

深層特徴を用いたスケッチに基づく漫画検索

Sketch-based Manga retrieval using deep features

成田 嶺^{*1}

Rei Narita

小川 徹^{*1}

Toru Ogawa

松井 勇祐^{*2}

Yusuke Matsui

山崎 俊彦^{*1}

Toshihiko Yamasaki

相澤 清晴^{*1}

Kiyoharu Aizawa

^{*1}東京大学

The University of Tokyo

^{*2}国立情報学研究所

National Institute of Informatics

We present a method to extract feature vectors from sketches and Manga images by combining two CNNs for sketch-based Manga retrieval. One is trained on a large amount of Manga face images and the other is trained on without screentones. We showed that fine-tuning AlexNet pre-trained on ImageNet significantly improves accuracy of Manga-to-Manga and sketch-to-Manga character retrieval.

1. はじめに

漫画は日本を代表するコンテンツ産業であり、国内の市場規模は2015年で4,437億円にもなる。スマートフォン・タブレットの普及により、漫画文化にも電子化の波が押し寄せるなか、大量の漫画のデータから目当ての情報を効率よく検索するシステムというのは、欠かせないものである。しかし、現在の検索は著者名、ジャンルなどのキーワードによるものに限られている。そこで、画像を用いて漫画を検索することができれば、キーワードのみを用いた検索では出来ないような細かいページの中身の検索を行うことが出来る。

このような画像を用いた検索としては、松井ら[1]によって、スケッチを用いた手法が提案されている。スケッチはスマートフォン・タブレットと相性がよく、手元にクエリとなるような画像がない場合でも直感的な検索を行うことができる。

松井ら[1]の手法では認識や検索にエッジに注目した特徴量を利用していたが、エッジを用いた手法ではうまく検索できない場合がある。例えば、見た目は似ていない画像でも、特徴量が似ている場合があり、ノイズのような画像が検索されてしまう。また、同じキャラクターでも書き方や角度が異なると同じキャラクターであると認識できないといった問題や違うキャラクターを細かく表現できないといった問題がある。

本論文では、Convolutional Neural Network (CNN) により抽出した深層特徴を利用することを提案する、これにより、スケッチによる漫画検索、特に顔検索において、上記の問題を解決し、検索の精度を向上させることを目標とする。CNNは近年、画像認識や検索分野で目覚ましい成果を残している。一方で、学習には大規模データセットとメタデータが必要である。漫画画像を考えると、著作権の問題から、大規模データセット作成が難しく、CNN技術を適用できなかった。学術利用可能な大規模漫画データセット、Manga109 [1] [2] を構築した。これにより、CNN技術を漫画にも適用することが可能になった。本論文では、この大規模漫画データセットを用いることで検索のパフォーマンスが大幅に向上することを示す。

2. 関連研究

2.1 漫画研究とデータセット

漫画に関する研究は、電子コミックの興隆を背景に多岐にわたって行われていた。しかし、研究で漫画の画像を利用する際は作者に許諾を得る必要があり、そのため、大量の漫画のデータの入手、利用が困難であるという問題があった。そこで、我々はManga109 [2] というデータセットを構築した [1]。このデータセットは様々な年代、作者、ジャンルからなる計109冊、計21,142ページの漫画作品から成っている。また、このデータセット内の作品は学術利用の範囲で作者の利用許諾を受けている。さらに、藤本ら[3]によりこのManga109に、キャラクター・テキスト・コマ割りのメタデータが追加された。これによりラベルが付いた漫画画像領域の情報が大量に得られるようになり、それを用いて機械学習を行うことが可能となった。

2.2 スケッチに基づく漫画検索

画像の検索を行うには、クエリとデータセットの類似度を評価するために画像を特徴量のベクトルで表す必要がある。松井らは局所領域の輝度勾配を量子化し作ったヒストグラムであるHOG (Histogram of Oriented Gradients) を特徴量とした。漫画のスクリーントーンを除去した [4] のちに特徴量を求めることで、ノイズを抑制し輪郭を強調した。また、スケッチによる検索結果を新たにクエリとして検索するRelevance feedbackと呼ばれるシステムを提案している。

