

デジタルクロージングに対する AI デザインツール開発

高石圭人[†] 徳井直生[†]
慶応義塾大学[†]

概要: 本稿では 3DCG 空間内で使用する衣服であるデジタルクロージングに対する AI を使用したデザインツールを提案する。デジタルクロージングは XR 技術をはじめとした 3DCG 技術の発展やメタバース概念の普及によって、バーチャル試着や 3DCG 空間内のアバタに使用する用途として注目を集めている。3DCG 空間内でも現実空間と同じようにアイデンティティを表現するためのメディアとしてデジタルクロージングは有用であると考えられる。デジタルクロージングの利用がより一般的になった際、3DCG に関する専門的な知識を有さないユーザーでも直感的に個人が望むデザイン編集が可能なツールに対する需要が高まると考えられる。本稿ではテンプレートメッシュ、デザインを指定するテキスト、Inpaint によって生成されたマスクを入力とする、デジタルクロージングに対する AI を使用したデザインツールを提案する。また本稿で提案する AI を使用したデザインツールがユーザーの創造性を拡張するような形でデザイン編集を行うことが可能か検証、考察を行う。

1. はじめに

昨今、XR 技術を初めとした 3DCG 技術の目覚ましい発展により VRChat、Cluster などのメタバースと呼ばれている新しいプラットフォームが普及している。このような 3DCG 空間内でアバタが着用する衣服の 3D モデルであるデジタルクロージングは注目を集めている[1]。現実空間において衣服は社会階級の表現や、自分自身のアイデンティティを表現するための視覚的なメディアとして重要されており、3DCG 空間でも同様な役割を果たすと考えられる[2]。近い将来、より VRChat、Cluster などのメタバース空間が一般化し、アバタの利用が普及した際に 3DCG、デザイン、衣服に関する専門的な知識を有さないユーザーでも直感的に個人が望むようなデジタルクロージングのデザインを編集することを可能とするツールに対する需要が高まると考えられる。

昨今では StableDiffusion[3]、DALL·E2[4]などの Text-to-Image、Image-to-Image の AI モデルをベースとしたツールの普及によって、イラストやデザインに関する専門的な知識を有さないユーザーでも簡単に望んだイラスト生成を行うことが可能となった。しかし、デジタルクロージングに対するデザイン編集は Blender や CLO3D などが存在するが、3DCG や裁縫に関する専門的な知識が必要であり、そのような知識を有さないユーザーには容易にデザインを編集することは難しい。また、AI モデルを活用した直感的なデザイン編集ツールもまだほとんど提案されていない。

本稿では 3DCG、裁縫などに関する知識を有さないユーザーでも直感的にデジタルクロージングに対するデザイン編集を可能とするツール開発を目的とする。また、3DCG に対する機械学習は GPU リソースを多く必要とするため、デザインの編集操作を行うためのユーザーインターフェ

ース(図 1)とバックエンドとして実際にメッシュの変形を行う機械学習処理に分離し、ユーザーインターフェースであるフロントアプリのみをユーザーへ配布することで、高価な GPU を必要とせずツールの使用を可能とすることを目指す。本稿では、テンプレートメッシュと自然言語を入力することによってメッシュを変形させることが出来る Text2Mesh[5] に Inpaint 機能を追加したモデルをバックエンドとして使用し、Unity を用いて実装されたデザイン編集を行うためのフロントアプリを通してデジタルクロージングに対する AI 活用したデザイン編集ツールの提案を行う。また、AI を活用し、制作を行うことはユーザーに対して専門的な知識を要求せず、直感的な操作を可能とするだけでなく、ユーザーの意図を汲み取りつつもユーザーが意図していなかった結果をしばしば出力することがある。このような AI ツールが既存の制作のワークフロー内に介入することによって、ユーザーが本来持っている創造性を拡張することが出来るのではないかと考えられ、その点について実験と考察を行った。

