

畳み込みニューラルネットワークによる 漫画における顔の年代別特徴量推移の検出

Detecting Temporal Facial Feature Transitions in Manga Using Convolutional Neural Networks

塚本 玲^{1*} 赤木 芽¹ 江草 遼平¹

Rei TSUKAMOTO¹, Kaya AKAGI¹, and Ryohei EGUSA¹

¹ 千葉商科大学

¹Chiba University of Commerce

Abstract: This study aims to quantitatively demonstrate the characteristics and transitions of facial expressions in manga characters across different decades using clustering based on Convolutional Neural Networks (CNNs). Additionally, feature maps were generated for character face images from each decade to interpret the features emphasized by CNN. In this study, we utilized the Manga109 dataset, which consists of 109 volumes published between the 1970s and 2010s, as constructed by Aizawa et al. (2018). Clustering and feature map generation were conducted on manga from the 1970s to the 2000s. The results confirmed a transition in facial feature representations over the decades. Furthermore, the results suggested that CNN primarily distinguishes different decades based on the depiction of eyes and noses.

研究背景・目的

図 1 は 1970,80,90,2000 年代を代表する集英社の少女漫画の表紙画像である。図 1 はあくまで年代を代表する 1 つの漫画であるが、漫画におけるキャラクターの絵柄が年代によって一定の傾向を持って変化していることは広く認識されているところである。



図 1 集英社を代表する年代別少女漫画
(画像出典:池田(1972),池野(1982),神尾(1992),矢沢(2001)より筆者編集)

漫画におけるキャラクターの絵柄に関して,成田他(2017)は CNN(畳み込みニューラルネットワーク)を利用して漫画ごとの絵柄の質を定量的に判別することに成功している。また,坪田・相澤(2019)は,CNN

によって得られた特徴量を基にキャラクターのクラスタリングを実現している。いずれの研究も,CNNによって漫画の絵柄を識別することには成功している。しかし,これらの研究は漫画の 1 キャラクター,及び漫画ごとの特徴を対象としており,年代というより大きな区分で特徴を識別しているわけではない。理論的には同様の手法で年代別の識別が可能であることを示すことは,漫画が年代別に一定の特徴を持っていることを定量的に実証することが可能であるといえる。

本研究では,これらの先行研究を参考に,対象をキャラ・漫画から年代ごとに変化し,1)年代ごとの絵柄の特徴を識別することが可能であることを示す,2)年代ごとの特徴が存在した場合にそれぞれの特徴を明らかにすることの 2 点を目指す。

手法

データ

データセットとして,相澤他(2018)による

* 連絡先: 千葉商科大学

〒272-0827 千葉県市川市国府台1丁目3-1

E-mail: c210323@st.cuc.ac.jp

Manga109(ver.2023/12/07)を用いる。Manga109 は、1970 年代から 2010 年代までに公開された日本のプロの漫画家による漫画 109 冊が収録されている学術漫画データセットである。内訳として、70 年代が 7 冊、80 年代が 24 冊、90 年代が 45 冊、00 年代が 32 冊、10 年代が 1 冊である。Manga109 は作品事にアノテーションが付与されており、キャラクターid、キャラクターの顔、全身、活字テキスト及び一部の描き文字、コマ枠にそれぞれ、矩形領域と、活字テキスト及び一部の描き文字である場合はテキスト内容が付与されている。

分析手法

本研究ではデータ数の僅少な 10 年代を除く 00 年代までの漫画を研究対象とする。アノテーションを基に抽出されたキャラクターの顔画像データ 46092 枚について、70 年代から 00 年代までの 10 年代ごとの 4 クラスにラベリングする。データのサイズは 224×224 ピクセルに統一し、またデータに多様性を持たせるために画像反転の処理を加え、その他にテンソル変換や正規化処理を施している。

