SAND2017-10633C

Data Visualization Saliency Model:

A Tool for Evaluating Abstract Data Visualizations

Laura Matzen, Michael Haass, Kristin Divis, Zhiyuan Wang & Andy Wilson

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Background

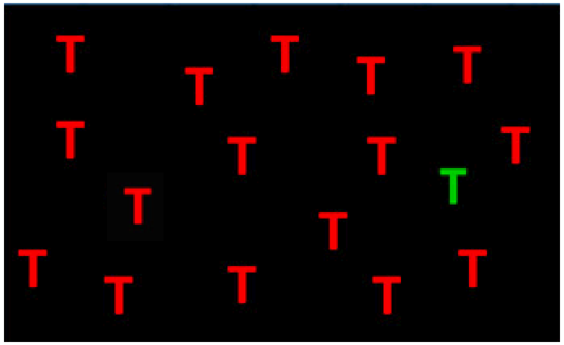
- Analysts often rely on data visualizations when making high-consequence decisions, but little is known about how to evaluate a visualization's effectiveness for an end user

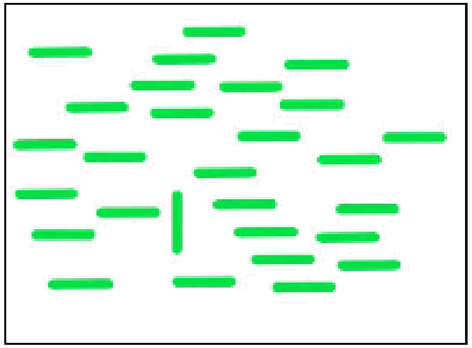
- The field of visual analytics is calling for the creation of models of human cognitive processing that can address this gap and advance our understanding of how humans reason about data visualizations.

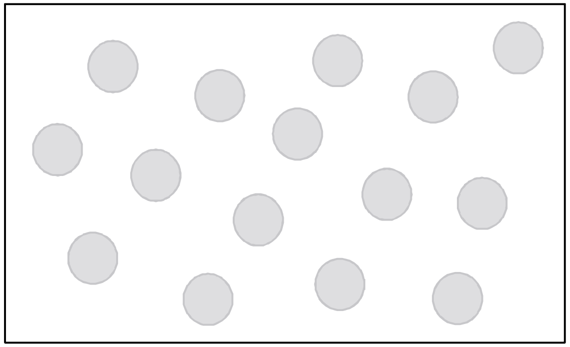
Bottom-up versus top-down visual processing

- Two parallel neural processes that guide visual processing
  - Bottom-up = stimulus-driven visual attention
  - Top-down = goal-oriented visual attention

- Bottom-up attention is captured *automatically* by the physical properties of a stimulus
  - Color, shape, orientation, motion



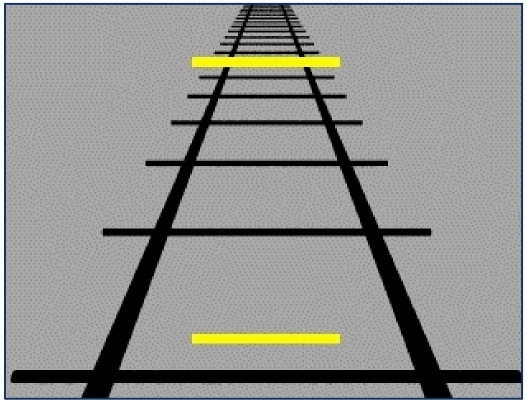


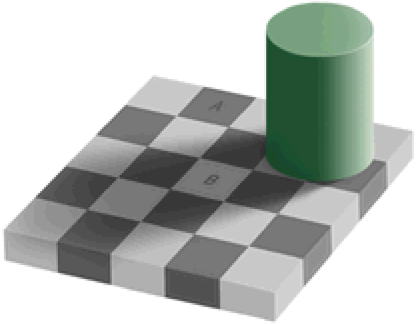


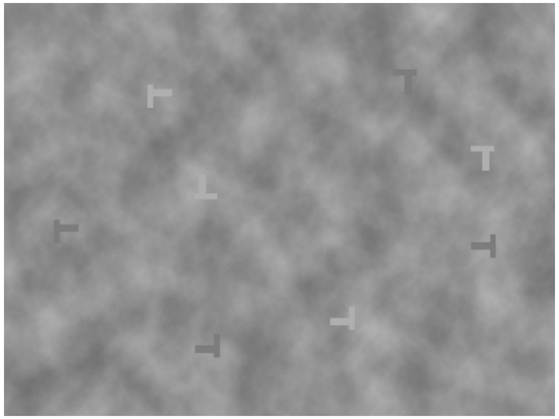
Bottom-up versus top-down visual processing

- Two parallel neural processes that guide visual processing
  - Bottom-up = stimulus-driven visual attention
  - Top-down = goal-oriented visual attention

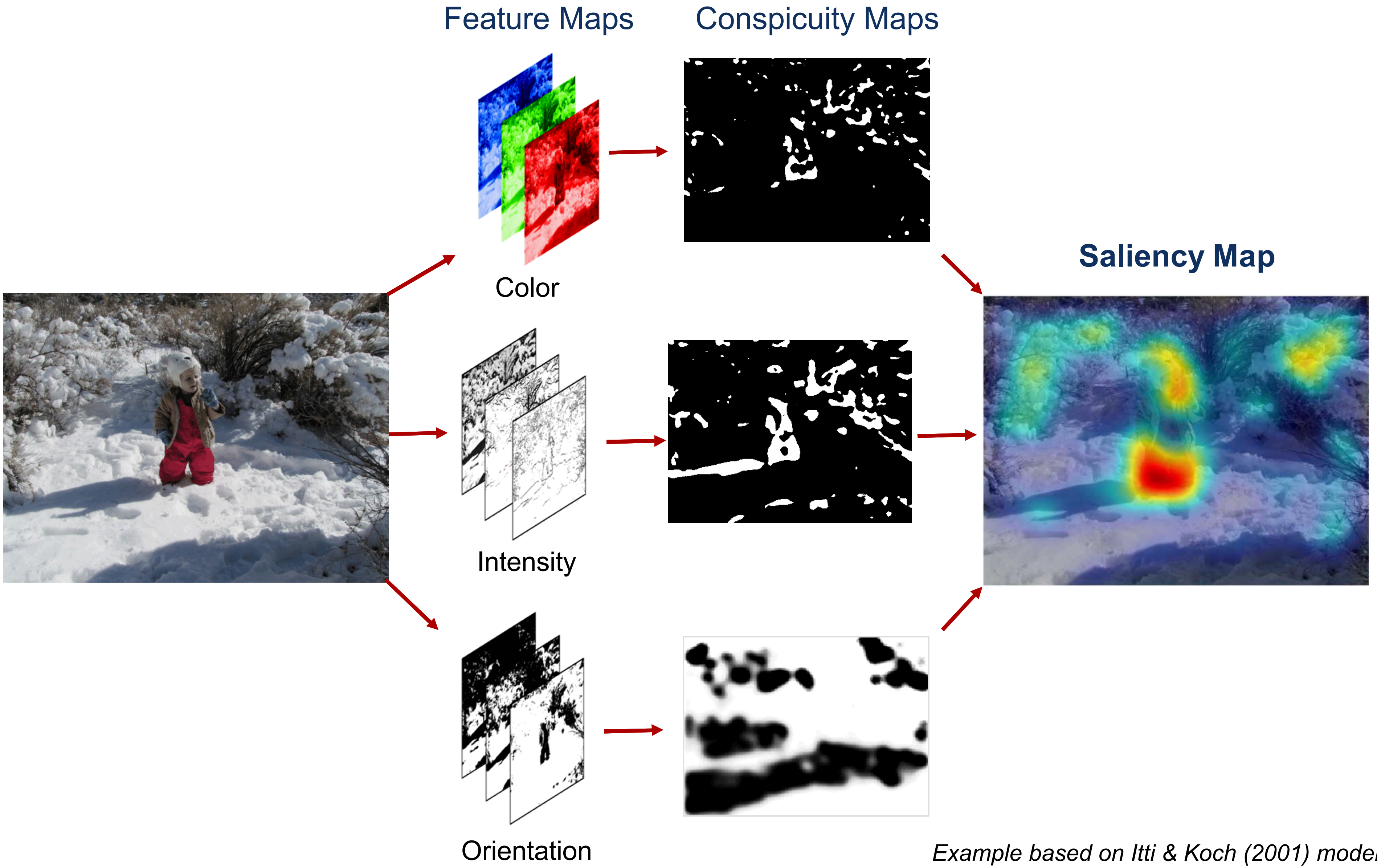
- Top-down attention is allocated *voluntarily* according to the viewer's goals and expectations
  - Current goal, past experience, cognitive load







Bottom-up visual saliency can be modeled

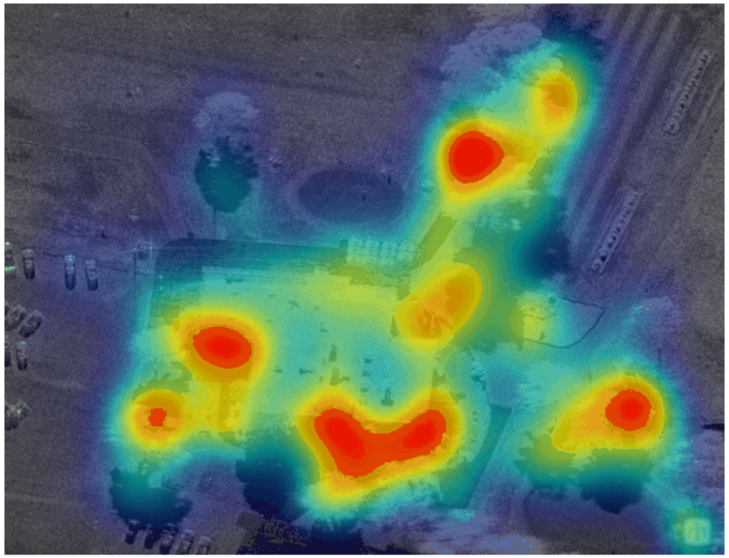


Example based on Itti & Koch (2001) model

Saliency models could be a useful tool for evaluating visualizations

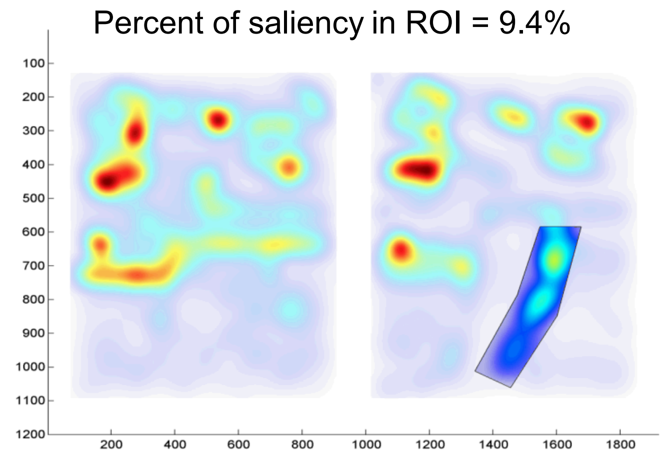
- Will the design draw the viewer's attention to the most important information? (Jänicke & Chen, 2010)
  - Does the bottom-up visual saliency support the viewer's top-down goals?

- This approach works well for scene-like visualizations
  - Spatial properties and features similar to photographs (Matzen et al., 2016)



Visual attention in data visualizations

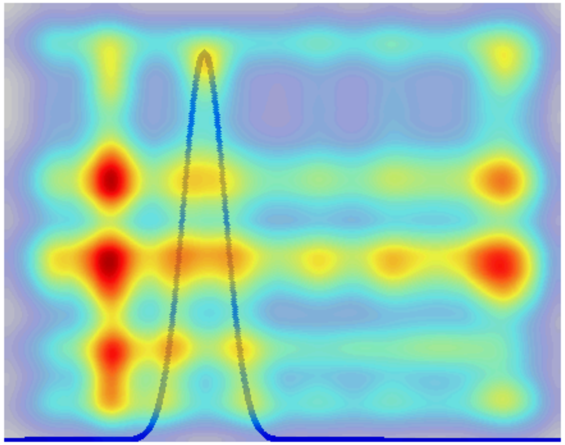
- Ideally, a visualization would draw the viewers' attention to the most important information for their task
  - Information that is important should also be visually salient!



- Maps of visual saliency could provide metrics for iterative evaluation during the design process
  - Designer can assess the match between top-down goals and bottom-up saliency

However…

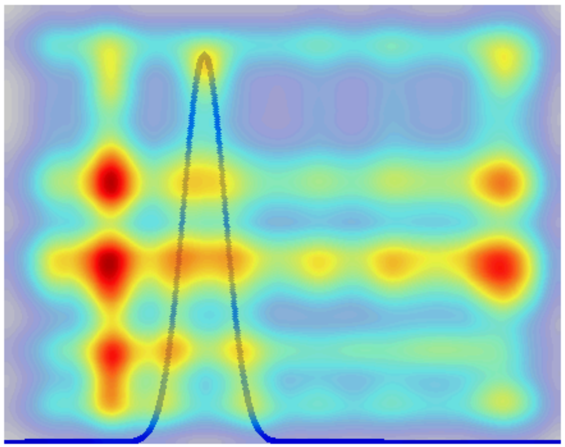
- Existing saliency models fail for abstract visualizations!



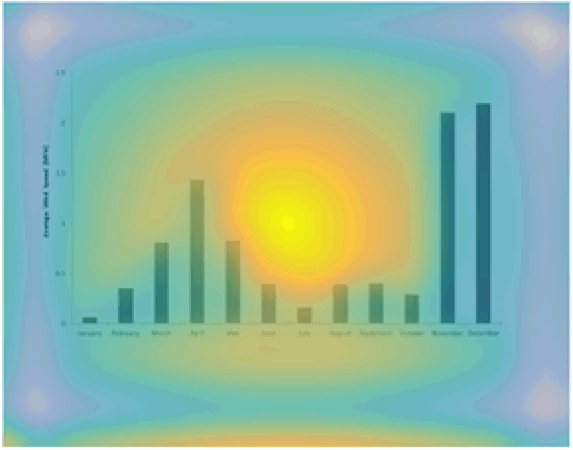
Itti & Koch Model
(Itti & Koch, 2001)

However…

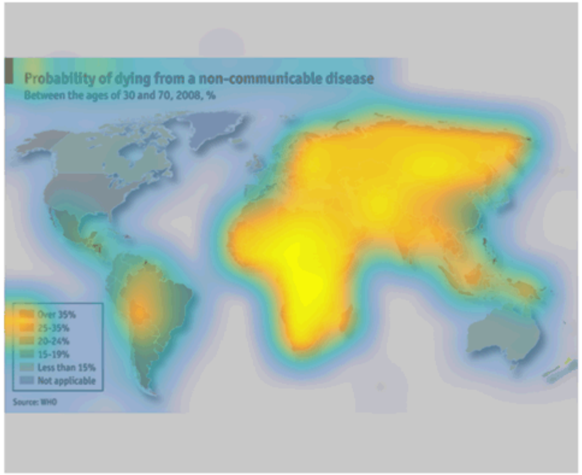
Existing saliency models fail for abstract visualizations!



