

Deep Learning を用いた画像生成 AI の建築都市デザイン分野への適用可能性

○大野 耕太郎*¹ 山田 悟史*²

キーワード：ジェネレーティブデザイン、GAN、人工知能、深層学習、機械学習

1. はじめに

近年、DeepLearning を基盤とする人工知能 (AI) が急速に社会的な関心を集め、自動運転や医療診断など様々な分野で活用が期待されている。その期待を反映するように DeepLearning 分野では日々新たな技術や応用に関する研究が発表されている。それらの研究において多く引用されている技術の一つに、本研究も着目している「敵対的生成ネットワーク (Generative Adversarial Network、以降は GAN と呼称)」がある。GAN は当時モントリオール大学の学生であったイアン・グッドフェロー (Ian Goodfellow) 氏が考案したコンテンツ生成型の DeepLearning である¹⁾。画像認識の分野においては教師あり学習が一般的であり、予め任意の画像にラベルを付与し、そのラベルを推論するように特徴量を学習する画像分類が代表的な例である。一方、GAN は教師なし学習 (unsupervised learning) である。学習対象のデータにはデータの状態を示すデータ以外の情報は与えられておらず、何らかの方法でデータ間の関連性を導き出す学習である。このような特定のラベル等では表現できない情報の関連性を学習できる性質を活かした顕著な例に画像生成がある。

画像生成を例に GAN の概要を示す。GAN は、ジェネレーター (generator) とディスクリミネーター (discriminator) という 2 つのニューラルネットワークで構成されている。ジェネレーターの役割は与えられた画像に対してランダムな n 次元のノイズを発生させ、元画像とは異なる画像を生成することである。ディスクリミネーターの役割は、生成された画像が本物かどうかを判別することである。そしてジェネレーターが出力した画像に対してディスクリミネーターが返す誤差が小さくなるように両者を学習させる。この際にジェネレーター側が賢過ぎると、ディスクリミネーターが偽物であることを示す誤差を適切に出力出来なくなり、生成される画像の精度が上がらない。逆にディスクリミネーター側が賢過ぎると、生成された画像が全て偽物と判別され、誤差を使った学習が進まずやはり生成される画像の精度が上がらない (過学習)。そのため、過学習が生じないように両者を交互に学習させる。言わば競い合うように精度の高い画像を生成させる仕組みである。このような学習の仕組みから敵対的生成ネットワークと呼ばれている。近年では、GAN の

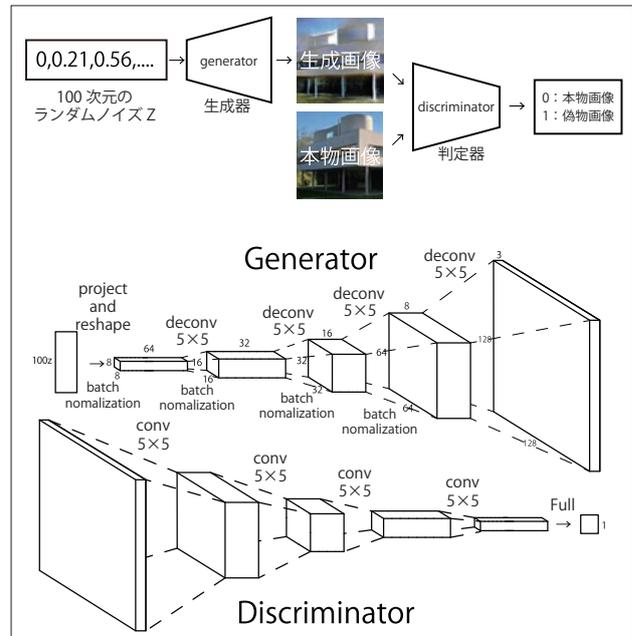


図1 DCGAN のネットワーク図

仕組みを応用した $\text{pix}2\text{pix}$ ^{注1)} や、立体的なデータを生成する 3DGAN^{注2)} などが提案されており、自動生成に対する DeepLearning の活用が高い注目を集めている。

