Vol.2017-MPS-112 No.15
2017/2/27

近代書籍用フォントの自動生成

竹本有紀† 上坂和美† 石川由羽† 高田雅美† 城和貴†

概要： 国立国会図書館では，近代書籍と呼ばれる明治から昭和初期にかけて刊行されていた書籍群を画像データとしてアーカイブ化し Web 上で一般公開している. 近代書籍の利便性を高めるために画像データの早急なテキスト化が望まれており，近代書籍に特化した認識手法が提案されているが，これには学習データとして膨大な種類の近代書籍漢字の画像が必要となる．学習データの収集には甚大なコストがかかり，また収集できる漢字の種類や個数には限界がある．学習データが十分に収集できなければ，文字認識の際に未知の漢字に遭遇し誤認識が生じる可能性がある．本稿では，近代書籍に特化した認識手法のためのフォントセットを自動生成する手法を提案する．ディープラーニングにより近代書籍から切り出した文字のフォントを学習し，明朝体の文字画像から近代書籍文字の特定のフォントに変換するフィルタを生成する．

キーワード：フォント生成，ディープラーニング，ニューラルネットワーク，近代書籍

Automatic Font Generator for Early-Modern Printed Books

YUKI TAKEMOTO† KAZUMI KOUSAKA† YU ISHIKAWA†
MASAMI TAKATA† KAZUKI JOE†

1. はじめに

国立国会図書館[1]では，近代書籍と呼ばれる明治から昭和初期にかけて刊行された図書・雑誌をデジタル化し，平成 14 年から近代デジタルライブライリーとして Web 上で一般公開している．近代書籍は哲学，歴史，文学や芸術など幅広い分野にわたり，また現在は絶版になっているものも多いため非常に貴重な資料である．インターネットに接続できる環境であれば，近代デジタルライブラリーを利用することで，図書館に足を運ぶことなく国立国会図書館の所有する資料を閲覧することができる．さらに，Web 上で公開されている資料はデジタルデータであるため，貴重な資料の劣化や破損，紛失を防ぐことができ，図書館で一般公開が困難な書籍も閲覧可能となる．近代デジタルライブライリーは平成 28 年 5 月にサービスを終了し,国立国会図書館デジタルコレクション[2]に統合され引き続き閲覧可能となっている．

公開されている資料は書籍をページごとにマイクロフィルムに撮影した画像データであり，文書内容を対象としたテキスト検索はできない．そこで，資料の利便性を高めるために文書内容の早急なテキスト化が望まれるが，公開されている資料の数は膨大であるため手作業でのテキスト化は効率的ではない．

現代の一般的な書籍であれば，光学文字認識（Optical Character Recognition，OCR）ソフトウェアの利用により書籍の画像データからテキストデータへの自動的な変換が可能である．一方，近代書籍は異字体・旧字体を多く含み，出版者や出版年代によって異なる種類の活字が使用されており，市販の OCR ソフトウェアを用いて正確な認識を行うことはできない．この問題を解決するために，様々な出版者や出版年代の活字に対応可能な近代書籍に特化した多フォント活字認識手法の研究[3][4][5]が進められている．この認識手法の精度を向上するためには学習データを増やす必要がある．

学習データの収集の主な方法としては，近代書籍の画像データから文字画像の切り出しが行われている．しかしながら，近代書籍で扱われる活字の種類は出版社や出版年代によって異なり，およそ 2 万種類のフォントが存在している.漢字によっては使用頻度の低さから収集が困難であり，また収集できたとしても近代書籍の画像データは印刷物であることから印刷時の滲みやかすれにより学習データとして不適切な場合もある．学習データとして不適切な文字画像の例を図 1 に示す. 左の文字画像はインクの滲みによって文字の細部が判別困難となっている．右の文字画像は印刷のかすれによって文字の一部が欠けてしまっている．よって，近代書籍の画像データから文字画像を切り出す以外に学習データを効率よく収集する手法が必要である．そこで本稿では，近代書籍用 OCR のためのフォントセットを自動生成する手法を提案する．フォントの持つ固有の特徴をディープラーニングによって学習し，現在容易に入手可能なフォントに学習した特徴を反映して変換することで，特定の出版社や出版年代のフォントを生成することが目標である．

