Vol.2021-MPS-134 No.13
2021/7/27

近代書籍フォントの推定生成

竹本有紀[1]　石川由羽[2]　髙田雅美[1]　城和貴[1]

概要：明治から昭和初期に刊行された近代書籍は，現代の OCR 技術によるテキスト化が困難である．近代書籍に特化した文字認識手法では，学習データの不足が課題となっている．この問題を解決するため，近代書籍で用いられるフォントの特徴を持つ文字画像の自動生成を目指している．本稿では，より精度の高い文字画像の生成を目指し，StarGAN v2 を用いた近代書籍フォントの生成を行う．生成された画像から，入力画像に適した現代フォントの種類と StarGAN v2 の近代書籍フォントの生成性能について検討する．

キーワード：フォント生成，ディープラーニング，敵対的生成ネットワーク，近代書籍

Estimated Font Generation for Early-Modern Japanese Printed Books

YUKI TAKEMOTO[†1]　YU ISHIKAWA[†2]
MASAMI TAKATA[†1]　KAZUKI JOE[†1]

1. はじめに

国立国会図書館では，Web 上で書籍のデジタルデータを一般公開している[1]．その中に含まれる近代書籍は，明治から昭和初期に刊行された書籍である．当時の文化や歴史を知る上で重要な資料としても利用される．ところが，Web 上で公開されているのは書籍を撮影した画像データである．テキストデータであれば文書内容に対して文字列検索を行い，必要な情報を瞬時に得られる．しかし，画像データには文字列検索ができない．利便性を向上させるため，画像データの早急なテキスト化が求められている．

光学文字認識（Optical Character Recognition, OCR）技術の発達により，画像からのテキスト抽出は盛んに行われている．しかしながら，近代書籍は既存の OCR 技術によるテキスト化が困難である．そこで，近代書籍に特化した多フォント活字認識手法が提案されている[2-4]．畳み込みニューラルネットワークの導入によって近代書籍の文字認識率は向上している[5]が，まだ実用化に十分な精度ではない．

近代書籍文字認識の認識率向上への課題として，学習データの不足が挙げられている．学習データは，近代書籍の画像データから手動で 1 文字ずつ切り出される．しかしながら，書籍からの収集には限界がある．書籍に含まれる全ての文字の中で，出現頻度がk番目となる文字の割合は$1/k$に比例する．これは Zipf の法則である．出現頻度の低い文字が用いられる回数は低く，書籍の画像データから見つけるのが困難である．さらに，近代書籍は活版印刷であることから，インクににじみやかすれが生じる．収集した文字画像が学習データに利用できない場合もある．以上のことから，学習データの数を増やすためには，書籍から収集する以外の方法を見つける必要がある．

新たな学習データの収集方法として，近代書籍文字画像の自動生成に取り組んでいる[6,7]．入手が容易な現代フォントの文字画像からフォント変換を行い，近代書籍で特定の出版社・出版年代に用いられたフォントと同じ特徴を持つ文字画像を生成可能である．しかしながら，生成される画像は文字と背景の境界が不明瞭である．そこで，より精度の高い画像を生成するために，フォント変換を行うネットワークの構造を改善する必要がある．さらに，フォント変換の元となる現代フォントの種類についても検討する必要がある．これまでは，フォント変換がニューラルネットワークにより正確に行われていることを確認するため，入力画像の現代フォントに，近代書籍フォントとの違いが大きいゴシック体を用いている．入力に用いる現代フォントを変更すれば，より精度の高い近代書籍フォントの文字画像を生成できる可能性がある．本稿では，近代書籍フォントの生成に適した入力画像の現代フォントの種類とネットワークの構造を検証するため，StarGAN v2 を用いた近代書籍フォントの生成を行う．StarGAN v2 は，鮮明な画像を生成できる敵対的生成ネットワークを用いて，複数ドメイン間における画像変換が可能である．この特徴を生かし，複数の現代フォントから近代書籍フォントへ変換した画像を生成する．生成された文字画像を比較し，入力画像に適した現代フォントと StarGAN v2 による近代書籍フォントの生成性能について検討する．

1 奈良女子大学
Nara Women's University
2 滋賀大学
Shiga University

以下に本稿の構成を示す．まず，2 章では既存研究を紹介し，3 章では本稿で用いる StarGAN v2 について説明する．次に，4 章で StarGAN v2 を用いた近代書籍フォントの生成実験の方法と結果を述べる．