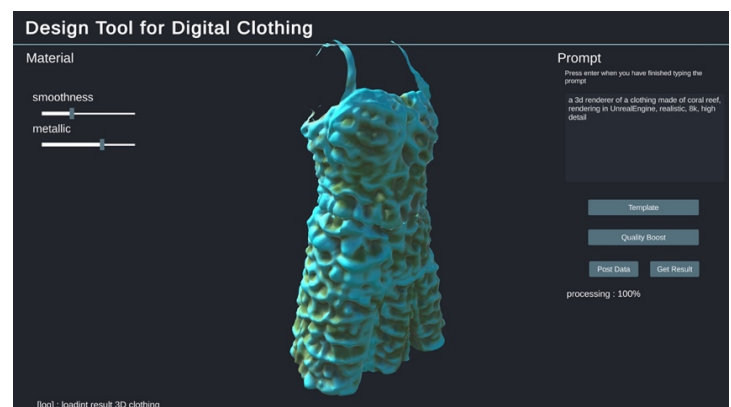


図 1 ユーザーインターフェース

2. 関連研究

2.1 衣服に対するデザインツール

スケッチをベースとした衣服に対するデザインツール既に提案されている[6][7]. Wibowoら[6]による手法ではガイドのマネキンに対して提案されたタンジブルツールを使用して、3DCG空間上でデザインされた衣服の制作過程を確認しつつ現実世界の衣服のパターンを制作することを支援している. また、何ら[7]による手法では、衣服の単視点のスケッチ画像から符号なし距離関数を生成することでデジタルクロージングのデザイン支援システムが提案されている. 本稿ではスケッチではなく、テキストを入力としたより容易に操作が可能なデジタルクロージングに対するデザインツールを提案する.

2.2 Text を入力とした画像生成

CLIP[8]はインターネットから収集した4億の画像とテキストデータをデータセットとして、画像の共同埋め込み空間を学習することに成功し、テキストと画像の類似度を測ることを可能とした. StableDiffusion, DALL・E2などのText-to-Imageモデルは生成された画像とガイドテキストの類似度をCLIPによって算出し、類似度が高まるように生成モデルのパラメータが最適化される. このようなText-to-Imageモデルを利用した画像生成サービスを使用することで、デザインの知識や、イラストなどを描く技術がないユーザーでもテキストを用いて直感的に高品質な画像を出力することが可能となった.

2.3 Text を入力とした 3D モデル形状操作

Text2Meshは入力としてテンプレートメッシュとテキストを入力とする. テンプレートメッシュを複数視点からレンダリングした画像 I_θ に対して、射影変換やクロップなどのデータオーギュメンテーション φ を敵応する. それらによって結果とガイドテキスト T をCLIP Encoder E を用いて共同埋め込み空間内のベクトルに変換し、類似度を算出する. そして算出された類似度 L_{sim} を高めるように微分可能レンダラを通して、テンプレートメッシュを変形、変色を行うために使用されるテンプレートメッシュからの差分を表現するテンソルに対して最適化が実行される. このText2Meshを応用することによって本稿で提案するデザインツールの実装を行った.

$$\hat{S} = \frac{1}{n_\theta} \sum_{\theta} E(\varphi(I_\theta)) \quad (1)$$

$$\Phi_{target} = E(T_{target}) \quad (2)$$

$$L_{sim} = \text{dot}(\hat{S}, \Phi_{target}) \quad (3)$$

2.4 Prompt Engineering

Prompt Engineering[9]とはエンドタスクに対して望ましい情報が伝達されるように言語モデルに対して入力 Prompt を探索するタスクのことである. また、Text-to-Imageの領域では、文脈とは関係なく、“UnrealEngine”, “8K”, “Octane Renderer”などといった特定の Prompt を追加するだけで出力される画像の品質が向上する Magic Prompt[8]が存在している.

3. システム

3Dデータを用いた機械学習は多くのGPUリソースを必要とするため、高スペックなGPUを所有していないユーザーでもツールの使用を可能とするため、図2に示すようにAWS上のデータベースであるS3を仲介として、デザインの指定を行うフロントアプリとText2Meshによって変形を行うバックエンドにシステム分離した.

