

マンガ分析のための類似度に基づくキャラクタ顔画像クラスタリングの検討

添田, 碧人
九州大学工学部電気情報工学科

伊東, 栄典
九州大学情報基盤研究開発センター

<https://hdl.handle.net/2324/7329779>

出版情報 : JSAI Technical Report, SIG-KBS. 133, pp.39-44, 2024-12-05. The Japanese Society for Artificial Intelligence

バージョン :

権利関係 : Notice for the use of this material. The copyright of this material is retained by the Japanese Society for Artificial Intelligence (JSAI). This material is published on this web site with the agreement of the author(s) and the JSAI. Please be complied with Copyright Law of Japan if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) The





マンガ分析のための類似度に基づく キャラクター顔画像クラスタリングの検討

A study on similarity-based clustering of manga character facial images

添田 碧人^{1*} 伊東 栄典^{2†}
Aoto Soeda¹ Eisuke Ito²

¹ 九州大学工学部電気情報工学科

¹ School of Engineering, Kyushu University

² 九州大学情報基盤研究開発センター

² Research Institute for IT, Kyushu University

Abstract: Abstract (English) The prevalence of pirated manga and unauthorized manga videos, infringing on intellectual property rights, is increasing significantly. To address this issue, we aim to develop methods for preventing manga piracy by identifying key information such as the author, work title, volume number, and episode number from manga page images, as well as by classifying character images. Our previous work focused on character face part extraction from manga page images. Using the Manga109 dataset, we trained a machine learning model to extract face part and evaluated its accuracy in identifying character face parts.

In this paper, we attempt to cluster characters that appear in unknown manga page images using the face part extraction model which we created. We apply these models to manga page images, and the extracted face images are clustered using unsupervised learning to separate characters. This study contributes to advancing automated character analysis and character-based identification in manga images.

1 はじめに

近年、著作権を無視した海賊版マンガサイトや最新話マンガのスライド動画投稿が問題になっている。マンガのページをそのまま用いるもの以外に、著作権侵害を判定するAIを出し抜くために画像を意図的にずらしたり、一部分だけを切り出して連続的に表示する形で転載するものも出現している。後者のような工夫を加えた著作権を侵害画像および動画を検出するには、マンガ画像に含まれる登場キャラクターを判別し、さらにはキャラクター名からタイトル名や作者の判別が必要である。キャラクターの登場順と登場頻度を使えば巻数や話数の特定も可能かもしれない。

海賊版マンガ抑止のため、我々はマンガのページ画像からマンガのタイトル・巻数・話数を推定する判別

手法を研究してきた[1]。高精度でのページ画像判別を実現したけれども、提案手法は未知のマンガの判別できない。そこで次に、ページ画像に出現するキャラクターの判別を試みた。マンガ原作を多数作成した小池一夫氏が述べるように[2]、マンガにおいて登場キャラクターは非常に重要である。実際、有名マンガのキャラクターであれば、マンガの読者は一目見ただけで、それが誰であるか識別できている。人間のようなキャラクター判別機能を持つ判別機を実現できれば、未知のマンガ画像の判別も可能になる。

そこでマンガページ画像からのキャラクター顔部分抽出を試みた[3]。モデルの機械学習および評価にはManga109データセット[4]を用いた。その結果、約80%の精度で顔部分抽出を実現した。また顔部分画像からの、キャラクター判別も試みた[5]。ResNetおよびDensNetを用いることで、Manga109データセットのマンガに登場するキャラクターを高い精度で判別できた。

しかしながら、これまで検討してきた方法には限界がある。まずはタイトル数およびキャラクター数の膨大さである。日本のマンガ文化は広大で歴史も長いため、

*連絡先：九州大学工学部電気情報工学科
〒819-0395 福岡市西区元岡744
E-mail: soeda.aoto.467@s.kyushu-u.ac.jp

†連絡先：九州大学情報基盤研究開発センター
〒819-0395 福岡県福岡市西区元岡744
E-mail: ito.eisuke.523@m.kyushu-u.ac.jp

出版されたマンガの作品数は膨大で、したがってマンガに登場するキャラクター数も膨大である。全ての登場キャラクターについて、教師あり学習を適用することは現実的ではない。また過去に出現していない新キャラクターを判別する方法が無い。