本稿では，第 2 章でディープラーニングについて述べ，第 3 章においてディープラーニングを用いたフォントの自動生成の手法を提案する．第 4 章では提案する手法によってフォントを自動生成した結果について述べる．

† 奈良女子大学
Nara Women's University

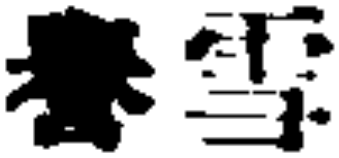


図 1　学習データとして不適切な文字画像の例

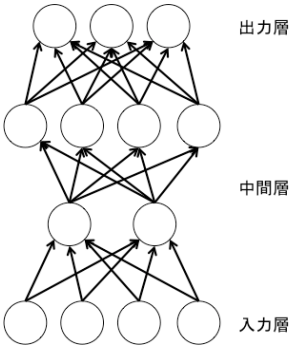


図 2　多層構造のニューラルネットワーク

2. ディープラーニング

ディープラーニング[6]は多層構造のニューラルネットワークを用いた機械学習の一手法である．多層構造のニューラルネットワークを図 2 に示す．画像や音声などの入力を中間層の各層へ順番に伝播し，1 層経るごとに学習が進んでいく．この過程で入力の持つ特徴を抽出し，パターン認識や画像生成などを行うことができる．

多層構造のニューラルネットワークにおいて，入力を$\boldsymbol{x}=\boldsymbol{h^0}$とすると，第$k$層の各ユニット$\boldsymbol{h}^k=\left[h_1^k,\dots,h_i^k,\dots\right]$への入力は，第$k-1$層のユニットへの入力$\boldsymbol{h}^{k-1}$から次のように計算される．

$$h_i^k=f\left(b_i^k+{w_i^k}^{\mathrm{T}}\boldsymbol{h}^{k-1}\right) \quad (1)$$

ここで，$\boldsymbol{W}=[w_1,\dots,]^T$とすると

$$\boldsymbol{h}^k=f(\mathbf{b}^k+\boldsymbol{W}^k\boldsymbol{h}^{k-1}) \quad (2)$$

と表される．$\boldsymbol{W}^k$は各層間の結合の重み，$\mathbf{b}^k$はバイアスである．関数fには何らかの非線形性を持つものが用いられる．本稿では，ReLU 関数を用いる．ReLU 関数は次の式で表される正規化線形関数である．

$$f(x)=max(0,x) \quad (3)$$

ニューラルネットワークは入力から出力に至るひとつの関数を実現し，各層間の結合の重み$\boldsymbol{W}^k$とバイアス$\mathbf{b}^k$がパラメータとなる．これらのパラメータをθとすると，θは訓練データを用いた学習によって決定し，未知の関数をニューラルネットワークに再現させることが学習の目的である．そのためにニューラルネットワークの出力と訓練データの誤差を計算し，この誤差が最小となるようにパラメータを更新する．$\theta^{(i)}$を$\theta^{(i+1)}$に更新する式の例を以下に示す．

$$\theta^{(i+1)}=\theta^{(i)}-\alpha\nabla E_k \quad (4)$$

このとき，∇E_kはニューラルネットワークの出力と訓練データの誤差であり，αは学習率と呼ばれるパラメータである．この値が大きいほど更新量が大きくなる．

このパラメータの更新を繰り返して学習することでニューラルネットワークは訓練データを再現することができるようになる．

ディープラーニングは様々な研究で用いられている．Le ら[7]は YouTube からランダムに切り出した大量の画像を訓練データとし，大規模なニューラルネットワークによる教師なし学習で人や猫の顔に反応するニューロンを自動生成した．マイクロソフトは畳み込みと再起のニューラルネットワークを用いて自動音声認識を行い，その精度は人間と同等に達したと報告している[8]．ディープラーニングでは画像や音声などが持つ特徴を手動で設計する必要がなく，学習によって必要な特徴が自動的に発見される．それにより高い性能を有し，ディープラーニングへの期待が高まっている．本稿では近代書籍の文字画像を訓練データとし，近代書籍のフォントが持つ特徴の学習を目指す．

