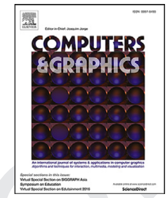
Computers & Graphics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag



Technical Section

# Sketch recognition with few examples☆

Q1 Kemal Tugrul Yesilbek*, T. Metin Sezgin

*Koc University, Istanbul 34450, Turkey*

A R T I C L E I N F O

*Article history:*
Received 23 December 2016
Revised 8 August 2017
Accepted 30 August 2017
Available online xxx

*Keywords:*
Sketch recognition
Learning from few examples
Self-learning

A B S T R A C T

Sketch recognition is the task of converting hand-drawn digital ink into symbolic computer representations. Since the early days of sketch recognition, the bulk of the work in the field focused on building accurate recognition algorithms for specific domains, and well defined data sets. Recognition methods explored so far have been developed and evaluated using standard machine learning pipelines and have consequently been built over many simplifying assumptions. For example, existing frameworks assume the presence of a fixed set of symbol classes, and the availability of plenty of annotated examples. However, in practice, these assumptions do not hold. In reality, the designer of a sketch recognition system starts with no labeled data at all, and faces the burden of data annotation. In this work, we propose to alleviate the burden of annotation by building systems that can learn from very few labeled examples, and large amounts of unlabeled data. Our systems perform self-learning by automatically extending a very small set of labeled examples with new examples extracted from unlabeled sketches. The end result is a sufficiently large set of labeled training data, which can subsequently be used to train classifiers. We present four self-learning methods with varying levels of implementation difficulty and runtime complexities. One of these methods leverages contextual co-occurrence patterns to build verifiably more diverse set of training instances. Rigorous experiments with large sets of data demonstrate that this novel approach based on exploiting contextual information leads to significant leaps in recognition performance. As a side contribution, we also demonstrate the utility of bagging for sketch recognition in imbalanced data sets with few positive examples and many outliers.





## 1. Introduction

Hand-drawn sketches are ubiquitous in design, arts, education and entertainment. More recently sketching has also been receiving attention as a natural human-computer interaction modality as seen from the continually increasing body of work on automated sketch recognition.

Sketch recognition is defined as the task of segmenting a full sketch into individual groups of ink representing domain symbols, and assigning labels denoting classes. State of the art approaches to sketch recognition are predominantly based on machine learning technologies. However, the development and evaluation of these algorithms have traditionally been carried out with strong assumptions that do not hold in practice.

For example, it is generally assumed that sufficiently large set of annotated symbols are readily available for training classifiers. In practice, however, such data is generally unavailable. Moving into a new domain requires the designer of the sketch recognition system to create an annotated data set. This is done either by collecting isolated instances of symbols from users [1–5], or by annotating full sketches [6,7] (i.e., sketches consisting of multiple symbols). Both cases require substantial annotation effort. In this paper, we propose methods for training sketch recognizers using only a few (1–3) labeled examples. We do so by leveraging large sets of unlabeled examples. This ability of the proposed framework allows users of the system to define their own classes for an unlabeled data set on-fly, which offers great flexibility.

Although our main contribution addresses learning with few examples, our setup also challenges other assumptions in the field. It is generally assumed that recognizers will only be tested on symbols strictly within the domain of interest. This assumption manifests itself through the use of crisp multi-class data sets, or in the form of drawing instructions for users where they are first briefed about the set of available domain objects, and told not to use any symbols outside this restricted set. Hence, evaluation results in the literature are all reported in a multi-class classification setting where the knowledge of all classes are available. However, real drawings usually contain a large number of objects, marks, and writing that are irrelevant for the domain, and act as outliers. The learning framework we describe explicitly abstains from crisp

☆ This article was recommended for publication by Prof. J. Jorge

* Corresponding author.
*E-mail address:* kyesilbek@ku.edu.tr (K.T. Yesilbek).

https://doi.org/10.1016/j.cag.2017.08.016

data assumptions, and is evaluated with realistic sketch data containing many outliers.

Our approach is technically a semi-supervised method performing *self-learning*. Self-learning refers to using some amount of labeled data to label unlabeled instances, and training a classifier with the extended set of labeled instances. Generally self-learners start with an initial seed set of 10 or more labeled examples per class, and extend the training data. However, we target very few examples (1–3 labeled examples). This results in two main challenges. First, with only 1–3 items in the initial list of labeled examples, it becomes essential that any additional items brought into the list do indeed belong to the correct class. Even a few incorrectly labeled examples can cause catastrophic drops in recognizer performance. Second, it is extremely important to ensure that the additional labeled items are not too similar to the existing examples. New labeled examples help only if they are diverse and carry variations. We show that a context-based selection criterion promotes diversity. The key insight that we bring is to give precedence to candidate examples that not only have the appearance of the class of interest, but also appear in contexts that are typically observed for the object of interest. This scheme favors diversity.

