視写課題の自動採点へ向けた
子供らしい文字の自動生成による OCR 精度の向上

関 歩実　滕 昱璇　野坂 瞭太　大井 淳司　松崎 拓也
東京理科大学 理学部第一部 応用数学科
{1423518,1420068,1420078,1420020}@ed.tus.ac.jp　matuzaki@rs.tus.ac.jp

概要

本研究では，児童による文章の視写結果の自動採点へ向けた基礎技術開発を行った．自動採点の前提として，お手本の文字列と視写結果の文字画像列を正確にアラインし，視写の誤りを検知する必要がある．そして，正確なアラインメントのためには，児童の手書き文字に対する OCR の精度を向上させることが必要である．そこで本研究では，文字のスタイル変換手法を利用して，児童の手書き文字を大量に生成し，OCR の訓練データとして利用することを試みた．結果として，大人の手書き文字のみを訓練データとした場合と比べて，児童の手書き文字の認識精度が向上し，これに伴ってお手本と視写結果のアラインメント精度も向上した[1)]．

1　はじめに

小学校において板書を書き写す作業が授業時間内に終わらない児童が一定数存在する．故に，小学校から大学までを通じ，穴埋めプリントやスライドを用いた授業形式が増えている．結果として，板書を写す速度が一層遅くなっていると考えられる．

我々は，視写課題の実施をサポートするシステムを開発することで，書き写す能力の向上を狙う．システムは，視写対象の教科書テキストと視写結果の文字画像を入力とし，誤字・脱字等についてフィードバックを返すものを想定している．視写の効果に関しては，これまでに以下のような報告がある．江川 [1] は，大学院生を調査対象に，視写課題を行うことで作文能力の向上を示した．Askov と Greff [2] は，幼稚園児と小学 2 年生を調査対象に，速記に用いる文字の視写またはなぞり書きの練習を行った結果，なぞり書きよりも視写がより文字の再現性に効果があると示した．中島 [3] は，母語が日本語ではない日本語学習者を調査対象に，視写課題によって 1 分あたり正しく書き写せる文字数が向上することを示した．以上の研究結果から，教科書の文章を書き写す課題を課すことで児童たちが板書の視写に要する時間が短縮され，授業時間内により高い学習効果を得られると期待できる．

視写結果に関するフィードバックのために，お手本テキストと視写結果の各文字の正確なアラインメントをとる必要があり，その前提として，OCR（光学文字認識）の精度を高めることが必要である．ディープラーニングによる現在の OCR は，活字の認識に関しては実用レベルの精度に至っている．一方，手書き文字に対しても認識精度は高いものの，字形の特徴によっては識別できない場合が存在する [4]．児童の書く文字は字形が多様である上に，児童の手書き文字のデータを大量に用意することは難しい．そこで本研究では，児童の字を自動生成して学習データとして用いることで，OCR の精度向上を試みる．類似の試みとして Kitagawa ら [5] の研究があるが，児童の手書き文字らしさに注目したスタイル変換の応用は新しい試みである．

2　手法

2.1　自動文字画像生成

文字画像生成に関し拡散モデルに基づく手法と GAN に基づく手法を比較した．順に説明する．

2.1.1　拡散モデルと部品埋め込みを用いた文字生成

この項で説明する手法は Nikolaidou らの拡散モデルによる英単語画像生成手法をベースにした [6]．Nikolaidou らのモデルは英字の埋め込みの列とスタイルの埋め込みを入力とし，潜在空間上の拡散モデルによって指定したスタイルによる英単語画像を生成する．我々はこのモデルを用いて英単語の代わり

