

Where do you look on these images?



The squares shows where 15 observers looked
in eye tracking experiments

Understanding and predicting where people look in images

Tilke Judd

with advisors Frédo Durand and Antonio Torralba
committee Aude Oliva and Bill Freeman











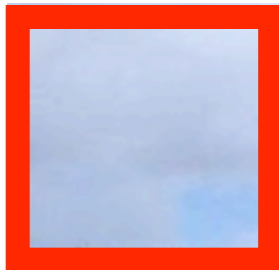








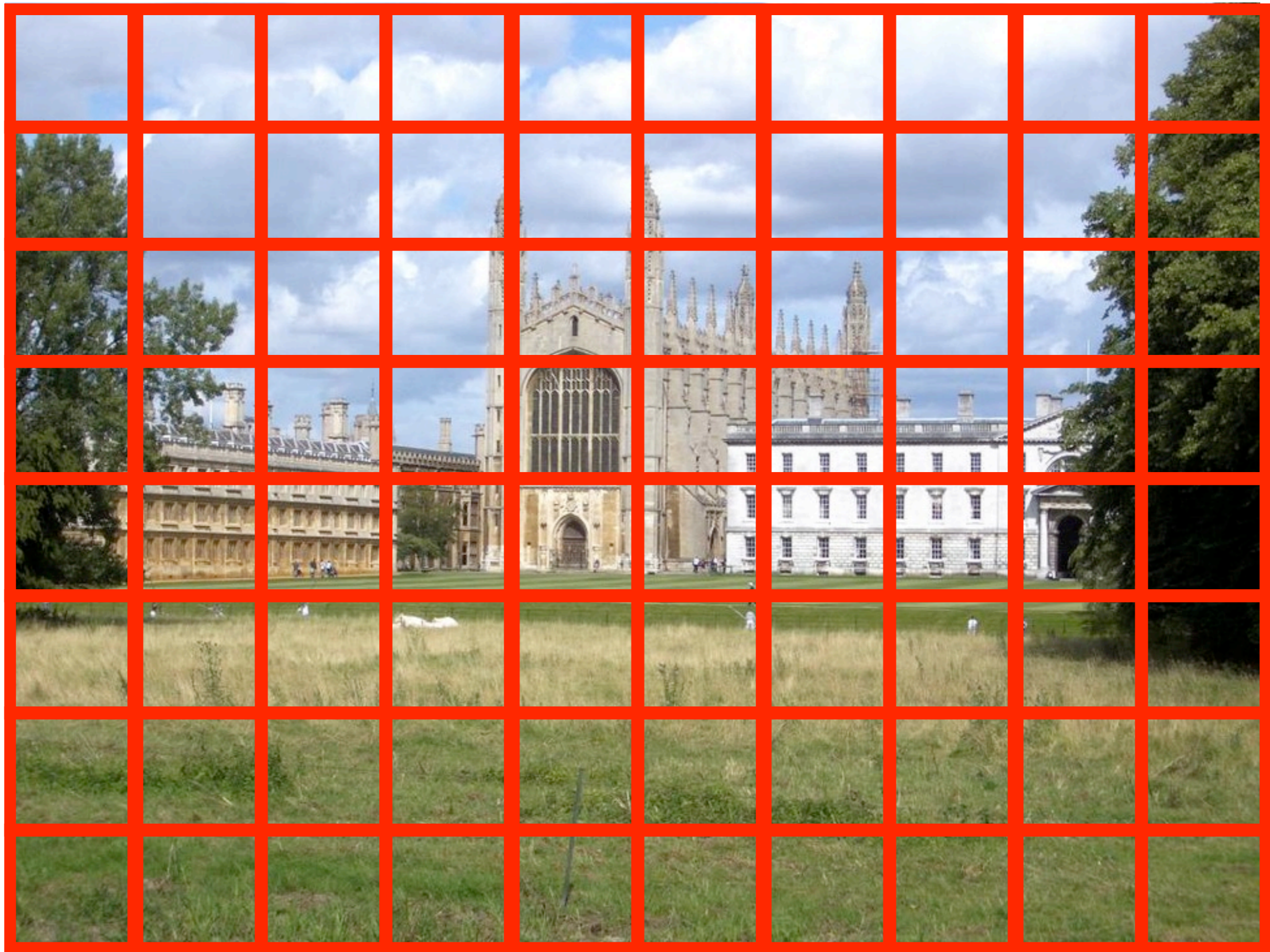




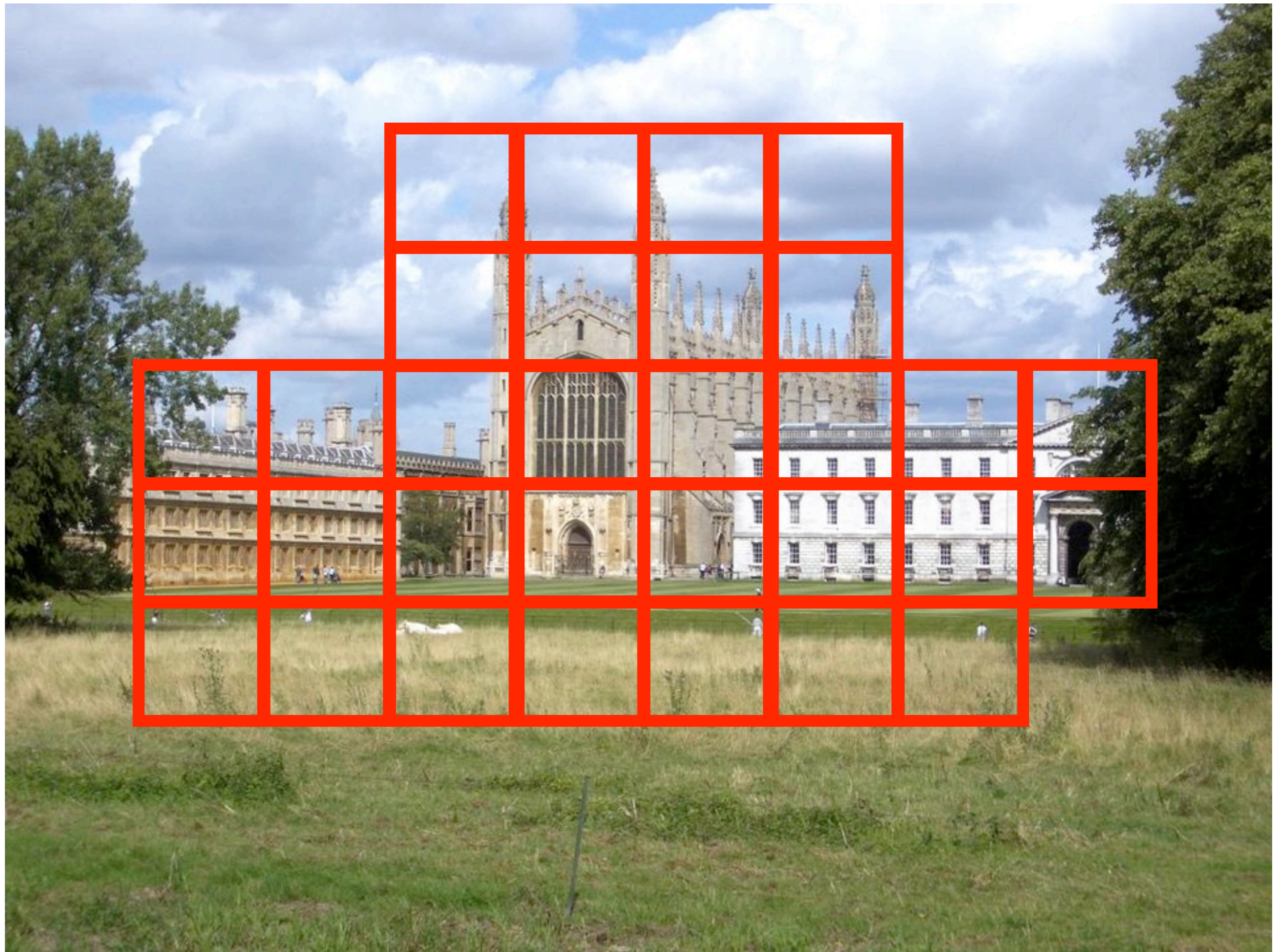












What is common to both these situations?

need to *prioritize the visual information*
and *decide what is most important*

These are applications of research we do in

Saliency and Visual Attention

Understanding and predicting where people look in images

Tilke Judd

with advisors Frédo Durand and Antonio Torralba
and collaborators Krista Ehinger and Aude Oliva

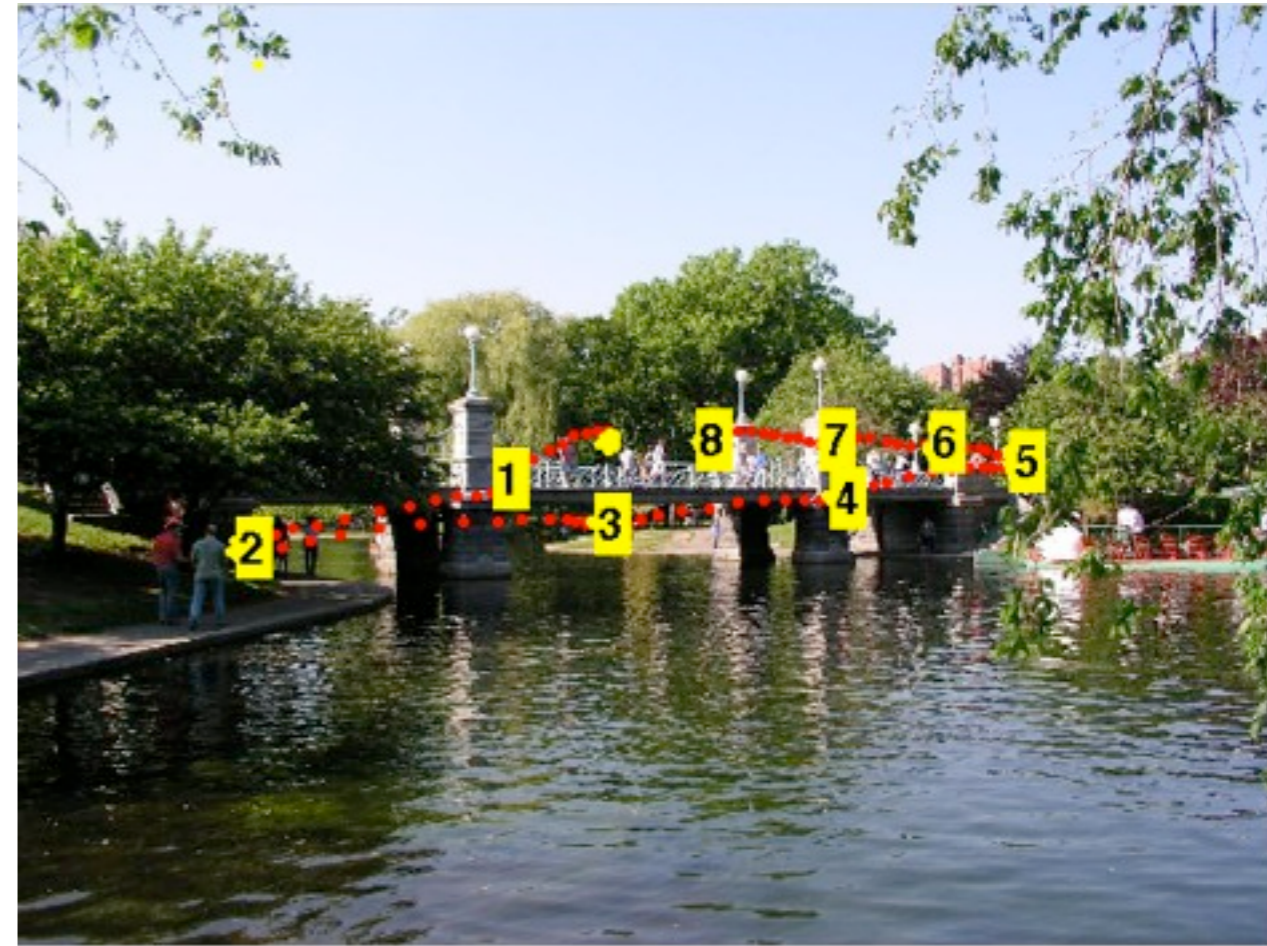
Understanding attention enables applications in computer graphics & vision, design

- image cropping / thumbnailing
- image and video compression
- non photorealistic rendering
- scene understanding
- advertising and package design
- web usability
- localization / recognition
- object detection
- navigational assistance
- robot active vision
- surveillance systems
- assistive technology for blind or low-vision people

Human visual system has developed *selective attention* through evolution



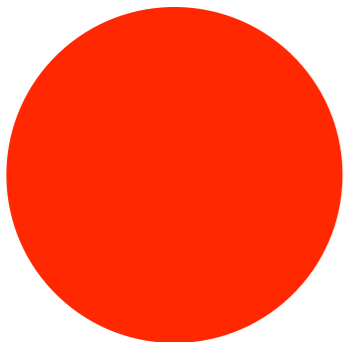
Anstis 1998



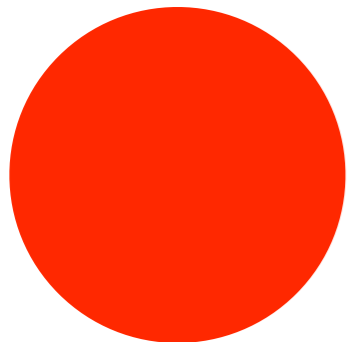
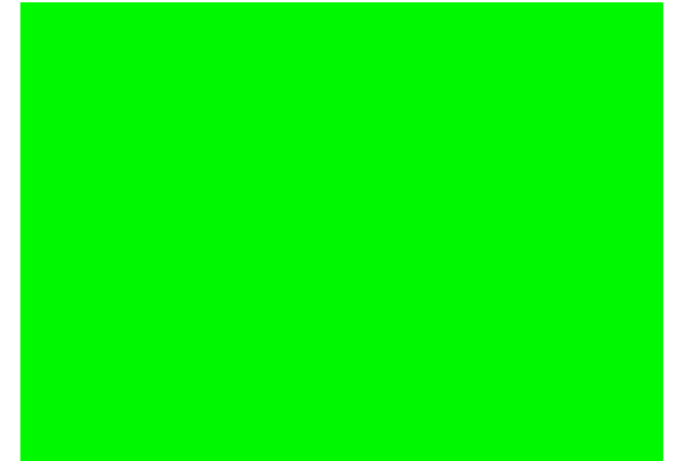
Fixations and **saccades** of an example scanpath

Where we move our eyes is dictated
by two mechanisms

Where we move our eyes is dictated
by two mechanisms

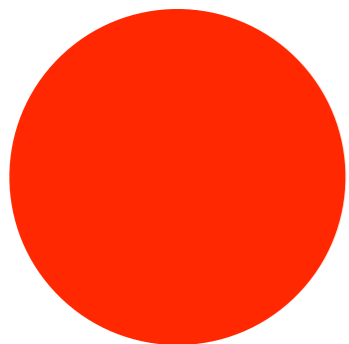
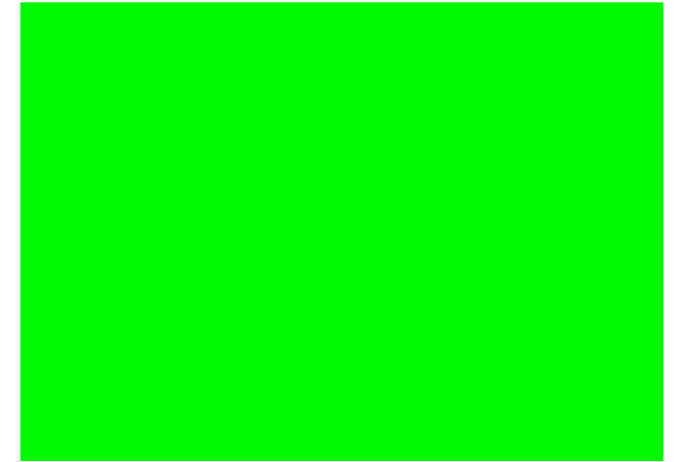


Where we move our eyes is dictated
by two mechanisms



Where we move our eyes is dictated by two mechanisms

- Bottom-up mechanisms



Where we move our eyes is dictated by two mechanisms

- Bottom-up mechanisms
- Top-down mechanisms

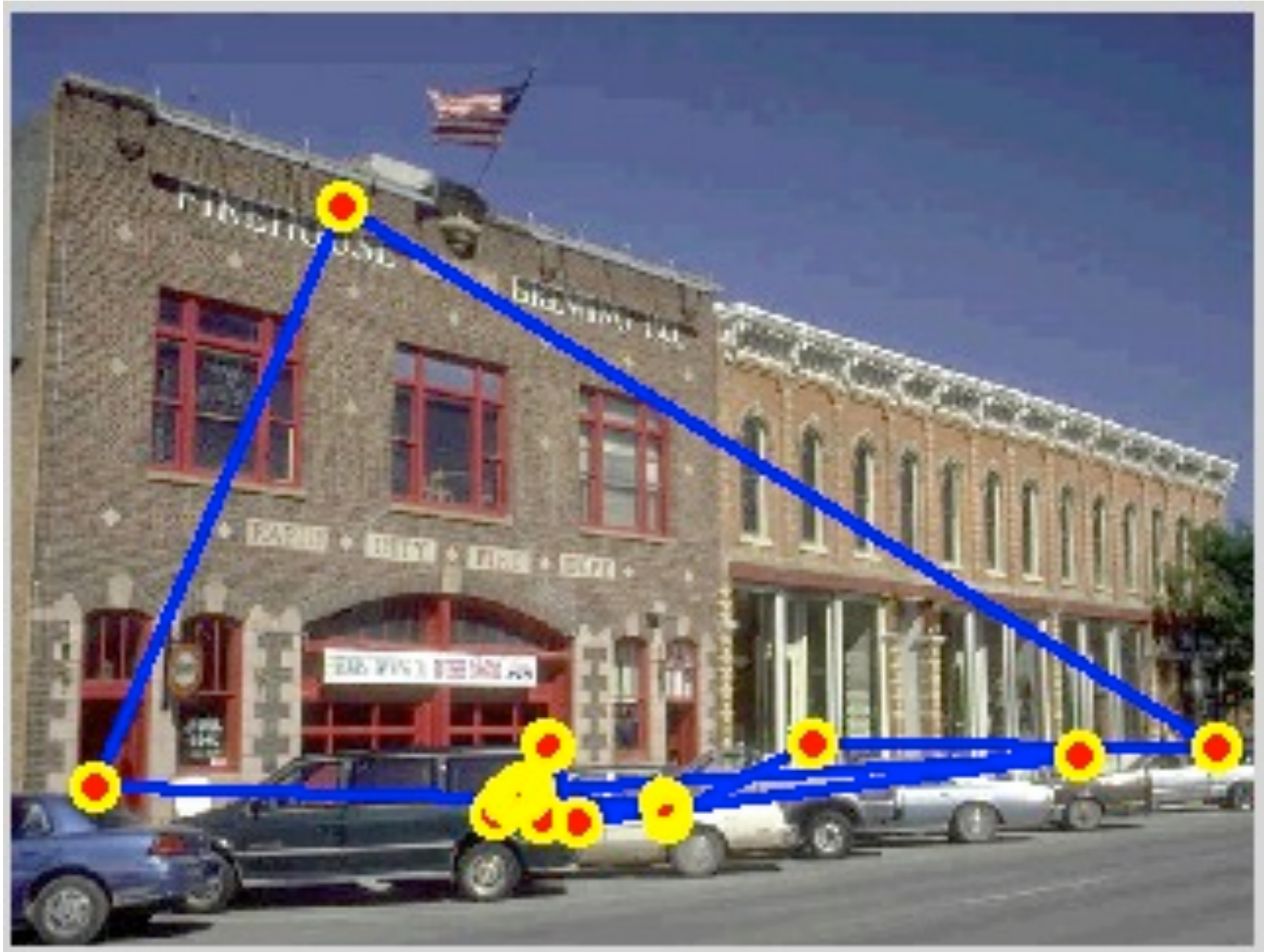
Find the pedestrian



There is no pedestrian



you likely looked here



this is where someone else looked in an experiment



we look here because of our top-down semantic understanding of the scene: humans are on the ground

Where we move our eyes is dictated by two mechanisms

- Bottom-up mechanisms
- Top-down mechanisms
 - semantic understanding
 - memories, state
 - task

Researchers create *computational models of visual attention* to predict where people look



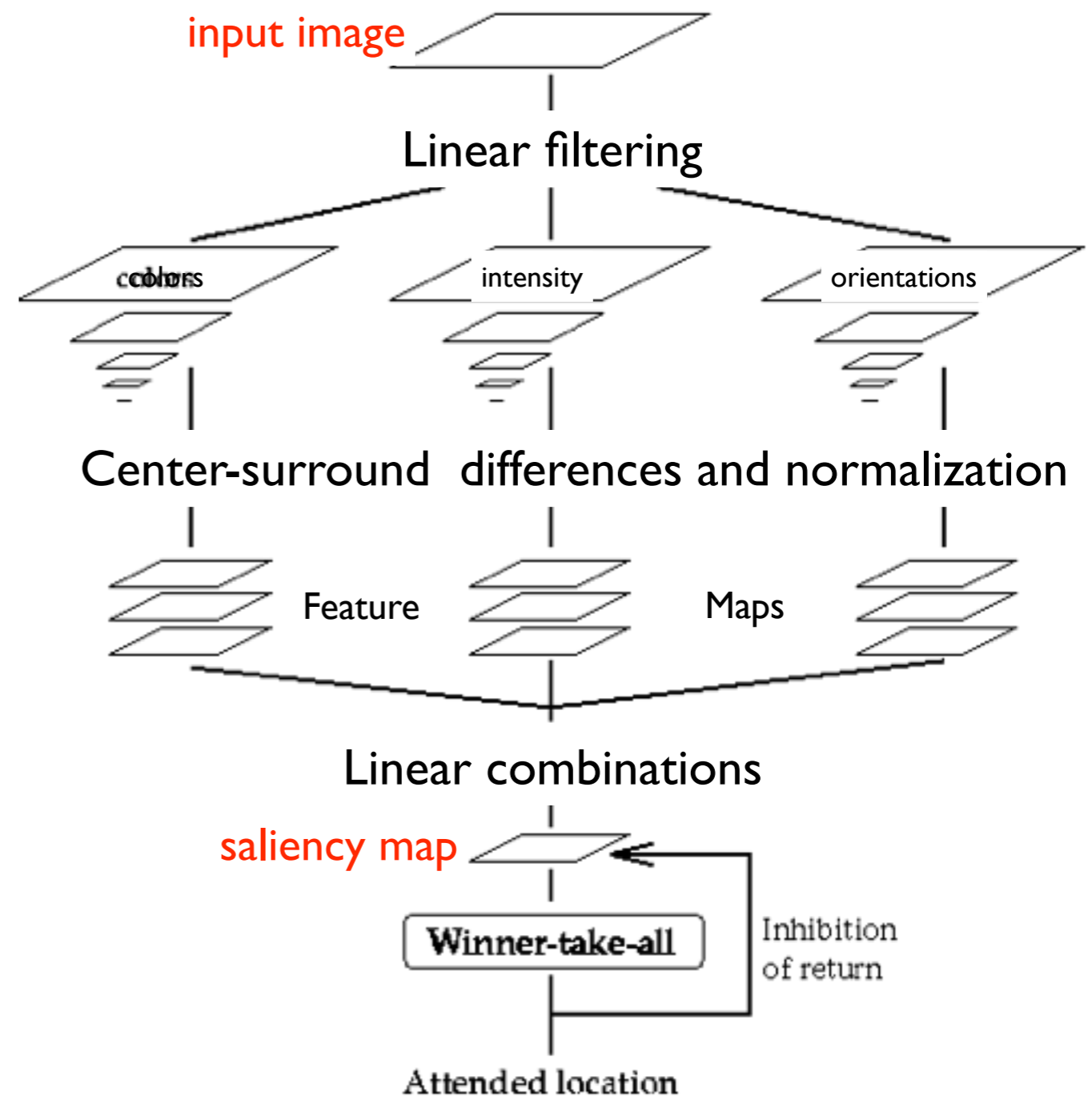
Image



Saliency Map

Common models of saliency based on bottom up features

- based on biologically plausible filters
- mimic the human visual system
- measure intensity, illumination, contrast
- several parameters need tuning



Itti and Koch model

People do not always look where bottom up models predict



Many models of saliency have been introduced

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Hou and Zhang (2007)

Rosenholtz (1999)

Itti and Baldi (2006)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)
Graphical Model

Avraham and Lindenbaum (2009)
Esaliency

Bruce and Tsotsos (2009)
Information theoretic approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

Itti and Baldi (2006)
“Surprise” model

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Torralba et al. (2006)

Zhang et al. (2008)

Kanan et al. (2009)

Which one is the best?

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Hou and Zhang (2007)

Rosenholtz (1999)

Itti and Baldi (2006)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)
Graphical Model

Avraham and Lindenbaum (2009)
Esaliency

Bruce and Tsotsos (2009)
Information theoretic approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

Itti and Baldi (2006)
“Surprise” model

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Torralba et al. (2006)

Zhang et al. (2008)

Kanan et al. (2009)

Which one is the best?

Biologically Inspired

Itti and Koch (1998)

Cerf et al. (2007)

Zhang (2007)

sen et al. (2000)

Itti and Baldi (2006)

Le Meur et al. (2006)

Seo & Milanfar (2009)

Zhang & Cottrell (2008)
SUN model

Goferman et al. (2009)

Achanta (2010)

Mathematically Inspired

Heral et al. (2007)

Graphical Model

Avraham and Lindenbaum (2009)

Esaliency

uce and fso et al. (2009)

at in the re-approach

Kienzle et al., (2007)

Gao and Vasconulos (2005)

ia and Bal (2006)

“Surp” mod

Navalpakkam and Itti (2005)

Elazary and Itti (2010)

Add top-down features

Ehinger et al., (2009)
(search task)

Oliva et al. (2003)

Talbot et al. (2000)

ha et al. (2008)

Kanan et al. (2009)

Benchmark needed

Issues with the state of the art:

Issues with the state of the art:

Models have too many parameters
top-down information not well integrated

Issues with the state of the art:

Models have too many parameters
top-down information not well integrated

Too many models, no good comparison

Issues with the state of the art:

Models have too many parameters
top-down information not well integrated

Too many models, no good comparison

We do not understand human visual
attention under different styles of
images

This thesis offers the following contributions

Models have too many parameters
top-down information not well integrated

model using machine learning

Too many models, no good comparison

benchmark

We do not understand human visual
attention under different styles of
images

study of fixations on
low-res images

How do we do this?

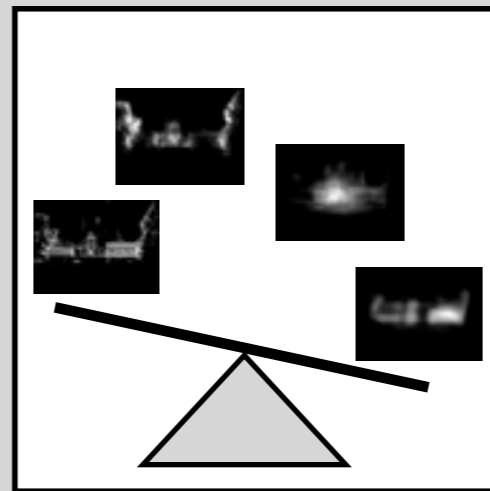
- Go back to ground truth eye tracking data
- Use data to learn a model
- Use data to evaluate success metrics for models
- Use data to compare human fixations on variations of images

Understanding and predicting where people look

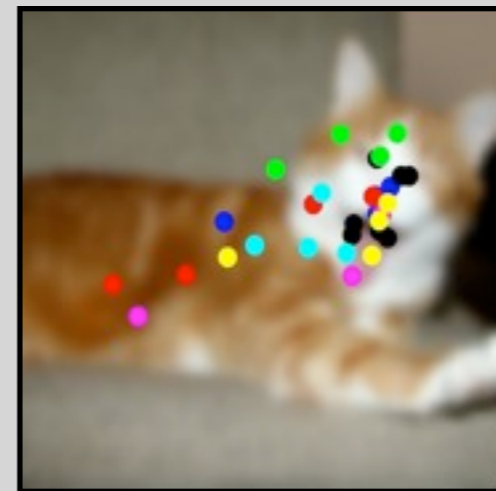
Introduction



Learning a model



Benchmarking models of saliency



Fixations on low-resolution images

Conclusion

Our goal is to learn where people look directly from eye tracking data

- Step 1: collect eye tracking data set
- Step 2: learn the model

We collected 1003 natural images



Natural images of objects and scenes downloaded from Flickr and LabelMe

We run eye tracking experiments



[Photo Credit: Jason Dorfman CSAIL website]

We run eye tracking experiments



user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

We run eye tracking experiments



eye tracker
measures location
of eye fixation
several times a
second.

user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

We run eye tracking experiments



eye tracker
measures location
of eye fixation
several times a
second.

user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

We run eye tracking experiments



user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

eye tracker
measures location
of eye fixation
several times a
second.

We run eye tracking experiments



screen resolution
1280x1024

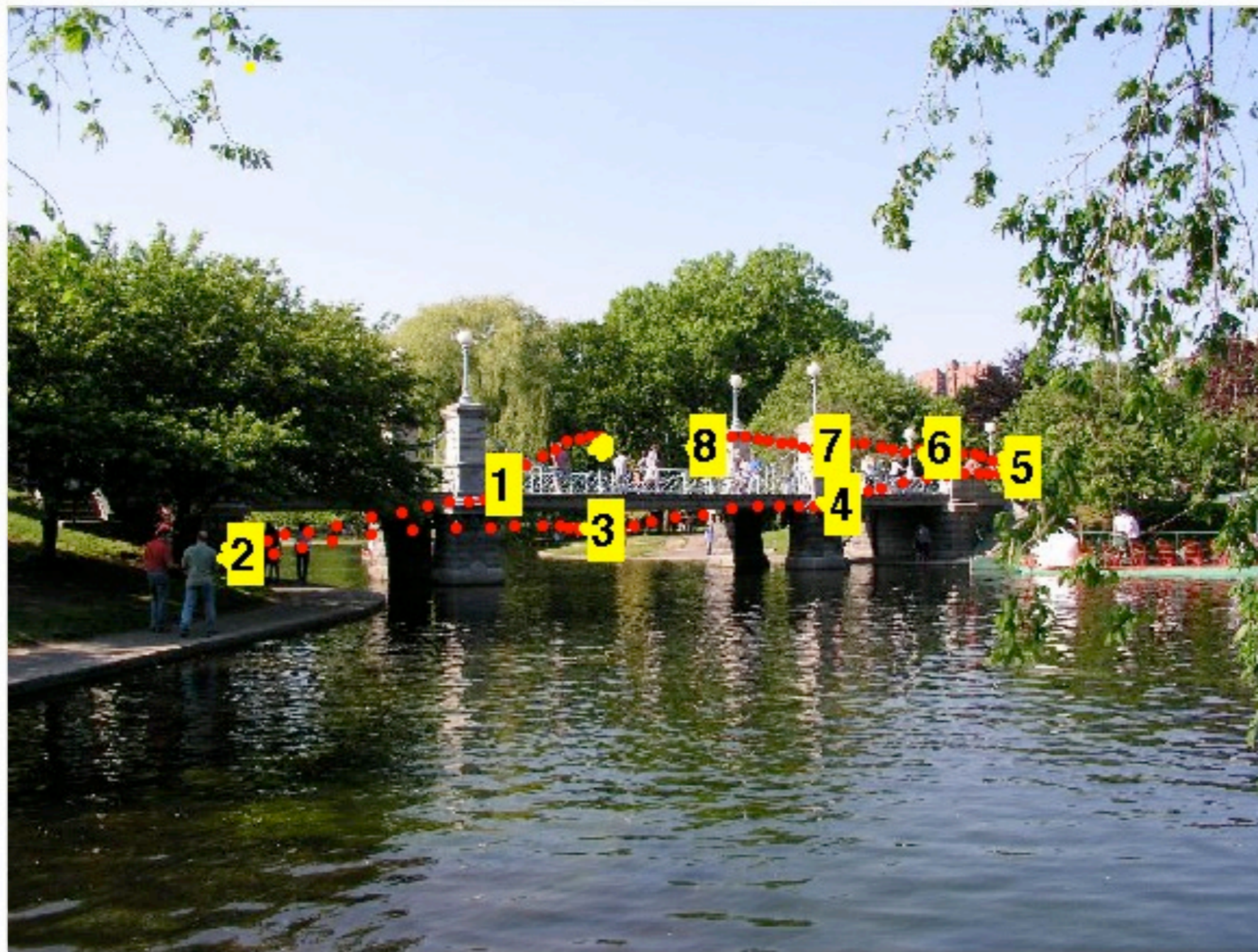
each image shown
for 3 seconds

eye tracker
measures location
of eye fixation
several times a
second.

