Abstract—In this paper, we present a subcategory-aware recognition framework to boost category level object classification performance. Different from the existing monolithic model approaches, we aim to automatically leverage the embedded subcategory structure to assist the further category level recognition. Motivated by the observation of considerable intra-class diversities and inter-class ambiguities in many current object classification datasets, we explicitly split data into subcategories by ambiguity guided subcategory mining. The resulting subcategories are seamlessly integrated into the state-of-the-art detection assisted classification framework. More specifically, we build the instance affinity graph by combining both intra-class similarity and inter-class ambiguity. Visual subcategories, which correspond to the dense subgraphs, are detected by the graph shift algorithm. We then train an individual model for each subcategory rather than attempt to represent an object category with a monolithic model. Related samples, which are informative for subcategory classification, are utilized to regularize each subcategory model. Finally the responses from subcategory models are aggregated by subcategory-aware kernel regression. The extensive experiments over the PASCAL VOC 2007 and PASCAL VOC 2010 databases show the state-of-the-art performance from our framework.

Index Terms—Object Classification, Subcategory Mining, Related Samples, Contextualization.

I. INTRODUCTION

Visual categorization is a core problem in computer vision. Bag-of-Words (BoW) approaches to category level classification advanced significantly during the past few years [1], [2], [3], [4], [5]. This framework utilizes the local feature extraction, feature encoding and feature pooling pipeline to generate global image representations. Each object category is then represented by a monolithic model, such as a support vector machine classifier. However, the large intra-class diversities induced by pose, viewpoint and appearance variations [6] make it difficult to build an accurate monolithic model for each category, especially when there are many ambiguous samples. For example, the chair category in Figure 1 includes three obvious subcategories, namely, sofa-like chairs, rigid-material chairs and common chairs. In feature space, these subcategories are essentially far away from each other. Furthermore, the ambiguous sofa-like chairs look more like sofas than common chairs. In this case, representing all chairs with a monolithic model will weaken the model separating capacity and cannot distinguish sofas from chairs. Hence, it is intuitively beneficial to model each subcategory independently. These considerable intra-class diversities and inter-class ambiguities are common in the challenging real world datasets [7], [8], which makes the subcategory awareness necessary.

To effectively employ the subcategory information for category level classification in a principled way, the first step is to mine the subcategory structure automatically. At first glance, clustering all training data of an object category with intra-class similarity seems to be a natural strategy, since objects belonging to the same subcategory should intuitively have
larger similarity in terms of appearance and shape. However, in the context of generic object classification, subcategories mined with only intra-class visual similarity cues are unnecessary to be optimal due to the ignorance of valuable inter-class information [9]. More specifically, if the samples are clustered by standard clustering methods, we are unable to utilize the valuable inter-class information to handle the ambiguous samples. Then all ambiguous samples, which often lie near the decision boundary, may be grouped together and preserve the original complicated decision boundary. On the contrary, with the assistance of inter-class information ambiguous samples can be grouped into proper subcategories, which leads to easier subproblems and further improves the overall performance. For instance, the chair category and other categories in Figure 1 have non-linear decision boundary.

By noting the ambiguous chair sample distribution near the decision boundary, these chairs should be intuitively divided into separate subcategories. The proper split as indicated in Figure 1 will make all subcategories linearly separable from other categories, which is only achievable with the assistance of inter-class information. The above observation inspires us to propose an ambiguity guided subcategory mining approach to explore the intrinsic subcategory structure embedded in each category.

With mined subcategories, designing an effective strategy to train subcategory classifiers tailored for category level classification is not trivial. A naive approach is assigning the samples for the mined subcategory as positive samples and samples in the other categories as negative samples. However, such an approach ignores the informative related samples (samples from other subcategories of the same category) and is unstable for some subcategories with small number of samples. Instead, we propose to employ the related samples under the “Universum” SVM framework [10], which can stabilize and regularize the subcategory classifier to further boost the category level performance.

Overall, with subcategory awareness we can boost category level classification by subcategory-aware object classification (SAOC). As indicated in Figure 1, we split data into subcategories by ambiguity guided subcategory mining and train an individual model for each subcategory. During subcategory classifier training, besides positive and negative samples we further leverage the related samples to regularize the subcategory classifier for better fitting the overall category level data distribution. Since the diversities in each subcategory and ambiguities between subcategories and other categories are reduced, more accurate shape-based [11], [12]/appearance-based [13], [14] detectors and foreground classification model [5] can be built, which fits nicely with the state-of-the-art detection assisted classification framework [15], [16]. The final classification results are generated by aggregating subcategory responses through subcategory-aware kernel regression.

The main contributions of this paper are summarized as follows.

- We propose a novel ambiguity guided subcategory mining approach, which gracefully integrates the intra-class similarity and inter-class ambiguity for effective subcategory mining.
- We design an effective strategy to employ “related samples” under the “Universum” SVM framework. Such informative related samples will fine-tune the subcategory classifiers to be more suitable for category level classification.
- We provide a subcategory-aware object classification framework based on the detection assisted classification scheme [15], [16] to demonstrate how to effectively employ the subcategory information for visual recognition. Our ambiguity guided subcategory mining approach can be seamlessly integrated into such framework. Utilizing mined subcategories can improve both detection and classification performance and allow more effective subcategory level interaction in the fusion model. The state-of-the-art classification results on the PASCAL VOC datasets verify the effectiveness of our new framework.

The rest of the paper is organized as follows. Section II briefly reviews the related literature. Section III describes the overview of the proposed subcategory aware classification framework. Detailed explanation of subcategory mining and subcategory classification with related samples is presented in Section IV and Section V. Extensive experiments are conducted in Section VI. Section VII concludes the paper and discusses ideas for future work.

A previous version of this work has been published in [17]. This version adds: (1) new concept of “related samples” and strategies to utilize them; (2) a more detailed description of the subcategory-aware object classification framework; (3) an in-depth analysis of the various parameters of our system; (4) more detailed experimental results, especially about comparison with different strategies for employing “related samples”.

II. RELATED WORK

Current leading detection assisted classification framework relies on the cooperation of many recognition techniques, such as classification, detection and even segmentation. A detailed review of all the fields is beyond the scope of this paper, hence we only focus on the topics that are most related to the proposed framework.

