# Group-Sensitive Multiple Kernel Learning for Object Categorization

Jingjing Yang[1, 2, 3, *], Yuanning Li[1, 2, 3, *], Yonghong Tian[3], Lingyu Duan[3], Wen Gao[3, 1]
[1]Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, China
[2]Graduate University, Chinese Academy of Sciences, China
[3]National Engineering Laboratory for Video Technology, School of EE & CS, Peking University, China
{jjyang, ynli}@jdl.ac.cn, {yhtian, lingyu, wgao}@pku.edu.cn

## Abstract

*In this paper, we propose a group-sensitive multiple kernel learning (GS-MKL) method to accommodate the intra-class diversity and the inter-class correlation for object categorization. By introducing an intermediate representation "group" between images and object categories, GS-MKL attempts to find appropriate kernel combination for each group to get a finer depiction of object categories. For each category, images within a group share a set of kernel weights while images from different groups may employ distinct sets of kernel weights. In GS-MKL, such group-sensitive kernel combinations together with the multi-kernels based classifier are optimized in a joint manner to seek a trade-off between capturing the diversity and keeping the invariance for each category. Extensive experiments show that our proposed GS-MKL method has achieved encouraging performance over three challenging datasets.*

## 1. Introduction

Recently, various learning approaches have been developed to improve object categorization [1, 2, 3, 4, 8, 11, 20]. Significant improvements have been achieved over several public datasets, such as Caltech, Pascal VOC and ImageCLEF. However, object categorization is still a challenging task. The essential reason lies in that the images within a category would exhibit diversity while the images from distinct categories would produce correlations in low-level visual attributes (e.g. color, texture, and shape). We may refer to such phenomena as "intra-class diversity" and "inter-class correlation" in this study.

Fig.1 illustrates the example images from WikiPediaMM dataset [9]. Given the category "bridges", positive samples can be grouped into three sub-categories, each of which produces distinct visual appearance. On the other hand, negative samples from other categories (e.g., "buildings", "cities by night") may exhibit similar visual attributes to some samples of "bridge". We argue that, to elegantly and robustly categorize objects over extensive image data-sets, it is meaningful to effectively model both intra-class diversity and inter-class correlation.

*First two authors contributed equally to this work.

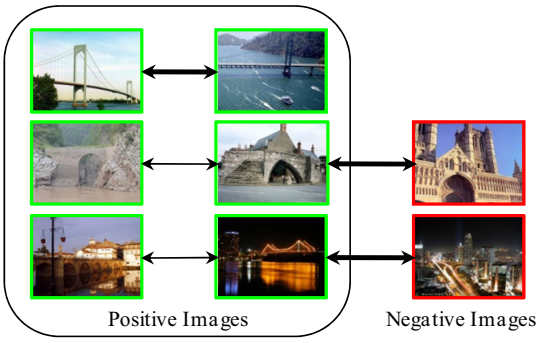


Fig. 1. Illustration of intra-class diversity and inter-class correlation of object "bridges". Double-headed arrows stand for the visual correlation between images. A thicker arrow indicates stronger visual correlation.

Various classifiers based on distance metrics (e.g. [17]) or kernels (e.g. [3, 5]), which aim to maximize inter-class distance (or interval), have been applied to object categorization. In particular, multiple kernel learning (MKL) methods [10, 22, 30] have shown great advantages in this task recently (e.g. [5, 18]). Instead of using a single kernel in support vector machine (SVM) [21], MKL learns an optimal kernel combination and the associated classifier simultaneously, providing an effective way of fusing informative features and kernels. However, these methods basically adopt a uniform similarity measure over the whole input space. When a category exhibits high variation as well as correlation with other categories in appearance, they are difficult to cope with the complexity of data distribution.

On the other hand, several sample-based methods [11, 19, 20, 29] have been proposed to capture the characteristics of individual samples. For example, a sample-specific ensemble kernel learning method is proposed in [29] to explore the relative contributions of distinct kernels for each sample. In practice, such methods have yielded promising discriminative power. But expensive computation is incurred to learn sample-based similarity measures. More importantly, heavily respecting individual samples may overwhelm the intrinsic properties of a category so as to make the classifier less reliable.

In this paper, we attempt to introduce an intermediate



2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

representation "*group*" between object categories and individual images to seek a trade-off between capturing the diversity and keeping the invariance for each category in training classifiers. Given an object category, the image samples with similar visual appearance are clustered into a group so that the intra-class diversity can be represented by a set of groups. On the other hand, inter-class correlation can be represented by the correlation between the groups from different categories.

Consequently, we incorporate group into the MKL framework and propose a group-sensitive multiple kernel learning (GS-MKL) method for object categorization. In GS-MKL, the image-to-image similarity is represented as a weighted combination of multi-kernels, where the kernel weights not only depend on the corresponding kernel functions, but also on the groups that two comparing images belong to. Instead of a uniform or sample-specific similarity measure, such group-sensitive similarity measure is shown effective in dealing with both intra-class diversity and inter-class correlation.

In GS-MKL, the group-sensitive kernel weights together with the associated classifier are jointly optimized by a gradient descent wrapping canonical SVM solver [30]. Over three datasets (i.e., Caltech101, Pascal VOC2007 and WikipediaMM), we have shown that GS-MKL can significantly alleviate the negative effects of intra-class diversity and inter-class correlation, coming up with a more robust discriminative power for object categorization.