Itti & Koch Model
(Itti & Koch, 2001)

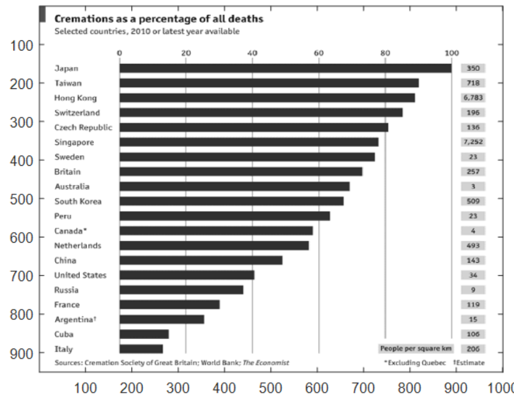


Ensembles of Deep Networks Model (eDN)
(Vig, Dorr & Cox, 2014)

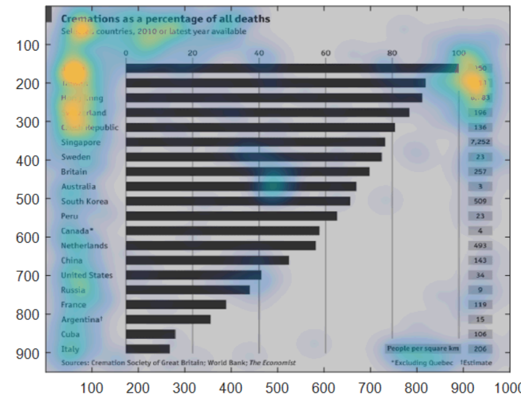


Boolean Map-Based Saliency Model (BMS)
(Zhang & Sclaroff, 2015)

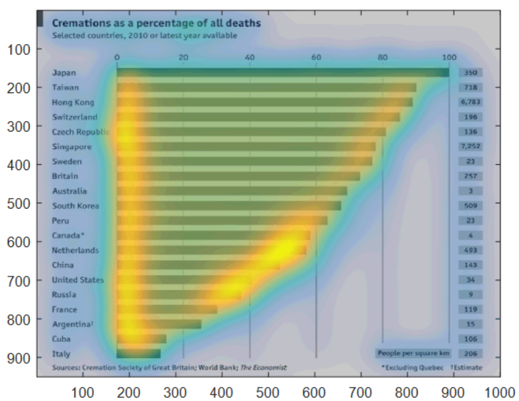
Visualization*



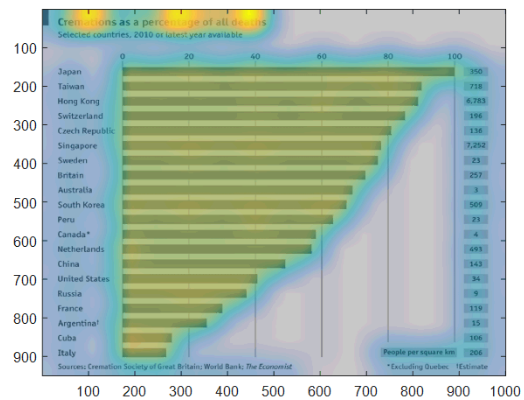
Human Subjects Fixation Map



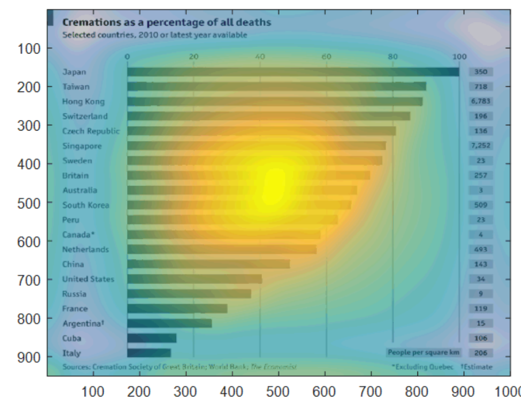
Itti & Koch



BMS



eDN

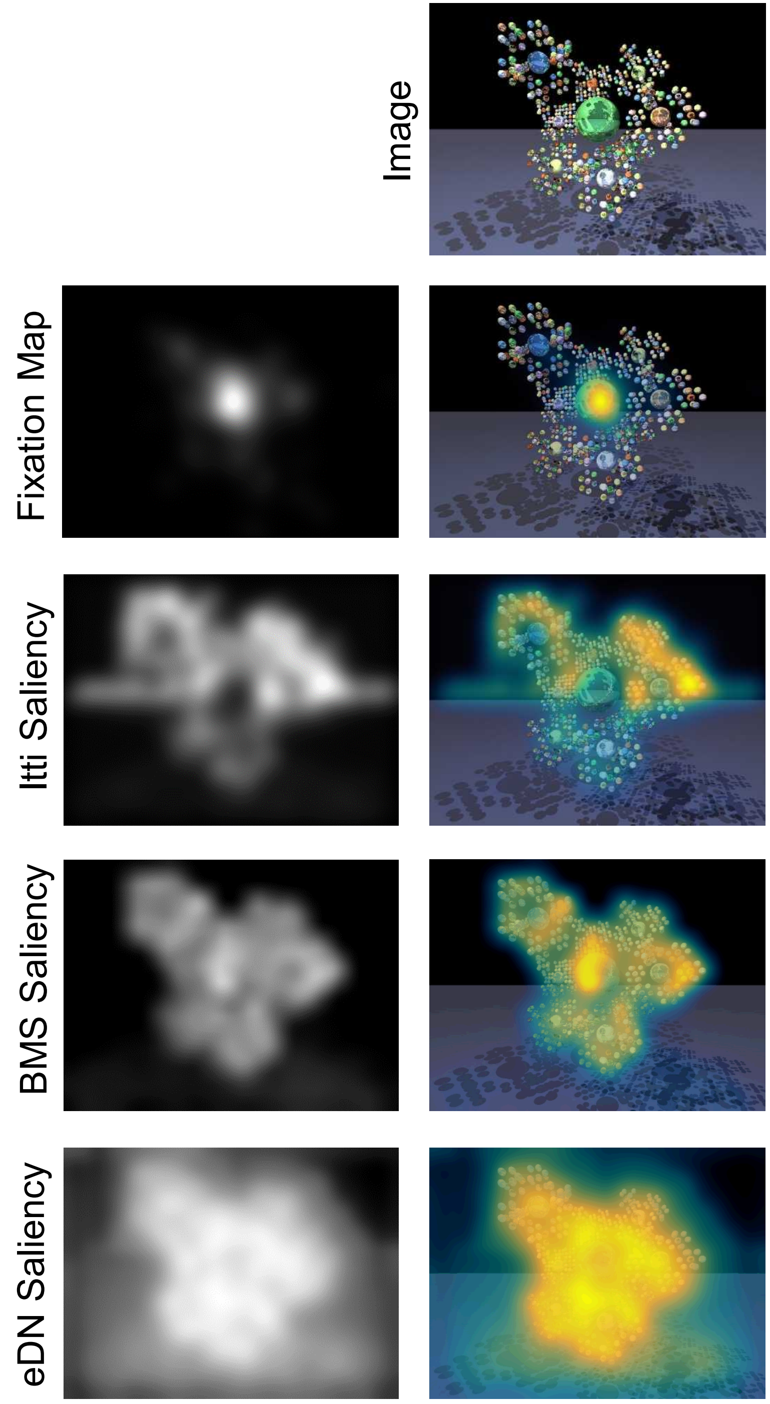


*From MASSVIS Dataset, http://massvis.mit.edu

A note on metrics

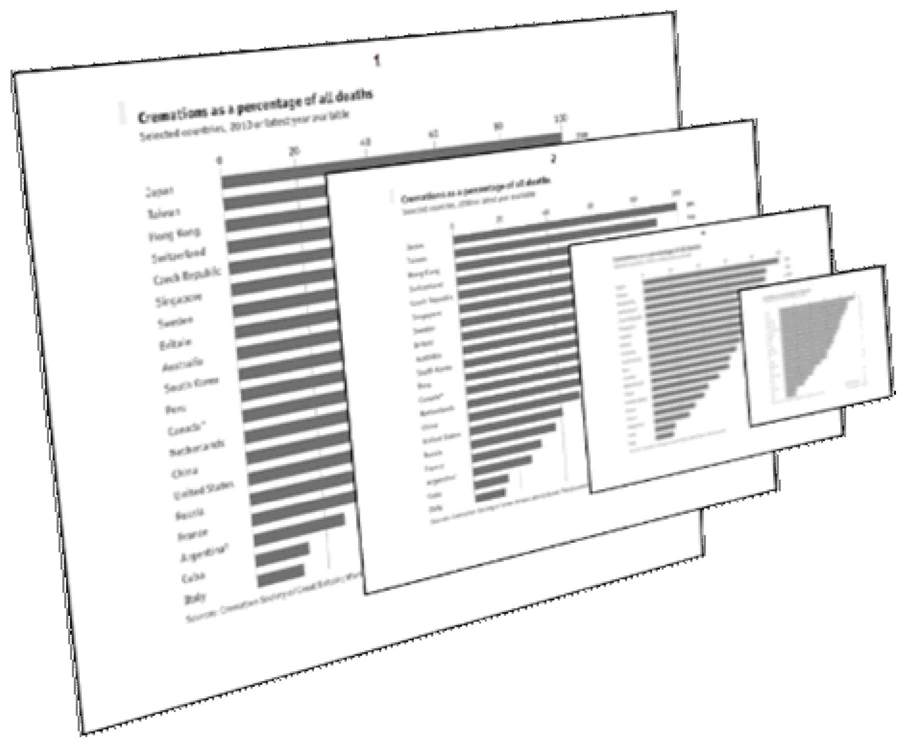
- MIT Saliency Benchmark project (saliency.mit.edu) uses 8 metrics to assess the performance of saliency models by comparing them to maps of human fixations

- Location-based metrics
  - Area under the ROC Curve (AUC)-Judd
  - AUC-Borji
  - Shuffled AUC
- Distribution-based metrics
  - Similarity (SIM)
  - Earth Mover’s Distance (EMD)
  - Pearson’s Correlation Coefficient (CC)
  - Kullback-Leibler Divergence (KL)
- Value-based metric
  - Normalized Scanpath Saliency (NSS)



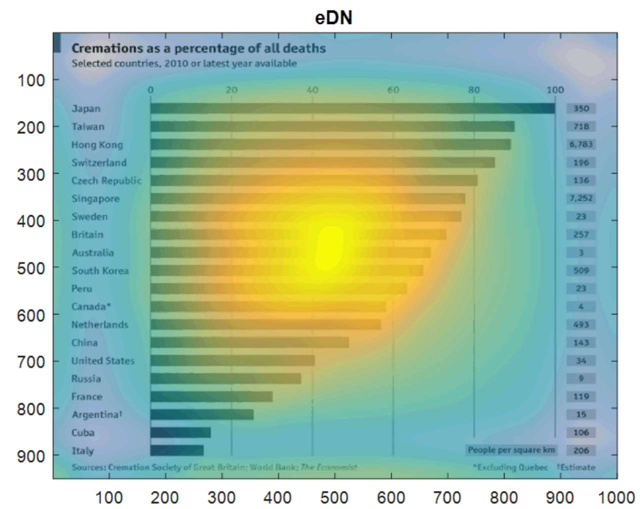
Why do existing models fail for vis?

- Inappropriate spatial scales and weighting
  - Visualizations have features that are very small relative to the extent of the image
    - Input images are resized and smoothed, eliminating fine details



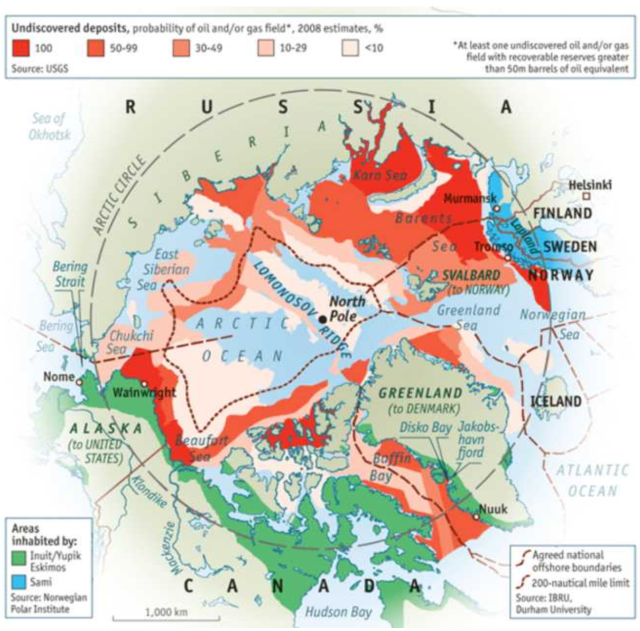
Why do existing models fail for vis?

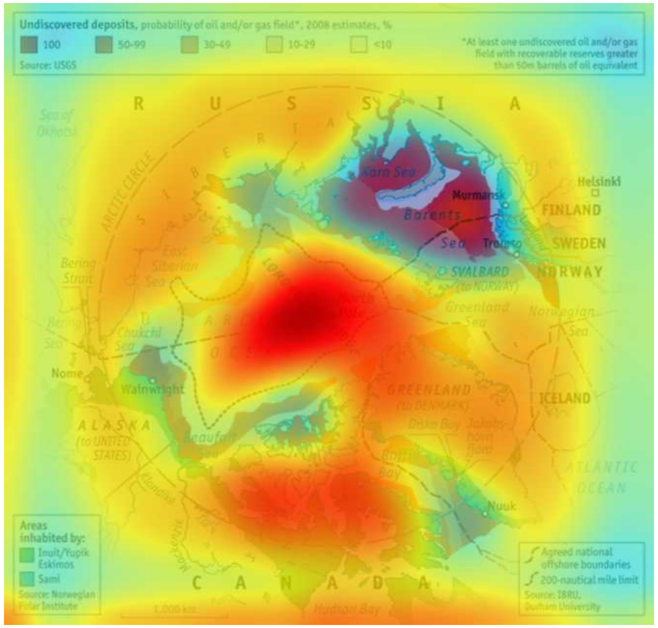
- Inappropriate spatial scales and weighting
  - Visualizations have features that are very small relative to the extent of the image
    - Input images are resized and smoothed, eliminating fine details
  - Center bias incorporated into some models does not hold for vis



Why do existing models fail for vis?

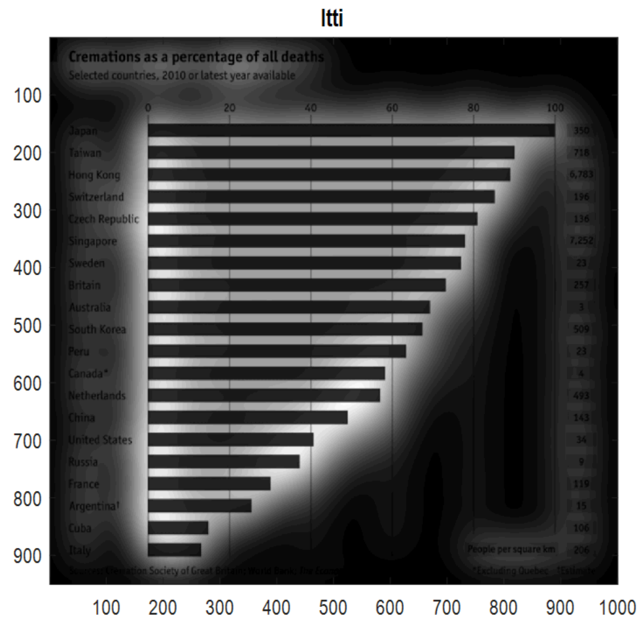
- Inadequate feature sets
  - RGB color space does not correspond well to human color perception

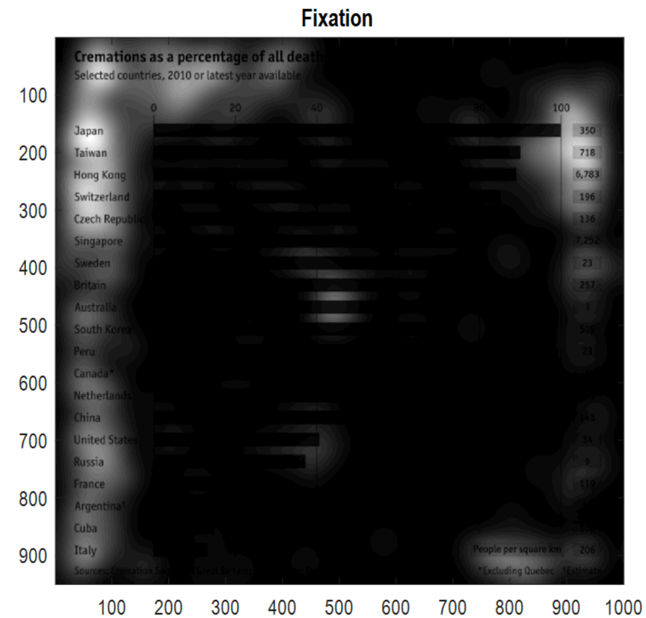




Why do existing models fail for vis?

