H-001

Deep Learning 特徴量を用いたマンガキャラクター顔画像の分類
Classification of Manga Character Facial Images Using Deep Learning Features

柳澤 秀彰† 渡辺 裕†
Hideaki Yanagisawa Hiroshi Watanabe

1. はじめに

電子コミックは電子書籍市場の売上の 8 割を占める重要なコンテンツである．現在の電子コミックの多くは単純に紙媒体のマンガをスキャンして電子化したものとなっている．このとき，画像からセリフ内容やキャラクターといったコンテンツ情報を抽出してメタデータを付与することで作品検索や要約の自動生成などの多様なサービスに利用することが可能である．しかし，手動でのマンガコンテンツの抽出作業は多大なコストがかかるために，マンガ画像から自動的にコンテンツの切り出しと認識を行うシステムの構築が必要となる．本稿では，未知のマンガ画像を対象としたキャラクター顔画像の自動分類を目的として，X-means 法および Deep Learning 特徴量を用いた顔画像のクラスタリング手法について検討を行った．

2. 関連研究

対象のマンガ作品に関する事前知識を持たない状態において，入力画像から主要キャラクターと思しき人物を同定する手法について，長尾らはキャラクターの顔領域より計算した SURF 特徴量を Bag-of-Visual-Words (BoVW)に変換し，K-means 法によってクラスタリングする手法を提案した[1]．K-means 法は事前に分割するクラスタの数を指定する必要があるが，この手法では多数のクラスタに分類したとき主要キャラクターはデータ数の大きいクラスタに含まれると報告している．我々はクラスタ数を自動決定するクラスタリング手法である X-means 法のキャラクター分類への適用について検討を行った[2]．その結果，k-means 法のクラスタ数を暫定的に固定してクラスタリングを行った場合より主要キャラクターの抽出精度が向上することを確認した．一方，Deep Learning によって生成される特徴量を用いた画像クラスタリング手法として，菊田らは Faster R-CNN の中間層から抽出される特徴量マップを用いた新築戸建外観画像のクラスタリング手法を提案した[3]．

マンガ画像が写真のような自然画像と異なる点として，2 値の線画で構成されることから認識に利用できる特徴が少ないことが挙げられる．本稿では，Deep Learning の画像特徴量の利用がマンガキャラクター顔画像の分類に対しても有効であるかどうかについて検討を行う．

3. クラスタリング手法

本稿ではマンガキャラクター顔画像のクラスタリング手法として，Faster R-CNN の畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)の中間層から抽出される特徴量マップを X-means に入力する手法を提案する．

3.1.1 Faster R-CNN

CNN によって生成される画像特徴量を利用した物体検出手法として，Girshick らは R-CNN(Regions with CNN features)を提案した[4]．R-CNN の物体検出の流れは以下のようになる．1) はじめに入力画像に対して Selective Search による物体の候補領域抽出を行う．2) 次に抽出された各候補領域を CNN に入力して画像特徴量を計算する．3) 最後に出力された特徴量に対して SVM によるクラス分類を行い，対象物体と判定された候補領域を表示する．R-CNN の問題点として，Selective Search と CNN の処理が分かれているため計算時間がかかる点が挙げられる．Ren らは候補領域の抽出と CNN の計算を一つの処理で行うよう改良した手法として Faster R-CNN を提案した[5]．Faster R-CNN の物体検出は以下のようになる．1) はじめに画像を CNN に入力し，CNN の中間層において画像全体の特徴量マップを作成する．2) 特徴量マップを CNN 内の次層にある Regions Proposal Network (RPN)に入力して物体候補の推定を行い，候補領域を特徴量マップに射影する．3) 得られた各候補領域に対してクラス分類を行い，対象物体を検出する．

本研究で使用する画像特徴量は，RPN に投入する直前の conv5_3 層の出力を使用する．出力される特徴量マップの次元は VGG16 のモデルでは$(512, h_{out}, w_{out})$となる．ここで，512 はフィルタの数，h_{out}とw_{out}は特徴量マップの高さと幅の次元で$h_{out} = ceil(h_{init}/16)$となり$w_{out}$も同様である．このとき$h_{init}, w_{init}$は入力画像の次元，$ceil$は天井関数を示している．

3.1.2 X-means 法

X-means 法は K-means 法の逐次繰り返しとベイズ情報量基準(BIC)による分割停止基準を用いて最適なクラスタ数を決定する手法である[6]．BIC の値は以下の式(1)より求められる．

$$BIC = -2\ln L + k \cdot \ln(n) \quad (1)$$

式(1)において，L はモデルにおける尤度関数の最大値，k はモデルのパラメータの個数，n は標本のサイズを表す．また第 1 項はモデルへの当てはまりの良さを，第 2 項はモデルの複雑さに対するペナルティを示す．X-means 法によるクラスタリング手順は次のようになる．1) はじめに(k = 2)として k-means の 2 クラス分類を行い，入力データを 2 つのクラスタに分割する．2) 生成された各クラスタに対して 2-means を行い，分割後のクラスタの BIC が分割前よりも小さいとき分割は適切であるとして採用する．3) 新たに生成されたクラスタに対して 2)の処理を繰り返し，全てのクラスタについて分割後の BIC が分割前より大きくなる状態となったとき，適切なクラスタ数に分割が行われたと見なす．

文献[2]において，100 次元以上の BoVW を x-means に入力したとき，式(1)の第 2 項の値が第 1 項よりも過大に計算されるために分割されるクラスタの数が想定される結果よりも少なくなることが確認できた．このため本稿では第 2 項の値を小さくすることで，適切なクラスタ個数が得られるように調整を行う．