本研究では、CNN(ConvNeXt) を用いて画像認識モデルを実装する。ConvNeXt は Vision Transformer を参考にしつつ、ResNet(ResidualNetwork)を基盤として作られた CNN であり、生成 AI が台頭し始める 2022 年以前の CNN 画像認識モデルとしては最高峰の性能を有している。成田(2017)では、CNN モデルとして AlexNet が使用されているが、本研究ではより優れたモデルである ConvNeXt を用いて研究を行う。

学習においてはリサイズ等の前処理を施した顔画像データについて、8 割を教師データ、2 割を評価データとして学習を行い、モデルによる分類の精度について確認を行う。

ConvNeXt の実施におけるハイパーパラメータは表 1 の通りである。

表 1 CNN ハイパーパラメータ

モデル	ConvNeXt_Tiny
損失関数	クロスエントロピー損失
オプティマイザ	Adam
学習率	0.0001
バッチサイズ	32
エポック数	8

その後、モデルが注目する特徴を捉えるために、CNN(ConvNeXt)の最終層の出力特徴量を PCA 及び t-sne によって次元削減し特徴量の可視化を実施する。また、Grad-CAM を用いて特徴マップを作成し、予測の根拠についての説明を試みる。

結果

学習

図 2 及び図 3 は、それぞれ学習データ(train)、評価用データ(val)における予測の正答率と損失関数の返り値を示したグラフである。CNN ではモデルの性能を向上させるため繰り返し学習を行う。その際、データを一巡して学習する回数をエポック数と呼ぶ。エポック数を増やすほど一般的に性能が高まるが、数が多過ぎる場合には過学習が起こり、かえってデータの正答率が下がってしまう場合がある。したがって、分析に先立って適切なエポック数を検証した。

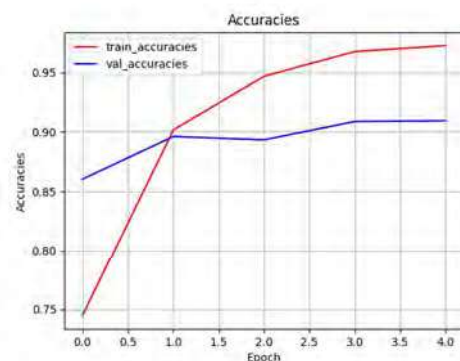


図 2 epoch 数別の正答率

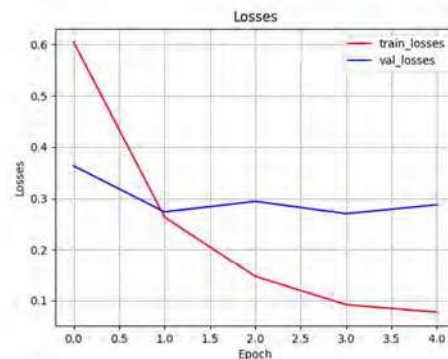


図 3 epoch 数別の損失

図 2 から本研究に使用する学習データの予測正答率については、epoch=3 程度で val_accuracies が 0.9 以上と十分な性能を発揮している。また、その後の推移も安定的でありこれ以上の大きな改善は見込めない。過学習による正答率の低下も発生していない。損失関数に関しても同様の動きをしており、epoch=3 で十分な学習がなされているとみなせる。したがって、以下の研究ではこの epoch=3 における学習結果を利用する。

特徴量のクラスタリング

CNN によって出力された特徴量を取得し、クラス

タリングによる特徴量の可視化を行った。クラスタリングを行う際、CNN によって出力された特徴量は次元が膨大すぎるため、次元削減を行い人間が解釈可能な次元にまで情報の圧縮を行った。次元削減の手法として、PCA 法と t-sne 法の二つの手法を利用し比較した。

図 4、図 5 はそれぞれ、PCA、t-sne による次元削減された特徴量のクラスタリング結果を表している。また、図中にはそれぞれの点に対応する漫画画像を付

記している。

図 4 図 5 より、二つの次元削減手法どちらも左から右にかけて弧を描くように年代ごとの識別が行われていることが見て取れる。年代ごとの識別性能の高さは、t-sne の方により強く表れている、一方で PCA 法では 70 年代と 80 年代との間に比較的類似性が現れていることが分かった。以降識別性能の高い PCA によって次元削減された特徴量及びクラスタリングの結果を分析対象として利用する。

