

DeepLabv3+の出力ラベルと色情報分布を用いた風景画像検索

平賀圭伍 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

類似画像検索を行う手法として、様々な手法が提案されている。その中の一つとして、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 [1] が提案されている。これは画像認識を学習した学習済みの畳み込みニューラルネットワーク [2] である VGG16[3] の出力層の手前の層の出力を特徴ベクトルとして利用し検索する手法である。しかし、この手法では、画像認識を学習したモデルを用いているため、空や森など形が定まらないようなものを含む画像に対して特徴ベクトルが適切に生成できていないことがあるという問題がある。

それに対し、画素単位でラベル付けする Semantic Segmentation の結果と色情報とを用いて画像が検索する手法 [4][5] が提案されている。Semantic Segmentation を行うモデルとして、文献 [4] では PSPNet[6] が、文献 [5] では DeepLabv3+[7] が用いられている。この手法では、Semantic Segmentation の結果の出力ラベルと各画素の HSV 色空間における座標を特徴量として画像を関連付けてデータベースに保存しておき、検索キーとデータベース中の画像の特徴量を比較することで類似した画像の検索を実現している。しかし、これらの手法では、色情報に関して画素単位で距離を計算しているため、形が少し違っただけで類似した画像として検索されないことがあるという問題がある。

本研究は、DeepLabv3+の出力ラベルと色情報の分布を用いた風景画像検索を提案する。提案手法では、情報も画素単位ではなく領域単位で比較するものとし、Deep Lab v3+の出力ラベルと HSV 色空間における座標のヒストグラムを特徴量として利用する。

2 DeepLabv3+

DeepLabv3+[7] はエンコーダとデコーダから構成されるモデルであり、エンコーダは Xception[8] と呼ばれる畳み込みニューラルネットワーク [2] を改良した

モデルと Atrous 空間ピラミッドプーリングから構成されている。また、デコーダで元のサイズにもどす際に、Xception の中間層の特徴情報とデコーダの特徴情報を組み込み合わせることで画像全体の位置関係を考慮しながらデコードを行うことができる。画像が入力としてエンコーダに与えられるとを行った結果が Semantic Segmentation を行い、画素単位でラベル付けした結果がデコーダから出力される。

3 DeepLabv3+の出力ラベルと色情報の分布を用いた風景画像検索

提案手法では、画像を矩形の領域に分割し、領域ごとの特徴量を比較すること類似した画像の検索を行う。画素単位でなく領域単位で比較することで、オブジェクトの位置が多少異なっても似ている画像であると判断できるようにしている。特徴量としては、DeepLabv3+[7] の出力ラベルと HSV 色空間における座標を用いる画像の特徴量と画像を関連付けてデータベースに保存しておき、検索キーとデータベース中の画像の特徴量を比較することで類似した画像の検索を実現する。検索の際には、領域ごとの特徴量としてとらえるため、出力ラベル、色情報ともに領域内のヒストグラムで比較することになる。

3.1 HSV 色空間における座標の分布のヒストグラムによる表現

HSV 色空間における座標の分布をヒストグラムで表現する場合に、 x, y, z のそれぞれを別のヒストグラムとして表現すると、異なる色の分布でも同じヒストグラムになることがある。

x, y, z を個別に扱うのではなく、1つのヒストグラムとして表現できれば解決できるが、そのままでは x, y, z を1つの軸として扱うことができず、横軸の階級を決めることができない。それに対しては、色空間における座標 x, y, z をクラスタリングすることで似た色をまとめて1つの階級として扱うことが可能になる。提案手法では、 K -means++法 [9] を用いて HSV 色空間上の座標に対してクラスタリングを行って階級を決定し、領域ごとのヒストグラムを作成する。

Landscape Image Search using Output Labels of DeepLabv3+ and Color Information Distribution
Keigo Hiraga and Osana Yuko(Tokyo University of Technology, osana@stf.teu.ac.jp)

