

# 敵対的生成ネットワークを用いた日本古典籍修復に関する試み

金子 隼大 (立命館大学大学院 理工学研究科 修士一回生)

E-mail [ri0099xx@ed.ritsumei.ac.jp](mailto:ri0099xx@ed.ritsumei.ac.jp)

孟 林(立命館大学 理工学部 准教授)

## 要旨

日本古典籍は、多くの文化的、歴史的価値を有し、多くの研究者が深層学習などの技術を用いて、日本古典籍の翻刻に取り組んでいる。しかし、古典籍は長い年月を経て、劣化などの影響で、文字のかすれ、汚れ、欠損を生じ、自動翻刻の難しさを増している。本研究は、初めて、敵対的生成ネットワーク(GAN)を用いて劣化文字の自動修復を試み、古典籍の自動翻刻に貢献する。提案手法では、DTN, cycleGAN, discoGAN, UNIT などの複数の GAN 手法を用いて、教師無し画像変換を行い、劣化の修復に取り組む。そして、定量的評価により、GAN を用いた劣化修復の可能性を示す。

## abstract

Early Japanese Books recode a larger amount of historical information, thus attracting researchers to re-organize these books. However, due to the aging processing, some characters are blurred, dirty, and often partial, which increases the difficulty of automatic document re-organization. We challenge to repair the damaged character by applying the Generative Adversarial Network (GAN), for assisting book re-organization. In detail, unsupervised Image-to-Image translation of GAN methods such as DTN, cycleGAN, discoGAN, and UNIT are employed. The quantitative evaluation has proved the feasibility of the proposal.

## 1. 序論

日本には数多くの古典籍が存在している。これらの古典籍は、先人達の知恵により構築され、日本の政治・経済・文化などの多くの情報が記載されている。これらは、日本および世界文化遺産の一部として重要な役割を持つ。従って、これまでに、研究者らは、コンピュータ技術により古典籍の整理を行ってきた。例として、立命館大学アート・リサーチセンターの「古典籍ポータルデータベース」<sup>1)</sup>、国文学研究資料館の「日本古典籍総合日本の古典籍の総合目録データベース」<sup>2)</sup>、早稲田大学の「古典籍総合データベース」<sup>3)</sup>などが挙げられる。これらのデータベースは、古典籍のデジタル画像保存による文化の継承に大きく貢献してきた。さらに、研究者らは、古典籍のデジタル画像化による保存だけでなく、これらの古典籍を翻刻し、現存文字のデジタル整理を目指している。

しかし、多くの古典籍は、くずし字という、現在使用されていない文字を使用しているため、翻刻が容易ではない。従って、大量な古典籍の自動解析を目指し、研究者らは、深層学習という最先端な人工知能技術を用

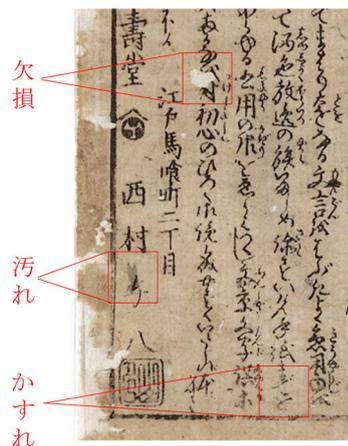


図 1. 日本古典籍とその劣化の例

い、文字の自動抽出と認識に取り組んできた<sup>4-6,19)</sup>。現状、深層学習を用いた文字の抽出と認識で、高い精度の認識率を達成している。しかし、これらの古典籍は、百年以上の年月を経て、劣化による文字の欠損や汚れ、かすれなどが生じている。図1には、日本古典籍と欠損、汚れ、およびかすれを示す。これらの劣化は、

