Human Data and Prediction Pipeline Developed in the Paper:

Asking for Help with the Right Question by Predicting Human Visual Performance

Hong (Herbert) Cai and Yasamin Mostofi

(Presented at Robotics: Science and Systems 2016)

If you have any questions or comments regarding the data and/or the trained models, please contact Herbert Cai at hcai@ece.ucsb.edu.

Credits and Usage

The code/data are owned by UCSB and can be used for academic purposes only.

If you have used this data and/or code for your work, please refer the readers to the data and code website at its DOI address: https://doi.org/10.21229/M9V66F, and also cite the following paper:

> H. Cai and Y. Mostofi, "Asking for Help with the Right Question by Predicting Human Visual Performance," Robotics: Science and Systems (RSS), June 2016.

> @inproceedings{CaiMostofi_RSS16_Ask,
> title={Asking for Help with the Right Question by Predicting Human Visual Performance},
> author={H. Cai and Y. Mostofi},
> booktitle={Proceedings of Robotics: Science and Systems},
> year={2016}}

1. Data Set:

There is a total of 3000 images collected from NOAA, SUN and PASCAL VOC. Each image contains at least one human. The images are put into 3 categories: 1) easy images where the human presence is obvious, 2) images where the human is in a cluttered environment, and 3) images where the size of the human is small (due to distance). We have also manually darkened the images to simulate night time scenarios.

- The images can be found in the folder "images".

2. Training and Validation Data

We used 2400 (80%) of the images for training and 600 (20%) for validation.

- The files specifying the training set and validation set can be found in the folder "caffe_models".

- "train.txt" and "val.txt" specify the training and validation cases respectively.

- "train_balance.txt" is the training file where we have oversampled the images with human performance below 0.9 to avoid biasing the predictor (see the paper for more details). This is the file used for training the Convolutional Neural Network (CNN).

3. Trained Caffe Model for Human Performance Prediction

This is the human performance predictor model. It takes as input a 256x256 image and outputs the probability that a person is able to find the human in the image.

- "human_prediction_iter_50000.caffemodel" is the trained model and can be found in the folder "caffe_models". The prototxt files for training and deploying the model can also be found in the folder.

- *Note: Naturally, the trained model works the best if the input images share similar characteristics with the training images. For instance, the prediction may not be as accurate if the type of input images is too different from the type of training images, in terms of performance degradation factors. In such cases, the network can be fine-tuned or re-trained with more images that can capture the new previously-unseen characteristics.*

4. Pre-trained Model Used for Initialization

This model, which is based on the AlexNet architecture [1], takes as input a 256x256 image and performs a binary classification indicating whether a human is present in the image. It is used as the initialization for the training of our human predictor network.

- "human_detection_iter_150000.caffemodel" is the pre-trained model and can be found in the folder "caffe_models".

References:

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2012.