

マンガキャラクタ画像に基づく作者および作品の機械的判別に関する考察

霜出, 秀太
九州大学工学部電気情報工学科

添田, 碧人
九州大学工学部電気情報工学科

伊東, 栄典
九州大学情報基盤研究開発センター

<https://hdl.handle.net/2324/7329778>

出版情報 : JSAI Technical Report, SIG-KBS. 133, pp.33-38, 2024-12-05. The Japanese Society for Artificial Intelligence

バージョン :

権利関係 : Notice for the use of this material. The copyright of this material is retained by the Japanese Society for Artificial Intelligence (JSAI). This material is published on this web site with the agreement of the author(s) and the JSAI. Please be complied with Copyright Law of Japan if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) The



マンガキャラクター画像に基づく作者および作品の機械的判別に関する考察

A study on classification of manga authors and titles based on manga character classification

霜出秀太^{1*} 添田碧人¹ 伊東栄典^{1,2}
Shimoide Shuta¹ Soeda Aoto¹ Ito Eisuke^{1,2}

¹ 九州大学工学部電気情報工学科

¹ School of Engineering, Kyushu University

² 九州大学情報基盤研究開発センター

² Research Institute for IT, Kyushu University

Abstract: In recent years, the unauthorized use of manga images has posed a significant issue, causing substantial harm due to the infringement of rights. This study aims to address the unauthorized reproduction of manga, specifically focusing on methods to identify manga pages based on characters appearing within them. Our approach involved recognizing and extracting character face regions as individual images, followed by character identification using machine learning techniques for image classification. These results are significant for identifying manga pages, as they enable the organization of characters' appearances and their frequency throughout the manga content.

1 はじめに

近年、著作権を無視した海賊版マンガサイトや最新話マンガのスライド動画投稿が問題になっている。特にここ数年では、マンガのページをそのまま転載したものだけでなく、著作権侵害を判定するAIを出し抜くために画像を意図的にずらしたり、一部分だけを切り出して連続的に表示する形式で転載するものも出現している。後者のような無断転載画像も著作権侵害であると正しく認識するには、画像内のキャラクター情報を拾い集めて作者やタイトルを断定し、さらにその登場順や登場頻度を整理することで巻数とページ数を特定する必要がある。

海賊版マンガサイトおよびマンガ動画の無断投稿を抑止するため、我々は今までマンガのページ画像からマンガのタイトル・巻数・話数を推定する判別手法を研究してきた[1]。ページ画像だけでなく、切り取った一部分のマンガ画像からでも、切り抜き元マンガの情報を機械的に抽出したい。マンガ原作を多数作成した小池一夫氏が述べるように、マンガにおいて登場キャラクターは非常に重要である[2]。我々は有名なマンガの

キャラクターであれば、一目見ただけで、それが誰であるか識別できる。

そこで本論文では、マンガのページ画像からキャラクターの顔部分を検出して画像として取り出し、それらのマンガの登場キャラクター画像から、それがどのマンガの誰なのかを判定する手法を検討する。本稿ではまず用いたデータ、キャラクター顔部分検出に適用した手法および検出精度、顔画像からのキャラクター判別に適用した手法および検出精度の順に報告する。その後、これらの手法を用いて実際に市販のマンガ内のキャラクターについてキャラクター判別を行うことができるのか検証する。最後に、この手法がキャラクター判別以外の情報特定にも利用できるのかを検証する。

2 Manga109 データセット

マンガページからのキャラクター顔部分判別機と、キャラクター顔画像からのキャラクター判別機を機械学習で作成するためには、なるべく沢山の「マンガページ画像」と登場する「キャラクター顔画像」、その画像の「キャラクター名」を持つデータを用意する必要がある。機械学習用のデータとして、本研究ではManga109 データセットを用いた。Manga109 は、日本漫画のメディア処理

*連絡先: (所属機関名)
(所属機関住所)
E-mail:

表 1: Manga109 の XML 注釈数

項目	数値
巻数	109
ページ数	10602
注釈	10130
顔	118593
全身	157234
台詞	147887
コマ	103850

の学術研究使用のために、東京大学の相澤・山崎・松井研究室によりとりまとめられたデータセットである [4, 5]。Manga109 は、日本のプロ漫画家により描かれて出版された 109 作品（タイトル）について、1 冊ずつ選ばれた計 109 冊のマンガ画像を保持する。109 冊のマンガは 1970 年代～2010 年代に出版された漫画で、幅広い対象読者層やジャンルを網羅している。

