

Efficient Evaluation of SVM Classifiers using Error Space Encoding

Nisarg Raval
 Department of Computer Science
 Duke University

Rashmi Tonge, C. V. Jawahar
 Center for Visual Information Technology
 IIT Hyderabad

Abstract—Many computer vision tasks require efficient evaluation of Support Vector Machine (SVM) classifiers on large image databases. Our goal is to efficiently evaluate SVM classifiers on a large number of images. We propose a novel Error Space Encoding (ESE) scheme for SVM evaluation which utilizes large number of classifiers already evaluated on the similar data set. We model this problem as an encoding of a novel classifier (query) in terms of the existing classifiers (query logs).

With sufficiently large query logs, we show that ESE performs far better than any other existing encoding schemes. With this method we are able to retrieve nearly 100% correct top-k images from a dataset of 1 Million images spanning across 1000 categories. We also demonstrate application of our method in terms of relevance feedback and query expansion mechanism and show that our method achieves the same accuracy 90 times faster than exhaustive SVM evaluations.

I. INTRODUCTION

With the advent of sophisticated features like SIFT [8], representations like Bag of Words [14] and powerful classifiers like Support Vector Machine (SVM), it is now possible to capture the visual content more effectively. This is often modeled as visual categorization. Most of the successful categorization schemes use SVMs, where the category (query) is represented as a hyperplane (w) which separates images of a specific category from the rest of the images. We are interested in efficient evaluation of linear SVM classifiers over a large set of images that returns k highest scoring (top-k) images as a result. Problems such as discriminative mining of concepts [13], incremental learning for interactive vision [12], large scale image retrieval [7] etc. often require evaluation of multiple SVM classifiers on large datasets.

Exhaustively evaluating query w on a large database is computationally inefficient. It becomes even more difficult as the number of categories grow rapidly. Even for linear SVMs, it takes $\mathcal{O}(md)$ operations to evaluate a query on the database of m images represented by d dimensional features. As many state of the art feature representations have very high dimensions (e.g. Bag of Words), we are interested in exploring an evaluation scheme whose computational complexity is independent of d without compromising the accuracy.

One approach to reduce the SVM evaluation time is pruning of the images to retrieve top-k results [10]. Although the method works very well in case of binary features, it doesn't perform well on popular real or integer valued features. Song *et al.* [15] proposed *sparselet* model based on sparse coding

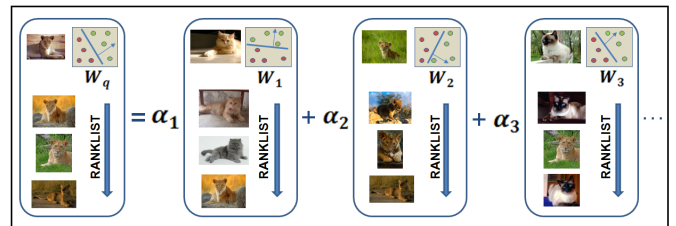


Fig. 1: Conceptual View of our Approach. The idea is to represent a new query (SVM hyperplane) in terms of previously evaluated queries ($w_q = \sum_{i=1}^p \alpha_i w_i$). Then, pre-computed rank lists can be efficiently merged to retrieve results. Thus, we avoid evaluating SVM query on entire database which is computationally intensive task.

of the part filters. The number of classes to be considered in such a setting is still quite limited, and often known apriori. Raval *et al.* [11] proposed a method to learn *eigen queries* using principal component analysis and used them for efficient image retrieval. However, this method fails to handle queries which are not in the span of the limited number of eigen vectors considered.