Object Classification. Many state-of-the-art image classification systems follow the popular local feature extraction-coding-pooling pipeline [1]. First, local features like HOG [11], SIFT [18] and LBP [19] are extracted on the dense grids or sparse interest points. They are then encoded with a predefined visual dictionary by vector quantization (VQ), locally-constrained linear coding (LLC) [3] or Fisher kernel (FK) [4], [3]. Finally the encoded vectors are pooled together to form the image-level representation [2], [5]. Much research on image classification has been focused on improving this pipeline [3], [4], [20]. Some recent works [16], [15], [14], [13], [21], [22], [23] begin to investigate out of this pipeline. Harzallah et al. [15] introduced the pioneering work for detection and classification contextualization, the extension of which leads to the state-of-the-art results [16], [5], [21]. Segmentation results [24] have also been employed to boost the classification performance [13], [14]. However, all the
above methods train a monolithic model for each category, and there are few works analyzing the data structure embedded in each category. In our work, we show that properly splitting the data into subcategories will boost the performance of the state-of-the-art pipeline. Another line of work design a large number of weakly trained classifiers and treat the output of these classifiers as image descriptor [22], [23]. Such weak classifiers are usually obtained from semantic annotation, such as visual concepts, and bear the mid-level information to some extent. Unlike these methods, our work automatically discovery the structure embedded in each category without relying on manual annotation.

Object Detection. Object detection [12] is another central problem in object recognition, which is complementary to object classification [16], [15]. As most standard semantic categories do not form coherent visual categories, mixture models are proposed and have become the standard approach for object detection [25], [12]. Early works only investigate heuristics based on meta-data or manual labels such as bounding box aspect ratio [12], object scale [26], object viewpoint [27] and part labels [28] to group the positive samples into clusters. However, each of these methods has its own limitations and ignores other more general intra-class variations such as appearance and shape variance [6], [29]. Malisiewicz et al. [6] handled the intra-class variation by training a separate model for each positive instance, which inevitably reduces the generalization capacity of each model. Some recent works begin to investigate the visual subcategory structure embedded in each category [30], [29], [31], [25], [32], [33], which leads to considerable improvement in object detection performance. Gu et al. [29] grouped the samples into components based on the key point and mask annotations. Aghazadeh et al. [32] built a similarity graph based on intra-class information and utilized spectral clustering to split the data. In contrast to our method, these methods either require manual annotation or are fragile to outliers corresponding to highly occluded or strange samples. Furthermore, most of the previous works focus on object detection and are not suitable for object classification. Finally, these methods discard the inter-class information during data grouping, which is critical for object classification.

Locally Adaptive Classifiers. When the data has a complex non-linear structure, locally adaptive classifiers are usually superior to the use of a single global classifier [34], [35], [9]. Kim and Kittler placed the local classifiers at the clusters obtained by the K-means clustering algorithm [35]. Instead of placing the classifiers based on the data distribution only, Dai et al. [9] proposed a responsibility mixture model that uses the uncertainty associated with the classification at each training sample. Using this model, the local classifiers are placed near the decision boundary where they are most effective. Hoai and Zisserman [36] learn sub-categories by investigating a weakly supervised approach using both positive and negative samples of the category. In this work, we borrow the idea of uncertainty piloted classification and propose an ambiguity guided subcategory mining approach under the graph shift [37] framework.

III. SUBCATEGORY-AWARE OBJECT CLASSIFICATION

Our subcategory-aware object classification (SAOC) framework relies on the automatically mined subcategory information to boost the category level recognition. In this section we mainly demonstrate how to effectively utilize the mined subcategory information in current leading detection assisted classification scheme. Details on subcategory mining and strategy of training a subcategory classifier are shown in the sequel sections.

The diagrammatic flowchart of our SAOC framework is depicted in Figure 2. The whole framework consists of three main components - detection, classification and fusion models. We will first introduce each component of the framework and then emphasize how subcategory information fits into each step.

Fig. 2: Diagrammatic flowchart of the proposed subcategory-aware object classification framework. Given a testing image, they are first processed by each learnt subcategory model including detection and classification models. Then the responses from all subcategory models are fed into the fusion model to generate the final category level classification results.
A. Classification Model

For classification, we follow the state-of-the-art Generalized Hierarchical Matching (GHM) pipeline [5] and train a classifier for each subcategory individually. GHM generalizes the Spatial Pyramid Matching by allowing image adaptive pooling instead of pre-defined grid-based pooling. Both the detection confidence map and saliency map have shown to be effective to guide the pooling process for certain datasets [5]. In this work, since we focus on the scenarios where background is usually cluttered and many of the concerned object classes may co-occur in a single image, detection confidence maps are employed as the side information for GHM. The details for classifier training are explained in Section V.

B. Detection Model

Detection and classification are two strongly correlated and complementary tasks. Most leading classification systems employ the detection techniques to some extent. In our framework, the raw detection results are fed into the final fusion model as middle level features as well as provide the confidence map for the GHM pooling. Specifically, each subcategory is characterized by one shape-based sliding window detector [12], [38] and one appearance-based selective window detector [39], [13], respectively. The usage of two detectors is to guarantee both high precision and high recall on object detection since none of the detectors can achieve this alone and they complement each other.

C. Fusion Model

The fusion model mainly aims to: (1) boost the classification performance by complementary detection results, (2) utilize the context of all categories for reweighting, and (3) fuse the subcategory level results into final category level results. All of these are achieved by kernel regression. First, we construct a middle level representation for each training/testing image by concatenating classification scores and the leading two detection scores from each subcategory model. The final category level classification results are then obtained by performing Gaussian kernel regression on this representation. Without sophisticated models and complicated postprocessing [40], [16], our subcategory-aware kernel regression is very efficient and still performs well experimentally.

D. Subcategory Awareness

Subcategory awareness, which benefits each model separately and then boosts the overall performance of the framework, plays a critical role in extending current detection assisted classification framework.

- The subcategory information can be used to initialize both detection and classification models to better handle the rich intra-class diversities in challenging datasets. Less diversity in each subcategory will lead to a simpler learning problem, which can be better characterized by current state-of-the-art models, such as the Deformable Part based Model (DPM) for detection and the foreground BoW models involved in GHM.

IV. AMBIGUITY GUIDED SUBCATEGORY MINING

In this section, we will introduce how to find the subcategories by our ambiguity guided subcategory mining approach as illustrated in Figure 3. Before digging into details, we first summarize the notations used in this work. For a classification problem, a training set of $M$ samples is given and represented by the matrix $X = [x_1, x_2, \ldots, x_M] \in \mathbb{R}^{d \times M}$. The class label of $x_i$ is $c_i \in \{1, 2, \ldots, N_c\}$, where $N_c$ is the number of classes. We also denote the number of samples belonging to the $c$th class by $n_c$, and the corresponding index set of samples by $\pi_c$. 