本論文の対象は、著作権侵害の海賊版コンテンツの検出である。そのため、マンガ画像内部の注釈データは存在しないものの、マンガに対するメタデータは利用できるとする。つまり作者やタイトル情報は利用できる。あるタイトルに登場する主要キャラクターが誰かは、キャラクター毎の出現頻度で推定できる。ページ画像から自動抽出した顔部分画像群を、顔画像情報だけを用いてクラスタリングできれば、キャラクター毎の顔画像集合を自動作成できる。本論文では、以前に作成したキャラクターの顔部分抽出モデル [3] と、キャラクター判別モデル [5] を用いて、マンガの登場キャラクターをクラスタリングを試みる。クラスタリングには画像のベクトル化が必要である。

2 関連研究

Sharma らは 2024 年の論文 [6] で、過去 10 年間の漫画における画像セグメンテーション、分類、認識手法を体系的にレビューしている。この研究では 72 件の論文を分析し、漫画認識 (CR) 技術の現状を評価しています。主な発見として、選定された研究の 75% がコンピュータビジョン技術を使用し、13.88% が非 CV 技術、11.11% がハイブリッド技術を採用していると示している。また、対象論文の 50% で Manga109 と eBDtheque が用いられていると述べている。著者らは、CR 研究における精度基準とデータセットの標準化を強調している。

Aukkapinyo らはマンガ画像からのキャラクター顔検出を試みている [7]。自然画像の顔検出と異なり、マンガキャラクターの顔検出タスクでは、作者の描画スタイルの多様性により検出が困難になる。彼らは Mask R-CNN と転移学習を用いたモデルを提案している。学習時間とリソースを削減しつつ、高い精度での検出を実現している。MS COCO データセットからの事前学習済みの重みを使用し、限られたデータで平均 87% の精度を達成している。

キャラクター顔画像のクラスタリングについても研究されている。柳澤らは、主要キャラクターの抽出を目的としたキャラクター画像のクラスタリングを試みている [8]。キャラクター分類モデル (VGG16, VGG19, Inception V3, Xception, ResNet50) を用いて、キャラクター顔画像をベクトル化する。その後、ベクトルを次元削減 (PCA, KPCA, t-SNE) している。得られた次元削減後のベクトルを DB-SCAN によりクラスタリングしている。その結果、ResNet によるベクトル化、t-SNE による次元削減の場合が最

も良い結果であると述べている。本論文で我々は、基本的に柳澤らが示した手順によるクラスタリングを試みている。

相澤らは、顔特徴量によるクラスタリングをおこなっている [9, 10]。相澤らは、一般的な漫画の顔認識の特徴量を学習し、その後、ページやコマ情報を用いて擬似的に作成された正・負のペアを用いて個別漫画に対して深層距離学習を適用している。これにより個々の漫画に特化した顔特徴量を作成でき、高い精度で顔のクラスタリングが可能になる。実験の結果、Normalized Mutual Information (NMI) と精度 (Accuracy) がそれぞれ 0.08 および 0.16 改善されたと述べている。

3 キャラクターの顔部分特定

本研究で用いた、マンガ登場キャラクターの顔部分抽出について述べる。論文 [3] で我々は、キャラクターの顔部分抽出手法として、animeface (Viola-Jones) [11], animeface (CascadeClassifier) [12], Faster R-CNN (Detectron2) [13] の 3 つを比較した。

animeface (Viola-Jones) は、OpenCV の Viola-Jones 物体検出フレームワークに基づいたカスケード分類器の一種で、nagadomi 氏により作成・提供されている。マンガやアニメなどの二次元画像からの顔画像検出に特化した分類モデルで、単純な特徴と効率的な学習アルゴリズムを組み合わせることで、高速かつ高精度な物体検出を実現している。

animeface (CascadeClassifier) は、シンプルで高速なアニメ顔検出に適したモジュールである。このモデルは OpenCV の CascadeClassifier をベースに開発されており、アニメ顔の特徴を学習したカスケード分類器を用いて、画像からアニメ顔を検出する。

Faster R-CNN は、画像内の物体を検出する深層学習モデルの一種で、特にリアルタイムな物体検出に置いて高い性能を発揮する。CNN の高い特徴抽出能力と、Region Proposal Network による効率的な物体候補領域の生成により高精度が実現されている。なお、Detectron2 は、Facebook AI Research が開発した PyTorch ベースの物体検出ライブラリである。

3 種類の検出器についてキャラクターの検出精度を比較した。Manga109 データセットの images を入力として用いて、出力として得られたキャラクターの顔の位置情報と、annotation に記載されたキャラクターの顔の位置情報とで、重なり具合 (IoU: Intersection over Union) を算出する。ここで IoU が指定値 (今回は 0.5) を超えていたものを検出成功とした。

