

PSPNet の出力ラベルの分布と色情報を用いた風景画像検索

李泰英 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

畳み込みニューラルネットワーク [1] を用いて類似画像検索を行う手法として、画像分類を学習させた畳み込みニューラルネットワークである VGG16[2] などを利用する手法がいくつも提案されている。これらの手法では、学習済みの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が画像全体の特徴を反映していると考え、検索を行う際の特徴量として利用されている。しかしながら、画像分類を学習した畳み込みニューラルネットワークでは、画像中に含まれる含まれる 1 つのオブジェクトを認識するように学習が行われているため、複数のオブジェクトを含む風景画像の検索などでは必ずしも適切な特徴量とはなっていない可能性がある。また、風景画像中の検索を行う場合には、画像にどのようなオブジェクトが含まれているかだけでなく、画像のどのあたりにどのようなオブジェクトがあるかという情報も重要な意味をもつこともある。それに対しては、画像を複数の領域に分割し、それぞれ畳み込みニューラルネットワークに入力して特徴量を得ることで画像内での位置情報も考慮した検索を行おうとする試みも行われている [3]。

また、Semantic Segmentation の結果を利用して類似画像検索を行うとする方法も提案されている。そのような手法の 1 つとして、HSV 色空間における座標と PSPNet (Pyramid Scene Parsing Network)[4] の出力ラベルを利用した類似画像検索 [5] が提案されている。しかしながら、この手法では、(1) 画素単位でラベルの比較を行っているため、形状が少しでも違うと似ていないと判断されてしまう、(2) 似たようなラベルであっても厳密に一致していないと似ていないと判断されてしまう、(3) 似たような構図の画像であっても明るさの差によって色が似ていないと判断されてしまうなどの問題がある。

本研究では、PSPNet の出力ラベルの分布と色情報を用いた風景画像検索を提案する。この手法は、従来の HSV 色空間における座標と PSPNet から出力され

るラベルを利用した類似画像検索 [5] に基づいた手法である。形状が少しでも違うと似ていないと判断されてしまう問題に対しては、画素ごとではなく領域ごとにラベルの分布を比較することで解決する。また、似たようなラベルをグループ化することで、ラベルが厳密に一致しないと似ていないと判断されてしまう問題を、特徴ベクトルの距離の計算において明度の影響が小さくなるように重み付けすることで明るさの差によって色が似ていないの判断されてしまう問題を解決する。

2 PSPNet の出力ラベルの分布と色情報を用いた風景画像検索

ここでは、提案する PSPNet の出力ラベルの分布と色情報を用いた風景画像検索について説明する。

2.1 画像蓄積過程

提案手法では画像に対応する特徴量をあらかじめ抽出し、画像と関連付けてデータベースに保存しておく。特徴量としては、各画素の HSV 色空間における座標と PSPNet の出力ラベルを使用する。

(1) 各画素の HSV 色空間における座標

1 つ目の特徴量としては、円錐型の HSV 色空間における各画素の座標を用いる。人が似ていると感じる色どうしは HSV 色空間上で近い位置にあると考えられるため、提案手法では色の類似度を把握するのに RGB の値ではなく HSV 色空間での座標を利用する。

画素 i の HSV 色空間における座標 x_i, y_i, z_i は

$$x_i = s_i v_i \cos(h_i) \quad (1)$$

$$y_i = s_i v_i \sin(h_i) \quad (2)$$

$$z_i = v_i \quad (3)$$

で与えられる。ここで、 h_i は画素 i の色相 (Hue : H)、 s_i は画素 i の彩度 (Saturation : S)、 v_i は画素 i の明度 (Value : V) である。

Landscape Image Retrieval Using PSPNet Output Label Distribution and Color Information
Lee Taeyeng and Osana Yuko (Tokyo University of Technology, osana@stf.teu.ac.jp)

(2) PSPNet の出力ラベル

2つ目の特徴量としては PSPNet の出力ラベルを用いる。PSPNet では画素単位で画像を分割し、ラベリングすることができる。従来手法 [5] では、画素ごとのラベルをそのまま特徴量として用いていたため、似たようなラベルであっても厳密に一致していないと似ていないと判断されてしまうという問題があった。提案手法では、

{Road, Path, Sidewalk} {Path}
 {Bench, Armchair, Chair} {Chair}

のように類似したラベルをグループ化したものを用いることでこの問題を解決する。

2.2 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

$$D^{(key,p)} = D_{label}^{(key,p)} + \alpha D_{color}^{(key,p)} \quad (4)$$

のように与えられる。ここで、 α は重み付け係数である。

検索画キーと画像 p との色に関する距離 $D_{color}^{(key,p)}$ は画素ごとの HSV 色空間の座標の差であり

$$D_{color}^{(key,p)} = \frac{1}{8 + \beta} \sum_{i=1}^N \left\{ \left(x_i^{key} - x_i^{(p)} \right)^2 + \left(y_i^{key} - y_i^{(p)} \right)^2 + \beta \left(z_i^{key} - z_i^{(p)} \right)^2 \right\} \quad (5)$$

で与えられる。ここで、 β は明度に関する重み付け係数である。 β の値を小さく設定することで、明度に対応する z 座標の影響を小さくすることができ、 x, y 座標の値を重視できるようになる。このようにすることで、似たような構図の画像であっても明るさの差によって色が似ていないと判断されてしまうという問題が軽減されることを期待している。なお、式 (5) において $8 + \beta$ で割っているのは、各画素の距離の値が $0 \sim 1$ になるように正規化するためである。

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

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

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

領域の分割方法は、すべてのピクセルが 1 つの領域にのみ属するように分割する方法と複数の領域に属するように分割する方法について検討する。また、領域のサイズについては実験により検索精度が高くなるように決定する。

3 計算機実験

計算機実験を行い、提案システムにおいて従来システム [5] よりも高い精度 (F 値 0.567) で検索が行えることを確認した。

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>, 2015 (2022/07/22 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] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia : “Pyramid scene parsing network,” <https://arxiv.org/abs/1612.01105>, 2017 (2022/12/22 参照).
- [5] 興津 祐人, 長名優子 : “HSV 色空間における座標と PSPNet から出力されるラベルを利用した風景画像検索,” 情報処理学会第 84 回全国大会, 2022.