

Caps : 深層学習を用いた書道作品における書体変換 —篆書の文字認識と行書の作品生成—

二ノ宮 梢平¹⁾ 藤代 一成²⁾(正会員)

1) 慶應義塾大学 大学院理工学研究科 2) 慶應義塾大学 理工学部

Caps: Script Conversion in Calligraphic Works Using Deep Learning —Recognition of seal script characters and generation of running script works—

Shohei Ninomiya¹⁾ Issei Fujishiro²⁾

1) Graduate School of Science and Technology, Keio University

2) Faculty of Science and Technology, Keio University

{shohei.ninomiya, fuji}@fj.ics.keio.ac.jp

概要

書道は日本の伝統的な文化である一方で、書道展は経験者の間で完結している。その要因の一つとして、鑑賞が書道の習熟度合いに大きく依存していることが挙げられる。そこで、本研究では書道初心者を対象にした鑑賞支援として、篆書の文字を認識し、行書の書体へと変換するシステム Caps (Calligraphy appreciation system) を開発してきた。篆書の文字認識には、畳込みニューラルネットワークを用いた手法を導入した。漢字が偏旁冠脚といったパーツから構成されるという特徴を利用して、書道作品に用いられる難読文字の認識を実現する。これにより、少ないデータ数と少ないクラス分類で、より性能の高い文字認識が可能になる。一方、行書体への変換には、敵対的生成ネットワークを用いた手法を導入した。文字骨格に対する筆跡の付与を学習したモデルを用いて、任意の漢字による書道作品画像を生成する。ここで、筆画情報と筆順情報をもった骨格画像を用いることで、効果的な筆跡の学習を実現した。評価実験を通じて、上記の二種類の手法を実装した Caps により、書道作品に用いられる漢字の特定から書道作品画像の生成まで、直観的な操作に基づく鑑賞体験が可能になることを確認した。

キーワード：文字認識、画像生成、深層学習、書道作品。

Abstract

Although calligraphy is a traditional Japanese culture, calligraphy exhibitions are being nearly monopolized by experienced calligraphers. It is primarily because appreciation depends largely on the level of calligraphic skills. In order to support the calligraphy beginners, we have developed in this study, a system called Caps (Calligraphy appreciation system) that provides seal script character recognition and conversion to running script. For recognizing seal script characters, a convolutional neural network-based method was introduced. Respecting the feature that kanji is composed of parts such as the radicals, unreadable characters used in calligraphic works can be recognized. This results in higher performance character recognition with small numbers of data and classifications. For converting to running script, a generative adversarial network-based method was introduced. A calligraphic work with arbitrary characters can be generated by a trained model that assigns calligraphic styles to the character skeleton. A skeleton image with stroke order and attribute information was used to achieve effective learning of calligraphic styles. It was empirically proven that Caps, which incorporates these two methods, enabled an intuitive appreciation experience ranging from recognizing the characters used in calligraphic works to generating calligraphic work in a different style.

Keywords: Character recognition, image generation, deep learning, calligraphic works.

1 序論

書道は日本の伝統的な文化である一方で、書道展は経験者の間で完結している。2018年に実施された日本書道ユネスコ登録推進協議会による調査 [1] では、書道展に会場した未就学児から高校生の 87.2% が小学生以下であり、72.2% が書道教室に通ったことがある事実が示された。この調査結果からも若い年代に書道が定着していないことがわかる。その要因の一つとして、鑑賞の際に作品から得られる情報量が書道の習熟度合いに依存し、初心者にとって鑑賞が困難であることが挙げられる。そこで、本論文では、書道に興味がある一方で、鑑賞の手立てを知らない初心者を対象にした鑑賞支援システム Caps (Calligraphy appreciation system) を提案する。

経験者と初心者の作品から得られる情報量の大きな違いとして、文字情報と書道表現情報の二種類が挙げられる。書道の作品に用いられる書体は、篆書、隸書、草書、行書、楷書の五種類 [2] に大別される。同じ漢字でも書体によって字形が異なり、識字には専門知識の習得が必要となる。実際、書道初心者だけでなく経験者でも、すべての書体を識字できる人は多くない。現状、読みや部首、画数などの漢字情報をもとに、ある書体での表現を調べることができる反面、漢字画像からその漢字の情報を導くことは困難である。また、書道表現情報として、書道作品には筆遣いや滲み、擦れ、空間の疎密など、書家の意匠が凝らされている。それらを汲み取ることも鑑賞の醍醐味である。しかし、書道表現への感覚は、古典を手本にして書く臨書や、古典の書風を模して書く倣書といった修練を積んで身につくものであり、経験者はそれらの知識や経験に基づいて作品を鑑賞する。例えば、現代の書家は作品制作において、古典から書風を抽出し、自身の作品に落とし込むことが多々あり、経験者はその古典の書風を想起しながら鑑賞することがある。すなわち、書家や古典の知識を有さない初心者にとって、作品のみから書道表現の情報を取得することは難しい。

以上より、文字情報と書道表現情報における習熟度依存が、初心者にとっての作品鑑賞を困難にする大きな要因であると考えられる。そこで、本研究では鑑賞支援として、文字認識と他の書体への変換を提供するシステム Caps を設計し、直観的な操作に基づく鑑賞体験を可能にする。文字認識には、畳込みニューラルネットワーク (CNN:

