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

2019 年 2 月

柳澤　秀彰

Hideaki YANAGISAWA

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

2019 年 2 月

早稲田大学大学院　基幹理工学研究科

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

柳澤　秀彰

Hideaki YANAGISAWA

目次

第 1 章

序論

1.1　研究の背景

スマートフォンやタブレットなどのディジタル端末の普及によって，電子書籍はより身近なコンテンツとなっている．出版科学研究所より発表された 2017 年度の漫画市場規模調査では，電子媒体での漫画単行本の売上は前年比 17.2%増となる 1,711 億円を記録したと報告している[1]．これは，2017 年度の紙媒体の単行本の売上である 1,666 億円を上回る金額である．このことから，ユーザの漫画の利用形態が，紙媒体から電子媒体へと移行していることが確認できる．電子コミックは画像を含んだ構造化文章とできることから，拡張性と応用性に秀でるという利点を持つ．例として，従来の漫画の枠にとらわれない表現方法の創出や，ユーザの環境に最適化した購読補助機能の提供といった様々なサービスの提供が可能である．しかし，現在普及している電子コミックの多くは単に紙媒体の漫画をスキャンして電子化したコンテンツであり，電子コミックの可能性を十分に活用できる状態にはない．そこで，漫画画像にメタデータのタグ付けを行うことによる電子コミックの構造化が提案されている．このとき，メタデータ抽出の作業を効率化するために，漫画画像から内容を認識する技術が必要となる．電子コミックをより活用するための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，世界的に研究が行われている[2].

1.2　漫画画像の特徴

本研究では，日本で一般的に流通している漫画作品を認識の対象とする．漫画画像の例を図 1-1 に示す．漫画はコマと呼ばれる枠によって区切られた領域の中に「キャラクタ・背景・フキダシ・セリフ・オノマトペ」といった要素を描き込むことで構成される．人物のセリフや思考はフキダシと呼ばれる枠中に文字で書かれ，フキダシの形状や文字の書体によって語調を表す．擬音語・擬態語は，手書きの書き文字として絵の中に書かれることが多く，また細々としたセリフなども書き文字で書かれることがある． 雑誌や単行本として刊行される漫画はカラーよりもモノクロのものが多い．このようなモノクロの漫画画像は，白黒の 2 値からなる線画と，ベタと呼ばれる黒く塗りつぶされた領域，スクリーントーンと呼ばれる一定のパターンが印刷された領域から構成される．本

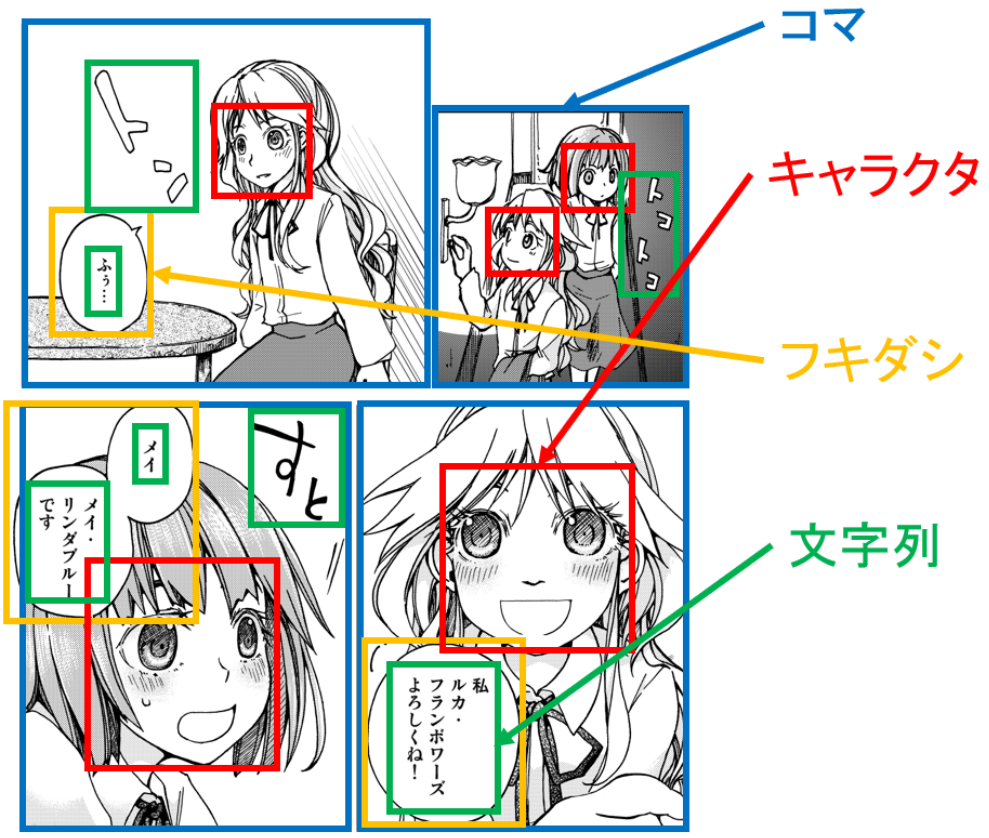


図 1-1: 漫画画像の構造例（漫画画像は文献[3]より著者の許可を得て抜粋）

研究では，セリフとオノマトペの両方を「文字列」として扱い，漫画の主要な構成要素である「コマ・キャラクタ・フキダシ・文字列」のことを総称して「漫画オブジェクト」と定義する．漫画画像と自然画像との違いとして，漫画画像では陰影の変化が省略されるため，画素間の輝度変化が大きい領域(エッジ成分)と輝度がほとんど変化しない平坦な領域が多いことが挙げられる．また，漫画は表現上の自由度が大きいことから，漫画オブジェクトは場面ごとにさまざまな形態で描かれるという特徴がある．

1.3　漫画画像からの内容理解

漫画画像からの内容理解は，図 1-2 に示すような三つの工程に分けることができる．まず，「オブジェクトの検出」では漫画 1 ページから漫画オブジェクトを検出し，ページ内のコマの配置やコマに含まれるキャラクタ・フキダシ等の情報を取得する．次に，「シーン理解」では検出されたオブジェクトに対して詳細な情報の解析を行う．例としては，キャラクタ名の同定や文字列の内容の認識などが挙げられる．最後に，「ストーリー理解」では複数のページにおけるオブジェクト情報を統合し，構造化することで，全体的な漫画内容の理解を行う．例としては，コマ順序の認識や，登場キャラクタの

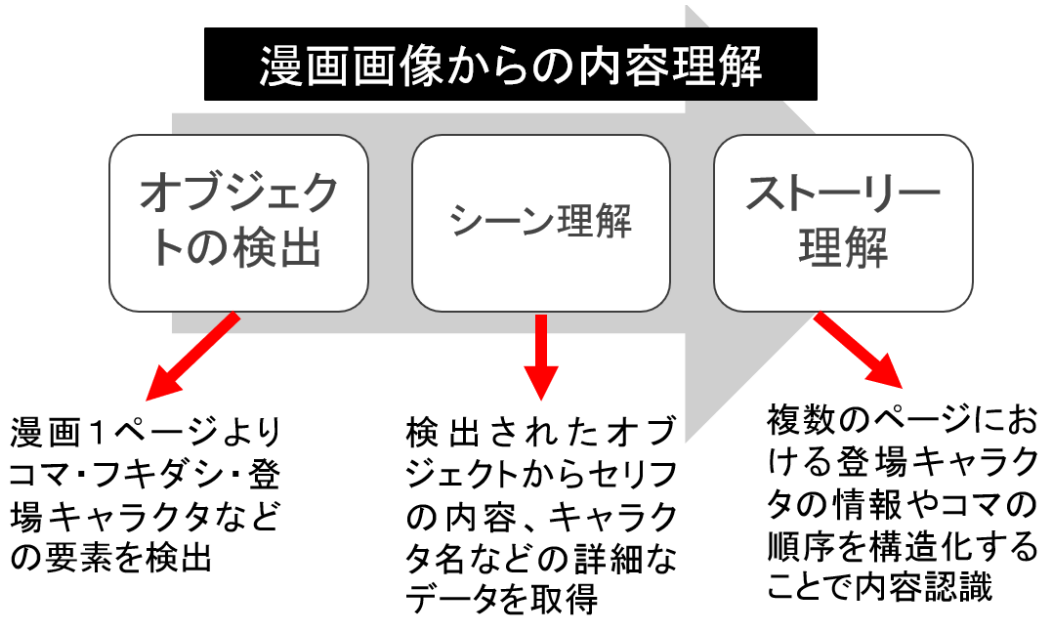


図 1-2: 漫画画像からの内容理解の工程

関連付けなどが挙げられる．漫画内容を理解するための基礎技術として「オブジェクトの検出」と「シーン理解」の工程に着目したとき，それぞれ以下のような課題が存在する．

1.3.1 漫画オブジェクトの検出における課題

漫画オブジェクトを対象とした検出について，以下のような研究が行われている．キャラクタの検出において, 新井らや石井らは Haar-Like や Histograms of Oriented Gradients (HOG)といった局所特徴量を手掛かりに画像から顔領域のマッチングを行う手法を提案している[4, 5]．次に，コマの検出において，野中らは「コマは矩形領域で表現されることが多い」という知見に基づいてページ内から矩形領域を検出することでコマを特定する手法を提案している[6]．さらに，フキダシの検出において，田中らは「フキダシは文字列を内包する」というルールに基づき，AdaBoost で文字領域を検出してその周辺領域について Support Vector Machine (SVM)による認識を行うことでフキダシ領域を検出する手法を提案している[7]．また，Arai らは Blob 検出によって，画素の連結した領域より一定以上の大きさのものを抽出することで, コマとフキダシを同時に検出している[8]．これらの従来手法に共通した課題として, 画像の幾何学的な解析に基づいてオブジェクトを検出するため，特異な特徴を持つオブジェクト(e.g. 極端にデフォルメされたキャラクタ，コマの枠線に他のオブジェクトがオーバラップした表現)に対して検出が困難であることが挙げられる．したがって，高精度な漫画オブジェクト検出を実現す

るために，漫画表現の多様性に対応可能な検出システムの構築が必要となる．

1.3.2　漫画のシーン理解における課題

漫画のシーン理解における課題の一つに，キャラクタの認識がある．漫画のストーリーを理解するためには，それぞれのコマの中にどのキャラクタが登場しているかという情報が必要である．このとき，対象の漫画作品について事前知識を持たない場合には，漫画内に登場するキャラクタの総数といった情報を利用できないという問題がある．したがって，漫画オブジェクト検出からシーン理解までの工程を自動化するために，正解ラベルを使用しない教師なしでのキャラクタの認識技術が必要となる．

1.4　本論文の目的

本論文では，漫画画像の解析によるストーリー理解を目的として漫画オブジェクト情報の抽出について研究を行う．具体的な内容としては，画像からの漫画メタデータの検出及び，登場キャラクタの認識について述べる．

先述のように従来の漫画オブジェクト検出では，場面ごとのオブジェクトの形状変化を捉えることができないという問題がある．この問題を解決するために，物体のパーツの位置変動に頑強な物体検出手法である Deformable Part Model (DPM)や，画像特徴を自動的に生成する手法である Convolutional Neural Network (CNN)の適用を提案し，漫画オブジェクト検出精度の向上について検討する．

また，未知の漫画画像を対象に教師なしでのキャラクタの認識を行う手法として，クラスタリングを用いて，キャラクタ画像を類似した画像ごとに分類する方法が考えられる．ただし，クラスタリングを行うためには最適なクラスタ数を設定する必要がある．本研究では，クラスタ数を自動的に決定するクラスタリング手法である x-means 法や Density-Based Spatial Clustering of Applications with Noise (DBSCAN)の適用を検討することで，キャラクタ分類の自動化を目指す．

1.5　本論文の構成

本章以降の構成を述べ，漫画内容の理解におけるそれぞれの位置づけを図 1-3 に示す．

第 1 章「序論」は本章であり，本論文の背景と課題，研究の目的について述べた．

第 2 章「局所特徴量を用いたキャラクタ検出」では，従来手法を改良したキャラクタ検出手法として，DPM の適用を試みる．はじめに，DPM の概要について述べる．次に，従来の顔検出モデルと通常の DPM についてキャラクタ顔画像の検出精度を比較することで，キャラクタ顔検出における DPM の有効性を評価する．さらに， 漫画キャラクタのマルチビュー顔検出において最適な DPM のパラメータ設定を求めるため，パラメータの変化が検出率に及ぼす影響について調査する．最後に，実験結果に対する考察から，DPM を用いたキャラクタ検出における課題を明らかにする．

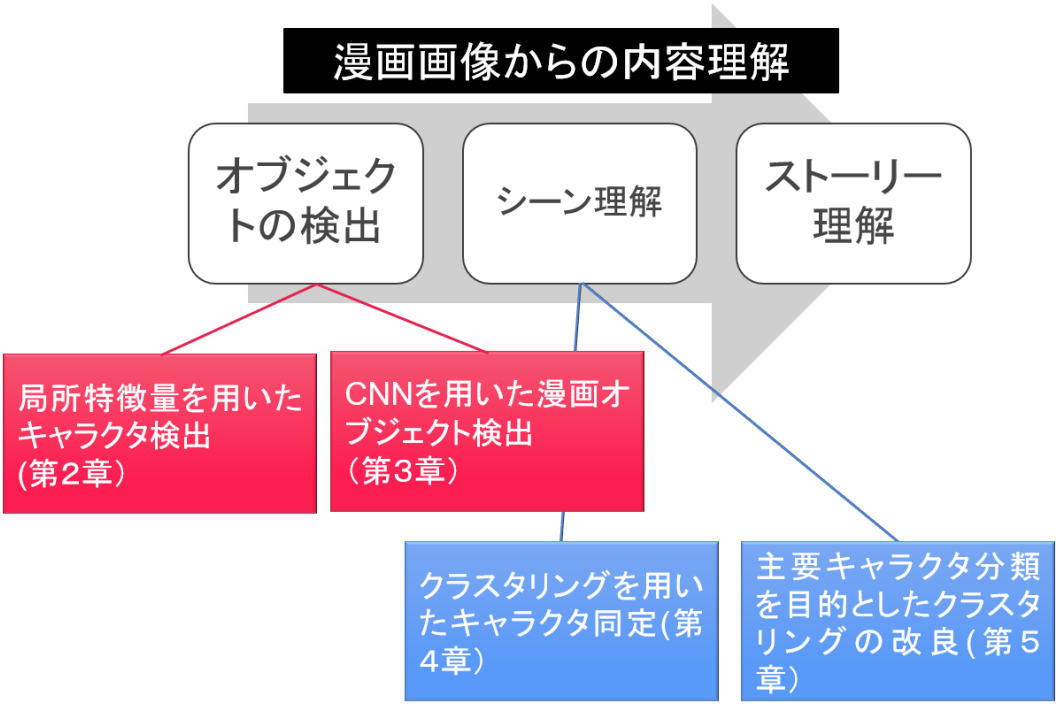


図 1-3: 本論文の構成

第 3 章「CNN を用いた漫画オブジェクトの検出」では, 第 2 章において示された DPM の問題点を改良した手法として，CNN を用いた漫画オブジェクト検出を提案する．はじめに，CNN の概要について述べる．次に，CNN を用いた物体検出手法である Regions with CNN features (R-CNN)のアルゴリズムについて述べる．そして，R-CNN を改良した物体検出手法について述べる．次に，キャラクタ顔検出における Fast R-CNN と DPM の比較から，CNN の有効性を評価する．さらに，キャラクタに加えてコマとフキダシを検出対象とすることで，コマ内容を認識するアルゴリズムについて評価する．最後に，CNN を用いた物体検出手法である Fast R-CNN，Faster R-CNN，Single Shot Multibox Detector (SSD)の 3 種類について，漫画オブジェクト検出の精度を比較することで，漫画画像に有効な候補領域の抽出方法を求める．

第 4 章「クラスタリングを用いたキャラクタ同定」では，教師なしでの主要キャラクタの同定を目的として，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する．はじめに，キャラクタ同定における従来手法の概要を述べ，問題点を明らかにする．次に，提案手法の概要を述べる．そして，主要キャラクタの同定における従来手法と提案手法の比較から，提案手法の有効性を評価する．最後に，キャラクタ顔画像の分類における課題を明らかにする．

第 5 章「主要キャラクタ分類を目的としたクラスタリングの改良」では，第 4 章のクラスタリング手法を改良することで, 主要キャラクタ顔画像を複数のクラスタに分類する手法を提案する．はじめに，CNN を用いたキャラクタ顔画像からの特徴抽出につい

て述べる．次に，特徴量の次元削減について述べる．さらに，クラスタリング手法である DBSCAN の概要と，そのパラメータの決定方法について述べる．そして，大規模データセットを学習した CNN モデルと DBSCAN を使用した主要キャラクタの抽出手法を評価する．さらに，顔画像の分類精度を向上させる試みとして，画像処理による背景除去がキャラクタ分類に及ぼす影響について検討する．最後に，キャラクタ顔画像を学習した CNN モデルと DBSCAN を用いた主要キャラクタの分類について評価する．

第 6 章「結論」では，本論文における研究成果を総括し，結論を述べる．

第 2 章

局所特徴量を用いたキャラクタ検出

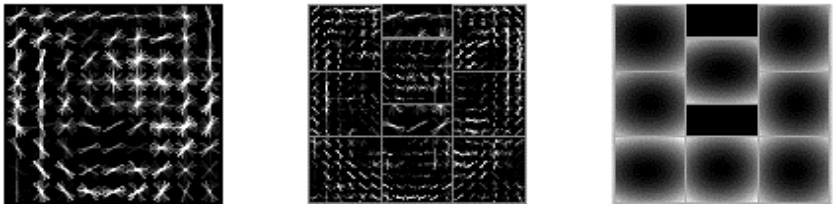
2.1 まえがき

漫画画像は 2 値の線画を主体として構成されることや, 独自のデフォルメ表現を含むといった理由から，一般物体と同様の手法で認識を行うことが困難であり，独自な特徴を持つ対象となっている．従来のキャラクタ顔検出では，漫画画像にエッジ成分が多く含まれるという特徴に着目して，Histograms of Oriented Gradients (HOG)等の局所特徴量を用いた検出を行っている．しかし，顔画像の形状変化に対応できず，十分な検出精度は得られていない．本章では，キャラクタ顔画像の検出精度の向上を目指して，Deformable Part Model (DPM)の適用を試みる．

まず，提案手法で使用する DPM について概要を述べる．次に，従来手法と DPM の比較からキャラクタ顔検出における DPM の有効性を評価する．さらに，キャラクタのマルチビュー顔検出における DPM の最適なパラメータ設定を求めるため，DPM のパラメータの変化が検出率に及ぼす影響について調査する．最後に実験結果の考察より，キャラクタ検出における DPM の課題を明らかにする．

2.2 Deformable Part Model

DPM は 2008 年に Felzenszalb らによって提案された物体検出手法である[9, 10]．DPM の基本概念は，従来のキャラクタ検出手法と同様に，物体の形状を HOG 特徴量で表現し，Support Vector Machine (SVM)を用いて対象物体の形状を捉えるフィルタを学習するものである．ただし，DPM の検出モデルは物体の全体の形状を捉える「ルートフィルタ」と，物体の各パーツの形状を捉える複数の「パートフィルタ」の 2 種類のフィルタによって構成される．DPM の物体検出モデルの例を図 2-1 に示す．DPM は対象物体を複数のパーツを持つモデルとして表現し，全体及び各パーツの形状の妥当性と，パーツの相対的な位置関係から，検出スコアを求める．また，DPM は画像を物体のアスペクト比によって分類し, アスペクト比の異なる物体に対応した検出モデルを学習することが可能である. これによって生成された複数のコンポーネントからなる検出モデルを使用することで，物体の角度の変化に対応できる．以下に DPM の詳細を述べる．



ルートフィルタ　　パートフィルタ　　ルートフィルタ上のパートフィルタの配置

図 2-1: DPM の物体検出モデルの例

2.2.1　Histograms of Oriented Gradients (HOG)

HOG 特徴量は，人物検出を目的として 2005 年に Dalal らによって提案された画像特徴量である[11]．この特徴量は画像の局所領域の輝度の勾配方向をヒストグラム化したものであり，幾何学的変換に強く，照明の変動に頑健であるという利点を持つ．HOG 特徴量の計算過程は以下のようになる．

1. 画像の各ピクセルから輝度の勾配方向と勾配強度を算出する．
2. セル領域ごとにヒストグラムを求める．
3. ブロック領域ごとに正規化し，特徴量を抽出する．

1 ~ 3 の処理の詳細について，以下で説明する．

1. 輝度の勾配方向と勾配強度の算出

グレースケール画像の各ピクセルの輝度値から勾配方向と強度を算出する．x, y をあるピクセルの座標として，x軸方向とy軸方向におけるピクセルの輝度変化の値をそれぞれ$f_x(x,y)$，$f_y(x,y)$とする．この値を用いて，輝度変化の勾配方向$\theta(x,y)$および勾配の強度$m(x,y)$を式(2.1)，式(2.2)より求める．

$$\theta(x,y) = \tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2.1}$$

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \tag{2.2}$$

2. ヒストグラムの作成

複数のピクセルからなるセル領域を設定し，勾配方向の量子化を行うことで一つのセルにおける勾配強度のヒストグラムを作成する．本研究では，1 セルを 8×8 ピクセルからなる領域と設定し，勾配方向を 0°から 160°にかけて 20°刻みで 9 方向に量子化した勾配強度ヒストグラムを求める．

3. ブロック領域での正規化

複数のセルからなるブロック領域を設定し, ブロックごとに勾配強度の正規化を行うことで特徴量を求める．本研究では，1 ブロックを 2×2 セルからなる領域と定める．したがって，n番目の勾配方向ヒストグラムを$v(n)$としたとき，正規化の計算は式(2.3)のようになる．

$$v(n) = \frac{v(n)}{\sqrt{(\sum_{k=1}^{2\times2\times9} v(k)^2) + 1}} \tag{2.3}$$

2.2.2 HOG ピラミッド

DPM は物体のスケール変化に対応するために，解像度の異なる HOG 特徴マップの集合から物体検出を行う．まず，1 枚の入力画像から解像度の異なる画像の集合である画像ピラミッドを作成する．次に，画像ピラミッド内のそれぞれのレベルの画像について HOG 特徴量を計算することで HOG ピラミッドを求める．HOG ピラミッドの上層ではスケールの小さい画像によって大域的な荒い形状が表現され, 下層ではスケールの大きい画像によって局所的な細かい形状が表現される．

2.2.3 フィルタ

入力された HOG 特徴に対する重みフィルタ F は，w×h×9×4 個のベクトルである．ここで，w と h は検出ウィンドウの高さと横幅を表す．HOG ピラミッドにおける画素の位置を表すベクトルを$p = (x, y, l)$とおく．lは HOG ピラミッドの解像度のレベルを示す. HOG ピラミッド H のある位置pにおいて, w×h ピクセルからなるブロック内の HOG 特徴量を$\phi(H, p, w, h)$と示す．検出ウィンドウにおけるフィルタ F のスコアは，重みを持ったベクトルと特徴量の内積であり， $F \cdot \phi(H, p, w, h) = F \cdot \phi(H, p)$と表される．

2.2.4 可変パーツ

DPM の検出モデルにおいて，ルートフィルタは検出ウィンドウと同等と定義する．パートフィルタにおける HOG 特徴量のセルのサイズは，ルートフィルタにおけるセルのサイズの半分に設定する. これは, ルートフィルタが画像の大域的なエッジに着目し, パートフィルタでは各パーツの局所的な特徴に着目することで, 物体の姿勢変化を吸収できるという考えに基づく．n個のパーツから構成される物体のモデルをルートフィルタF_0及び，パートモデル$(P_1, \cdots, P_n)$から表す．このとき，各パートモデルのパラメータは$P_i = (F_i, v_i, s_i, a_i, b_i)$と表される．ここで，$F_i$は$i$番目のパートフィルタ，$v_i$はルートフィルタを基準としたときの$i$番目のパートフィルタの中心座標のデフォルトの相対位置を示す 2 次元ベクトル，s_iはi番目のパートフィルタの中心点を定める際の許容範囲を

