6M-07

アニメのシーン解析のためのキャラクター表情識別

藤波 広凪† 中島 克人‡

東京電機大学未来科学部情報メディア学科†‡

1 はじめに

現在までに多くの作品が存在するアニメ作品は，日本だけでなく海外からも注目されている．作品も増えていく中で，印象に残ったシーンを探して再度視聴したり，類似作品を探したりする事は容易ではない．しかし，アニメの各シーンがキャラクター達の表情や行動で識別されていれば，検索だけでなく要約等も容易になるであろう．そこで，このアニメシーン解析の一助とするため，まずキャラクターの表情識別を試みた．今回は，種々のアニメキャラクターの顔画像を 7 種類の表情別に収集して学習データセットとし，これにより表情識別モデルを構築して，その精度を評価した．

2 関連研究

キャラクターの表情を識別する研究の 1 つに，岩崎らによる深層学習を用いた表情識別がある[1]．彼らは，いくつかのマンガ作品から登場キャラクターの顔だけを切り出し，それを 7 表情と 5 表情で識別した学習データセットをそれぞれ用い，深層学習と HOG+SVM を比較する形で，キャラクターの表情識別精度を評価した．その結果，データセットに横顔や眼鏡をかけた顔は含まないが，画風によっては深層学習を用いて 90%を超える精度が示された．従って，マンガに比べて画風の差が比較的少なく，しかも色情報が利用可能なアニメでは，更に高い表情識別精度が期待できる．

3 実装

類似作品やシーンの検索を行うためには，動画のシーン分割や，シーン内の表情および行動の識別等の様々な工程を経て，予め各シーンの分類をしておく必要がある．今回はアニメ 4 作品を選択し，そのキャラクター達の表情識別の可能性検証を行うため，データセットの作成やそれによる学習，そして，識別精度の評価を行った．

3.1 データセットの作成

表情識別の対象として，今回は 3D モデルで作成されたようなものではなく，目が大きく口や鼻が小さく作画された人間のキャラクターの顔とした．使用する作品は，「Charlotte」，「とある科学の超電磁砲 S」，「この素晴らしい世界に祝福を！」，「さくら荘のペットな彼女」の 4 作品である．まずこれらの作中から LBP 特徴による顔検出器[2]で顔部分を切り出す．顔ではない部分の誤検出は排除した上で，それらの顔画像を表情別に手動で振り分けた．識別数は Ekman らの提唱する人間の基本的な 6 表情[3]に Natural (自然な表情) を加えた 7 表情である．なお，振り分けの際にはアニメの各シーンの内容も参考にしている．

表 1 にデータセット内の表情毎の枚数を示す．なお，検出できた顔であれば，正面の顔だけでなく眼鏡をかけたものや横顔も含めており，手動によるものは一切含めていない．

表 1 表情の内訳

表情	Angry	Disgust	Fear	Happy	Natural	Sad	Surprise
枚数	190	194	194	194	190	194	182

3.2 学習と検出・識別

学習には検出と識別を同時に行う検出器の中で比較的高精度で高速とされる YOLOv5[4]を使用した．画像は学習前にあらかじめ解像度を変換したものを使用する．データセット内の訓練画像は 896 枚，検証画像は 266 枚，テスト画像は 175 枚で，識別は前節の 7 表情で行う．

4 評価

今回はデータセットのような顔だけを切り出した画像 (以下，顔画像) とアニメからランダムに選択したフレーム (以下，シーン画像) の 2 種類に対して実験・評価を行った．評価は検出率と各表情の識別精度の 2 つとする．検出率は画像中から顔が検出できた割合であり，識別精度は，顔として検出されたものの内，どれだけ正確に表情を識別できているかを表すものである．

4.1 顔画像の検出と表情識別

まずテスト画像の各表情 25 枚，合計 175 枚に対し，顔検出と表情識別の評価を行った．その結果，顔を顔として検出できなかったものが 7 枚有ったため，検出率は 96% となった．顔検出された 168 枚に対する表情識別精度を表 2 の 7 クラス混同行列で示す．

Facial Expression Identification of Anime Characters for Scene Analysis

† Hironagi Fujinami・Tokyo Denki University

‡ Katsuto Nakajima・Tokyo Denki University

表 2 顔画像での識別精度

正解ラベル＼推論結果	Angry	Disgust	Fear	Happy	Natural	Sad	Surprise
Angry	64.0%	10.0%	14.0%	0.0%	4.0%	8.0%	0.0%
Disgust	6.0%	58.0%	4.0%	0.0%	2.0%	30.0%	0.0%
Fear	13.6%	0.0%	65.9%	6.8%	2.3%	6.8%	4.6%
Happy	12.0%	2.0%	6.0%	66.0%	4.0%	4.0%	6.0%
Natural	4.2%	0.0%	4.2%	2.1%	89.6%	0.0%	0.0%
Sad	0.0%	4.2%	22.9%	0.0%	8.3%	64.6%	0.0%
Surprise	6.5%	0.0%	6.5%	2.2%	6.5%	4.4%	73.9%