- Inadequate feature sets
  - RGB color space does not correspond well to human color perception
  - Don’t account for attention to text



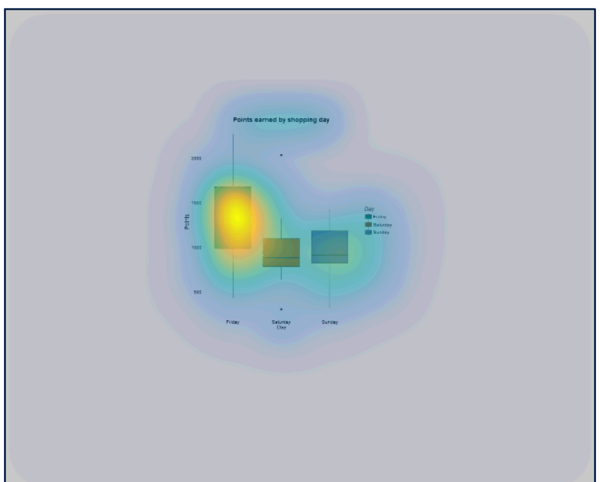


Data Visualization Saliency (DVS) Model

- DVS model provides a significantly better match to human fixation data than prior saliency models
  - Greater than 1 SD improvement for most metrics!

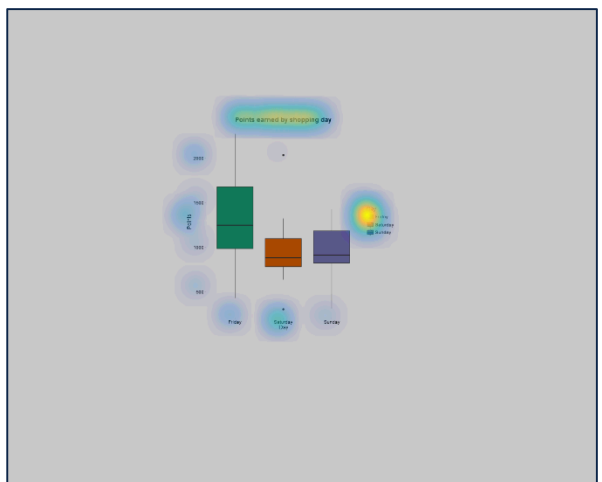
- Weighted combination of two components:

Modified Itti Saliency Map



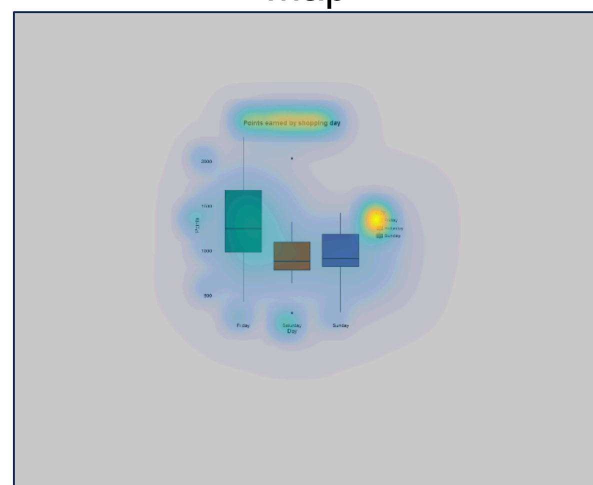
+

Text Saliency Map



=

Data Visualization Saliency Map

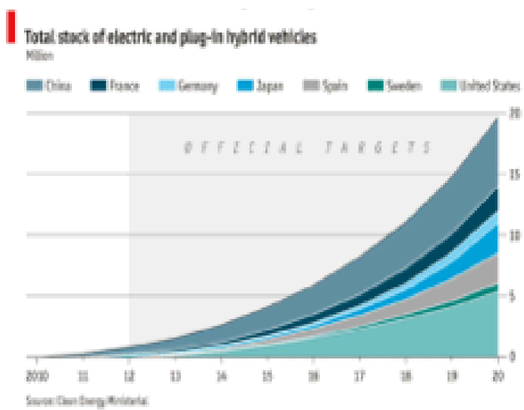


Modified Itti model

- Used model as implemented in the Graph-Vased Visual Saliency code (Harel, Koch & Perona, 2006)
  - Inspired by structure and function of V1 area of human brain
  - Uses color, intensity, and orientation as features
  - Performed best on abstract vis (relative to other models tested)
    - Still had poorer performance for vis than for natural scenes
- Color map changed from RGB to CIE LAB
  - Better approximation of human color perception
  - Change led to 2-15% improvement in performance, depending on the metric

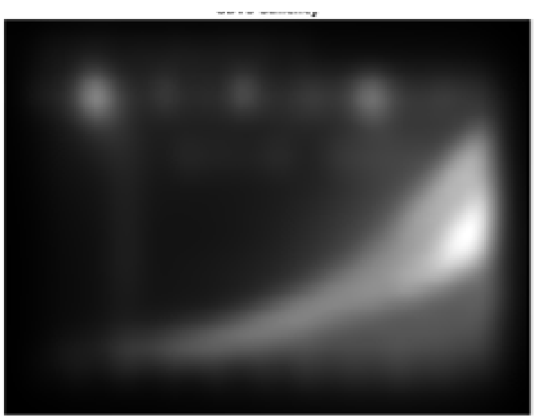
Text Saliency Map

- Implements a hybrid of several published text detection algorithms in the form of a feature map
  - Produces a continuous, probabilistic output that can be incorporated into a saliency map



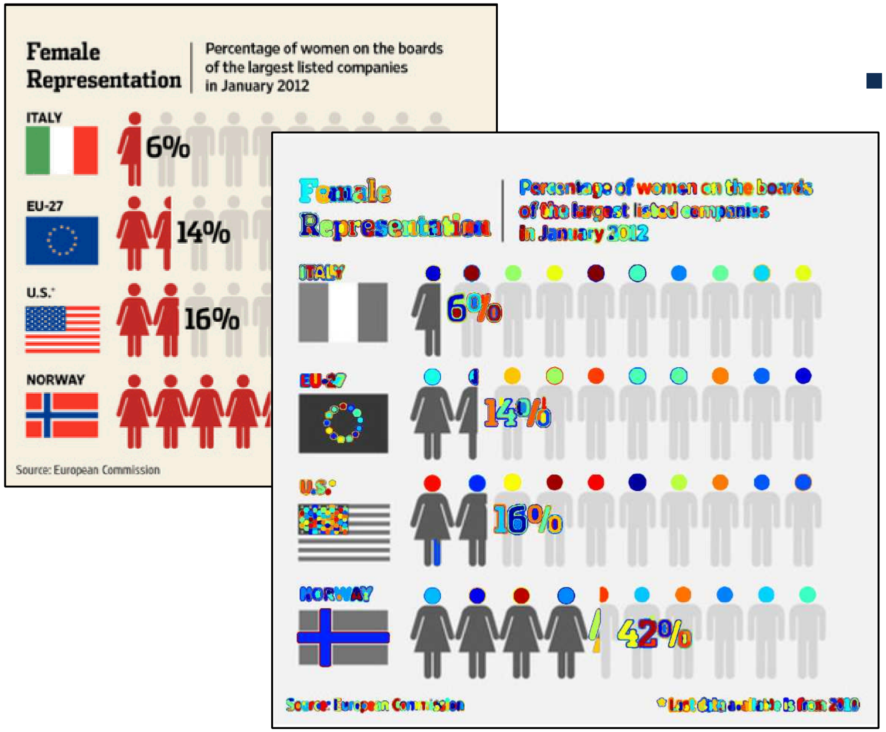
Original Image

Text Saliency Map



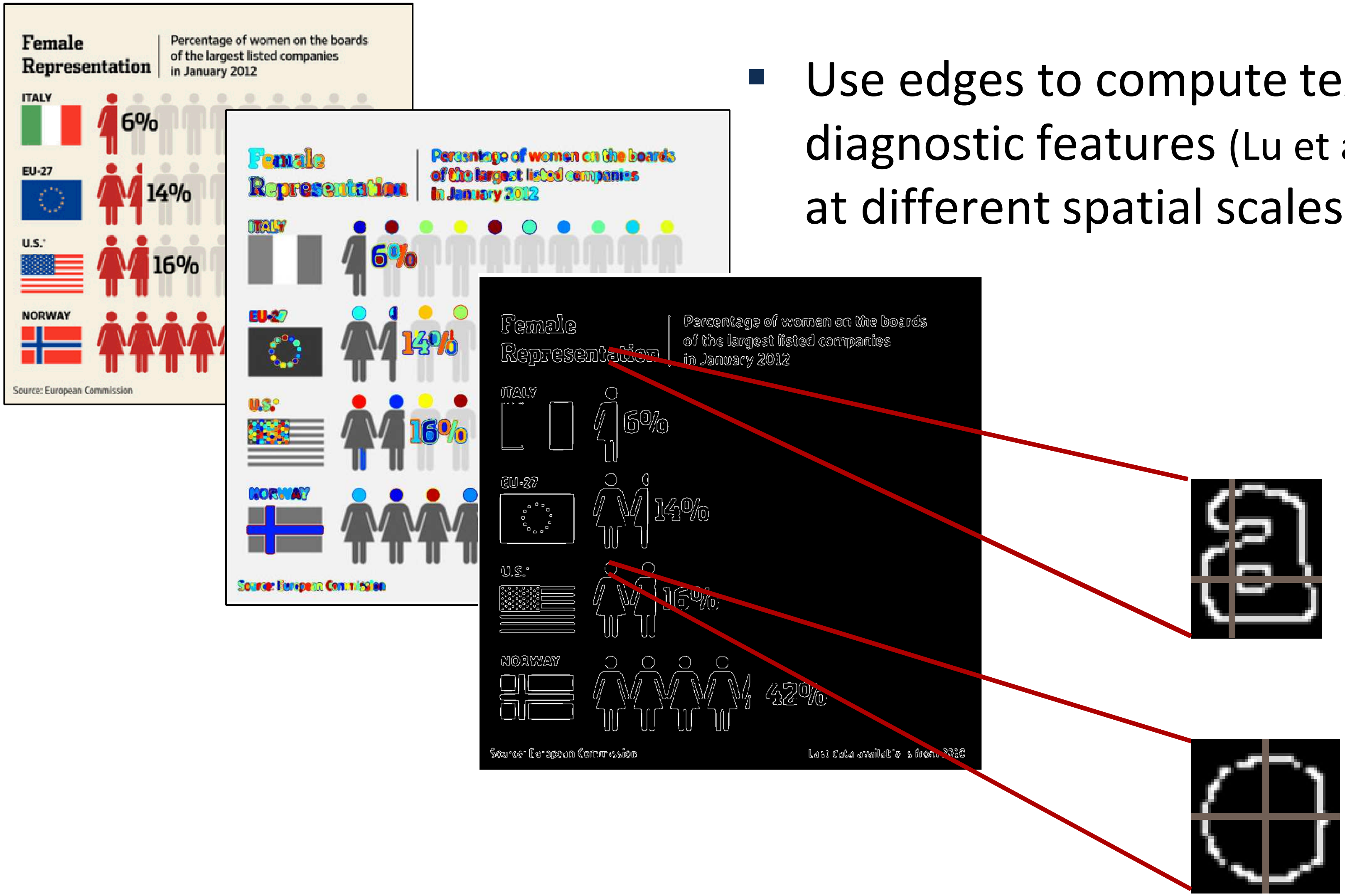
Modified Itti Map

Calculating the Text Saliency Map



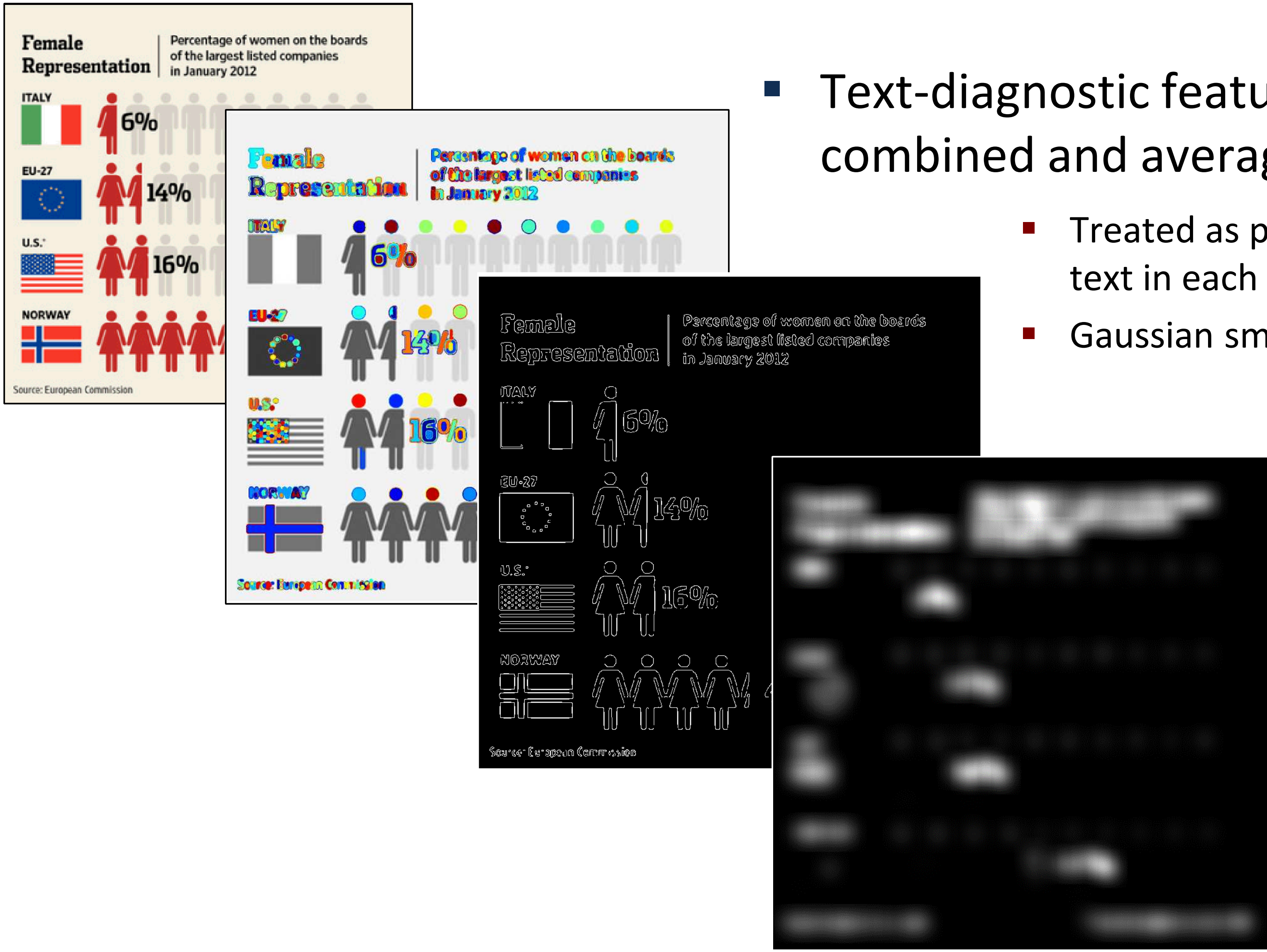
- Find Maximally Stable Extremal Regions (MSER; Matas et al., 2004)
  - Filter according to basic criteria like aspect ratio and stroke width variation
  - Use these regions as a mask to filter the image, then extract edges

