古代文字フォントの画像データに基づく手書き篆文文字の検索支援

李　康穎（立命館大学 情報理工学研究科）
Batjargal Biligsaikhan（立命館大学 総合科学技術研究機構）
前田 亮（立命館大学 情報理工学部）

篆文は今でも広く使用されている古代の書体である．しかしその字形は抽象性が高いため，専門的な知識を持たない人にとって篆文字形を読むのは非常に難しい．また，現存の篆文の記載に使用する文献資源が極めて少ない状況であり，利用できる篆文の画像データも不足しているため，篆文に基づく OCR システムもほとんど存在しない．本研究は小篆の篆文書体を研究対象とし，白川静の漢字研究の成果に基づく「白川フォント」中の小篆書体の文字と「説文解字フォント」中の白川フォントと対応する文字を Generative Adversarial Network に基づく zi2zi モデルの入力データとして利用し，生成された画像の変形処理を行う．処理した後の画像ファイルを訓練データとして使用し，文字認識の実験を行った．本論文では，生成した画像を手書き篆文の認識モデルの訓練データとして用いる手法を提案し，今後の手書き篆文文字の検索支援の実現に向けた基礎を構築することを目指す．

Retrieval of handwritten characters of the Seal script using the image data of ancient character typeface

Kangying Li (Graduate School of Information Science and Engineering, Ritsumeikan University)
Biligsaikhan Batjargal (Research Organization of Science and Technology, Ritsumeikan University)
Akira Maeda (College of Information Science and Engineering, Ritsumeikan University)

The Seal script is a one of the widely used ancient Chinese typeface in nowadays. However, most people today cannot read the Seal script unless they have expert knowledge to read the glyph form. Moreover, there is virtually no OCR system for Seal script due to the limited availability of historical documents in the Seal script and a small number of image dataset. In this paper, we propose a method to employ generated image data as training data in a character recognition model of handwritten Seal script, and aim to construct a retrieval system to support further search of handwritten characters of the Seal script. This research focuses on typeface of the Seal script and, we build augmented image data by inputting 1) the characters of the Seal script in Shirakawa font, a typeface based on the results of Emeritus Shirakawa Shizuka's ancient Chinese characters' research, and 2) the corresponding characters in Shuowen Jiezi font; into the zi2zi, a Conditional Adversarial Networks model. We conducted experiments in the recognition of handwritten characters by using the generated images as training data.

１．まえがき

漢字は世界屈指の古代文字であり，現在も日本や中国など東アジアの広い範囲で日常的に使われている．既存の古代文字のスキャン画像を閲覧可能な検索サービス「木簡画像データベース・木簡字典」および「電子くずし字字典データベース」連携検索システム[1]では，飛鳥・奈良時代から江戸時代に至るまでの文字を認識し，解読することができる．日本古典籍字形データセット[2]では，国文学研究資料館所蔵古典籍に書かれた 86,176 文字が公開され，機械学習用データなどへの活用が可能である．しかし，さらに時代が古い甲骨文，金文，篆文などで書かれた古典籍の字形データは少なく，それらの古代文字を認識し，解読できるシステムの構築が期待される．

図１に示すように，象形文字である金文から更に字形の整理が進み，後世の漢字のように部首へ

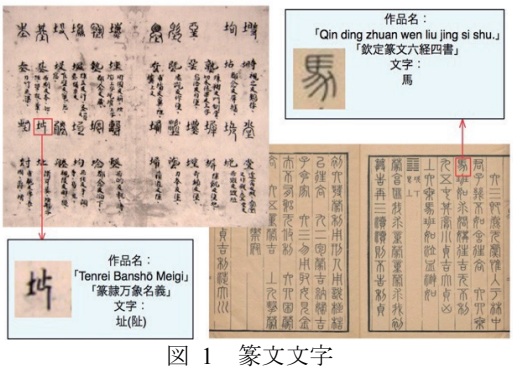

図 1　篆文文字
Figure 1　The seal script

の分割が容易なのが特徴である篆文が生まれ，これは現在でも印章などに用いられることが多い．篆文は古代文字の中では最も息が長い書体であり，専門知識を持たない一般人にはその字形を読み解くのは難しい．日本の古代著作『篆隷万象名義』は題名の中に「篆」という文字が含まれるが，しかし本文の中ではほぼ楷書で書かれており，存在する篆文の数が少ない．中国の古代著作『欽定篆文六経四書』は篆文で書かれているが，しかし楷書体の注釈が無く，これは現代人にとって閲読するのが非常に難しい．篆文に基づく文献の文字認識システムを実現できれば，現代人と古代文字の文化の間の距離を短縮することができる．

