

〔論 文〕

画像生成 AI による プロジェクションマッピング動画制作 ——Stable Diffusion Deforum の活用と制作成果——

赤 井 良 行

要 旨

本研究では、プロジェクションマッピングの映像制作における Stable Diffusion Deforum のような生成 AI の活用について論じた。生成 AI により初学者でも高品質な映像を短時間で制作できるメリットがある一方で、言語化の難しさや動きの表現が課題となった。しかしながら、生成 AI はクリエイターの創造性を補助するツールとして機能し、人間と AI の緊密な協働により、新しい映像表現が生まれると期待される。従来のスキルに加え、新たな能力が求められることから、映像制作の形態や評価基準に変革をもたらす可能性がある。

キーワード：プロジェクションマッピング, 生成 AI, 映像制作

1. はじめに

近年、人工知能の急速な進歩により、生成 AI (Generative Artificial Intelligence) による画像生成技術は急速な進歩を遂げている。2015 年 Google によって開発された DeepDream¹⁾ の発表はアートおよびデザインの世界でもセンセーショナルなものであった。特に、Foster The People のミュージックビデオ Doing It for the Money²⁾ に使用され話題になった。DeepDream の特徴は、画像認識モデルがイメージ内の特定のパターンを検出し、それを強調して幻想的な視覚効果を生成する技術であった。また使用にはプログラミングの知識が要求され直感的に操作することはできなかった。

それに対し、2022 年に登場した Midjourney³⁾ や Stable Diffusion⁴⁾ に代表される、テキスト入力に基づく画像生成 (text-to-image) はそれ以前のシステムと大きく異なり、初学者でも容易に扱えるようになり、芸術的な表現から科学的な探索まで、幅広い領域で革新

的なツールとして活用されている。

その中でも、Stable Diffusion Deforum (以下 Deforum)⁵⁾ は Stable Diffusion の追加機能として開発された映像生成プラグイン (plug-in) で、Linkin Park のミュージックビデオ Healing Foot⁶⁾ などにも使用されている。Deforum は、従来の映像制作手法よりも直感的であり、専門的な知識やスキルを必要とせず、初学者でもプロフェッショナルなレベルの高品質な映像を生成できる可能性がある。

その一方で、Deforum は複雑な AI アルゴリズムを使用しており、その背後にある理論や原理を理解する必要がある、初学者にとっては理解が難しい場合がある。

そこで本論文では、プロジェクションマッピングの映像制作を中心に、Deforum の活用方法に焦点を当てる。Deforum は専門的な知識やスキルを必要とせず、初学者でもプロフェッショナルなレベルの高品質な映像を生成できる可能性がある。

一方で、Deforum は複雑な AI アルゴリズム

を使用しており、その背後にある理論や原理を理解する必要があり、初学者にとっては理解が難しい場合がある。

そこで本論文では、プロジェクションマッピングの映像制作を中心に、Deforumの活用方法に焦点を当てる。

論文では、2023年度9月と2月に、大阪府松原市にある大型ショッピング施設セブンパーク天美で行ったプロジェクションマッピングで使った映像制作をもとに検証を行う。

2. Stable Diffusionの仕組み

Stable Diffusionは、拡散過程を通じて学習モデルを構築し、逆拡散過程を経て画像を生成する。拡散過程では、鮮明な画像に徐々にノイズを加え、最終的には完全なノイズ画像を生成する。このノイズ画像を1つのオブジェクト(例えばイヌ)に対して複数の画像を生成することで、学習モデルを構築する。一方、画像生成の過程では、逆拡散過程により、ランダムに生成されたノイズ画像からノイズを少しずつ取り除き、鮮明な画像を生成する⁷⁾。

3. 生成AIと著作権問題

現在、AIによって生成された画像の著作権問題がしばしば問題視されている。この問題の核心は、学習モデルに使用される画像にある。特定のイラストレーターや画家の画像を用いて学習モデルを構築すれば、その作家の作風・画風などを模倣した画像を生成することが可能である。

