

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

千葉商科大学 商経学部 塚本 玲

指導教員: 赤木 茅, 江草 遼平

※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)」の一環として実施されている
また,千葉商科大学基盤教育機構の助成を受けている.

研究背景・目的

- 昔の漫画から現在の漫画にかけて、絵柄が変化してきている
- 絵柄の推移に関する定量的な分析は少ない
- 年代ごとの特徴の識別は可能なのか
- 年代ごとの特徴があった場合、それぞれの特徴を明らかにできないか



池田(1972),池野(1982),神尾(1992),矢沢(2001)

先行研究と目的

- 成田, 小川, 松井, 山崎, 相澤 (2017) 深層特徴を用いたスケッチに基づく漫画検索
 - Manga109データを利用し漫画の顔画像をキャラクタ毎にCNN(AlexNet)を利用して識別
 - CNNによって漫画ごとの特徴は分析可能である一方で年代ごとの漫画など複数漫画の特徴に関しては扱っていない
- 坪田, 相澤 (2019) 個別漫画への顔特徴量の適応によるクラスタリング
 - CNN(ResNet-50)から得られた特徴量をk-means法によってクラスタリングすることで, 個別の漫画キャラクターごとのクラスタリング
 - キャラごとの正負ペアをラベルに付加することで精度向上
- これらの手法を参考に, キャラ・漫画から, 年代ごとにCNN特徴量の抽出, クラスタリングを行い, 年代ごとの特徴を把握する.

対象データ

- Manga109

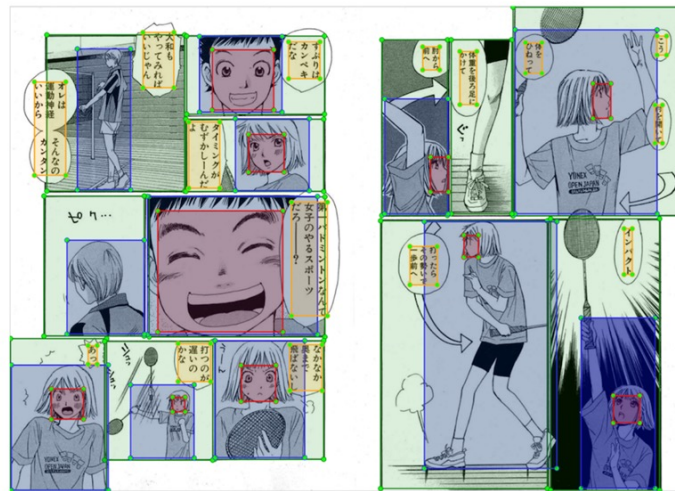
- 日本漫画のメディア処理の学術研究使用を目的とし、相澤ら(2018)によって構築
- 日本のプロの漫画家によって描かれた1970年代～2010年代までの漫画109冊で構成
- 幅広いジャンルや対象読者層の漫画が収録
- 本研究では、作品数の少ない2010年代を除いた2000年代までの作品を研究対象とする



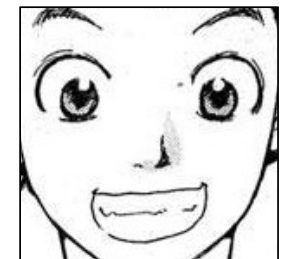
データ処理

- Manga109からキャラクターの顔画像をアノテーションを基に抽出
- 漫画顔画像を年代別にラベリング
- そのうち8割を学習用,2割をテストデータとして使用

年代	冊数	顔画像数
1970	7	3205
1980	24	9152
1990	45	19057
2000	32	14678
合計	108	46092



YamatoNoHane ©Saki Kaori



分析の流れ

データ変換(前処理)

畳み込みニューラルネットワーク(ConvNeXt)

- エポック数の決定

モデル最終層の出力特徴量を取得・特徴量を次元削減

クラスタリングによる特徴量の可視化

特徴マップの作製

データ変換

リサイズ (Resize)

- 画像のサイズを揃えます.今回は224 × 224 ピクセルに統一

画像反転 (Horizontal Flip)