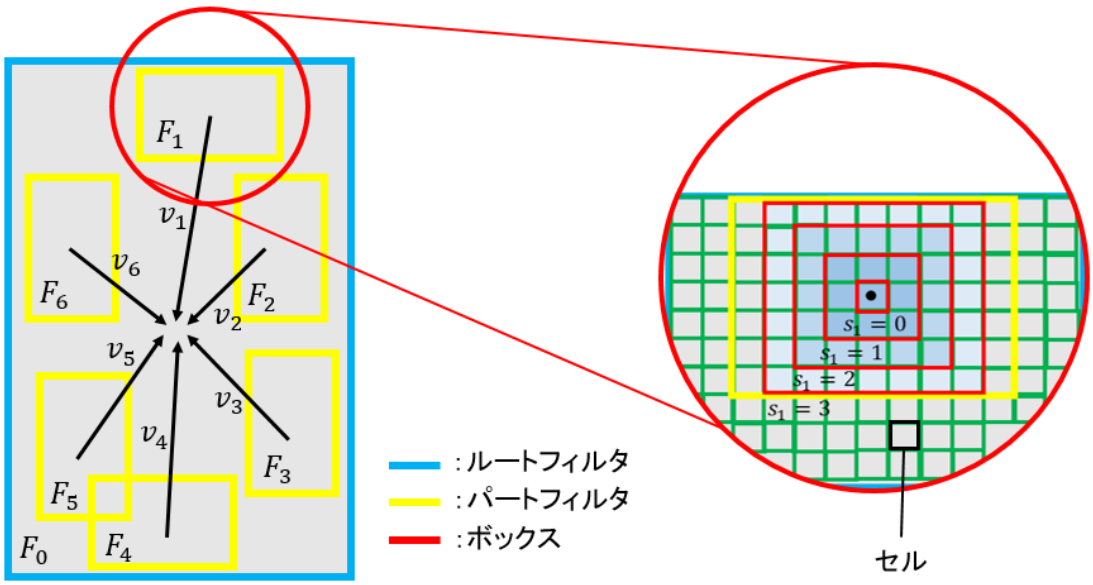


図 2-2: パートモデルの概要（自発表[12]より引用）

表すボックスのサイズ，a_i, b_iは，i番目のパートフィルタにおける 2 次元ベクトルによる係数を表す．パートモデルの概要を図 2-2 に示す．パートモデルを評価する潜在変数を$z = (p_1, \cdots, p_n)$とする．zのスコアは，各フィルタのスコアと，パートフィルタとルートフィルタの位置関係より，式(2.4)，式(2.5)で与えられる．

$$score(z) = \sum_{i=0}^{n} F_i \cdot \phi(H, p_i) - \sum_{i=1}^{n} a_i \cdot (\tilde{x}_i, \tilde{y}_i) + b_i \cdot (\tilde{x}_i^2, \tilde{y}_i^2) \tag{2.4}$$

$$(\tilde{x}_i, \tilde{y}_i) = ((x_i, y_i) - 2(x, y) + v_i)/s_i \tag{2.5}$$

式(2.5)において，第 1 項は，各パートフィルタの重みと，HOG 特徴ベクトルの内積をとったフィルタのスコアの合計を表し，第 2 項はパートフィルタとルートフィルタの相対的な位置関係と距離を表す．式(2.5)はパートフィルタの配置のスコアである$(\tilde{x}_i, \tilde{y}_i)$を求める式であり，$i$番目のパートフィルタの中心座標$(x_i, y_i)$，ルートフィルタの中心座標$(x, y)$と$v_i$，$s_i$より計算される．このとき，パートフィルタはルートフィルタの 2 倍の解像度を持つため，距離関係を元に戻すために，(x, y)を 2 倍にして計算する．$\tilde{x}_i$と$\tilde{y}_i$は共に-1 から 1 の値をとる．

2.2.5 検出

DPM の検出処理は，画像全体に対してスライディングウィンドウを走査し，各ルート位置におけるスコアを計算することで行われる．このとき，式(2.4)のスコアが最大となるパートフィルタの組み合わせを求め，スコアが閾値以上となった箇所を物体として

検出する．各パートフィルタのスコアは独立に求めることができるため，それぞれのパートフィルタの最大値を求めることでルート位置p_oにおけるスコアの最大値を計算することが可能である．

2.3　DPM の有効性の評価

従来手法をルートフィルタのみを使用する検出モデルを従来手法と定めて，通常の DPM との比較より，キャラクタ顔検出における DPM の有効性を評価する．本実験において，DPM のアルゴリズムは voc-relese5[13]を使用する．

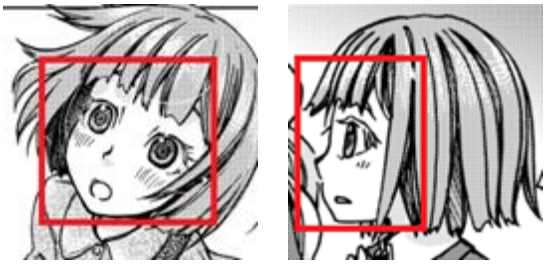
2.3.1　データセットの設定

検出器の学習及び評価に使用するデータセットには，作者の異なる漫画 10 作品を使用する. それぞれの作品から登場キャラクタを無作為に抽出したポジティブサンプルと，キャラクタ顔領域を含まない画像であるネガティブサンプルを用意する．本実験では，使用する漫画画像の一部に Manga109[14, 15]データセットにおいて公開された画像を使用する．本実験では，アノテーションの設定を簡略化するため，顔領域を含む画像と含まない画像をそれぞれ切り出して 200×200 pixel にリサイズした画像を作成する. また，ポジティブサンプルでは, 切り出された画像に対して顔領域のバウンディングボックスを記述するアノテーションを付与する. ポジティブサンプルはキャラクタの顔が正面向きに描かれている「正面顔」と，キャラクタの顔が横向きに描かれている「横顔」の 2 種類に分類する．ポジティブサンプルとネガティブサンプルの例をそれぞれ図 2-3，図 2-4 に示す．図 2-3 において，赤枠はキャラクタ顔領域としてアノテーションで指定された領域を示す.

本実験では，正面向きの顔画像のみを検出対象として，漫画 10 作品より「正面顔」を 100 枚ずつ抽出したポジティブサンプル 1000 枚と，作品を問わずにランダムに抽出したネガティブサンプル 2000 枚を学習に使用する．また，検出器の評価に使用する画像は，全て学習画像とは異なる画像として，漫画 10 作品より「正面顔」50 枚ずつを抽出したポジティブサンプル 500 枚と，ランダムに抽出したネガティブサンプル 2000 枚を使用する．

2.3.2　評価指標

本実験における検出器の評価には, PASCAL VOC の Precision-Recall プロトコル[17]を適用する. 顔として検出された領域とアノテーションに記載されたバウンディングボックスが 50%以上オーバラップしている場合に，Ture Positive と判定する．また，検出された領域とバウンディングボックスとのオーバラップが 50%未満の場合は False Positive と判定する．さらに，バウンディングボックスで指定された顔領域の中で検出されなかったものを False Negative とする. Precision と Recall の値は, True Positive, False Positive,



正面顔　　横顔

図 2-3: ポジティブサンプルの例（自発表[16]より引用，画像は文献[3]より著者の許可を得て抜粋）

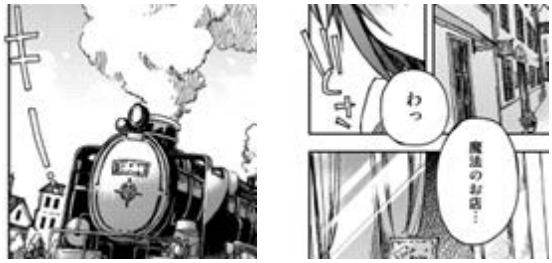


図 2-4: ネガティブサンプルの例（自発表[16]より引用，画像は文献[3]より著者の許可を得て抜粋）

False Negative の個数より，それぞれ式(2.6)，式(2.7)から求められる．

$$Precison = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2.6}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2.7}$$

式(2.4)の検出スコアに対する閾値を変化させることで，それぞれの閾値における Precision，Recall の値を算出し，Precision と Recall の値の変化を P-R 曲線として図示する．また，Recall = $0.0, 0.1, \cdots, 1.0$の 11 点に対応する Precision の平均値を Average Precision (AP)として評価指標に使用する．

2.3.3　パートモデルの有効性の確認

DPM による検出モデルが従来の単体のフィルタのみを使用する手法より有効であることを評価するため，ルートフィルタのみを使用する検出モデルと通常の DPM の検出精度を比較する．本実験では，正面向きの顔のみを検出対象として，検出モデルのコン

ポーネント数は 2 個に設定する．また，DPM モデルについて，パートフィルタの個数はデフォルトである 6 個に設定する．重複して検出されたバウンディングボックスを統合するための閾値である Non-Maximum Suppression (NMS)の値は 0.3 に設定する．

実験結果を図 2-5 に示す． P-R 曲線において，Recall が約 0.4 以上のときに DPM の検出率は従来手法を上回る結果となった．これは，パートモデルを使用することによって物体の誤検出が減少した一方で，形状変化の大きな物体に対する検出率が低下するトレードオフが発生したためであると考えられる．AP では，従来手法の 60.2%に対して DPM は 65.2%を示し，総合的な検出率において DPM が従来手法を上回ることを確認した．

2.4 DPM の最適化

DPM を顔検出に適用した研究として，Orozco らは DPM を用いた人物のマルチビュー顔検出を行っている[18]．この研究において，Orozco らは DPM の検出モデルにおけるコンポーネントの個数とパートフィルタの個数はそれぞれ数が増えるほど Precision が上昇し Recall が低下するトレードオフの関係にあると説明している．そして，人物の顔検出における実験では，コンポーネント 4 個，パートフィルタ 6 個の構成が最適となったと報告している．しかし，実在の人物とは異なる特徴を持つ漫画キャラクタに対しては異なる傾向を示す可能性が考えられる．したがって，DPM のパラメータ変化による検出率の影響について調査することで漫画キャラクタ顔検出に最適な設定を検討する．本研究では，NMS，コンポーネントの個数，パートフィルタの個数の 3 種類のパラメータについて検討する．本実験で使用する DPM のアルゴリズム及び評価指標は，2.3 節と同様とする．

2.4.1 データセットの設定

本実験で使用するデータセットは，2.3.1 項で作成した画像を使用する．ここで，2.4.2 項の実験では，正面向きの顔画像のみを検出の対象として，2.3.1 項で使用したものと同様の構成のデータセットを使用する．一方，2.4.3 項及び 2.4,4 項の実験では，キャラクタのマルチビュー顔検出を目的として，「横顔」をポジティブサンプルに追加したものを使用する．したがって，検出器の学習に使用する画像は，漫画 10 作品よりそれぞれ「正面顔」と「横顔」を 100 枚ずつ抽出したポジティブサンプル 2000 枚と，作品を問わずにランダムに切り出したネガティブサンプル 2000 枚を使用する．また，検出器の評価に使用する画像は，漫画 10 作品より「正面顔」と「横顔」50 枚ずつを抽出したポジティブサンプル 1000 枚と，作品を問わず切り出されたネガティブサンプル 2000 枚を使用する．

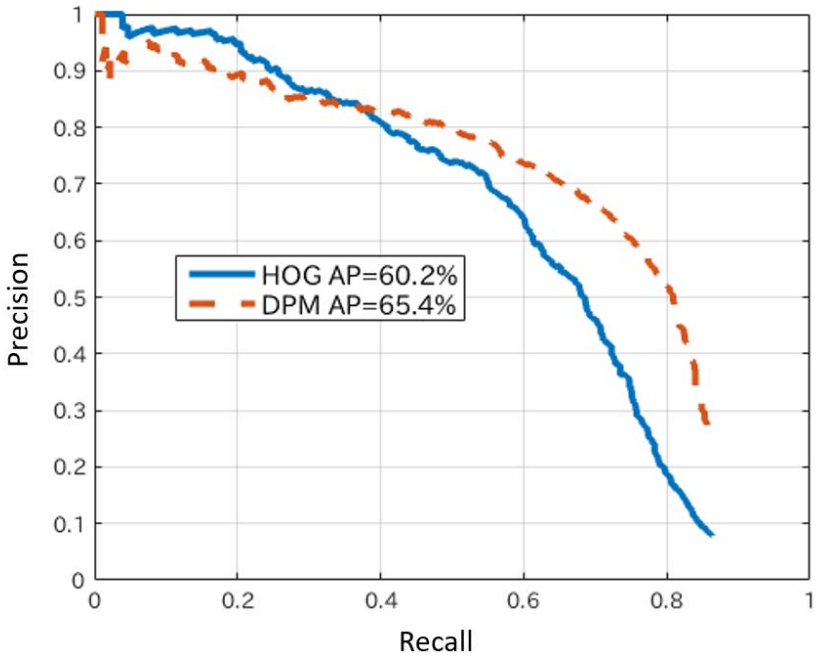


図 2-5: 正面顔の検出における従来手法と DPM の比較（自発表[16]より引用）

2.4.2　NMS の最適化

DPM を用いた顔検出における，NMS の最適値について検討する．ルートフィルタのコンポーネントを 2 個，パートフィルタを 6 個とした検出モデルについて，NMS を 0.1 ~ 0.5 に変化させた検出結果を比較した．実験結果を図 2-6 に示す．この結果より，NMS の値が低い場合は Recall が低下し，高い場合は Precision が低下するトレードオフが確認できた．AP の値は NMS を 0.3 に設定したとき最も高くなった．

2.4.3　コンポーネントの個数の最適化

多視点でのキャラクタ顔検出を対象とした場合に最適となる DPM のコンポーネント数について検討する． NMS を 0. 3，パートフィルタの個数を 6 として，コンポーネント数を 2 ~ 20 まで変化させた検出モデルによる検出結果を比較した．実験結果を図 2-7 に示す．実験結果より，コンポーネントの個数が 4 ~ 18 の範囲においては P-R 曲線に大きな変化は見られず，2 や 20 に設定した場合には他のモデルより検出率が低下することが確認できた．この理由として，正面向きの顔と横向きの顔に対応可能な検出モデルを生成するために，最低でも 4 個以上のコンポーネントが必要であることと，コンポーネント数が多い場合にはトレードオフによって Recall の低下が大きくなることが原因であると考えられる．AP の値はコンポーネント数が 10 のとき最大となった．

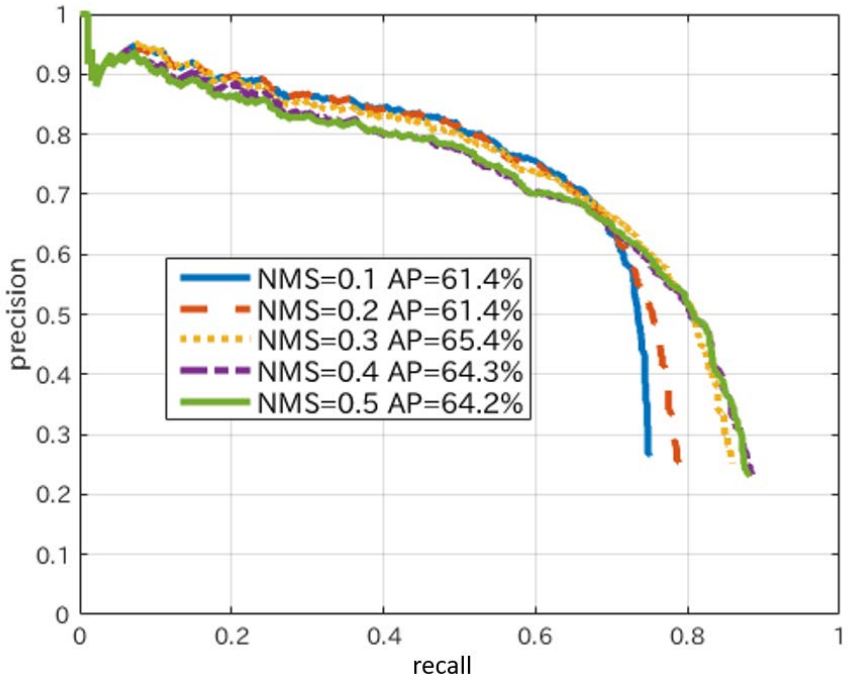


図 2-6: NMS と検出率の関係（自発表[16]より引用）

2.4.4　パートフィルタの個数の最適化

最適なパートフィルタの個数について検討する．NMS を 0.3，ルートフィルタのコンポーネントを 10 個と設定し，パートフィルタを 1 ~ 16 個の範囲で変化させた検出モデルについて比較を行った．実験結果を図 2-8，図 2-9 に示す．実験結果からはパートフィルタ数と検出精度について明確な相互関係は確認できなかった．この理由について，生成された顔検出モデルからの考察を次の項で述べる．AP の値はパートフィルタの個数を 11 に設定した場合に，最高で 77.6%となった．

2.4.5　実験結果の考察

2.4.4 項の実験より，ルートフィルタのコンポーネントを 10 個，パートフィルタを 11 個に設定した場合における，DPM の検出モデルのフィルタを可視化した画像を図 2-10 に示す．図では 10 個のコンポーネントのうち片側の向きに対応する 5 種類のフィルタを表しており，(a)はルートフィルタ，(b)はパートフィルタをそれぞれ可視化した画像である．

通常のマルチビュー顔画像を対象とした検出モデルでは，対象物体のバウンディングボックスをアスペクト比で分類することで顔の角度変化に対応したモデルが生成される．しかし，本実験で生成された検出モデルでは，正面顔と横顔に対応したルートフィルタが複数生成されている様子が確認できる．これは，デフォルメ表現を含む漫画キャ

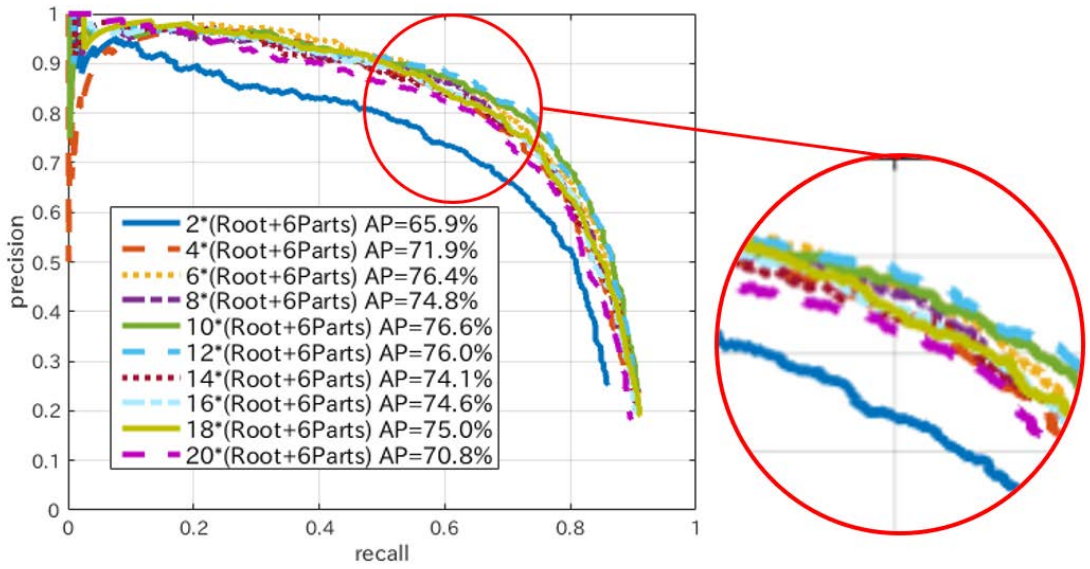


図 2-7: 検出モデルのコンポーネントの個数と検出率の関係（グラフは自発表[16]より引用）

ラクタは人物ごとの顔画像のアスペクト比の変化が実在の人物より大きいためであると推測できる．したがって，DPM のキャラクタ検出モデルの設計において，学習に使用する漫画画像の影響が通常の顔検出よりも大きいことが分かった．

また，パートフィルタについて，図では各フィルタが目や鼻といった顔パーツではなく，顔の輪郭部分を捉えるように配置されていることが確認できる．この理由として，現実の顔画像では目や鼻といった顔パーツの形状にある程度の共通性が存在するが，漫画画像では顔パーツがより多様な形式で表現されることが考えられる．また，漫画では顔の全体が一様のエッジ強度を持つ線画で表現されるため，輪郭部分の影響が自然画像よりも強いことも影響していると考えられる．これらのことから，HOG 特徴量では顔パーツの形状を捉えることが困難であることが推測できる．したがって，検出モデルにおいて顔パーツへのパートフィルタの割り当てが明確でないために，2.4.4 項の実験においてパートフィルタ数による検出率の変動が小さくなったと考察できる．

以上より，漫画キャラクタに最適な DPM 検出モデルを構築するためには，学習データの影響の考慮及び，顔の向きやキャラクタごとのアスペクト比の変化に対応した複数のモデルを用意する必要があるといえる．

2.5 むすび

本章では，キャラクタ顔画像の形状変化に対応できる検出手法として，DPM の漫画

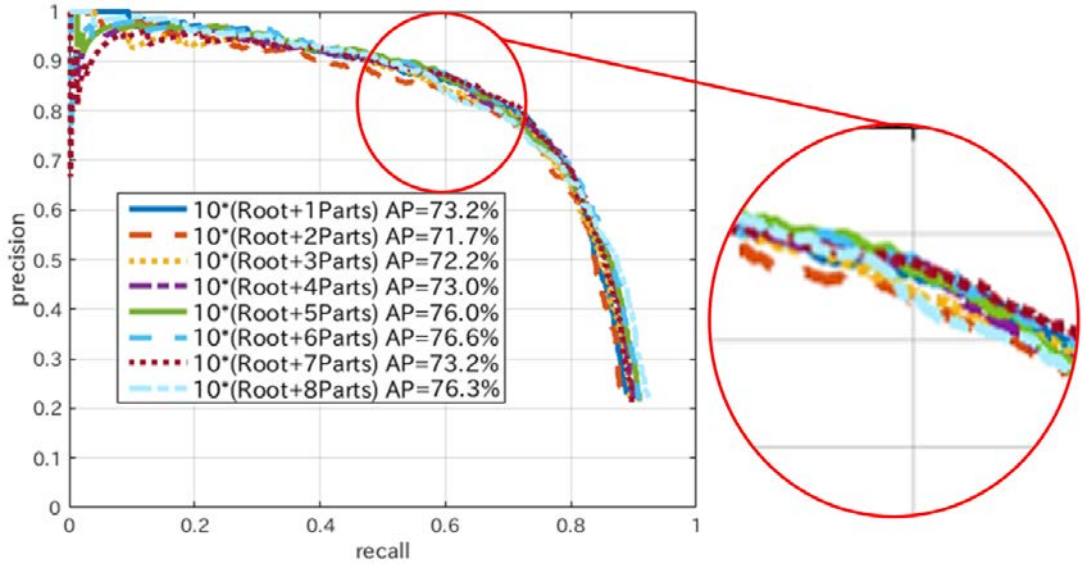


図 2-8: パートフィルタの個数と検出率の関係（その 1）（グラフは自発表[16]より引用）

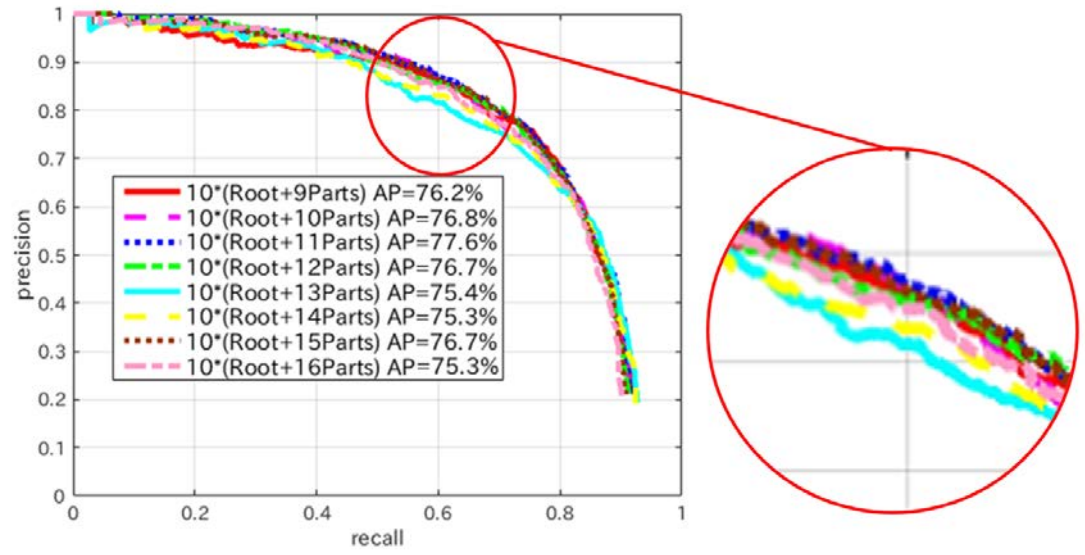


図 2-9: パートフィルタの個数と検出率の関係（その 2）（グラフは自発表[16]より引用）

キャラクタ検出への適用について検討した．

はじめに，提案手法で使用する技術として，DPM の概要について述べた．次に，ルートフィルタのみを使用する検出モデルと通常の DPM についてキャラクタ顔画像の検出精度を比較することで，キャラクタ顔検出に対する DPM の有効性を評価した．実験