With explorations in large visual hierarchies of concepts, semantic relationship in feature space has become very important [4]. We argue that evaluation of certain classifiers (categories) can give us useful information about evaluation of novel (both related and distant) categories. In this work, we approximate the new query classifier with the help of a set of already evaluated similar classifiers. The motivation behind our work is in the long tail distribution of categories (queries) [17]. With very few distinct queries, one can answer similar new queries without much computation. For example, even though two queries “mountain” and “hills” are different, one can be approximated by another without significantly affecting their results. Also, there may be cases where the result of a new query can be easily computed using the results of few related existing queries. For example, results of a query “garden” can be computed using queries like “flowers”, “lawn”, “plants” etc. We also address the problem of rare queries which are very different from those in the query logs.

We propose an efficient technique to evaluate linear SVMs on large image databases by exploiting the history of the previous SVM evaluations. The complexity of our algorithm

is independent of the dimensionality of data (d). Intuitively, we transform the query from d dimensional feature space to p dimensional concept space. Here $p \ll d$, because the number of distinct concepts needed to approximate a novel query are very less [17]. The knowledge of these concepts (classifiers) can be gained through previous queries (query logs). Results (rank lists) of a set of classifiers in query logs are precomputed on the database. To evaluate a novel query, we first transform the query into concept space by representing it in terms of known classifiers (basis classifiers). Then, result (rank list) of the novel query is efficiently computed by merging precomputed rank lists of basis classifiers (see Figure 1). The proposed approach can be categorized as a special kind of ensemble method [18].

II. EFFICIENT EVALUATION OF LINEAR SVMs

We propose fast and scalable evaluation of SVM classifiers on a large number of images. We start with an assumption that the system has enough access to reasonably large query logs, i.e. a large set of popular queries from the past.

A. Query Logs and Indexing

As a running example, we consider a database which contains 1500 images from 20 categories of Caltech256 [6]. The database is called Var20 and is described in [3]. We use dense SIFT [8] based Bag of Words representation with a vocabulary size of 1K visual words. We consider the query logs as a set of n distinct SVM classifiers $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$. All these classifiers are evaluated on the entire database which contains m images $\mathbf{x}_1, \dots, \mathbf{x}_m$. The SVM scores, $s_{ij} = \mathbf{w}_i^T \mathbf{x}_j$, are the confidence of the j^{th} image about the i^{th} concept. We initialize the query logs with 5K classifiers built for 10 categories. These classifiers are built using linear SVM [2] by randomly selecting 20 positive images and 100 negative images. Similarly, we generate 800 queries, out of which 500 are either duplicate or almost same (*Repeated Queries*) and rest 300 are the ones for which similar queries are present in the query logs (*Frequent Queries*). For the rest 10 categories, we generate another 200 queries for which there are no similar queries available in the query log (*Rare Queries*).

B. Representation of Novel Queries

Given a novel query \mathbf{w}_q , we want to represent it as a linear combination of existing queries.

$$\mathbf{w}_q = \sum_{i=1}^p \alpha_i \mathbf{w}_i \quad (1)$$

In such a setting, there are many questions to be answered. How to select p ? How do we obtain a set of useful \mathbf{w}_i for a given query? How do we compute α_i ? For efficiency, we want p to be small and α to be sparse. The optimal α can then be found by solving the below problem.

$$\underset{\alpha}{\text{minimize}} \quad \|\mathbf{w}_q - \sum_{i=1}^p \alpha_i \mathbf{w}_i\|$$

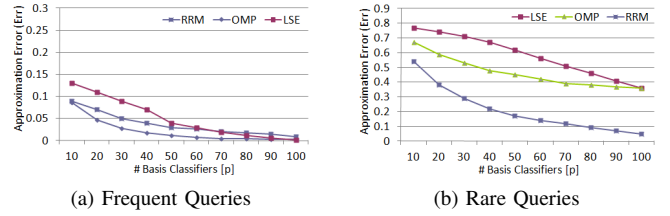


Fig. 2: Comparison of approximation error using various methods. In case of frequent queries all LSE, OMP and RRM gives similar results. But in case of rare queries RRM gives far better approximation compared to LSE/OMP.