Learning from few examples also poses a data imbalance challenge. The number of positive examples are multiple orders of magnitude smaller than the number of negative and unlabeled examples. We address this issue through bagging (bootstrap aggregation).

Finally, we successfully adopt a Viola–Jones-like filtering scheme to speed up the self-learning process for large data sets. The filtering acts as a conservative rejection mechanism that excludes irrelevant unlabeled instances from the self-learning pipeline.

The focus on learning from very few examples distinguishes our work from others. The context-based self learning method is our main contribution. We demonstrate the utility of this approach through its ability to accurately select diverse examples for training sketch recognizers. Successful incorporation of bagging and conservative rejection serve as two additional contributions.

In the rest of the paper, we first put our work into perspective by discussing the related work from the sketch recognition domain. Since the use of realistic data is one of the core contributions of our work, we describe the in-the-wild sketch data set that we use in Section 3. We measure the feasibility of self-learning through many repeated experiments designed to mimic what would have happened if the process had started with various initial conditions. The Experimental Setup section describes the end-to-end pipeline for self learning, including the details of data preparation, and metrics for performance measurement. Section 5 describes the details of our context-based self-learning algorithm, along with three others. We report our findings in the Results section, conclude with a discussion of the main findings, a summary of our contributions and directions for future work.

## 2. Related work

The historical progression of interest in sketch recognition started with investigation of knowledge-based and model-based recognition systems with no elements of machine learning [8–11]. The focus later shifted to approaches based on machine learning. These methods proved to be superior, and the field enjoyed steady progress in feature representations and recognition architectures. It is only recently that the interest has shifted to alleviating the difficulties associated with approaches based on machine learning. Below we discuss how our work fits in this vast body of work on sketch recognition.

The early work on sketch recognition focused on building rule-based recognition algorithms. These approaches combined structural descriptions of symbols with efficient matching algorithms and rule-based interpretation architectures for recognition [8–11]. Rather than learning from examples, they use knowledge based object models. For example, Mahoney and Fromherz [8] propose structural descriptions that describe domain objects in terms of connections and constraints defined over line segments, and use sub graph isomorphism for recognition [8]. Sezgin and Davis propose automatic generation of recognizer code from structural descriptions of domain objects [9]. Hammond takes the idea of structural descriptions further by defining a formal symbol representation language [11] and a perceptually inspired method for generating object descriptions from single hand-drawn examples [10]. The work of Veselova and Davis is in the same spirit as ours in the attempt to learn from few examples, however we operate within a machine-learning-based framework, and try to exploit unlabeled data.

With the development of powerful feature representations for sketches, recognition frameworks based on machine learning gained dominance [2,3,12–14]. These methods were developed and evaluated within the standard train/validate/test machine learning pipeline, and our work aims to address the limitations induced by the assumptions of these systems. These and many others [1,4,5] assume fully labeled training data sets consisting of isolated hand-drawn symbols instances. They assume a predetermined set of object categories, and focus on performance indicators measured over isolated symbols or scenes consisting of domain objects only. In contrast, we focus on learning from few examples, while symbols are not isolated (i.e., there exists multiple symbols in a sketch), and exploiting unlabeled data. Most of the work supporting sketch scenes with multiple objects assume that each object is drawn with a single stroke [15,16]. While this assumption both reduces the complexity and increases the success rate of the techniques, it forces users to change their sketching style which affects usability negatively. To address this issue, our system follows a fragment-and-combine approach similar to Alvarado and Davis [17].

The most relevant pieces of work to ours are those that try to exploit unlabeled examples [15,18,19]. All these systems assume a small seed set of labeled examples, and try to extend the number of labeled instances by automatically labeling unlabeled examples with the user in the loop. Technically these methods are active learning approaches, since they require user supervision. They starts with a low number of labeled instances, and allow the labeling of the mis-recognized instances [15], or ask for specific instances to be labeled [19] by the user. Unlike these, we do not rely on the user for labeling. We start with very few labeled instances and continue in a fully automated fashion. This makes the problem more challenging, since no user intervention is possible in case of errors in automatic instance labeling. Furthermore, these approaches mostly assume that the unlabeled data is already segmented, an assumption we explicitly avoid.

Within the machine learning and computer vision literature, there are plenty of approaches for zero shot learning, one shot learning, and transfer learning [20]. These approaches rely on attributes that serve as reusable models of object properties. Models for new objects are subsequently defined in terms of the previously learned attributes [20,21]. Examples of work along these lines in the sketch recognition community include the work of Alvarado and Davis [22,23]. They model subparts of domain objects using distributions over features and reuse this information to build generative graphical models. These approaches have been disadvantaged by high computational requirements, and lower recognition rates compared to the learning-based approaches that came later (e.g., [3,12,24]). Furthermore, the inherently sparse, and ambiguous nature of sketches renders the tuning process of these generative models an art.