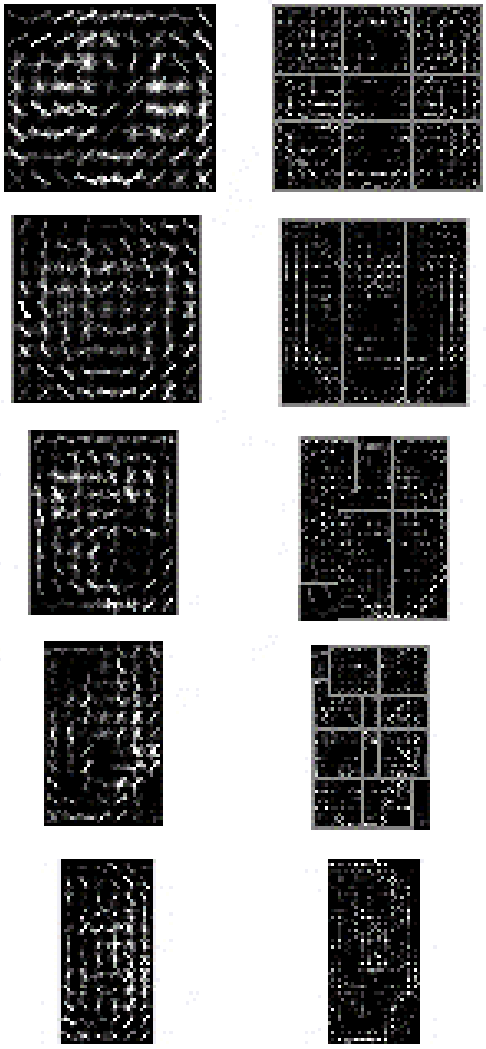


図 2-10: 実験において生成された DPM 検出モデル（自発表[16]より引用）

結果より，DPM の AP は 65.4%を示し，従来手法の 60.2%を上回った．

次に，キャラクタ顔画像のマルチビュー顔検出を対象として，最適な DPM の検出モデルの構成を検討した．実験結果より，最適な検出モデルを設定することで，最大で 77.6%の AP を達成できることが示された．

最後に，実験結果における考察を述べ，DPM を多様なキャラクタ顔画像に対応させるには，多数のフィルタを用いた複雑な検出モデルの設計が必要であるという課題を明らかにした．

第 3 章

CNN を用いた漫画オブジェクトの検出

3.1 まえがき

本章では，漫画オブジェクトの検出における Convolutional Neural Network (CNN)の適用を提案する．2 章では，HOG に代表されるハンドクラフト特徴量は既定の形状した表現できないために，多様な形状を持つ漫画キャラクタへの対応が困難であるという知見が得られた．CNN は，近年の物体認識タスクにおいて高い認識率を達成している手法である．この手法は，画像認識に最適な特徴量を自動的に学習するため，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つ．しかし，CNN の詳細な画像認識のメカニズムはブラックボックスであるため，一般画像と異なる特徴を持つ漫画画像への有効性は未知である．そこで，漫画オブジェクトに対する CNN の有効性を評価する．

はじめに，CNN の概要について述べる．次に，CNN を用いた物体検出手法である R-CNN のアルゴリズムについて述べる．そして，R-CNN を改良した物体検出手法について述べる．次に，キャラクタ顔画像に対して最適化した DPM と Fast R-CNN との比較から，CNN 導入の有効性を評価する．さらに，キャラクタに加えてコマとフキダシ検出対象とした Faster R-CNN の適用を評価し，コマ内容を認識するアルゴリズムを検討する．最後に，CNN を用いた物体検出法の比較を行うことで，漫画オブジェクトに対して有効な候補領域の抽出方法について考察する．

3.2 畳み込みニューラルネットワーク

Convolutional Neural Network (CNN)は，脳の視覚情報処理の構造を模したネットワークを持つ多層パーセプトロンの一種である． CNN の処理の流れを図 3-1 に示す．CNN の処理は多段接続された複数の処理ユニットを通して行われる．CNN の中間層は畳み込み層とプーリング層の組み合わせから構成されており，各ユニットの入出力は特徴マップと呼ばれる複数個の 2 次元データである．まず，入力画像に対して重みフィルタの畳み込み処理を行い，特徴マップを出力する．次に，出力された特徴マップに対してプーリング処理を行い，新たな特徴マップを得る．この処理を繰り返すことで CNN は入力画像から特徴量を求める．入力に近い層ではエッジや線などの単純なパーツが抽出さ

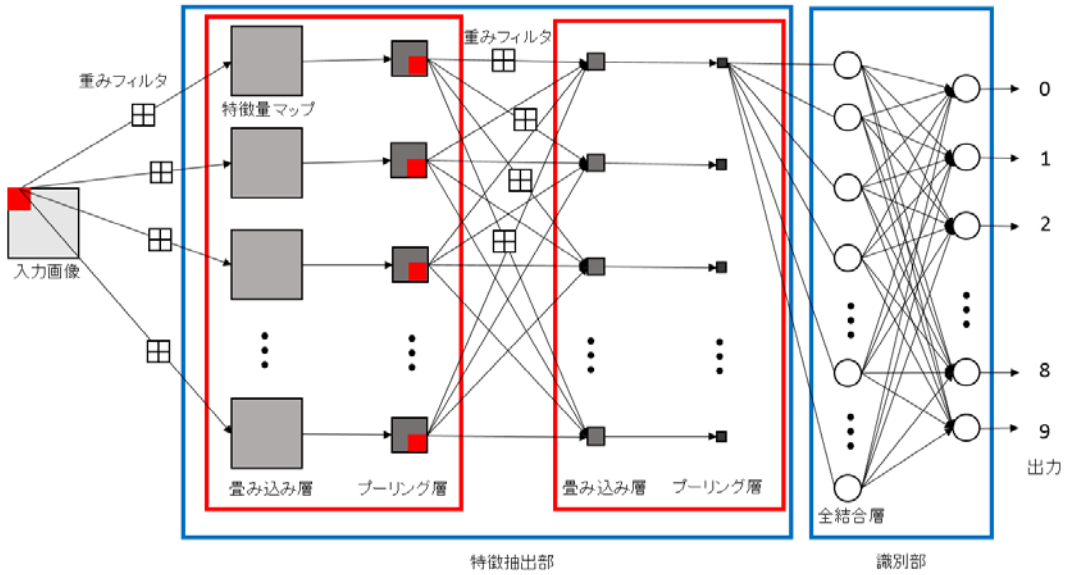


図 3-1: CNN の概要（自発表[12]より引用）

れ，畳み込みとプーリングを繰り返すことで特徴同士がまとめ上げられて顔や物などを表現する複雑で抽象的な特徴量が生成される．最後に，得られた特徴マップを識別部に入力してクラス分類を行う．

CNN の学習には，教師付き学習を前提として，誤差逆伝播法を用いた勾配降下最適化法が用いられる．通常の多層パーセプトロンは，各層間の重みが全結合した構造を持つため，中間層の数が多い場合には学習時に誤差の勾配が拡散してしまうという問題がある．CNN は層間を局所的に結合することによって，複数の中間層を持つ場合の学習を可能としている．

3.2.1 畳み込み層

畳み込み層は入力画像に対して重みフィルタとの内積をとり，重みフィルタを画像上で走査して計算処理を繰り返すことで特徴マップを出力する．この処理によって，ピクセルベースではなく領域ベースでの特徴抽出が行われるため，画像の移動や変形などに頑強な特徴が得られる．フィルタの重みは，誤差逆伝播による勾配降下最適化法によって更新される．また，CNN は一つの入力画像に対して複数のフィルタを用いて異なる特徴マップを生成することで，様々な画像特徴を捉えることができる．

3.2.2 プーリング層

プーリング層は畳み込み層の直後に置かれる層であり，入力された特徴マップの小領域から値を抽出して新たな特徴マップを生成する．プーリング処理には二つの目的があり，一つは，ユニット数を削減することで調整するパラメータの数を削減するためであ

る．もう一つは，画像のどの位置に対してフィルタの応答が強かったという情報を一部捨てることで，画像内に存在する特徴の微小な位置変化に対して普遍性を得るためである．プーリング処理は畳み込み層の隣接している 2×2 ユニットについて行われる．プーリング処理には，ユニット内の最大値を出力する「最大プーリング」，ユニットの平均値を出力する「平均プーリング」，ユニットのp乗平均偏差を出力する Lp プーリングの３種類が存在する．

3.2.3　全結合層

全結合層は中間層の最後に置かれる全結合した多層パーセプトロンによって構成される層であり，特徴抽出部を通して生成された特徴マップを一つのノードに結合して特徴変数を出力する．その後，各ユニットの特徴変数を出力層に入力する．出力層は正解クラスの数と同数のユニットを持つ層であり， softmax 関数を用いて入力値を確率に変換することで，入力画像がそれぞれのクラスに一致しているか判定する．

3.3　Regions with CNN features (R-CNN)

CNN の特徴抽出を物体検出に応用したアルゴリズムとして，2015 年に Girshick らは R-CNN を提案した[19]．R-CNN の物体検出の流れは以下の三つの工程に分けられる．はじめに，入力画像から物体の候補となる領域を検出し，切り出しを行う．次に，抽出された領域の候補をそれぞれ CNN に入力して特徴量を計算する．最後に，それぞれの候補領域について何が映っているか特徴量よりクラス分類を行い，対象物体の存在する領域を検出する．それぞれの工程について，以下に詳細を述べる．

3.3.1　物体候補領域の抽出

従来の物体検出では，画像から認識を行う領域を切り出すために，スライディングウィンドウが用いられていた．これは，様々なサイズ・アスペクト比を持つ矩形フィルタで画像全体を走査し，総当り的に領域を切り出す手法である．しかし，スライディングウィンドウには，処理の対象となる領域が非常に多くなることや，対応できる形状やサイズに制限があるといった問題がある．そこで，一つの物体を表している可能性の高い領域の候補を画像から検出するアルゴリズムを用いて候補領域の切り出しを行い，物体認識器に入力することで計算量を削減する手法が提案されている．R-CNN は，Uijlings らによって提案された Selective Search[20]を利用して候補領域の抽出を行う手法である．

Selective Search は，ボトムアップ型の階層的セグメンテーションによって，あらゆる位置やスケールに対応した候補領域を抽出することが可能なアルゴリズムである．はじめに，Efficient Graph-Based Image Segmentation[21]と呼ばれるアルゴリズムによって初期のセグメンテーションを行う．このアルゴリズムは，画像中の各画素を一つのノードとした木から，輝度が類似しているノードを纏めていくことで類似した画素をセグメン

テーションする．次に，セグメンテーションによって得られた各小領域について，色特徴・テクスチャ特徴・小領域の面積・小領域の外接矩形の四つの特徴を複合した特徴量を算出する．そして，特徴量の類似度が最も高い近接領域を統合し，小領域の外接矩形を候補領域として抽出する．この統合処理を 1 枚の画像となるまで繰り返すことで，最終的に 2000 個程度の候補領域が得られる．

3.3.2　CNN 特徴量の計算

抽出された候補領域を CNN に入力して特徴量の計算を行う．このとき，候補領域の周辺の領域の情報を付け加えるために，Selective Search で抽出された候補領域より少し大きい領域（リサイズ後のサイズで周囲 16 画素分）について切り出しを行い，227×227 画素にリサイズして CNN に入力する．

3.3.3　候補領域のクラス分類

R-CNN では，CNN の出力層でクラス分類を行う代わりに，生成された特徴量を線形 SVM に入力してクラスを識別する．この理由は，ニューラルネットワークでクラス分類を行うためには大規模な学習データが必要となるが，線形 SVM を用いて CNN 特徴量を分類する場合には少量の学習データからでも高精度な分類が可能であるためであると説明されている[22].

多クラスの識別には，物体のクラスごとに学習した複数の線形 SVM を使用する．識別結果が複数のクラスについてオーバラップした場合には，NMS によって SVM のスコアが小さいクラスを除去する．Selective Search と CNN の特徴量は複数のクラスに共通して計算できるため，クラス依存の計算は線形 SVM の識別と NMS だけで効率的に計算できる．候補領域が物体として認識された後は，CNN によって計算された特徴量から境界ボックス回帰を行うことで，バウンディングボックスの位置を修正する．

3.4　R-CNN の改良

近年では，R-CNN のアルゴリズムをベースとした様々な改良手法が提案されている．方針の一つに，画像からの候補領域の抽出方法を改良することで計算量の削減と位置推定の精度を向上させるアプローチがある．しかし，これらの手法は主に一般物体検出を目的として設計されていることから，一般物体と異なる特徴を持つ漫画画像に対する有効性は未知である．そこで本研究では，異なる候補領域抽出アルゴリズムを使用している手法を比較することで，漫画画像に対する有効性を検証する．

3.4.1　Fast R-CNN

R-CNN は Selective Search によって抽出された全ての候補領域について CNN の計算処理を行うため，処理時間が膨大になるという問題を持つ．この問題に対して，計算量

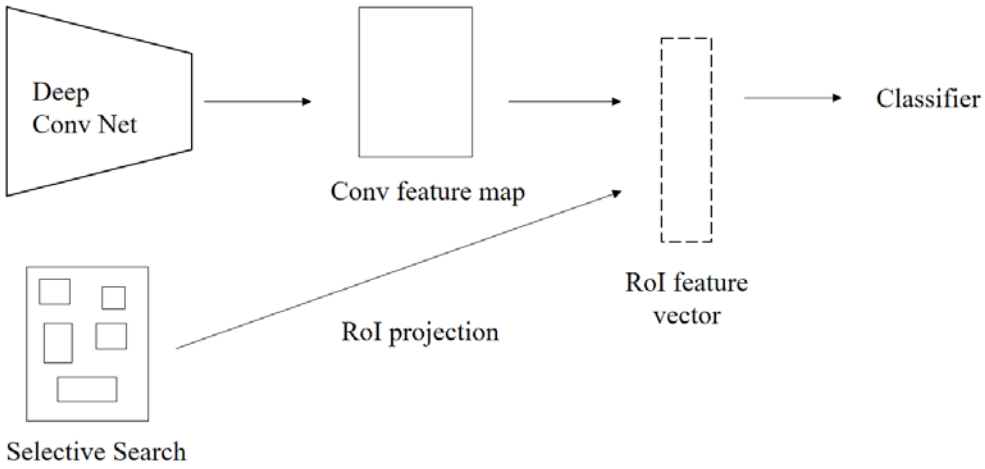


図 3-2: Fast R-CNN の概要（自発表[24]より引用）

を削減することで高速化を目指したアルゴリズムとして Fast R-CNN が提案されている[23].

Fast R-CNN による物体検出の流れを図 3-2 に示す．はじめに，画像全体を CNN に入力し，特徴抽出部において任意サイズの特徴マップを計算する．次に，Selective Search によって求めた候補領域を特徴マップ上に射影し，候補領域についてプーリング処理を行う．このとき，任意サイズの候補領域のプーリングによって固定サイズの出力を得るために，RoI Pooling と呼ばれるアルゴリズムが用いられる．RoI Pooling は，抽出された候補領域をプーリング層の特徴マップと同じサイズのブロックに等分割する．そして，領域内のピクセルをそれぞれいずれかのブロックに割り当て，ブロック内の最大値又は平均値をとることで固定サイズの出力を得る．候補領域の射影を行った後は通常の R-CNN と同様に，候補領域について物体クラスの分類問題と境界ボックス回帰問題を解く．Fast R-CNN の学習は通常の CNN と同様に，誤差逆伝播によって重みを更新する．ただし，Multi-task loss という学習技術により，候補領域の抽出処理について同時に更新することで，検出モデル全体の end-to-end な学習を可能としている．

3.4.2　Faster R-CNN

Faster R-CNN は, Fast R-CNN の改良として Ren らによって提案された手法である[25]. Fast R-CNN は R-CNN からの高速化を実現したが，候補領域の抽出に Selective Search を使用していることが更なる高速化におけるボトルネックとなっていた. この問題について，Faster R-CNN は候補領域の検出処理自体を CNN で行うことによる解決を図っている．

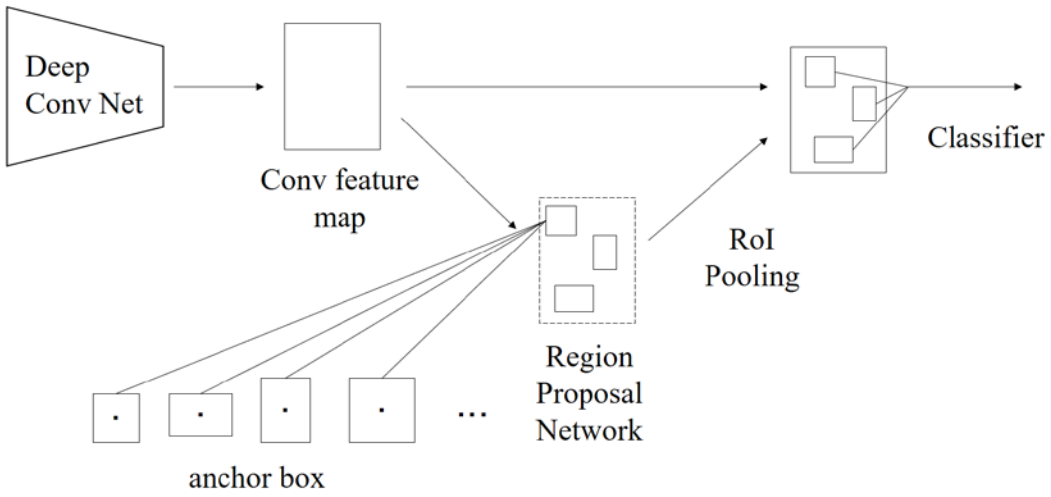


図 3-3: Faster R-CNN の概要　(自発表[24]より引用)

Faster R-CNN による物体検出の流れを図 3-3 に示す．Faster R-CNN の処理のうち，入力画像全体から特徴マップを計算する処置と，抽出された候補領域を特徴マップに射影して物体のクラス判定を行う処理は，Fast R-CNN と同様である．Faster R-CNN のアルゴリズムに特異な点として，特徴マップからの候補領域抽出に Regions Proposal Network (RPN)を用いることが挙げられる．RPN は CNN 内に組み込まれたネットワークであり，特徴マップに対してフィルタの畳み込み処理を行い，出力された特徴量について物体か否かの判定と矩形回帰の計算を行う．この処理は，特徴マップ上をスライディングウィンドウで走査し，抽出された小領域に対して判定を行うという処理と同等である．このとき，物体のアスペクト比やスケールの変化に対応するために，RPN は複数の外接矩形からなるアンカーを使用する．デフォルトの設定では，3 種類のスケールとアスペクト比からなる 9 種類のアンカーが設定されている．アンカーによって，一つの検出ウィンドウからそれぞれの矩形に対するスコアと矩形回帰を一度に計算することが可能である．RPN の導入によって Faster R-CNN は Fast R-CNN からの高速化を実現し，さらに，候補領域の抽出方法が改良されたことによって検出精度も向上している．

3.4.3　Single Shot MultiBox Detector (SSD)

Faster R-CNN の提案によって，単一の CNN による物体検出が可能となった．しかし，特徴量の計算と RPN による候補領域の検出処理が分かれていることからネットワーク構造が複雑であり，リアルタイムで検出処理を行うには計算速度がまだ不十分であるという問題がある．SSD は，ネットワークの処理を単純化することでリアルタイムでの検出を目指したアルゴリズムであり，Faster R-CNN と同等の検出精度を持ちつつ，大幅な

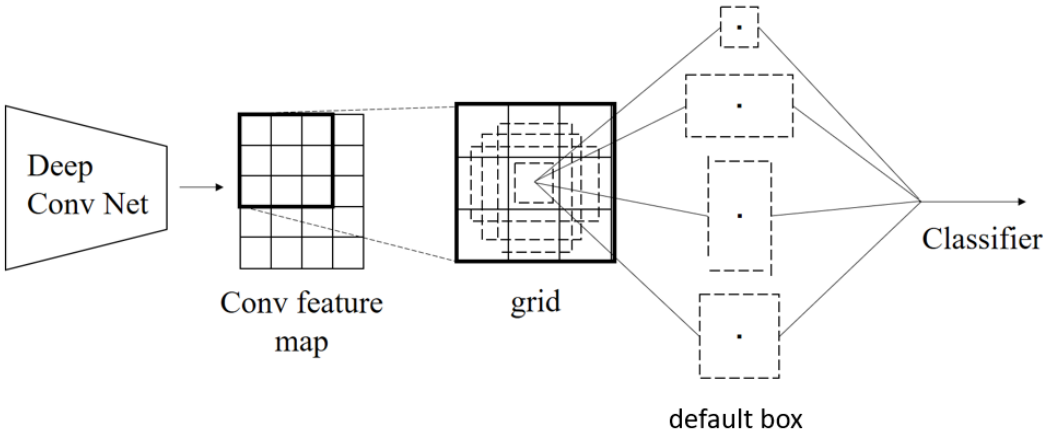


図 3-4: Single Shot MultiBox Detector の概要

処理速度の向上を達成している[26].

SSD による物体検出の概要を図 3-4 に示す．SSD は画像から特徴マップを計算した後に，特徴マップをグリッドで分割し，それぞれの分割領域に対してデフォルトボックスを適用する．その後，畳み込み処理によって特徴マップのスケールを縮小して再び分割処理を行う．これを繰り返すことで，ネットワークの下段に進むほど大きなグリッドで特徴マップが分割され，物体検出が行われることとなる．デフォルトボックスはアスペクト比の異なるバウンディングボックスのセットであり，一つの検出ウィンドウからそれぞれのバウンディングボックスに対するカテゴリのスコアと矩形回帰を予測する．この処理は Faster R-CNN におけるアンカーボックスと類似している．ただし，スケールの異なる物体を検出するために，アンカーボックスはスケールの異なる矩形を用意して対応する一方，デフォルトボックスは特徴マップの分割スケールを変化させることで対応するという点において異なる．

3.5　提案手法の評価

漫画キャラクタに対する CNN の有効性を評価するため，Fast R-CNN と DPM のキャラクタ顔検出精度を比較する．本実験の評価基準には 2.3.2 項で述べた PASCAL VOC の Precision-Recall プロトコルを使用する．

3.5.1　検出器の設定

DPM の実装は voc-release5 を使用する．Fast R-CNN の実装は girshickICCV15fastrcnn [27]を使用し，ニューラルネットワークのアークテクチャには vgg_cnn_m_1024[28]を使用する．vgg_cnn_m_1024 は 5 層の畳み込み層と 3 層の全結合層から構成される 8 層の CNN であり，第 7 層で出力される特徴量は 1024 次元である．またディープラーニング

では，大規模データセットについて学習したモデルを初期重みとして，対象のデータセットについて再度学習するファインチューニングと呼ばれる処理により，効果的な学習が可能であることが知られている．したがって本研究では，一般物体データセットである ImageNet で学習を行ったニューラルネットワークモデルを元に，ファインチューニングを行うことで検出器を学習する．Fast R-CNN のパラメータは NMS を 0.1，学習の反復回数を 40000 回，バッチサイズを 128 に設定する．

3.5.2　データセットの設定

本実験では複数の漫画作品についてキャラクタを検出することを目的として，「ドラえもん」[29]，「ブラック・ジャック」[30]，「名探偵コナン」[31]，「SLAM DUNK」[32]の 4 作品に登場するキャラクタの顔領域を抽出した画像を検出の対象とする．アノテーションの指定を簡略化するため，元の漫画からキャラクター人のみを含む領域をそれぞれ切り出して使用する．ポジティブサンプルはキャラクタの顔領域周辺を切り出して，200×200 ピクセルにリサイズした画像とする．このとき，顔領域のバウンディングボックスを記述するアノテーションを付与する．また，顔領域のうち両目が描かれている角度のものを「正面顔」，片目のみが描かれている角度のものを「横顔」，コマの枠線やオブジェクトによって顔の一部が隠れているものを「隠れ顔」とそれぞれ定義する．「正面顔」，「横顔」及び「隠れ顔」として定義された画像の例を図 3-5，図 3-6，図 3-7 に示す．それぞれの図において，赤枠はバウンディングボックスで指定する領域を示す．ネガティブサンプルは，先述のマンガ作品からキャラクタの顔が含まれない領域を無作為に切り出して 200×200 ピクセルにリサイズした画像を使用する．ネガティブサンプルとして定義される画像の例を図 3-8 に示す．

3.5.3　データセットに対する DPM の最適化

学習データについて最適な DPM のパラメータ設定を求める．学習に使用するデータセットの内容を表 3-1 に，評価に使用するデータセットの内容を表 3-2 に示す．それぞれのデータセットは全て異なる画像で構成される．

まず，最適なルートフィルタのコンポーネントの数を求める．パートフィルタの個数を 8 個，NMS を 0.5 に設定し，ルートフィルタのコンポーネント数を 2 個，4 個，6 個に設定した検出モデルについて検出率を比較した．実験結果を図 3-9 に示す．この結果より，コンポーネント数を 4 としたとき AP は 88.0%となり，最も高い値が得られることが分かった．

次に，最適なパートフィルタの数を求める．コンポーネントの個数を 4 個，NMS を 0.5 に設定し，パートフィルタの個数を 3〜8 個の範囲で変動させたモデルについて検出結果を比較した．実験結果を図 3-10 に示す．この結果より，パートフィルタの個数を 4 個に設定したとき AP は最高で 88.2%を示すことが分かった．

