

One-shot Learning of Sketch Categories with Co-regularized Sparse Coding

Yonggang Qi¹, Wei-Shi Zheng², Tao Xiang³, Yi-Zhe Song³,
Honggang Zhang¹, and Jun Guo¹

¹ School of Information and Communication Engineering, BUPT, Beijing, China

² School of Information Science and Technology, Sun Yat-sen University, China

³ School of EECS, Queen Mary, University of London, London E1 4NS, U.K.

Abstract. Categorizing free-hand human sketches has profound implications in applications such as human computer interaction and image retrieval. The task is non-trivial due to the iconic nature of sketches, signified by large variances in both appearance and structure when compared with photographs. Prior works often utilize off-the-shelf low-level features and assume the availability of a large training set, rendering them sensitive towards abstraction and less scalable to new categories. To overcome this limitation, we propose a transfer learning framework which enables one-shot learning of sketch categories. The framework is based on a novel co-regularized sparse coding model which exploits common/shareable parts among human sketches of seen categories and transfer them to unseen categories. We contribute a new dataset consisting of 7,760 human segmented sketches from 97 object categories. Extensive experiments reveal that the proposed method can classify unseen sketch categories given just one training sample with a 33.04% accuracy, offering a two-fold improvement over baselines.

1 Introduction

Sketch is used to render the visual world since prehistoric times. Closely correlated with the increasing availability of digital touch-screen devices, research on human sketches has begun to return to the center stage with important applications such as sketch-based image retrieval (SBIR) and sketch recognition.

Sketches are intuitive to humans and descriptive in nature. They can conveniently capture object pose, configuration and fine appearance details. It was shown recently that although humans are highly capable of identifying sketches, it remains a very challenging task for computers [1]. Automatically recognizing sketches is difficult because: (i) sketches are often highly abstract in representation compared with photographs, e.g. a photo of a person can be sketched as a stick-man, (ii) sketches are hand-drawn by people with different levels of artistic skills, as a result they often do not conform precisely to natural image boundaries, (iii) sketches lack visual cues (e.g. color and texture) commonly used in image understanding.

Most prior works [1–3] address the sketch recognition problem following a standard supervised learning pipeline widely adopted for object recognition. That is, first a large number (e.g. hundreds) of labeled instances are collected for each

class; then followed by feature extraction and finally learning a classifier. Due to the unique characteristics of sketches described above, many existing works focus on designing features specifically engineered for sketches [4–6]. However, one critical problem has largely been ignored – it is extremely difficult to collect sufficient training samples, especially for large number of visual categories. Recently the problem of lack of training data has attracted increasing attentions for natural images due to the need for large scale learning of thousands of visual categories [7]. In particular, many works exploit the idea of transfer learning using an intermediate level semantic representation such as attributes [8] so that recognition can be achieved even without any training samples, i.e. zero-shot. It thus comes as a surprise that no one has so far considered this lack of training data problem for sketch recognition, because this problem is much more acute for sketches – while almost unlimited number of images can be found for each visual category on media-sharing sites such as Flickr, much fewer sketches are uploaded and made available on the Internet.

In this paper, we address this lack of training data problem by developing a novel one-shot learning framework. Our framework enables the learning of a sketch classifier using only one training sample for each class. Similar to previous one-shot learning work [9], the framework takes advantage of knowledge transferred from previously learned categories, no matter how different these categories might be. In particular, we make use of common sketch parts learned from an auxiliary set labeled by human and utilize them in a sparse coding based one-shot learning framework. Our underlying hypothesis is that common parts exist among sketches from distinct object categories (e.g. wings of ‘bird’ and ‘airplane’). The common parts can then be learned as a set of sparse codes from the auxiliary set and used as transferrable knowledge to help learn a classification model for the target classes. We importantly introduce a novel co-regularized sparse coding algorithm where the sparse coding models for both the auxiliary set and target set are learned jointly. The objective is to make sure that the resulting sparse representation agrees as much as possible between the two sets.