Our main contributions are summarized as follows:

- We have proposed a group-sensitive multiple kernel learning method GS-MKL for robust object categorization, where both intra-class diversity and inter-class correlation are taken into account.
- We formulate GS-MKL in a general and flexible learning framework. When the group number declines to one, GS-MKL is reduced to canonical MKL. When the group number reaches up to the number of training images, GS-MKL becomes a sample-specific MKL.
- We have achieved promising results comparable to the state-of-the-art results on Caltech101 and Pascal VOC2007, and significant improvements over canonical MKL across three datasets.

The remainder of this paper is organized as follows. In Section 2 we brief the related work. In Section 3, we introduce the GS-MKL framework for object categorization. The GS-MKL learning algorithm is presented in Section 4. We present the experimental results in Section 5. Finally we conclude this paper in Section 6.

## 2. Related Works

In computer vision, many research efforts have been devoted to characterizing visual statistics for a number of object categories in the past decades. Kernel based method is one of attractive research areas for object categorization in recent years. Diverse kernels such as pyramid matching kernel (PMK) [15], spatial pyramid matching kernel (SPK) [3], proximity distribution kernel (PDK) [16] and chi-square kernel [33] are delicately designed to compute the similarity of image pair on certain features that represent particular visual characteristics.

Recently, multi-kernel based classifiers have been introduced into object categorization yielding promising results. In [5, 18], multiple features (e.g., appearance, shape) are employed and kernels (e.g., PMK and SPK with different hyper-parameters) are linearly combined in MKL framework. Like the canonical MKL [10], these methods adopt a uniform kernel combination strategy over the whole input space. However, in the presence of significant intra-class diversity and inter-class correlation, they may be difficult to deal with complex data distribution and suffer a degraded performance.

More recently, sample-specific MKL methods are proposed in [23] by adopting a sample-specific kernel weighting strategy. The basic idea is that kernel weights not only depend on the kernel functions but also on the samples. Compared with canonical MKL, a sample-specific MKL tend to reflect the relative importance of different kernels at each sample rather than at the level of object category. Despite of some performance improvements, learning too many parameters may lead to the expensive computation as well as the high risk of over-fitting. Although our proposed GS-MKL and other methods [5, 18, 23] reviewed above are all extended from MKL framework, GS-MKL provides a mechanism of evaluating multi-kernels over groups (sub-categories).

In addition, GS-MKL is different from classifier ensemble methods which train multiple classifiers separately using different data subsets or features and then combine the classifiers to obtain better performance. Although our method also partitions training data into groups, GS-MKL learns a single classifier based on the group-sensitive kernel combinations which adapt with the local data distributions of object sub-categories. Also, GS-MKL couples feature/kernel weighting and classifier leaning in a joint optimization problem.

## 3. Group-Sensitive MKL Framework

Let $D_L = \{x_i, y_i\}_{i=1}^N$ be a training image dataset, where $x_i$ denotes the $i^{th}$ sample and $y_i = \{\pm 1\}$ denotes the binary label for a given object category, and $N$ is the number of training samples. Based on the labeled dataset $D_L$, we aim to train a multi-kernels based classifier with a decision function $f(x)$ to predict the object category of an unlabeled image $x$.

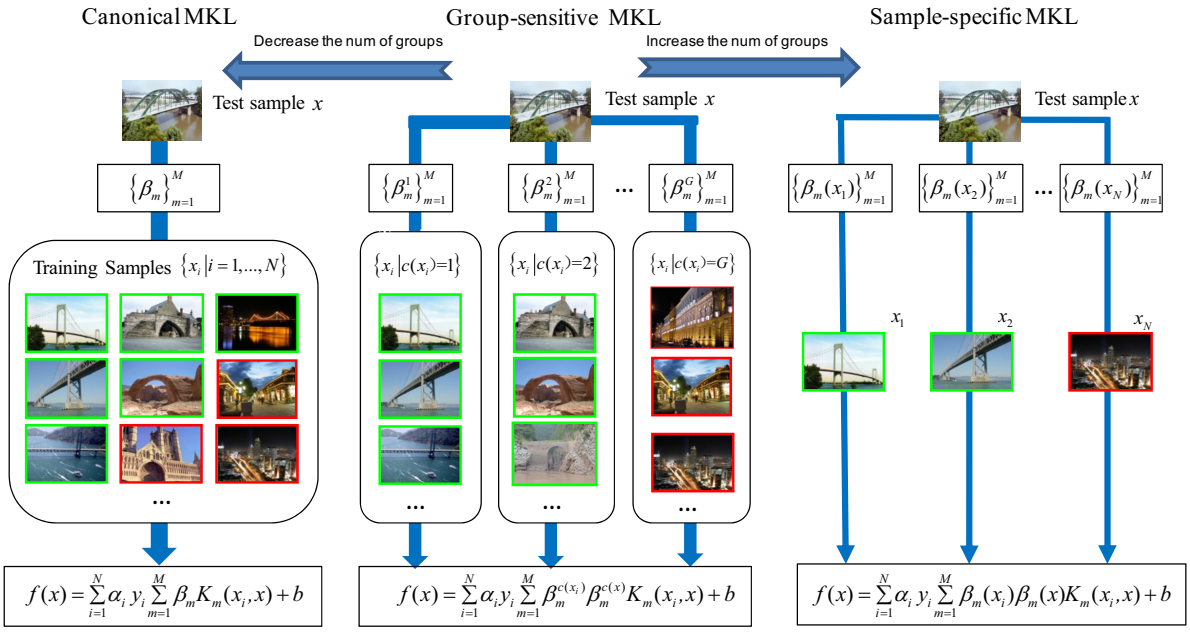


Fig. 2.Three paradigms of object categorization using (a) Canonical MKL, (b) Group-sensitive MKL, and (c) Sample-specific MKL. In the figure, images with green bounding boxes are positive samples while those with red bounding boxes are negative samples for "bridge". Note that sample-specific MKL will learn two sets of kernel weights even for two images with quite similar appearance (e.g. $x_1$and $x_2$).

