

生成AIにおける視覚的雰囲気分析 新世界を対象として

Analysis of Visual Atmosphere in Generative AI For the Shinsekai

○梅津 憂剛^{*1}, 村山 滉大^{*1}
Yugo UMEZU^{*1}, Kota MURAYAMA^{*1}

*1 大阪産業大学大学院工学研究科県境デザイン専攻 修士 (工学)
Master of Engineering in Environmental Design, Graduate School of Engineering, Osaka Sangyo University

キーワード: 生成AI; Stable Diffusion; 特徴抽出; 深層学習
Keywords: Generative AI; stable diffusion; feature extraction; deep learning

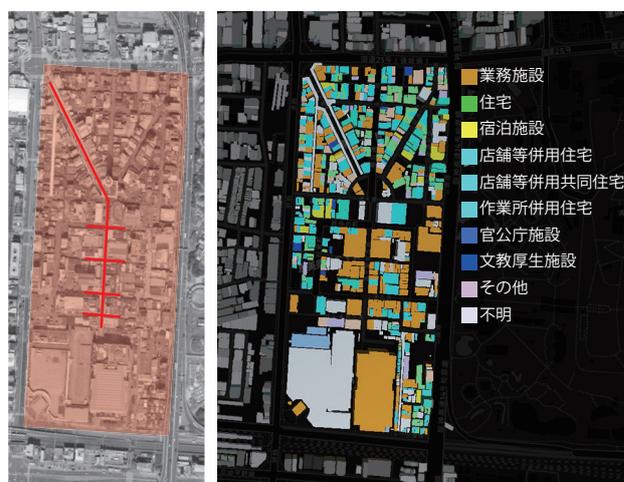
1. 研究の背景と目しんせかい

過去1年間にわたり、生成AIの進化は人々の創造性を刺激し、新たな可能性を開拓してきた。この生成AIの能力は、都市や住宅の未来の形状をどう描き出すのか、という問いを新たに投げかけている。気候変動や資源の問題、移住の現実を背景に、ただデザインや機能性だけでなく、都市や住宅がどのように持続可能で、再生可能な形を取るべきかという視点が必要とされている。この問いに対する答えとして、均一化を促進するスマートシティのアプローチとは異なる、新しい都市の形が求められている。

こうした中、大阪の新世界という地域は独自の歴史と文化を持ち続けてきた。だが、近年の開発やユニバーサルデザインの導入により、その特色が失われつつあるという危機感がある。地域のアイデンティティ、その「新世界らしさ」をどう維持し、発展させていくかは、現在の都市開発の大きな課題となっている¹⁾。

新世界は、その独自の文化と歴史を持つ地域である。その町並みは独特な雰囲気などから構成されており、私たちはその町並みを見ると、新世界であることを認識するだろう。しかし、その人間が空間を知覚する際の、雰囲気とはなにか。また、画像生成AIやChatGPTといった様々なAIが登場し発展が目覚ましい中、AIが作り出す雰囲気と人が想像する雰囲気とはどのような差が存在するのかは非常に興味のある部分である。

このプロジェクトの目的は、生成AIを活用して「新世界らしさ」を維持・発展させる手法を探ることだ。具体的には、Stable Diffusionという画像生成AIを使用して、画像からテキストへの変換とその分析を行い、何が「新世界らしさ」を形成しているのかを明らかにする。その画像を比較すると利点を用い、その場所の住民たちが主体となり、自らの手で地域の建築スタイルを維持・発展



a.調査ルート b.施設内訳
図1 新世界の調査ルートと施設内訳

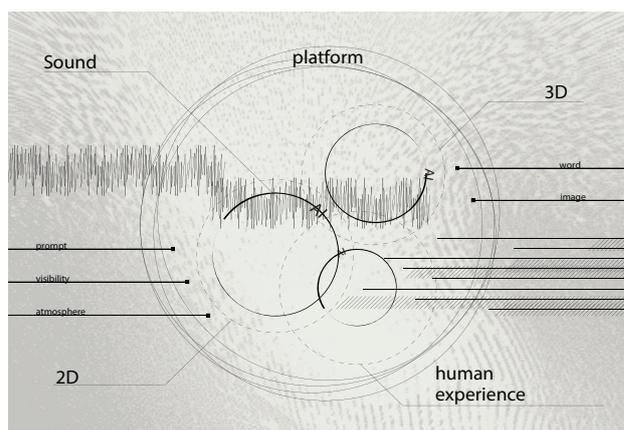


図2 分析の全体図

させる方向性を探求するツールとして、我々はAIを活用できるのではないかと考える。

さらに、今後の展望として、この手法を用いて私たちだけでなく、住民自身が都市計画に参加し、自らの地域の未来を描くことを目指す。これにより、住民主体のボトムアップの都市計画が可能となるだろう。

