

修士論文

深層学習を用いたマンガ画像の
セマンティックセグメンテーション

令和3年度修了

三重大学大学院 工学研究科 情報工学専攻
ヒューマンインターフェース研究室

照喜名 祐希

はじめに

現在、電子書籍として出版される書籍の市場規模は大きく、2018 年度では 2826 億円の規模であり、マンガはその中でも 2387 億円を占める重要なコンテンツである [1]。電子書籍の発展とともに、多くのマンガが電子書籍となり出版されている。その中でもデジタルデータとしての形を持たない古い作品は、紙面をスキャンしてデジタル画像とし、その画像に加筆や修正などを行ったうえで掲載する形式をとる。この時、スキャンで得たデジタル画像は最初からデジタルデータとして描かれるアニメーションのような構成要素ごとのレイヤーデータを持たないため、電子書籍化の際に書き足しや着色などの編集を行う際は、非常に専門性の高い作業が必要となる。

また、キャラクターやものがアニメーションで現れたり動いたりするモーショコミックや、VR 機器を用いてマンガを視聴する 3D コミックを制作するとき、キャラクターをはじめとする各要素を切り出す作業が必要となる。従って、キャラクターやフキダシなどの各要素を検出し、詳細な領域を自動抽出することで、従来の編集作業にかかる時間を削減することが可能となる。更に、自動で切り抜いた画像データをタグ付けなどに使用することで、副次的に電子書籍の利便性の向上に繋がることが考えられる。

これまでのマンガ画像の自動抽出の研究には、柳澤らのマンガキャラクターを対象とした多視点顔検出の研究 [2] などがある。文献 [2] は R-CNN を用いてキャラクターの顔を検出しているが、外接矩形を検出するに留まり、画素レベルでの範囲の指定には至っていない。

そこで本研究では、マンガ中の要素のセマンティックセグメンテーションおよびその精度向上を目的とし、画像から物体を検出するための深層学習の一種である YOLO v3 と、セマンティックセグメンテーションで画素ごとにラベル付けを行う DeepLab v3+を用いたマンガ画像中のキャラクター範囲抽出手法を提案する。

本手法は、マンガ画像を入力し、抽出した構成要素を赤画素、その他の要素を黒画素とする 2 値画像を出力する。抽出する構成要素はキャラクター、フキダシ、コマ割りであり、各要素の抽出は互いに独立に行う。この手法は YOLO v3 による外接枠の検出・前処理と DeepLab v3+ による画素単位での要素抽出からなる。

実験には、東京大学メディア情報処理研究室により提供されるマンガデータセット MANGA109[5] から作者の異なる 109 作品のマンガを選択して用いる。また、公開されているマンガ画像中の詳細なアノテーションデータが存在しないため、独自に 5 作品、278 枚の画像に付与した。

実験の結果、キャラクター領域では F 値 88 %，フキダシ領域では F 値 97 %，コマ割り領域では DeepLab v3+単独で F 値 92 %を得ることができ，各要素において十分な精度を得ることに成功した。

本論文では，1 章で研究背景と研究目的，2 章で提案手法の詳細，3 章で実験と考察，4 章でまとめと今後の課題について述べる。

目次

はじめに	i
第 1 章 序論	1
1.1 研究背景	1
1.2 先行研究	1
1.3 研究目的	2
第 2 章 提案手法	4
2.1 提案手法の流れ	4
第 3 章 評価実験	10
3.1 データセット	10
3.2 学習	16
3.3 評価方法	16
3.4 YOLO v3 による外接矩形検出	16
3.5 DeepLab v3+ による領域抽出	19
第 4 章 結言	29
4.1 まとめ	29
4.2 今後の課題	29
付録 A 付録	30
A.1 プログラムと実験データ	30
A.2 実行方法	32
謝辞	33

第 1 章

序論

本章では、1.1 節でキャラクター領域抽出の研究背景を、1.2 節で先行研究とその課題について述べる。

1.1 研究背景

現在、電子書籍として出版される書籍の市場規模は大きく、2018 年度では 2826 億円の規模であり、マンガはその中でも 2387 億円を占める重要なコンテンツである [1]。電子書籍の発展とともに、多くのマンガが電子書籍となり出版されている。その中でもデジタルデータとしての形を持たない古い作品は、紙面をスキャンしてデジタル画像とし、その画像を加筆や修正などを行ったうえで掲載する形式をとる。

この時、スキャンで得たデジタル画像は最初からデジタルデータとして描かれるアニメーションのような構成要素ごとのレイヤーデータを持たないため、電子書籍化のために書き足しや着色などの編集を行う際は、非常に専門性の高い作業が必要となる。また、キャラクターやものがアニメーションで現れたり動いたりするモーションコミックや、VR 機器を用いてマンガを視聴する 3D コミックを制作するとき、キャラクターをはじめとする各要素を切り出す作業が必要となる。従って、キャラクターやフキダシなどの各要素を検出し、詳細な領域を自動抽出することで、従来の編集作業にかかる時間を削減することが可能となる。更に、自動で切り抜いた画像データをタグ付けなどに使用することで、副次的に電子書籍の利便性の向上に繋がることが考えられる。

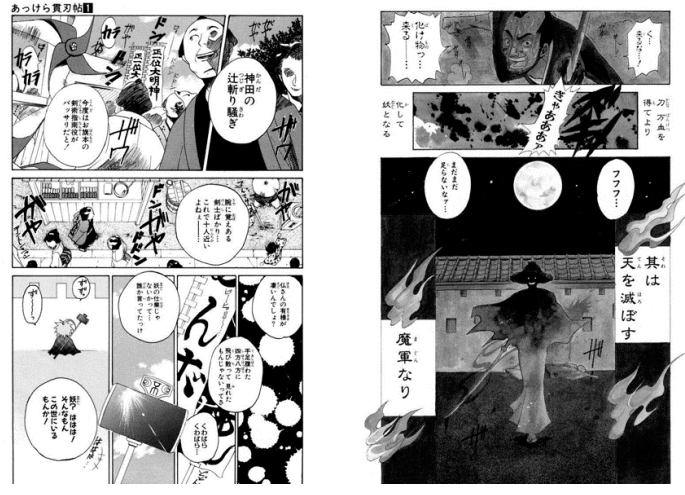
1.2 先行研究

先行研究としては、柳澤らのマンガキャラクターを対象とした多視点顔検出の研究 [2] などがある。文献 [2] では、Regions with Convolutional Neural Networks(R-CNN) を用いてキャラクターの顔を検出しているが、マンガ画像中のキャラクターの顔の外接矩形のみ検出するため、画素レベルでの範囲の指定には至っていない。