Calculating the Text Saliency Map



- Use edges to compute text-diagnostic features (Lu et al., 2015) at different spatial scales

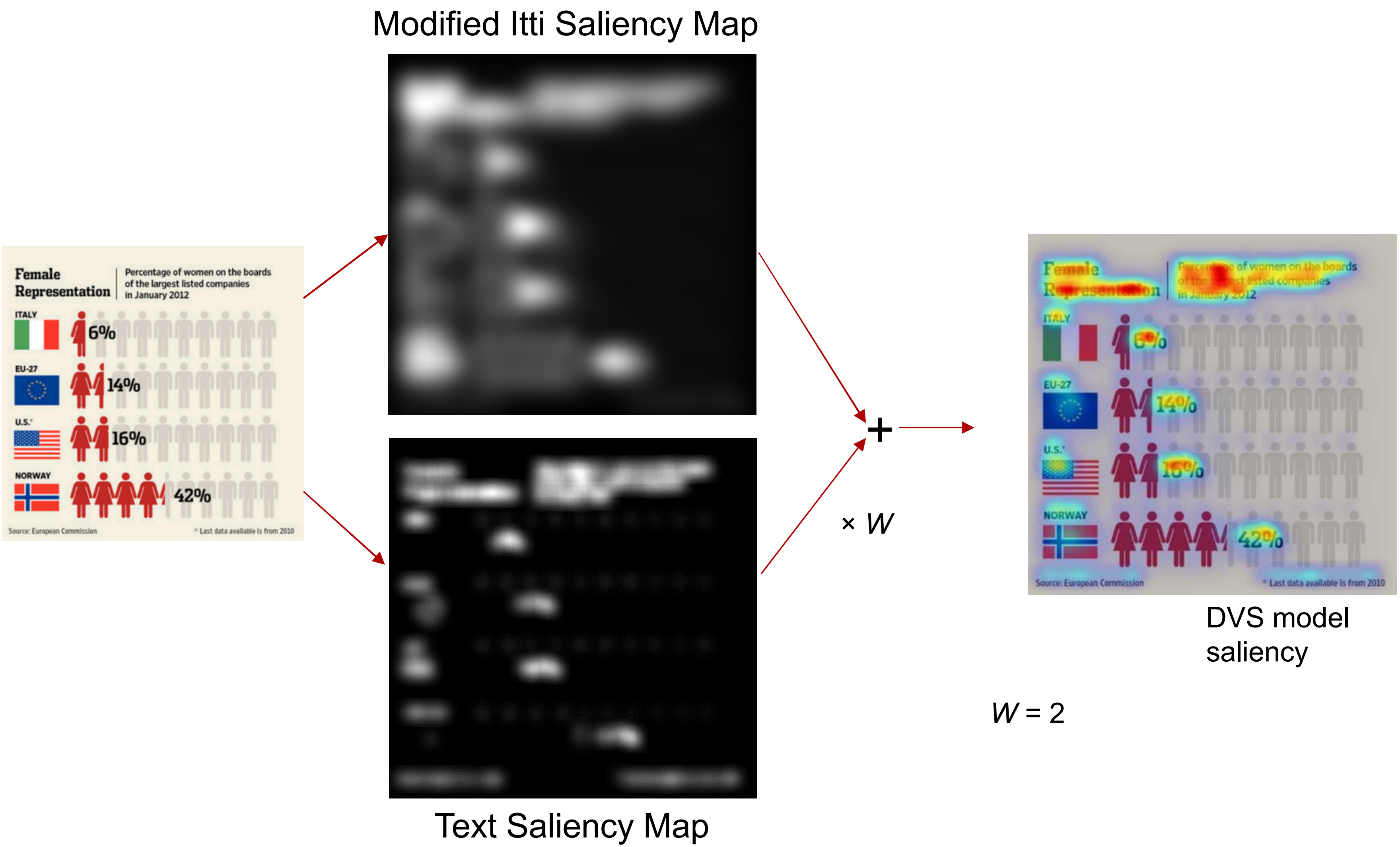
Calculating the Text Saliency Map



- Text-diagnostic feature values are combined and averaged
  - Treated as probability of text in each region
  - Gaussian smoothing applied

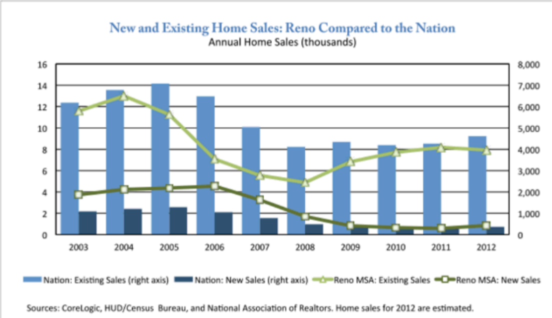
Text Saliency Map

Weighted Combination of Maps

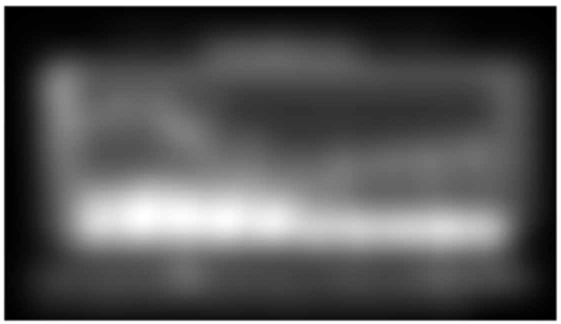


W = 2

Weighted Combination of Maps



Modified Itti Saliency Map



+

Text Saliency Map

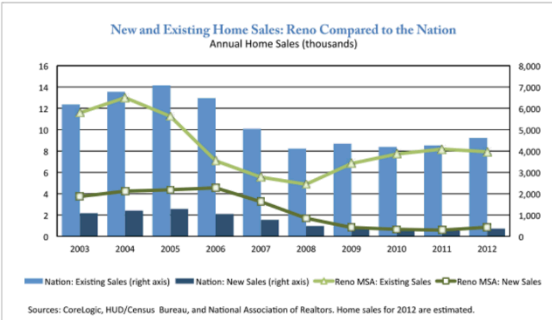
× *W* →

Data Visualization Saliency (DVS) Map

What should *W* be?

- Systematically varied *W* and compared results to eye tracking data from MASSVIS dataset
  - 392 visualizations
- Used all 8 metrics to assess match between DVS maps and fixation maps

Weighted Combination of Maps



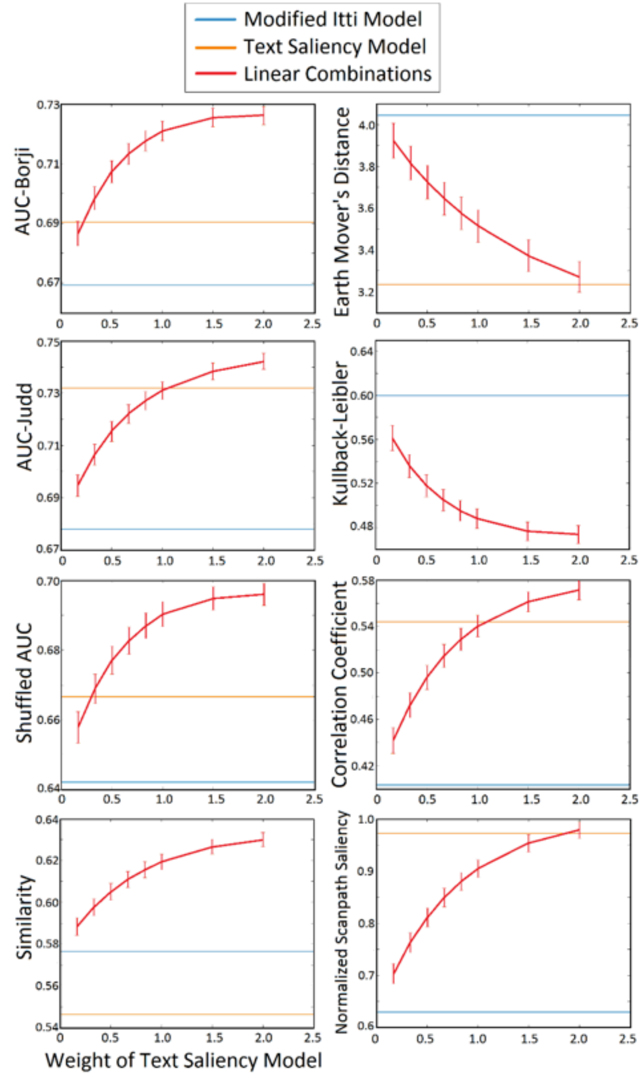
Modified Itti Saliency Map

+

× *W*

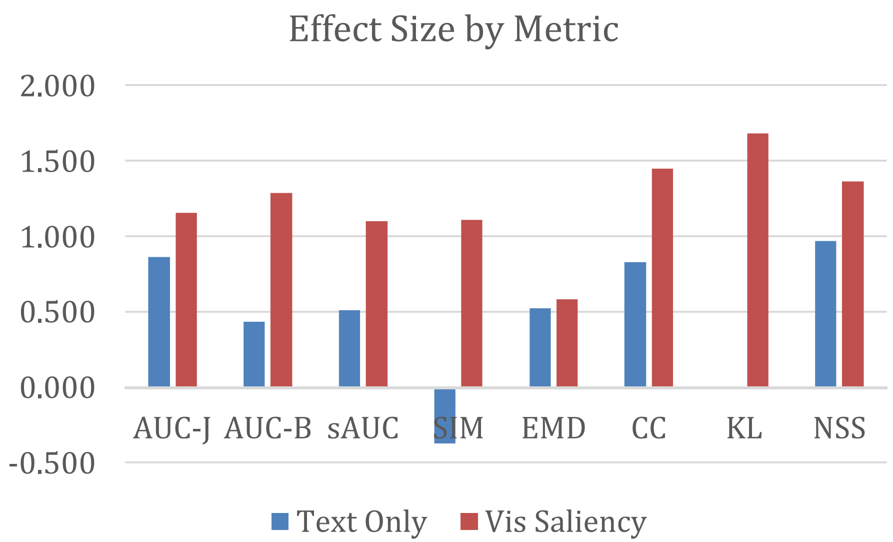
Text Saliency Map

W = 2



How does the model perform?

- Final, weighted DVS model compared to Itti, BMS, and eDN models for MASSVIS dataset (392 visualizations)
  - DVS model performed significantly better than other models for all 8 metrics
- Improvement is large
  - Relative to original Itti model, improvement was >1 SD for 7 of 8 metrics



How does the model perform?

- Final, weighted DVS model compared to Itti, BMS, and eDN models for MASSVIS dataset (392 visualizations)
  - DVS model performed significantly better than other models for all 8 metrics
- Improvement is large
  - Relative to original Itti model, improvement was >1 SD for 7 of 8 metrics

- Limitations of this comparison:
  - MASSVIS data set collected during a memory task, 10 sec viewing time
    - Saliency models usually compared to fixation maps from free viewing tasks with shorter viewing times (3-5 sec)
  - DVS model weights were optimized using this data set
    - Unfair comparison for other models?

Limitations of this comparison

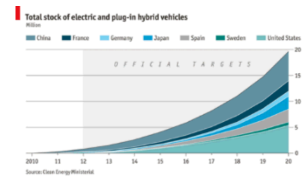
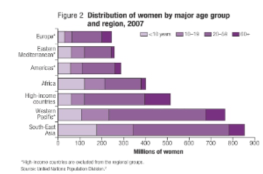
- MASSVIS data set collected during a memory task, 10 sec viewing time
  - Atypical for saliency map assessments
- DVS model weights were optimized using this data set
  - Unfair comparison for other models?

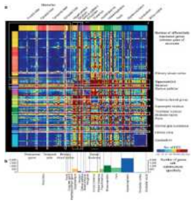
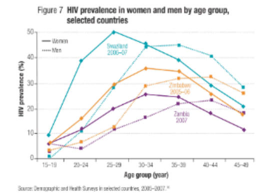
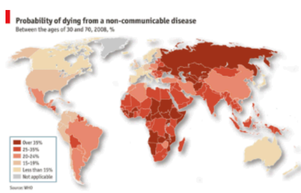
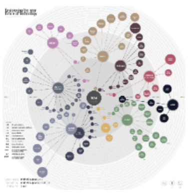
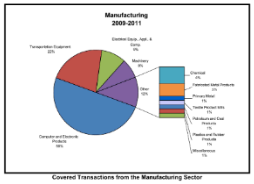
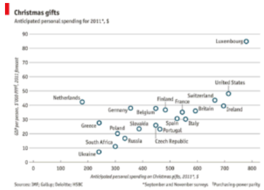
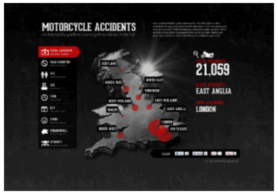
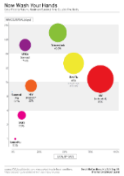
- Collected new eye tracking data set:
  - 30 participants
  - Free viewing task
  - Each image presented for 5 seconds
  - Stimuli from MIT Saliency Benchmark

Parameters typical for saliency modeling research

Stimuli

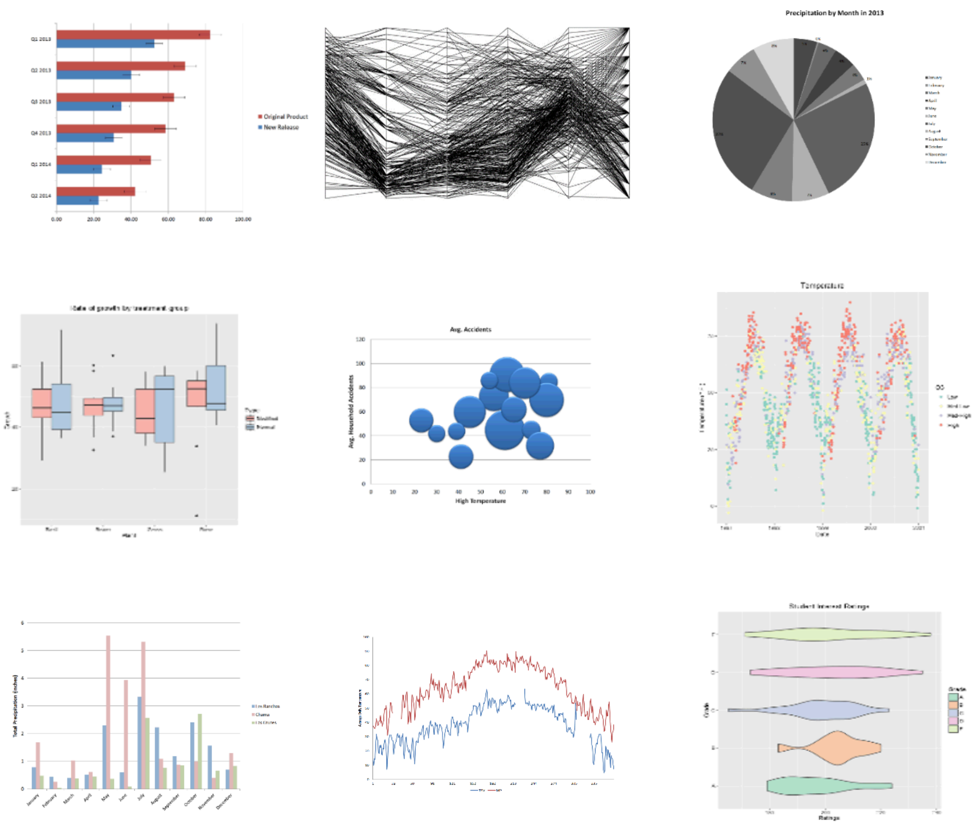
- Four sets of stimuli (108 images):
  - 35 data visualizations from MASSVIS set
    - 4 Area Plots
    - 4 Bar Charts
    - 1 Bubble Plot
    - 4 Column Charts
    - 3 Correlation Plots
    - 3 Line Graphs
    - 2 Maps
    - 3 Network Diagrams
    - 3 Pie Charts
    - 5 Scatter Plots
    - 3 Infographics

