H-022

CNN を用いた撮影都市の推定のための色制御
Picture Color Control for Location Estimation using CNN

杉山 和弘†1 Kazuhiro Sugiyama
黒木 修隆†1 Nobutaka Kuroki
沼 昌宏†1 Masahiro Numa

1. はじめに

近年，画像や動画の共有サービスの普及により，様々な都市の風景画像や映像を投稿，取得できるようになった．これらの情報は鑑賞だけでなく，景観シミュレーションや地域文化の調査にも利用可能である．しかし，撮影地点が不明なデータについては，手作業または自動で地域の情報を付与する必要がある．そこで本稿では CNN を用いて撮影都市を推定する手法を提案する．

2. VGG16 による画像分類

2.1 概要

図 1 に VGG16[1]のネットワーク構造を示す．VGG16 は汎用的に利用可能な画像の分類器である．学習の際には，1 枚の入力画像につき 1 つの出力ラベルを教える．しかし，その出力ラベルの根拠となる領域を教えることはできない．

2.2 問題点

我々は VGG16 を用いて写真の撮影都市を推定する実験を行った． 図 2 は推定に失敗した画像について， Grad-CAM[2]を用いて VGG16 の着目領域を可視化したものである．この例では，空や雲などの都市ごとに違いがあまり見られない要素に注目していることがわかる．このように VGG16 では，画像中の本質的ではない特徴で都市推定を行うことがあり，誤認識が発生していた．

3. 提案手法

空や雲の学習を抑制し， 都市全体で頻出する人工物に着目させることで，認識率の向上を図る手法を提案する．
提案手法では 2 種類の CNN を用意し， 一つは通常の画像で, もう一つは色を制御した画像のみで学習を行う．

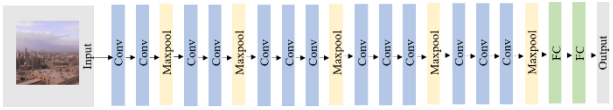
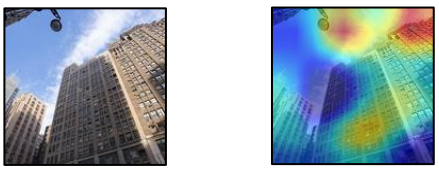


図 1 VGG16 のネットワーク構造



(a) 入力画像 (b) 着目箇所

図 2 VGG16 が誤認識した例

3.1 VGG16 をベースとするCNN_A

提案手法では VGG16 を一部変更したネットワーク構造を用いた．各畳み込み層と活性化関数の前にバッチ正規化層を挿入し， 全結合層の 1 層目直前の MaxPooling から Global AveragePoolong に変更した． このモデルをカラー画像で訓練したものを「CNN_A」と定義する．

3.2 色の制御を用いたCNN_B

図 3 に色を制御した画像を示す． 空に代表される青色は都市推定の手がかりにはなりにくい考えられるため， HSV 空間を用いて色を制御する．原画像を I, 原画像の色相成分を I_Hとするとき，次の条件,

$$0 < I_H < 160 \text{ and } 260 < I_H < 360 \quad (1)$$

を満たす青色を抽出する．そして，この領域の彩度成分$I_s = 0$ とすることで制御画像を生成する． この制御画像で訓練したものを「CNN_B」と定義する． なお, 推論時には図 4(a)のように, 学習時と同様の色制御を行う．

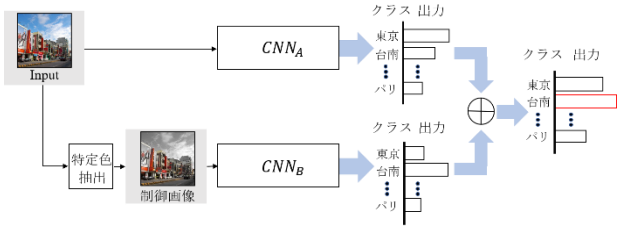
3.3 CNN_AとCNN_Bの融合

CNN_AとCNN_Bの二つの CNN を用いて推論を行う手法を「CNN_A+CNN_B」と定義する． この手法では，図 4(b)のように出力を足し合わせた後に判定する．

(a) 原画像 (b) 制御画像

図 3 色を制御した画像

(a) CNN_Bの推論段階の流れ



(b) CNN_A+CNN_Bの推論段階の流れ

図 4 各手法における推論段階の流れ

†1 神戸大学大学院工学研究科,
Graduate School of Engineering, Kobe University

表 1　各都市に対する認識率

都市名	認識率 [%]		
	CNN_A	CNN_B	$CNN_A + CNN_B$
バルセロナ	56.6	**83.3**	**83.3**
カイロ	60.0	**73.3**	66.6
京都	76.6	76.6	**80.0**
ニューヨーク	76.6	**89.9**	83.3
パリ	56.6	**69.9**	66.6
東京	73.3	76.6	**80.0**
チュニス	**89.9**	53.3	80.0
台南	83.3	**93.3**	**93.3**
ハンブルグ	**46.6**	23.3	43.3
平均	68.8	71.1	**75.1**