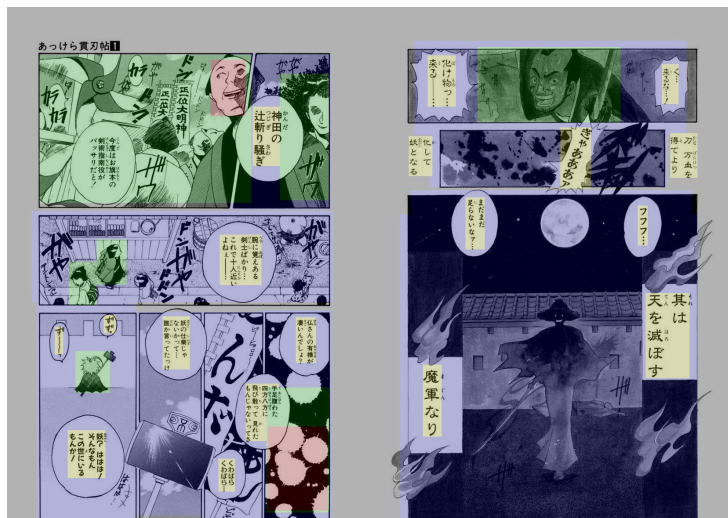
1.3 研究目的

本研究は、マンガ中の詳細なキャラクター、フキダシ、コマ割りの各領域抽出およびその精度向上を目的とし、画像から物体を検出する Object Detection を行う深層学習の一種である YOLO v3[3] と、画素ごとにラベル付けを行う Semantic Segmentation を行う DeepLab v3+[4] を用いたマンガ画像中の要素範囲抽出を提案する。

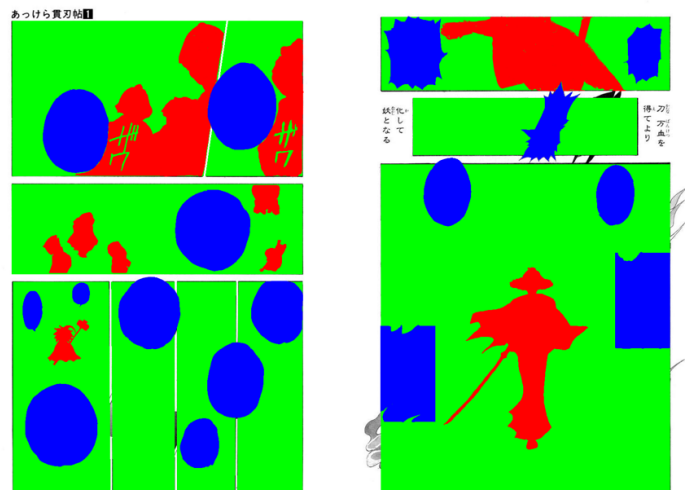
YOLO v3 では図 1.1(b) のように、マンガ画像からキャラクターの顔 (赤)、キャラクターの全身 (緑)、セリフなどの文章 (黄)、コマ割り (青) の外接矩形を検出する。各領域の画素単位での抽出には DeepLab v3+を用い、図 1.1(b) の情報を元にして作成した前処理情報を入力し、図 1.1(c) のようなマンガ画像中のキャラクター領域を赤、セリフを青、コマ割りを緑、それ以外を黒画素とする画像を出力する。



(a) マンガ画像



(b) 各領域の外接矩形の正解画像



(c) 各領域の正解画像

図 1.1: マンガ構成要素のセマンティックセグメンテーションの一例

第 2 章

提案手法

本章では、本研究の提案手法について説明する。この手法は YOLO v3 による前処理と、DeepLab v3+ による領域抽出からなる。2.1.1 節で YOLO v3 による前処理について、2.1.2 節で DeepLab v3+ について説明する。

2.1 提案手法の流れ

本手法では、マンガ画像を入力し、キャラクター領域の抽出では赤画素がキャラクターを、黒画素が背景を表す 2 値画像を出力する。フキダシ、コマ割りにについても同様の手法で 2 値画像を出力する。この手法は YOLO v3 による前処理と、DeepLab v3+ によるピクセル単位での領域抽出からなる。以下でそれぞれの説明を行う。

2.1.1 YOLO v3 による前処理

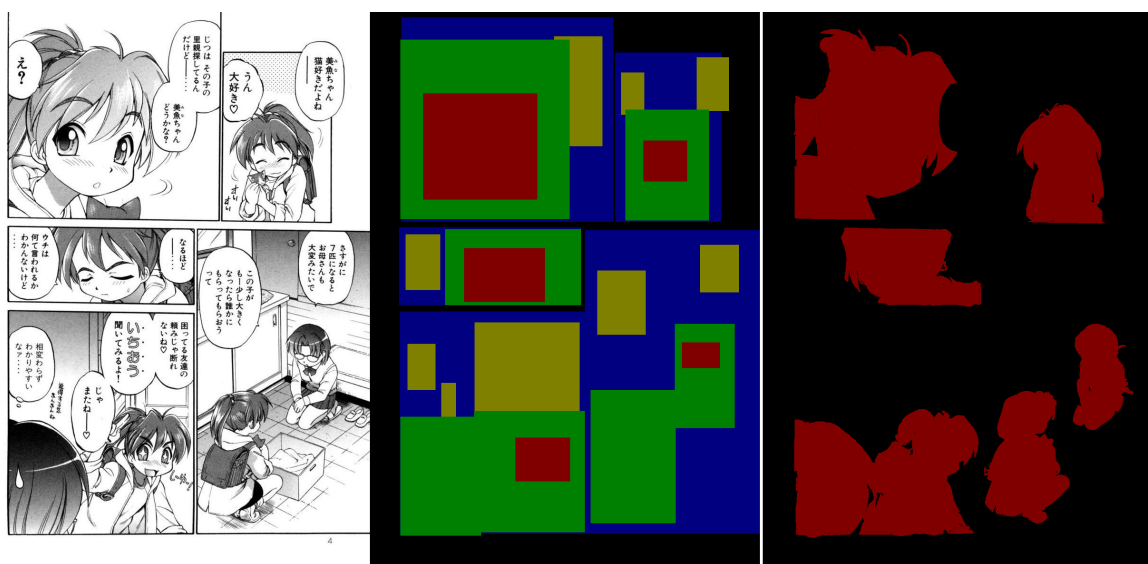
YOLO v3 は、すべて畳み込み (Convolution) 層で構成されている Convolutional Neural Network(CNN) の一種である。CNN は畳み込み層で画像の局所的な特徴を抽出し、得た特徴マップをプーリングを行うことで縮小する。CNN においては全結合層を用いるが、FCN では畳み込み層のみを用いる。そして逆畳み込みを行うことで入力画像と同じ大きさまで画像を拡大する。YOLO v3 のこのネットワークは 53 個の畳み込み層を持ち、Darknet-53 と呼ばれている。

