

畳み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

柳澤 秀彰 山下 拓朗 渡辺 裕

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

〒169-0072 東京都新宿区大久保 3-14-9 シルマンホール 401 号室

E-mail: bule-cosmo@ruri.waseda.jp

あらまし マンガ画像を web 上で活用するために、フキダシ・キャラクター等のオブジェクトを検出しメタデータを作成する手法が研究されている。このときマンガオブジェクトに対して一括検出を行う方法が重要となる。従来研究において、我々は畳み込みニューラルネットワークを用いた物体検出手法がマンガオブジェクト検出に対して有効性があることを確認した。本稿では、様々な形態のマンガオブジェクトに対する検出結果について検証を行い、CNN 検出器が多様な形状を持つマンガオブジェクトに対して認識を行うメカニズムについて検討する。

キーワード CNN, Faster R-CNN, マンガ, 物体検出

1. はじめに

マンガ画像は線画とテキストから構成される比較的シンプルなモデルである。しかし、キャラクターやフキダシといったマンガを構成するオブジェクトは作品ごとに様々な形状で表現されており、画像認識の上でマンガは独特な特徴を持つ対象である。我々はマンガオブジェクトの検出において畳み込みニューラルネットワーク (CNN) を用いた物体検出モデルである Faster R-CNN の適用について実験を行った [1]。その結果、複数のマンガ作品を対象としてフキダシ 95.9%、キャラクター 86.8% という精度での検出が可能であることを確認した。本稿では、様々な形状のマンガオブジェクトに対する検出結果を比較することで、CNN がマンガオブジェクトを認識する仕組みについて検討する。

2. Faster R-CNN

本稿におけるマンガオブジェクト検出には Faster R-CNN [2] を使用する。本手法は Girshick らの提案した R-CNN [3] を改良した手法であり、物体の候補領域の検出を CNN 内の Region Proposal Network (RPN) で実行することによって検出速度と学習の効率を向上させている。本研究に用いる CNN モデルは、VGG16 [4] をマンガ画像についてファインチューンしたものを使用する。ファインチューンには Manga109 [5] の画像のうち、作者の異なる作品 19 作から 100 枚ずつを抜き出した合計 1900 枚の画像を使用する。全ての学習画像に対してフキダシおよびキャラクターの顔領域を記述したアノテーションを作成し、各マンガオブジェクトの検出器を個別に生成した。このとき、学習の反復回数は 80000 回に設定した。第 3 章と第 4 章の検出実験では、候補領域の閾値が 0.5 以上の領域を物体と見なして検出を行う。

3. フキダシの認識

フキダシはマンガにおけるセリフの表現に用いら

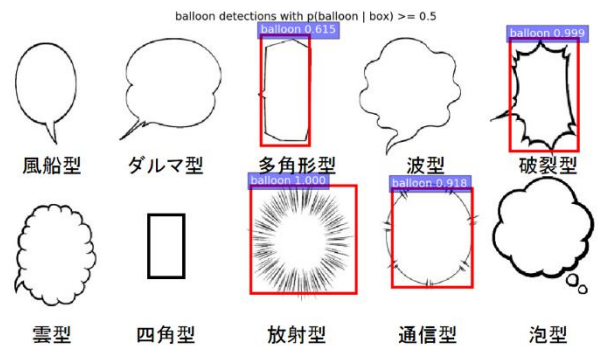


図 1: 文字を含まないフキダシの検出

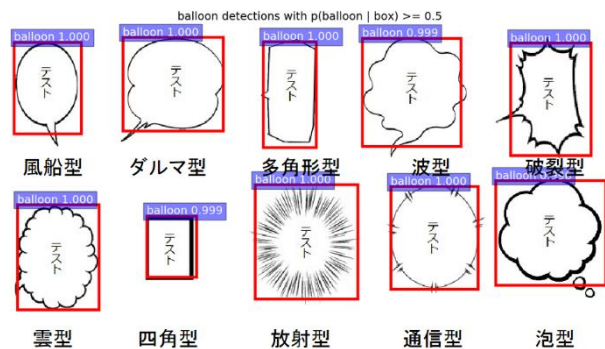


図 2: 文字を含むフキダシの検出

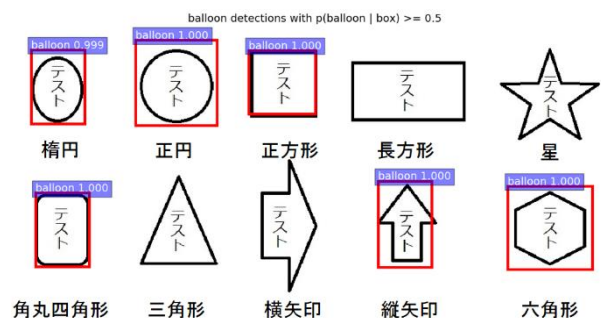


図 3: 文字を含む図形の検出

れるオブジェクトであり、セリフの内容や状況に応じて形状が使い分けされる。IDIA.JP [6]では、フキダシの形状を「風船型」「ダルマ型」「多角形型」「波型」「破裂型」「雲型」「四角型」「放射型」「通信型」「泡型」「象徴型」「無形型」の12種類に分類している。初めにこの中の10種類のフキダシについて、内部に文字がない画像と文字が存在する画像に対して検出を行った。実験に用いるフキダシは、IDIA.JP および吹き出し素材専門サイト「フキダシデザイン」[7]から引用している。実験結果を図1, 2に示す。内部に文字が存在する場合は全てのフキダシが検出されたのに対して、文字が存在しない場合には「多角形型」「破裂型」「放射型」「通信型」の4種類のみが検出された。