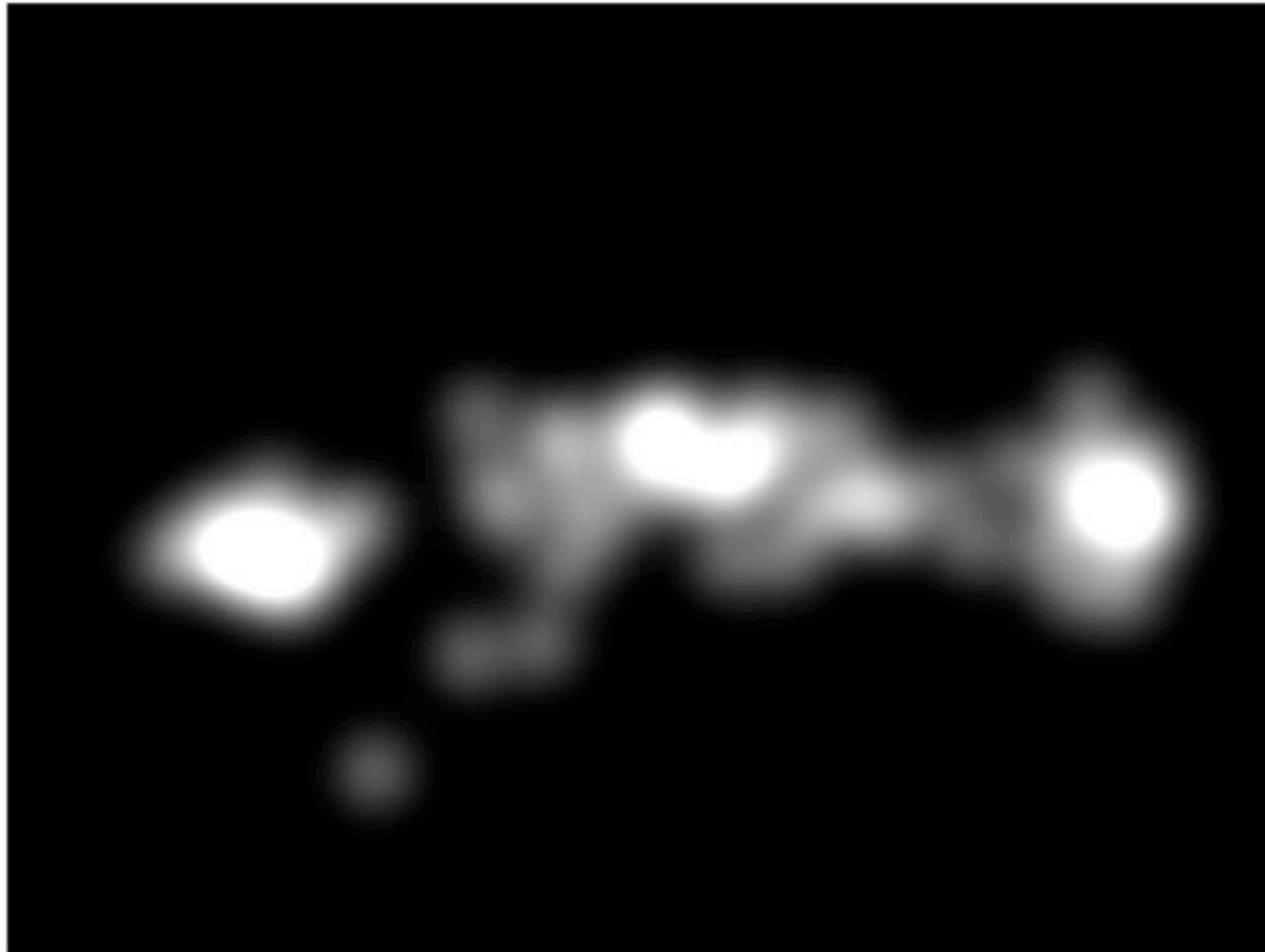
user rests head in chin rest

[Photo Credit: Jason Dorfman CSAIL website]

Fixations for one observer

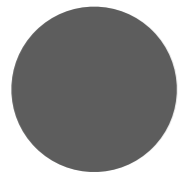


Fixation map created from gaussian convolution over fixations

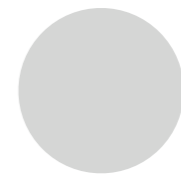


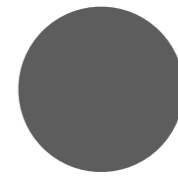
Fixation map

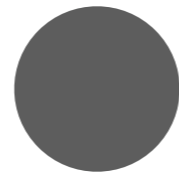
You do the experiment

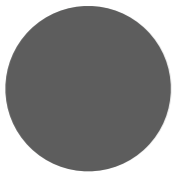
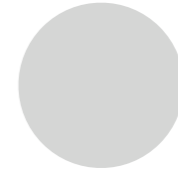


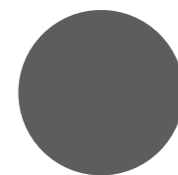
Look here
Shine pointer here











You will see a series of images
Look closely at each one











FULL HOUSE 美食苑

OPEN

FULL HOUSE 美食苑
BUFFET
ONLY €9.90
ALL U CAN EAT
Eat as much as u can





You have completed the experiment









FULL HOUSE 美食苑

OPEN

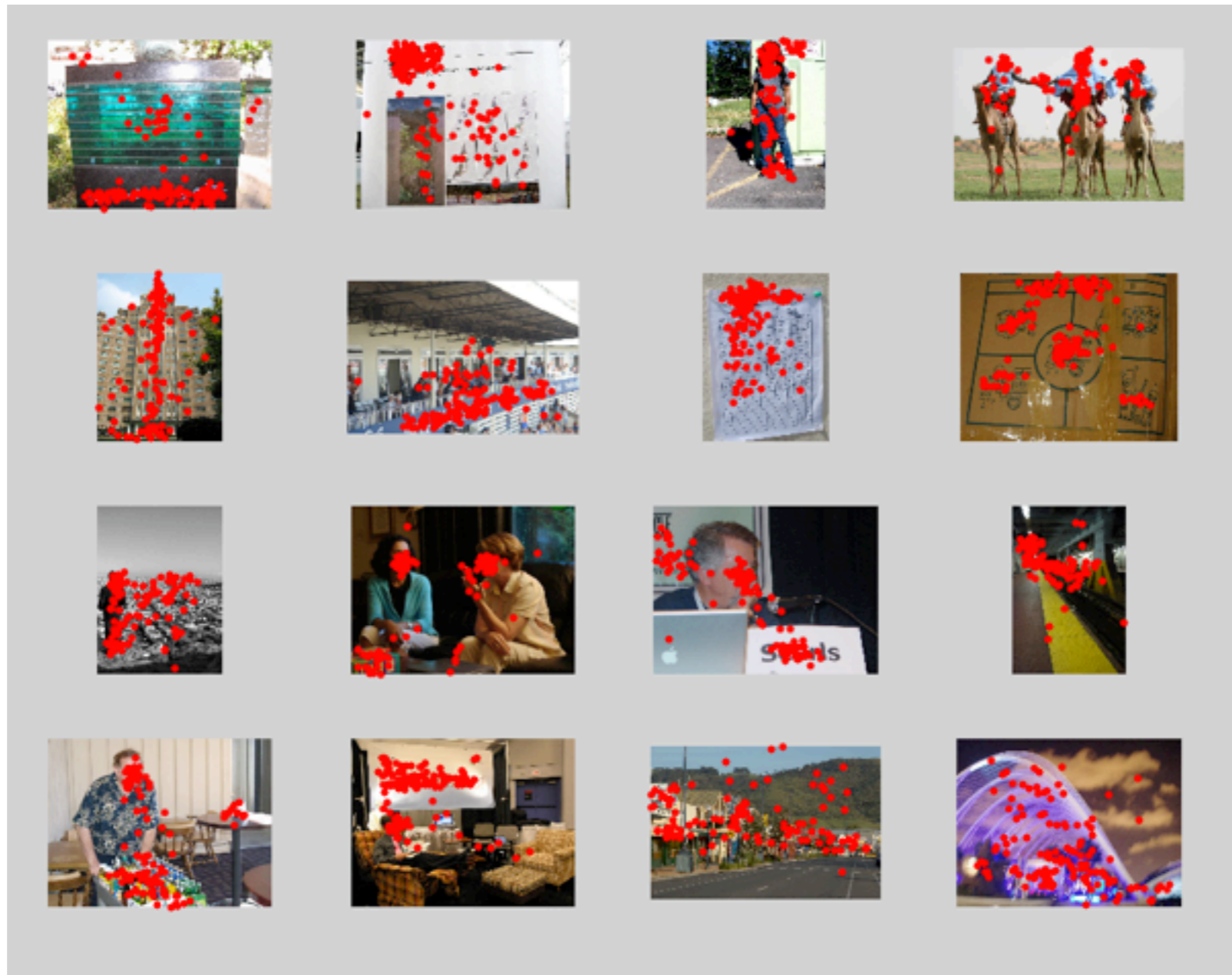
FULL HOUSE 美食苑
BUFFET
ONLY €9.90
ALL U CAN EAT
Eat as much as u can





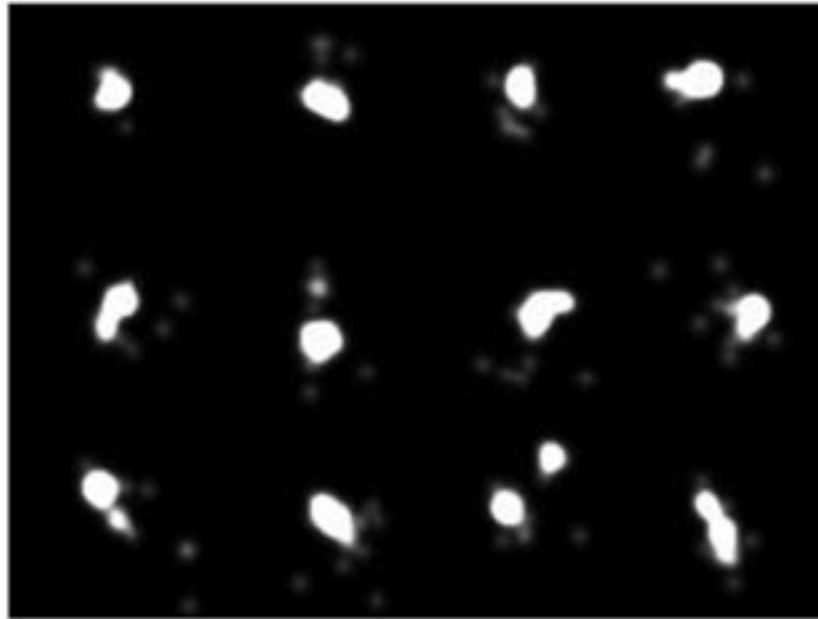


We have data from 15 observers on 1003 images

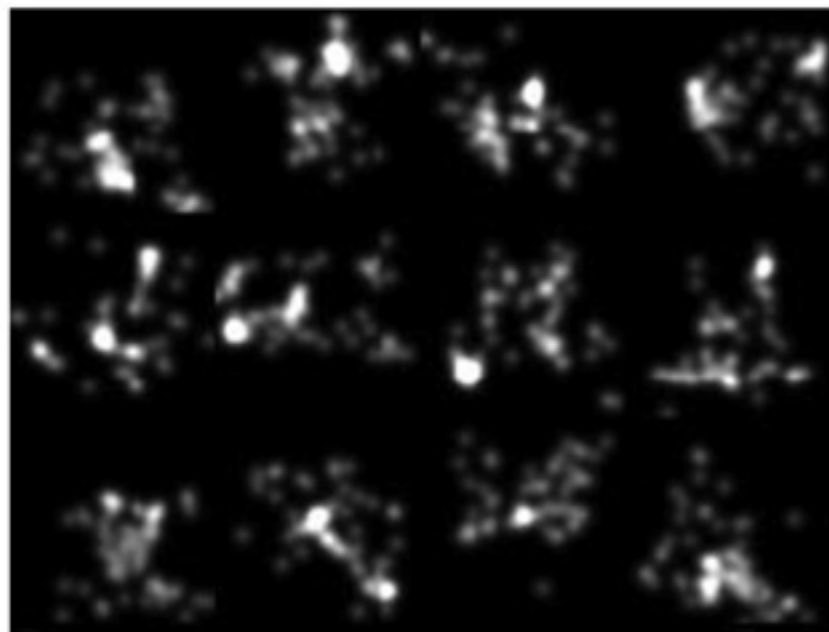


[MIT 2009 data set]

Fixation consistency depends on the image content



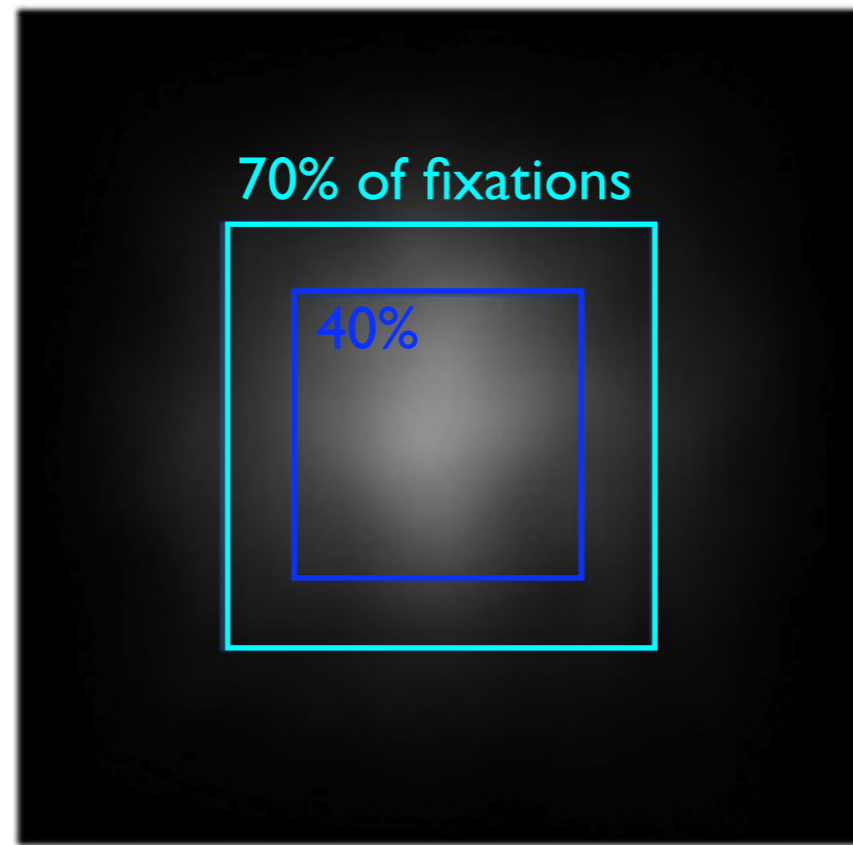
Low entropy saliency maps



High entropy saliency maps



Average human fixations are biased towards the center

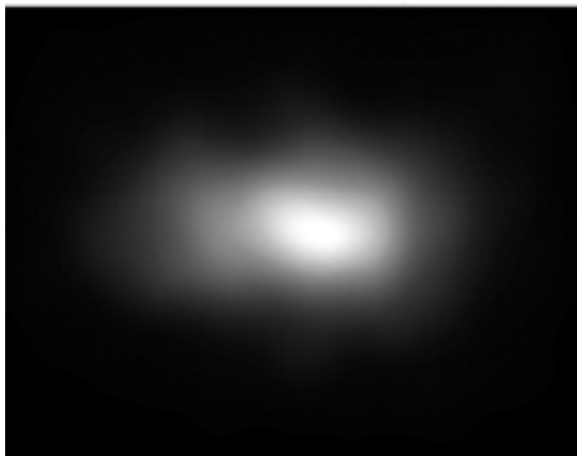


Avg of all saliency maps

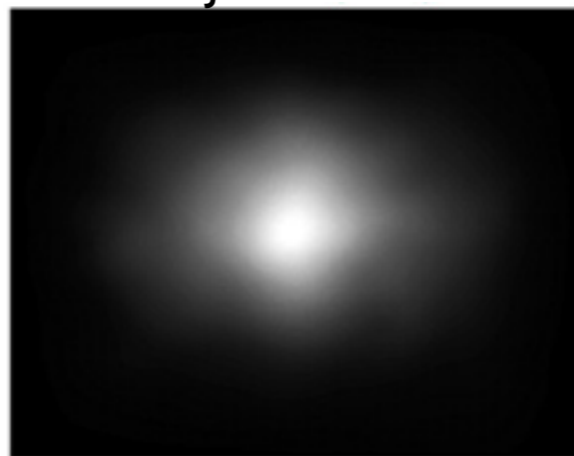
Why are fixations center biased?

- photographer bias
- viewing strategy

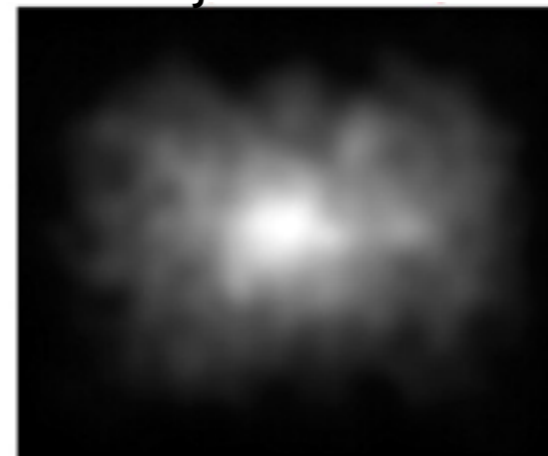
Bruce and Tsotsos 2005



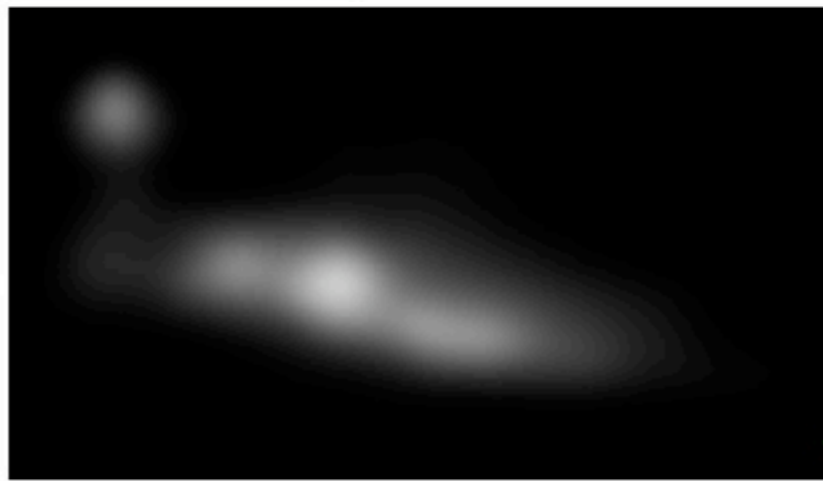
Judd 2009



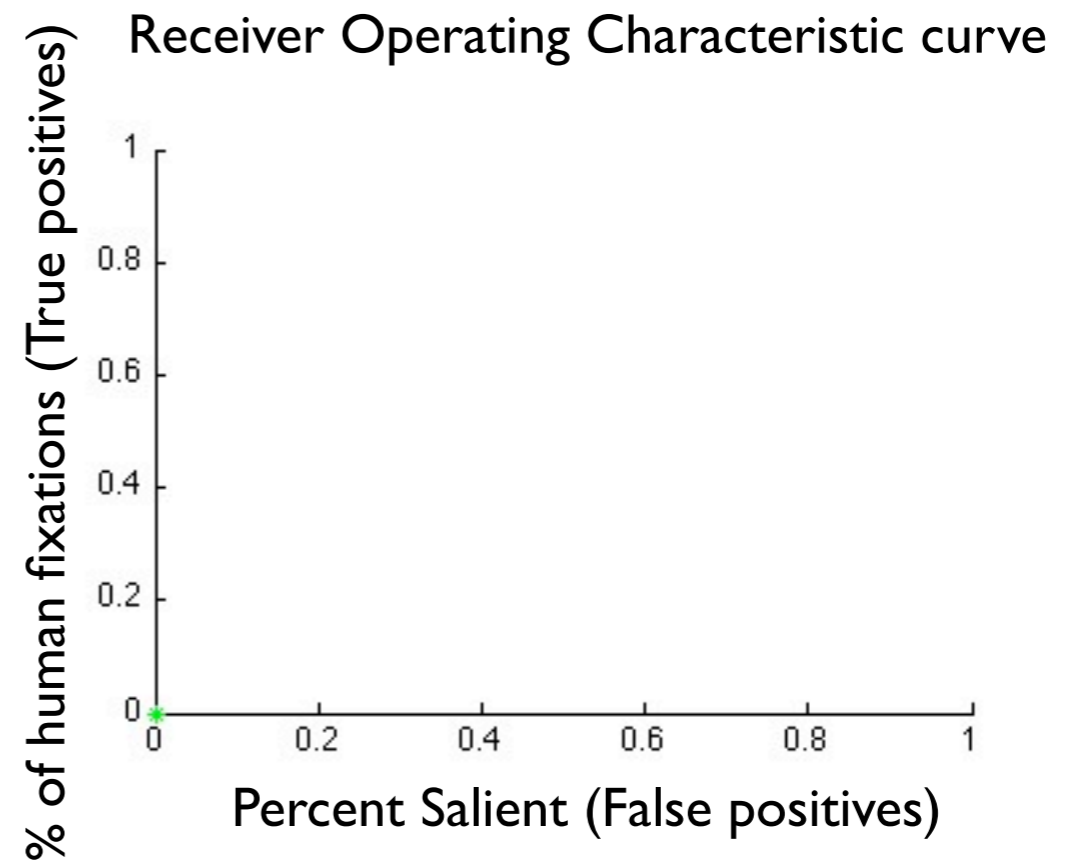
Judd 2011



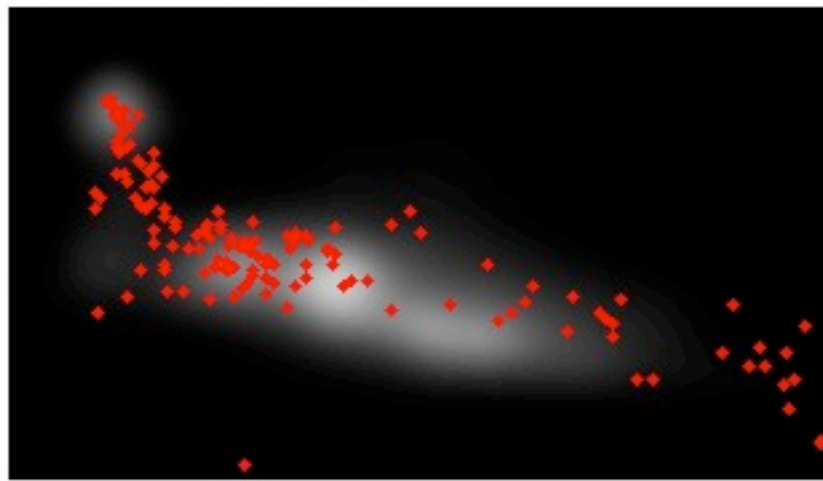
We use a Receiver Operating Characteristic (ROC) curve to measure performance



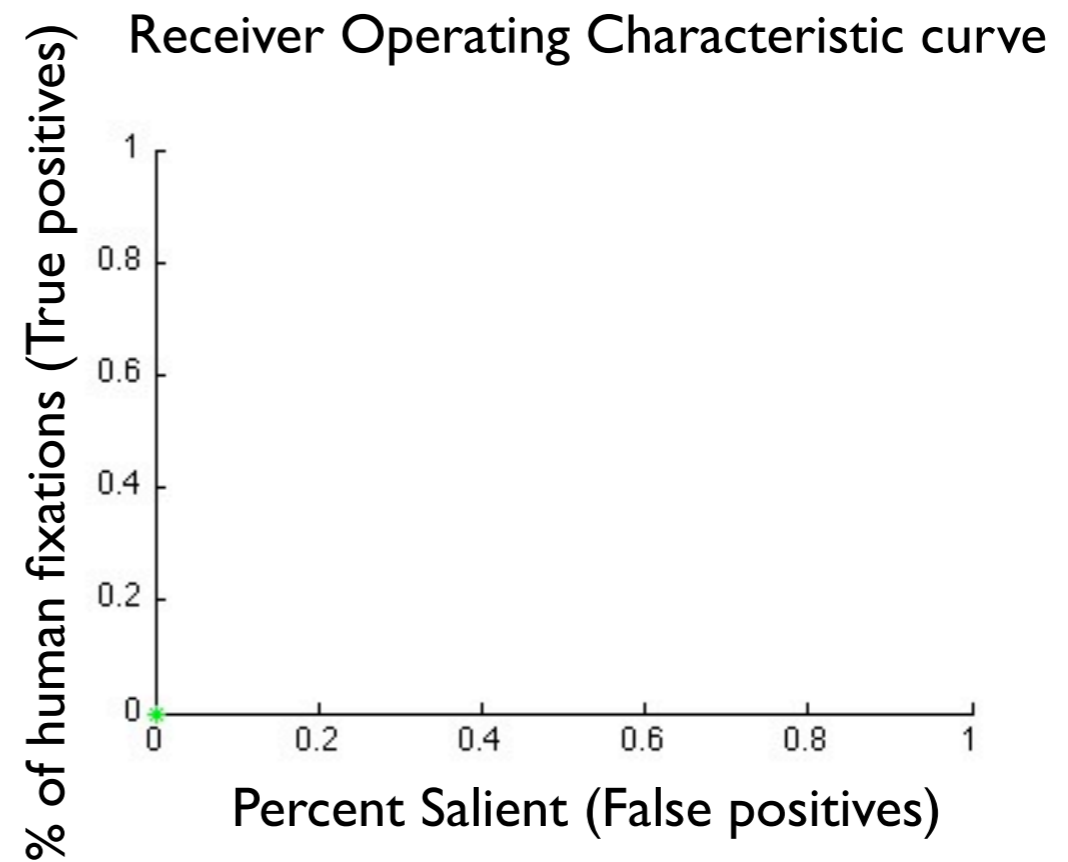
Saliency Map



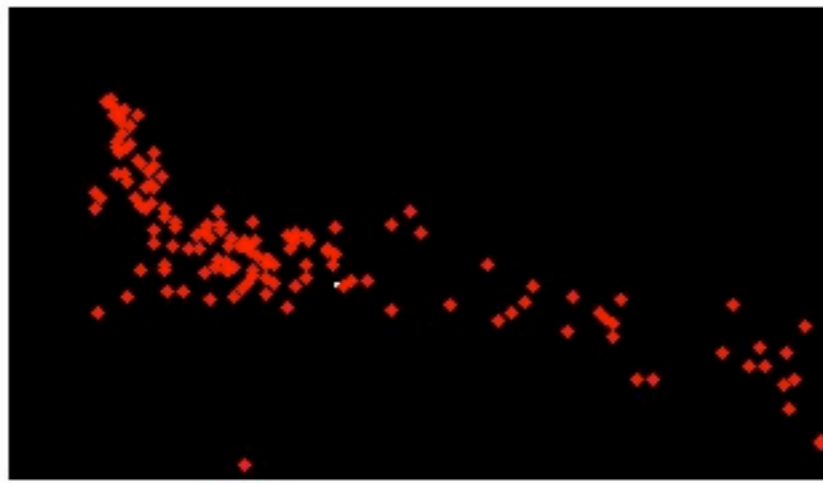
We use an ROC curve to measure the performance of a saliency map



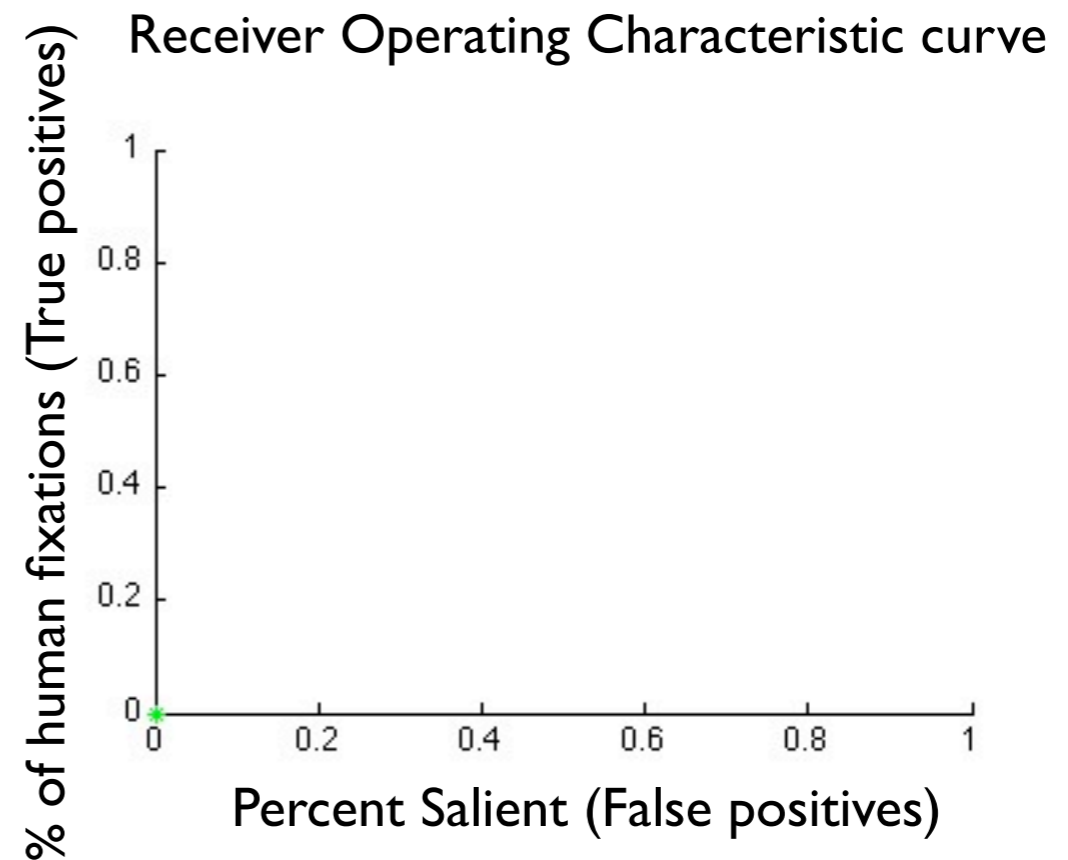
Human Fixations



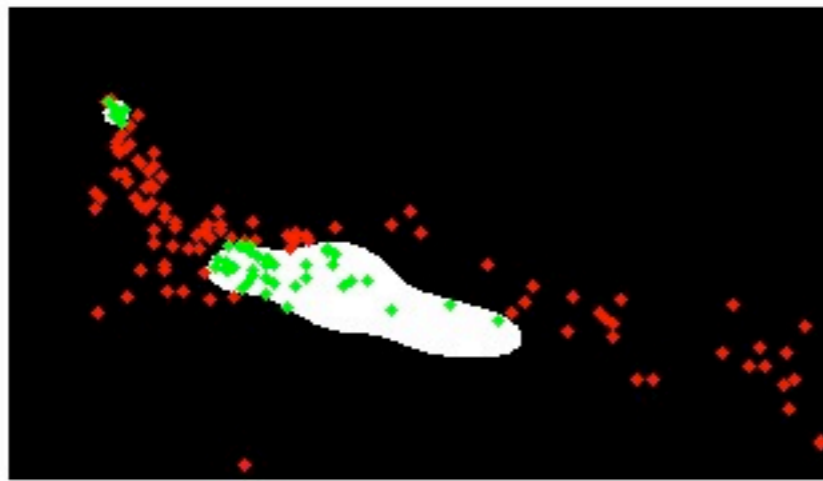
We use an ROC curve to measure the performance of a saliency map



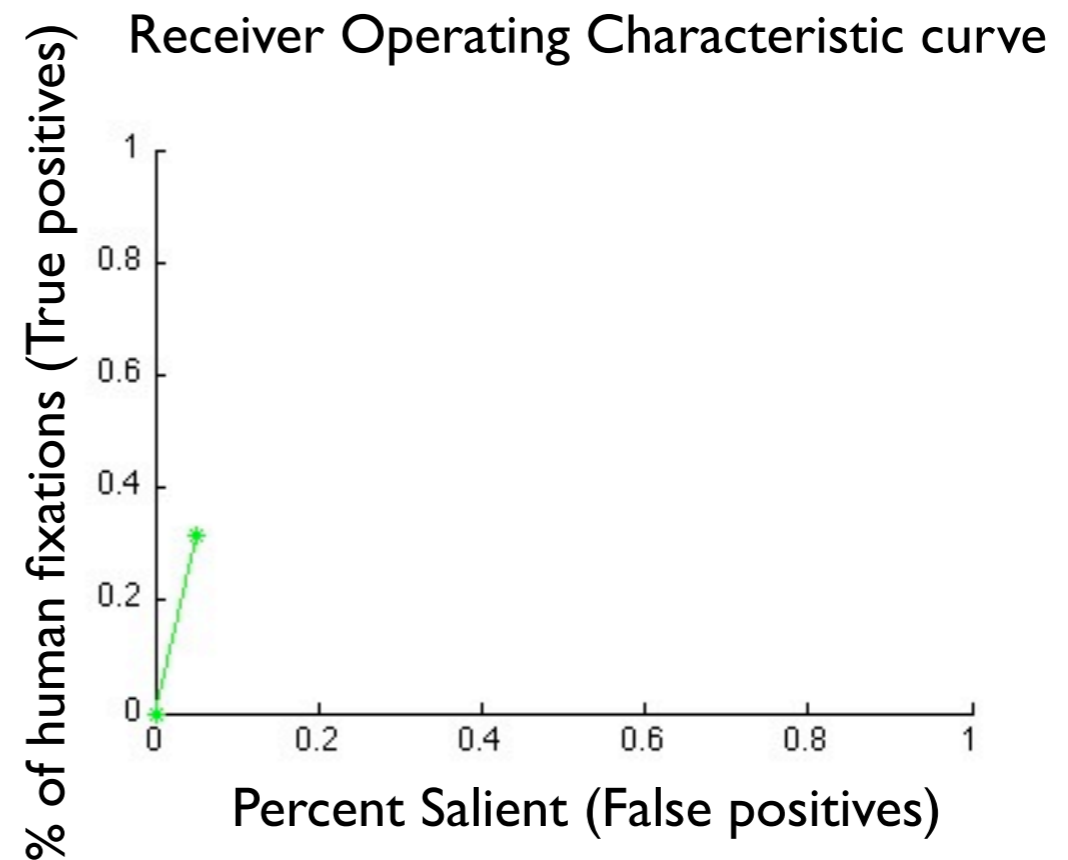
Thresholded Saliency Map



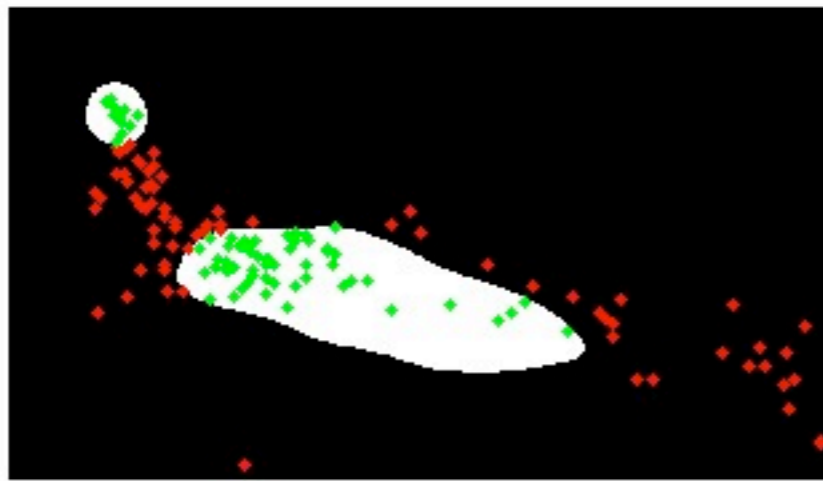
We use an ROC curve to measure the performance of a saliency map



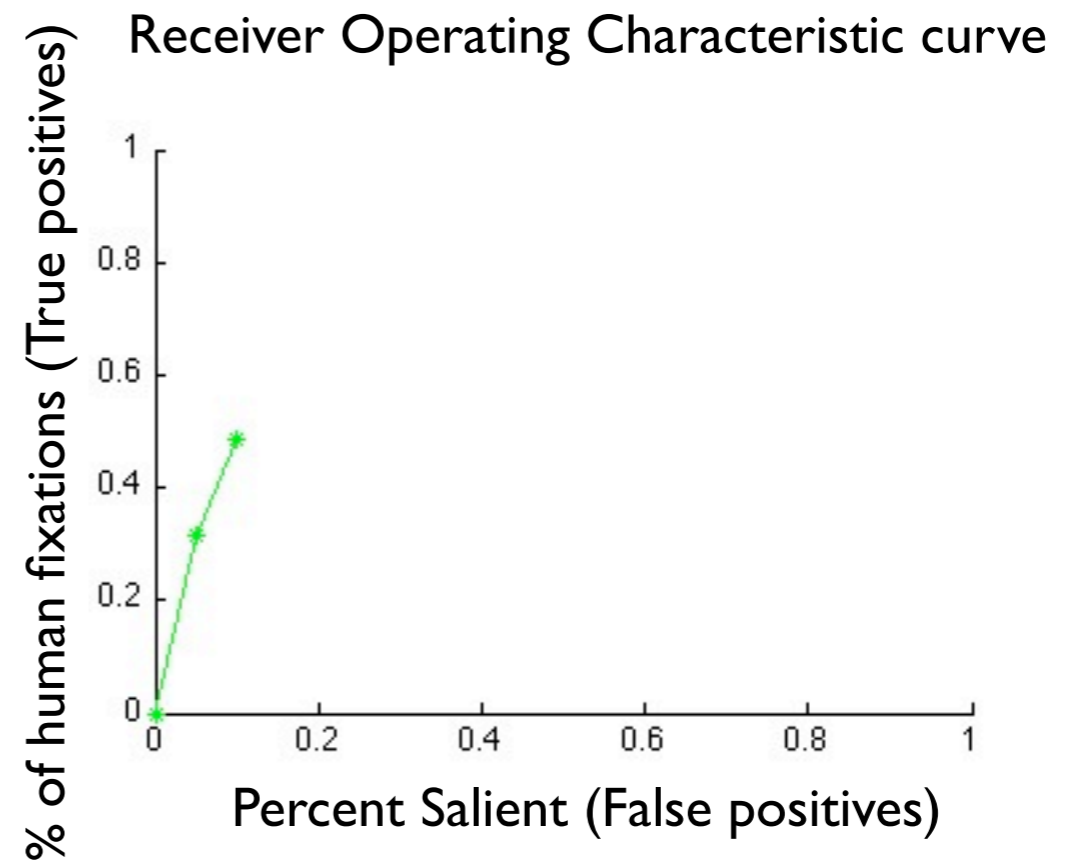
Thresholded Saliency Map



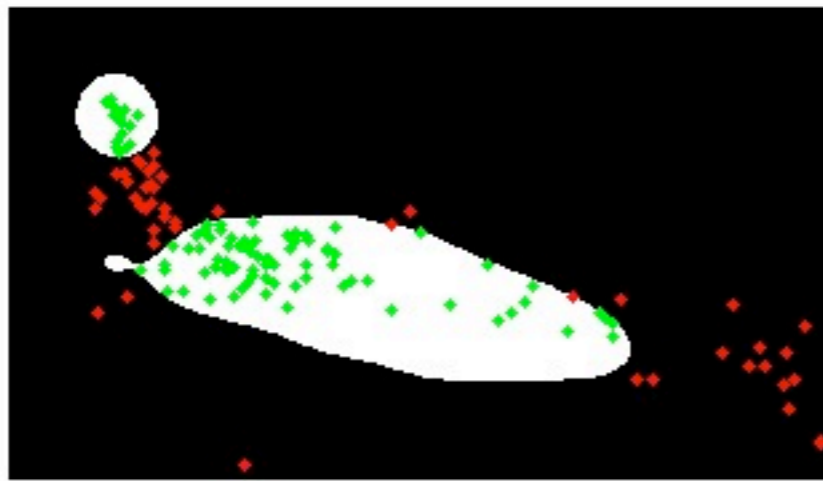
We use an ROC curve to measure the performance of a saliency map



