

Manga109 データセットに対する擬音語アノテーション Onomatopoeia annotation for Manga 109 dataset

加藤 ななみ^{*1}, 若林 哲史^{*1}, 大山 航^{*2}

Nanami Kato^{*1}, Tetsushi Wakabayashi^{*1}, Wataru Ohyama^{*2}

^{*1} 三重大学大学院 工学研究科, 三重県津市栗真町屋町 1577

^{*1} Mie University, Kurimamachiya-cho Tsu city, Mie

^{*2} 九州大学大学院システム情報科学研究院, 福岡市西区元岡 744

^{*2} Kyushu University, 744, Motooka, Nishi-ku, Fukuoka-shi

概要: 日本の漫画は世界に広く知られ漫画翻訳の需要が高まっている。また電子漫画の需要増加により電子漫画ならではの付加価値を持たせる試みが行われている。オノマトペを自動で抽出することにより手間・費用の削減が期待できる。本研究では、漫画データセット“Manga109”で提供されているアノテーションデータに含まれていない、擬音語・擬態語に対するアノテーションデータを作成した。それを用いて機械学習しオノマトペを抽出する実験を行い、約 60%の F 値が得られた。

Abstract: Japanese comics attract a large population all over the world. For rapid distribution to the world, the contents in the comics should be translated fastly and effectively, but it is tedious and time-consuming jobs. Adopting machine learning based digital image processing technologies may be a promising solution for the situation. However, the lack of enough amount dataset leads difficulty to develop such systems. In this research, the authors propose semantic annotation dataset for onomatopoeia in Japanese comics database "Manga109". This dataset can be used for training machine learning algorithms for extracting or recognizing hand-writing onomatopoeia. The results of the initial trial using the created annotations for training U-Net, a neural network architecture for region extraction, show that about 60% of F-measure for trained comic books.

キーワード: Manga109, オノマトペ, 深層学習, U-Net

Keywords: Manga109, onomatopoeia, Deep Learning, U-Net

1. はじめに

漫画は日本を代表するメディアコンテンツの一つであり、世界中で広く親しまれている。漫画は日本文化を世界に発信する手段として注目が集まっており、各国の言語で出版するために漫画翻訳の需要が高い。漫画は、擬態語や擬声語を表現する手書き文字であるオノマトペを含む。オノマトペは、その効果で聴覚の

イメージに訴えかけ、臨場感を高め、漫画の表現を広げる重要な要素である。漫画の翻訳は、漫画画像に対し、日本語の文字を消す、セリフなどの文を翻訳する、翻訳した文字を画像内に配置する、という流れで行われ、人手や時間がかかる。そのため、外国語版においてオノマトペは翻訳されない場合もある。

オノマトペの自動検出が実現できれば、翻訳の手間と費用の削減が期待できる。しかし、オノマトペの自動検出は以下のような困難を含む。第一にオノマトペは背景と一体化していることがある、第二にページの端で文字が途切れていることがある、第三に書き方が作者や場面によってさまざまである。これらが原因となり、従来の文字認識技術では検出が困難である。近年では、漫画解析の研究の進展に伴って、コマを検出できる成果[1]や、スマートフォン上で動作するアプリケーションソフトウェアが存在する[2]。しかしオノマトペの高精度な自動検出は実現されていない。

また、電子書籍市場の発展により、電子漫画の需要も増加している。電子書籍の国内市場規模は年々拡大しており、その8割を電子漫画が占めている[3]。電子漫画では、オノマトペに動きをつけるなど、電子漫画ならではの付加価値を持たせる試みが行われている[4][5]。このようなコンテンツ制作においても、漫画からオノマトペを自動検出し、ページ上に重畳した別レイヤーに配置して漫画画像に階層構造を導入すれば、利便性の向上が期待できる。

近年、深層学習に代表される機械学習技術が様々な分野に導入されている。特に深層学習においては、学習用データの量と質が性能に大きな影響を与える。高い性能を得るためには、適用タスクに適した学習用データセットを収集する必要がある。研究に利用可能な代表的な漫画画像データセットである Manga109[6]には種々のアノテーションデータも含まれているが、オノマトペに対するアノテーションは含まれていない。また、筆者らの知る限り、漫画中のオノマトペに対して適切にアノテーションが付与されたデータセットも存在しない。本研究では、Manga109 に含まれるオノマトペに対するアノテーションデータを作成した。また、領域検出を行うニューラルネットワークの一種である、U-Netを用いた漫画画像中のオノマトペ検出手法を提案する。図1のように、さまざまなオノマトペをピクセル単位で検出し、オノマトペのみを含む画像を出力する。



