

ディープラーニングを用いた マンガにおける人物の表情識別

岩崎 萌子^{1,a)} 廣井 裕一^{1,b)} 伊藤 勇太^{1,c)} 杉浦 裕太^{1,d)} 杉本 麻樹^{1,e)}

概要: マンガの画風は作者によって様々である。画風の違いに対して頑健に、マンガにおける人物の表情識別ができれば、マンガの内容の解析や描画作業支援のアプリケーションへの有用性等が期待される。本研究では、画風の違いに対して頑健な、マンガにおける人物の表情識別においてディープラーニングを用いたアプローチを提案する。また、識別精度の評価実験を行い、提案手法では従来手法よりも高精度に識別が可能であることを示した。

Facial Expression Classification of Characters in Comics Using Deep Learning

MOEKO IWASAKI^{1,a)} YUICHI HIROI^{1,b)} YUTA ITOH^{1,c)} YUTA SUGIURA^{1,d)} MAKI SUGIMOTO^{1,e)}

Abstract: Painting style of comics differs by the author. Therefore, robust classification of facial expression of comic characters in regard to the differences in the style, is expected to be used to analyse content or support drawing of comics. In this paper, we propose an approach using deep learning for facial expression classification of human characters in comics, which is robust to the style of painting. We conducted experiments to measure the recognition accuracy and the result showed that our proposed method was able to recognize more accurately compared to the conventional method.

1. はじめに

我々の日常生活においてマンガは身近なものである。マンガにおける人物の表情は、見る者に様々な感情を伝えるのに効果的である。

マンガの画風は作者によって様々であるが、画風の違いに対して頑健にマンガにおける人物の表情を解析する手法があれば、電子書籍等のデータ化されたマンガの内容の解析や、マンガに描かれた人物の表情によって効果的なエフェクトを付与する作業支援ツール等のアプリケーションへの応用が考えられる。しかしながら、従来研究では特定の作者のマンガにおける人物の検出や識別までは行われて

いるが、画風の違いに対応した人物の検出や表情までは考慮されていない。

一方で、画像認識の分野においてディープラーニングと呼ばれる、深層ニューラルネットワークを用いた機械学習が注目を集めている。従来、画像認識等のパターン認識は、認識に有効な特徴ベクトルを抽出し、それらをクラスラベルに対応付けることにより行っていた。しかし、ディープラーニングを用いることで、入力データのもつ複雑で潜在的な特徴を学習によって自動的に獲得することが可能となった。これによりディープラーニングは画像処理の問題においてはほぼ例外なく従来方法を上回る性能を示すことが報告されている。

以上を踏まえ、本研究では、任意のマンガにおける人物に対する表情識別においてディープラーニングを用いたアプローチを提案し、画風の違いに対しても頑健な性能を発揮するかを検証する。

¹ 慶應義塾大学

Keio University

a) iwasaki@imlab.ics.keio.ac.jp

b) y.hiroi@imlab.ics.keio.ac.jp

c) itoh@imlab.ics.keio.ac.jp

d) sugiura@imlab.ics.keio.ac.jp

e) sugimoto@imlab.ics.keio.ac.jp

2. 関連研究

2.1 画像特徴量によるマンガにおける人物の識別

マンガにおける人物の識別を行う研究として、石井らによる登場人物を識別する研究 [1] がある。この研究では、画像の局所領域における輝度の勾配方向の分布の組み合わせである HOG(Histograms of Oriented Gradients) 特徴量 [2] と教師あり学習を用いるパターン認識手法の 1 つである SVM(Support Vector Machine) を用いて特定の登場人物を対象とし、他の登場人物の顔画像を織り交ぜたデータセットを用いて識別を行っている。新井らはマンガ画像からの正面顔検出を行う研究 [3] を同じく HOG 特徴量を用いて行っている。しかし、この研究ではマンガの画風の違いによって検出結果は大きく左右されている。

2.2 画像特徴量による人間の表情識別

画像特徴量を用いることによって人間の表情を識別することができる。Valstar らは人間の目を検出し前処理を行うことによって、画像の局所領域における輝度の分布の組み合わせである LBP(Local Binary Pattern) 特徴量を用いて人間の表情識別を行っている [4]。このような画像特徴量を用いた表情認識は、人間の実写画像に対しては有効な手法であるといえるが、画風の違いや描かれる角度等により造形の違うマンガにおける人物に対しては適用が難しいと考えられる。

