

顕著情報を加えた DCNN による高山植物認識システム

根岸 知誠[†]

山梨大学 大学院医工農学総合教育部[†]

服部 元信[‡]

山梨大学 大学院総合研究部[‡]

1 はじめに

高山植物とは、高山帯に生息している植物を指し、日本に生息している種類は約 440 種である。本研究では、Deep Convolutional Neural Network (DCNN) を用いて高山植物の認識システムを構築することで、煩雑な種類の同定を自動で行うことを最終的な目的としている。

人が植物の種類を同定することを前提に高山植物画像を見ると、植物以外の背景情報を差し置いて、植物の花弁や葉の色、形状に集中することが考えられる。また、注視される部分は一般的に種類の同定に重要な部分であることが考えられる。顕著性マップは、このような画像に対して人の注視が予測される領域を画像として表現できる。

実験では、植物画像認識の一般的なタスクとして Oxford Flower Dataset を実験データに設定し、画像内の重要だと思われる領域情報を顕著性マップによってネットワークに与えて学習を行うことで、認識システムの性能の向上を図った。

2 DCNN

DCNN は、主に画像認識に応用される順伝播型の多層ニューラルネットワークである。画像処理の基本的な演算によって特徴抽出を行う中間層、中間層からの出力情報から認識結果を出力する識別層から構成される(図 1)。中間層は convolution 層 → pooling 層 → normalize 層という処理を繰り返し行い、入力された画像の特徴表現を獲得している。各層は、複数の画像状の平面構造を持つ素子から成り立っている。

- convolution 層: 畳み込み演算が行われ、画像上の特徴的な濃淡パターンの検出を行う
- pooling 層: 最大プーリングを行い、情報圧縮と局所変形に対する応答の頑健性の向上を図る
- normalize 層: 出力の正規化を行い、一連の処理のあとの出力が過剰に大きくなることを抑制する

この DCNN では損失関数として softmax 関数を用い、誤差逆伝播法 (Back-propagation: BP) によって学習を行う。

Alpine Plants Recognition with Deep Convolutional Neural Network using Saliency

[†] Negishi Tomoaki, University of Yamanashi

[‡] Hattori Motonobu, University of Yamanashi

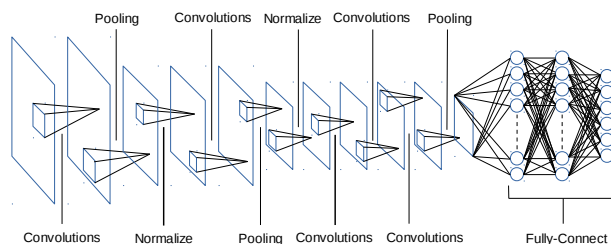


図 1: DCNN の概要

3 顕著性マップ

人は外界から来る情報が重要であるか、そうでないかを選択する能力があるため、画像を見た時にはその主成分を認識することができる。植物画像の場合、背景情報を差し置いて、植物の花弁や花の色、形状に集中し、種類の同定に必要な情報を獲得することができる。このように、画像の中で人の注視が集まりやすい場所を推定し注視の予測される確率を画像として表現したものを顕著性マップ(図 2: 中央)と呼ぶ。

Liu ら[1]は、DCNN による水田害虫画像の分類性能を向上させるために、顕著情報の予測を利用した学習データベースの作成を行った。実験では、画像から抽出された顕著領域が含まれる範囲を切り取ることで新しい画像データベースとして作成し、これによって学習を行うことで分類性能を向上させている。

Liu らの研究では、顕著性マップの予測に global contrast based salient region detection (GC Saliency) [2] が採用された。GC Saliency は周囲に比べて高い色の差が見られる領域を顕著領域として算出することが可能で、水田害虫画像においては、認識対象である水田害虫が背景に比べてコントラストの強い色領域を持つという特徴から効果を発揮している。本研究でも、植物の花弁と背景に同様の関係が確認できたため、この方法を採用した。

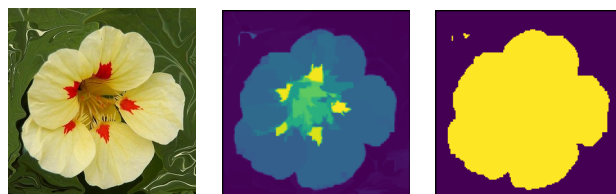


図 2: 顕著性マップの例(左:オリジナル,中央:顕著性マップ,右:顕著性マップを 2 値化したもの)