表 3 に判別精度を示す。なお、表 3 の以降、animeface (Viola-Jones) を af-VJ、animeface (CascadeClassifier) を af-CC、Faster R-CNN を F.R-CNN と表す。

表 1: 判別精度 (IoU=0.5)

Method	Prec.	Rec.	F1-s.	Acc.
af-VJ	0.9813	0.1502	0.2648	0.1865
af-CC	0.9927	0.3387	0.5051	0.4937
F_R-CNN	0.7661	0.8306	0.7971	0.8112

af-VJ と af-CC はどちらも Precision は高いものの Recall は低い。検出画像の多くは顔部分であるものの、取り逃がした顔部分も多いことを示す。学習させた F_R-CNN では Recall と Precision の数値が高い。

図 1 と図 2 に検出結果を示す。図 1 は Manga109 に含まれるマンガの画像で、図 2 は、Manga109 に含まれない「推しの子」の 7 巻 [14] の p.177 である。af-VJ は内の顔画像を検出できていない。af-CC モデルは、真正面で顔の大部分が描写されている画像のみを検出している。横顔やフキダシ等で顔が隠れている画像は検出できない。F_R-CNN モデルは、真正面の顔だけでなく、横顔、顔の半分がフキダシで隠れているもの、顔が小さく描写されているものも検出している。

F_R-CNN モデルは、画像の検出部分と共に、その正確さを確信度 (%) で示す。今回は確信度 50% 以上の画像を対象に表の数値を算出した。しきい値となる確信度の値を変化させて評価すれば、顔でない部分に誤反応している検出候補を消去できるため、さらに精度が上がる可能性がある。また、Faster R-CNN には確信度以外にもパラメータが多数ある。より強いモデル作成のため、パラメータ変化による評価指標の変化も検討したい。

4 キャラクタ顔画像クラスタリングの検討

4.1 画像のベクトル化

クラスタリングの対象が数値ベクトルで表現されている場合、様々なクラスタリング手法が適用できる。キャラクタの画像群をクラスタリングするには、画像をベクトルで表現する必要がある。先行研究で用いられているベクトル化は、画像分類用の CNN モデルを用いるベクトル化である。入力画像に対して、分類モデルが作成する最終ベクトルを、画像の特徴ベクトルとして用いる。本研究でも、DenseNet121 を用いて研究室内で作成した画像分類モデルを用いる。

4.2 クラスタリング手法

古くからあるクラスタリング手法は K-Means である。K-means では与えた整数 K 個のクラスターを作成

する。しかしクラスタリング対象のデータ群を知らない場合、適切なクラスタ数 K を与えることは難しい。そこで K-means を拡張した X-means や G-means が提案されている。X-means では、クラスタ数 K を逐次増加させつつ K-means を適用する。できた複数のクラスタ群のうち、情報量 BIC (Bayesian Information Criterion) により最適なクラスタ数を決定する手法である。

G-means も K-means 法の拡張であり、G-means でもクラスタ数を自動で決定する。K-means 法で初期クラスタリングを行い、各クラスタ内のデータがガウス分布に従うか統計検定を行う。この際用いる検定手法には Anderson-Darling 検定などがある。クラスタがガウス分布に従わない場合、そのクラスタをさらに分割する。

他にも DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [15] もある。DBSCAN は密度に基づくクラスタリング手法の一つで、データの分布密度に基づいてクラスタを形成する。ある空間に点集合が与えられた場合、互いに多くの隣接点を持つ集合をグループ化し、最近接点が遠い点を外れ値とする。DBSCAN では最初にクラスタ数を与える必要がなく、かつ外れ値にもロバストであるという利点がある。

機械的に作成したクラスタを評価する尺度に、Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index も有る。Silhouette Score は各データポイントが所属するクラスタにどれだけ適切に割り当てられるかを評価する尺度である。このスコアは -1 から 1 の範囲を取り、値が高いほど良いクラスタリングとする。Davies-Bouldin Index はクラスタ間の分離とクラスタ内の凝集を比較することでクラスタの品質を評価する。この指数は値が低いほど良いクラスタリングを示します。Calinski-Harabasz Index は、クラスタ内の分散とクラスタ間の分散の比率を評価する尺度で、値が高いほど良いクラスタリングと評価する。