![Ambiguity guided subcategory mining approach. First instance affinity graph is built by combining both intra-class similarity and inter-class ambiguity. Then dense subgraphs are detected within the affinity graph by performing graph shift. Each detected dense subgraph corresponds to a certain subcategory.](image-url)
A. Similarity Modeling

In this work, we define the appearance similarity as the Gaussian similarity between classification features \( \exp \{-||x_i - x_j||^2/\delta^2\} \), where \( \delta^2 \) is the empirical variance of \( x \). Though it is a common similarity metric for object classification, appearance similarity only is not enough for our SAOC framework, as in SAOC classification and detection are closely integrated. Subcategory mining only based on appearance similarity may lead to poor detectors, which in turn harms the overall performance. Hence detection and classification feature spaces ought to be taken into count simultaneously for similarity calculation.

The HOG based sliding window methods are the dominant approaches for object detection, which concatenate all the local gradients to form the window representation. These grid based HOG representations roughly capture object shapes and thus are sensitive to highly cluttered backgrounds and misalignments. Directly computing distance in concatenated HOG feature space often leads to poor results due to image misalignments [6]. To better measure the shape similarity between samples, we train a separate Exemplar-SVM detector[6], [41] for each positive sample. The misalignments can thus be partially handled by sliding the detector. The calibrated detection scores [6] are defined as the pair-wise shape similarity.

The final instance similarity is defined by fusing the appearance similarity and pair-wise shape similarity. More specifically, we denote the appearance similarity as \( S(A)_{i,j} \) and the pair-wise shape similarity as \( S(P)_{i,j} \). Both \( S(A) \) and \( S(P) \) are normalized to \([0, 1]\). The final instance similarity is defined as \( S_{i,j} = S(A)_{i,j} \times S(P)_{i,j} \).

B. Ambiguity Modeling

As discussed above, inter-class information is crucial for object classification. Dai et al. [9] have shown that placing local classifiers near the decision boundary instead of based on the data distribution only leads to better performance. This is intuitive as even there are many subcategories spreading separately in the feature space, if none of subcategories are close to samples of other categories, a single classifier may be enough to correctly classify all these subcategories. On the contrary, if some subcategories are near the decision boundary, separate classifiers should be trained for these ambiguous subcategories. Otherwise the ambiguous subcategories may decrease the classification performance of categories near the decision boundary.

As ambiguity is critical for object classification, subcategory mining should be guided by ambiguity instead of only relying on intra-class data distribution. Before introducing how to combine sample similarity and ambiguity into a unified framework, we need to first explicitly define the ambiguity measure. Here, we consider the L-nearest neighbours\(^1\) of a particular sample \( x_i \). If most of its neighbours share the same class label as \( x_i \), the classification of \( x_i \) should be easy. Otherwise, \( x_i \) will be ambiguous and likely to be classified incorrectly. We thus define the ambiguity \( A(x_i) \) of a training sample \( x_i \) as:

\[
A(x_i) = \frac{\sum_{j \in N_i^L} S_{i,j}}{\sum_{j \in N_i^L} S_{i,j}},
\]

where \( N_i^L \) is the index set of the L-nearest neighbours of \( x_i \). From the definition, a large \( A(x_i) \) means that the neighbouring samples are likely to be of different classes, and hence the classification of \( x_i \) is more uncertain. On the contrary, a small \( A(x_i) \) indicates that more neighbouring samples share the same class label of \( x_i \). Note that computing the ambiguity relies on not only the intra-class information but also the inter-class formation. The ambiguity will be high for those training samples lying close to the decision boundary, and thus such samples should be more likely to form a separate subcategory.

C. Subcategory Mining by Graph Shift

Intuitively, the subcategory mining algorithm is expected to satisfy the following three properties. (1) It should be compatible with graph representation. Many similarity metrics are defined based on pair-wise relation, such as our pair-wise shape similarity. Hence, non-graph based algorithms, such as mean shift, k-means and [36], may not be suitable due to the lack of ability to directly utilize the pair-wise information. (2) It is able to utilize the informative inter-class ambiguities. Clustering methods based on only intra-class data distribution may fail to detect the ambiguous subcategories on the decision boundary and lead to subcategories imperfect for classification. Hence the expected algorithm should be able to adaptively cluster the data guided by ambiguity. (3) It should be robust to outliers. Some samples, such as highly occluded or strange images, may not belong to any subcategory. Methods insisting on partitioning all the input data into coherent groups without explicit outlier handling may fail to find the true subcategory structure.

The traditional partitioning methods, such as k-means and spectral clustering methods, are not expected to always work well for subcategory mining due to their insistence on partitioning all the input data and inability to integrate the inter-class information. Hence we need a more effective algorithm satisfying the above three properties. The graph shift algorithm [37], which is efficient and robust for graph node seeking, appears to be particularly suitable for our subcategory mining problem as it directly works on graph, allows one to extract as many clusters as desired, and leaves the outlier points ungrouped. More importantly, the ambiguity can be seamlessly integrated into the graph shift framework. The graph shift algorithm shares the similar spirit with mean shift [42] algorithm and evolves through iterative expansion and shrink procedures. The main difference is that mean shift operates directly on the feature space, while graph shift operates on the affinity graph. The simulation results for comparing our ambiguity guided graph shift (AGS) with kmeans and spectral clustering are provided in Figure 4, from which we can see that our AGS can lead to subcategories more suitable for boosting classification.

Formally, we define an individual graph \( G = (V, A) \) for each category. \( V = \{v_1, \ldots, v_n\} \) is the vertex set, which

---

\(^1\)In the experiments, we simply use \( L = n_c/10 \) for the cth class.
represents the positive samples for the corresponding category. A is a symmetric matrix with non-negative elements. The diagonal elements of A represent the ambiguity of the samples while the non-diagonal element measures the similarity between samples. The modes of a graph G are defined as local maximizers of graph density function $g(y) = y^T A y$, $y \in \Delta^n$, where $\Delta^n = \{ y \in \mathbb{R}^n : y \geq 0 \text{ and } \|y\|_1 = 1 \}$. More specifically, in this paper sample similarity and ambiguity are integrated and encoded as the edge weights of a graph, whose nodes represent the instances of the specific object category. Hence subcategories should correspond to those strongly connected subgraphs. All such strongly connected subgraphs correspond to large local maxima of $g(y)$ over simplex, which is an approximate measure of the average affinity score of these subgraphs.