画像に加え 1 冊毎に作品のメタデータがあり、かつ全ページ画像に注釈 (annotation) が付与されている。メタデータには作品タイトル・作者名・出版社・出版年に加え、登場する全キャラクターのリスト、各ページ（見開き）のリストも含む。各ページの注釈は、登場キャラクターの顔領域、キャラクターの全身領域、印字文字のテキスト、手書き文字のテキスト、コマ枠の領域情報を含む。メタデータと注釈は XML 形式で記述されている。アノテーションデータから欲しい対象物の座標データを取得し、画像データから該当部を取り出すことで任意のオブジェクトの切り抜き、保存が可能である。表 1 に注釈の統計を示す。

3 マンガページ画像からのキャラクター顔部分検出

この章では、マンガのページ画像から、そのページに登場するすべてのキャラクターの顔部分を検出するための手法について検討する。

3.1 従来研究

マンガ画像からキャラクターの顔部分を検出する研究は現在様々な手法を下に行われている。マンガ描写はその著者により大きく異なる。同じ作品内のキャラクターでもデフォルメされてその見た目が異なる場合もあるので、そのようなキャラクターの描写の変化に柔軟に対応できる顔検出器の作成が望まれる。過去研究では陳らが幾何学的変化に強く対応できる HOG(Histograms of Oriented Gradients) 特徴量と AdaBoost を利用したマンガキャラクター検出を提案した [10]。柳沢らは DPM(Deformable Part Model)、R-CNN での顔検出や、Faster R-CNN

を用いたキャラクターおよびコマ、フキダシ、文字列の検出の有効性を示した [11][12]。

本研究では、Faster R-CNN と公開されている 2 つの 2 次元顔検出分類器を用いて、マンガページからキャラクターの顔の位置情報を検出するモデルの作成を作成した。

3.2 キャラクター顔部分検出手法

顔部分の検出に使うモデルとして、以下の 3 つを検証する。

animeface (Viola-Jones) モデルは OpenCV の Viola-Jones 物体検出フレームワークに基づいたカスケード分類器の一種で、nagadomi 氏により作成・提供されている [8]。

animeface (CascadeClassifier) モデルは、シンプルで高速なアニメ顔検出に適したモジュールである。OpenCV の CascadeClassifier をベースに開発されており、アニメ顔の特徴を学習したカスケード分類器を用いて、画像からアニメ顔を検出する [9]。

Faster R-CNN (Detectron2) は、画像内の物体を検出する真相学習モデルの一種で、特にリアルタイムな物体検出において高い性能を発揮する。Detectron2 は、Facebook AI Research が開発した、PyTorch ベースの物体検出ライブラリである。

3.3 キャラクター顔部分検出実験

3 種類の検出器について、マンガキャラクターに対する検出率を比較する。Manga109 データセットの images を入力として用いて、出力として得られたキャラクターの顔の位置情報と、annotation に記載されたキャラクターの顔の位置情報を比較し、その重なり具合 (IoU:Intersection over Union) が指定値を超えていたものを検出成功例とする。入力の画像例と検出の様子を以下に示す。なお、表では、animeface(Viola-Jones) を af-VJ、animeface(CascadeClassifier) を af-CC、Faster R-CNN を F.R-CNN と表す。



図 1: Manga109 データセットにおける各検出器の結果の比較 (左:af-VJ, 中央:af-CC, 右:F.R-CNN)

公開されていた検出器はどちらも Precision は高いが、Recall が低いという傾向が見られた。これは、検

表 2: 判別精度 (IoU=0.5)

Method	Prec.	Rec.	F-s.	AP.
af-VJ	0.9813	0.1502	0.2648	0.1865
af-CC	0.9927	0.3387	0.5051	0.4937
F_R-CNN	0.7661	0.8306	0.7971	0.8112