More specifically, the proposed one-shot co-regularized sparse coding (OCSC) approach has two stages. First, sharable basis (e.g. bird wings for replacing wings of airplane) are discovered from auxiliary set for each novel sketch category in target set. Secondly, considering which categories an unknown sketch most relevant to, the sketch is encoded via the proposed co-regularized sparse coding algorithm that enforces the resulting sparse representation to use as much as possible the relevant basis discovered during the first stage (e.g. suppose we know the given unknown sketch is an ‘apple’, it would likely to be encoded by the basis of ‘tomato’ and ‘peach’ in the auxiliary set, because of their similar looking). To perform categorization, we employ a sparse representation classifier (SRC) [10].

The contributions of this work can be summarized as follows: (i) As far as we know, this is the first work on one-shot learning for sketch recognition. (ii) We introduce a novel transfer learning framework based on co-regularized sparse coding. (iii) We create a new dataset containing over 7760 sketches in

97 categories with manually labeled parts, based on the 20,000 human sketches dataset [1].

2 Related Works

Sketch recognition There exist plenty of works on sketch recognition [1, 5], most of which employ a bag-of-visual-words (BoVW) representation coupled with local features. Some of the features can be commonly found in the vision literature [2], while others are specifically engineered for sketches [4, 6]. Of all features tested, it was shown that Histogram of Oriented Gradient (HOG) based features are among the most effective ones [2, 3]. Despite being useful, unstructured local features are often incapable of capturing the relatively high degree of intra-class variance and inter-class ambiguity associated with human sketches. Yi et al. [5] tackled this problem by proposing a novel mid-level sketch representation in the form of a star-graph that encapsulates local features to encode holistic object structure, thereby offering the state-of-the-art performance to date on the 20,000 sketches dataset [1]. Nonetheless, existing approaches often assume the availability of a large number of training data, which seriously limits their scalability to new categories.

One-shot learning Although one-shot sketch recognition is an unstudied topic, one-shot learning has been exploited in related vision topics [9, 11]. Recent works on attributes learning [12, 13] demonstrate that human-defined shareable attributes can be an intuitive mechanism for transfer learning. In this work, we utilize human segmented sketch parts as the shareable components between seen and unseen categories. This is drastically different from the previous one-shot learning works in terms of how transfer learning is enabled: (i) Compared with the general part-shape prior knowledge based transfer learning [9, 11], the human segmented parts provide much stronger constraints and thus are much more informative. (ii) Compared with the semantic attribute based approaches [12, 13], our human segmented parts do not have to conform to a visual concept ontology, which are more data-driven, and are regularized by the target data to provide more discriminative information. In other words, without relying on a human defined ontology, it is more flexible and can be used for recognizing many more new visual categories.

Sparse Coding For transferring part-based source sketch dictionary information, a co-regularized sparse coding model is developed. Sparse coding has been widely used in image classification [14, 15], face recognition [16], visual tracking [17] and many other computer vision areas. However, there is no previous work on the use of sparse coding for sketch recognition, despite the fact that sparse coding is intrinsically appropriate for mining sharable parts from human segmented sketches. Importantly, we propose a novel co-regularized sparse coding model and apply it to the new problem of one-shot sketch recognition.

3 One-shot Co-regularized Sparse Coding Algorithm

Given the source/auxiliary dictionary consisting of sketch parts and target dictionary consisting of one-shot target sketches, we propose an one-shot co-regularized

sparse coding (OCSC) approach to obtain sparse representation of an unknown sketch, followed by a sparse representation classifier (SRC) to classify it.

3.1 Notations

We denote matrix $A \in \mathbb{R}^{d \times n}$ the source dictionary representing n sketch parts in a d -dimensional space. The target dictionary is denoted as $B \in \mathbb{R}^{d \times m}$ representing m one-shot instances in the same space, where $B = \{b_1, b_2, \dots, b_j, \dots, b_m\}$, and vector $b_j \in \mathbb{R}^d$ represents the one-shot instance of j -th target category. Therefore, given any unknown sketch y , two different sparse representations, i.e. $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^m$, are obtained based on A and B , respectively.