Since the modes are local maximizers of $g(y)$, to find these modes, we need to solve following standard quadratic optimization problem (StQP) [43]:

$$\begin{align*}
\text{maximize} & \quad g(y) = y^T A y \\
\text{subject to} & \quad y \in \Delta^n.
\end{align*}$$  \hspace{1cm} (2)

Replicator dynamics, which arises in evolutionary game theory, is the most popular method to find the local maxima of StQP (2). Given an initialization $y(0)$, corresponding local solution $y^*$ of StQP (2) can be efficiently computed by the discrete-time version of first-order replicator equation, which has the following form:

$$y_i(t + 1) = y_i(t) \frac{(Ay(t))_i}{y(t)^T A y(t)}, \quad i = 1, \ldots, n. \hspace{1cm} (3)$$

It can be observed that the simplex $\Delta^n$ is invariant under these dynamics, which means that every trajectory starting in $\Delta^n$ will remain in $\Delta^n$. Moreover, it has been proven in [44] that, when $A$ is symmetric and with non-negative entries, the objective function $g(y) = y^T A y$ strictly increases along any non-constant trajectory of Eqn. (3), and its asymptotically stable points are in one-to-one correspondence with strict local solutions of StQP (2). One of the main drawbacks of replicator dynamics is that it can only drop vertices and be easily trapped in any local maximum. The graph shift algorithm provides a complementary neighbourhood expansion procedure to expand the supporting vertices [37]. The replicator dynamics and the neighbourhood expansion procedure thus have complementary properties, the combination of which leads to better performance. In addition, as the diagonal elements may prevent the expansion to other vertices with no diagonal elements, vertices with large diagonal elements tend to form a local subgraph.

Like mean shift algorithm, the graph shift algorithm starts from each individual sample and evolves towards the mode of $G$. The samples reaching the same mode are grouped as a cluster. Each large cluster corresponds to one subcategory, while small clusters usually result from noises and/or outliers.

V. SUBCATEGORY CLASSIFICATION WITH RELATED SAMPLES

With the subcategory mining results, the following step is to construct subcategory classifiers tailored for category level classification. One intuitive approach is to employ standard binary SVM while treating samples in the target subcategory as positive samples and samples in other categories as negative samples. However, this strategy may lead to suboptimal results for the final category level classification due to several reasons. First, this hard separation of the whole training samples may result in limited samples for some subcategories, which will lead to unstable classifiers. Second, this approach is unable to exploit other informative samples in the same category. As our main goal is to construct classifiers suitable for category level classification instead of for accurate subcategory classification, classifiers only relying on the samples in the target subcategory may decrease the final category level performance.

To overcome the difficulty mentioned above, we propose the concept of “related samples”. For a target subcategory, related samples are defined as samples from other subcategories of the same category. Though unlabeled, the related samples should be informative for classification. A prominent example for utilizing unlabeled data is semi-supervised learning [45], where an additional set of unlabeled data are assumed to follow the same distribution as the training inputs. However, for our subcategory classification problem, related samples should have different distribution from either positive or negative samples. In other word, these related samples, which are considered potentially helpful for classification, should represents a third class. We note that the related samples can be viewed as a special form of “Universum” set as in [10]. Hence, we employ the “Universum” SVM framework [10], [46] for subcategory classification.

“Universum” SVM is an extension of the standard SVM by introducing the “Universum” set. Let $D = \{ (x_i, y_i) \mid x_i \in \mathbb{R}^p, y_i \in \{-1, 1\} \}_{i=1}^{n} \cup U \}$ be the set of labeled examples and let $U = \{ (x_j) \mid x_j \in \mathbb{R}^{p'} \}_{j=1}^{m}$ denote the set of related samples. $H_a[t]$ is the hinge loss ($H_a[t] = \max(0, a - t)$) and $L_\epsilon[t]$ is $\epsilon$-insensitive loss ($L_\epsilon[t] = \max(0, |t| - \epsilon)$). Besides penalizing the wrongly classified samples in $D$, we also bring...
Noting that the labeled samples with high confidence, the proposed related sample augmented approach (lower figure) will not make a strong assertion about the labels of related samples, which is beneficial for the final category level classification.

VI. EXPERIMENTS

A. Datasets and Metrics

Many datasets exist for object classification. Unfortunately, most of them, such as Oxford Flowers [48], Caltech 101 [49] and Caltech 256 [50], are single-label, object-centric and deficient in variance of pose and appearance, which makes these datasets insufficient to represent the visual world. In addition, these datasets are near saturation and not discriminative enough to distinguish different leading algorithms. Hence, we validate the proposed framework on the challenging PASCAL Visual Object Challenge (VOC) datasets [7], [51], [40], which provide a common evaluation platform for both object classification and detection. These datasets are extremely challenging since the images are crawled from the real-world photo sharing website and the objects contained vary significantly in size, pose, viewpoint and appearance. VOC 2007 and 2010 datasets, which contain 9,963 and 21,738 images respectively, are selected for experiments. The two datasets contain 20 object classes and are divided into “train”, “val” and “test” subsets. We conduct our experiments on the “trainval” and “test” splits. The employed evaluation metric is Average Precision (AP) and mean of Average Precision (mAP). We follow the standard PASCAL VOC comp1 test protocol for classification and PASCAL VOC comp3 test protocol for detection.

In the following experiments, we first show our ambiguity guided subcategory mining results for the bus and chair categories in Section VI-B. We then extensively compare different subcategory mining methods and subcategory classifier training strategies using VOC 2007 “trainval/test” datasets (i.e., “trainval” set for training and “test” set for test) for proof of concept and ease of parameter tuning in Section VI-C and VI-D. Finally, we evaluate the optimal configuration of our method on 2010 “trainval/test” datasets and compare with the state-of-the-art performance ever reported in Section VI-E.

B. Ambiguity Guided Subcategory Mining Results

It has been shown that models trained by “clean” subsets of images usually perform better than trained with all images [25]. The importance of “clean” training data suggests that it is critical to cluster training data into “clean” subsets and remove outliers simultaneously. Figure 6 displays our subcategory mining results for bus and chair categories. Each row on the left side shows one discovered subcategory while right side images are detected as outliers and left ungrouped.

For the bus category, the first 3 subcategories correspond to 3 different views of buses. This is mainly due to the discriminative pair-wise shape similarity for different views of buses, as the Exemplar-SVM works well for the categories with common rigid shapes. We note the shape and appearance of the last subcategory show much larger diversity than other subcategories. Though these images are not very similar to each other, the strong ambiguity with the person category still guides them to form a separate subcategory.
For chairs, there are no common rigid shapes as buses and the shapes of various chairs are very diverse, which leads to much noisier pair-wise shape similarity. Hence the subcategory mining results should be the combination effects of both appearance similarity and shape similarity, which can be observed from the discovered subcategories. Some subcategories may not have common shapes, but have similar local patterns. For example, chairs of the 2nd subcategory all have the stripe-like patterns. We note again the last detected subcategory looks like sofas. Besides being different from other chair subcategories, the ambiguity with sofa is also one of the main reasons that these images form a separate subcategory.