© 木野陽

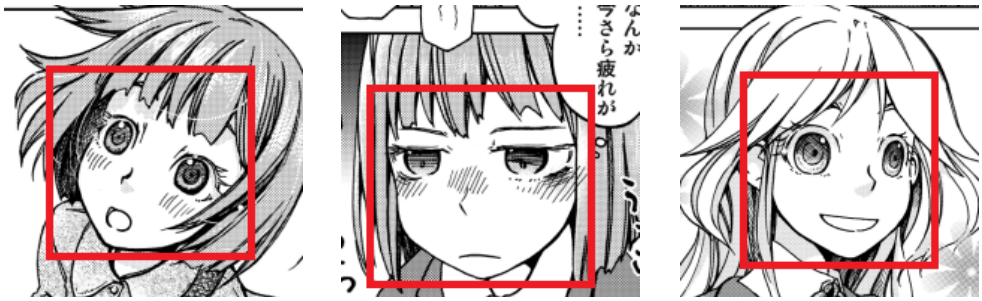


図 3-5: 正面顔の例（画像は文献[3]より著者の許可を得て抜粋）

© 木野陽

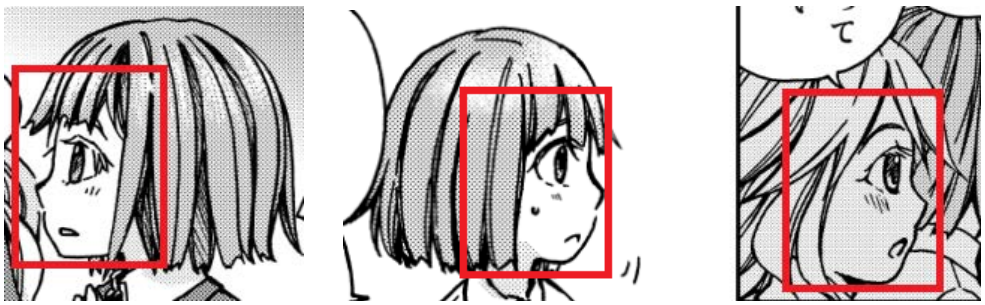


図 3-6: 横顔の例（画像は文献[3]より著者の許可を得て抜粋）

© 木野陽

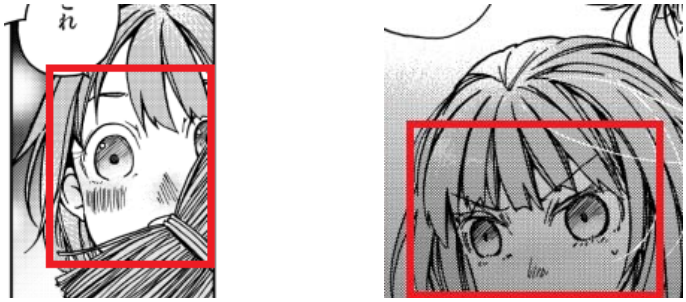


図 3-7: 隠れ顔の例（画像は文献[3]より著者の許可を得て抜粋）

© 木野陽

図 3-8: ネガティブサンプルの例（画像は文献[3]より著者の許可を得て抜粋）

表 3-1: R-CNN と DPM の学習に使用する漫画画像（自発表[12]より引用）

作品タイトル	ポジティブサンプル		ネガティブサンプル
	正面顔	横顔	
"ドラえもん"	100	50	1000
"ブラック・ジャック"	100	50	
"名探偵コナン"	100	50	
"SLAM DUNK"	100	50	
合計	400	200	1000

表 3-2: R-CNN と DPM のテストに使用する漫画画像（自発表[12]より引用）

作品タイトル	ポジティブサンプル			ネガティブサンプル
	正面顔	隠れ顔	横顔	
“ドラえもん”	90	10	50	2000
“ブラック・ジャック”	90	10	50	
“名探偵コナン”	90	10	50	
“SLAM DUNK”	90	10	50	
合計	360	40	200	2000

これらの結果から，本実験における最適な DPM の検出モデルを，コンポーネントの個数を 4 個，パートフィルタの個数を 4 個としたモデルと定めた．

3.5.4　DPM と Fast R-CNN の検出精度の比較

3.5.3 項で求めた DPM の検出モデルと Fast R-CNN の比較により，CNN の漫画キャラクタに対する有効性を確認する．検出器の学習と評価には 3.5.3 項の実験と同様のデータセットを使用する．

実験結果を図 3-11 に示す．P-R 曲線より，Fast R-CNN の検出精度が Recall において DPM を大きく上回ることが示された．このことより，DPM では検出することができなかった顔画像が，Fast R-CNN を用いることで検出可能になったことが分かる．AP の値は，DPM の 87.8%に対して，Fast R-CNN は 90.8%を示した．P-R 曲線に比べて Fast R-

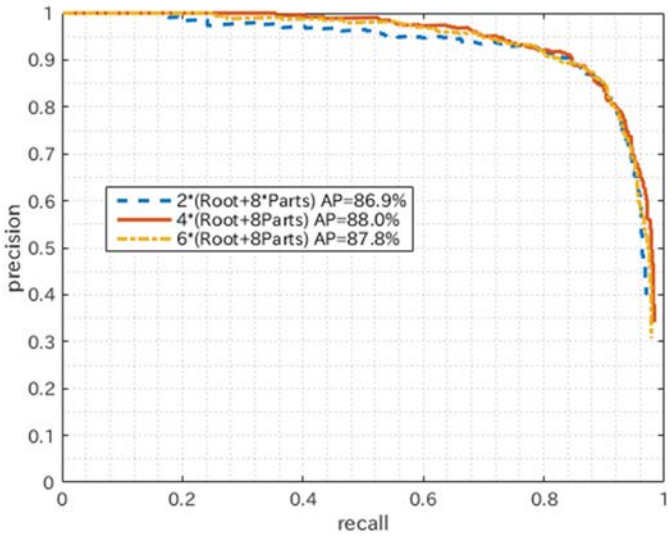


図 3-9: ルートフィルタのコンポーネントの個数と検出率の関係（自発表[12]より引用）

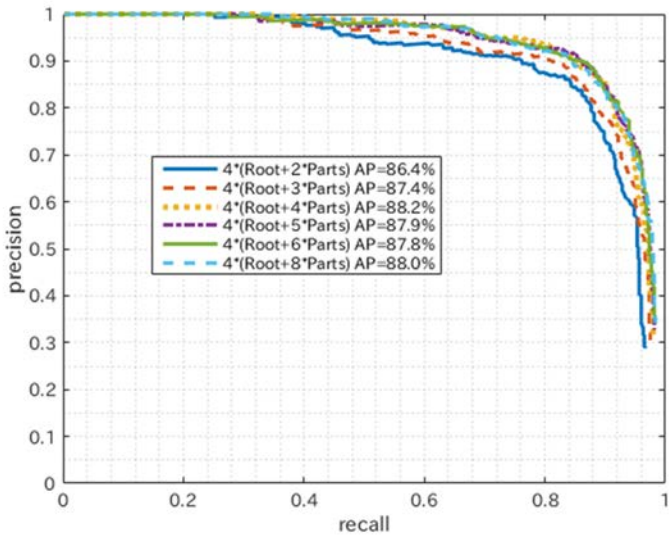


図 3-10: パートフィルタの個数と検出率の関係（自発表[12]より引用）

CNN と DPM の AP の差が小さい理由として，Fast R-CNN が Recall = 1.0 となる検出結果を持たないため，Recall =0.0, 0.1, ⋯, 0.9の範囲における Precision の値のみが反映されたことが考えられる．検出結果の比較では Fast R-CNN の検出率が DPM を上回ることが確認でき，Fast R-CNN が DPM より簡潔な検出モデルによってキャラクタのマルチビュー顔検出に対応可能であることが示された．

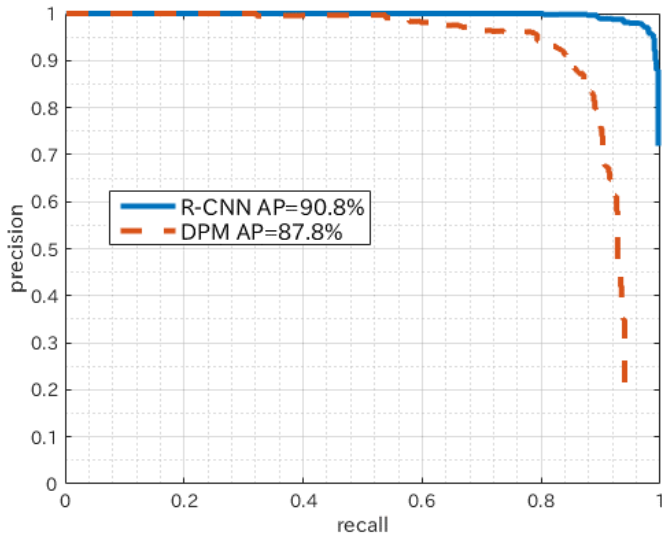


図 3-11: Fast R-CNN と最適化した DPM の比較（自発表[12]より引用）

3.6　漫画オブジェクトの検出

漫画からコマやフキダシといったオブジェクトを検出する手法として，従来研究では，画像から特定の基準以上の大きさを持つ blob 領域を検出し，blob 領域と連結した画素を抽出する手法が提案されている[7]．このような幾何学的解析に基づいた手法は，コマやキャラクタが途切れのない線で囲まれているというルールを前提としている．しかし，実際の漫画オブジェクトは様々な特殊な表現が存在しており，例えばコマ同士が重なっているといった場合に従来手法ではオブジェクトを正確に検出することができない．したがって，多様な漫画オブジェクトに対応可能な手法として，Faster R-CNN を用いた漫画オブジェクトの検出を提案する．また，漫画内容を認識するために，オブジェクトの検出結果からそれぞれのコマに含まれるオブジェクトの内容を取得するシステムを提案し，その有効性を評価する．

3.6.1　データセット

本実験において学習と評価に使用するデータセットは，Manga109 より公開されている漫画画像を使用する．学習セットは，作者の異なる 20 作品からそれぞれ 100 ページずつをランダムに抽出し，各ページに対して「コマ」，「フキダシ」，「キャラクタ」の 3 種類の漫画オブジェクトのアノテーションを付与した画像合計 2000 枚を使用する．テストセットは，学習セットに使用したものとは異なる漫画 5 作品からそれぞれ 30 ページを抽出した合計 150 枚の画像を使用する．

3.6.2　検出器のパラメータ設定

Faster R-CNN のアルゴリズムは[33]で公開されているプログラムを使用する．ここで，

複数クラスの漫画オブジェクトについて同時に学習した場合において, 検出率が低下するという問題が報告されている[34]．これは，キャラクタとコマのように，あるクラスが別のクラスの内部に存在するという関係が漫画オブジェクトにおいて成り立つことによるものと推測される．このため，通常の多クラス検出と同様の学習を行った場合に領域のクラス割り当てに不具合が発生することが検出率の低下の原因であると考えられる．本実験では，各手法の個々のオブジェクトに対する有効性を求めることを目的とすることから, それぞれのオブジェクトについて対象物体か否かの 2 クラス分類を行う 3 種類の検出器を作成して評価を行う．CNN モデルには vgg_cnn_m_1024 を使用し，NMS を 0.3，バッチサイズを 128 と設定する．

3.6.3　学習回数と検出率の関係

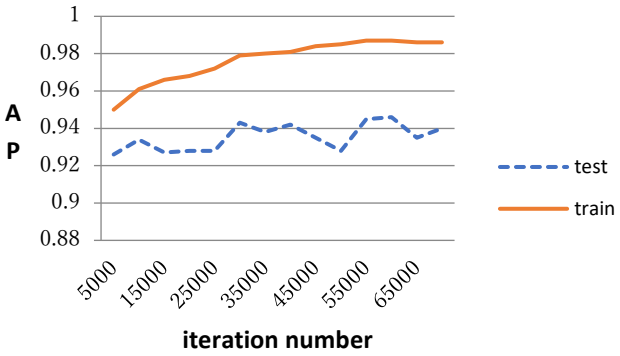
学習セットとテストセットを対象として，CNN 学習の反復回数によるそれぞれの漫画オブジェクト検出器の検出率の変化を調査する．反復回数による AP の変化を表したグラフを図 3-12 に示す．実験結果より，反復回数が 70000 回以上の時，学習セットのそれぞれのオブジェクトに対する検出率は収束することが確認できた．

3.6.4　閾値設定による漫画オブジェクトの検出

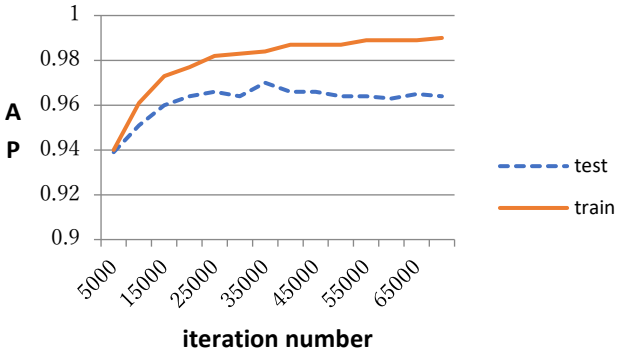
それぞれ 70000 回の学習を行った検出器を対象として, 検出スコアの閾値を設定して検出を行った場合におけるオブジェクトの検出結果を求める．コマ検出における検出スコアの閾値を 0.6 に設定し，フキダシとキャラクタ検出の閾値は 0.8 に定める．各オブジェクトの検出結果を表 3-3 に示す．この表において，”Total”は対象の漫画画像に含まれる各オブジェクトの総数を示す．また，”TP”, “FN”, “FP”はそれぞれ True Positive, False Negative, False Positive の値，R と P は Recall と Precision の値を示す．従来手法[7]によるコマ割り及びフキダシの検出結果を表 3-4 に示す．この結果より，Faster R-CNN を用いた漫画オブジェクト検出が従来手法の検出精度を上回ることが示された. 従来手法及び Faster R-CNN によるコマ検出結果の例を図 3-13 に示す．図よりコマが重なって描かれている場合において，従来手法ではコマを正確に検出できない一方で，Faster R-CNN ではそれぞれのコマを検出可能であることが確認できた．

3.6.5　コマ内容の認識

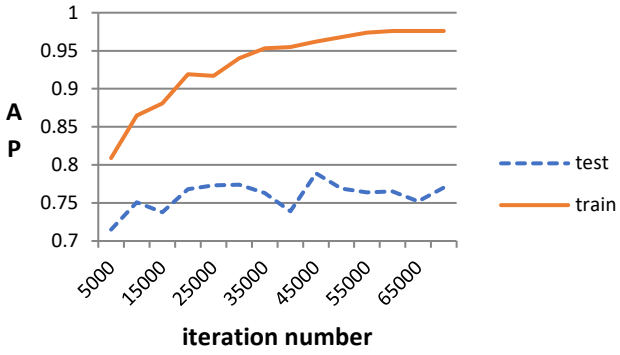
オブジェクトの位置情報より, 一つのコマ内に存在するフキダシとキャラクタの数を認識するシステムについて検討する. 検出されたコマ領域のバウンディングボックスに対して，50%以上重なっているフキダシ又はキャラクタのバウンディングボックスが存在するとき，該当のオブジェクトはそのコマに含まれているとする．ここで一部の漫画では，図 3-14 の例のようにあるコマが別のコマの内部に存在するため，同一のオブジェクトが複数のコマとオーバラップするという問題が見られる. この問題を解決するた



(a) コマ検出率の変化



(b) フキダシ検出率の変化



(c) キャラクタ検出率の変化

図 3-12: 学習の反復回数による検出率の変化（自発表[35]より引用）

め，検出されたコマ画像をソートし，ソート順がより後ろのコマへオブジェクトを所属させるアルゴリズムを提案する．コマのソートの順番は，コマ枠の上辺がページの上側にあるものから順番にソートするものとする．このとき，上辺が同じ位置にあるコマが

表 3-3: テストセットに含まれる漫画オブジェクトの個数と Faster R-CNN による検出率（自発表[35]より引用）

	Total	*TP*	*FN*	*FP*	*R* (%)	*P* (%)
Panel	859	770	90	40	89.5	95.1
Balloon	1190	1161	29	42	97.6	96.5
Character	937	803	134	50	85.7	94.1

表 3-4: テストセットに含まれる漫画オブジェクトの個数と従来手法[7]による検出率（自発表[35]より引用）

	Total	*TP*	*FN*	*FP*	*R* (%)	*P* (%)
Panel	859	481	378	183	56.0	72.4
Balloon	1190	790	400	650	66.4	54.9

複数存在する場合は，枠の右辺がページの右側にあるものから順にしてソートする．漫画画像 1 ページに対するコマのソーティングの例を図 3-15 に示す．

テストセットに含まれる 5 作品を対象に，コマ内容の認識を行った実験結果を表 3-5 に示す．この表において，”B”はテストセットに含まれる全てのコマのうち，フキダシの数を正確に抽出できたコマの割合，”C”はキャラクタの数を正確に抽出できたコマの割合，”B+C”はフキダシとキャラクタの数を両方とも正確に抽出できたコマの割合を示す．実験結果より，Comic E では他の作品と比較してコマの認識率が低くなった．この理由として，Comic E では図 3-16 に見られるような複雑な形状のコマ割りが多く存在するため，コマ自体の検出率が低下したことが挙げられる．また，本実験では画像 2000 枚という比較的少数のデータセットで検出器を学習しているため，特異な形状を持つオブジェクトへの対応が十分でなかったことが考えられる．したがって，今後はより大規模なデータセットについて学習することで，このような特異な表現に対する有効性を検討する必要がある．

3.7　漫画オブジェクト検出精度の比較

R-CNN のアルゴリズムをベースとして，候補領域の抽出方法を改良した物体検出手法が複数提案されている．これらの候補領域抽出の漫画画像に対する有効性を評価するため，漫画オブジェクトに対する検出精度を比較する．本実験では，Fast R-CNN，Faster

(a) 従来手法[7]によるコマ検出結果

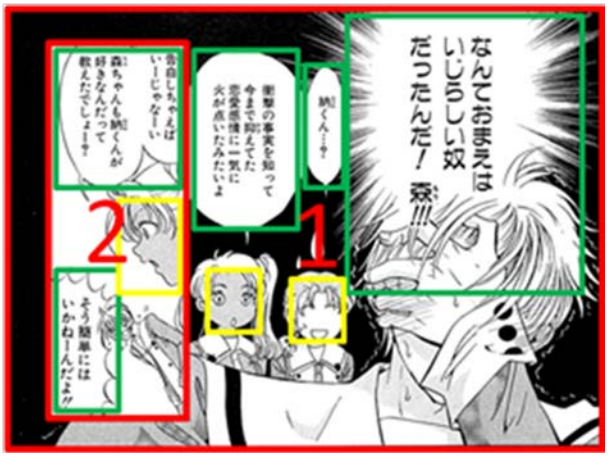
(b) Faster R-CNN によるコマ検出結果

図 3-13: 従来手法[7]と Faster R-CNN におけるコマの検出結果の比較（自発表[35]より引用）

R-CNN，SSD の 3 種類の手法について検討する．

3.7.1 データセット

本実験において，学習と評価に使用するデータセットは Manga109 より公開されている漫画画像を使用する．学習セットは，作者の異なる 19 作品よりそれぞれ 100 ページ



- コマ 1 は，フキダシ 3 個，キャラクタ 2 個を含む
- コマ 2 は，フキダシ 2 個，キャラクタ 1 個を含む

図 3-14: コマ内容の認識の例（自発表[35]より引用）

図 3-15: 漫画画像内のコマのソーティングの例（自発表[35]より引用）

表 3-5: 漫画オブジェクトの検出によるコマ内容の認識結果（自発表[35]より引用）

	B (%)	*C* (%)	*B* + *C* (%)
Comic A	83.0	74.5	68.1
Comic B	91.4	89.8	84.9
Comic C	81.7	72.8	66.3
Comic D	94.6	69.0	65.2
Comic E	62.3	62.9	52.8

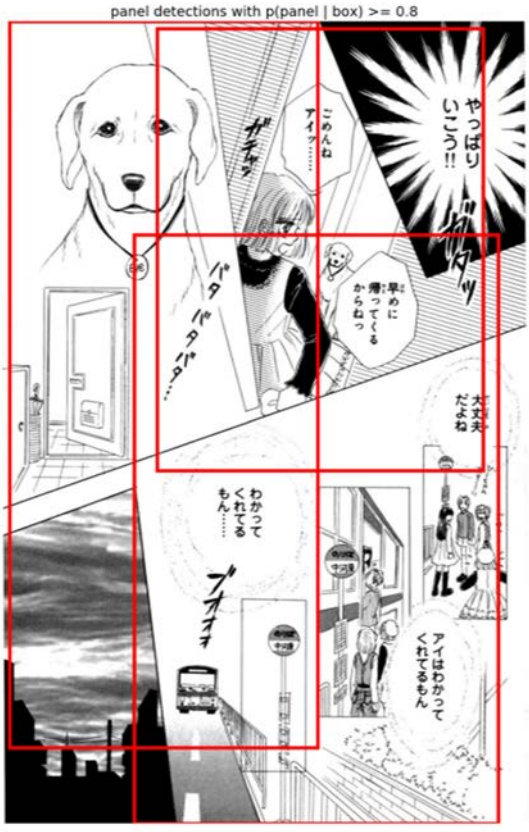


図 3-16: Faster R-CNN によるコマ検出に失敗した例（自発表[35]より引用）

ずつを抽出し，「コマ」，「フキダシ」，「キャラクタ」，「文字列」の 4 種類の要素についてそれぞれアノテーションを付与した 1900 枚の画像を使用する．テストセットは，学習セットに使用したものとは異なる漫画 5 作品からそれぞれ 30 ページを抽出した合計 150 枚の画像を使用する．

表 3-6: 漫画オブジェクトに対する 3 種類の検出器の比較（自発表[24]より引用）

	Fast R-CNN	***Faster R-CNN***	***SSD***
Panel layout	0.959	0.953	0.897
Speech balloon	0.969	0.961	0.907
Character face	0.810	0.816	0.765
Text	0.740	0.898	0.866
mAP	0.870	0.910	0.859

3.7.2　検出器のパラメータ設定

Fast R-CNN, Faster R-CNN, SSD のアルゴリズムはそれぞれ，[27]，[33]，[36]において公開されているプログラムを使用する．3.3 節の実験と同様に，それぞれのオブジェクトについて対象物体か否かの 2 クラス分類を行う 4 種類の検出器を学習して検出精度を比較する. 各手法の CNN モデルには VGG16[37]を使用し, 学習回数はそれぞれ 70000 回に設定する． VGG16 は 13 層の畳み込み層と 3 層の全結合層から構成されるネットワークであり，全結合層は 4096 個のユニットを持つ 2 層と出力層で構成される．その他のパラメータに関しては公開されているプログラムのデフォルトのものを使用する.

3.7.3　実験結果と考察

漫画オブジェクトの各クラスに対する AP と，4 クラスの AP の平均値である mean Average Precision (mAP)を表 3-6 に示す．コマとフキダシの検出においては，Fast R-CNN が Faster R-CNN を上回る結果を示した．一方，キャラクタと文字列の検出においては Faster R-CNN が Fast R-CNN を上回った．また，SSD の検出精度は他の手法よりも低くなった．3 種類の手法によるコマ検出の例を図 3-17，図 3-18，図 3-19 に，キャラクタ検出の例を図 3-20，図 3-21，図 3-22 に示す.

コマとフキダシに対する Fast R-CNN の検出率が高くなった理由として，この二つのオブジェクトが線で囲まれて描かれていることから，Selective Search によるセグメンテーションによって正確な抽出が可能であったためと推測できる．一方で，キャラクタや文字列といったオブジェクトには明確な枠線が存在しないため, セグメンテーションによる正確な領域抽出が困難であることが，検出精度の低下に繋がったと考えられる.