しかしながら、著作権法の基本的な考え方は、作風・画風自体は保護の対象とはならないとされている。一方で、生成された画像の著作権は、人によって描かれた場合と同様に「類似性」および「依拠性」により判断され、場合によっては著作権侵害と見なされる可能性がある⁸⁾。したがって、公開を必要とする作品を生成する場合、学習モデルの選択には十分注意を払う必要がある。

4. Deforumの操作手順

Deforumの基本的な操作手順は以下のとおりである。ノイズを生成するための数値「Seed」(ノイズ生成の種)、ノイズを除去する手法「Sampling method」、画像の動きを決定する「Motion Parameters」、使用する学習モデル選択、そして映像の内容を指定する「プロンプト」の設定によって映像の生成を行う。

このうちSeed、学習モデル、Sampling methodが画風や画像の品質を決めるMotion Parametersは映像生成に必要な動きを設定する項目である。ここでいう動きとはキャラクターなどの動きではなく、絵巻物 右から左に鑑賞していく感覚に近い。この際プロンプトによって指定された内容に沿って画像を移行していくためアニメーション効果が生まれる。

Motion Parametersには、X軸およびY軸の移動と、前後の移動を設定するZoom(拡大・縮小)がある。Zoomによる変化を図1に示す。

図1で示した画像は、学生が2023年9月に実施したハロウィンをテーマにした映像の一部である。画像生成に使用したプロンプトは表1

表1：使用したプロンプト

"0" : "Pumpkin field, autumn, lively, harvest, style by pénélope, style by georg Hallencleben, simple illustration"
"20" : "Children, fun, Style by Dick Bruna, simple illustration",
"40" : "Children, pumpkin, digging up, smiles, Style by Dick Bruna, simple illustration 1",
"80" : "Jack Olantin, children, open the door, Style by Dick Bruna, simple illustration",
"100" : "Jack Olantin, Magic, Bats, Makai, Style by Dick Bruna, simple illustration",
"120" : "Jack Olantin, Bats, Children, Dancing, Makai, Style by Dick Bruna, simple illustration"

に示すとおりである。

図 1 の画像は、表 1 内の“0”のプロンプトによって生成されたものである。0 フレームから 20 フレームの間で、最初に設定した内容から 2 つ目の内容に画像が移行していく。これと同時に、画像の崩壊や、不適切な内容や望ましくない要素が生成されるのを防ぐため、Negative Prompt の設定を行う。