Convolutional Neural Network) を用いた手法を導入した。漢字をパーツの組合せと解釈しパーツごとに認識することで、漢字単位の学習に比べクラス数を削減する。ここでのパーツとは、偏旁冠脚といった漢字の構成要素をさす。また、本研究では、最も古い書体で字形が不安定な篆書を対象とすることで、本手法の有用性を示す。一方、他の書体への変換では、敵対的生成ネットワーク (GAN: Generative Adversarial Network) を用いた画像生成手法を導入した。文字骨格に書家の筆跡特徴を付与するモデルにより、任意の漢字による書道作品画像を生成する。これは、作品鑑賞において、経験や知識の差による比較材料の有無を補完すると同時に、倣書に通じる体験により、美的感覚や洞察力の向上を促進すると考えられる。本研究では、楷書に類似した字形でありながら、楷書に比べて流れるような筆致特徴をもつ行書を変換対象とした。なかでも、特に多くの現代書家に大きな影響を及ぼしており、作品鑑賞において重要な情報である王羲之 (303–361) と米芾 (1051–1107) の行書を選択した。

2 関連研究

本章では、書道文字を対象とした文字認識、画像生成、および描画システムの先行研究について説明し、本研究の位置づけを明確にする。

2.1 書道文字の画像認識

日本の古文書や古典籍に用いられるくずし字を対象とした光学的文字認識 (OCR: Optical Character Recognition) 手法として、北本ら [3] の研究が挙げられる。くずし字データセットを構築し、オープンデータを公開し、OCR アルゴリズムのコンテストを開催した。コンテスト上位者の手法では、CNN が用いられ、入力から出力まで一貫したくずし字認識を実現した。

漢字を対象とした研究として、X. Zhang ら [4] は、画像からストローク幅の平均、標準偏差、アスペクト比、重心、傾斜のバランスといったスタイルの情報を抽出することで、スタイルの定量的な尺度を構築し、線形分類器で書体を識別する手法を提案している。また、J. Zhang ら [5] は、深層学習により筆記スタイルの特徴を抽出し、サポートベクタマシンを用いて筆記スタイルの一貫性を検出する手法を提案している。この手法により、書体の分類に加え、書家の分類が実現した。

2.2 書道文字の画像生成

画像生成および画像変換において, Isola ら [6] が提案した Pix2Pix は強力な効果を発揮する手法であり, フォントの自動生成にも多く採り入れられている.

Gao ら [7] は複数のネットワークを組み合わせた AGIS-Net というフォント生成手法を提案している. 二種類のエンコーダで標準の文字形状とフォントスタイルの特徴をそれぞれ抽出し, 二種類の並列デコーダによりフォントの形状とテクスチャを復元する. また, フォントスタイル, テクスチャ, 局所的な特徴を対象とする三種類の識別器を設け, 高品質なフォントの生成を実現した.

Miao ら [8] は, Pix2Pix のエンコーダとデコーダの間に RDB (Residual Dense Block) を導入することで, 勾配消失を軽減し, フォントの特徴を効率的に伝播する手法を提案している. また, 損失関数として Wasserstein 距離を用いることでより安定した学習を可能にしている. また, 二段階の変換を施す手法として, Jiang ら [9] の SCFont が挙げられる. 第一段階で標準フォントの骨格から対象フォントの骨格に変形した後, その骨格に対して対象フォントのスタイルを肉付けすることで, 字形および筆跡の特徴を付与する.

2.3 書道文字の描画システム

書道表現を自在に描画することが困難な人を対象にした書道描画システムとして, 内平ら [10] のサンプリング書道が挙げられる. 書道作品の筆跡をサンプリングして描画に反映する機能により, ユーザは容易に多彩な表現の筆跡を描画することが可能になる.

2.4 本論文の位置づけ

2.1 節および 2.2 節より, 深層学習は書道文字に対する文字認識および画像生成に有効な技術であるといえる. しかし, 2.1 節で挙げた研究において, 認識の対象としている古文書などのくずし字や書道の書体は, クラスが少数かつ十分なデータが利用可能なのに対して, 書道作品の漢字の分類はクラスが膨大であり, さらに書跡が十分に現存しない. 2.2 節で紹介した画像生成手法は高品質なフォント生成を実現している. そこで, Jiang らの骨格への肉付け, および Gao らの特徴ごとのモデルの組合せを, 書道作品の筆跡特徴の学習に応用する. そして, 2.3 節で紹介したサンプリング書道が特定の筆跡を反映するのに対して, Caps の描画機能は GAN を用いることで, 書家の総合的な特徴である書風を反映する.

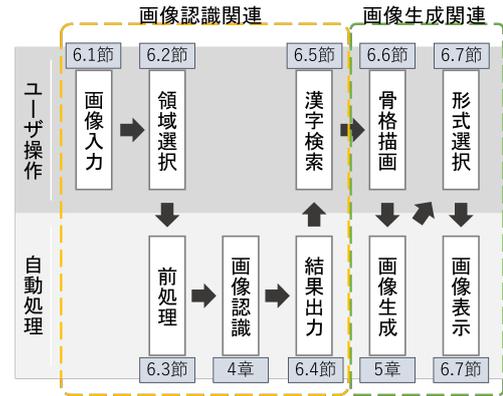


図 1: Caps における処理の流れと論文記載箇所

本論文では, CNN を用いた, 篆書作品のパーツ単位での画像認識と, GAN を用いた, 行書作品の画像生成を統合した鑑賞支援システム Caps を提案する. 書道文字の特定および書道表現の提示により, 書道作品の効果的な鑑賞の支援を目指す.