2. 既存研究

画像生成は，機械学習において注目されている分野の 1 つである．敵対的生成ネットワーク(Generative Adversarial Networks, GAN)の登場により，生成される画像の精度は飛躍的に向上している．GAN は生成器と識別器の 2 つのニューラルネットワークによって構成される．識別器は与えられた画像の真偽，つまり，学習データの本物の画像か生成された偽物の画像か正確な判別を目指す．生成器は，識別器が学習データの画像と判断するような精度の高い画像の生成を目指す．Pix2pix[8]は入力画像と目標画像をペアとする学習データを用いて，入力画像から目標画像に変換された画像の生成を学習する．学習データに用いる画像によって様々な画像の変換が可能となっている．CycleGAN[9]は，さらに生成画像を入力画像へと再変換する過程も学習することで，入力画像と目標画像のペアを作成する必要がない．これを応用した StarGAN[10]は，複数のドメイン間における画像変換が可能となっている．ドメインとは，画像が持つ様々な属性に対し，同じ属性に属する画像の集合である．従来の生成器は，1 対 1 のドメイン間で画像変換を行う．k個のドメイン間で画像変換を行う場合には，$k(k-1)$個の生成器を学習しなければならない．StarGAN は，生成器の入力にドメイン情報を与え，識別器でドメインのクラス分類を行う．これにより，1 つの生成器によって複数ドメイン間の画像変換が可能となる．StarGAN では，学習データに対して属性のラベル付けを行わなければならない．このラベル付けを行わずに複数ドメイン間の画像変換を可能にしたのが StarGAN v2[11]である．学習データは大まかなドメインに分けるだけで良い．例えば，人の顔の画像を学習データとする場合，StarGAN では性別，髪の色，表情，肌の色などの属性に対し，学習データの画像がどの属性に該当するのかラベルで与えなければならない．一方，StarGAN v2 では，男女の性別で分けるだけで．髪色や表情，肌の色といった詳細な属性は，学習によって獲得する．本稿では，StarGAN v2 のこの特性を活かしてフォント変換を行う．学習データに用いる文字画像は，フォントをドメインとして，複数の現代フォントから特定の近代書籍フォントへの変換を目指す．StarGAN v2 の詳細は 3 章で説明する．

3. StarGAN v2

StarGAN v2 は，複数のドメインに分けた画像の集合を用いて 1 つの生成器を学習する．この生成器は，特定のドメインの画像から他の任意のドメインへ変換した画像を生成可能である．StarGAN v2 の概要を図 1 に示す．生成器Gは，

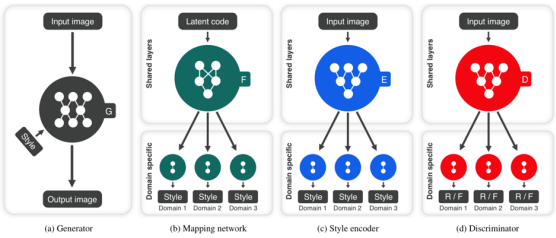


図 1　Overview of StarGAN v2

入力画像xからスタイルコードsを用いて画像$G(x,s)$を出力する．スタイルコードsはマッピングネットワークもしくはスタイルエンコーダから得られる．マッピングネットワークFは潜在コード$z \in Z$からスタイルコード$s = F_y(z)$を生成する．スタイルエンコーダEは，入力画像xからスタイルコード$s = E_y(x)$を抽出する．このとき，yは変換後の目標となるドメインである．ドメインに対応したスタイルコードを用いることで，生成器にドメインの情報を与えずに目標のドメインに変換した画像が生成できる．識別器Dは入力された画像の属するドメインと，画像の真偽を判定する．

StarGAN v2 の学習について述べる．学習データとして与えられる画像とドメインの集合をそれぞれ$\mathcal{X}$, $\mathcal{Y}$とする．ドメイン$y \in \mathcal{Y}$に属する画像$\mathbf{x} \in \mathcal{X}$に対して，StarGAN v2 は以下の工程を行う．

I. スタイルコードの生成
II. 敵対的生成
III. スタイルコードの再構築
IV. 生成器の正規化
V. 生成画像の再変換

まず，マッピングネットワークFは目標ドメイン$\tilde{y}$へ変換するための目標スタイルコード$\tilde{\mathbf{s}} = F_{\tilde{y}}(\mathbf{z})$を生成する．このとき，潜在コード$\mathbf{z} \in \mathcal{Z}$と目標ドメイン$\tilde{y} \in \mathcal{Y}$はランダムにサンプリングされる．