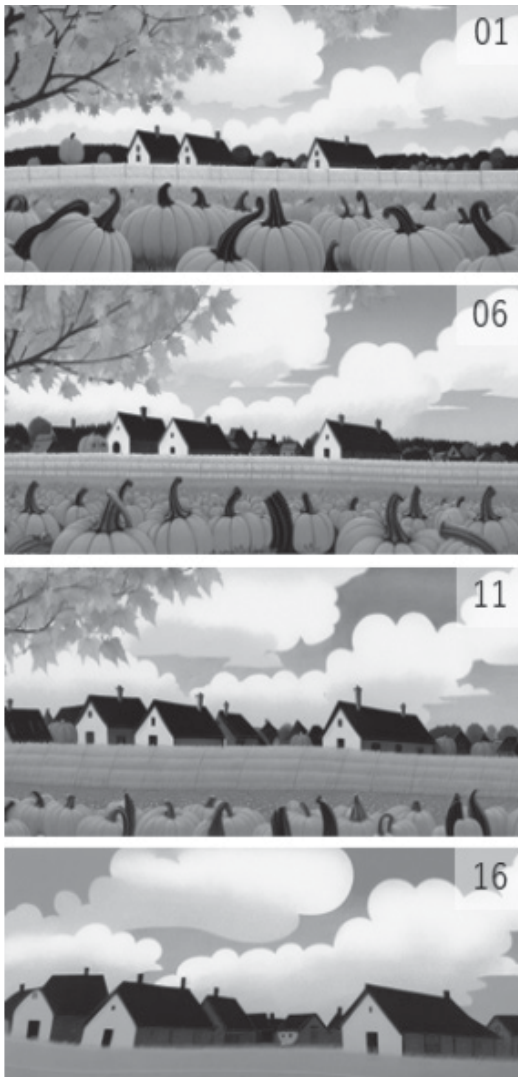


図 1：設定値“Zoom”：“0：(1.04)”1 フレームから 16 フレームの変化

5. 映像制作環境と構成

Stable Diffusion を快適に動作させるには、VRAM (Video Random Access Memory) が 10GB を超える GPU (Graphics Processing Unit) を搭載したパソコンの使用が推奨される⁴⁾。そのため、GPU を搭載していないノートパソコンでは、生成が不安定になる可能性がある。そこで本制作では、Google Research の Google Colab (正式名称は Colaboratory) を利用した。Colab はブラウザ上で Python を実行でき、クラウド上の GPU を無料で使用できる。すべての Stable Diffusion の操作がクラウド上で実行されるため、性能の低いパソコンでも実行が可能となる。また、常に Colab で実行可能な最新の Deforum が公開されており、プログラミング知識がなくても利用できる。プロンプト制作には、Google Spreadsheet で Google 翻訳の関数を呼び出し、日本語から自動で英語のプロンプトを作成するシステムを構築した。同時に、映像生成に必要なフレーム番号などを追加する機能も同一システム上で提供した。このシステムの導入により、プロンプトの作成プロセスが大幅に効率化され、ユーザーは翻訳や英単語検索の手間を省き、迅速かつ正確に英語のプロンプトを生成できるようになった。また、Deforum での映像生成に必要な情報を簡単に追加できるため、作業の効率も向上した。このようなシステムを導入したことで、英語に自信のないユーザーや専門的な知識が必要な場合でも、プロンプトの作成が容易になり、クリエイティブなアイデアの実現が促進されるようになった。

6. Deforum を活用したプロジェクションマッピング映像制作と投影

本章では実際に制作を行い、投影した映像について述べる。

1) 投影場所の相性と選定

プロジェクションマッピングでは、通常、立体物の形状や構造に合わせて映像を制作し、その立体物に映像を投影（マッピング）する。そのため映像は構造物の形状を意識して制作する必要がある。しかし、本研究で使用したDeforumは、生成される映像が矩形に限定されるため、生成画像内でオブジェクトの位置を完全に制御することができない。そのため、複雑な形状の立体物にマッピングする場合、中心となるオブジェクトが投影対象のアウトラインから外れてしまう可能性がある。マスクなどの機能を利用すれば、投影対象の形状に合わせた映像生成は可能である。しかし、本制作の参加者は映像制作経験がないため、できる限り複雑な処理は避ける必要があった。そこで、大型の矩形壁面を投影面とするセブンパーク天美投影場所に選定した。同施設は大規模ショッピングモールであり、集客イベントとしてプロジェクションマッピングを定期的実施している。2023年度は9月にハロウィン、2月にバレンタインをテーマにしたプロジェクションマッピングを実施した。

実施したプロジェクションマッピングの映像では、ハロウィンで投影した20分21秒の映像のうち9分52秒、バレンタインの映像12分44秒のうち6分37秒をDeforumによって生成した映像を使用した。

バレンタインの投影様子を図2に示す。

2) Deforumを用いた映像制作プロセス

Deforumのプロンプトでは生成したい映像のテーマを決定しなくてはならない。例えば今回の制作ケースでは、プロジェクションマッピングのテーマ“Halloween”や“Valentine’s Day”がそれにあたる。また同時に“animation style”のようにスタイルを決定するワードを加えることにより、安定した画像の生成が行える。このスタイルには著名な映画や画家を指定することもできる。この場合、第2章で述べたように、生成に用いる学習モデルの選択を慎重に行わな



図2：バレンタイン投影（2024/2/12）

いと、著作権侵害に至る可能性がある。プロンプトの制作には何度か試行錯誤しながら調整していく必要がある。生成される映像が望みどおりであるかどうかを確認し、理想の映像になるよう修正を加えることで、イメージした映像を生成できるようになる。