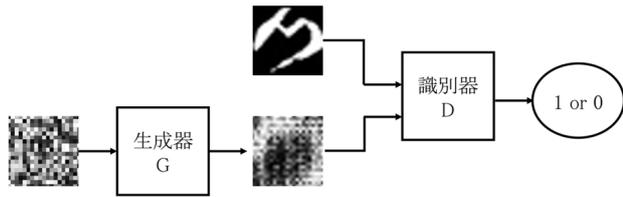


図 2. GAN の基本構造

古典籍における文字の自動整理を、より困難にしている。

近年、人工知能技術の進化は、画像の修復においても大きな発展を遂げている。中でも、注目されているのは、敵対的生成ネットワーク(Generative Adversarial Networks)<sup>7)</sup>という技術である。その中でも、GAN を用いたブラインド・フェイス・レストレーション<sup>8)</sup>に成功しているため、GAN を用いた日本古典籍の修復も可能であると考えられる。従って、本研究は、初めて、GAN という最先端な人工知能技術を用いて、日本古典籍の修復を試みる。そして、その可能性と問題点を示す。

本論文は以下の構成となる。2 章では、複数の既存の GAN モデルを紹介する。また、それぞれの GAN モデルが古典籍の修復に適しているかを議論する。3 章は評価方法と評価指標の定義で、4 章はそれらの指標を用いた実験結果を示す。ここでは、評価基準と評価方法を説明したうえで、GAN モデルによる提案手法の有効性を示す。5 章は本論文をまとめ、今後の課題について言及する。

## 2. GAN による画像変換

### 2-1. GAN (Generative Adversarial Network)

GAN は敵対的生成ネットワークと呼ばれる、識別器  $D$  と生成器  $G$  から構成される生成モデルである。図 2 には、GAN の基本構造を示す。識別器  $D$  は、入力された画像が、本物のデータの画像  $x$  であるか、生成器  $G$  から生成された画像  $G(z)$  かの判別を行う。ここで、 $z$  は生成器  $G$  に入力されるノイズを表す。対して、生成器  $G$  は、識別器  $D$  が、本物の画像  $x$  であると判別するような画像  $G(z)$  を生成することを目指す。目的関数は、以下の式で表される。

$$\mathcal{L}_{GAN}(G, D) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad [1]$$

従って、GAN は  $\arg \min_G \max_D \mathcal{L}_{GAN}(G, D)$  を目的とする。そして、これらの生成器  $G$  と識別器  $D$  のネットワークを競わせながら学習を進めることで、生成器  $G$  が、より本物に近い画像を生成することが可能になる。

### 2-2. Pix2Pix

Pix2Pix<sup>9)</sup>は、ノイズの代わりに、画像を生成器に入力することで、教師有り画像変換を実現したモデルであ

る。入力画像は、変換を行いたい画像と、変換後の目的とする画像とのペアとする。そして、条件付き GAN<sup>10)</sup>を用いることで、入力画像から選択的にペアの目的画像を生成する。

Pix2Pix は、ペア画像を用いることで、高精度の画像変換を実現する。しかし、全てのデータに対し、ペア画像を揃えることは困難である。特に、くずし字の劣化の場合、人の手によって、大量の修復済み画像を作成する必要がある。よって、本研究では、ペア画像を必要としない、教師無し画像変換に着目する。

### 2-3. Domain Transfer Network (DTN)

教師無し画像変換では、生成される画像が入力画像の特徴を維持している必要がある。しかし、画像を入力とした GAN による学習だけでは、入力に関係のない画像が生成されてしまう。DTN<sup>10)</sup>は、入力関数  $f$  を定義し、生成器  $G$  に対し、 $G = g \circ f$  かつ  $f(x) \sim f(G(x))$  という条件を加えることで、入力画像と生成画像の特徴を一致させる画像変換を行った。 $f$  は、事前に正解ラベルを用いて、入力画像の分類を行うことで、入力画像の特徴を抽出する機能を持つ。

### 2-4. cycleGAN

cycleGAN<sup>11)</sup>は、ペア画像を必要としない、双方向の画像変換が可能なモデルである。cycleGAN は、それぞれ 2 つずつの生成器  $G, F$  と識別器  $D_x, D_y$  から構成される。生成器  $G, F$  は、2 つの異なるドメイン  $X, Y$  に属する画像に対し、 $G: X \rightarrow Y, F: Y \rightarrow X$  の変換を行うことを目的とする。同時に、 $F(G(x)) \approx x$  かつ  $G(F(y)) \approx y$  を成り立たせる、サイクル一貫性損失を導入することにより、入力画像に対応した画像が生成される。以下に、サイクル一貫性損失の式を示す。

$$\mathcal{L}_{cyc}(G, F) = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1] \quad [2]$$

