Cognitive Computer Vision

Computer Exercise

Tuesdays 14.15-15.45, Room D-118/119

Dr. Mikko Lauri
Email: lauri@informatik.uni-hamburg.de

SS 2018

Contents

Eye fixation prediction

Eye fixation prediction

Given an input RGB image I, output a eye fixation prediction map P that highlights regions that are likely to attract human attention.

Figure: Left: input image I, right: output image P

Eye fixation prediction

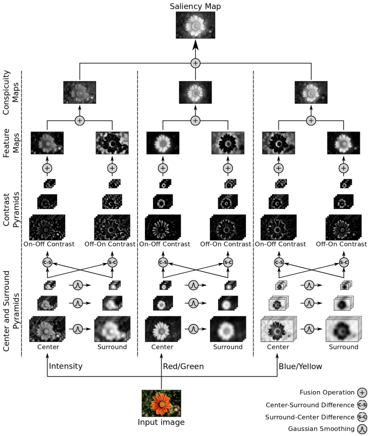


Figure: VOCUS2 [Frintrop et al., 2015]

Eye fixation prediction

- Deep learning based systems now state-of-art
- Use multi-level features to predict eye fixations

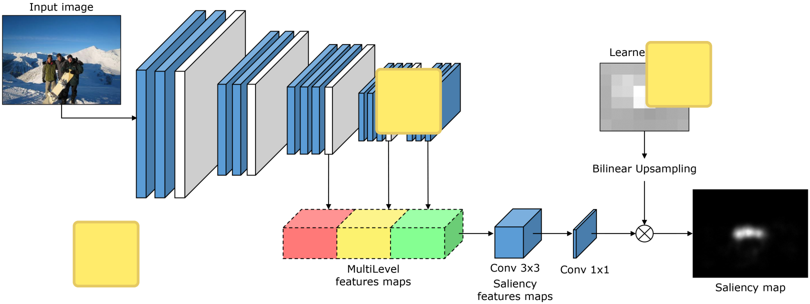
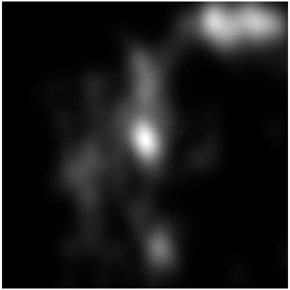


Figure: MLNET [Cornia et al., 2016]

Center bias

- Humans have a tendency to gaze at the center of an image.
- Many models take this into account by explicitly adding a center bias.



Evaluating eye fixation predictions

- Prediction image M is compared to ground truth image G
- Both images interpreted as probability mass functions: value M_i of pixel i denotes probability that human will fixate at i.
    - First step in evaluation: normalize both images so that their pixel values sum to 1.
- Many evaluation metrics available [Bylinskii et al., 2016], we focus on Kullback-Leibler divergence (KLD):

$$KLD(P, G) = \sum_i G_i \log \left(\frac{G_i}{P_i} \right).$$

- KLD is a non-symmetric difference between probability mass functions (pmfs): small KLD $\equiv$ pmfs are similar

Loss function?

- Term $\log\left(\frac{G_i}{P_i}\right)$ in KLD is minimized if $P_i = G_i$: pixelwise similarity of prediction to ground truth is desired, use e.g., Euclidean norm of difference $||P_i - G_i||$
- Since P will be interpreted as pmf, its absolute scale does not matter: normalize P by dividing it by $\max_i P_i$
- Most ground truth pixels are near zero, but still important to predict non-zeros correctly: Cornia et al. [2016] suggests weighting pixel i by $\alpha - G_i$ where $\alpha > 1$ to give higher weight to non-zeros in ground truth
- Suggested loss function for N images:

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\alpha - G_i}\cdot\left(\frac{\phi(P_i)}{\max_i \phi(P_i)} - G_i\right)\right\|^2$$

Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? *arXiv preprint arXiv:1604.03605*, 2016.

Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. A Deep Multi-Level Network for Saliency Prediction. In *International Conference on Pattern Recognition (ICPR)*, 2016.

Simone Frintrop, Thomas Werner, and German Martin Garcia. Traditional saliency reloaded: A good old model in new shape. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.