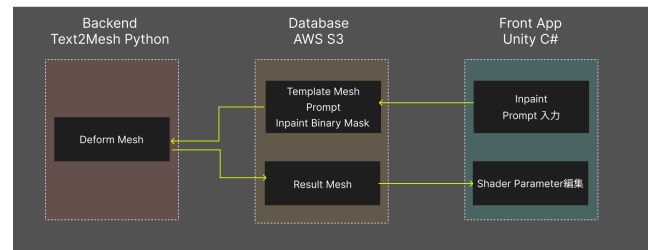


図2 システムアーキテクチャ

3.1 Inpaint

Text2Meshは入力されたテンプレートメッシュの全てのポリゴンに対してテキストに沿った形状変形が実行される. 図3に示されている通り、黄色矩形内はブリーツのように上手く変形が行われているが、青い矩形内のワンピースの肩の細い紐部分のような、細かい形状が入力とされた場合、メッシュ形状が破綻する傾向が高く、洋服らしさを保つことが難しい. そのため変形箇所を明示的に指定できるようなInpaintを実装した.



図3 Text2Meshの変形によってメッシュが崩壊してしまう例

3.1.1 Unity フロントアプリ

フロントアプリ上では、図4のようなリアルタイムでのデジタルクロージングに対する Inpaint 機能の実装を行った。扱う3Dモデルの頂点数は数万に及ぶため、リアルタイムでの Inpaint を実現する為に GPU 上で高速に動作する Compute Shader を用いて Inpaint 機能の実装を行った。



図4 Inpaint が行われている様子

3Dモデルに対する Inpaint は式(4)のように3Dモデルの頂点座標 V をカメラ外部行列 M_{view} 、カメラ内部行列 $*M_{proj}$ によって、投影後の2次元平面上の座標空間へ変換し、同じ座標空間でのマウス座標と比較を行なっている。そしてカメラに対して表面を向いている頂点群に対してのみ Inpaint を実行可能にするために、3Dモデルの法線ベクトルと原点からのカメラベクトルの内積が正の場合のみ Inpaint される対象として扱っている。

$$distance(M_{proj} * M_{view} * V, mousePosition) \quad (4)$$

3.1.2 Text2Mesh

Text2Meshのネットワークがバイナリマスクを追加の入力として受け取れるようにネットワークを変更し、変形、変色を表現するために使用されるプレートメッシュからの差分を表現するテンソルに対してバイナリマスクを敵応することによって変形、変色箇所を Inpaint によって制御することを可能とした。

3.2 Prompt 支援

Blender や CLO3D などといった専門的な知識をもとにある程度決まった道筋のものデザイン編集を行うツールに比べて、テキストによるデザインツールは専門的な知識を必要とせず、直感的に使用できる。しかし、テキストによるデザインの指定方法は形式がより柔軟かつ、拡張性が高

く、ガイドのバリエーションは無限に考えられる。そのため、望んだ生成結果を得る為に、テキストガイドの自由度の高さがユーザーに対して操作の難易度を高めているとも考えられる。

3.2.1 Template Prompt

本ツールでは、ユーザーが一から Prompt を思考する難易度を下げる為、事前に生成実験を行った中で生成結果の品質が高い傾向にあった Prompt の構文をデータベースに保存し、Template というボタンを押すたびに毎回ランダムに保存されていた図5に示されているような Prompt が出力されるようなシステムの実装を行なった。

- (1) "a clothing like a witch wears"
- (2) "a clothing made of coral reef"
- (3) "3d renderer of a luxurious clothing like a queen would wear"

図5 template として出力される prompt 例

3.2.2 Magic Prompt

文脈とは関係なく、特定の Prompt を追加するだけで出力の品質が向上する Magic Prompt[10]が存在している。しかし、この Magic Prompt の動作はまだあまり詳しくわかっていない。その為、ユーザーに対してこのような不確かな手法について悩ませるのではなく、付け加えるだけで品質が上がるという形で Quality Boost というボタンを押すことによって Magic Prompt が文末に追加されるようなシステムの実装を行なった。

3.3 質感調整

Text2Meshの出力では頂点の移動を表現するテンソルと頂点カラーを表現するテンソルのみが得られる。しかし、デジタルクロージングを含む3Dモデルのデザインを検討する際には色だけではなく、金属らしさ(metallic パラメータが1.0に近づくほど金属らしい)、なめらかさ(smooth パラメータが1.0に近づくほどなめらか)などをはじめとした質感、も重要である。そのため、本ツール内では、生成結果の質感をリアルタイムで変更できるように Shader パラメータの編集機能を実装した。例として図6、図7は"a shiny clothing formed by squid and octopus legs"という Prompt により生成された結果である。同じ頂点色でも図6のような Shader パラメータではマットな質感で"shiny"らしさや、"squid and octopus legs"といった濡れた質感は感じられないが、図7のようなパラメータの元では Prompt の特徴が反映されているように結果を確認することが出来る。