### 3.1. Canonical MKL

SVMs have been proven to be efficient tools for solving classification problems. However, the discriminative power of SVMs heavily relies on kernel selection which is generally accomplished by cross-validation. Instead of selecting a single kernel, MKL [10] learns a convex kernel combination and the associated classifier simultaneously. The combination of multi-kernels is defined as follows:

$$K(x_i,x)=\sum_{m=1}^{M}\beta_m K_m(x_i,x) \tag{1}$$

with $\sum_{m=1}^{M}\beta_m=1$ and $\beta_m\geq 0\ \forall m$, where $M$ is the total number of kernels, $K_m$ is a positive definite kernel associated with a reproducing kernel Hilbert space (RKHS), and $\{\beta_m\}_{m=1}^M$ are kernel weights which are optimized during training. Each $K_m$ can employ different kernel functions and use different feature subsets or data representations.

For binary classification, the decision function of canonical MKL is given as follows:

$$f(x)=\sum_{i=1}^{N}\alpha_i y_i \sum_{m=1}^{M}\beta_m K_m(x_i,x)+b, \tag{2}$$

where $\{\alpha_i\}_{i=1}^N$ and $b$ are the coefficients of the classifier, corresponding to the Lagrange multipliers and the bias in the canonical SVM problem. In MKL, the coefficients $\alpha_i$ and the kernel weights $\beta_m$ can be learnt in a joint optimization problem (details can be found in [30]).

### 3.2. GS-MKL

As shown in Fig. 2a, canonical MKL employs a uniform kernel combination over the whole input space. Instead of learning a global kernel combination, GS-MKL learns a set of group-sensitive kernel combinations to adapt with the complexity of data distribution.

As shown in Fig. 2b, images from the same category are clustered into different groups by a pre-process (see Sec. 5.3 for details). Then the kernel weights in GS-MKL not only depend on the kernel functions, but also on the groups that the two images belong to. Let $c(x)$ and $c(x_i)$ be the group ids of image $x$ and $x_i$ respectively. The combined kernel form in Eqn. 1 can be rewritten as:

$$K(x_i,x)=\sum_{m=1}^{M}\beta_m^{c(x_i)}\beta_m^{c(x)}K_m(x_i,x)\,, \tag{3}$$

where $\beta_m^{c(x)}$ and $\beta_m^{c(x_i)}$ are *group-sensitive kernel weights* of $x$ and $x_i$. Let $G$ denote the total group number, then $\beta_m^{c(x)}\in\{\beta_m^1,...,\beta_m^g,...,\beta_m^G\}$ for $m\in(1,...,M)$. Accordingly, the decision function in Eqn. 2 can be reformulated as:

$$f(x)=\sum_{i=1}^{N}\alpha_i y_i \sum_{m=1}^{M}\beta_m^{c(x_i)}\beta_m^{c(x)}K_m(x_i,x)+b, \tag{4}$$

where the coefficients $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_N]^T$ and bias $b$ have similar meanings as in canonical MKL. This decision function can be derived from the GS-MKL primal problem in Sec. 4.1. Compared with $M$ kernel weights in the canonical MKL case, the number of group-sensitive kernel

weights gets rise up to $G \times M$. The coefficients and the group-sensitive kernel weights are optimized in a joint manner, which will be shown in Sec. 4.

### 3.3. Connection with Other MKL Methods

In this part, we show that GS-MKL can be generalized to canonical MKL and sample-specific MKL.

In the special case of $G = 1$, all samples belong to one group and share a unique set of kernel weights $\{\beta_m^1\}_{m=1}^M$. In this case, the GS-MKL is simplified to canonical MKL (see Fig. 2a), where $\beta_m^1$ in Eqn. 4 equals to the square root of $\beta_m$ in Eqn. 2.

In the case of $G = N$, each training sample belongs to an individual group and thus an sample-specific kernel weighting strategy is employed . In this way, $\beta_m^{c(x)}$ only depends on the kernel function and the sample $x$. We also note that $\beta_m^{c(x)}$ is equivalent to $\beta_m(x)$ in Localized MKL (LMKL) [23]. Correspondingly, the decision function is:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i \sum_{m=1}^{M} \beta_m(x_i)\beta_m(x) K_m(x_i, x) + b, \quad (5)$$

The number of group-sensitive kernel weights then reaches up to $N \times M$ where $N \gg G$. In this case, GS-MKL scales up to sample-specific MKL (see Fig. 2c).

