

文字認識器訓練のための手書き文字生成

北川 智樹* レオ チー シャン†,‡ 西崎 博光†

山梨大学工学部* 山梨大学大学院医工農学総合教育部† Artibrains 合同会社‡

1 はじめに

手書きで書かれた書類を撮影し、その画像から文字を読み取りテキスト化する技術に Optical Character Recognition (OCR) がある。近年では、深層学習を利用した Artificial Intelligence (AI)-OCR 技術が実用化されている。AI-OCR 技術で高精度な文字認識器を訓練させるには大量の手書き文字データが必要である。しかし、手書き文字を大量に収集するには、人手や時間など大きなコストがかかる。また、本研究で対象とする日本語は、他言語と比べて、ひらがな、カタカナ、漢字と文字の種類数が多いことから、データ収集が容易ではない。

本研究では、深層生成モデルを用いて文字画像を生成することで文字認識器の訓練データを増やし、文字認識器の認識性能向上を図ることを目的とする。

深層生成モデルとして本研究では Autoencoder (AE) [1] を用いる。また AE の訓練方法に Y-Autoencoder (Y-AE) [2], モデルに Adaptive Instance Normalization (AdaIN) [3] 層を組み込み、既存の Y-AE だけでは生成できなかったカタカナ、ひらがなの文字画像を生成できるようにした。生成された文字画像を手書き文字認識器の訓練データに加え、手書き文字認識性能向上に対する文字画像生成の有効性を確認した。実験の結果、文字認識器の訓練データに生成画像を加えることで、加えない場合に比べて認識性能の向上を確認でき、本手法が文字認識器訓練のための手書き文字生成として有効であることが確認できた。

2 Y-Autoencoder (Y-AE)

Y-AE とは、Patacchiola ら [2] によって提案された条件付き AE である。本研究の文字生成を適応させた Y-AE の概要を図 1 に示す。一般的な AE との違いは、入力 x に対して Encoder が特徴量 i と何の文字かの認識結果 e を出力する点、Decoder への入力が特徴量 i とラベル y となっている点である。Y-AE の訓練時、Encoder は文字認識と特徴量圧縮を同時に最適化することにより、文字の形で分離

可能な特徴量の出力を学習する。Decoder は 2 つのブランチに分かれ、左のブランチでは Decoder に特徴量 i と入力 x のラベル y を与えて画像を生成し、生成画像 $\hat{x}_L$ と入力画像 x との間に平均二乗誤差 (MSE) で制約を加え、文字の復元性を担保しながらラベルに対する文字の形を学習する。右のブランチでは Decoder に特徴量とランダムなラベル $\tilde{y}$ を与えて画像を生成し、もう一度 Encoder に生成画像を入力し、認識結果 $\hat{e}_R$ を得る。認識結果 $\hat{e}_R$ とランダムなラベル $\tilde{y}$ が同じ値になるように制約を与えることで特徴量を考慮した他の文字への変化を学習する。また、左右のブランチで生成された画像を Encoder に入力し、生成画像の特徴量 $\hat{i}_L, \hat{i}_R$ に対して MSE で制約を与え、左右のブランチの生成画像の特徴量の一貫性を担保する。

以上により、Y-AE は図 2 に示すように、Encoder で特徴量を得て、任意のラベルと得た特徴量を Decoder に入力することで特徴量を考慮した文字生成ができる。

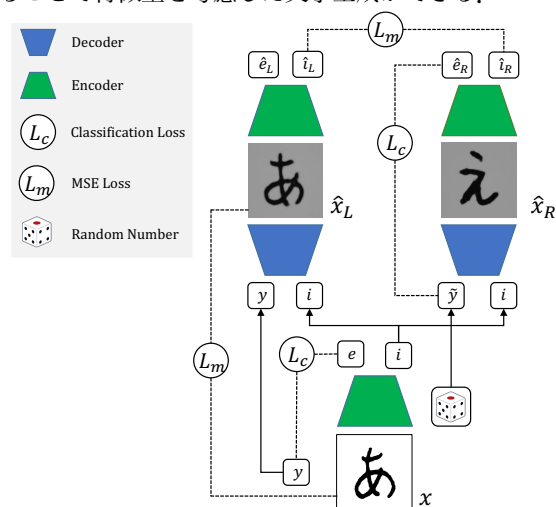


図 1 Y-Autoencoder (Y-AE) の概要

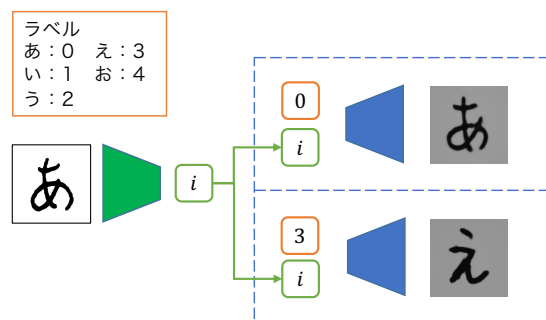


図 2 生成時の Y-Autoencoder (Y-AE)