図 6 Smoothness:0.0, Metallic:0.0



図 7 Smoothness:0.65, Metallic:0.65

4. 実験と評価

本研究では、3DCG、映像などの制作経験がある6名の参加者を対象に実験を行った。初めにツールの操作方法を説明し、同じテキスト入力で Inpaint なし、Inpaint ありの場合の2サンプルを1セットとして3セット(合計36サンプル)、本ツールを用いたデザイン編集を行う実験を行った。実験後に本稿で提案するツールによって、生成されたデジタルクロージングに対して、出力に対する満足度、出力に対して想像通りの結果が得られたかという図8のような5段階(1~5)のスコアリングとフリーコメントによる評価を実施した。

- 出力に対する満足度
- 出力を制作などにおいて全く使える兆しが見えない
 - 気に入らない点が多く、修正する制作などの場において使用できる
 - ある程度修正すれば制作の場において使用できる
 - 気に入らない点が少なく、少量の修正で制作などの場において使用できる
 - 出力をそのまま制作の現場で使用できる

- 出力に対して想像通りの結果が得られたか
- 全く想像とは異なる出力が得られた
 - 殆ど想像とは異なる出力が得られた
 - 想像していた雰囲気と同じ出力が得られた
 - 殆ど想像とは異なる出力が得られた
 - 全く想像とは異なる出力が得られた

図 8 評価指標詳細

生成 AI モデルを制作ワークフローに介入させることで、図中央図のように自分が指定した要求を担保しつつも、出力が自分の想像とは異なることがあり、新しいアイデアの獲得などにつながることもある。これらのことから AI 活用した Creative Support Tool がユーザ本来の創造性を拡張することにつながることもあると考えられる。それに対して、図8の右端の図のように Blender や Adobe などの従

来の Creative Support Tool の出力は大きく予想から外れることは少なく、出力はユーザの想像内に収まるものがほとんどである。本稿で提案しているツールのように AI を活用したツールがユーザの創造性を拡張するか検証を行うために、出力に対して想像通りの結果が得られたかという指標を設けた。また、想像と離れているだけで、見当違いな出力が生成される場合もあるため、出力に対する満足度という指標を設け、これら2つの指標によって評価を行った。

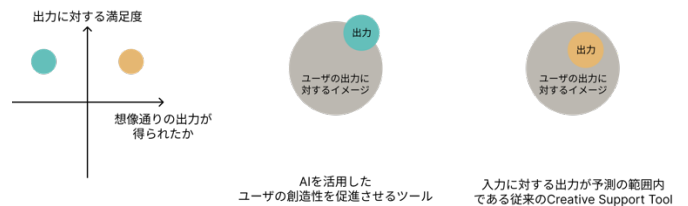


図 9 Creative Support Tool

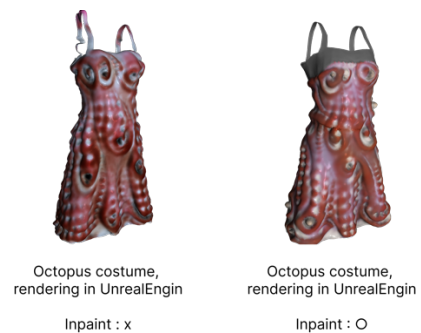
4.1 結果

Inpaint の有無に関わらず、出力に対して想像通りの結果が得られたかという項目に関する結果を確認すると、ユーザの予測とは異なるような出力が得られる傾向を確認することが出来た。また、Inpaint を行うことにより変形箇所を細かに指定できることによって出力に対する満足度が向上しているがわかった。

	出力に対する満足度	出力に対して想像通りの結果が得られたか
Inpaint なし	3.39 ± 1.02	2.39 ± 1.13
Inpaint あり	3.89 ± 0.99	2.44 ± 1.36