3.2 Modeling

Basis Discovery The problem of sharable basis discovery is casted into a sparse coding problem. For a target category, sharable basis is discovered from source dictionary by finding the non-zero entries of the sparse representation of its corresponding one shot target example. The intuition is that, the selected parts in source dictionary which are able to perfectly reconstruct the one-shot target instance sketch, should be also qualified to reconstruct other sketches in the same target category. Therefore, given the one-shot target instance b_j , the basis v_j of j -th target category is obtained by:

$$\min_{v_j} \frac{1}{2} \|b_j - Av_j\|_2^2 + \sigma \|v_j\|_1, \quad s.t. \quad v_{ji} \geq 0 \quad (1)$$

where v_{ji} is the i -th entry of v_j , and v_j is a n -dimensional vector whose non-zero entries indicate the relevance between the sketch parts in source dictionary A and the j -th target category; in other words, entries in v_j can be considered as the probabilities of the source sketch parts being relevant to the j -th target category, if a normalization is further imposed.

Co-regularized Sparse Representation Given an unknown sketch y , we firstly determine the relevance between y and each of the target categories by obtaining a sparse representation β according to the target dictionary B :

$$\min_{\beta} \frac{1}{2} \|y - B\beta\|_2^2 + \gamma \|\beta\|_1, \quad s.t. \quad \beta_j \geq 0 \quad (2)$$

where each entries of β indicates the relevance between y and each of the target categories. Secondly, based on the relevance, the co-regularized sparse representation α is obtained by:

$$\min_{\alpha} \frac{1}{2} \|y - A\alpha\|_2^2 + \sigma \|\alpha\|_1 - \frac{\lambda}{m} \langle V^T \alpha, \beta \rangle, \quad s.t. \quad \alpha_i \geq 0, \beta_j \geq 0 \quad (3)$$

where $V = \{v_1, v_2, \dots, v_j, \dots, v_m\}$ is constructed by Eq. (1), and $\langle V^T \alpha, \beta \rangle = \sum_{j=1}^m (\langle v_j, \alpha \rangle \times \beta_j)$. According to the role of v_j in Eq. (1), $\langle v_j, \alpha \rangle$ indicates how strong the resulting sparse representation α is linked to the j -th target category, and the penalty $\langle V^T \alpha, \beta \rangle$ is to guide the learning of α such that the response on the entries relevant to the j -th target category (non-zero entries of v_j) should agree with the response on the corresponding entry in β , i.e. β_j . E.g. $\beta_j = 0$ means the unknown target sketch should be irrelevant to the j -th

target category according to Eq. (2) and therefore the corresponding coefficients in α should be set to 0. The co-regularization process is the reason why we call our model co-regularized sparse coding. We also name the novel penalty $\langle V^T \alpha, \beta \rangle$ as ‘guidance term’, which controls the strength of the co-regularization, therefore the amount of knowledge transferred between the source and target sets.

We address the optimization of Criterion Eq. (3) in the next section by reformulating it as a quadratic program (QP) and further derive an equivalent linear complementary problem (LCP), such that an efficient principle pivoting algorithm can be used to solve the problem.

3.3 Optimization Algorithm of OCSC

Here we give details on the optimization algorithm of our one-shot co-sparse-coding (OCSC) model. To simplify the notations in Eq. (3), we set $g(\alpha) = \langle V^T \alpha, \beta \rangle$. We then re-formulate the problem in Eq. (3) as the following quadratic program:

$$\min_{\alpha} \frac{1}{2} \alpha^T A^T A \alpha + (\sigma - A^T y)^T \alpha - \frac{\lambda}{m} g(\alpha) \quad s.t. \quad \alpha_i \geq 0 \quad (4)$$

Since $A^T A$ is a positive semidefinite matrix, this quadratic program in Eq. (4) is convex, where Karush-Kuhn-Tucker optimal conditions constitute the following monotone linear complementary problem [16]:

$$\delta = A^T A \alpha - A^T y + \sigma - \frac{\lambda}{m} g'(\alpha), \quad \delta \geq 0, \alpha \geq 0, \alpha^T \delta = 0. \quad (5)$$