出したものはほぼ全て顔であったが、ページに登場するすべての顔のうち検出できた顔の割合が低いということを表している。それに対して Faster R-CNN では、Recall と Precision の両方で高い値を出している。これらの数値は図 1 の検出結果の様子からも見てとれる。

af-VJ モデルに関しては、ページ内の顔画像を全く検出できていないことがわかる。af-CC モデルは、真正面かつパーツのほとんどの部分が描写されている顔のみ検出しており、横顔やフキダシ等でほとんど隠れてしまっている顔の場合は検出できない傾向にある。F_R-CNN モデルについては、真正面の顔だけでなく、横顔や領域の半分がフキダシで隠れている顔、小さく描写されている顔まで検出に成功している。

またこのモデルは、顔を検出したとき、その検出結果がどれくらい正確なのかを「確信度」という指標で示している。今回は確信度 50 % 以上のものを使用して判別精度を算出したが、この指標を調整して性能を評価することで顔でない部分に誤反応している検出候補を除外することができる。それによりさらに精度が上がると思われる。

キャラクタ顔部分検出実験の結論として、キャラクタの顔部分特定は、公開されている分類器よりも Faster R-CNN の detectron2 を用いて学習させた分類器のほうが顔検出に適しているということが示された。

4 キャラクタ顔画像からのキャラクタ判別

この章では、マンガページ内に登場するキャラクタ顔画像があらかじめ用意できることを前提とし、それらの画像からキャラクタを機械的に判別するための手法について検討する。

4.1 データの整理

Manga109 データセットには、顔部分座標とキャラクタ識別子が注釈として提供されている。このデータを対象に、複数手法でキャラクタ判別を試した。元の XML 注釈の顔部分座標では、キャラクタの頭部（髪型）が含まれない場合が多い。そこで、本稿では Manga109 の「顔部分」の画像群と、顔部分から上および左右に

拡張した画像も試した。図 2 に顔部分画像の抽出方法を示す。

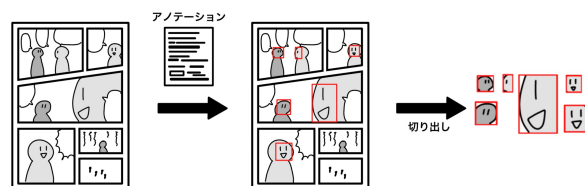


図 2: 顔部分画像の抽出

Manga109 の注釈 XML に記載されている顔座標の通りに顔部分を切り出した画像群を A（切り出し A）、上左右を拡張して顔部分を切り出した画像群を B（切り出し B）とする。切り出し B では、上に 20%、左に 10%、右に 10% のピクセルを増やして切り出した。図 3 に A と B の切り出し方法を示す。

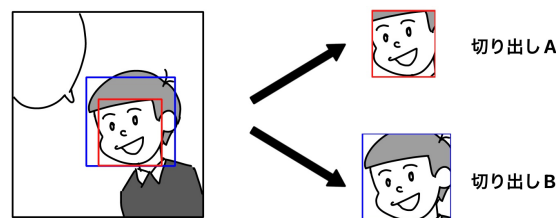


図 3: 顔画像の切り出し方

Manga109 の注釈で抽出できる顔部分画像の解像度が低い（領域が小さい）キャラクタや、切り取られた画像が少ないキャラクタは、学習や分類に適していない。そこで、顔領域が 10000px 以上の画像数が 50 以上であるキャラクタの画像のみに限定した。表 6 に、学習に用いた切り出し画像の数、キャラクタ数、などを示す。

表 3: 学習データ数

項目	数値
顔画像数	5,348 個
キャラクタ数	61 人
キャラクタ事平均画像数	87.67 個
作品あたりのキャラクタ数	3.59 人

4.2 用いた画像判別器

マンガに登場するキャラクタの顔画像を学習させるための CNN アーキテクチャとして、画像認識の分野で高い性能を発揮する ResNet[6] と DenseNet の 2 つを用いた。ResNet では、ある層に与えられた信号を、上位層の出力に追加するスキップ接続により深いネッ

トワークを訓練できるようにしている。DenseNet は ResNet から開発されており、DenseNet は前の各層からの出力すべてが後の層への入力として用いられる。

ResNet には層の数により ResNet34, ResNet50, ResNet101 などの種類がある。本論文では ResNet50 のみを試した。DenseNet も層の数により DenseNet121, DenseNet169, DenseNet201 などの種類が有る。本論文では DenseNet121 のみを試した。