3 システム全体の流れ

漢字の特定および書道作品の表示までの処理の流れの全体像を図 1 に示す. 漢字の特定に至るまでの流れとして, ユーザは入力画像に対して画像認識の対象領域を選択するだけで, 画像認識に適した画像への変換や, 学習済みの画像認識モデルへの入力などは自動で処理され, パーツの存在確率が表示される. そして, 構成パーツによる検索機能で漢字を絞り込むことができる. その結果をふまえて, ユーザが任意の骨格を描画することで, 学習済みの画像生成モデルを介して, 書道表現が付与された作品画像を得る. また, 画像認識機能と画像生成機能はそれぞれ独立で利用することもできる.

4 画像認識

本章では, Caps の書道文字の特定に用いた画像認識モデルおよびモデル学習用に作成したデータセット [11, 12] について説明する.

4.1 モデル

画像認識を行うための学習アルゴリズムとして, CNN を用いた. また, 少ない学習データに対して, 画像認識の精度を上げるために, 学習済みモデルの VGG16 [13] をファインチューニングした. また, VGG16 モデルは出力層が 1,000 クラスで構成されているため, 本手法におけるクラス数に変更した. 入力データは VGG16 モデ

表 1: 漢字パーツのデータセットの構成

クラス	画像枚数
いと (へん)	624
おんな (へん)	304
き (へん)	567
くさかんむり	1,281
こころ	491
さんずい・みず	629
たけかんむり	602
てへん	345
しんじょう	254
合計	5,097

ルと同様に、 224×224 画素の画像とし、出力は画像に含まれる漢字パーツの存在確率とする。偏と旁、冠と脚といったパーツの組合せによって構成されるという事実は漢字の大きな特徴であり、これを利用することで少ないデータ数と少ないクラス分類でより性能の高い画像認識が可能になる。

4.2 データセット

本手法では、文字認識の対象とする書体を篆書に限定する。篆書が用いられていた時代は古く、同じ漢字でもパーツの配置が不規則であったり、パーツの形状が大きく変化したりと、文字として不完全な特徴がある。また、篆書は現存するサンプル数が少なく、文字の種類も現代よりはるかに少ない。以上の二点をふまえ、篆書はパーツごとに解釈する本手法の有用性を示すのに適しているといえる。そして、表 1 に示すデータセットを作成した。

4.3 モデルの学習

モデル学習において、損失関数に交差エントロピー誤差を、最適化手法に Momentum SGD を用いた。また、各層のパラメタの学習率は、入力に近い順に 16 個の畳込み層で 0.01%、2 個の全結合層で 0.05%、そして最後の全結合層で 0.1% とした。下位層の学習率を小さくし、上位層の学習率を大きく設定したのは、CNN の特性として下位層でエッジやプロブなどの汎用的な特徴を抽出し、上位層では学習データに特化した専用の特徴を抽出するからである。そして、学習のエポック数は複数回の試行に基づき 3 回が適切と判断した。

5 画像生成

本章では、書道表現提示のための画像生成アルゴリズムについて、作成したモデルのアーキテクチャおよび学

習用データセットについて述べる。本章で言及するモデルのアーキテクチャを図 2 に示す。

5.1 モデル

書道表現の提示として、骨格画像に対して書家の筆跡特徴を肉付けする画像変換を行う。そのために画像生成アルゴリズムとして Pix2Pix を用いた。また、骨格に肉付けを施す目的で生成器 $G1$ 、画像の品質およびサイズを大きくする目的で $G2$ をそれぞれ設けた。 $G1$ の入力データは 128×128 画素の骨格画像で、出力データは骨格画像に書家の筆跡特徴を肉付けした 128×128 画素の画像となる。また、筆画情報と筆順情報を含む二種類の骨格画像を用意し、それぞれに対する $G1$ を学習する。そして、 $G1$ の出力データを結合して $G2$ の入力とし、 256×256 画素の書道画像を出力する。

5.2 データセット

前節のモデル学習に、骨格画像と書道画像のペアデータセットを用いた。書道画像として、王羲之と米芾の行書各 500 文字の画像を用意し、骨格画像を手動で作成した。また、書道画像に影響する骨格の情報として、筆順と筆画が挙げられる。例えば、筆順からは連続する筆跡間の関係性、筆画からは筆遣いの違いが反映されると考えられる。そこで、筆順情報と筆画情報を線の色に対応させた二種類の骨格画像を作成した。各対応関係を表 2 に示す。ここで、筆順の列の数字は、漢字の画数を表し、赤色から桃色に至る着色を繰り返した。また、筆画の分類は永字八法をもとに、行書に合わせて右上方向の線を横方向の線と統一し、行書特有の連綿の分類を追加した。色は色相環における等間隔の八色を用いた。筆順にも八色を用いたことで、前後の筆順情報に加え、長期的な墨量の変化を特徴付けることができると考える。