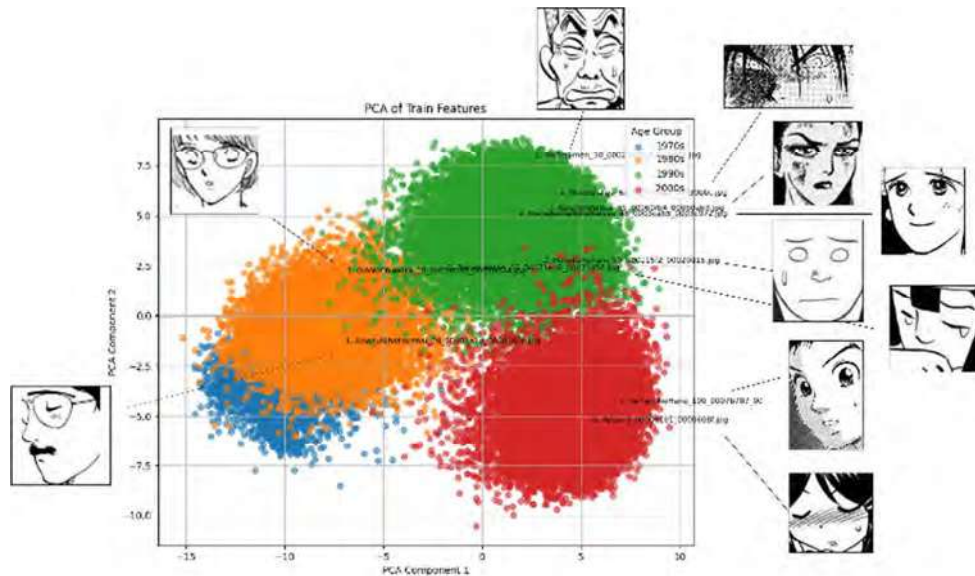


図 4 PCA によるクラスタリング

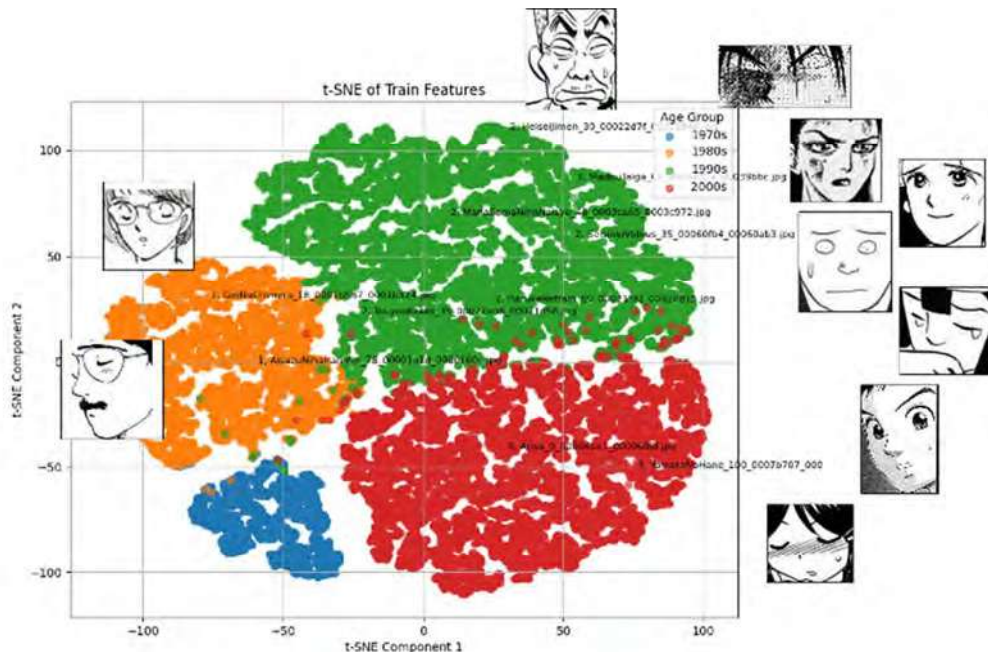


図 5 t-sne によるクラスタリング

予測の傾向

図 6 は評価データにおける正解ラベルとモデルの予測結果の列相対度数を表している。年代の自交点（対角線）が正答率であり、それ以外の行に対する数値が誤って行の年代を選択した割合を表す。