Stimuli

- Four sets of stimuli (108 images):
  - 35 data visualizations from MASSVIS set
  - 27 new data visualizations
    - 3 of each of 9 common vis types:
      - – Bar charts
      - – Box plots
      - – Bubble graphs
      - – Column charts
      - – Line graphs
      - – Parallel coordinates plots
      - – Pie charts
      - – Scatter plots
      - – Violin plots



Stimuli

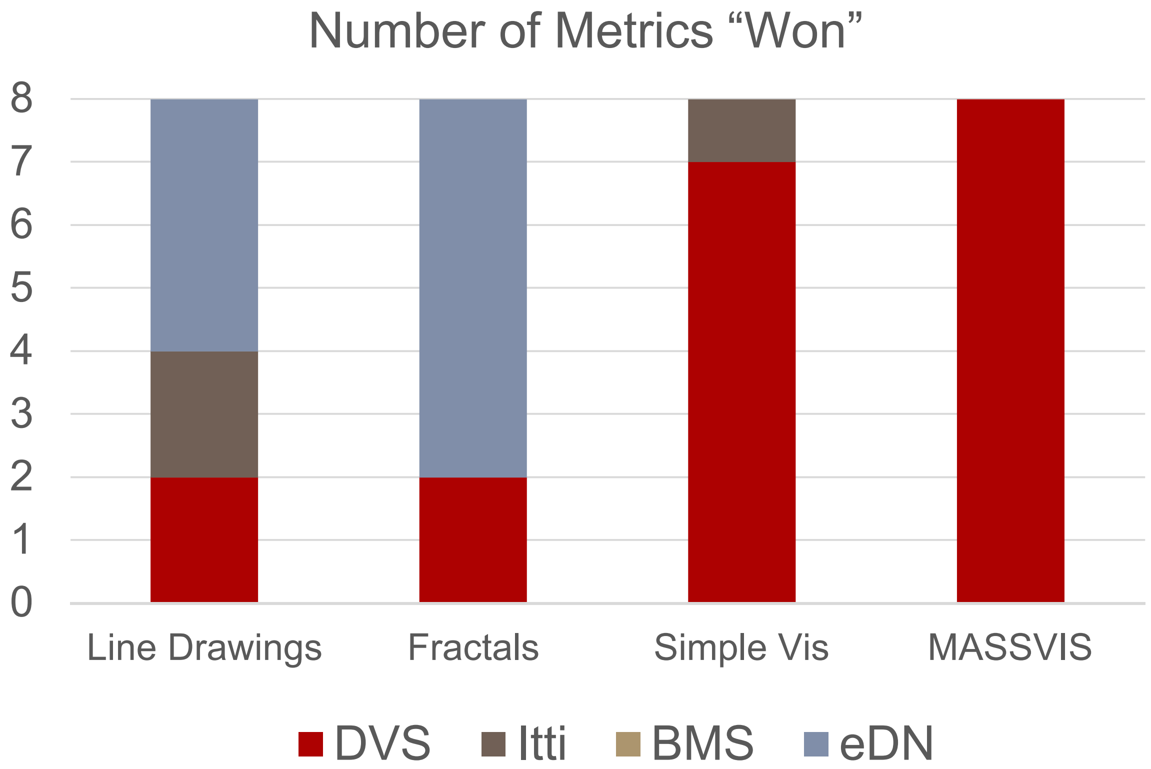
- Four sets of stimuli (108 images):
  - 35 data visualizations from MASSVIS set
  - 27 new data visualizations
    - 3 of each of 9 common vis types
  - Line Drawings from MIT Saliency Benchmark
  - Fractals from MIT Saliency Benchmark

Stimuli

- Four sets of stimuli (108 images):
  - 35 data visualizations from MASSVIS set
  - 27 new data visualizations
    - 3 of each of 9 common vis types
  - Line Drawings from MIT Saliency Benchmark
  - Fractals from MIT Saliency Benchmark

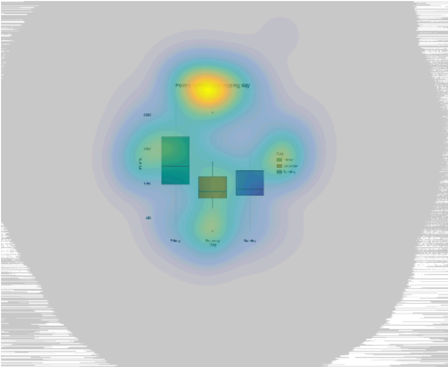
- 8 metrics used to compare eye tracking data collected in this experiment to:
  - DVS Maps
  - Itti Saliency Maps
  - BMS Saliency Maps
  - eDN Saliency Maps

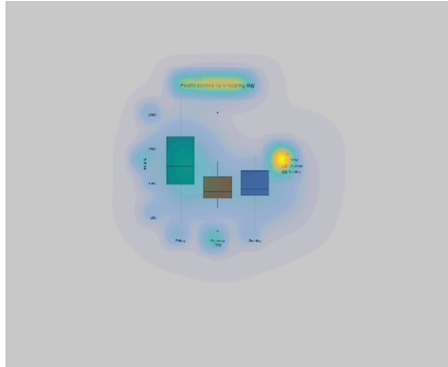
Results

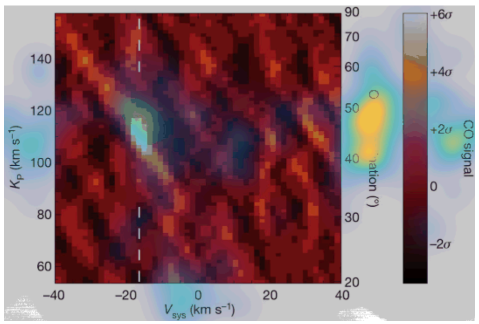


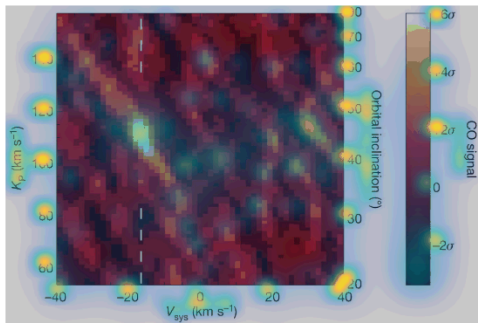
- Simple Vis and MASSVIS stim combined for statistical analysis
  - DVS scores were significantly better than all other models for 7 of 8 metrics
  - For the 8th metric (AUC-Borji), the DVS model's performance was significantly higher than BMS and eDN, but not Itti
- Match between DVS and fixation data approaches match between two sets of fixation data (our study and MASSVIS)

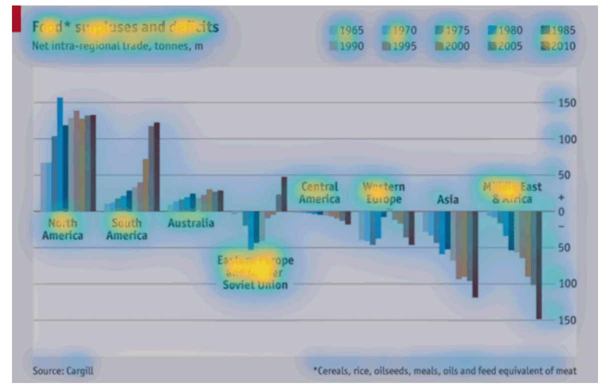
Sandia National Laboratories

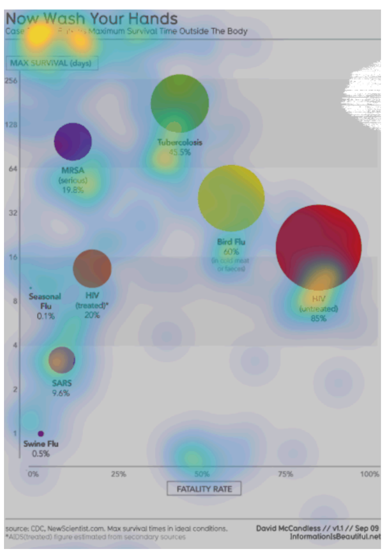


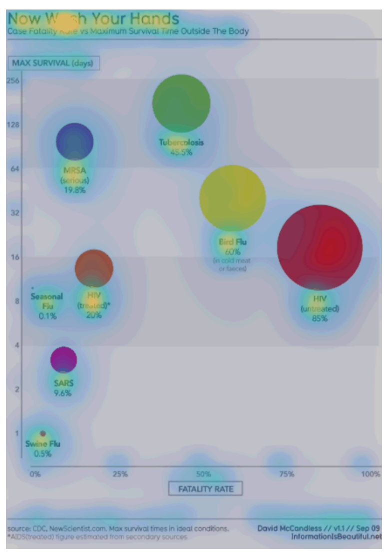


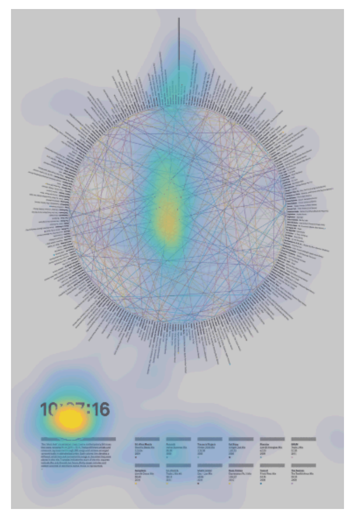


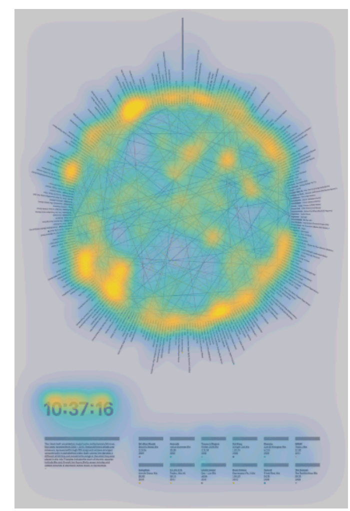






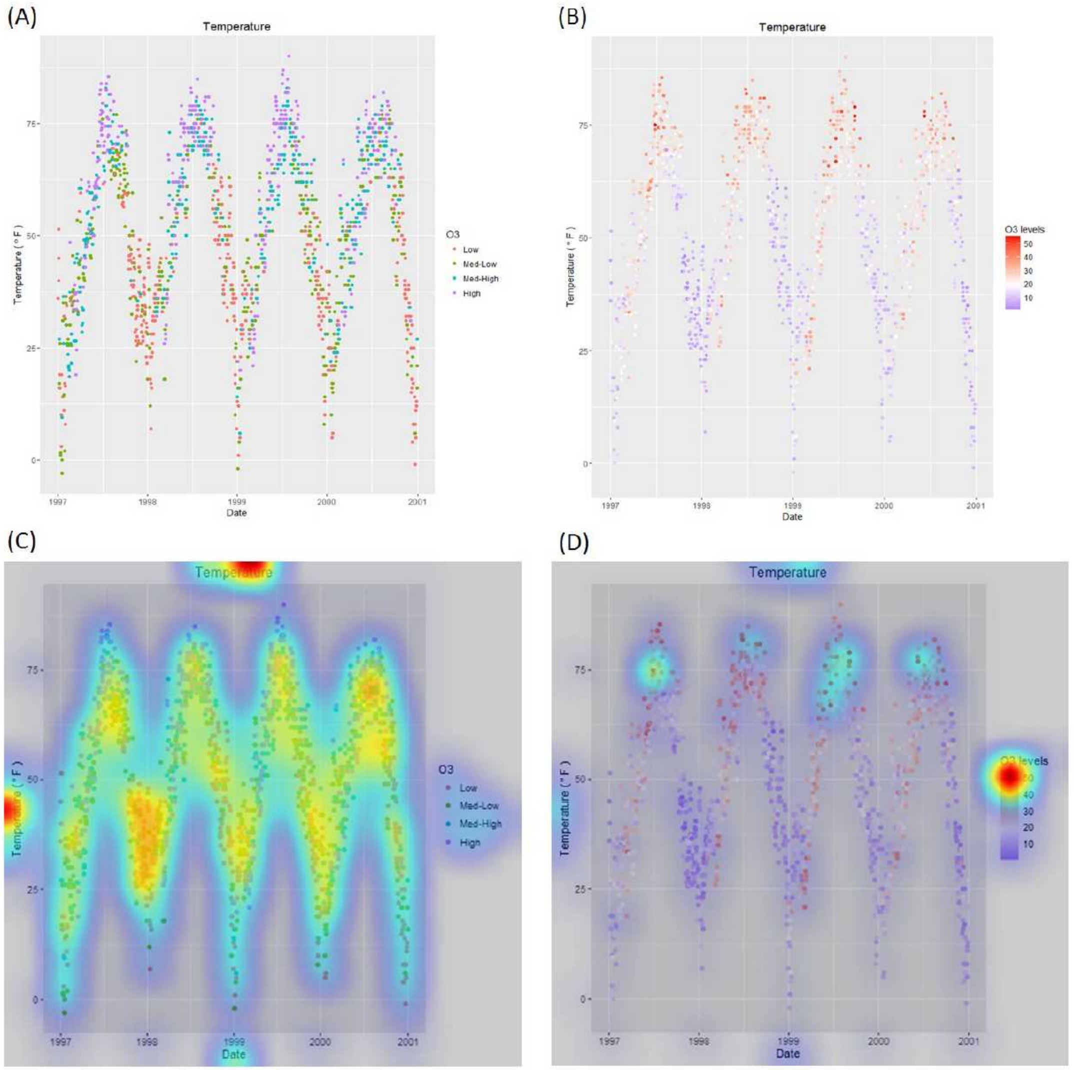






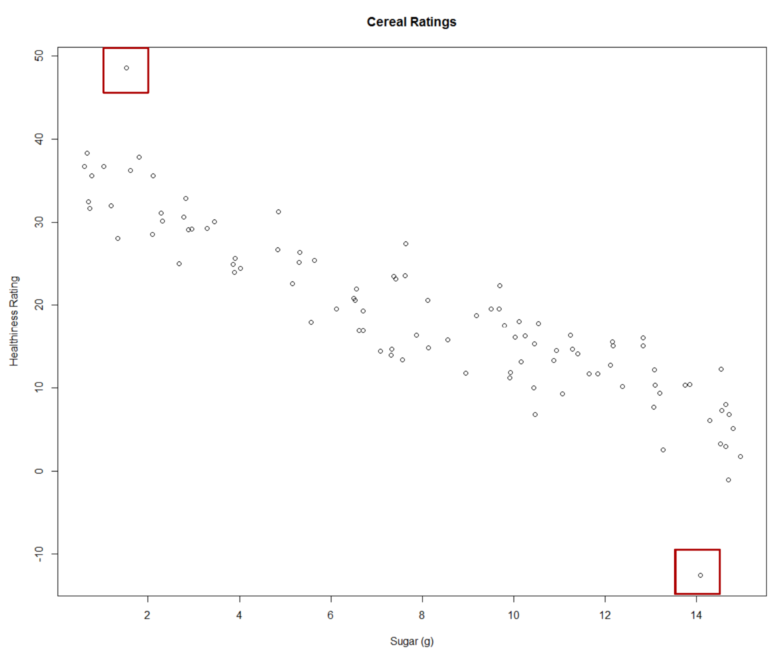
Examples of applying DVS model - Qualitative

