New directions in saliency research: Developments in architectures, datasets, and evaluation

Tutorial Overview

Zoya Bylinskii & Tilke Judd, ECCV Oct. 8, 2016
Ali Borji  
Center for Research in Computer Vision, University of Central Florida

Zoya Bylinskii  
Computer Science and Artificial Intelligence Lab, MIT

Tilke Judd  
Product Manager, Google  
Previously: MIT
Judd et al. “A Benchmark of Computational Models of Saliency to Predict Human Fixations” [MIT Tech Report 2012]
# mit saliency benchmark

## mit saliency benchmark results: mit300

### images

300 benchmark images (the fixations from 39 viewers per image are not public such that no model can be trained using this data set).

### model performances

**Model Visualizations**

66 models, 5 baselines, 8 metrics, and counting...

Performance numbers prior to September 25, 2014.

Matlab code for the metrics we use.

Sorted by: [AUC-Judd] metric

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Published</th>
<th>Code</th>
<th>AUC-Judd</th>
<th>SIM</th>
<th>EMD</th>
<th>AUC-Borji</th>
<th>sAUC</th>
<th>CC</th>
<th>NSS</th>
<th>KL</th>
<th>Date tested</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline: infinite humans</td>
<td></td>
<td>0.92</td>
<td>1</td>
<td>0</td>
<td>0.88</td>
<td>0.81</td>
<td>1</td>
<td>3.26</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep Gaze 2</td>
<td>Matthias Küpper, Lucas Theis, Matthias Bethge. <em>Deep Gaze 2: Boosting Saliency Prediction with Feature Maps Trained on ImageNet</em> [arXiv 2016]</td>
<td></td>
<td>0.86 (0.84)</td>
<td>0.46 (0.43)</td>
<td>3.96 (4.52)</td>
<td>0.86 (0.83)</td>
<td>0.72 (0.77)</td>
<td>0.92 (0.45)</td>
<td>1.29 (1.18)</td>
<td>0.96 (1.04)</td>
<td>first tested: 26/11/2015 last tested: 13/09/2016 maps from authors (model without center bias in parentheses)</td>
<td></td>
</tr>
<tr>
<td>SALICON</td>
<td>Xun Huang, Chengyao Shen, Xavier Boix, Qi Zhao</td>
<td></td>
<td>0.87</td>
<td>0.60</td>
<td>2.62</td>
<td>0.85</td>
<td>0.74</td>
<td>0.74</td>
<td>2.12</td>
<td>0.54</td>
<td>first tested: 19/11/2014 last tested: 15/11/2015 maps from authors</td>
<td></td>
</tr>
<tr>
<td>DeepFix</td>
<td>Srinivas S S Kruthivas, Kumar Ayush, R. Venkatesh Babu. DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations [arXiv 2015]</td>
<td></td>
<td>0.87</td>
<td>0.67</td>
<td>2.04</td>
<td>0.80</td>
<td>0.71</td>
<td>0.78</td>
<td>2.26</td>
<td>0.63</td>
<td>first tested: 02/10/2015 last tested: 02/10/2015 maps from authors</td>
<td></td>
</tr>
<tr>
<td>Deep Saliency</td>
<td>Nian Liu et al. [unpublished]</td>
<td></td>
<td>0.87</td>
<td>0.68</td>
<td>2.17</td>
<td>0.79</td>
<td>0.72</td>
<td>0.80</td>
<td>2.35</td>
<td>0.55</td>
<td>first tested: 16/03/2016 last tested: 27/07/2016</td>
<td></td>
</tr>
</tbody>
</table>
mit saliency benchmark

2011
- 10 saliency models
- 3 baslines
- 3 metrics (AUC, SIM, EMD)
- 1 dataset

2016
- 66 saliency models
- 5 baslines
- 8 metrics
- 2 datasets
mit saliency benchmark

mit saliency benchmark results: mit300

images
300 benchmark images (the fixations from 30 viewers per image are not public such that no model can be trained using this dataset).

model performances
Model Visualizations
66 models, 5 baselines, 8 metrics, and counting...
Performance numbers prior to September 25, 2014.
Matlab code for the metrics we use.
Sorted by: [ AUC-Judd 4 metric

