

早稲田大学大学院 基幹理工学研究科

# 博士論文概要

## 論文題目

ストーリー理解を目的とした漫画オブジェクトの抽出

Manga Object Extraction towards Story Understanding

申請者

柳澤	秀彰
Hideaki	YANAGISAWA

情報理工・情報通信専攻 オーディオビジュアル情報処理研究

2018年12月

スマートフォンやタブレット等のデジタル端末の普及に伴い、電子書籍はより身近なコンテンツとなっている。出版科学研究所の発表した2017年度の漫画市場規模は、電子コミックの売上が紙媒体の漫画の売上を上回ったと報告している。このことから、ユーザの利用形態が従来の紙媒体の漫画から電子コミックへと移行していることがうかがえる。電子コミックはデジタル端末上でのリアルタイム処理が可能であることから、高い拡張性と応用性を有している。しかし、現状の電子コミックの多くは単純に紙媒体をスキャンすることで電子データ化された静的なコンテンツであるため、電子コミックの可能性を十分に活用できる状態にない。したがって、漫画画像から内容を認識することによって、メタデータを取得する技術が必要となる。このような電子コミックをより活用する技術や、その応用可能性を追求する研究は「コミック工学」と称され、国内外で研究が行われている。

漫画画像から内容を認識する工程は大きく以下の3段階に分けられる。1)「オブジェクト検出」：画像から「コマ」、「フキダシ」、「登場キャラクター」などの構成要素を検出する、2)「シーン理解」：検出された要素からキャラクター名や台詞の内容など詳細な情報を認識する、3)「ストーリー理解」：ページごとのオブジェクト情報を統合してコマの順序や登場キャラクターの関連付けを行う。本研究では、「オブジェクト検出」と「シーン理解」の工程に着目し、漫画画像からのオブジェクト検出手法およびキャラクターの同定手法を提案する。

2章では、従来手法の改良によるキャラクター顔検出の高性能化を提案した。漫画画像の解析における課題として、漫画は線画を主体とした2値の画像で表されることから、自然画像と比べて利用できる画像特徴が限定されていることがある。またもう一つの課題として、漫画オブジェクトは場面に応じて様々な形態で表現させるため、1種類のオブジェクトを包括して認識することが難しいということがある。特にキャラクターは種類ごとの特徴変化が大きく、漫画に特有なデフォルメ表現をもって描かれることから、同一のキャラクター間でも形状の変動が大きい。従来手法では、画像の輝度勾配の変化を捉えるHOG特徴量の利用を提案しているが、実用上において誤認識が多いという問題があった。我々は、従来手法の検出率低下の原因として、手描きで表現される漫画キャラクターは場面ごとにパーツ位置に変動があるためであると推測した。したがって、キャラクター顔検出の高性能化のために、物体のパーツ位置の変動に頑強な物体検出モデルであるDeformable Part Model (DPM)の適用を提案した。はじめに、キャラクター検出におけるDPMのパートモデル導入の有効性を評価するため、正面向きのキャラクター顔画像を対象としてルートフィルタのみを用いた検出モデルと通常のDPMとの比較を行った。実験結果では、パートモデルの導入によって5.2ポイントの性能向上が示された。次に、キャラクターのマルチビュー顔検出に最適化した検出モデルを求めるために、DPMの各パラメータを変化させたモデルの比較を行っ

た。実験結果では、ルートフィルタのコンポーネント数を 10, パートフィルタ数を 11 と設定したモデルにおいて、最大に 77.6%の AP が示された。検出モデルの変化における考察では、多様なキャラクタの顔画像に対応するために多数のフィルタの生成が必要であることが分かった。このことから、キャラクタ顔検出において DPM の高性能化を考えたとき、複雑な検出モデルが必要になるという課題が明らかになった。