次に，生成器Gは画像$\mathbf{x}$と目標スタイルコード$\tilde{\mathbf{s}}$を用いて画像$G(\mathbf{x}, \tilde{\mathbf{s}})$を生成する．画像$\mathbf{x}$と$G(\mathbf{x}, \tilde{\mathbf{s}})$に対し，識別器$D$を用いて画像が属するドメインと画像の真偽を判定する．識別器はドメインyの画像$\mathbf{x}$が入力されると，$D_y(\mathbf{x})$を出力する．ここで，マッピングネットワークはドメイン$\tilde{y}$の画像が持つ特徴を表すスタイルコードの生成を学習する．生成器は，識別器がドメイン$\tilde{y}$に属する本物であると判断するような画像の生成を学習する．これらの学習には，式 1 で表される敵対的損失を用いる．

$$\mathcal{L}_{adv} = \mathbb{E}_{\mathbf{x},y}\left[\log D_y(\mathbf{x})\right] + \mathbb{E}_{\mathbf{x},\tilde{y},\mathbf{z}}\left[\log\left(1 - D_{\tilde{y}}\left(G(\mathbf{x}, \tilde{\mathbf{s}})\right)\right)\right] \quad (1)$$

スタイルエンコーダEは，生成画像$G(\mathbf{x},\tilde{\mathbf{s}})$からスタイルコード$E_{\tilde{y}}(G(\mathbf{x},\tilde{\mathbf{s}}))$を抽出する．このとき，スタイル再構築損失を用いて，画像$G(\mathbf{x},\tilde{\mathbf{s}})$の生成に必要なスタイルコード$\tilde{\mathbf{s}}$の再構築を学習する．スタイル再構築損失は式 2 に示す．

$$\mathcal{L}_{sty} = \mathbb{E}_{\mathbf{x},\tilde{y},\mathbf{z}}\left[\left\|\tilde{\mathbf{s}} - E_{\tilde{y}}(G(\mathbf{x},\tilde{\mathbf{s}}))\right\|_1\right] \quad (2)$$

この学習により，スタイルエンコーダは参照画像からドメイン変換に必要なスタイルコードを抽出可能になる．

また，多様な画像の生成を可能にするため，式 3 に示す多様性過敏損失を用いて生成器の正規化を行う．

$$\mathcal{L}_{ds} = \mathbb{E}_{\mathbf{x},\tilde{y},\mathbf{z_1},\mathbf{z_2}}[\|G(\mathbf{x},\widetilde{s_1}) - G(\mathbf{x},\widetilde{s_2})\|_1] \quad (3)$$

ここで，$\widetilde{s_1}$と$\widetilde{s_2}$はそれぞれ，ランダムな潜在コード$\mathbf{z_1}$，$\mathbf{z_2}$を用いてマッピングネットワークから生成されたスタイルコードである．本来の正規化では，$G(\mathbf{x},\widetilde{s_1}) - G(\mathbf{x},\widetilde{s_2})$を$\mathbf{z_1} - \mathbf{z_2}$で除算する．そうすると，$\mathbf{z_1}$と$\mathbf{z_2}$の差が微小であるときに損失の値が大きくなり，学習が不安定になる．そのため，この損失では正規化の分母の項を省略しているが，正規化の本質的な部分は変化しない．

最後に，画像$G(\mathbf{x},\tilde{\mathbf{s}})$をドメイン$y$に再変換する．スタイルエンコーダ$E$から画像$\mathbf{x}$の推定スタイルコード$\hat{\mathbf{s}} = E_y(\mathbf{x})$を生成し，画像$G(\mathbf{x},\tilde{\mathbf{s}})$と推定スタイルコード$\hat{\mathbf{s}}$から生成画像$G(G(\mathbf{x},\tilde{\mathbf{s}}),\hat{\mathbf{s}})$を得る．生成器は，式 4 に示す循環一致損失を用いて，画像$\mathbf{x}$と一致する画像$G(G(\mathbf{x},\tilde{\mathbf{s}}),\hat{\mathbf{s}})$を生成できるように学習する．これは，画像$\mathbf{x}$の持つドメインに依存しない特徴が画像変換によって失われるのを防ぐためである．

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{x},y,\tilde{y},\mathbf{z}}[\|\mathbf{x} - G(G(\mathbf{x},\tilde{\mathbf{s}}),\hat{\mathbf{s}})\|_1] \quad (4)$$

以上の学習を行う StarGAN v2 の目的関数を式 5 に示す．

$$\min_{G,F,E}\max_{D}\mathcal{L}_{adv} + \lambda_{sty}\mathcal{L}_{sty} - \lambda_{ds}\mathcal{L}_{ds} + \lambda_{cyc}\mathcal{L}_{cyc} \quad (5)$$