2.3 ディープラーニングを用いた画像識別

ディープラーニングを用いることで、自動的に特徴を抽出し画像識別を行うことができる。ディープラーニングによる画像識別において高い性能を示すネットワーク構造の一つとして Lin らの Network In Network(NIN)[5] がある。NIN では、一般的な畳み込みニューラルネットワーク(Convolutional Neural Network:CNN) における畳み込み層の部分に多層パーセプトロンを配置した構造をもつ。また、NIN では、通常 CNN で出力層の前に配置される全結合層の代わりに、大域平均プーリング(Global Average Pooling) 層を配置している。NIN を用いた画像認識では CIFAR-10 や CIFAR-100 等の画像データセットについて高い識別精度を示している。

2.4 ディープラーニングを用いた人物の表情識別

ディープラーニングを用いることによって人間の表情を識別することができる。Yu らは 5 つの畳み込み層・3 つのプーリング層・3 つの全結合層をもった CNN によるディープラーニングを用いて、画像ベースで顔検出および 7 表情の識別を行っている [6]。結果として、表情によってばらつきはあるが、最高で 8 割程度の精度で表情識別を可能とし

ている。

ディープラーニングを用いることで、画像特徴量を用いるよりも諸条件の変化に対して頑健に識別ができることが期待される。そこで、本研究ではディープラーニングを用いて、画風の違いに対して頑健に、マンガにおける人物の表情識別を行うことを目標とする。

3. 実装

3.1 データセットの作成

本研究においては、マンガをスキャンした画像から顔を切り出し、表情ラベルを割り振ることで学習およびテストに用いるデータセットを作成した。スキャンしたマンガには 6 つの全て作者の異なる作品(A:乙嫁語り・B:新世紀エヴァンゲリオン・C:キュレーター・D:ブラックジャック)によるしく・E:ダンジョン飯・F:ベルセルク)を用い、表情ラベルの振り分けは Ekman らの提唱する [7] 人間の基本的な 6 表情に Neutral を加えた、Angry(怒り)・Disgust(嫌悪)・Fear(恐怖)・Happy(幸福)・Neutral(中立)・Sad(悲しみ)・Surprise(驚き)の 7 表情にて行った。また、本データセット作成においては人物の横顔や眼鏡をかけている人物は除外した。作成したデータセットにおける顔画像の枚数の内訳を表 1 に示す。

表 1 作成したマンガの顔画像データセットの内訳

作品 \ 表情	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Total
A	98	188	142	632	1170	202	320	2752
B	134	120	240	96	1384	70	38	2082
C	38	244	64	100	118	14	38	616
D	146	72	176	102	574	108	128	1306
E	156	346	144	216	864	202	294	2222
F	60	52	120	110	180	44	88	654
Total	632	1022	886	1256	4290	640	906	9632

3.2 ネットワーク構造

本研究では、学習及び識別に用いるネットワークとして 2.3 で示した NIN を用いる。また、プーリングの手法として max pooling, 活性化関数として Relu[8] を用い、ネットワークの過学習を防ぐ手法である Dropout[9] の割合は 0.5 としている。

3.3 入力と出力

任意の顔画像を入力とし、ディープラーニングを用いて表情を識別し、表情ラベルを出力する。この際、学習およびテストに用いるデータセットの構成を変えて実験を行った。実験の詳細は 4 にて後述する。

4. 実験

提案手法を用いた表情識別について精度評価実験を行った。また、比較手法として HOG 特徴量と SVM を用いた手法によって同様の実験を行った。

4.1 7表情での識別

全作品のデータを学習およびテストデータに用いて、7表情での識別精度を評価した。7表情での識別に用いたデータセットの詳細を表2に示す。

表2 7表情の識別でのデータセット構成

表情ラベル	Angry/Disgust/Fear/Happy/Neutral/Sad/Surprise
学習データ数	各ラベル約400枚
テストデータ数	各ラベル約120枚

提案手法および比較手法における7表情での識別精度を表すConfusion Matrixをそれぞれ図1および図2に示す。

識別結果

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	92.86%	1.59%	3.97%	0.00%	0.79%	0.79%	0.00%
Disgust	21.74%	2.61%	16.52%	2.61%	8.70%	42.61%	5.22%
Fear	3.42%	3.42%	82.88%	0.00%	3.42%	2.74%	4.11%
Happy	8.22%	0.00%	6.85%	65.75%	6.16%	10.27%	2.74%
Neutral	5.56%	0.00%	8.33%	6.25%	53.47%	19.44%	6.94%
Sad	2.07%	0.69%	2.07%	0.00%	0.69%	93.79%	0.69%
Surprise	9.91%	0.90%	26.13%	0.00%	18.92%	18.92%	25.23%