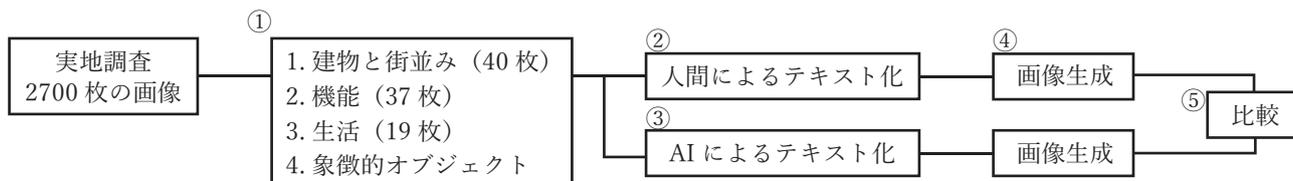


図3 分析手順の概略図

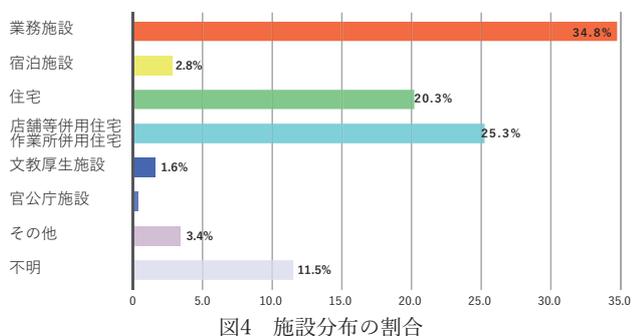


図4 施設分布の割合

2. 実地調査について

新世界の独自の文化と歴史、そして「新世界らしさ」という概念を具体的に把握するための第一歩として、実地調査を実施し、その雰囲気を直接感じ取ることとした。本稿の目的である「新世界らしさ」をデジタルで捉えるため、街並みの画像撮影を実施した。同時に、より高度な分析を目的として、iPad Pro の 3D Scanner App を活用し、街並みの 3D スキャンも進めた。調査日は 2023 年 8 月 2 日を選定し、新世界の昼夜の雰囲気の違いをより詳細に捉えるため、午前と午後の 2 回にわたって撮影を実施した。

撮影ルートは、新世界のシンボルである通天閣からスパワールドまでの主要ルートを基準に選定し、さらに撮影者が特に「新世界らしい」と感じた場所も追加で撮影を行った(図1)。

3. 分析手順について

新世界の「新世界らしさ」を具体的に捉えるために、実地調査で撮影した画像を中心に分析を進めていく。特に、1 章で取り上げた生成 AI の活用と、2 章での実地調査結果の双方を踏まえた独自の方法を試みる。このプロジェクトでは、「Stable Diffusion」と「Midjourney」という 2 つの先進的な画像生成 AI を使用する。分析の手順は以下のとおりである。①撮影した画像を人間の感性で「建物と街並み」、「機能」、「生活」、「象徴的オブジェクト」の 4 種類に分類する。②分類した画像から、新世界の独特な雰囲気や歴史的背景に基づく状況や経験からの言葉を人間が抽出し、リストアップする。③Stable Diffusion を

用いて、画像から AI が感じ取る「新世界らしさ」を形成する要素を言葉として抽出させる。④②の人間が抽出したテキストと③の AI によるテキストをそれぞれ、画像生成 AI のプロンプトとして活用し、新しいイメージを生成する。⑤生成されたイメージの比較から、人間と AI の「新世界らしさ」の認識の差異を探る。

以上の手順により、新世界の特色や背景に基づいた「新世界らしさ」の深層的な理解を目指す(図2, 3)。

4. 分析結果

図4における施設の割合を詳しく観察すると、新世界の街並みの 35% が業務施設で占められていることが明らかとなった。この点は、筆者らの実地調査の際にも確認され、撮影された写真の中には飲み屋などの業務施設が頻繁に登場していた。これにより、新世界特有の雰囲気や印象は、業務施設によって大きく形成されていることが理解された。