3. ディープラーニングによるフォントの自動生成

フォントには，とめ，はね，はらいなどの部分に固有の特徴が現れると考えられる．この特徴をディープラーニングによって学習し，現在容易に入手可能なフォントに学習した特徴を反映して変換するニューラルネットワークを構築することで，特定の出版社や出版年代のフォントを生成する．

提案するフォントの自動生成手法は以下の通りである．このフローチャートを図 3 に示す．

(1)　教師データの読み込み
(2)　フォントの特徴の学習
　(a)　画像の生成
　(b)　誤差の算出
　(c)　誤差の逆伝播
　(d)　パラメータの更新
　(e)　(b)に戻る
(3)　フォントの生成
　(a)　元となる画像の読み込み
　(b)　画像の生成

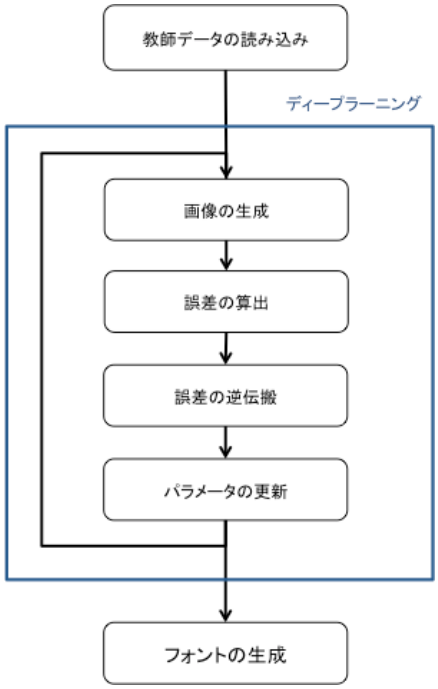


図 3　フォントの自動生成アルゴリズム

手順(1)では教師データとなる画像の読み込みを行う．手順(2)(a)でニューラルネットワークによって画像を生成し，手順(2)(b)で生成された文字画像と教師データの近代書籍の文字画像から誤差を算出する．手順(2)(c)では手順(2)(b)で算出した誤差をニューラルネットワークに逆伝播する．手順(2)(d)でパラメータを更新して手順(2)(a)に戻る．学習終了後は手順(3)(a)でフォント生成の元となる文字画像を読み込み，手順(3)(b)では学習によって得られたニューラルネットワークでフォントを生成する．それぞれの手順の詳細については 3.1 節から 3.6 節で述べる．

3.1 教師データの読み込み

教師データはフォント変換の元となる文字画像と近代書籍の特定フォントの文字画像である．それぞれ 1 枚ずつの画像データを 2 枚セットで 1 組の教師データとする．本稿では元となる画像を明朝体の文字画像とする．明朝体は現在広く普及しているフォントの 1 つであり，いかなる作業環境においても容易に入手可能である．近代書籍の特定フォントの文字画像は，同一のフォントが使われている近代書籍の画像データから切り出したものである．読み込んだ教師データは 2 値化し，縦と横のサイズを同じものにする．これを訓練データとテストデータに分ける．訓練データによって学習したニューラルネットワークの精度をテストデータによって確認するためである．

3.2 画像の生成

元となる文字画像を入力とし，フォントを変換した画像を出力とするようなニューラルネットワークを構築する．設計したニューラルネットワークを図 4 に示す．ニューラルネットワークはそれぞれ 4 つの畳み込み層と逆畳み込み層