表 2: 骨格情報と色の関係

色 (r, g, b)	筆順	筆画
赤色 (255, 255, 255)	1, 9, ...	左払い
黄色 (255, 223, 128)	2, 10, ...	横
黄緑色 (192, 255, 128)	3, 11, ...	点
緑色 (128, 255, 160)	4, 12, ...	左下
水色 (128, 255, 255)	5, 13, ...	はね
青色 (128, 160, 255)	6, 14, ...	連綿
紫色 (192, 128, 255)	7, 15, ...	右払い
桃色 (255, 128, 223)	8, 16, ...	縦

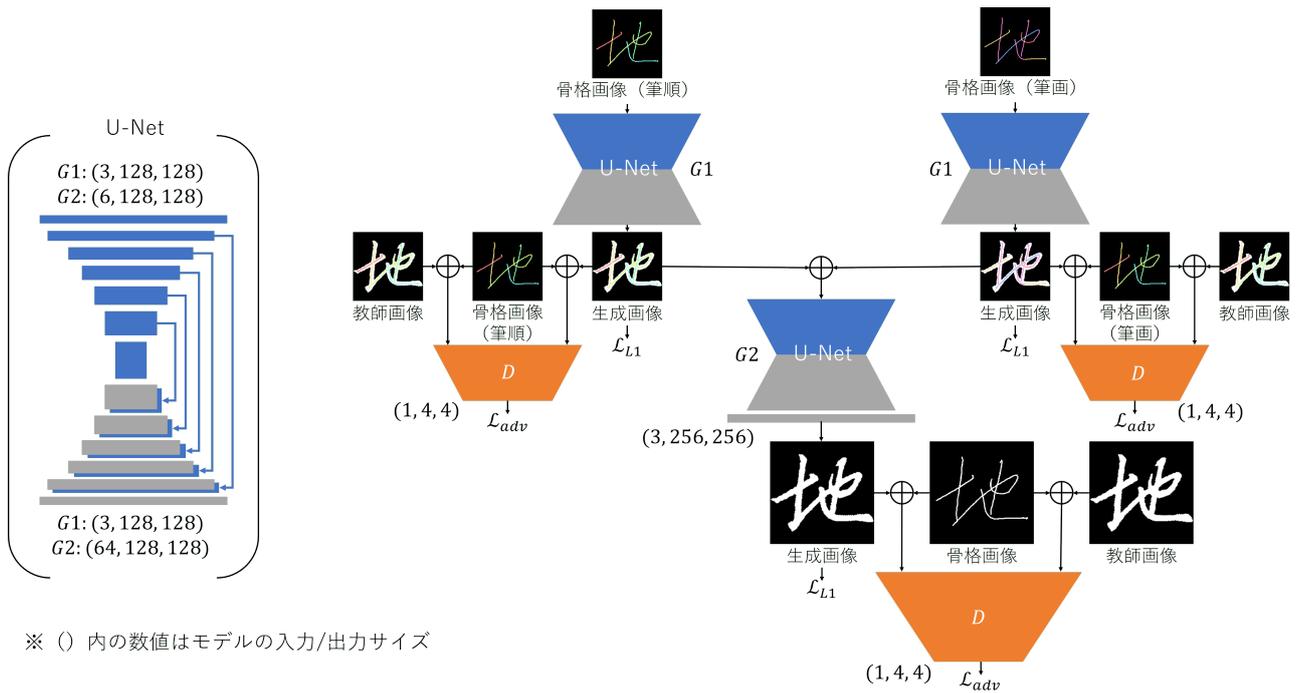


図 2: 筆跡特徴を付与する画像生成モデルのアーキテクチャ

5.3 モデルの学習

Pix2Pix の学習における目的関数は以下の式で表される。

$$G^* = \arg \min_G \max_D \mathcal{L}_{adv}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (1)$$

ここで、 $\mathcal{L}_{adv}(G, D)$ は敵対的損失、 $\mathcal{L}_{L1}(G)$ は L1 損失であり、それぞれ以下の式で表される。

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_x [\log(1 - D(x, G(x)))], \quad (2)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y} [\|y - G(x)\|_1], \quad (3)$$

ここで、 $G(x)$ は入力画像 x を入力とした生成器の生成画像、 $D(x, y)$ は入力画像 x と教師画像 y による識別器の出力値、 $D(x, G(x))$ は入力画像 x と生成画像 $G(x)$ による識別器の出力値を表す。

モデル学習において、式 (2) の実現のため、二値交差エントロピー誤差を、式 (3) には平均絶対誤差を用いた。また、最適化手法として Adam を用いた。そして、学習のエポック数は複数回の試行に基づき $G1$ 、 $G2$ ともに 50 回が適切と判断した。 $G1$ を学習した後、学習済みの $G1$ から得られた出力画像を用いて $G2$ を学習した。

さらに、筆順と筆画の影響を観察するために、両者の情報を入力としたモデルに加え、それぞれ一方を入力として学習したモデルの計三種類を作成した。

6 システム

Caps は、図 3 に示すような GUI を介して、ユーザに直観的な操作性を提供する。設計に用いたプログラミング言語は Python3.7.6 で、GUI ライブラリには kivy1.11.1 を使用した。本章では、前述の画像認識モデルおよび画像生成モデルを実装した Caps における処理全体の流れを順に説明する。