The approximation error Err is defined as

$$Err = (\mathbf{w}_q - \sum_{i=1}^p \alpha_i \mathbf{w}_i)^T (\mathbf{w}_q - \sum_{i=1}^p \alpha_i \mathbf{w}_i) \quad (2)$$

Various popular methods like Least Squared Error (LSE) [16] or Orthogonal Matching Pursuit (OMP) [9] solve this problem optimally. However, in our case we also need to select useful set \mathbf{W}^P from \mathcal{W} . Thus, the above objective function will change as shown below.

$$\underset{\alpha, \mathbf{W}^P}{\text{minimize}} \quad \|\mathbf{w}_q - \sum_{i=1}^p \alpha_i \mathbf{w}_i\|, \mathbf{w}_i \in \mathbf{W}^P$$

Many approximate solutions exist to solve this problem by fixing \mathbf{W}^P . For example, one might apply principal component analysis (PCA) on query logs and fix \mathbf{W}^P as top eigen vectors [11]. Instead of PCA, other methods like linear discriminant analysis (LDA) [5] can also be used to achieve the same goal. However, in all these methods \mathbf{W}^P is computed independently of the query \mathbf{w}_q and the same \mathbf{W}^P is used to approximate all future queries. This generalization results into poor approximation¹ for many queries which are out of span of \mathbf{W}^P . We claim that a new query can be better approximated using similar queries from the past. Therefore, we find the suitable \mathbf{W}^P for every query based on the p nearest neighbours of \mathbf{w}_q from the query logs (\mathcal{W}). Selecting nearest neighbours results into smaller coefficients (during approximation) which leads to faster convergence.

Since we fix \mathbf{W}^P , the problem becomes that of minimizing α , which can be solved using techniques like LSE/OMP as discussed above. As the dimensionality d of the space could be much higher than the concepts present in the database (i.e. $p \ll d$), this approximation could be very poor. One direction to reduce the error rate is to increase p as shown in Figure 2. This draws an interesting conclusion; if similar queries exist in the query log, the approximation is more accurate. While for completely different queries, the approximation could be very poor due to lack of representative basis classifiers in the query neighbourhood.

To improve the approximation for rare queries, we propose a different technique of query approximation – Recursive

¹For the same number of basis classifiers approximation error (Err) using PCA is more than 4 times the Err using our method.

Algorithm 1 Recursive Residual Minimization (RRM)

Input: Query \mathbf{w}_q , error threshold ϵ , #Neighbours l **Output:** Approximated Query \mathbf{w}'_q

- 1: $\mathbf{w}'_q = 0$
 - 2: $r = \mathbf{w}_q$
 - 3: **while** $\|r\| \geq \epsilon$ **do**
 - 4: $\mathbf{w}_i = \text{Neighbours}(r), i = 1 \rightarrow l$, where \mathbf{w}_i is i^{th} nearest neighbour of r
 - 5: Find α_i such that $\|r - \sum_{i=1}^l \alpha_i \cdot \mathbf{w}_i\|$ is minimized
 - 6: $r = r - \sum_{i=1}^l \alpha_i \cdot \mathbf{w}_i$
 - 7: $\mathbf{w}'_q = \mathbf{w}'_q + \sum_{i=1}^l \alpha_i \cdot \mathbf{w}_i$
 - 8: **end while**
-

Residual Minimization (RRM). RRM solves the above minimization problem by finding best \mathbf{w} and its corresponding coefficient together. The idea is to minimize the error by recursively approximating the residue r (difference between query and its approximation) with its l nearest neighbours, where $l \ll p$. For a given query \mathbf{w}_q finding approximated query \mathbf{w}'_q using RRM is outlined in Algorithm 1.