図 10 検証結果

4.2 生成結果



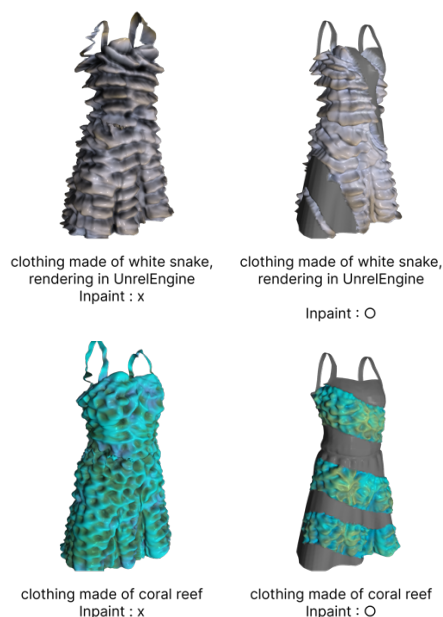


図 11 生成結果

4.3 考察

Inpaintの有無に関わらずユーザの想像から結果が外れた要因は入力された少ないテキストに対してユーザが複雑なコンテキストを詰め込んでいるためだと考えられる。例えば、“clothing made of white snake”という入力して得られたデジタルクロージングは蛇の鱗のような形状によって構成されていた。しかし、ユーザに対するフリーコメントでは「蛇の頭のような形状は見られなかった」という回答があった。これはユーザが“snake”というひと単語に対して蛇の頭まで含めた想像を含めていることに対して、AIに対しては“snake’s head”などと明示的に説明する必要があり、その差分によって想像とは異なる結果が生まれるのではないかと考えられる。Inpaintにより満足度が向上した理由は2点考えられる。1点目は前述した通り、破綻してしまうようなメッシュの問題箇所に対して事前に対策が行えるという点である。2点目はメッシュ全体でガイドテキストを表現するよりもInpaintを行うことによって指定された少ない限定箇所のみで変形を行なった方がその箇所に特徴が凝縮され視覚的に特徴をユーザが認識しやすいという点が考えられ、そのような点が満足度向上に繋がったと考えられる。

5. 結論と今後の展望

本稿ではテンプレートメッシュ、デザインを指定するテキスト、Inpaintを入力とする、デジタルクロージングに対するAIを使用したデザインツールを提案した。また、AIを用いたデザインツールがユーザの創造性を拡張する可能性があるということが検証によりわかった。

今後の研究の展望として、AIを用いることによってユーザの創造性を拡張する利点を活かしたアイデア発散ベースのモードと、既に生成したいデザインが決定されている際に使用する、ユーザの想像により近い出力するモードを切り替えられるようにすることがツールとして必要であると考えられる。

参考文献

- [1]. Joy, Annamma, et al. “Digital future of luxury brands: Metaverse, digital fashion, and non-fungible tokens.” *Strategic Change* 31.3 (2022): 337–343.
- [2]. Makryniotis, Thomas. “Fashion and costume design in electronic entertainment—bridging the gap between character and fashion design.” *Fashion Practice* 10.1 (2018): 99–118.
- [3]. Rombach, Robin, et al. “High-resolution image synthesis with latent diffusion models.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [4]. Ramesh, Aditya, et al. “Hierarchical text-conditional image generation with clip latents.” *arXiv preprint arXiv:2204.06125* (2022).
- [5]. Michel, Oscar, et al. “Text2mesh: Text-driven neural stylization for meshes.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [6]. Wibowo, Amy, et al. “DressUp: a 3D interface for clothing design with a physical mannequin.” *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 2012.
- [7]. 何毅, 謝浩然, and 宮田一乘. “手描きスケッチ入力による衣服 3D モデルの生成手法.” *研究報告コンピュータグラフィックスとビジュアル情報学 (CG)* 2022.1 (2022): 1–9.
- [8]. Radford, Alec, et al. “Learning transferable visual models from natural language supervision.” *International Conference on Machine Learning*. PMLR, 2021.
- [9]. Reynolds, Laria, and Kyle McDonell. “Prompt programming for large language models: Beyond the few-shot paradigm.” *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021.
- [10]. Aran Komatsuzaki. 2021. When you generate images with VQGAN CLIP, the image quality dramatically improves if you add “unreal engine” to your prompt. People are now calling this “unreal engine trick” lol.e.g. “the angel of air. unreal engine” pic.twitter.com/G4xBgVLYiv. <https://twitter.com/arankomatsuzaki/status/1399471244760649729>