

絵柄を考慮したキャラクタイラスト検索における背景除去による精度の向上

小沢伸介 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

近年、pixiv[1] や Twitter[2] などの SNS にキャラクタイラストを投稿する人も多く、それらを目にする機会も増加している。しかし、多くのキャラクタイラストの中から、自分の好みの絵柄のイラストを探し出そうとすると、絵柄を表す適切な名称が分からないことが多く、望んだ通りのイラストを検索することは難しい。それに対しては、キャラクタイラストをキーとし、それに類似した絵柄のキャラクタイラストを検索できるようなシステムが有効であると考えられる。

キャラクタイラストの検索を行う手法として、色の出現頻度に着目した画風を考慮したキャラクタイラスト検索 [3] が提案されている。この手法では、画風の特徴が彩度と明度の傾向、目の領域の形状と配色、顔の領域の配色などに表れると考えて検索を実現している。入力データとしてイラスト全体の特徴をとらえる彩度と明度のヒストグラム、YOLO (You Only Look Once)[4] によって抽出した目の領域の RGB 画像、YOLO によって抽出した顔領域の色相のヒストグラムの 3 つを利用している。それらを入力データとして、画風に基づいて分類したどのグループに属するかを出力するように畳み込みニューラルネットワーク [5] の学習を行う。学習を行った畳み込みニューラルネットワークの出力層の一つ前の層の出力を特徴ベクトルとして利用し、検索キー画像の特徴ベクトルと類似した特徴ベクトルを持つイラストを出力することで検索を行う。しかしながら、顔の領域の色相は同じグループでもキャラクタによって異なる可能性がある。また、検索の際に背景も考慮されてしまっているが、背景自体はキャラクタの絵柄とは関係がない。

そこで、本研究では絵柄を考慮したキャラクタイラスト検索における背景除去による精度の向上を提案する。提案システムでは、入力画像に対して背景除去を行い、彩度と明度のヒストグラム、顔領域の彩度と明度のヒストグラム、目領域の RGB を特徴ベクトルの生成に用いる畳み込みニューラルネットワークへの入

力に用いることで検索精度の向上を目指す。

2 remove.bg

remove.bg[6] は、ブラウザ上で画像をアップロードすることで入力画像の背景部分を検出して除去することができるサービスである。背景除去を行う詳しい仕組みは公開されていないが、次のように処理されていると考えられている [7]。

- (1) Semantic Segmentation を用いて画像を前景、背景、そのどちらかの三つに粗く分割した trimap を作成する。
- (2) 元画像と trimap を入力して Image Matting を用いてマスクを取得する。このマスクには前景か背景のどちらかに分類された画像情報が保存されている。
- (3) 元画像とマスクをかけ合わせて背景を除去する。

提案システムでは、remove.bg を用いて入力データの背景を除去した画像を入力として用いる。

3 You Only Look Once (YOLO)

YOLO[4] は畳み込みニューラルネットワークを用いた物体の検出と識別を同時に行うことのできる物体検出アルゴリズムである。物体の検出と識別を同時に行うことで高速な処理を実現している。YOLO では、まず入力画像をグリッド状の領域に分割し、分割された領域がどの物体のカテゴリに含まれているかを識別する。そして物体と予測されるものを領域で囲み、認識結果として出力する。提案システムでは、YOLO を用いてキャラクタイラストから顔領域の抽出を行う。

4 絵柄を考慮したキャラクタイラスト検索における背景除去による精度の向上

ここでは、提案する絵柄を考慮したキャラクタイラスト検索における背景除去による精度の向上について説明する。提案システムでは絵柄に基づくグループに

Accuracy Improvement by Removing Background in Character Illustration Search considering Patterns
Nobuyuki Ozawa and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