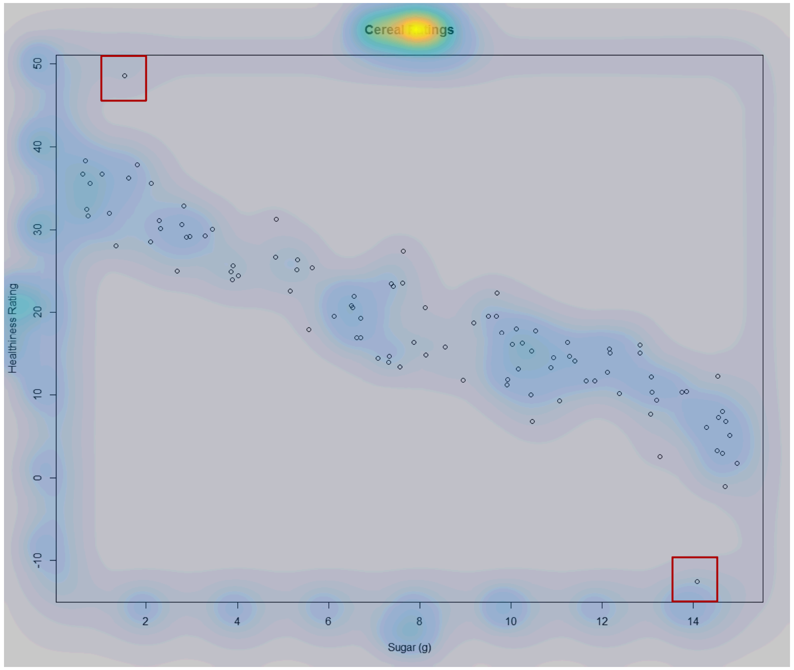
- Same data plotted two ways
  - Default ggplot2
  - Diverging color scheme
- Which to choose?
  - Default colors are equally salient, draw attention to overall shape
  - Diverging color map draws attention to the highest values



Examples of applying DVS model – Qualitative or Quantitative

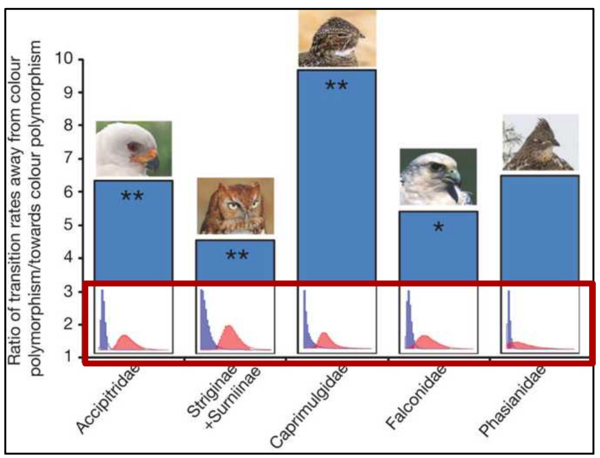
- Saliency maps can be compared to a relevancy map defined by the vis designer (Jänicke and Chen, 2010)
  - Comparisons can be done categorically or using one (or more) metrics
  - Can also define regions of interest and calculate the percentage of saliency

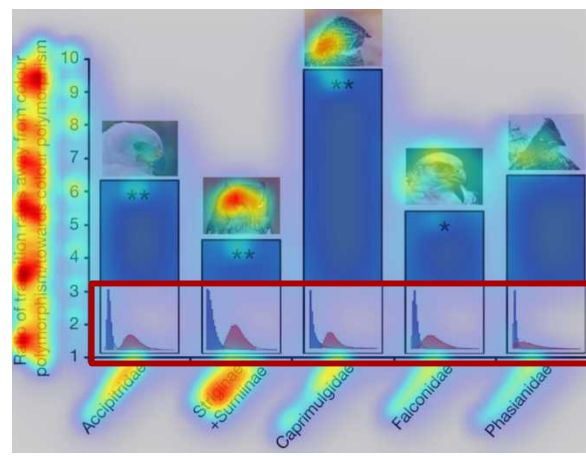




Examples of applying DVS model – Qualitative or Quantitative

- Saliency maps can be compared to a relevancy map defined by the vis designer (Jänicke and Chen, 2010)
  - Comparisons can be done categorically or using one (or more) metrics
  - Can also define regions of interest and calculate the percentage of saliency





Summary

- The Data Visualization Saliency Model can provide predictions of which visual features and regions of a vis are most likely to draw the viewer's attention
- We suggest that DVS maps would be a useful tool for conducting qualitative or quantitative evaluations during the design process
  - Could be particularly useful for assessing emphasis effects
- Incorporating a better color map improved the Itti model's performance in general
- Adding a text map as an additional feature dramatically improved model performance for data visualizations
  - In some ways, this is adding a top-down component…

- Future directions:
  - Investigate additional features to capture other common elements of data visualizations (glyphs, clusters, etc.)
  - Investigate addition of Gestalt-like features, similar to BMS model

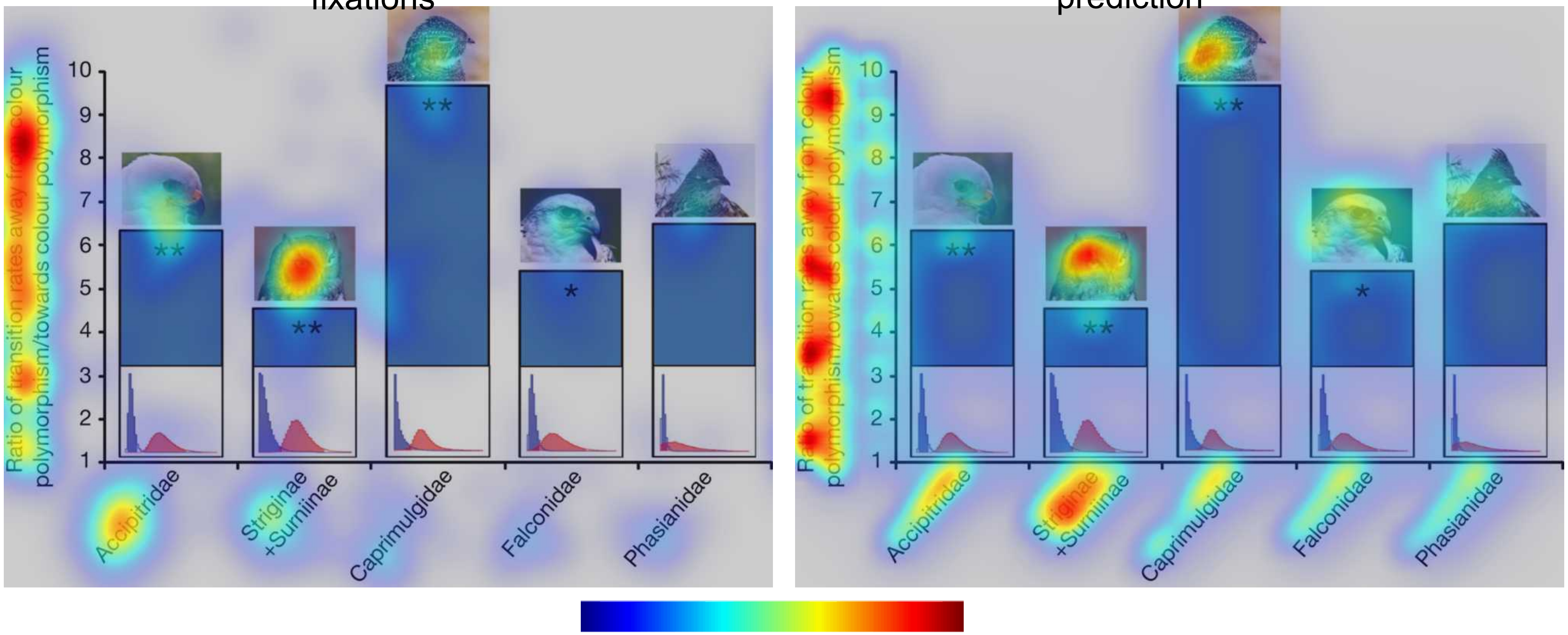
Limitations and Future Directions

- DVS model currently only applies to static images
  - Motion strongly captures human attention and many visualizations incorporate motion
  - Adding motion detection algorithms would extend the utility of the model
- Spatial scaling is still problematic
  - Fine details (other than text) can be lost due to resizing and smoothing
  - Future work: Allowing larger input images, exploring the impact of changing the scales at which the feature maps are calculated
- Focus on bottom-up processing
  - Inclusion of text as a feature adds a top-down component
  - Other top-down features could be added, but this could reduce the generalizability of the model
  - Future work: Incorporate Gestalt-based features

Backup Slides

Actual fixations

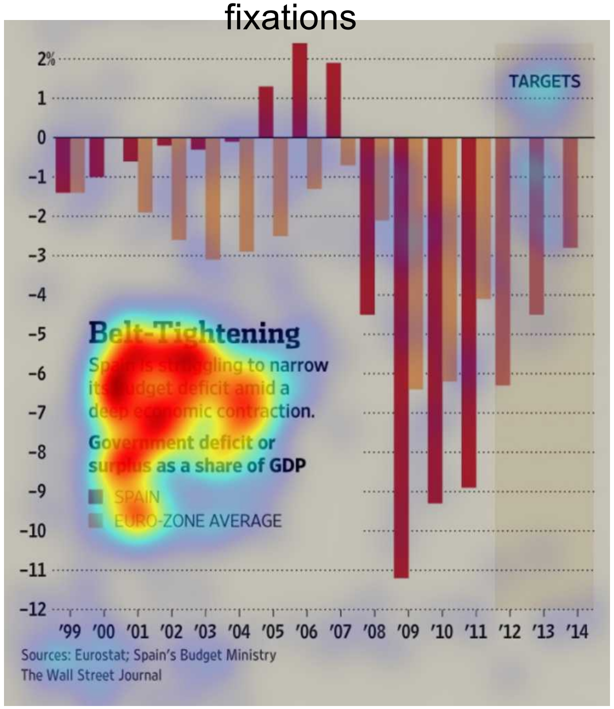
DVS model prediction



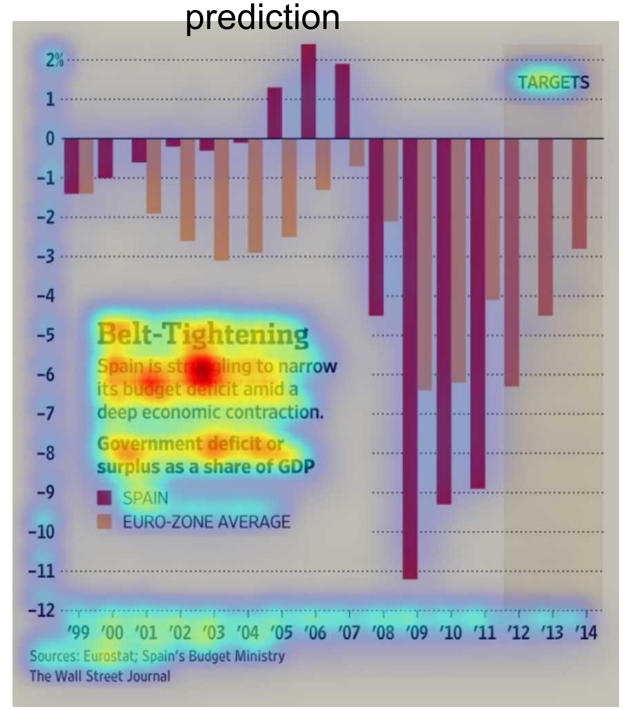
Least fixated

Most fixated

Actual fixations



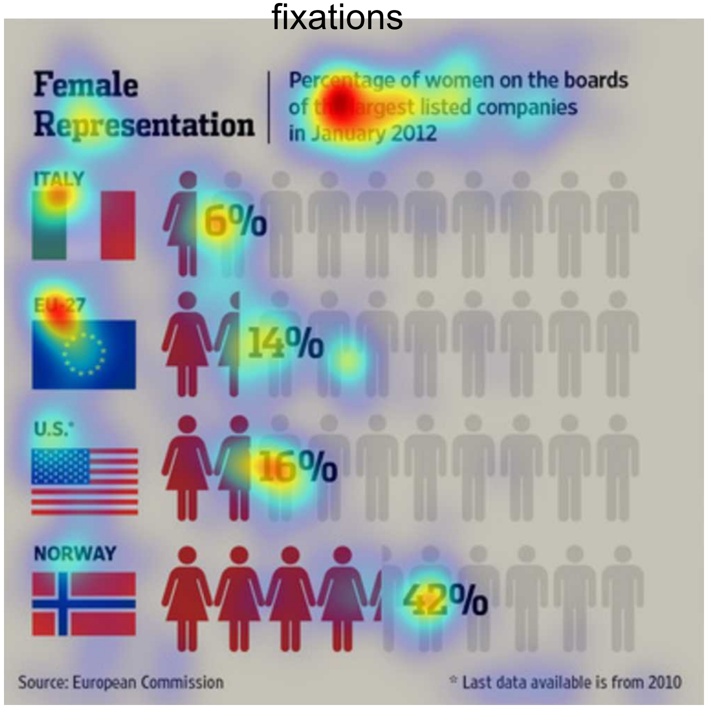
DVS model prediction



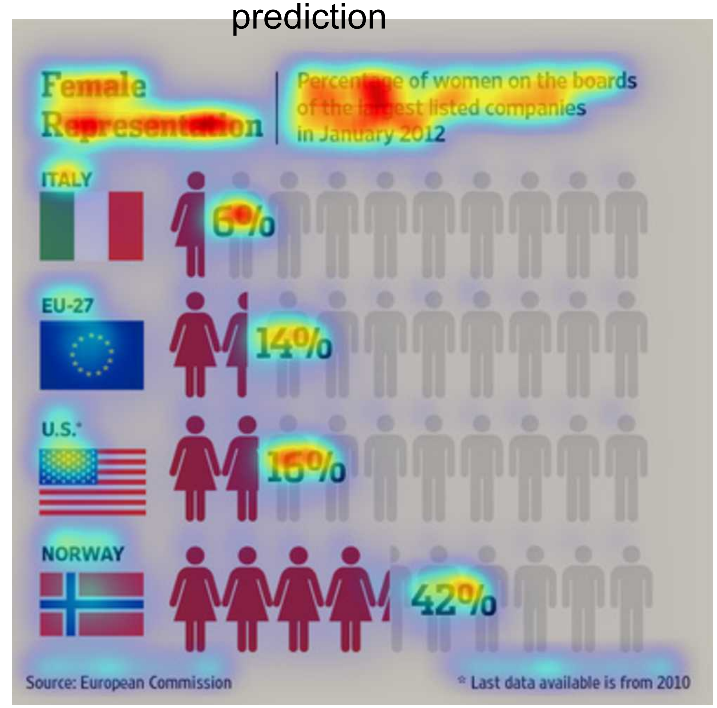
Least fixated

Most fixated

Actual fixations



DVS model prediction



Least fixated

Most fixated

A closer look at attention to text in data visualizations

- Analyzed two sets of eye tracking data:

