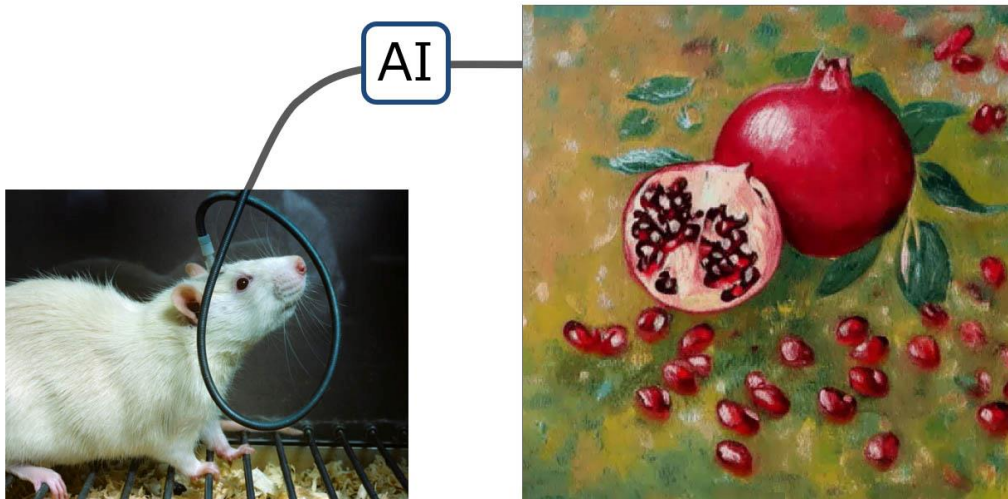


ネズミの脳で絵を描くことに成功

—脳と AI を接続することで実現—

発表のポイント

- ◆潜在拡散モデルとラットの脳波を組み合わせ、画像を出力するシステムを実装しました。
- ◆画像の元となる潜在ベクトルにラット皮質の局所場電位を採用することで、ラットの状態に応じて、刻一刻とリアルタイムで画像を出力させました。
- ◆この手法は、脳活動のみならず、心臓や腸蠕動などの他のバイオ信号や、風や波などの自然現象といった、あらゆる時系列信号に応用することができます。



ネズミの大脳皮質の神経活動から画像をリアルタイムで生成

概要

東京大学大学院薬学系研究科の山城皓太郎大学院生、池谷裕二教授らの研究グループは、ネズミ（ラット）の大脳皮質から局所場電位（脳波の一種）（注1）を記録し、これをもとに人工知能にイラストを描かせるシステムを構築しました。絵は、その時の脳の状態に応じて、途切れなく次々と絵が更新されていきます。今回は潜在拡散モデル（注2）という人工知能を活用しています。潜在拡散モデルは、ノイズから画像を作りだすことができる生成 AI のひとつです。今回はこのノイズ成分として神経活動を活用しました。描かれる画像のジャンルは使用するモデルを変えることで自由に設定することができます。脳活動と人工知能を融合した描画システムは前例がなく、今回が世界初の実施例となります。この技術は、脳だけでなく、心臓や腸蠕動などの他のバイオ信号や、風や波などの自然現象といった、あらゆる時系列信号に応用することができます。今回の技術は芸術の創作手法における新しいジャンルを拓くことが期待されます。描画の様子は https://ikegaya.jp/rat_brain_drawing.mp4 でご覧いただけます。

本研究成果は日本時間 2024 年 9 月 7 日の PLOS ONE 誌（オンライン版）に掲載されました。

発表内容

研究の背景

2022年8月に、たとえば「馬に乗る宇宙飛行士 (An astronaut riding a horse)」のような文章を入力すると、それに応じた画像を出力する「Stable Diffusion」がオープンソースとして Stability AI 社から公開されました。このモデルが公開される前にも、DALL-E や Midjourney のような、文章から画像を生成するサービスは存在していましたが、これらのサービスに利用されているモデルは一般公開されていません。一方で、Stable Diffusion モデルはオープンソース化され、各ローカル環境で画像を生成できるようになりました。また十分な性能のパソコンさえあれば、モデルそのものをカスタマイズし、自在に画像を生成することも可能になりました。最近では、このように個人で作成したモデルを共有するコミュニティも発達し、様々なジャンルの画像を生成することが可能になりました。今回構築したシステムでは、ラットの脳皮質から記録した局所場電位を直接リアルタイムで入力できるように、Stable Diffusion モデルを改造したフレームワークを作りました。

研究の内容

拡散モデルは、画像に加えられたノイズを除去し、元々の画像を復元することを学習します。その結果、純粋なノイズから画像を復元することが可能となります。学習過程では、拡散モデルは用意された画像セットの特徴を学習します。このモデルにガウスノイズなどのランダムなノイズを入力することで、学習した画像のスタイルに沿った全く新しい画像を生成することが可能になります。