Here $g'(\alpha) \in \mathbb{R}^n$ is given by the differential of $g(\alpha)$ over α , and the i -th entry $g'(\alpha)_i = \beta_1 v_{1i} + \beta_2 v_{2i} + \dots + \beta_m v_{mi}$. In our problem, the matrix $A^T A$ is always positive definite, so the convex problem in Eq. (4) and the monotone LCP in Eq. (5) thus have unique solutions for each vector y . Next, we describe how a complementary solution can be obtained. Let F and G be two subsets of $\{1, \dots, n\}$ such that $F \cup G = \{1, \dots, n\}$ and $F \cap G = \emptyset$. Then consider the following partition of the matrix A : $A = [A_F, A_G]$, where $A_F \in \mathbb{R}^{d \times |F|}$, $A_G \in \mathbb{R}^{d \times |G|}$, and $|F|$ and $|G|$ are the numbers of F and G , respectively. Based on the partition we reformulate Eq. (5) as the following form:

$$\begin{bmatrix} \delta_F \\ \delta_G \end{bmatrix} = \begin{bmatrix} A_F^T A_F & A_F^T A_G \\ A_G^T A_F & A_G^T A_G \end{bmatrix} \begin{bmatrix} \alpha_F \\ \alpha_G \end{bmatrix} - \begin{bmatrix} A_F^T y \\ A_G^T y \end{bmatrix} + \begin{bmatrix} \sigma_F - \frac{\lambda}{m} g'(\alpha)_F \\ \sigma_G - \frac{\lambda}{m} g'(\alpha)_G \end{bmatrix} \quad (6)$$

where $\alpha_F, \delta_F, \sigma_F, g'(\alpha)_F \in \mathbb{R}^{|F|}$, $\alpha_G, \delta_G, \sigma_G, g'(\alpha)_G \in \mathbb{R}^{|G|}$, $\alpha = (\alpha_F, \alpha_G)$, and $\delta = (\delta_F, \delta_G)$. A complementary basic solution is obtained by setting $\alpha_G = 0$ and $\delta_F = 0$ in Eq. (6), and we can compute the values of the basic variables α_F and δ_G by :

$$\min_{\alpha_F \in \mathbb{R}^{|F|}} \frac{1}{2} \|A_F \alpha_F - y\|_2^2 + \sigma \sum_{i \in F} \alpha_i - \frac{\lambda}{m} \langle g'(\alpha)_F, \alpha_F \rangle \quad (7)$$

$$\delta_G = A_G^T (A_F^T \alpha_F - y) + \sigma_G - \frac{\lambda}{m} g'(\alpha)_G \quad (8)$$

Finally the optimal solution is given by setting $\alpha = (\alpha_F, 0)$ and $\delta = (0, \delta_G)$. Please refer to [18] for more details.

3.4 Classification based on OCSC

In Eq. (3), we aim to reconstruct a test sketch y using parts in source dictionary as well as possible, and parts belonging to the same class of y shall be expected to contribute the most during reconstruction. Therefore, we design a class specific reconstruction classifier similar to the sparse classifier proposed by [10]. More specifically, for each class c , let $\chi_c : \mathbb{R}^n \rightarrow \mathbb{R}^{n_c}$ be a function which selects the coefficients belonging to class c , i.e. $\chi_c(\alpha) \in \mathbb{R}^{n_c}$ is a vector whose entries are the entries in α corresponding to class c . Thus the unknown sketch y is reconstructed as $\hat{y}_c = A_c \chi_c(\alpha)$ only by using the coefficients associated with class c . To this end, y can be classified by assigning it to the class c corresponding to the minimal Earth Mover’s Distance (EMD) between y and \hat{y}_c , which has shown to be suitable for many pattern recognition problems for matching patterns represented as features [19]:

$$\min r_c(y) = EMD(y, \hat{y}_c) \quad (10)$$

4 Experiments

We evaluate the proposed one-shot co-regularized sparse coding (OCSC) algorithm under a sketch recognition framework, and offer comparisons against four standard non-transfer-learning-based alternatives⁴, namely template matching (TM), support vector machine (SVM), SVM with bag-of-words (SVM+BOW) and sparse coding with sparse representation classifier (SC+SRC).