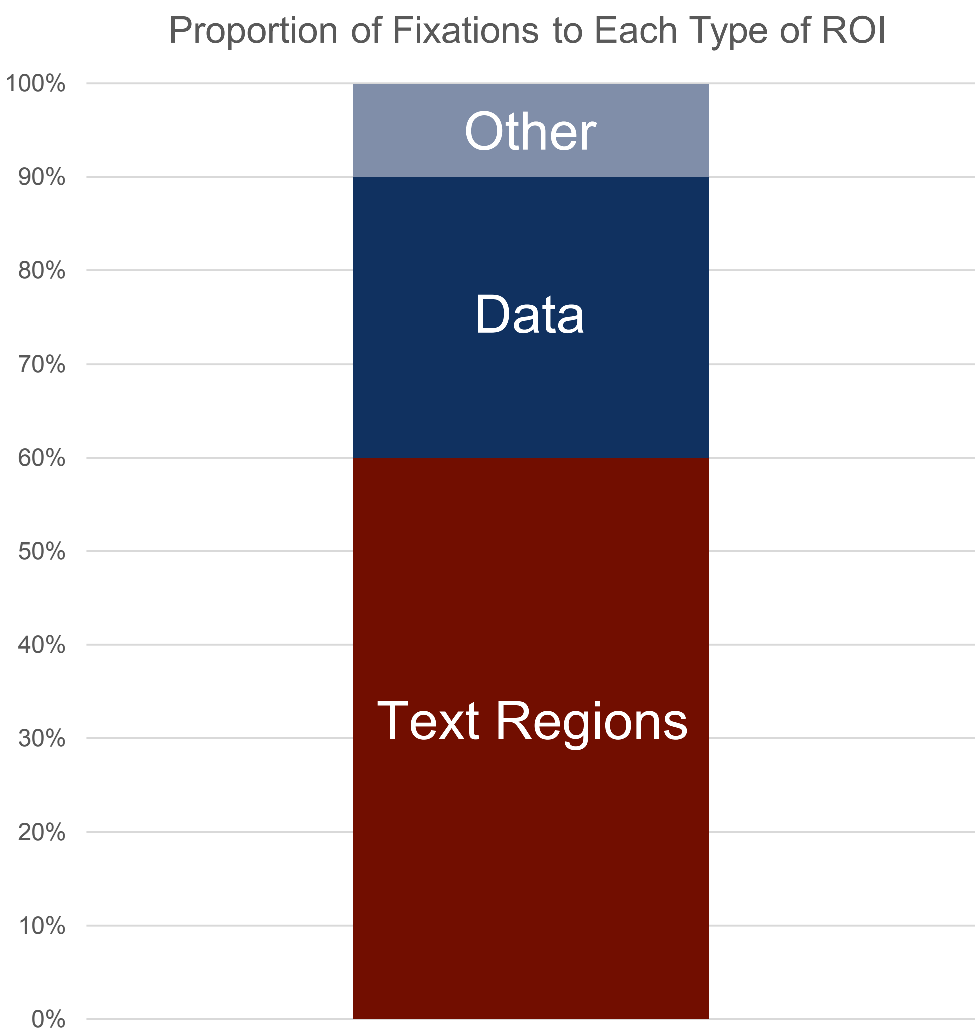
- MASSVIS dataset (Borkin et al., 2013)
  - Stimuli are visualizations collected from “the wild” (magazines, government reports, etc.)
  - Memory task, each stimulus viewed for 10 seconds
    - This task differs from tasks commonly used to evaluate saliency models…

- Collected a new eye tracking dataset
  - Subset of stimuli from the MASSVIS set
  - Created new stimuli -- common types of visualizations
  - Free viewing task, each stimulus viewed for 5 seconds

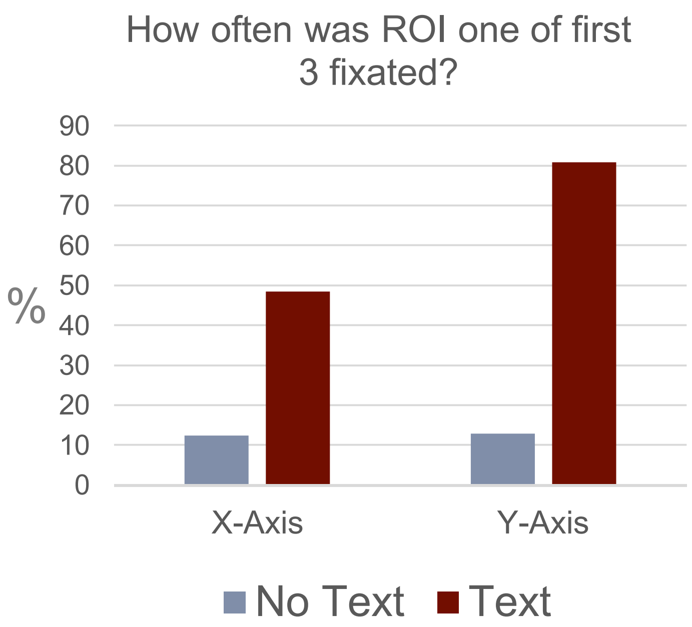
MASSVIS dataset

- http://massvis.mit.edu
- 184 visualizations with corresponding eye tracking data
- Subset of 35 visualizations used for our analysis
- An average of 16 viewers per visualization

Results



Proportion of Fixations to Each Type of ROI
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Other
Data
Text Regions



How often was ROI one of first 3 fixated?
%
90
80
70
60
50
40
30
20
10
0
X-Axis
Y-Axis
No Text
Text

Discussion

- **Performance of visual saliency models should improve if text is incorporated as a feature**
  - Viewers devote a great deal of attention to the text in data visualizations
  - Proportion of fixations to text is equal to or greater than proportion to data itself

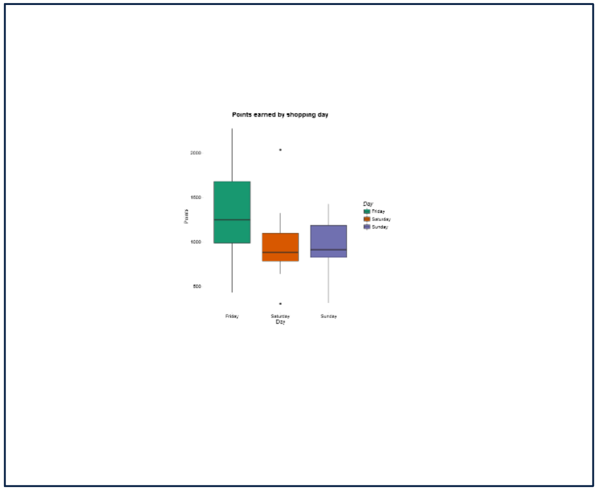
- Text draws attention automatically
  - It is processed involuntarily (cf. Logan, 1997)
  - Reading requires multiple fixations
  - Relatively small size requires multiple fixations (cf. Legge et al., 1997)

- Limitations
  - Free viewing task
  - Lack of domain experience

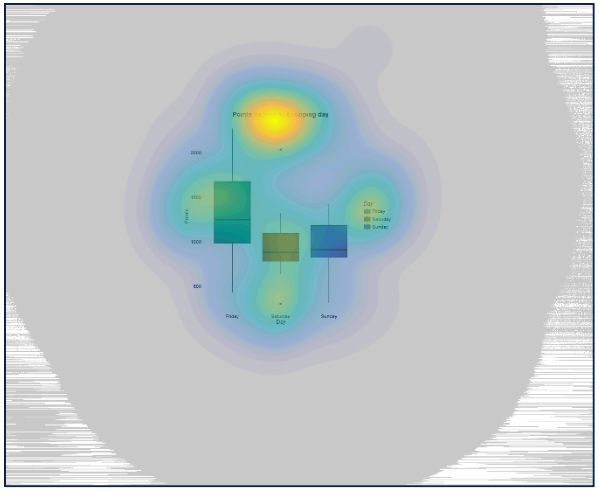
Subsequent Work

- Data Visualization Saliency Model (DVS)

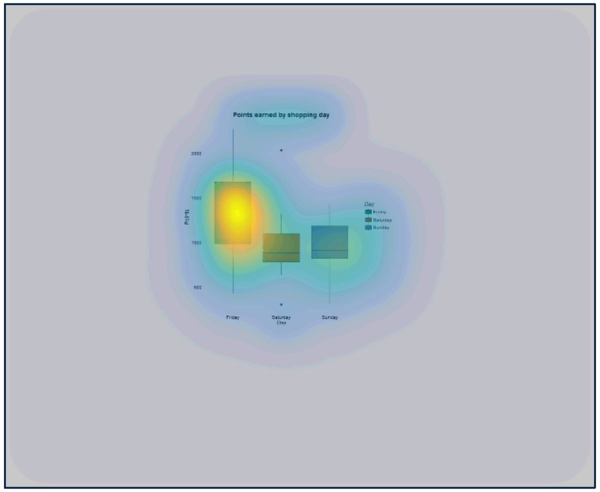
Original Vis



Fixation Map

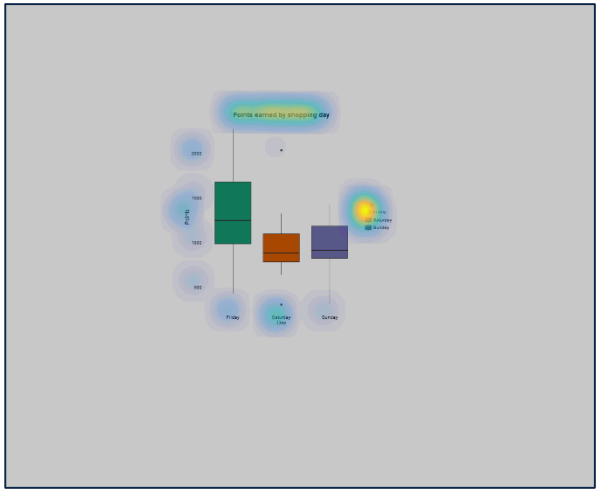


Modified Itti Saliency Map



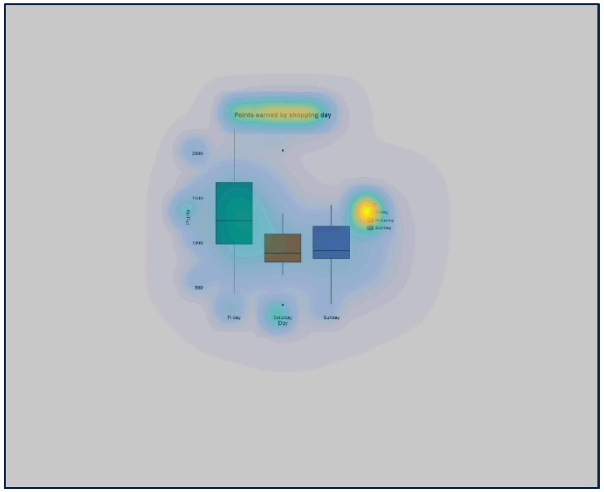
+

Text Saliency Map



=

Data Visualization Saliency Map



Thank you!

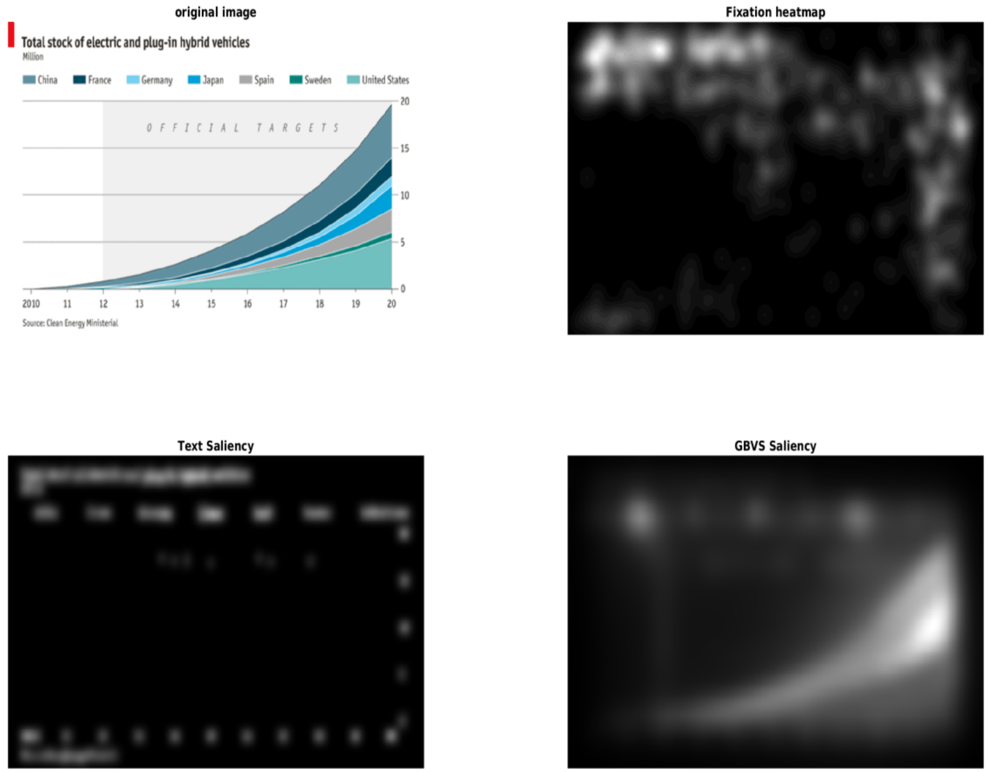
- Questions?