Stable Diffusion のような文章から画像を生成するモデルでは、入力された文章をノイズから画像を復元するときの指針とすることで、文章に従った画像を復元することができます(図1)。つまり、文章から画像を生成する拡散モデルでは、生成の元となるノイズと、指示文章の二つの入力を受けとり、文章に従ってノイズから画像を作り出します。一方で、指示文章をあえて入力しないことで、入力ノイズのみに依存した画像も生成することが可能です。

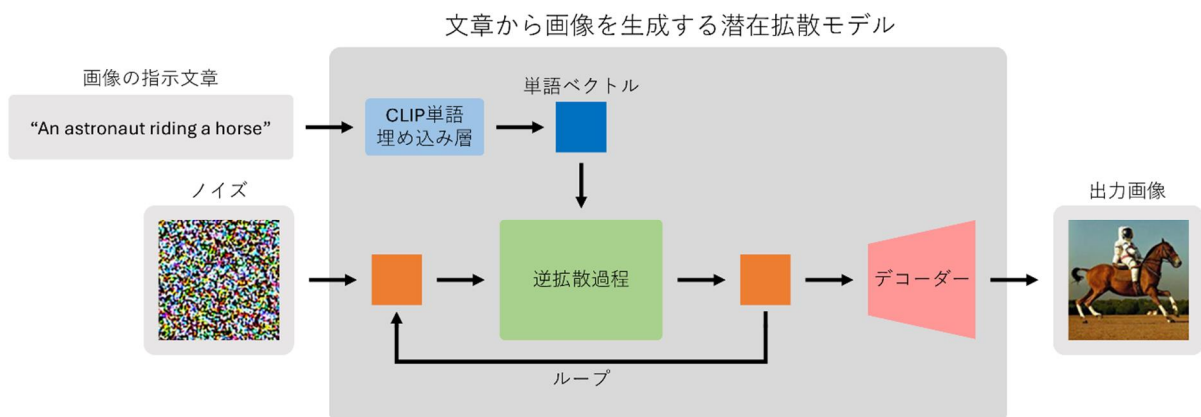


図1：潜在拡散モデルの仕組み

生成したい画像の指示文と、ノイズ（橙色）を入力とする。指示文は、文字埋め込み層によって単語ベクトル（青）に変換される。それぞれが拡散モデルに入力され、ノイズから画像が復元される。最後に潜在空間のベクトルがデコーダによって画像の次元に拡張され、画像として出力される。

今回のシステムでは、Stable Diffusion の指示文章は入力せず、画像の元となるノイズの代わりにラットの脳皮質から記録した局所場電位を入力しています(図2)。局所場電位は、

ニューロンが発する信号を記録したものであるため、波の形をした時系列データです。このままでは、Stable Diffusionに入力することができないため、局所場電位の次元を Stable Diffusion のデータ形式に合うように圧縮しました。

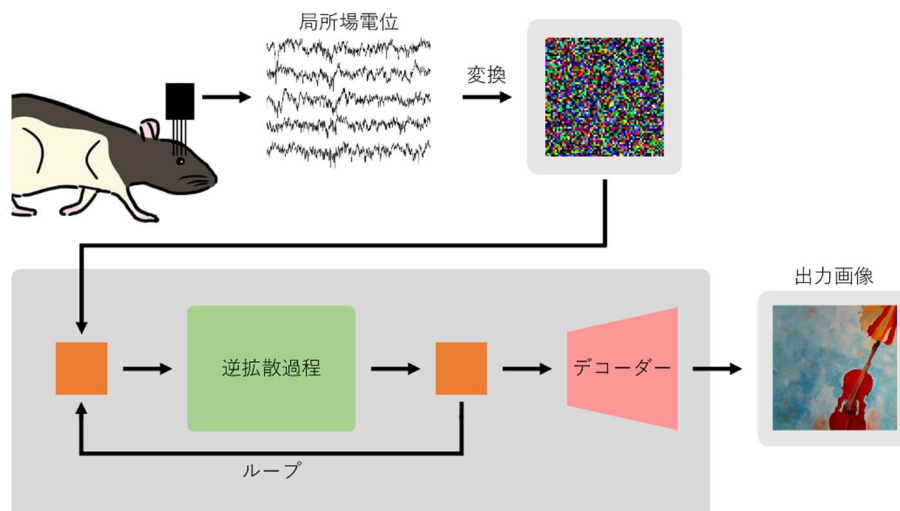


図 2：実装したシステムの構造

ラットの大脳皮質から記録した局所場電位の変換方式。1秒間の局所場電位を潜在ベクトルの形に変形する。局所場電位を変換して作った潜在ベクトルを潜在拡散モデルに入力することで、画像が得られる。なお、指示文章は入力しない。

局所場電位は1秒の区間を1/30秒ずつずらして切り取り、その時間窓に応じて1/30秒ごとに画像が生成されます。局所場電位は時系列データなので、1/30秒前の脳波と現在の脳波は類似しています。これにより画像間の移行がスムーズになり、脳波の変化に応じて徐々に変化する画像が生成されます（図3）。