学習モデル構築のプログラムは Python 言語を用いて作成した。機械学習および画像判別精度の出力部分は、生成 AI の ChatGPT（無料版）が回答したコードを援用して実装した。ライブラリとして Tensorflow と Keras を用いている。

4.3 作成した学習モデル

ResNet50 を用いて、切り出し A と B の顔画像の分類を学習させたモデルをそれぞれ R-A, R-B とする。DenseNet121 を用いたものを、それぞれ D-A, D-B とする。これらの学習モデルを表 4 に示す。

表 4: 学習モデル

画像群	ResNet	DenseNets
切り出し A	R-A	D-A
切り出し B	R-B	D-B

4.4 分類精度

表 5 に、4 つのモデルの Recall, Precision, F1-Score, Accuracy を示す。

表 5: 分類精度

モデル	Recall	Precision	F1-Score	Accuracy
R-A	0.8393	0.8705	0.8462	0.8575
R-B	0.8826	0.8981	0.8864	0.8904
D-A	0.8921	0.9053	0.8951	0.9050
D-B	0.9181	0.9284	0.9205	0.9269

表 5 を見ると、ResNet50 と DenseNet121 を比較すると、切り出し A と B のどちらでも DenseNet121 の方が精度が良い。次にキャラクターの顔部分の切り出し方について、座標通りの切り出し方 A と、上左右に拡張した切り抜き B を比較する。ResNet50 を用いたモデルと、DenseNet121 を使用したモデルの両方で切り出し B の方が精度が良い。

4.5 キャラクタ判別の精度に関する考察

ResNet よりも DenseNet の方が精度が良くなる理由および切り出し A より B の方が精度が良くなる理由を調べるために、4 つのモデルが、キャラクター顔画像のどの部分に着目しているのか調べる。そのために各モデルの attention-heatmap を作成した。

図を見ると、DenseNet は ResNet と比べて全体的にキャラクターの目に attention が集まっている。さらに、拡張して髪の毛部分も含めて切り出した画像群を学習した D-B は主に前髪と目を含めた部分に attention が集まっている。上左右を拡張したことでキャラクターの目以外の情報を特徴として取り入れることができたため精度が向上したと考えられる。

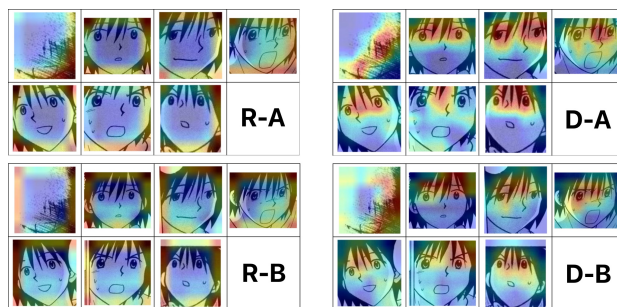


図 4: 各モデルの attention-heatmap 例

キャラクター判別実験の結論として、キャラクター判別の手法としては、DenseNet を使用して髪の毛等を含めた顔画像を学習させたモデルを使用すると最も精度が高くなることがわかった。

5 実際のマンガ画像での顔部分認識とキャラクター判別

5.1 実際のマンガへの適用

これまでで説明した手法を、注釈などが存在しない実際のマンガに対して適用できるのかを試した。実際のマンガとして、福本伸行氏の作品『賭博目次録カイジ』[13] の 1～3 巻を使用した。3、4 章で述べた実験結果をもとにした手法を使って顔部分の検出とキャラクター分類学習モデルの作成を行い、『賭博目次録カイジ』に登場するキャラクターを判別できるのか検証した。さらに、キャラクター分類から枠組みを広げたタイトル分類、二値分類を行うことができるのかを試した。