映像制作初期の段階で、学生たちが使用したプロンプトには、“Children, it looks fun”や“a child to make Jack-O’-Lantern”のように文による表現が見られた。

また“fun”, “enjoy”, “happily”など感情を表す単語, “walking”, “make”, “calling”のように動作を表す単語が多く使われていた。

Deforumの特徴はStable Diffusionの画像生

成機能に動きを加え、連続した画像を生成するところにある。ベースになる Stable Diffusion の学習モデルは静止画を用いて作成されている。例えば、イヌの画像にタグ付けされる単語は“dog”であり、楽しそうに走っているイヌの画像も、寝ているイヌの画像も同じ“dog”なのである。例えば、芋掘りの様子を生成するためにプロンプトに“digging potatoes”を用いた場合、“digging”は無視され、大量のジャガイモの画像が生成されることがある。そこで、“soil”、“potatoes”、“farmer”といったように、画像に収めたい対象をプロンプトに組み入れるなどの調整が必要となる。

実際に映像生成に使用された調整後のプロンプトを確認すると、“Halloween party, children, Jack-O’-Lantern, dance”, “Castle, bat, magic world, children, dance”といったように、名詞による表現に変化していた。また、楽しい状況の表現には、形容詞ではなく、“party”のように楽しいことを連想させる名詞や、“bright colors”, “colorful”のように色彩による表現が行われていた。このように画像生成 AI を用いて感情などを表現する場合、形容詞を用いるよりも、表現したい感情に結びつく記号を持った名詞を用いるほうが有効的であることがわかる。

生成 AI によるアートではしばしば創造性を制限や、人の感性や表現力の損失がしばしば懸念される。しかしながら、この記号の選択は明らかにプロンプト作成において人の感性や表現力が介在するものであり、新しい創造性の可能性であるのではないかと考えられる。

3) Deforum の制約と克服策

本制作では、初学者による試みであったため、Deforum の機能をフルに使用できなかった。そのため特殊な形状の映像の生成には至らなかった。この点に関しては今後の課題として大きなものではないと考えられる。

しかしながら、意図したとおりの映像を生成するためにはかなりの制約がある。例えば、学

習モデルの制限がある。学習モデル構築には基本となるデータセットが使用されている。これらの多くはインターネット上に公開された画像から構築されている。そのため、非常に特殊なまたはニッチな画像に対しては、適切な画像を生成するのが難しい。一例として、“大仏”がある。“大仏”から連想されるイメージは、西日本では奈良、東日本では鎌倉が一般的であろう。“Great Buddha”を用いて画像を生成すると、鎌倉スタイルもしくは東南アジアスタイルの大仏が生成されることがほとんどである。その理由は明白で、“Great Buddha”を用いて Google などの画像検索を行うと、この 2 つのスタイルの画像が圧倒的に多いからである。

もちろんオリジナルの学習モデルを構築することも可能ではあるが、汎用性の低い特殊な学習モデルになってしまい、非効率である。また、キャラクターの動作や感情表現が難しい課題もある。

これらの問題を克服する手法として、より多くの画像を生成し、「Seed」, 「Sampling method」, 「プロンプト」といった画像生成に必要なパラメーターのデータベース化が考えられる。Stable Diffusion では同じパラメーターを用いて画像の生成を行った場合、ほぼ同じ画像の再現が可能である。

また、言語化が困難なプロンプト作成には Stable Diffusion に搭載された Interrogate CLIP や Interrogate といった画像からプロンプトを抽出する機能の活用が考えられる。

7. 課題と今後の展望

本研究で使用した Deforum は、アニメーション制作未経験の初学者でもプロフェッショナルなレベルの高品質なアニメーション映像を生成できる。その反面、言語に基づく画像生成のため、言葉選びが重要である。特に、生成 AI は動詞を理解しにくく、生成される映像は静止画の連続で構成されるため、キャラクターを動かすなど動きを制御するには工夫が必要である。同

