Guang-hui Song, Xiao-gang Jin, Gen-lang Chen, Yan Nie, 2016. Two-level hierarchical feature learning for image classification. *Frontiers of Information Technology & Electronic Engineering*, **17**(9):897-906.
http://dx.doi.org/10.1631/FITEE.1500346

Two-level hierarchical feature learning for image classification

Key words: Transfer learning, Feature learning, Deep convolutional neural network, Hierarchical classification, Spectral clustering

Corresponding author: Gang-jin Xiao
E-mail: xiaogangj@cise.zju.edu.cn
ORCID: http://orcid.org/0000-0002-7787-7228

Motivation

- At present, image datasets have a growing sample size and image category in the real world. The similarity is different among different categories, with some categories being more difficult to distinguish than others.
- To distinguish highly similar categories, more specific features are required so that the classifier can improve the classification performance.
- Adopting the hierarchical classification and feature learning methods to solve the above problem is natural. However, so far, only limited studies have focused on how to combine deep feature learning with hierarchical classification to improve classification accuracy.

Main idea

- Inspired by the idea of transfer learning, we have come up with two questions: whether the general feature extractor can be adequate for distinguishing the categories with high similarity; how to extract more specific features using the feature learning method to improve the classification performance.
- The general features are extracted from the general network model, and the specific features are extracted from the corresponding specific network model.
- Our proposed method effectively increases the classification accuracy in comparison with flat multiple classification methods.

Method

1. The deep features of each image consist of two parts, namely general features and specific features. They are also called two-level hierarchical features.
2. Three deep CNN models pre-trained on ImageNet, AlexNet, CaffeNet, and VGG-16 net are used as the base models for transfer learning.
3. Experiments using the Caltech-256, Oxford Flower-102, and Tasmania Coral Point Count datasets demonstrate that the expression ability of the deep features resulting from two-level hierarchical feature learning is powerful.

