

自然言語処理を活用した設計支援のための プロンプト生成と画像生成精度向上に関する研究

日大生産工（院） ○肖 宇力
日大生産工 岩田 伸一郎

1. はじめに

近年、テキストから画像を生成するAIモデルの発展により、建築設計における要望や意図を視覚的に提示することが容易になっている。特に、Stable Diffusionなどの生成モデルは、自然言語で入力されたプロンプトをもとに多様な建築空間イメージを出力できるため、設計初期段階における意思疎通手段として注目されている。また、ChatGPTなどの自然言語生成AIを併用することで、建築主の曖昧な要求から適切なプロンプトを自動生成し、それをもとに空間の雰囲気や素材感を可視化することも可能となった。

一方で、こうしたテキスト生成型AIを活用した建築可視化には、生成結果がプロンプトの文法構成・語順・語彙選択・形容詞や構造語の使い方などに大きく依存するという課題が存在する。たとえ同一の設計意図をもとにしても、プロンプトの順序や重み付け¹⁾の違いによって生成される空間の印象は大きく異なる。特に建築空間の生成においては「壁面材」「天井の質感」「採光方向」「空間スケール」といった要素が、画像全体のリアリティに強く影響を与えるため、これらをどのように入力すれば生成結果の安定性に直結する。

本研究の目的は、自然言語によって記述されたプロンプトの構成要素や表現の違いが、Stable Diffusionによる建築空間画像の生成結果にどのような影響を与えるかを明らかにすることである。

2. 研究方法

2.1 研究の概要

本研究では、画像から画像生成に必要なプロンプトの構成要素である語順、語彙、重み付けなどについて研究し、生成結果を安定化させるためのプロンプト修正および操作手法を検討し、最適化手法を提案する。

本論では画像生成について必要なプロンプトを段取り入力し、プロンプトの変化によって生成結果にどのような結果を与えるのかについて検証し、分析を行う。より精度の高く、実用可能な画像を生成できるためのプロンプトの関連性を判明し、生成結果に影響を与える箇所を着目し、プロンプト変化による影響を判明する。また、拡張機能であるControlNet²⁾を追加し、入力画像と同じスケール感と画角の画像を生成することを目指す。

2.2 研究フロー

研究のフローはFig.1に示したように、プロンプトの語彙・重み付けの順から検証を行い、プロンプトの変化が生成結果に与える影響を分析する。

2.3 Stable Diffusion 実装環境の設定

2.3.1 Stable Diffusion ComfyUI 環境設定

本研究では、Stable Diffusionを基盤とする画像生成環境としてComfyUIを採用し、その実行環境においてPyTorch、Transformers、およびXFormersを使用した。

PyTorchはモデル計算の基盤としてGPU最適化を行い、Transformersはプロンプト文の解析と埋め込み生成を担い、XFormersは生成処理の高速化を実現した。

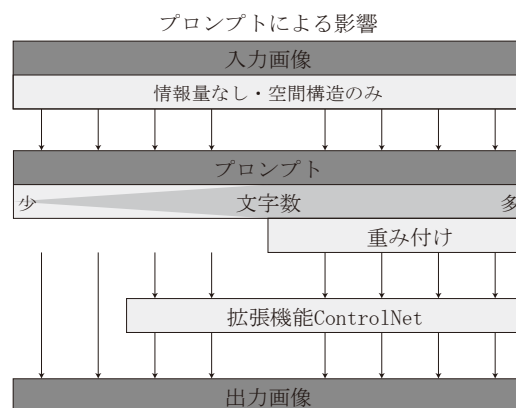


Fig.1 研究のフロー

2.3.2 チェックポイントモデルの設定

本研究で使用したチェックポイントモデルは、Realistic Vision V6.0 B1であり、フォトリアリスティックな質感再現に優れたチェックポイント統合モデルである。人物の顔や肌の描写、屋内外の光表現、陰影の自然さに強みを持ち、リアルな構図生成を得意とため、建築内観のリアルレンダリング、ControlNet との併用による照明・質感表現の最適化であると判断した。

2.3.3 ControlNet の設定とモデル選定

本研究では、Stable Diffusion の拡張機能である ControlNet を使用し、画像構成の安定化と入力画像との整合性を図った。使用モデルは control_v11p_sd15_scribble_fp16 であり、スケッチ図の情報を基に構図を制御するタイプのモデルである。これにより、建築内観の輪郭や奥行き、家具などの主要構成要素を保持しながら、照明や質感のみを変化させる生成が可能となった。