6.1 画像入力

Caps は、ローカルフォルダ内の任意の画像を入力とする。図 3(A)の「画像を選択」ボタンからファイル選択ができる。



図 3: Caps の GUI の画面構成



図 8: 骨格の描画

6.6 骨格の描画

Caps では、ユーザが任意の骨格を描画し、画像生成モデルへの入力とすることで、その骨格に書家の筆跡特徴が付与された画像が提示される。ユーザは図 3(D)において、図 8 のように八色のパレットから筆画に応じた色を選択し描画する。ここで、ユーザ自身が骨格を描画することにより、骨格情報の試行錯誤を通して、長期的な鑑賞支援としての美的感覚や洞察力の向上が期待できる。同時に、一画ごとに「線を保存」ボタンを押すことで、表 2 に示す配色で筆順情報を保存する。また、領域選択と同様、「描画を戻る」ボタンで 1 回分の描画の取消しが可能で、「描画を消去」ボタンで全描画の消去ができる。

6.7 作品の形式選択と表示

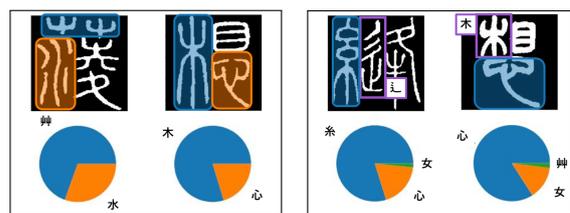
書道の画仙紙の一般的なサイズとして、半紙サイズ(縦横比約 334:243) や半切サイズ(縦横比約 272:69) が挙げられる。そこで、Caps では 1 文字での表示に加え、半紙に 2 行 6 文字と半切に 2 行 16 文字の表示形式を設けた。画像生成モデルおよび表示形式を選択することで作品表現を観察することができる(図 3(E))。

7 結果と評価

本章では、Caps に実装した画像認識モデルと画像生成モデルの性能を評価する。学習の実行環境は Google Colaboratory (GPU: NVIDIA Tesla K80, CPU: Intel Xeon Processors @ 2.30GHz) で、深層学習フレームワークには PyTorch1.8.1+cu101 を用いた。

7.1 画像認識の性能

4 章のモデルについて、データ拡張以前の画像を学習用と分けて検証したところ、97.6% の正答率を示した。検証では存在確率が最大のもののみを比較した。一枚の画像に対して一個のラベルを付与したデータセットを構築したため、複数のパーツを含む漢字に対して、付与されたラベルと異なるパーツに大きい確率を表す場合があ



(a) 成功例

(b) 失敗例

図 9: 二種類のパーツを含む画像の画像認識の例



図 10: 形状が異なるパーツの画像認識の例

る。そこで、Caps では上位の三候補を提示する。図 9 に二つのパーツを含む漢字の画像認識例を示す。図 9(a) のように両パーツとも検出される場合と、図 9(b) のように片方のみ認識され、他方は上位三候補に現れない場合がある。これは、一対一対応のラベル付けデータを入力としたことに起因すると考える。

また、パーツの変形や移動に対する画像認識結果として、図 10 に「しんにょう」の例を示す。篆書のしんにょうは、偏のように漢字の左に位置する場合や、一部が分離し漢字の下に潜り込む場合があるが、両者とも高い確率で検出されている。これは、手法に CNN を用いたことの大きな便益であるといえる。