4.3 次元削減

CNN の画像認識モデルが出力する特徴ベクトルは数千次元の場合も有る。高次元空間では、あるデータ点から見た他のデータ点群が、遠い距離に位置することになり、効果的なクラスタリングを妨げる。次元削減手法として、主成分分析 (PCA: Primary Component Analysis)、カーネル主成分分析 (Kernel PCA)、t-SNE[16] が存在する。これらの手法の適用を検討する。



図 1: 検出結果の比較 (Manga109):af-VJ,af-CC,F_R-CNN

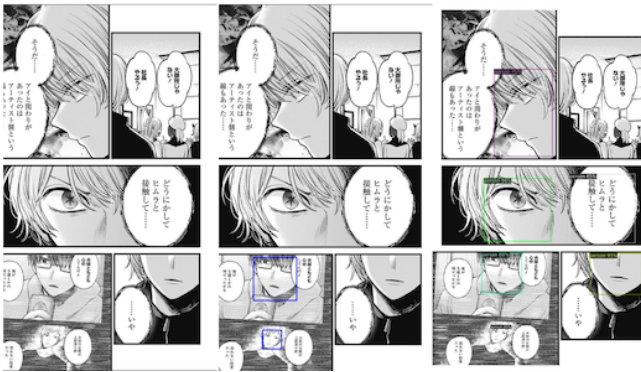


図 2: 検出結果の比較 (推しの子 7, p.177):af-VJ,af-CC,F_R-CNN

5 クラスタリング実験

キャラクター顔画像群のクラスタリングを試みた。ここでは対象データ、手法および結果を示す。

5.1 対象データ

表2に、クラスタリング対象データを示す。Manga109 データセットに含まれるキャラクター顔部分データのうち、画素数が1万 px 以上、かつ顔部分が50枚以上のキャラを対象とした。

表 2: クラスタリング対象

項目	数値
顔画像数	6,533 個
キャラクター数	61 人
キャラクター事平均画像数	107.01 個
作品あたりのキャラクター数	3.59 人

図3に、キャラクター毎の画像数を示す。縦軸が画像数で、横軸が画像数を降順に並べた場合の順位である。図の通り、画像数は一様ではなく、べき分布に近いものになっている。マンガでは、主人公の登場頻度が高

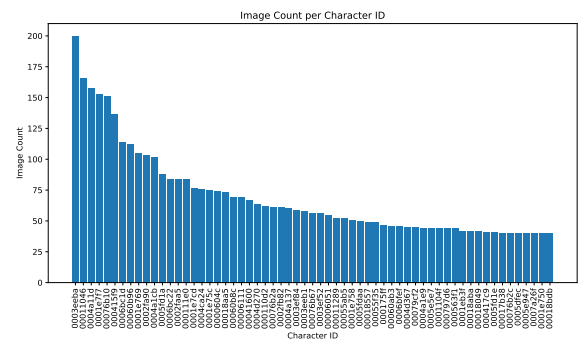


図 3: 検出結果の比較

いの比べ、他のキャラクターは頻度が少ない。そのためべき分布に近い分布になる。

5.2 用いた手法

DenseNet121 用いて研究室で学習させたキャラクター分類モデルを用いて、各キャラクターをベクトル化した。ベクトルの長さは1024である。

次に次元削減として、PCAにより1~100次元に次元削減したベクトルデータも作成した。次元削減したデータ群と、次元削減しないままの1024次元データも用意し、4つの場合を比較した。t-SNEによる次元削減も試したものの、良い結果は得られなかった。そのため、本論文ではt-SNEによる結果は示さない。

最後にクラスタリング手法として、DBSCANを拡張したHDBSCAN[?]と、X-meansを試した。X-meansではクラスタ数Kを1~65の範囲で試した。キャラクター数は61人であるため、我々が期待する結果のクラスタ数は61に収まると良いけれど、画像によりキャラクターと認識されない場合もあるため、61より多い数まで調べた。また、X-meansのクラスタ数評価尺度として、Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index を算出した。

5.3 結果 1:HDBSCAN

画像の特徴ベクトルを、1024次元から t-SNE で 2,3次元に削減し、HDBSCAN でクラスタリングした。結果を図 4, 5 に示す。得られたクラスはキャラクタ毎に分類されておらず、よく分からない結果になった。