既存の多数の篆書体フォントは美観の追求に主眼が置かれており，「造字」が頻繁に行なわれているため，これらと歴史の上の篆文字形は極めて大きな相違が存在する．したがって，これらのフォントは学術研究には不向きである．そのため本研究では，使用するフォントに対して選別を行い，古代著作を参照し作成したフォントを実験データに用いる．図 2 に示すように，立命館大学白川静記念東洋文字文化研究所の研究プロジェクトでは，篆文フォント 2,590 字を含む古代文字フォント「白川フォント」およびその検索システム[3]を開発し，公開している．

図 2 白川フォント検索システム
Figure 2　Search system for Shirakawa characters

手書きの楷書，草書などと比べ，篆書体は，字の書き方に個人のこだわりが見受けられる．しかしながら，同じ書体を書くのに，書く者個人のスタイルや特徴に差はそれほど大きくないため，本研究は，篆書体の認識用データの自動生成が可能という仮説を立て，そしてその効果を検証する実験を行った，その流れを図 3 に示す．

「白川フォント」の検索システムで公開されている篆文フォントデータと「說文解字 True Type 字型」[15]を用いて手書き篆文の文字認識システムの構築を試みる．

本研究の成果は，古代文字に関する学術研究に有用であるだけでなく，たとえば小学生の漢字教育に適用することができ，あるいは古代文字の愛好者や書道学者，専門知識を持たない一般人に直観的な文字読解の支援を提供する．

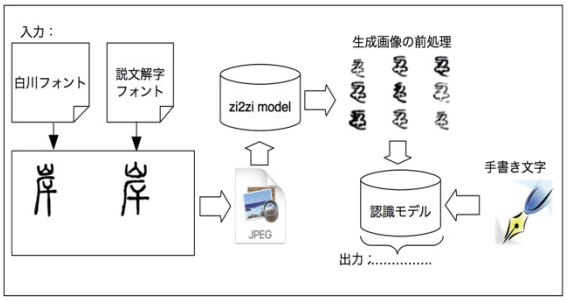

図 3 提案手法の流れ
Figure 3　Flow chart of proposed method

2．関連研究

2.1 文字認識

文字認識の手法は，画像に含まれる文章や手書き文字を認識し，テキストを抽出するだけでなく，その内容に関連する情報をユーザにフィードバックすることができる．例えば，Luo らによる研究[4]では，文字の構造特徴を抽出する手法を用い，中国の手書き漢字を認識する実験を行った．Deepa らは，Zernike モーメントと対角線の特徴を利用し，タミルの手書き文字を認識する手法[5]を提案した．古代文字を対象として文字認識を行う研究もある．石井ら[6]は，テンプレートマッチングによって甲骨文字の認識を行い，認識のためのデータベースを構築した．そして，特徴量の分析，直線の抽出などの手法を用いた認識実験を行った．

2.2 手書き風文字の生成

近年，手書き風文字の生成に関する研究が注目されている．Graves は，認識実験のための手書き文字生成器を作成した．脳神経回路を模したニューラルネットワーク（NN）を用いて「手書き風」画像を生成する手法は多くみられるが[7]，結果画像がペンで書かれたイメージであり，古文書などにあるような毛筆で書いた文字とはスタイルが異なる．毛筆の字に変換できる手法として，Tian によるプロジェクト[8]において，与えられた訓練データから生成モデルを推定し，2 つのニューラルネットワークを競い合わせることで画像生成をさせるアルゴリズム GAN（Generative Adversarial Network）により構築した zi2zi モデルがある．Chang らは新しいフォントを綺麗に生成できるモデル[9]を提案した．

2.3 GAN の生成データを用いた認識実験

Ghosh ら[10]は MNIST 手書き数字のデータセットを生成する実験を行い，その結果により GAN の文字認識領域においての有効性を示した．渡部ら[11]は，訓練データを増やして猫の画像を研究対象にし，DCGAN（Deep Convolutional GAN）で訓練データを生成した．その中で，選別された品質の比較的高いデータに対してデータ拡張操作を施し，またこれらのデータを使って，猫の品種分類実験を行った．Creswell ら[12]は，Merchant