### 2-5. discoGAN

discoGAN<sup>12)</sup>は、cycleGAN と同様に、ペア画像を必要としない双方向の画像変換手法である。入力画像と、2 方向の変換を経由した画像が一致するように、再構成誤差(式 3)を設けることで、入力画像に対応した生成画像が得られる。ただし、実装では、平均二乗誤差を用いて、再構成誤差を求める。

$$\mathcal{L}_{constA} = d(G_{BA} \circ G_{AB}(x_A), x_A) \quad [3]$$

### 2-6. UNIT

画像変換において、GAN は、生成画像が目的のドメインに属することを目的とする。従って、1 つの入力画像に対して、敵対的損失を最適化する画像が無数に存在することになる。cycleGAN や discoGAN では、入力画像に対し、変換と逆変換を行った画像を、元の入



図 3. KMNIST に含まれる劣化画像

力画像と一致させることで生成画像を限定した。UNIT<sup>13)</sup>は、異なるドメインの画像が共有潜在空間  $z$  を有すると仮定することで、これの対処をした。異なるドメインの画像  $x_1, x_2$  に対し、2 つの変分オートエンコー<sup>14)</sup>を使用して、 $z = E_1(x_1) = E_2(x_2)$  になるように学習し、共有潜在空間  $z$  を求める。そして、 $z$  を入力とする、 $D_1, D_2, G_{1 \rightarrow 2}, G_{2 \rightarrow 1}$  による双方向の変換を行う。

### 3. 修復方法と評価指標の定義

#### 3-1. データセット

本研究は、データセットに KMNIST<sup>15)</sup>を用いる。KMNIST とは、10 種計 7 万枚のくずし字からなるデータセットである。はじめに、7 万枚のくずし字画像から、900 枚の劣化画像を選択し、劣化画像データセットを作成する。図 3 には、劣化くずし文字の例を示す。実験において、劣化画像データセットは、700 枚を劣化画像の訓練データとし、残りの 200 枚をテストデータとする。そして、劣化画像を取り除いた残りの画像を、無劣化画像とする。

#### 3-2. 評価指標

GAN の定量的評価方法については、今も多くの議論がなされている。GAN-test<sup>16)</sup>では、GAN の訓練データを用いて訓練された分類器に、GAN で生成された画像を再分類させることで、GAN の精度を評価した。本研究では、テスト用の画像と分類器を用いて、文字修復に適した GAN の評価を行う。また、統計的な評価手

法を用いて、分類器に依存しない評価も行う。

#### 3-2.1. 分類器を基準とした評価指数

本研究では、分類器を、畳み込み層 2 層、全結合層 3 層から構成された LeNet<sup>17)</sup> に、KMNIST のおよそ 6 万枚の訓練データを学習させ、作成する。この分類器は KMNIST のテストデータに対し、約 90% の認識精度を有した。

本提案では、分類器を基準にし、それぞれの修復率、再現率、正解率を定義し、GAN による劣化文字修復の評価指数とする。

##### ・ 分類器を基準とした修復率

図 4-(a) は分類器を基準とした修復率の評価方法を示す。初めに、テスト用の劣化画像を分類器によって分類する。次に、分類器が正しく分類できなかった劣化画像に対して、GAN による修復を行う。そして、修復された画像を再度同じ分類器を用いて分類する。修復率は、正しく分類できなかった画像に対し、修復した後に、再分類し、正しく分類できるようになった画像枚数の割合とする。

##### ・ 分類器を基準とした再現率

日本古典籍は、歴史的かつ文化的な価値を有している。よって、可能な限り、元の状態を維持した修復が求められる。図 4-(b) は分類器を基準とした再現率の評価方法を示す。テスト用の劣化画像の内、分類器が正しく分類できた画像を対象とする。これらの画像を GAN によって修復し、再び分類を行う。分類器が、修復した画像を再び分類できた画像の割合を求める。これを、GAN による修復の再現率とする。入力画像の特徴を損なうような変換を行った場合、再現率は減少することになる。