C. Subcategory Mining Method Comparison

We extensively evaluate the effectiveness of different subcategory mining approaches on the VOC 2007 dataset, as the ground-truth of its testing set is released. To allow direct comparison with other popular works [4], [20], [5], we only implement a simplified SOAC framework. More specifically, we choose the state-of-the-art FVGHM method [5] as the classification pipeline (dense SIFT feature [18] with FK coding [4] plus GHM pooling [2], [5]) and the customized DPM [52] as object detector. The only difference between customized DPM and the standard DPM is the model initialization step. Unlike standard DPM, which utilize the aspect ratio to cluster training samples into different groups, DPM-spectral, DPM-GS and DPM-AGS replace the aspect ratio based initialization with spectral clustering, graph shift and ambiguity guided graph shift mining results, respectively. For the standard DPM, we use the publicly available implementation with the default settings (8 parts) [52]. As detection assisted classification has become a standard approach for classification on PASCAL VOC. We augment FVGHM with detection context information as in [16] and utilize the resulting FVGHM-CTX as the starting point to evaluate different subcategory mining methods. Dense SIFT is extracted using multiple scales setting (spatial bins are set as 4, 6, 8, 10) with step 4. The size of Gaussian Mixture Model in FK is set to 256. For GHM [5], we construct the hierarchical structure with three-level clusters, each of which includes 1, 2, 4 nodes respectively. One-vs-All SVM is learnt for each category/subcategory. For our graph shift based approach, the subcategory number is determined by the expansion size (the number of selected nearest neighbors for the expansion stage [37]). In experiments the expansion size is decided by cross-validation, and the subcategory number is generally from 2 to 5. For fair comparison, We did not compare with the non-graph based approaches, such as k-
TABLE I: Classification results (AP in %) comparison for different subcategory mining approaches on VOC 2007. For each category, the winner is shown in bold font.

<table>
<thead>
<tr>
<th>Model</th>
<th>plane</th>
<th>bike</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>table</th>
<th>dog</th>
<th>horse</th>
<th>person</th>
<th>plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV [4]</td>
<td>55.7</td>
<td>58.8</td>
<td>79.0</td>
<td>60.4</td>
<td>69.4</td>
<td>55.6</td>
<td>55.6</td>
<td>55.6</td>
<td>41.8</td>
<td>56.3</td>
<td>41.7</td>
<td>76.3</td>
<td>64.4</td>
<td>82.7</td>
<td>28.3</td>
<td>39.7</td>
<td>56.6</td>
<td>79.7</td>
<td>71.5</td>
<td>58.3</td>
</tr>
<tr>
<td>FVGHM [5]</td>
<td>67.1</td>
<td>58.6</td>
<td>79.4</td>
<td>60.4</td>
<td>67.2</td>
<td>55.8</td>
<td>58.4</td>
<td>62.8</td>
<td>55.4</td>
<td>56.2</td>
<td>51.1</td>
<td>81.4</td>
<td>71.5</td>
<td>78.5</td>
<td>36.4</td>
<td>55.3</td>
<td>63.0</td>
<td>80.6</td>
<td>75.8</td>
<td>64.7</td>
</tr>
<tr>
<td>FVGHM-CTX</td>
<td>78.5</td>
<td>50.4</td>
<td>75.9</td>
<td>55.4</td>
<td>75.1</td>
<td>78.6</td>
<td>67.2</td>
<td>58.4</td>
<td>63.0</td>
<td>60.0</td>
<td>47.3</td>
<td>83.0</td>
<td>76.3</td>
<td>90.5</td>
<td>44.9</td>
<td>59.6</td>
<td>63.2</td>
<td>83.5</td>
<td>68.9</td>
<td>68.3</td>
</tr>
<tr>
<td>FVGHM-CTX-spectral</td>
<td>81.2</td>
<td>81.2</td>
<td>73.5</td>
<td>56.2</td>
<td>76.5</td>
<td>88.5</td>
<td>67.8</td>
<td>58.0</td>
<td>60.1</td>
<td>61.7</td>
<td>48.1</td>
<td>85.1</td>
<td>77.8</td>
<td>90.7</td>
<td>45.5</td>
<td>60.6</td>
<td>64.4</td>
<td>84.3</td>
<td>92.2</td>
<td>69.4</td>
</tr>
<tr>
<td>FVGHM-CTX-GS</td>
<td>81.8</td>
<td>82.8</td>
<td>74.1</td>
<td>56.5</td>
<td>77.2</td>
<td>88.7</td>
<td>68.4</td>
<td>59.4</td>
<td>61.3</td>
<td>63.0</td>
<td>49.8</td>
<td>84.9</td>
<td>80.0</td>
<td>91.3</td>
<td>47.7</td>
<td>61.3</td>
<td>65.9</td>
<td>85.7</td>
<td>70.8</td>
<td>70.4</td>
</tr>
<tr>
<td>FVGHM-CTX-AGS</td>
<td>92.2</td>
<td>83.0</td>
<td>58.7</td>
<td>76.1</td>
<td>56.4</td>
<td>77.5</td>
<td>85.8</td>
<td>69.1</td>
<td>62.2</td>
<td>61.8</td>
<td>51.3</td>
<td>85.4</td>
<td>80.2</td>
<td>91.1</td>
<td>48.1</td>
<td>61.7</td>
<td>67.7</td>
<td>86.5</td>
<td>79.0</td>
<td>71.4</td>
</tr>
</tbody>
</table>

TABLE II: Detection results (AP in %) comparison for different subcategory mining approaches on VOC 2007.