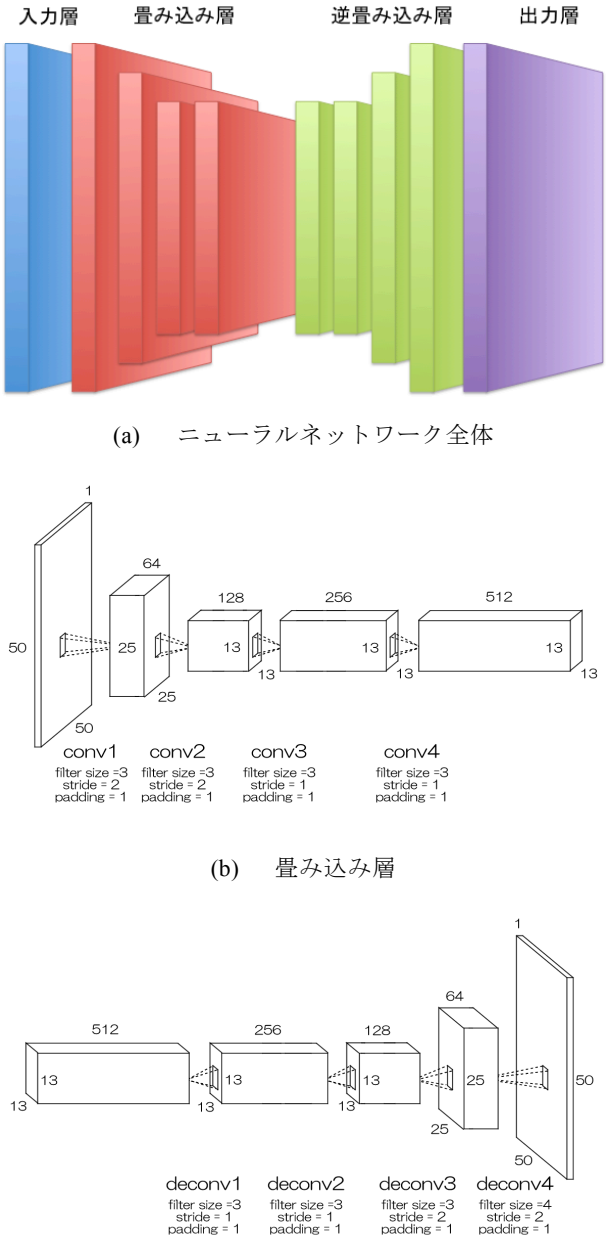


図 4　フォントを自動生成するニューラルネットワーク

によって構成される．図 4 (a)はニューラルネットワーク全体の図であり，図 4 (b)，図 4 (c)はそれぞれニューラルネットワーク内の畳み込み層，逆畳み込み層の図である．畳み込み層では各 4 層に dropout 関数を適用する．dropout 関数は各層間の結合において一定の割合でランダムにユニットを使用しないようにするもので，これにより過学習を防ぐ．過学習とは訓練データの特徴を過剰に学習してしまい，訓練データに対する精度は高い一方で訓練データにはない未知のデータに対する精度が低い状態である．本稿では dropout を行う割合は 90%とする．この割合は予備実験として dropout の割合を変化させてニューラルネットワークに

よる学習を行った結果，過学習が起こることがなく学習に成功した割合である．また，訓練データとテストデータはミニバッチに分けて学習を行う．分けられたミニバッチをランダムな順番で学習することで，訓練データのノイズによる変動を抑える効果があり，より精度の高い結果を得ることができる．

3.3 誤差の算出

本稿におけるニューラルネットワークの学習のために用いる誤差は，生成画像と訓練データの近代書籍の文字画像における各画素値の二乗平均誤差とする．画像の各画素値の誤差が最小となるように学習を行うことで，目標となる近代書籍の特定フォントの生成を目指す．