4.1 Datasets and Features

Datasets – A total number of 7760 sketches from 97 categories (80 sketches per category) are first collected from the largest human sketches to date [1]. We then ask annotators to manually label semantic parts of each sketch. The experiment is largely unconstrained where the annotators were free to segment the sketch based on their own subjective criteria. Considering the large number of sketches to annotate, 10 annotators are employed.

Features – Histogram of Oriented Gradient (HOG), the most effective descriptor for sketch according to [5, 1], is employed to encode parts of sketches. Because of the redundancy of the original parts data, we further apply K-means to extract 256 most common parts. A similar practice is also used in [20] to obtain tokens.

4.2 Experimental settings

Among the 97 categories labeled by human, 77 are randomly selected as the source categories and the rest 20 categories reserved for testing. Our goal is to identify which target category an unknown sketch belongs to, given only one instance of each target categories. That is, a total number of 6160 sketches from 77 categories are utilized to form the source part dictionary and 20 sketches from each of the rest 20 categories to form one shot instances, and $20 \times 79 = 1580$ sketches used for testing. For any sketch image, after scaling it into size of 256×256 ,

⁴ Note that few existing transfer learning works handles cross-dataset transfer and none is designed for transferring from human segmented source sketches to one-shot target sketches.

HOG feature is extracted with two main parameters: cell size of 32 pixels and orientation of 9; hence a $8 \times 8 \times 36 = 2304$ dimensional feature vector is obtained as representation for a sketch image.

Since this is the first attempt to perform sketch classification using one-shot learning, there are very few existing work to compare against. Instead, we consider four standard alternatives for comparison:

Template Matching (TM) – where we use the single given sketch instance per target class as template and measure the distance to every unknown sketch, then assign it to the corresponding class with minimal template matching distance. Specifically, each of the 20 one-shot instance sketches is represented by a 2304 dimensional HOG feature vector, which serves as a template to classify the rest 1580 sketches by measuring the EMD distance.

SVM – where all one-shot instances are used for training SVM classifiers to classify unknown sketches. Here we directly use the extracted HOG features of the same training sketches to train classifiers to classify the same 1580 testing sketches same as in **TM**.

SVM+BOW – same as in [1], we employ HOG formed bag-of-words (BOW) as features to train classifiers, which is the most popular strategy for sketch recognition. In particular, we randomly sample 784 local features for each one-shot instance sketch, hence get totally $784 \times 20 = 15,680$ samples to form a 500 visual words vocabulary by Kmeans. Then it results in a 500 dimensional histogram of visual words to represent a sketch for training and testing.

SC+SRC – where all one-shot instances form a dictionary, and the standard sparse coding (SC) algorithm is employed to produce representation for an unknown sketch, followed by a sparse representation classifier (SRC) to classify.

4.3 Results and discussions

We run our one-shot sketch recognition experiment 10 times by randomly sampling 77 source and 20 target categories each time, and report the average recognition accuracy in Table 1. It can be seen that the proposed OCSC method achieves an overall 33.04% classification accuracy, and outperforms SVM, SVM+BOW, TM and SC+SRC by 28.87%, 28.24%, 18.03% and 13.19%, respectively. This result show clearly that useful knowledge has been transferred from the source dataset to help one-shot classification.

SVM	SVM+BOW	TM	SC+SRC	OCSC (Ours)
4.17%	4.80%	15.01%	19.85%	33.04%

Table 1. Comparison on Sketch Recognition Results

Fig. 1 shows an example confusion matrix for our method. It suggests that we can achieve very good classification results on some relatively complex categories even with just one sketch instance, e.g. ‘wheel’ (97.47%), ‘t-shirt’ (91.14%), ‘pumpkin’ (84.81%), ‘wine-bottle’ (83.54%), and ‘computer-monitor’ (74.68%). However, it performs poorly on some other categories, such as ‘snake’ (16.46%), ‘syringe’ (17.72%), ‘banana’ (13.9%), and ‘sword’ (6.33%). We perform less well on these categories because they exhibit relatively simple holistic structures