図5と図6は、2章で収集された「建物と街並み」のカテゴリーにおける画像から言葉を抽出し、それをプロンプトとして使用して画像生成を行ったものである。人間の感覚と AI の分析力の違いが顕著に表れている。図5における人間によるプロンプトから生成された画像を詳細に検討すると、我々が新世界の風情や歴史を背景に持つ「新世界らしさ」を言語で捉える際の独自性が確認できる。人間が選定したプロンプト内では、形容詞と名詞が組み合わせられた詳細な表現が多用されている。対照的に、図6に示される AI によって作成されたプロンプトからの画像では、言語のシンプルさが特徴的である。単一の名詞が中心的に使用され、形容詞の使用は限定的であることが確認される。つまり、AI が対象をストレートかつシンプルに捉える能力に優れていることが明らかとなった。一方、人間は複雑な情景や状況を言語で捉え、表現することの豊かさを持っていることが確認できた。

さらに、同様のプロンプトを使用して Stable Diffusion による画像生成も試みた。人間が作成したプロンプトを用いて生成された画像(図5)を詳しく観察すると、特徴的な看板や文字の存在、派手な色彩の使用など、新世界の特徴を強く感じさせる要素が見受けられた。全体の生



Tsutenkaku Tower, signboard, kushikatsu, Tekkonkinkreet, stripes, balloon, vivid color, shopping street, chaos, streetlight, beer bottle, outdoor unit, electric wire, tacky font, old signboard, beer basket, handwriting, mixed electric wire, tanned wall, bare bulb, dim bar, electric board, Ganz Osaka version, sale Handwritten croquettes, Shinsekai not Shinsekai, lanterns, tanned yellow, road signs, songs, leopard print, lion bags, Osaka ladies, dark blue store, handwritten price tags, yellow tiles miscellaneous display space tangled extension cables black rusty hangers, upturned noren, tattered ceiling, A bathhouse in town, a bicycle with a bag on it, a bulletin board, kanji and hiragana characters, a fresh smile, a Kohaku No. 5 store, a dirty floor, a food stall spreading out in the back, a standing drink, seven gods of good fortune, Kabuki, a symbol of Osaka, a three-dimensional sign, a doll-shaped lantern, a golden Billiken, the word Yokozuna, a snack sign, a sad back alley, a primary-color eaves tent, a high-rise with an advertisement sticking out of it utility poles, high-rises spreading out in the back, condominiums Shinsekai Choukai Rengokai, slurry glass, countless stickers, spray art, shutters, colored cones, AKIRA, target practice, glittering restaurants, sumo signs, Katsushika Hokusai-style signs, garbage accumulation, mixed pipes, no live bullets, Osaka's best is Japan's best, restrained hanafuda (Japanese playing cards), Red lattice ceiling, sweets lined up on the dais, many signs, Billiken-sama looking down on the people below, signboard covered with giant kushikatsu, people eating in line, bicycle thieves

図5 人間が作成したプロンプトによって生成



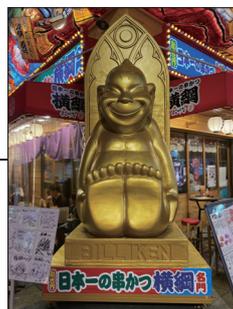
Osaka city, Shinsekai, Tsutenkaku Tower, Japan, Scenery, no humans, sign, outdoors, building, shop, ground vehicle, door, multiple boys, window, black hair, day, shirt, city, plant, sky, 1boy, 1girl, bicycle, box, food, 6+boys, indoors, pants, road, white shirt, air conditioner, blue sky, bottle, graffiti, hat, industrial pipe, multiple girls, real world location, shoes, shorts, tree, architecture, bowl, brown hair, chair, Chinese text, cityscape, east asian architecture, lantern, male focus, night, noodles, paper lantern, school uniform, solo, stairs, street, aircraft, arm up, bag, banner, barefoot, basket, bicycle basket, black footwear, book, bush, cable, card board box, cloud, computer, cup, cyberpunk, faucet, fence, fish, gun, high heels, neon lights, plate, ramen, realistic, redshirt, refrigerator, restaurant, rice, sandals, science fiction, shelf, shorthair, short sleeves, sink, sneakers, standing, table, train station, utility pole, weapon

図6 AIが作成したプロンプトによって生成

人間によるプロンプト



Osaka giant golden Billiken sitting in the center of the town --ar 3:4



AIによるプロンプト



a golden buddha statue sitting outside a tokyo market, in the style of playful cartooning, goblin academia, bold outlines, mischievous feline motif, smilecore, bold curves, die brücke --ar 3:4

図7 人間が作成したプロンプトによって生成 (左) と AI が作成したプロンプトによって生成 (右)