- Contact:
  - lematze@sandia.gov

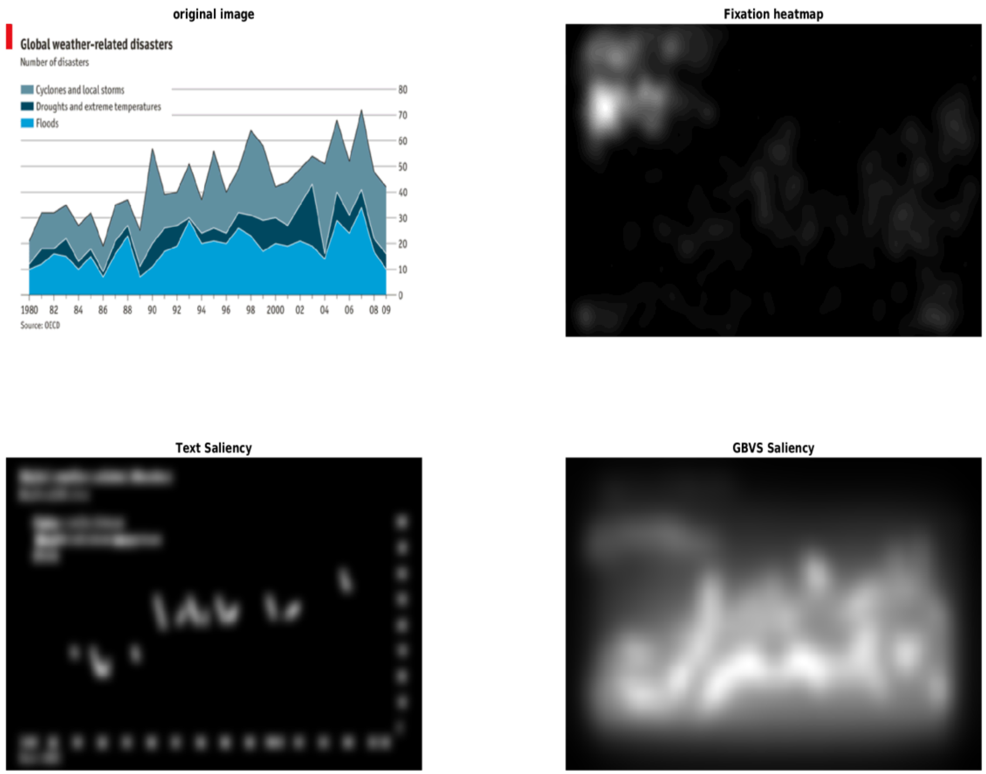
Text Detector, First Cut

hybrid of several already published text detecting algorithms

- Maximally Stable Extremal Regions (MSER), proposed by Matas et al.
- Filter according to basic criteria like aspect ratio and stroke width variation
- Use these regions as a mask to filter the image, then extract edges
- Compute two text-specific features at different aspect ratios and spatial scales
    - proposed by Lu, S., Chen, T., Tian, S., Lim, J. H., & Tan, C. L. (2015).
- Combine and average to create text saliency map

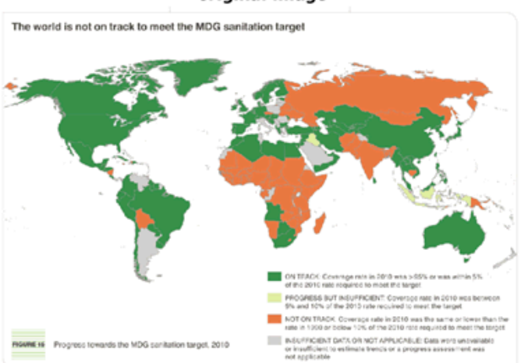


Text Detector: Some False Positives…Stay Tuned



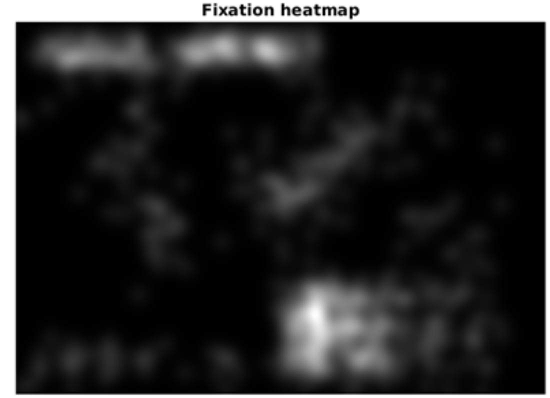
- Modifications to Itti & Koch model to create Vis Saliency Model (VSM):
  - Implemented text detector
    - Match to fixation maps greatly improved!
  - Color space
    - RGB converted to CIE LAB, a color map more aligned with human color perception

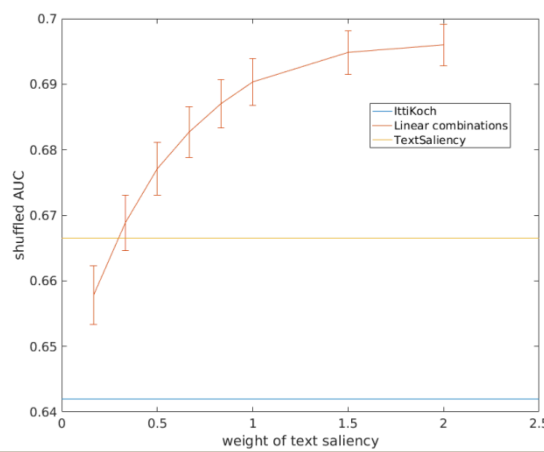
original image



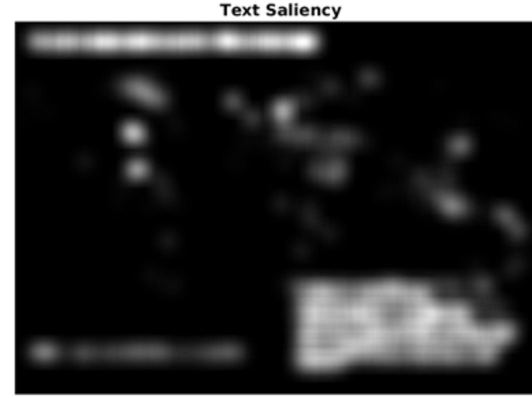
GBVS Saliency

Fixation heatmap





Text Saliency



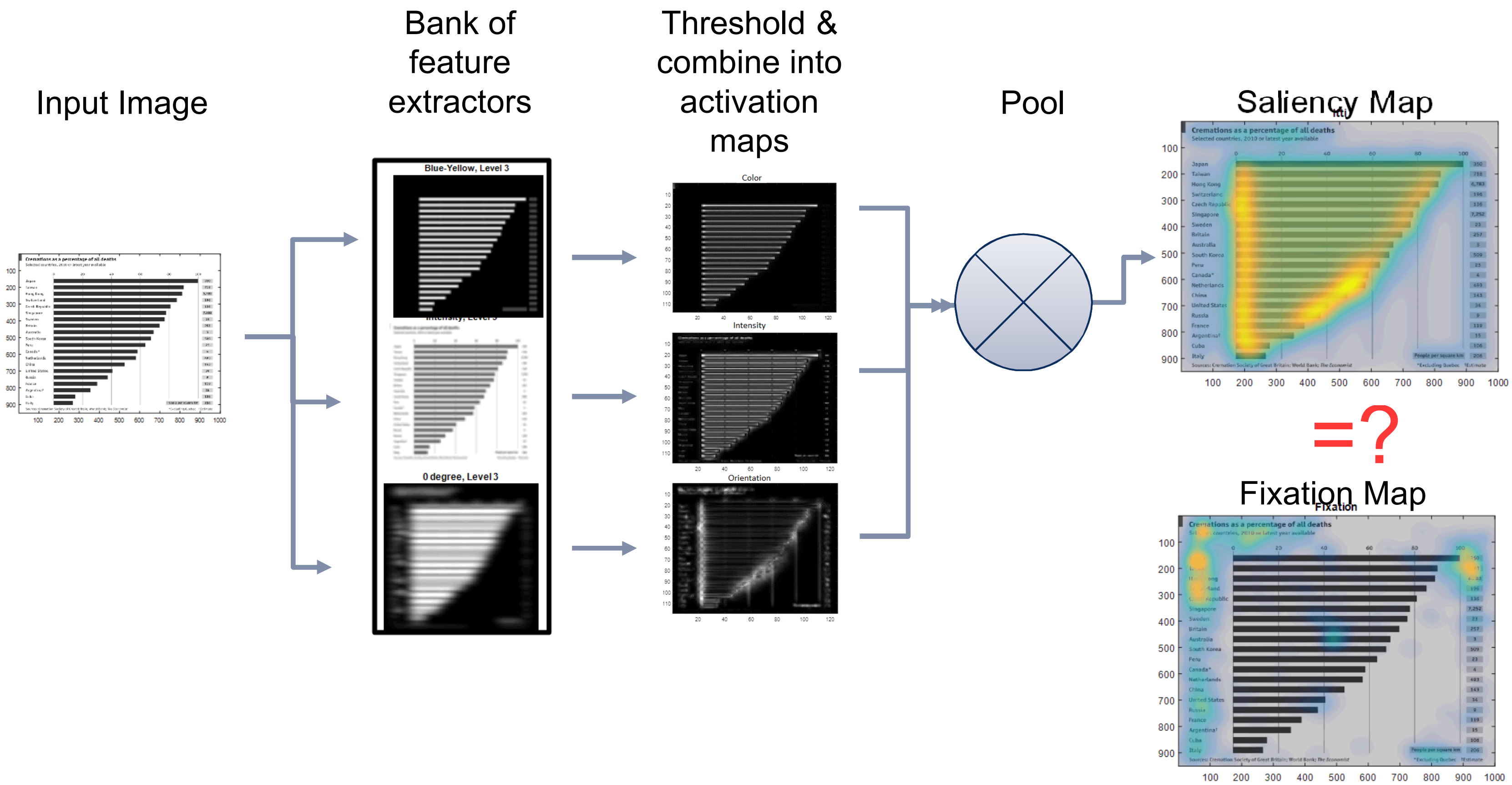
Conclusion

- *Our goal is to create a tool that will give designers a useful evaluation metric for their visualizations*

- We're developing principled methods and measures to
  - Be sure that new techniques are really better than old ones
  - Know the strengths and weaknesses of each tool
    - When to use which tool

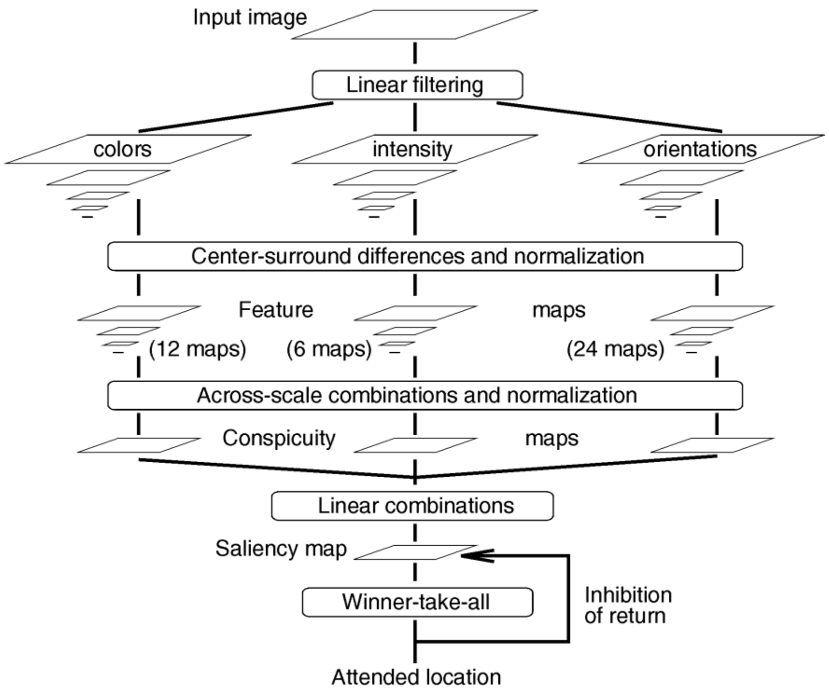
- We have focused on visual saliency because it is a general metric that can be applied to any type of image from any domain
- If a designer has a sense of what information is most important from a top-down perspective, she can then assess the visual salience to determine whether or not the most important features are also highly salient from a bottom-up perspective

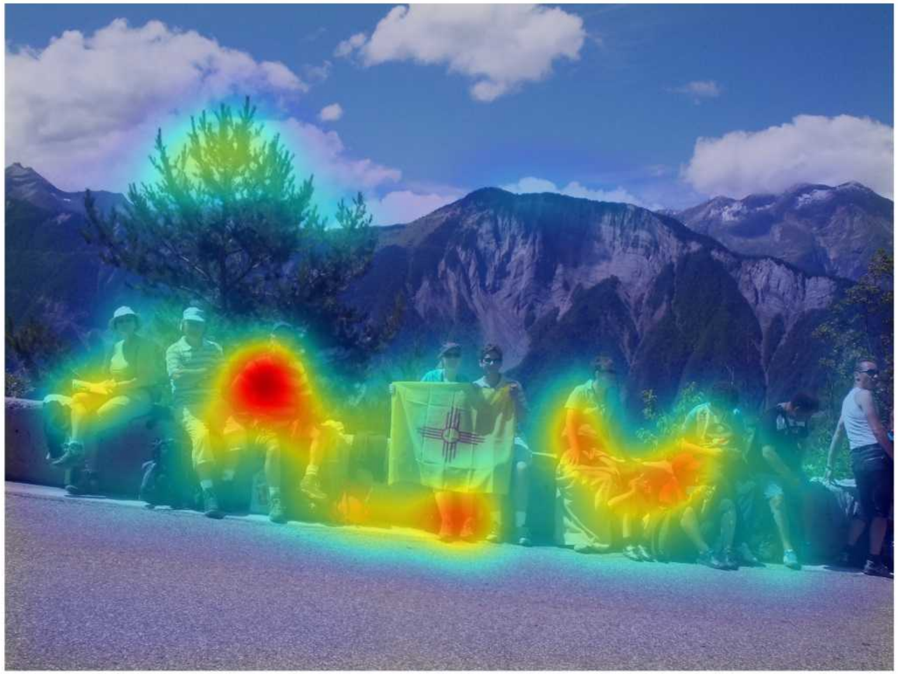
Modeling Visual Saliency: Establishing Baseline of Performance



Existing Saliency Models for Natural Scenes

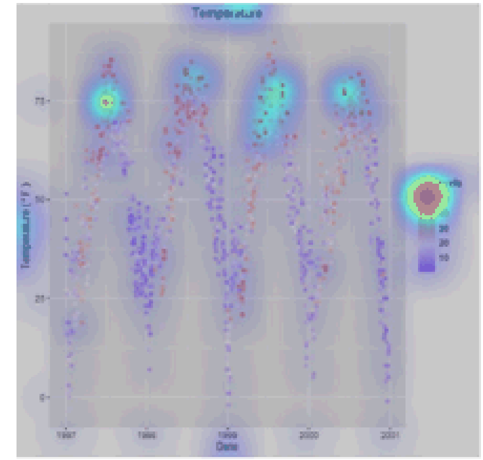
- Models based on neural architecture can predict where people will look in natural scenes
  - Would be a useful tool for assessing vis designs – do they draw bottom-up attention as the designer intends?



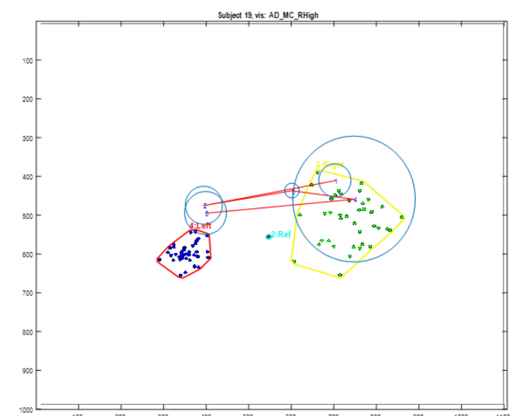


Project Goals

- **Develop models for assessing the bottom-up visual saliency of data visualizations.**
  - Ideally, a vis will draw the viewers' attention to the most important information for their task



- **Conduct experiments to characterize common top-down sensemaking strategies employed by users of visualizations.**
  - Studies using eye tracking to investigate how analysts navigate through abstract information
  - Expansion of the "Value of Vis" framework (Stasko, Georgia Tech)



- Intro
- ~~Existing saliency models~~
  - ~~Why they don't work~~
- DVS model
  - ~~Modified Itti~~
  - ~~Text Saliency map~~
  - ~~Weighted combo~~
  - ~~Comparison to other models on MASSVIS~~
- Testing model performance on a new data set
  - MASSVIS, new vis, fractals, line drawings
- How to apply the model
- Discussion
- Limitations
- Future Directions