本研究は、このような AI による画像生成の建築・都市のデザイン分野に対する適用可能性を試行するものである。生成を試行したのは街並みと建築物の外観パースである。街並みについては、画像生成に際して筆者らが意図した印象を持った街並みが生成されたかを被験者実験で検証した。外観パースについては、特定の建築物を模倣した画像を生成するだけでなく、デザインを演算したような新たなデザインを持った外観パースの生成を試行した。

2. 街並み画像の生成

2.1 研究対象とデータセット

本章の目的は、街並みらしい画像が生成できるかに加え、AI 作成者の意図を反映した街並みを生成できるかを検証することである。学習データとして選定した街並みは、和風な街並みとして「京都「産寧坂」」、西洋的な街並みとして「スコットランド「エディンバラ」」を選定した。画像は Google のストリートビューからスクリーンショット (200 枚) に反転による水増しを行った 400 枚である。図 2 に学習データ画像の一例を示す。



図2 教師画像の一例（上：京都 下：エディンバラ）

生成した画像に対する印象調査は、建築系学科に所属する大学生32名を対象に行った。被験者実験では、GANで生成した画像であることを告げずに画像を被験者に提示し、画像が何を表現していると思うか、またそれらの画像に対してどのようなイメージを抱くかを自由記述方式で回答して貰った。提示した画像は予め選定したそれぞれ4枚である。

2.2 学習モデルについて

今回用いた学習モデルはDCGAN⁵⁾である。モデルの構築には深層学習のフレームワークであるchainer^{注3)}を使用した。モデルの構造は図1の通りである。ジェネレーター側ではDeconvolutionによりアップサンプリングを行う。今回のモデルでは100次元のランダムノイズから最終的に128×128ピクセルの画像を生成している。ディスクリミネーターではストライド2のConvolutionによりダウンサンプリングを行う。こちらでは逆に128×128ピクセルの画像から最終的に本物か偽物かの2クラス分類をしている。また学習層の各所にはBatch Normalization^{注4)}層を挟んでいる。

2.3 DCGANによる街並み画像の生成

GANを用いて生成した画像の一部を図3に示す。画像を見ると京都の街並み生成画像では軒や瓦屋根、柳の木などの特徴をAIが学習し、描画に反映させているように読み取れる。またエディンバラの街並み生成画像では、茶色のレンガ壁や屋根の形などの描写ができていくように読み取れる。

2.4 生成画像に対する人間の評価

アンケートで得られた記述の抜粋を表1に示す。これをテキストマイニングの手法の一つであるワードクラウド^{注5)}を用いて分析した結果が図4である。この図を見ると、京都の画像群では、「和風」や「古風」といった言葉が、エディンバラの画像群では「ヨーロッパ」や「西洋」といった言葉においてスコアが高い。このことからAI作成者の意図・実際の街並みに近似な印象