## 4. Learning GS-MKL Based Classifier

### 4.1. The GS-MKL Primal Problem

In GS-MKL, sample $x$ is transformed via mappings $\{\phi_m(x) \mapsto \mathbb{R}^{d_m}\}_{m=1}^M$ from the input space into $M$ feature spaces $(\phi_1(x), \cdots, \phi_M(x))$, where $d_m$ denotes the dimensionality of the $m^{th}$ feature space. Each feature map is associated with a weight vector $\mathbf{w}_m$. To allow the combination of kernels as expressed by Eqn. 3, the decision function of canonical MKL in Eqn. 2 can be rewritten as follows:

$$f(x) = \sum_{m=1}^{M} \beta_m^{c(x)} \langle \mathbf{w}_m, \phi_m(\mathbf{x}) \rangle + b. \quad (6)$$

Inspired by SVM [21], training can be implemented by solving the following optimization problem, which maximizes the margin between positive and negative classes as well as minimizes the classification error.

$$\min_{\mathbf{w}_m, b, \xi, \beta} \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{w}_m\|^2 + C \sum_{i=1}^{N} \xi_i,$$
$$s.t.\ y_i \Big( \sum_{m=1}^{M} \beta_m^{c(x_i)} \langle \mathbf{w}_m, \phi_m(x_i) \rangle + b \Big) \geq 1 - \xi_i \ \forall i\ , \quad (7)$$
$$\xi_i \geq 0 \ \forall i$$

$\|\mathbf{w}_m\|^2$ is a regularization term which is inversely related to margin, $\sum_{i=1}^{N} \xi_i$ measures the total classification error, and $C$ is the misclassification penalty. The optimal $C$ can be obtained by cross-validation.

### 4.2. The GS-MKL Dual Problem

Through introducing Lagrange multipliers $\{\alpha_i\}_{i=1}^N$ into the above inequalities constraint in Eqn. 7, and formulating the Lagrangian dual function which satisfies the Karush-Kuhn-Tucker(KKT) condition [10], the former optimization problem reduces to a max-min problem as follows:

$\max_{\boldsymbol{\beta}} \min_{\boldsymbol{\alpha}} J$, where

$$J = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \Big( \sum_{m=1}^{M} \beta_m^{c(x_i)} \beta_m^{c(x_j)} K_m(x_i, x_j) \Big) - \sum_{i=1}^{N} \alpha_i, \quad (8)$$
$$\text{s.t.} \sum_{i=1}^{N} \alpha_i y_i = 0,\ 0 \leq \alpha_i \leq C\ \forall i,$$

This max-min problem is the GS-MKL dual problem. $J$ is a multi-object function for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. When $\boldsymbol{\beta}$ is fixed, minimizing $J$ over the coefficient $\boldsymbol{\alpha}$ is meant to minimize the global classification error and maximize the margin between positive and negative classes. When $\boldsymbol{\alpha}$ is fixed, maximizing $J$ over the group-sensitive kernel weights $\boldsymbol{\beta}$ is meant to maximize the intra-class similarity and minimize the inter-class similarity simultaneously.

### 4.3. Optimization Algorithm

Similar to the parameter learning in canonical MKL, we adopt a two-stage alternant optimization approach.

*4.3.1 The computation of $\boldsymbol{\alpha}$ given $\boldsymbol{\beta}$*

Fixing $\boldsymbol{\beta}$, the classifier coefficient $\boldsymbol{\alpha}$ can be estimated by minimizing $J$ under the constraint $0 \leq \alpha_i \leq C,\ \forall i$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$. Minimization of $J$ is identical to solve the canonical SVM dual problem with the kernel combination in Eqn. 3. Consequently, minimizing $J$ over $\boldsymbol{\alpha}$ can be easily implemented as there exist several efficient SVM solvers.

*4.3.2 The computation of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$*

To optimize the group-sensitive kernel weights $\boldsymbol{\beta}$ with a fixed $\boldsymbol{\alpha}$, the objective function can be expressed as:

$$J(\boldsymbol{\beta}) = \sum_{g=1}^{G} \sum_{g'=1}^{G} \sum_{m=1}^{M} \beta_m^g \beta_m^{g'} S_m^{gg'}(\alpha) - \sum_{i=1}^{N} \alpha_i, \quad (9)$$

where

$$S_m^{gg'}(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{\{i | c(x_i) = g\}} \sum_{\{j | c(x_j) = g'\}} \alpha_i y_i \alpha_j y_j K_m(x_i, x_j). \quad (10)$$

When $G$=1, $S_m^{gg'}$ corresponds to $S_k(\boldsymbol{\alpha})$ in canonical MKL [22]. When $G > 1$, the samples within a group have the

same label ( $\{\pm 1\}$ ) based on the assumption that the intermediate representation *group* is introduced to capture the locality of each sub-category. In this case, $S_m^{gg'}$ stands for the correlation of group $g$ and $g'$ over the $m^{th}$ kernel function. When $g$ and $g'$ have the same label, maximizing $J$ over $\boldsymbol{\beta}$ is to maximize the intra-class similarity. When $g$ and $g'$ have different labels, maximizing $J$ over $\boldsymbol{\beta}$ is to minimize the inter-class similarity. Correspondingly, the optimization of $J$ over $\boldsymbol{\beta}$ can be rewritten as:

$$\max_{\boldsymbol{\beta}} \sum_{g=1}^{G}\sum_{g'=1}^{G}\sum_{m=1}^{M} \beta_m^g \beta_m^{g'} S_m^{gg'}(\boldsymbol{\alpha}) \tag{11}$$