- 水平方向に画像を反転させることで,データに多様性を付加

テンソル(Tensor)変換

- Tensor:データの多次元配列
- 画像の場合は,高さ,幅,チャンネル数(RGBカラー等の色)の3次元テンソル

正規化(Normalize)

- 画像データを特定の範囲や分布に変換してモデルが効率的に学習できるようにする手法
- CNNで学習される画像データは,0から255の範囲のピクセル値(色や濃淡の数値)で表現
- ばらつきが大きすぎると学習が不安定になるため,0から1の範囲に変換

ConvNeXt

- Meta(旧Facebook)によって発表されたモデル
- Vision Transformer (ViT) を参考にしつつResNet (Residual Network) を基盤として作られたCNN
- 2022年時点(生成AI以前)でのCNN画像認識モデルとしては最高性能
- PyTorch上で実装

先行研究

本研究

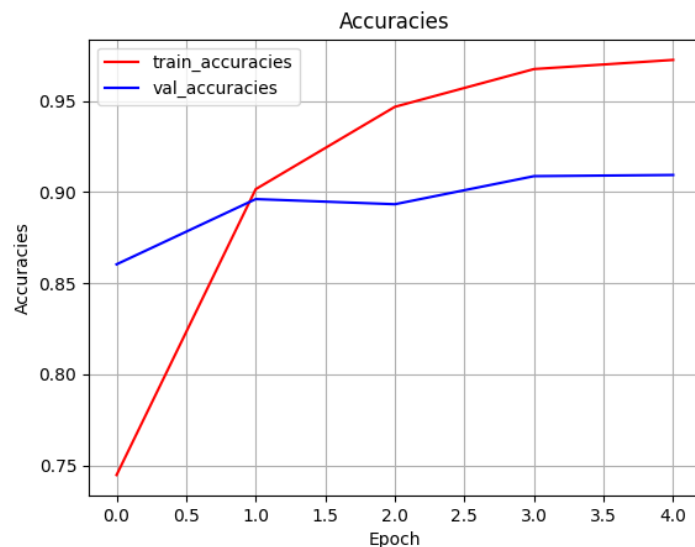
年代	モデル名	特徴・概要
1998	LeNet	手書き数字認識に使われた初期のCNNモデル. 主にMNISTデータセットで使用.
2012	AlexNet	ImageNetで初めて畳み込みニューラルネットワークを使用し, トップ5エラー率を大幅に改善.
2014	VGGNet	小さな3x3のフィルタを多層に積み重ねたアーキテクチャ.VGG16とVGG19が特に有名.
2014	GoogLeNet (Inception)	“Inception Module”を採用し, 計算効率と精度を両立.
2015	ResNet	残差ブロックを使用し, 非常に深いネットワークのトレーニングを可能に.ImageNetで優勝.
2016	DenseNet	各層がすべての前層からの入力を受け取る密結合アーキテクチャ.
2017	Xception	畳み込みの代わりに「深さ方向の畳み込み (Depthwise Separable Convolutions)」を使用.
2017	MobileNet	軽量かつ効率的なCNNアーキテクチャ. モバイルデバイス向けに最適化.
2020	Vision Transformer (ViT)	画像をパッチに分割し, トランスフォーマーアーキテクチャを使用したモデル.
2021	Swin Transformer	階層的なトランスフォーマーアーキテクチャで, 局所的な窓 (ウィンドウ) を用いた画像認識モデル.
2022	ConvNeXt	Vision Transformerのアイデアを取り入れた畳み込みネットワークの進化版.

