Human vs. Computer in Scene and Object Recognition

Cheng Guan

Jun 28, 2018

Abstract

Over the past several decades, computer and primate vision research have led to a number of models (some specific to a problem, other more general) and valuable experimental data. Here, in order to help research work focus on the most difficult problems, and set up a bridge between computer and human vision,, the authors define a battery of 5 tests that measure the gap between human and machine performances in several dimensions (generalization across scene categories, generalization from images to edge maps and line drawings, invariance to rotation and scaling, local/global information with jumbled images, and object recognition performance).The authors measure model accuracy and the correlation between model and human error patterns.

1. Introduction

The rapid development of the computer vision community Recently in several fields. Restricted cases (*e.g.*, wherre variability is low), computers even perform well. Human tasks, such as frontal-view face recognition,fingerprint recognition, change detection, etc. The trend is to get more and more unbiased datasets.(*e.g.*, ImageNet, SUN, Flicker, LabelMe), constructing features / algorithms.Starting from these data, the design is appropriate Scores are used to measure progress. The success of the past has been created .The authors hope one day they can solve the difficulties of vision without having humans in the picture. Several previous studies, under the names of humans in the loop, human debugging, finding weak links (and often using Amazon Mechanical Turk) [4, 5, 1, 3, 2], have used humans to estimate the relative strengths and weaknesses of different algorithm components (*e.g.*, [1]).

Here, the authors take a more systematic approach to compare 14 computer vision models on 7 datasets using 5 different tests. They focus on two difficult central visual problems, namely object and scene recognition. The current computer vision model have difficulty on these issues, human beings have solved these problems almost effortless-
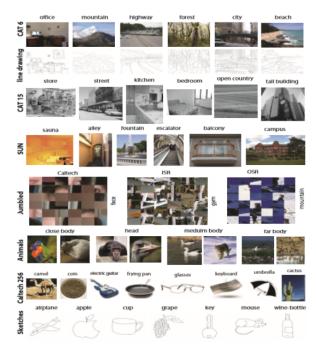

Figure 1. Sample images from scene and object datasets used in this study

ly.

The authors organize 5 tests: The first two regard scene categorization using color photographs and line drawings. The third test addresses invariance properties of models on animal vs. non-animal recognition.

2. Elements of the Comparison

Statistics of 7 datasets (5 for scenes and 2 for objects) used here are described (Fig. 1), followed by their model evaluation part. See supplementary for details on human data gathering protocols on datasets.They are 6-CAT, 15-CAT,SUN, Animals, Jumbled images, Caltech-256 and Sketch images. I will describe the 7 datasets in the next writing.

References

[1] F. Fleuret, T. Li, C. Dubout, E. K. Wampler, S. Yantis, and D. Geman. Comparing machines and humans on a visual categorization test. *PNAS*, 2011. 1

[2] D. Hoiem, Y. Chodpathumwan, and Q. Dai. Diagnosing error in object detectors. In *ECCV*, 2012. 1

[3] R. Mottaghi, S. Fidler, J. Yao, R. Urtasun, and D. Parikh. Analyzing semantic segmentation using hybrid human-machine CRFs. In *CVPR*, 2013. 1

[4] D. Parikh and C. L. Zitnick. Finding the weakest link in person detectors. In *CVPR*, 2011. 1

[5] D. Parikh, C. L. Zitnick, and T. Chen. Tiny images: The roles of appearance and contextual information for machine and human object recognition. *IEEE TPAMI*, 2012. 1