時に、生成される映像は修練によって一定程度コントロールできるが、ギャンブル的要素も大きい。

しかしながら、ミュージックビデオやプロジェクトンマッピングのように映像美を要求される反面、ストーリー性を強く要求されない映像作品制作においては、生成AIによる映像生成に技術習得時間を割り当てることで高品質なアニメーション映像を生成できるメリットがある。このため、プロジェクトンマッピングにおいて生成AIを活用することは有用であると考えられる。また、従来のAdobe AnimateやAfter Effectsと組み合わせることで、コントロールされた質の高い映像を提供できると期待される。

さらに、2024年2月にOpenAI社によって公開された動画生成AI「Sora」⁹⁾のような生成AIが今後続々と公開されることが期待される。これにより、Deformが不得意とするモーションもより詳細に表現できる可能性があると考えられる。

8. 考察

プロジェクトンマッピングのような、映像美が重視される分野では、生成AIの強みが発揮されることが考えられる。本制作では、矩形の画面に投影することを前提に制作を行った。また、ハロウィンやバレンタインといったテーマに沿って生成を行ったため、具象的な映像となった。しかしながら抽象的で幻想的な映像表現は、生成AIを使用すれば、初学者でも容易に生成することができる。これらの映像を複雑な形状の構造物に合わせて投影を行ったり、構造物内の形状にマッピングしたりすることで複雑なプロジェクトンマッピングにも十分応用可能であると確認できた。短期間に大量の映像を生成できるメリットは、新しい表現領域を切り拓く革新的な手法として期待できる。

9. まとめ

かつて、写真がフィルムに収められた時代、写真家に要求されたセンスは、現場での光や空気からフィルムに収められるであろうイメージを想像し、適切な画像を可能な限り収めるための設定を割り出して撮影する現場での力であった。さらに、現像の段階でも、限られた枚数の中から最高の1枚を選択し、現像後に現れるイメージする力と技術も求められた。

現代のデジタル時代において、写真家に求められる能力は、現場において選択可能なシチュエーションを記録し、大量のデータの中から最高の1枚を選択し、より効果的にフォトタッチを行うセンスに変化した。

映像制作においても、短時間に大量の映像を生成し、その中から最適な映像を選択することが映像クリエイターにとって必要な能力になることが十分に予想される。また同時に、テクニカルなスキルよりも、企画力やディレクション能力、AIとのコミュニケーション能力が重視されるようになって考えられる。

特に、初学者の場合、自分の能力と比較し制作可能な作品を選択してしまう傾向がある。プロジェクトンマッピングのように、公共の場所で展示・投影する作品の場合、そこに来場する観客が作品に要求するものは、制作時間の長さや、制作者の努力ではない。彼らが要求するものは、映像美やエンターテインメント性など作品の質である。生成AIによって大量に生成した映像の中から最適なシーンを選択できる能力を育成することは、映像制作に必要なアプリケーションの習得や、高度な制作スキルを得るよりも、心理的および時間的ストレスが軽減されると考えられる。AIを使えば一定水準の映像制作が可能になり、制作の敷居が下がり、初学者でもクオリティーの高い映像制作に挑戦しやすくなる。また、企画やディレクションに十分な時間を割くことができ、よりクオリティーの高い作品を展示・投影できると考えられる。さらに、映像企画の段階での導入はアイデアの円

滑な具現化に繋がるメリットがあり、想像することのできなかった抽象的で幻想的な映像のアイデアを偶然得ることも十分あり得る。

AIは究極的にはクリエイターの創造性を補助するツールとして機能すると考えられる。クリエイティブ活動において人は自分の経験や記憶の中からアイデアを創造する。言い換えれば、過去の作品の模倣と改変の繰り返しによって新しいアイデアは生み出されてきた。AIによる創造物は、過去にはなかった映像世界や芸術表現を生み出す可能性がある。

AIの創造力が進化すれば、人間とAIが緊密に協働して作品を制作する機会が増えていくと予想される。人間がコンセプトを立て、AIが下準備を行い、人間がブラッシュアップを加える、といった具合に役割分担しながら創作を行うプロセスが一般化する可能性は容易に予想される。