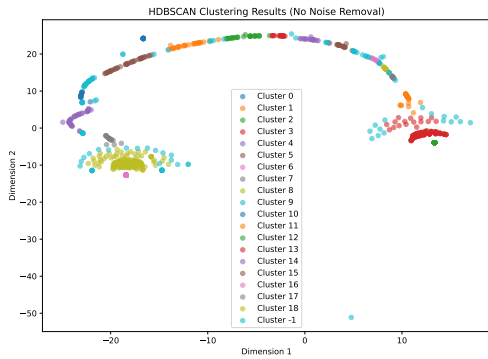


図 4: PCA 2次元, HDBSCAN

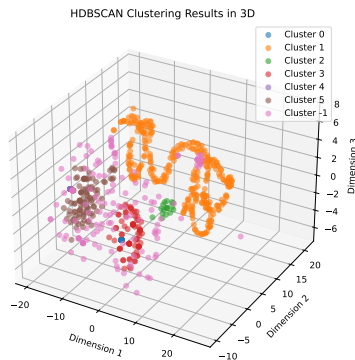


図 5: PCA 3次元, HDBSCAN

5.4 結果 2:X-means

次に X-means によるクラスタリングを試みた。図 6 に、PCA による次元数を 10, 20, ... 100 に削減した場合における、クラス数ごと評価値 (Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index) を示す。図を見ると、どの次元数でもクラス数が少ない場合に評価尺度の値が大きい。そのため、期待していたキャラクタ毎のクラスができていない。

6 おわりに

著作権侵害の海賊版コンテンツの検出のため、マンガ画像に登場するキャラクタ判別が必要である。本論文では最初にマンガページ画像からの登場キャラクタの顔部分抽出について述べた。Manga109 データセットのデータで、Faster R-CNN による顔部分抽出モデルを学習させた。その結果、約 80%の精度で顔部分を抽出できた。

顔画像が抽出できても、それがどのキャラクタであるかを知る必要がある。顔画像から、どのキャラクタかを判別する手助けとして、クラスタリングによる画像群のグループ分割が望ましい。そこでクラスタリング手法を検討し、実際に Manga109 データセットに適用してみた。事前に Manga109 データセットで学習した DenseNet121 モデルによるキャラクタ分類器を用いて、顔画像をベクトル化した。ベクトル型のデータ群を、次元削減し、HDBSCAN と X-means でクラスタリングしてみた。実験したところ、良いクラスライブラリ結果は得られなかった。今後はベクトル化手法、次元削減法、クラスタリング手法について詳細に調べ、より良いクラスタリングを実現したい。

今回は学習済みのキャラクタでクラスタリングを試みたけれど、海賊版マンガ判別には、教師あり学習で学習していないキャラクタも判別できると良い。未学習のキャラクタでもクラスタリングできる方法を考えていきたい。

謝辞

本研究は、JSPS 科研費 24K15654 の助成を受けた。

参考文献

- [1] 山仲一颯, 伊東栄典: 大規模マンガ画像判別手法の検討, 火の国情報シンポジウム 2024, 情報処理学会, pp. A4-1 (2024).
- [2] 小池一夫: 小池一夫のキャラクター新論 ソーシャルメディアが動かすキャラクターの力, ゴマブックス (2017).
- [3] 添田碧人, 伊東栄典: マンガ顔画像からのキャラクタ判別に関する検討, 2024 年度 (第 77 回) 電気・情報関係学会九州支部連合大会, Vol. 09-1P-09, pp. 186-187 (2024).
- [4] Aizawa, K., Fujimoto, A., Otsubo, A., Ogawa, T., Matsui, Y., Tsubota, K. and Ikuta, H.: Building a Manga Dataset “Manga109” with Annotations for Multimedia Applications, *IEEE MultiMedia*, Vol. 27, No. 2, pp. 8-18 (online), doi:10.1109/mmul.2020.2987895 (2020).
- [5] 霜出秀太, 伊東栄典: マンガ顔画像からのキャラクタ判別に関する検討, 2024 年度 (第 77 回) 電気・情報関係学会九州支部連合大会, Vol. 09-1P-10, pp.188-189 (2024).

