局所特徴の共起表現の埋め込みと Fisher Vector を用いた詳細画像カテゴリ識別

中山 英樹[1,a)]

1. はじめに

詳細画像カテゴリ識別 (fine-grained visual categorization, FGVC) は，対象物体のドメインを限定する代わりに視覚的に極めて類似したカテゴリの多クラス識別を行うものであり，動植物の種類識別などが代表的な例として挙げられる [10], [12], [13]．これを可能とするためには，クラス間の微小な差異まで抽出できる強力な画像特徴量を用いる必要がある．我々は，教師付次元圧縮手法を用いて隣接する局所特徴量の共起表現を低次元へ圧縮し，これを新しい局所特徴量として Fisher Vector [11] のような最新の bag-of-visual-words (BoVW) ベース手法に用いるアプローチを提案した [8]．本手法は標準的なベンチマークで高い識別精度を得るとともに，ImageCLEF 2013 [5] の plant identification challenge (NaturalBackground task) において第 1 位となった．本稿では，手法の概要と実験結果の一部について報告を行う．

局所特徴の共起情報が識別に有効なことは広く知られている [3], [14] が，単純に共起をとるだけでは爆発的に次元数が大きくなる点が問題となる．最新の強力な BoVW ベース手法の多くは，最終的な画像特徴ベクトルの次元数が局所特徴量の次元数に比例する [6], [11] ため，両者を併用するためには，高次元の共起表現をできる限り判別的な情報を残しつつ圧縮する必要がある．

提案手法では，教師付次元圧縮手法により高次元の共起表現を一般的な局所特徴のサイズまで圧縮するが，この際に画像全体のカテゴリラベルを局所特徴レベルで共通に用いる近似的なアプローチをとる．これは，画像中の大部分の領域が対象となるカテゴリと関連していることを期待するものであり，一般的な画像識別問題においては必ずしも適切ではない．しかしながら FGVC においては，識別のためにユーザが対象物の接写や切り出しまで行う半自動の用途が多いと考えられ，この文脈では比較的妥当性のある前提であると言える．

[1] 東京大学 大学院情報理工学系研究科 創造情報学専攻 〒 113–0033 東京都文京区本郷 7–3–1
[a)] nakayama@ci.i.u-tokyo.ac.jp

2. 提案手法

画像中の各点 (x, y) における局所特徴を $\boldsymbol{v}_{(x,y)}$ と表記する．まず，空間的に隣接する局所特徴の要素同士の積を以下に示す $\boldsymbol{p}^c_{(x,y)}$ のように列挙し，共起情報を明示的にとりこむ．ここで，c は考慮する隣接局所特徴の数である．隣接局所特徴を用いない場合は，次のように自身の要素間の共起のみ扱う．

$$\boldsymbol{p}^0_{(x,y)} = \begin{pmatrix} \boldsymbol{v}_{(x,y)} \\ upperVec\left(\boldsymbol{v}_{(x,y)}\boldsymbol{v}^T_{(x,y)}\right) \end{pmatrix}. \tag{1}$$

ここで，$upperVec()$ は上三角行列の要素を全て列挙したベクトルを示す．左右二つの隣接局所特徴まで用いる場合は，次のようになる．

$$\boldsymbol{p}^2_{(x,y)} = \begin{pmatrix} \boldsymbol{v}_{(x,y)} \\ upperVec\left(\boldsymbol{v}_{(x,y)}\boldsymbol{v}^T_{(x,y)}\right) \\ Vec\left(\boldsymbol{v}_{(x,y)}\boldsymbol{v}^T_{(x-\delta,y)}\right) \\ Vec\left(\boldsymbol{v}_{(x,y)}\boldsymbol{v}^T_{(x+\delta,y)}\right) \end{pmatrix}. \tag{2}$$

ここで，δ はオフセットのパラメータ (本稿では 20 に固定)，$Vec()$ は行列の要素全てを列挙したベクトルを示す．同様にして、さらに多くの隣接局所特徴との共起を取り込むことができる．本稿では，高々 4 つまでを考慮する．このようにして得た共起表現ベクトル $\boldsymbol{p}$ は数千から数万次元に及ぶ．これに画像全体のラベルを教師として正準相関分析 (CCA) [4] を適用し，最も判別的な 64 次元を抽出する．これを新たな局所特徴と解釈し，BoVW や Fisher Vector に適用して最終的な画像特徴ベクトルを得る．

3. 実験

SIFT, C-SIFT, opponent-SIFT, self-similarity の 4 つの局所特徴記述子からそれぞれ提案手法による特徴ベクトルの抽出と識別器の学習を行う．各局所特徴は dense sam-

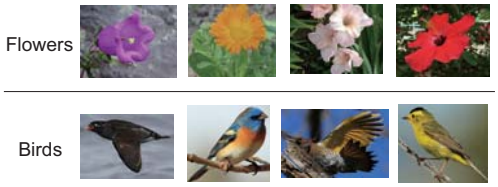

図 1 Images from FGVC benchmark datasets. Top: Oxford-Flower102 [10]. Bottom: Caltech-Bird200-2010 [12].

表 1 Classification performance on FGVC datasets (%).