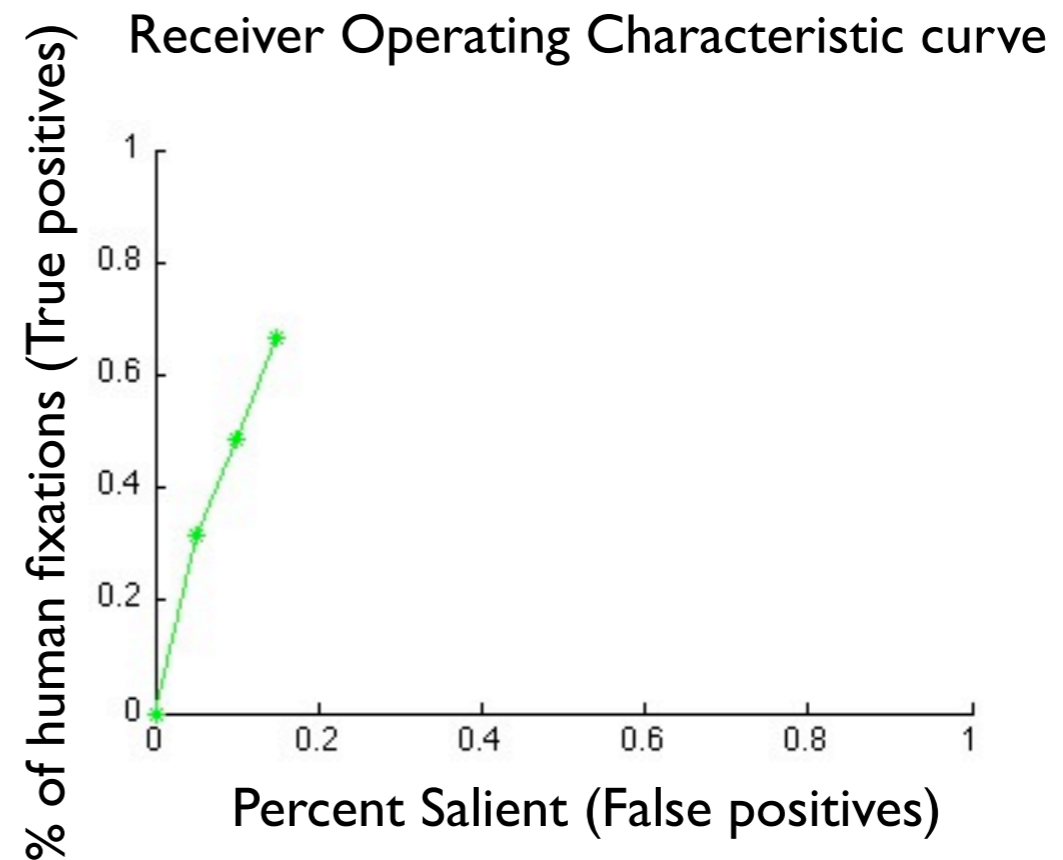
Thresholded Saliency Map



We use an ROC curve to measure the performance of a saliency map

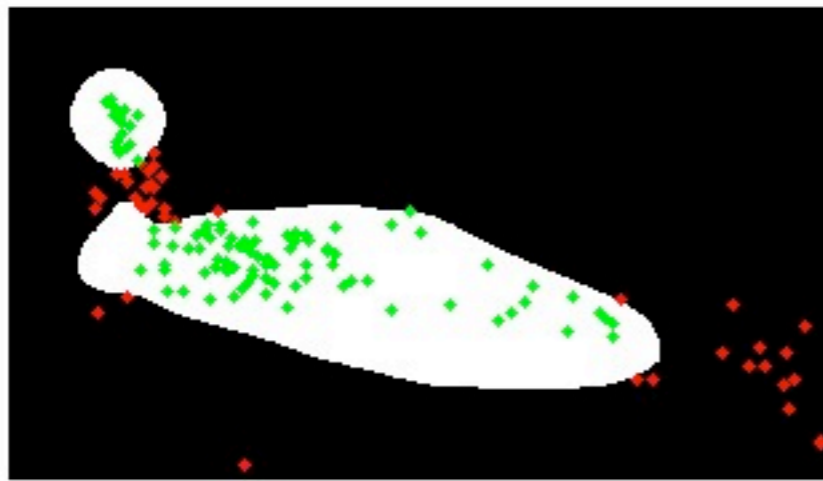


Thresholded Saliency Map

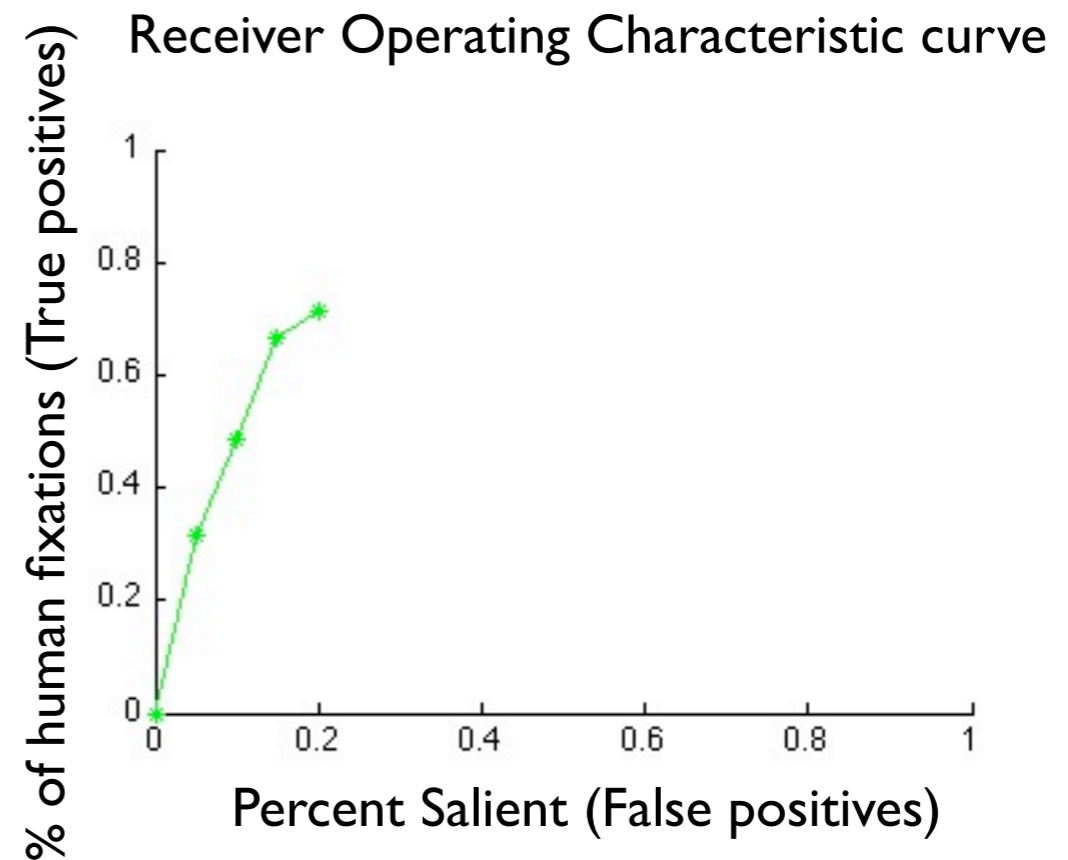


We calculate the percentage of fixations that lie within the salient portion of the map

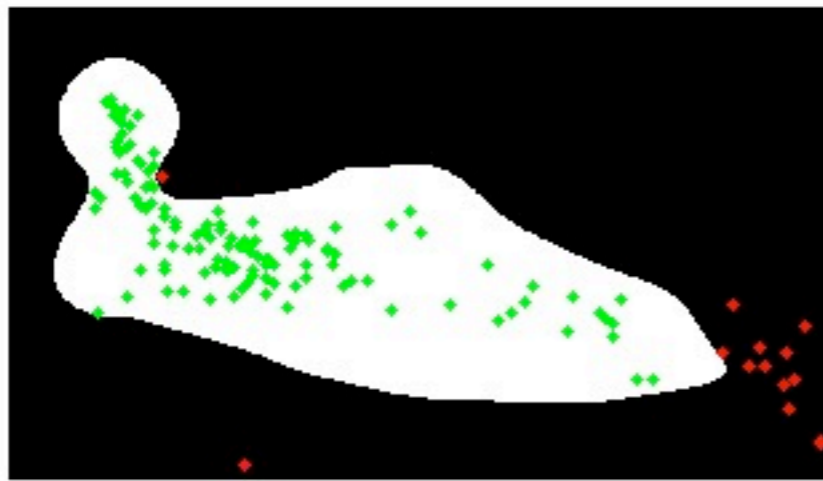
We use an ROC curve to measure the performance of a saliency map



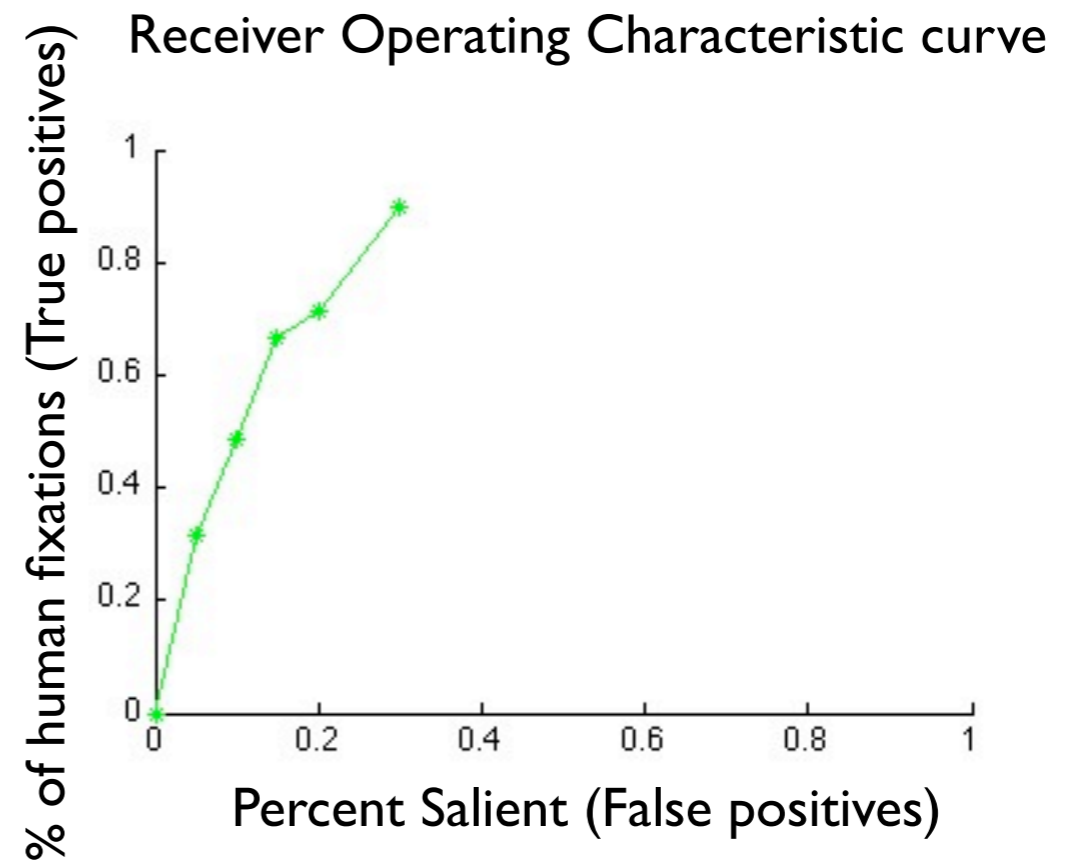
Thresholded Saliency Map



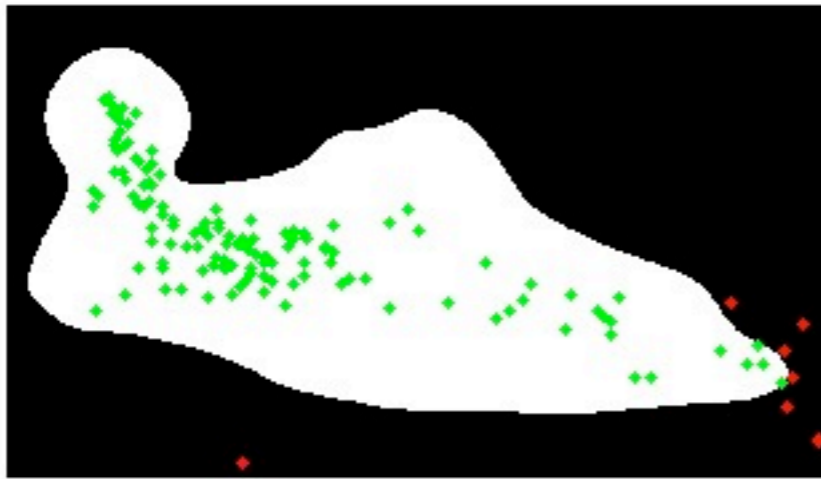
We use an ROC curve to measure the performance of a saliency map



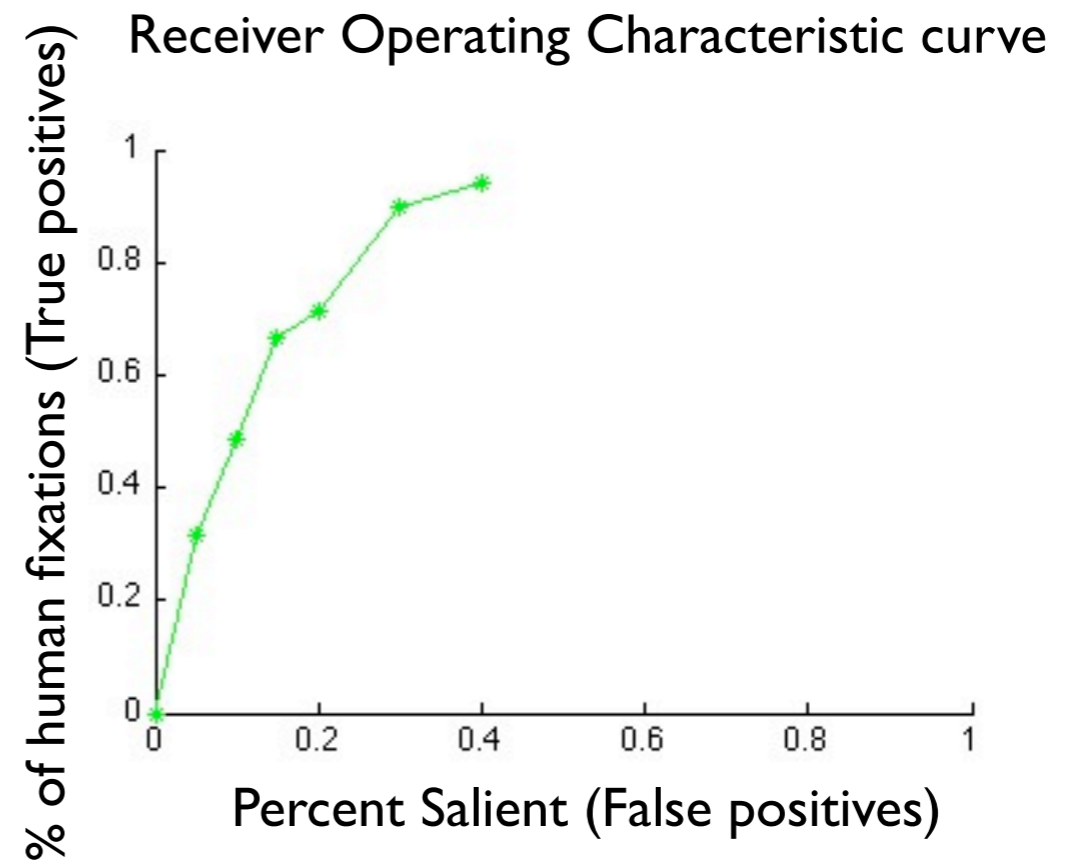
Thresholded Saliency Map



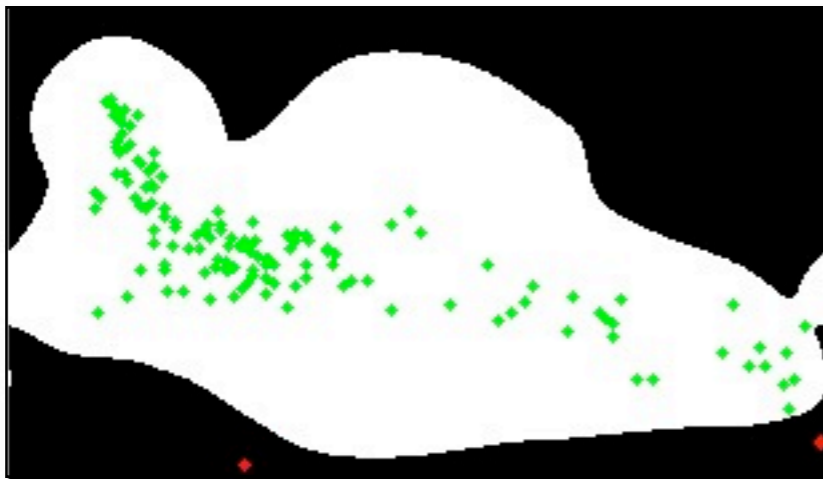
We use an ROC curve to measure the performance of a saliency map



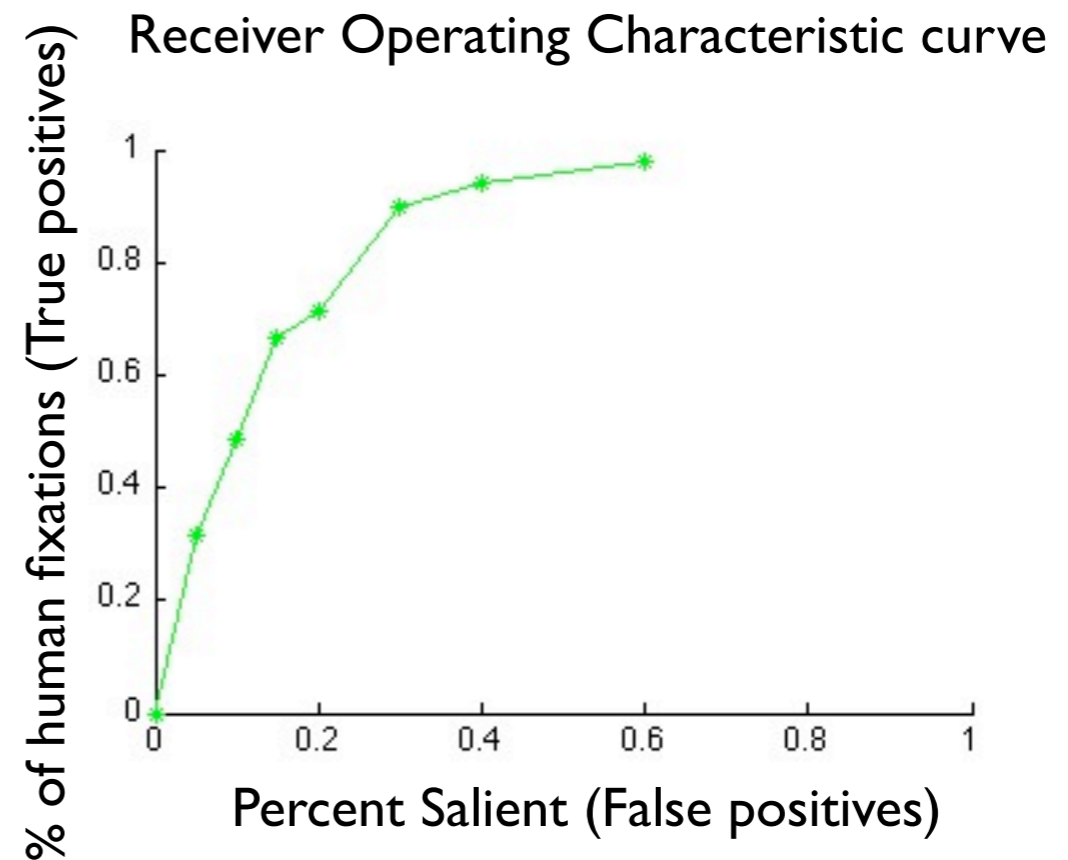
Thresholded Saliency Map



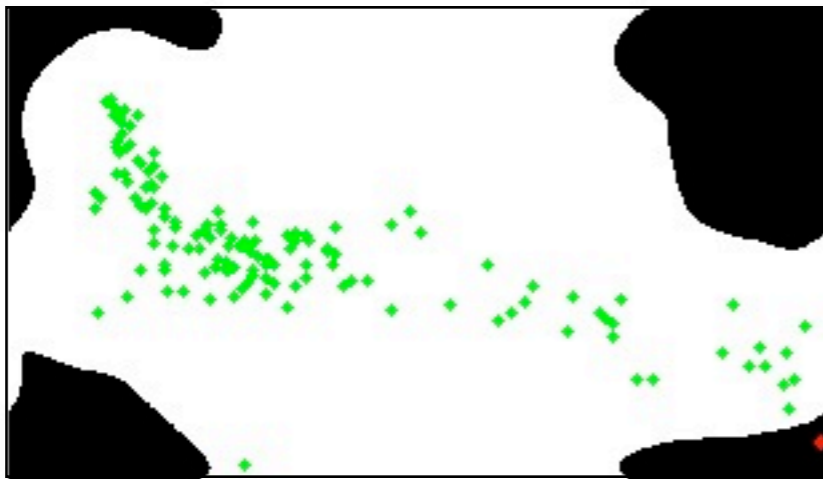
We use an ROC curve to measure the performance of a saliency map



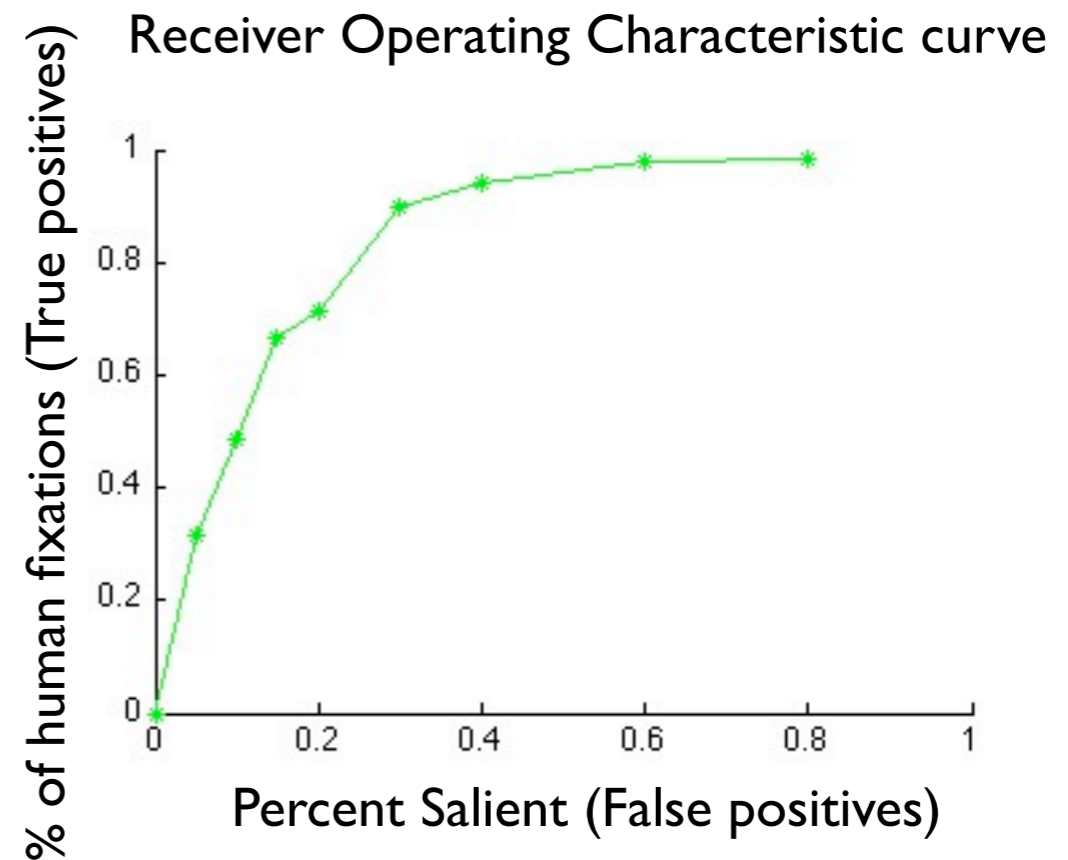
Thresholded Saliency Map



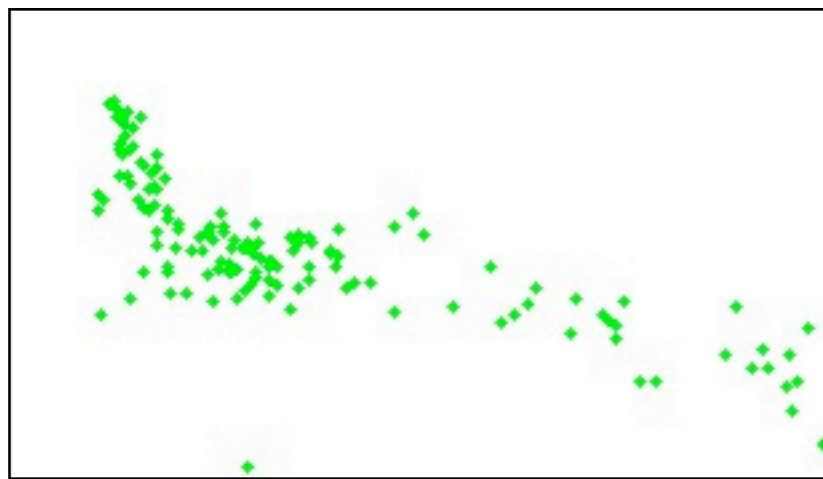
We use an ROC curve to measure the performance of a saliency map



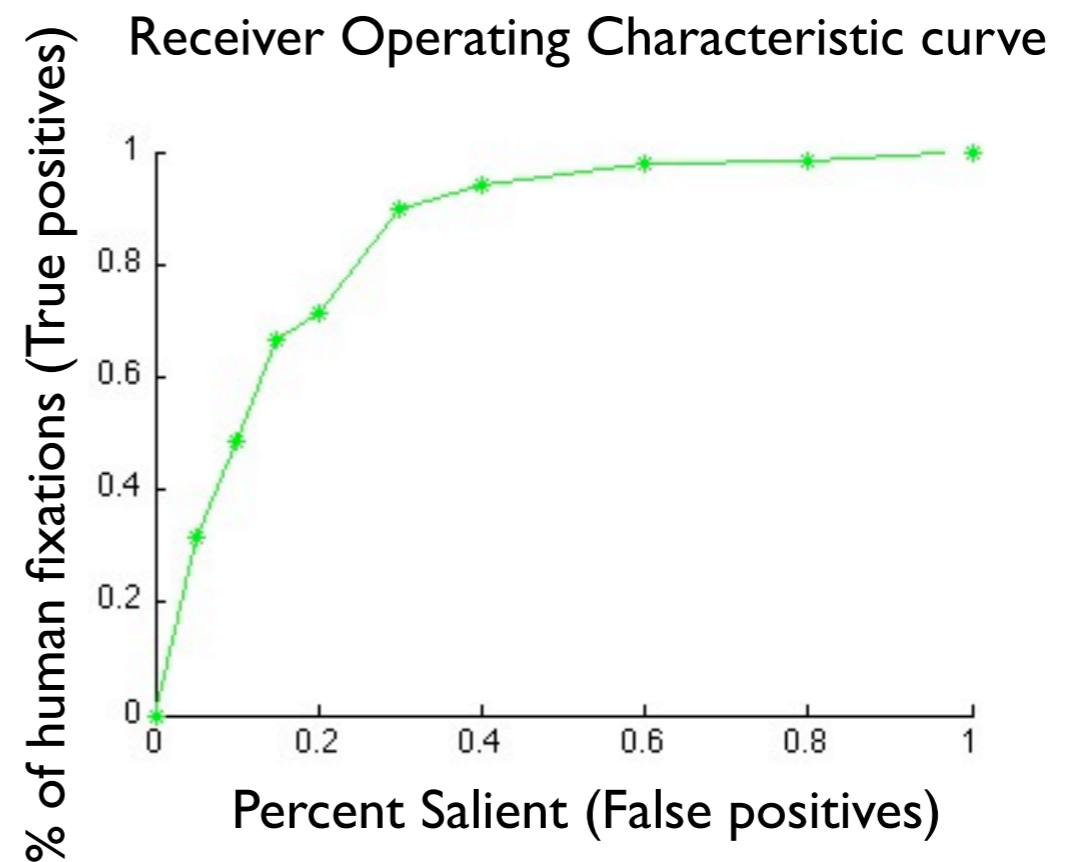
Thresholded Saliency Map



We use an ROC curve to measure the performance of a saliency map

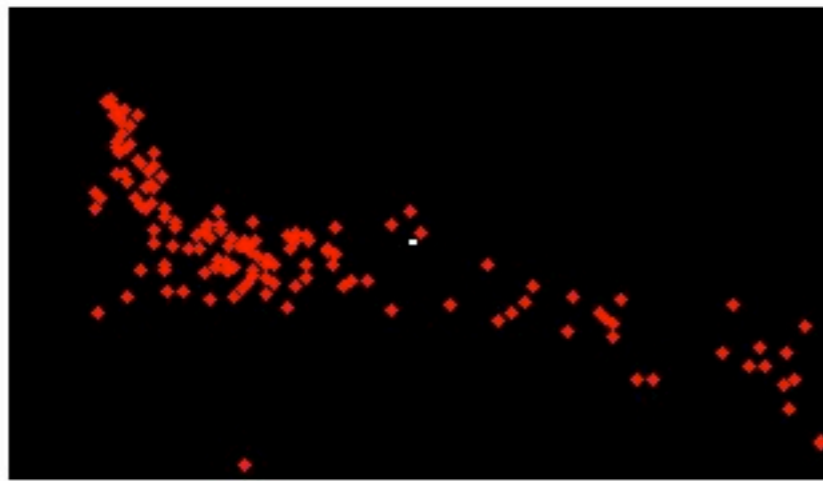


Thresholded Saliency Map

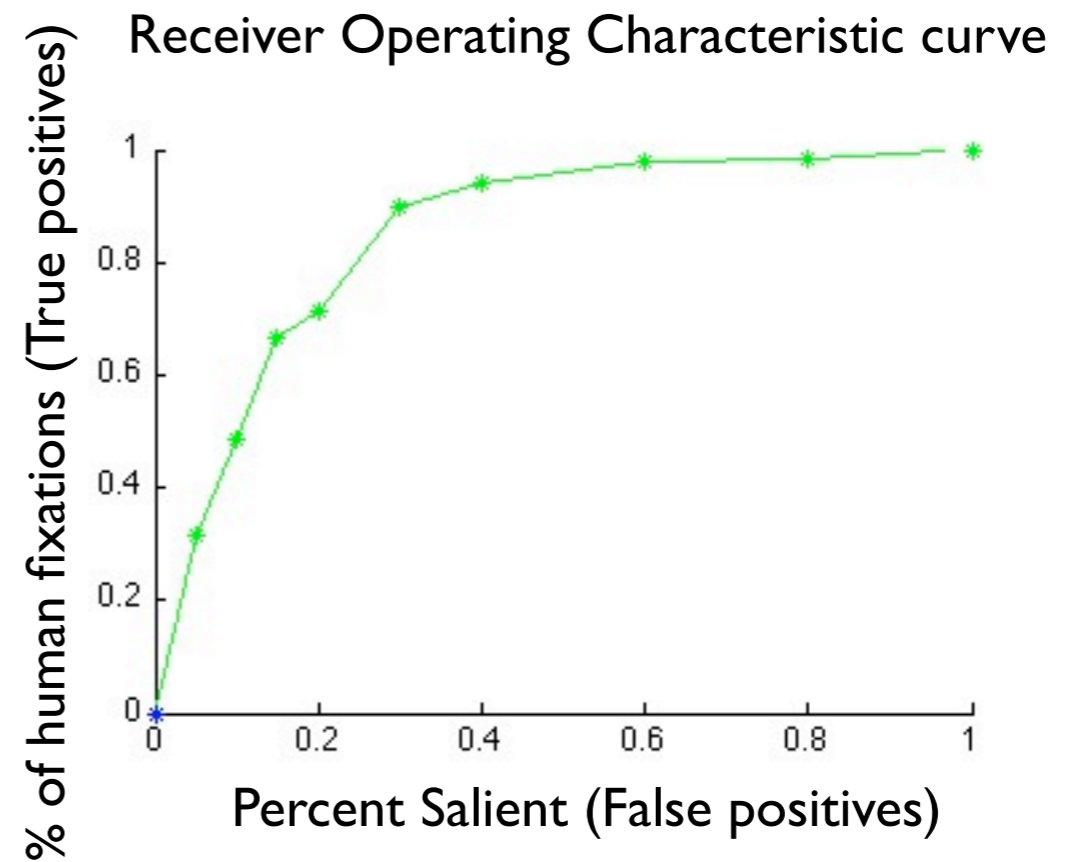


ROC curve always starts at 0
ends at 1

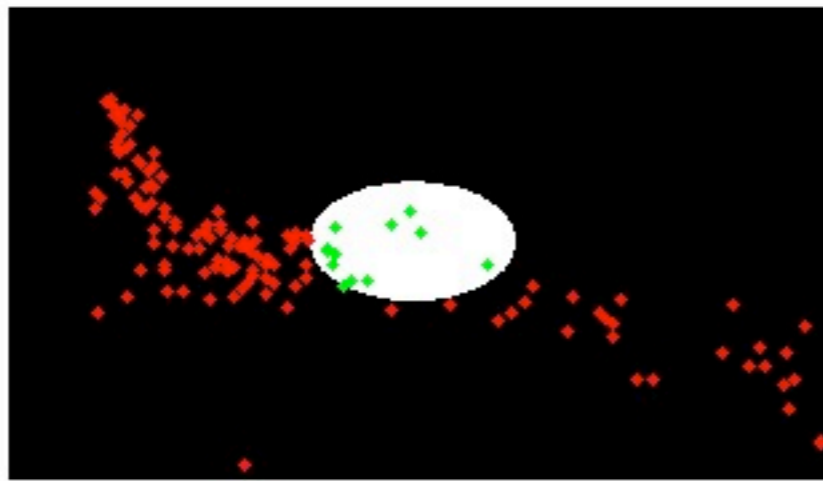
We use an ROC curve to measure the performance of a saliency map



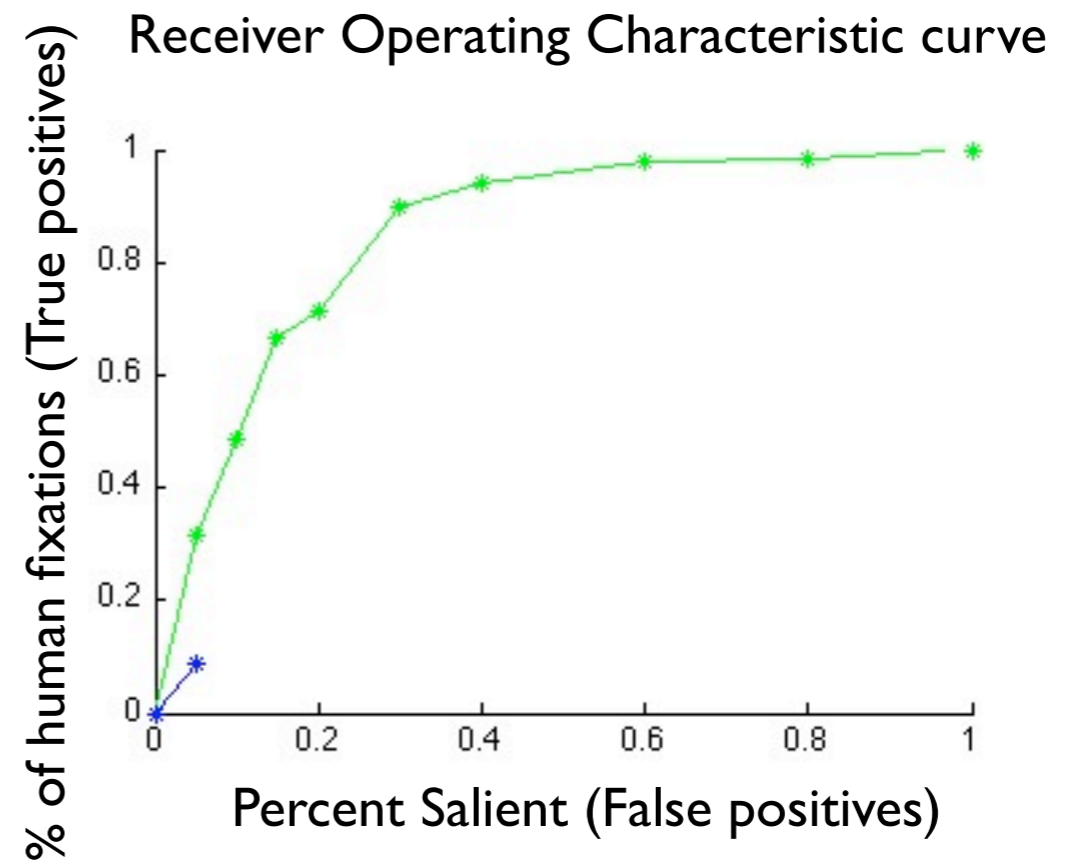
Thresholded Center Map



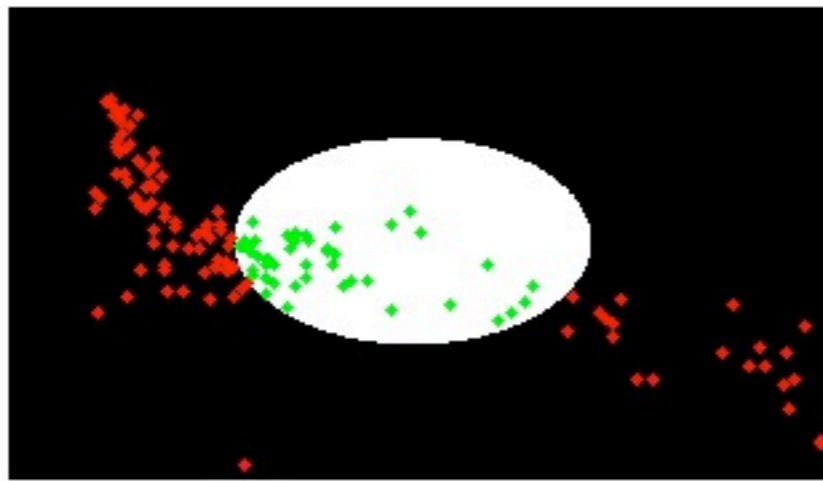
We use an ROC curve to measure the performance of a saliency map