図 6 から、それぞれの年代の正答率は、70 年代が 91%、80 年代が 89%、90 年代が 90%、00 年代が 94%という結果となった。いずれの年代においても 90%前後の正答率を出せており、図 5 の PCA によるクラスタリングの結果とも整合的である。

特徴マップによる特徴量の可視化

CNN が画像を分析する際、どの部位に注目しているのかを可視化するために特徴マップを作製した。CNN における最終識別層の出力に対して Grad_CAM によって、CNN の判断基準を人間にも解釈しやすくしたものが図 7 である。

図 7 から、どの年代においても共通して注目されているのは、目や眉、眉間や鼻などの部位である。このことから年代ごとの特徴の違いは、主にこれらの部位における描き方によって現れるものと推察できる。

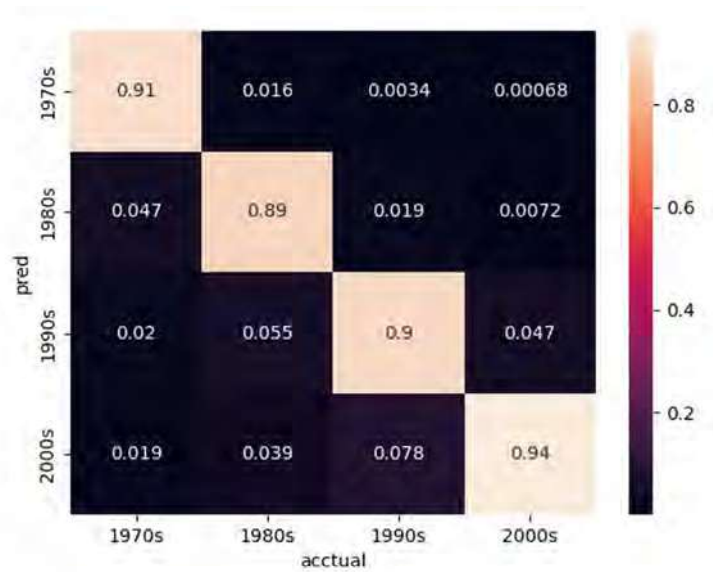


図 6 予測と正解の列相対度数

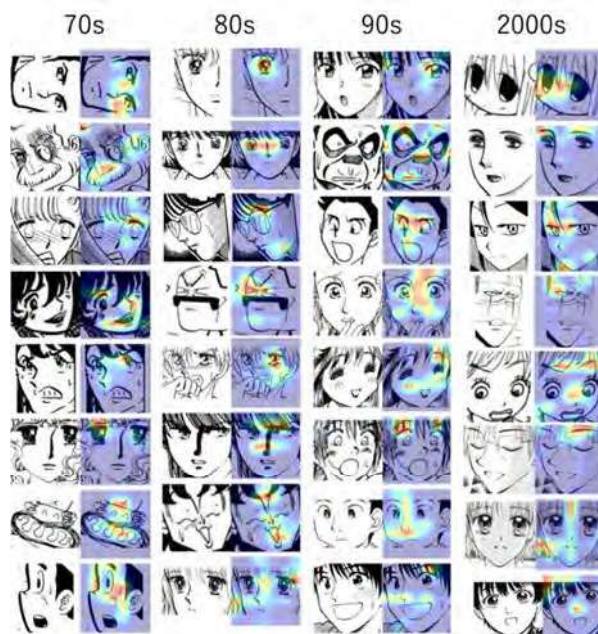


図 7 Grad_CAM による特徴マップの可視化画像

考察

t-sne によるクラスタリングを行った結果、明確なクラスが確認され、それぞれの年代における予測精度も高いことから、1970,80,90,00 年代ごとに識別可能な絵柄の特徴が存在することが明らかとなり、本稿の第一の目的は達成されたと言える。