3章では、畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を用いた漫画オブジェクト検出手法を提案した。2章より、HOG のような従来の特徴量は特徴表現が既定であることから多様な形状を持つ漫画オブジェクトに対応することが困難であるという問題が示された。ここで、近年の画像認識に用いられる CNN は学習の過程で入力画像の認識に最適な特徴量を自動的に生成するため、人間の事前知識に基づいた特徴量の設計を必要としないという利点を持っている。ただし、CNN による特徴量の学習過程はブラックボックスであることから、一般物体とは異なる特徴を持つ漫画画像への有効性を検証した。はじめに、CNN を用いた物体検出手法である fast R-CNN と DPM との比較から、キャラクタのマルチビュー顔検出における CNN の有効性を示した。次に、検出の対象をキャラクタ以外に「コマ」や「フキダシ」といった要素に広げることで、個々のコマが含むオブジェクト情報を認識するアルゴリズムを提案した。物体検出器に faster R-CNN を用いて実験を行い、コマやフキダシのような不定形なオブジェクトに対して 90%以上の精度で検出が可能であることを確認した。また、CNN の物体検出の性能は入力画像からの候補領域の切り出しに影響されることから、異なる切り出し方法を採用している fast R-CNN, faster R-CNN, SSD の三つの手法について漫画画像に対する有効性の比較を行った。実験結果では、fast R-CNN はキャラクタや文字列など境界線が明確でないオブジェクトに対して検出率が低下することが分かった。また、SSD は入力画像に対する比率が小さなオブジェクトに対して検出率が低下する傾向が見られた。このことから、漫画オブジェクトの検出において faster R-CNN の候補領域切り出しアルゴリズムが有効であることを示した。

4章では、漫画キャラクタ顔画像から主要キャラクタを抽出する手法を提案した。登場キャラクタのメタデータを取得するためには、キャラクタの位置を検出した後に、それがどのキャラクタであるか同定する作業が必要となる。このとき、未知の漫画画像を対象とする場合には教師なしでのキャラクタ分類が必要となる。従来研究では、キャラクタ顔画像の SURF 特徴量を k-means 法でクラスタリングする手法が提案されている。しかし、k-means 法の実行には生成するクラスタ数を事前に決定する必要がある。このとき、登場キャラクタの数が未知の画像に対してはクラスタ数を決め打ちで決定することになるため、分類精度が低下するという問題がある。この問題を解決するために、キャラクタ顔画像を

x-means 法でクラスタリングする手法を提案した。x-means 法は、ベイズ情報量基準 (Bayesian Information Criterion: BIC) による停止基準を満たすまで k-means 法によるクラスタの分割を逐次繰り返す手法である。通常の x-means による処理は完全に自動化されているが、理想的なクラスタリング結果を得るために、新たに BIC に任意のパラメータを設定することで、適切なクラスタリング結果を得ることができるか検討を行った。従来手法との比較では、最適な BIC のパラメータが設定されているとき、提案手法が従来手法を上回ることを確認した。このことより、クラスタ数を自動決定するクラスタリングアルゴリズムの適用によって、顔画像分類精度の高性能化が可能であることを示した。

5章では、4章で提案したクラスタリング手法の問題点を改良することで、主要キャラクタを分類する手法を提案した。4章における提案手法の問題として、SURF 特徴量による特徴記述は局所的なテクスチャの類似した画像に対して誤分類が発生する点が挙げられる。この問題について、大規模データセットを学習した CNN は物体認識において汎化性能を持つことから、CNN の中間出力を特徴量に使用することを考えた。もう一つの問題は、x-means 法は全てのデータが特定のクラスタに属するよう分類を行うが、登場回数の少ないキャラクタを独自のクラスタとして分離することが困難であるため、それらの顔画像がノイズとなってクラスタの純度を低下させる点がある。これを解決するため、類似データのみをクラスタとして抽出し、外れ値を許容するクラスタリングアルゴリズムである Density-Based Spatial Clustering of Applications with Noise (DBSCAN) の適用を考えた。はじめに、一般 CNN モデルと DBSCAN を用いたクラスタリング実験では、ノイズ除去における DBSCAN の有効性を確認し、また DBSCAN の適用において入力データの局所的な類似度に基づいた次元削減が有効であるという知見を得た。次に、分類精度を向上させるための試みとして、キャラクタ顔画像で学習を行った CNN の適用と、キャラクタ顔画像からの背景領域の除去について検討を行った。最後に、これまでの実験を基にキャラクタ画像を事前学習した CNN モデルと DBSCAN によるキャラクタ顔画像の分類手法を提案し、さらに DBSCAN のパラメータの自動決定方法について検討した。実験結果では、NMI において最大で 49.5% の分類精度を示し、分類の自動化では 45.7% の精度を達成した。