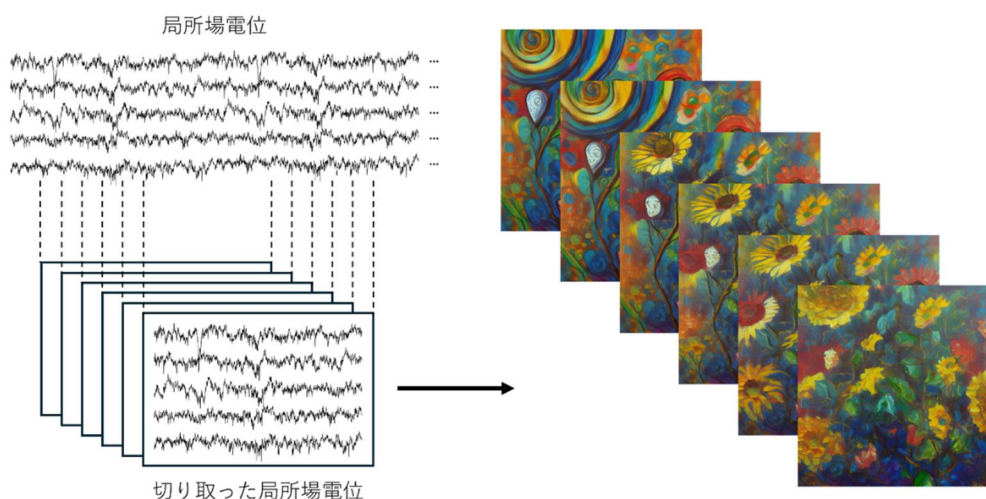


図 3：連続的に変化する画像の実装

1秒間の局所場電位を1/30ごとに切り出し、潜在ベクトルに変換した。時間的に隣り合う潜在ベクトルが類似しているため、潜在拡散モデルを通して得られた連続する画像はなめらかな推移を示す。

今後の展望

画像の元となるノイズの代わりに局所場電位を利用することで、ラットの大脳皮質の神経活動に応じた画像が出力されるシステムを実装しました。Stable Diffusion はノイズのほかに指示文章を入力として受けとります。この文章の内容にラットの内部状態を反映させることができれば、ラットが興味を持っているときには明るい雰囲気の写真、眠たいときには静かな雰囲気の画像、といったようにラットの「気分」に応じて画像を生成できるようになります。

今回開発された手法は、原理的に、神経活動だけでなく、心臓や腸蠕動などの多くのバイオ信号、風や波などの自然現象といった、あらゆる時系列信号に応用することができます。この技術は、芸術の創作手法における新しいジャンルを拓くことが期待されます。

発表者・研究者等情報

東京大学大学院薬学系研究科

池谷 裕二 教授

山城 皓太郎 博士課程

論文情報

雑誌名：PLOS ONE (9月6日オンライン版)

題名：Diffusion model-based image generation from rat brain activity

著者名：Yamashiro, K., Matsumoto, N., Ikegaya, Y.*

DOI：10.1371/journal.pone.0309709

URL：<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0309709>

研究助成

本研究は、「戦略的創造研究推進事業ERATO池谷脳AI融合プロジェクト(課題番号:JPMJER1801)」、「科学研究費補助金(課題番号:22K21353)」、「AMED CREST(課題番号:24wm0625401h0001)」、「Beyond AI 研究推進機構」の支援により実施されました。

用語解説

(注1) 局所場電位

局所場電位(LFP, local field potential)とは、脳波の縮小版のようなもので、脳の電氣的活動を局所レベルで測定したものです。LFPは、ニューロンが発生する電氣信号を検出する電極を使用して測定します。電極を脳の様々な部位に設置することで、その局部で起きている電氣的活動を拾い上げ、その場所のニューロンがどのような活動をしているかを大雑把に知ることができます。

(注2) 潜在拡散モデル

ノイズを起点とし、徐々にノイズを除去していくことで、データへと変換していくモデルを拡散モデルと呼びます。ノイズを追加してデータを劣化させる関数と、このノイズを除去するための単純な画像復元ネットワークという2つのパーツから構築されています。このモデルには様々な応用が考えられますが、画像生成の分野では、OpenAI社のOpenAIを皮切りに、2022年に強力なルーツが発表されました。今回の研究では、Stability AI社のStable diffusionを用いています。

問合せ先

〈研究に関する問合せ〉

東京大学大学院薬学系研究科

教授 池谷 裕二 (いけがや ゆうじ)

Tel : 03-5841-4780 E-mail : ikegaya@mol.f.u-tokyo.ac.jp

〈報道に関する問合せ〉

東京大学大学院薬学系研究科 庶務チーム

Tel : 03-5841-4702 E-mail : shomu@mol.f.u-tokyo.ac.jp