2.3 Deep learning と画像処理

Deep learning とは多層のニューラルネットワークを用いて分類、回帰等の問題を解く手法である。

CNNは中間層に入力画像の特徴量が自動的に現れるため、予め学習が行われたCNNを特徴抽出器として用いることが可能である。大規模画像データセットであるImageNetで事前に学習されたCNNは汎用性のある特徴抽出器として用いられる。また、CNNの学習をさせる際も、ImageNetなどの大規模画像データセットで学習済みのCNNのパラメータを初期値とすることで、一から学習させるよりも精度が向上することが知られている [5]。

連絡先: 東京大学 情報理工学系研究科

〒113-8656 東京都文京区本郷 7-3-1

TEL.03-5841-6761

E-mail: narita@hal.t.u-tokyo.ac.jp

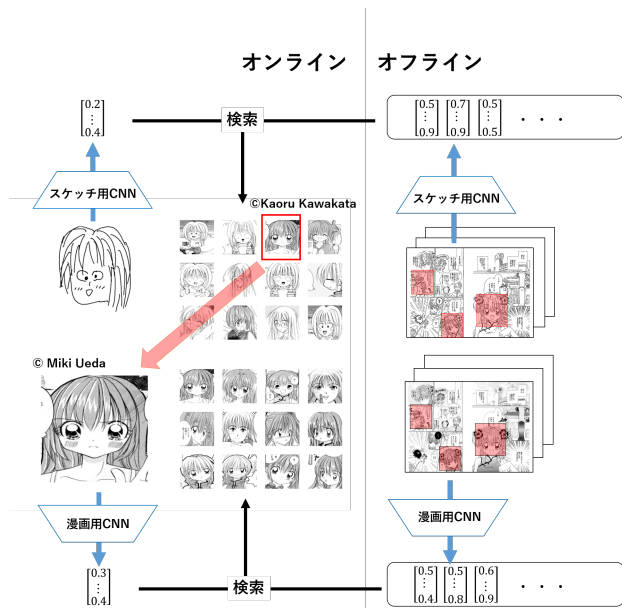


Fig. 1: 検索システムの概略図.

3. 提案手法

3.1 概要

スケッチによる検索を行い、検索結果をさらにクエリとして利用するという検索システムにおいて、クエリに応じて2種類のCNNを特徴抽出器として使い分ける手法を提案する。概略図をFig. 1に示す。提案手法では、クエリとして「スケッチ (ユーザがシステム上で描いた絵)」を使う場合 (Fig. 1の上部)、および「漫画画像そのもの」を使う場合 (Fig. 1の下部) がある。ユーザはまず自分が描いたスケッチを用いて検索を行う。結果を見て、もし自分の意図に合致する検索結果があれば、それをクエリとして再利用することができる (Fig. 1の中心部の赤い矢印)。ここで、クエリがスケッチの場合、スクリーントーンを除去した漫画画像で学習したCNN (以下、スケッチ用CNN) を用いて特徴抽出を行う (スケッチ-漫画検索)。これは、スクリーントーンを除去後の漫画画像がスケッチをよく近似するという考えに基づく。一方、クエリが漫画画像の場合は、スクリーントーンを除去しない漫画画像で学習したCNN (以下、漫画用CNN) で特徴抽出を行う (漫画-漫画検索)。これは、漫画画像に含まれるスクリーントーンの情報を活かすためである。

検索用データベースとして、まず初めに検索対象のページから検索候補領域を Selective Search [6] により求める。次に全ての矩形領域に対し、以下の二通りの方法で特徴ベクトルを抽出する。

- (1) スクリーントーンを除去しスケッチ用CNNで特徴抽出。
- (2) スクリーントーンを除去せず、漫画用CNNで特徴抽出。

クエリとしてスケッチが入力された場合、スケッチ用CNNで特徴ベクトルを抽出し、(1)で得られたベクトル全てとのコサイン類似度を計算し、その大きさにランキングを得る。Relevance feedbackにより、クエリに漫画画像が入力された場合、漫画用CNNで特徴ベクトルを抽出し、(2)で得られたベクトル全てとのコサイン類似度を計算し、その大きさにランキングを得る。

