拡散モデルを用いた脳内情報解読への取り組み

石﨑 文都　(指導教員：小林 一郎)

1　はじめに

近年，人工知能研究において，深層学習を使用してヒト脳活動から視覚体験を表現する多くの研究が行われている．Deep Neural Network（DNN）を用いた研究では，脳内の視覚情報の階層的な表現と DNN の間に相同性があることが明らかになり [1]，DNN の階層から視覚的特徴を組み合わせることでヒト脳活動から知覚内容の可視化が可能であることが確認された [2].

本研究では，視覚体験に関する脳活動デコーディングに注目して，脳活動データから画像特徴量を予測してヒトが何を見ているかを読み取ることを課題に掲げており，さらに拡散モデルを用いて予測特徴量からの画像生成を行うことで高精細かつ意味的に妥当な画像を出力する手法の開発を試みる.

2　研究概要

2.1　提案手法

本研究で提案する, 拡散モデルを用いた脳活動デコーディングの手法の概要を図 1 に示す．画像を閲覧している時に取得した脳活動データから刺激画像の特徴量を予測するために逆符号化モデルを構築する．逆符号化モデルは，予測画像特徴量と刺激画像特徴量が近づくように線形回帰を用いて重みを学習させる．そして，拡散モデルである Stable Diffusion [3] を使用して，予測した画像特徴量から刺激画像を再構成させる.

2.2　逆符号化モデル

ヒトの脳活動パターンは，体験する認知状態や行動などの刺激内容に応じて変化する．そのため，脳活動を解読することで，ヒトが実際に受け取っている内容を読み取ることができると考えられている．逆符号化モデル [4] は，特定の脳活動が計測されたときに，ヒトがどのような体験をしているのか推定するモデルである [5].

本研究では，刺激画像の内容を脳活動から予測することで対応関係をモデル化した.

2.3　拡散モデル

拡散モデル [6] は，生成モデルの一種である．一般に，ノイズからデータへの変換は難しいが，データをノイズへ変換することの逆変換として考えれば，容易に実現可能であるため, 拡散モデルでは, データに徐々にノイズを付与していく過程（拡散過程）を考え，拡散過程を逆方向に辿ってノイズの除去を繰り返すこと（逆拡散過程）によってデータを生成する．拡散過程で学習は必要なく，逆拡散過程でのデータ生成時のみ学習が必要となる.

本研究では，学習済みの重み[1]を利用して画像生成を行った.

Stable Diffusion はテキストで条件付けた高精細な画像を生成する拡散モデルである．高画質化や画像の一部を加工するなどの画像生成を可能にした Latent Diffusion Model（LDM）[3] をベースにしており, Variational Autoencoder（VAE）でピクセル画像を潜在的な埋め込み表現に変換し，低次元の潜在空間を利用することで効率的な学習が可能である.

3　実験

3.1　実験設定

使用データ　本研究では，Natural Object Dataset（NOD） [7] を用いた．NOD は，30 名の被験者から得られた 57,120 枚の ImageNet と MS-COCO の自然画像に対する反応を含む大規模 fMRI データセットである．本研究では，全実験に参加した 9 名の被験者のデータのうち，ImageNet 画像（合計 36,000 枚）による試行をデータとして用いた．被験者全体の画像中に提示された最も顕著な物体が生物であるかどうかの平均認識精度は 83.7% であり，sub-02 を除くすべての被験者が同程度の良好な成績を示した.

特徴量の抽出　逆符号化モデルの学習時に使用する画像特徴量は, Stable Diffusion 内の事前学習済みの VAE Encoder [8] の潜在層から抽出した.

逆符号化モデルの構築　脳活動データと VAE Encoder で抽出した画像特徴量を使用し，符号化モデルを構築した．脳活動の時系列を説明変数として画像特徴量の時系列を予測するモデルをリッジ回帰により学習した[2]．その際，神経活動に伴う血流の増加の反応時間（Hemodynamic response）を考慮し，fMRI で観測された脳活動データとその時系列の 2, 4, 6 秒前の特徴量と回帰を行なった．また，チャンクを 100 として訓練データをシャッフルした上で 10 分割交差検証を行い，平均の相関係数が最も良くなる正則化項を採用した.

