Non-Local Deep Features for Salient Object Detection

Zhiming Luo\textsuperscript{1,2}, Akshaya Mishra\textsuperscript{1}, Andrew Achkar\textsuperscript{3}, Justin Eichel\textsuperscript{3}, Shaozi Li\textsuperscript{1}, Pierre-Marc Jodoin\textsuperscript{2}

\textsuperscript{1}Xiamen University, China
\textsuperscript{2}University of Sherbrooke, Canada
\textsuperscript{3}Miovision Technologies Inc., Canada

Goal
Highlight the most relevant objects in an image (saliency).

Proposed method
- Novel 4 × 5 multiresolution CNN grid structure
- Contrast features
- Loss function inspired by the Mumford-Shah functional
- No CRF, no superpixels

Outcome
- Top performing method on 6 datasets
- Real-time, high performance saliency detection.

Mumford-Shah Function [5]

$$ F = \sum \alpha f(v) d\nu + \sum \beta \int \chi_{\text{boundary}} d\nu $$

Bayesian statistical approximation [6]

$$ F \approx \sum \alpha \log \left( p(v | f(v)) \right) d\nu + \sum \beta \int d\nu $$

Final loss function

$$ F \approx \sum \alpha \int h(y(v) | f(v)) d\nu + \sum \beta (1 - \text{IoU}(C_y, C)) $$

Cross-Entropy Loss

$$ \hat{y}(v) = p(y(v) = c) = \frac{1}{\sum \alpha \int h(y(v) | f(v)) d\nu + \sum \beta (1 - \text{IoU}(C_y, C))} $$

Local & global features

$$ H_y(g(v), \hat{y}(v)) = \frac{1}{N} \sum_{i \in [1:N]} \{ (\hat{y}(v) = c)(\log(\hat{y}(v) = c) \} $$

IoU Boundary Loss (Dice Loss)

$$ \text{IoU Loss} = 1 - \frac{\text{IoU}(C_y, C)}{|C_y| + |C|} $$

Introduction

Network Architecture

Results and Evaluation