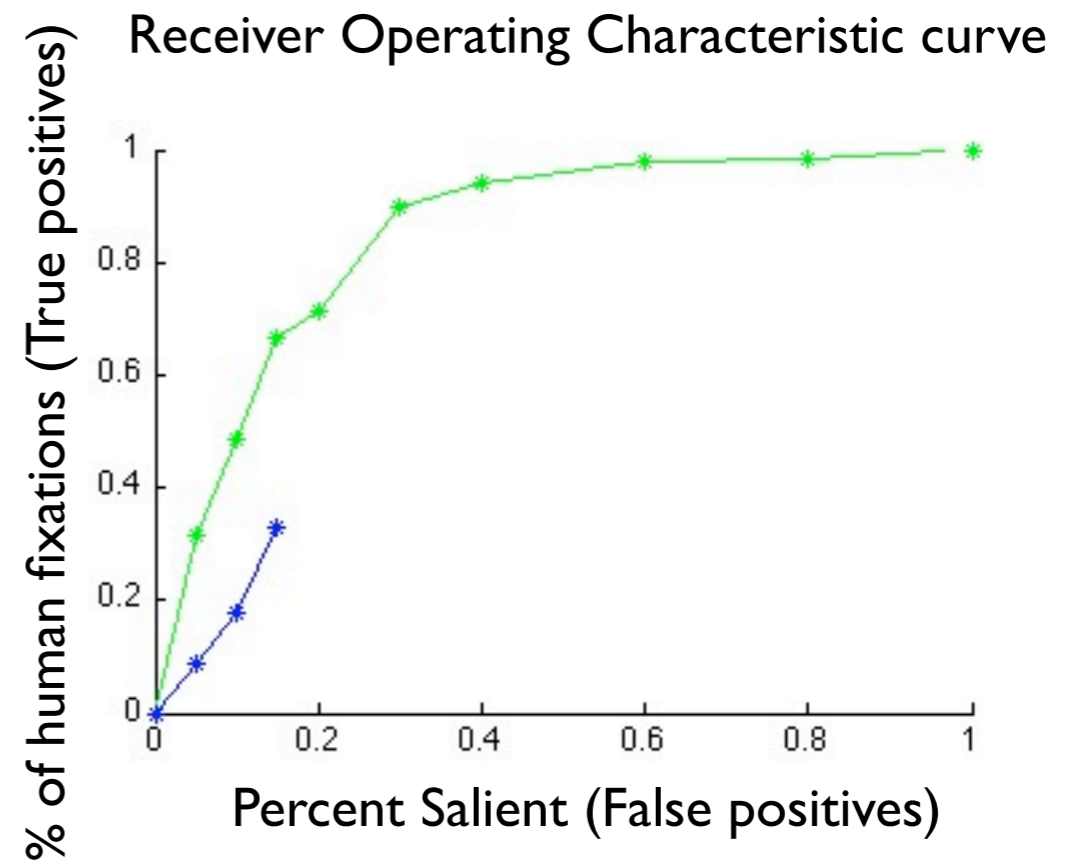
Thresholded Center Map



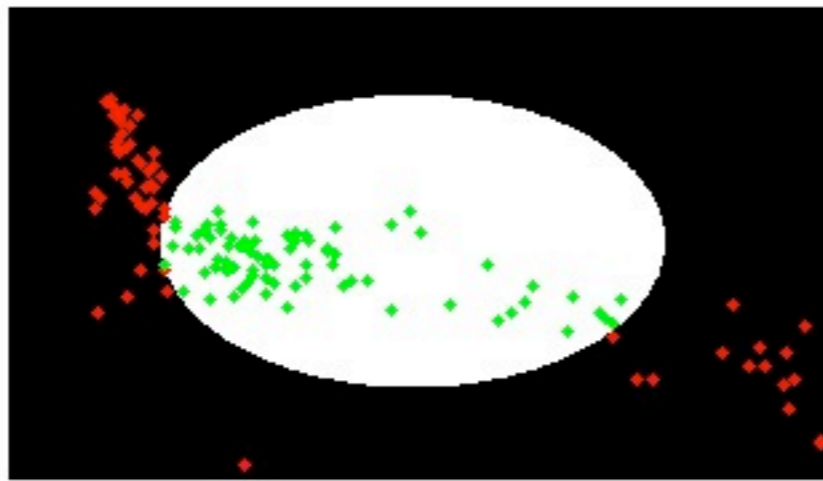
We use an ROC curve to measure the performance of a saliency map



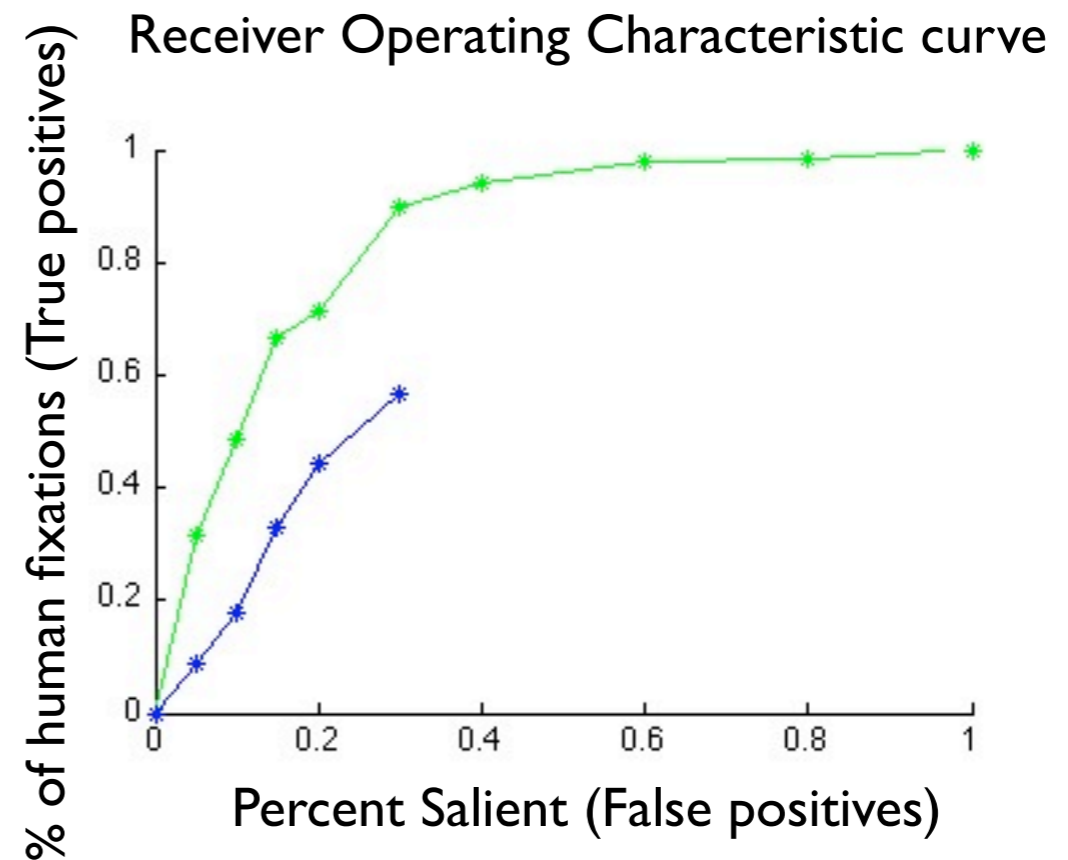
Thresholded Center Map



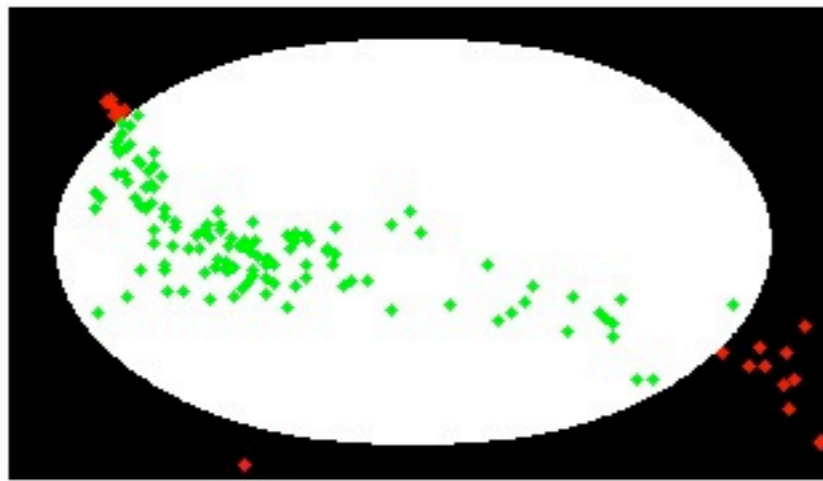
We use an ROC curve to measure the performance of a saliency map



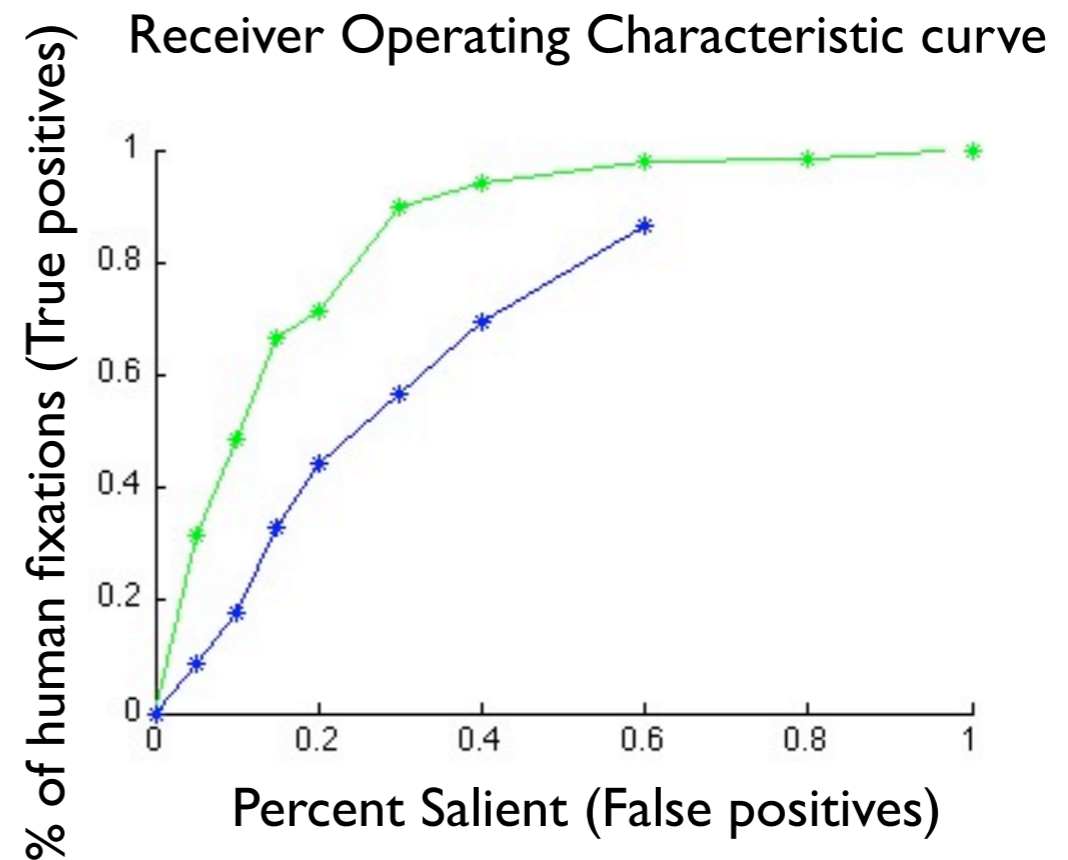
Thresholded Center Map



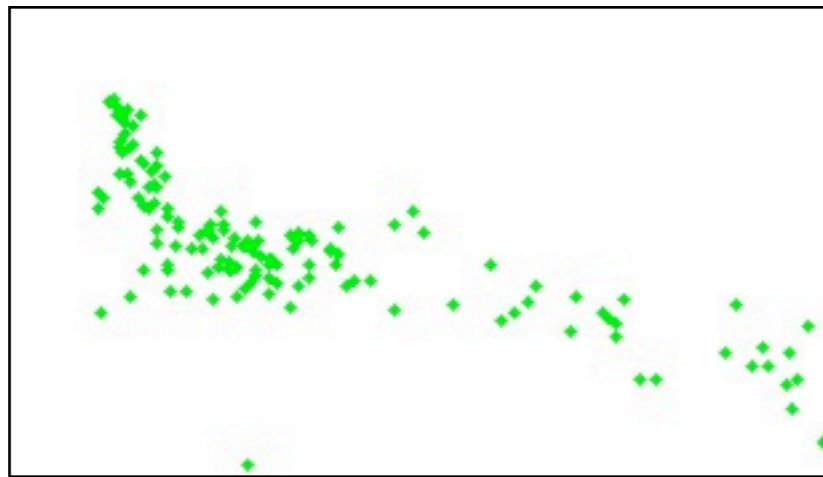
We use an ROC curve to measure the performance of a saliency map



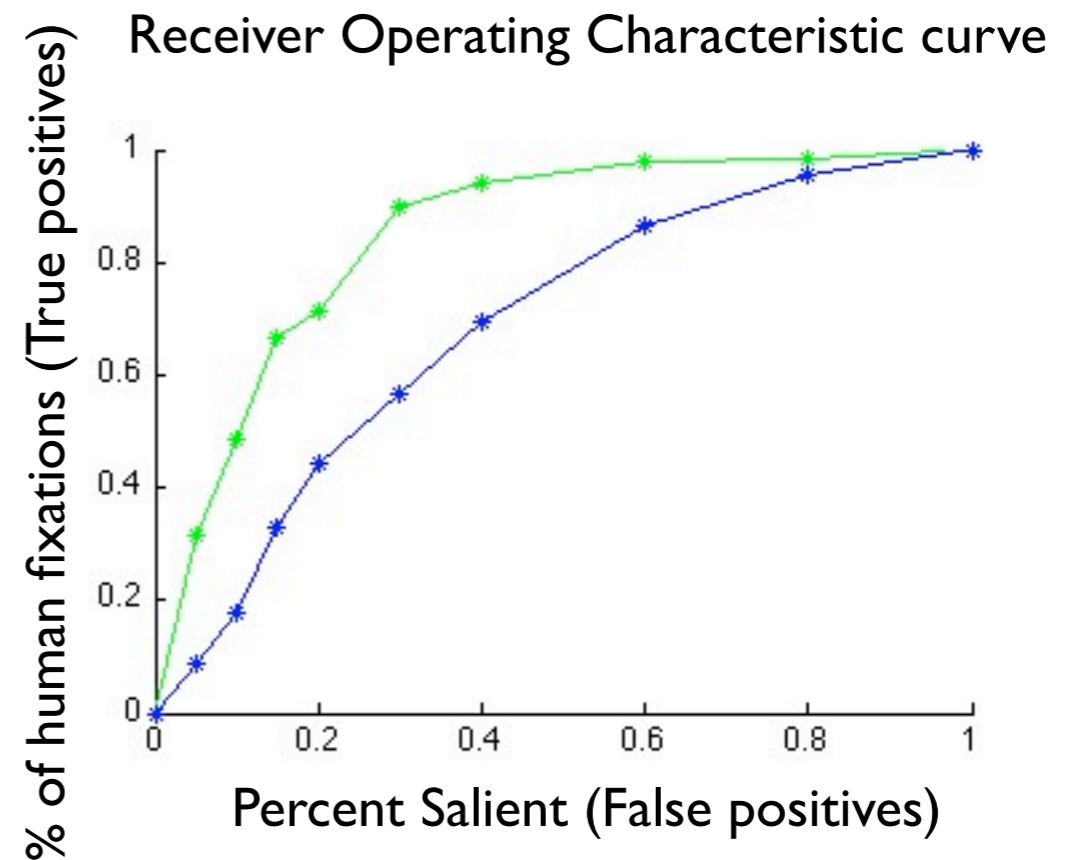
Thresholded Center Map



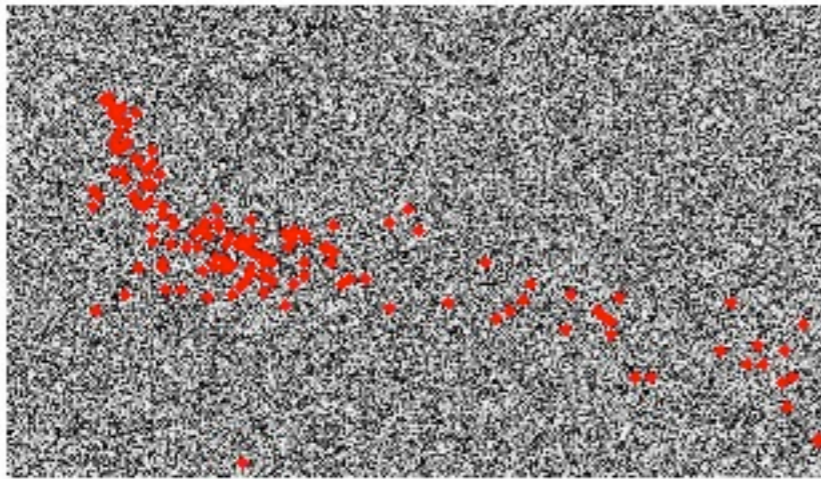
We use an ROC curve to measure the performance of a saliency map



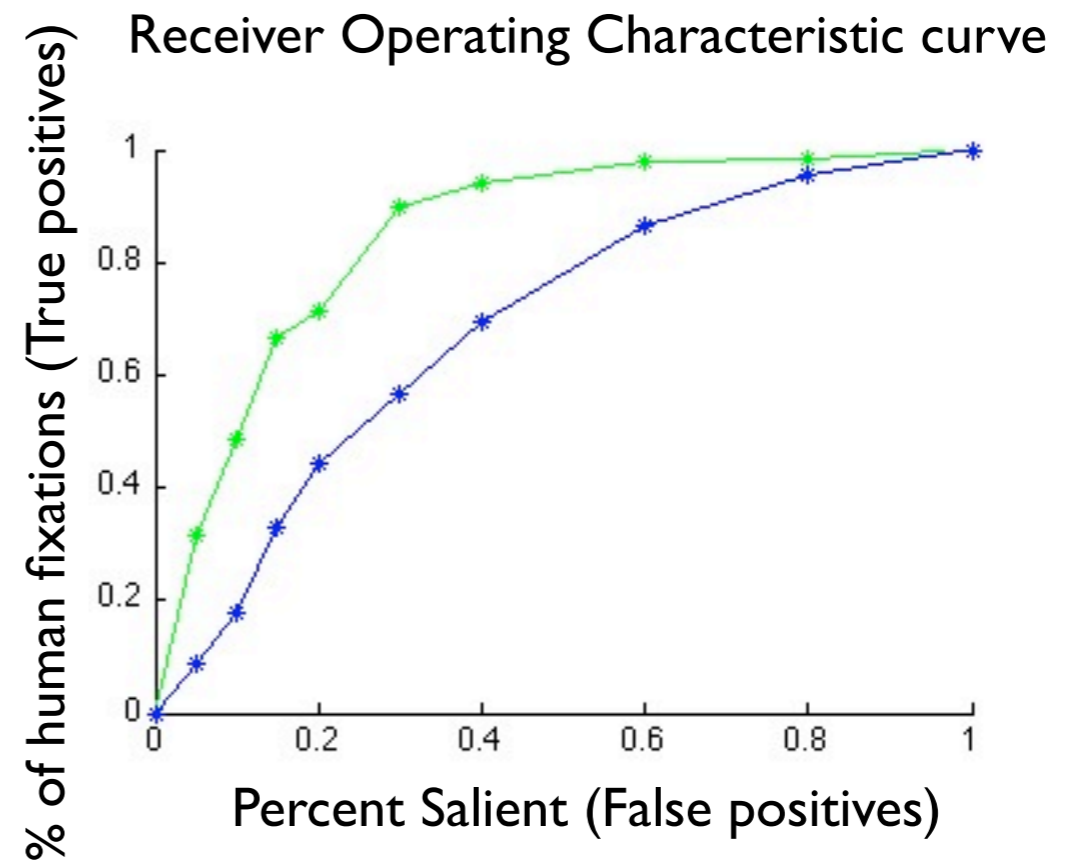
Thresholded Center Map



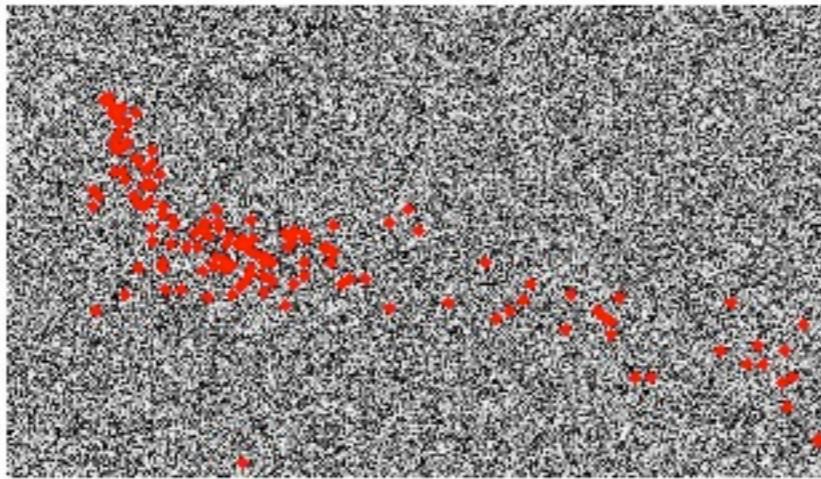
We use an ROC curve to measure the performance of a saliency map



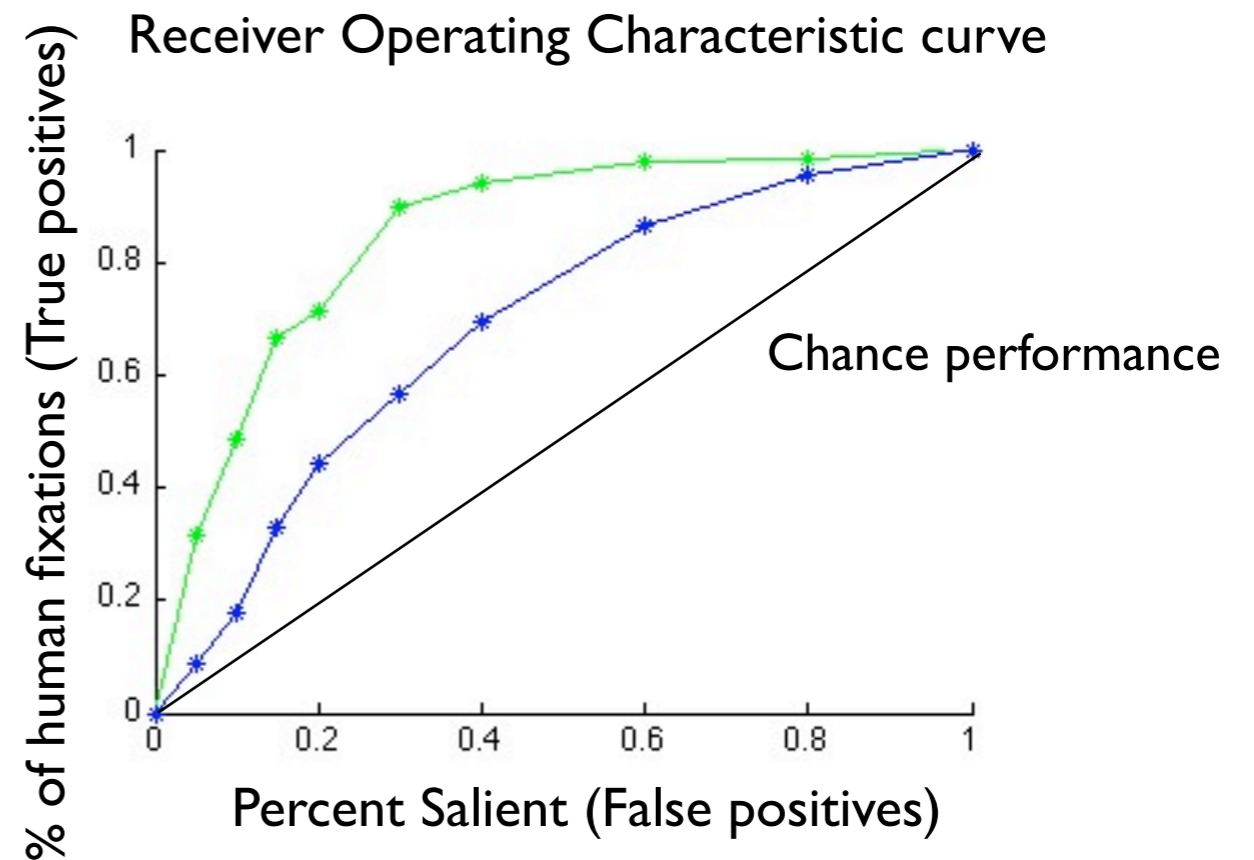
Random Noise map



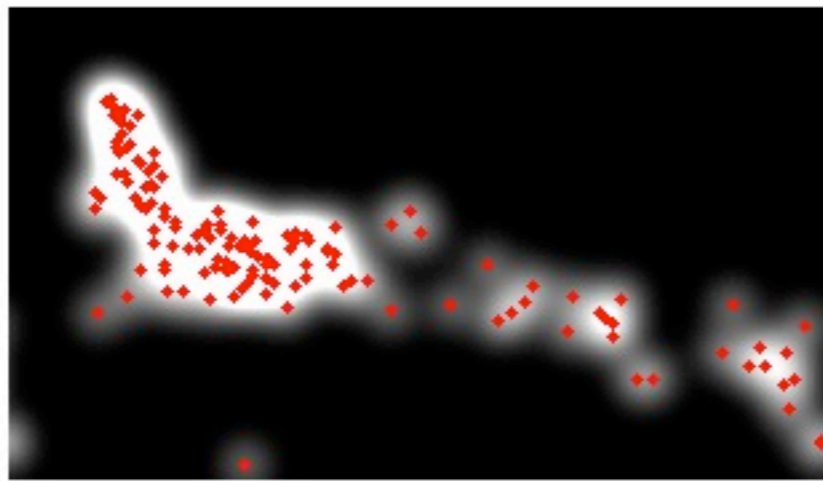
We use an ROC curve to measure the performance of a saliency map



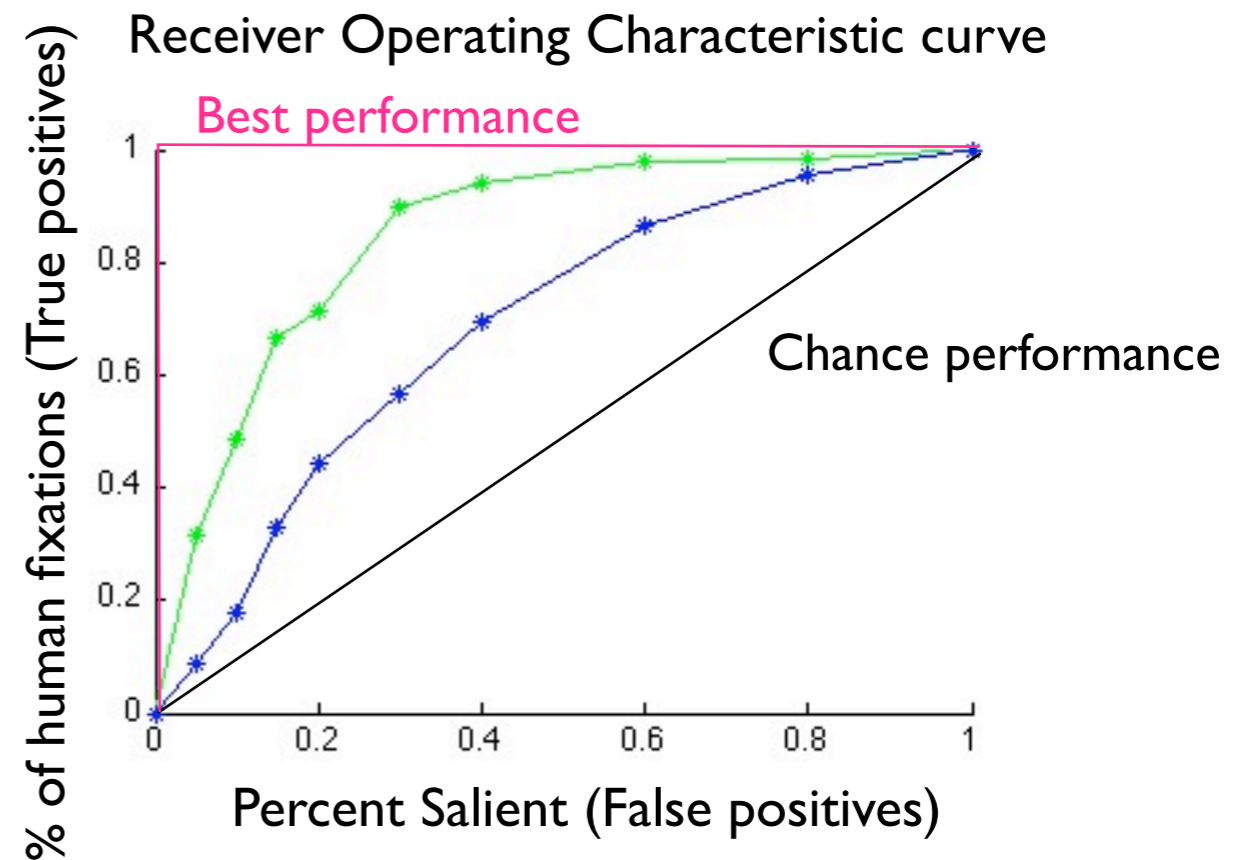
Random Noise map



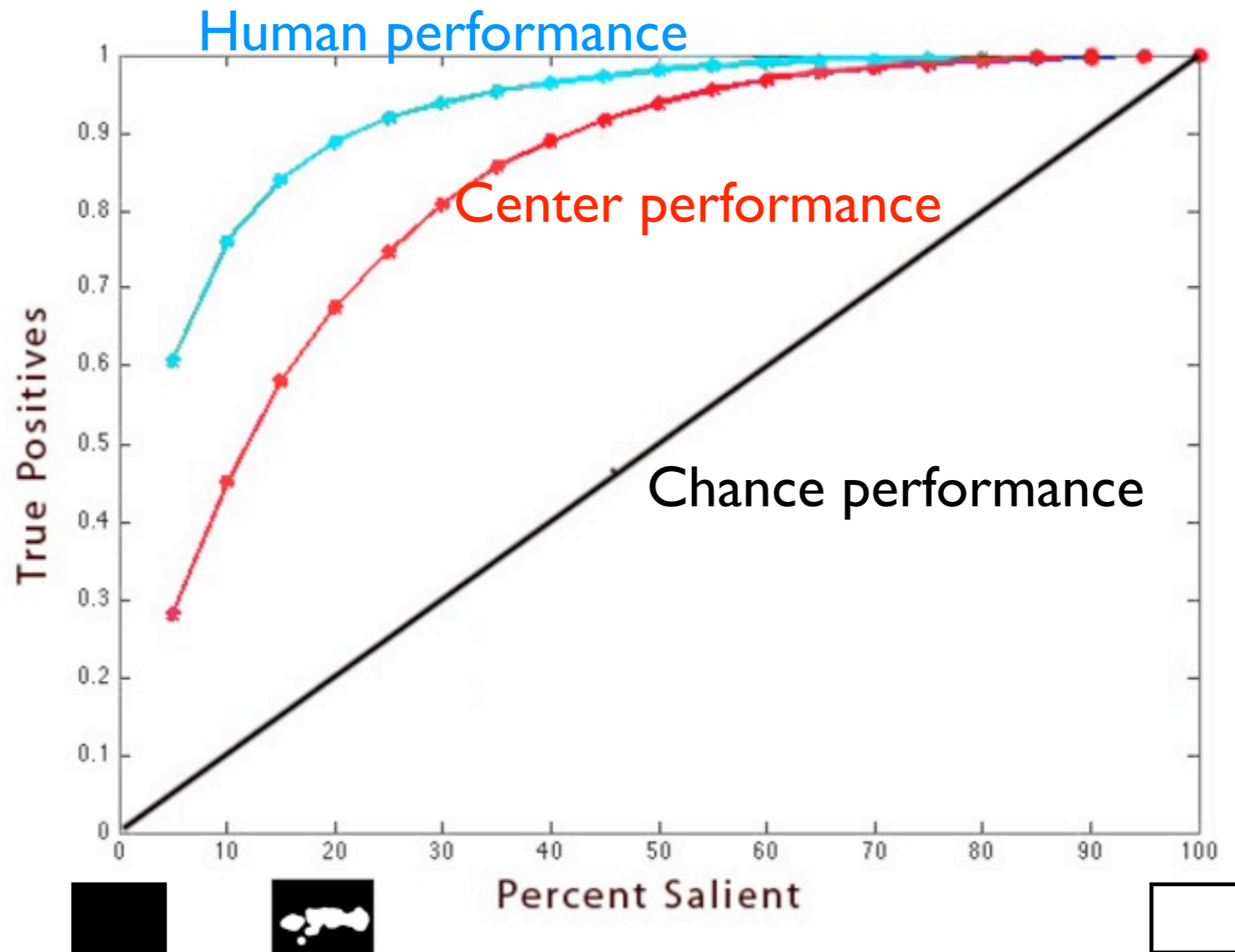
We use an ROC curve to measure the performance of a saliency map



Perfect saliency map



Humans are good predictors of where other humans will look



Our goal is to learn where people look directly from eye tracking data

- Step 1: collect eye tracking data set
- Step 2: learn the model

We compute low and high level image features for each image...