4 実験

DCNN モデルに顕著性マップの情報を導入する実験を行った。本来の画像の入力と並行して、顕著情報を加えた画像を入力として与えて fine-tuning を行う。顕著情報を加える方法として、画像に対して顕著性マップを乗算する方法を取った。乗算は元の画像に対して顕著領域の情報を反映する簡単な方法であると考えられ、この画像によって分類に有用な特徴を強く学習することを期待している。顕著性マップによる情報の有無による認識率を比較し、認識性能への影響を確認した。

4.1 実験データ

実験データとして Oxford Flower Dataset を利用した。このデータセットは 102 種類の植物画像クラスから成り、各クラスは 40 から 258 枚の画像で構成されている。全体の画像枚数は 8189 枚であり、この中から重複の無いようにそれぞれ訓練データに 1020 枚、テストデータに 1020 枚用いることで実験を 8 回試行し、その平均で比較を行った。

4.2 実験方法

実験に用いたネットワーク構造を図 3 に示す。ベースとなる DCNN モデルには VGG19[3]を採用した。自然画像を用いて事前学習した DCNN として、ImageNet データセットで学習済みのモデルを用いた。ネットワークの各 Max Pooling 層の前に Batch Normalization (BN) を設置し、事前学習時に用いられていた平均画像の減算と Drop Out を除去している。初期学習率は 0.001、学習回数 200 回ごとに 0.96 倍することで学習率を減衰させた。

乗算する顕著性マップは GC Saliency から出力されるオリジナルのもの、と顕著領域と背景で 2 値化したものを用意した。顕著性マップの出力は [0,1] の値を取り、顕著であるとされる領域ほど高い値で表現される。2 値化の処理では、0.2 をしきい値として 0 と 1 に分割した。

上記によって顕著情報が加えられた画像の入力を行うため、事前学習された DCNN の重みを保持したまま中間層をコピーし、元のネットワークと並列に設置する。入力された画像は、それぞれの中間層を通して特徴抽出される。中間層から出力された特徴量 (図 3: FC1, FC saliency) は連結され (図 3: Concatenate), 次の全結合層へと渡される。

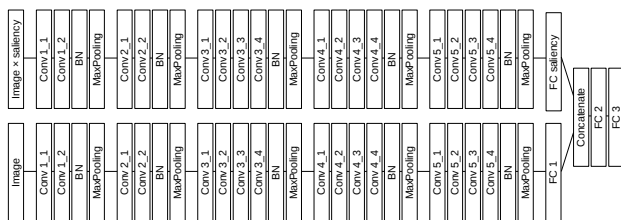


図 3: 実験に用いたネットワーク構造

4.3 実験結果

実験結果を表 1 に示す。fine-tuning では、誤差が十分に下がると学習回数によって認識率にばらつきができることが実験的に確認できた。そこで、fine-tuning を学習回数 10000 回まで続け、200 回ごとにテストデータの認識精度を確認した。その中で一番高い精度を最高認識率として比較を行った。

表 1: 交差検証による最高認識率の平均

再学習方法	平均認識率(%)	標準偏差
顕著情報なし	89.87	1.05
2 値化なし	90.16	0.74
2 値化あり	90.65	1.16

5 考察

表 1 から、顕著情報なしの場合に比べて認識率が向上している。顕著情報を用いた手法はどちらの認識率の上昇も微小ではあるが、2 値化の場合では、有意な差を確認することができた ($p < 0.01$)。2 値化なしの場合、領域のセグメンテーションに依存した顕著値によって掛けられた画像がネットワークに渡されるため、DCNN による形状特徴の抽出に対して悪く影響したことが考えられる一方、2 値化ありの場合、顕著であると算出された画像内の領域がそのまま渡されるため、DCNN による自然な処理が行えたのではないかと考えられる。

元の画像に併せて、画像内の顕著領域をネットワークに与えることで、植物画像の分類に有用な特徴抽出が強化され、認識性能を向上できたのではないかと考えられる。

6 まとめ

本研究では、DCNN を用いた高山植物認識システムの構築を目的とし、認識性能の向上を図った。認識に重要な領域の学習強化を期待し、顕著性マップを用いた学習と認識を行った。一般的な問題として植物画像を対象に行った実験では、顕著情報の付与によって認識性能が向上することを確認できた。

参考文献

- [1] Liu, Ziyi, et al. "Localization and Classification of Paddy Field Pests using a Saliency Map and Deep Convolutional Neural Network." Scientific reports 6, 20410. 2016.
- [2] Cheng, Ming-Ming, et al. "Global contrast based salient region detection." IEEE Transactions on Pattern Analysis and Machine Intelligence 37.3. pp. 569-582. 2015.
- [3] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556. 2014.