We compare the approximation process with (i) LSE (ii) OMP and (iii) RRM in Figure 2. For a frequent query, all three methods give low approximation error, which is obvious since the representative classifiers are available in the query log. But when the query is rare, most of the basis classifiers are unrepresentative as they are usually very far from the query in the feature space. Therefore, both LSE and OMP can not guarantee an optimal solution with the nearest neighbours. In our scheme, instead of relying on neighbours of the query (as they can be very far), we find the neighbours of residues and using them we approximate the residues recursively. This recursion process is guaranteed to converge since at each step we are minimizing the approximation error Err . Hence, for rare queries our scheme achieves very low approximation error. However, for frequent queries LSE/OMP performs slightly better than RRM because it considers all neighbours at once and find globally optimal solution.

To take advantage of both the methods we consider a novel technique called Error Space Encoding (ESE). ESE chooses the optimum approximation method on the fly based on the error as shown in Algorithm 2. ESE ensures low approximation error Err by selecting a method apt for the nature of the query. At the same time it avoids unnecessary computations required for optimization.

C. Evaluation of Novel Queries

Once, \mathbf{w}_q is approximated using ESE, the SVM score for any image \mathbf{x}_j can be efficiently computed using query logs and precomputed scores (Section II-A) as shown below.

$$s_{qj} = \mathbf{w}_q^T \mathbf{x}_j = \sum_{i=1}^p \alpha_i \mathbf{w}_i^T \mathbf{x}_j = \sum_{i=1}^p \alpha_i s_{ij}$$

The scoring requires dot product of two p dimensional vectors – coefficients (α) and precomputed scores (s). Exhaustive evaluation of SVM requires dot product of two

Algorithm 2 Error Space Encoding (ESE)

Input: Query \mathbf{w}_q , distance thresholds ϵ_1 and ϵ_2 , #Basis Classifiers p , #Neighbours l **Output:** Approximated Query \mathbf{w}'_q

- 1: $\mathbf{w} =$ nearest neighbour of \mathbf{w}_q from the indexed query log
 - 2: $dist = \|\mathbf{w} - \mathbf{w}_q\|$
 - 3: **if** $dist < \epsilon_1$ **then**
 - 4: $\mathbf{w}'_q = \mathbf{w}$
 - 5: **else if** $dist < \epsilon_2$ **then**
 - 6: Find p nearest neighbours of \mathbf{w}_q from query log
 - 7: Use LSE/OMP to find \mathbf{w}'_q by solving Equation 1
 - 8: **else**
 - 9: Use RRM to find \mathbf{w}'_q (Algorithm 1)
 - 10: **end if**
-

d dimensional vectors. Thus, we gain $\mathcal{O}(\frac{d}{p})$ speedup by approximating query using p basis classifiers.

When we encounter an extremely rare query which is out of span of query logs, we add it to the query logs as it represents a completely new concept. Due to such active learning we can compute the scores for future queries much efficiently. Similarly we remove the least frequent classifiers from query logs periodically. This process of active learning does not require any additional computation and hence can be performed on the fly. Thus we develop a complete solution leading to *Life-long Learning*.

D. Speedup

The proposed ESE technique intelligently chooses the optimal method out of Nearest Neighbour (NN), LSE and RRM such that it maximizes both accuracy and speedup. Thus, depending on the query, the speedup varies based on the selected method. For example, if the query is very close to one of the queries in the query log (less than ϵ_1 distance apart) the time to evaluate query is only that of finding nearest neighbours from indexed query log. Hence, the overall speedup over all queries is the ratio of the time taken by exhaustive method to the time taken by the corresponding methods.

According to power law distribution of queries [17], most of the queries come from a small set of distinct queries and they also repeat very often. For all such queries our system gives very high speedup as it only requires to perform nearest neighbour search over small number of basis classifiers as opposed to exhaustively evaluating SVM over large database. We simulate this scenario by evaluating previously constructed 1000 queries and measure the overall speedup. We find that ESE is 34 times faster than exhaustive SVM evaluation with comparable accuracy. In addition to ESE various pruning techniques like *Subset Search* [11] and *Top-K Ranking* [10], can be used to achieve high speedup. However, in this paper we focus only on the encoding scheme itself.