アートの世界は常に新しい技術の登場に伴いスクラップ・アンド・ビルドを繰り返し進化してきた。AIは、アートの形態や制作プロセスだけではなく、評価基準に大きな変革をもたらし、本質的な価値とは何かを問い直すきっかけを生むことであろう。

付 録

- ・2023年9月23日 セブンパーク天美 プロジェクションマッピング
<https://youtu.be/nTZBlwOBpSA>
- ・投影動画 (オリジナル)
<https://youtu.be/VEzHZu3oD1s>
- ・2024年2月12日 セブンパーク天美 プロジェクションマッピング
<https://youtu.be/CHUWMXg1G9U>
- ・投影動画 (オリジナル)
<https://youtu.be/AciOB322z2Y>
- ・2023年12月23日ならら プロジェクションマッピング (吹奏楽部演奏の背景にDeformで制作した映像を投影)
<https://youtu.be/ZFxxULA6Whw>

謝 辞

本研究にあたりまして、阪南大学において、プロジェクションマッピングの分野で先駆的な功績を残された

田上博司学長に、深く感謝申し上げます。着任当初、私にはプロジェクションマッピングに関する知識が全くありませんでした。しかしながら、学長からお受け継ぎしたMacBook Air内の貴重な資料や、田上ゼミ出身の学生たちの経験や知識のおかげで、プロジェクトを成功裏に遂行することができました。心より感謝の意を表する次第です。

参考文献

- 1) Alexander Mordvintsev, Ohad Fried, Mike Tyka, Christopher Olah. (2015年6月18日). Inceptionism: Going Deeper into Neural Networks. 参照日: 2024年3月4日, 参照先: Google Research: <https://blog.research.google/2015/06/inceptionism-going-deeper-into-neural.html>
- 2) People The Foster. (2017年8月12日). Foster The People - Doing It for the Money (Video). 参照日: 2024年3月4日, 参照先: YouTube: <https://www.youtube.com/watch?v=dJ1VorN9Cl0>
- 3) Midjourney. Midjourney. 参照日: 2024年3月4日, 参照先: <https://www.midjourney.com/home>
- 4) CompVis. CompVis/stable-diffusion: A latent text-to-image diffusion model. 参照日: 2024年3月4日, 参照先: github.com: <https://github.com/CompVis/stable-diffusion>
- 5) Deforum Studio. (2022年11月24日). deforum-stable-diffusion. 参照日: 2024年3月4日, 参照先: GitHub: <https://github.com/deforum-art/deforum-stable-diffusion>
- 6) Linkin Park. (2023年4月7日). Linkin Park - Healing Foot (Official Audio). 参照日: 2024年3月4日, 参照先: YouTube: <https://www.youtube.com/watch?v=wsCdnXu4Nxc>
- 7) Robin Rombach, Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, Andreas. (2022年4月13日). High-Resolution Image Synthesis with Latent Diffusion Models. 参照日: 2024年3月6日, 参照先: arXiv: <https://arxiv.org/abs/2112.10752>
- 8) 文化庁著作権課. (2023年6月19日). AIと著作権. 参照日: 2024年3月4日, 参照先: 令和5年度著作権セミナー: https://www.bunka.go.jp/seisaku/chosakuken/pdf/93903601_01.pdf
- 9) Sora. 参照日: 2024年3月6日, 参照先: OpenAI Sora: <https://openai.com/sora>
- 10) 中川 譲 (2023) 画像生成 AI による作画表現の実用性と教育への応用, その正当性と法的妥当性について, 東京工芸大学芸術学部紀要, 29, pp. 23-29
- 11) 瀧雅音, 山路和希, 小笠原賛, 高橋創, 早瀬幸彦,

酒井義幸, 古川智之, 海老原靖子, 伊藤央, 高木友博 (2023) Stable Diffusion を用いたファサードデ

ザインの編集, 人工知能学会全国大会論文集 第37回 (2023), 1N5GS1002-1N5GS1002