Hand-written Character Image Generation for Character Recognizer Training

* Tomoki Kitagawa, Faculty of Engineering, University of Yamanashi

† Leow Chee Siang, Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi

‡ Artibrains LLC

† Nishizaki Hiromitsu, Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi

3 Adaptive Instance Normalization (AdaIN)

AdaIN[3]とは画像のスタイル変換で使われる層である。式(1)で定義されるように、本研究の文字生成における任意の文字の形であるコンテンツ c の特徴量の平均と分散を文字生成における文字の筆跡であるスタイル s の特徴量の平均と分散に合わせることで、学習時に一切のパラメータを持たないでスタイル変換を可能にした。

$$AdaIN(c, s) = \sigma(s) \left(\frac{c - \mu(c)}{\sigma(c)} \right) + \mu(s) \quad (1)$$

4 Y-AE を用いた文字生成

本研究では Y-AE の Decoder に AdaIN 層を組み込んだ生成モデルで文字生成を行った。

4.1 文字生成モデルの訓練

1枚のカタカナ画像から46種類のカタカナ画像を生成するモデル、1枚のひらがな画像から46種類のひらがな画像を生成するモデルの2つを訓練した。訓練に使用したデータは ETL 文字データベース [4] の9つの手書き文字コーパスのうち、カタカナ生成モデルは ETL5 のカタカナ各208枚の計9,568枚、ひらがな生成モデルは ETL9 のひらがな各200枚の計9,200枚を使用して訓練された。

4.2 文字画像生成

カタカナ生成モデルの訓練に使用したカタカナ画像、ひらがな生成モデルの訓練に使用したひらがな画像から、生成モデルで約86万枚の文字画像を生成した。実際に生成させた例を図3に示す。背景が白の画像を Y-AE に入力し、ラベルに応じて46種類の文字画像が生成されていることが分かる。また入力する画像によって生成される文字画像が異なることも見て取れる。

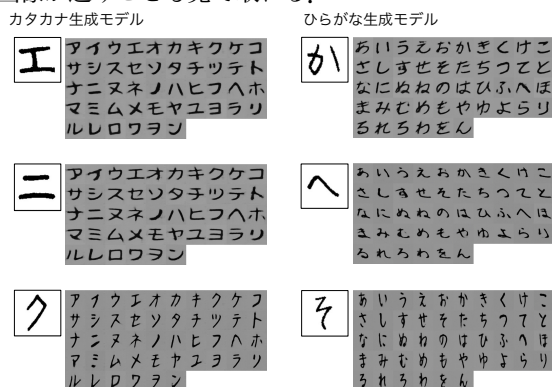


図3 生成画像

5 手書き文字認識実験

生成された文字画像を文字認識器の訓練に使用し、手書き文字認識率の評価を行う。

5.1 実験条件

カタカナ46種類、ひらがな46種類、計92種類の文字認識器を訓練した。ベースラインは手書き文字画像だけを提案手法は手書き文字と Y-AE による生成画像から訓

表1 文字認識器の精度

	F1 スコア
ベースライン (ETL5 + ETL9)	0.7089
提案手法 (ETL5 + ETL9 + 生成画像)	0.8602

練したモデルで、ETL の検証データに対する F1 スコアで提案手法の有効性を調べた。文字認識器のモデル構造は VGG16[5] とし、同じイテレーション回数で訓練したモデルで比較した。また、他のハイパパラメータもベースラインと提案手法で一致させた。ベースラインの訓練データは Y-AE の訓練・生成に使用した ETL9 と ETL5 とし、提案手法の訓練データはベースラインの訓練データと Y-AE による生成画像全てとした。検証データは ETL1 からランダムに各カタカナ画像を200枚、ETL7 からランダムに各ひらがな画像を200枚を選んだ。すなわち、オープンなデータで検証している。

5.2 実験結果

表1に、既存のデータセットのみを用いて訓練した文字認識器と、Y-AE による生成画像を加えて訓練した文字認識器の認識結果を示す。ベースラインに比べて提案手法は検証データに対する F1 スコアが 0.1513 pt と大幅に向上した。このことから、Y-AE によって生成された文字画像を加えて訓練することで、手書き文字認識性能が大幅に向上することが確認できた。

6 おわりに

本研究は、文字認識の精度向上を目的とし、深層生成モデルによって手書き文字認識の訓練に必要な文字画像を生成することで文字認識器の訓練データを増やす手法を提案した。手書き文字認識実験の結果、生成された文字画像を文字認識器の訓練データに加えることで文字認識器の認識性能が向上したことから、深層生成モデルによる文字生成が手書き文字認識性能の向上に有効であることが確認できた。

今後の展望として、漢字の生成に取り組み、より複雑な形の文字でより多くの文字の種類が扱える文字認識器に対して、生成画像を訓練データに加えることによって認識性能向上を目指す。

参考文献

- [1] G. E. Hinton, R. P. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol.313, no.5786, pp.504-507, 2006.
- [2] M. Patacchiola, P. Fox-Roberts, E. Rosten, "Y-Autoencoders: disentangling latent representations via sequential-encoding," Pattern Recognition Letters, vol.140, pp.59-65, 2020.
- [3] X. Huang, S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," Proc. of the 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1510 - 1519, 2017.
- [4] 独立行政法人産業技術総合研究所, "ETL 文字データベース" <http://etl1cdb.db.aist.go.jp/>, (2022.1.6 参照)
- [5] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Proc. of ICLR 2015, pp.1-14, 2015.