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Published</th>
<th>Code</th>
<th>AUC-Judd</th>
<th>SIM</th>
<th>EMD</th>
<th>AUC-Borji</th>
<th>sAUC</th>
<th>CC</th>
<th>NSS</th>
<th>KL</th>
<th>Date tested [key]</th>
<th>Sample [img]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline: infinite humans [?]</td>
<td></td>
<td>0.92</td>
<td>1</td>
<td>0</td>
<td>0.88</td>
<td>0.81</td>
<td>3.28</td>
<td>0</td>
<td></td>
<td></td>
<td>first tested: 26/11/2015 last tested: 13/09/2016 maps from authors (model without center bias in parentheses)</td>
<td></td>
</tr>
<tr>
<td>Deep Gaze 2</td>
<td>Matthias Kümmерer, Lucas Theis, Matthias Bethge. Deep Gaze 2: Boosting Saliency Prediction with Feature Maps Trained on ImageNet [arXiv 2014]</td>
<td>0.88</td>
<td>0.46</td>
<td>3.98</td>
<td>0.86</td>
<td>0.72</td>
<td>0.52</td>
<td>1.29</td>
<td></td>
<td>0.95</td>
<td>first tested: 19/11/2014 last tested: 15/11/2015 maps from authors</td>
<td></td>
</tr>
<tr>
<td>SALICON</td>
<td>Xu Huang, Chengyao Shen, Xavier Boix, Qi Zhao</td>
<td>0.87</td>
<td>0.60</td>
<td>2.62</td>
<td>0.85</td>
<td>0.74</td>
<td>0.74</td>
<td>2.12</td>
<td></td>
<td>0.54</td>
<td>first tested: 02/10/2015 last tested: 02/10/2015 maps from authors</td>
<td></td>
</tr>
<tr>
<td>DeepFix</td>
<td>Srinivas S S Kruthiventi, Kumar Ayush, R. Venkatesh Babu. DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations [arXiv 2015]</td>
<td>0.87</td>
<td>0.67</td>
<td>2.04</td>
<td>0.80</td>
<td>0.71</td>
<td>0.78</td>
<td>2.26</td>
<td></td>
<td>0.63</td>
<td>first tested: 10/03/2016 last tested: 27/07/2015 maps from authors</td>
<td></td>
</tr>
<tr>
<td>Deep Spatial Contextual Long-term Recurrent Convolutional</td>
<td>Nian Liu et al. [unpublished]</td>
<td>0.87</td>
<td>0.68</td>
<td>2.17</td>
<td>0.79</td>
<td>0.72</td>
<td>0.80</td>
<td>2.35</td>
<td></td>
<td>0.55</td>
<td>first tested: 14/03/2015 last tested: 27/07/2015 maps from authors</td>
<td></td>
</tr>
</tbody>
</table>
tracking progress with the **MIT Saliency Benchmark**

- 66 models evaluated on MIT benchmark (as of Sept 30, 2016)
- top **8/10** are neural networks
- neural networks make up **25%** of all submissions (17 models since 2014)
What are the common architectures used for neural network models of saliency? How are these models trained?
“Deep networks for saliency map prediction”
Naila Murray

Tutorial overview

architectures

loss functions & evaluation

datasets

applications

semantic segmentation
object detection
object proposals
image clustering & retrieval
cognitive saliency
tracking progress with the MIT Saliency Benchmark
66 models evaluated on MIT benchmark
using 8 evaluation metrics
on 2 datasets (MIT300, CAT2000)
tracking progress with the MIT Saliency Benchmark

gradual improvement over time but saturating
tracking progress with the **MIT Saliency Benchmark**
large boosts in performance in recent years driven by neural network architectures
As model performances continue to approach ground truth, how can we have more meaningful evaluation?
“Evaluating saliency models in a probabilistic framework”
Matthias Kümmerer

Tutorial overview
How can we collect enough training data for neural network models of saliency?
New **data collection** methodologies

**Webcam-based**

- **Xu et al. [ArXiv 2015]**
  - **iSUN**
  
- **Papoutsaki et al. [IJCAI 2016]**

- **Krafka et al. [CVPR 2016]**
New data collection methodologies

**Webcam-based**

- **Xu et al. [ArXiv 2015]** iSUN
- **Papoutsaki et al. [IJCAI 2016]**
- **Krafka et al. [CVPR 2016]**

**Click-based**

- **Jiang et al. [CVPR 2015]** SALICON
- **Kim et al. [CHI EA 2015]**
What kind of new applications are possible with state-of-the-art saliency architectures and big data?
“Saliency for image understanding and manipulation”
Ming-Ming Cheng
Is saliency solved?
“Towards cognitive saliency”
Zoya Bylinskii

Is saliency solved? NO!
“Towards cognitive saliency”
Zoya Bylinskii
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 9:15</td>
<td>Intro &amp; Overview: New directions in saliency</td>
</tr>
<tr>
<td></td>
<td>Tutorial organizers</td>
</tr>
<tr>
<td>9:15 - 9:45</td>
<td>Deep networks for saliency map prediction</td>
</tr>
<tr>
<td></td>
<td>Naila Murray, Xerox Research Centre Europe</td>
</tr>
<tr>
<td>9:45 - 10:15</td>
<td>Evaluating saliency models in a probabilistic framework</td>
</tr>
<tr>
<td></td>
<td>Matthias Kümmerer, University of Tuebingen</td>
</tr>
<tr>
<td>10:15 - 10:45</td>
<td>Saliency for image understanding and manipulation</td>
</tr>
<tr>
<td></td>
<td>Ming-Ming Cheng, Nankai University</td>
</tr>
<tr>
<td>10:45 - 11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00 - 11:30</td>
<td>Towards cognitive saliency</td>
</tr>
<tr>
<td></td>
<td>Zoya Bylinskii, Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>11:30 - 12:00</td>
<td>Research Panel</td>
</tr>
</tbody>
</table>