ここで，λ_{sty}，λ_{ds}，λ_{cyc}はハイパーパラメータである．

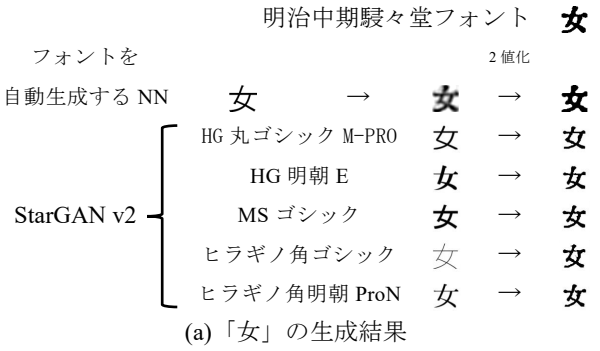
この StarGAN v2 を用いて，複数の現代フォントから特定の近代書籍フォントへのフォント変換を試みる．フォントの種類でドメイン分けされた文字画像の集合で StarGAN v2 を学習する．これにより，1 つの生成器から，複数の現代フォントの文字画像を近代書籍フォントへ変換した文字画像の生成が可能になると考えられる．StarGAN v2 を用いたフォントの自動生成実験の詳細は 4 章で述べる．

4. StarGAN v2 を用いたフォント自動生成実験

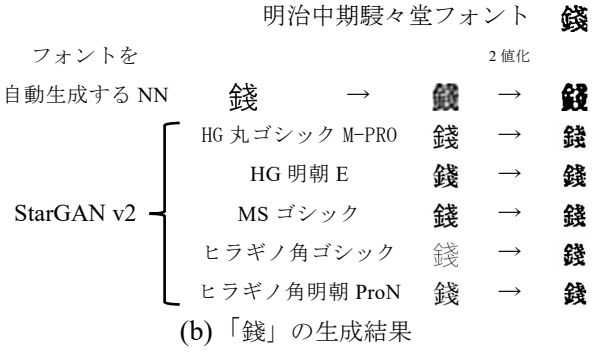
3 章で紹介した StarGAN v2 を用いて，複数の現代フォントを元に近代書籍フォントの自動生成を試みる．複数の現代フォントから生成された生成画像を比較し，近代書籍フォントの自動生成に適した現代フォントと StarGAN v2 による近代書籍フォントの生成性能を検討する．

表 1　各フォントの文字画像数

フォント名	文字画像数
HG 丸ゴシック M-PRO	1294
HG 明朝 E	1294
MS ゴシック	1297
ヒラギノ角ゴシック	1293
ヒラギノ角明朝 ProN	1296
明治中期駸々堂フォント	1297



(a)「女」の生成結果



(b)「錢」の生成結果

図 2　フォントを自動生成する NN と StarGANv2 による生成画像

学習に用いる文字画像のデータセットについて述べる．使用する現代フォントは，HG 丸ゴシック M-PRO，HG 明朝 E，MS ゴシック，ヒラギノ角ゴシック，ヒラギノ角明朝 ProN の 5 種類である．目標とする近代書籍フォントは，明治中期に駸々堂から出版された書籍に用いられているフォントとする．これを明治中期駸々堂フォントと呼ぶ．それぞれのフォントに対する文字画像の枚数を表 1 に示す．フォントによって枚数が異なるのは，現代フォントの中には近代書籍で用いられる旧字体や異字体に対応していない場合があるためである．全てのフォントで文字画像が揃っている 1200 の文字を StarGAN v2 に学習させ，残りの文字で未学習の文字に対するフォント生成の性能を確認する．

フォント自動生成の実験結果を述べる．図 2 に，フォントを自動生成する NN と StarGAN v2 から自動生成された文字画像を示す．フォントを自動生成する NN による生成

図 3　文字線の欠損が著しい生成画像

表 2　近代書籍文字画像と各生成画像から求めた
パワースペクトルの RMSE の平均値

		RMSE
フォントを自動生成する NN		0.9210
StarGAN v2	HG 丸ゴシック M-PRO	0.9150
	HG 明朝 E	0.9070
	MS ゴシック	0.9110
	ヒラギノ角ゴシック	0.9148
	ヒラギノ角明朝 ProN	0.9091