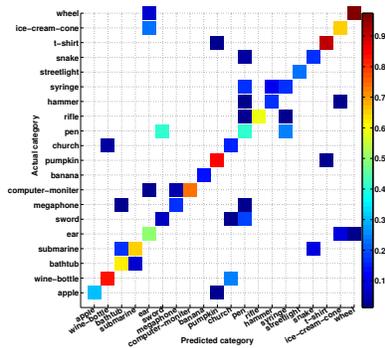
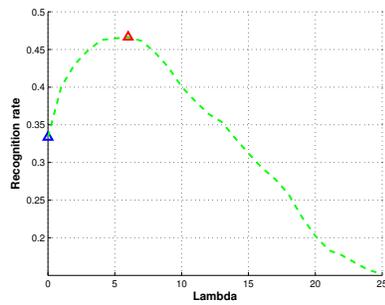
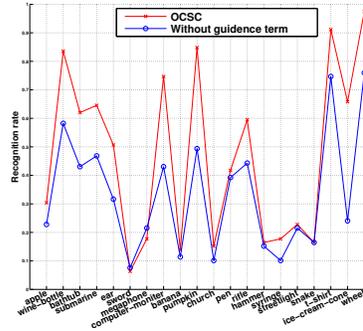


Fig. 1. Classification confusion matrix for randomly selected target categories showing the classification capacity of our proposed OCSC algorithm. Diagonal entries indicate classification accuracy for each class. Non-diagonal entries stands for how many sketches was incorrectly classified, and which categories they were classified to. We just show the top mistakes of each category classification for clarity.



(a) Impact of the weight of guidance term



(b) Result comparison

Fig. 2. Effect of guidance term. The left figure shows the trend of recognition rate when varying the weight of the guidance term. The right figure shows the comparison of each category with the parameter setting corresponding to the red and blue triangles in the left figure.

(hence share more common parts) that cause the classifier to confuse one with another. For example, 40.51% swords go to the pen category, 24.05% syringes are recognized as pen, and 18.98% pens go to the sword category.

Effect of guidance term The guidance term, $g(\alpha) = \langle V^T \alpha, \beta \rangle$ in Eq. (3), is an important penalty in the proposed OSCS framework. In particular, its weight λ controls how strong this constraint is to enforce the use of the corresponding basis to encode a sketch, e.g. in the case of small λ , a sketch would be encoded with the optimal parts by searching all the words in source part dictionary which

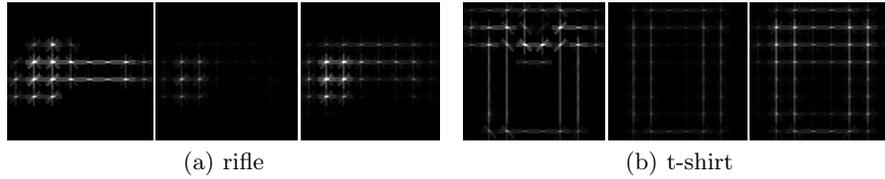


Fig. 3. Original feature maps and reconstructions on (a) ‘rifle’ and (b) ‘t-shirt’. From left to the right: the original feature maps, reconstructions without guidance term, reconstructions by our proposed method.

leads to precise reconstruction. In contrast, in the case of large λ , it would be encoded by a subset of predetermined parts, i.e., the parts of relevant categories given by Eq. (1) and Eq. (2). Fig. 2(a) illustrates how overall the recognition rate changes while increasing the value of λ from 0 to 25, with the other two parameters fixed. It clearly shows that there is a steep climb in recognition rate before the optimal value $\lambda = 6$ is reached (indicated by red triangle in Fig. 2(a)). Afterwards, performance drops steadily while approaching 0% when $\lambda = 25$. Such a reduction of recognition rate reflects the trade-off between guidance term and the regression term, $\|y - A\alpha\|_2^2$, in Eq. (3). That is, too large a weight on guidance term will make the regression problem ill-conditioned that consequently impacts the overall classification accuracy. Note that it becomes the standard sparse coding problem when removing the guidance term in Eq. (3), which is also equivalent to setting $\lambda = 0$ (blue triangle in Fig. 2(a)).