3.4 パラメータの更新

本稿では，学習の最適化手法として Adam[9]を用いる．Adam の更新式は以下のように表される．

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \quad (5)$$
$$\theta_t = \theta_{t-1} - \alpha\, \hat{m}_t / \sqrt{\hat{v}_t} \quad (6)$$
$$\hat{m}_t = m_t / \left(1 - {\beta_1}^t\right) \quad (7)$$
$$\hat{v}_t = v_t / \left(1 - {\beta_2}^t\right) \quad (8)$$
$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (9)$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) {g_t}^2 \quad (10)$$

ここで，g_tは目的関数$f_t(\theta_{t-1})$の勾配であり，θは学習するパラメータを表す．αは学習率，m_tとv_tはそれぞれ勾配の1次モーメントと2次モーメントの概算値である．β_1，β_2は過去の勾配による影響を減衰させるパラメータである．この更新式によりパラメータを更新する．

3.5 ニューラルネットワークの精度

訓練データによって学習したニューラルネットワークは，パラメータを更新する前にテストデータによって精度を確認しておく．テストデータのうち，元となる文字画像を入力としたニューラルネットワークの出力と，テストデータの近代書籍の文字画像の二乗平均誤差を求める．この値の平均が最も小さくなるニューラルネットワークが最も精度の高いニューラルネットワークとなる．

3.6 フォントの生成

テストデータの誤差の平均が最も小さいニューラルネットワークを用いる．このニューラルネットワークに新たな明朝体の文字画像を入力として与えると，近代書籍文字の特定フォントに変換した画像が自動生成される．

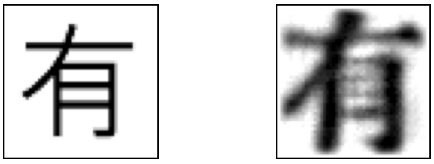
4. フォントの生成実験

本稿で提案するニューラルネットワークの有用性を検証するために実験を行う．

4.1 実験方法

今回の実験で学習する近代書籍のフォントとして，大正時代に日吉堂から出版された書籍の画像データから切り出

図 5　明朝体の文字画像（左）と日吉堂大正フォントの文字画像（右）



(a) 元となる明朝体の文字画像　(b) 自動生成された文字画像

(c) 日吉堂大正フォントの文字画像

図 6　ニューラルネットワークによって自動生成された文字画像の例

された文字画像を用いる．これ以降はこのフォントを日吉堂大正フォントと呼ぶものとする．明朝体と日吉堂大正フォントの文字画像 1 枚ずつ，計 2 枚の画像データを 1 セットの教師データとする．教師データの例を図 5 に示す．図 5 の右図は明朝体の文字画像，左の図は日吉堂大正フォントの文字画像である．用意した画像データセットは 1407 種類であり，このうち 1000 種類の漢字をランダムに選び，ニューラルネットワークの学習のための教師データとする．教師データはさらに訓練データとテストデータに分けられる．今回の実験では 900 種類を訓練データとし，100 種類をテストデータとする．学習時の dropout 率は 90%，ミニバッチのサイズは 100 である．

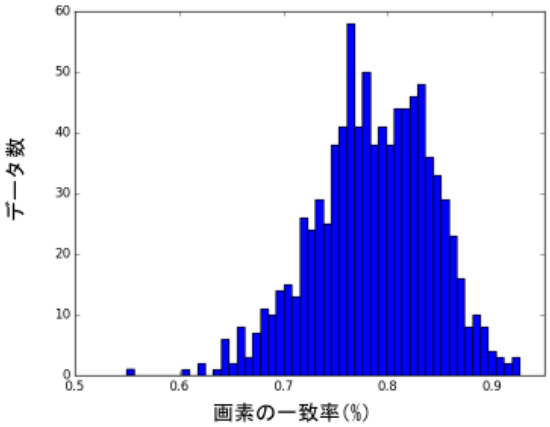
学習終了後は学習によって得られたニューラルネットワークを用いてフォントを自動生成する．訓練データとして用いた 900 種類と，訓練データとして用いなかった残りの 507 種類を合わせた 1407 種類すべての明朝体の文字画像からフォントを自動生成する．生成された文字画像と日吉堂大正フォントの文字画像で画素値の比較を行い，ニューラルネットワークの有用性を検証する．