9 名の被験者のデータのうち，被験者ごとの平均認識精度の高かった sub-06 のデータのみ，同程度の認識精度を示した sub-08 のデータも使用した 2 名分のデータ，認識精度の低かった sub-02 を除いた 8 名分のデータ，9 名のデータをそれぞれ使用した合計 12 個の逆符号化モデルを構築した.

画像生成　Stable Diffusion の事前学習済み拡散モデル[1]と VAE Decoder [8] を使用し，逆符号化モデルにより得られた予測画像特徴量から画像を生成した.

3.2　実験結果

脳活動データから生成した画像の例を以下に示す．また，刺激画像の画像特徴量と予測した画像特徴量でピアソンの積率相関係数を求め，それにより各モデルの予測精度を評価した.

図 2 に，各評価データの脳活動を用いて生成した画像の例を示す．被験者数が増えるにつれて，画像がぼ

[1] https://huggingface.co/CompVis/stable-diffusion-v-1-4-original で提供されているものを用いた.

[2] リッジ回帰は https://github.com/alexhuth/ridge で提供されているものを用いた.

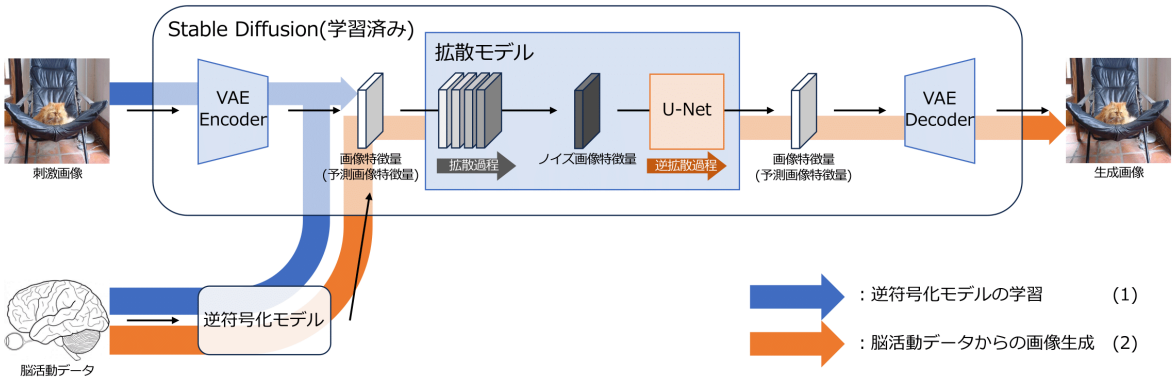

図 1: 本研究の概要図

やけることが確認された. なお，相関係数からは相関が見られなかった.

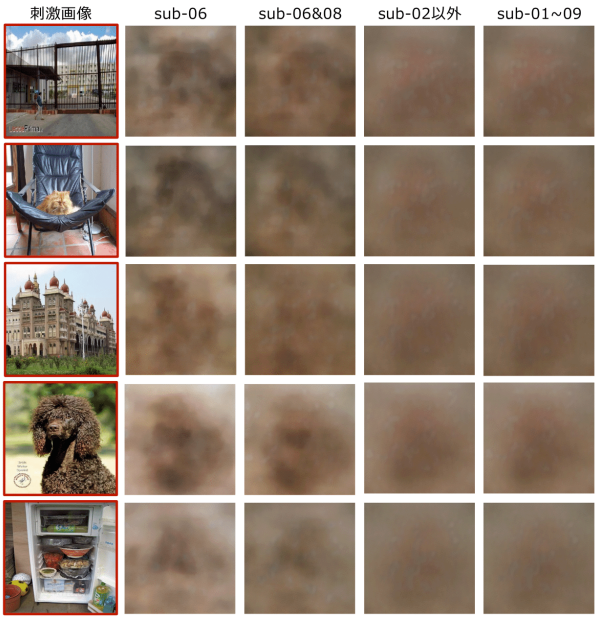

図 2: 評価データから生成した画像とその時閲覧していた画像. 図は 6 秒前の特徴量と回帰を行った例.