4. 実験

本実験では，研究用マンガ画像データベースの Manga109 [7]に公開されているマンガ画像より，作者の異なるマンガ 3 作品 (BEMADER・P，ぶらり鉄扇捕物帳，爆裂！かんふー娘) をテストに使用する．各作品について横顔を含めたキャラクター顔画像 300 枚ずつを抽出し，画像サイズを$(width, height) = (200, 200)$に正規化したデータセットを対象として X-means 法によるクラスタリングを行った．使用する特徴量は 1) SURF 特徴量，2) 一般物体検出用の CNN モデルである VGG16 [8]を用いた特徴量マップ，3) VGG16 に対してマンガ画像 2000 枚を用いてファインチューニングを行った CNN の特徴量マップの 3 種類を使用した．このとき，SURF 特徴量では 100 次元に変換した BoVW を X-means への入力として，X-means の分割停止基準を以下のように設定した．

$$BIC = -2\ln L + 0.3 \cdot k \cdot \ln(n) \qquad (2)$$

また，2 種類の CNN 特徴量マップは特異値分解によって 100 次元に次元圧縮したデータを X-means の入力とした．

実験結果について，各クラスタ内に最も多く含まれるキャラクターの画像数がクラスタ全体に占める割合(純度)を求めた．マンガ 3 作品に対するクラスタリングにおいて生成されたクラスタの個数を表 1，クラスタの平均純度を表 2 に示す．表 1，2 について VGG16(ft)はファインチューニングされた VGG16 モデルによる結果を示す．表 2 より CNN 特徴量を使用したときのクラスタの純度は，SURF 特徴量を用いた場合よりも低下することが確認できた．VGG16 モデルを使用したときのクラスタリング結果の例を図 1 に示す．SURF 特徴量を使用した場合には，主に線描のパターンや塗りつぶしなどのテクスチャの種類によって画像の分類が行われる傾向が見られた．一方で CNN 特徴量を使用した場合には，画像は顔の向きや画像に占める大きさといった見え方の違いによって分類される傾向が確認できた．このことから，CNN 特徴量による画像認識ではキャラクター間の顔特徴の変化よりも顔の見え方の違いが重視されるため，キャラクター分類の精度が SURF 特徴量よりも低くなる結果が得られたと推測する．また，CNN モデルのファインチューニングを行ったときの純度は，以前の CNN モデルのものより低下した．この理由について，使用した画像枚数が少ないためマンガ画像の特徴を十分に学習できていないことや，異なるマンガ作品の画風について学習したことが対象となる顔画像の認識に悪影響を及ぼしていることなどが考えられる．

表 1　X-means クラスタリングで生成されたクラスタ個数

	SURF	VGG16	VGG16 (ft)
BEMADER・P	12	9	9
ぶらり鉄扇捕物帳	9	8	7
爆裂！かんふー娘	8	8	6

表 2　X-means クラスタリングの平均純度

	SURF	VGG16	VGG16 (ft)
BEMADER・P	0.588	0.531	0.355
ぶらり鉄扇捕物帳	0.629	0.491	0.387
爆裂！かんふー娘	0.714	0.671	0.485

図 1　VGG16 を使用したクラスタリング結果の例

5. おわりに

本稿では Deep Learning 特徴量を用いた X-means クラスタリングによるマンガキャラクター顔画像の分類について検討を行った．実験結果より，従来の CNN モデルを使用した場合では，キャラクターの分類について SURF 特徴量よりも精度が低下することが確認できた．今後の改善点として，キャラクター間の特徴を捉えるように強化学習を行う CNN モデルを特徴抽出に使用することや，顔領域の形状やフキダシといった他のコンテンツ情報を利用したキャラクター分類手法を使用することで分類性能を向上できると考えられる．

†早稲田大学大学院基幹理工学研究科情報通信専攻，
Graduate School of Fundamental Science and Engineering,
Waseda University

謝辞

本研究は JSPS 科研費 17K00511 の助成を受けたものである．

参考文献

[1] 長尾, 渡辺, “コミックにおける主要キャラクター同定の検討”, 電子情報通信学会総合大会, D-21-3, (2016).
[2] 柳澤, 渡辺, “X-means 法を用いたマンガキャラクターの自動分類に関する検討”, 電子情報通信学会総合大会, D-12-40, (2017).
[3] 菊田, 野村, 李, 小林, 神津, “Deep Learning 技術をベースとした異常画像検出”, 第 30 回人工知能学会全国大会, 1A4-OS-27b-1, (2016).
[4] R. Girshick, J. Donahue, T. Darrell, J. Malik, “Rich feature hierarches for accurate object detection and semantic segmentation,” in IEEE Conference on Computer Vision and Pattern Recognition, (2014).
[5] S. Ren, K. He, R. Girshick, J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Advances in Neural Information Processing Systems (NIPS), (2015).
[6] D. Pelleg, A. Moore, “X–means: Extending K-means with efficient estimation of the number of clusters” In Proceedings of the 17th International Conf. on Machine Learning, pp.727-734, (2000).
[7] Y.Matsui, K.Ito, Y. Aramaki, T.Yamasaki, K. Aizawa: “Sketch-based Manga Retrieval using Manga109 Dataset”, arXiv:1510.04389 , (2015).
[8] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, In *arXiv preprint arXiv:1409.1556*, (2014).