4.2 実験結果

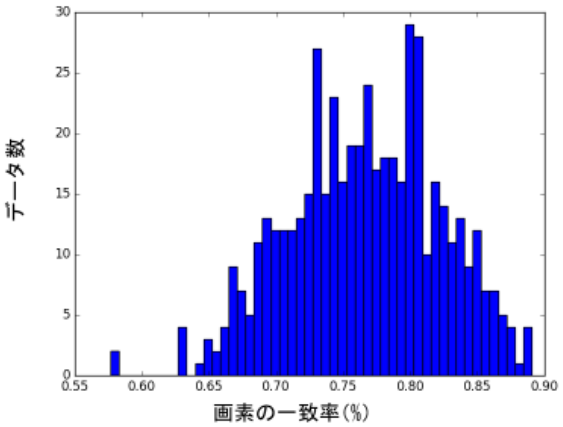
学習によって得られた最も精度の高いニューラルネット

表 1　自動生成された文字画像と日吉堂大正フォントの文字画像の画素の一致率（%）

	平均	分散	最小値	最大値
訓練データ	78.60	0.003148	54.96	92.60
未知データ	76.58	0.003157	57.68	89.04



(a)　訓練データの文字から自動生成したフォントと日吉堂大正フォントの一致率のヒストグラム



(b)　未知データの文字から自動生成したフォントと日吉堂大正フォントの一致率のヒストグラム

図 7　画素の一致率のヒストグラム

ワークを用いて自動生成された文字画像の例を図 6 に示す．図 6 (a)はフォント生成の元となる明朝体の文字画像，図 6 (b)はニューラルネットワークによって自動生成された文字画像，図 6 (c)は日吉堂大正フォントの文字画像である．自動生成された文字画像の，とめ，はね，はらいなどの部分に，元となる明朝体にはなかった，日吉堂大正フォント固有と思われる特徴が表れていることが目視によって確認できる．

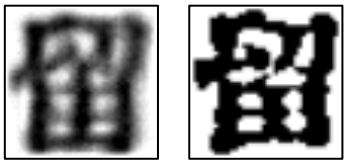


図 8　未知データから生成された文字画像のうち，画素の一致率が最も高い文字画像（左）と日吉堂大正フォントの文字画像（右）

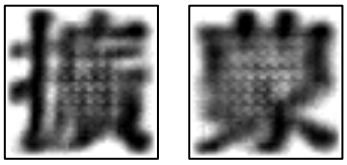


図 9　細部が再現できていない文字画像の例

自動生成された文字画像と日吉堂大正フォントの文字画像の画素の一致率を求めた結果を表 1 に示す．画素の一致率は，自動生成されたフォントの文字画像と日吉堂大正フォントの文字画像を 2 値化し，画像のすべての画素値を 1 ピクセルごとに比較して 2 枚の画像の画素値が一致した割合である．

自動生成されたフォントの文字画像は，ニューラルネットワークの学習のための訓練データとして用いる 900 種類の漢字と，訓練データとして用いない未知データである 507 種類の漢字に分けて比較を行う．訓練データの文字から自動生成したフォントと日吉堂大正フォントの画素の一致率の平均は 78.60%, 未知データの文字から自動生成したフォントと日吉堂大正フォントの画素の一致率の平均は約 76.58%である．それぞれのデータにおける画素の一致率のヒストグラムを図 7 に示す．図 7 (a)は訓練データの文字から自動生成したフォントと日吉堂大正フォントの一致率のヒストグラム, 図 7 (b)は未知データの文字から自動生成したフォントと日吉堂大正フォントの一致率のヒストグラムである．グラフの横軸が自動生成したフォントと日吉堂大正フォントの一致率，縦軸が各一致率のデータの個数である．