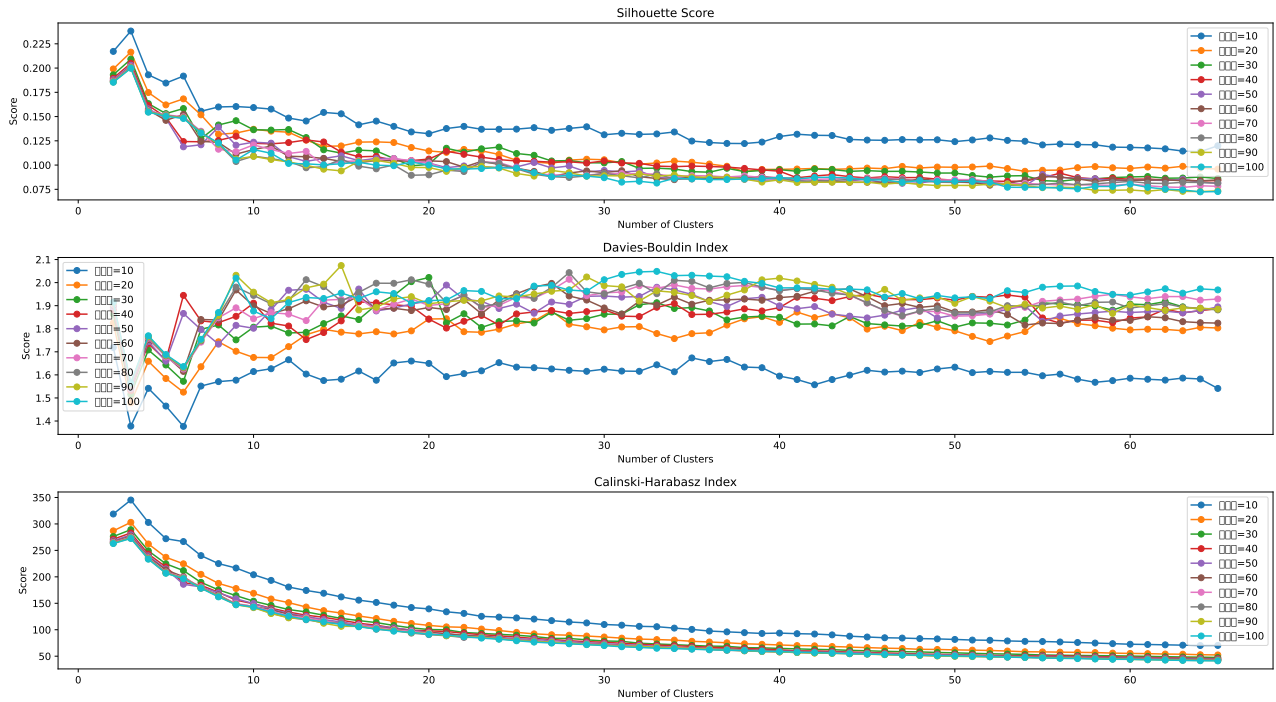


図 6: クラスタの評価: PCA 10,20,...,100 次元, X-means (K=1,...,65)

- [6] Sharma, R. and Kukreja, V.: Image segmentation, classification and recognition methods for comics: A decade systematic literature review, *Engineering Applications of Artificial Intelligence*, Vol. 131 (online), doi10.1016/j.engappai.2023.107715 (2024). Cited by: 3.
- [7] Aukkapinyo, K. and Hotta, S.: Manga Face Detection on Various Drawing Styles Using Region Proposals-Based CNN, *Science and Technology Asia*, Vol. 28, No. 1, pp.120–135 (2023).
- [8] 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 (オンライン), doi10.3169/itej.73.199 (2019).
- [9] Tsubota, K., Ogawa, T., Yamasaki, T. and Aizawa, K.: Adaptation of manga face representation for accurate clustering, *SA'18: SIG-GRAPH Asia 2018 Posters*, pp.1–2 (online), doi10.1145/3283289.3283319 (2018).
- [10] 坪田亘記, 相澤清春: 個別漫画への顔特徴量の適応によるクラスタリング, *日本画像学会誌*, Vol. 58, No. 5, pp.523–527 (2019).
- [11] nagadomi: lbpcascade animeface, https://github.com/nagadomi/lbpcascade_animeface. Accessed: 2024-08-05.
- [12] OpenCV: OpenCV haarcascades, <https://github.com/opencv/opencv/tree/master/data/haarcascades>. Accessed: 2024-10-31.
- [13] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *Advances in Neural Information Processing Systems*, Vol. 28, pp.91–99 (2015).
- [14] 赤坂アカ, 横槍メンゴ: 推しの子 7, 集英社 (2020).
- [15] Ester, M., Kriegel, H.-P., Sander, J. and Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pp.226–231 (1996).
- [16] van der Maaten, L. and Hinton, G.: Visualizing data using t-SNE, *Journal of Machine Learning Research*, Vol. 9, pp.2579–2605 (online), doi10.1145/1390155.1390283 (2008).