本手法では、DeepLab v3+ による Semantic Segmentation の前処理として YOLO v3 による Bounding Box Detection を行う。その理由は、図 2.1(b) のような使用したマンガ画像のデータセットと対となる外接矩形のアノテーションが付属していたためである。マンガ内に存在する要素は多様であり、少ないサンプル数でのキャラクターなどの領域の抽出は困難であると考えられる。Bounding Box Detection に使用可能な正解データは 10465 枚存在することに対して、Semantic Segmentation に使用した正解データは 238 枚と少ない。

本手法では、予め YOLO v3 によって検出したキャラクターの顔、キャラクターの全身、セリフなどの文章、コマ割りの 4 要素を検出し、その外接矩形を DeepLab v3+ に入力するこ

とで、領域抽出の補助となると考えたためである。学習を行った YOLO v3 によりキャラクターの体、文章の位置、コマ割りそれぞれの位置に矩形を描画した画像を出力し、その画像を利用して DeepLab v3+ に入力する画像に前処理を行う。

前処理とは、図 2.2 のように、マンガ画像をグレースケール画像に変換して得た各ピクセルの値を半分にした上で RGB チャンネルのうち B にコピーし、さらに抽出する領域に対応するキャラクターやフキダシ、コマ割りそれぞれの矩形領域を RGB チャンネルのうち R にコピーすることを指す。図 2.3(a) は、キャラクター領域の抽出に inputs する画像であり、YOLO v3 がキャラクターの体の位置を予測した外接矩形領域を紫、その他を青で描画したものである。図 2.3(b) は、フキダシ領域の抽出に inputs する画像であり、YOLO v3 が文章の位置を予測した外接矩形領域を紫、その他を青で描画したものである。図 2.3(c) は、コマ割り領域の抽出に inputs する画像であり、YOLO v3 がコマ割りの体の位置を予測した外接矩形領域を紫、その他を青で描画したものである。



(a) マンガ画像

(b) 外接矩形の正解画像

(c) キャラクター領域の正解画像

図 2.1: 研究目的

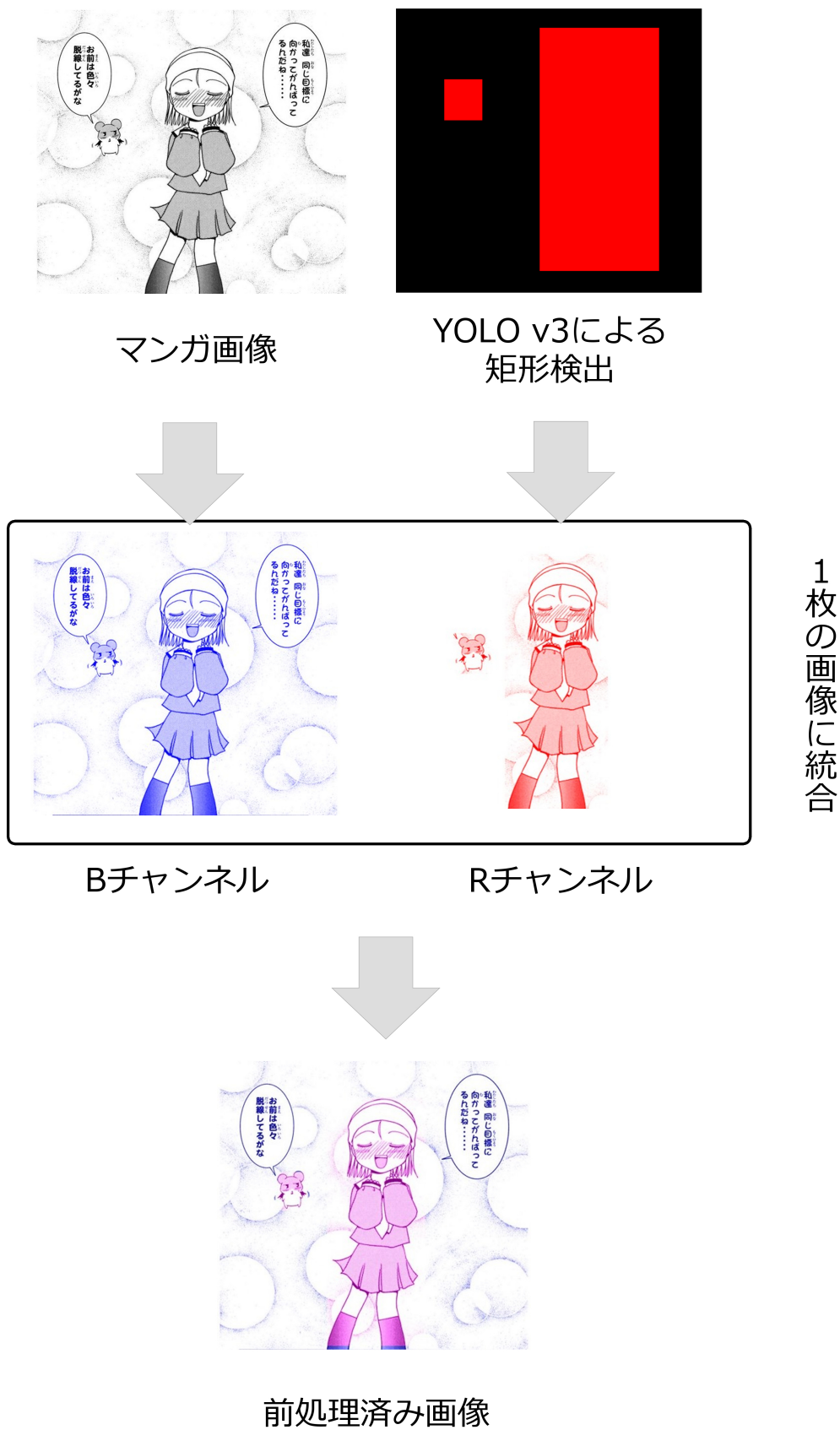


図 2.2: 前処理の工程



(a)



(b)



(c)

図 2.3: 前処理画像の例

2.1.2 DeepLab v3+による領域抽出

DeepLab v3+ は、すべて畳み込み（Convolution）層で構成されている Convolutional Neural Network(CNN) を基にした Deep Convolutional Neural Network（DCNN）と称されるネットワークの一種である。