SSD の検出精度が低くなった理由には，図 3-22 で見られるように小さなオブジェクトに対する未検出が多かったことが挙げられる. SSD が小さな物体を対象とした分類タスクを苦手とすることは文献[26]においても指摘されている．小さな物体に対する SSD の検出精度を向上させるには, 小さな物体の検出処理を学習できるようにデータセット

を拡張することや，検出ボックスの敷き詰め方を改良するといった操作が必要となる．

3.8 むすび

本章では，漫画オブジェクトの検出における CNN の適用について検討を行った．

はじめに，本章の実験に使用する CNN 及び，CNN を用いた物体検出手法について概要を述べた．

次に，キャラクタ顔検出を対象として，顔検出に最適化した DPM と Fast R-CNN の検出精度を比較することで漫画画像に対する CNN の有効性を評価した．実験結果では，DPM の AP が 87.7%であるのに対して，CNN では 90.7%となった．このことから，CNN の導入により従来よりもシンプルな検出モデルでマルチビュー顔検出に対応できることが示された．

さらに，コマ割りとフキダシを含んだ漫画オブジェクトの検出に対して Faster R-CNN を適用し，従来手法との比較とコマ内容の認識について検討を行った．実験結果より，CNN がコマ割りやフキダシのような不定形な物体の検出に対しても有効であり，従来手法よりも柔軟な検出が可能であることを確認した．

最後に，CNN を用いた物体検出法を比較した．その結果，4 種類の漫画オブジェクトに対して Faster R-CNN が 91.0%の mAP を示し，他の手法より有効であることを確認した．

図 3-17: Fast R-CNN によるコマ検出の例（自発表[24]より引用）

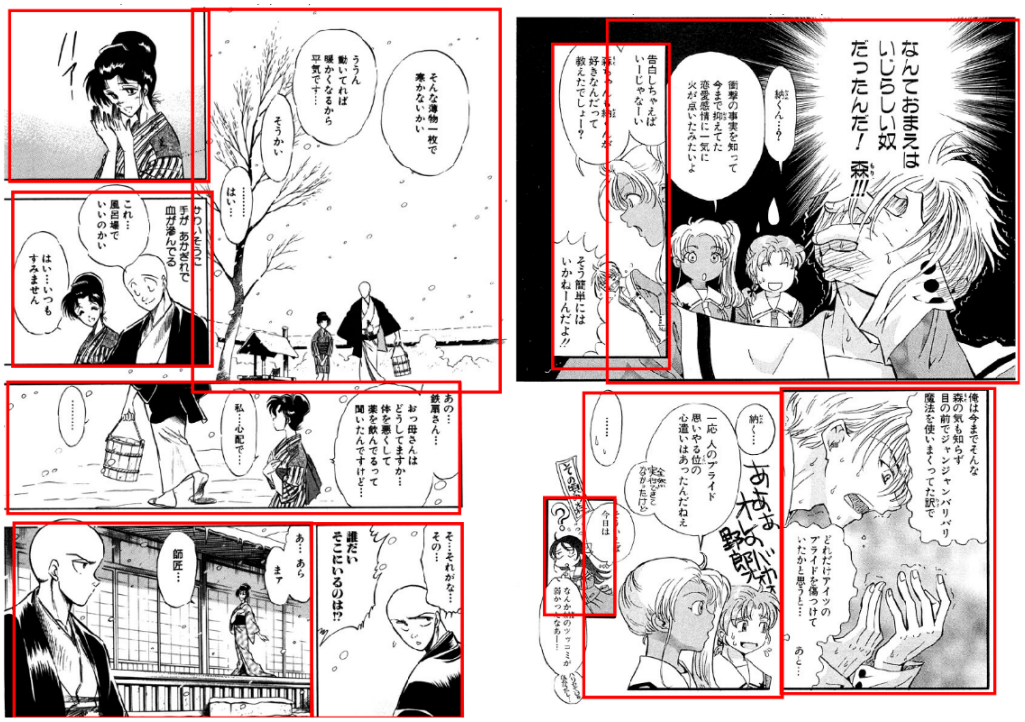


図 3-18: Faster R-CNN によるコマ検出の例（自発表[24]より引用）

図 3-19: SSD によるコマ検出の例（自発表[24]より引用）

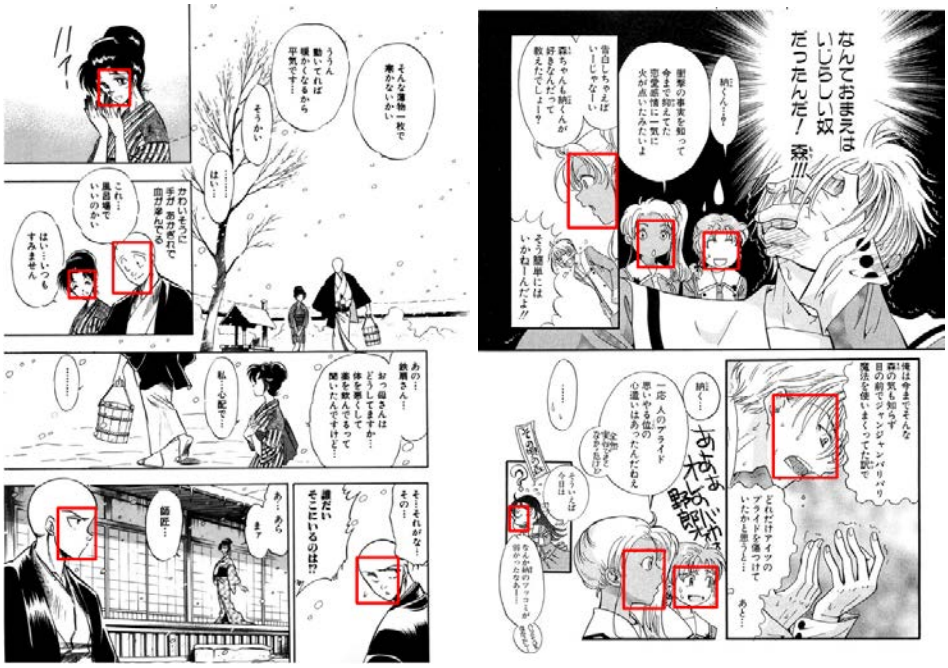


図 3-20: Fast R-CNN によるキャラクタ検出の例（自発表[24]より引用）

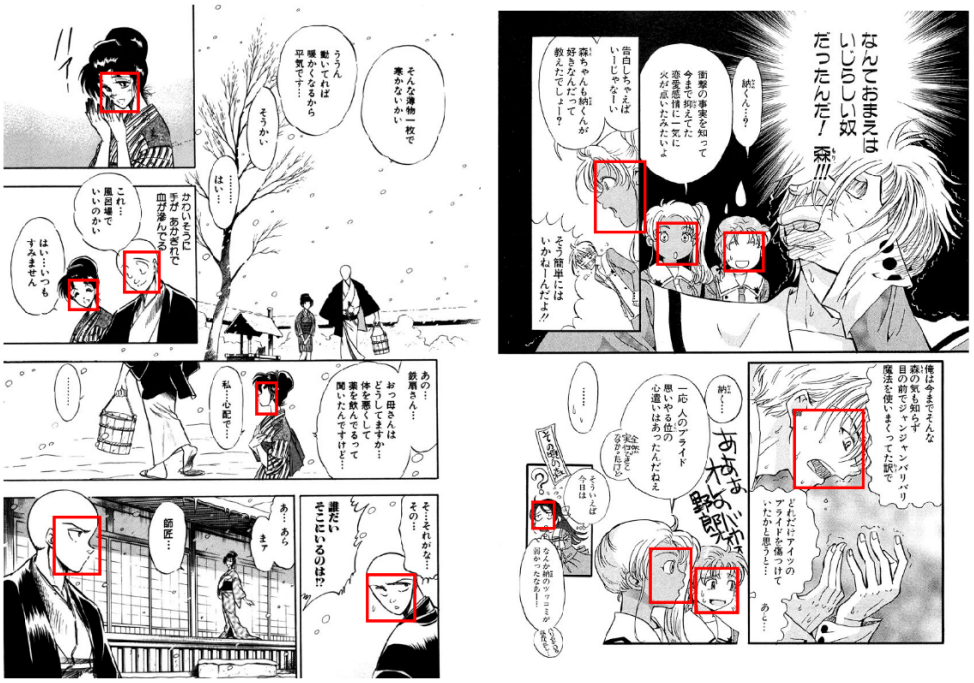


図 3-21: Faster R-CNN によるキャラクタ検出の例（自発表[24]より引用）

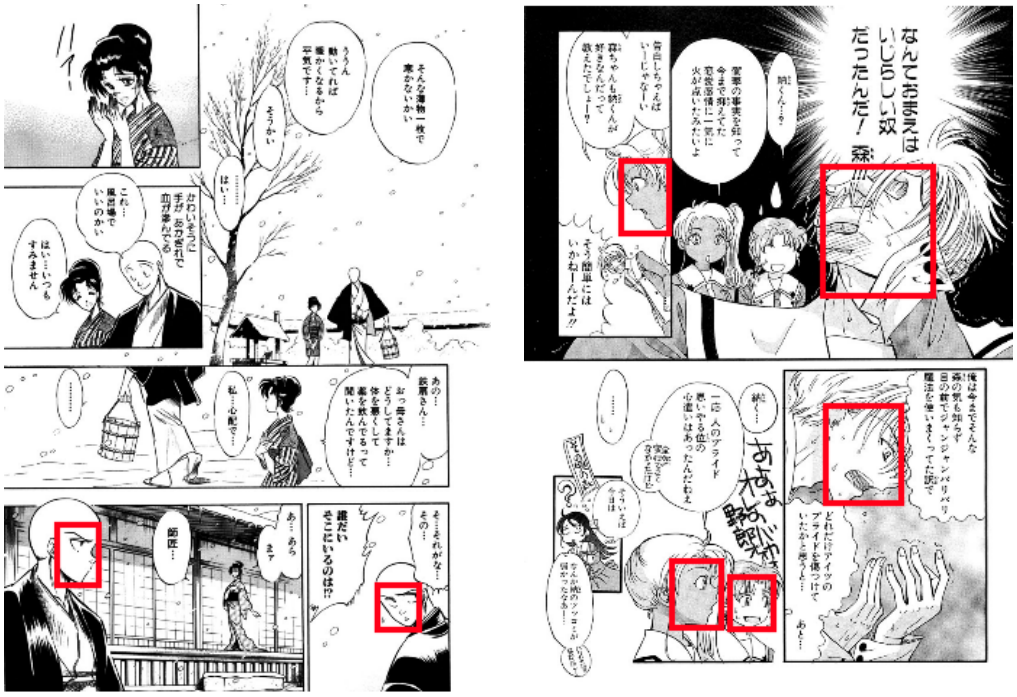


図 3-22: SSD によるキャラクタ検出の例（自発表[24]より引用）

第 4 章

クラスタリングを用いたキャラクタ同定

4.1 まえがき

本章では，x-means 法を用いた漫画キャラクタのクラスタリングによって，漫画の登場キャラクタを同定する手法を提案する．漫画キャラクタの認識は，画像からのキャラクタ位置の検出と，検出されたキャラクタ画像からキャラクタを同定する二つの処理によって成り立つ．このとき，未知の漫画画像からキャラクタを認識するには，事前知識に依存しない教師なしでのキャラクタ同定技術が必要となる．従来研究では，キャラクタ顔画像を k-means 法でクラスタリングすることで類似した画像を抽出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設定する必要があるため，登場キャラクタ数が未知の場合に最適なクラスタ数を定めることが困難であるという問題がある．本研究では，クラスタ数の設定を必要としない主要キャラクタの抽出を達成するために，x-means 法を用いたキャラクタ顔画像のクラスタリングを検討する．

まず，キャラクタ顔画像クラスタリングの従来手法の概要とその問題点について述べる，次に，提案手法の概要と提案手法で用いる技術について説明する．さらに，キャラクタ顔画像からの主要キャラクタの抽出精度について従来手法との比較から評価を行う．最後に，実験結果よりキャラクタ分類における課題を明らかにする．

4.2 主要キャラクタ同定の従来手法

既知の情報に依存しない漫画画像からの主要キャラクタの同定について，長尾らはクラスタリングによってキャラクタ顔画像を類似した特徴を持つクラスタに分類することで，主要キャラクタを抽出する手法を提案した[38]，従来手法の概要を以下に示す．

1. キャラクタ顔画像について Speeded-Up Robust Features (SURF)特徴量を計算する．
2. 算出された SURF 特徴量を Bag-of-Visual-Words (BoVW)に変換する．
3. k-means 法で BoVW のクラスタリングを行い，最もデータの数が多いクラスタに主要キャラクタが含まれるとして抽出する，

ここで，3.における k-means 法のクラスタリングの実行には，あらかじめ何個のクラスタに分類するかユーザが設定する必要がある．しかし，作品に関する事前知識が存在し

ない未知の漫画画像を対象とする場合には, 作品に何種類のキャラクタが登場するかという情報を事前に得られないという問題がある．従来手法では，この問題についてクラスタ数を登場キャラクタ数に対して十分に高い値に設定するという対応をとっている．しかし，登場キャラクタ数に対してクラスタ数が過大である場合には，正しく主要キャラクタを抽出することができない．そこで，より高精度なキャラクタ分類を実現するために，クラスタ数を自動決定する手法が必要となる．

4.3 提案手法

本研究では, クラスタ数の設定を必要としない主要キャラクタ抽出を実現するために, x-mean 法を用いたキャラクタ顔画像のクラスタリング手法を提案する．x-means 法は，k-mean 法によるクラスタリングの逐次繰り返しによって，最適なクラスタ数を自動的に求める手法である．提案手法の流れを以下に示す.

1. キャラクタ顔領域を切り出した画像を入力として，SURF 特徴量を計算する
2. SURF 特徴量を 500 次元の BoVW に変換する．
3. 算出された BoVW を x-means の入力として，クラスタリングを実行する．
4. 分割されたクラスタの中で Bayesian Information Criterion (BIC) [39]の値が最も小さいクラスタが，主要キャラクタの顔画像の集合であるとして抽出する．

以下の項では提案手法で使用する技術について詳細を述べる．

4.3.1 Speeded-UP Robust Features (SURF)

Lowe らは画像の照明変化や回転, 拡大縮小に不変な特徴量として SIFT(Scale Invariant Feature Transform)を提案した[40]．しかし，SIFT は計算コストが高いため，高速な計算処理に不向きであるという問題を持つ．そこで，SIFT の特徴を維持したまま高速に計算処理が可能な特徴量として SURF が提案された[41]．SURF 特徴量の計算処理は，画像からの特徴点検出と，各特徴点における特徴記述の処理に分けられる．

1. 特徴点の検出

SIFT は画像のスケール空間から画素の極値を求めることで特徴点を検出する．SIFT のアルゴリズムではガウシアンの差分(Difference of Gaussian: DoG)を用いてスケール空間を計算する．しかし，この計算処理は比較的遅いことから，SURF では高速化のためにガウシアンフィルタをボックスフィルタで近似して計算を行う．次に，検出された特徴点について, 極値が小さなものとエッジ上に存在するものを削除することで有効な特徴点を絞り込む.

2. 特徴量の計算

抽出された各特徴点について特徴量を計算する．まず，特徴点の周りの輝度変化より HOG 特徴量と同様の輝度勾配ヒストグラムを作成し，最も輝度変化が大きい向きを特

徴点の方向として求める．次に，特徴点の方向をベースとして輝度勾配ヒストグラムを再度求める．この操作によって回転に不変な特徴量が得られる．SURF では特徴量の周辺を 4×4 ブロックに分割し, 8 方向の輝度勾配を求めるため, 特徴量は 128 次元となる．さらに，特徴量を正規化することで照明変化に不変な特徴を求める．

4.3.2　Bag-of-Visual Words (BoVW)

BoVW は自然言語処理において用いられる Bag-of-Words (BoW)を画像認識に適用した手法であり，Bag-of-Keypoints (BoK)や，Bag-of-Features (BoF)とも表記される[42]. BoVW は局所特徴を各ベクトルに対応させることで, 画像における局所特徴の登場頻度をヒストグラム化した特徴ベクトルを作成する．この処理によって，複雑な画像特徴を一つのベクトルで表現することができるため, 特徴量の分類やクラスタリングにおいて有用な手法である．

BoVW は以下の手順で求められる．はじめに，認識対象の画像から局所特徴を抽出する．次に，抽出した特徴量をクラスタリングによってk個のクラスタに分類する．最後にクラスタリングされたk個のクラスタについてセントロイドを求め，それぞれのセントロイドとなるベクトルを Visual Word とする．本研究では，局所特徴量に SURF 特徴量, 特徴量のクラスタリング手法に k-means 法を使用し，クラスタ数$k = 500$として 500 個の Visual Words を求める．k-means 法の詳細については次の 4.3.3 項で述べる．

4.3.3　k-means 法

k-means 法は，非階層型クラスタリングのアルゴリズムであり，クラスタの平均を用いてデータを与えられたクラスタ数k個に分類する[43]. k-means 法の流れを以下に示す.

1. 初期状態として，各データ$x_i(x_1, x_2, \cdots, x_n)$にランダムにクラスタを割り当てる．
2. 割り当てたデータをもとに各クラスタの中心$V_j(j = 1, \cdots, k)$を計算する．
3. 各x_iと各V_jとの距離を求め，新たにx_iを最も中心との距離が近いクラスタに割り当てる．
4. 上記の処理で全てのデータのクラスタの割り当てが変化しなかった場合，又は変化量が一定の閾値を下回った場合には，クラスタリングが収束したと見なし処理を終了する．それ以外の場合は新たに割り当てられたクラスタから再度V_jを計算し，3.の処理を繰り返す.

k-means 法によるクラスタリング結果は，最初のクラスタのランダムな割り振りに依存する．この問題に対して，初期のクラスタの中心はなるべく離れている方がよいという考えに基づき, 重み付き確立分布を用いてクラスタの初期値を決定するアルゴリズムとして k-means++法が提案されている．本研究では，この k-means++法による初期化を用いて k-means クラスタリングを実行する．

4.3.4 x-means 法

x-means 法は k-means 法の拡張として Pelleg らによって提案されたアルゴリズムである[44]．この手法は，BIC による分割停止基準を設けて k-means 法のクラスタ分割処理を逐次繰り替えることで，最適なクラスタ数を自動決定することが可能である．本研究では，石岡らによって改良された x-means のアルゴリズムを使用する[45]．このアルゴリズムは，入力データの個数をn，次元数をpとしたとき，以下のように示される．

1. 十分に小さなクラスタ数k_0(特に指定しなければ 2)を定め，$k = k_0$として k-means を適用し，初期のクラスタ$C_1, C_2, \cdots, C_{k_0}$を得る．
2. クラスタ$C_i (i = 1, 2, \cdots, k_0)$について BIC を計算する．
3. クラスタC_iに対して$k = 2$として k-means を適用する．分割後のクラスタを$C_i{}^1, C_i{}^2$として，2 分割モデルの BIC の値を BIC’として求める．
4. $\mathrm{BIC} > \mathrm{BIC}'$ならば，2 分割モデルをより好ましいと判断し，2 分割の処理を継続する．$\mathrm{BIC} \leq \mathrm{BIC}'$ならば，2 分割しないモデルをより好ましいと判断し，処理を停止する．
5. 全てのクラスタC_iについて 2 分割の処理が終了したとき，クラスタリングを終了して得られたクラスタを出力する．

分割前のクラスタC_iにおける BIC の値は，C_iに含まれるデータx_iのp変量正規分布を式(4.1)と仮定して，式(4.2)から求められる．

$$f(\theta_i;\, x) = (2\pi)^{-p/2}|V_i|^{-1/2} exp\left[-\frac{1}{2}(x-\mu_i)^t {V_i}^{-1}(x-\mu_i)\right] \tag{4.1}$$

$$BIC = -2\log L\left(\widehat{\theta_i}; x_i \in C_i\right) + q\log n_i \tag{4.2}$$

ここで，$\widehat{\theta_i} = \left[\widehat{\mu_i}, \widehat{V_i}\right]$は$p$変量正規分布の最尤推定値とする．$\mu_i$は$p$次の平均値ベクトル，$V_i$は$p \times p$の分散・共分散行列である．$q$はパラメータ空間の次元数で，$V_i$の共分散を無視すれば(0 とおけば)，$q = 2p$である．

分割後のクラスタ$C_i{}^1, C_i{}^2$における BIC’の計算は以下のようになる．$C_i{}^1, C_i{}^2$のそれぞれに対して，パラメータ$\theta_i{}^1, \theta_i{}^2$をもつ$p$変量正規分布を仮定し，2 分割モデルにおいてデータx_iの従う確率密度を式(4.3)とおく．

$$x_i \sim \alpha_i\left[f\left(\theta_i{}^1; x\right)\right]^{\delta_i}\left[f\left(\theta_i{}^2; x\right)\right]^{1-\delta_i} \tag{4.3}$$

ここで，δ_iの値はx_iが$C_i{}^1$に含まれるとき 1，$C_i{}^2$に含まれるとき 0 である．また，α_iは式(4.3)を確率密度とするための基準化定数であり，この手法では以下の式(4.4)，式(4.5)によって近似された値が用いられる．

$$\alpha_i = 0.5/K(\beta_i) \tag{4.4}$$

$$\beta_i = \sqrt{\frac{\|\mu_1 - \mu_2\|^2}{|V_1| + |V_2|}} \tag{4.5}$$

このとき，$K(\cdot)$は標準正規化分布の下側確率とし，β_iは$f\left(\theta_i{}^1; x_i\right)$と$f\left(\theta_i{}^2; x_i\right)$の分離の

程度を表す指標である．BIC’の値は以下の式(4.6)で求められる．

$$BIC' = -2\log L\left(\widehat{\theta'_i}; x_i \in C_i\right) + q'\log n_i \tag{4.6}$$

ここで，$\widehat{\theta'_i} = \left[\widehat{\theta_i^1}, \widehat{\theta_i^2}\right]$は，二つの$p$変量正規分布の最尤推定値である．共分散を無視すれば，各pに対して平均と分散の二つのパラメータが存在するため，パラメータ空間の次元は$q' = 2 \times 2p = 4p$となる．

提案手法では, BoVW によって得られた 500 次元の特徴ベクトルに対する x-means 法の適用を考える．このとき，高次元データを x-means の入力とした場合には，式(4.2)及び式(4.6)における第 2 項の重みが大きく計算されるため，第 1 項の正規分布の影響が小さくなり，クラスタの分割処理が早い段階で終了してしまうという問題がある．したがって，提案手法では BIC と BIC’の値を，変数$c(c > 1)$を用いて式(4.7), 式(4.8)によって求める．

$$BIC = -2\log L\left(\widehat{\theta_i}; x_i \in C_i\right) + \frac{q\log n_i}{c} \tag{4.7}$$

$$BIC' = -2\log L\left(\widehat{\theta'_i}; x_i \in C_i\right) + \frac{q'\log n_i}{c} \tag{4.8}$$

また，提案手法は BIC の値が最も低いクラスタにはより類似度の高い顔画像が集中していると仮定して，式(4.7)によって求められた BIC の値が最も低いクラスタを主要キャラクタの顔画像集合として抽出する．ただし，データを 1 個のみ含むクラスタは抽出対象から除外する．

4.4 提案手法の評価

主要キャラクタ同定における提案手法の有効性を従来手法との比較より評価する．

4.4.1 テストセットの設定

クラスタリングの評価には Manga109 の漫画画像を使用する．作者の異なる漫画 3 作品よりそれぞれランダムにキャラクタ顔領域を切り出した画像 100 枚を用意し，これらを 200×200 ピクセルに正規化したデータセットを作成する．各作品において，データセットに顔画像が 10 枚以上含まれるキャラクタを主要キャラクタとして，登場回数が多い順にキャラクタ A~D と定義する．また，それ以外のキャラクタは全て「その他」と定義する．テストセットの内容を表 4-1 に示す.

4.4.2 パラメータ設定

従来手法のパラメータ設定は文献[47]を参考に，BoVW の次元数を 500，k-means 法によって分割するクラスタ数kを$k = 30$と設定する．また，提案手法は式(4.7)，式(4.8)におけるパラメータcを$c = 10, 20, 30$と設定し，それぞれのクラスタリング結果を比較す

表 4-1: テストセットの主要キャラクタ枚数（自発表[46]より引用）

タイトル	A	B	C	D	その他
BEMADER・P	35	13	11	−	41
ぶらり鉄扇捕物帳	52	35	−	−	13
爆裂！かんふー娘	27	19	17	16	21

表 4-2: 抽出されたクラスタの purity

	BEMADER・P	ぶらり鉄扇捕物帳	爆裂！かんふー娘
従来手法[40]	0.660	0.749	0.776
提案手法($c = 10$)	**0.661**	0.804	**0.794**
提案手法($c = 20$)	0.605	0.804	0.664
提案手法($c = 30$)	0.514	**0.879**	0.513

る．

4.4.3 実験結果

従来手法及び提案手法によって主要キャラクタを含むとして抽出されたクラスタについて，purity を比較することで評価する．対象のクラスタに含まれる顔画像の集合をC_i，正解クラスをA_hとしたとき，クラスタの purity P_iは式(4.9)より求められる．本実験では正解クラスのラベルを A ~ D からなる$h \in a, b, c, d$とおき，「その他」のクラスは正解ラベルに含まないものとする．

$$P_i = \frac{1}{|C_i|} \max_h |C_i \cap A_h| \tag{4.9}$$

各手法によって抽出されたクラスタの purity を表 4-2 に示す．実験結果より$c = 10$と設定した場合に，3 作品全てに対して提案手法の purity が従来手法を上回ることを確認した．提案手法($c = 10$)によって抽出されたクラスタについて，最大となったクラスの画像を Positive，それ以外のクラスの画像を Negative としたとき，True Positive，False Negative，False Positive，True Negative に分類した例をそれぞれ図 4-1，図 4-2，図 4-3 に示す．この結果より，キャラクタ顔画像のクラスタリングはキャラクタの髪や背景などのテクスチャに着目して分類される傾向が見られた．また，主要キャラクタと髪のテクスチャが類似したキャラクタが存在するとき，False Positive として抽出されることが確認できた．

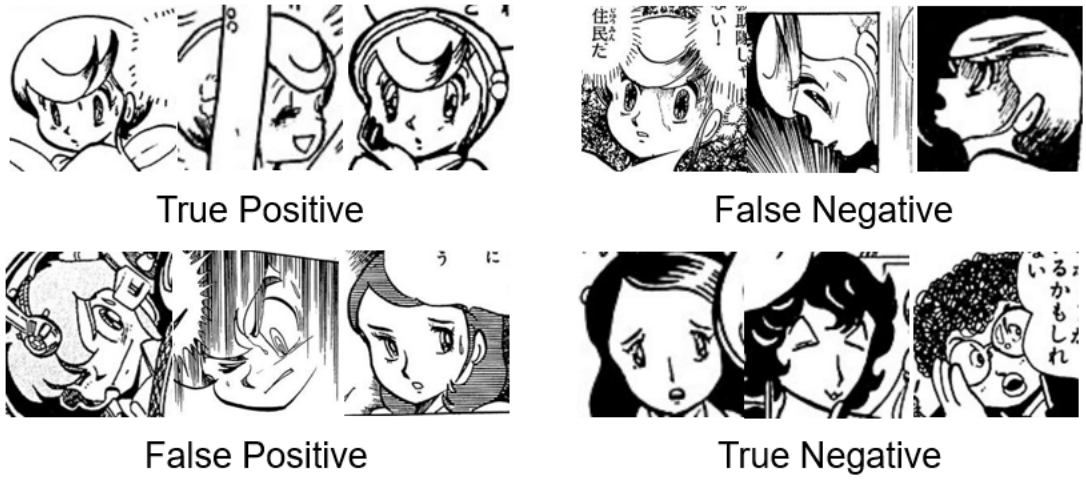


図 4-1: 提案手法(c=10)による「BEMADER・P」からのキャラクタ顔画像抽出結果

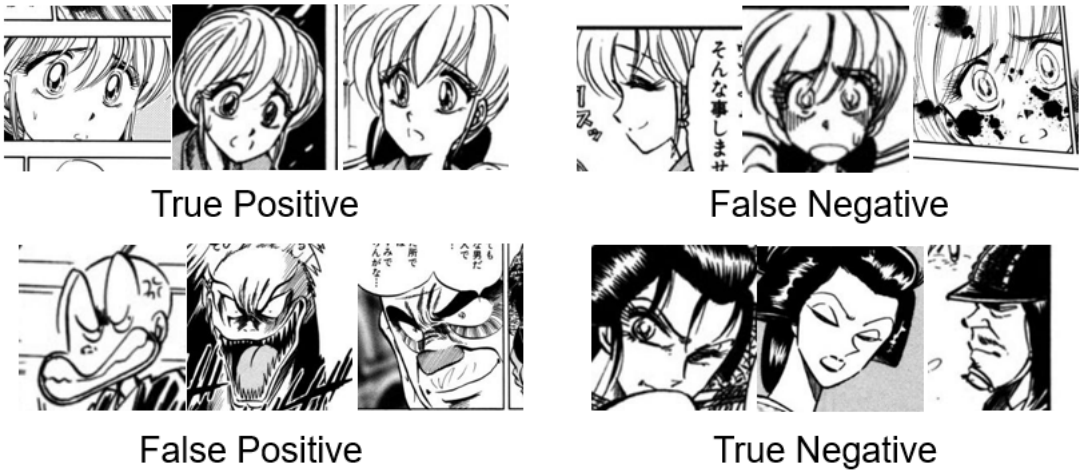


図 4-2: 提案手法(c=10)による「ぶらり鉄扇捕物帳」からのキャラクタ顔画像抽出結果

4.5 キャラクタの分類における課題

実験結果より, 提案手法が特定の主要キャラクタを抽出するにあたってある程度の有効性を持つことが確認できた．しかし，全ての主要キャラクタの分類を考えたとき，提案手法には以下の問題が存在する．

図 4-3: 提案手法(c=10)による「爆裂！かんふー娘」からのキャラクタ顔画像抽出結果

4.5.1 x-means 法のパラメータ設定の問題

登場キャラクタを分類するにおいては, キャラクタ顔画像が登場キャラクタの種類に近い値で分割されることが理想的なクラスタリング結果であるといえる. 提案手法では, BIC の計算式に調整可能なパラメータcを設けることでクラスタ数の調整を可能とした. しかし，パラメータcの値と生成されるクラスタ数の関係性は明確でないため，実際のクラスタリング結果から最適な値を検討する必要がある．このため，キャラクタ分類による登場キャラクタリストの自動生成を考えたとき, 有効なクラスタリング結果を得られるパラメータを設定することが困難であるという問題がある.