2.4 入力画像の設定

本研究では、入力画像を一律 1280 × 720 ピクセルに統一し、画像に含めている情報はシンプルな空間の輪郭だけのレンダリング画像とスケッチ画像を使用する。

3. プロンプトによる影響と考察

3.1 使用するプロンプト

Positive Prompt:

office interior (主題), glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, door on the left wall (要素)

Negative Prompt:

(sketch:1.4, drawing:1.3, cartoon:1.3)

3.2 異なるプロンプトによる結果および考察

入力画像の内容及びプロンプト構成の違いが生成に影響を与えることを明らかにした。


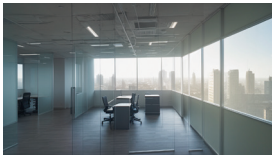



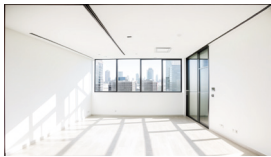
入力画像 \ プロンプト要素	主題のみ	主題 + 要素	主題 + 要素 + ネガティブプロンプト
レンダリング画像			
スケッチ図			

Fig. 2 異なるプロンプトによる生成結果






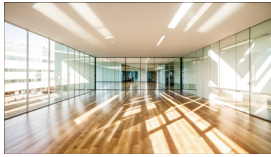
入力画像 \ ControlNet 種類	Canny	Depth	Canny + Depth
レンダリング画像			
スケッチ図			

Fig. 3 ControlNet 追加による生成結果

実験では、画像生成の目標である主題、構成要素及びネガティブプロンプトを段階的にプロンプトへ追加し、構図の安定性と自由性がどのように変化するかを観察した。

まず、「主題」のみの場合では、床・壁・天井を推測するができ、自由性が高く、構図の生成種類は豊かだが、生成するたびにランダムがあり、素材や反射表現が乏しく、精度が低いため、実用性が低いと判明した。

次に「主題＋要素」の場合では、「真ん中と右側にはガラス張り」、「ガラス越しの陽光」「ガラスばりの外の景色を都会のビル」というプロンプトを与えることにより、画像のリアルティが向上し、特にガラス要素や外部景色を指定した場合、モデルが学習過程で蓄積した都市景観・建築構図の知識を反映した。前提的に、自由性が下がったが、生成を制御し、生成したいものを生成することができた。

さらに「主題＋要素＋ネガティブプロンプト」の場合では、自由性が低い、ランダムな生成が制限されることによって、空間において不自然なものの生成をしなくなった。

3.3 ControlNet 追加による結果および考察

生成結果を Fig. 3 に示す。

Canny、Depth、及び Canny＋Depth の併用の 3 種類で比較した。各モデルが画像構成の安定性や空間表現に与える影響を検証した。

まず、Canny の場合、エッジ抽出による明確な輪郭線が生成過程で強調され、床や天井などの建築的構造の正確性が高く維持されたが、光の配置や影の位置では不自然があり、奥行に関する情報が足りないと考えられる。

次に、Depth の場合、入力画像の奥行き分布に基づいて生成が誘導されるため、スケール感や光の拡散が自然に再現されたが、ガラス枠や開口部の線的構造が歪む傾向から、輪郭に対する認識がやや曖昧と考えられる。

さらに Canny＋Depth を併用した場合、構造の正確性と奥行き表現が両立し、最も安定した生成結果が得られた。空間の質感や奥行きが向上し、生成画像をより制御しやすくなり、実用性を向上した。

3.4 プロンプト重み付けによる結果及び考察

3.4.1 プロンプトに材質を追加

Positive Prompt:

office interior, wooden floor(木製の床), white wall(白い壁), concrete ceiling(コンクリート天井), glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk(オフィスデスク), door on the left wall

Negative Prompt:

(sketch:1.4, drawing:1.3, cartoon:1.3)

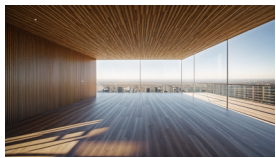
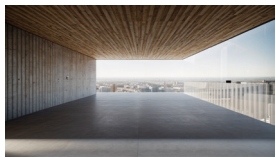
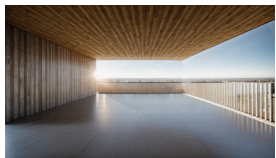

変化点	プロンプト	プロンプト
壁、天井、床に材質、机を追加する	Positive Prompt: office interior, wooden floor, white wall, concrete ceiling, glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
壁と天井に「()」を追加する	Positive Prompt: office interior, wooden floor, (white wall), (concrete ceiling), glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
壁と天井に「f:1.4」を追加する	Positive Prompt: office interior, wooden floor, white wall:1.4, concrete ceiling:1.4, glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
壁と天井に「(*:1.4)」を追加する	Positive Prompt: office interior, wooden floor, (white wall:1.4), (concrete ceiling:1.4), glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	

