

深層学習を用いたイラスト顔画像からの情報抽出

三坂 蒼海[†] 犬塚 玲那[†] 岩野 公司[†]東京都市大学[†]

1. はじめに

近年、ウェブサイトを通じて様々な漫画やイラストが公開されるようになった。利用者は多数の作品を閲覧することができるが、その中から好みの作品を見つけるためには検索機能の充実が望まれる。しかし、製作者が作品を公開する際に、必ずしも検索用のメタ情報を十分に付与しているとは限らないため、自動的な情報の付与が必要である。このような観点から、イラスト画像からの作者の同定[1]や、キャラクターの分類[2]などの検討が行われている。

本研究では、100枚のイラスト顔画像で構成されるデータベースを構築し、各画像から「作者」と「キャラクターの性格」の2つのメタ情報を、畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) を用いて抽出する手法について検討する。共通のデータに対する両者の抽出性能がどの程度になるかを調査する。

2. 使用するイラスト顔画像データベース

2.1 イラスト顔画像の作成

「イラスト顔画像から検索に必要となる様々な属性情報を自動抽出する」実験に利用することを想定した画像データベースを構築する。具体的には、イラスト画像メーカーである「Picrew[3]」を用い、そのメーカーに登録されている10作者について、それぞれ10種類の男性キャラクターを作成し、合計100枚の画像データベースとする。その際、漫画等に登場しそうな10種類の性格（やんちゃ、まじめ、弱気、強気、熱血、温厚、爽やか、性悪、変わり者、甘えん坊）を設定し、10種類のキャラクターがそれぞれに当てはまるように顔画像（解像度600×600）を作成した。図1に、ある作者で作成した2キャラクターの顔画像の例を示す。左は「熱血」になるように作成したキャラクターで、右は「弱気」になるように作成したキャラクターである。背景や着衣、装飾品（眼鏡など）、表情などの影響を除外するため、これらの描き

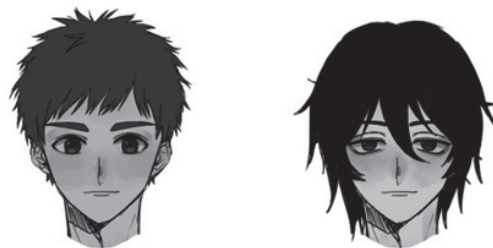


図1 作成した画像の例（左：熱血，右：弱気）

込みは行っていない。また、色の違いによる影響も除外するため、グレースケール化を行った。なお、CNNの学習時には、学習データに対して11段階で濃度変更処理を施し、11倍にデータ拡張を行う。

2.2 人間による性格の評価

作成した顔画像を複数の被験者に提示し、「そのキャラクターが10個の性格それぞれにどの程度あてはまると思うか」を5段階（1: 全くそう思わない～5: とてもそう思う）で評価してもらった。被験者の総数は36名であり、1つの顔画像あたり、5～9名で評価を行っている。例えば、図1の左の顔画像の平均評価値は「やんちゃ 4.6, まじめ 4.0, 弱気 1.1, 強気 4.0, 熱血 4.9, 温厚 4.1, 爽やか 4.4, 性悪 1.4, 変わり者 1.6, 甘えん坊 1.9」となっている。

3. 深層学習に基づくメタ情報の抽出

3.1 作者識別の方法と性能評価

作者識別用のCNNとして、VGG16[4]とResNet50[5]を使用する。初期ネットワークには、大規模データベースImageNet[6]で事前学習されたものを使用する。出力ノード数を10（作者数）に変更した上で、イラスト顔画像を用いて転移学習することで作者識別用のCNNを構築する。学習時の最適化アルゴリズムには確率的勾配降下法を用い、バッチサイズは100、エポック数は300、学習率は0.0001とした。

転移学習時に、全ての層のパラメータを最適化する方法と、全結合層～出力層のみを最適化する方法が考えられる。VGG16, ResNet50のそれ

Metadata extraction from face illustrations using deep learning
Ami Misaka[†], Rena Inuduka[†], Koji Iwano[†], [†]Tokyo City University

表 1 作者識別性能の比較

CNNの種類	最適化の対象	作者識別率
VGG16	全層	0.94
	全結合層のみ	0.92
ResNet50	全層	0.99
	全結合層のみ	0.57

それぞれについて、最適化の対象範囲を変えたときの作者識別率の比較を表 1 に示す。この実験は、キャラクターの性格に対する 10 分割の交差検証（ある 1 つの性格の顔画像を評価に、残り 9 種類の性格の顔画像を学習に利用）で実施した。結果をみると、ResNet50 の全層を対象に転移学習したときに最高の性能となり、非常に高い識別率 (99%) になることが確認できた。