4.5.2 サブキャラクタのクラスタリングにおける問題

一冊の漫画に登場するキャラクタには，複数回登場する主要キャラクタの他に，1~2 回しか登場しないサブキャラクタが存在する．k-means 法や x-means 法によるクラスタリングでは, 入力された全てのデータがいずれかのクラスタに所属するように画像を分類する．しかし，画像枚数が数枚しか存在しない全てのサブキャラクタを個別のクラスタに割り当てることは困難である．このため，実際のクラスタリングではサブキャラクタが他の主要キャラクタと同じクラスタに割り当てられてしまい，クラスタの purity が低下する要因となっている．また，キャラクタの誤分類を改善するために，特徴量の改良や画像からの背景領域の除去についても検討が必要である.

4.6 まとめ

本章では，x-means 法を用いた主要キャラクタ同定手法の改良を提案した.

はじめに，キャラクタ顔画像クラスタリングの従来手法について概要を説明し，課題点を述べた．

次に，提案手法の概要を述べ，提案手法で使用する SURF 特徴量，BoVW，k-means 法，x-means 法について詳細を説明した．

さらに，従来手法との比較による提案手法の評価を行った．実験結果より，提案手法のパラメータ設定が適切であるとき，主要キャラクタ顔画像の抽出において提案手法の精度が従来手法を上回ることを示した．

最後に，キャラクタ顔画像の分類における課題を明らかにした．

第 5 章

主要キャラクタ顔画像の分類

5.1 まえがき

本章では，登場キャラクタリストの生成を目的として，キャラクタの顔画像から複数の主要キャラクタを分類する手法を提案する．4 章における実験では，未知の漫画画像を対象とした顔画像クラスタリングにおいて，クラスタ数の自動決定が分類精度の向上に有効であることが確認できた．しかし，主要キャラクタの分類への応用を考えたとき，x-means 法のパラメータ設定や，サブキャラクタへの対応といった課題が存在する．そこで，画像の特徴抽出とクラスタリング手法についてそれぞれ改良することで，主要キャラクタの分類手法を提案する．

はじめに，キャラクタ顔画像の特徴表現として，CNN の出力する深層特徴量の利用について述べる．次に，特徴量の次元削減について述べる．さらに，ノイズに頑強なクラスタリング手法である DBSCAN の概要を説明し，DBSCAN のパラメータの決定方法について述べる．次に，一般データセットについて学習した CNN モデルを特徴抽出器とした場合における，顔画像クラスタリングの有効性を評価する．さらに，顔画像の分類精度を向上させるための試みとして，顔画像の背景領域の除去によるクラスタリング精度の変化について検討する．最後に，キャラクタ顔画像について学習した CNN モデルを特徴抽出器するクラスタリング手法を提案し，有効性を評価する．

5.2 キャラクタ顔画像の特徴表現

4 章で提案した SURF 特徴量によるキャラクタ顔画像の認識では，髪型や背景など局所的な特徴が類似した画像について誤分類が行われるといった問題が確認された．この問題を解決するため，キャラクタ間の特徴をより正確に捉えることのできる特徴量が必要である．

大規模データセットを用いて学習された CNN のパラメータは高い汎用性を持ち，強力な特徴抽出器として関連する他タスクへ転用可能であることが知られている[48]．Guérin らは，ImageNet を用いて学習を行った CNN の中間層より得られる深層特徴量と古典的なクラスタリングアルゴリズムを組み合わせた画像分類手法を提案し，画像クラスタリングにおける最先端手法より高い精度を示したことを報告している[49]．このことから，CNN の出力を画像特徴量として使用することで，SURF 特徴量よりも有効な特

徴記述が可能であることが期待できる．また，成田らは漫画から特定のキャラクタを検索するシステムにおいて，キャラクタ顔画像と名前のペアを学習した CNN の出力を特徴量として使用することを提案している[50]．この研究より，キャラクタ顔画像について CNN を学習することで，漫画キャラクタ全般の認識に有効な特徴量を得ることが期待できる．

そこで，本研究では ImageNet を事前学習した CNN モデル及び，キャラクタ顔画像でファインチューニングを行った CNN モデルについて，キャラクタ顔画像の特徴抽出器としての有効性を検証する．

5.3 次元削減

高次元空間のデータを対象とした解析では，データの次元数が大きくなるほどデータ点同士の距離差が小さくなり，クラスタリングが困難となる球面集中現象と呼ばれる問題が発生する．この問題を解決するために，高次元データの関係性を保ったまま低次元に変換する処理である次元削減が用いられる．本研究では，以下の 4 種類の次元削減手法について検討する．

5.3.1 主成分分析 (Principal Component Analysis: PCA)

PCA は，多数の変数を持つデータをできるだけ情報が損失しないように少数個の変数で表現する手法である[51]．これは，データを縮約して低次元データ化するという意味を持つことから，PCA は次元削減にも用いられる．PCA による次元削減はデータの基底に対して直交変換を行い，低次元の新たな座標系を得る操作である．新たな座標系の成分は，データの分散を最大化するものから順に第 1 主成分，第 2 主成分，…と決定される．主成分の導出過程を以下に示す．

まず，データの個数をn，次元数をpとした入力データ$X = (x_1, x_2, \dots, x_n)$を考えたとき，$X$の分散共分散行列$S$は式(5.1)で示される．

$$S = \frac{1}{n}(X - \bar{X})^T(X - \bar{X}) \tag{5.1}$$

ただし，$\bar{X}$はXの標本平均である．次に，データXをp次元の単位ベクトルωによって低次元の座標系Yに射影する．

$$Y = X\omega \tag{5.2}$$

このとき，Yの分散共分散行列S_Yは，Xの分散共分散行列を使って以下のように求められる．

$$S_Y = \frac{1}{n}\omega^T S\omega \tag{5.3}$$

したがって，データを射影したときの分散が最大となる係数ベクトルωを求める問題は，$\omega^T S\omega$の最大化問題と等しい．ここで，ωが単位ベクトルという条件での最大化を考

えると，ラグランジュの未定常数法より，以下の標本分散共分散行列Sの固有値問題として解くことができる．ただし，λはラグランジュ乗数である．

$$S\omega = \lambda\omega \tag{5.4}$$

式(5.4)の固有値問題を解いて得られたp個の固有値を，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$としたとき，第 1 主成分，第 2 主成分はそれぞれ固有値$\lambda_1, \lambda_2, \ldots$に対応する固有ベクトル$\omega_1, \omega_2, \ldots$となる．

5.3.2　カーネル主成分分析 (Kernel PCA)

PCA はデータの線形関係に基づいて変換するため，データ集合がガウス分布に従う場合には有効だが，非線形な構造を持つ集合に対しては効果的な圧縮ができない．この問題に対して，Kernel PCA はカーネル法によってデータを一度高次元空間へ射影してから PCA を適用することで，非線形の変換を行う手法である[52]．カーネル法による高次元空間への変換について，以下に詳細を示す．

p次元の変数ベクトル$x = \left(x_1, x_2, \ldots, x_p\right)^T$を特徴空間に射影して，$r$次元の変数ベクトル$\Phi(x) = \left(\phi_1(x), \phi_2(x), \ldots, \phi_r(x)\right)^T$を得るとする．ただし，$r \gg p$である．特徴空間上に射影された$n$個のデータからなるデータ行列を以下のように表す．

$$Z_c = \left(\Phi_c(x_1), \Phi_c(x_2), \ldots, \Phi_c(x_n)\right)^T \tag{5.5}$$

このとき，データ行列Z_cの標本分散共分散行列S_cは以下のようになる．

$$S_c = \frac{1}{n} Z_c {Z_c}^T \tag{5.6}$$

ただし，Z_cはデータの中央化が行われた値であるとする．この特徴空間上のデータに対して PCA を実行すると，S_cの固有値問題$S_c\omega = \lambda\omega$となる．ここで，特徴空間上のデータ間の内積に基づく行列K_cを考える．

$$K_c = Z_c {Z_c}^T = \begin{bmatrix} \Phi_c(x_1)^T\Phi_c(x_1) & \Phi_c(x_1)^T\Phi_c(x_2) & \ldots & \Phi_c(x_1)^T\Phi_c(x_n) \\ \Phi_c(x_2)^T\Phi_c(x_1) & \Phi_c(x_2)^T\Phi_c(x_2) & \ldots & \Phi_c(x_2)^T\Phi_c(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_c(x_n)^T\Phi_c(x_1) & \Phi_c(x_n)^T\Phi_c(x_2) & \ldots & \Phi_c(x_n)^T\Phi_c(x_n) \end{bmatrix} \tag{5.7}$$

行列S_cの固有値問題を行列K_cで置き換えると，以下に示すK_cの固有値問題で表される．

$$K_c\alpha = n\lambda\alpha \tag{5.8}$$

ここで，αは固有ベクトルであり，$n\lambda\alpha^T\alpha = 1$を満たす．入力データの次元が高いほど，より高次の特徴空間へ射影する必要があるため，データ間の内積$\Phi_c(x_i)^T\Phi_c\left(x_j\right)$の計算量は増加する．しかし，特徴空間への射影にカーネル法K_cを用いるとき，内積の計算において式(5.9)の関係が成り立つため，計算量を入力空間の次元数に抑えることができる．

$$K_c\left(x_i, x_j\right) = \Phi_c(x_i)^T\Phi_c\left(x_j\right) \tag{5.9}$$

カーネル関数には，Gaussian 関数や Sigmoid 関数といった種類が存在するが，本研究

ではコサイン類似度を使用する．コサイン類似度の定式を式(5.10)に示す.

$$K(x,\ y) = \frac{xy^T}{\|x\|\|y\|} \tag{5.10}$$

5.3.3　t-Distributed Stochastic Neighbor Embedding (t-SNE)

PCA は，データ全体の分散に基づいて次元削減を行うことため，低次元空間の表現は「類似しないデータを遠くに配置する」ことを優先したものとなる．これに対して，t-SNE は局所的なデータ間の特性に基づいて次元削減を行うことで，「類似したデータを近くに配置する」ことを優先した表現を得る手法である[53]．t-SNE はデータの 2～3 次元への圧縮に効果的であり，主にデータの可視化に利用される．t-SNE の計算過程は以下のようになる．

高次元空間上のデータ$X = (x_1, x_2, \cdots x_n)$の低次元空間上のデータ$Y = (y_1, y_2, \cdots y_n)$への変換を考える．高次元空間上の点$x_i$から$x_j$の近さを条件付き確率$p_{j|i}$で以下のように表す.

$$p_{j|i} = \frac{exp\left(-\|x_i - x_j\|^2 / 2{\sigma_i}^2\right)}{\sum_{k \neq i} exp(-\|x_i - x_k\|^2 / 2{\sigma_i}^2)} \tag{5.11}$$

t-SNE では 2 点の距離に対称性を持たせるため，以下の同時分布p_{ij}を高次元空間におけるデータの類似度と定める．

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \tag{5.12}$$

次に，x_i, x_jに対応する低次元空間上の 2 点y_i, y_jの類似度q_{ij}を求める．低次元空間の類似度は，自由度 1 の student の t-分布を用いて以下で表される．

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq i}(1 + \|y_k - y_i\|^2)^{-1}} \tag{5.13}$$

ここで正規分布の代わりに裾の重い t-分布を使用することで，2 点間の距離が遠い場合にはより遠くに配置されるような低次元空間の表現が得られる．圧縮前の確率分布p_{ij}と圧縮後の確率分布q_{ij}について，それぞれカルバック・ライブラー(KL)情報量を計算し，二つの KL 情報量の差を最小化するようにy_i, y_jの値を更新する．この処理は，以下の損失関数Cを最小とするy_iの値を確率的勾配降下法で求めることで成される．

$$C = KL(P \parallel Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{5.14}$$

また，式(5.11)におけるパラメータσ_iは，以下の式(5.15)において$Perp(P_i)$の値がユーザの指定する perplexity の値と等しくなるように決定される．

$$Perp(P_i) = 2^{H(P_i)} \tag{5.15}$$

$$H(P_i) = -\sum_j p_{j|i} \log p_{j|i} \tag{5.16}$$

実用上における perplexity の値は，5～50 の範囲が推奨される．本研究では，perplexity の値を 30，反復回数を 1000 回と定めて実験を行う．

5.3.4 Uniform Manifold Approximation and Projection (UMAP)

UMAP はリーマン幾何学とダイス的位相幾何学に基づく理論的枠組みから構成される次元削減手法である[54]．この手法は t-SNE と同等のデータ圧縮を行うこと目的としており，t-SNE より高速な計算処理が可能である．また，t-SNE と異なり UMAP の次元削減は確かな数学的理論が保証されている．UMP のアルゴリズムは以下の三つの仮定に基づく．1)データはリーマン多様体上に一様に分布している．2)リーマン計量は局所定数である．3)多様体は局所連結である．これらの仮定より，データセットをリーマン多様体で近似することが可能である．さらに，多様体からファジー位相表現へ変換し，ファジー位相表現において，データセットとの距離が最短となる低次元データの表現を求める．

はじめに，実空間$\mathbb{R}^n$上のデータが乗っているリーマン多様体(M, g)を推定する．リーマン計量は局所定数であるという仮定を用いると，多様体上の距離d_Mは，ユークリッド空間の距離$d_{\mathbb{R}^n}$の定数倍として表すことができる．さらに，現実データXについて，$X_i \in X$の周りでは点はX_iとだけ連結していると仮定したとき，X_iの周りで見る 2 点X_j, X_kの距離d_iは以下のように定義される．

$$d_i(X_j, X_k) = \begin{cases} 0, & if\ j = k \\ d_M(X_j, X_k) - \rho, & if\ j = i \\ \infty, & o/w \end{cases} \tag{5.17}$$

ここで，ρはX_iと最近傍の点とのd_Mの値である．

次に，ファジー位相空間について考える．通常の集合Xは，要素xが属しているかどうかの情報のみを持つため，次のような membership 関数で表される．

$$\mu_X(x) = \begin{cases} 1, & if\ x \in X \\ 0, & o/w \end{cases} \tag{5.18}$$

一方，ファジーな集合Pでは，この値を$(0, 1]$に拡張して表現する．

$$\mu_P(x) = strength\ of\ menbership(x \in X) \tag{5.19}$$

ここでPは，「属する強さ」$0 < a \leq 1$を与えた場合に，その強さで属する値全てを集めた集合$X_a = P(a)$を返すような関数と見なせる．ファジー空間から距離空間への対応付け$FinReal$は，$([n], a) \in \Delta \times I$を$\Delta_a^n$と書いたとき，以下のように表される．

$$FinReal(\Delta_a^n) = (\{x_1, \cdots, x_n\}, d_a) \tag{5.20}$$

ただし，d_aは次のように定義される．

$$d_a(x_i, x_j) = \begin{cases} -\log(a), & if\ i \neq j \\ 0, & o/w \end{cases} \tag{5.21}$$

距離空間からファジー空間への対応付けを行う逆向きの関数$FinSing$を考えると，次のように定義される．

$$FinSing(Y)(\Delta_a^n) = \mathrm{Hom}(FinReal(\Delta_a^n), Y) \tag{5.22}$$

現実空間のデータXは先述したように距離空間に変換できるため，$FinSing$によってファジー空間で表現することが可能である．したがって，データ $X = (X_1, X_2, \cdots X_N) \subset \mathbb{R}^n$のファジー位相表現は以下で表される．

$$\bigcup_{i=1}^{N} FinSing(X, d_i) \tag{5.23}$$

UMAP の次元削減の処理は，高次元のデータXに対応する低次元データを$Y = (Y_1, Y_2, \cdots Y_N) \subset \mathbb{R}^d, (d \ll n)$としたとき，それぞれをファジー空間に変換して，そのクロスエントロピーを距離と定める．ここで，ファジー集合Pの membership 関数は以下で与えられる．

$$\mu(x) = \sup\{a \in (0,1] | x \in P(a)\} \tag{5.24}$$

したがって，同じ定義域$\mathrm{A} = \bigcup_{a\in(0,1]} P(a)$における，高次元データと低次元データの membership 関数をμ, νとすると，距離は次のように求められる．

$$C\big((A,\mu),(A,\nu)\big) = \sum_{a\in A} \mu(a) \log\left(\frac{\mu(a)}{\nu(a)}\right) + \big(1-\mu(a)\big) \log\left(\frac{\mu(a)}{\nu(a)}\right) \tag{5.25}$$

UMAP は入力データX が与えられたとき，t-SNE と同様に確率的勾配降下法を用いて，式(5.25)のクロスエントロピーCを最小にするような低次元表現Yを学習することで，局所的な類似度に基づいた低次元表現を得る．本研究では，学習の反復回数は 500 回と定める．

5.4 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN はある空間に点集合が与えられたとき，多くの隣接点を持つ点の集合を一つのクラスタとして抽出するクラスタリングアルゴリズムである[55]. DBSCAN は与えられたデータ点を，コア点，到達可能点，外れ値の 3 種類に分類する．コア点は自身の半径 ε 以内に minPts 個以上のデータ点が存在する点, 到達可能点は半径 ε 以内に minPts 個未満のデータ点しか存在しないが半径 ε 以内に 1 個以上のコア点が存在する点, 外れ値は半径 ε 以内に存在するデータ点が minPts 個未満かつ半径 ε 以内にコア点が存在しない点と定義する．全てのデータ点について分類を行った後に，お互いに半径 ε 以内に存在するコア点の集合をクラスタとして, 隣接している到達可能点をそれぞれのクラスタに割り当てることでクラスタリングを行う．DBSCAN は k-means 法と異なり，事前にクラスタ数を決定する必要がないほか, 外れ値に対してロバストであるという利点が

ある．

5.4.1　DBSCAN のパラメータ決定

DBSCANの実行にはminPtsとεの二つのパラメータを正しく設定することが重要である．minPtsは望まれる最小のクラスタサイズであり，有効なクラスタリングを行うためには3以上の値に設定すればよい．ただし，より高い値に設定する方がノイズデータの分離に効果的である．

εはデータ点の結びつきを決定するパラメータである．εが非常に小さい場合にはデータの大部分は外れ値としてクラスタリングされない．また，大きな値の場合にはクラスタは併合され，データの大多数は同一のクラスタに存在することとなる．εの値を求める方法として，$k = \mathrm{minPts}$とおき，各データ点におけるk番目の最近傍点への距離をプロットしたグラフの解析が考えられる． Soniらは点群の密度が一様でないデータセットを対象として，クラスタリングに有効なεの値を求める手法として，Automatic Generation of Eps for DBSCAN (AGED)を提案している[56]．AGEDのアルゴリズムの概要を以下に示す．まず，次元数dのデータ点xについて，1 ~ k番目の近傍点までのユークリッド距離を計算し，その平均値をAverage Local Density Function (ADEN)として求める．

$$ADEN(x, y_1, \cdots, y_k) = \frac{\sum_{j=1}^{k} \sqrt{\sum_{i=1}^{d} \left(x_{(i)} - y_{(i)_j}\right)^2}}{k} \tag{5.26}$$

次に，全てのデータ点より求めたADENの値を正規化し，ビニングによって10個のバケットを持つヒストグラムに変換する．最後に，作成されたヒストグラムからデータ数がk以上となるバケットを抽出し，それぞれのバケットにおけるADENの平均値をεの候補値とする．AGEDは上記の過程によって得られたεの候補値をDBSCANクラスタリングに適用し，その結果を検証することでクラスタリングに最適な値を決定する．

本研究では，クラスタリングの自動化を目的として，AGEDによって求められた候補から一意なεの値を決定する方法を検討する．漫画1冊に登場する主要キャラクタの割合はそれぞれ異なることから，それぞれのデータセットについてデータ間距離の傾向からεを決定する方法を考えた．顔画像から抽出した特徴量より求めたADENのグラフと，AGEDによって提案されたεの候補値を示した例を図5-1に示す．図における青線は，各データ点におけるADENの値をソートした結果を表し，赤線はAGEDによって求められたεの候補値を示す．ここで，キャラクタ顔画像について理想的な特徴抽出が行われている場合には，それぞれの主要キャラクタの画像間の類似度は近くなり，主要キャラクタとサブキャラクタの間の類似度は遠くなると推定できる．したがって，ADENグラフにおいて勾配変化が起きる点が，主要キャラクタとサブキャラクタとの間でのデータ分

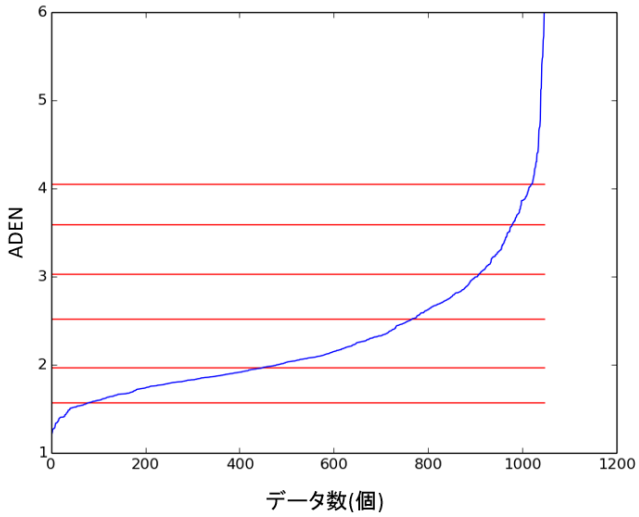


図 5-1: 漫画 1 冊の顔画像より算出した特徴量の ADEN グラフ（文献[57]より引用）

布の変化に相当すると仮定した．この仮定に基づき，隣接するAGEDのε候補についてそれぞれ傾きを求め，2点間の傾きが全ての候補間の傾きの平均値より大きく変化する箇所が最適なεの値であるとして抽出する手法を提案する．

5.5　一般 CNN モデルを用いた DBSCAN クラスタリング

主要キャラクタの分類について，一般画像データベースに対して学習を行ったCNNモデルとDBSCANによるクラスタリングの有効性を評価する．

5.5.1　特徴抽出器の設定

本実験では，ImageNetを学習した重みを適用したCNNモデルを特徴抽出器とする．CNNモデルはVGG16，VGG19[37]，InceptionV3[58]，Xception[59]，ResNet50[60]の5種類を使用する．文献[49]に基づき，各CNNモデルの最終層の一つ前の層の出力を特徴量として使用する．