III. EXPERIMENTS

We evaluate the proposed methods on two image datasets – Caltech256 [6] and Large Scale Visual Recognition Challenge 2010 (ImageNet) [1]. First, we measure the performance

Iterations	Recall@10			Discrepancy		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
10	58.1	61.8	69	0.31	0.29	0.20
30	73.2	89.2	91.9	0.18	0.05	0.03
50	86.2	93.8	96.5	0.09	0.02	0.02
80	90.9	96.9	99.1	0.03	0.02	0.006
100	94.8	98.9	99.2	0.02	0.009	0.004

TABLE I: Retrieval Quality of RRM using rare queries on Var20. Increasing neighbourhood gives better approximation as more relevant basis classifiers are being used for approximation. The low value of Discrepancy shows that incorrectly retrieved images are actually part of the query concept.

with respect to various parameters on a smaller subset of Caltech256 – Var20. Then, we show the scalability of our approach by performing experiments on ImageNet. In case of ImageNet, we use publicly available features – dense SIFT [8] with histogram of Bag of Words having vocabulary size of 1K. Finally, we discuss how our method can be useful for various computer vision tasks like relevance feedback and query expansion.

A. Performance Analysis

In this section, we analyze the performance of our method on Var20. The database construction and query log generation is described in Section II-A. In addition to the learned 5K SVM hyperplanes, we use 5K randomly generated hyperplanes to construct sufficiently large query log. The *error* threshold ϵ_1 and ϵ_2 can be determined empirically by running few frequent and rare queries on the training data. In our experiments we set $\epsilon_1 = 0.1$ and $\epsilon_2 = 0.3$ and we use cosine distance. In all our experiments we use $k = 10$ unless otherwise stated. We retrieve $2k$ images and then perform SVM evaluation on those to retrieve top-k results. The measurements used to evaluate the performance of our methods are as follows.

Recall@top-k: We construct the *truthset* by exhaustively evaluating query on all the images and retrieving top-k results. Recall@top-k is the percentage of images from *truthset* that are correctly retrieved by our method. Recall@top-k shows how accurate our results are with respect to the baseline (exhaustive evaluation of SVM query).

Discrepancy (Desc): We use discrepancy as a measure of the quality of our top-k retrieved results. It is the difference between the average distance from the query hyperplane of the images in *truthset* and the images we retrieved. Having low discrepancy shows that, even though retrieved results may not be exactly same as *truthset*, they are comparable in terms of their SVM scores.

Rare Queries: To see how our method (RRM) performs on rare queries we run multiple rare queries on Var20 and measure Recall@10 and Discrepancy. We vary the number of basis classifiers $p = iter \times l$ by varying number of iterations *iter* and the neighbourhood size *l*. The results are summarized in Table I. Note, that increasing p results into increase in Recall@10 which is obvious as more basis classifiers are used to reduce the residues. With only 100

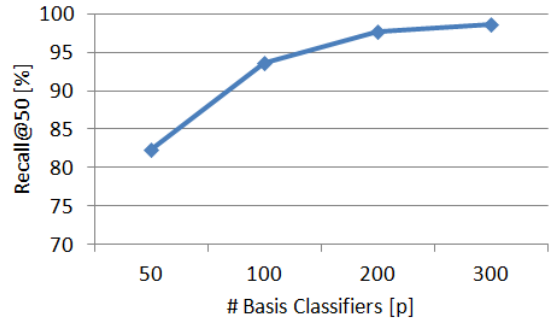


Fig. 3: Performance of ESE on more than 1 million images with 1000 categories.

iterations and neighbourhood size of 3, RRM achieves 99.2% Recall@10 with only 0.004 Discrepancy. The small number of Discrepancy in results shows that the incorrectly retrieved images are indeed part of a query concept and for all practical purposes the method is as good as exact.