画像は文字と背景の境界が不明瞭であるため，画像生成後に 2 値化処理を行う．StarGAN v2 による生成画像は，図 2(a)のように画数の少ない文字では，文字の骨格が入力の現代フォントに依存する．文字の骨格が近代書籍フォントに最も近いのはヒラギノ角明朝 ProN である．図 2(b)のように画数の多い文字では，文字の骨格に差が少なく，いずれのフォントからも近代書籍フォントに近い文字画像が生成されている．フォントを自動生成する NN による生成画像は，画数が多い文字では文字線が潰れるが，StarGAN v2 による生成画像は, 文字の細部まで正確に再現されている．しかしながら，StarGAN v2 からは図 3 のように，文字線の欠損が著しい画像も生成される．正確な文字画像を生成できなかった数は，ヒラギノ角ゴシックを入力とした場合に最も多く，HG 明朝 E を入力とした場合に最も少ない．この原因は近代書籍フォントとの文字線の太さの違いであると考えられる．以上のことから，近代書籍フォントの生成に適しているのは，文字の骨格や文字線の太さが近代書籍フォントに近いものであるといえる．

次に, 近代書籍文字画像と各生成画像のパワースペクトルを比較した結果を示す．パワースペクトルは画像の持つ空間周波数成分の分布を表し，同じ特徴を持つ画像同士はパワースペクトルの差が小さい．各画像の余白を削除して画像サイズを 64×64px に統一した後，2 次元フーリエ変換を行い，その絶対値を求めることで各画像のパワースペクトルを得る．近代書籍文字画像と StarGAN v2 およびフォントを自動生成する NN から生成された画像のパワースペクトルを求め, その対数の平均平方二乗誤差(RMSE)を平均した値を表 2 に示す．StarGANv2 から生成された画像は，現代フォント 5 種類全ての場合において，フォントを自動生成する NN から生成された画像よりも RMSE の平均値が小さい．このことから，StarGAN v2 の方が近代書籍フォントの持つ特徴をより正確に再現した画像が生成されているといえる．

5. まとめ

本稿では，より精度の高い近代書籍フォントの自動生成を目指し，StarGAN v2 を用いて複数の現代フォントから近代書籍フォントの自動生成を試みる．その結果，5 種類の現代フォントから特定の近代書籍フォントの生成が可能であることが分かる．しかしながら，生成画像の文字の骨格は入力に用いた現代フォントに依存する．今後は，文字の骨格をフォントの特徴の 1 つとして学習させ，文字の骨格も近代書籍フォントに近い文字画像の生成を目指す.

謝辞　本研究は MEXT 科研費 JP20H04483 の助成を受けたものです.

参考文献

[1] 国立国会図書館デジタルコレクション．http://dl.ndl.go.jp.（参照 2021-06-01）.
[2] Ishikawa C., Ashida N., Enomoto Y., M. Takata, Kimesawa T. and Joe K.. Recognition of Multi-Fonts Character in Early-Modern Printed Books. Proceedings of International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA09), 2009, Vol. Ⅱ, pp. 728-734.
[3] Fukuo M., Enomoto Y., Yoshii N., Takata M., Kimesawa T. and Joe K.. EvaluaTion of the SVM based Multi-Fonts Kanji Character Recognition Method for Early-Modern Japanese Printed Books. Proceedings of International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA2011), 2011, Vol. Ⅱ, pp. 727-732.
[4] 粟津妙華，上坂和美，高田雅美，城和貴．近代書籍を対象とした多フォント活字認識手法，情報処理学会論文誌．数理モデル化と応用(TOM)，2016，Vol. 9(2)，pp. 33-40.
[5] Yasunami S., Takemoto Y., Ishikawa Y., Takata M. and Joe K.. Applying CNNs to Early-Modern Printed Japanese Character Recognition. Proceedings of the 2019 International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA'19), 2019, pp.189-195.
[6] Takemoto Y., Ishikawa Y., Takata M., Joe K.. Automatic Font Generation for Early-Modern Japanese Printed Books. Proceedings of The 2018 International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA2018), 2018, Vol. Ⅰ, pp. 326-332.
[7] Y. Takemoto, Y. Ishikawa, M. Takata and K. Joe: Structure of Neural Network Automatically Generating Fonts for Early-Modern Japanese Printed Books，Proceedings of International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA'19), 2019, pp.182-188.
[8] Isola P., Zhu J.-Y., Zhou T. and Efros A. A.. Image-to-Image Translation with Conditional Adversarial Networks. Computer Vision and Pattern Recognition (CVPR), 2017, pp.1125-1134.
[9] J.-Y. Zhu, T. Park, P. Isola, and Efros A. A.. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. International Conference on Computer Vision(ICCV), 2017, pp.2223-2232.
[10] Choi Y., Choi M., Kim M., Ha J.-W., Kim S., and Choo J.. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. Computer Vision and Pattern Recognition (CVPR), 2018, pp.8789-8797.
[11] Choi Y., Uh Y., Yoo J. and Ha J.-W.. StarGAN v2: Diverse Image Synthesis for Multiple Domains. Computer Vision and Pattern Recognition (CVPR), 2020, pp.8188-8197.