To gain more insight into the usefulness of the guidance term, we offer a set of per-category recognition results on 20 categories in Fig. 2(b). It shows that, under the optimal parameters, the guidance term generally improve performance on all categories except ‘sword’ and ‘megaphone’. In particular, we can observe a jump in recognition rate on categories such as ‘ice-cream-cone’ (from 24.05% to 65.82%), ‘computer-monitor’ (from 43.04% to 74.68%) and ‘pumpkin’ (from 49.37 to 84.81%), with those of ‘t-shirt’ and ‘wheel’ lifted to over 90%.

Fig. 3 shows qualitative examples of feature maps and reconstructions with and without the guidance term, for ‘rifle’ and ‘t-shirt’. It can be seen that better reconstructions can be obtained for both categories, especially for salient parts (e.g. ‘barrel’ of ‘rifle’) and object contours (e.g. outline of ‘t-shirt’).

5 Conclusion

We have studied the problem of one-shot learning of sketch categories, via a novel co-regularized sparse coding framework. We also demonstrated how shared sketch parts can be used within this framework as a semantic level descriptor, as opposed to the rigid grid-level features commonly used in the literature. A key contributing factor towards the superiority of our work is the introduction of a guidance term within our one-shot learning formulation that enforces sparse representations of sketches to agree on a set of predetermined basis. Our experiments on a human labeled dataset of 7,760 sketches show a two-fold improvement over baselines.

Acknowledgment

This work was partially supported by National Natural Science Foundation of China under Grant No.61273217, 61175011 and 61171193, the 111 project under Grant No.B08004.

References

1. Eitz, M., Hays, J., Alexa, M.: How do humans sketch objects? *ACM Trans. Graph.* **31** (2012) 44
2. Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M.: Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *IEEE Trans. Vis. Comput. Graph.* **17** (2011) 1624–1636
3. Hu, R., Collomosse, J.P.: A performance evaluation of gradient field hog descriptor for sketch based image retrieval. Volume 117. (2013) 790–806
4. Hu, R., Barnard, M., Collomosse, J.: Gradient field descriptor for sketch based retrieval and localization. In: *ICIP*. (2010) 1025–1028
5. Li, Y., Song, Y.Z., Gong, S.: Sketch recognition by ensemble matching of structured features. In: *BMVC*. (2013)
6. Cao, X., Zhang, H., Liu, S., Guo, X., Lin, L.: Sym-fish: A symmetry-aware flip invariant sketch histogram shape descriptor. In: *ICCV*. (2013) 313–320
7. Frome, A., Corrado, G.S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., Mikolov, T.: Devise: A deep visual-semantic embedding model. In: *NIPS*. (2013) 2121–2129
8. Lampert, C.H., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: *CVPR*. (2009) 951–958
9. Li, F.F., Fergus, R., Perona, P.: One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.* **28** (2006) 594–611
10. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** (2009) 210–227
11. Tommasi, T., Caputo, B.: The more you know, the less you learn: From knowledge transfer to one-shot learning of object categories. In: *BMVC*. (2009) 1–11
12. Fu, Y., Hospedales, T.M., Xiang, T., Gong, S.: Learning multimodal latent attributes. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** (2014) 303–316
13. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.A.: Describing objects by their attributes. In: *CVPR*. (2009) 1778–1785
14. Gangeh, M.J., Ghodsi, A., Kamel, M.S.: Kernelized supervised dictionary learning. *IEEE Transactions on Signal Processing* **61** (2013) 4753–4767
15. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T.S., Gong, Y.: Locality-constrained linear coding for image classification. In: *CVPR*. (2010) 3360–3367
16. He, R., Zheng, W.S., Hu, B.G.: Maximum correntropy criterion for robust face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** (2011) 1561–1576
17. Zhang, S., Yao, H., Sun, X., Lu, X.: Sparse coding based visual tracking: Review and experimental comparison. *Pattern Recognition* **46** (2013) 1772–1788
18. Portugal, L.F., Judice, J.J., Vicente, L.N.: A comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables. *Mathematics of Computation* **63** (1994) 625–643
19. Grauman, K., Darrell, T.: Fast contour matching using approximate earth mover’s distance. In: *CVPR* (1). (2004) 220–227
20. Lim, J.J., Zitnick, C.L., Dollár, P.: Sketch tokens: A learned mid-level representation for contour and object detection. In: *CVPR*. (2013) 3158–3165