<table>
<thead>
<tr>
<th>Model</th>
<th>plane</th>
<th>bike</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>table</th>
<th>dog</th>
<th>horse</th>
<th>person</th>
<th>plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-SVM [6]</td>
<td>20.8</td>
<td>48.0</td>
<td>7.7</td>
<td>14.3</td>
<td>13.1</td>
<td>59.7</td>
<td>41.1</td>
<td>5.2</td>
<td>11.6</td>
<td>18.6</td>
<td>11.1</td>
<td>3.1</td>
<td>44.7</td>
<td>39.4</td>
<td>16.9</td>
<td>11.2</td>
<td>22.6</td>
<td>17.0</td>
<td>36.9</td>
<td>30.0</td>
</tr>
<tr>
<td>MC [29]</td>
<td>33.4</td>
<td>37.0</td>
<td>15.0</td>
<td>15.0</td>
<td>22.6</td>
<td>33.1</td>
<td>49.3</td>
<td>32.8</td>
<td>11.5</td>
<td>35.8</td>
<td>17.8</td>
<td>16.3</td>
<td>43.6</td>
<td>38.2</td>
<td>29.8</td>
<td>11.6</td>
<td>33.3</td>
<td>23.5</td>
<td>30.2</td>
<td>29.0</td>
</tr>
<tr>
<td>DPM [52]</td>
<td>28.9</td>
<td>59.5</td>
<td>10.0</td>
<td>15.2</td>
<td>25.5</td>
<td>49.6</td>
<td>57.9</td>
<td>19.3</td>
<td>22.4</td>
<td>25.2</td>
<td>23.8</td>
<td>11.1</td>
<td>56.8</td>
<td>48.7</td>
<td>41.9</td>
<td>12.2</td>
<td>17.8</td>
<td>33.6</td>
<td>45.1</td>
<td>41.6</td>
</tr>
<tr>
<td>DPM-spectral</td>
<td>32.9</td>
<td>60.3</td>
<td>9.6</td>
<td>15.9</td>
<td>29.2</td>
<td>52.6</td>
<td>58.1</td>
<td>21.0</td>
<td>24.1</td>
<td>23.0</td>
<td>33.2</td>
<td>19.3</td>
<td>58.2</td>
<td>48.1</td>
<td>37.6</td>
<td>11.9</td>
<td>21.3</td>
<td>35.3</td>
<td>48.6</td>
<td>43.1</td>
</tr>
<tr>
<td>DPM-GS</td>
<td>34.4</td>
<td>60.1</td>
<td>11.4</td>
<td>17.5</td>
<td>29.9</td>
<td>53.0</td>
<td>58.9</td>
<td>23.7</td>
<td>22.3</td>
<td>25.8</td>
<td>30.3</td>
<td>12.6</td>
<td>60.8</td>
<td>49.2</td>
<td>42.6</td>
<td>13.3</td>
<td>22.9</td>
<td>37.0</td>
<td>50.2</td>
<td>45.4</td>
</tr>
<tr>
<td>DPM-AGS</td>
<td>34.7</td>
<td>61.4</td>
<td>11.5</td>
<td>18.6</td>
<td>30.0</td>
<td>53.8</td>
<td>58.3</td>
<td>24.7</td>
<td>24.7</td>
<td>26.8</td>
<td>31.4</td>
<td>13.8</td>
<td>61.4</td>
<td>49.2</td>
<td>42.2</td>
<td>12.9</td>
<td>23.9</td>
<td>38.5</td>
<td>50.8</td>
<td>45.2</td>
</tr>
</tbody>
</table>

means and [36], as they are difficult to directly utilize our pairwise shape similarity. Spectral clustering, the representative graph based partition method, is chosen for comparison. We extensively evaluate spectral clustering with the cluster number from 2 to 5 and report the best results.

The detailed classification results are shown in Table I. It can be concluded from the table that:

- Subcategory awareness does improve the performance of current detection assisted classification framework. Subcategory information provides an effective approach to decompose the original difficult problem into several easier sub-problems. Such simplified sub-problem can be better captured by current classification methods, which then improves the overall performance. Even with the naive spectral clustering for category mining, we can still boost the state-of-the-art classification performance.
- Our ambiguity guided graph shift approach is effective for subcategory mining. The resulting subcategories can obviously improve the classification performance; By adaptively grouping the samples into subcategories and rejecting the outliers, our ambiguity guided graph shift approach performs much better than the spectral clustering.
- Ambiguity is informative for subcategories mining. The sample ambiguity implicitly provides information about other categories and enables the algorithm to focus on the samples near the decision boundary, which are more important to the classification problem. With the assistance of sample ambiguity, the graph shift algorithm can obtain better results for 17 out of 20 categories.

Figure 7 shows some exemplar results for the baseline method (FVGHM-CTX) and the proposed algorithm (FVGHM-CTX-AGS) from the VOC “test” set. It can be observed that the monolithic model (FVGHM-CTX) fails to recognize many samples due to the variance of pose, view point and appearance. On the contract, such samples can be successfully recognized by some subcategory classifiers. The less diversities in each subcategory will make the corresponding classifier more reliable and accurate. The final subcategory-aware classifier, which fuses the responses from all subcategory classifiers, can successfully recognize more samples than the baseline method.

As object detection is an inseparable component of our SAOC framework, we also show the intermediate detection results in Table II. Besides standard DPM, we add two more baselines, which also use the multiple components/models for object detection [29], [6]. When compared with other leading techniques in subcategory based detection, our method obtains the best results for most categories, achieving superior performance on categories with rigid shape or high ambiguity. We note the MC [29], which requires manually labelling the pose of each image, performs quite well on articulated categories. The inferior performance of our ambiguity guided mining framework on articulated categories is mainly due to the limited discriminative ability of current similarity metric.

The number of subcategories: We have proposed the ambiguity guided graph shift for subcategory mining and verified its effectiveness. Here we evaluate the influence of the number of subcategories. Particularly, we select the bus, chair and horse category as representative.

From Figure 8, the optimal number of subcategories depends on the characteristics of the specific category. We can summarize the observations for the different categories as follows:

- For small number of subcategories (K) the performance gradually increases with increasing K, but stabilizes around K = 4. As there are large variation for samples in each category due to pose, viewpoint and appearance variance, properly dividing them into subcategory will lead to easier sub-problems and thus improve the overall performance.
- Further increasing K may decrease the performance. One of the reasons for such decrease is the lack of data. Larger K will lead to fewer samples in each subcategory. Such small number of samples may be insufficient for training a reliable subcategory model and hurt the overall
Baseline Classifier  | Subcategory Classifier  | Subcategory-aware Classifier

Fig. 7: Exemplar results for the baseline method (FVGHM-CTX) and FVGHM-CTX-AGS from the VOC 2007 “test” set. The classification results are compared by the confidence scores for each classifier. The blue and green bars represent the baseline classifier and the subcategory classifiers, respectively. The subcategory-aware classifier, which fuses the scores of all subcategory classifiers to obtain the final score, is represented by the red bar. For better viewing, please see original colour pdf file.

As the running time increases with $K$ and $K = 5$ is large enough to get the optimal performance for most categories, we select the best $K$ from 2 to 5 for the balance of accuracy and speed.

D. Subcategory Classifier Training Strategy Comparison

In this subsection, we evaluate different strategies for training subcategory classifiers. We compare the related samples augmented approach described in Section V with two baseline strategies. For the target subcategory, the first strategy assigns the samples in this subcategory as positive samples and the samples belonging to the other categories as the negative samples. This is the approach used in Subsection VI-C. The
TABLE III: Classification results (AP in %) comparison for different subcategory classifier training strategies on VOC 2007. For each category, the winner is shown in bold font.