本研究では、漫画のストーリー理解を目的として、漫画オブジェクトを認識する手法の提案とその改善を行った。本研究の成果として、多様な漫画オブジェクトの検出において CNN を用いた物体検出が有効性を持つことを確認した。この技術はキャラクタの表情やフキダシの種類など、オブジェクトのより詳細な認識へと応用が期待できる。また、事前知識の存在しない漫画画像を対象とした、キャラクタの自動分類について基礎となる技術の構築を行った。

## 早稲田大学 博士（工学） 学位申請 研究業績書

氏名 柳澤秀彰 印

(2019 年 1 月 現在)

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
論文	<p>[1]. ○<u>柳澤秀彰</u>, 山下拓朗, 渡辺裕, “主要キャラクターの抽出を目的とした漫画キャラクター画像のクラスタリング”, 映像情報メディア学会誌, vol.73, No.1, pp.199-204, Jan. 2019.</p> <p>[2]. ○<u>H. Yanagisawa</u>, D. Ishii and H. Watanabe, “Face Detection for Comic Images Using the Deformable Part Model”, IIEEJ Transactions on Image Electronics and Visual Computing, Vol.4, No.2, pp.95-100, Dec. 2016.</p>
国際会議 (査読付き)	<p>[1]. ○<u>H. Yanagisawa</u>, T. Yamashita, H. Watanabe, “Manga Character Clustering with DBSCAN using Fine-Tuned CNN Model”, International Workshop on Advanced Image Technology and International Forum on Medical Imaging in Asia (IWAIT-IFMIA 2019), No.202, Jan. 2019.</p> <p>[2]. ○<u>H. Yanagisawa</u>, T. Yamashita, H. Watanabe, “A Study on Object Detection Method from Manga Images Using CNN”, International Workshop on Advanced Image Technology (IWAIT2018), No.16, pp.1-4, Jan. 2018.</p> <p>[3]. ○<u>H. Yanagisawa</u>, H. Watanabe, “Recognition of Panel Structure in Comic Images Using Faster R-CNN”, International Workshop on Image Electronics and Visual Computing 2017 (IEVC2017), No. 4C-2, Mar. 2017.</p> <p>[4]. ○<u>H. Yanagisawa</u>, D. Ishii, H. Watanabe, “Face detection for comic images with deformable part model”, The 4th IIEEJ International Workshop on Image Electronics and Visual Computing 2014 (IEVC2014), 4A-1, Oct. 2014.</p>
国内研究会	<p>[1]. <u>柳澤秀彰</u>, 山下拓朗, 渡辺裕, “畳込みニューラルネットによるマンガオブジェクト認識メカニズムの一検討”, 電子情報通信学会, パターン認識・メディア理解研究会, PRMU2017-79, Oct. 2017.</p> <p>[2]. 石井大祐, <u>柳澤秀彰</u>, 三原鉄也, 永森光晴, 渡辺裕, “マンガの構成要素に基づく自動シーン分割処理に関する一検討”, 情報処理学会 AVM 研究会研究報告, Vol.2014-AVM87, No.15, pp.1-4, Dec. 2014.</p>
シンポジウム	<p>[1]. 山下拓朗, <u>柳澤秀彰</u>, 渡辺裕, “深層学習福笑い”, 2017年画像符号化シンポジウム・2017年映像メディア処理シンポジウム (PCSJ/IMPS2017), P-5-2, Nov. 2017.</p> <p>[2]. <u>柳澤秀彰</u>, 渡辺裕, “CNNを用いたマンガオブジェクト検出手法の比較”, 2017年画像符号化シンポジウム・2017年映像メディア処理シンポジウム (PCSJ/IMPS2017), P-1-10, Nov. 2017.</p>