Marks 画像に対して DCGAN を使ってデータ拡張を適用する研究を行い，これらのデータを使って Merchant Marks のデータベースに対する検索支援を提供した．Tran ら[13]は，撮影角度変換を施した人の画像の生成や人の顔認識におけるデータ生成支援の提供に GAN を使用した．

3．提案手法

近年，文字認識に幅広く応用されている CNN（Convolutional Neural Network）を使用し，手書き篆文文字の認識を行う．CNN では，画像の局所的な特徴を抽出できる畳み込み層と局所的な特徴をまとめあげる処理をするプーリング層を用い，画像が持つ情報量を大幅に圧縮して抽象化する．さらに抽象化された画像を用い，画像を自動分類する手法である．

本研究で提案する手法は，(1) フォントデータから「手書きスタイル」文字画像の自動生成，(2) 生成した画像データの前処理，(3) 入力手書き画像の前処理，(4) 前処理をした画像を用いた篆文手書き画像の認識の四つのステップによって構成される．

3.1 手書きスタイル画像の取得

Tian らが提案した zii2zi モデル[8]の構造は図 4 の通りである．

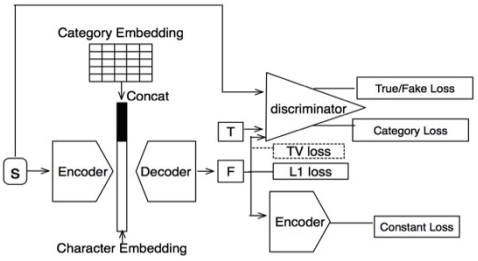

図 4　zi2zi モデル
Figure 4　The zi2zi model

このモデルは画像から画像を生成するモデル pix2pix[14]を基に，フォント生成モデルとして最適化したものである．Google のゼロショット GNMT（Google's Neural Machine Translation）を基に，カテゴリ埋め込みは，訓練できないガウスノイズによる文字埋込みにスタイル埋め込みを連結することで，同時に複数のフォントスタイルの学習が実現できる．エンコーダ層は同じ文字を同じベクトルにマップする．デコーダ層は 2 種類の文字とスタイル埋め込みを使用してターゲット文字を生成する．事前に訓練されたスタイル埋め込みはすでにエンコーダ層に固定されたので，ソースフォント「S」を入力する．ターゲットフォント「T」は識別器の教師データとして利用する．本研究はスタイルが混合される過程で形成された画像の「スタイル不安定性」という特徴を捉える．この特徴は人が標準字をまねて文字を書く時，書いた文字のスタイルに不安定性があることと類似していると考え，一部の画像を抽出し，次のステップの処理サンプルとした．そしてこの一部のサンプルを使って，単独で認識実験を行った．

「白川フォント」の篆文 2,590 字を利用し，これらを画像形式に変換する．これらの篆文画像を生成モデルの生成ターゲット画像とする．台湾における「CNS11643 中文標準交換碼全字庫」（Master Ideographs Seeker for CNS 11643 Chinese Standard Interchange Code）[15]では，最古の部首別漢字字典「説文解字」に基づく Unicode で作成した篆文 6,721 字を含むオープンソースのフォント「說文解字 True Type 字型」を公開している．研究などの目的であれば無料で使えるため，この篆文フォントを利用し，画像形式に変換し，ソース画像として，モデルに入力する．取得された画像データの「忍」を例として図 5 に示す．

図 5　フォントから変換された篆文画像
Figure 5　A Seal script character extracted from a font

左側は白川フォントから生成されたターゲット画像，右側は「說文解字 True Type 字型」から抽出されたソース画像である．文字のスタイルを学習することで，新しいスタイルの文字を生成することが，モデルを最適化する目的である．本研究では，学習の過程で形成された画像を抽出して利用する．

抽出した字の一部は図 6 の通りで，ソースフォントと比べ，字の構造にも構成にも違いが現れた．

図 6　フォントから生成された「手書き風」篆文文字
Figure 6　Newly generated “handwritten-like” Seal script characters