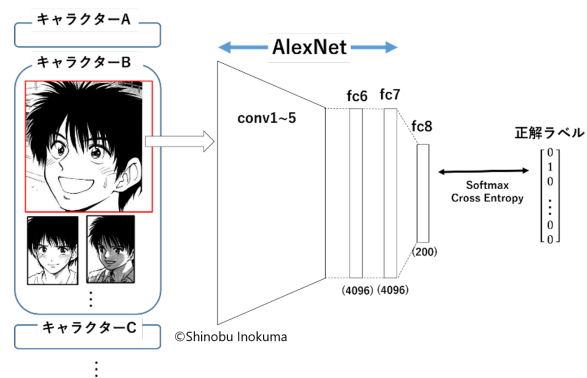


Fig. 2: 学習用のネットワーク

学習に用いる顔画像のデータセットは前述の Manga109 を利用する。メタデータにおいて顔領域の定義は、上下は眉の上から顎の下、左右は耳までとなっているが、キャラクターの構成要素として髪の毛の占める割合は大きいと考えられる。そこで、上下左右ともに一定割合で拡大した後、スクリーントーン除去前、除去後ともに切り出し、顔画像データセットを作成した。

3.2 実装

オープンソースの Deep learning のフレームワークである Caffe [7] を用いて CNN を実装した。CNN の構造としては、ベースラインである AlexNet [8] を使用する。顔画像データセットの中から、70枚以上あるキャラクターを計70冊の漫画から200人選び、キャラクターをクラスとして多クラス分類問題として学習を行う (Fig. 2)。それぞれのキャラクターごとに7割を学習用、3割をテスト用画像とする。合計すると学習用が23,422枚、テスト用が9,156枚となる。ImageNet事前学習済みモデルを初期値として fine-tuning を行う。学習が収束した時点でのモデルを特徴抽出器として利用し、fc6層の出力を特徴ベクトルとする。学習が収束した時点の Test Accuracy は、スクリーントーンを除去した画像の場合約77%、スクリーントーンを除去しない画像の場合は約84%であった。

4. 漫画画像をクエリとした検索実験

4.1 目的

漫画画像をクエリとした検索において、提案手法の特徴抽出器が、学習に用いていない未知の漫画に対して有効かどうか、また、既存手法である HOG への優位性の確認のため、顔画像のみのデータセットを用いて検索の実験を行った。

4.2 方法

顔画像データセットのうち、CNNの学習に用いていない漫画7冊に登場する顔画像5379枚をデータセットとした。6人のキャラクターを選び、顔画像のうち各50枚をクエリとし、クエリと同じキャラクターの顔画像を ground truth とした (Table 1)。検索は、クエリの特徴ベクトルとデータセットの特徴ベクトル間のコサイン類似度を計算し、その値が小さい順に結果のランキングを得ることにより行った。そして各クエリごとに Average Precision (AP) を計算し、各キャラクターごとに平均し、mean Average Precision (mAP) を最終的に算出した mAP は検索の評価に良く用いられる指標である。提案手法としては、スクリーントーンを除去せず漫画用CNNに入力し、fc6層の出力4,096次元を特徴ベクトルとして抽出した。比較する既存手法としては、HOGを利用し、スクリー

character	Manga	#GT	#query
芹	BakuretsuKungFuGirl	140	50
ベルモンド	Belmondo	179	50
ロラン	Belmondo	119	50
中里圭太	GOOD_KISS_Ver2	125	50
浦島景太郎	LoveHina_vol14	240	50
鳥羽撫子	YamatoNoHane	100	50

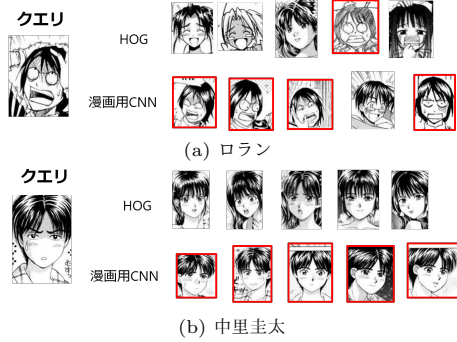


Fig. 3: 検索結果の例. 赤い枠は検索が成功した例である.