このネットワークの特徴としては、空間ピラミッドプーリングモジュールを持ち、画像中の異なるスケールの特徴をプーリングすることで多くの情報を得られるという点が挙げられる。図 2.4 のように、予め YOLO v3 によって予測された外接矩形を含むデータを DeepLab v3+ に入力し、ピクセル単位での詳細な領域を抽出する。

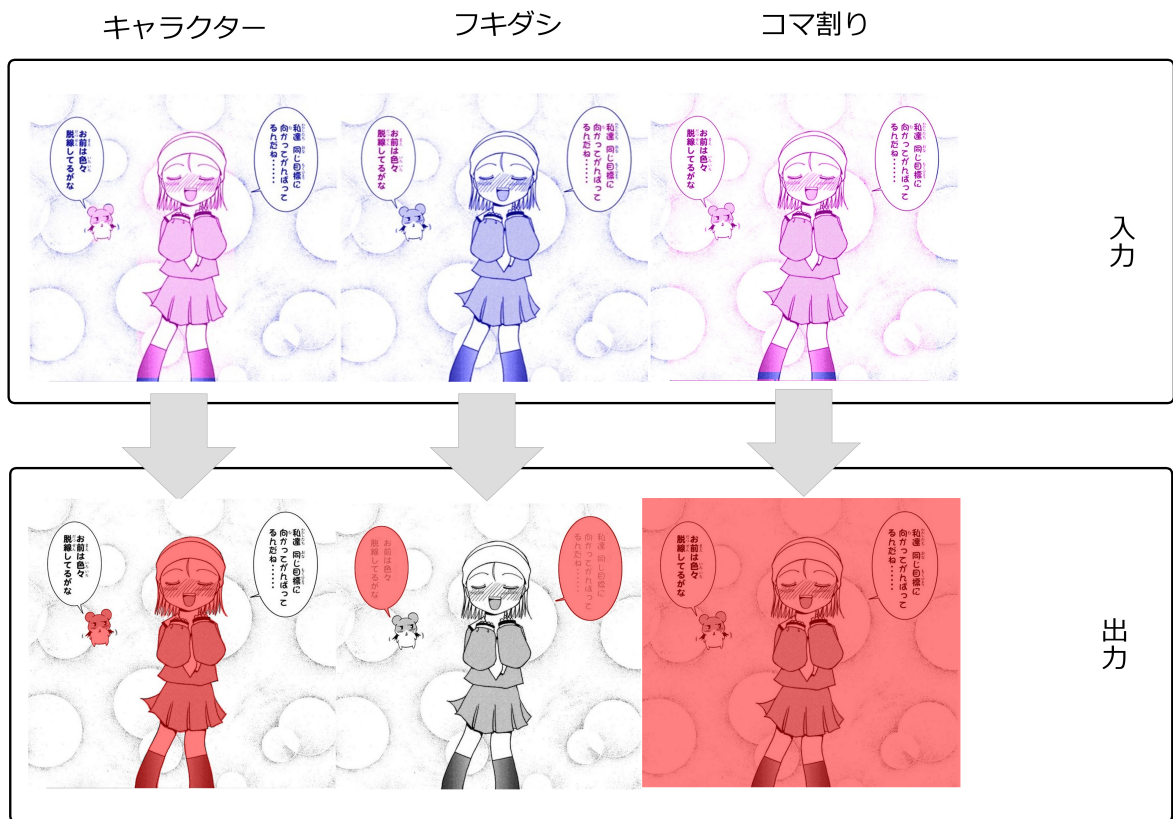


図 2.4: 領域抽出の例

第 3 章

評価実験

本章では，第 2 章で示した手法を用いた，実際のマンガ画像によるキャラクター領域の抽出について述べる．3.1 節では実験に用いたデータセットを，3.2 節で各ネットワークの学習パラメータについて，3.3 節で評価方法について，3.4 節，3.5 節で YOLO v3 による外接矩形検出の結果とその考察について，3.6 節，3.7 節で DeepLab v3+ によるキャラクター領域，フキダシ領域，コマ割り領域抽出の結果とその考察について述べる．

3.1 データセット

データセットは，東京大学メディア情報処理研究室より提供されている Manga109[5] を使用した．マンガの見開き 2 ページが 1 枚の画像になっており，サイズはすべて横 1654 × 縦 1170pixel である．YOLO v3 に対しては 105 冊分の画像 10227 枚を学習用に，4 冊分の画像 238 枚を予測用として用いた．DeepLab v3+ に対しては YOLO v3 で予測用に用いたマンガ画像 238 枚中，195 枚を学習用，23 枚を検証用，20 枚を評価用として用いた．マンガ画像の要素領域抽出時，DeepLab v3+ に入力したマンガのタイトルとその画像の枚数は表 3.1 のとおりである．

また，以上の実験では，DeepLab v3+ に使用するデータには，学習用データと検証用データ，評価用データ間にそれぞれ異なるデータではあるが同一の著者のマンガ画像データが存在する．よって，特定の著者の描画パターンに最適化されている可能性がある．提案手法の汎化性能を調べるため，YOLO v3 や DeepLab v3+ の学習に使用したマンガとは異なる著者のマンガ，すなわち Manga109 上のデータに存在しないマンガを使用した実験も行う．この実験には，自由な二次利用の可能な漫画「ブラックジャックによろしく」を使用する．

未知の作者の評価実験では，YOLO v3 のうち Manga109 より 105 冊分の画像 10227 枚を学習用に，4 冊分の画像 238 枚を予測用として用いた．DeepLab v3+ には YOLO v3 で予測用に用いたマンガ画像 238 枚中，218 枚を学習用，20 枚を検証用，「ブラックジャックによろしく」より 40 枚を評価用として用いた．DeepLab v3+ に入力するマンガのタイトルとその画像の枚数は表 3.2 のとおりである．

表 3.1: DeepLab v3+に入力する既知の作者のマンガ画像データ内訳

	学習用	検証用	評価用	合計
ねこだま [6]	92	11	10	113
あくはむ [7]	62	7	6	75
あっけら貫刃帖 [8]	30	4	3	37
学園ノイズ [9]	11	1	1	13
合計	195	23	20	238

表 3.2: DeepLab v3+に入力する未知の作者のマンガ画像データ内訳

	学習用	検証用	評価用	合計
ねこだま [6]	103	10		113
あくはむ [7]	69	6		75
あっけら貫刃帖 [8]	34	3		37
学園ノイズ [9]	12	1		13
ブラックジャックによろしく [10]			40	40
合計	218	20	40	278

3.1.1 キャラクターの定義

正解画像を作成する際に着色したキャラクター領域について言及する。

マンガにおけるキャラクターの定義は曖昧であり、動物やロボットなど人間の形でないキャラクター、ストーリー上重要でないキャラクターなどが同マンガ内に混在する。また、手や足などキャラクターの一部だけが描かれている場合もある。一貫性のない正解データは検出に影響を及ぼすため、基準を設定し、それに従って着色することが重要であると考えた。