Major results

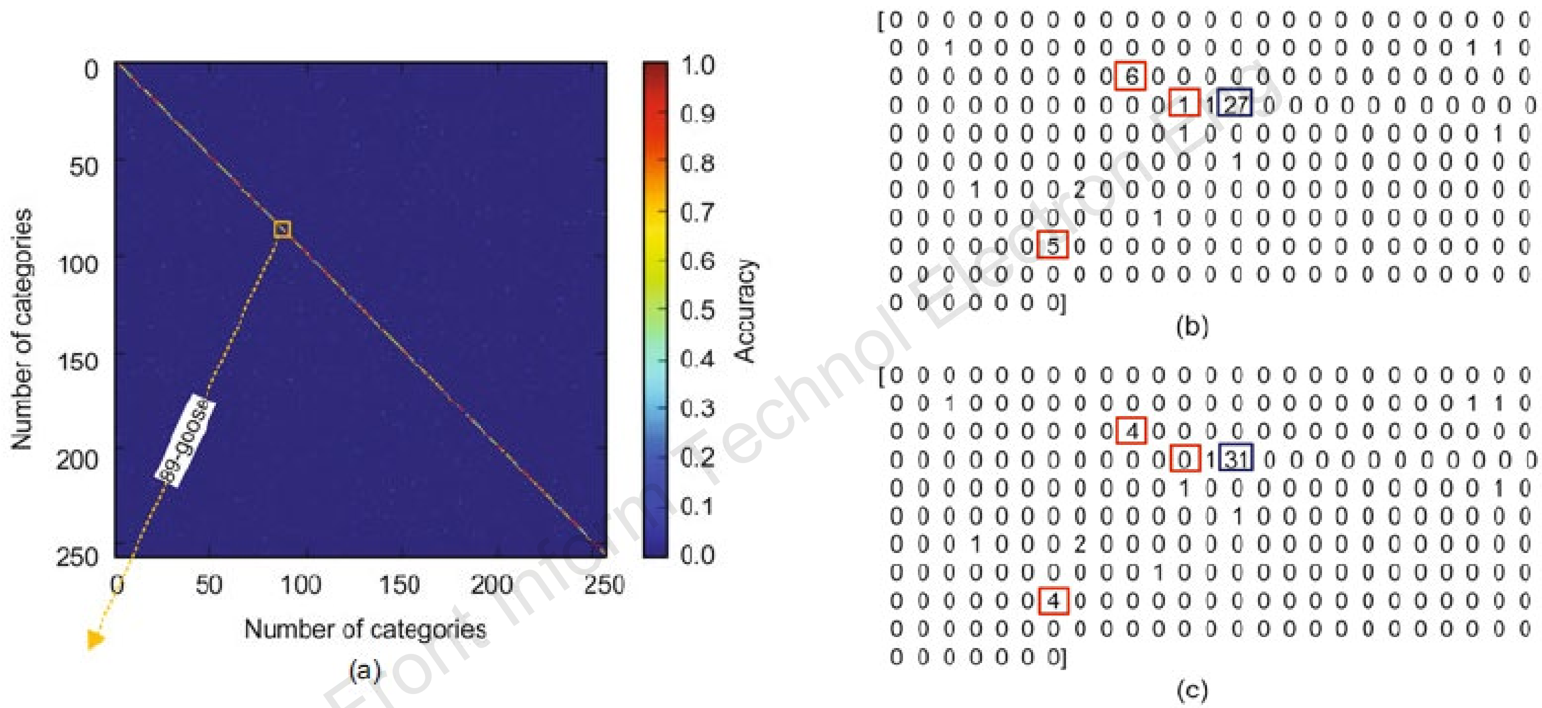

Fig. 3 The confusion matrix showing that the highly similar categories are easily misclassified: (a) the confusion matrix is generated by the result of pre-classification; (b) the classification result of category 89-goose is extracted from the confusion matrix (the number with a blue box indicates that the sample number is classified correctly and the number with a red box indicates that the sample number is misclassified); (c) the classification result of category 89- goose is extracted from the confusion matrix after using our proposed method (the sample number of misclassification among the highly similar categories is significantly reduced).

Major results

Table 1 Comparison of the performance between our two-level HFL method with optimal k value and two baseline methods on the Caltech-256 dataset

Method	acc (%)	mAP (%)
AlexNet-base (fc6)	73.17	70.87
AlexNet-ft (fc6)	73.49	71.20
AlexNet-base (fc7)	73.54	71.57
AlexNet-ft (fc7)	74.03	72.08
AlexNet-HFL (fc7, k=4)	**74.65**	**72.56**
CaffeNet-base (fc6)	73.69	71.72
CaffeNet-ft (fc6)	74.19	72.30
CaffeNet-base (fc7)	73.80	72.16
CaffeNet-ft (fc7)	74.18	72.58
CaffeNet-HFL (fc7, k=4)	**74.66**	**72.92**
VGG-16 net-base (fc6)	81.31	80.77
VGG-16 net-ft (fc6)	81.90	81.37
VGG-16 net-base (fc7)	79.41	79.69
VGG-16 net-ft (fc7)	80.70	80.61
VGG-16 net-HFL (fc6, k=4)	**82.45**	**81.90**

acc: classification accuracy on all images; mAP: mean accuracy per category. Bold numbers represent the optimal values in the corresponding test

Major results

- Comparison of different feature extraction methods on the Tasmania Coral Point Count datasets.

Table 3 Comparison of classification performance based on deep features and conventional hand-designed features in various sizes patches

Method	Size (pixels)	mAP (%)
PCA	63×63	80.13
GLCM	95×95	74.82
LBP	31×31	71.35
CaffeNet-ft (fc6)	60×60	89.90
	80×80	90.59
	100×100	91.24
	120×120	**92.11**
	140×140	91.51
	160×160	91.55

The mAP averaging over five data splits is used as the measurement scale. PCA: principal component analysis; GLCM: gray-level co-occurrence matrix; LBP: local binary pattern. mAP: mean accuracy per category. The bold number represents the optimal value

Table 4 Performance of our proposed method with different k values (patch size: 120×120 pixels)

Method	mAP (%)
CaffeNet-ft	92.11
CaffeNet-HFL (fc6, k=2)	**93.76**
CaffeNet-HFL (fc6, k=3)	92.52
CaffeNet-HFL (fc6, k=4)	92.02

mAP: mean accuracy per category. The bold number represents the optimal value

Conclusions

- We proposed a two-level HFL framework based on transfer learning to solve the misclassification problem in highly similar categories.
- The specific deep features are gradually learned using the two-level HFL method to improve the classification performance.
- In the future, we will conduct the experiments with a large-scale dataset to verify the adaptability of the proposed method. We will also explore a more suitable clustering approach for highly similar categories to further improve the classification accuracy.