Large Scale SVM Evaluation: To measure the scalability of our method we scale our dataset Var20 by adding 1 million images from ImageNet spanning across 1000 categories. We also add 100K randomly generated hyperplanes to query log to scale it accordingly. The results of top-50 retrievals are shown in Figure 3. Even for such a large scale data we achieve nearly 100% Recall@50 with only 300 basis classifiers.

B. Comparison

In this section we show how the proposed method performs in comparison to the existing techniques. First, we compare various approximation schemes in terms of their retrieval quality. We evaluated 100 known queries and 100 unknown queries on Var20 using three methods LSE, RRM and ESE. We vary the number of basis classifiers from 10 to 100 to measure the Recall@10 as shown in Figure 4. The figures shows that both RRM and ESE give consistently high Recall@10 compared to LSE for the same number of basis classifiers. Also, ESE achieves slightly better Recall@10 as it dynamically selects either LSE or OMP to gain low approximation error.

Next, we compare ESE with recent technique of approximation through *eigen queries* [11]. The method approximate the query using LSE, which performs poorly on rare queries (as shown in Section II-B). On the other hand our method approximates the repeated queries with the nearest query in query logs without extra computations. We compare our method with [11] on Var20 dataset by executing 100 frequent and 100 rare queries. We vary the number of basis classifiers/eigen queries and compute average Recall@10 over all queries. With same number of basis classifiers and eigen queries, ESE achieves 20% more Recall@10 compared to [11]. Also, updating eigen queries to incorporate rare queries is computationally intensive task because eigen decomposition has to be performed on entire set of basis classifiers. While, in case of ESE the updating process is

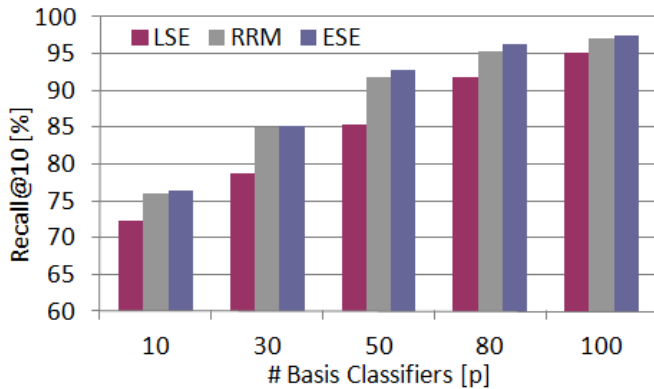


Fig. 4: Comparison of retrieval quality of various encoding methods on Var20. ESE performs consistently better than both LSE and RRM by ensuring minimum approximation error.

as simple as adding rare queries directly to the set of basis classifiers and does not require any additional computation.

Finally, we compare our method with Top-K Ranking [10]. The idea behind Top-K Ranking is to compute partial SVM scores at each step and update upper and lower bounds of the SVM scores for every image. Based on updated bounds the images which can not be in top-k results are pruned at every iteration. Although, Top-K Ranking is generic and can be applied to any kind of features, it is efficient only for binary features. We show that in case of real/integer valued features the pruning rate is very slow because of the slow updates on bounds in each iteration. On the database of 1500 images (Var20) with 1K dimensional features, our method is 10 times faster than Top-K Ranking and achieves 100% accuracy in terms of Recall@10. The speedup is due to the fact that, on an average Top-K Ranking accesses 970 dimensions out of 1K and the pruning rate is very slow in each iteration due to real valued features. Note, that Top-K Ranking is an exact solution while ours is an approximate one.

C. Analysis of Query Approximation

Since, we represent query classifier as a linear combination of the basis classifiers, several questions may arise - What basis classifiers are used to represent a specific query classifier? How well we can represent a query? In this section, we give qualitative results and try to answer these questions.