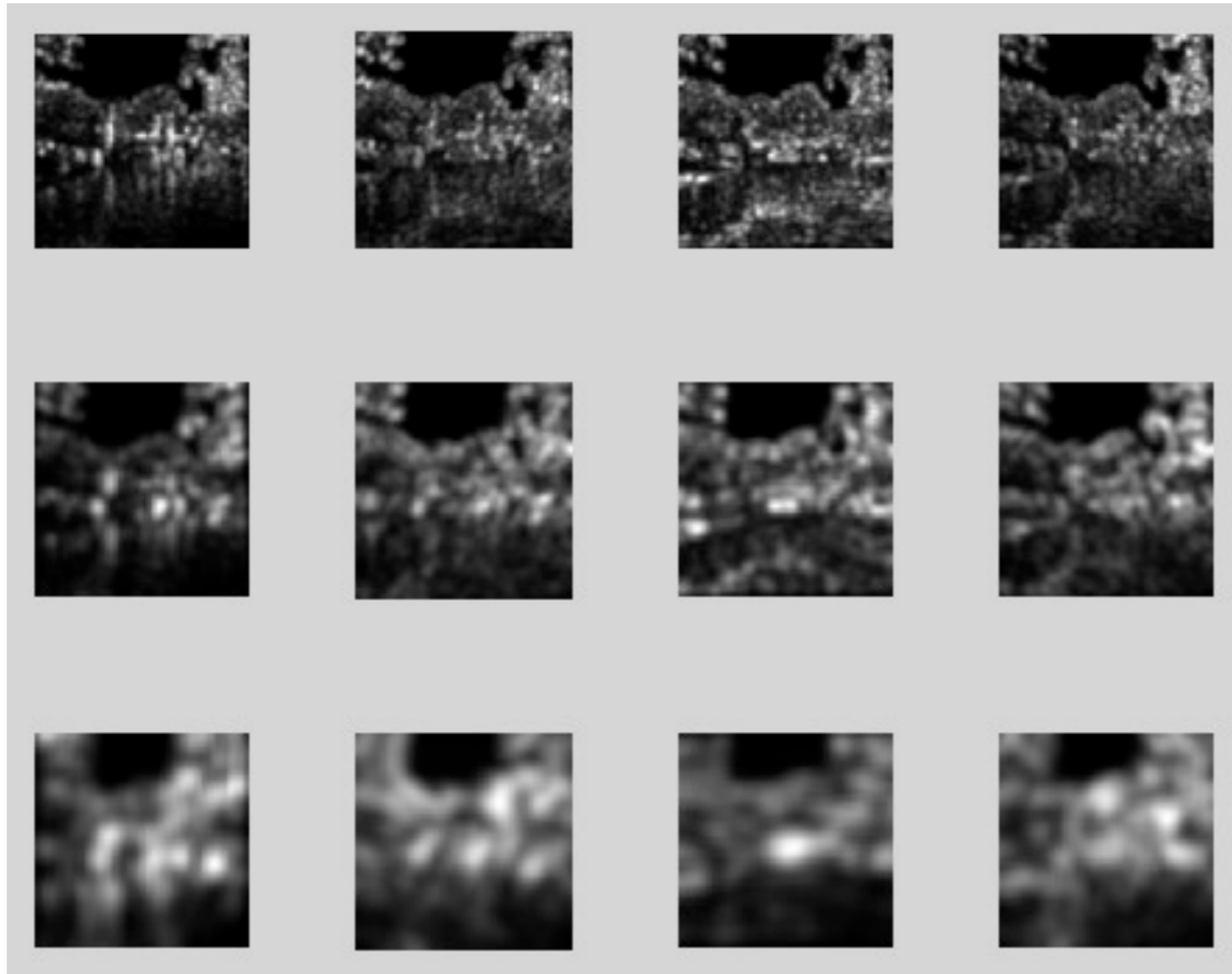


We compute low level image features

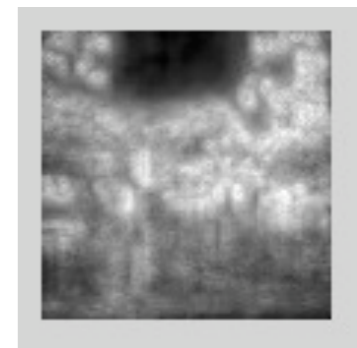
Use subbands of the steerable pyramid filters
(Simoncelli and Freeman 1995)

4 orientations

3 scales



Simple saliency map
from Torralba based
on steerable filters



We compute low level image features



We compute low level image features



As calculated by Itti and Koch's saliency method

We compute low level image features

Red

Green

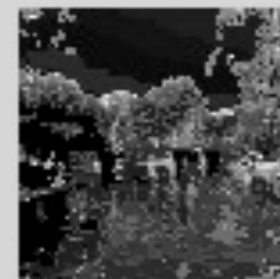
Blue



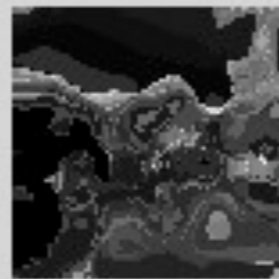
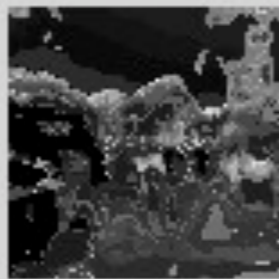
Color channels



