3∼V4	11.9 / 0.108	12.0 / 0.112	11.5 / 0.097
V1∼V4	**12.1 / 0.119**	**12.0 / 0.115**	**11.9 / 0.116**
Left	11.6 / 0.108	11.4 / 0.100	11.7 / 0.108
Right	11.7 / 0.099	11.8 / 0.110	11.7 / 0.106

また、複数フレームの入力が実際に効果的な手法なのかどうかを調べるため，単一フレームの入力を行うものと，複数フレームとその長さを変更したものを複数用意し，学習・比較を行う．これによって得られた結果を表 3 に示す．ここでの時間長 1 は，単位フレームを入力した場合を表す．

5. 結論

ニューラルネットワークで GAN の学習法を用い，脳活動から動的視覚刺激の再構成を試みた．これにより，ニューラルネットワークを用いたモデルは，ランダムサンプリングや最近

表 3: 扱う時間長の変化による PSNR/SSIM の比較

時間長 [s]	実験対象者 1	実験対象者 2	実験対象者 3
1	11.9 / 0.114	**12.0** / 0.112	11.9 / **0.118**
3	**12.1** / **0.119**	**12.0** / **0.115**	11.9 / 0.116
5	**12.1** / 0.114	11.9 / 0.110	**12.0** / 0.112
7	11.9 / 0.110	11.9 / 0.104	11.8 / 0.114
9	11.9 / 0.104	11.7 / 0.101	11.6 / 0.100

傍法よりも高いスコアを得ることを確認した．また，GAN を利用したモデルは GAN を用いないモデルよりも高い SSIM を出すことに成功した．

使用する脳領野を変化させたときは，全ての領野を使用するのが最もよい結果となった．また，入力のフレーム長に関しては，時間長 $d = 3$ のときがよい傾向になることが確認できた．これにより，再構成にとって有効な BOLD 信号は，視覚刺激から約 4 ～ 6 秒後に変化とともに発生していると考えることができる．

参考文献

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems (NIPS)*, 2014.

[2] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, Vol. 33, No. 4, pp. 1104–1116, 2006.

[3] 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*, Vol. 60, No. 5, pp. 915–929, 2008.

[4] Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, Vol. 63, No. 6, pp. 902–915, 2009.

[5] Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Gallant lab natural movie 4t fmri data. crcns.org., 2014.

[6] 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*, Vol. 21, No. 19, pp. 1641–1646, 2011.

[7] Guohua Shen, Kshitij Dwivedi, Kei Majima, Tomoyasu Horikawa, and Yukiyasu Kamitani. End-to-end deep image reconstruction from human brain activity. *bioRxiv preprint*, 2018.

[8] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Trans. Img. Proc.*, Vol. 13, No. 4, pp. 600–612, April 2004.

[9] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *International Conference on Learning Representations (ICLR)*, 2018.

[10] Takeru Miyato and Masanori Koyama. cgans with projection discriminator. *International Conference on Learning Representation (ICLR)*, 2018.

[11] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in Neural Information Processing Systems (NIPS)*, 2017.