1）　第 1〜第 3 著者は同等の貢献をした．

に日本語一文字を生成する．

文字埋め込みは次の 2 種類を比較する．

文字単位の文字埋め込み　すべての文字に異なる埋め込みを与える．

部品単位の文字埋め込み　漢字には部首のように共通する形を持つものがある．この特性から，漢字を適切に分解して埋め込み表現を作ると，異なる漢字の訓練データから共通部分を抽出するように学習でき，より正しい字形で高品質な生成が可能になると考えられる．この分解を KanjiVG [7] を用いて行う．KanjiVG は各漢字のストローク集合を部分－全体関係を表す木構造で格納している[2)]．まず木構造の各頂点の子孫のストロークからなる部分を画像として出力し，この画像を基に漢字を一定の大きさ以上の部品列に分解する．また，画像を K-Means 法でクラスタリングする．部品の種類の埋め込みとクラスター番号の埋め込みを結合して部品の埋め込みを作り，部品の埋め込みの列を漢字の埋め込みとする．この方式では Nikolaidou らの設定における英単語に漢字が，英字に部品が対応する．学習時には，部品の埋め込みが正確にその部品を表現できるよう，部品のみの画像も訓練データに加える．

Nikolaidou らは書き手の筆跡を真似た英単語画像の生成を目的としており，それぞれの書き手の ID を個別の「スタイル」とみなしている．これに対し，OCR の訓練にはできるだけ多様な画像を用意できることが望ましいので，本研究では訓練データを分類し「大人の手書き」「児童の手書き」「活字」の 3 つの ID を「スタイル」とみなす．拡散モデルは生成過程で確率分布からのサンプリングを多数行うため，「児童の手書き」のように粗い分類に対応するスタイルを指定することで様々な画像が生成できると期待できる．

2.1.2　GAN と「部首」データを用いた文字生成

この項では Kong ら [8] の文字生成手法を基にした手書き文字生成に関して説明する．この生成手法では，入力された 2 つの文字画像 C および S に対し，C に書かれた文字を S のスタイルで書いたような文字画像を生成する．訓練時には，生成された文字か否かの real/fake 識別やスタイルの分類による生成器へのフィードバックに加え，漢字構成記述文字列 (IDS) に従って分解された各文字の構成要素に対しても識別 (real/fake 識別) やスタイルの分類を行う．例えば，「作」という字は IDS によって「亻」＋「乍」に分解される．単純に GAN を用いたモデルと比較して，文字の細部についても識別器からのフィードバックを受けることでより高品質な文字画像の生成が可能となる．

「部首」データの作成　日本語の漢字の分解に関しては，IDS データ [9] を使用する．さらに本研究では，ひらがな間のループ部分の形状の類似性を捉えるため，ひらがなに対しても分解データを作成し，訓練に使用する．これは，予備的な検討においてループ部分を上手く表現できない現象が観察されたためである．

ひらがなの分解に関しては，以下の通りに行う．

- 濁点を記号「゛」，半濁点を「゜」で表し，「べ→へ＋゛」，「ぺ→へ＋゜」のように分解する．
- 「す」や「む」のように上から入り下へ抜けるループを含む文字：ループ部分を「゜」で表し，「す→す＋゜」，「む→む＋゜」のようにループ形状の共通性を表す．
- 「め」と「ぬ」，「れ」と「ね」，「ろ」と「る」：それぞれ付加されるループ部分を「α」や「゜」で表し，ぬ→め＋α，ね→れ＋α，る→ろ＋゜のように分解する．
- 「は」「ほ」「ま」のように上から入り右へ抜けて終わるループを含む文字：ループ部分を「〆」で表し，「ほ→ほ＋〆」のように表す．

2.2　生成文字画像を加えた OCR の訓練

2.2.1　OCR モデルと訓練データの水増し

日本語の文字認識に関する Tsai の報告 [4] を参考にして，OCR モデルを作成した．具体的には論文 [4] で M7_1 と呼ばれているモデルを使用した．これは，4 層の畳み込みネットワークに，全字種 (3088 種) に対するスコアを出力する 3 層の全結合層を付加したモデルである．

OCR モデルの精度向上のため，各文字の画像のバリエーションを増やすとともに，文字ごとの訓練データのサンプル数の偏りを解消し，文字認識のバランスを改善する．具体的には，画像をランダムに 0.5〜1.5 倍に拡大縮小し，-30〜30 度で回転させることによりデータの水増しを行う．