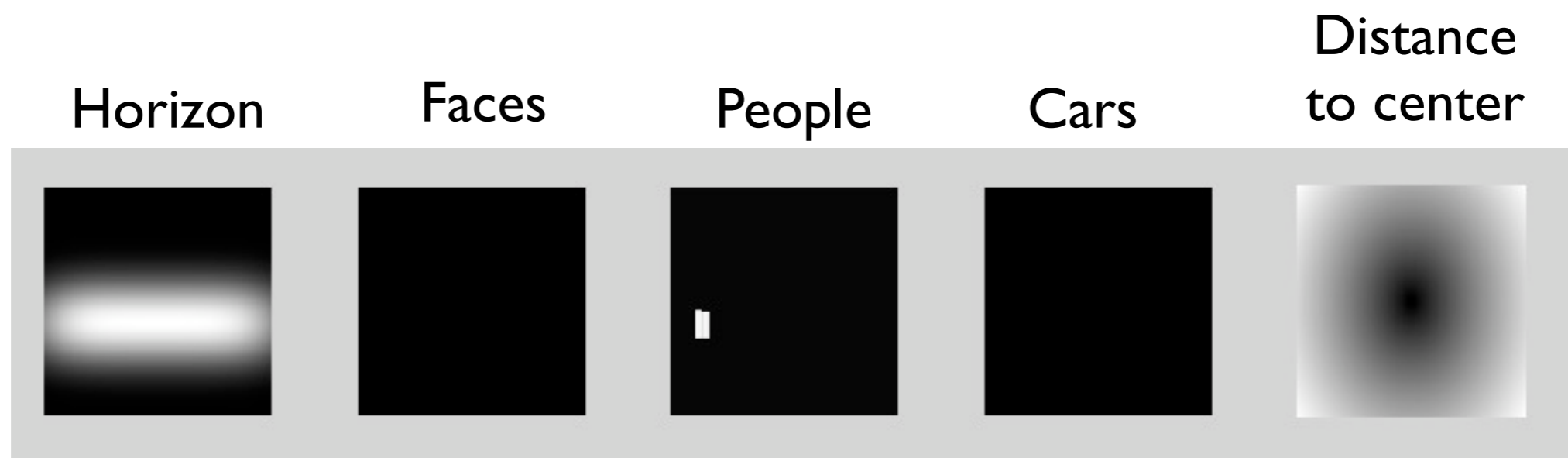
Probability of colors



Compute probability of colors from 3D color histograms at different scales



We compute high level image features



We compute high level image features

we train a
horizon detector
from gist features



Horizon



Faces



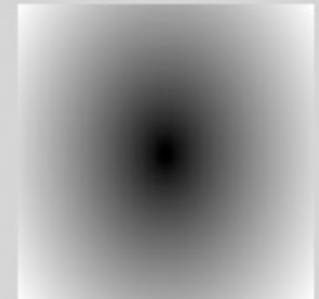
People



Cars

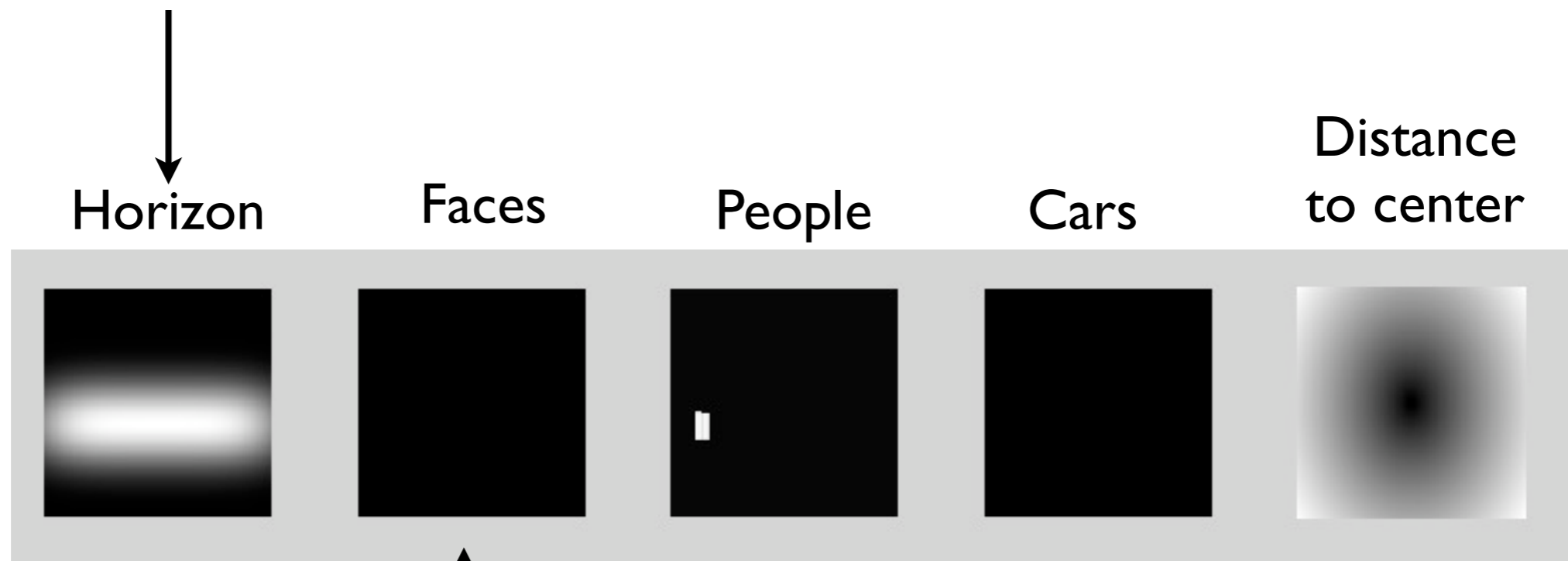


Distance
to center



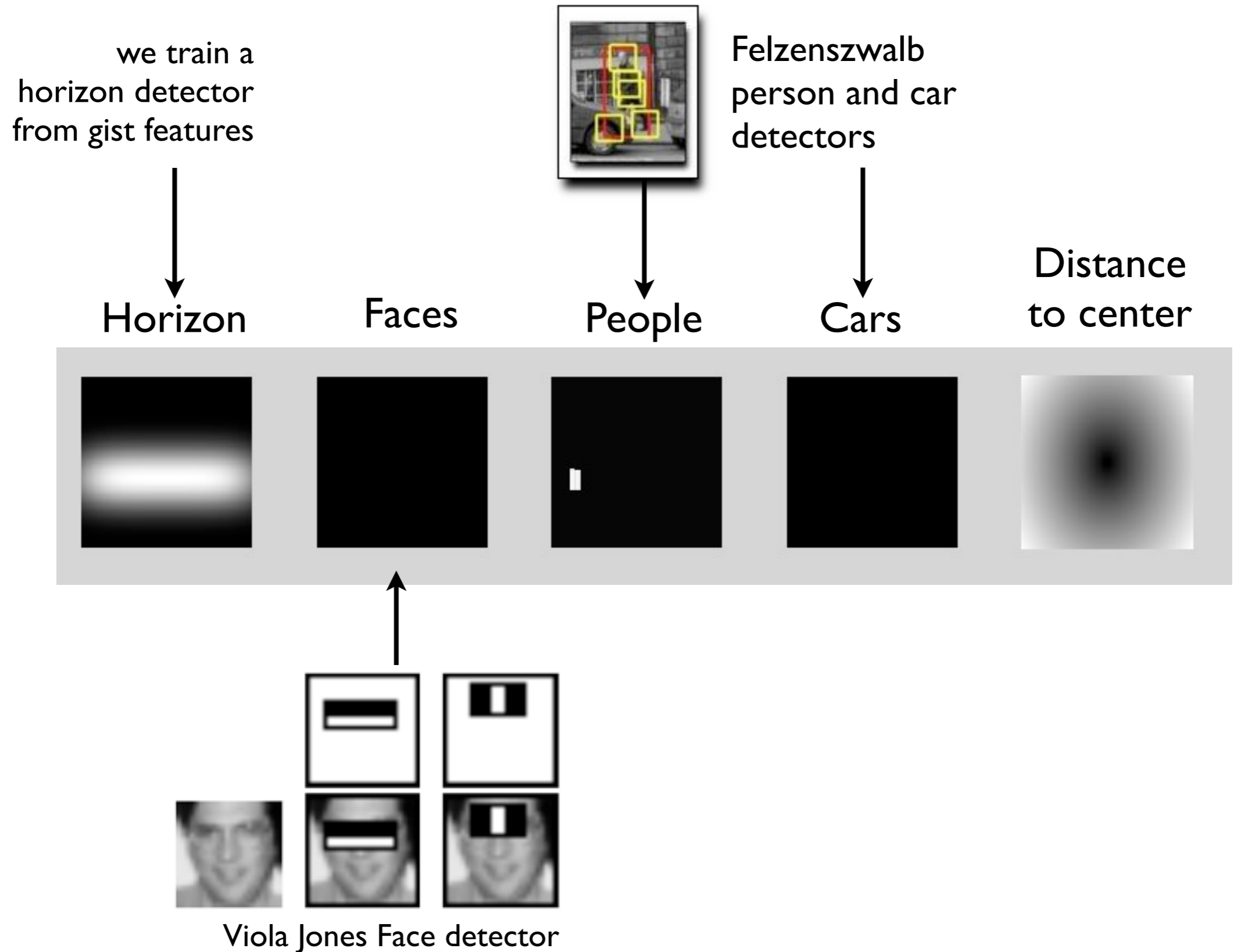
We compute high level image features

we train a
horizon detector
from gist features

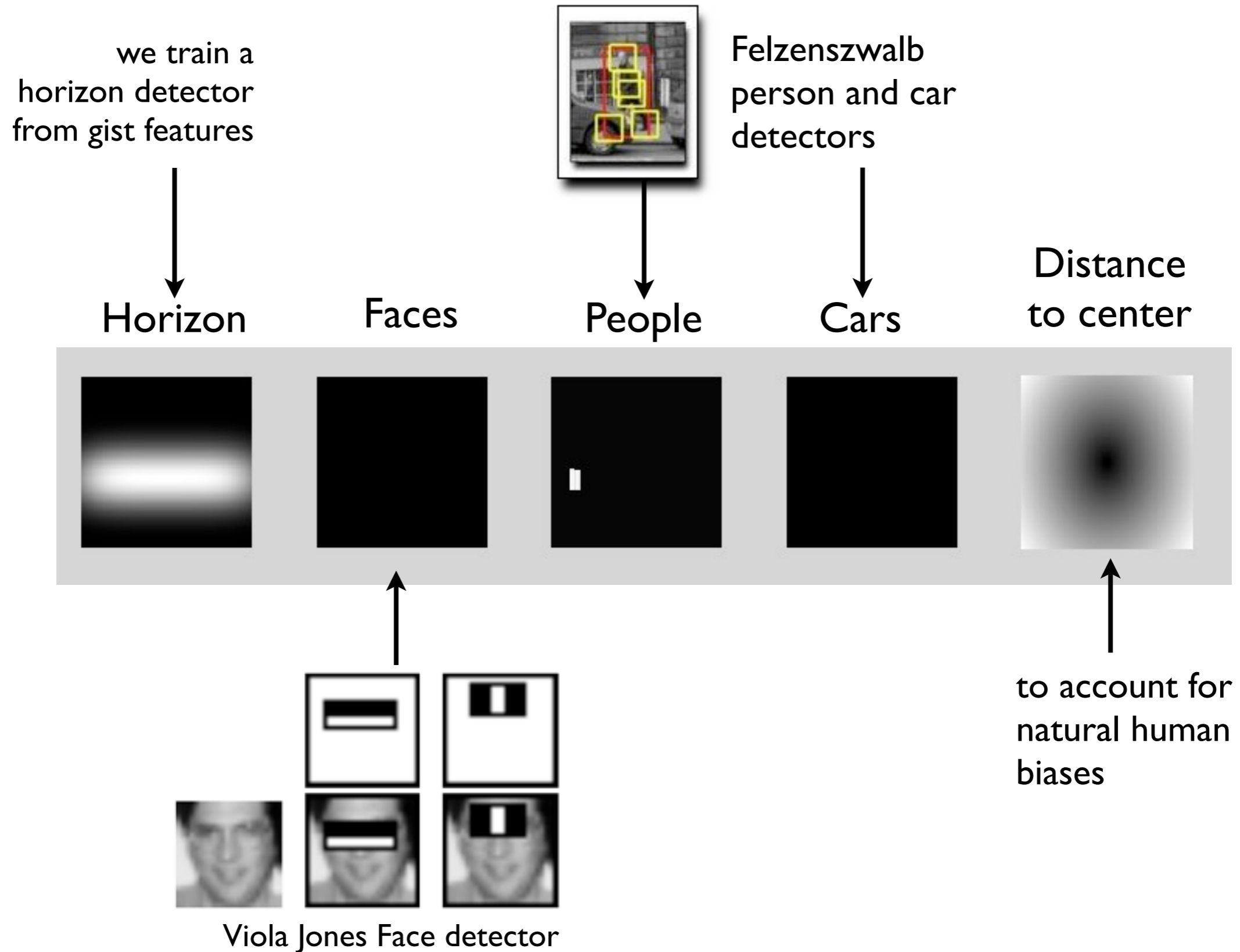


Viola Jones Face detector

We compute high level image features



We compute high level image features

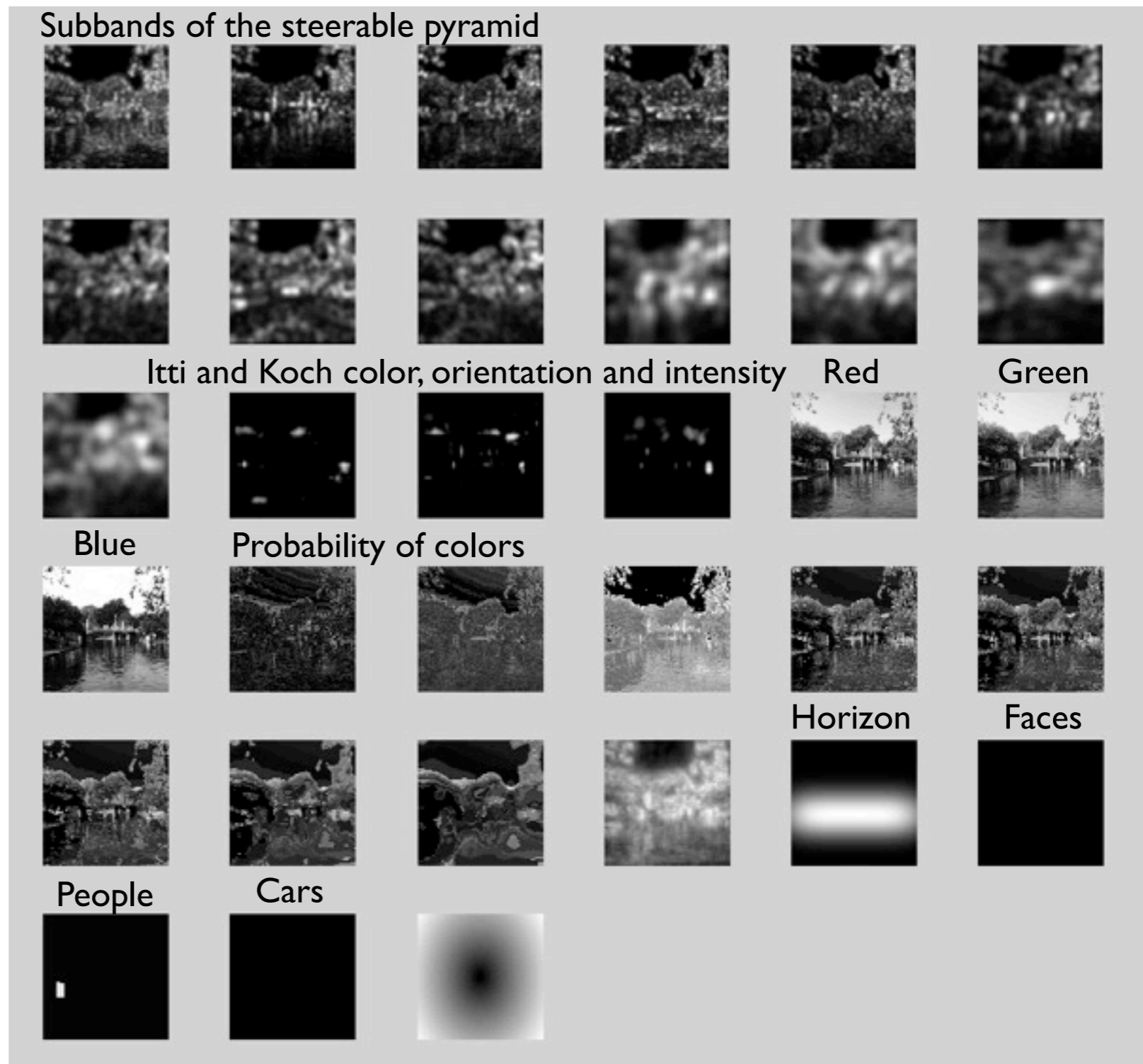


Each image had a stack of 33 features

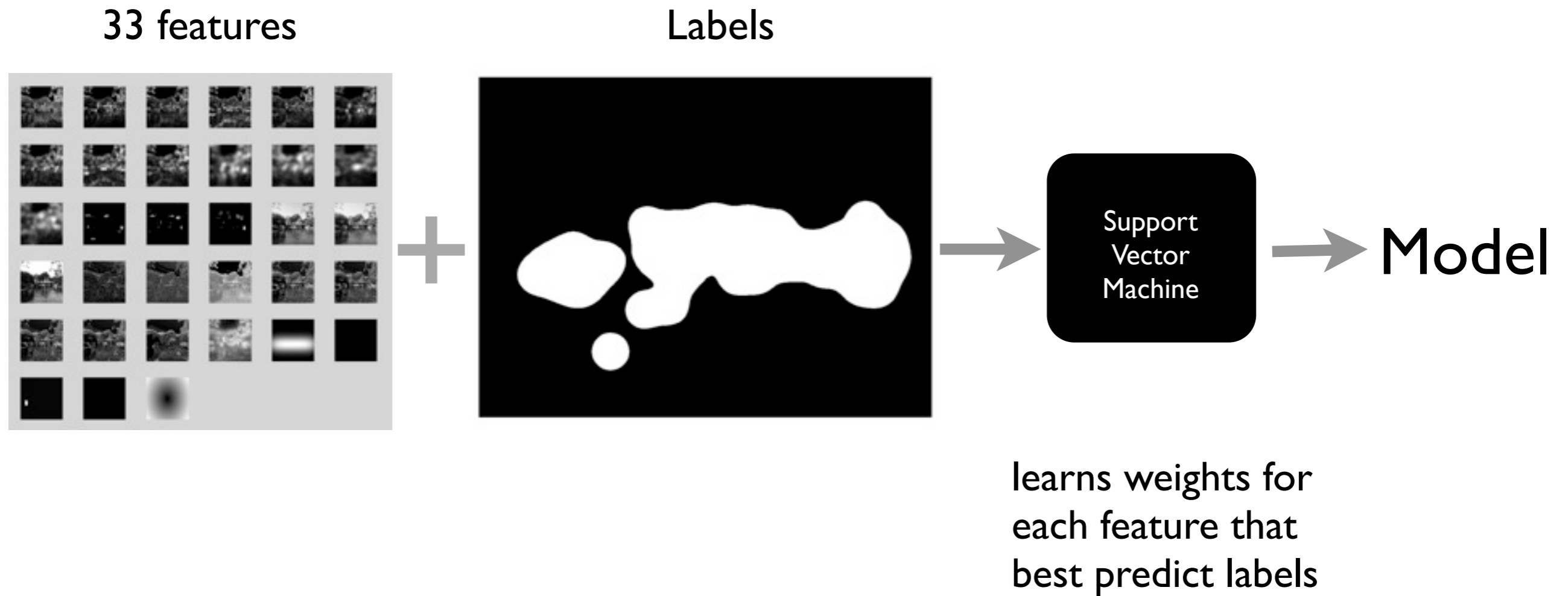
Image



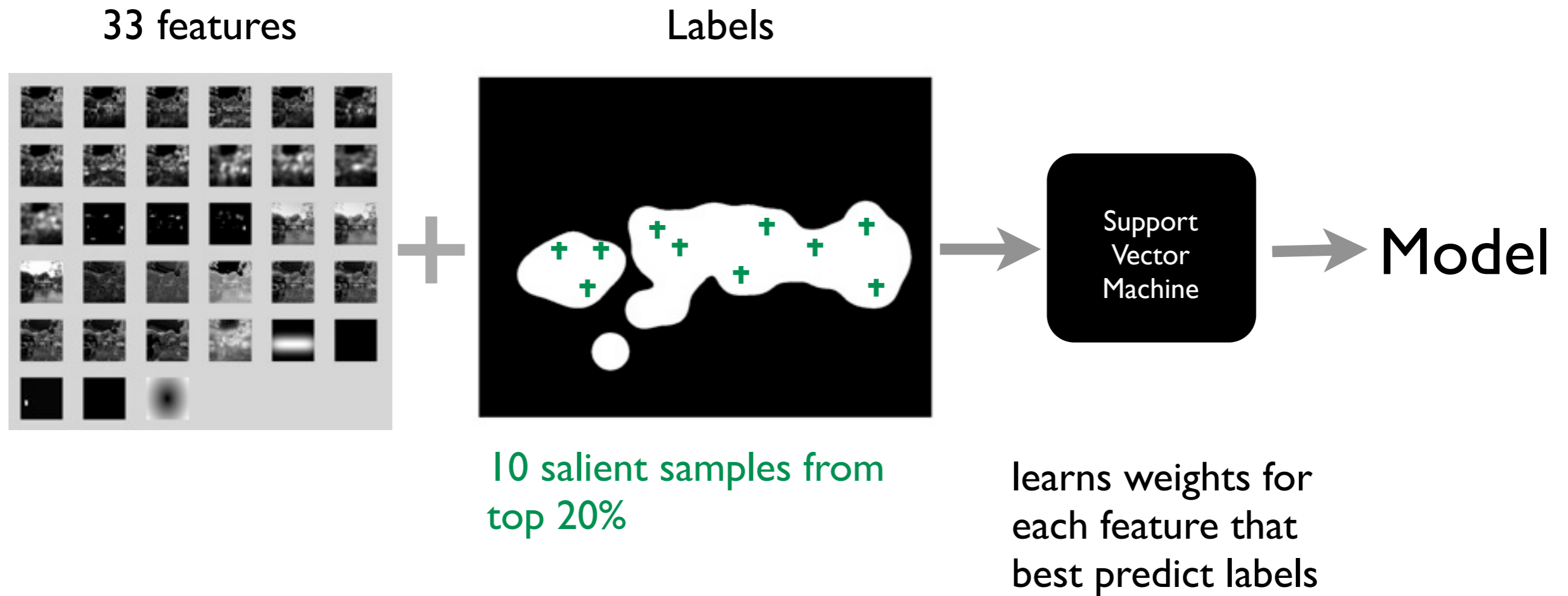
Features



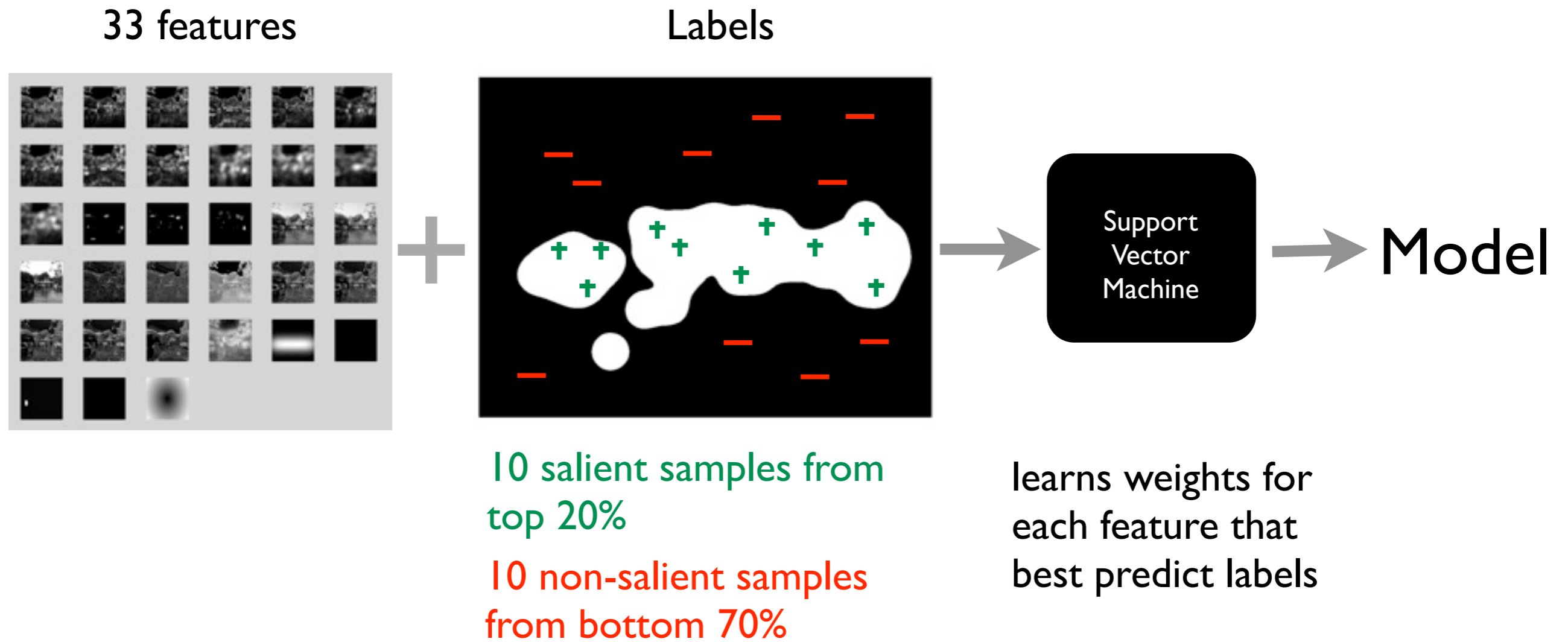
We learn several models of saliency



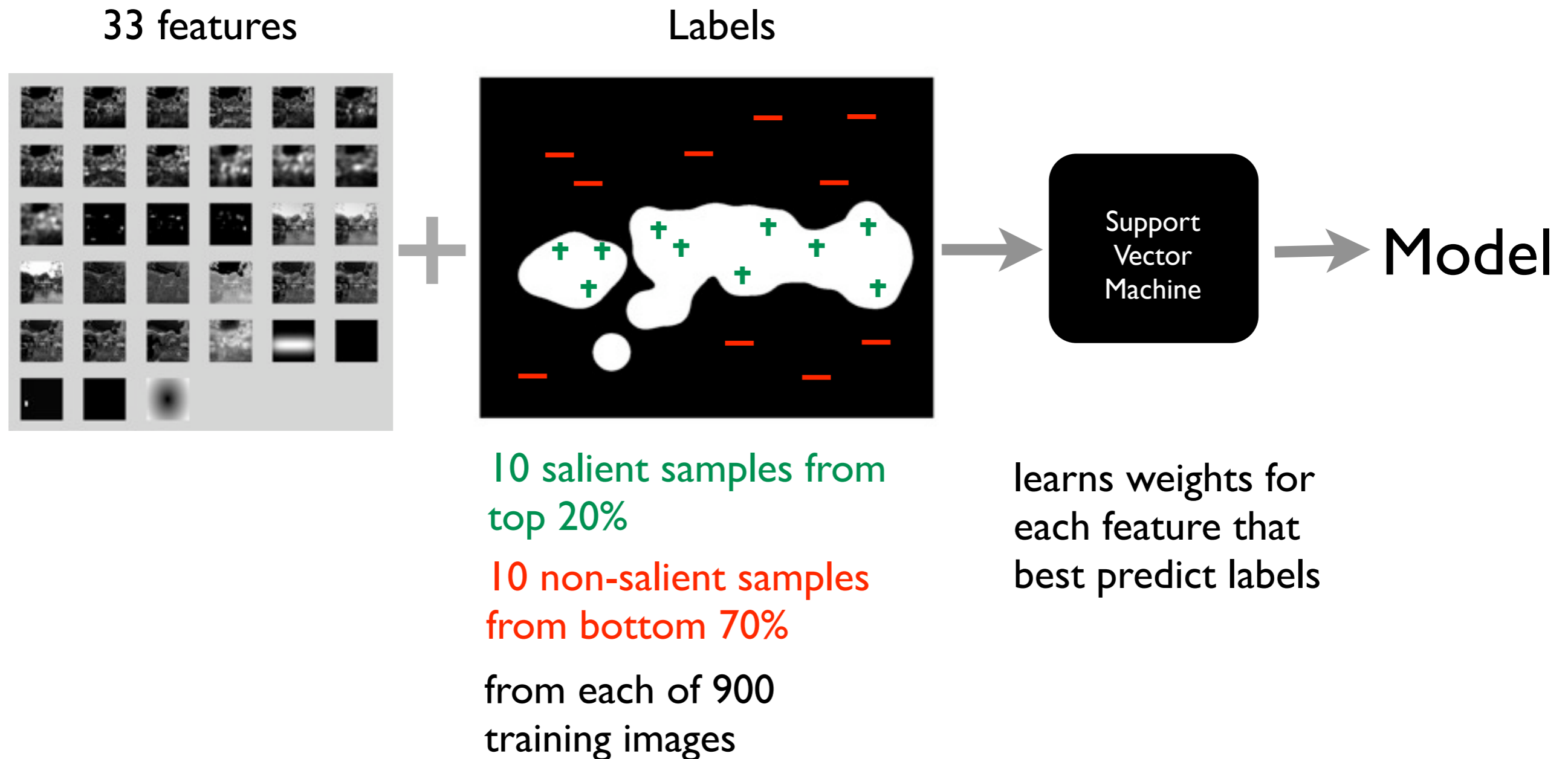
We learn several models of saliency



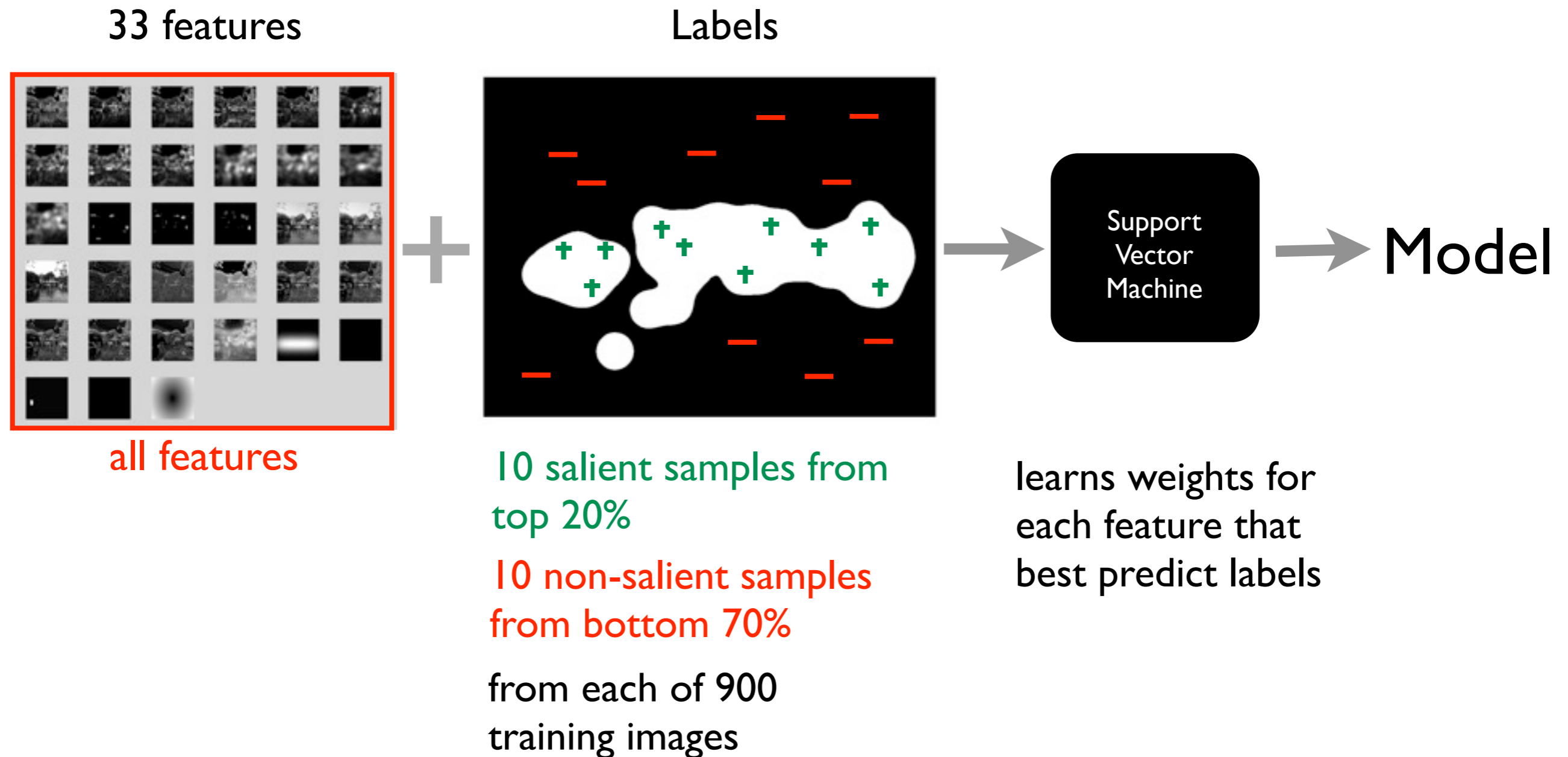
We learn several models of saliency



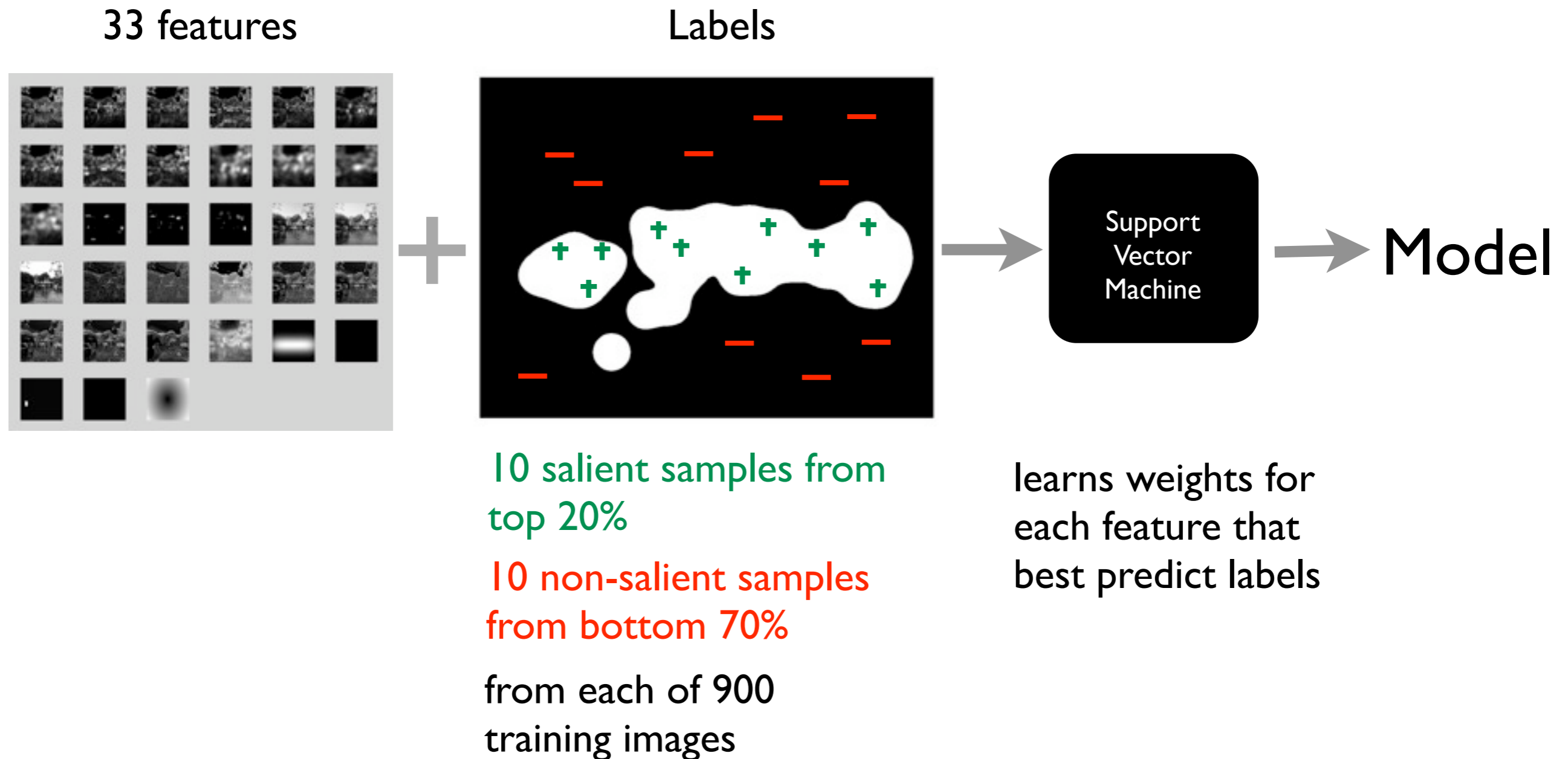
We learn several models of saliency



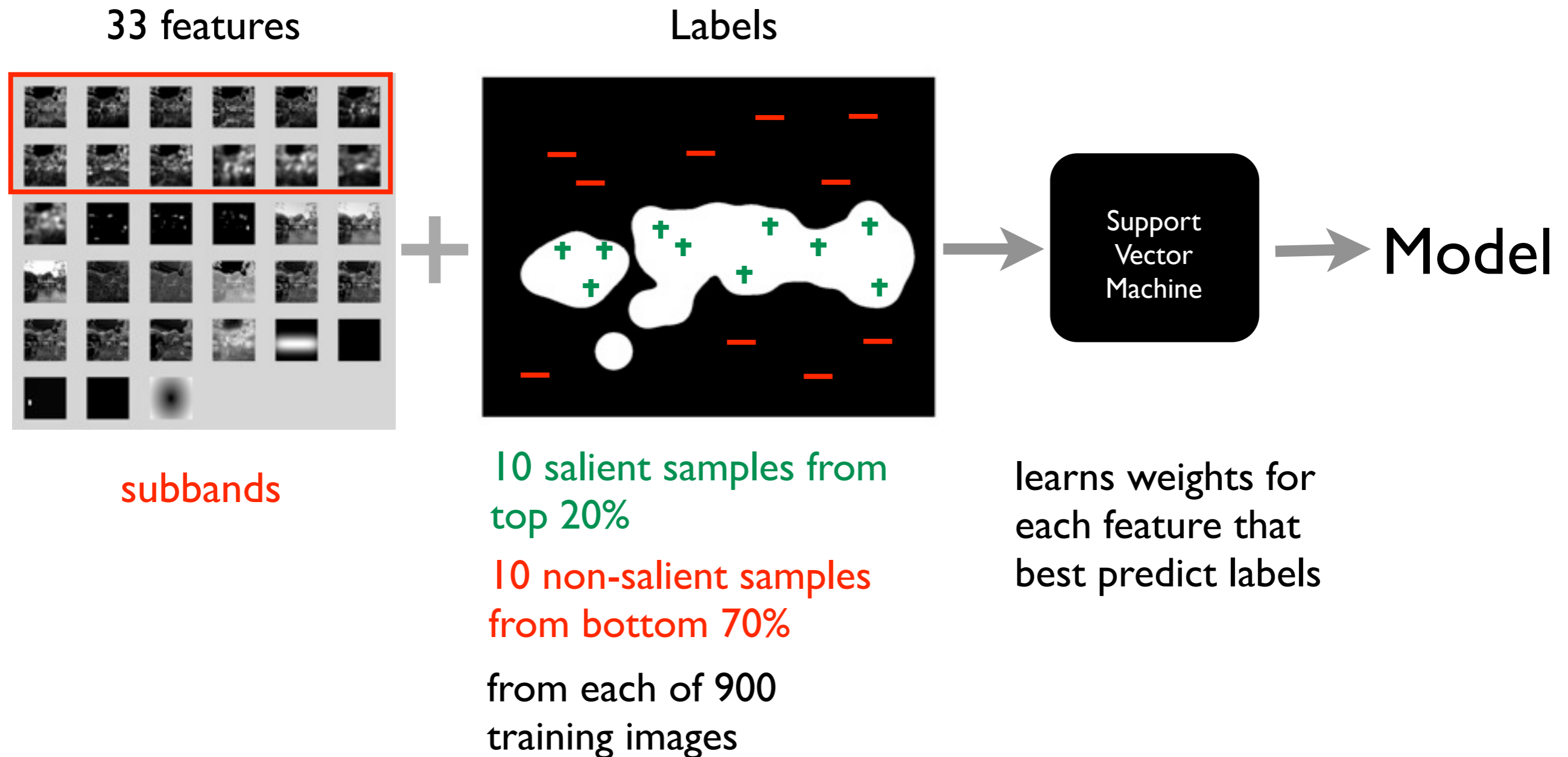
We learn several models of saliency



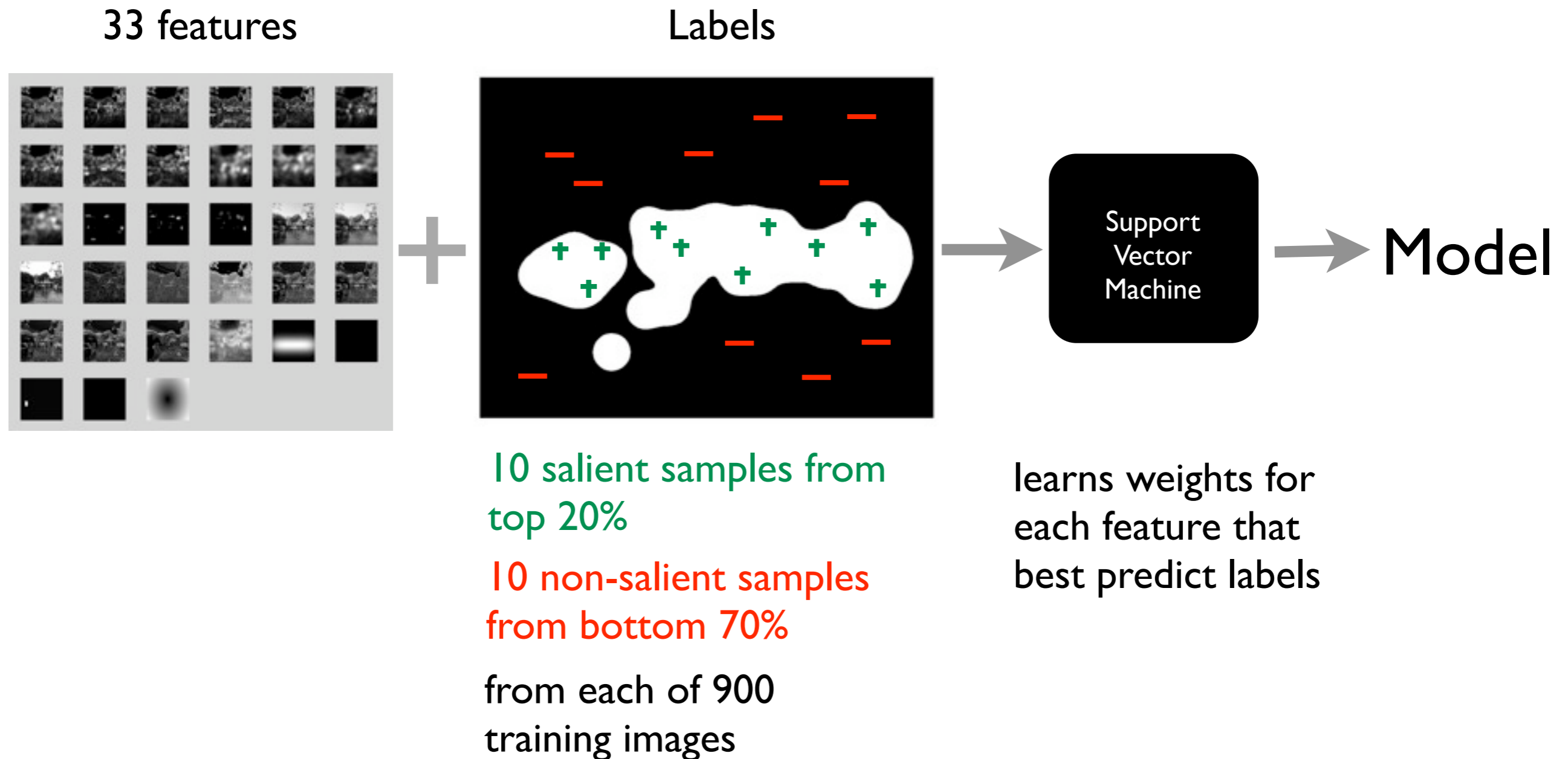
We learn several models of saliency



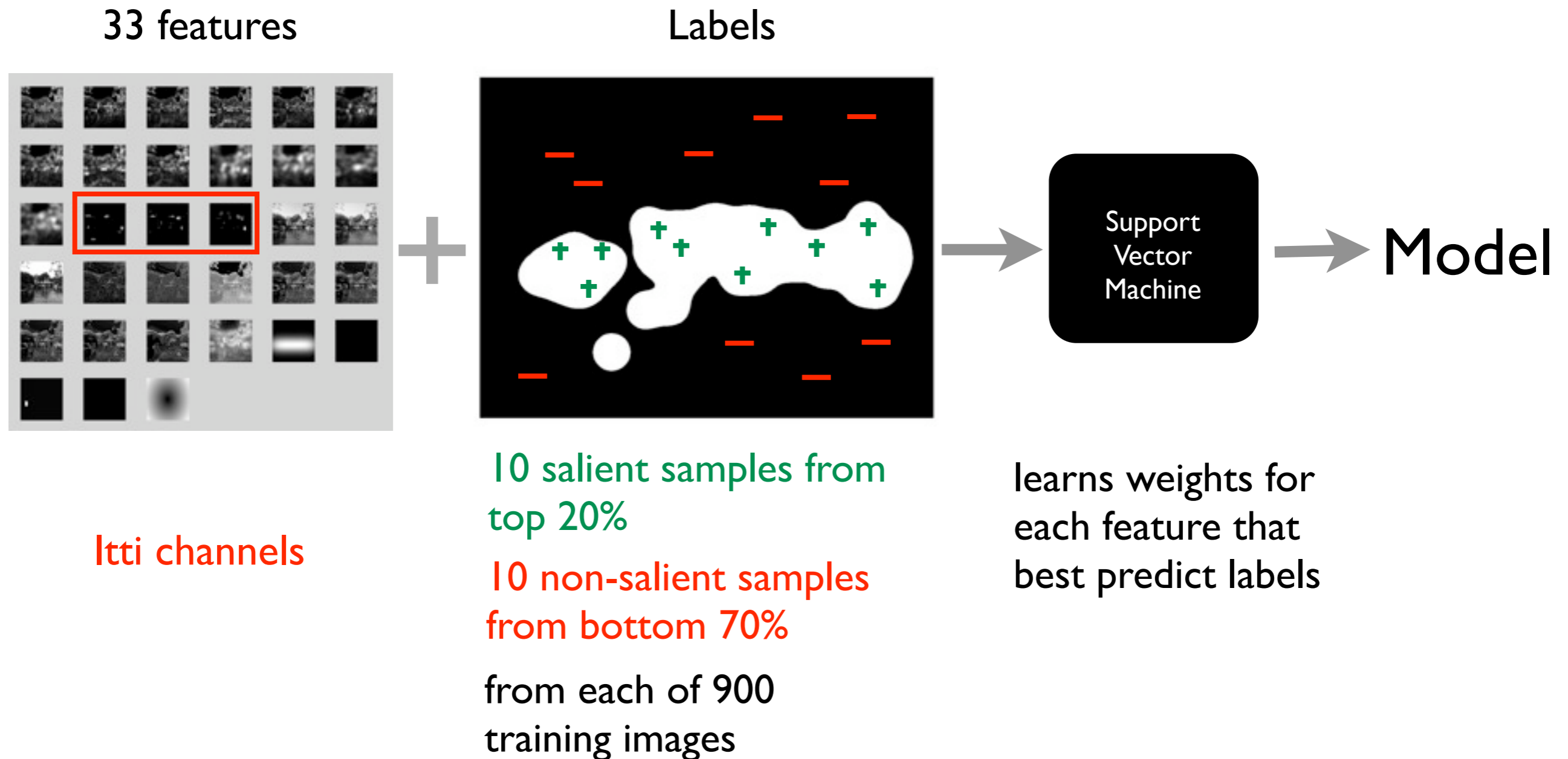
We learn several models of saliency



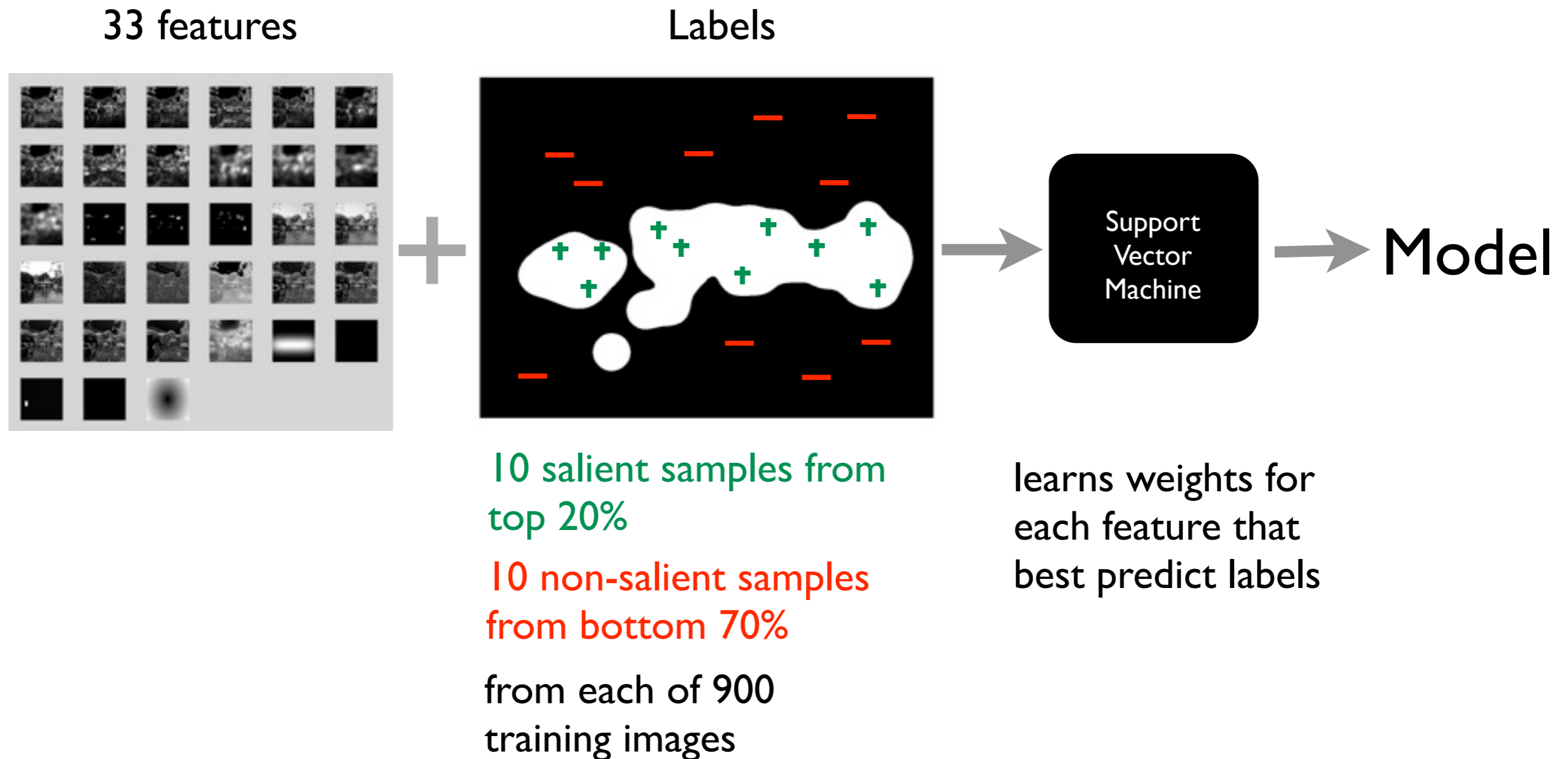
We learn several models of saliency



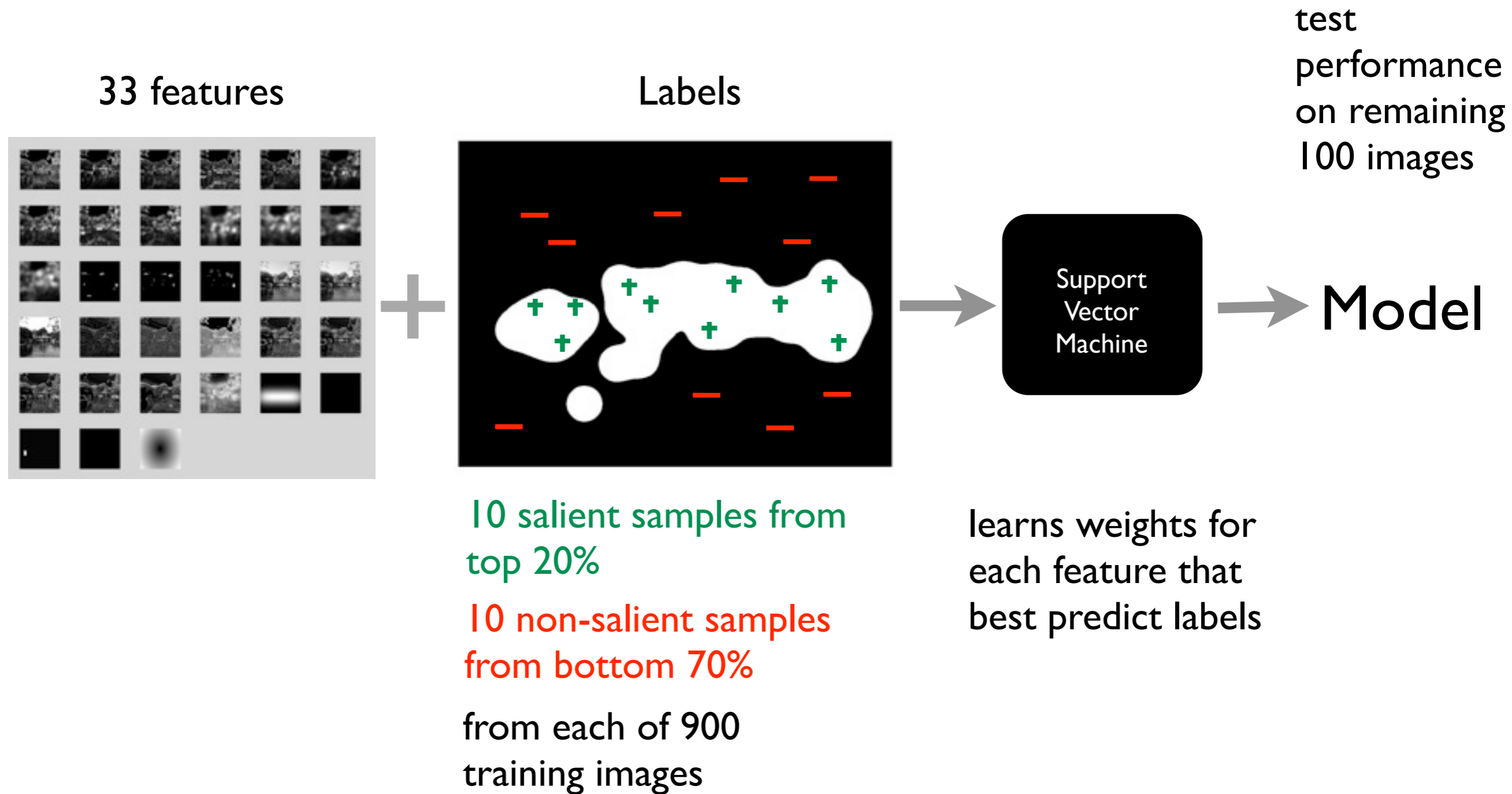
We learn several models of saliency



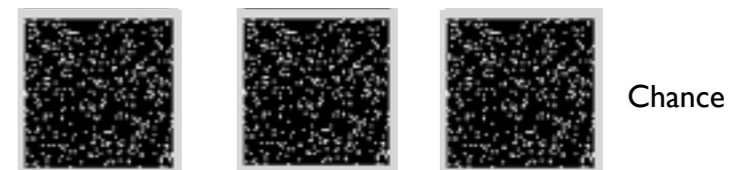
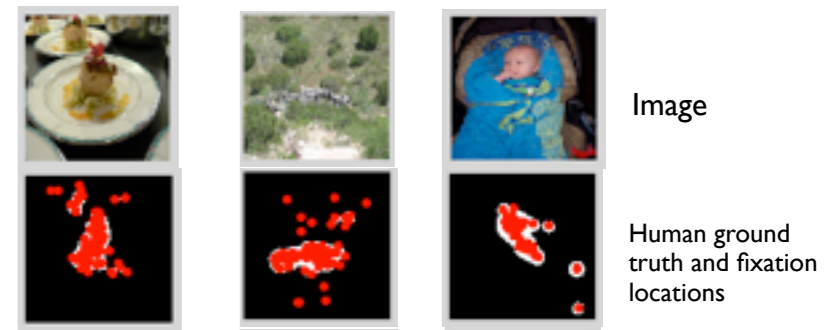
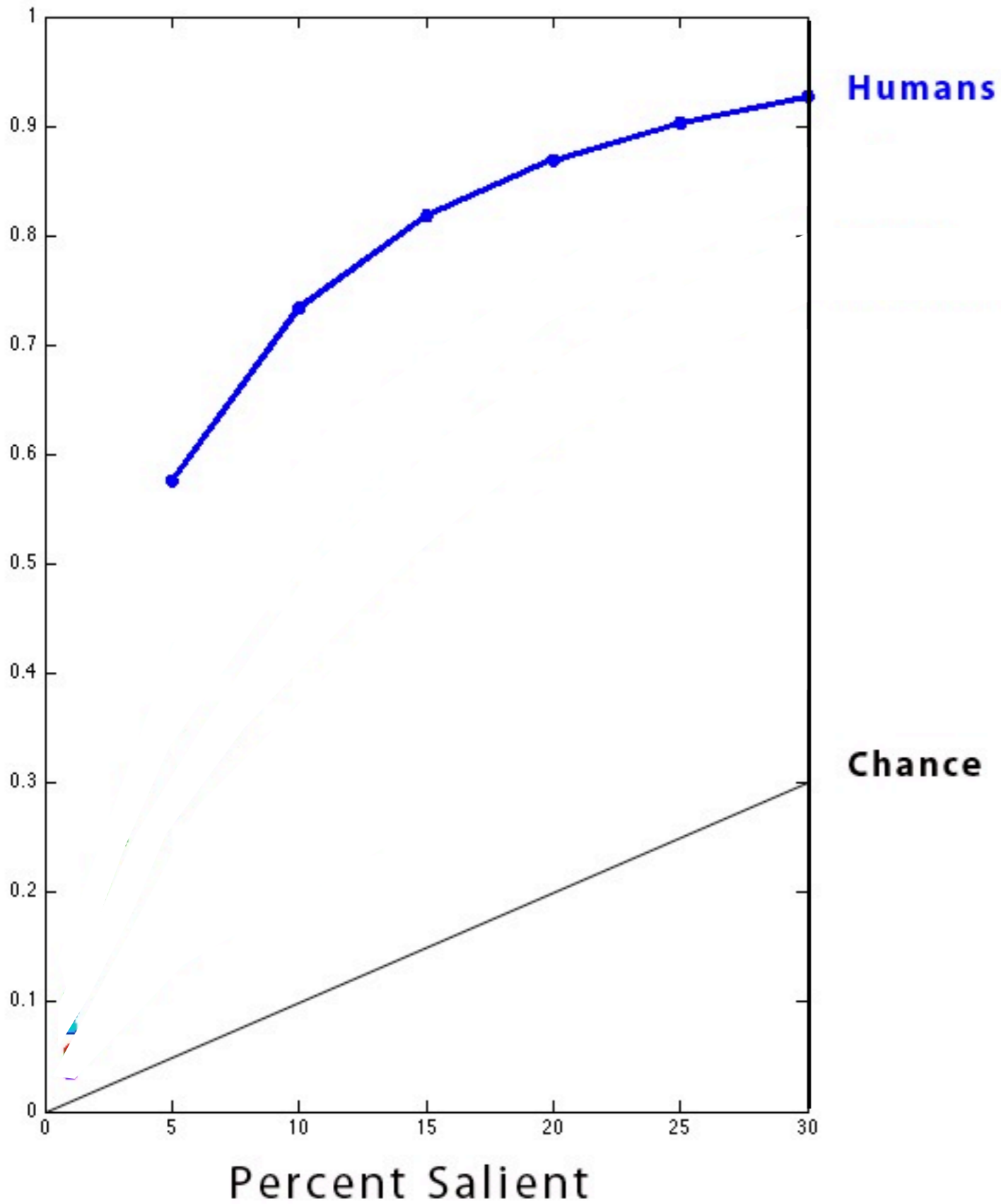
We learn several models of saliency



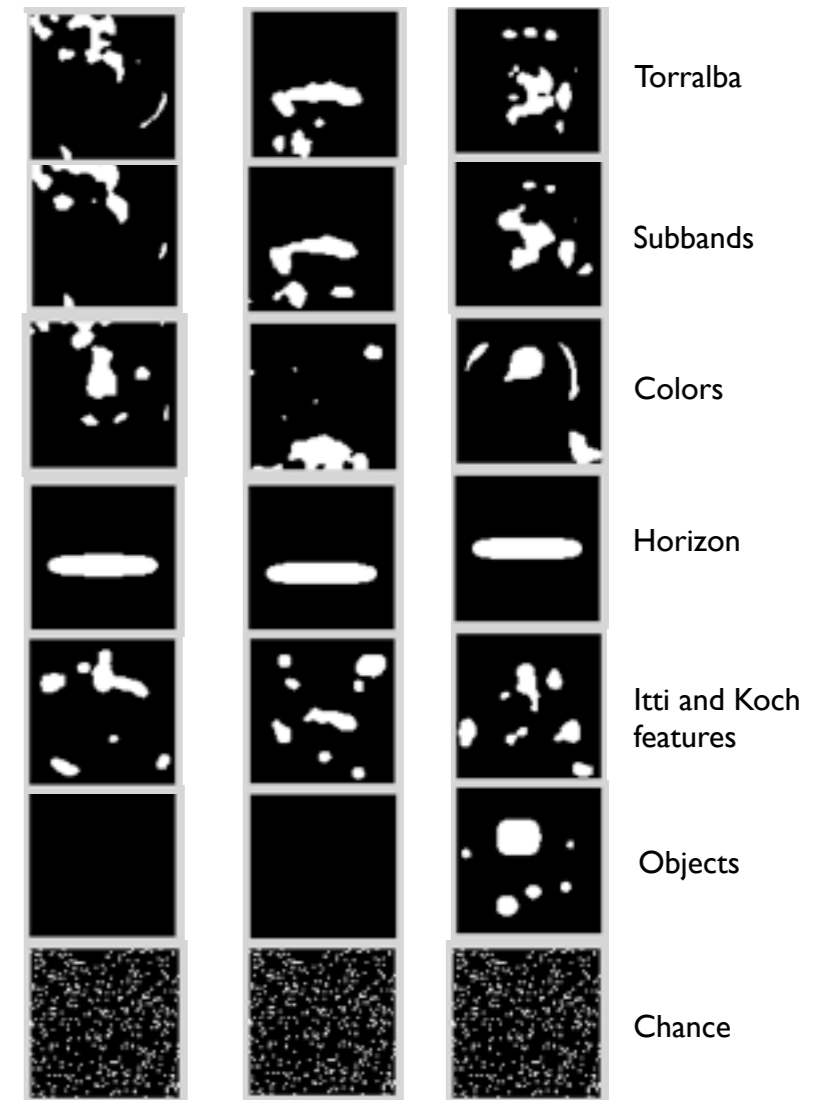
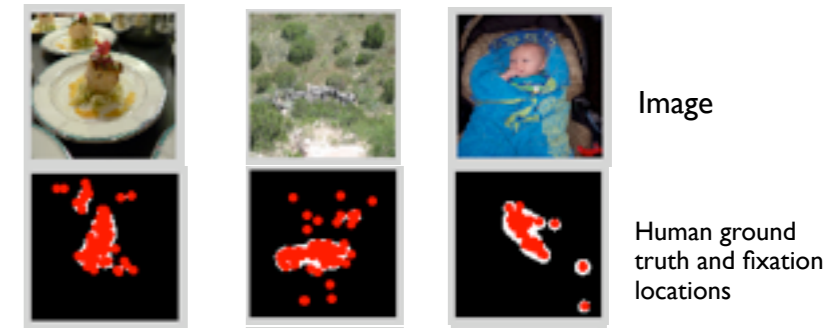
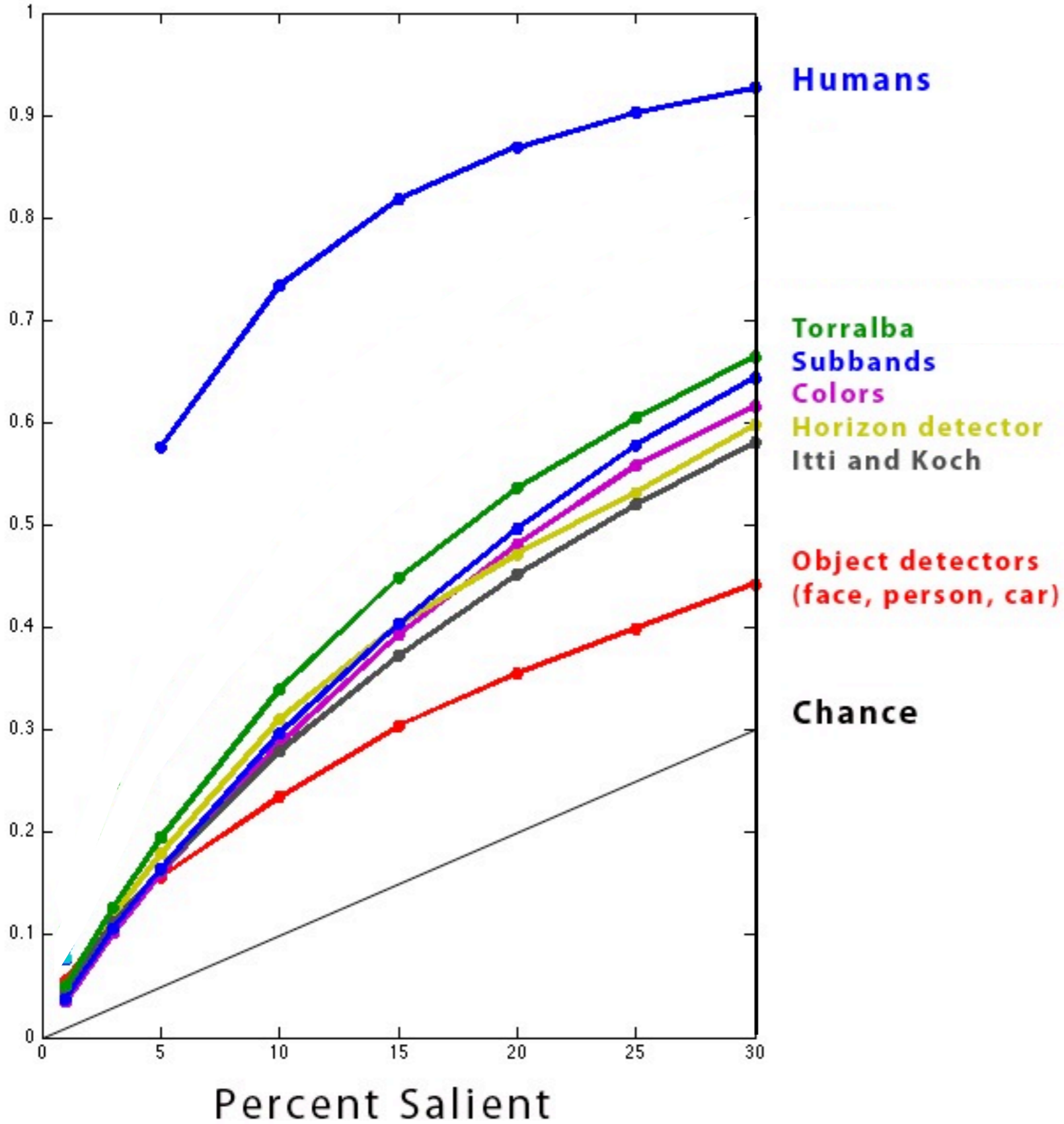
We learn several models of saliency