<table>
<thead>
<tr>
<th></th>
<th>plane</th>
<th>bike</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>cat</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>table</th>
<th>dog</th>
<th>horse</th>
<th>motor</th>
<th>person</th>
<th>plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVGHM-CTX-ASM</td>
<td>82.2</td>
<td>83.0</td>
<td>58.4</td>
<td>76.1</td>
<td>56.4</td>
<td>77.5</td>
<td>88.8</td>
<td>69.1</td>
<td>62.2</td>
<td>61.8</td>
<td>64.2</td>
<td>51.3</td>
<td>85.4</td>
<td>80.2</td>
<td>91.1</td>
<td>48.1</td>
<td>67.7</td>
<td>86.3</td>
<td>70.9</td>
<td>71.1</td>
<td>0.81</td>
</tr>
<tr>
<td>FVGHM-CTX-ASM-2</td>
<td>76.9</td>
<td>79.0</td>
<td>55.9</td>
<td>72.7</td>
<td>53.6</td>
<td>66.4</td>
<td>88.4</td>
<td>67.8</td>
<td>60.6</td>
<td>59.9</td>
<td>61.3</td>
<td>50.4</td>
<td>79.9</td>
<td>74.3</td>
<td>87.8</td>
<td>44.8</td>
<td>59.2</td>
<td>62.2</td>
<td>83.7</td>
<td>69.7</td>
<td>0.82</td>
</tr>
<tr>
<td>FVGHM-CTX-ASM-RS</td>
<td>82.6</td>
<td>85.3</td>
<td>58.2</td>
<td>78.5</td>
<td>57.7</td>
<td>79.2</td>
<td>88.6</td>
<td>70.4</td>
<td>63.8</td>
<td>64.1</td>
<td>65.4</td>
<td>53.7</td>
<td>86.1</td>
<td>80.6</td>
<td>90.8</td>
<td>48.9</td>
<td>63.4</td>
<td>69.7</td>
<td>87.8</td>
<td>71.6</td>
<td>0.83</td>
</tr>
</tbody>
</table>

second strategy assigns the samples in this subcategory as positive samples and all other samples as the negative samples. The difference between two baseline strategies lies in how to handle the related samples. The first strategy simply abandons them while the other one assigns them as the negative samples. We use the same experiment setting as in Subsection VI-C and the experimental results are shown in Table III. The penalty parameters \( C_D \) and \( C_U \) in Eqn. 4 are decided by cross-validation. From the Table III, it can be observed that:

- The baseline strategy 2 (FVGHM-CTX-AGS-2: assign samples for the target subcategory as the positive samples and all other images as negative samples) leads to the worst results. This is intuitive as our concern is category level classification. However, because the samples from the same category are usually more similar, this strategy will make subcategory classifier focus on the boundary between the target subcategory and the other subcategory of the same category instead of the boundary between the target subcategory and other categories. Hence, the final subcategory classifier is not discriminative for the category level classification.
- Unlike the baseline strategy 1 (FVGHM-CTX-ASM), which abandons the informative related samples, the proposed related samples augmented approach (FVGHM-CTX-ASM-RS) effectively utilize them under the “Universe” SVM framework. These related samples are effectively exploited to tune the classifier for category level classification, especially for the subcategory with small number of samples, which leads to the best performance.

E. Comparison with the State-of-the-arts

In this section we compare the performance of the proposed SAOC framework with the reported state-of-the-art results on the VOC 2010 dataset. To obtain the state-of-the-art performance, we conduct the experiments with more complicated setting. For classification, we extract dense SIFT, HOG, color moment and LBP features in a multi-scale setting. All these features are encoded with VQ, LLC and FK [20] and then pooled by GHM. The pooling results are concatenated to form the final image representation. During SVM training, \( \chi^2 \) and linear kernel is employed for VQ/LC and FK, respectively. For object detection, we train one shape-based detector and one appearance-based object detector for each object category. The augmented DPM [38], [16] employing both the HOG and LBP features is adopted as the shape-based model. For appearance-based approach [39], [13], we sample 4000 sub-windows of different sizes and scales, and perform the BoW based object detector on these sub-windows. The number of subcategories is also determined by cross-validation as mentioned above.

We compare with the best known VOC 2010 performance from several recent papers and the released results from the VOC 2010 challenge [40], which are all obtained through the combinations of multiple methods in order to obtain better performance. The comparison results are presented in Table IV, from which it can be observed that:

- Our proposed method outperforms the competing methods on all 20 object categories. We note that all the leading classification methods combine object classification and object detection to achieve higher accuracy. However, most of the previous methods simply fuse the outputs of a monolithic classification model and a monolithic detection at category level. This limitation prevents them from grasping the informative subcategory structure and the interaction among the subcategories. By properly...
TABLE IV: Classification results from the proposed framework with comparison to other leading methods on VOC 2010.

<table>
<thead>
<tr>
<th>plane</th>
<th>bike</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>table</th>
<th>dog</th>
<th>horse</th>
<th>motor</th>
<th>person</th>
<th>plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.3</td>
<td>77.0</td>
<td>65.3</td>
<td>75.0</td>
<td>53.7</td>
<td>85.9</td>
<td>80.4</td>
<td>74.6</td>
<td>62.9</td>
<td>66.2</td>
<td>54.1</td>
<td>66.8</td>
<td>76.1</td>
<td>81.7</td>
<td>89.9</td>
<td>41.6</td>
<td>66.3</td>
<td>57.0</td>
<td>85.0</td>
<td>74.3</td>
</tr>
<tr>
<td>93.3</td>
<td>72.9</td>
<td>69.9</td>
<td>77.2</td>
<td>47.9</td>
<td>85.6</td>
<td>79.7</td>
<td>79.4</td>
<td>61.7</td>
<td>56.6</td>
<td>61.1</td>
<td>71.1</td>
<td>76.7</td>
<td>79.3</td>
<td>86.8</td>
<td>38.1</td>
<td>63.9</td>
<td>55.8</td>
<td>87.5</td>
<td>72.9</td>
</tr>
<tr>
<td>ContextISVM [16]</td>
<td>93.1</td>
<td>89.9</td>
<td>75.2</td>
<td>77.1</td>
<td>54.3</td>
<td>85.3</td>
<td>80.7</td>
<td>78.9</td>
<td>64.5</td>
<td>68.4</td>
<td>64.1</td>
<td>70.3</td>
<td>81.5</td>
<td>83.9</td>
<td>91.5</td>
<td>48.9</td>
<td>72.6</td>
<td>58.2</td>
<td>87.8</td>
</tr>
<tr>
<td>GHM ObjHierarchy [3]</td>
<td>94.3</td>
<td>81.3</td>
<td>77.2</td>
<td>80.3</td>
<td>56.3</td>
<td>87.3</td>
<td>83.8</td>
<td>82.2</td>
<td>65.8</td>
<td>53.7</td>
<td>61.0</td>
<td>75.9</td>
<td>82.3</td>
<td>86.5</td>
<td>92.0</td>
<td>51.7</td>
<td>75.1</td>
<td>65.3</td>
<td>89.9</td>
</tr>
<tr>
<td>PVGHM-CTX-AGS</td>
<td>95.9</td>
<td>85.2</td>
<td>79.0</td>
<td>84.9</td>
<td>58.1</td>
<td>87.4</td>
<td>84.3</td>
<td>83.4</td>
<td>70.2</td>
<td>75.2</td>
<td>78.0</td>
<td>85.7</td>
<td>88.4</td>
<td>92.1</td>
<td>52.7</td>
<td>78.5</td>
<td>67.8</td>
<td>93.0</td>
<td>77.4</td>
</tr>
<tr>
<td>PVGHM-CTX-AGS-RS</td>
<td>96.4</td>
<td>84.8</td>
<td>87.3</td>
<td>85.5</td>
<td>57.0</td>
<td>91.7</td>
<td>85.6</td>
<td>85.1</td>
<td>77.7</td>
<td>79.3</td>
<td>80.6</td>
<td>81.0</td>
<td>86.4</td>
<td>89.4</td>
<td>92.9</td>
<td>54.2</td>
<td>78.0</td>
<td>66.5</td>
<td>93.1</td>
</tr>
</tbody>
</table>

