

# Faster R-CNN を用いたマンガ画像からのメタデータ抽出

## Metadata Extraction from Comic Images Using Faster R-CNN

柳澤 秀彰<sup>†</sup> 渡辺 裕<sup>†</sup>

Hideaki YANAGISAWA<sup>†</sup> and Hiroshi WATANABE<sup>†</sup>

<sup>†</sup> 早稲田大学大学院 基幹理工学研究科 情報通信専攻

<sup>†</sup> Graduate School of Fundamental Science and Engineering, Waseda University

**Abstract** In recent years, various new services are proposed to improve the convenience of e-comics. In order to commercialize such services, there is need for the technique of automatically metadata extraction from comic images. In this paper we studied extraction of character, text, balloon, and panel using Faster R-CNN.

### 1. はじめに

電子コミックは電子書籍市場全体の売上の約 8 割を占める重要なコンテンツである。今後さらに電子書籍の普及が進むと予想される中、膨大なデータの中から所望する作品を得るために、メタデータを用いた高度な検索手法や、自動要約の作成といった機能が求められている。

現在の電子コミックの多くは、紙媒体のマンガをスキャンして電子化したものであり、メタデータ抽出を手動で行う必要がある。従って、作業の効率化のために、マンガ画像からメタデータを自動的に抽出する技術が必要である。本論文では、マンガの構成要素として、キャラクター、フキダシ、オノマトペを含む文字列、コマ割りの 4 種類について、Faster R-CNN による検出について検討した。

### 2. 関連研究

マンガ画像からのコマの検出について、石井らは濃度勾配の方向を利用してコマの分割線を同定する手法を提案した[1]。野中らはマンガのコマは矩形で表現されることが多いという特徴を利用して、画像内から矩形領域を検出し、コマを特定する手法を提案した[2]。

フキダシの検出について、田中らは Ada Boost を用いてページ内の文字領域を特定し、その領域をもとにフキダシの候補を検出する手法を提案した[3]。

これらの手法はコマおよびフキダシは線で囲まれた領域にあるという幾何学的特徴に基づき検出を行っている。しかし、実際のマンガにおけるコマやフキダシの表現方法は多様であり、異なる特徴を持つものに対しては検出率が低下するという問題がある。

一方、キャラクターの抽出について石井らは HOG 特徴量について機械学習を行うことでキャラクター顔領域を検出する手法を提案している[4]。我々は畳み込みニューラルネットワークを用いた検出手法である

R-CNN がキャラクター顔検出に有効であることを確認した[5]。

本研究では、R-CNN の改良手法である Faster R-CNN について、キャラクターおよびコマ、フキダシ、文字列の検出における有効性を検討し、マンガ内の多様なメタデータに対応可能な抽出手法の実現を目指す。

### 3. Faster R-CNN

畳み込みニューラルネットワーク (Convolutional Neural Network) を用いた一般物体検出手法として、Girshick らは R-CNN を提案した[6]。R-CNN の物体検出の手順は以下ようになる。初めに、Selective Search を用いて入力画像から物体の候補領域を抽出する。次に、抜き出した候補領域をそれぞれ規定の大きさにリサイズし、CNN に入力する。更に、CNN から出力された特徴量について SVM でクラス分類を行なう。最後に、矩形の座標を回帰することで候補領域の位置を補正する。R-CNN は抽出された候補領域についてそれぞれ特徴量の計算を行なうため、検出処理に時間がかかるという問題がある。Ren らは R-CNN の改良として、物体候補領域検出処理を CNN で行う手法である Faster R-CNN を提案した[7]。Faster R-CNN は候補領域の検出について、後段の CNN と計算を共通化することにより、検出処理の更なる高速化を実現した。また候補領域抽出からクラス分類にかけて End-to-end で学習を行うことで、既存手法を上回る検出精度を示している。

### 4. 実験

コマ割り、フキダシ、文字列、キャラクター顔領域について Faster R-CNN の学習を行ない、マンガ画像に対する検出精度の評価を行なった。本稿では、Mnga109 <http://www.manga109.org/index.php> より学術目的の為に使用許可を頂いたマンガを実験に使用した[8]。Faster R-CNN のアルゴリズムは、