本実験では、着色するキャラクター領域を以下のいずれかに当てはまるものとする。

- 名称がマンガで言及されているキャラクターである
- 図 3.1(a) のような、頭部または全身のどちらか半分以上が描かれたキャラクターである
- 図 3.1(b) のような、キャラクターの衣類、もしくは持っている道具である

ただし、図 3.1(c) のように、名称がマンガで言及されず、かつ描かれた全体の長辺が 100 ピクセル以下のキャラクターは背景であるとみなし、いずれも着色の対象外とする。



(a)



(b)



(c)

図 3.1: キャラクター領域の正解の例

3.1.2 フキダシの定義

正解画像を作成する際に着色したフキダシ領域について言及する。

フキダシには閉じていないもの、形状や内容が特殊なものがあり、定義によって領域を明確にすることが必要である。

- 図 3.2(a) のように、中心に活字・手書き文字・オノマトペのいずれかが入る
- 図 3.2(b) のように、閉じていないフキダシは、閉領域をつくる最短の直線によって囲まれた領域を正解領域として扱う
- 図 3.2(c) のように、絵が干渉しているフキダシはフキダシ部分のみを正解領域とする
- 図 3.2(d) のように、文字などがフキダシからはみ出した場合、はみ出た文字部分も着色する



(a)



(b)



(c)



(d)

図 3.2: フキダシ領域の正解の例

3.1.3 コマ割りの定義

コマ割りの定義については、マンガ画像の各シーンの最も外側に位置する、単純な図形で表された閉領域群をコマ割り領域とする。

3.2 学習

YOLO v3 の損失関数には二乗誤差損失を、最適化手法には Adaptive moment estimation(Adam) を用いた。また、学習はバッチサイズ 8, エポック数 184 で行った。DeepLab v3+ のバックボーンは ResNet を、損失関数は交差エントロピーを、最適化手法は SGD を用いた。DeepLab v3+ では、すべての実験において統一してバッチサイズ 4, エポック数 200 で学習した。

3.3 評価方法

提案手法の評価は、正解画像と予測画像をピクセル単位で比較した際に算出される適合率、再現率、F 値で行う。適合率 (Precision), 再現率 (Recall), F 値 (F-measure) は以下の式で定義される。

$$Precision = \frac{TP}{TP + FP} \quad (3.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.2)$$

$$F - measure = 2 \frac{Precision * Recall}{Precision + Recall} \quad (3.3)$$

キャラクター要素の抽出の場合、

- キャラクターである画素をキャラクターとして検出した画素数が TP
- 非キャラクターである画素をキャラクターとして検出した画素数が FP
- キャラクターである画素を非キャラクターとして検出した画素数が FN

となる。フキダシ領域、コマ割り領域に関しても、キャラクター領域と同様の区分で評価を行う。

3.4 YOLO v3 による外接矩形検出

3.4.1 キャラクター、フキダシ、コマ割りの外接矩形検出の結果と考察

表 3.3 と図 3.3 に実験結果を示す。出力画像の矩形領域が赤がキャラクターの顔、緑がキャラクターの全身、黄がセリフなどの文章、青がコマ割りを表している。キャラクターの顔は

62 %, キャラクターの全身は 79 %, 文章は 80 %, コマ割りは 90 % の F 値を得ることができた. 図 3.3(a)(b) に示すように, ほとんどのキャラクターなどの要素を検出することができた. 図 3.3(b) のように, 人間以外のキャラクターや, 小さく描かれたキャラクターも認識することができている. 一方, 図 3.3(c) のように, 読者から見て横や斜め向きに描かれたキャラクターの顔を検出できていないときがあるが, これは学習用画像が縦向きのものばかりであるためだと考えられる.



(a)



(b)



(c)

図 3.3: 外接矩形検出の一例

表 3.3: 外接矩形検出の結果

	キャラクターの顔	キャラクターの全身	セリフなど文章	コマ割り
適合率	0.69	0.75	0.82	0.86
再現率	0.56	0.83	0.78	0.94
F 値	0.62	0.79	0.80	0.90

3.5 DeepLab v3+ による領域抽出

3.5.1 同一著者マンガ間の要素領域抽出

学習用、検証用データと評価用データの間には同一著者のデータが存在する。

(1) キャラクター領域抽出の結果と考察

表 3.4 と図 3.4, 3.5 に実験結果を示す。出力画像の赤画素がキャラクターとして抽出された領域を表している。表 3.4 より、本手法において、ほとんどのキャラクターを詳細に抽出することができ、YOLO v3 の前処理を行うことで 88 % の F 値を、前処理なしで 86 % の F 値を得ることができた。図 3.4 に示すように、図に含まれているキャラクターの大部分を精密に抽出することができている。図 3.5(a) より、人間以外のキャラクターも抽出することに成功し、図 3.5(b) より、フキダシなど非キャラクター領域を避けて抽出を行っていることがわかる。反面、曖昧に認識されたキャラクターも複数存在した。学習データが少ないため、曖昧に抽出された領域が存在すると考えられる。

(2) フキダシ領域抽出の結果と考察

表 3.5 と図 3.6, 3.7 に実験結果を示す。出力画像の赤画素領域がフキダシとして抽出された画素を表している。表 3.5 より、本手法において、ほぼ全てのフキダシを明確に抽出することに成功し、YOLO v3 の前処理を行うことで 97 % の F 値を、前処理なしで 96 % の F 値を得ることができた。図 3.6, 図 3.7 より、様々な形のフキダシを漏れなく正確に抽出できていることがわかる。