精度が最も高かったのは Natural で約 90%となっている．表情が顔に大きく表れるものとの区別がしやすいためと考えられるが，表情変化の小さい Sad や Surprise さえも Natural と推定される事が多かった．2 番目の高精度となった Surprise は，同様に大きな表情表現となり得る Angry や Fear 等に推定するミスが散見される．最も識別精度が低かったのは嫌悪や軽蔑といったネガティブな表情である Disgust で，精度は 60%に満たなかった． Disgust は事前予想した Angry との混同がそれ程多くなく，意外にも Sad との混同が 30%もあった．一方，Sad は Fear に識別される割合が 23%もあり，これら 3 表情の区別は画像認識的な課題であると同時に，アニメータにとってもこれらを描き分けるのが腕の見せ所とも言えよう．

7 表情全体での平均識別精度は 68.8%となり，シーン解析にこの 7 表情を用いるとすればやや物足らない結果となった．

4.2 シーン画像からの顔検出と表情識別

続いて，シーン画像に対する顔の検出と表情識別を行った．テストのためのシーン画像はデータセットにない作品から 50 枚選択した．各表情の数は表 3 に示す．

表 3 シーン画像内に出現する表情の内訳

表情	Angry	Disgust	Fear	Happy	Natural	Sad	Surprise
枚数	12	9	9	10	13	11	9

顔検出率と表情識別精度を示す混同行列はそれぞれ表 4 および表 5 に示す．

シーン画像に対する顔の検出率は 6 割程度となった．この原因は 2 つ考えられる．

1 つは顔の大きさのバリエーションである．アニメにおいて，登場人物の顔が小さく複数人映る場面と 1 人の顔が大きく映る場面が頻繁に入れ替わる事が少なくない．そのため，顔検出器の検出サイズの許容範囲の逸脱による検出ミスが発生する．例えば今回は，検出時の設定を 1120×1120 px で行えば小さい顔は検出できるが，大きい顔は未検出となり，416×416 px で行えばその逆となった．そのため，深層ネットワークの改良が望まれる．

もう 1 つの原因は顔の傾きである．シーン画像においては学習用の顔画像にはあまり含めていなかったような角度の上下の向き，あるいは横向きの顔の画像がしばしば現れ，検出を困難にしていた．しかし，この問題は学習データの増強によって改善可能であろう．

表 4 シーン画像での検出率

		推論結果	
		検出あり	検出なし
真値	顔である	61.64%	38.36%
	顔でない	0.00%	—

表 5 シーン画像での識別精度

正解ラベル＼推論結果	Angry	Disgust	Fear	Happy	Natural	Sad	Surprise
Angry	66.7%	0.0%	16.7%	0.0%	0.0%	0.0%	16.7%
Disgust	28.6%	71.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Fear	0.0%	20.0%	60.0%	0.0%	0.0%	20.0%	0.0%
Happy	16.7%	0.0%	0.0%	50.0%	33.3%	0.0%	0.0%
Natural	10.0%	10.0%	0.0%	0.0%	80.0%	0.0%	0.0%
Sad	0.0%	0.0%	20.0%	0.0%	0.0%	80.0%	0.0%
Surprise	0.0%	0.0%	0.0%	0.0%	33.3%	0.0%	66.7%

表情識別の精度については，顔画像と同様に Natural が高い精度を，そして，Fear が低めの精度を示した．7 表情全体での平均識別精度は 68.5%となり，顔画像と同程度であった．

なお，顔画像で高い精度であった Surprise がシーン画像ではそれ程高くなく，顔画像で低かった Disgust はシーン画像では高めの精度を示している．このような混同の様子の違いは，テスト用のシーン画像が少ない事や，学習画像とシーン画像のアニメの作風の違い等が影響したと考える．

5 まとめと今後の課題

アニメのシーン解析を目的として，深層学習によるキャラクターの表情識別を行った．その結果，7 表情の識別が平均で 70%程度可能である事が分かったが，表情ごとに 90～50%のバラツキがあり，混同しやすい表情を整理する必要がある．また，シーン画像では，顔の向きや大きさの影響による顔検出のミスや誤識別があるため，学習データの追加等が今後の課題となる．シーン解析に向けて，表情の時間遷移の認識等も更なる課題である．

参考文献

[1] 岩崎, 他,“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,” エンタテインメントコンピューティングシンポジウム論文集, 2016.
[2] “lbpcascade_animeface,” https://github.com/nagadomi/lbpcascade_animeface, 2014, 2020/8 参照.
[3] P.Ekman, et al., “Constants across cultures in the face and emotion”, Journal of personality and social psychology 17.2, 1971.
[4] G.Jocher, ”Yolov5,” https://github.com/ultralytics/yolov5, 2020, 2020/12 参照.