3.2 画像検索過程

画像検索過程では、検索キーとデータベース内の画像の特徴量を比較し、その距離がしきい値以下であれば類似画像として出力する。検索キーとデータベース内の画像 p の特徴量の距離 $D^{(key,p)}$ はラベルに関する距離 $D_{label}^{(key,p)}$ と色に関する距離 $D_{color}^{(key,p)}$ との和で

$$D^{(key,p)} = D_{label}^{(key,p)} + \alpha D_{color}^{(key,p)} \quad (1)$$

のように与えられる。ここで、 α は重み付け係数である。

ラベルに関する距離は、領域を矩形に分割し、それぞれの領域におけるラベルの分布が類似しているかを表す JS (Jensen-Shannon) ダイバージェンスを用いて表現する。JS ダイバージェンスは2つの確率分布が類似しているほど小さくなり、同じ分布のときには0となる。検索キーと画像 p とのラベルに関する距離 $D_{label}^{(key,p)}$ は

$$D_{label}^{(key,p)} = \frac{1}{M} \sum_{j=1}^M D_{JS} \left(P_{label,j}^{key} \parallel P_{label,j}^{(p)} \right) \quad (2)$$

で与えられる。ここで、 M は領域数である。ここで、 $D_{JS} \left(P_{label,j}^{(key)} \parallel P_{label,j}^{(p)} \right)$ は検索キーの領域 j のラベル分布 $P_{label,j}^{key}$ と画像 p の領域 j のラベル分布 $P_{label,j}^{(p)}$ の JS ダイバージェンスであり

$$D_{JS} \left(P_{label,j}^{key} \parallel P_{label,j}^{(p)} \right) = \frac{1}{2} D_{KL} \left(P_{label,j}^{key} \parallel P_{label,j}^{(key,p)} \right) + \frac{1}{2} D_{KL} \left(P_{label,j}^{(p)} \parallel P_{label,j}^{(key,p)} \right) \quad (3)$$

で与えられる。ここで、 $P_{label,j}^{(key,p)}$ は分布 $P_{label,j}^{key}$ と $P_{label,j}^{(p)}$ の平均であり

$$P_{label,j}^{(key,p)} = \frac{1}{2} \left(P_{label,j}^{key} + P_{label,j}^{(p)} \right) \quad (4)$$

である。また、 $D_{KL}(P, Q)$ は分布 P, Q の KL (Kullback-Leibler) ダイバージェンスであり

$$D_{KL}(P, Q) = P \log \left(\frac{P}{Q} \right) \quad (5)$$

で与えられる。

検索キーと画像 p との色に関する距離 $D_{color}^{(key,p)}$ は

$$D_{color}^{(key,p)} = \frac{1}{M} \sum_{j=1}^M D_{JS} \left(P_{color,j}^{key} \parallel P_{color,j}^{(p)} \right) \quad (6)$$

で与えられる。これも、ラベルの分布に関する距離と同様に求められる。

4 計算機実験

計算機実験を行い、提案手法において色情報とラベルの分布を考慮した検索が行えることを確認した。

参考文献

- [1] 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.
- [2] 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.
- [3] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane and M. Jagersand: “U2-Net: Going deeper with nested U-structure for salient object detection,” Pattern Recognition, Vol.106, 2020.
- [4] 李泰英, 長名優子: “PSPNet の出力ラベルの分布と色情報を用いた風景画像検索,” 情報処理学会第84回全国大会, 2022.
- [5] 高松晴仁, 長名優子: “Deep Lab v3+の出力ラベルと色情報を用いた風景画像検索,” 情報処理学会第84回全国大会, 2023.
- [6] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia: “Pyramid scene parsing network,” <https://arxiv.org/pdf/1612.01105.pdf>, 2017 (2024/01/07 参照).
- [7] L. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam: “Encoder-decoder with atrous separable convolution for semantic image segmentation,” <https://arxiv.org/pdf/1802.02611.pdf>, 2018 (2024/01/07 参照).
- [8] F. Chollet: “Xception: deep learning with depthwise separable convolutions,” <https://arxiv.org/pdf/1610.02357.pdf>, 2017 (2024/01/07 参照).
- [9] D. Arthur and S. Vassilvitskii: “K-means++: the advantages of careful seeding,” Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, pp.1027–1035, 2007.