図1 7表情での識別結果：提案手法

識別結果

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	22.22%	12.70%	34.92%	3.17%	11.90%	2.38%	12.70%
Disgust	10.32%	0.79%	3.17%	15.87%	19.05%	18.25%	32.54%
Fear	12.31%	9.23%	47.69%	6.92%	6.15%	2.31%	15.38%
Happy	5.65%	0.00%	3.23%	45.97%	5.65%	6.45%	33.06%
Neutral	6.25%	0.00%	2.34%	21.09%	35.94%	7.03%	27.34%
Sad	12.60%	8.66%	22.83%	12.60%	8.66%	14.96%	19.69%
Surprise	7.87%	10.24%	28.35%	8.66%	3.94%	4.72%	36.22%

図2 7表情での識別結果：HOG+SVM

7表情での識別では、提案手法においてAngry・Fear・Sadの3表情において80%以上の精度で識別ができた。HappyおよびNeutralにおいては50%程度の識別精度となったが、他の表情との混同が起きている割合は最高で20%程度である。一方で、DisgustおよびSurpriseにおいては識別精度よりも他の表情であると予測された割合の方が高いという結果となった。DisgustではSadと、SurpriseではFear・Neutral・Sadと主に混同が起きている。この原因としては、データセットの作成時に手動でラベルの振り分けを行った時点で表情を明確に分類することが困難であったことが考えられる。

比較手法においては、7表情における識別精度は全て50%未満であり、Surprise以外の全ての表情において提案

手法よりも低い値となっている。加えて、比較手法ではどの表情に対してもSurpriseまたはFearとの混同が起こりやすいことが分かる。

また、7表情における識別結果を踏まえ、混同の多かったDisgustおよびSurpriseを除いた5表情において同様に識別精度の評価を行った。

4.2 5表情での識別

全作品のデータを学習およびテストデータに用いて、5表情での識別精度を評価した。5表情での識別に用いたデータセットの詳細を表3に示す。

表3 5表情の識別でのデータセット構成

表情ラベル	Angry/Fear/Happy/Neutral/Sad
学習データ数	各ラベル約500枚
テストデータ数	各ラベル約150枚

提案手法および比較手法における5表情での識別精度を表すConfusion Matrixをそれぞれ図3および図4に示す。

識別結果

	Angry	Fear	Happy	Neutral	Sad
Angry	99.35%	0.00%	0.00%	0.65%	0.00%
Fear	0.63%	99.37%	0.00%	0.00%	0.00%
Happy	3.80%	5.70%	72.78%	4.43%	13.29%
Neutral	2.60%	6.49%	3.25%	61.04%	26.62%
Sad	0.00%	0.63%	0.63%	0.00%	98.73%

図3 5表情での識別結果：提案手法

識別結果

	Angry	Fear	Happy	Neutral	Sad
Angry	27.27%	36.36%	9.74%	20.13%	6.49%
Fear	8.86%	46.20%	22.15%	18.99%	3.80%
Happy	10.83%	3.18%	63.69%	10.19%	12.10%
Neutral	2.61%	1.31%	27.45%	59.48%	9.15%
Sad	10.19%	28.66%	31.21%	17.20%	12.74%

図4 5表情での識別結果：HOG+SVM

5表情での識別では、提案手法において全ての表情において識別精度が60%以上となった。また、7表情の識別結果と比べると、5表情においてはHappyおよびNeutralをはじめとして全ての表情において識別精度が向上している。これは、除いた2表情に起因する混同が排除されたためであると考えられる。

比較手法においては、5表情においても全ての表情にお

いて識別精度は提案手法よりも低い値となっており、同時に他の表情との混同も提案手法に比べて起こりやすいことが分かる。

以上を踏まえ、全作品のデータを学習およびテストデータに用いた識別において、7 表情および 5 表情のいずれにおいても提案手法の方が比較手法よりも高精度での識別が可能であるといえる。