図 11: 骨格の組合せによる生成画像の比較と書家の比較



(a) ニノ宮揮毫

(b) 王義之

(c) 米芾

図 12: 複数文字の画像生成例。半紙に「芸術科学会」を描画

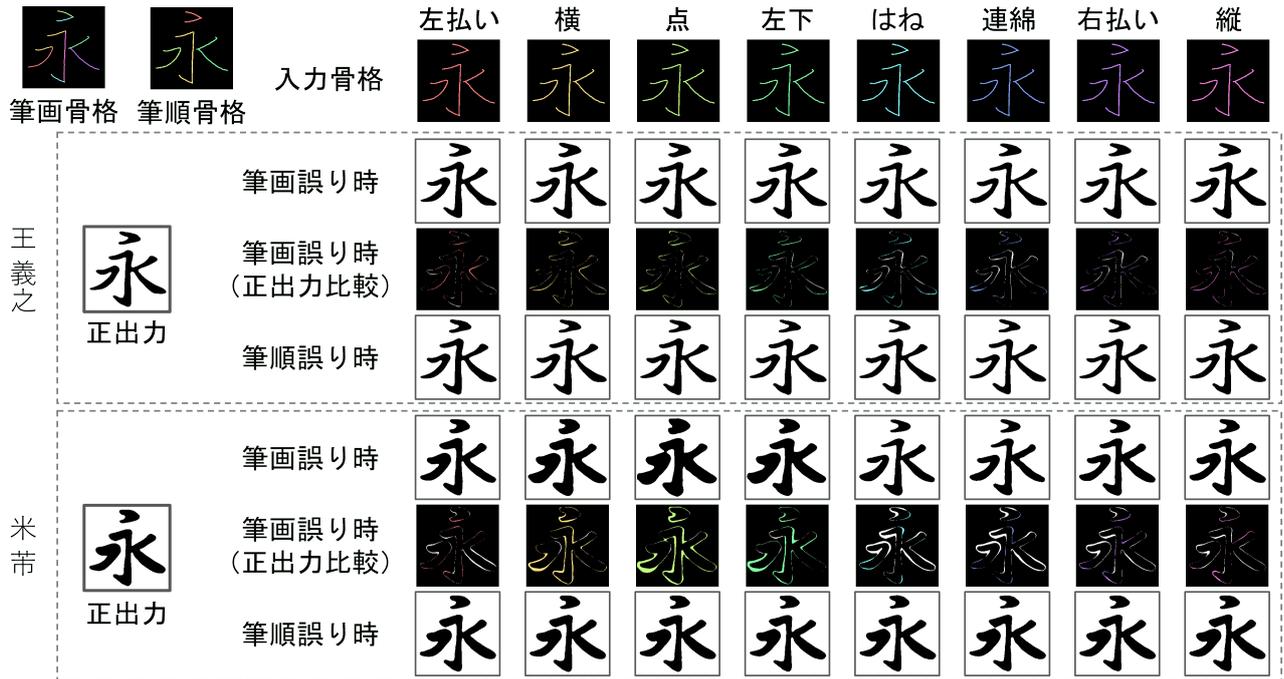


図 13: 骨格の描画を誤った際の出力例

7.2 生成画像の品質

5章で言及した三種類のモデルによる画像生成結果を図 11 に示す。学習用とは別に用意した検証用の骨格画像を入力とした。モデルの種類による比較として、筆画情報と筆順情報の両方を入力としたモデルと、筆画情報のみを入力としたモデルは線の特徴が現れているのに対して、筆順情報のみを入力とした場合は線に揺らぎが生じたり、筆跡特徴が減衰したりしている。また、Caps に導入した筆画情報と筆順情報を組み合わせるモデルが、筆画のみのモデルに対して優位な点は観察されなかった。これは、利用したデータセットが、筆順によって現れる特徴を十分に含んでいなかったためであると考えられる。例えば、墨量の減少に伴うかすれ表現などは筆順に関係している。また、書家による比較として、王義之はゆったりのびやかで流れるような筆致が反映され、米芾はたっぷりしなやかで抑揚に富んだ筆致が反映されている。これらの特徴は、ユーザが骨格を入力した場合の出力結果の図 12 にも観察される。しかし、第一著者（二ノ宮）が揮毫した図 12(a) に対して、生成画像は筆画ごとに表現が類似しており、入筆の角度や筆の状態、墨量による線質の変化に多様性が乏しい。これは、GAN のモデルが学習する過程で平均的な特徴を大きく学習し、外れた特徴が欠落することに起因すると考えられる。

また、入力骨格描画で色の選択を誤った場合の出力例を図 13 に示す。ここで、筆画筆順ともに正しい場合の出力を「正出力」とし、筆画を誤った場合と筆順を誤った場合の出力を羅列している。また、正出力比較は、正出力と誤り時の出力を重ね合わせて一致した部分は黒色、正出力のみの部分は白色、誤り時の出力のみの部分はその骨格の色で着色した画像で表している。図 13 において、筆順情報を誤った場合の出力に大きな違いはみられない。これは前述の通り、今回用いたデータセットが、筆順情報に依存する特徴が少ないためであると考えられる。筆画情報を誤った場合は、その色に依存する特徴が出力に現れた。例えば、黄緑色（点）による誤り時は、全体にわたって正出力を膨張させたような出力になり、水色（はね）による誤り時は、端に向けて鋭くなる線が多く見られる。また、左払いやはねを誤った場合は、正出力に比べて線の始まりが細く、終わりが太くなっている。すなわち、骨格に筆画情報をもたせることで、線の特徴を区別した学習ができており、特に、類似した骨格である左払いと左下や、短い線と点、はねなどをユーザの意図通りに出力に反映するには、筆画情報の区別が必要不可欠であるといえる。

表 3: 被験者情報と Caps 使用時の篆書 5 文字の識字時間

被験者	性別	年齢 (歳)	書道歴 (年)						仮名	書道展に行く頻度	識字時間
			篆書	隸書	草書	行書	楷書	行書			
A	男	20	0	0	0	0	0	2	減多に行かない	2分 25秒	
B	男	22	0	0	0	0	0	3	減多に行かない	3分 56秒	
C	男	23	0	0	0	0	1	3	減多に行かない	3分 54秒	
D	男	22	0	3	0	0	0.5	0	年間 20 回程度	2分 47秒	
E	男	22	0	0	3	3	6	0	年間 3 回程度	2分 21秒	
F	女	21	0	0	0.5	8	10	0	年間 3 回程度	2分 45秒	
G	女	21	0	1	2	10	16	7	減多に行かない	3分 27秒	
H	男	21	1	0	0	0	3	0	減多に行かない	2文字可読	
I	男	21	0.5	0.5	0.5	0.5	2.5	1	年間 5 回程度	4文字可読	
J	男	24	2	12	12	12	18	1	年間 20 回程度	全文字可読	

7.3 Caps の有用性

Caps の有用性を評価するために、慶應義塾大学の書道サークルに所属する初心者および経験者 10 名 (20 歳から 24 歳までの女性 2 名と男性 8 名) A から J にインタビューを行った。篆書を読める、あるいは経験から調べられる 3 名 (H, I, J) を除く 7 名に、Caps の画像認識を実際に利用してもらい、篆書 5 文字を特定するまでの時間を計測した。また、全員に Caps の操作性と満足度をインタビューした。被験者の基本情報および篆書 5 文字の識字時間を表 3 に示す。篆書を読むことができない被験者も全員、Caps を用いることで 5 文字全ての識字ができた。