3.2 キャラクターの性格推定の方法と性能評価

性格推定の CNN についても、VGG16 と ResNet50 を使用する。ネットワークは、「顔画像を入力し、それぞれの性格の（被験者による）平均評価値を推定結果として出力する」ものとして構築する（それぞれの性格ごとに、CNN を個別に構築する）。作者識別と同じく、ImageNet で事前学習済みの CNN を転移学習するが、出力層は 1 ノードに変換する。また、各全結合層の前後に過学習防止のためのドロップアウト層（ドロップアウト率 0.2）を挿入する。評価実験は、作者に対する 10 分割の交差検証（ある作者のキャラクターの性格推定を行う際に、他の 9 作者の画像で学習する）で実施する。転移学習の対象はネットワークの全層のパラメータとし、最適化アルゴリズムには確率的勾配降下法を、損失関数には平均二乗誤差を使用する。また、学習時のバッチサイズは 90、エポック数は 150、学習率は 0.0001 とした。

図 2 に、人間（被験者）が評価した際の正解（平均評価値）に対する平均二乗誤差と、機械（VGG16 と ResNet50）で推定した際の正解に対する誤差を、キャラクターの性格ごとに示す。この結果を見ると、機械の方が人間よりも小さい誤差で推定できており、特に VGG16 の方が ResNet50 よりも推定誤差が小さいことがわかる。また、人間の評価の誤差が小さい性格（例えば「熱血」）において、機械でも推定誤差が小さい傾向が読み取れる。

図 3 に 36 名の各被験者の評価の誤差と、機械の推定誤差を昇順で並べたときの結果を示す。これを見ると、機械による推定が人間の上位の評価と同等であることが確認できる。

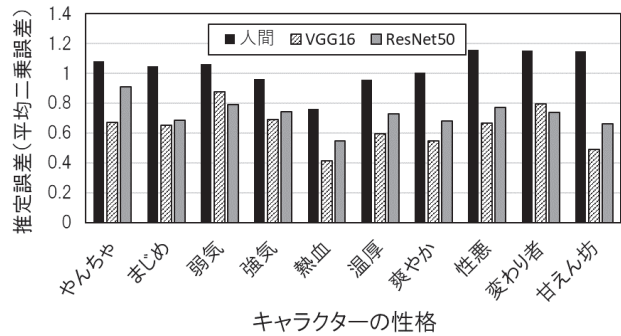
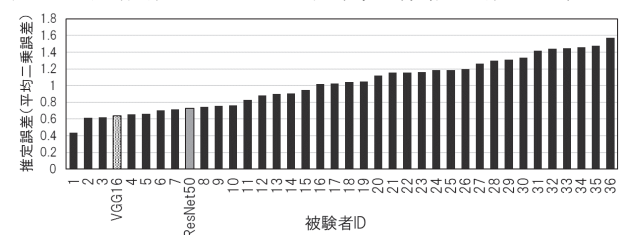


図 2 キャラクターごとの性格推定の誤差

図 3 性格推定における人間と機械の推定誤差の関係



関係

4. まとめ

本研究では、CNN (VGG16, ResNet50) を用いたイラスト顔画像からの作者識別とキャラクターの性格推定について、それぞれ性能評価を行った。実験の結果、作者識別については ResNet50 で 99% という高い性能が確認された。性格推定については VGG16 の方が ResNet50 よりも推定誤差が小さく、人間の上位の評価と同等であることが確認された。

今後の課題としては、性格評価の被験者数を増やすことによる信頼性の向上や、異なる深層学習ベースの画像認識手法を利用した際の性能検証などを行う必要がある。

参考文献

- [1] 板持ら, “イラストの作者同定アルゴリズムの提案,” 情報処理学会第 74 回全国大会, pp. 209-210, 2012.
- [2] 柳澤, 渡辺, “Deep Learning 特徴量を用いたマンガキャラクター顔画像の分類,” 第 16 回情報科学技術フォーラム, vol.3, pp.185-186, 2017.
- [3] <https://picrew.me>
- [4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Proc. International Conference on Learning Representations, 2015.
- [5] K. He, et al., “Deep Residual Learning for Image Recognition,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
- [6] <https://www.image-net.org>