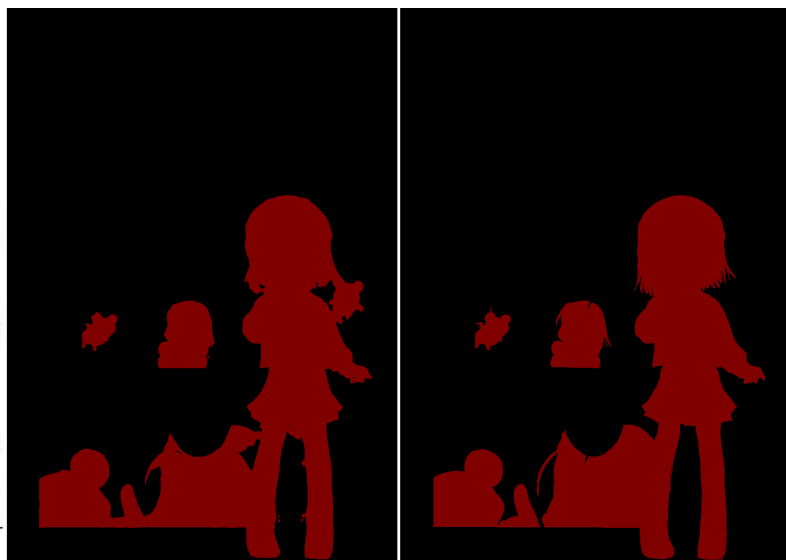
(3) コマ割り領域抽出の結果と考察

表 3.6 と図 3.8, 3.9 に実験結果を示す。出力画像の赤画素領域がコマ割りとして抽出された画素を表している。表 3.6 より、本手法において、YOLO v3 の前処理を行うことで 82 % の F 値を、前処理なしで 92 % の F 値を得ることができた。図 3.6 では、単純な形のコマ割りでもわずかに枠を外れて抽出を行ってしまっており、精度が高いとは言えない。また、図 3.7 では、コマ内にて少量の抽出漏れが存在している。

本抽出においては、YOLO v3 による前処理を行わない手法の方が前処理を行う手法よりも精度が高い結果となった。考えられる理由として、YOLO v3 は境界線が複雑な要素の提出に対して有効であるが、コマ割りの構造は単純かつ画像の最外部に位置し他要素に内包されていないことなどが挙げられる。