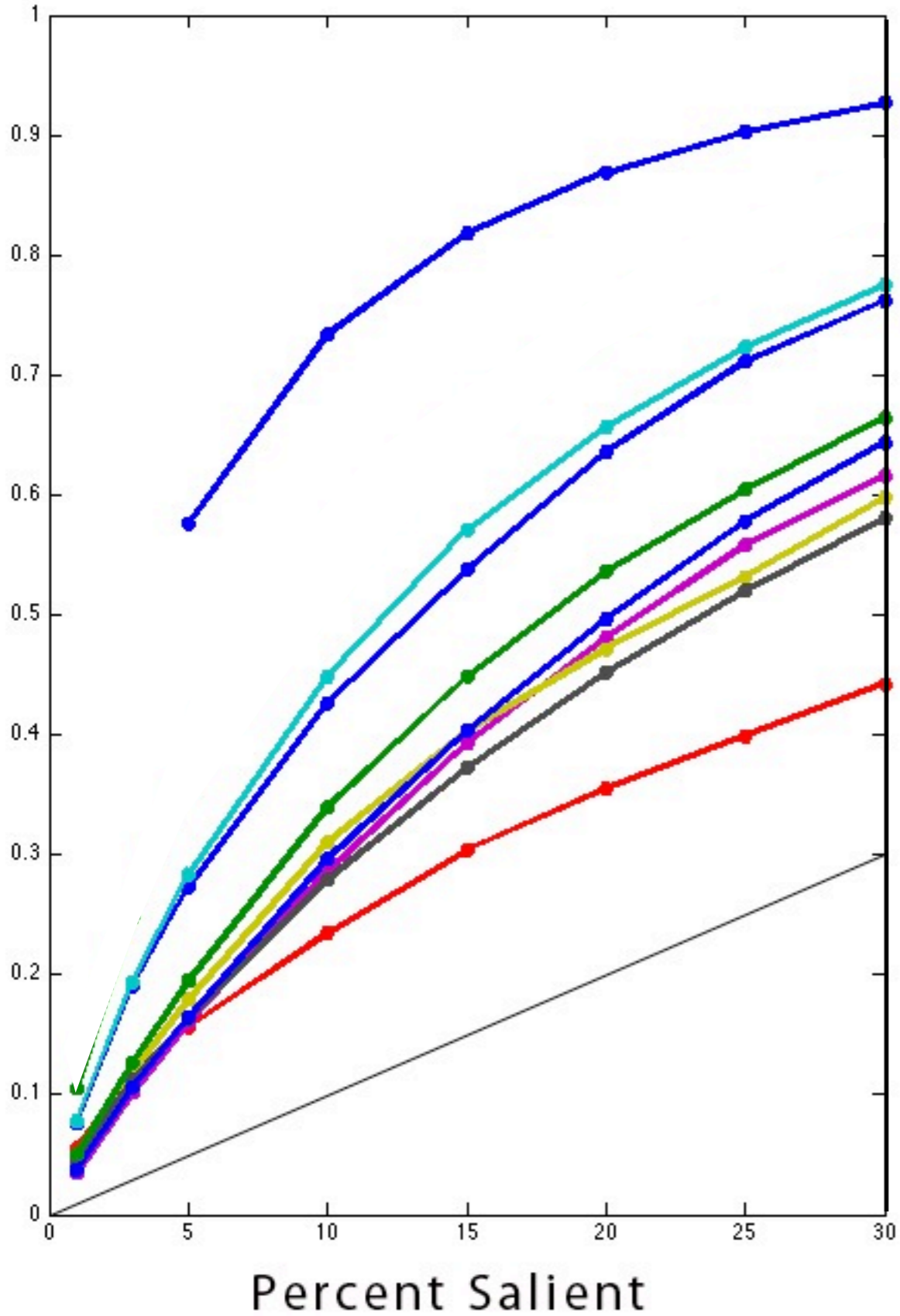
ROC curve shows performance



ROC curve shows performance



ROC curve shows performance



Humans

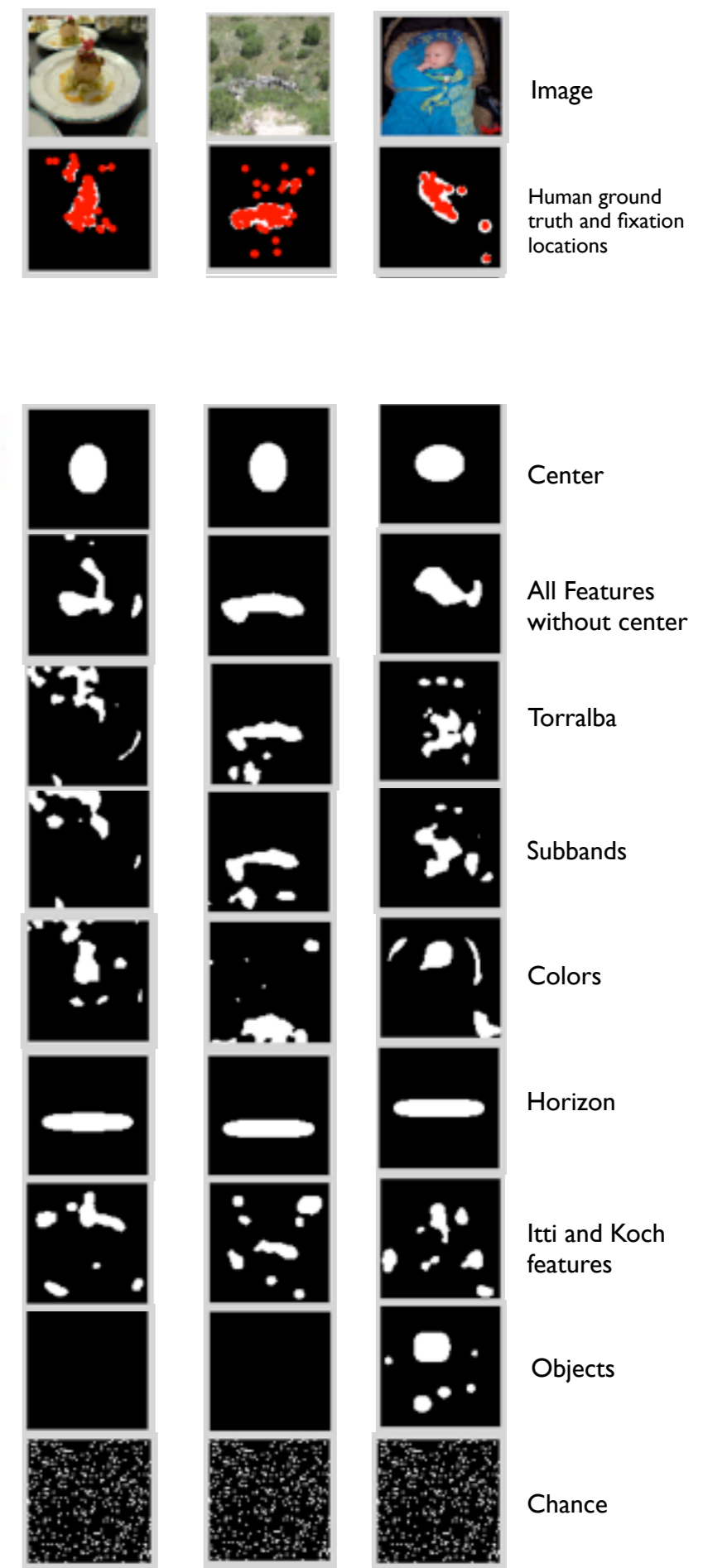
Center
All features without center

Torralba
Subbands
Colors

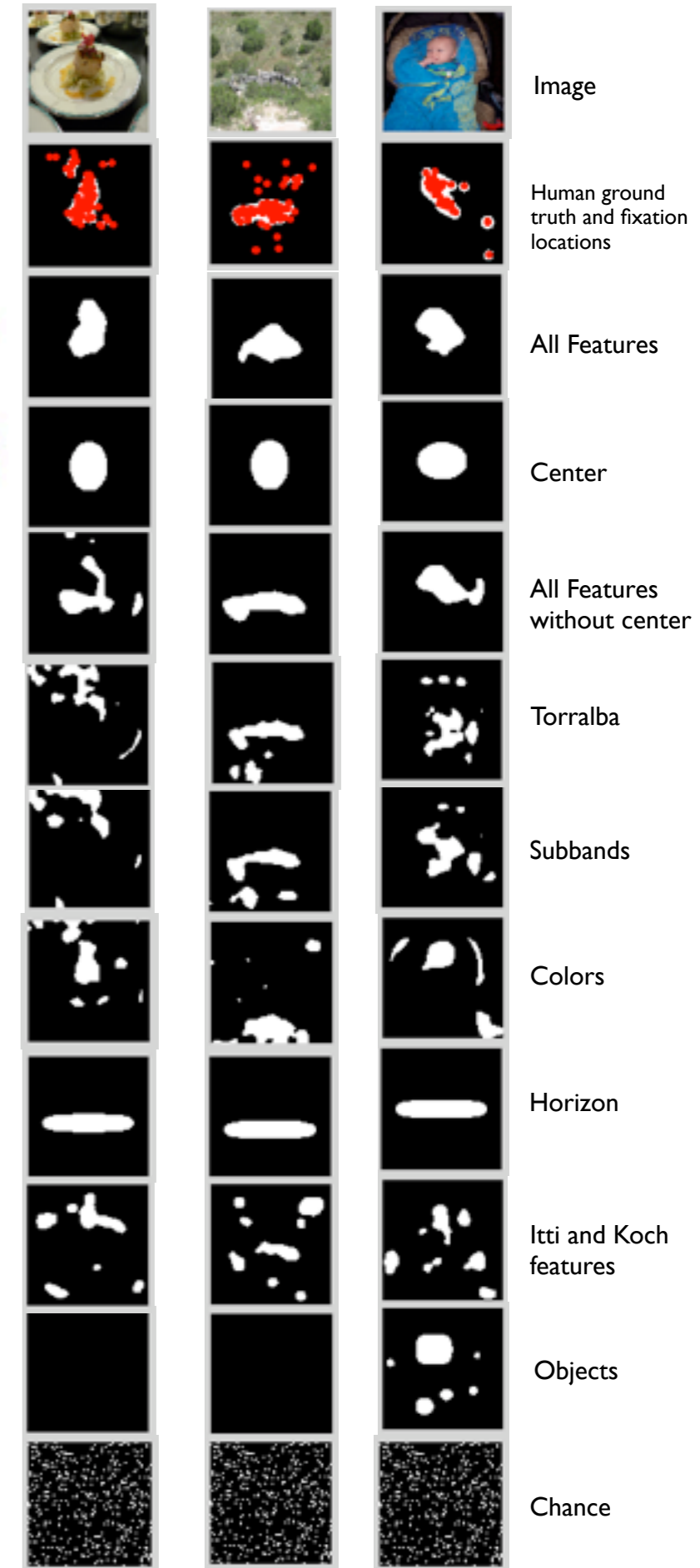
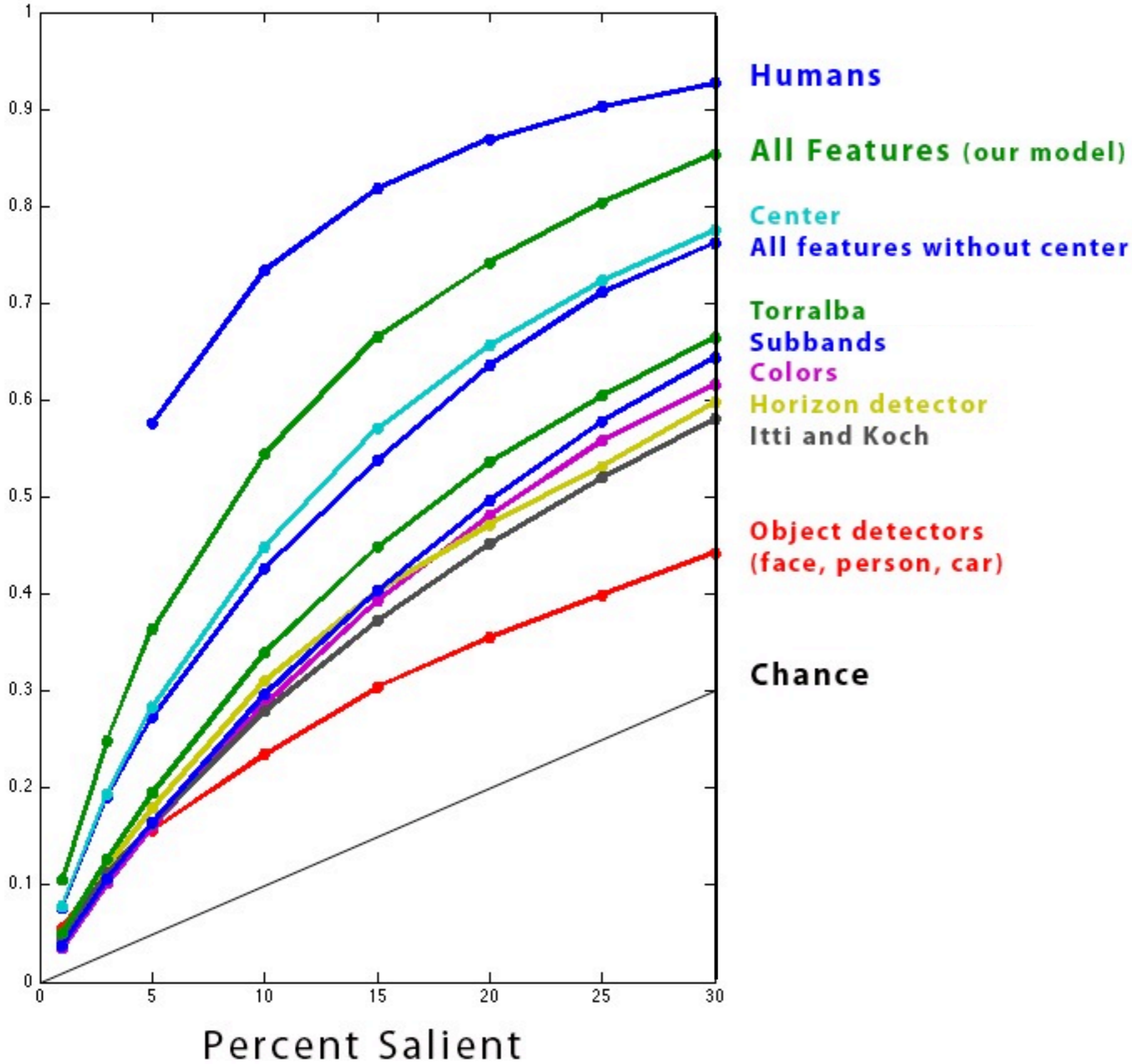
Horizon detector
Itti and Koch

Object detectors
(face, person, car)

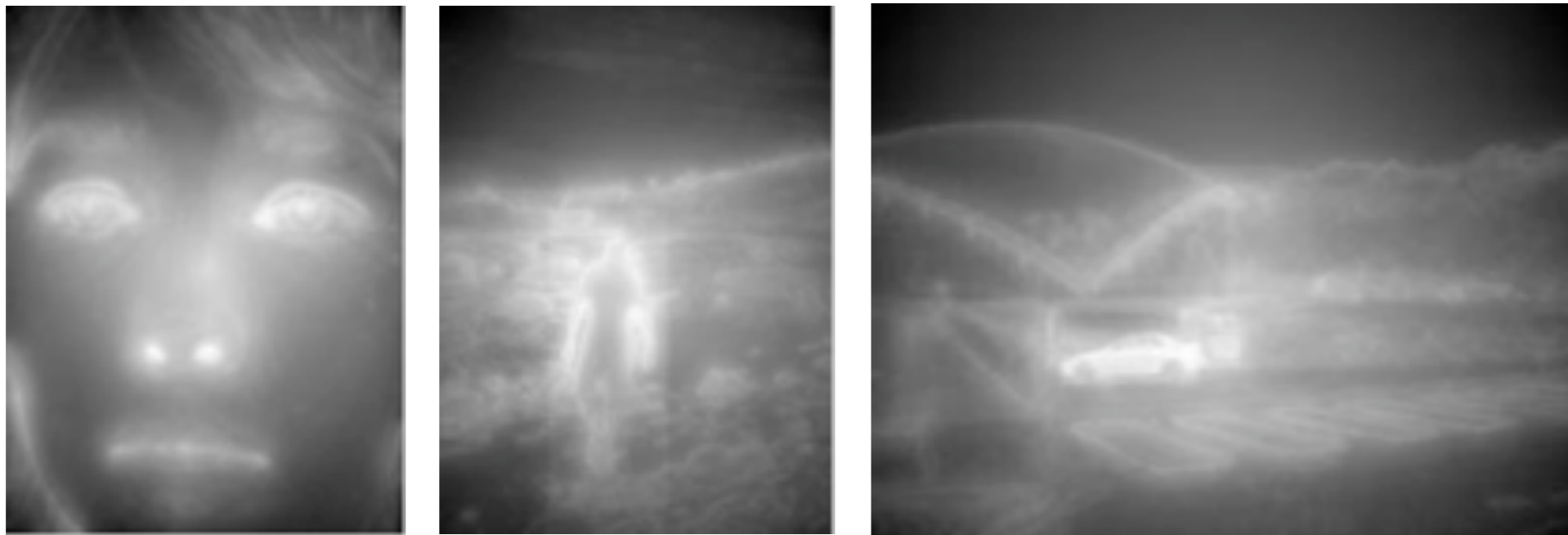
Chance



ROC curve shows performance



Some example saliency maps



This is where our model predicts people will look.



This is where people actually looked.

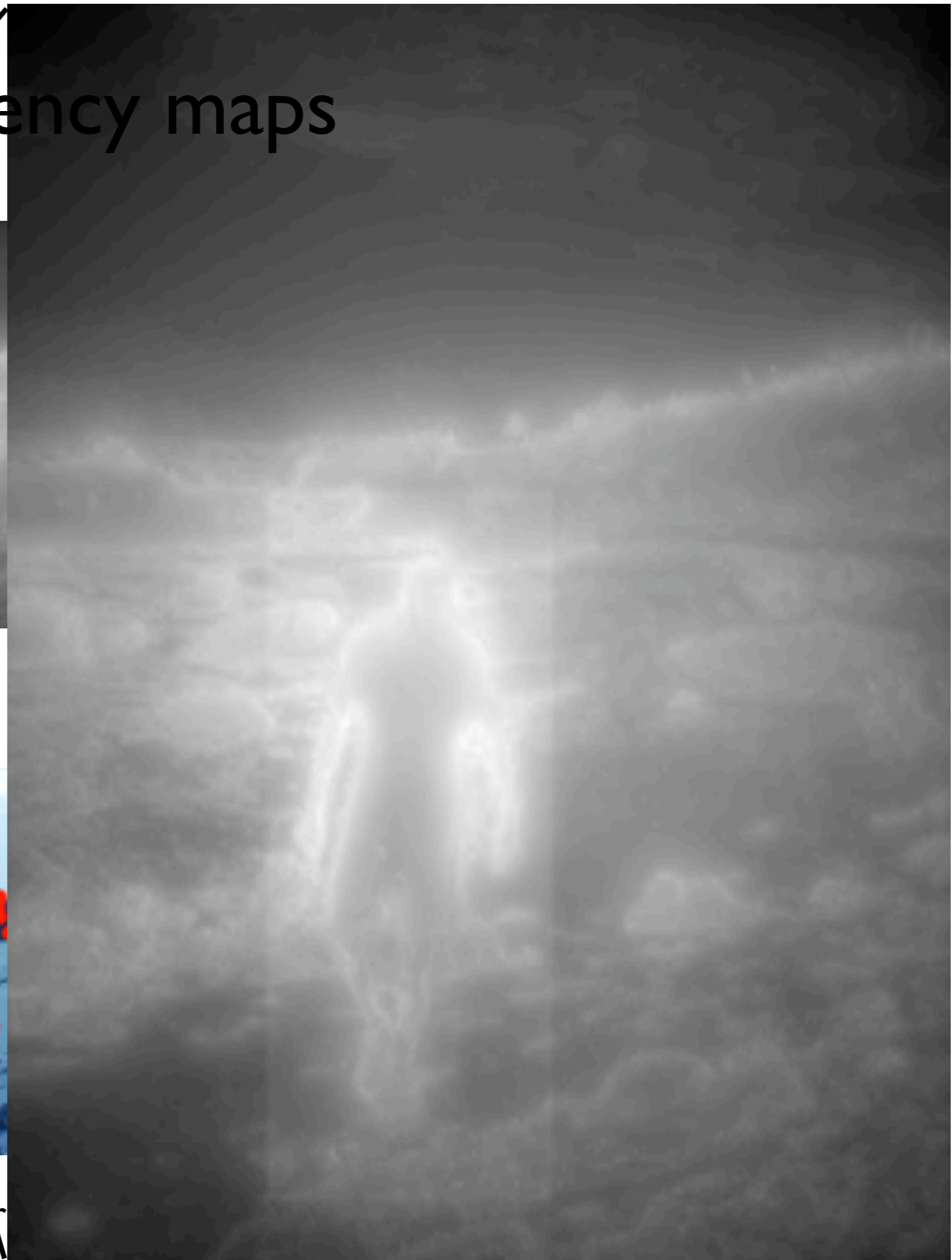
Some example saliency maps



This is where our

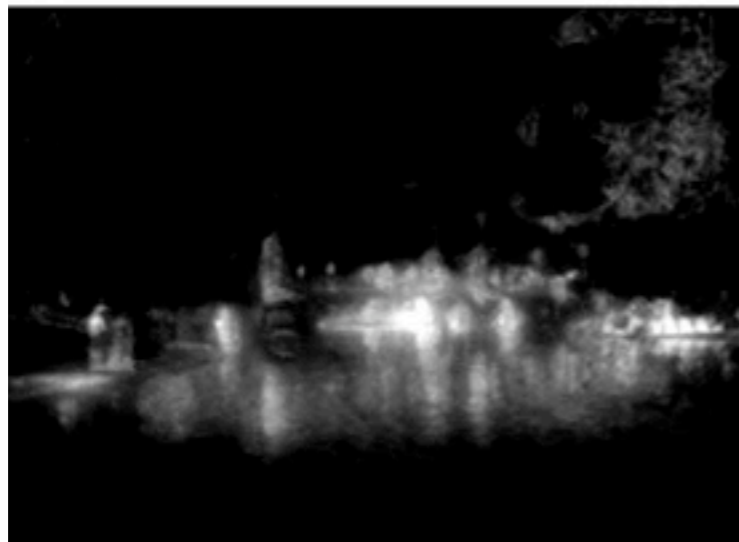


This is where



We demonstrate an application for non-photorealistic rendering

input image



our saliency map



Rendering of image with more detail at salient locations

based on technique by DeCarlo and Santella [2002]

Summary of learning work

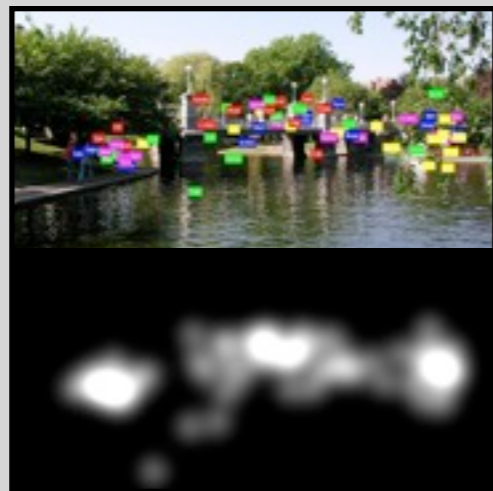
- **Have large collection of eye tracking data**
- **Learned a model of saliency**

Weights learned from ground truth data, not tuned
Found center feature to be surprisingly high -- strongest feature
Subbands and Color features next strongest. Objects useful
Our model reaches 88% way to human performance
- **Future work**

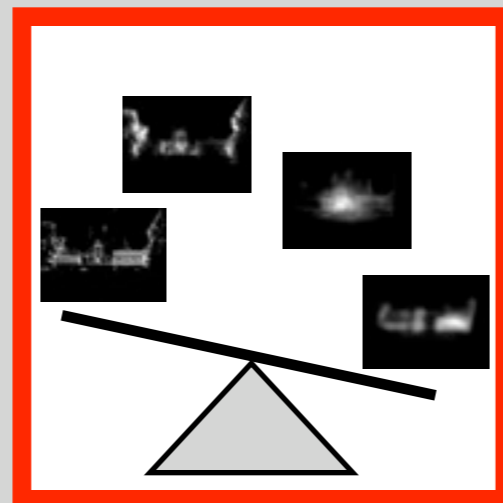
Enhance model by trying other features, text detector

Understanding and predicting where people look

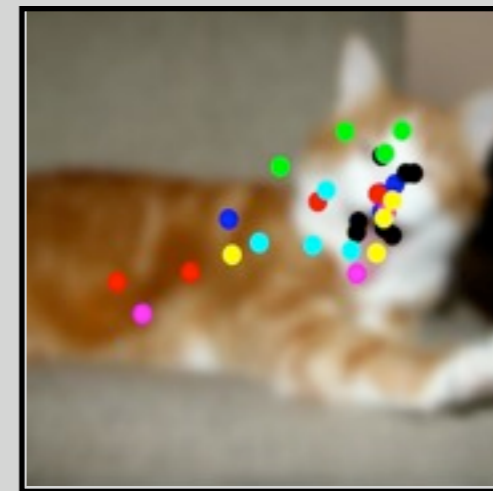
Introduction



Learning a model



Benchmarking models of saliency



Fixations on low-resolution images

Conclusion

Many saliency models, hard to compare them

- **Each model introduced with different goal**
 - mimic the human visual system
 - correlate with human fixations
 - segment a salient object from the background
 - tell the story of the image
- **Each evaluated on different data sets**
 - possibly few images, few observers
 - compared against a few (1-3) other models
- **Use different metrics of evaluation**
 - ROC curve, Similarity, Earth movers distance, etc
 - Ability to align with human annotated bounding boxes

We contribute a benchmark

- **Compare 10 modern models**
all have code available online
made in last 5 years
- **Evaluated on *ability to predict human fixations***
on new data set
300 images, 20 observers
- **Use 3 metrics of evaluation**
ROC curve, Similarity, Earth movers distance

Benchmark data set: 300 images seen by 20 people



[MIT benchmark data set]

Benchmark 10 different saliency models



Original Image



Human Fixations



Human Fixation Map



Achanta



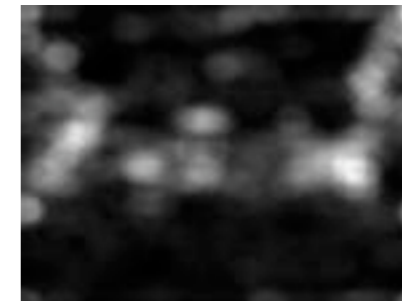
Bruce & Tsotsos



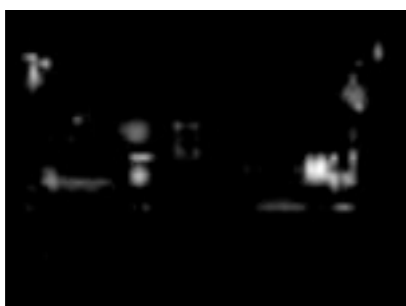
Context Aware



GBVS



Hao&Zhang



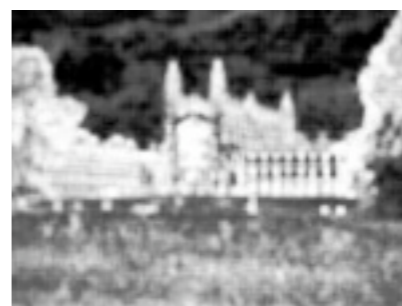
Itti & Koch



Itti & Koch2



Torralba



SUN



Judd

Equalize the amount of salient pixels in each map via histogram matching



Original Image



Human Fixations

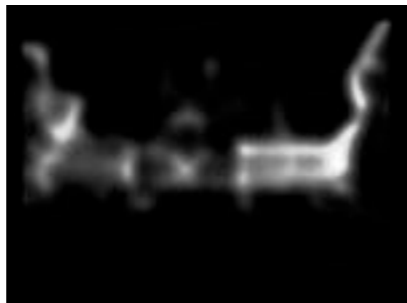


Human Fixation Map

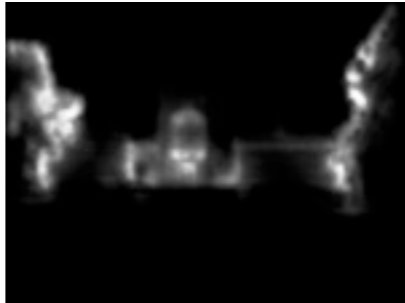
This is for visualization purposes only-- it does not affect ROC performance



Achanta



Bruce & Tsotsos



Context Aware



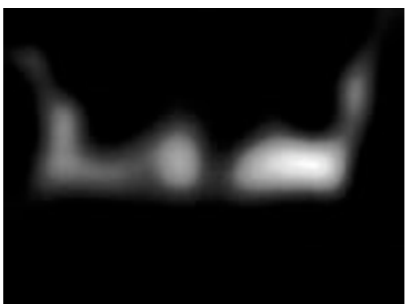
GBVS



Hao&Zhang



Itti & Koch



Itti & Koch2



Torralba



SUN



Judd

We compare to the performance of baselines



Original Image



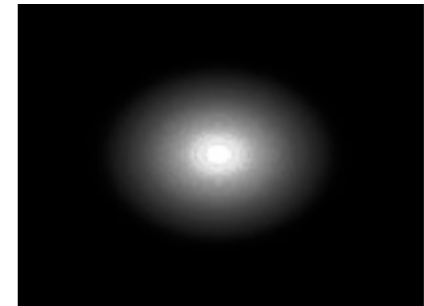
Human Fixations



Human Fixation Map



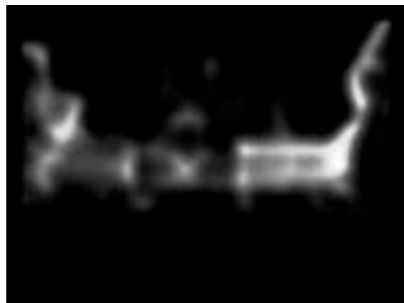
Chance baseline



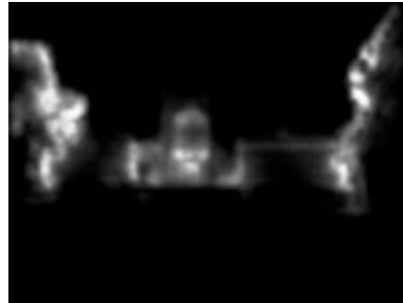
Center baseline



Achanta



Bruce & Tsotsos



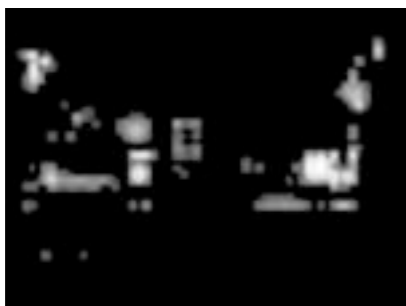
Context Aware



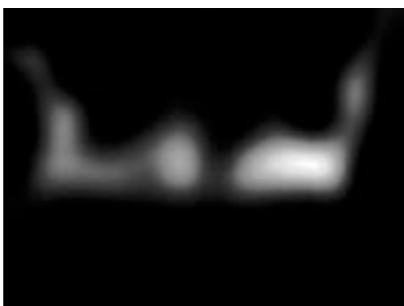
GBVS



Hao&Zhang



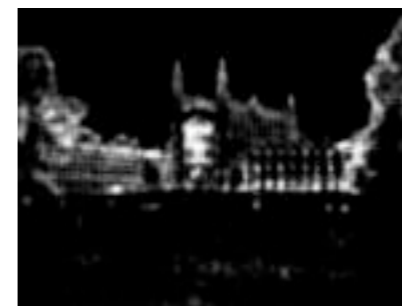
Itti & Koch



Itti & Koch2



Torralba



SUN

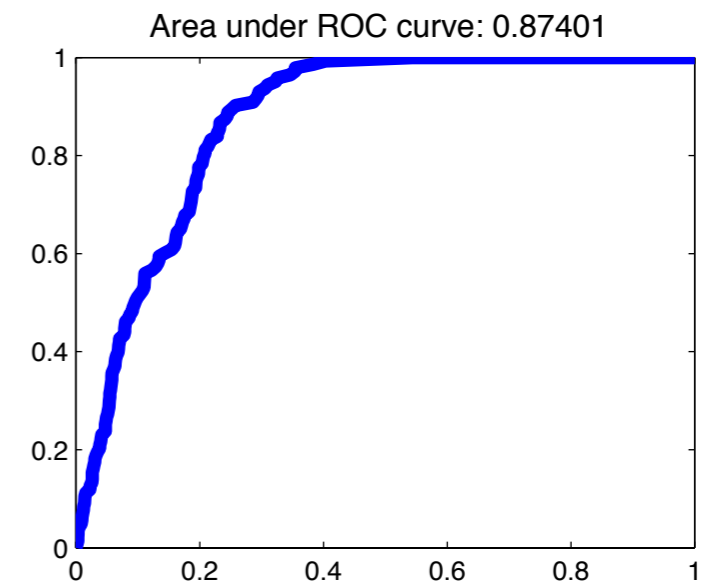
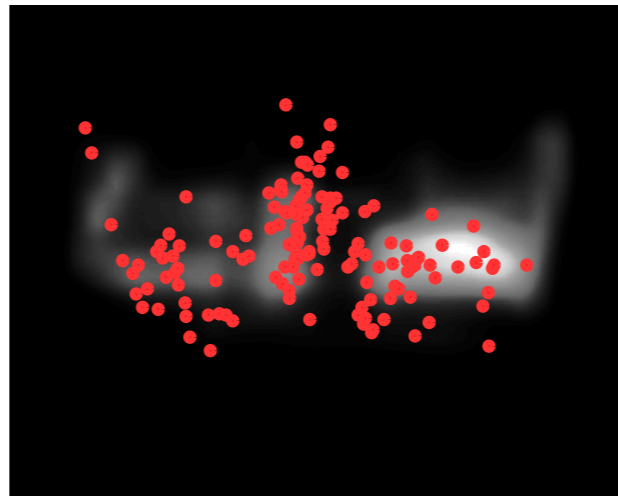


Judd

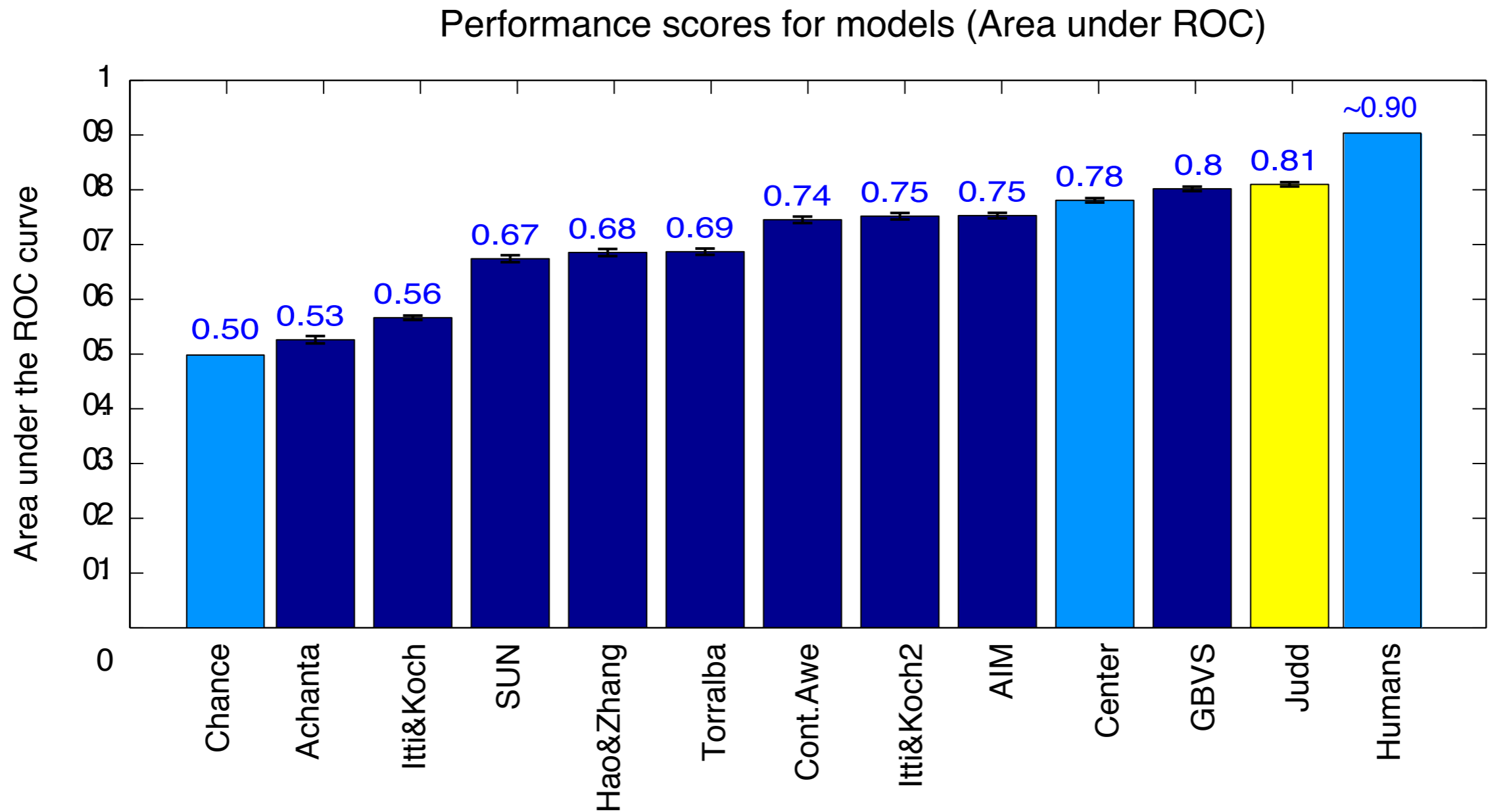
Performance of each model measured by the area under the ROC curve