##### ・ 修復による分類器の正解率の差分

テスト用の劣化画像を、分類器によって分類し、正解率を求める。次に、同画像に対し、GAN による修復を行い、得られた修復後の画像を再び分類することで修復前後の正解率の差分を取る。すなわち、下記の式 4 となる。

(修復による正解率の差分)

$$= (\text{修復後の正解率}) - (\text{修復前の正解率}) \quad [4]$$

再現率は、修復前に分類できていた画像を維持することで最大となる。すなわち、入力された劣化画像に対し、GAN が修復を行わず、同じ画像を出力した場合も再現率は最大となる。しかし、この場合、修復率は最小となる。よって、本指標により、修復率と再現率の総合的な評価を行う。

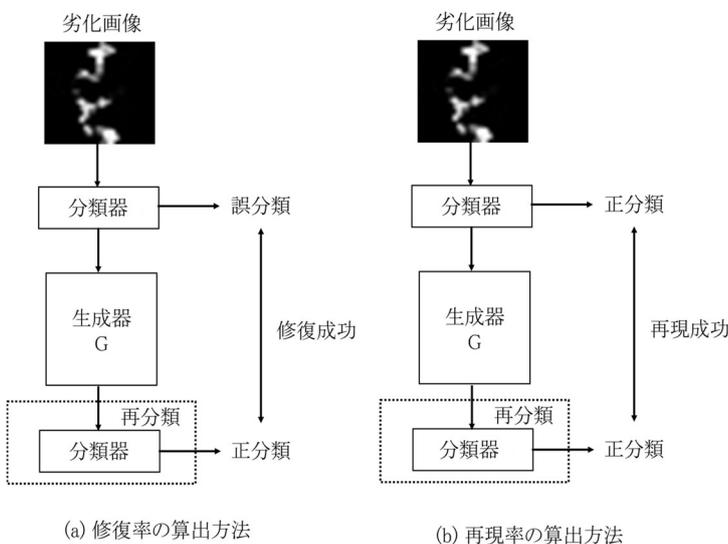


図 4. 分類器を用いた評価指数の算出方法

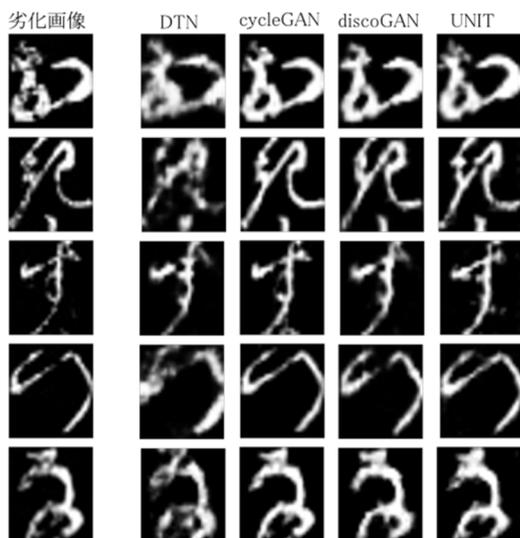


図 5. GAN による劣化の修復結果

### 3-2-2. 変換前後の画像の統計的類似度

最後に、Sliced Wasserstein Distance (SWD)を用いて評価を行う。GAN の性能指標としての SWD<sup>18)</sup>では、訓練画像と生成画像の SWD を計測し、画像の統計的な類似度を測定する。ただし、比較する画像同士の SWD の値が小さいほど、類似度は高い。本研究では、訓練用の無劣化画像(B)と、劣化を修正した画像(B')の SWD を算出し、修復した画像が、どれだけ無劣化画像に類似しているかを測る。また、cycleGAN、discoGAN、UNIT に関しては、劣化修復と同時に、無劣化画像(A)から劣化画像(A')の変換を行うため、A と A'、B と B'の両方の SWD を測定し、これによるモデルの比較を行う。