表 3.4: キャラクター領域抽出の結果

	DeepLab v3+ & YOLO v3	DeepLab v3+
適合率	0.88	0.83
再現率	0.87	0.87
F 値	0.88	0.86



(a) マンガ画像

(b) キャラクター領域の予測画像 (c) キャラクター領域の正解画像

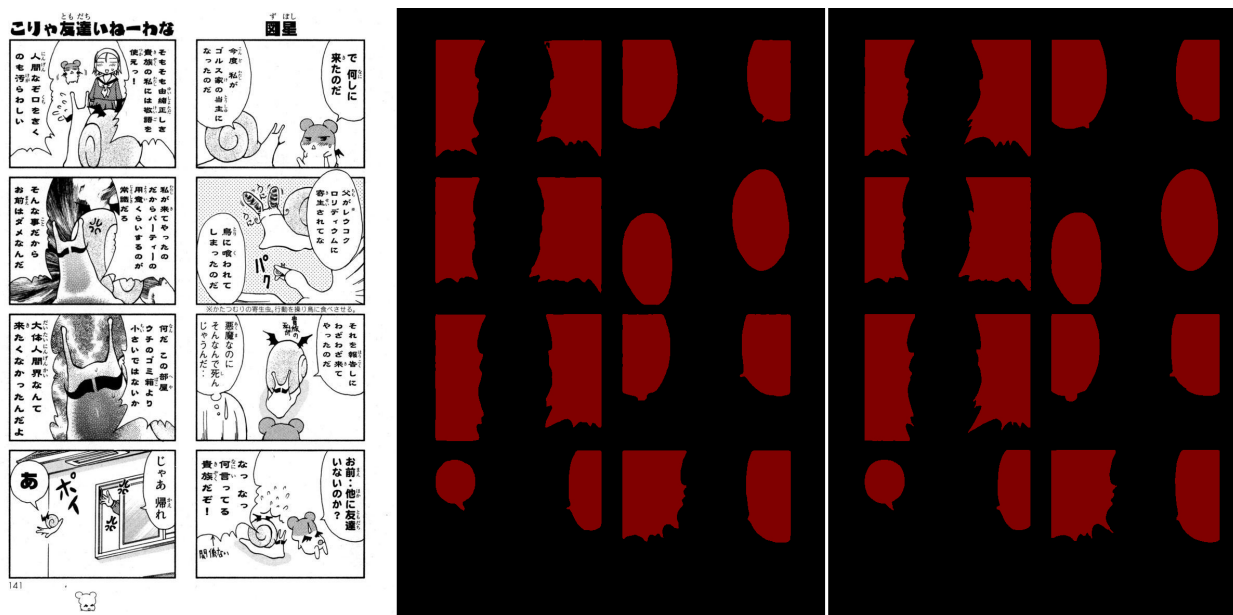
図 3.4: キャラクター領域抽出画像の比較



図 3.5: キャラクター領域抽出画像の一例

表 3.5: フキダシ領域抽出の結果

	DeepLab v3+ & YOLO v3	DeepLab v3+
適合率	0.97	0.96
再現率	0.97	0.96
F 値	0.97	0.96



(a) マンガ画像

(b) フキダシ領域の予測画像

(c) フキダシ領域の正解画像

図 3.6: フキダシ領域抽出画像の比較



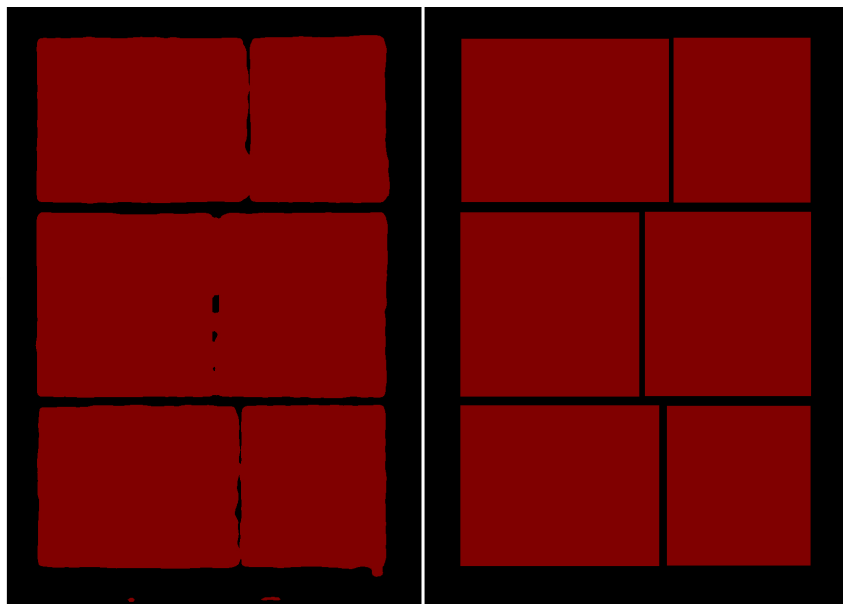
図 3.7: フキダシ領域抽出画像の一例

表 3.6: コマ割り領域抽出の結果

	DeepLab v3+ & YOLO v3	DeepLab v3+
適合率	0.98	0.98
再現率	0.71	0.87
F 値	0.82	0.92



(a) マンガ画像



(b) コマ割り領域の予測画像 (c) コマ割り領域の正解画像

図 3.8: コマ割り領域抽出画像の比較

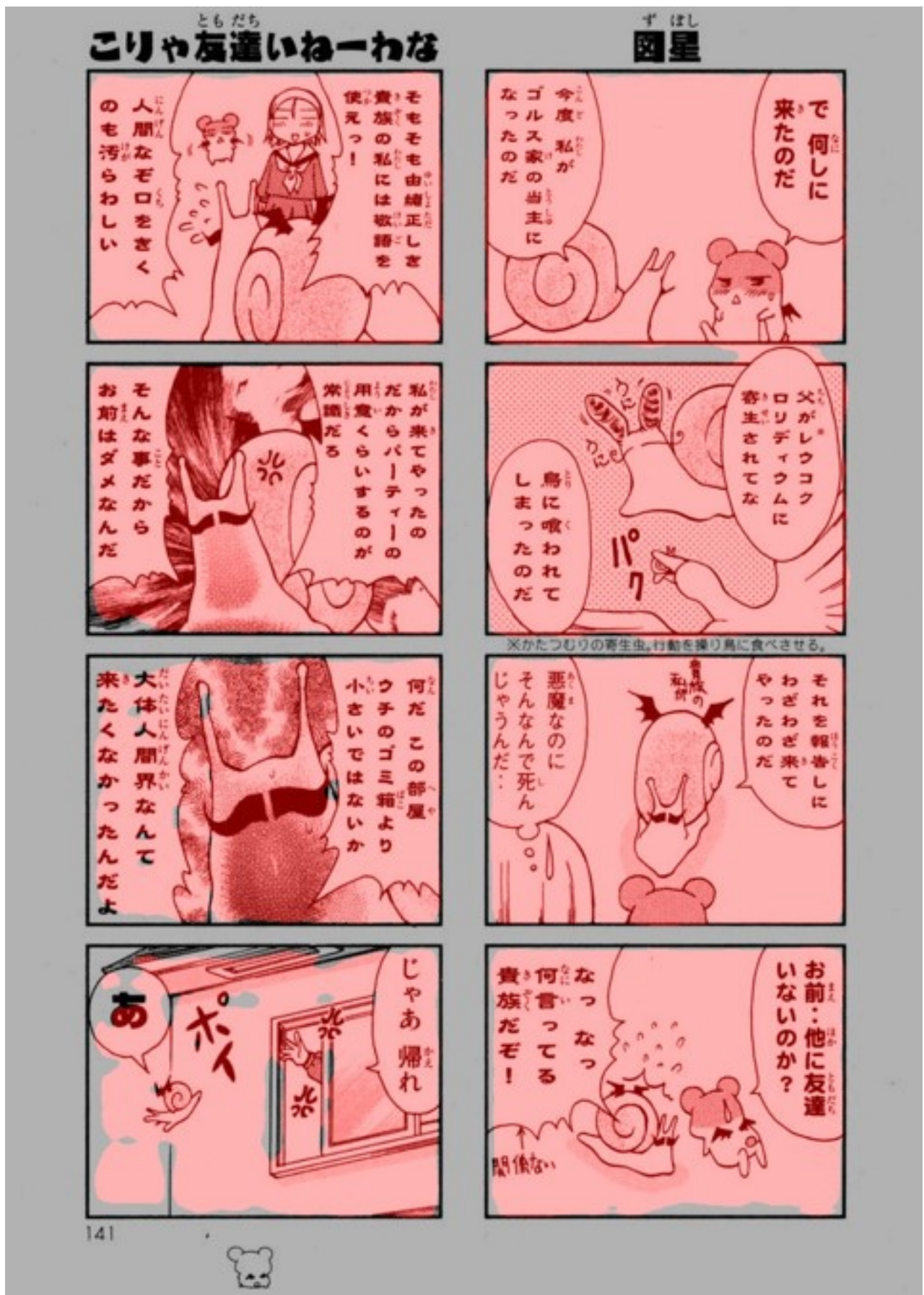


図 3.9: コマ割り領域抽出画像の一例

3.5.2 異なる著者間の要素領域抽出

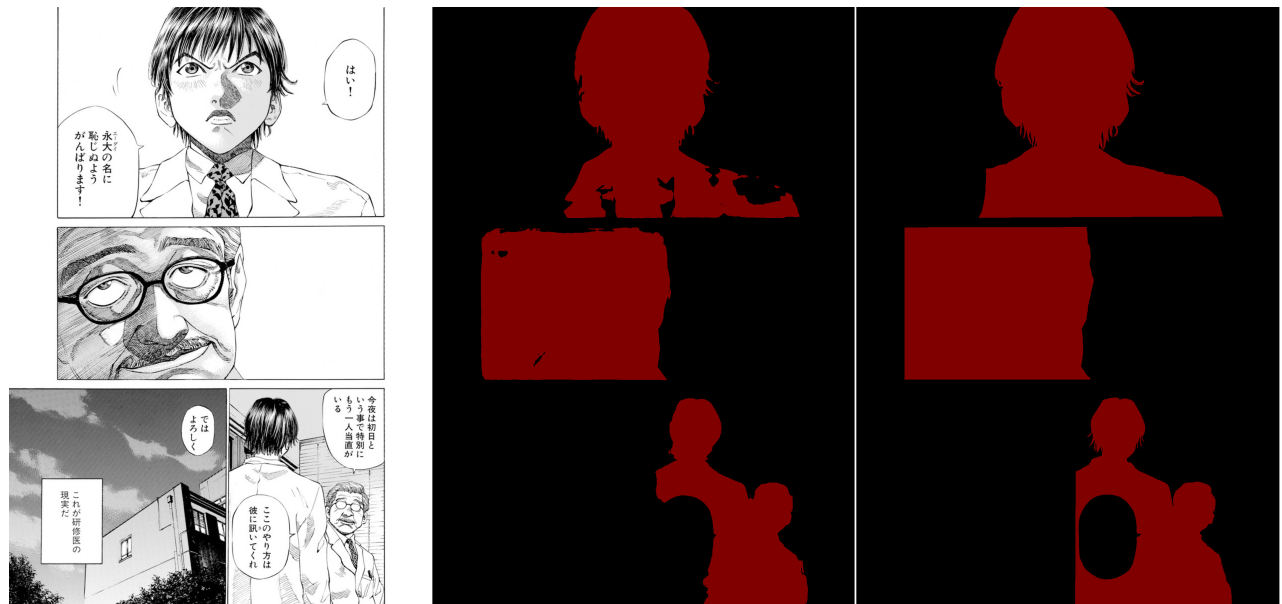
(1) キャラクター領域抽出の結果と考察

表 3.7 と図 3.10, 3.11 に実験結果を示す。表 3.7 に示すように、本手法において、YOLO v3 の前処理を行うことで 81 % の F 値を、前処理なしでも同様に 81 % の F 値を得ることができた。出力画像の赤画素領域がキャラクターとして抽出された画素を表している。図 3.10 を見ると、大まかなキャラクター要素は抽出できているが、細部は曖昧な箇所が多く、輪郭を捉えられていない点が目立つ。