次に、より単純な形状の10種類の図形の内部に文字が存在する画像について検出を行った。実験結果を図3に示す。実験結果では「楕円」「正円」「正方形」「角丸四角形」「縦矢印」「六角形」の6種類の図形がフキダシとして検出された。

以上の結果から考察すると、Faster R-CNNのフキダシ検出においてはまず「放射型」「通信型」といった独特な形状を持つフキダシに対して文字の存在に関わらずに検出することが分かった。次に文字列とその周囲に図形が存在することを認識することで、フキダシが検出される。しかしこの時に「星形」や「三角形」など文字に対して十分な余白が存在しない図形や、「長方形」のような余白が過大な図形に対してはフキダシとして検出されないことが確認できた。

4. キャラクターの認識

マンガのキャラクターは作者によって外見の変化が特に大きなオブジェクトであり、デフォルトめされた人間のほか、動物や無機物など作品によって様々なキャラクターが定義される。本研究では、検出器がどのような基準でキャラクター顔画像を認識しているか検討を行った。

Manga109より検出器の学習に使用していない6作品についてキャラクターの正面顔画像、横向き顔画像をそれぞれ12枚ずつ切り出した画像を元画像として実験に使用する。元画像から口、目、顔全体を消去した画像をそれぞれ作成し、顔検出精度を比較した。実験画像の例を図4に示す。顔検出の結果を表1に示す。口を消去した画像は元画像から検出率が変わらない一方、目を消去した画像と顔全体を消去した画像はそれぞれ検出率の低下を確認した。また、検出率の低下は正面より横向き画像の方が大きくなった。この結果から、顔検出に目の存在が影響しており、横顔は正面顔と比較して検出に利用できる特徴が少ないことが分かった。しかし、顔パーツ全てを消去した場合においても60%以上の検出が行われることから、顔検出には顔

© 佐佐木 あつし

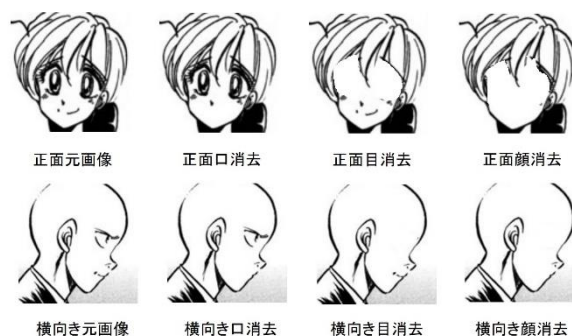


図4: 顔検出実験に使用する画像例

表1: キャラクター顔画像の検出結果(%)

	元画像	口消去	目消去	顔消去
正面	98.6	98.6	91.7	79.6
横向き	93.1	93.1	77.8	66.7

パーツのみならず顔の周辺領域の特徴も利用されていることが考えられる。

5. まとめ

本稿では、マンガオブジェクトを対象としたFaster R-CNNの物体認識メカニズムの理解を目的として、フキダシおよびキャラクター顔画像の検出実験を行った。実験結果からFaster R-CNNはオブジェクトの内部だけでなく周辺領域の特徴も利用して検出を行っていることが示された。

謝辞

本研究はJSPS科研費17K00511の助成を受けたものである。

文献

- [1] H. Yanagisawa, and H. Watanabe, Recognition of Panel Structure in Comic Images Using Faster R-CNN, Proc of International Workshop on Image Electronics and Visual Computing (IEVC), Mar.2017.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Advances in Neural Information Processing System (NIPS), pp.1-9, 2015.
- [3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarches for accurate object detection and semantic segmentation, in IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [4] K. Simonyan and A. Zisserman, Very deep convolutional networks for largescale image recognition, arXiv preprint arXiv:1409.1556, 2014.
- [5] Y.Matsui, K.Ito, Y.Aramaki, A.Fujimoto, T.Ogawa, T.Yamasaki, and K.Aizawa, "Sketch-based Manga Retrieval using Manga109 Dataset", Multimedia Tools and Applications, Springer US, pp. 1-28, Nov.2016.
- [6] 吉田貴之(2012), 「IDIA.JP」, <<https://www.idia.jp/>>, 2017年9月16日アクセス.
- [7] TopeconHeroes, 「吹き出し素材専門サイト「フキダシデザイン」」, <<http://fukidesign.com/>>, 2017年9月16日アクセス.