## 4. 評価実験

本研究は GPU 上で GAN モデルの学習と評価を行う。ハードウェア環境について、GPU が RTX3080 Ti、CPU が Intel®Core™i9-10900CPU@2.80GHz となる。ソフトウェア環境について、OS が、Ubuntu 21.10、機械学習フレームワークが Pytorch、プログラミング言語が python となる。

### 4-1. 評価結果

図 5 には、劣化画像の例と、それぞれの GAN モデルによる修正を行った結果を示す。図 5 から、GAN による画像変換を行うことで、劣化画像に見られた字のかすれを修正できていることが確認できる。特に、目視では、cycleGAN、discoGAN、UNIT が、より精度の高い修復を実現できている。

さらに、我々は、定義した評価指標による定量的な評価を行う。表1は、GAN の修復率、再現率、正解率を表す。ただし、これら3つの項目を、100 エポックまで10 エポック毎に計測し、最も良い値が得られたときを採用した。

表 1. 学習済み分類モデルによる評価結果

	DTN	cycleGAN	discoGAN	UNIT
修復率	25.00% (12/48)	25.00% (12/48)	29.17% (14/48)	37.5% (18/48)
再現率	63.23% (98/155)	97.42% (151/155)	95.48% (148/155)	94.84% (147/155)
正解率の差分	-27.6% (48.8%)	+2.4% (78.8%)	+1.9% (78.3%)	+1.9% (78.3%)

表 2. 元画像と生成画像における SWD

		DTN	cycleGAN	discoGAN	UNIT
1	A, A'	-	1	1.07	1.14
	B, B'	-	147.41	157.56	168.53
2	A, A'	-	1.18	1.25	1.55
	B, B'	-	173.70	183.99	228.22
2	A, A'	-	1.71	1.90	1.79
	B, B'	3.88	1	1.08	1.87
		442.13	113.90	123.04	212.60

\*A は劣化画像、A'は無劣化画像から劣化画像に変換した文字、Bは無劣化画像、B'は劣化を修復した画像を表す。  
1は、各モデルのA,A'のSWDが最小のときのA,A'とB,B'のSWD、2は、B,B'のSWDが最小のときのA,A'とB,B'のSWDを表す。ただし、上段の数値は、1、2の場合それぞれにおける最小のSWDとの比を表す。

表 1 を見ると、修復率に関しては、UNIT が最も高く 37.5%であった。一方、再現率は、cycleGAN の 97.42%が最も高く、UNIT の 94.84%が最小となった。正解率では、cycleGAN が他のモデルを上回る結果となった。

表 2 は、各 GAN のモデルによる、元の画像と、変換後の画像の SWD を表す。表 2 から、cycleGAN による修復で最も低い SWD が得られていることが分かる。さらに、cycleGAN は、無劣化から劣化への変換も、最も低い SWD 値を得られた。

### 4-2. 考察

実験結果から、学習済みモデルの正解率と SWD の値から判断して、cycleGAN が劣化文字の修復に最も適切なモデルであると考えられる。しかし、修復率に関しては、他のモデルに比べて低い値となっている。このことから、入力された画像の特徴を維持しようとするほど、修復を加えることが制限されていると考えられる。

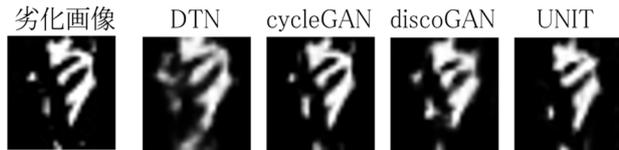


図 6. GAN による修復が行われない画像

また、欠損のある画像に対しては、GAN による修復が行われない恒等変換が確認できる(図6)。この原因として、欠損の修復は、かすれ文字の修復に比べ、多くの変更を必要とするからであると考えられる。

現時点で GAN を用いた劣化文字修復の可能性を調べたため、簡易的な分類用モデルを使用している。今後、より複雑なものによる評価を実現する予定である。

## 5. まとめ

本研究では、初めて、GAN による教師無し画像変換を応用し、劣化文字の修復を試みた。結果は、かすれ文字に対し、かすれが無いかのような画像への修復に成功した。中でも、cycleGAN を用いることで、より精度の高い劣化修復が行えることが分かった。しかし、欠損の大きな文字に対しては、修復が行われていなかったため、文字が完全に損なわれているような画像の修復を可能にすることが今後の課題である。