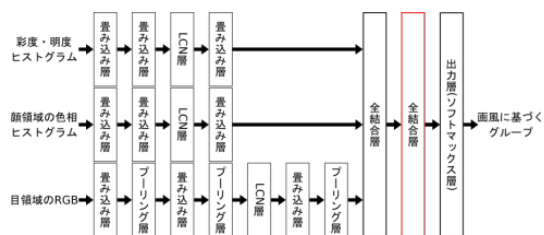


図1: 提案手法で用いる畳み込みニューラルネットワークの構造

分類するように学習を行った畳み込みニューラルネットワークを用いて検索を行う。畳み込みニューラルネットワークの入力には、彩度と明度のヒストグラム、顔領域の彩度と明度のヒストグラム、目領域のRGBを用いる。入力画像は remove.bg を用いて背景を除去しておく。顔領域は入力画像から YOLO を用いて抽出する。これらを入力とし、絵柄に応じたクラスに分類されるように学習を行う。検索の際には、畳み込みニューラルネットワークの出力層の一つ前の層の出力を特徴ベクトルとして用いる。検索キーの特徴ベクトルと類似した特徴ベクトルを持つ画像が検索結果として出力される。

4.1 構造

提案システムでは、図1のような構造の畳み込みニューラルネットワークを用いる。入力画像は事前に remove.bg を用いて背景を除去しておく。入力には、彩度と明度のヒストグラム、顔領域の彩度と明度のヒストグラム、目領域のRGBを用い、全結合層の手前まで、それぞれ独立した構造で処理されるようにしている。顔領域はYOLOを用いて抽出する。全体と顔領域の彩度と明度のヒストグラムは色の塗り方の特徴を抽出するため、畳み込み層と局所コントラスト正規化層を組み合わせた構造にしている。ヒストグラムにおいては位置情報がとても重要なため、位置ずれを許容する役割を持っているプーリング層は使用しない。目領域のRGBの値は、目の形状を特徴として抽出したいため、従来の畳み込みニューラルネットワークと同様に畳み込み層とプーリング層を組み合わせた構造の後に、局所コントラスト正規化層を挿入し、その後に更に畳み込み層とプーリング層を挿入した構造にしている。入力ごとに異なる構造で処理をした後、全結合層によって全体の情報を統合し、出力層であるソフトマックス層によって絵柄ごとに分類された各クラスに属する確率が出力される。ここでのクラスは作者ごとに分けるものとしている。

4.2 学習

畳み込みニューラルネットワークの学習は画像が各クラスに属する確率を出力するように確率的勾配降下法を用いて行う。学習済みの畳み込みニューラルネットワークに対して画像を入力したときの出力層の手前の全結合層の出力を特徴ベクトルとし、画像と特徴ベクトルの組み合わせをデータベースに保存しておく。

4.3 検索

提案システムにおいて検索を行う際には、学習済みの畳み込みニューラルネットワークに対して検索キー画像を入力し、特徴ベクトルを求める。データベース内の画像の特徴ベクトルと検索キーの特徴ベクトルとのユークリッド距離を計算し、しきい値未満である画像を検索結果として出力する。

5 計算機実験

従来システム[3]と同様の画像を対象として実験を行い、提案システムにおいてすべての画像に対してではないものの背景除去を行うことにより検索精度の向上が見られることを確認した。

参考文献

- [1] pixiv 公式サイト, <https://www.pixiv.net/>, (2022.01.05 参照).
- [2] Twitter 公式サイト, <https://twitter.com/>, (2022.01.05 参照).
- [3] 古西魁登, 長名優子: “色の出現頻度に着目した画風を考慮したキャラクタイラスト検索,” 情報処理学会第83回全国大会, 2021.
- [4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi: “You Only Look Once: Unified, real-time object detection,” <https://arxiv.org/abs/1506.02640>, (2022.01.05 参照).
- [5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, Vol.86, No.11, pp.2278–2324, 1998.
- [6] remove.bg 公式サイト: <https://www.remove.bg/>, (2022.01.05 参照).
- [7] Deep Learning で背景除去をしてみる: <https://tech.fusic.co.jp/posts/2020-01-20-remove-background/>, (2022.01.05 参照).