3.2 生成画像の処理

zi2zi モデルから生成した画像の構造に，ある程度の一致性があり，同一の文字では，たとえばストロークの間隔などの属性に比較的高い一致性が見られた．また，本研究ではランダムな画像処理をある程度加えることで，その形態的な構造を変え，生成した画像の効果を合わせて，手書き篆文特徴に近い画像を生成したあと，これを訓練データとする．

3.2.1 データオーギュメンテーション

本研究では，基本的なランダム形態変化，すなわち，ランダム化アフィン変換（affine transformation），ランダム化射影変換（projective transformation），膨張化処理（dilation），収縮化処理（erosion）の 4 種類の方法を使って，前節で説明した手順で得た画像に拡張作業を施した．その中で，膨張化処理，収縮化処理の二つの変換作業は，異なる太さのペンで書いた文字に適応するため，筆画の幅特徴を調整するのに用いる．そして，ランダム化アフィン変換，ランダム化射影変換の二つの変換作業を行うことで，字にある程度の傾斜，また変形のような特徴を生成し，構造上の特徴を変えた．この手順によって，手書き字の写真を文字認識する際に，撮影時の角度で生じた変形をカバーでき，また個人の書き方の習慣により起こり得る手書き字の特徴の差もカバーできると考えられる．

3.2.2 ランダム化アフィン変換

ランダム化アフィン変換処理では，パラメータを使用し，字形が崩れないように，パラメータ変化の範囲をコントロールする．変換した結果の例を図 7 に示す．

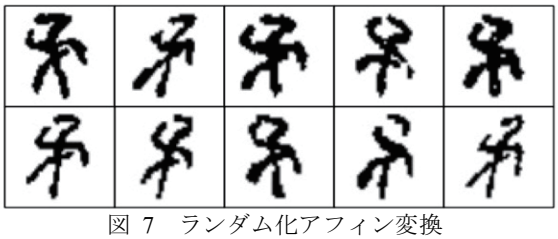

図 7　ランダム化アフィン変換
Figure 7　Randomized affine transformation

3.2.3 ランダム化射影変換

射影変換処理により，元の画像の四つの頂点座標は，ランダムでランダム域内の写像となる．頂点座標の写像域が重なると，写像の字形が崩れる．字形の構造を維持できるよう，ランダム域を図 8 に示すように設定した．

zi2zi モデルの出力画像を 64×64 ピクセルの画像に変換し，ランダム域の大きさを 10×10 ピクセルに設定した．

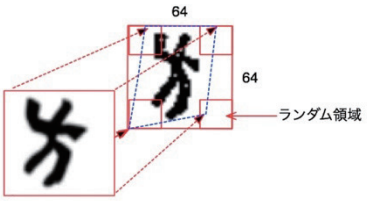

図 8　ランダム化射影変換
Figure 8　Randomized projective transformation

3.2.4 正規化処理

生成画像を形態変化したあと，文字の領域を決めて正規化する必要がある，正規化の流れは図 9 に示すように三つのステップがある．

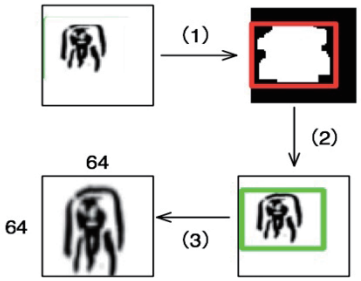

図 9　正規化処理
Figure 9　Normalization process

1）まず，ソーベルフィルタを使い，入力画像のグレースケール図の x 方向の微分を以下で計算する．

$$\frac{\partial}{\partial x} f(x,\ y) \approx\ f(x+1,\ y-1) - f(x-1,\ y-1) + 2f(x+1,\ y) - 2f(x-1,\ y) + f(x+1,\ y+1) - f(x-1,\ y+1) \quad (1)$$

処理した画像を二値画像に変換する．膨張操作と収縮操作の核関数を調整し，そして輪郭を突き出させるために膨張化処理を一回行い，その後，罫線などを除去するため，収縮処理を一回行う．