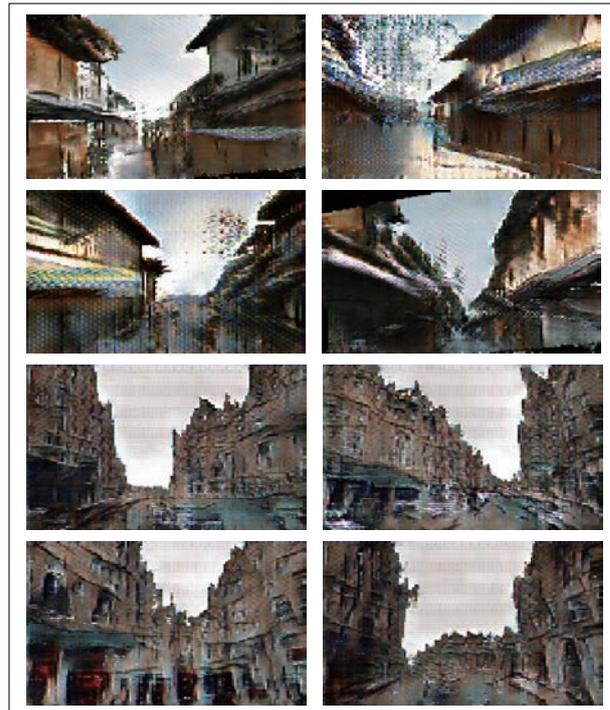


図3 生成画像の一例（上：京都 下：エディンバラ）

表1 アンケート結果の一例

京都		
番号	何に見えたか？	イメージ
1	住宅	粗い、汚い、ドット
2	家	汚い、目に黒そう
3	家と街路	ぼやぼや、日本風、明るい
4	町屋	京都、和、癒し、日本風、柳、石畳、木
5	京都の昔の通り	みにくい
6	古い家の通り	古い、時代劇的
7	民家	日本的な
8	古い家	趣がある
9	日本の古風な街並み	京都、古風、あたたかい、夕方
10	日本の家	粗い、和風

エディンバラ		
番号	何に見えたか？	イメージ
1	住宅	ヨーロッパ、石、レンガ、汚い、粗い
2	ヨーロッパの建物	汚い、目に黒そう
3	西洋風の街並み	くもり、ぼやぼや、重い
4	ヨーロッパ	洋、ヨーロッパ、パリ、城
5	パリの街	みにくい
6	外国の建築	重たい感じ、とげとげ
7	建物の並び	画質が粗い、細部がわからない、ヨーロッパっぽい
8	ヨーロッパ建築	高い
9	川のある街	暗い、西洋風、古風、芸術的、高い
10	建物の並び	壮大だ、ごつごつ、粗い、まっぼい

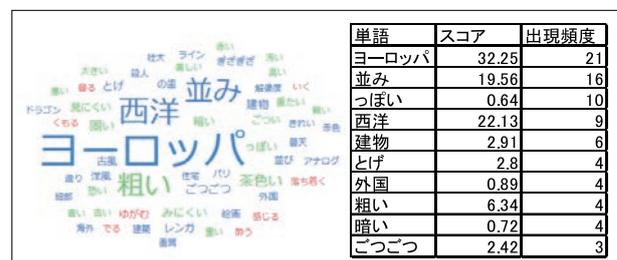


図4 ワードクラウド結果



図5 学習させた教師画像の一例

を被験者に抱かせる画像を生成できていると考えられる。しかしながら「粗い」などの言葉のスコアも高い。このことからGANで生成した画像には解像度やリアリティなどの面において課題があると考えられる。

3. 建築物の外観画像の生成

3.1 特定の建築物のデザイン生成・演算

次に建築への適用である。前章では1つの街路の画像のみを学習させ、その街路の特徴を持った画像の生成を行った。GANでは複数の種類の画像を学習させその中間の特徴を持つ画像を生成するといったことも可能である（本稿ではこのように画像を生成することを演算と呼称する）。そこで本章では、特定の建築家の単一の建物の外観の学習を行い、その建築家の意匠的な特徴を学習したコンテンツ生成AIを作成することが可能かを検証する。次に、複数の建築物の外観を学習し演算した外観を生成可能かを検証する。本稿が対象とした建築家は「ル・コルビュジェ」である。具体的な建築物は「サヴォア邸」「ロンシャンの礼拝堂」「フェミニの教会」「ユニテダビタシオン」の外観写真である。画像枚数はインターネット上で収集した画像に反転による水増し画像加えた計800枚である。図5に学習データの一部を示す。

単一の建物の外観を学習させたときに生成された画像を図6に示す。画像を視覚的に確認すると、学習に用いた建築物の外観の特徴を持っている。しかし、類似であるがそのものではない新たな外観が生成されていることが読み取れる。