### [謝辞]

本研究は、立命館大学アート・リサーチセンターの助成を受けたものである。ここで、謝意を表する。

### [注]

- 1) 立命館大学, 古典籍ポータルデータベース, <https://www.arc.ritsumei.ac.jp/j/database/index.html> (2022. 06. 01 accessed)
- 2) 国文学研究資料館, 日本古典籍総合日本の古典籍の総合目録データベース, <https://base1.nijl.ac.jp/~tkoten/> (2022. 06. 01 accessed)
- 3) 早稲田大学, 古典籍総合データベース, <http://wul.waseda.ac.jp/kotenseki/> (2022. 06. 01 accessed)
- 4) Bing Lyu, Hengyi Li, Ami Tanaka, and Lin Meng, "The Early Japanese Books Reorganization by Combining Image processing and Deep Learning," CAAI Transactions on Intelligence Technology, 2022. 5
- 5) Alex Lamb, Tarin Clanuwat, Asanobu Kitamoto, "KuroNet: Regularized Residual U-Nets for End-to-End Kuzushiji Character Recognition," SN Computer Science, Vol.1, No.177, pp. 1-15, 2020. 5.
- 6) Tarin Clanuwat, Alex Lamb, Asanobu Kitamoto,

- "KuroNet: Pre-Modern Japanese Kuzushiji Character Recognition with Deep Learning," 15th International Conference on Document Analysis and Recognition (ICDAR2019), pp. 607-614, 2019. 9.
- 7) Goodfellow, Ian and Pouget-Abadie, Jean and Mirza, Mehdi and Xu, Bing and Warde-Farley, David and Ozair, Sherjil and Courville, Aaron and Bengio, Yoshua "Generative adversarial nets," Advances in neural information processing systems 27, 2014.
- 8) Tao Yang, Peiran Ren, Xuansong Xie and Lei. Zhang, "GAN Prior Embedded Network for Blind Face Restoration in the Wild," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 672-681, 2021.
- 9) Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- 10) Taigman, Yaniv and Polyak, Adam and Wolf, Lior, "Unsupervised cross-domain image generation," arXiv preprint arXiv:1611.02200, 2016.
- 11) Zhu, Jun-Yan and Park, Taesung and Isola, Phillip and Efros, Alexei A., "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proceedings of the IEEE international conference on computer vision, pp. 2223-2232, 2017.
- 12) Kim, Taeksoo and Cha, Moonsu and Kim, Hyunsoo and Lee, Jung Kwon and Kim, Jiwon, "Learning to discover cross-domain relations with generative adversarial networks," International conference on machine learning, pp. 1857-1865, 2017, PMLR.
- 13) Liu, Ming-Yu, Thomas Breuel, and Jan Kautz, "Unsupervised image-to-image translation networks," Advances in neural information processing systems, 30, 2017.
- 14) Kingma, Diederik P and Welling, Max, "Auto-encoding variational bayes", arXiv preprint arXiv:1312.6114, 2013.
- 15) Tarin Clanuwat and Mikel Bober-Irizar and Asanobu Kitamoto and Alex Lamb and Kazuaki Yamamoto and David Ha, "Deep Learning for Classical Japanese Literature", arXiv:1812.01718, 2018.
- 16) Shmelkov, Konstantin, Cordelia Schmid, and Karteek Alahari, "How good is my GAN?," Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- 17) LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. "Object recognition with gradient-based learning," Shape, contour and grouping in computer vision. Springer, Berlin, Heidelberg, 1999, pp. 319-345.
- 18) Karras, T., Aila, T., Laine, S., & Lehtinen, J, "Progressive growing of gans for improved quality, stability, and variation," arXiv preprint arXiv:1710.10196, 2017.

- 19) Xuebin Yue, Hengyi Li, Yoshiyuki Fujikawa, and Lin Meng, “Dynamic Dataset Augmentation for Deep Learning-based Oracle Bone Inscriptions Recognition,” ACM Journal on Computing and Cultural Heritage, ACM. 2022. 4.