Cats and Dogs

Muhammad Hashim Javed 21100271

1 Paper trivia

This paper is combined effort of Department of Engineering Science, University of Oxford, UK, and Center of Visual IT, IIIT, Hyderabad, India, written by Omkar M Parkhi [1] , Andrea Vedaldi[1], Andrew Zisserman[1], and C.V. Jawahar[2]. Paper is 8 pages long and contains both theoretical and experimental tracks.

2 Problem description

Paper addresses the problem of classification of thirty seven different breeds of cats and dogs[3]. This problem is important because 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, Catster, Dogster. Given a image of cat or dog, proposed solution tells the breed of the animal. Furthurmore many bot detection methods of microsoft use cats and dogs pictures. They are based on assumption that the out of a batch of twelve images of pets, any machine would predict incorrectly the family of at least one of them. If any bot classifier has accuracy 95%, it will roughly fool the test half of the time.
Previoulsy many categorization research was focused on categorization of fairly distinct objetcs[4]. For similar objects it is difficult to categorize them, therefore categorization of cats and dogs is particularly challenging.

3 Dataset

Authors used Oxford-IIIT Pet dataset for training, validation and testing. This dataset was gathered particularly for this paper and available publically. For furthur testing authors used the ASIRRA dataset of Microsoft Research (MSR). Oxford-IIIT dataset id collection of 7,349 images, 200 images for each breed. ASIRRA dataset contain 24,990 images publically avaliable out of several million.

4 Main techniques of the paper

This section is divided into two parts.

[1]University of Oxford, U.K
[2]IIIT, Hyderabad, India
[3]12 cat breeds and 25 dog breeds
[4]e.g, airplanes and cats

4.1 A model for breed discrimination

The breed of a pet affects its shape, fur type, and color. This model focuseson capturing the pet shape and the appearance of its fur. The model also involves automatically segmenting the pet from the image background.
In shape model, author uses the deformable part model[5] to get the shape of body, particularly head[6]. In appearence model, author extracts SIFT[7] descriptors to compute the visual words, words are then represented as bag-of-words model[6]. In automatic segmentation part, the foreground abd background regions needed for computing the appearence descriptors are obtained automatically using the grab-cut segmentation technique[8,6].

4.2 Experiments

Before experiment author establishes the difficulty of dataset, by comparing it with other available Fine Grained Visual Categorization datasets using publically avaliable VLFeat BoW classification code, results are given below.

Dataset	Mean Classification Accuracy
Oxford-IIIT Pet Dataset	38.45%
UCSD-Caltech Birds	6.91%
Oxford-Flowers102	53.71%

Experiments are furthur divided into Pet family descrimination and Pet breed descrimination, both. In each type of classification shape and appearence are used separately and together to train the dataset.

5 Evaluation and results

In above experiments, the performance is measured as average per-class classification accuracy. It uses the same parameter when comparing with other state of the art methods. On ASIRRA dataset, best accuracy reported is 82.7%, and this paper gives accuracy of 92.9%.

Method	Mean Class. Accuracy
Golle *et al.* [25]	82.7%
This paper (Shape only)	92.9%

[5]Deformable Part is a discriminatively trained, multi-scale model for image training that aim at making possible the effective use of more latent information such as hierarchical (grammar) models and models involving latent three dimensional pose

[6]This can be done through avaliable python libraries

[7]The scale-invariant feature transform (SIFT) is a feature detection algorithm in computer vision to detect and describe local features in images

[8]GrabCut is an image segmentation method based on graph cuts

Furthermore paper compares the results with different layout type and alssification problem and all the result are given below.

This result shows as we increase information in Appearence the accuracy increases, similarly including

.	Shape	Appearance		Classification Accuracy (%)				
		layout type	using ground truth	family	breed (S. 4.2)		both (S. 4.3)	
				(S. 4.1)	cat	dog	hierarchical	flat
1	✓	–	–	94.21	NA	NA	NA	NA
2	–	Image	–	82.56	52.01	40.59	NA	39.64
3	–	Image+Head	–	85.06	60.37	52.10	NA	51.23
4	–	Image+Head+Body	–	87.78	64.27	54.31	NA	54.05
5	–	Image+Head+Body	✓	88.68	66.12	57.29	NA	56.60
6	✓	Image	–	94.88	50.27	42.94	42.29	43.30
7	✓	Image+Head	–	95.07	59.11	54.56	52.78	54.03
8	✓	Image+Head+Body	–	94.89	63.48	55.68	55.26	56.68
9	✓	Image+Head+Body	✓	95.37	66.07	59.18	57.77	59.21

shape and ground truth information in training also increases the accuracy. It is also noteable that in classificaion of breed and family both, flat classification[9] gives better results then hierarchical classification in which the family is decided first,and then the breed is decided.

6 Conclusion

As we know, designed dataset was meant to be challenging for machine and it is same dataset used to detect bots by microsoft, the accuracy[10] achieved by this model, makes it a state of art model.

[9] in which a 37-class SVM is learned directly

[10] This model can fool microsoft bot detection 44% of time