Fig. 4 プロンプト重み付けによる生成結果①





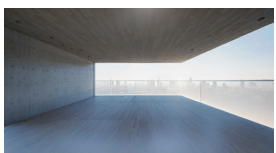
変化点	プロンプト	プロンプト
床に 「(**:1.1)」 を追加する	Positive Prompt: office interior, (wooden floor:1.1), (white wall:1.4), (concrete ceiling:1.4), glass wall on the center and right side, sunlight through glass, bright daylight, city buildings outside, office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
壁と天井を 「(**:1.3)」 ガラスを 「(**:1.6)」 景色を 「(**:1.4)」 に変更する	Positive Prompt: office interior, (wooden floor:1.1), (white wall:1.3), (concrete ceiling:1.3), (glass wall on the center and right side:1.6), sunlight through glass, bright daylight, (city buildings outside:1.4), office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
床を 「(**:1.2)」 天井を 「(**:1.2)」 に変更する	Positive Prompt: office interior, (wooden floor:1.2), (white wall:1.3), (concrete ceiling:1.2), (glass wall on the center and right side:1.6), sunlight through glass, bright daylight, (city buildings outside:1.4), office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
光を 「(**:1.3)」 ガラスを 「(**:1.4)」 景色を 「(**:1.2)」 に変更する	Positive Prompt: office interior, (wooden floor:1.2), (white wall:1.3), (concrete ceiling:1.2), (glass wall on the center and right side:1.4), (sunlight through glass:1.3), bright daylight, (city buildings outside:1.2), office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	
光を 「(**:1.4)」 景色を 「(**:1.3)」 に変更する	Positive Prompt: office interior, (wooden floor:1.2), (white wall:1.3), (concrete ceiling:1.2), (glass wall on the center and right side:1.4), (sunlight through glass:1.4), bright daylight, (city buildings outside:1.3), office desk, door on the left wall Negative Prompt: (sketch:1.4, drawing:1.3, cartoon:1.3)	

Fig. 5 プロンプト重み付けによる生成結果②

3.4.2 結果および考察

生成結果を Fig. 4 と Fig. 5 に示す。

生成結果からプロンプトの重み付けが生成に大きな影響を与えることを明らかにした。

検討対象は、各構成要素（床・壁・天井・ガラス・光・景色など）に対して「()」「(:1.4)」「(:1.4)」などの強調表現を段階的に追加した場合の違いである。これにより、モデルがどの要素を優先的に解釈・生成するか、またその結果として生じる空間的・視覚的变化を明らかにした。

まず、壁や天井などの構成要素に「()」または「(:1.4)」などをつけることで、生成の優先度が明確化され、床と異なるの材質を生成することができたが、指定した材質通りに生成できなかった。まだ、この段階では、室内の形状やスケール感が保持されつつ、明確な差異が見られなかった。

次に、「(white wall:1.4)」のように括弧と数値を組み合わせた場合、壁や天井などの構成要素を強調しすぎてしまい、床に対して生成の影響を与えてしまい、室内の形状を崩

れしまった。そのため、床にも「(**:1.4)」を追加し、ガラス張りや光、景色が消えたが、室内の形状を安定させた。

さらに、ガラス張りや光、景色にも少量な重み付けたところ、室内の形状を安定が保たれ、ガラス張りや景色が生成することができた。また。過剰な重さをつけると、生成物が完全に乱れてしまい、床や壁などの場所に影響を与えたが判明した。

参考文献

- 1) Amir Hertz and Ron Mokady and Jay Tenenbaum and Kfir Aberman and Yael Pritch and Daniel Cohen-Or:Prompt-to-Prompt Image Editing with Cross Attention Control, 2022
- 2) Alec Radford and Jong Wook Kim, Chris Hallacy and Aditya Ramesh and Gabriel Goh and Sandhini Agarwal and Girish Sastry and Amanda Askell and Pamela Mishkin and Jack Clark and Gretchen Krueger and Ilya Sutskeve:Learning Transferable Visual Models From Natural Language Supervision, 2021
- 3) Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser and Björn Ommer:High-Resolution Image Synthesis with Latent Diffusion Models, 2021

注釈

- 注 1) Stable Diffusion などの画像生成 AI において、プロンプト内の特定の語句や要素の重要度を数値的に調整する操作を指す。
- 注 2) Stable Diffusion の派生拡張技術であり、ロンプトのみでは指示できない複雑な構図を再現した画像などを生成することが可能になる機能である。