2)　例えば，「線」は「糸」と「泉」から成り立ち，さらに「泉」は「白」と「水」から成り立つ．「白」はさらに上部の点と「日」から成り立つ．

表 1: 訓練データとテストデータ

	データの数			書き手
	ひらがな	カタカナ	漢字	の数
視写結果 (訓練)	14,168	214	3,719	218
視写結果 (テスト)	7,214	117	1,949	113
ETL (訓練)	10,867	9,537	533,700	5,236
ETL (テスト)	1,208	1,071	59,300	582

2.3 OCR を用いたお手本と視写結果のアラインメント

お手本のテキストと視写結果の文字画像の列のアラインメントを自動的に行う方法を説明する．まず，お手本の視写結果に OCR を適用し，視写結果の各文字がお手本の各文字クラスである確率を得る．

次に，「お手本の文字列」と「視写結果の各文字画像列」の間の編集距離を最小にするアラインメントを求める．一般的な編集距離は，与えられた 2 つの文字列のうち一方に対して「文字の挿入」「文字の削除」「文字の置換」の操作を何回繰り返すことで他方の文字列と一致するかを測ったものである．本研究では，視写結果中のある文字画像 i をお手本中の文字 c に置換するコストを，OCR モデルによる画像 i に対する文字クラス c の確率 $p(c|i)$ の対数としている．「文字画像の挿入」「お手本の文字の削除」のコストは，いずれも $\log 10^{-3}$ とした．

3 実験

本節では，拡散モデルと GAN で 生成した文字画像を使用し，FID を用いて児童の手書きスタイルの表現を評価する．また，OCR の精度と，お手本と視写結果のアラインメントの精度を評価する．

3.1 使用データ

データセットとして小 1 から 小 6 まで計 338 人の視写結果と，大人の手書き文字として ETL 文字データベース [10] [3] を使用した．画像の大きさは 64×64 とした．訓練データとテストデータへの分割は表 1 の通りである．

3.2 文字画像の生成例

文字単位拡散モデルは訓練データから ETL の漢字データを 133,281 枚までに制限したもので訓練し，「児童の手書き」スタイルで各文字 500 枚生成した．GAN では，ひらがな，カタカナ，漢字をそれぞれ分けて訓練を行った．漢字に関しては，ETL の漢字データ 133,281 枚と児童の漢字データを訓練に使用した．ひらがな，漢字の生成では，訓練に使用した児童の手書き文字を書き手ごとに各 2〜3 枚スタイルとして計 500 枚使用し，各文字 500 枚生成した．カタカナの生成では，児童の訓練データ 214 枚をスタイルとして使用し，各文字 500 枚生成した．これらの生成例を図 1 に示す．また，GAN のひらがなの生成において分解データを使用しない場合と使用した場合の生成例を図 3 に示す．

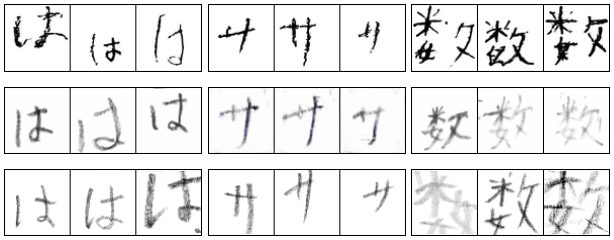

図 1: 文字単位拡散モデル (上) と GAN (中) の「は」「サ」「数」の生成例，及び児童の視写結果の例 (下)

図 2: 文字単位拡散モデル (左) と部品単位拡散モデル (右) の「類」の生成例．「類」は ETL に含まれるが児童の視写結果に含まれない

部品単位拡散モデルは文字単位拡散モデルで使用した漢字の訓練データと KanjiVG の部品画像 39,856 枚で訓練し，「児童の手書き」スタイルで各漢字 500 枚生成した．文字単位拡散モデルに比べて字形を保って生成できているものが見られる (図 2).