ハイパーパラメータ

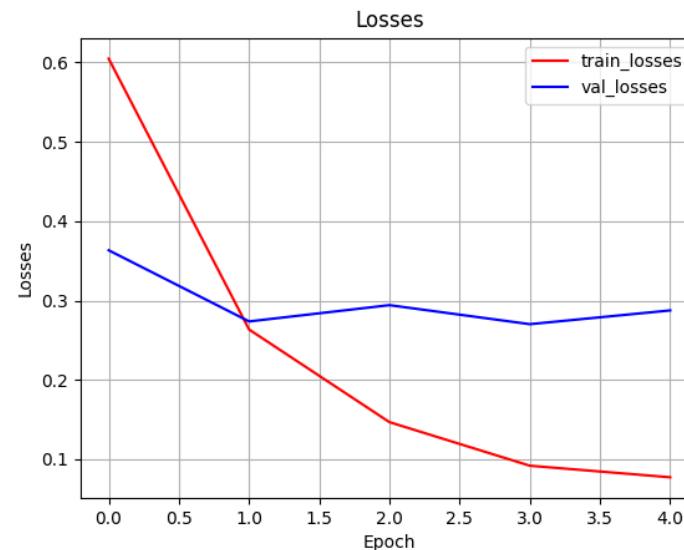
モデル	ConvNeXt_Tiny
損失関数	クロスエントロピー損失
オプティマイザ	Adam
学習率	0.0001
バッチサイズ	32
エポック数	4

結果:エポック数の推移

テストデータにおける予測の正答率



損失関数(クロスエントロピー損失)



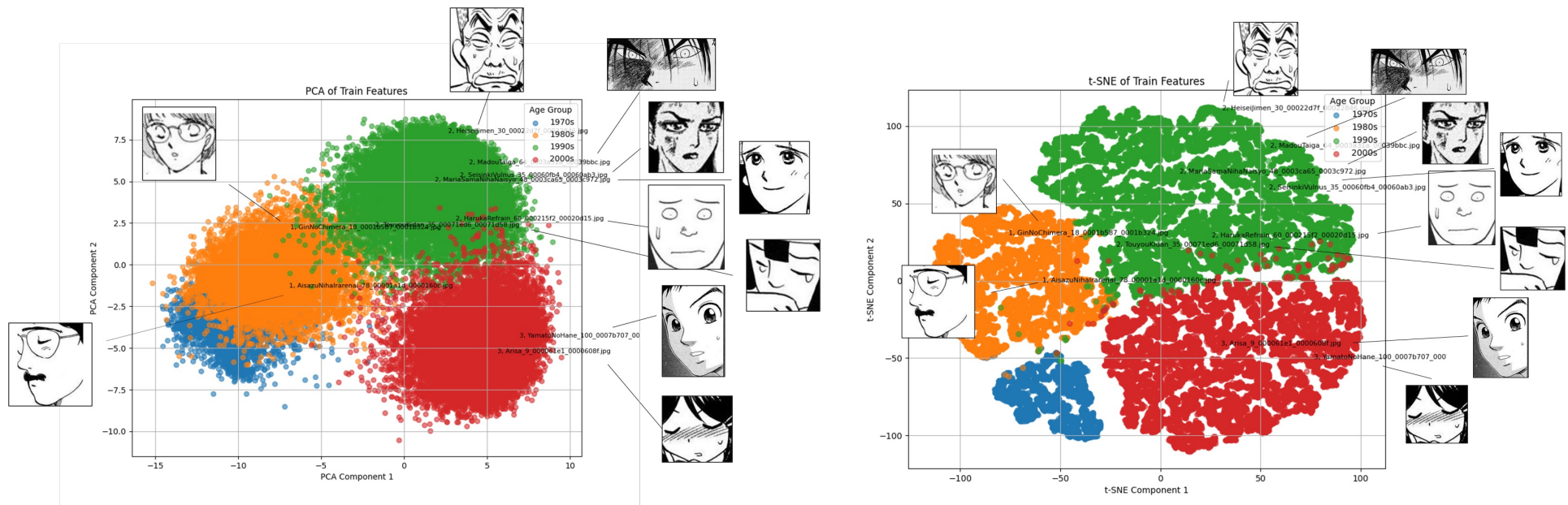
Epoch=3 程度
が適当

以下 Epoch=3
における学習結果
を利用

- Epoch=3程度で学習データで十分な性能
- Epoch =3 程度であれば過学習は起きていない (テストデータの正答率が下がっていない)