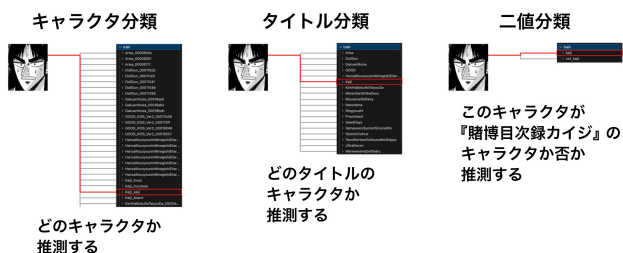


図 5: キャラクタ分類・タイトル分類・二値分類

5.2 キャラクタ顔部分検出を利用したデータセットの作成

『賭博目次録カイジ』に登場するキャラクターの顔部分について、Faster R-CNN の detectron2 を用いて検出を行った。のちに教師データとして使用するデータセットの作成にこの検出を利用するため、顔でない部分を誤検出したパターンをできるだけ排除したい。したがって、推論時に 90 %以上の確率で顔であると認識したもののみを検出するようにした。また、これまでに実験で用いたデータ形式に合わせるために、検出した顔部分に対して縦横それぞれ拡張して切り出し、キャラクターごとに手動で分別した。

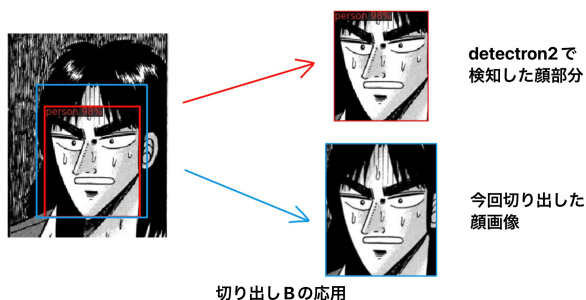


図 6: 顔画像の切り出し方

Manga109 から作成したデータセットと同じように、10000px 以上のピクセル数を持つ顔画像が 50 枚以上存在するキャラクターのみをこれまでのデータセットに追加することで、Manga109 の 61 キャラクタに『賭博目次録カイジ』の 4 キャラクタを加えた新規のデータセットを作成した。

表 6: 新規追加した学習データ数

項目	数値
顔画像数	6,233 個
キャラクター数	65 人
キャラクター事平均画像数	95.89 個
作品あたりのキャラクター数	3.61 人

5.3 実際の漫画のキャラクター分類

注釈が用意されている Manga109 ではなく、detectron2 を用いて実際のマンガから切り出した顔画像でもキャラクター分類がうまくいくのかを確かめるために新規のデータセットを使用した。先述した実験と同様に DenseNet を使用してキャラクター顔画像分類学習モデルを作成し、その精度を確かめた。

表 7: カイジ追加版分類精度

指標	数値
Recall	0.9048
Precision	0.9143
F1-Score	0.9054
Accuracy	0.9245

表を見ると、Manga109 のみを使用した場合に比べて精度は劣るものの、高い精度でキャラクター判別が行えることがわかった。したがって、detectron2 を用いて実際のマンガから切り出した顔画像でもキャラクター分類を高精度で行うことができることがわかった。

5.4 キャラクタ断定までの段階設定

注釈がなくてもキャラクター分類が高精度で行うことができることは検証できたが、作成したモデルだけでは判別の誤りを 0 にすることはできない。しかし、マンガページやコマから顔画像を切り出して測定することを考えると、同作品のキャラクターを複数与えて元のマンガの情報を抽出することになるため、複数の顔画像から多数決形式でタイトルを先に推定することでキャラクターを判別する際に選択肢を減らし、間違いを減少させることができると考えられる。したがって、まずキャラクター顔画像から作者判別モデルを用いて作者を断定し、次に該当する作者専用の二値分類モデルを用いてタイトルを断定し、最後に該当タイトル専用のキャラクター分類モデルを用いてキャラクターを断定する段階方式を考案した。

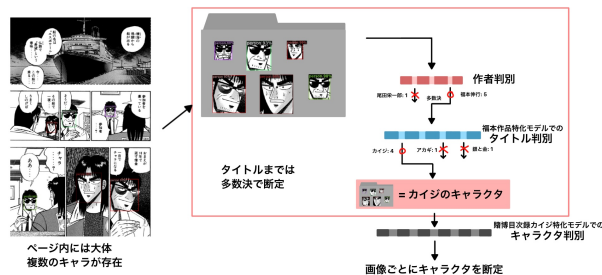


図 7: キャラクタ特定までの段階

今回はその足がかりとして、作者分類を最終目標としたタイトル分類モデルと、特定タイトルか否かの二値分類モデルを作成してその精度を検証した。

5.5 タイトル判別/特定タイトルか否かの二値判別

キャラクターの顔画像から、抽出した Manga109 に『賭博目次録カイジ』を加えた全 18 作品のどの作品にそのキャラが登場するかどうかを推測するタイトル判別と、『賭博目次録カイジ』か否かを推測する二値判別ができるかどうか検証した。DenseNet と顔画像は同じものを用いたが、データセットの構造を変えてタイトルごとに学習し直したモデル、二値分類として学習し直したモデルのそれぞれを作成して精度を確かめた。

