PASCAL-S Dataset Instructions

Yin Li Xiaodi Hou

April 8, 2015

1 Introduction

The document describes the folder and file structure of our dataset package used in our paper [8]. Our PASCAL-S dataset is built on the validation set of the PASCAL VOC 2010 segmentation challenge. It contains 850 images of natural images with both eye tracking data and salient object annotations. Our salient object annotations extend the 20 classes in VOC by using the full segmentation masks from [9]. As we have discussed in the paper, by using images from existing dataset, we will be able to minimize the risk of dataset design bias.

For the convenience of other users, we have also included major datasets in a format that is similar to PASCAL-S. The other **4** fixations datasets are: Bruce [4], Cerf [5], imgSal [7], and Judd [6]. The other **3** salient object segmentation datasets are: Bruce [4, 3], FT [1], and imgSal [7]. Please cite the corresponding papers if you use these datasets.

2 Data Collection

In the fixation experiment, 8 subjects were instructed to perform a "free-viewing" task to explore the images. Each image was presented for 2 seconds, and eye-tracking re-calibration was performed on every 25 images. The eye gaze data was acquired using Eyelink 1000 eye-tracker, at a sampling rate of $125Hz$.

In the salient object segmentation experiment, we use the full segmentation masks provided in [9] to crop out all objects in the image (not limited to 20 classes used in PASCAL-VOC). We then conduct an experiment of 12 subjects to label the salient objects. Given an image, a subject was asked to highlight the salient objects by clicking on them. There was no time limitation or constraints on the number of objects one can choose. Similar to our fixation experiment, the instruction of labeling salient objects was intentionally kept vague.

The segmentation of images adhere to the following rules:

1. We do not intentionally label parts of the image (e.g. faces of a person).
2. Disconnected regions of the same object are labeled separately.
3. We use solid regions to approximate hollow objects, such as bike wheels.

3 Data Format

3.1 Image data

Images of each dataset are renamed and stored them in a separate folder in:
`datasets/imgs/`.
The images is renamed into a numerical order as 1.jpg, 2.jpg ...

3.2 Fixation data

The fixation data of all 5 fixation datasets are stored in `datasets/fixations/`. For instance, PASCAL-S fixation data is stored in `pascalFix.mat` as a 850×3 cell matrix, `fixCell`. Each cell element in `fixCell` is a $N \times 3$ matrix, storing the information of all N fixations of an image. The rows of fixation data are stored in the following order:

Fixation Height	Fixation Width	Subject Number

`pascalSize.mat` keeps the resolution for each image in our PASCAL-S dataset. Both the fixation and resolution data are indexed based on the renamed files. For example, the third cell of fixation data and the third row of the resolution data present the fixation and the size of image file 3.jpg. Both `pascalFix.mat` and `pascalSize.mat` are required to reconstruction the fixation map. Other datasets follows a similar naming convention.

3.3 Salient object masks

We have included salient object masks for our PASCAL-S dataset.
`datasets/masks/pascal/`
The masks is normalized from percentage of agreement (0-1) to 0-255 and stored in png format. We threshold the masks at 0.5 to get the binary mask as discussed in [8]. In addition, we have included salient object masks for Bruce [4, 3], ImgSal [7], and FT [1]. For ImgSal and FT, we provide binary masks in png format. Note we have fixed a few annotation errors in FT, e.g. tiny segments being marked as salient objects. For Bruce dataset, the png file stores the per pixel raw count of annotations, e.g. the number of subjects that mark the pixel as salient object.

3.4 Full object segmentation data

The full segmentation of all 850 images of PASCAL-S is stored in:
`datasets/segments/pascal/isoCell.mat`
as an 850×1 cell array `isoCell`. Each cell element in `isoCell` consists of an indexed bitmap $\mathbf{x} \in \mathbb{Z}^{H \times W}$. Specifically $x_{i,j} = 0$ denotes "void" regions. Usually a "void" region appears at the gap between two objects. There is no particular ordering for index $x > 0$.

Our salient object annotation tool (Matlab) is also included in the folder. The raw data of clickings can be found in `datasets/segments/pascal/data`. For more segmentations on full PASCAL dataset, check [9].

4 Results

The results of 7 fixation methods are available on all 6 datasets included in the package. In addition, the results of 3 state-of-the-art salient object segmentation algorithms are included for FT, imgSal and PASCAL-S. Our results (CPMC+GBVS) are also included for these 3 datasets. The results are located at
`algmaps/dataset_name/alg_name/`

5 Benchmark Code

In addition, we provide separate benchmark code (with GPU support) for fixation prediction (sAUC score) and salient object segmentation (F-measure). The code is locatd at `benchmark/`. You can try `benchPR` and `benchAUC` as the demo. Please refer to the comments as the document, as they should be easy to follow.

6 Code

Our implementation of the paper can be found at `code/`. We provide detailed comments in the code. Again please refer to the comments as the document. Here is a quick instruction for running the code.

- Setup the environment using `setup_env.m`
- We include a demo for the full training and testing pipeline in `salobj_train_demo.m`
- To use a pretrained model, you can try our script in `salobj_test_demo.m`

- We provide a pre-trained model using full PASCAL-S dataset in the folder `code/models`

Our implementation is slightly different than what is reported in the paper. Here is a brief change log.

- We replaced CPMC with more recent MCG [2] for better efficiency
- We fixed a bug in previous benchmark by setting $\beta^2 = 0.3$. This bug increased all the scores reported in our paper. It did not change the ranking of the algorithms or their relative performance.
- We used a different set of features than our CVPR paper. Most notably, we add boundary strength features similar to MCG [2] and delete expensive geometry features such as convex area and solidity.
- We replace random forest regression with classification, which leads to much faster training with marginal performance lost.

This version of implementation has similar results as in our CVPR paper.

In addition, our code package includes the following third-party libraries

- MCG with pre-trained model available at the link
- Piotr Dollar's toolbox, which is required by MCG and also used for random forest. The toolbox is available at the link
- GBVS code for fixation prediction, which is available at the link

We thank the authors of these libraries for providing their code. For latest version of our code, please check our public repository. Bug reports are expected to be submitted via this repository.

7 Citation

If you make use of the PASCAL-S dataset, please cite our paper as follows.

```
@inproceedings{secrets2014li,
  title={The Secrets of Salient Object Segmentation},
  author={Li, Yin and Hou, Xiaodi and
    Koch, Christof and Rehg, James and Yuille, Alan},
  booktitle={CVPR, 2014, IEEE Conference on},
  year={2014}
}
```

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597–1604. IEEE, 2009.

[2] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Computer Vision and Pattern Recognition*. IEEE, 2014.

[3] A. Borji, D. N. Sihite, and L. Itti. What stands out in a scene? a study of human explicit saliency judgment. *Vision research*, 91:62–77, 2013.

[4] N. Bruce and J. Tsotsos. Saliency based on information maximization. In *Advances in neural information processing systems*, pages 155–162, 2005.

[5] M. Cerf, J. Harel, W. Einhäuser, and C. Koch. Predicting human gaze using low-level saliency combined with face detection. In *Advances in neural information processing systems*, pages 241–248, 2008.

[6] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.

[7] J. Li, M. D. Levine, X. An, X. Xu, and H. He. Visual saliency based on scale-space analysis in the frequency domain. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(4):996–1010, 2013.

[8] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 280–287. IEEE, 2014.

[9] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille. The role of context for object detection and semantic segmentation in the wild. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 891–898. IEEE, 2014.