

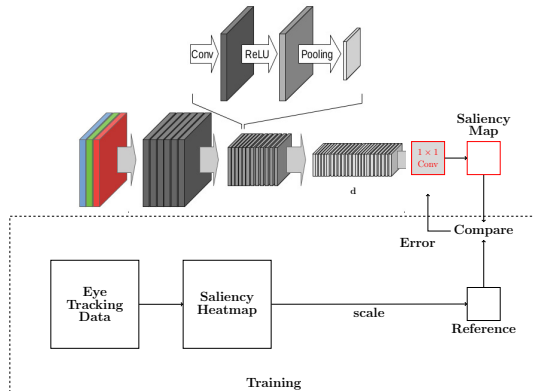
Biological Vision and Applications

Module 05-08: NN based attention models

Hiranmay Ghosh



Basic Architecture



- Bottom-up or top-down attention?

Attention and object detection

- Use CNN pre-trained for object detection
 - ▶ Not enough training data for saliency
 - ▶ **Objects lead to saliency**
- In neural network based architectures
 - ▶ Attention and object detection complement each other
 - ▶ Find salient locations (where objects are likely to be there)
 - ▶ Detect objects at those locations

Soft attention vs. hard attention

- Soft attention
 - ▶ Graded saliency values for different image locations
 - ▶ Fixation traverses from location with highest saliency to lowest
- Hard attention
 - ▶ Binary saliency values
 - ▶ Fixation at the region with saliency
- NN based attention models generally use hard attention
 - ▶ One or very few “salient” objects in a scene
 - ▶ A binary classifier (SVM / Softmax) is added at the end

Objects are salient

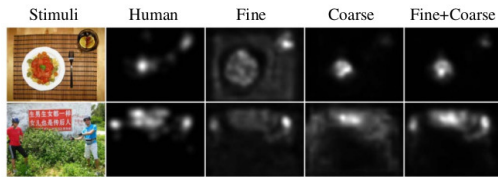
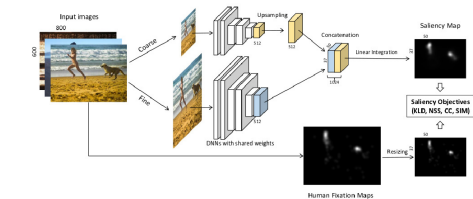
Saliency of nearby regions should be close to each other

- The approach discussed computes location based saliency
- The image can be divided into 'superpixels'
 - ▶ Areas of near uniform color/texture
- Adjust saliency values to encourage locations in nearby superpixels to have homogeneous saliency
 - ▶ Something like graph-cut algorithm
 - ▶ Minimize $\sum_i (s_i^{new} - s_i)^2 + \sum_{i,j} w_{ij} (s_i^{new} - s_j^{new})^2$
 - ▶ Weights w_{ij} depend on physical distance
 - ▶ Optimal weights are learned



Multi-scale analysis

SALICON: Saliency in Context



- Coarse level captures context; fine level captures local contrasts
- Usually 2 or 3 levels of resolution is found to be sufficient

Quiz 05-08

End of Module 05-08