We use ImageNet as it has many categories with hierarchical category structure which gives intuitive meaning to the representation. We use a subset of ImageNet hierarchy which comes under *Vehicle* category as shown in Figure 5a. The query log contains classifiers for every category under *Vehicle*. Next, we construct a novel query which represents a category *Car*, by selecting positive images from every sub-category of *Car* and negatives from other categories. The constructed query is approximated by both LSE and ESE. The high valued coefficients with their respective basis classifiers are shown in Figure 5b. We can see from the figure that both LSE and ESE give higher weights to the classifiers which belongs to

Method	Precision@10			Time (second)
	<i>iter</i> 1	<i>iter</i> 3	<i>iter</i> 5	
Exhaustive	70	78.33	81.66	0.0271
ESE	70	78.33	81.66	0.0003

TABLE II: Performance of Query Expansion using ESE on Var20 dataset. ESE is 90 times faster than exhaustive SVM evaluation while achieving same accuracy.

the sub-categories of *Car*. This is intuitive because query is made of images from those categories so in turn it is broken down into those categories. Note, that ESE has slightly better approximation compared to LSE, as in case of LSE *Pickup Truck* is used to approximate the query. However, the weight assigned to it is relatively low.

The categories which are similar to the query category are given higher weights and dissimilar ones are given lower weights. In case of queries which are out of span of the query log, the approximation may be poor. Those cases can be handled by adding rare query into the query log without any overhead. Thus, the proposed system is designed to adapt to the trending queries by continuously learning, and improving the performance (both in accuracy and time) significantly.

D. Relevance Feedback and Query Expansion

Relevance feedback is a popular mechanism to narrow down the gap between retrieved and intended results. Often user provides feedback by specifying correct images in the retrieved results. Based on the feedback the system modifies the search query and provides refined results to the user. Usually many feedback iterations with such refinement are required to capture the real intent of the user. On every iteration, the modified query is evaluated on the entire database which is very time consuming. We propose an efficient relevance feedback mechanism using ESE to efficiently refine the results of a modified query.

If \mathbf{w}_i is the query in i^{th} iteration, then $\delta\mathbf{w} = \mathbf{w}_i - \mathbf{w}_{i-1}$ is an update in i^{th} iteration. We can model this $\delta\mathbf{w}$ as a residue in our method. This will allow us to efficiently evaluate \mathbf{w}_i , just by computing approximation of $\delta\mathbf{w}$ and using precomputed scores of \mathbf{w}_{i-1} . The advantage of evaluating $\delta\mathbf{w}$ instead of \mathbf{w}_i , is that the modification $\|\delta\mathbf{w}\|$ on every iteration will be much smaller than $\|\mathbf{w}_i\|$. This will enable accurate approximation with very less number of basis classifiers p which will in turn result into high speedup.

We tested ESE in the relevance feedback framework using a subset of ImageNet hierarchy which comes under *Vehicle* category. The results of four selected iterations with their ranked list is shown in Figure 6. The initial query is generated by randomly selecting positive images of various car types and negatives images which do not belong to any sub-category of car. We simulate the user's feedback where user intends to retrieve images of *Race Car*. At every iteration we keep adding 5 images of a *Race Car* for this purpose. With only 10 iterations we were able to get *Race Car* in all top-k results. It is 50 times faster than exhaustive evaluation over all 10 iterations as it uses only 10 basis classifiers in each iteration.