Note that the optimization problem in Eqn. 11 is not convex. Inspired by [23], instead of solving $\boldsymbol{\beta}$ directly, we use a normalized exponential weighting function to approximate the nonnegative $\boldsymbol{\beta}$ . Particularly, $\boldsymbol{\beta}$ is determined by statistical property of the group and the parameters of the function which are also learned from data. In this paper, such weighting function is defined as:

$$\beta_m^g = \frac{\exp(a_m^g K_m^g + b_m^g)}{\sum_{m'=1}^{M} \exp(a_{m'}^g K_{m'}^g + b_{m'}^g)}, \tag{12}$$

where $a_m^g$ and $b_m^g$ are the parameters of the function, and $K_m^g$ corresponds to a certain statistical property for the $g^{th}$ group over the $m^{th}$ kernel function. Let $n_g$ be the number of samples in the $g^{th}$ group. In this paper, we define $K_m^g$ as:

$$K_m^g = \frac{\sum_{\{i|c(x_i)=g\}}\sum_{\{j|c(x_j)=g\}} K_m(x_i, x_j)}{n_g^2} \tag{13}$$

As stated in [31], $J(\boldsymbol{\beta})$ is differentiable if the SVM solution is unique. Such condition can be guaranteed by the fact that all kernel matrices are strictly positive definite. Thus, we take derivatives of $J(\boldsymbol{\beta})$ w.r.t. $a_m^g, b_m^g$, and use gradient-descent method to train the weighting function:

$$\frac{\partial J(\boldsymbol{\beta})}{\partial a_m^g} = 2\sum_{l=1}^{M}\left(\sum_{i=1}^{G}(\beta_l^i S_l^{ig}(\boldsymbol{\alpha}))\beta_m^g K_m^g(\delta_m^l - \beta_l^g)\right) \tag{14}$$

$$\frac{\partial J(\boldsymbol{\beta})}{\partial b_m^g} = 2\sum_{l=1}^{M}\left(\sum_{i=1}^{G}(\beta_l^i S_l^{ig}(\boldsymbol{\alpha}))\beta_m^g (\delta_m^l - \beta_l^g)\right) \tag{15}$$

where $\delta_m^l$ is 1 if $l = m$ and 0 otherwise. After updating the parameters of the weighting function, we get a new $\boldsymbol{\beta}$ and then solve a single kernel SVM as in Sec. 4.3.1.

*4.3.3 Summarization of GS-MKL optimization process*

The optimization algorithm of GS-MKL is summarized in Alg. 1. The termination criteria can be the consistency of $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ between two consecutive steps, or a predefined iteration upper bound.

In Alg. 1, the step size of each iteration, $\gamma^{(t)}$ and $\lambda^{(t)}$, can be fixed as a small constant or determined with a line search method which needs additional canonical SVM optimizations for better convergence. Optimizing the classifier coefficients and group-sensitive kernel weights is a gradient descent wrapping canonical SVM solvent process. Note that the proposed algorithm does not guarantee convergence to global optimum and the initial parameters $a_m^g$ and $b_m^g$ may affect the solution quality.

Algorithm 1: GS-MKL Optimization Process

1: Initialize $a_m^g$ and $b_m^g$ with small random numbers for $g = 1,...,G$ and $m = 1,...,M$ .

2: **while** the termination criterion is not met **do**

3: Calculate kernel weights $\boldsymbol{\beta}$ as Eqn.12

4: Calculate $K(x_i, x_j) = \sum_{m=1}^{M} \beta_m^{c(x_i)} \beta_m^{c(x_j)} K_m(x_i, x_j)$

5: Solve $\boldsymbol{\alpha}$ using the canonical SVM with $K(x_i, x_j)$

6: $a_m^g \Leftarrow a_m^g + \gamma^{(t)} \frac{\partial J(\boldsymbol{\beta})}{\partial a_m^g}$ for $g = 1,...G$ and $m = 1,...M$

7: $b_m^g \Leftarrow b_m^g + \lambda^{(t)} \frac{\partial J(\boldsymbol{\beta})}{\partial b_m^g}$ for $g = 1,...G$ and $m = 1,...M$

8: **end while**

## 5. Experiments

In the experiments, we treat object categorization as the multi-class classification problem in the one-vs.-all setting. As we assume that no prior knowledge is available about the image data distribution, we empirically evaluate the optimal grouping strategy and determine a proper group number. And then we evaluate the performance of our proposed GS-MKL on three datasets.

### 5.1. Dataset

Extensive experiments are performed on Caltech101 [6], Pascal VOC2007 [7] and WikipediaMM [9] datasets. Caltech101 involves 102 object categories, where each category containing 31 to 800 images. Pascal VOC2007 consists of 20 object categories, where 2501 images taken in real-world are provided for training, 2510 for validation and 4952 for test. WikipediaMM dataset contains some 150,000 real-world web images from Wikipedia that cover 75 topics. In our experiment, 33 topics, each of which contains more than 60 positive samples, are employed. Note that some topics not only share similar visual appearances, but also produce semantic correlations, e.g., “house architecture” versus “gothic cathedral” and “military aircraft” versus “civil aircraft”. Compared with Caltech101, Pascal VOC2007 and WikiPediaMM exhibit higher intra-class diversity and inter-class correlation with more background clutter but less alignment.

## 5.2. Features and Kernels

Several feature descriptors are involved in our experiments. Two local appearance features (dense-color-SIFT (DCSIFT) and dense-SIFT (DSIFT) [3]), two shape features (self-similarity (SS) [32] and pyramid histogram of orientated gradients (PHOG) [33]), and one texture feature (Gabor feature) are used. In particular, DCSIFT is computed in CIE-lab 3-channels over a square patch of radius with the spacing of r. We take r = 4, 8 and 12 pixels to allow scalability. Likewise DSIFT and Gabor feature are calculated in gray channel. SS descriptor is used to capture a correlation map of a 5×5 patch with its neighbors at every 5th pixel. The correlation map is quantized into 10 orientations and 3 radial bins to form a 30 dim descriptor. We employ k-means to quantize these descriptors to obtain codebooks of size k (say, 400) respectively.

