

マンガ顔画像からのキャラクタ判別に関する検討

霜出, 秀太
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/2324/7329573>

出版情報 : Record of Joint Conference of Electrical and Electronics Engineers in Kyushu. 77 (09-1P-09), pp.188-189, 2024-09-19. Committee of Joint Conference of Electrical, Electronics and Information Engineers in Kyushu

バージョン :

権利関係 : © 2024 Committee of Joint Conference of Electrical, Electronics and Information Engineers in Kyushu



マンガ顔画像からのキャラクタ判別に関する検討

霜出 秀太* 伊東 栄典**

(九州大学 *工学部電気情報工学科 **情報基盤研究開発センター)

* shimoida.shuta.112@s.kyushu-u.ac.jp, ** ito.eisuke.523@m.kyushu-u.ac.jp

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制するため、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、マンガの登場キャラクタ画像から、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。なお、ページ画像からの顔部分抽出については、文献 [3] に有るような手法を用いるものとする。本稿では、用いたデータと、適用した手法および検出精度について報告する。

2 Manga109 とキャラクタ顔部分画像抽出

マンガのキャラクタ顔部分画像からキャラクタ名を判別する判別機を機械学習で作成するためには、なるべく沢山の「キャラクタ顔画像」と、その画像の「キャラクタ名」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、登場キャラクタの顔領域、キャラクタの全身領域、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

表 1: Manga109 の XML 注釈数

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

存が可能である。表 1 に注釈の統計を示す。

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 1 に顔部分画像の抽出方法を示す。

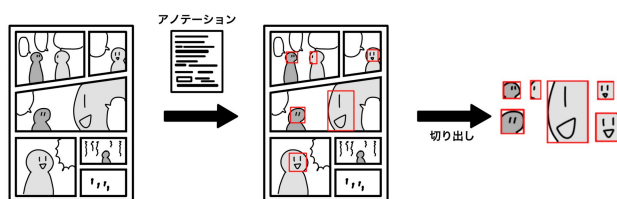


図 1: 顔部分画像の抽出

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 2 に A と B の切り出し方法を示す。

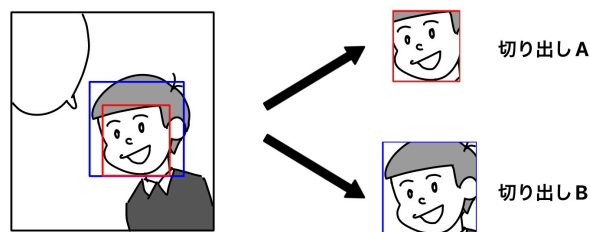


図 2: 顔画像の切り出し方

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 2 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。

3 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れるための 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

表 2: 学習データ数

	切り出し A
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

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。

3.1 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 3 に示す。

表 3: 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 分類精度

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

表 4: 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 4 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。

5 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。

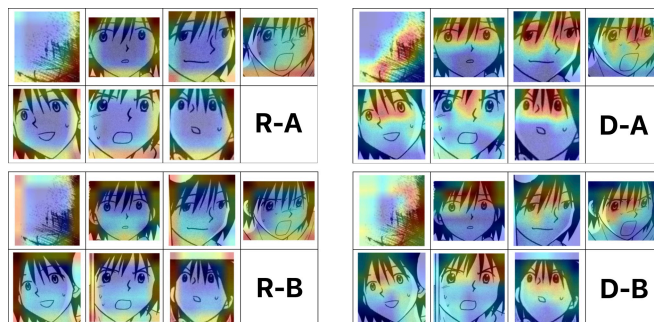


図 3: 各モデルの attention-heatmap 例

6 おわりに

本論文では、マンガに登場するキャラクタの顔部分画像から、キャラクタの判別を試みた。Manga109 データセットに含まれるキャラクタ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。

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>