(a) 入力画像 (b) 出力画像

図1 オノマトペ抽出の目標



図2 オノマトペと類似した手書き文字の例

2. Manga109 データセットに対する擬音語アノテーション

2.1. Manga109 について

Manga109 は、様々なジャンル、年代の 109 作品の漫画 2114 ページを含む漫画画像データセットである。漫画の見開き2ページ分を、サイズ 1654×1170 画素のグレースケール画像ファイル1個として収録している。

Manga109 に含まれるすべての漫画にはアノテーションが付与されている。アノテーションデータは漫画 1 作品に対し 1 個の xml 形式ファイルが対応している。アノテーションはキャラクターの顔、キャラクターの全身、コマ枠、テキストの 4 種類がつけられており、全てに id と矩形領域 (xmin, ymin, xmax, ymax) が付与されている。種類別の固有情報として、キャラクターの顔とキャラクターの全身には漫画内で統一されたキャラクター id が、テキストにはテキストの内容が与えられている。コマ枠は固有情報を持たない。テキストはセリフ、モノローグ、状況説明を表す活字と一部の手書き文字に対して付与されており、オノマトペには付与されていない。

2. 2. オノマトペに対するアノテーション

本研究では、Manga109 でアノテーションが付与されていないものも含む全ての手書き文字に対してアノテーションを付与する。漫画中の手書き文字には本来の擬音語、擬態語以外にも含むが、本研究では以下の理由によりアノテーションの対象とした。

1. 図2(a), (b)に含まれるセリフとオノマトペのように、類似した書体で描写される場合が多い。
2. 図2(c)のように吹き出し内のセリフが手書き書体で描画されている場合があり、形状特徴のみを用いる検出の段階において、これらを切り分けることは困難である。
3. 図2(d)のように、作者のコラムがコマ内に手書きにより記述される場合がある。漫画の翻訳においてはこれらの記述も翻訳の対象となる。

アノテーション付与作業は目視により手作業で行った。本研究で付与したアノテーションは以下の形式である。アノテーションの例を図 3 に示す

- ・ 漫画 1 ページに対し、1 画像ファイルに対応させる。
- ・ 画像のサイズはすべて 827×1170 画素である。
- ・ 画像中で、オノマトペ、それ以外の手書き文字に属する画素をそれぞれ赤色、青色で表す。

現時点で、以下に示す 5 作品のアノテーションが完了している。

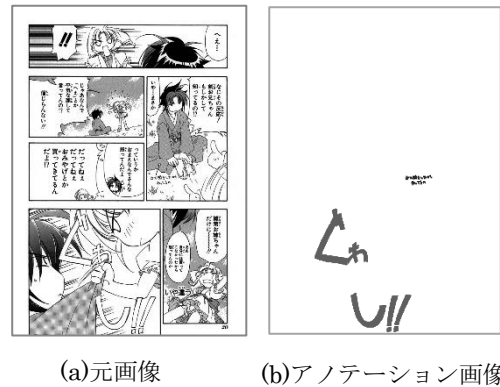


図 3 アノテーション画像の例

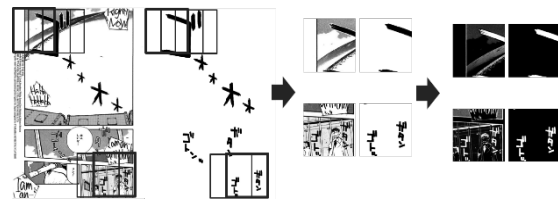


図 4 前処理

- ・ 4コマ漫画:あくはむ 1巻 [8]
- ・ バトル漫画:学園ノイズ 1巻 [9]
- ・ 恋愛漫画 1:ゆめ色クッキング 1巻 [10]
- ・ 恋愛漫画 2:ひなぎく見参! 一本桜町編 1巻 [11]
- ・ SF漫画:ヒーリング・プラネット [12]

3. 深層学習による漫画中のオノマトペの抽出

3. 1. 抽出手法の流れ

本手法では、漫画画像を入力し、オノマトペまたは手書き文字のみを含む 2 値画像を出力する。この手法は前処理、U-Net によるオノマトペ抽出、後処理からなる。

前処理では、漫画1ページの画像を白黒反転し、以下の条件でお互いが一部重畳する小領域に分割する。小領域のサイズは256×256画素、小領域分割の間隔は縦横それぞれ100画素である。分割により得られる小領域の数は1ページあたり60個である。