7.3.1 画像認識の操作性

操作性について次のような感想が挙げられた。なお、感想末尾の括弧内に発言者を示す。

- 操作が簡単でわかりやすい (B, C, E, F, I, J)
- パーツ検索が使いやすかった (A, C, F)
- ボタンを押す操作が多い (A, G, H)

「操作が簡単」という意見はほぼ全員から得られ、2, 3 文字目からは操作に慣れ始めている様子だった。また、全てのパーツを認識できない場合は、検出されたパーツをもとにパーツ検索機能で補完できていた。しかし、領域選択におけるボタン入力に煩わしさを示す意見も挙げられた。システムの改善が必要である。

7.3.2 画像認識の満足度

満足度としては次のような感想が得られた。

- 勉強になる (A, G)
- 自力で読むときと同じ感覚 (H, J)

- 書道展で文字が読めないときに使いたい (E, I)
- 対応パーツが少ない (C, F, H, J)
- いっぺんに認識して欲しい (A, J)

「勉強になる」という意見からも、自身でパーツを確認しながら絞り込んでいくことで識字能力の向上が見込めると考えられる。また、篆書経験者自身も Caps と同じようにパーツごとに読み解いていくという意見が得られ、Caps が合理的であり、初心者に適切な識字体験を提供しているといえる。また、展覧会に行く被験者から、識字に慣れた経験に基づいて Caps の利便性を好評する感想も挙げられた。一方で、全員が認識されないパーツがあることに対する不満や戸惑いを示した。また、全ての文字を一度に認識できることを期待する意見も得られた。今回の計測結果から、例えば漢詩 40 文字の作品を Caps で認識する場合、20 分以上の時間を要することが予想される。これらは重要な課題であり、8.1 節で詳しく言及する。

7.3.3 画像生成の操作性

骨格入力の操作性について次のような感想が挙げられた。

- おもしろかった (D, E)
- 筆順を保存し忘れる (A, C, H)
- ひと息で書きたかった (A, F, G, I)

特徴を考えながら描画する作業に対して、ゲーム感覚で楽しいと感じる人と、煩わしいと感じる人がいた。また、煩わしさを示した人の多くに筆順の保存し忘れが度々観察された。さらに、行書経験者の多くが線を分けて書くことに違和感を示した。このように、現状の骨格入力に

は多くの改善点があることが確認された。この点に関しても、今後の課題として8.2節で言及する。

7.3.4 画像生成の満足度

生成結果の満足度については次のような感想が得られた。

- 線を書いたら書道文字が現れるのが良い (A, C, H)
- 中国の書家になった気分 (F)
- 字形が一緒でも線質で雰囲気が変わった (G)
- 作品づくりのイメージがしやすくなる (I, J)
- 太い線と細い線の変化が乏しい (G, H, J)
- もっと書道らしさが欲しい (D, E, I)
- 初心者最初の勉強には役立ちそう (G, I)

行書の経験が浅い人は、線の入力だけで書道文字が生成されることにおもしろさと手軽さを感じていた。また、作品や生成画像を単体で観察するより、それらを比較した場合の方が多くの感想を述べることができた。Capsの目的である初心者に対する比較材料としての役割を果たしているといえる。また、経験者からは、生成画像の品質の向上を求める意見が挙げられた一方で、現状でも初心者の入門には役立ちそうという意見もあった。さらに、経験者は、鑑賞だけでなく作品制作における有用性も感じていた。作品制作では、作品に用いる文字を古典から集める集字の工程があるが、過去の書家書いていない文字も多くあり、そのような場合にもCapsの画像生成が有用であると考えられる。

8 結論と今後の課題

本論文では、文字画像をパーツごとに画像認識する手法と、文字骨格に対して書体の特徴を肉付けした画像の生成手法を提案した。画像認識では、一漢字ークラス対応の文字認識に対して、少ないデータ数とクラス数で、より汎用的な画像認識の処理環境が実現した。画像生成では、筆順および筆画情報を付与した骨格画像を入力して学習したモデルによって、任意の骨格に対する書家の筆跡特徴の付与が可能になった。そして、提案手法を実装したCapsについて、書道作品に用いられる漢字の特定から書道作品画像の生成までの一連の流れを示した。

本手法について、今後取り組むべき課題を以下に提示する。

8.1 複数文字への適用

書道作品は一枚の紙面に複数の文字が書かれていることが多い。よって、画像認識手法として、作品全体の画像を対象に、そこに含まれる複数のパーツを識別することが望ましい。そこで、R-CNN[15]などの物体検出手法を採り入れることを検討する。

8.2 骨格入力におけるユーザの負担軽減

現状の骨格入力はユーザへの負担が少なくない。解決策として、骨格のデータセットの作成や骨格の自動抽出機能の実装、筆画および筆順情報の半自動抽出機能の実装などを検討する。