図 3.11 では、顔の大部分が抽出できずに欠けてしまっている。抽出に使用したマンガは学習に使用したマンガと画風が大きく異なり、かつ横顔の学習のサンプル数が少ないことが原因であると考えられる。

表 3.7: キャラクター領域抽出の結果

	DeepLab v3+ & YOLO v3	DeepLab v3+
適合率	0.92	0.91
再現率	0.71	0.73
F 値	0.81	0.81



(a) マンガ画像

(b) キャラクター領域の予測画像 (c) キャラクター領域の正解画像

図 3.10: キャラクター領域抽出画像の比較



図 3.11: キャラクター領域抽出画像の一例

第 4 章

結言

4.1 まとめ

本研究では、YOLO v3 と DeepLab v3+ を併用する二段階のマンガ画像中の構成要素自動抽出手法を提案した。キャラクター要素、フキダシ要素に対し、二段階の構成要素抽出手法を用いて抽出精度を改善することができた。

既知の著者での抽出実験では YOLO v3 の学習に 105 冊、DeepLab v3+ の学習に 4 冊のマンガを使用し、実験結果としてキャラクター領域は F 値 88 %、フキダシ領域は F 値 97 %、コマ割り領域は YOLO v3 を併用せずに F 値 92 % を得た。また、未知の著者の抽出実験においては、F 値 81 % を得たが、DeepLab v3+ のみでの抽出と YOLO v3 と DeepLab v3+ 双方を利用した抽出で抽出精度は変わらなかった。

4.2 今後の課題

本実験では多くのキャラクター領域、フキダシ領域、コマ割り領域を高い精度を保ちながら抽出することに成功した反面、特異な構図のマンガ画像に対して精度を発揮できない側面が見受けられた。したがって、正解画像のバリエーションを増やし、様々な構図に対応できるように学習を行いたい。

また、オノマトペの抽出も行い、更に多くの種類のマンガ画像の構成要素を判別可能な抽出器を作成したい。

付録 A

付録

A.1 プログラムと実験データ

本研究に関するプログラム，データをすべて以下のディレクトリに置く．

- /net/nfs2/export/home/teruki/Research

A.1.1 主要ディレクトリ

- chara_yolo_deeplab : YOLO v3 により前処理されたマンガ画像を入力してキャラクター領域の学習に使用する DeepLab v3+
 - VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
 - run : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ
- chara_vanila_deeplab : 前処理されていないマンガ画像を入力してキャラクター領域の学習に使用する DeepLab v3+
 - VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
 - run : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ
- self_yolo_deeplab : YOLO v3 により前処理されたマンガ画像を入力してフキダシ領域の学習に使用する DeepLab v3+
 - VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
 - run : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ
- self_vanila_deeplab : 前処理されていないマンガ画像を入力してフキダシ領域の学習に

使用する DeepLab v3+

- VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
- run : 作成した重みパラメータとその学習時のステータスのデータ
- images : 評価用画像データとその正解領域画像データ
- frame_yolo_deeplab : YOLO v3 により前処理されたマンガ画像を入力してコマ割り領域の学習に使用する DeepLab v3+
 - VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
 - run : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ
- frame_vanila_deeplab : 前処理されていないマンガ画像を入力してコマ割り領域の学習に使用する DeepLab v3+
 - VOCdevkit : DeepLab v3+ の学習に使用するマンガ画像データとその正解画像データ
 - run : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ
- keras_yolo3_master : YOLO v3 と矩形アノテーションデータ
 - VOCdevkit : YOLO v3 の学習に使用するマンガ画像とその正解画像データ
 - model_data : 作成した重みパラメータとその学習時のステータスのデータ
 - images : 評価用画像データとその正解領域画像データ

A.1.2 ソースプログラムの一覧

- yolo_deeplab
 - train.py : 学習を行うプログラム
 - demo.py : 作成した重みパラメータで出力を行うプログラム
- vanila_deeplab
 - train.py : 学習を行うプログラム
 - demo.py : 作成した重みパラメータで出力を行うプログラム
- keras_yolo3_master
 - train.py : 学習を行うプログラム
 - yolo.py : 作成した重みパラメータで出力を行うプログラム

A.2 実行方法

本研究では実験に GPU を用いた。また, Python はバージョン 3.7.3 を使用した。オプションについてはソースプログラムを参照のこと。

- DeepLab v3+
 - train.py : 学習
 - demo.py : 抽出
 - calc.py : 評価
- YOLO v3
 - train.py : 学習
 - yolo.py : 検出
 - calc_face.py : キャラクターの顔の検出の評価
 - calc_body.py : キャラクターの全身の検出の評価
 - calc_text.py : セリフなど文章の検出の評価
 - calc_frame.py : コマ割りの検出の評価

謝辞

研究を進めるにあたり，多くの助言をいただきました若林哲史教授，盛田健人助教，白井伸宙助教，副査をしていただきました松岡真如准教授に深く感謝いたします。また，多くの助言や添削，指導など，お世話になりました研究室の先輩方，ともに切磋琢磨した同期の皆様に感謝いたします。最後に，大学生活を支えてくれた両親，家族に今一度深く感謝の意を表明し，本論の締めといたします。

参考文献

- [1] インプレス総合研究所, ”電子書籍ビジネス調査報告書 2019”
- [2] 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”, 2016.
- [3] Joseph Redmon, Ali Farhadi, ”YOLOv3: An Incremental Improvement” 2018.
- [4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam, ”Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, 2018.
- [5] Toru Ogawa, Atsushi Otsubo, Rei Narita, Yusuke Matsui, Toshihiko Yamasaki, Kiyoharu Aizawa, Object Detection for Comics using Manga109 Annotations, arXiv, 2018. 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, Springer, 2016.
- [6] えびふらい, ねこだま, Jコミックテラス, 2005.
- [7] 新居さとし, あくはむ, 第1巻, 講談社, 2005.
- [8] 小林ゆき, あっけら貫刃帖, 第1巻, 集英社, 2002.
- [9] 猪原大介, 学園ノイズ, 第1巻, 東京三世社, 2003.
- [10] 佐藤秀峰, ブラックジャックによろしく, 第1巻, 2002.