

顔領域に着目した漫画の著者推定に関する研究 A Method to Estimate Author of Manga Focused on Face Area

笠原 尚弥[†] 大野 将樹[‡] 獅々堀 正幹[‡]
Naoya Kasahara Masaki Oono Masami Shishibori

1. はじめに

漫画は日本が世界に誇るコンテンツ産業であり国内の市場規模は 2016 年度で 4,456 億円にもものぼる[1]。そのため日本は漫画大国といわれ漫画画像が多く存在している。読者が漫画画像を閲覧した際に、著者名や作品名などの関連情報を得られれば漫画をより楽しむことができる。漫画画像に関連した情報を知りたい場合、キーワードを検索により様々な情報を得ることができるが、適切にキーワードを選択することは容易ではない。

本研究では、畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) に基づいて顔領域の特徴を学習し、漫画画像の著者を推定する手法を提案する。

2. 顔領域に着目する理由

顔領域は、シーン毎の違いが最も小さくキャラクターの服装や髪形による影響が少ないこと、また、全てのシーンに共通して現れる要素であるため、著者を識別するための要素として有効である。

3. 関連研究

板持ら[2]は、イラスト画像の著者を同定する手法を提案した。SIFT 特徴量に基づいてイラスト画像をクラスタリングし Bag-of-Features 表現(BoF)に変換する。顔画像特徴量を用いる場合、イラスト画像に対して顔認識を行い顔部分のみを抽出した画像の BoF をイラスト画像の BoF に対する重みとして使用する。そして最終的に出力される BoF をもとに SVM を用いて識別器を作成する。

板持らの手法は顔以外の領域も使用しているが、本研究は顔部分のみに着目して著者を推定する点に特徴がある。板持らは特徴量に SIFT、識別機に SVM を用いているが、本研究では CNN を用いる点に相違がある。

4. 提案手法

提案手法の概要を図 1 に示す。複数人の著者の漫画画像に対して顔検出を行い、生成された画像を用いて CNN で学習を行い、学習モデルを作成する。次に、著者が未知の漫画画像を入力とし、入力画像に対して顔検出を行い、生成された画像に対して、学習モデルを用いて著者の推定を行う。今回、顔検出は手動で行った。

CNN は、VGG16 の ImageNet 事前学習モデルを初期値として fine-tuning を行い学習が収束した時点でのモデルを特徴抽出器として利用した。

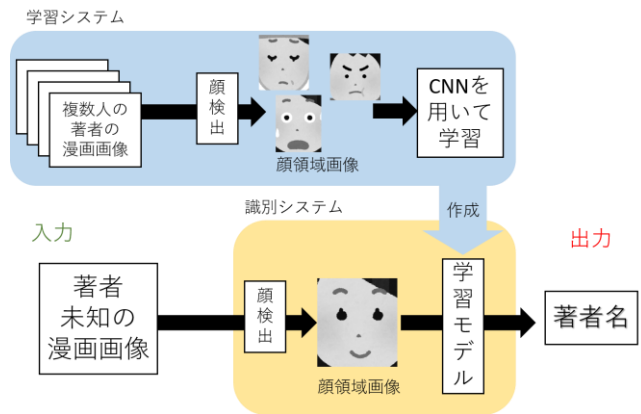


図 1: 提案システムの概要

5. 実験

5.1 データセットの作成

オンラインブックストアから得た漫画画像から顔領域を手手でトリミングし、顔画像を収集した。その際、顔領域画像について、画像の上端は眉の上、下端は口元の下まで、左右の端は目の横まで、顔が正面を向いておりどちらかの瞳が確認できる画像であると定義付けを行った。収集した漫画画像は 5 人の著者から成る合計 3,829 枚であり、作品ごとや同じ作品でも登場するキャラクターごとにグループを分け、著者ごとに 5 つずつのグループに分割した。

5.2 方法

データセットの各グループから 50 枚ずつを学習モデルの作成に用い、残りの画像を識別に用いた。学習では、学習データとテストデータに同じキャラクターが存在する事によってキャラクター推定の実験とならないようにするため、データセットのグループごとに分割し K 分割交差検証を行い、5 つの学習モデルを作成した。識別では、全ての学習モデルで著者推定を行い、各学習モデルで得られた予測値に対しクラスごとに平均を取って算出した予測値を元に予測結果を出力する。提案手法の有効性を検証するための比較手法として、画像ごとに HOG 特徴量を利用した特徴ベクトルを抽出しそれらを Support Vector Machine(SVM) を用いて分類する手法を用いた。

5.3 評価

識別結果を用いてクラスごとに再現率・適合率・F 値を求めた。F 値は、再現率と適合率の調和平均である。

$$\text{再現率} = \frac{\text{正しく分類した画像枚数}}{\text{真のクラスの画像枚数}} \quad (1)$$

$$\text{適合率} = \frac{\text{正しく分類した画像枚数}}{\text{クラスに分類された画像枚数}} \quad (2)$$

[†] 徳島大学大学院先端技術科学教育部 Graduate School of Advanced Technology and Science, Tokushima University

