2N-03

色の出現頻度に着目した画風を考慮したキャラクターイラストの検索

古西魁登　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

近年の携帯型デバイスの発展によってインターネットはより身近な存在となっている。それに伴い、pixiv[1] や Twitter[2] といった SNS を通じてキャラクターイラストに触れる機会も増加している。しかし、多くのキャラクターイラストの中から、自分の好みの画風のイラストを探し出そうとすると、画風を表す適切な名称が存在するとは限らないため、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、一枚のキャラクターイラストをキーとし、それに類似した画風のキャラクター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。

キャラクターイラストの検索を行う手法として、畳み込みニューラルネットワークを用いたタッチの類似性を考慮したキャラクターイラストの検索 [3] が提案されている。この手法では、このシステムでは、画風の特徴が彩度・明度の傾向、また目の形状などに表れると考えて検索を実現している。画像の彩度と明度のヒストグラムと YOLO (You Only Look Once)[4] によって抽出した目の領域の画像を入力とし、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学習した畳み込みニューラルネットワークを用いている。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手前の全結合層の出力を特徴ベクトルとして利用し、キー画像の特徴ベクトルと類似した特徴ベクトルをもつイラストを出力することで検索を行う。しかしながら、この手法では、使われている色、色相の傾向は考慮されていないという問題がある。

そこで本研究では、色の出現頻度に着目した画風を考慮したキャラクターイラストの検索を提案する。このシステムでは、従来のシステムで入力として用いている彩度・明度のヒストグラムと目領域の画像に加えて、顔領域の色相のヒストグラムを入力として用いることで検索精度の向上を目指す。顔領域は目領域と同様に YOLO を使って抽出する。なお、色相のヒストグラムを顔領域に限定して考えるのは、構図や服装などの違いによる影響を受けにくくするためである。

Character Illustration Retrieval based on Touch considering Frequency of Color Appearance with Convolutional Neural Network

Kaito Konishi and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

2　色の出現頻度に着目した画風を考慮したキャラクターイラストの検索

ここでは、提案する色の出現頻度に着目した画風を考慮したキャラクターイラストの検索について説明する。提案システムでは画風に基づくグループに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・明度のヒストグラム、目の領域、顔の領域の色相のヒストグラムを用いる。目と顔の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、画風に着目したグループに分類されるように学習を行う。検索の際には、画像の特徴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もつ画像を検索結果として出力される。

2.1　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。入力には、彩度・明度のヒストグラム、目の領域、顔の領域の色相のヒストグラムを用い、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なっている。目の領域と顔の領域は YOLO を用いて抽出する。彩度と明度のヒストグラムと顔の領域の色相のヒストグラムを入力として与える部分では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なっている。ヒストグラムにおいてはデータの位置情報 (階級の情報) が重要なため、情報の位置に対する感度を低下させる働きを持つプーリング層は利用しない。目の領域を入力とする部分では、目の特徴として、色の塗り方だけでなく目の形状も特徴として捉えるため、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、その後ろにさらに畳み込み層とプーリング層を挿入した構造となっている。入力ごとに構造の異なる部分での処理が終わった後、全結合層によって全

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

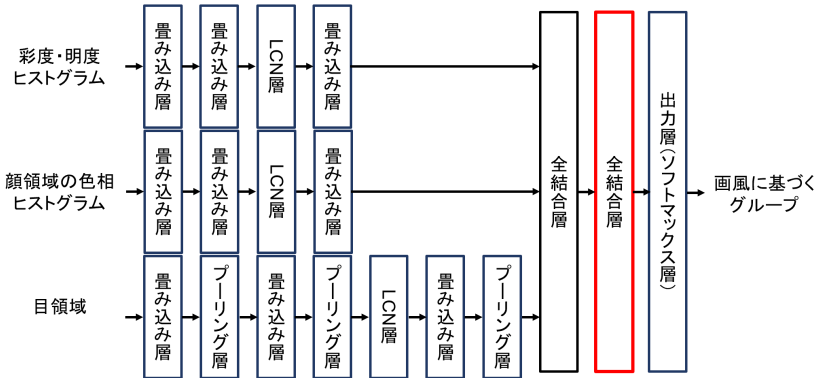


図 1: 提案システムで用いる畳み込みニューラルネットワークの構造

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画風ごとに分けら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わけるものとしている。

2.2 学習

畳み込みニューラルネットワークは、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学習を行う。学習は入力データに対する出力と教師信号との誤差が少なくなるように重みやしきい値のようなパラメータを後進することで行われる。確率的勾配降下法では、学習データの一部をランダムに抜き出してミニバッチを作成し、ミニバッチごとに計算した誤差関数に基づきパラメータの更新を行う。重みの更新は

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \frac{\partial E_s(\boldsymbol{w})}{\partial \boldsymbol{w}} \tag{1}$$

のように行う。ここで、$\boldsymbol{w}$ は重みやしきい値などのパラメータ、η は学習係数を表す。分類問題を学習させる際には誤差関数には交差エントロピー誤差関数が用いられる。ミニバッチ s に関する交差エントロピー誤差関数 $E_s(\boldsymbol{w})$ は以下のように与えられる。

$$E_s(\boldsymbol{w}) = \sum_{p \in C_s} \sum_{i=0}^{N^{(L)}-1} d_i^{(p)} \log x_i^{(L,p)} \tag{2}$$

ここで、C^s はミニバッチ s に含まれるデータの集合、$N^{(L)}$ は出力層である第 L 層のニューロン数、$d^{(p)}$ はミニバッチ s 内に含まれるデータ p の教師信号の i 番目の成分、$x_i^{(L,p)}$ はミニバッチ s 内のデータ p を畳み込みニューラルネットワークに入力した際の第 L 層のニューロン i の出力を表している。

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 (図 1 の赤枠の部分) 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。

2.3 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、距離がしきい値よりも小さいものを検索結果として出力する。

3 計算機実験

500 枚 (20 枚 ×25 グループ) のキャラクタイラストの画像を学習させ、実験を行い、従来システム [3] よりも高い精度で検索が行えることを確認した。

参考文献

[1] pixiv 公式サイト, https://www.pixiv.net/, (2021.01.06 参照)

[2] Twitter 公式サイト, https://twitter.com/, (2021.01.06 参照)

[3] 松井大樹, 長名優子 : “畳み込みニューラルネットワークを用いたタッチを考慮したキャラクターイラストの検索,” 情報処理学会第 82 回大会, 2020.

[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, (2021.01.06 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.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.