2）図の中の輪郭を抽出し，その輪郭の面積を計

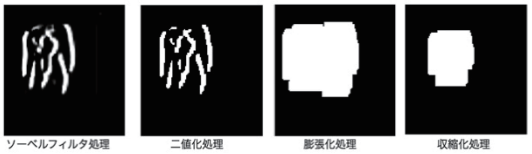

図 10　テキスト領域の選択
Figure 10 Selecting text area

算し，面積が小さすぎる輪郭を取り除く．図の中で最小面積の正方形域を見つけてから，四つの頂点の座標を保存する，その流れを図 10 に示す．
3）最後に，実験に必要な 64×64 ピクセルというサイズの基準に合わせて，リサイズを行う．

3.2.5 さまざまな背景との合成

実際の古代文献の場合では，通常損傷がひどく，ノイズが比較的少ない二値化画像を得るには，元の画像に多距離画像化，スムーズ化，二値化処理，膨張収縮処理などの多くの処理が必要である．近年，ディープラーニングによる画像認識技術の応用で，未処理の背景がある画像は訓練データとして直接入力できるため，本研究では文字と数多くの石刻や画仙紙，木簡，普通紙などの背景の逆融合処理を行い，そして画像の画素画質の相違をカバーするため，ガウスぼかし技術を使ってランダム処理を行う．出力画像を図 11 に示す．

図 11　背景との合成
Figure 11　Synthesis with background

3.3 手書き画像の処理

3.3.1 ノイズの除去

まずは，グレースケール画像に変換し，さらに，二値化画像変換をもとに，膨張収縮の手法を使用し，ノイズを除去する．

3.3.2 正規化処理

ノイズ除去した二値画像に，生成データと同じ提案方法を用いて正規化処理を行う．収集した手書きデータにおける処理効果の例を図 12 に示す．

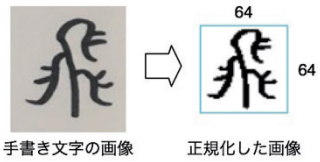

図 12　テスト用手書き画像の処理
Figure 12 Processing of handwritten image

3.4 手書き篆文文字の認識

前の段階で収集した篆文文字に対する手書き文字認識を行う．

3.4.1 モデル

CNN に基づき構成された LeNet5[16]モデルを基にした訓練モデルは図 13 の通りである．図 13 に示すように，入力は，64×64 ピクセルの大きさを持ったパッチ画像であり，具体的には，畳み込み層とプーリング層を複数連ね，その後に一つの全結合層を配置し，最終層のクラス分けでは認識する篆文文字の総数となる．

４．実験

本研究では，生成データを訓練データの訓練モデルとして利用し，手書きデータの文字認識実験テストを行った．「白川フォント」の 1,000 字をランダムに選び，画像に変換する．その中で，ランダムで 300 字を選び取り，それを 300 種類の分類学習に用いた．分類学習ごとの訓練データ数はおよそ画像 300～500 枚である．