For PHOG, two spatial pyramid kernels of gradient orientation are calculated to measure the image similarity in shape. PHOG-180degree employs 20 orientation bins and PHOG-360degree uses 40 orientation bins. For the other feature descriptors, we implement two kernel functions (i.e., SPK [3] and PDK [16]). For SPK, an image is divided into cells and the features from the spatially corresponding cells are matched across two images. The resulting kernel is a weighted combination of histogram intersections from coarse cells to fine cells. A 4-level pyramid is used with the grid sizes of 8×8, 4×4, 2×2 and 1×1 respectively. For PDK, local feature distributions of the *K*-nearest neighbors are matched across two images. The resulting kernel combines the local feature distributions at multiple scales, e.g. $K = 1, \ldots, k$, where $k$ is set to (8, 16, 32) ranging from the finest to the coarsest neighborhood.

## 5.3. Sensitivity Study of Grouping

In our experiments, grouping is a pre-processing step for GS-MKL. We have tried out two methods, k-means and probabilistic latent semantic analysis (pLSA) [25], to cluster images from each category into groups. Other grouping methods can be utilized. There is no prior knowledge about the number of sub-classes in an object category and the optimal number of groups for GS-MKL. Hence, we empirically identify the optimal group numbers for three datasets. For each category, images are clustered into $N_g$ groups (from 1 to 5). For Caltech101 and WikipediaMM, 20 images are randomly selected for training and 10 images for validation to find the optimal group number $N_g$ for each object category. For Pascal VOC2007, 2501 training images and 2510 validation images are employed to find the optimal $N_g$ for each object category.

In Tab.1, we list the best categorization results over the validation set and the corresponding mean group number for k-means and pLSA on three datasets. Clearly, $N_g$ ranging from 2.3 to 4.4, relates to the intra-class diversity of the dataset. From the table, we can see that pLSA outperform k-means slightly over three datasets. As a generative method, pLSA does not need explicit distance measure, which seems more robust against the distance based method. In the following experiments, we employ pLSA to group images with the optimized $N_g$ derived from validation.

Table1. Comparison of two grouping methods

| Dataset | Grouping method | Best Performance | Mean $N_g$ Per category |
|---|---|---|---|
| Caltech101 | k-means | 78.6 | 2.5 |
| | pLSA | 80.4 | 2.3 |
| Pascal VOC2007 | k-means | 54.2 | 4.2 |
| | pLSA | 56.7 | 3.8 |
| WikipediaMM | k-means | 60.0 | 4.4 |
| | pLSA | 61.2 | 3.9 |

## 5.4. Experiment Results

### *5.4.1 Experiment on Caltech101*

In this set of experiments, we randomly select $N_{train}$ and $N_{test}$ images for training and test respectively, where $N_{train}$={10, 15, 20, 25, 30} and $N_{test}$=15. We compare our GS-MKL approach with several recent methods [2-4, 6, 11, 14, 18, 26, 27]. As shown in Fig. 3, GS-MKL has achieved promising results comparable to the top performances of the state-of-the-art methods.

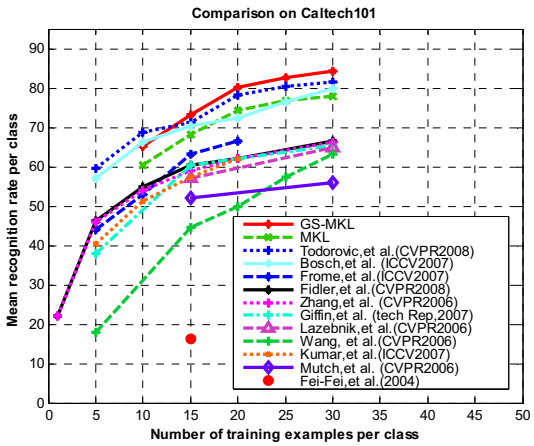


Fig. 3. Performance of GS-MKL and other recent methods on Caltech101dataset. GS-MKL: number of training samples (mean recognition rate), 10 (65.1), 15 (73.2), 20 (80.1), 25 (82.7), 30 (84.3).

From Fig. 3, we note that when $N_{train}$=10, GS-MKL obtains the performance of 65.1%, which is a bit lower than the best one (69.5%) [26]. This may attribute to the inefficacy of grouping methods when training samples are too sparse. Compared with other methods, GS-MKL has obtained better performance when $N_{train}$>10. When $N_{train}$ is

set to 30, the mean recognition rate reaches up to 84.3%, achieving a significant increase by 7.9% over canonical MKL (implemented as [30]). This shows that further optimization of kernel weights over groups yields better feature (kernel) combination for object categorization.

*5.4.2 Experiment on Pascal VOC 2007*

In this set of experiments, we employ 5011 images for training and 4952 for test respectively. Tab. 2 compares the performances of GS-MKL to canonical MKL and some other recently published methods [8, 28, 34, 35]. It is worthy of note that the approach INRIA_genetic [8] obtained the best performance in the Pascal VOC2007 challenge. The official performance metric Average Precision (AP) [7] is used to evaluate the performance.