3.2 複数の建築物の演算

単一の外観の生成において一定の精度が確認されたことを受けて、次に「サヴォア邸」と「ロンシャンの



図6 単体での画像生成の一例

教会」の両方の外観を学習させた。その結果として生成された画像の一例が図7である。生成画像を見ると、「ロンシャンの礼拝堂のような外形にサヴォア邸の立面を演算したような外観」が生成されている（左下の画像）。また「ロンシャンの礼拝堂の屋根とサヴォア邸の立面を演算したような外観（左上と右上）」も生成されている。次にさらに複雑な演算が可能かを検証するために、ユニテダビタシオンとフェミニの教会を追加して学習させた。生成画像の一部を図8に示す。生成画像を見ると、右上の画像は「稜線の左側にフェミニの教会を思わせる形状を持ち、右側にはロンシャンの礼拝堂の形状を思わせる形状有している。かつ立面の左側はユニテダビタシオンの立面、右側はフェミニの教会を思わせる外観」のように見える。他にも「フェミニの教会のような外形にサヴォア邸の立面を演算したような外観（左下）」が生成されている。以上のように、視覚的な特徴を言語化するのは難しいが学習データの外観の特徴を継承した外観デザインが作られているように読み取れる。しかし、特にディテールの描写に課題がある。ロンシャンの礼拝堂の窓が描写されていない点が一例である。また建築物の外観とは人間が判断できないような画像も生成された。これらの課題の大きな要因は、学習枚数の少なさが挙げられる。これは、既に高精度な画像生成しているGANの発表事例と比較して顕著に少ないことから分かる。

4. 考察とまとめ

DCGANを用いた画像生成AIの建築都市デザイン分野への適用可能性を検討した。以降にまとめる。

まず、AI作成者が意図した印象に応じた街並み画像を生成することができた。このことは、デザイナー

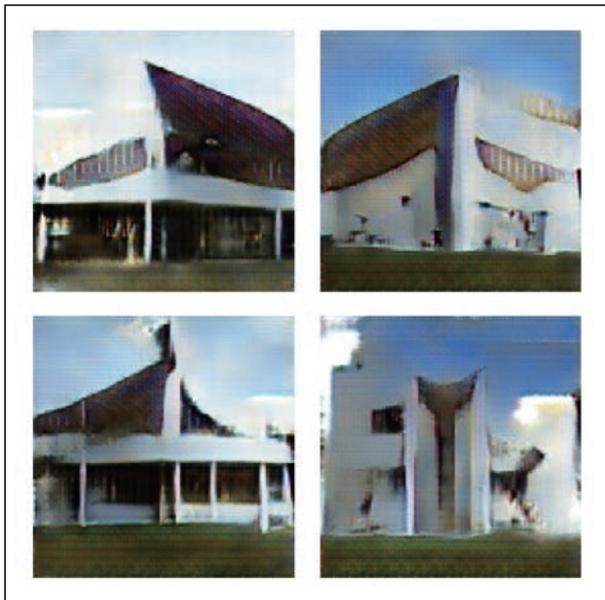


図7 2つの建物を同時に学習させた結果

の意図を反映したデザイン生成AIの建築都市デザイン分野での適用可能性の一端を示したと言える。

次に建築物単体の外観の生成と演算についてである。この試みについては、生成された画像に対する被験者実験などの検証が未実施である。そのため筆者らの視覚的印象に拠るが、単一の外観については高い精度を持った外観を生成できたと考えられる。そしてデザインの演算である。上記と同じく検証が未実施のため筆者らの視覚的印象に拠るが、各建築物の特徴が感じられる外観を生成できたと考えられる。

以上のように本稿ではDCGANを用いた画像生成AIの建築都市デザイン分野への適用可能性について以上のような成果を得たが発展的余地も多い。まずは学習枚数を増やして精度の検証を行う必要がある。次に高解像度の画像生成を可能にしディテールの描写の可能性を検討する必要がある。また検証についても、機械的な判断・被験者実験の方法を検討する必要がある。加えて、多様な対象に対して検証し適用可能性の汎用性の高さを把握することも課題である。