- Epoch= 3の時点でのテストデータの値がEpoch= 4よりも低い
- 学習データはEpoch= 3で十分な性能

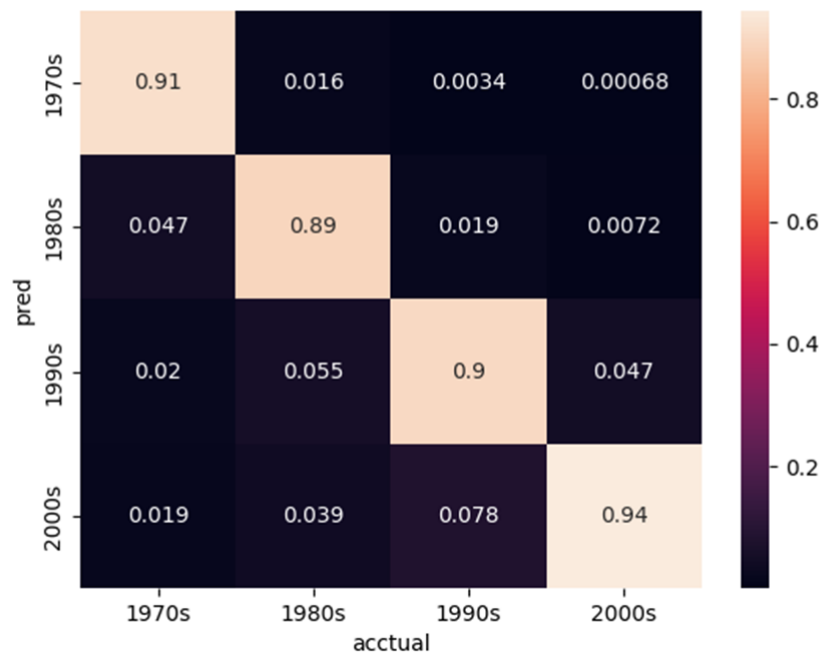
結果:特徴量のクラスタリング



- どちらも左から右へ弧を描くように年代が推移している
- 年代ごとの識別性能はt-sneが高い
- PCAは70年代と80年代に比較的類似性が現れている

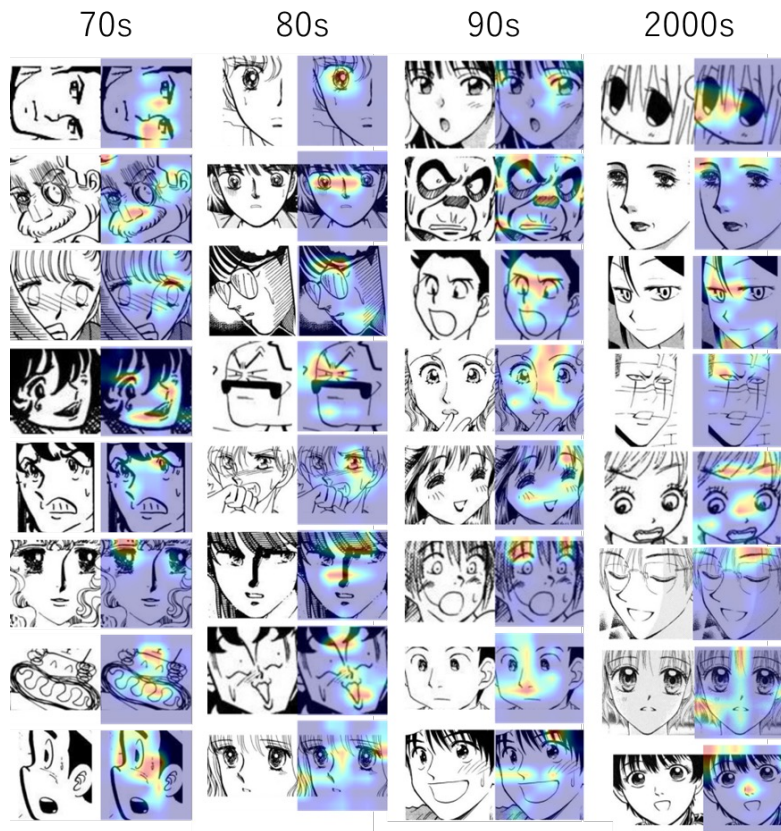
結果:予測の傾向

- テストデータにおける,予測と正解の列相対度数



- 70年代の正答率は91%
 - 80年代の正答率は89%
 - 90年代の正答率は90%
 - 00年代の正答率は94%
-
- どの年代の正答率も90%前後と良好
 - 誤答は近い年代に対して多い