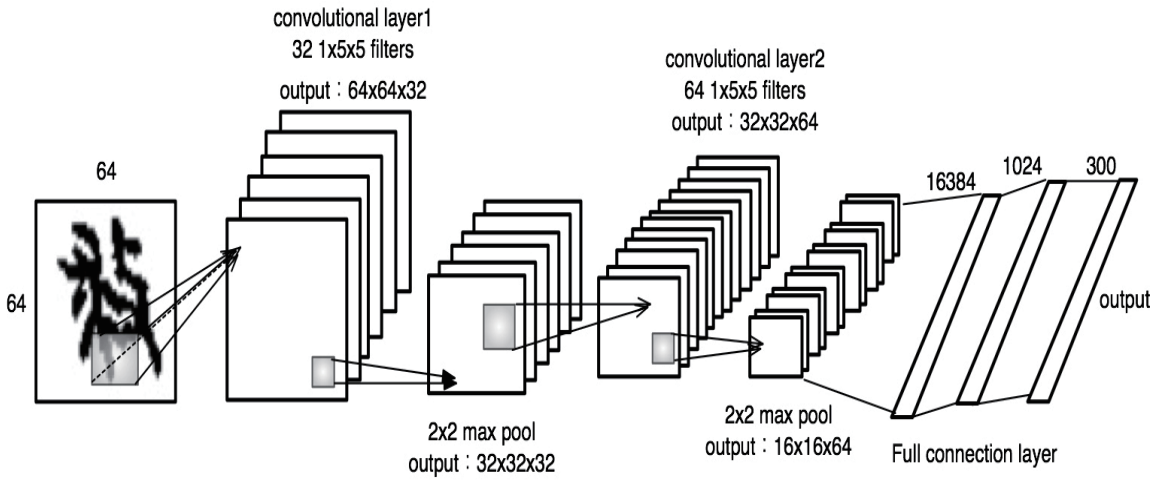

図 13　提案モデル
Figure 13　The proposed model

また，テスト用手書きデータ画像 300 枚を収集した．テスト用の前処理あり手書きデータの一部を図 14 に示す．

図 14　テスト画像
Figure 14　Test data

訓練データに 4 種類の異なる処理を施すという条件の下に，手書き字 300 個の認識実験を行った．評価の手法を式（2）に示す：

正解率 (%) ＝正解数/テストデータ数×100　　(2)

実験結果は表 1 の通りとなった．

表 1　実験結果
Table 1　Experimental results

手法	正解率 (%)
前処理なし	60. 00
+形態変換	73. 33
+背景	63. 60
+形態変換＋背景	77. 00

５．まとめ

本研究では，フォントから生成された文字画像を用い，「白川フォント」において公開されている篆文文字に対する認識手法を提案した．今後の課題として，訓練ネットワークのよりよい調整と，分類可能な文字総数の拡大を検討している．また，フォントから「手書き風」画像へ変換できる新たな手法を検討する．

認識率を高められるように，文字認識のモデルに対してより質的な分析と評価を行うことが必要となる．

参考文献

1)　奈良文化財研究所，東京大学史料編纂所:『木簡画像データベース・木簡字典』『電子くずし字字典データベース』連携検索システム，入手先〈http://r-jiten.Nabunken.go.jp〉（参照 2017-5-16）
2)　人文学オープンデータ共同利用センター: 日本古典籍字形データセット, 入手先〈 http://codh.Rois.ac.jp/char-shape/〉（参照 2017-5-1）
3)　立命館大学白川静記念東洋文字文化研究所: 白川フォント，入手先〈http://www.ritsumei.ac.jp/acd/re/k-rsc/sio/shirakawa/index.html〉（参照 2017-10-15）
4)　Luo, Y., Xia, R., and Abdulghafour, M. :Offline Chinese Handwriting Character Recognition through Feature Extraction. In Proceedings of 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV 2016), pp. 394-398 (2016)
5)　Deepa, A., Rao, R. R.：An Efficient Offline Tamil Handwritten Character Recognition System using Zernike Moments and Diagonal-based features. International Journal of Applied Engineering Research 11(4), pp. 2607-2610 (2016)
6)　石井康史, 藤川佳之, 孟林，山崎勝弘：特徴量を用いた甲骨文字の候補テンプレート抽出と認識. 情報処理学会第 78 回全国大会講演論文集, 2016(1), pp. 211-212 (2016)
7)　Graves, A.：Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850 (2013)
8)　Tian, Y.C. :zi2zi Master Chinese Calligraphy with Conditional Adversarial Network: available from〈https://github.com/kaonashi-tyc/zi2zi〉（accessed 2017-7-1）
9)　Chang, J., Gu, Y. : Chinese Typography Transfer. arXiv preprint arXiv:1707.04904 (2017)
10)　Ghosh, A., Bhattacharya, B. and Chowdhury, S. B. R. : Handwriting Profiling Using Generative Adversarial Networks. In Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017), pp. 4927-4928 (2017)
11)　渡部 宏樹，渡辺 裕：DCGAN を用いたデータオーギュメンテーションによる猫の品種認識について ，2016 年映像情報メディア学会年次大会 (2016)
12)　Creswell, A., Bharath, A. A. : Adversarial Training for Sketch Retrieval. In European Conference on Computer Vision. pp. 798-809 (2016)
13)　Tran, L., Yin, X., and Liu, X.: Disentangled representation learning GAN for pose-invariant face recognition. In Proceedings of Computer Vision and Pattern Recognition Conference (CVPR 2017),Vol. 4, No. 5, pp. 7 (2017)
14)　Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A.: Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004 (2016)
15)　台灣地區行政院主計處電子處理資料中心: 說文解字 True Type 字型，入手先〈http://www.cns11643.gov.tw/MAIDB/welcome.do〉（参照 2017-5-1）
16)　LeCun, Y.: LeNet-5, convolutional neural networks. available from:〈http://yann. lecun. com/exdb/lenet〉（accessed 2017-6-2）