	Flowers	Birds
4 desc. (PCA64)	81.6	23.9
4 desc. (Ours)	**87.2**	28.1
8 desc. (PCA64 + Ours)	85.7	**28.8**
Previous Work	85.6 [2]	28.2 [13]
	80.0 [7]	26.7 [2]
	76.3 [9]	26.4 [1]

pling により抽出し，PCA を用いて 64 次元へ圧縮する[*1]．これに提案手法を適用し，共起ベクトルを CCA により 64 次元へ圧縮し，埋め込み局所特徴を算出する．最終的に，これを Fisher Vector へ加工し，ロジスティック回帰により識別器を構築する．各識別器が出力するスコアの重み付平均により識別結果を決定する．

3.1 Oxford-Flower & Caltech-Bird データセット

まず，FGVC において標準的に用いられるベンチマークである，Oxford-Flower102 データセット [10] と Caltech-Bird200-2010 データセット [12] (図 1) を用いて評価を行う．それぞれ，102 クラスの花の画像，200 クラスの鳥の画像からなるデータセットであり，クラスごとの識別正解率の平均を評価指標とする．

表 1 に結果を示す．提案手法は，PCA により圧縮した局所特徴をそのまま用いる一般的な Fisher Vector のスコア (PCA64) を大きく改善させており，埋め込みの効果が表れていることが分かる．同時に，全ての先行研究を上回る高い識別精度を得る結果となった．

3.2 ImageCLEF'13 Plant Identification

ImageCLEF [5] はコンペティション型のワークショップであり，本年度の plant identification challenge では画像中の葉 (Leaf)・花 (Flower)・果実 (Fruit)・幹 (Stem)・全景 (Entire) の各部位から植物の種類を識別し，Mean Average Precision (MAP) により評価を行う (図 2)．ここでは，自然背景を前提としたより困難なタスクである NaturalBackground task について結果を報告する．コンペティションにおける最終的な評価結果を，第 2 位・第 3 位のチームとあわせ表 2 に示す．提案手法は，全体で第 1 位，5 つのサブカテゴリのうち 4 つで第 1 位となる良好な結果を得た．

[*1] Self-similarity 特徴を除く

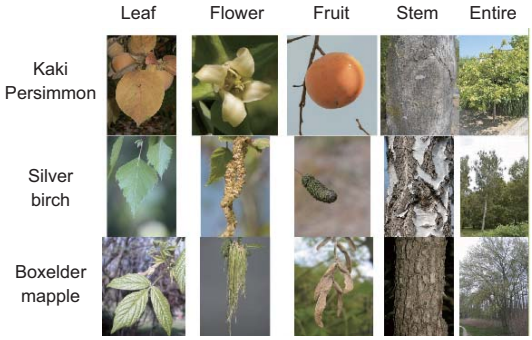

図 2 Images from ImageCLEF'13 plant identifiation challenge.

表 2 Results on ImageCLEF'13 plant identification challege (NaturalBackground task). MAP scores are evaluated.

	Leaf	Flower	Fruit	Stem	Entire	All
Ours	**0.275**	0.472	**0.311**	**0.253**	**0.297**	**0.393**
INRIA	0.272	**0.494**	0.260	0.240	0.274	0.385
Sabanci & Okan	0.049	0.223	0.194	0.106	0.174	0.181

参考文献

[1] Bo, L. and Fox, D.: Kernel descriptors for visual recognition, *Proc. NIPS* (2010).
[2] Chai, Y., Rahtu, E. and Lempitsky, V.: TriCoS: A tri-level class-discriminative co-segmentation method for image classification, *Proc. ECCV*, pp. 794–807 (2012).
[3] Harada, T. and Kuniyoshi, Y.: Graphical Gaussian vector for image categorization, *Proc. NIPS* (2012).
[4] Hotelling, H.: Relations between two sets of variants, *Biometrika*, Vol. 28, pp. 321–377 (1936).
[5] ImageCLEF 2013: http://imageclef.org/2013/.
[6] Jégou, H., Douze, M., Schmid, C. and Pérez, P.: Aggregating local descriptors into a compact image representation, *Proc. IEEE CVPR*, pp. 3304–3311 (2010).
[7] Lempitsky, V. and Zisserman, A.: BiCoS: A bi-level co-segmentation method for image classification, *Proc. IEEE ICCV*, pp. 2579–2586 (2011).
[8] Nakayama, H.: Augmenting descriptors for fine-grained visual categorization using polynomial embedding, *Proc. IEEE ICME* (2013).
[9] Nilsback, M.-E.: An automatic visual flora: segmentation and classification of flower images, PhD Thesis, University of Oxford (2009).
[10] Nilsback, M.-E. and Zisserman, A.: Automated flower classification over a large number of classes, *Proc. ICVGIP*, pp. 722–729 (2008).
[11] Perronnin, F., Sánchez, J. and Mensink, T.: Improving the Fisher kernel for large-scale image classification, *Proc. ECCV* (2010).
[12] Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S. and Perona, P.: Caltech-UCSD birds 200, Technical Report CNS-TR-2010-001, California Institute of Technology (2010).
[13] Yang, S., Bo, L., Wang, J. and Shapiro, L.: Unsupervised template learning for fine-grained object recognition, *Proc. NIPS* (2012).
[14] 山内悠嗣，山下隆義，藤吉弘亘：Boosting に基づく特徴量の共起表現による人検出，電子情報通信学会論文誌，Vol. J92-D, No. 1, pp. 1125–1134 (2009).