Reducing the Semantic Gap in Saliency

Prediction by Adapting Deep Neural Networks

Motivation

- Conventional saliency models rely on image statistics to predict human fixations
- They perform well unless images have a strong semantic content (i.e. meaning)
- New approach tries to fine tune Object Recognition DNNs
- Object Recognition DNNs are inherently very good at extracting meaning from images
- Another key component is to integrate information at different image scales
- > The new approach surpasses all state of the art methods on all datasets and metrics

System Design **Trained Object detection DNN** Upsampling Input images Saliency Map 800 Concatenation Coarse 009 Linear Integration Extra saliency layers 2 Streams Saliency Objectives (KLD, NSS, CC, SIM) **Human Fixation Maps** DNNs with shared weights Resizing E Fine Size = Coarse Size x 2 (double resolution) Evaluation sAUC scores for all datasets 0.68 0.88 0.78 0.86 0.66 0.76 0.82 0.64 0.74 0.8 0.72 0.62 0.76 0.7 0.74 0.6 0.68 0.720.66 0.7 (a) OSIE (b) MIT1003 (c) NUSEF 0.84 0.74 0.7 0.82 0.8 0.72 0.68 0.78 0.7 0.76 0.66 0.74 0.68 0.640.72 0.7 0.66 TIGHT SUNDCUSTOR CAS GOD Judd Judd TIENS PROCUENT (d) PASCAL-S (f) Toronto (e) FIFA ntuition RARE LG BMS AWS Human AIMOurs Fixed Fine-tuned 400 300 200 QunN 100 -0.06 -0.04-0.02 0.02 0.04

Weight Value

Poster was adapted by Caus Danu based on the following paper

Huang, Xun, et al. "Salicon: Reducing the semantic gap in saliency prediction by adapting deep

neural networks." Proceedings of the IEEE International Conference on Computer Vision. 2015.