

# Faster R-CNN を用いたマンガの構造解析

## Structural Analysis of Comic Images using Faster R-CNN

柳澤秀彰<sup>†</sup>      渡辺 裕<sup>†</sup>  
Hideaki Yanagisawa<sup>†</sup>      Hiroshi Watanabe<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<sup>†</sup>

**Abstract:** 電子コミック作成の効率化を目的として、マンガ画像からコマ、フキダシ、キャラクターといった構成要素を自動抽出する研究が取り組まれている。従来手法では線画の幾何学的特徴を利用してコマやフキダシを抽出している。しかし、マンガには多様な表現があることから、特定の幾何学的特徴であらゆる要素に対応することは難しい。本研究では、ニューラルネットワークを用いた機械学習による物体検出手法である Faster R-CNN によって要素を抽出することで、幅広い表現に対応可能なマンガ構成要素の実現を目指し、1 ページに含まれるコマの内容を認識する手法を提案した。実験結果より、平均で 67.5% のコマに対して内容を正確に認識できることを確認した。

## 1 研究背景

電子コミックは電子書籍市場の約 8 割を占めており、今後の出版業界における重要なコンテンツである。近年では、電子コミックの利便性の向上を目的として、マンガ内容のメタデータを用いた作品の検索やストーリー要約の生成といったサービスが提案されている。しかし、現状の電子コミックは紙媒体のマンガをスキャンしたものがほとんどであり、人間が手でメタデータの付与を行なう必要がある。従って、メタデータ付与作業の人的負担を低減するために、自動的にメタデータを取得する技術が求められている。

## 2 関連研究

マンガ画像からのコマ検出およびフキダシの検出について、従来手法では、線検出や連結した画素の抽出などの幾何学的特徴に基づく手法で要素の抽出を行なっている。しかし、マンガにおけるコマやフキダシの表現方法は多種多様であり、特定の幾何学特徴であらゆる表現に対応することは困難である。

一方、マンガに登場するキャラクターは他の要素と比較して特徴の変化が大きいため、画像特徴量を用いた機械学習による検出手法が提案されている。我々は、畳み込みニューラルネットワーク(CNN)を用いた物体検出手法の Fast R-CNN をキャラクター顔検出に適用することによって、HOG 特徴量を上回る検出率が得られることを確認した [1]。

© 佐々木あつし

© 南澤久佳



図1 コマの配置と番号付けの例

## 3 マンガ構造解析

本研究では、Fast R-CNN の改良手法である Faster R-CNN を使用してマンガ構成要素の検出を行ない、検出された領域の情報を基にコマの内容を自動的に取得する手法を提案する [2]。

Faster R-CNN の学習には、マンガ画像および画像に含まれるコマ、フキダシ、キャラクター顔領域の外接矩形を指定したアノテーションを使用する。そして、個々の要素について学習を行ない、3 種類の検出器を作成した。

続いて、コマ内容を取得する手順を示す。初めに、画像からコマの検出を行ない、検出された領域について並べ替えを行なう。並べ替えのルールとして、矩形の上辺が上側にあるコマから順に番号を付ける。また、上辺の高さが同じコマは右側にあるものから先に番

号を付ける。コマの配置と番号付けの例を図 1 に示す。このとき、Faster R-CNN によって検出されるコマの位置ずれを補正するために、上下方向 50 ピクセルごとに正規化を行なう。次に、フキダシとキャラクターの検出を行ない、検出された領域が 50%以上重なっているコマ領域に関連付ける。また、検出された領域が複数のコマ領域について重なっている場合には、番号が後ろのコマ領域に関連付ける。最後に各コマ領域について関連付けられているフキダシとキャラクターの個数を取得する。

## 4 実験

Faster R-CNN を用いたマンガ構成要素の検出率および、提案手法によるコマ内容の認識率について検討した。本実験において、学習および評価に使用するマンガ画像是研究用のマンガ画像データベースである Manga 109 において公開された画像を使用した [3]。CNN のアーキテクチャには VGG\_CNN\_M\_1024 を使用した[4]。また、学習の反復回数は 70000 回に設定した。学習には作者の異なるマンガ 20 作品から各 100 ページずつを抜き出した画像 2000 枚をデータセットとした。性能評価には学習用データセットとは異なるマンガ 5 作品 A ~ E について各 30 ページを使用した。検出率の評価において、検出された領域がアノテーションで指定した領域に 50%以上重なっている場合に正しく検出が行われたと見なす。

初めに、テスト画像 150 枚に対して、Faster R-CNN の閾値と F 値の変化を求めた。実験結果を表 1 に示す。実験結果より、コマ検出では閾値 0.6 のとき最も F 値が高くなり、フキダシとキャラクターにおいては閾値 0.8 のとき最も F 値が高くなることを確認できた。

続いて、前の実験で得られた閾値を基に、テストセットの 5 作品ごとに対するコマ内容の認識率を求めた。認識率は画像に含まれる全てのコマのうち、コマの位置およびフキダシとキャラクターの個数が正しく取得できたコマの割合とする。実験結果を表 2 に示す。認識率は最も高い作品 B で 85.5%、最も低い作品 E で 53.4%となった。コマの認識に失敗として、デフォルメされた顔や隠れの存在する顔に対して検出に失敗したことや、複雑なコマ割りのページについて正しくコマが認識されなかったことなどが挙げられる。

表 1 閾値による F 値の変化

閾値	コマ (%)	フキダシ (%)	キャラクター (%)
0.9	94.16	95.38	81.94
0.8	94.45	95.44	82.05
0.7	94.43	95.30	81.82
0.6	94.53	95.05	81.40
0.5	94.44	94.74	81.29

表 2 作品ごとのコマ内容認識率

マンガ作品	認識率 (%)
A	68.1
B	84.9
C	66.3
D	65.2
E	52.8

## 5 まとめ

本稿では、Faster R-CNN を用いたマンガ構成要素の検出とコマ内容の認識について検討した。その結果、平均で 67.5%のコマに対して正しく内容を認識できることが可能であった。

今後の課題として、本実験において正しく検出できなかったコマやキャラクターに対する検出精度の改善が挙げられる。また、自動要約生成に適したメタデータを得るために、検出されたキャラクターについて主要キャラクターの判定を行う技術が必要である。

## 参考文献

- [1] 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
- [2] S. Ren, K. He, R. Girshick, J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Advances in Neural Information Processing Systems (NIPS), 2015
- [3] Y.Matsui, K.Ito, Y. Aramaki, T.Yamasaki, K. Aizawa: “Sketch-based Manga Retrieval using Manga109 Dataset”, arXiv:1510.04389, 2015
- [4] S. Farfadi, M. Saberian: “Multi-view Face Detection Using Deep Convolutional Neural Networks”, arXiv:1502.02766, 2015

早稲田大学大学院 基幹理工学研究科 情報通信専攻  
渡辺研究室  
〒169-0072 東京都新宿区大久保 3-14-9  
早大シルマンホール 401  
Phone: 03-5286-2509, Fax: 000-111-2223  
E-mail: bule-cosmo@ruri.waseda.jp