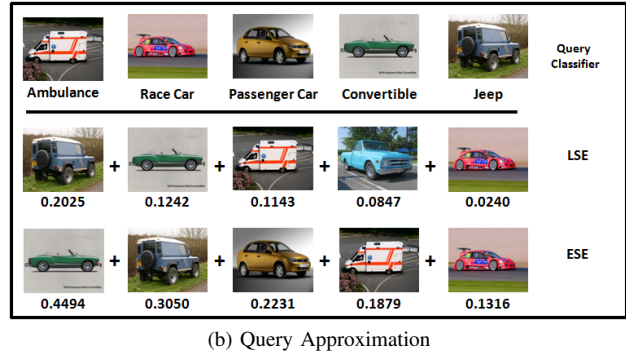
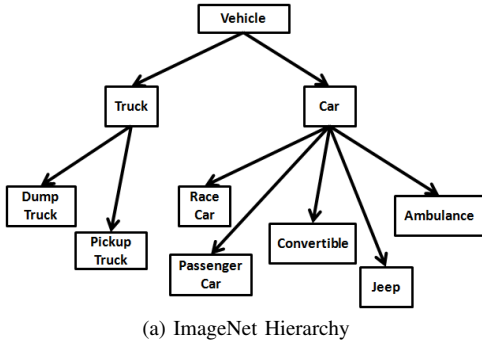


Fig. 5: Figure 5a shows the subset of ImageNet hierarchy used for the experiments. Figure 5b shows an example of a Query Approximation for a specific category of *Car*. The query category *Car* is approximated by the classifiers of its sub-categories.



Fig. 6: Relevance Feedback using ESE. We show top-5 results for few iterations, where 5 positive images are added in each iteration. The top-k results on every iteration change according to the specified feedback for intended category *Race Car*.

The proposed method ESE can also be used in more general setting of Query Expansion (QE). Instead of user feedback, QE uses top-k retrieved results of every iteration to refine the query. In each iteration, top-k results are considered to be correct and added as positive examples of the query. For the same reason explained above, ESE can be used to efficiently perform QE as the change in every iteration will be very small resulting into high speedup with comparable accuracy. We use various categories from Var20 dataset as queries and perform QE using ESE. Total 5 iterations are performed for every query and in each iteration 5 top ranked images are added as positive examples. Results averaged over all queries are summarized in Table II. While we achieve same accuracy for both ESE and exhaustive evaluation, ESE is 90 times faster than the baseline method.

IV. CONCLUSIONS

We achieve faster evaluation of novel SVM queries with high accuracy using past SVM queries. We propose RRM encoding scheme which gives near exact approximation even for rare queries. We also introduce a *complete* encoding scheme, ESE giving “best-of-all” performance. We get nearly 100% correct images in top-k, much faster than exhaustive evaluation. We also show contribution of our method towards important applications such as Large Scale SVM Evaluations, Relevance Feedback and Query Expansion.

REFERENCES

- [1] J. D. A. Berg and L. Fei-Fei. Large scale visual recognition challenge, 2010.
- [2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011.
- [3] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. *JMLR*, 2010.
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. 2000.
- [6] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical report, California Institute of Technology, 2007.
- [7] H. Jegou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *TPAMI*, 2011.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.
- [9] Y. Pati, R. Rezaifar, and P. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers*, 1993.
- [10] M. Rastegari, C. Fang, and L. Torresani. Scalable object-class retrieval with approximate and top-k ranking. *ICCV*, 2011.
- [11] N. Raval, R. Tonge, and C. V. Jawahar. Image retrieval using eigen queries. In *ACCV*, 2012.
- [12] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental learning for robust visual tracking. *IJCV*, 2008.
- [13] S. Singh, A. Gupta, and A. A. Efros. Unsupervised discovery of mid-level discriminative patches. In *ECCV*, 2012.
- [14] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. *ICCV*, 2003.
- [15] H. Song, S. Zickler, T. Althoff, R. Girshick, M. Fritz, C. Geyer, P. Felzenszwalb, and T. Darrell. Sparselet Models for Efficient Multiclass Object Detection. In *ECCV*, 2012.
- [16] G. Strang. Introduction to applied mathematics. 1986.
- [17] A. Torralba, R. Fergus, and W. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *TPAMI*, 2008.
- [18] Z.-H. Zhou. *Ensemble Methods: Foundations and Algorithms*. 2012.