トーンをの除去後 4 方向*8*8 の 256 次元を特徴ベクトルとして抽出した.

4.3 結果

各キャラクターの mAP を Table 2 に示す. このように全てのキャラクターにおいて mAP が飛躍的に向上しており, 提案手法の漫画用 CNN は学習に用いていない漫画に対しても有効な特徴量を抽出することが可能であることが分かった. 検索の例を Fig. 3 に示す. このように, HOG ではキャラクターごとの違いを表現出来ていないのに対し, 提案手法ではクエリと同一のキャラクターが上位に検索されていることが分かる.

5. スケッチによる検索実験

5.1 目的

実際の漫画検索を想定し, 被験者に集まってもらいスケッチを書いてもらう. 集まったスケッチをクエリとして漫画を検索し, 提案手法であるスケッチ用 CNN と既存手法である HOG のスコアを比較する.

5.2 方法

クエリは, 被験者 7 名から集めた 3 キャラクターのスケッチを用いる. 被験者に見せた画像は Fig. 4 の通りである. 特徴があるキャラクターを選んだため, スケッチ難易度は比較的低いと考えられる. また, 実際に集まったスケッチの例は Fig. 5 のようになっている.

キャラクター	HOG	漫画用 CNN (提案手法)
芹	0.04	0.24
ベルモンド	0.04	0.52
ロラン	0.04	0.47
中里圭太	0.04	0.25
浦島景太郎	0.07	0.47
鳥羽撫子	0.04	0.34



Fig. 4: 被験者に見せた画像

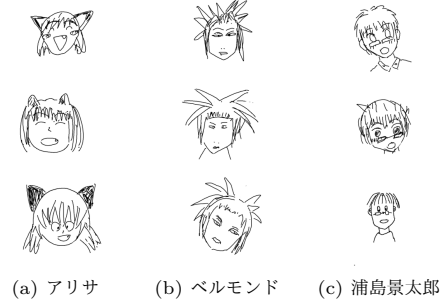


Fig. 5: 実際に書いてもらったスケッチ.

1. あるキャラクターの漫画の画像を 30 秒間見せる
2. 画像を隠し, 1 分 30 秒でキャラクターのスケッチをしてもらう. アプリケーションはペイントを使用し 512*512 の白いキャンバスを用意する.
3. 2,3 を繰り返す.

CNN の学習に用いていない漫画 7 冊に対し, 見開き 1 ページにつき 200 個, 計 137,800 個の窓候補を Selective Search を用いて選出しデータセットとする. これらの窓候補のうち, データベース中の人手でつけた顔領域との IoU (領域の被り率) が 0.5 以上あるものを ground truth とする (Table 3). 検索は, クエリの特徴ベクトルとデータセットの特徴ベクトル間のコサイン類似度を計算し, その値が小さい順に結果のランキングを得ることにより行った. そして各クエリごとに, Top100 までの AP を計算し, 各キャラクターごとに平均し, mAP@100 を求めた. 提案手法としては, スクリーントーンを除去した後スケッチ用 CNN に入力し, fc6 層の出力 4,096 次元を特徴ベクトルとして抽出した. 比較する既存手法としては, HOG を利用し, スクリーントーンを除去した後 4 方向*8*8 の 256 次元を特徴ベクトルとして抽出した.

5.3 結果

各キャラクターの mAP は Table 4, 検索結果上位 12 件の例は Fig. 6 のようになった. 各キャラクターごとに結果に差が出ていることが分かる.

アリサとベルモンドでは HOG と比較して提案手法では高い精度が出ていることが分かる. 特に, ベルモンドでは HOG では顔すらも認識できていないが, 提案手法ではしっかりと正解のキャラクターが検索出来ている.