5.5.2　データセット

本実験に使用する漫画画像は，Manga109で公開されている作品の中で作者の異なる11冊を使用する．データセットに用意されたアノテーションに従い，バウンディングボックスのサイズが30×30 pixel以上の顔領域を切り出してテストセットを作成する．このとき，元のアノテーションにはキャラクタの髪や耳の領域が含まれないため，顔領域の横幅，縦幅をそれぞれ均等に2倍に拡張した矩形領域を切り出す，また，矩形領域がページからはみ出す場合には，矩形をページに収まるようクリッピングする．次に，顔画

表 5-1: テストセットにおけるクラス数と画像数（自発表[57]より引用）

タイトル	クラス数	画像数
ARMS	8	319 (122)
愛さずにはいられない	8	972 (141)
あっけら貫刃帖	10	704 (155)
あくはむ	8	1047 (64)
青すぎる春	10	552 (80)
天晴れ！カッポーレ	8	864 (75)
ありさ 2	8	960 (156)
BEMADER・P	14	1226 (123)
爆裂！かんふー娘	11	1036 (172)
ベルモンド	12	847 (44)
ラブひな 14 巻	11	1168 (43)

像をキャラクタ別に分類し, 画像枚数が20枚以上のものを主要キャラクタとして個別のクラスに割り当てる．そして，画像枚数が20枚未満のキャラクタは全て「その他」のクラスに割り当てる．テストセットの内容を表5-1に示す．この表において，クラス数はそれぞれ1個の「その他」のクラスを含んだ値であり，括弧内は「その他」のクラスに属する画像数を示す．CNNへの入力時には，各画像のサイズを224×224にリサイズする．

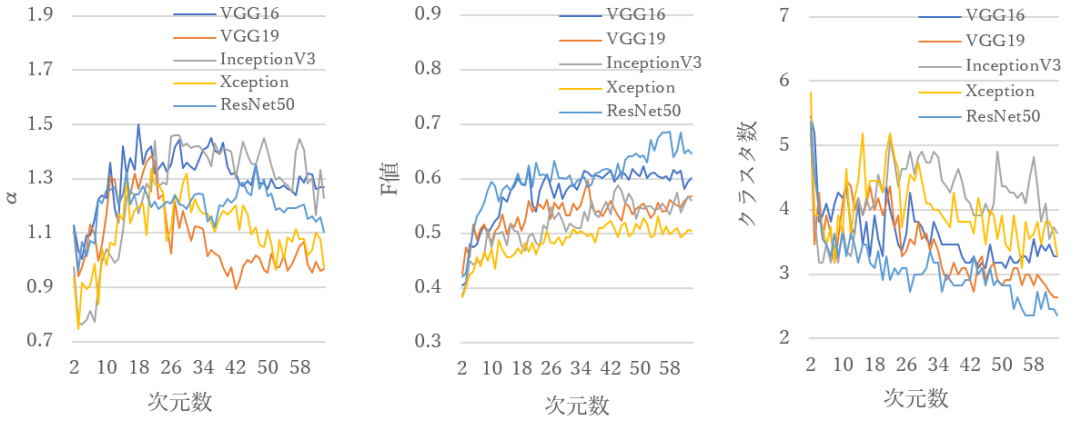
5.5.3 評価基準

DBSCANによるクラスタリングでは，抽出されるクラスタの他に外れ値が存在する．本実験におけるクラスタリング精度の評価では, クラスタとして抽出されたデータを対象として，以下のpurityとinverse purityを求める．

$$purity(\Omega, \mathbb{C}) = \frac{1}{N}\sum_{k} \max_{j} \left|\omega_k \cap c_j\right| \tag{5.27}$$

$$inverse\ purity(\Omega, \mathbb{C}) = \frac{1}{N}\sum_{j} \max_{k} \left|\omega_k \cap c_j\right| \tag{5.28}$$

ここで, Nは得られたクラスタの総データ数, $\Omega = \{\omega_1, \omega_2, \cdots, \omega_k\}$は生成されたクラスタの集合，$\mathbb{C} = \{c_1, c_2, \cdots c_j\}$は正解クラスの集合を表す．このとき，「その他」に属する画像はノイズデータとして，$\mathbb{C}$から除外している．purityは「異なるキャラクタが同じクラスタに入らない」ことを評価する基準，inverse purityは「同じキャラクタが異なるクラスタに入らない」こと評価する基準である．生成されたクラスタの正確さを評価する



(a) 次元数とαの変化 (b) 次元数と F 値の変化 (c) 次元数とクラスタ数の変化

図 5-2: データの次元数を 2~64 次元に変化させたときのクラスタリング結果の変化（自発表[57]より引用）

指標として，purityとinverse purityの調和平均であるF値を参照する．

$$F値 = \frac{1}{0.5 \cdot \left(\frac{1}{purity} + \frac{1}{inverse\ purity}\right)} \tag{5.29}$$

生成されるクラスタの個数が少ないほどinverse purityの値は大きくなりやすいため，同時にF値も大きく計算される傾向にある．ここで，本研究は主要キャラクタの抽出によるリストの生成を目的とすることから，可能な限り多数の主要キャラクタがそれぞれ異なるクラスタとして抽出される状態が望ましいクラスタリング結果といえる．したがって，purityとinverse purityに加えて生成されたクラスタ数を考慮した独自の評価基準αを定め，DBSCANクラスタリングの評価において最も重視する基準とおく．

$$\alpha = クラスタ数 \times purity \times inverse\ purity \tag{5.30}$$

5.5.4 データ次元数の評価

データの次元数がクラスタリング結果に及ぼす影響について評価する．本実験では，次元削減手法にKernel PCAを使用し，2次元から64次元まで変化させた場合におけるクラスタリング結果を比較する．DBSCANのパラメータは，minPtsを20と設定し，εの値をADENの最小値から最大値に到達するまで1/100の間隔で変化させて，それぞれαの値が最大となるクラスタリング結果を比較する．実験結果を図5-2に示す．この結果より，次元数が高いほどF値は大きくなり，生成されるクラスタの数は少なくなる傾向が確認できた．このような結果が得られた理由として，高次元データではデータ間の差異をより詳しく表現できる反面，球面集中現象の影響により，類似したデータをクラスタとして

表 5-2: 特徴抽出器として使用する CNN のモデルと次元削減の手法によるクラスタリング結果の変化.（その 1）

		次元削減なし	PCA 64 次元	PCA 16 次元	PCA 2 次元
VGG16	α	0.711	0.814	0.911	0.900
	F 値	0.634	0.585	0.503	0.427
	クラスタ数	1.46	2.18	3.00	3.82
VGG19	α	0.665	0.837	0.939	0.868
	F 値	0.580	0.561	0.514	0.411
	クラスタ数	1.82	2.36	3.27	4.55
InceptionV3	α	0.429	0.536	0.634	0.718
	F 値	0.519	0.544	0.503	0.368
	クラスタ数	1.27	1.36	2.00	4.09
Xception	α	0.519	0.612	0.711	0.925
	F 値	0.555	0.536	0.505	0.400
	クラスタ数	1.36	1.73	2.46	4.64
ResNet50	α	0.828	0.961	1.145	1.135
	F 値	**0.646**	0.596	0.560	0.440
	クラスタ数	1.82	2.36	3.35	5.82

抽出することは難しいためであると考えられる．また，αの値は16次元付近でピークをとることが分かった.

5.5.5　画像特徴量と次元削減の評価

特徴抽出器として使用するCNNモデルと次元削減手法の組み合わせによるクラスタリング結果の変化について評価する．次元圧縮を行わない元データ，PCAによって64次元，16次元，2次元へ圧縮したデータ，Kernel PCA (KPCA)によって64次元，16次元，2次元へ圧縮したデータ，t-SNEによって2次元へ圧縮したデータの8種類について比較する．DBSCANのパラメータ設定は，5.5.4項と同様とする．また，t-SNEについては乱数の影響を考慮して10回のクラスタリングを行った平均値を求める.

実験結果を表5-2，表5-3に示す．5.5.4項と同様に，全体的にデータの次元数が高いほどF値が高く, クラスタ数は少なくなる傾向が見られた. 次元削減の手法を比較すると，t-SNEを用いたとき，他の手法によって2次元に圧縮した場合と同等以上のF値でより多くのクラスタを得ることができることを確認した．このことから，DBSCANの適用にお

表 5-3: 特徴抽出器として使用する CNN のモデルと次元削減の手法によるクラスタリング結果の変化.（その 2）

		KPCA 64 次元	KPCA 16 次元	KPCA 2 次元	t-SNE 2 次元
VGG16	α	1.267	1.376	1.128	1.896
	F 値	0.602	0.586	0.405	0.494
	クラスタ数	3.273	3.82	5.455	7.41
VGG19	α	0.966	1.242	1.114	1.848
	F 値	0.567	0.511	0.424	0.483
	クラスタ数	2.636	4.09	5.273	7.54
InceptionV3	α	1.230	1.255	0.975	1.560
	F 値	0.559	0.504	0.386	0.471
	クラスタ数	3.636	4.18	5.273	6.33
Xception	α	0.973	1.135	0.939	1.506
	F 値	0.505	0.461	0.382	0.436
	クラスタ数	3.27	4.18	5.82	7.24
ResNet50	α	1.103	1.206	1.120	**2.149**
	F 値	**0.646**	0.592	0.420	0.499
	クラスタ数	2.36	3.45	5.36	**7.98**

いて， PCAのようなデータ全体の分散に着目した圧縮より，データの局所的な類似度に着目した圧縮の方が有効であることが示された.

特徴量の比較では，ResNet50を用いたとき最もαの値が高くなった．ResNet50によってクラスタとして抽出された顔画像の例を図5-3に示す．図において，(a)の左側と中央の画像は同一のキャラクタだが，右側の画像は異なるキャラクタである．また，(b)はそれぞれ異なるキャラクタである. 深層特徴量によるクラスタリングに共通した傾向として, キャラクタの髪や背景のテクスチャの類似性に基づいて顔画像の分類を行うことが確認できた．このことから，異なるキャラクタが類似した髪型や背景を持つ場合に誤ったクラスタが作られる例が見られた．また，ResNet50とそれ以外のCNNによる結果を比較すると, ResNet50では髪に着目したクラスタが抽出される割合が他のCNNよりも高いことが，クラスタリングの精度を向上させたと考察できる.

5.5.6 DBSCAN のパラメータ設定

AGEDによって求められるεの候補から最適な値を決定する方法を検討する．5.5.5項

(a) 髪型が類似した画像を抽出したクラスタの例

(b) 背景が類似した画像を抽出したクラスタの例

図 5-3: 一般 CNN モデルによって同じクラスタとして抽出された顔画像の例（自発表[57]より引用）

表 5-4: パラメータεの決定方法によるクラスタリング結果の変化（自発表[57]より引用）

	理想値	AGED 理想値	AGED 平均値	AGED 中央値	提案手法	シルエット
purity	**0.543**	0.533	0.492	0.507	0.498	0.319
inverse purity	0.482	0.510	0.517	0.537	0.472	**0.757**
クラスタ数	**7.98**	6.64	6.12	5.96	6.29	2.15
α	**2.149**	1.821	1.634	1.621	1.743	0.519
F 値	0.499	0.506	0.481	**0.501**	0.468	0.435

の実験において最も高いαの値を示したResNet50とt-SNEの組み合わせを用いて検証を行う．5.5.5項の実験で求めたαの値が最大となる結果を「理想値」とおく．また，AGEDによって求められるεの候補の中で, クラスタリング結果のαの値を最大とするものを選択した結果を「AGED理想値」とおく．隣り合ったAGEDの候補間の傾きを求め，傾きが候補間全体の傾きの平均値より大きく変化する点をεに設定する方法を「提案手法」とする．さらに，提案手法の有効性を確認するため，AGEDの候補の平均値をεに設定する方法を 「AGED平均値」，中央値をεに設定する方法を「AGED中央値」として比較

した．また，文献[61]において提案されているDBSCANのパラメータの自動決定の方法として，クラスタリング結果のシルエットスコアを最小化するεを最適値とする手法を「シルエット」として合わせて評価する．

各漫画作品について10回のクラスタリングを行った平均値を表5-4に示す．実験結果より，提案手法を平均値や中央値に設定した場合と比較すると，inverse purityが低い反面，ほぼ同等のpurityでより多くのクラスタを抽出することが可能であり，αの値は高くなることが示された．このことから，データ間距離の勾配変化からεを決定する方法が有効性を持つことを確認した．また，シルエットスコアを使用した場合では少数のクラスタのみが抽出され，purityの値も低くなることから今回のクラスタリングには適さないことが分かった．

5.6 クラスタリングにおける背景除去の影響

画像処理によってキャラクタ顔画像から背景の除去を行った場合におけるクラスタリング結果の変化について調査する．本実験では，特徴抽出器として，漫画キャラクタを学習したCNNモデルを使用する．キャラクタ顔画像について異なる画像処理を適用した画像を用意し，k-means法によるクラスタリング結果を比較することで，背景削除による画像特徴表現の変化を評価する．

5.6.1 特徴抽出器の設定

ImageNetの初期重みを適用したVGG16モデルを用意し，漫画キャラクタと名前のペアについて学習を行ったモデルを生成する．学習に用いるデータセットの詳細は5.6.2項で述べる．学習の反復回数は200 epochに設定する．生成されたモデルについて，5.5.1項と同様に，最終層の一つ前の層の出力を特徴量とする．抽出された特徴量をPCAによって100次元に次元削減したデータを，k-meansの入力とする．

5.6.2 データセット

CNNモデルの学習及びクラスタリングの評価には，Manga109の画像を使用する．5.5.1項と同様にアノテーションデータに従ってキャラクタ顔画像を切り出す．学習セットには作者の異なる93作品に含まれるキャラクタのうち，10回以上登場するキャラクタを抽出し，別々のクラスとして分類したデータセットを作成する．学習セットの画像数は83,186枚，クラス数は1,222個である．

テストセットは，学習セットとは異なる漫画11作品より，10回以上登場するキャラクタを抜き出したセットと30回以上登場するキャラクタを抜き出したセットの2種類を作成する．2種類の画像セットの内容を表5-5に示す．さらに，2種類の画像セットについてそれぞれ以下の3種類の画像処理を適用した画像を生成する．パターン1は，学習セットと同様の顔領域切り出しを行った画像，パターン2は，パターン1で切り出した領域につ

図 5-4: 3 種類の画像処理を行ったキャラクタ顔画像の例（自発表[62]より引用）

表 5-5: 登場回数 10 回以上のキャラクタ及び登場回数が 30 回以上のキャラクタを抽出した画像セットの内容

タイトル	登場回数が 10 回以上のキャラクタ		登場回数が 30 回以上のキャラクタ	
	クラス数	画像数	クラス数	画像数
ARMS	13	286	2	74
愛さずにはいられない	10	885	6	811
あっけら貫刃帖	11	578	7	506
あくはむ	8	999	6	960
青すぎる春	11	493	6	400
天晴れ！カッポーレ	10	826	6	764
ありさ [2]	10	856	5	756
BEMADER・P	15	1130	13	1103
爆裂！かんふー娘	14	915	9	840
ベルモンド	12	819	8	719
ラブひな 14 巻	10	1125	9	1098

いて, キャラクタの全身のアノテーションに従って背景のトリミング処理を行った画像, パターン3は, パターン2の画像に対してSelective Searchのセグメンテーションを適用し, 顔領域のアノテーションとオーバラップした部分を持つセグメンテーション領域だけを抽出した画像とする．それぞれのパターンによって抽出された画像の例を図5-4に示す．これによって6種類のテストセットを作成し，クラスタリング結果を評価する．

表 5-6: 10 回以上登場するキャラクタを抜き出したテストセットに対するクラスタリング結果（自発表[62]より引用）

	パターン 1	パターン 2	パターン 3
F 値	**0.462**	0.449	0.434
NMI	**0.435**	0.419	0.407

表 5-7: 30 回以上登場するキャラクタを抜き出したテストセットに対するクラスタリング結果（自発表[62]より引用）

	パターン 1	パターン 2	パターン 3
F 値	0.557	**0.558**	0.540
NMI	0.432	**0.443**	0.418

5.6.3 評価基準

本実験では，データセット全体のクラスタリングを目的として，k-meansで分類された全クラスタを対象としてF値及び，正規化相互情報収集量(NMI)を求める．purity, inverse purity, F値の導出式はそれぞれ式(5.27)，式(5.28)，式(5.29)と同様である．NMIの計算式は式(5.31)のようになる．

$$NMI(\Omega, \mathbb{C}) = \frac{I(\Omega, \mathbb{C})}{[H(\Omega) + H(\mathbb{C})]/2} \tag{5.31}$$

ここで，$\Omega = \{\omega_1, \omega_2, \cdots, \omega_k\}$はクラスタのラベル，$\mathbb{C} = \{c_1, c_2, \cdots c_j\}$はデータセットの正解ラベルを表す.

5.6.4 背景削除の評価

k-means法のクラス数kの値をテストセットの正解クラス数と同数に設定し，6種類のテストセットに対するクラスタリング結果を比較した．実験結果を表5-5, 表5-6に示す．実際にクラリングされた画像群を確認すると，パターン1の切り出しを行った場合には，図5-5のように異なるキャラクタ顔画像が特徴的な背景情報を持っているときに，誤って同一のクラスタとして抽出する例が見られた．一方，パターン2の処理を行った場合では，これらの画像は別々のクラスタに分けられた．このことから，画像の背景削除を行うことで，誤分類の改善に効果があることが確認できた．しかし，全体的なクラスタ

図 5-5: パターン 1 の処理を行った画像のうち，誤って一つのクラスタとして抽出された画像例

リング精度を見ると，表5-5で示されたように10回以上登場するキャラクタを対象とした場合では，パターン2の精度がパターン1より低下していることが分かる．

「ベルモンド」に10回以上登場するキャラクタを対象にパターン1の画像切り出しを行い，CNNの出力した特徴量をt-SNEで可視化した例を図5-6に示す．同様にパターン2の処理を行い，特徴量を可視化した例を図5-7に示す．二つの図において，同色のプロットはそれぞれ同じキャラクタを示す．この図より，パターン2では，パターン1よりクラスごとのデータの分散が大きくなっていることが確認できた．この理由としては，図5-5のような顔が一部しか入っていない画像から背景を除去した場合に，キャラクタの認識に十分な画像特徴を含んでいないため，他の画像との類似度が低いと判断されたためであると考えられる．

また，パターン3の画像群に対してはどちらのテストセットにおいてもクラスタリング精度は低くなった．この理由として，図5-8に見られるような顔領域の付近に効果線や枠線が描かれている画像や，髪などの輪郭が顔領域から離れて存在する画像に対して提案手法では正確な背景削除を行うことができなかったことが考えられる．

5.7　ファインチューニング済み CNN を用いた DBSCAN クラスタリング

データセット全体のクラスタリングを目的として，5.5節で提案したクラスタリング手法を改良し，キャラクタ顔画像でファインチューニングを行ったCNNモデルを特徴抽出器として使用する手法の有効性を評価する．

5.7.1　特徴抽出器の設定

特徴抽出器には，5.6.1項で述べたものと同様の学習済みCNNモデルを使用する．

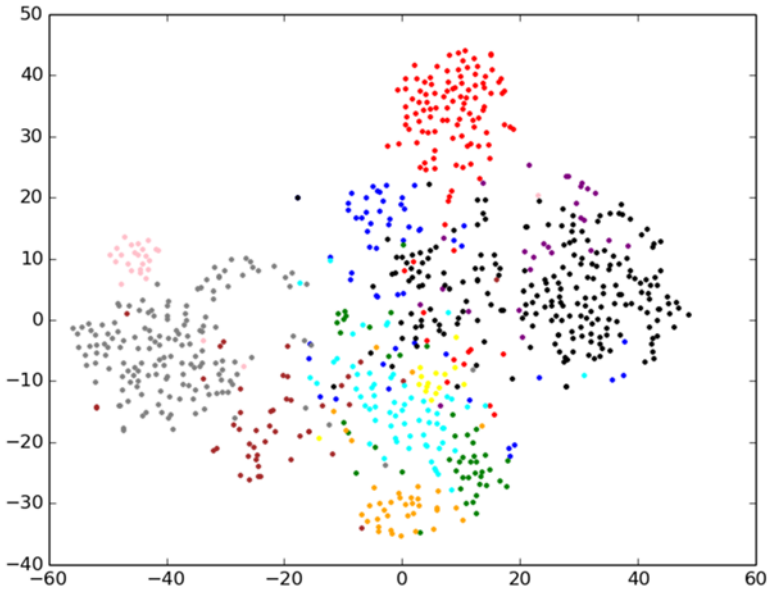


図 5-6: 「ベルモンド」に 10 回以上登場するキャラクタの顔画像についてパターン 1 の処理を行い，特徴量を可視化した画像

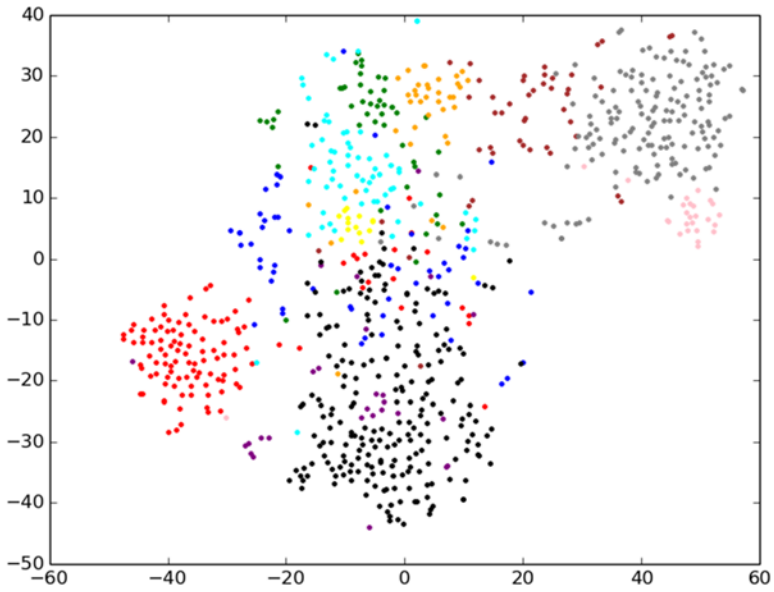


図 5-7: 「ベルモンド」に 10 回以上登場するキャラクタの顔画像についてパターン 2 の処理を行い，特徴量を可視化した画像

図 5-8: パターン 3 の処理を行った画像のうち，背景の除去に失敗した画像例

表 5-8: テストセットにおけるクラス数と画像数

タイトル	クラス数	画像数
ARMS	14	319 (33)
愛さずにはいられない	11	972 (87)
あっけら貫刃帖	12	704 (126)
あくはむ	9	1047 (48)
青すぎる春	12	552 (60)
天晴れ！カッポーレ	11	864 (38)
ありさ [2]	11	960 (104)
BEMADER・P	16	1226 (96)
爆裂！かんふー娘	15	1036 (121)
ベルモンド	13	847 (28)
ラブひな 14 巻	11	1168 (43)

5.7.2 データセット

評価に使用するデータセットの画像は5.5.2項で使用した画像と同様とする．ただし，詳細なキャラクタのクラスタリングを目的として，本実験では10回以上登場するキャラクタを主要キャラクタと設定してクラス分類を行う．これによって生成されたデータセットの内容を表5-8に示す．表においてクラス数はそれぞれ1個の「その他」のクラスを含み，括弧内は「その他」のクラスに属する画像数を示す．

5.7.3 評価基準

本実験では，データセットに含まれる顔画像全体の分類を目的として，5.6節と同様にF値とNMIから評価を行う．ただし，F値とNMIの算出において「その他」のクラスを正解クラスの一つとして扱い，DBSCANで外れ値と判断された画像群を一つのクラスタと

表 5-9: 次元削減の手法によるクラスタリング結果の変化（その 1）（自発表[63]より引用）

	Original	*PCA 128*	*PCA 64*	*PCA 32*	*PCA 2*
F-measure	0.442	0.449	0.453	0.466	0.489
NMI	0.122	0.140	0.152	0.168	0.240

表 5-10: 次元削減の手法によるクラスタリング結果の変化（その 2）（自発表[63]より引用）

	KPCA 128	*KPCA 64*	*KPCA 32*	*KPCA 2*
F-measure	0.514	0.525	0.522	0.514
NMI	0.293	0.311	0.327	0.300

表 5-11: 次元削減の手法によるクラスタリング結果の変化（その 3）（自発表[63]より引用）

	t-SNE 2	*UMAP 128*	*UMAP 64*	*UMAP 32*	*UMAP 2*
F-measure	0.579	0.595	0.595	0.593	**0.597**
NMI	0.419	0.452	0.452	0.451	**0.455**

見なす．したがって，式(5.27)，式(5.28)におけるNの値は入力画像の総数と等しいとする．

5.7.4　画像特徴量と次元削減の評価

CNN特徴量に対する次元削減によるクラスタリング結果の変化について評価を行う．次元削減の設定は，次元削減を行わないデータを「Original」，PCAによって128次元，64次元，32次元，2次元に圧縮したデータを「PCA 128, 64, 32, 2」，Kernel PCAによって128次元，64次元，32次元，2次元に圧縮したデータを「KPCA 128, 64, 32, 2」，t-SNEで