成画像を見ると、これらの特徴が一貫して反映されていることが確認できる。対照的に、AI が作成したプロンプトを使用して生成された画像（図 6）は、特定の雰囲気や統一性に欠けるものであった。新世界の特徴的な要素は、この AI 生成の画像群にはあまり顕著には現れていなかった。この結果から、人間が経験や感覚に基づいて選んだプロンプトは、新世界の特定の雰囲気や特徴を引き出すのに適していることが示された。一方、AI が独自に抽出したプロンプトでは、その地域の特徴や雰囲気を十分に反映するのは難しい可能性が考えられる。これは、人間の感覚や経験が、特定の文化や風土を理解する上で不可欠であることがわかった。

続いて、「象徴的オブジェクト」のカテゴリーの画像についても同様の分析を試みた(図7)。このカテゴリーでは、主にオブジェクトの特性に焦点を絞る形での分析が求められるため、Stable Diffusion よりも Midjourney が適していると判断された。テキストへの変換の際に確認された違いは、「建物や街並み」のカテゴリーと似ており、人間の解析はより情景を具体的にとらえる姿勢が見受けられた。一方、AI の解析は、「建物や街並み」とは異なり、形容詞を活用したり、幅広い解釈を持つような傾向が見られた。しかしながら、AI がどのようにオブジェクトを解釈するかは非常に変動が大きく、その結果、出力されるプロンプトの差異が顕著になっていた。生成された画像を比較すると、人間が生成したプロンプトに基づく画像は、元の画像との一致度が高く、対象の特性をよく捉えていたことが確認できる。一方、AI が生成したプロンプトを基にした画像は、部分的には元の画像の特性を捉えている箇所もあるものの、解釈のずれから生じる顕著な差異が見受けられ、全体としては十分な精度を持っているとは言えない結果となっていた。この結果を通して、人間の解析と AI の解析の間には明確なギャップが存在することが再確認され、特に象徴的なオブジェクトの解釈においてはその差異が大きく影響してくることが理解される。

5. まとめ

本研究では、新世界の街並みとその雰囲気を理解するための詳細な分析を行った。その主な結果として以下の点を挙げる事ができる。

① 実地調査の成果

筆者らは新世界での実地調査を通じて、多くの写真と 3D スキャンデータを収集した。これらのデータは、新世界の日常と特有の雰囲気を捉える上で非常に有効であった。

② 施設の割合と雰囲気

新世界の街並みの約 35% が業務施設で占められており、

これが新世界特有の雰囲気を大きく形成していることが明らかとなった。

③ 画像とテキストの解析

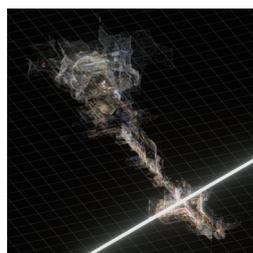
「Stable Diffusion」や「Midjourney」といった画像生成 AI を活用し、画像とテキストの解析を行った結果、人間と AI の認識の違いや解釈の幅についての洞察を得ることができた。

④ 人間と AI の違い

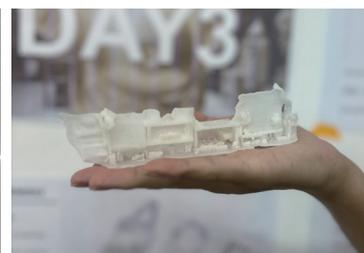
人間は情景や背景を総合的に捉える傾向があり、形容詞や名詞を複雑に組み合わせることで表現することが確認された。一方、AI はよりシンプルで直接的な表現に偏りがちであることが示された。

本稿では画像生成 AI において、人間と AI の可能なことの領域を明確に示すことができた。これらの結果を通じて、新世界の独特な雰囲気や文化について、さらに、人間と AI の間での認識や解釈の違いについても、本研究が新たな示唆を提供していると考えられる。

また、本研究で収集した 3D スキャンデータは、地域の雰囲気や景観の保全に有効に役立つと考えられる。これからの新世界の都市計画の見直しに際して、この 3D データは重要な情報源として位置づけられるだろう。さらに、本研究で明らかとなった AI との関連性を活かし、新たな研究の方向性やスタディを模索することが期待される。この点に関する詳細な検証は、今後の研究で進めていく予定である（画像 1, 2）。



画像 1 3D モデル



画像 2 3D プリント

[参考文献]

1) 『複製技術時代の芸術』, ヴァルター・ベンヤミン (佐々木基一ほか訳), 晶文社, 1997