次に、学習済みの U-Net によりオノマトペを抽出する。256×256画素の小画像を入力とし、オノマトペまたは手書き文字(以下、表記の簡単化のためにオノマトペと表記する)を前景(白画素)として含む2値画像を出力する。

後処理として、前処理の小領域切り出しと逆の要領で各小領域を重ね合せて繋ぎ、もとの1ページの画像と同じサイズに戻す。小領域の重畳部分は画素値を加算し、255を超えた場合は255とする。1ページ分が1枚にまとめられたのち2値化を行う。

3. 2. U-Net

U-Net [7]は医用画像処理のために開発されたセマンティックセグメンテーションのためのニューラルネットワークである。位置ずれや大きさの違いに頑健かつ、原画像における位置情報を復元するため、セグメンテーションタスクで高い精度を誇る。畳み込みにより、層が深くなるほど局所的な情報となり全体的な位置情報は曖昧になるため、位置ずれや大きさの違いに頑健になる。しかしセマンティックセグメンテーションでは原画像における正確な位置情報を特定する必要があるため、曖昧となった位置情報を復元するため浅い層の情報を連結している。

図5に今回用いたU-Netの構成を示す。活性化関数には、Rectified Liner Unit(ReLU)を使用する。最終層の活性化関数には sigmoid を使用し、各画素がオノマトペまたは非オノマトペのそれぞれに属する確率を出力する。

3. 3. 実験

3. 3. 1. データセット

アニメーションを作成した5種の漫画のうち、3種の漫画を学習に、5種の漫画を評価用に使用する。各漫画それぞれ90画像を学習用、10画像を検証用、20画像を評価用に用いた。

3. 3. 2. 学習

提案手法におけるU-Netは、入力画像をオノマトペ、非オノマトペをそれぞれ前景、背景とする2値画像に変換する。前述した通りに小領域に分割した原画像と教師画像のそれぞれを与え、U-Netを学習する。

以下に示す2種類の学習用データセット構成について、オノマトペ検出性能を比較した。

1. 学習用とした3作品全体を用いて学習した1個のU-Net (全体学習と呼ぶ)

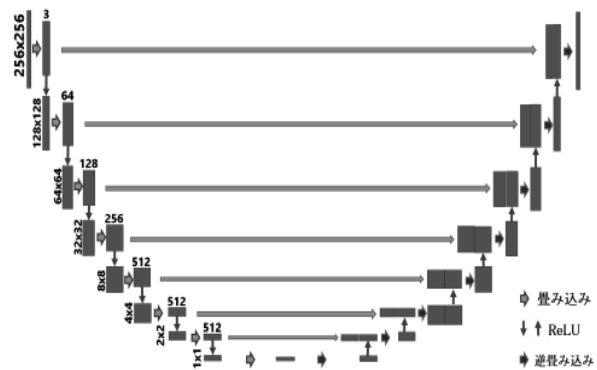
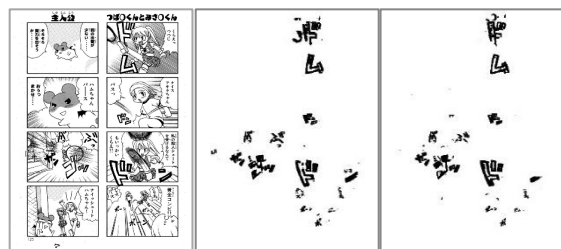


図 5 U-Net



(a) 4コマ漫画に対する検出結果、左から順に、原画像、個別学習の検出結果、全体学習の検出結果である。



(b) バトル漫画に対する検出結果、配置は(a)と同様である。



(c) 未学習作品に対する検出結果、左から順に、原画像、全体学習による検出結果である。

図 6 検出結果

2. それぞれの作品に対して個別に学習した複数の U-Net (個別学習と呼ぶ)

また, 全体学習した U-Net に未学習の漫画を入力した際の結果を検証する.

U-Net 学習時の損失関数としてダイス係数を, 最適化法として Adam を用いた. バッチサイズ 10, エポック数 100 で学習した.

3. 3. 3. 評価方法

提案手法によるオノマトペ検出の正確性を, 画素数で算出した再現性(R), 精度(P), および F 値(F)で評価した. R, P, F の定義を以下の式に示す.

$$R = \frac{TP}{TP + FP}$$

$$P = \frac{TP}{TP + FN}$$

$$F = \frac{2 \times P \times R}{R + P}$$

ここで, TP, FP, FN はオノマトペに属する画素に対して, それぞれ, 正しく検出, 過検出, 未検出した画素数である. 作品全体で TP, FP, FN を数え, R, P, F を求める.