また、t-sne のクラスタリング結果では、時計回りに弧を描くように年代の推移があることが確認できた。これは、年代ごとの特徴の変化が離散的ではなく、機械判読可能な特徴量が年代の推移と共に変化していることを表している。また、時計回りの変化が確認できることから t-sne による 2 次元への次元削減の範囲では、70 年代の絵柄の特定の特徴に対して 00 年代は回帰傾向にあるといえる。

本稿の第二の目的であるこれらの特徴の推移に関して、具体的にどのような特徴が年代ごとに推移しており、回帰しているのかを明らかにすることは機械的には実施できないため、CNN による出力を人間が解釈する必要がある。

CNN が最終的に画像のどの部分を識別に用いているかを可視化する特徴マップでは、年代ごとの特徴を判別するために、目や眉、眉間や鼻といった部位に注目していることが確認できた。このことから年代ごとの特徴の違いは主にこれらの部位に多く表れていると推察できる。年代ごとの具体的な特徴の違いについては、今回の研究では特定には至らなかったが、鼻や眉間などの部位ごとに詳細に区分していくことで、特徴の違いを特定できる可能性がある。クラスタリングの結果からは、右上に属する 90 年代の画像は鼻筋や皺など線が多く写実的な表現がなされている傾向がある。またトーンやベタなどが多用されており画像全体の印象が暗い。一方で、70 年代や 90 年代など画像の下側に属する画像は鼻筋が省略されており、デフォルメが多用されている傾向が見て取れる。このことから、70 年代から 00 年代にかけて、デフォルメ→写実→デフォルメといった回帰があることが予測される。ただし、このことを明らかにするためには、漫画表現に関する専門家の知識を借りながらより多数の画像で網羅的な検証が必要となるであろう。

今後の展望については、鼻や眉間などの部位に特徴が表れていること、デフォルメ調と劇画調との間で年代ごとに推移していることを仮説として、漫画の変遷に関しての有識者も交えた研究を実施する予定である。

謝辞

本研究は、「千葉商科大学・数理データサイエンス教育プログラム」における「特別講義(データサイエンス)」の一環であり、千葉商科大学 基盤教育機構による助成を受けている。

参考文献:

- [1] 池田理代子: ベルサイユのばら, 1, 集英社 (1972)
- [2] 池野恋: ときめきトゥナイト, 1, 集英社(1982)
- [3] 神尾葉子: 花より男子, 1, 集英社(1992)
- [4] 矢沢あい: NANA-ナナ-, 3, 集英社(2001)
- [5] 成田嶺, 小川徹, 松井勇佑, 山崎俊彦, 相澤清晴: 深層学習を用いたスケッチに基づく漫画検索, 人工知能学会全国大会論文集, Vol. JSAI2017, No. 31, セッション ID 3H1-OS-04a-2, (2017)
- [6] 坪田亘記, 相澤清晴: 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)
- [7] 相澤清晴, 松井勇佑, 藤本東, 大坪篤史, 小川徹: 学術漫画データセットの構築～Manga109～, 映像情報メディア学会誌, Vol. 72, No. 5, pp.358-362, (2018)
- [8] Xiaoran Q., Yafeng Z., Yonggang L., Siwei W., Yongtao W., and Zhi T.: Progressive deep feature learning for manga character recognition via unlabeled training data. In Proceedings of the ACM Turing Celebration Conference - China (ACM TURC '19), Article 91, pp.1–6, (2019)
- [9] Zhuang L., Hanzi M., Chao-Yuan W., Christoph F., Trevor D., and Saining X.: A ConvNet for the 2020s, arXiv:2201.03545, (2022)