

## PSPNet のラベルと自然物領域の色相を考慮した風景画像を対象とした類似画像検索

中村亮 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、スマートフォンなどのデジタル機器の普及や高性能化により画像を扱ったコンテンツが増加している。それに伴い、画像をキーとして検索を行う類似画像検索に関する研究も盛んに行われている。そのような手法の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に関する研究も行われている。この手法では、事前に画像認識の学習を行った VGG16[2] など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という手法がとられている。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、事前に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に画像を入力したときの出力層の1つ手前の層の出力をその画像の特徴を表す特徴ベクトルとして利用して、検索が実現される。画像と特徴ベクトルを結び付けてデータベースに保存しておき、検索キーとなる画像を畳み込みニューラルネットワークに入力して生成した検索キーに対応する特徴ベクトルをデータベース内の特徴ベクトルと比較することで類似している画像の検索を行う。しかし、画像分類の学習を行ったニューラルネットワークを利用しているため、特定のオブジェクトが含まれる画像の検索は得意であるが、風景画像などを対象とした類似画像検索などにおいて画像内での位置などを考慮した検索は困難であるという問題がある。それに対し、複数の領域に分けて入力し、領域ごとに特徴ベクトルを生成することで画像内の大まかな位置情報を考慮した検索方法が提案されている [3]。これは風景画像を対象とした検索手法であり、画像内のオブジェクトの位置情報を大まかに考慮することで検索精度の向上を目指している。しかしながら、どこにどのようなオブジェクトが含まれているかを正確には考慮できていないという問題がある。

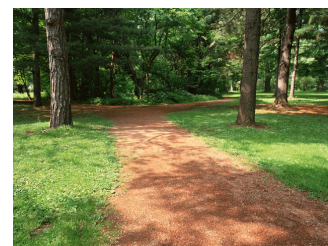
また、風景画像を対象とした検索において建物などの人工物が含まれている場合に色情相を考慮した検索を行うと検索精度が低下する可能性が指摘されて

いる。そのために、畳み込みニューラルネットワークを用いて風景画像中の人工物領域を特定しようとする試みも行われている [4]。それに対し、画像中に含まれるオブジェクトを画素単位で認識し、領域分割する Semantic Segmentation に関する研究が行われており、PSPNet[5] などの手法が提案されている。

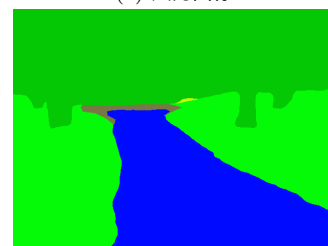
本研究では、PSPNet のラベルと自然物領域の色相を考慮した風景画像を対象とした類似画像検索を提案する。提案手法では、SegNet における領域分割の結果として得られる画素単位のカテゴリラベルと、そのカテゴリラベルから得られる情報に基づいて自然物領域に関してのみ色相を考慮して類似度を求めることで、類似画像検索を実現する。

## 2 PSPNet

PSPNet (Pyramid Scene Parsing Network)[5] は Encoder に ResNet101[6] の畳み込み層を利用し、Encoder と Decoder の間に空間ピラミッドプーリングモジュールを追加したものであり、画像分割を行うことができる。画像を入力すると画像に存在するオブジェクトに対して画素単位でラベリングを行い、画像分割と認識を同時に実現することができる。画像分割と認



(a) 入力画像



(b) 分割結果

図 1: PSPNet による画像分割の例

Similarity-based Image Retrieval for Landscape Images considering PSPNet Label and Hue of Natural Objects  
Ryo Nakamura and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

識の結果は図1のようにラベルごとに色分けされた画像として出力される。

### 3 PSPNetのラベルと自然物領域の色相を考慮した風景画像を対象とした類似画像検索

ここでは、提案するPSPNetのラベルと自然物領域の色相を考慮し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。提案手法では、SegNetの出力ラベルと自然物領域の色相をダウンサンプリングしたものを特徴量として利用する。画像とこの特徴量とを関連付けてデータベースに蓄積しておき、検索キーとデータベース中の特徴量を比較することで類似した画像を検索する。

検索を行う際には、特徴量間の距離を求め、その距離がしきい値以下であるような画像を検索結果として出力することになる。検索キーとデータベースの画像  $p$  の特徴量間の距離  $D^{(key,p)}$  は

$$D^{(key,p)} = D^{H(key,p)} + \alpha \left( 2|C^{UM(key,p)}| \right) \quad (1)$$

のように表すことができる。 $D^{H(key,p)}$  は色相に関する距離、 $C^{UM(key,p)}$  は検索キーと画像  $p$  でラベルが一致しない画素の集合、 $\alpha$  は重み付け係数である。 $2|C^{UM(key,p)}|$  はラベルが一致しない画素に関する距離を表しており、 $|C^{UM(key,p)}|$  に2を掛けているのは色相の角度が最も離れているときの距離の値2をペナルティとして扱っているからである。重み付け係数  $\alpha$  の値は実験を行った上で検索精度が最も高くなるように決定する。

検索キーと画像  $p$  の色相に関する距離  $D^{H(key,p)}$  は

$$D^{H(key,p)} = \sqrt{\sum_{i \in C^M(key,p) \cap C^N(key)} s_i^{(key,p)} + c_i^{(key,p)}} \quad (2)$$

$$s_i^{(key,p)} = \left( \sin(h_i^{key}) - \sin(h_i^{(p)}) \right)^2 \quad (3)$$

$$c_i^{(key,p)} = \left( \cos(h_i^{key}) - \cos(h_i^{(p)}) \right)^2 \quad (4)$$

で与えられる。ここで、 $C^M(key,p)$  は検索キーと画像  $p$  でラベルが一致した画素の集合、 $C^N(key)$  は検索キーにおける自然物の画素の集合、 $h_i^{key}$  は検索キーの画素  $i$  の色相、 $h_i^{(p)}$  はデータベース中の画像  $p$  の画素  $i$  の色相を表している。提案手法では、自然物領域であり、かつラベルが一致している画素に対してのみ色相を考慮して距離の計算を行う。なお、ここでは自然物領域、人工物領域という呼び方をしているが、厳密に

はラベルの内容とは一致しておらず、ラベルごとに色相を考慮するかどうかを割り当てている。

距離がしきい値以下の画像が検索キーと類似した画像として検索結果として表示される。なお、しきい値も実験を行った上で検索精度が高くなるような値に決定する。

### 4 計算機実験

文献 [3] において実験に用いられている 550 枚の画像を対象として検索実験を行った。提案手法において、ラベルを考慮せずに畳み込みニューラルネットワークを用いて検索を行った従来手法 [3] よりも高い精度で検索が行えることがわかった。

### 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,” <https://arxiv.org/pdf/1409.1556.pdf>, (2022年1月5日参照).
- [3] 矢森大貴, 長名優子 : “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,” 情報処理学会第83回全国大会, 2021.
- [4] 小川京太, 長名優子 : “畳み込みニューラルネットワークを用いた風景画像中の人工物領域の特定 – ファインチューニングの利用による精度の向上 –, ” 情報処理学会第83回全国大会, 2021.
- [5] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia : Pyramid Scene Parsing Network <https://arxiv.org/abs/1612.01105>, (2022年1月5日参照).
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun : <https://arxiv.org/abs/1512.03385>, (2022年1月5日参照).