employing the subcategory structure, we can improve the state-of-the-art performance by 2.1%.

- Related samples are informative for the category level classification. The proposed related samples enhanced approach can further boost the overall performance by 1.1%.

- Note that our methods can significantly improve the performance of rigid categories (bus, train) and ambiguous categories (sofa, chair). For the rigid categories, the proposed subcategory mining approach is able to split the data effectively, which leads to “clean” subcategories and boosts the performance. For ambiguous categories, our model can implicitly re-rank the results. The scores for subcategories without ambiguity are raised and scores for ambiguous subcategories are depressed (still larger than samples not belonging to the corresponding category), which will also improve the APM.

When measured with object detection, we can achieve the performance of 37.1% compared to the state-of-the-art results of 36.8% [40], which is obtained by much more complicated detection models than ours. As our framework focuses on classification, detailed detection results are omitted due to the space limitation.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an ambiguity guided subcategory mining and subcategory-aware object classification framework for object classification. We modeled the subcategory mining as a dense subgraph seeking problem. This general scheme allows us to gracefully embed intra-class similarity and inter-class ambiguity into a unified framework. The subcategories, which correspond to the dense subgraphs, can be effectively detected by the graph shift algorithm. Ambiguity guided subcategory mining results are then seamlessly integrated into the subcategory-aware detection assisted object classification framework. The usage of “relate samples” allows us to effectively tailor the subcategory classifiers for category level classification. Extensive experimental results on both PASCAL VOC 2007 and VOC2010 clearly demonstrated the proposed framework achieved the state-of-the-art performance.

In the future, we plan to further explore whether our ambiguity guided subcategory mining can be extended for object segmentation and also develop a more efficient and scalable version of current framework to handle bigger data.

ACKNOWLEDGMENT

This research is supported by the Singapore National Research Foundation under its International Research Centre @Singapore Funding Initiative and administered by the IDM Programme Office.

REFERENCES

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2355697, IEEE Transactions on Circuits and Systems for Video Technology


Jian Dong Dr. Jian Dong is now a Research Scientist of IBM Research, Australia. He was a Research Fellow of the Electrical and Computer Engineering Department of National University of Singapore in 2013. He received his B.E degree from University of Science and Technology of China (USTC) in 2010; and his Ph.D. degree in Electrical and Computer Engineering from the National University of Singapore in 2014. His research interests include computer vision and machine learning. He received the winner prizes of both classification and segmentation tasks in PASCAL VOC’12.

Qiang Chen Dr. Qiang Chen is now a Research Scientist of IBM Research, Australia. He was a Research Fellow of the Electrical and Computer Engineering Department of National University of Singapore in 2013. He received his B.E degree from University of Science and Technology of China (USTC) in 2010; and his Ph.D. degree in Electrical and Computer Engineering from the National University of Singapore in 2014. His research interests include computer vision and pattern recognition. He received the best student paper award PREMIA’12 and the winner prizes of the classification task in both PASCAL VOC’10 and PASCAL VOC’11, the honorable mention prize of the detection task in PASCAL VOC’10.

Jiashi Feng Jiashi Feng is a postdoc researcher working with Prof. Trevor Darrell at University of California, Berkeley. He received his phd degree from National University of Singapore (NUS) in 2014. He got his bachelor degree in automation from University of Science and Technology of China (USTC). He is interested in both computer vision and machine learning. In particular, his research works focus on object recognition, attributes learning, robust optimization, online and distributed learning.
Kui Jia
Kui Jia received the B.Eng. degree in marine engineering from Northwestern Polytechnical University, China, in 2001, the M.Eng. degree in electrical and computer engineering from National University of Singapore in 2003, and the Ph.D. degree in computer science from Queen Mary, University of London, London, U.K., in 2007. He is currently a Visiting Assistant Professor in the Faculty of Science and Technology, University of Macau. His research interests are in computer vision, machine learning, and image processing.

Zhongyang Huang
Mr. Zhongyang Huang received his Master degree in Information Engineering from Nanyang Technological University (Singapore) in 2001 and Bachelor degree in Biomedical Engineering from Shanghai Jiaotong University (China) in 1993 respectively. From 1994 to 1999, he worked as a senior engineer for medical apparatus development in medical image processing area with China-America Joint Venture Kang Ming Biomedical Engineering Ltd. in China. Since 2001 he has been working as a Senior Staff R&D Engineer of Panasonic Singapore Laboratories in Singapore. He received the winner prizes of the classification task in PASCAL VOC’10 / VOC’11. Mr. Huang has authored or co-authored 1 book chapter, 20 technical papers and holds over 15 granted patents with numerous others pending in related fields.

Shuicheng Yan
Dr. Yan Shuicheng is currently an Associate Professor in the Department of Electrical and Computer Engineering at National University of Singapore, and the founding lead of the Learning and Vision Research Group (http://www.lv-nus.org). Dr. Yan’s research areas include computer vision, multimedia and machine learning, and he has authored or co-authored nearly 300 technical papers over a wide range of research topics. He is an associate editor of IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT) and ACM Transactions on Intelligent Systems and Technology (ACM TIST), and has been serving as the guest editor of the special issues for TMM and CVIU. He received the Best Paper Awards from PCM’11, ACM MM’10, ICME’10 and ICIMCS’09, the winner prizes of the classification task in both PASCAL VOC’10 and PASCAL VOC’11, the honorable mention prize of the detection task in PASCAL VOC’10, 2010 TCSVT Best Associate Editor (BAE) Award, 2010 Young Faculty Research Award, 2011 Singapore Young Scientist Award, 2012 NUS Young Researcher Award, and the co-author of the best student paper awards of PREMIA’09, PREMIA’11 and PREMIA’12.