Table 2. Average Precision of GS-MKL and other methods on the Pascal VOC 2007 dataset

| categories | [8] | [34] | [28] | [35] | MKL | GS-MKL |
|---|---|---|---|---|---|---|
| aero plane | 77.5 | 63.0 | 65.0 | 65.0 | 74.1 | **79.4** |
| bicycle | **63.6** | 22.0 | 44.3 | 48.0 | 53.9 | 62.4 |
| bird | 56.1 | 14.0 | 48.6 | 44.0 | 46.6 | **58.5** |
| boat | **71.9** | 42.0 | 58.4 | 60.0 | 62.2 | 70.2 |
| bottle | 33.1 | 43.0 | 17.8 | 20.0 | 37.5 | **46.6** |
| bus | 60.6 | 50.0 | 46.4 | 49.0 | 55.6 | **62.3** |
| car | **78.0** | 62.0 | 63.2 | 70.0 | 70.7 | 75.6 |
| cat | **58.8** | 32.0 | 46.8 | 49.0 | 48.4 | 54.9 |
| chair | 53.5 | 37.0 | 42.2 | 50.0 | 54.0 | **63.8** |
| cow | **42.6** | 19.0 | 29.6 | 32.0 | 34.7 | 40.7 |
| dining table | 54.9 | 30.0 | 20.8 | 39.0 | 50.1 | **58.3** |
| dog | 45.8 | 29.0 | 37.7 | 40.0 | 40.7 | **51.6** |
| horse | 77.5 | 15.0 | 66.6 | 72.0 | 76.6 | **79.2** |
| motorbike | 64.0 | 31.0 | 50.3 | 59.0 | 59.8 | **68.1** |
| person | 85.9 | 43.0 | 78.1 | 81.0 | 82.5 | **87.1** |
| potted plant | 36.3 | 33.0 | 27.2 | 32.0 | 38.3 | **49.5** |
| sheep | 44.7 | 41.0 | 32.1 | 35.0 | 40 | **48.8** |
| sofa | 50.6 | 37.0 | 26.8 | 42.0 | 48.2 | **56.4** |
| train | **79.2** | 29.0 | 62.8 | 68.0 | 68.1 | 75.9 |
| TV monitor | 53.2 | **62.0** | 33.3 | 49.0 | 47.2 | 54.4 |
| Mean AP | 59.4 | 36.7 | 44.9 | 50.2 | 54.5 | **62.2** |

The mean AP of GS-MKL is 62.2%, which is better than that of [8, 28, 34, 35]. GS-MKL has obtained the best results for 13 out of 20 categories. Over 10% improvements are obtained for two categories (i.e., "chair" and "potted plant"). Such results show the advantage of GS-MKL in handling the intra-class variation on real world image data. Under the same experimental setting, GS-MKL obtains better results for all 20 categories and 14.1% improvement on MAP against canonical MKL. This demonstrates that GS-MKL has better discriminative power than MKL by taking into account the intra-class diversity.

*5.4.3 Experiment on WikipediaMM*

On WikipediaMM dataset, we further evaluate four typical multi-kernels based methods, i.e., unweighted multiple kernel (UMK) (equal kernel weights for multi-kernels), canonical MKL, sample-specific MKL (SS-MKL) (implemented as [23]) and our proposed GS-MKL. For each image category, $N_{train}$={10, 15, 20, 25, 30} images are randomly picked out for training and the remaining images for test.

The results of five runs are shown in Tab. 3. We can see that GS-MKL outperforms three other multi-kernels based methods significantly on the dataset of real web images. Compared with UMK and MKL, GS-MKL obtains different degrees of improvements. Such results may be attributed to the ability of GS-MKL in adapting with the intra-class diversity and inter-class correlation. Note that the result of SS-MKL is just slightly lower than that of GS-MKL when $N_{train}$<20; but their performance gap becomes larger with more training images. This shows that GS-MKL is more effective in seeking a trade-off between diversity and invariance within an object category.

Table3. Performance of four multi-kernels based methods on WikipediaMM.

| $N_{tr}$ | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|
| UMK | 38.9±0.7 | 42.0±0.6 | 44.8±0.5 | 47.0±0.5 | 49.2±0.4 |
| MKL | 45.0±1.0 | 50.1±0.8 | 54.3±0.8 | 56.1±0.7 | 58.2±0.6 |
| SS-MKL | 47.3±1.6 | 53.4±1.3 | 56.2±0.9 | 57.8±1.1 | 60.5±1.0 |
| GS-MKL | **49.2±1.2** | **56.6±1.0** | **61.0±1.0** | **64.3±0.8** | **67.6±0.9** |

### 5.5. Time complexity

We implemented GS-MKL in C++. In each iteration of algorithm 1, we need to solve a canonical SVM problem with the group-sensitive kernel weights optimized by a gradient descent method. The time complexity of the gradient calculation is ignorable compared to the SVM solver. As those in canonical SVM solvers, using hot-start (i.e., providing previous $\boldsymbol{\alpha}$ as input) may accelerate the training process. Given the convergence termination criteria, the number of iterations before convergence depends on the training data and the step sizes. During training each category over 5k image samples on Pascal VOC2007 , the canonical MKL needs about 20 minutes, and GS-MKL needs 40 to 60 minutes to converge on server (8 Corel 3.0 GHz, 8GB RAM).

## 6. Conclusion

In this paper, we argue that modeling intra-class diversity and inter-class correlation among images is essential to improve the discriminative power of an object categorization method. To this end, we have introduced an intermediate representation "group" in the MKL framework, and proposed a GS-MKL method to learn both the parameters of group-sensitive kernel weights and the

classifier in a joint manner. Our GS-MKL has yielded promising results over Caltech101, Pascal VOC2007 and WikipediaMM datasets based on existing visual features and kernels.