3.3 生成結果の FID による評価

画像の分布間の距離を測る指標 Fréchet Inception Distance (FID) [11] を用いて生成モデルの品質を評価した．前節で生成した各手法の結果と児童の視写結果のテストデータとの間で計測した FID を表 2 に示す．小さいほど児童が書いたような字が生成されたと考えられる．ひらがなに関しては，ETL の訓練データと比較して GAN の生成した文字画像が最も児童の手書き文字に近いといえる．

3.4 OCR 及びアラインメント精度

OCR モデルの訓練には，まず ETL と 3.2 節の生成結果を使用した．生成した結果は，各文字を 500 枚を生成し，550 枚まで水増した．各データの延べサンプル数を表 3 に示す．訓練データを以下の 5 つの

3) 9 つのコーパスのうち，ひらがなに ETL8G，カタカナに ETL5，漢字に ETL9G を使用した．

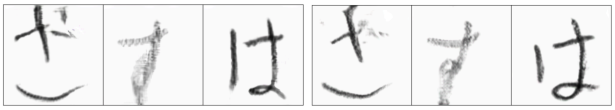

図 3: GAN の分解データなし (左) と分解データあり (右) の「ざ」「ね」「は」の生成例

表 2: 児童の視写結果のテストデータとの FID

	ひらがな	カタカナ	漢字	すべて
視写結果 (訓練)	2.18	41.30	6.31	1.87
ETL (訓練)	41.18	153.13	61.60	46.25
文字単位拡散モデル	22.58	61.30	32.93	31.87
部品単位拡散モデル	-	-	32.93	-
GAN と「部首」	7.97	242.77	19.98	19.43

パターンで組み合わせ，OCR モデルを訓練した：

A. ETL
B. ETL + 生成した児童の字 (文字単位拡散モデル)
C. ETL + 生成した児童の字 (部品単位拡散モデル)
D. ETL + 生成した児童の字 (ひらがな&カタカナ：文字単位拡散モデル，漢字：部品単位拡散モデル)
E. ETL + 生成した児童の字 (GAN)

更に，A〜E のそれぞれに児童の手書き文字 (本物) 延べ 17,987 画像を加え，モデル A'〜E' を訓練した．

訓練データに含まれない児童の字に対する OCR 及び視写結果のアラインメント精度を表 4 に示す．また，文字種類別の OCR の F1 スコアは表 5 に示す．比較として，大人の字のみ使用しているモデル A は，訓練データに含まれない大人の字に対する正解率が 99.05，F1-Score が 99.13 である．これは大人の字と児童の字の特徴がかなり異なり，訓練データに児童の字を含めることの重要性を示している．

大人の手書き文字のみで訓練したモデル A と比較すると，拡散モデルで生成した児童の文字画像も訓練データに含めたモデル D は，F1 スコアが 8.64 ポイント向上した．また，大人の文字に少量の児童の文字を加えることで F1 スコアは約 12 ポイント向上している (A → A') が，そこに拡散モデルによる生成結果を加えることで，さらに F1 スコアが約 7.5 ポイント向上している (A'→ B')．以上から，スタイル変換により生成された文字画像を訓練データに含めることで，OCR の精度が向上することを確認できた．

アラインメントの評価方法について説明する．お手本と視写結果の各文字を人手でアラインしたものを正解として，アルゴリズムの出力を評価した．具体的には，アラインメント結果において「文字の置換」によって対応づけられているお手本の文字と画像のペアに関する適合率と再現率から F1 値を計算した．置換による対応づけの内，お手本における空白文字と視写結果における空白画像のペアは，評価から除外している．

表 3: OCR の訓練データ

データ	サンプル数
ETL データベース	1,698,400
生成した児童の字 (文字単位拡散モデル)	85,250
生成した児童の字 (部品単位拡散モデル)	48,950
生成した児童の字 (GAN)	85,250
本物の児童の字	17,987