5. 考察

未知データの文字から自動生成したフォントの画素の一致率の平均値は，訓練データの文字から自動生成したフォントの画素の一致率の平均値と同じ程度の結果となった．このことから，学習によって得られたニューラルネットワークは，未知の文字画像に対しても訓練データから学習した近代書籍文字の特定フォントの特徴を再現したフォントを生成できることが分かる．

また，画素の一致率の平均値は訓練データ，未知データのいずれの場合でも 8 割程度である．近代書籍の文字画像

には滲みやかすれの他に，書籍を撮影した画像から切り出した文字画像であるが故の歪みを持つものがあるので，画素の一致には限界がある．しかしながら，目標である日吉堂大正フォントを再現できたと述べるに十分な割合とは言えない．未知データから生成された文字画像のうち，画素の一致率が最も高い文字画像を図 8 に示す．この例は日吉堂大正フォントの特徴を再現できているが，画数の多い漢字ではフォントの細部が不明瞭になる場合がある．その例を図 9 に示す．これは，ニューラルネットワークが訓練データの特徴を細部まで学習できていないためであると考えられる．画数の多い複雑な漢字は使用頻度が低いものも多く，また活版印刷の際の滲みやかすれの影響を受けやすいと考えられる．そのため，自動生成された文字画像を近代書籍に特化した文字認識手法の学習データとして利用できるようにするためには，画数の多い複雑な漢字でも学習データとして用いることができる精度でのフォントの自動生成が必要となる．

6. まとめ

本稿では，近代書籍に特化した文字認識手法に必要な学習データのフォントセットをディープラーニングによって自動生成する手法を提案した．ニューラルネットワークはそれぞれ 4 つの畳み込み層と逆畳み込み層によって構成され，フォント生成の元となる文字画像を入力として，訓練データから学習したフォント固有の特徴を反映した文字画像を出力として生成する．教師データとしては，フォント生成の元となる文字画像と，目標となる近代書籍の特定フォントの文字画像を用意する．近代書籍の文字画像は，近代書籍の画像データから 1 文字ずつ切り出された画像である．

構築したニューラルネットワークによってフォントを自動生成する実験では，明朝体の文字画像と大正時代の日吉堂の書籍から切り出した文字画像をそれぞれ 1407 種類ずつ用意し，このうち 900 種類を訓練データとしてディープラーニングによる学習を行った．これにより，大正時代の日吉堂のフォントが持つ固有の特徴を反映した文字画像を自動生成することに成功した．しかしながら，その精度は近代書籍に特化した文字認識手法の学習データに用いるには不十分である．今後はニューラルネットワークの改良を行い，より精度の高い文字画像の自動生成を目指す．

参考文献

[1] 国立国会図書館. http://www.ndl.go.jp. (参照 2017-01-30).
[2] 国立国会図書館デジタルコレクション. http://dl.ndl.go.jp. (参照 2017-01-30).
[3] Ishikawa, C., Ashida, N., Enomoto, Y., Takata, M., 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), Vol.Ⅱ, pp. 728-734(2009).
[4] 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 The 2011 International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA2011), Vol.Ⅱ, pp. 727-732(2011).
[5] 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)，Vol. 9(2)，pp. 33-40(2016)
[6] 岡谷貴之，齋藤真樹．ディープラーニング．情報処理学会研究報告，Vol.2013-CVIM-185, No.19, pp.1-6, 2013.
[7] Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, Andrew Y. Ng . Building High-level Features Using Large Scale Unsupervised Learning. ICML, 2012.
[8] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu and G. Zweig. Achieving Human Parity in Conversational Speech Recognition. MSR-TR-2016-71, 2016.
[9] Diederik P. Kingma, Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. ICLR, 2015.