

畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索における 入力による検索精度の影響への検討

石井隆朔 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

画像検索はキーワードから検索する手法と画像をキーとして検索する方法とに大きく分けることができる。画像をキーとして検索する方法の1つとして畳み込みニューラルネットワーク [1] を用いた類似画像検索が提案されている。この方法では、画像認識の学習を行った畳み込みニューラルネットワーク [2] を利用する転移学習を用いるのが一般的である [3][4]。検索は、その畳み込みニューラルネットワークの出力層の1つ前の層の出力を特徴ベクトルとして利用し、データベースの画像の特徴ベクトルと比較することで行う。しかし、従来の畳み込みニューラルネットワークを用いた類似画像検索では、多くの場合、画像中のどこにオブジェクトが含まれているかは考慮されないため、風景画像を対象とした位置なども含めて類似した画像を検索することは困難であるという問題がある。それに対し、風景画像を左上、右上、左下、右下、中央に分けて入力して検索を行うことで、検索精度が向上することが確認されている。画像そのものを入力するのではなく、画像を分割して入力することで検索精度が向上する可能性があることが確認されているが、この手法の方法で用いられて分割の方法が最適であるとは限らない。

そこで、本研究では、畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索において入力による検索精度の影響への検討を行う。

2 畳み込みニューラルネットワーク

畳み込みニューラルネットワークは何段もの深い階層構造を持つニューラルネットワークで、特に画像認識の分野で優れた性能を発揮することが知られている。構造は図1のようになっており、一般に畳み込み層とプーリング層の組み合わせを何度か繰り返したあ

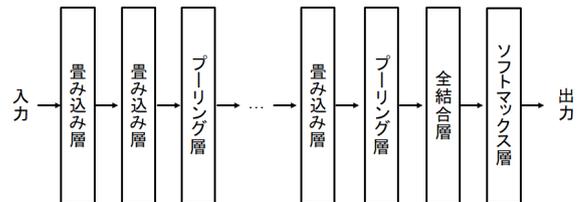


図1: 畳み込みニューラルネットワークの構造

とに、全結合層が配置される。

画像認識などの多クラス分類問題に用いる場合には出力層としてソフトマックス層が使われる。畳み込み層は、前の層の出力からフィルタが表す特徴的な濃淡構造を抽出するような働きをする。プーリング層には、畳み込み層で抽出した特徴の位置の感度を低下させ、対象とする特徴量の画像内での位置が多少変化した場合でも同じような値が出力する働きがある。全結合層では直前の層のすべての出力を結合し、ソフトマックス層ではそこまでの層で抽出した特徴に基づいてグループに分類する。

3 畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索

ここでは、本研究で扱う畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索について説明する。このシステムでは、画像認識のデータを学習済みの畳み込みニューラルネットワークを利用する転移学習を用いる。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、検索キーと類似した特徴ベクトルを持つ画像を検索結果として出力する。従来の手法では画像をそのまま入力として用いることが多かったが、本研究で扱う手法では画像を分割して畳み込みニューラルネットワークに入力する。こうすることで画像内のオブジェクトの位置の情報を考慮できるようにしている。図2に畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索のイメージを示す。

Examination of Effect of Search Accuracy on Input in Similarity-based Image Retrieval for Landscape Images using Convolutional Neural Networks
Ryusaku Ishii and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

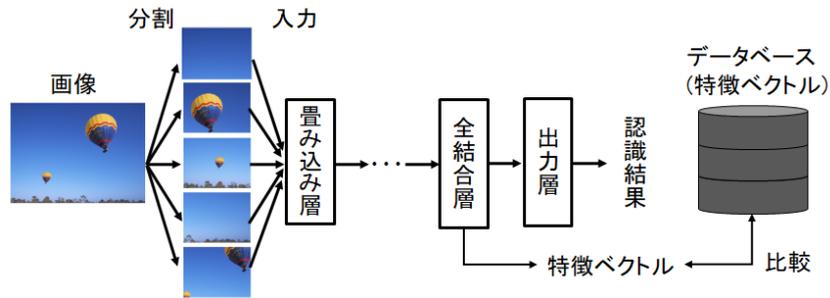


図 2: 風景画像を対象とした類似画像検索のイメージ

画像 p に対応する特徴ベクトル $\hat{c}^{(p)}$ は

$$\hat{c}^{(p)} = \frac{c^{(p)}}{|c^{(p)}|} \quad (1)$$

で与えられる。ここで、 $c^{(p)}$ は出力層の手前の全結合層の出力から生成する特徴ベクトルである。

従来の畳み込みニューラルネットワークを用いた類似画像検索手法とは違い、この手法では画像を上下左右、中央の 5 つの領域に分けて入力し、それぞれに対して出力層の手前の全結合層の出力から特徴ベクトルを求め、各領域に対する特徴ベクトルを並べて 1 つにまとめたものを特徴ベクトルとして用いることで、大まかな位置情報を保持した特徴ベクトルの生成を可能としている。特徴ベクトルは大きさが 1 になるように正規化したものを使用する。学習させた画像に対して生成した特徴ベクトルは画像と組み合わせてデータベースに保存しておく。検索時には、検索キーとして用いる画像を畳み込みニューラルネットワークに入力し、それに対する出力層の手前の全結合層の出力から特徴ベクトルを求める。検索キーの特徴ベクトルとデータベースに記録してある特徴ベクトルを比較し、ユークリッド距離の近い画像を検索結果として出力する。画像 p の特徴ベクトル $\hat{c}^{(p)}$ と検索キーの特徴ベクトル \hat{c}^{key} とのユークリッド距離 $d(\hat{c}^{(p)}, \hat{c}^{key})$ は

$$d(\hat{c}^{(p)}, \hat{c}^{key}) = \sqrt{\sum_{n=1}^{N^{(L-1)} N^{area}} (\hat{c}_n^{(p)} - \hat{c}_n^{key})^2} \quad (2)$$

で与えられる。ここで、 $N^{(L-1)}$ は出力層 (第 L 層) の直前の層のニューロン数、 N^{area} は入力する領域の数を表す。

4 計算機実験

これまでの研究では、画像全体を入力として用いる場合と、画像を 4 分割したものと中央の 5 つの画像を入力する場合について検討が行われている。今回は、

表 1: 検索精度の比較

入力	再現率	適合率	F 値
全体	0.548	0.910	0.603
4 分割	0.534	0.893	0.574
6 分割	0.521	0.914	0.576
9 分割	0.467	0.970	0.564
4 分割+中央	0.536	0.896	0.574
4 分割+全体	0.533	0.901	0.582
6 分割+全体	0.517	0.927	0.581

新たに 4 分割、6 分割、9 分割、4 分割した画像と全体の画像を組み合わせる場合、6 分割した画像と全体の画像を組み合わせる場合について検索精度がどのように変わってくるか検討を行った。その結果を表 1 に示す。

参考文献

- [1] 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.
- [2] K. Simonyan and A. Zisserman : “Very deep convolutional networks for large-scale image recognition,” Proceedings of International Conference on Learning Representation, 2015.
- [3] 中山英樹 : “深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習,” 電子情報通信学会音声研究会, 2015.
- [4] 内田祐介, 山下隆義 “畳み込みニューラルネットワークの研究動向,” パターン認識・メディア理解研究所, 2017.