4.3 未学習の画風に対する表情識別

4 作品のデータを学習・2 作品のデータをテストに用いて交差検証を行い、5 表情での識別精度を評価した。作品の分類は表 4 のように行った。

表 4 未学習の画風の識別でのデータセット構成

表情ラベル	Angry/Fear/Happy/Neutral/Sad
学習データに用いた作品	① ACEF ② ABDF ③ BCDE
テストデータに用いた作品	① BD ② CE ③ AF

提案手法および比較手法における 5 表情での識別精度を表す Confusion Matrix をそれぞれ図 5 および図 6 に示す。

識別結果

	Angry	Fear	Happy	Neutral	Sad
Angry	46.92	22.97%	11.10%	5.07%	13.94%
Fear	31.00%	38.47%	12.54%	5.89%	12.11%
Happy	17.55%	10.39%	38.28%	16.19%	17.59%
Neutral	11.30%	5.11%	9.06%	50.87%	23.66%
Sad	16.89%	13.27%	9.77%	25.01%	35.06%

図 5 未学習の画風の 5 表情での識別結果：提案手法

予測

	Angry	Fear	Happy	Neutral	Sad
Angry	33.50%	16.82%	6.03%	4.02%	39.62%
Fear	24.32%	21.78%	11.53%	8.15%	34.23%
Happy	20.50%	14.72%	23.66%	16.52%	24.60%
Neutral	21.47%	21.29%	14.59%	13.47%	29.17%
Sad	20.47%	18.49%	8.69%	2.97%	49.39%

図 6 未学習の画風の 5 表情での識別結果：HOG+SVM

未学習の画風の 5 表情での識別では、全作品のデータを学習およびテストデータに用いたとき識別精度の高かった 5 表情においては、提案手法において 30~50%程度の精度で識別ができた。さらに、全ての表情において正しい表情を予測した割合は、他の表情を予測した割合より高かった。しかし、特に Angry と Neutral 以外の 3 表情においては他の表情との混同が起こりやすいことが分かる。

一方、比較手法においては、予測される表情が Angry か Sad に偏っており、Sad 以外の 4 表情において正しい表情を予測した割合が、他の表情を予測した割合よりも低かった。比較手法においては学習済みの特徴量を未学習の画風のデータにおける識別の際に効果的に利用できていないことが考えられる。

未学習のデータにおける識別において、提案手法の方が比較手法よりも高精度での識別が可能であったが、その精度は学習済みの画風に対する識別よりも著しく低下した。

5. 結論と今後の課題

本研究では、ディープラーニングを用いて、学習済みの画風および未学習の画風のマンガに関して、人物の表情識別を行った。識別精度評価実験を行った結果、学習済みの画風および未学習の画風の両方において比較手法よりも高い精度で識別が可能であることを示した。未学習の画風での識別に関しては学習済みの画風に比べて識別精度が低かったが、今後はより多くの画風の作品を多くデータセットに加えて学習を行うことで精度を向上していくことが可能であると考えられる。

謝辞 本研究の一部は JSPS 科学研究費補助金 16H05870 の助成を受けたものである。

参考文献

- [1] 石井 大祐, 山崎 太一, 渡辺 裕: マンガ固有の特徴を利用したマンガ登場人物識別に関する一検討, 情報処理学会研究報告, 2013-AVM-80:NO.1, 2013.
- [2] Dalal, Navneet, Bill Triggs: *Histograms of oriented gradients for human detection.*, CVPR '05 Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.
- [3] 新井 俊宏, 松井 勇佑, 相澤 清晴: 漫画画像からの顔検出, 電子情報通信学会総合大会講演論文集 2012 年 情報・システム (2).161,2012.
- [4] Valstar MF, Mehu M, Bihan Jiang, Pantic M, Scherer K: *Meta-Analysis of the First Facial Expression Recognition Challenge.*, IEEE Trans Syst Man Cybern B Cybern, 2012.
- [5] Min Lin, Qiang Chen, Shuicheng Yan: *Network In Network*, arXiv:1312.4400, 2014.
- [6] Zhiding Yu, Cha Zhang: *Image based Static Facial Expression Recognition with Multiple Deep Network Learning*, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015.
- [7] Ekman, Paul, Wallace V. Friesen: *Constants across cultures in the face and emotion.*, Journal of personality and social psychology 17.2, 1971.
- [8] Nair, Vinod, Geoffrey E. Hinton: *Rectified linear units improve restricted boltzmann machines.*, In Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010.
- [9] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R Salakhutdinov: *Dropout: a simple way to prevent neural networks from overfitting*, The Journal of Machine Learning Research, 2014.