Image Classification by Hierarchical Spatial Pooling with Partial Least Squares Analysis

Jun Zhu[1]
junnyzhu@sjtu.edu.cn
Weijia Zou[1]
zouweijia@sjtu.edu.cn
Xiaokang Yang[1]
xkyang@sjtu.edu.cn
Rui Zhang[1]
zhang_rui@sjtu.edu.cn
Quan Zhou[2]
qzhou.lhi@gmail.com
Wenjun Zhang[1]
zhangwenjun@sjtu.edu.cn

[1] Institute of Image Communication and Information Processing
Shanghai Jiao Tong University
Shanghai, China

[2] Department of Electronics and Information Engineering
Huazhong University of Science and Technology
Wuhan, China

In recent image classification systems, spatial pooling is a key step to form image-level representation from patch-level local features. It captures meaningful statistical information of local feature codes over different ROIs, and achieves certain spatial invariance property to facilitate classification. On the spatial representation model, although the spatial pyramid is predominately used in image classification literature, its rigid structure may limit the resultant image representation from exploring richer spatial statistical information further.

Based on the tangram model, we construct a hierarchical ROI dictionary (called by HRD in this paper for short) for spatial pooling. Compared to rigid spatial pyramid model, it assembles the ROIs with more shape types, locations and scales, and is capable of retaining richer spatial statistical information. Besides, by taking advantage of mutual compositionality among ROIs, HRD can be inherently organized into a directed acyclic graph, and this derives an efficient hierarchical algorithm to facilitate spatial pooling. Besides, we further employ partial least squares (PLS) analysis for dimensionality reduction on the pooled features. It can capture the statistical relationship between pooled features and class labels for different visual words, and learn a more compact and discriminative image-level representation for classification. The experimental results demonstrate superiority of the proposed hierarchical pooling method w.r.t. spatial pyramid, on three benchmark datasets (Caltech-101, Caltech-256 and Scene-15) for image classification.

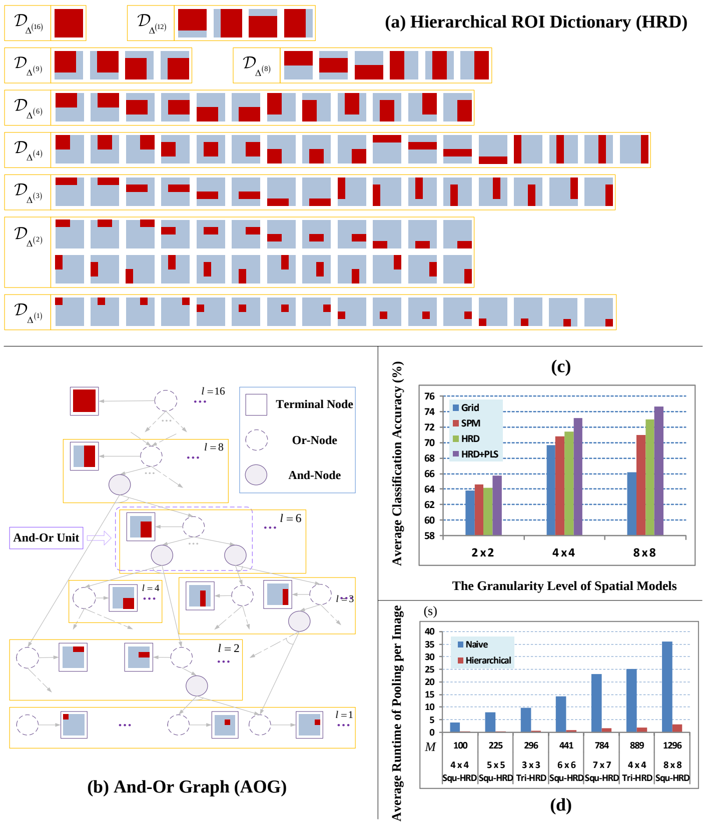

Figure 1: (a) Illustration on a 16-layer HRD. (b) Illustration of the associated AOG (only a portion of graph is shown for clarity). (c) Performance comparison on different spatial representation models (Caltech-101). (d) Comparison on runtime of spatial pooling algorithm.

1 Hierarchical ROI dictionary

In this paper, a layered dictionary of shape primitives (called *tans*) is constructed to quantize the spatial configuration space. Formally, the tan dictionary $\Delta = \bigcup_{l=1}^{L} \Delta^{(l)}$ is an union of L subsets. $\Delta^{(l)} = \{B_{(l,i)} \mid i = 1,2,\cdots,N_l\}$ denotes a set of tans for the l^{th} layer, where $B_{(l,i)}$ refers to the i^{th} tan. When placing each tan onto different locations in the image lattice, one tan $B_{(l,i)}$ may produce a set of multiple instances $\{\Lambda_{(l,i,j)} \mid j = 1,2,\cdots,J_{(l,i)}\}$, which makes up a HRD $\mathcal{D}_{\Delta}$ for spatial pooling:

$$\begin{aligned} &\mathcal{D}_{\Delta} = \textstyle\bigcup_{l=1}^{L} \mathcal{D}_{\Delta^{(l)}}, \\ \forall l, \quad &\mathcal{D}_{\Delta^{(l)}} = \{\Lambda_{(l,i,j)} \mid i = 1,2,\cdots,N_l \text{ and } j = 1,2,\cdots,J_{(l,i)}\}. \end{aligned}$$

Moreover, to describe the compositionality among tans, an associated And-Or graph (AOG) is accordingly built for organizing the ROIs in a deep hierarchy. The And-node represents that a tan can be composed by two smaller ones in layer belows, while the Or-node implies that it can be generated in alternative ways of shape composition. Fig.1(a) shows a 16-layer HRD for 4×4 grid, with an associated AOG illustrated in Fig.1(b).

2 Efficient Spatial Pooling in Deep Hierarchy

Based on the HRD, we can perform spatial pooling operation over the ROIs. Due to the over-completeness and increasing degree of freedom induced by recursive shape composition, the cardinality of HRD grows drastically with its granularity level so that direct spatial pooling operation on HRD is computationally demanding. However, the over-completeness and compositionality of the ROIs result in that each ROI in the HRD can be exactly composed by its child ones in the layers below. Considering the directed acyclic structure of associated AOG with deep hierarchy, we present an efficient algorithm for spatial pooling on HRD. Given a HRD and its associated AOG, the proposed pooling algorithm can be divided into two steps: I. For each ROI at the first layer, we directly compute its pooled feature from codes; II. For the other layers above, the pooled feature for each ROI is bottom-up propagated from its child nodes. Thus, most computational cost can be saved by taking advantage of recursive compositionality among the ROIs.

3 Learning Image Representation with PLS Analysis

Although the pooled features can be directly used for classification, for a HRD with large number of ROIs, it may produce a huge number of variables, which tend to be highly correlated and redundant. To obtain a more compact and discriminative image-level representation, we learn a PLS model for each visual word individually, to preserve class-specific discriminative information in the extracted representation. Finally, we perform dimension reduction via projecting pooled features onto the learned subspaces, obtaining a new image-level representation for classification.