表 5-12: UMAP による次元削減の前に他の次元削減を行ったデータに対するクラスタリング結果の変化（自発表[63]より引用）

	Original	*PCA 128*	*PCA 64*	*PCA 32*	*KPCA 128*	*KPCA 64*	*KPCA 32*
F-measure	0.597	0.602	0.604	0.608	0.632	**0.634**	0.626
NMI	0.455	0.463	0.466	0.468	0.490	**0.495**	0.486

2次元に圧縮したデータを「t-SNE 2」，UMAPで128次元，64次元，32次元，2次元に圧縮したデータを「UMAP 128, 64, 32, 2」とした14種類について比較する．DBSCANのパラメータは，minPtsを10と設定し，εの値をk距離グラフの最小値から最大値に到達するまで1/100の間隔で変化させ，αの値が最大になるクラスタリング結果をそれぞれ求めて比較する．また，t-SNEとUMAPについては乱数を考慮して10回のクラスタリングを行った平均値を求める．

実験結果を表5-9，表5-10，表5-11に示す．実験結果より，UMAPによる次元削減がt-SNEを上回るクラスタリング精度を示すことが確認できた．また，5.5節の実験と同様に2次元への圧縮がDBSCANクラスタリングにおいて良好な精度を示した．

次に，他の手法で一度次元削減したデータに対して，UMAPの2次元への削減を適用した場合について検討する．通常のUMAPを「Original」，PCAによって128次元，64次元，32次元に圧縮したデータを「PCA 128, 64, 32」，Kernel PCAによって128次元，64次元，32次元に圧縮したデータを「KPCA 128, 64, 32」として比較した．実験結果を表5-12に示す．この結果より，UMAPの次元圧縮を行う前にKernel PCAによって64次元に圧縮する処理を行うことで，クラスタリング精度の向上に効果があることを確認した．

5.7.5　DBSCAN のパラメータ設定

5.5.5項と同様に，AGEDによって求められるεの候補から最適な値を決定する方法を検討する．5.7.4項の実験において最もクラスタリング精度が高くなったKernel PCAとUMAPによって次元削減した特徴量について検証を行う．AGEDの候補間の傾きの平均値を基準値として，AGEDの候補間で傾きが基準値よりも大きく変化する点をεに設定する手法を「proposed」とおく．ただし，AGEDの候補の最小値をεに設定した場合では，明らかにクラスタリング精度が低くなる傾向が見られたため，最初の区間に対しては判定を行わないものとする．AGED候補の平均値をεに設定する手法を「average」，中央値をεに設定する手法を「median」として比較する．また，AGEDの候補からεの値を決定する提案手法が，固定された基準に基づいてεを決定する手法よりも有効であることを

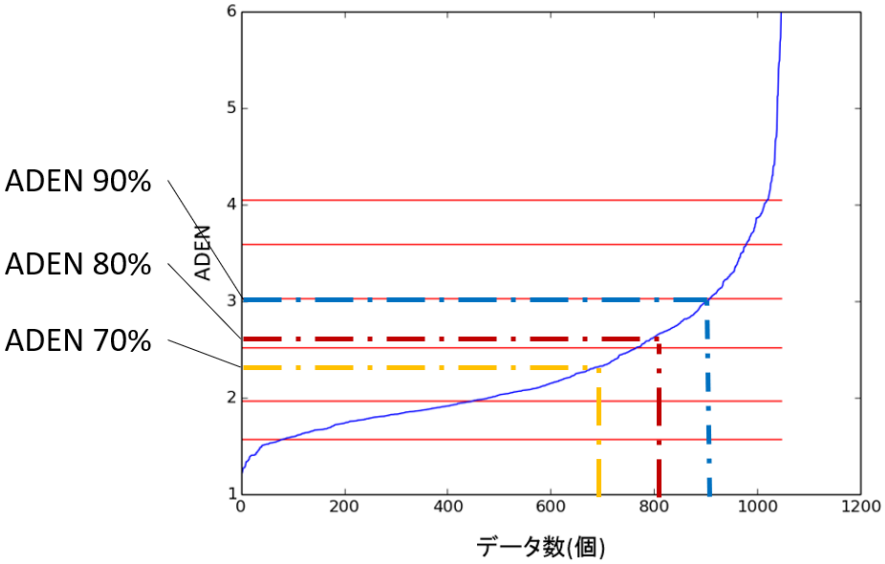


図 5-9: 固定された基準によるεの値の決定例

表 5-13: パラメータ ε の決定方法によるクラスタリング結果の変化（自発表[63]より引用）

	proposed	*median*	*average*	*k-dist 70%*	*k-dist 80%*	*k-dist 90%*
F-measure	**0.593**	0.503	0.510	0.509	0.491	0.574
NMI	**0.457**	0.402	0.408	0.376	0.415	0.452

確認するため，図5-9の例に示すようにデータ全体の70%，80%，90%に対応するADENの値をεに設定する手法と比較する．

それぞれの漫画作品に対して10回のクラスタリングを実行し，NMIとF値の平均値を求めた結果を表5-13に示す．実験結果より，提案手法は中央値や平均値をεに設定する方法よりも高い分類精度を示した．また，データ数に対して固定した基準からεを設定する方法と比較しても提案手法は優れた結果を示した．このことから，パラメータεの決定方法において提案手法が有効性を持つことを確認した．

5.8 まとめ

本章では，登場キャラクタリスト生成のためのキャラクタ顔画像の分類を目的として，第4章で提案したキャラクタ顔画像クラスタリングの改良について検討した．

はじめに，提案手法で使用する技術として，画像特徴量の記述，特徴量の次元削減，

DBSCANのアルゴリズムについて述べた.

次に, 一般物体データセットで学習を行ったCNNモデルを特徴抽出器とした手法について検討を行った. 実験結果より, DBSCANが主要キャラクタとサブキャラクタの分離に有効であることを確認し, またDBSCANの適用においてデータの局所的な関係に基づいた圧縮が有効であるという知見が得られた.

さらに, 顔画像からの背景除去によるクラスタリング精度の向上を考え, 異なる画像処理を行った顔画像に対するk-meansクラスタリングの精度変化を検証した. 実験結果では, 登場キャラクタ数が多い場合において, 背景除去によって新たに顔画像の誤分類が発生するケースが見られることが分かった.

最後に, これまでの実験から顔画像でファインチューニングを行ったCNNとDBSCANを用いたクラスタリング手法を提案した. 実験結果より, 提案手法は最大でF値 63.4%, NMI 49.5%という分類精度を示した. また, DBSCANのパラメータ自動推定について検討を行い, F値 59.3%, NMI 45.7%の精度での自動分類に成功した.

第 6 章

結論

本章では本論文の総括を述べ，今後の検討課題を明らかにする．

6.1 総括

本論文では，漫画のストーリー理解に必要となる漫画メタデータの自動抽出を目的として，漫画オブジェクトの検出モデルと，クラスタリングによるキャラクタ顔画像の分類手法を提案した．

第2章では，従来の局所特徴量を用いたキャラクタ顔検出の改良として，DPMの適用を検討した．まず，正面向きのキャラクタ顔画像に対する従来手法とDPMの検出率の比較を行い，キャラクタ顔検出におけるDPMの有効性を確認した．また，キャラクタのマルチビュー顔検出における最適なDPMのパラメータ設定について調査し，最大で77.6%の検出精度が示された．しかし実験結果の考察より，ハンドクラフトな特徴量であるHOGでは，顔画像の特徴変化の多様性を十分に吸収することができないという問題が明らかになった．このことから，DPMをキャラクタ顔検出に使用するにあたり，顔画像変化に対応した検出モデルの設計が必要であるという課題が示された．

第3章では，漫画オブジェクトの多様性に対応できる検出手法として，画像特徴を自動生成するCNNの適用を検討した．はじめに，キャラクタ顔検出を対象としてDPMと比較を行い，CNNが複雑な検出モデルの設定を必要とせずにDPMと同等以上の検出が可能であることを示した．次に，キャラクタに加えてコマやフキダシを対象とした検出について評価し，CNNが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も90%以上の精度での検出が可能であることを確認した．最後に，漫画画像に有効な候補領域の抽出方法を求めるため，CNNを用いた物体検出手法を比較した．実験結果より，Faster R-CNNのアルゴリズムが最も有効であることを確認し，4種類の漫画オブジェクトに対するmAPは91.0%となった．

第4章では，登場キャラクタの数といった事前知識に依存しないキャラクタ同定手法の構築を目的として，x-means法を用いたキャラクタ顔画像のクラスタリングを提案した．従来手法との比較より，x-meansのパラメータ設定が適切であるときに，提案手法が従来手法のクラスタリング精度を上回ることを確認した．このことから，未知の漫画キャラクタを対象としたクラスタリングにおいて，クラスタ数の自動決定が主要キャラ

クタ抽出精度の向上に有効であることを示した．ただし，主要キャラクタの分類において，提案手法には，パラメータ設定が複雑であることや，登場頻度の少ないキャラクタを分類することが困難であるといった課題が明らかとなった．

第5章では，主要キャラクタの分類を目的として，第4章で示された問題点を元にDBSCANを用いたキャラクタ顔画像のクラスタリングを提案した．はじめに，一般物体データセットを学習したCNNモデルを特徴抽出器したDBACANに適用について検討した．その結果，DBSCANが主要キャラクタとサブキャラクタの分離に有効であることを確認し，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，顔画像の分類精度を向上させるための試みとして，顔画像の背景領域の除去によるクラスタリング精度の変化について検討した．実験結果より，背景除去によって逆にクラスタリング精度が低下するケースが確認できた．このことから，正確な背景除去の処理に加え，キャラクタ分類により最適化した特徴量の設定が必要であるという課題が示された．最後に，キャラクタ顔画像でファインチューニングを行ったCNNモデルを特徴抽出器として使用したDBSCANクラスタリングを評価し，さらに，DBSCANのパラメータの自動決定方法を検討した．実験結果より，NMIにおいて最大で49.5%の分類精度が示され，45.7%の精度でキャラクタを自動分類することに成功した．

以上のように本論文では，ストーリー理解を目的とした漫画オブジェクト情報抽出のための技術として，漫画オブジェクトの検出とキャラクタ顔画像の分類について論じた．漫画オブジェクトの検出では，CNNを導入することによって，従来のハンドクラフト特徴では認識が困難なオブジェクトに対しても高精度な検出を行うことに成功した．本研究の成果は，キャラクタの表情やフキダシの種類といった，より詳細なオブジェクト情報の認識への応用が期待できる．キャラクタ顔画像の分類では，DBSCANクラスタリングの適用により，教師なしの状態で自動的に主要キャラクタを分類する手法を提案した．本研究の成果を基盤として，今後キャラクタ認識について更なる検討を行うことで，キャラクタ同定技術の構築が可能であると考えられる．

6.2　今後の課題

漫画のストーリー理解の実用化において残された課題として，以下が挙げられる．

漫画キャラクタの分類において，本研究ではキャラクタ顔画像についてファインチューニング行ったCNNモデルを特徴抽出器に用いたクラスタリング手法を提案した．しかし，実験結果ではテクスチャの類似した画像におけるキャラクタの誤分類が見られ，十分な認識精度を達成しているとはいえない．したがって，より高精度な特徴抽出器の設計が今後の課題となる．現在の手法が十分な精度が得られない理由として，本来一般物体認識を目的としたCNNモデルを転用していることから，漫画キャラクタの認識に最適な特徴抽出が行われていない可能性が考えられる．そこで今後の研究では，Auto Encoder

などCNN以外の特徴抽出についても検討することで, キャラクタ分類精度の向上を目指す.

また，本研究では漫画のシーン理解までの工程について検討を行ったが，ストーリー情報を取得するためには, コマの順序やシーンの移り変わりといった情報を認識する技術が必要となる．これは，抽出された各漫画オブジェクトの情報を利用して，オブジェクト間の関係を構造化することで達成される．従来研究では，漫画内容へのアクセスや漫画制作支援を目的としたメタデータのフレームワークや, 漫画の事前知識を利用して構造を理解する手法が提案されている[64, 65]．したがって，漫画のストーリー理解の自動化を実現するために, これらの従来研究との連携を考慮した漫画オブジェクト情報の抽出手法を構築する必要がある.

謝辞

本研究の実施にあたり素晴らしい実験環境を与えて下さり, 今日に至るまで終始変わらぬ懇切な御教示と御鞭撻を賜りました早稲田大学大学院 基幹理工学研究科 渡辺 裕 教授に深く感謝致します.

研究の方向性を初めとして, 数々の御指導ならびに御助言を賜りました亀山 渉 教授に心から感謝致します.

本論文の作成において, 貴重な御示唆を頂きました甲藤 二郎 教授に心より御礼申し上げます.

研究の方向性について御助言を下さり, また数々の貴重な御意見を頂いた電子通信大学大学院 情報理工学研究科 笠井 裕之 准教授に心から感謝致します.

本研究の機会を与えて下さり, また研究の進め方について丁寧な御指導を頂いた平成26年度 博士卒の石井 大祐 氏にはこの場を借りて深く感謝を申し上げます.

本研究を行うにあたり, 様々な御意見や御提案を頂いた渡辺研究室の皆様に御礼申し上げます. 特に, 実験用データセットの作成に御協力頂いた, 山下 拓朗 氏, 稲田 健太郎 氏に心より感謝致します.

本研究を行うにあたって, 漫画画像の提供及び論文への掲載を許可頂いた木野陽様 http://www.etheric-f.com/に心より御礼申し上げます.

最後に, 本研究と論文が完成するまでの長期にわたり, 暖かく見守って頂いた家族, 両親, 兄弟に心から感謝申し上げます.

2019年2月

参考文献

[1] 公益社団法人全国出版協会，“2017 年のコミック市場規模発表 紙＋電子で 2.8 % 減の 4,330 億円、紙は初の二桁減、電子は 17.2 % 増”, https://www.ajpea.or.jp/information/20180226/index.html，2018 年 2 月 26 日更新, (最終閲覧日 2018 年 11 月 29 日).

[2] 松下光範，“コミック工学のこれまでとこれから”，人工知能学会インテラクティブ情報アクセスと可視化マイニング研究会(第11回), SIG-AM-11-03, Nov. 2015.

[3] 木野陽，”ベリーベリークリームショコラ　ふたつのベリー”，2010.

[4] 新井俊宏，松井佑介，相澤清晴，“漫画画像からの顔検出”，電子情報通信学会総合大会論文集　2012年__情報・システム(2)，pp.161，Mar. 2012.

[5] 石井大祐, 渡辺祐, “マンガからの自動キャラクター位置検出に関する一検討”, 情報処理学会研究報告，Vol.2012-AVM-76, No.1, pp.1-5，Feb. 2012.

[6] 野中俊一郎，野沢拓也，羽場典久，“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 RESERCH & DEEVELOPMENT, No.57，pp.46-49，Mar. 2012.

[7] 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，Dec. 2010.

[8] Arai K, Tolle Herman, “Method for Real Time Text Extraction from Digital Manga Comic”, International Journal of Image Processing Vol 4, No. 6, pp. 669-676, Feb. 2011.

[9] P. Felzenszalb, R. Girshick, D. McAllester, D. Ramanan, “Object Detection with Discriminatively Trained Part Based Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, No.9, pp.1627-1645, Sept. 2010.

[10] P. Felzenszalb, D. McAllester, D. Ramanan, “A Discriminatively Trained, Multiscale, Deformable Part Model”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-8, June 2008.

[11] N. Dalal, B. Triggs, “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.886-893, June 2005.

[12] 柳澤秀彰，“マンガキャラクターを対象とした多視点顔検出の研究”，早稲田大学大学院基幹理工学研究科修士論文，Feb. 2016.

[13] P. Felzenszwalb, R. Girshick, D. McAllester, “Discriminatively Trained Deformable Part Models Version 5”, http://people.cs.uchicago, 2012, (最終閲覧日2018年11月29日).

[14] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, K. Aizawa, “Sketch-based Manga Retrieval using Manga109 Dataset”, Multimedia Tools and

Applications, Vol.76, Issue 20, pp.21811-21838, Nov. 2016.

[15] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, K. Aizawa, “Object Detection for Comics using Manga109 Annotations”, arXiv:1803.08670, Mar. 2018.

[16] H. Yanagisawa, D. Ishii, H. Watanabe, “Face Detection for Comic Images Using the Deformable Part Model”, IIEEJ (The Institute of Image Electronics Engineering of Japan) Transactions on Image Electronics and Visual Computing, Vol.4, No.2, pp.95-100, Oct. 2016.

[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, “The PASCAL Visual Object Classes (VOC) Challenge”, International Journal of Computer Vision, Vol.88, Issue 2, pp.303-338, June 2010.

[18] J. Orozco, B. Martinez, M. Pantic, "Empirical Analysis of Cascade Deformable Models for Multi-View Face Detection", Image and Vision Computing, Vol.42, pp.47-61, Oct. 2015.

[19] R. Girshick, J. Donahue, T. Darrell, J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, IEEE conference on Computer Vision and Pattern Recognition (CVPR), pp.580-587, Nov. 2013.

[20] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M. Smeulders, “Selective Search for Object Recognition”, International Journal of Computer Vision, vol.102, No.2, pp.154-171, Sept. 2013.

[21] P. Felzenszwalb, D. Huttenlocher, “Efficient Graph-Based Image Segmentation”, International Journal of Computer Vision, Vol.59, pp.167-181, Sept. 2004.

[22] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition", arXiv:1310.1531, Oct. 2013.

[23] R. Girshick, “Fast R-CNN”, International Conference on Computer Vision (ICCV), arXiv:1504.08083, Apr. 2015.

[24] H. Yanagisawa, 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.

[25] S. Ren, K. He, R. Girshick, J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks” in Advances in Neural Information Processing Systems, pp.91-99, June 2015.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, “SSD: Single shot multibox detector”, European Conference on Computer Vision (ECCV), pp.21-37, Sept. 2016.

[27] R. Girshick, “GitHub”, https://github.com/rbgirshick/fast-rcnn, 2015, (最終閲覧日

2018年11月29日).

[28] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, “Return of the Devil in the Details: Delving Deep into Convolutional Nets”, arXiv:1405.3531, May 2014.

[29] 藤子・F・不二雄，藤子プロ，“ドラえもん”，小学館，1974.

[30] 手塚治虫，“ブラック・ジャック”，秋田書店，1974.

[31] 青山剛昌，“名探偵コナン”，小学館，1994.

[32] 井上雄彦，“SLAM DUNK”，集英社，1990.

[33] R. Girshick, “GitHub”, https://github.com/rbgirshick/py-faster-rcnn, 2015, (最終閲覧日 2018 年 11 月 29 日).

[34] 小川徹，山崎俊彦，相澤清晴，“並列化された検出器による高精度漫画物体検出”，映像情報メディア学会技術報告，Vol.42，No.4，pp.293-298，Feb. 2018.

[35] H. Yanagisawa, 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.

[36] W. Liu, “GitHub”, https://github.com/weiliu89/caffe/tree/ssd, 2016, (最終閲覧日 2018年11月29日).

[37] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition”, arXiv:1409.1556, Sept. 2014.

[38] 長尾一輝, 渡辺裕, “コミックにおける主要キャラクター同定の検討“,電子情報通信学会総合大会, D-21-3, Mar. 2016.

[39] G. Schwarz, “Estimating the dimension of a model”, Ann. Statist, Vol.6, No.2, pp.461-464, Mar. 1978.

[40] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, Accepted for publication in the International Journal of Computer Vision, Vol.60, pp.91-110, Nov. 2004.

[41] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, “Speed-Up Robust Features (SURF)”, Computer Vision and Image Understanding 110, pp.346-359, June 2008.

[42] G. F. Pineda, H. Koga, T. Watanabe, “Scalable Object Discovery: A Hash-Based Approach to Clustering Co-occurring Visual Word”, IEICE Transactions, Vol.E94-D, Issue 10, pp.2024-2035, Oct. 2011.

[43] J. McQueen, "Some methods for classification and analysis of multivariate observations", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, pp.281-297, 1967.

[44] D. Pelleg, A. Moore, “X-means: Extracting K-means with Estimation of the Number of Clusters”, International Conference on Machine Learning (ICML)-2000, pp.727-734, Jan. 2000.

[45] 石岡恒憲，“クラスター数を自動決定するk-meansアルゴリズムの拡張について”，応用統計学，Vol.29, No.3, pp.141-149, Mar. 2001.

[46] 柳澤秀彰，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”，電子情報通信学会総合大会, D-12-40, Mar. 2017.

[47] 長尾一輝，“Bag-of-Visual Wordsを用いたコミックにおける主要キャラクター同定“，早稲田大学基幹理工学科卒業論文，Feb. 2016.

[48] 中山英樹, “深層畳み込みニューラルネットによる画像特徴抽出と転移学習”，電子情報通信学会技術研究報告，Vol.115，No.146，pp.55-59，July 2015.

[49] J. Giérin, O. Gibaru, E. Nyiri, S. Thiery, “CNN Features are also Great at Unsupervised Classification”, arXiv:1707.01700, July 2017.

[50] R. Narita, K. Tsubota, T. Yamasaki, K. Aizawa, “Sketch-based Manga Retrieval using Deep Features”, 2017 14th IAPR International Conference on Document Analysis and Recognition, pp.49-53, Nov. 2017.

[51] S. Wold, K. Esbensen, P. Geladi, “Principal Component Analysis”, Chemometrics and Intelligent Laboratory Systems, Vol.2, Issue1-3, pp.37-52, Aug. 1987.

[52] B. Schölkopf, A. Smola, K. R. Müller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”, Neural Computation, Vol.10, Issue 5, pp.1299-1319, July 1998.

[53] L. van der Maaten, G. Hinton, “Visualization Data using t-SNE”, Journal of Machine Learning 9, pp.2579-2605, Nov. 2008.

[54] L. Mclnnes, J. Healy, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”, arXiv:1802.03426, Feb. 2018.

[55] M. Ester, H. P. Kriegel, J. Sander, X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise” in Proceeding of 2nd International Conference on Knowledge Discovery and Data Mining, pp.226-231, Aug. 1996.

[56] N. Soni, N. Ganatra, “AGED (Automatic Generation of Eps for DBSCAN)”, International Journal of Computer Science and Information Security, Vol.14, No.5, pp536-559, May 2016.

[57] 柳澤秀彰，山下拓朗，渡辺裕，“主要キャラクタの抽出を目的とした漫画キャラクタ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, https://doi.org/10.3169/itej.73.199, Jan. 2019.

[58] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, “Rethinking the Inception Architecture for Computer Vision” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2818-2826, June 2016.

[59] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1800-1807, July 2017.

[60] H. Kaiming, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition" IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778. June 2016.

[61] E. Taskesen, M. J. T. Reinders, “2D Representation of Transcriptomes by t-SNE Exposes Relatedness between Human Tissues”, PLoS ONE 11(2): e0149853. https://doi.org/10.1371/journal.pone.0149853, Feb. 2016. (最終閲覧日 2018 年 11 月 29 日) .

[62] 柳澤秀彰，山下拓朗，渡辺裕，“CNNを用いた漫画キャラクタ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.

[63] H. Yanagisawa, 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.

[64] 三原鉄也，永森光晴，杉本重雄，“マンガメタデータフレームワークに基づくディジタルマンガのアクセスと制作の支援―ディジタル環境におけるマンガのメタデータの有効性の考察―”, 電子情報通信学会論文誌A, Vol.J98-A, No.1, pp.29-40，Jan. 2015.

[65] C. Rigaud, C. Guérin, D. Karatzas, J. C. Burie, J. M. Ogier: “Knowledge-driven understanding of images in comic books”, International Journal on Document Analysis and Recognition (IJDAR), Vol.18, Issue 3, pp.199-221, Apr. 2015.

図一覧

表一覧

研究業績

論文誌論文	○柳澤秀彰，山下拓朗，渡辺裕，“主要キャラクタの抽出を目的とした漫画キャラクタ画像のクラスタリング”映像情報メディア学会誌, vol.73, N0.1, pp.199-204, https://doi.org/10.3169/itej.73.199, Jan. 2019. ○H. Yanagisawa, 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.
査読付き国際会議	○ H. Yanagisawa, 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. ○H. Yanagisawa, 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. ○H. Yanagisawa, 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. ○H. Yanagisawa, 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.
国内研究会	柳澤秀彰，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. 石井大祐, 柳澤秀彰, 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.

シンポジウム	山下拓朗，柳澤秀彰，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. 柳澤秀彰，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. 柳澤秀彰，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. 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 (IMPS), I-4-12, pp.1-2, Nov. 2015.
国内大会	柳澤秀彰，山下拓朗，渡辺裕，“CNN を用いた漫画キャラクタ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. 山下拓朗，柳澤秀彰，渡辺裕，“深層学習を用いたマンガキャラクタ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. 柳澤秀彰，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. K. J. Ngeno, H. Yanagisawa, 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.

国内大会	柳澤秀彰，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. 柳澤秀彰, 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. 柳澤秀彰, 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. 柳澤秀彰, 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. 柳澤秀彰，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. 柳澤秀彰, 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. 陳明，柳澤秀彰，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. 柳澤秀彰, 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徴量の一検討”，映像情報メディア学会年次大会，17-9，Sept. 2014. M. Chen, H. Yanagisawa, 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.