In current implementation, grouping process is regarded as a pre-process followed by GS-MKL and the optimal group number is obtained over a validation set. In the future work, we will attempt to integrate the optimizations of three different stages (i.e., grouping, kernel combination and classifier learning). Additionally, we will employ more effective kernel functions and visual features in GS-MKL.

## 7. Acknowledgments

The work is supported by grants from Chinese NSF under contract No. 60605020 and No. 90820003, National Hi-Tech R&D Program (863) of China under contract 2006AA010105, and National Basic Research Program of China under contract No. 2009CB320906. Also this work is supported in part by the research fund from NLPR, Institute of Automation, Chinese Academy of Sciences, and Microsoft Research Asia Internet Services Theme. The authors would like to thank Hong Chang and Yu Su for their valuable suggestions.

## 8. References

[1] L. Fei-Fei, P. Perona. A bayesian hierarchical model for learning natural scene categories. In CVPR 2005.

[2] H. Zhang, A. Berg, M.Maire, J. Malik. SVM-KNN: discriminative nearset neighbor classification for visual category recognition. In CVPR, 2006.

[3] S. Lazebnik, C. Schmid, J. Ponce. Beyond bags of features: spatial pyramid matching for Recognizing Natural Scene Categories. In CVPR 2006.

[4] A. Bosch, A. Zisserman, X. Muoz. Image classification using random forests and ferns. In ICCV 2007.

[5] M. Varma, D. Ray. Learning the discriminative power-invariance trade-off. In ICCV 2007.

[6] L. Fei-Fei, R. Fergus, P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach testing on 101 object categories. In Workshop on Generative-Model Based Vision, CVPR, 2004.

[7] M. Everingham, L. VanGool, C. K. I. Williams, J. Winn, A. Zisserman. The PASCAL visual object classes challenge 2007 (VOC2007) results. http://www.Pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[8] M. Marszałek, C. Schmid, H. Harzallah, J. Weijer. Learning Object Representations for Visual Object Class Recognition. In Workshop on Visual Recognition Challenge, ICCV2007.

[9] www.imageclef.org/2008/wikipedia.

[10] F. R. Bach, G. R. G. Lanckriet, M. I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In ICML, 2004.

[11] A. Frome, Y. Singer, F.Sha, J. Malik. Learning globally-consistent Local distance functions for shape-based image retrieval and classification. In ICCV 2007.

[12] S. Fidler, M. Boben, A. Leonardis. Similarity-based cross-layered hierarchical representation for object categorization. In CVPR 2008

[13] J. Sivic, B. Russell, A. A. Efros, A. Zisserman. Discovering objects and their location in images. In ICCV 2005.

[14] G. Wang, Y. Zhang, L. Fei-Fei. Using dependent regions for object categorization in a generative framework. In CVPR, 2006.

[15] K. Grauman, T. Darrell. Pyramid match kernels: Discriminative classification with sets of image features. Tech. Report MIT CSAIL TR 2006-020, MIT, March 2006.

[16] L. Haibin, S. Soatto. Proximity distribution kernels for geometric context in category recognition. In ICCV 2007.

[17] K. Q. Weinberger, J. Blitzer, L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In NIPS 2005.

[18] A. Kumar, C. Sminc. Support kernel machines for object recognition. In ICCV 2007.

[19] O. Chum, A. Zisserman. An exemplar model for learning object classes. In CVPR2007.

[20] T. Malisiewicz, A. A. Efros. Recognition by association via learning per-exemplar distances. In CVPR 2008.

[21] J. Platt. Advances in Kernel Methods - Support Vector Learning, chapter Fast Training of Support Vector Machines using Sequential Minimal Optimization, pages 185–208. MIT Press, 1998.

[22] S. Sonnenburg, G. Raetsch, C.Schaefer, B.Scholkopf. Large scale multiple kernel learning. JLMR, (2006) 7,1531–1565.

[23] M. Gonen, E. Alpaydin. Localized multiple kernel learning. In ICML 2008.

[24] D. Lowe. Object recognition from local scale-invariant features. In ICCV 1999.

[25] T. Hofmann. Probabilistic latent semantic indexing. In ACM SIGIR 1998.

[26] S. Todorovic, N. Ahuja. Learning subcategory relevancies for category recognition. In CVPR2008.

[27] J. Mutch, D. G. Lowe. Multiclass object recognition with sparse, localized features. In CVPR, 2006.

[28] G.Wang, D.Hoiem, D.Forsyth. Learning Image Similarity from Flickr Groups Using Stochastic Intersection Kernel Machines. In MIR2008

[29] Y. Lin, T. Liu, C. Fuh. Local ensemble kernel learning for object category recognition. In CVPR 2007.

[30] A. Rakotomamonjy, F. Bach, Y. Grandvalet, S. Canu. SimpleMKL. JLMR, (2008) 9, 2491-2521.

[31] O. Chapelle, V.Vapnik, O. Bousquet, S. Mukherjee. Choosing multiple parameters for support vector machines. Machine Learning, (2002) 29, 131-159.

[32] E. Shechtman, M. Irani. Matching local self-similarities across images and videos. In CVPR 2007.

[33] A. Bosch, A. Zisserman, X. Munoz. Representing shape with a spatial pyramid kernel. In CIVR 2007.

[34] C. Galleguillos, A. Rabinovich, S. Belongie. Object Categorization using Co-Occurrence, Location and Appearance. In CVPR2008.

[35] F. Khan, J. Weijer, M. Vanrell. Top-Down Color Attention for Object Recognition. In ICCV2009.