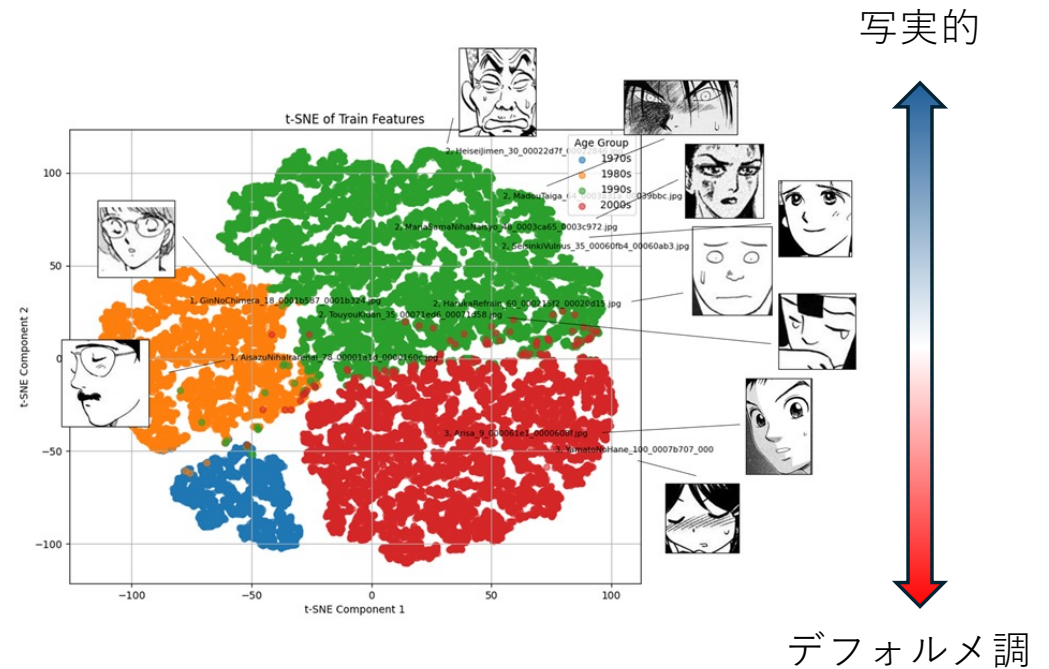
結果:特徴マップによる特徴量の可視化



- どの年代においても目や眉,眉間や鼻といった部位が注目されている
- 年代ごとの特徴の違いは上記の部位の描き方によって現れると推察できる

考察

- クラスタリングにより,年代ごとの特徴は識別可能
 - 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
 - 下側に属する70年代や90年代の画像は鼻筋などが省略されデフォルメ化
 - 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可能性
- 特徴マップでは,年代ごとの特徴が眉間や鼻などの部位に現れていた
 - 年代ごとの具体的な特徴の違いについては特定に至らなかった
 - 部位ごとに詳細に区分することで,特徴の違いを特定できる可能性



参考文献

- 成田 嶺, 小川 徹, 松井 勇介, 山崎 俊彦, 相澤 清晴 (2017): 深層特徴を用いたスケッチに基づく漫画検索, 第31回 人工知能学会
- 坪田 亘記, 相澤 清晴 (2019), 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, 58, 5, 5
- Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie (2022) :A ConvNet for the 2020s, arXiv:2201.03545
- 池田理代子: ベルサイユのばら, 1, 集英社 (1972)
- 池野恋: ときめきトゥナイト, 1. 集英社 (1982)
- 神尾葉子: 花より男子, 1, 集英社 (1992)
- 矢沢あい: NANA-ナナ-, 3, 集英社 (2001)