©島田ひろかず

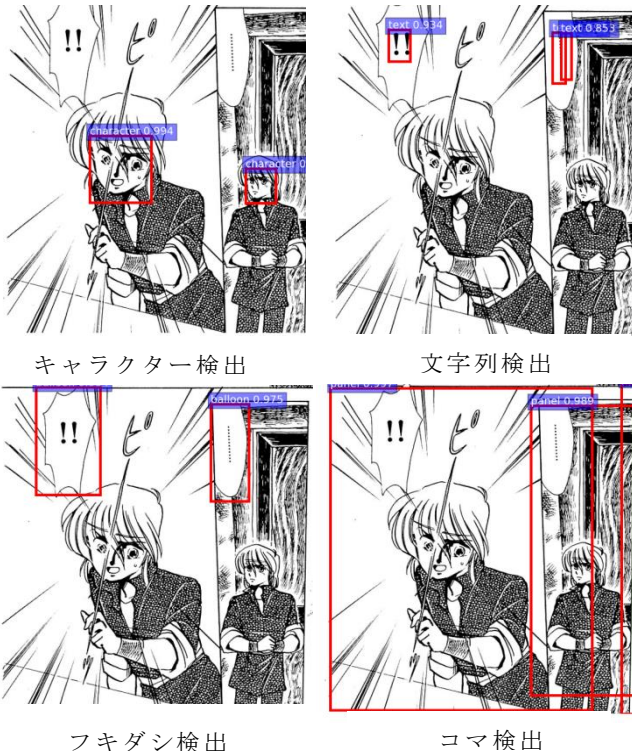


図 1: メタデータ検出例

<https://github.com/rbgirshick/py-faster-rcnn> より公開されているソースを使用し、ニューラルネットのアーキテクチャは VGG\_CNN\_M\_1024 を使用した[7]-[9].

検出器の学習では、作者の異なるマンガ 12 作品について各 100 ページを学習データセットとして使用した。4 種類のメタデータについて、個別に検出器の学習を行ない、それぞれバッチサイズは 256、学習回数は 40000 回に設定した。

検出器の評価では、学習に使用したものとは異なるマンガ 5 作品 A~E について各 100 ページをテストセットとして使用し、各作品における検出率を求めた。マンガ画像からのメタデータの検出例を図 1 に示す。図 1 において赤枠で示された領域がそれぞれ該当するメタデータとして検出された領域を示す。メタデータが検出された領域がアノテーションで指定された領域に 50%以上重なっている場合に正しく検出されたと見なし、PASCAL VOC2012 の評価手法に基づいて平均適合率(AP)を求めた[10]。各メタデータの検出率を表 1 に示す。

実験結果より、キャラクター検出は他のメタデータと比較して作品による影響を受けやすいことが確認できる。また、フキダシの背景が透過している場合や、1つのオブジェクトが複数のコマをまたいで描かれるなど学習データセット内での登場頻度が少ない表現に対しては正しく検出が行えないことが確認できた。

表 1: メタデータ検出結果(AP(%))

作品名	キャラクター	文字列	フキダシ	コマ割り
A	41.7	77.1	95.2	97.9
B	76.8	81.6	97.8	94.4
C	84.3	79.5	95.2	93.4
D	92.3	87.6	99.4	86.7
E	80.7	83.1	95.3	98.4

## 5. まとめ

本稿では、Faster R-CNN を用いたマンガメタデータ抽出について検討を行なった。実験結果より、コマ割りおよびフキダシの検出において、90%以上の検出率を確認した。今後の課題として、学習枚数の増加による検出率の変化について詳細な検討が必要である。

## 文 献

- [1] 石井大祐, 河村圭, 渡辺裕: “コミックのコマ分割処理に関する一検討”, 電子情報通信学会論文誌, Vol.J90-D, No.7, pp.1667-1670, (2007)
- [2] 野中俊一郎, 沢野拓也, 羽田典久: “コミックスキャン画像からの自動コマ検出を可能とする画像処理技術「GT-Scan」の開発”, FUJIFILM RESEARCH & DEVELOPMENT, No.57, pp.46-49, (2012)
- [3] 田中孝昌, 外山史, 宮道壽一, 東海林健二: “マンガ画像の吹き出し検出と分類”, 映像情報メディア学会誌, Vol.64, No.12, pp.1933-1939, (2010)
- [4] 石井大祐, 山崎太一, 渡辺裕: “マンガ上のキャラクター 識別に関する一検討”, 情報処理学会第 75 回全国大会 (分冊 2), pp.71-72 (2013)
- [5] 柳澤秀彰, 渡辺裕: “R-CNN を用いたマンガキャラクター検出に関する一検討”, 映像メディア処理シンポジウム, I-4-12, pp.1-2, (Nov. 2015)
- [6] R. Girshick, J. Donahue, T. Darrell, and J. Malik: “Rich feature hierarchies for accurate object detection and semantic segmentation,” in IEEE Conference on Computer Vision and Pattern Recognition, (2014)
- [7] S. Ren, K. He, R. Girshick, and J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks” Advances in Neural Information Processing System (NIPS), pp.1-9, (2015)
- [8] Y.Matsui, K.Ito, Y. Aramaki, T.Yamasaki, and K. Aizawa: “Sketch-based Manga Retrieval using Manga109 Dataset”, arXiv:1510.04389, pp.1-13, (2015)
- [9] S. Farfaded, and M. Saberian: “Multi-view Face Detection Usin Deep Convolutional Neural Networks”, arXiv:1502.02766, (2015)
- [10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman: “The PASCAL Visual Object Classes (VOC) Challenge”, IJCV, (2010)

† 早稲田大学大学院 基幹理工学研究科 情報通信専攻  
〒169-0072 東京都新宿区大久保 3-14-9 早大シルマンホール 401

TEL.03-5286-2509 E-mail: bule-cosmo@ruri.waseda.jp