注

- ※ 01) GANを利用した画像生成アルゴリズムの一種で、2つのペアの画像から画像間の関係を学習することで、1枚の画像からその関係を考慮した補間をしてペアの画像を生成する技術。詳細は参考文献2)を参照。
- ※ 02) 簡単な3Dモデルから徐々に本物に近い3Dデータを段階を踏みながら生成する技術。詳細は参考文献3)を参照。
- ※ 03) 主に日本国内で使用されているPreferred Networksの主導で開発が進められているニューラルネットワークの計算および学習を行うためのオープンソースソフトウェアライブラリである。詳細は参考文献4)を参照。
- ※ 04) Sargey IoffeとChristian Szegedyによって提案された手法であり、バッチ処理として入力される複数のデータに対してニューラルネットワークの内部データを正規化することにより過学習の抑制やモデルの学習速度を上げる効果がある。詳細は参考文献6)を参照
- ※ 05) 文章中で出現頻度が高い単語を検出しスコアを出し、その

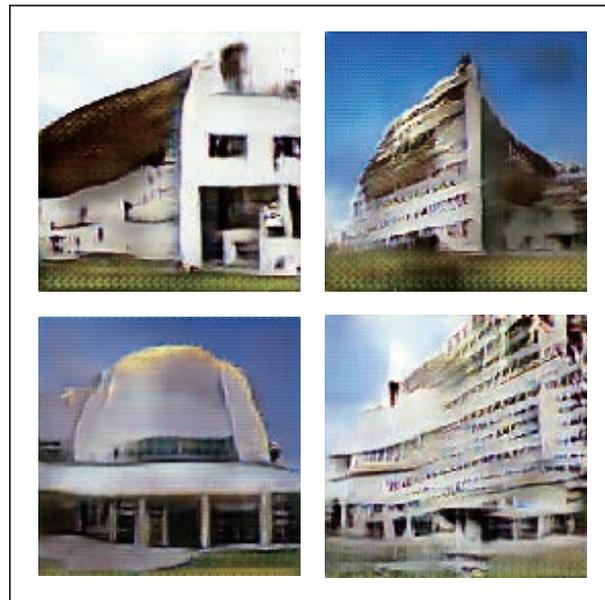


図8 4つの建物を同時に学習させた結果

スコアに応じた大きさで図示する手法。なおスコアとは単語における重要度のことであり、文章中に多く出現する言葉であっても一般的な文章で頻出しやすい単語はスコアが低くなり、逆に文章中でのみ頻出する単語はスコアが高くなるといった特徴を持つ。重要度の選定には一般的にTF-IDF法という統計処理が用いられる。TF-IDFはtf(英: Term Frequency、単語の出現頻度)とidf(英: Inverse Document Frequency、逆文書頻度)の二つの指標に基づいて計算される。参考文献7)を引用した。今回のアンケートではユーザーローカル テキストマイニングツールによる分析を行った。

$$tfidf_{ij} = tf_{ij} \cdot idf_i$$

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}$$

$$idf_i = \log \frac{|D|}{|\{d: d \ni t_i\}|}$$

$\sum_k n_{kj}$ は文書 d_j におけるすべての単語の出現回数の和

n_{ij} は文書 d_j における単語 t_i の出現回数

$|D|$ は総文書数、 $|\{d: d \ni t_i\}|$ は単語 t_i を含む文書数

参考文献

- 01) Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio: "Generative Adversarial Nets" <http://papers.nips.cc/paper/5423-generative-adversarial-nets>
- 02) <https://phillipi.github.io/pix2pix/>(参照 2018-06-15)
- 03) <http://3dgan.csail.mit.edu/>(参照 2018-06-15)
- 04) <https://chainer.org/>(参照 2017-12-28)
- 05) Alec Radford, Luke Metz, Soumith Chintala: "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" <https://arxiv.org/abs/1511.06434>
- 06) Sargey Ioffe, Christian Szegedy: "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift" <https://arxiv.org/abs/1502.03167>
- 07) <http://textmining.userlocal.jp/>
- 08) 坂本俊之: Chainerで作るコンテンツ自動生成AIプログラミング入門, 2017.2, C&R 研究所

* 1 立命館大学 理工学部 建築都市デザイン学科

* 2 立命館大学 理工学部 建築都市デザイン学科 講師・博士(工学)