3. 3. 4. 実験結果および考察

提案手法によるオノマトペ検出結果の例を図 6 に示す. 図中の黒画素がオノマトペに属する画素として検出された画素である. 図中の(a),(b)はそれぞれ, 4コマ漫画, バトル漫画に対するオノマトペ検出結果である. それぞれの図において, 左から順に, 元画像, 全体学習によるオノマトペ検出結果, 個別学習によるオノマトペ検出結果である.

(a),(b)の両方において, 原画像に含まれるほぼ全てのオノマトペが検出されている. 一方で, キャラクターの一部や, 吹き出し内の活字がオノマトペとして過検出されている.

表 1 学習漫画の検出結果

F 値	4コマ	バトル	恋愛漫画 1
全体学習	69.50%	56.70%	54.10%
個別学習	66.50%	42.10%	56.90%
適合率	4コマ	バトル	恋愛漫画 1
全体学習	71.30%	55.20%	44.20%
個別学習	69.40%	32.90%	63.90%
再現率	4コマ	バトル	恋愛漫画 1
全体学習	67.80%	58.30%	69.90%
個別学習	63.70%	58.30%	51.40%

表 2 未学習漫画の検出結果

F 値	恋愛漫画 2	SF 漫画
全体学習	46.38%	20.00%
適合率	恋愛漫画 2	SF 漫画
全体学習	44.87%	11.84%
再現率	恋愛漫画 2	SF 漫画
全体学習	47.99%	63.62%

表1に作品ごとのオノマトペ検出性能の, 学習方法による違いを比較した結果を示す. 4 コマ漫画とバトル漫画では, 全体学習の方がF値が大きく, 検出性能が向上した. 一方, 恋愛漫画では, 個別学習の方が大きいF値となったが, この理由は, 個別学習の方が再現率が大きくなったためである. 学習データセットに含まれない作品に対してオノマトペ検出を行なった結果を表2に, 例を図5(c)に示す. 6コマ目の“ふん”はある程度抽出されているが, 4 コマ目の“ぐる”は十分に抽出されていない. また, 恋愛漫画2とSF漫画は同作者による作品だが, 検出結果に26%もの差が出ている. 現時点では未学習の漫画作品に対して安定したオノマトペ抽出は困難である.

4. まとめ

Manga109 のアノテーションの対象に含まれていない, オノマトペを含む手書き文字のアノテーションを漫

画5巻分作成した。作成したアノテーションを用いて深層学習を行い、抽出する実験を行った結果、F値は約60%となった。対象の漫画を学習していない深層学習モデルでの抽出は困難である。

今後の研究課題は主に以下の2点である。第一に、現時点では学習用データとして利用できるオノマトペのアノテーション数が不足しているため、今後は、さらにアノテーション数を増加させてニューラルネットワークによる検出の性能を向上させる予定である。第二に、現時点では画像として作成しているアノテーションを、Manga109で提供されているXMLフォーマットで公開する予定である。

参考文献

- [1]石井大祐，河村圭，渡辺裕，“コミック画像のコマ分割処理における制御パラメータに関する検討”，電子情報通信学会パターン認識・メディア理解研究会 PRMU2009-34 (IE2009-43, MI2009-34), pp.187-192, May 2009.
- [2]ふ き だ し ズ ー ム ，
<https://japan.googleblog.com/2017/01/google-play.html>
- [3]インプレス総合研究所，“電子書籍ビジネス調査報告書 2018”
- [4]松下光範，今岡夏海，“デジタルコミック制作のための動的な音喩表現生成システム”，人工知能学会，IC1-OS4a-3, 2011.
- [5]佐藤剣太，中村聡史，鈴木正明，電子コミックの表現を豊かにする手書き文字アニメーション生成手法，第30回人工知能学会全国大会(JSAI2016)，vol. 4L4-4in2, p. 1-4, 2016.
- [6]Manga109, <http://www.manga109.org/>
- [7]Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015
- [8]新居さとし，あくはむ，第1巻，講談社，2005.
- [9]猪原大介，学園ノイズ，第1巻，東京三世社，2003.
- [10]くりた陸，ゆめ色クッキング，第1巻，講談社，1988.
- [11]桜野 みねね，ヒーリング・プラネット，第1巻，エ

ニックス，2000

- [12]桜野 みねね，ひなぎく見参！一本桜花町編，第1巻，東京三世社，2007