4.　実験と考察

4.1　実験内容

本実験では，CNN_A, CNN_B, および $CNN_A + CNN_B$ の 3 種類の手法の性能比較評価を行う．画像共有サイトの flickr [3] から，バルセロナ，カイロ，京都，ニューヨーク，パリ，東京，チュニス，台南，ハンブルグの 9 都市の画像をそれぞれ 270 枚ずつ取得した．いずれも 150×150 pixel のカラー画像である．学習用画像は各クラス 240 枚，評価用画像は各クラス 30 枚，学習回数は 110 epochs，バッチサイズは 32 とする．

なお, 学習用画像はデータオーギュメンテーションのため，左右反転画像を加え，さらに画像の一部をランダムでマスクする Random Erasing を施した．

4.2　結果と考察

表 1 に認識率の結果を示す．CNN_A に比べ，CNN_Bの平均認識率が 2.3 pt, また, CNN_A+CNN_Bの平均認識率が 6.3 pt 向上した．

図 5 はCNN_Aで不正解，CNN_A+CNN_Bで正解となった画像について Grad-CAM による着目領域の可視化を行ったものである．CNN_Aでは，都市ごとに違いがあまりない木や空などに着目していたり，他の都市と特徴が似ている要素に着目している場合に誤認識を起こしやすいことが確認できた．それに対して，CNN_Bではユニークな特徴を持つ建物などに着目していることが確認できた．このように青の色相を抽出し，その領域をグレースケール画像でマスクしたことにより，空や雲の学習を抑制でき，建物や都市全体を構成する人工物などの本質的な要素に着目することができた．さらに両者を融合した CNN_A+CNN_Bでは着目領域が分散することから，都市ごとの特徴をより多角的に捉えることができたと考えられる．

5.　まとめ

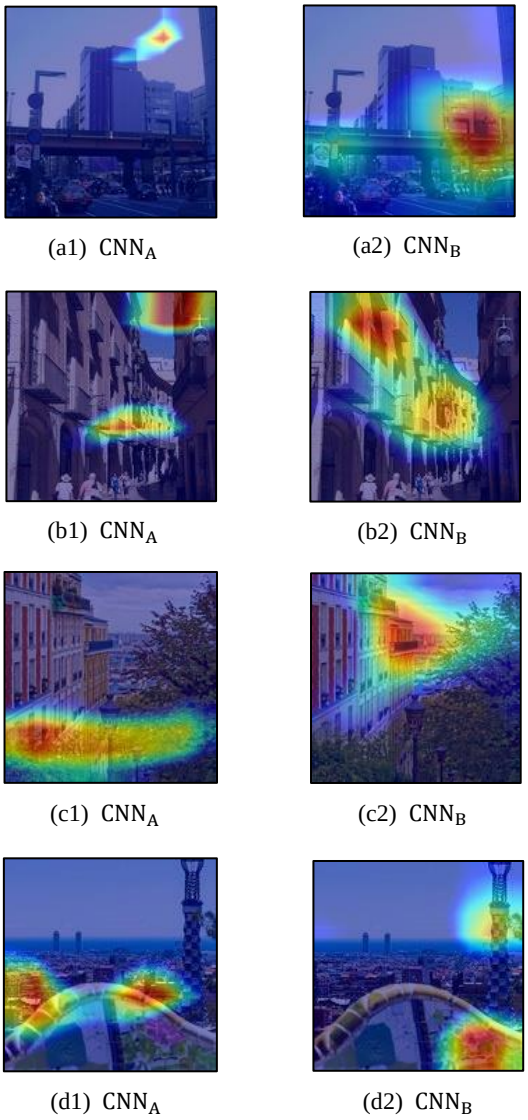
本論文では，CNN を用いて 1 枚の画像からそれが撮影された都市の推定を行うことを目的に，色の制御を行う学習方法を提案した．提案手法では特定の色の彩度を下げて学習することで，CNN の着目領域が偏らないように工夫した．実験の結果，通常の画像で学習する CNN_Aに比べて，CNN_Bの平均認識率は 2.3 pt 向上した．さらに，両者を融合したCNN_A+CNN_Bの平均認識率は 6.3 pt 向上した．これらの結果より，撮影都市の推定において着目領域を色で制御することの有用性を確認できた．今後の課題は抽出する特定色の範囲の検討である．

参考文献

[1] K.Simonyan, A.Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *The International Conference On Learning Representations (ICLR)*, 2015.

[2] R.Selvaraiu, M.Cogswell, A.Das, R..Vedantam, D.Parikh, D.Batra “Grad-CAM: Visual explanations from deep networks via gradient-based localization”, *The IEEE International Conference onComputer Vision(ICCV)*, 2017

[3] “flickr”, https://www.flickr.com/



(a1) CNN_A　(a2) CNN_B

(b1) CNN_A　(b2) CNN_B

(c1) CNN_A　(c2) CNN_B

(d1) CNN_A　(d2) CNN_B

図 5 Grad-CAM による着目領域の可視化