## 早稲田大学 博士（工学） 学位申請 研究業績書

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
シンポジウム	<p>[3]. <b>柳澤秀彰</b>, 渡辺裕, “Faster R-CNN を用いたマンガ画像の構造解析“, 2016 年画像符号化シンポジウム・2016 年映像メディア処理シンポジウム (PCSJ/IMPS 2016), No. P-2-10, Sept. 2016.</p> <p>[4]. <b>柳澤秀彰</b>, 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 (IMPS), I-4-12, ラフォーレ修善寺, pp.1-2, Nov. 2015.</p>
国内学会	<p>[1]. <b>柳澤秀彰</b>, 山下拓朗, 渡辺裕, “CNN を用いた漫画キャラクター顔画像クラスタリングシステムの改良に関する一検討“, 映像情報メディア学会年次大会, 12B-2, Aug. 2018.</p> <p>[2]. 山下拓朗, <b>柳澤秀彰</b>, 渡辺裕, “深層学習を用いたマンガキャラクターの検出における顔変形の影響評価“, 情報処理学会全国大会, 5Y-03, Mar. 2018.</p> <p>[3]. <b>柳澤秀彰</b>, 山下拓朗, 渡辺裕, “マンガキャラクター顔画像クラスタリングの改良における一検討“, 映像情報メディア学会冬季大会, 22B-8, Dec. 2017.</p> <p>[4]. K. J. Ngeno, <b>H. Yanagisawa</b>, H. Watanabe: “Ship Classification Using Faster Region Convolution Neural Network (Faster R-CNN) for Automatic Identification of Marine Vessels”, FIT2017(第 16 回科学技術フォーラム), H-039, Sept. 2017.</p> <p>[5]. <b>柳澤秀彰</b>, 渡辺裕, “Deep Learning 特徴量を用いたマンガキャラクター顔画像の分類“, FIT2017(第 16 回科学技術フォーラム), H-001, Sept. 2017.</p> <p>[6]. <b>柳澤秀彰</b>, 渡辺裕, “X-means 法を用いたマンガキャラクターの自動分類に関する検討“, IEICE 総合大会, D-12-40, Mar. 2017.</p> <p>[7]. <b>柳澤秀彰</b>, 渡辺裕, “Faster R-CNN を用いたマンガ画像からのメタデータ抽出“, 映像情報メディア学会年次大会, No.14B-1, Sept. 2016.</p> <p>[8]. <b>柳澤秀彰</b>, 渡辺裕, “マンガキャラクターのマルチビュー顔検出に関する検討“, 電子情報通信学会総合大会, D-11-12, Mar. 2016.</p> <p>[9]. <b>柳澤秀彰</b>, 渡辺裕, “マンガキャラクター検出における学習画像枚数の影響“, 映像情報メディア学会冬季大会, 23B-5, Dec. 2015.</p> <p>[10]. <b>柳澤秀彰</b>, 石井大祐, 渡辺裕, “マンガの複数キャラクターに対する顔検出率について“, 電子情報通信学会総合大会, D-12-31, Mar. 2015.</p> <p>[11]. 陳明, <b>柳澤秀彰</b>, 張傑, 石井大祐, 渡辺裕, “マンガにおける HOG+AdaBoost による顔画像検出の性能評価“, 映像情報メディア学会年次大会, 17-4, Sept. 2014.</p>

## 早稲田大学 博士（工学） 学位申請 研究業績書

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
国内学会	<p>[12]. <u>柳澤秀彰</u>, 石井大祐, 陳明, 渡辺裕, “マンガ画像からの顔検出におけるパーツ特微量の一検討”, 映像情報メディア学会年次大会, 17-9, Sept. 2014.</p> <p>[13]. M. Chen, <u>H. Yanagisawa</u>, D. Ishii, H. Watanabe: “A Note on Face Detection of Comic Image with Different Background,” 映像情報メディア学会年次大会, 17-9, Sept. 2014.</p>