[‡] 徳島大学大学院社会産業理工学研究部 Graduate School of Technology, Industrial and Social Science, Tokushima University

再現率・適合率・F 値それぞれの総和をクラス数で割った平均値を表 1 に示す。すべての数値において提案手法の方が比較手法よりも高い数値が得られた事から、提案手法の優位性が示された。

表 1: 実験の結果

	再現率	適合率	F 値
比較手法	0.58	0.54	0.53
提案手法	0.78	0.70	0.72

識別の一例をそれぞれ図 2 に示す。図 2 において元画像に記した著者名は真の著者名、Grad-CAM[3] 画像に記した著者名は各学習モデルで出力された予測結果の著者名、結果の著者名は識別結果の著者名を示している。それぞれの Grad-CAM 画像を見てみると、予測結果ごとに同じ部分の特徴を学習していることが多いことが分かった。例えば、画像 A では全てのモデルでキャラクターの鼻や口周辺の特徴が学習されている。また、画像 B では真の著者が画像 A と異なるが、画像 A と同様に全てのモデルでキャラクターの鼻や口周辺が学習されており、画像 C では著者 B と推定したモデルは全てキャラクターの額の特徴が学習されており、真の著者である著者 C と推定したモデルは全てキャラクターの鼻や口周辺の特徴が学習されている。この事から、画像ごとに正解と判断している部分が異なっていることが分かった。これらの事から提案手法の著者推定における法則性は確認できなかった。

元画像	モデル1	モデル2	モデル3	モデル4	モデル5	結果
著者A 画像A	著者A	著者A	著者A	著者A	著者A	著者A
著者B 画像B	著者B	著者B	著者B	著者B	著者B	著者B
著者C 画像C	著者C	著者B	著者B	著者B	著者C	著者B

図 2: 実験に対する出力例

5.4 顔画像に着目しない実験

顔領域に着目している理由を裏付けるために追加実験を行った。

画像の収集には Manga109[4][5]を利用し、漫画画像のコマ領域を切り出す事で画像を収集しデータセットを作成した。収集した漫画画像は、5人の著者から成る各著者 50枚ずつ合計 250枚である。各著者 50枚のうち 35枚を学習データ、15枚をテストデータとして使用し学習を行った。追加実験では、交差検証は行わない。比較手法は実験と同様の手法を用いる。追加実験で得られた結果を用いて実験と同様に、再現率・適合率・F 値それぞれの総和をクラス数で割った平均値を求める。その結果を表 2 に示す。全ての数値において提案手法の方が比較手法よりも高い数値が得られたため、追加実験においても提案手法の優位性が示された。

表 2: 追加実験の結果

	再現率	適合率	F 値
比較手法	0.43	0.62	0.38
提案手法	0.63	0.64	0.63

識別の一例を図 3 に示す。図 3 において、元画像に記した著者名は真の著者名、結果の著者名は識別結果の著者名を示している。正解例の画像は、大きく描かれたキャラクターの特徴が学習されていることが分かる。一方、不正解例の画像は、キャラクターの特徴が学習されておらず、コマの中にある小さなコマの特徴が学習されている。これらの結果から、著者が描くキャラクターに著者特有の特徴が多いことが分かった。

	元画像	Grad-CAM画像	結果
正解例	著者D 画像D		著者D
不正解例	著者E 画像E		著者D

図 3: 追加実験に対する出力例

表 1 と表 2 で示した結果を比較する。提案手法、比較手法のそれぞれともに実験の結果の方が高い数値が得られた。この事から、キャラクターの中でも顔部分は特に著者特有の個性が現れ、そこに着目したほうが著者推定の精度が高くなることが分かった。

6. おわりに

本研究では、ディープラーニングを用いて著者毎の特徴を抽出することによって漫画画像の著者推定を試みた。顔領域に着目するかしないかと、ディープラーニングを用いるか既存手法を用いるかという 2つの観点から結果の比較を行い、顔領域に着目した提案手法を用いた方がより高い F 値を得ることができた。今後の課題として、提案手法における著者推定の法則性を見つけることが挙げられる。また、顔領域の抽出を自動で行えるようにすれば提案するシステムがより現実的なものになると考えられる。

参考文献

- [1] アニメーションビジネス・ジャーナル, animationbusiness.info, 参照 2019 年 6 月.
- [2] 板持貴之, 三輪誠, 田浦健次朗, 近山隆, “イラストの作者同定アルゴリズム”, 情報処理学会第 74 回全国大会講演論文集, pp.209-211, 2012.
- [3] Ramprasaath R.Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, The IEEE International Conference on Computer Vision, pp.618-626, 2017.
- [4] Manga109, www.manga109.org, 参照 2019 年 6 月.
- [5] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Sketch-based Manga Retrieval using Manga109 Dataset, Multimedia Tools and Applications", Springer, Issue 20, Vol.76, pp.21811-21838, 2017.