Cats and Dogs

Omkar M Parkhi[1,2] Andrea Vedaldi[1] Andrew Zisserman[1] C. V. Jawahar[2]

[1]Department of Engineering Science,
University of Oxford,
United Kingdom
{omkar,vedaldi,az}@robots.ox.ac.uk

[2]Center for Visual Information Technology,
International Institute of Information Technology,
Hyderabad, India
jawahar@iiit.ac.in

Abstract

We investigate the fine grained object categorization problem of determining the breed of animal from an image. To this end we introduce a new annotated dataset of pets, the Oxford-IIIT-Pet dataset*, covering 37 different breeds of cats and dogs. The visual problem is very challenging as these animals, particularly cats, are very deformable and there can be quite subtle differences between the breeds.*

We make a number of contributions: first, we introduce a model to classify a pet breed automatically from an image. The model combines shape, captured by a deformable part model detecting the pet face, and appearance, captured by a bag-of-words model that describes the pet fur. Fitting the model involves automatically segmenting the animal in the image. Second, we compare two classification approaches: a hierarchical one, in which a pet is first assigned to the cat or dog family and then to a breed, and a flat one, in which the breed is obtained directly. We also investigate a number of animal and image orientated spatial layouts.

These models are very good: they beat all previously published results on the challenging ASIRRA test (cat vs dog discrimination). When applied to the task of discriminating the 37 different breeds of pets, the models obtain an average accuracy of about 59%, a very encouraging result considering the difficulty of the problem.

1. Introduction

Research on object category recognition has largely focused on the discrimination of well distinguished object categories (*e.g*, airplane vs cat). Most popular international benchmarks (*e.g*, Caltech-101 [22], Caltech-256 [26], PASCAL VOC [20]) contain a few dozen object classes that, for the most part, are visually dissimilar. Even in the much larger ImageNet database [18], categories are defined based on a high-level ontology and, as such, any visual similarity between them is more accidental than systematic. This work concentrates instead on the problem of *discriminating different breeds of cats and dogs*, a challenging example of fine grained object categorization in line with that of previous work on flower [15, 32, 33, 39] and animal and bird species [14, 27, 28, 43] categorization. The difficulty is in the fact that breeds may differ only by a few subtle phenotypic details that, due to the highly deformable nature of the bodies of such animals, can be difficult to measure automatically. Indeed, authors have often focused on cats and dogs as examples of highly deformable objects for which recognition and detection is particularly challenging [24, 29, 34, 45].

Beyond the technical interest of fine grained categorization, extracting information from images of pets has a practical side too. People devote a lot of attention to their domestic animals, as suggested by the large number of social networks dedicated to the sharing of images of cats and dogs: Pet Finder [11], Catster [4], Dogster [5], My Cat Space [9], My Dog Space [10], The International Cat Association [8] and several others [1, 2, 3, 12]. In fact, the bulk of the data used in this paper has been extracted from annotated images that users of these social sites post daily (Sect. 2). It is not unusual for owners to believe (and post) the incorrect breed for their pet, so having a method of automated classification could provide a gentle way of alerting them to such errors.

The first contribution of this paper is the introduction of a large annotated collection of images of 37 different breeds of cats and dogs (Sect. 2). It includes 12 cat breeds and 25 dog breeds. This data constitutes the benchmark for pet breed classification, and, due to its focus on fine grained categorization, is complementary to the standard object recognition benchmarks. The data, which is publicly available, comes with rich annotations: in addition to a breed label, each pet has a pixel level segmentation and a rectangle localising its head. A simple evaluation protocol, inspired by the PASCAL VOC challenge, is also proposed to enable the comparison of future methods on a common grounds (Sect. 2). This dataset is also complementary to the subset of ImageNet used in [27] for dogs, as it contains additional annotations, though for fewer breeds.