表 4: 児童の手書き文字に対する OCR 及び視写結果のアラインメント精度

Model	OCR		Alignment	Model	OCR		Alignment
	Accuracy	F1	F1		Accuracy	F1	F1
A	61.66	64.49	60.35	A'	78.52	76.11	60.96
B	69.21	71.62	61.96	B'	**86.29**	**83.64**	**67.36**
C	61.60	64.08	59.60	C'	84.31	81.03	66.36
D	**72.22**	**73.13**	**67.07**	D'	82.46	80.42	63.30
E	68.76	69.20	61.39	E'	82.39	77.57	64.88

表 5: 児童の手書き文字に対する OCR の文字種類別の F1 スコア

モデル	ひらがな	カタカナ	漢字	モデル	ひらがな	カタカナ	漢字
A	62.93	0.76	61.10	A'	80.25	43.85	74.50
B	69.59	30.25	70.24	B'	**86.95**	**65.53**	**85.08**
C	60.86	2.48	67.99	C'	85.63	45.57	81.83
D	**72.80**	**32.64**	**72.54**	D'	83.75	58.23	79.23
E	70.72	0.91	65.43	E'	84.53	50.23	76.26

表 4 より，児童の文字データを OCR 訓練に使った場合・使わない場合それぞれにおいて，生成した文字画像を訓練データに加えることで最も OCR 精度が向上したモデル (D および B') を用いた際のアラインメント精度が最も高いことが分かる．従って，生成文字画像による OCR の質向上は確かにアラインメント精度を高めたと言える．

4 おわりに

本研究では，視写課題の自動評価に向け，OCR の精度向上に取り組んだ．結果として，児童の手書き文字画像を生成し，OCR の訓練データに含めることで OCR 及びアラインメントの精度が向上したことが示された．アラインメントアルゴリズムにおけるコスト，特に「文字の挿入」と「文字の削除」のコストについては，パラメータの最適化により，より出力の精度を高めることが今後の課題である．

謝辞

本研究は JSPS 科研費 JP21H04416 の支援を受けたものである．

参考文献

[1] 江川克弘. 視写による作文学習の効果の検証 －大学院生を対象とした事例研究を通して－. 鳴門教育大学学校教育研究紀要, Vol. 32, pp. 209–217, 2018.

[2] Eunice N. Askov and Kasper N. Greff. Handwriting: Copying versus tracing as the most effective type of practice. **The Journal of Educational Research**, Vol. 69, No. 3, pp. 96–98, 1975.

[3] 中島由季子. 日本語学習における視写の効果. 国文目白, No. 59, pp. 16–32, 2020.

[4] Charlie Tsai. Recognizing handwritten japanese characters using deep convolutional neural networks. arXiv 2016.

[5] Tomoki Kitagawa, Chee Siang Leow, and Hiromitsu Nishizaki. Handwritten character generation using Y-autoencoder for character recognition model training. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 7344–7351, Marseille, France, June 2022. European Language Resources Association.

[6] Konstantina Nikolaidou, George Retsinas, Vincent Christlein, Mathias Seuret, Giorgos Sfikas, Elisa Barney Smith, Hamam Mokayed, and Marcus Liwicki. WordStylist: Styled Verbatim Handwritten Text Generation with Latent Diffusion Models. **arXiv preprint arXiv:2303.16576**, 2023.

[7] Ulrich Apel and Julien Quint. Building a graphetic dictionary for japanese kanji: Character look up based on brush strokes or stroke groups, and the display of kanji as path data. In **Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries**, ElectricDict '04, pp. 36–39, USA, 2004. Association for Computational Linguistics.

[8] Yuxin Kong, Canjie Luo, Weihong Ma, Qiyuan Zhu, Shenggao Zhu, Nicholas Yuan, and Lianwen Jin. Look closer to supervise better: One-shot font generation via component-based discriminator. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 13482–13491, June 2022.

[9] CHISE 漢字構造情報データベース,（2023-10 閲覧）. https://www.chise.org/ids/index.ja.html.

[10] 独立行政法人産業技術総合研究所. Etl 文字データベース,（2023-11 閲覧）.

[11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018.