8.3 書道表現の拡張

本手法では、画像生成における書道表現として、書家の筆跡特徴の肉付けを実装した。しかし、書家の特徴は漢字の字形にも現れる。よって、ユーザの描画した骨格に対して、肉付けをする前に字形の変形を施す手法を検討する必要がある。また、書道作品の表現として、一文字に対する表現だけでなく、紙面上での文字同士の関係性も挙げられる。例えば、滲みを含む文字の近くにかすれた文字を配置することで、立体感や奥行き感を演出する表現がある。大局的な書道表現の反映を実装することで、より多様な鑑賞材料を提供し、能動的な鑑賞を促進できると考える。

8.4 画像生成手法の改良

Capsの機能の一つに、書道作品から筆跡特徴を抽出し、任意の漢字に付与することを想定している。この機能は、初心者に今までにない鑑賞の楽しみを提供するだけでなく、経験者の作品制作支援としても有用であると考えられる。そこで、作品に含まれる少数の文字画像を入力とし、その筆跡特徴を抽出、付与できる手法を検討する。

8.5 データセットの拡張

現時点では、画像認識のクラス数が9個であり、限られた文字に対してのみ本手法が適用される。Capsを実用的なものにするためにはデータセットの拡張が必須である。また、画像生成では、8.3節および前節の課題に取り組むうえで、様々な種類の多量な学習データが必要となる。例えば、現時点で文字ごとのデータセットを作成しているが、大局的な特徴を扱うために、作品として複数文字で構成された画像を利用することが考えられる。

謝辞

本研究を進めるにあたり、深層学習の技術面での助言を頂いた藤代研究室の中山 雅紀氏と宮澤 篤氏に深く感謝する。本論文の一部は、科研費基盤研究(A)17H00737および21H04916の支援により実施された。

参考文献

- [1] 日本書道ユネスコ登録推進協議会調査委員会. 書道文化に関する基礎調査報告書. <http://www.shodoisan.jp/tk.html>. 参照: 2021-10-14.
- [2] 日本書道協会. 日常に生きる書体. https://www.u-can.co.jp/shodo/hand_write/study/in_the_life/. 参照: 2021-10-14.
- [3] 北本朝展, カラーヌワットタリン, 宮崎智, 山本和明. 文字データの分析-機械学習によるくずし字認識の可能性とそのインパクト-. 電子情報通信学会誌, Vol. 102, No. 6, pp. 563-568, 2019.
- [4] X. Zhang and G. Nagy. Computational method for calligraphic style representation and classification. *Journal of Electronic Imaging*, Vol. 24, No. 5, pp. 1-13, 2015.
- [5] J. Zhang, L. Guo, S. Yang, X. Sun, and X. Li. Detecting chinese calligraphy style consistency by deep learning and one-class SVM. In *Proceedings of the 2017 2nd International Conference on Image, Vision and Computing*, pp. 83-86, 2017.
- [6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125-1134, 2017.
- [7] Y. Gao, Y. Guo, Z. Lian, Y. Tang, and J. Xiao. Artistic glyph image synthesis via one-stage few-shot learning. *ACM Transactions on Graphics*, Vol. 38, No. 6, pp. 185:1-185:12, 2019.
- [8] Y. Miao, H. Jia, K. Tang, and W. Cheng. Chinese calligraphy generation based on residual dense network. In *Proceedings of the 2019 4th International Conference on Intelligent Information Processing*, pp. 508-512, 2019.
- [9] Y. Jiang, Z. Lian, Y. Tang, and J. Xiao. SCFont: Structure-guided chinese font generation via deep stacked networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 4015-4022, 2019.
- [10] 内平博貴, 宮下芳明. サンプリング書道: サンプラーのメタファを取り入れた書道による描画・閲覧手法の提案. 芸術科学会論文誌, Vol. 9, No. 1, pp. 10 - 19, 2010.
- [11] ニノ宮梢平, 中山雅紀, 宮澤篤, 藤代一成. 深層学習を用いた書道作品における書体翻訳. 情報処理学会第82回全国大会講演論文集, Vol. 4, pp. 135-136, 2020.
- [12] ニノ宮梢平, 中山雅紀, 宮澤篤, 藤代一成. 深層学習を用いた書道作品における書体翻訳に向けて-篆書の文字認識-. 映情学技報, Vol. 44, No. 10, pp. 75-78, 2020.
- [13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [14] Y. Okuda. kanjivg-radical. <https://github.com/yagays/kanjivg-radical/>. 参照: 2021-10-14.
- [15] Z. Cai and N. Vasconcelos. Cascade R-CNN: Delving into high quality object detection. In *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6154-6162, 2018.

二ノ宮 梢平



2020年慶應義塾大学理工学部情報工学科卒業。現在、同大学大学院理工学研究科開放環境科学専攻情報工学専修修士課程に在籍。書道を対象とした鑑賞支援システムCapsの研究開発に従事。映像情報メディア学会学生会員。

藤代 一成 (正会員)



東京大学，筑波大学，お茶の水女子大学，東北大学を経て，2009年より慶應義塾大学理工学部情報工学科教授。1988年理学博士（東京大学）。CG・可視化や知的環境メディアに関する研究に従事。第16回CG Japan Award受賞。本会では副会長，評議員，NICOGRAPH

International アドバイザリ委員等を歴任。日本工学会，情報処理学会フェロー，画像電子学会名誉会員，IEEE，ACM シニア会員。IEEE Visualization Academy 会員。現在，日本学術会議連携会員，可視化情報学会会長。