図 3 に，各訓練データの脳活動を用いて生成した画像の例を示す．被験者数が少ない結果ほど，概形をよりよく表現する画像を生成できており，刺激画像と生成画像が類似していることがわかった．次に相関係数の結果を示す．sub-06 のみの訓練データで学習した場合には 0.88，sub-06 と sub-08 の場合は，0.75 と，強い正の相関が見られた．また，sub-02 以外と 9 名の被験者のデータで学習した場合にも，0.45 と 0.44 で相関があり，相関係数においても類似性が高いことが示された.

4 まとめ

本研究では，脳活動から刺激画像の画像特徴量を予測する逆符号化モデルを構築し，高精細な画像の生成を可能にする Stable Diffusion を利用して，脳活動から画像再構成を行った.

その結果，訓練データで生成された画像において，刺激画像との類似性を確認されたが，評価データでは刺激画像に類似した画像を生成できたとは言い難い結果となった．今後は逆拡散過程においても画像特徴量

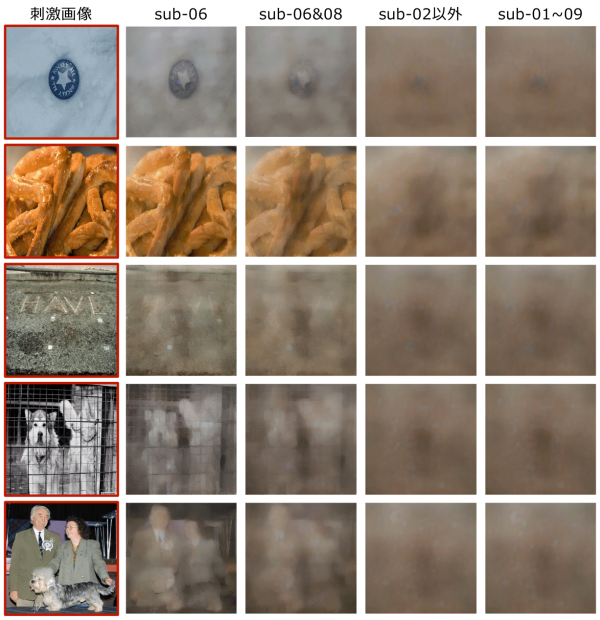

図 3: 訓練データから生成した画像とその時閲覧していた画像. 図は 6 秒前の特徴量と回帰を行った例.

を条件として活用することで，評価データでも高精度な画像再構成が行えるようにしたい.

参考文献

[1] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, Vol. 8, No. 1, p. 15037, 2017.

[2] Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani. Deep image reconstruction from human brain activity. *PLOS Computational Biology*, Vol. 15, No. 1, pp. 1–23, 01 2019.

[3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

[4] 西本伸志, 西田知史. 視覚と認知をつかさどる脳機能の定量的理解とその応用に関する研究. 情報通信研究機構ジャーナル, Vol. 64, No. 1, pp. 5–11, 2018.

[5] 宮脇陽一, 神谷之康. 《第 8 回》脳情報デコーディング技術とその応用. 計測と制御, Vol. 50, No. 10, pp. 888–894, 10 2011.

[6] 井尻善久. 生成 AI. コンピュータビジョン最前線 / 井尻善久 [ほか] 編, No. Summer 2023. 共立出版, 2023.

[7] Zhengxin Gong, Ming Zhou, Yuxuan Dai, Yushan Wen, Youyi Liu, and Zonglei Zhen. A large-scale fmri dataset for the visual processing of naturalistic scenes. *Scientific Data*, Vol. 10, No. 1, p. 559, 2023.

[8] Patrick Esser, Robin Rombach, and Björn Ommer. Taming transformers for high-resolution image synthesis, 2020.