表 8: タイトル分類精度

モデル	Recall	Precision	F1-Score	Accuracy
タイトル分類	0.9531	0.9552	0.9523	0.9731
二値分類	1.0000	0.9994	0.9997	0.9996

表を見ると、タイトル判別はキャラクター判別よりも高い精度で行うことができることがわかった。同様に、キャラクター別に整理されたデータセットがない新規のマンガ作品でも、detectron2 を用いて顔画像を切り出したものさえ用意すればタイトル判別をすぐに行うことができるかと予想される。

また二値分離については、『賭博目次録カイジ』か否かの判別はほぼ 100 % で当たるという高精度で行うことができるとわかった。これは、『賭博目次録カイジ』が特徴的な絵柄であり他の作品との見分けがつきやすいことによるものだと考えられる。同様の手法を用いることで、精度は劣るものの『賭博目次録カイジ』に限らず他の作品でも特定タイトルか否かを判別できることが期待できる。

6 おわりに

本論文では、マンガページからの顔部分検出とマンガに登場するキャラクターの顔部分画像からのキャラクターの判別を試みた。Manga109 データセットに含まれるキャラクターの顔画像を機械学習させて、判別した。ResNet, DenseNet を用いた場合の精度は DenseNet の方が高かった。また、拡張画像を使用した場合の方が精度が高くなることも確認できた。以上の手法を用いて、Manga109 に含まれない実際の作品でも高い精度でキャラクター判別できることがわかった。さらに、タイトル判別や特定タイトルか否かの判別も同様の手法で高精度で行うことができるとわかった。

謝辞

本研究は、JSPS 科研費 24K15654 の助成を受けた。

参考文献

- [1] 山仲一颯, 伊東栄典: 大規模マンガ画像判別手法の検討, 情報処理学会 火の国情報シンポジウム 2024, A4-1, Mar.14, 2024.
- [2] 小池一夫: 小池一夫のキャラクター新論 ソーシャルメディアが動かすキャラクターの力, ゴマブックス, 2017.
- [3] Wei-Ta Chu, Wei-Wei Li: Manga FaceNet: Face Detection in Manga based on Deep Neural Network, ICMR'17, pp.412–415, 2017.
- [4] Manga109, <http://www.manga109.org/ja/index.html> (accessed on 05-Aug-2024).
- [5] K. Aizawa, et.al.: Building a Manga Dataset “Manga109” with Annotations for Multimedia Applications, IEEE MultiMedia, vol.27, no.2, pp.8–18, 2020.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition, <https://doi.org/10.48550/arXiv.1512.03385>, 2015.
- [7] G.Huang, Z.Liu, L.van der Maaten, K.Q.Weinberger: Densely Connected Convolutional Networks. IEEE Conference on Pattern Recognition and Computer Vision (CVPR), 2016. <https://arxiv.org/abs/1608.06993>
- [8] nagadomi: lbpcascade_animeface, https://github.com/nagadomi/lbpcascade_animeface (accessed on 05-Aug-2024).
- [9] <https://github.com/opencv/opencv/tree/master/data/haarcascades> (accessed on 05-Aug-2024).
- [10] 陳 明, 柳澤 秀彰, 張 傑, 石井 大祐, 渡辺 裕: マンガにおける HOG+AdaBoost による顔検出の性能評価, 映像情報メディア学会年次大会, 17-4, Aug.31, 2014.
- [11] 柳澤 秀彰: マンガキャラクターを対象とした多視点顔検出の研究, 早稲田大学修士論文, Feb.1, 2016.
- [12] 柳澤 秀彰: ストーリー理解を目的とした漫画オブジェクトの抽出, 早稲田大学, 博士論文, 甲第 5715 号, 2019.
- [13] 福本伸行: 賭博目次録カイジ, 集英社.