一方, 浦島景太郎の場合, 提案手法では正解のキャラクターを検索出来ていない. 浦島景太郎を書いた他の検索結果を Fig.

character	Manga	#GT	#query
アリサ	Arisa	258	$7 \times 2 \times 2$
ベルモンド	Belmondo	134	$7 \times 2 \times 2$
浦島景太郎	LoveHina_vol14	158	$7 \times 2 \times 2$

キャラクター	HOG	スケッチ用 CNN (提案手法)
アリサ	0.09	0.26
ベルモンド	0.00	0.18
浦島景太郎	0.07	0.05

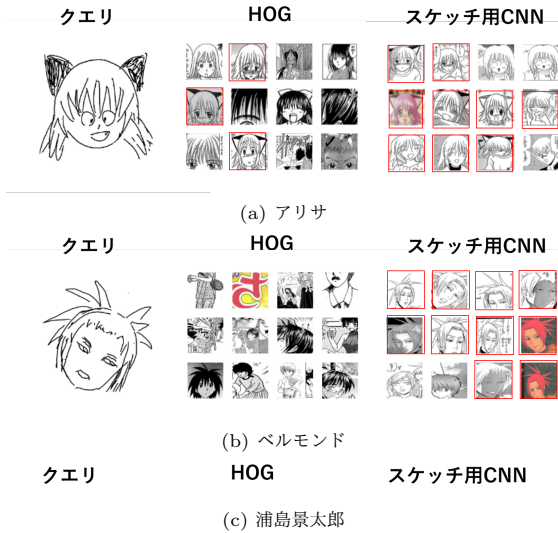


Fig. 6: キャラクターごとの検索結果. 赤い枠は検索が成功した例である.

7に示す. これを見ると, 全体的な髪の色や, 書き込み方が検索の際に大きなウェイトを占めていることが分かる.

浦島景太郎のスクリーントーンを除去した画像は Fig. 8 のようになっている. スクリーントーンがうまく除去できておらず, 線画から離れてしまっている. 提案手法では, スクリーントーンを除去したものでスケッチを近似できると考えているため, このような髪型のキャラクターを上手く検索することができないと考えられる.

6. 結論

本研究では, 漫画の直感的な検索方法としてスケッチによる検索を考え, 大量の画像を用いた CNN の学習を行った. スクリーンを除去した画像, 除去しない画像の 2 種類を用いて学習した 2 つの特徴抽出器による, スケッチによる新しい漫画検索システムの提案を行った. 漫画画像によるキャラクター検索の実験を行い, 既存手法に対し提案手法では平均で mAP が 0.34 向上することを示した. また, スケッチを用いたキャラクター検索の実験を行い, クエリのキャラクターによって既存手法と比べ mAP が 0.17 以上向上することを示した.

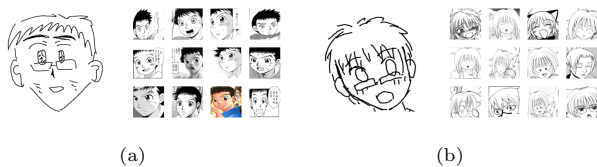


Fig. 7: 他のスケッチによる検索結果 (浦島景太郎)



Fig. 8: 浦島景太郎のスクリーントーン除去の様子. 黒いベタ塗りが残っており線画とはいえない.

謝辞

本研究の一部は戦略的情報通信研究開発推進事業 SCOPE の委託を受けた

参考文献

- [1] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, pp. 1–28, 2016. Springer.
- [2] Manga109. <http://www.manga109.org/>.
- [3] Azuma Fujimoto, Toru Ogawa, Kazuyoshi Yamamoto, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. Manga109 dataset and creation of metadata. In *MANPU*, pp. 2:1–2:5, 2016.
- [4] Kota Ito, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. Separation of manga line drawings and screentones. In *Eurographics (Short Papers)*, pp. 73–76, 2015.
- [5] Pulkit Agrawal, Ross Girshick, and Jitendra Malik. Analyzing the performance of multilayer neural networks for object recognition. In *ECCV*, pp. 329–344, 2014.
- [6] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *IJCV*, Vol. 104, No. 2, pp. 154–171, 2013.
- [7] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACMMM*, pp. 675–678, 2014.
- [8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, pp. 1097–1105, 2012.