SaliencyMap with fixations to be predicted



Performance of each model measured by the area under the ROC curve



Models perform better than other if

- they have better features
- they have more center bias
- model is blurrier

Increasing blurriness improves performance for some models

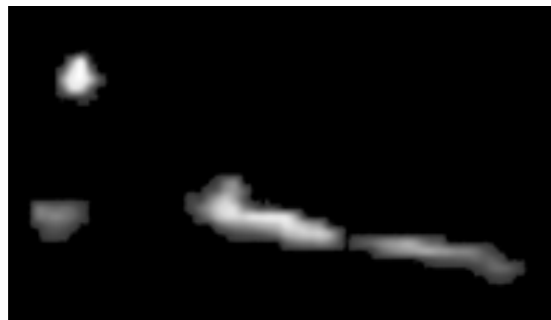
Image



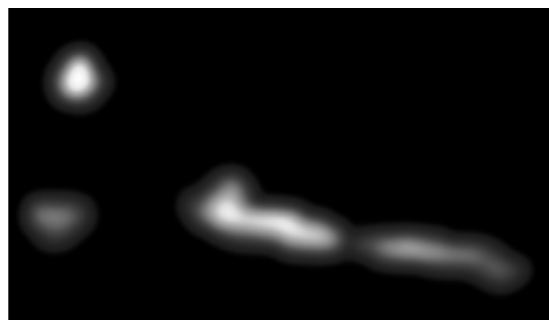
Fixations



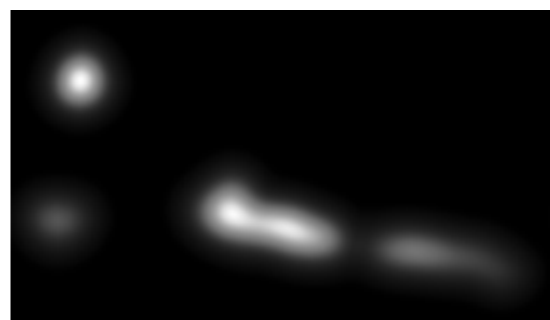
AUR= 0.563



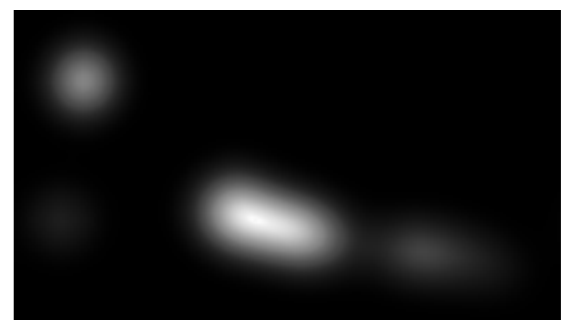
AUR=0.659



AUR=0.721



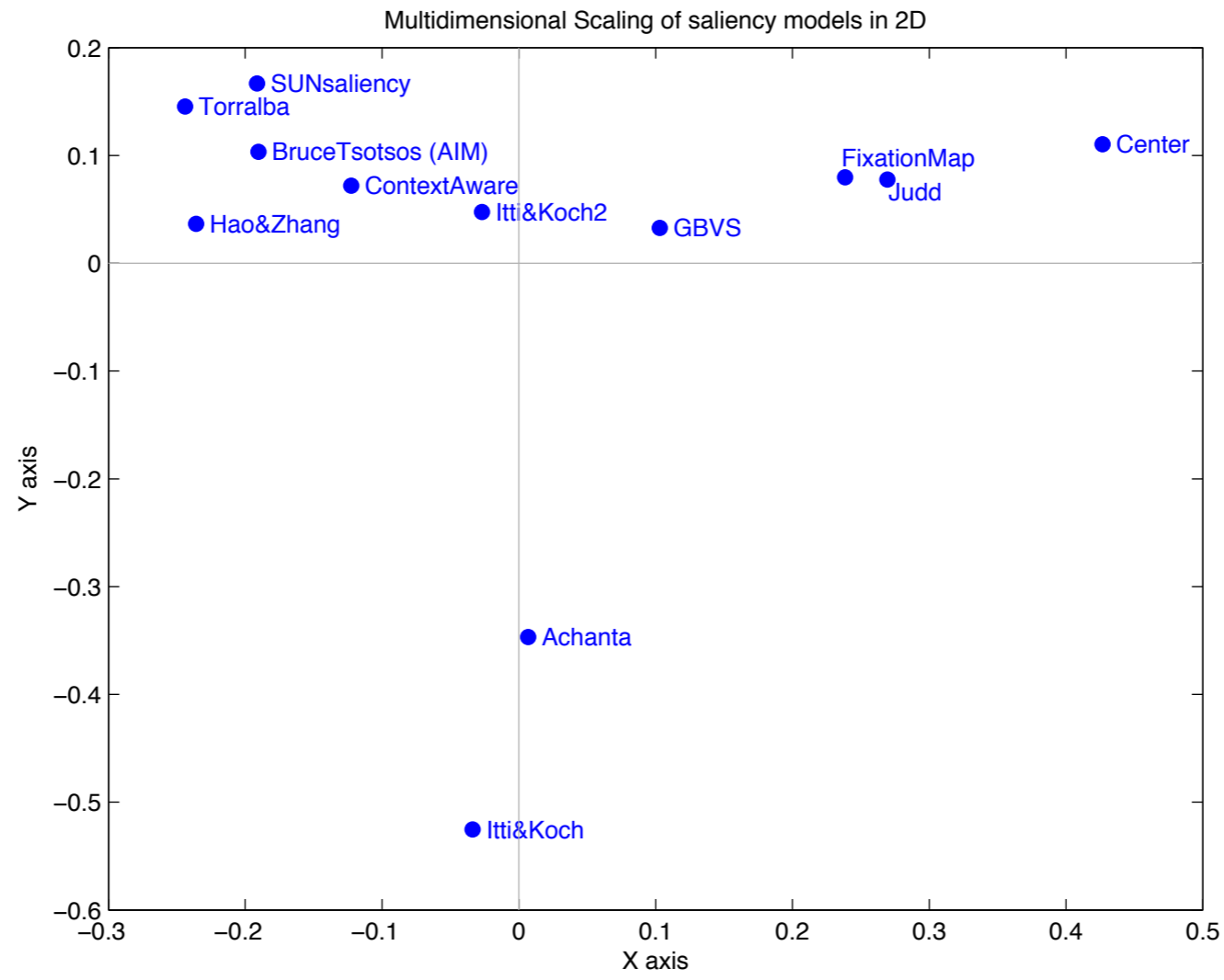
AUR=0.717



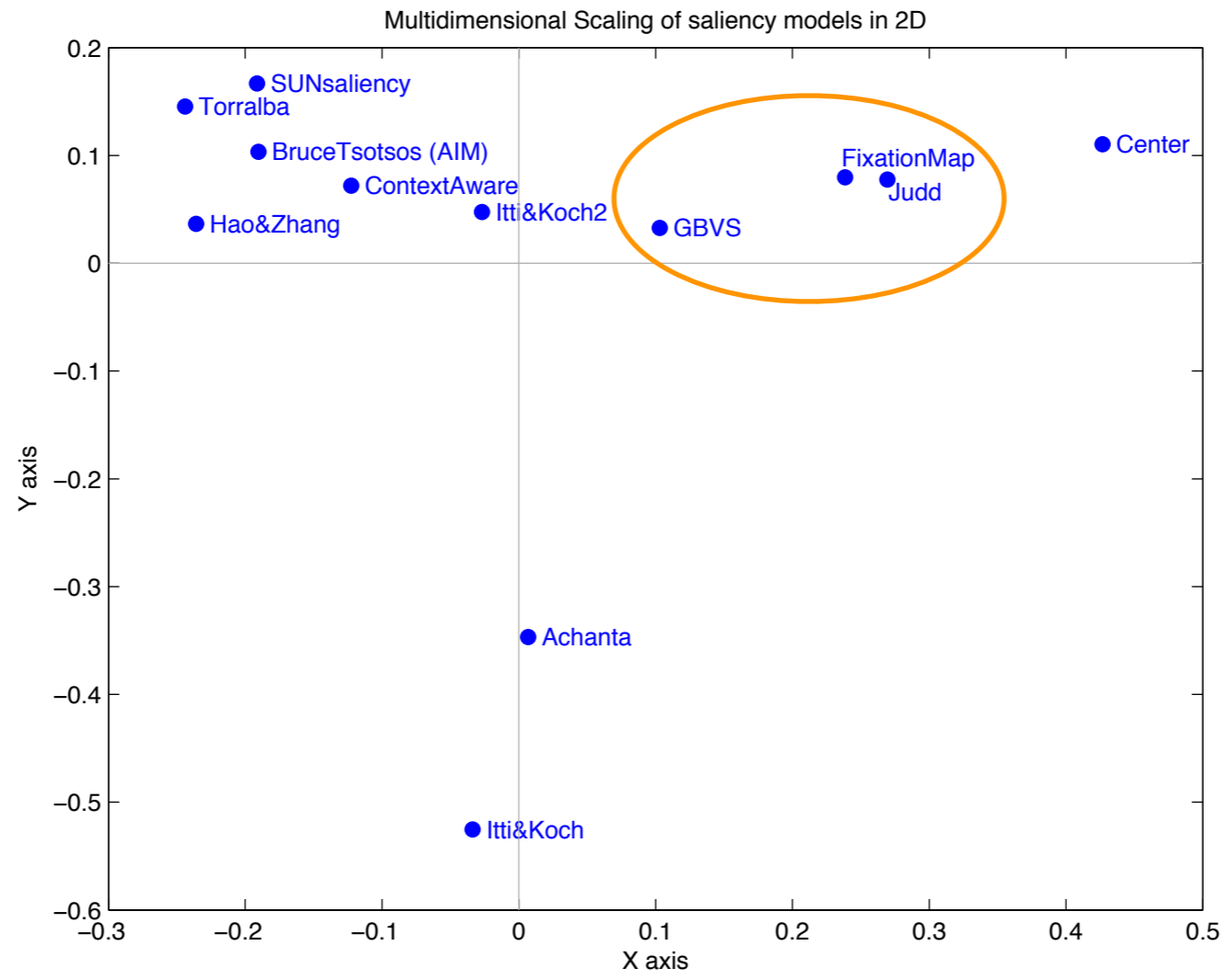
Itti & Koch

increasing blurriness \longrightarrow

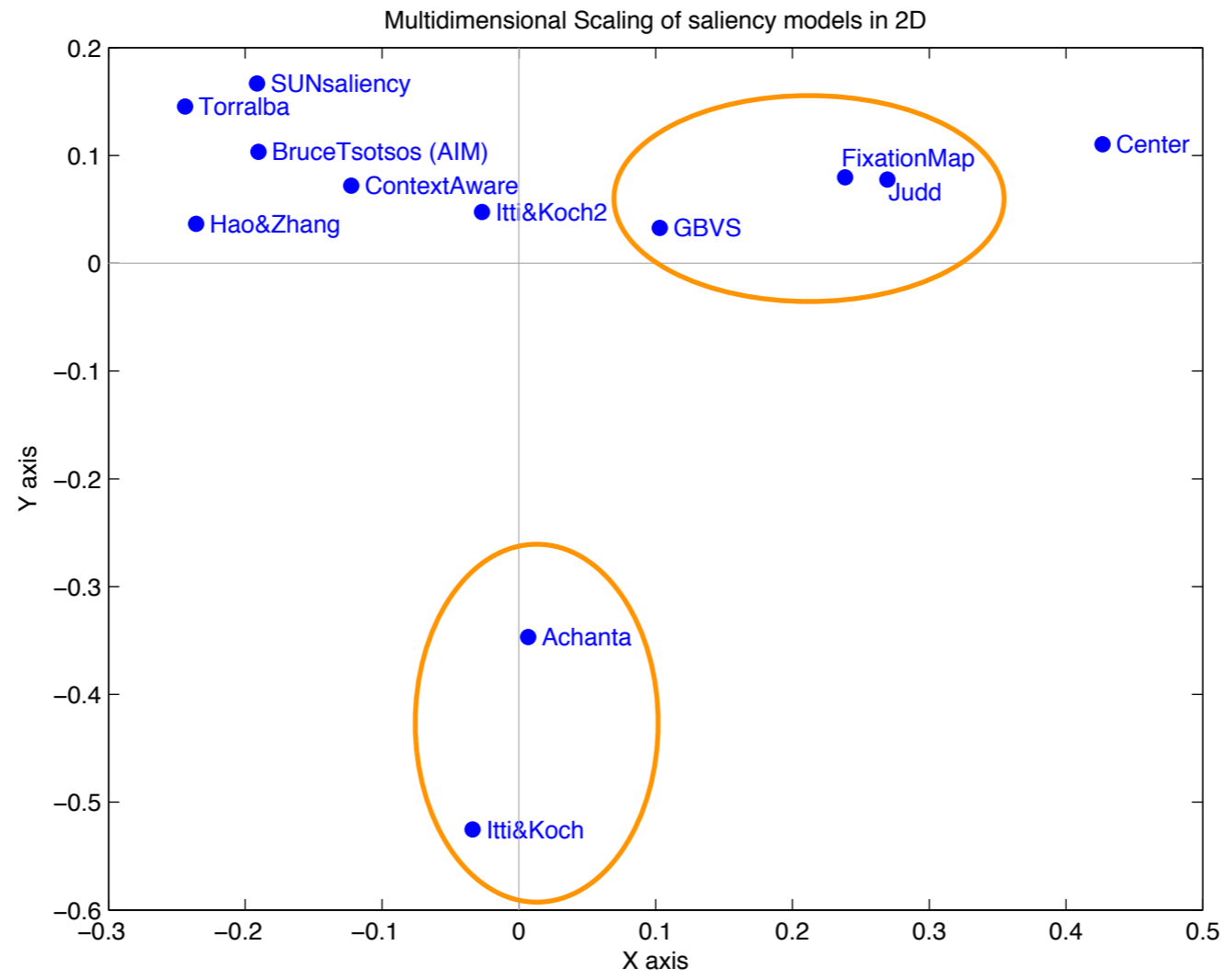
Plotting models in higher dimensional space shows relationships



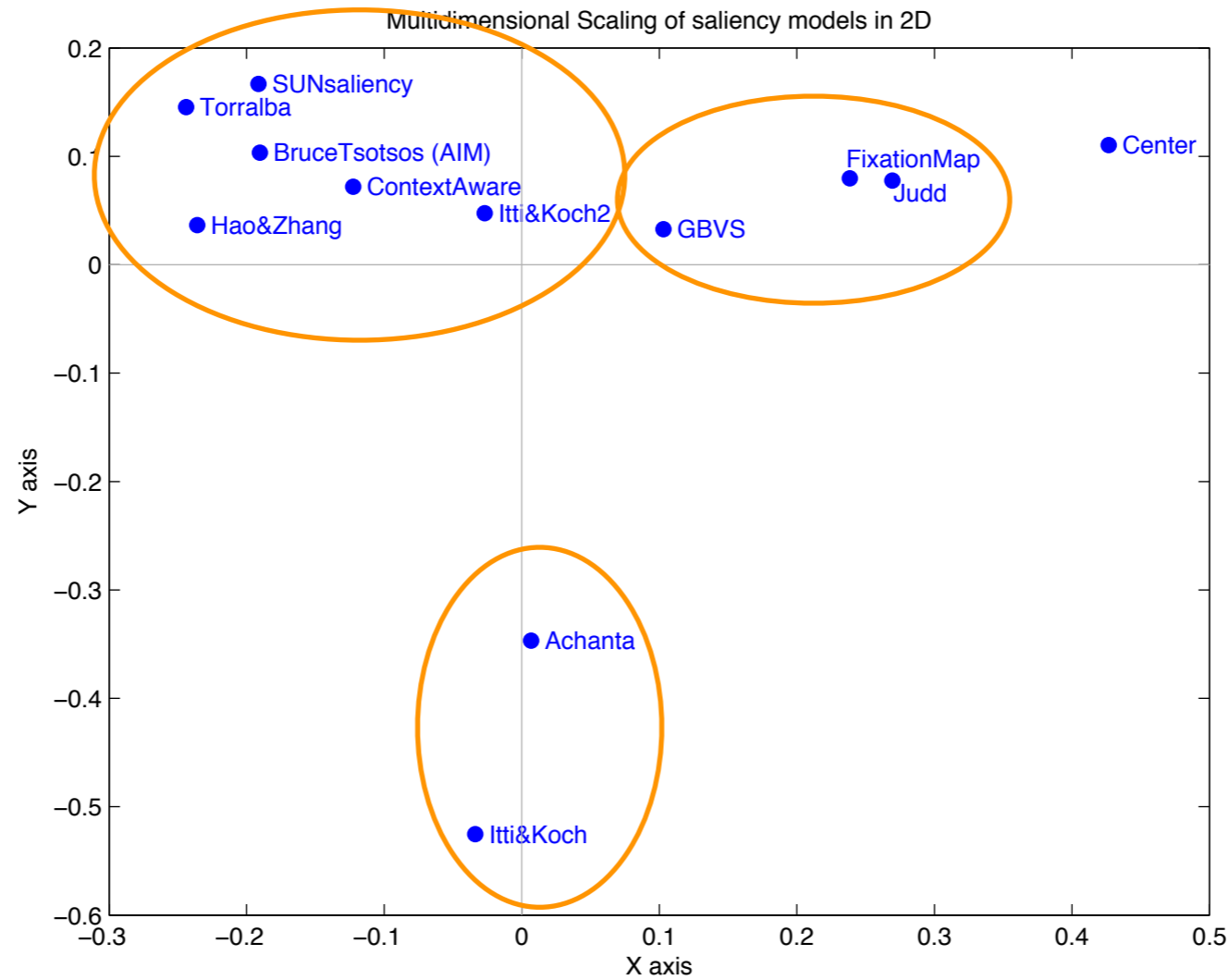
Plotting models in higher dimensional space shows relationships



Plotting models in higher dimensional space shows relationships



Plotting models in higher dimensional space shows relationships



We provide an online benchmark to evaluate future models

saliency benchmark

Which model of saliency best predicts where people look?

There are many computational models of visual attention created from a wide variety of different approaches that aim to predict where people look in images. Since each model is introduced by demonstrating performances on new images, we can not make immediate objective comparisons between the models. To alleviate this problem, we propose a benchmark data set, containing 300 natural images with eye tracking data from 20 viewers, to compare the performance of many available models. For each model of saliency, we calculate each model's performance at predicting ground truth fixations using three different metrics: a receiver operating characteristic, a similarity metric, and the Earth Mover's Distance and post the results here.

images

put [comparison](#) images and maps in.

model performances

Model Name	Related paper	Link to code	Area under ROC curve	Similarity	Earth mover's distance
Humans*		code	0.90	~0.65	~1.55
Judd		code	0.78	0.466	1.404
Graph Based Visual Saliency (GBVS)		code	0.772	0.447	1.455
Center*		code	0.745	0.412	1.714
AIM		code from Neil Bruce. look for AIM.zip	0.71	0.378	1.649
Itti&Koch2		code from the GBVS package	0.71	0.388	1.670
Context-Aware saliency		code	0.699	0.377	1.642
Torralba		code	0.628	0.319	1.644
Hao & Zhang		code	0.627	0.308	1.855
SUN saliency		code from Lingyun Zhang's site	0.615	0.321	1.691
Achanta		code	0.453	0.221	2.202
Chance*		code	0.44	0.307	1.714
Itti&Koch		code from the Saliency Toolbox	0.215	0.177	2.181

* Baseline models that we compare against

submit a new model

We measure performance under two other metrics and find similar results

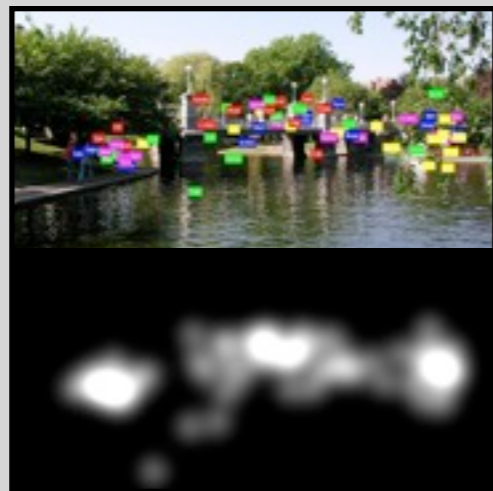
They send us saliency maps, we test for them

Summary of benchmark

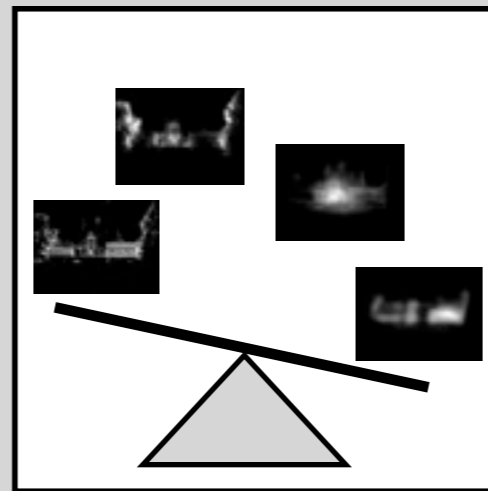
- Provide a consistent way to measure models against each other
- Currently Judd and GBVS models work best
- There is still a gap between models and human performance
- Want to optimize blur for each model and then compare models

Understanding and predicting where people look

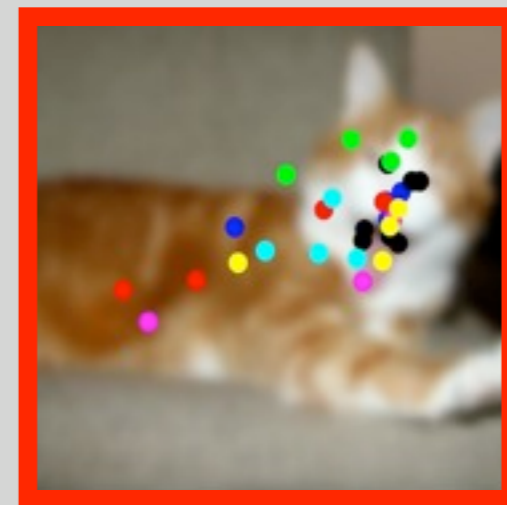
Introduction



Learning a model



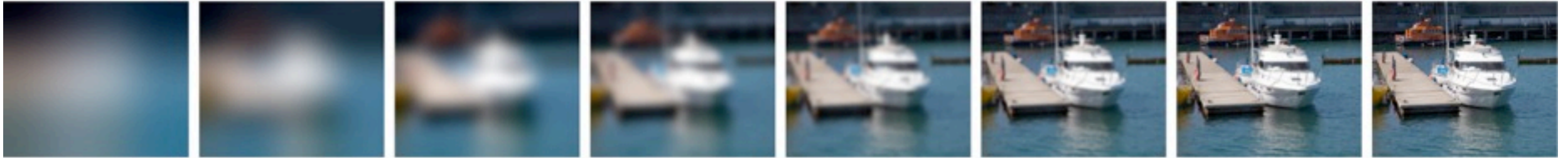
Benchmarking models of saliency



Fixations on low-resolution images

Conclusion

Motivation



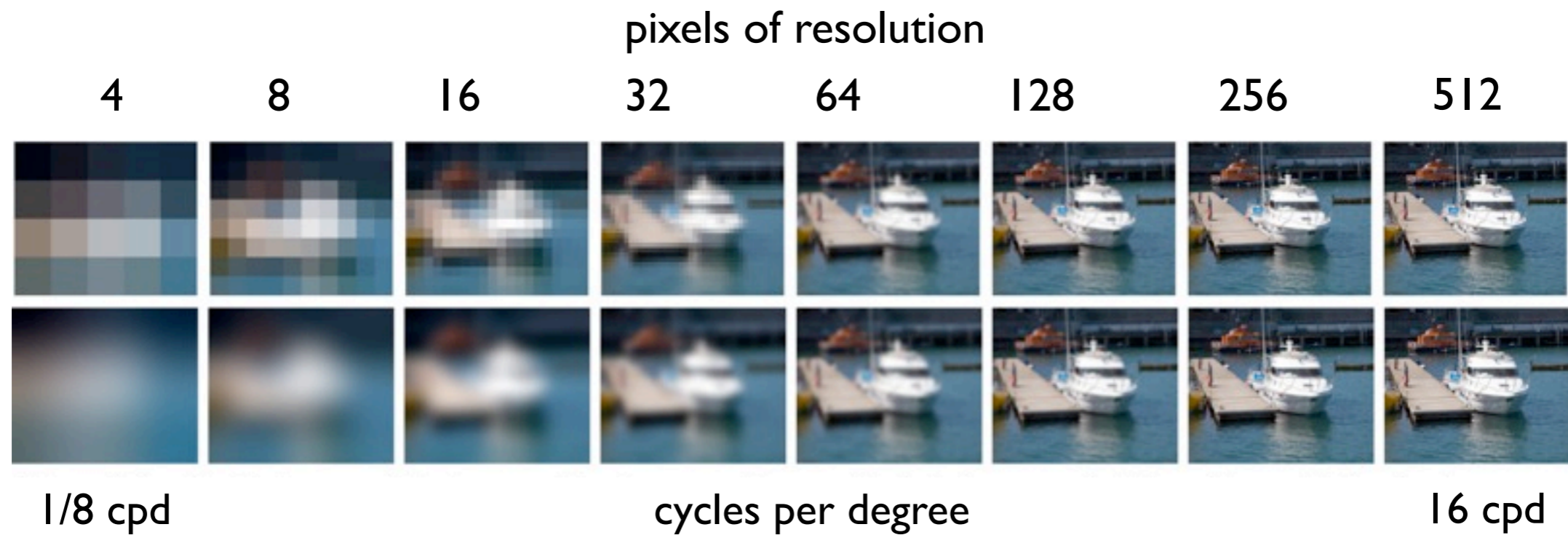
- Most eyetracking studies on high res images
- Torralba showed images understood as early as 32px
[Torralba 2009 How many pixels make an image?]
- Faces can be recognized as low as 16x16px
[Harmon & Jules 1973, Bachmann 1991, Schyns&Oliva 1997, Sinha et al 2006]



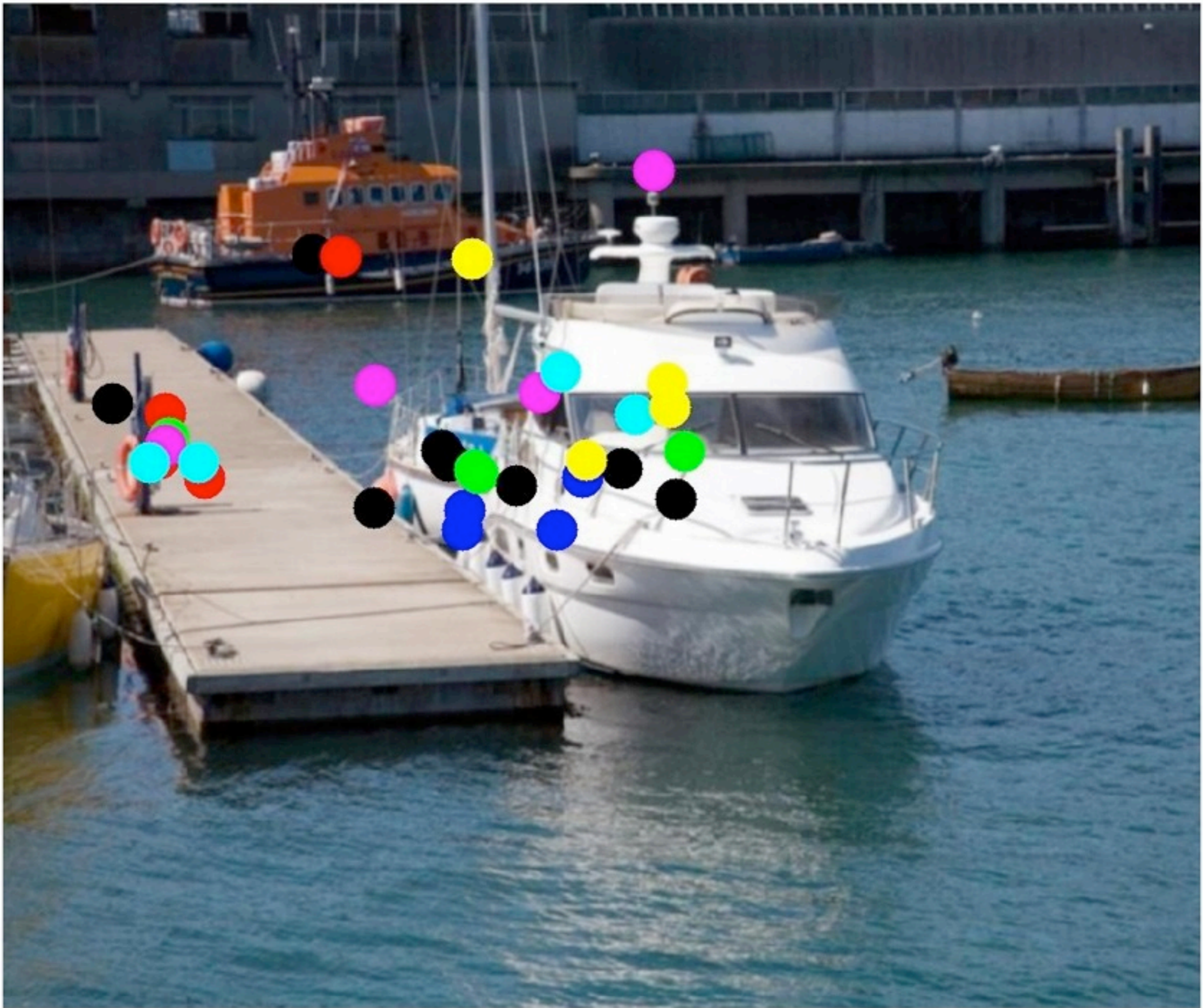
Motivating questions

- How does image resolution affect fixations?
- Are fixation patterns the same as you reduce resolution? how far can you go?
- Need for computational efficiency
- Models use high frequency features.
Is this really necessary?

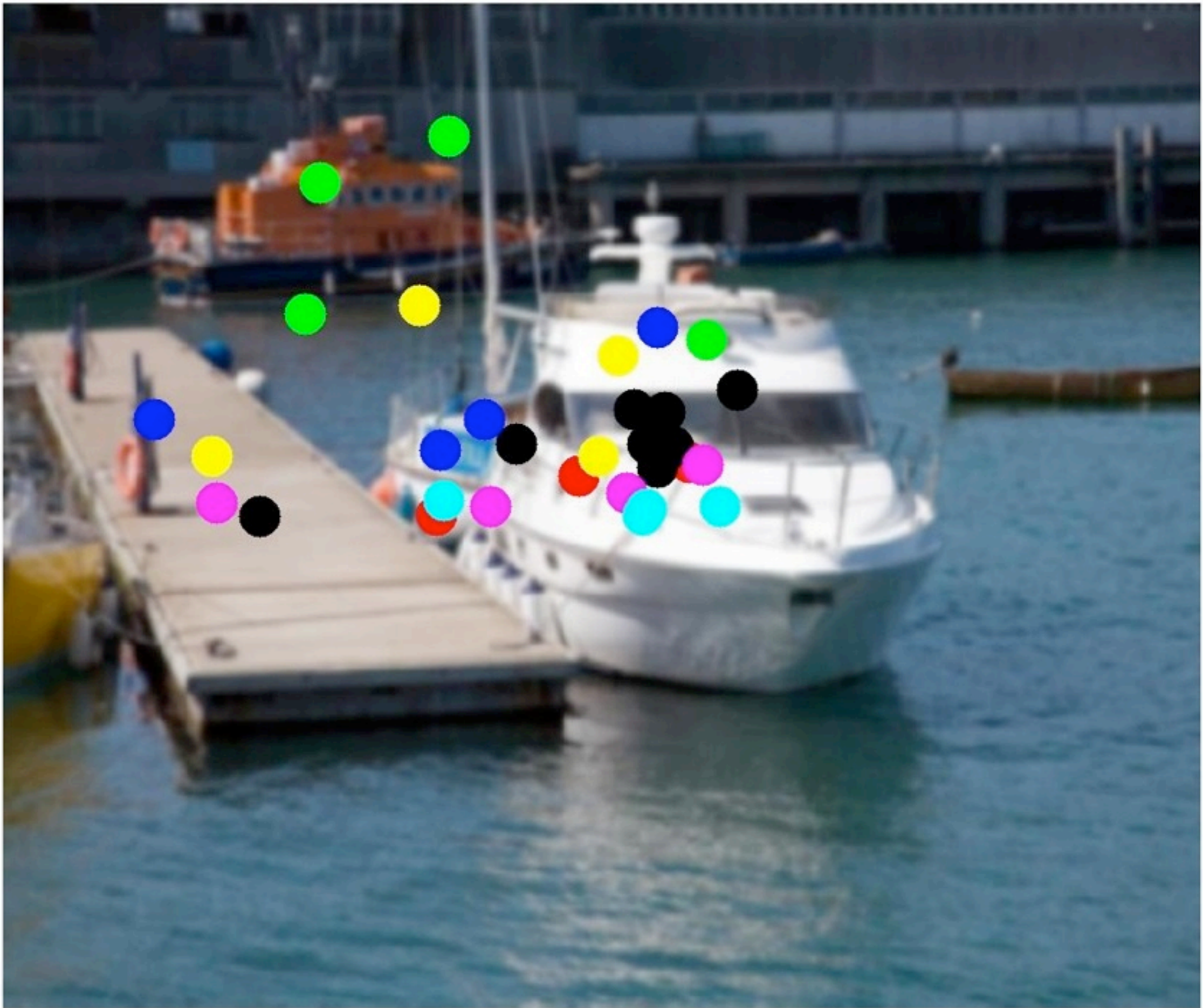
168 images were shown at
8 different resolutions
64 observers (8 per resolution)



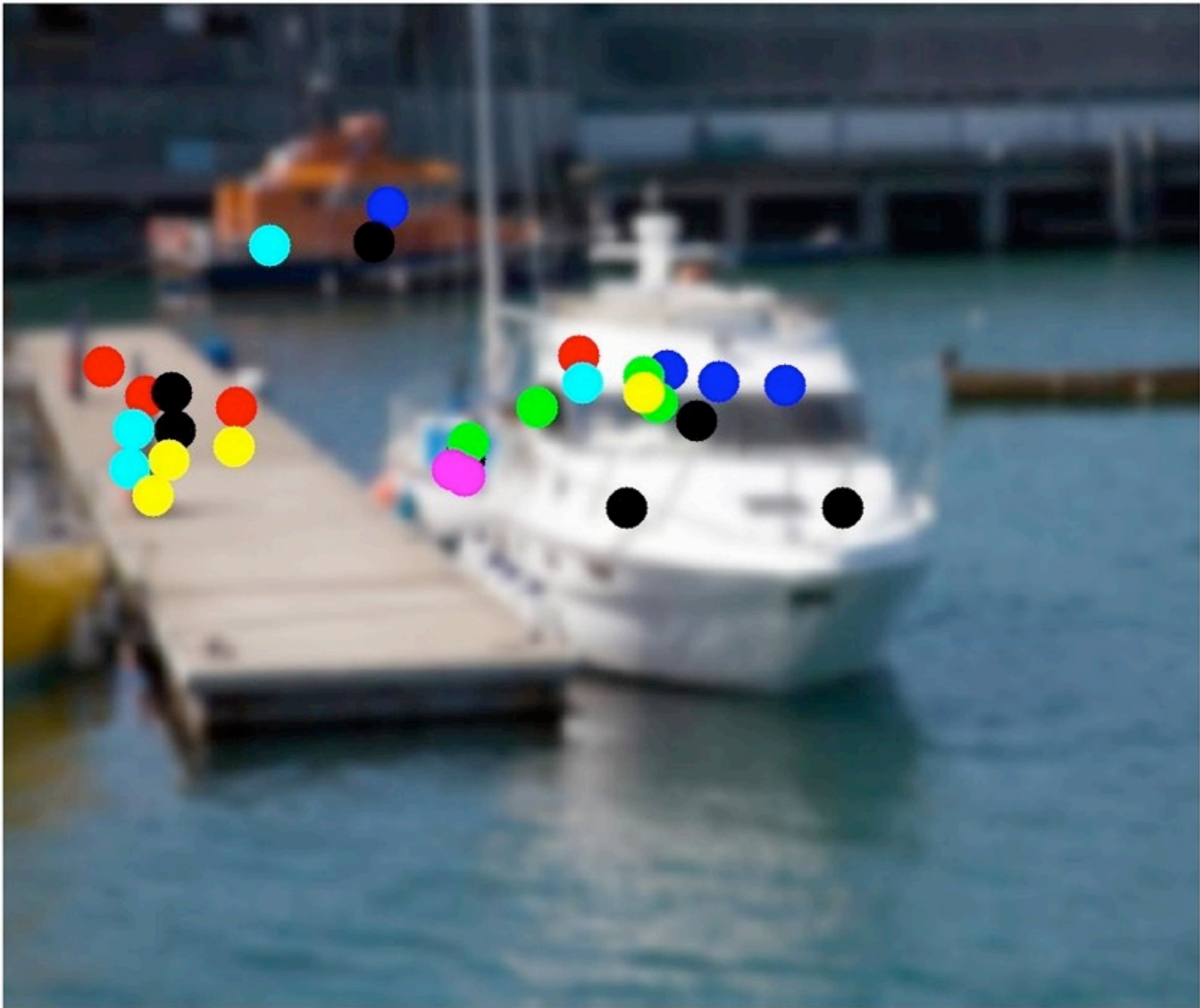
[Low-res data set]



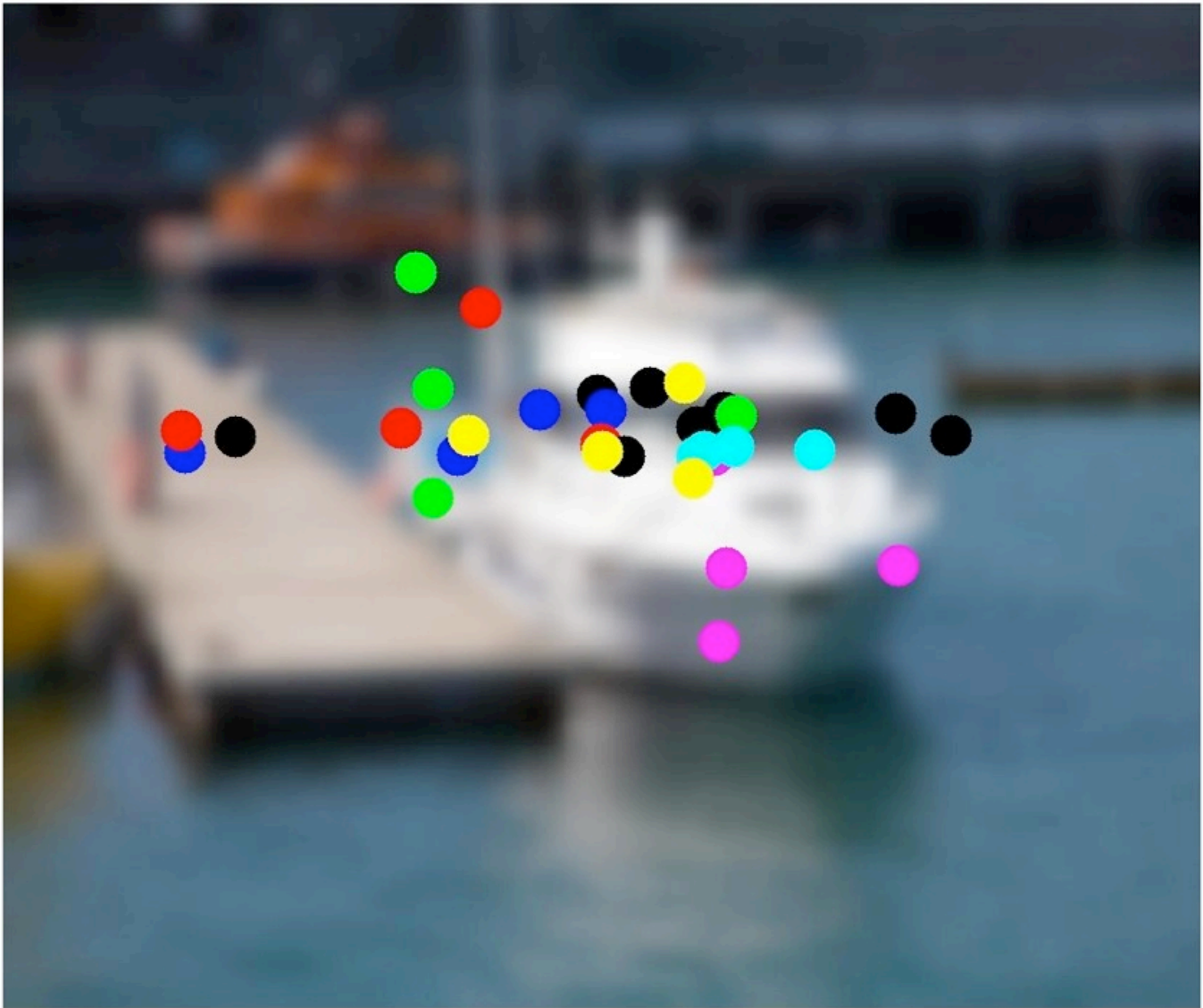
512px



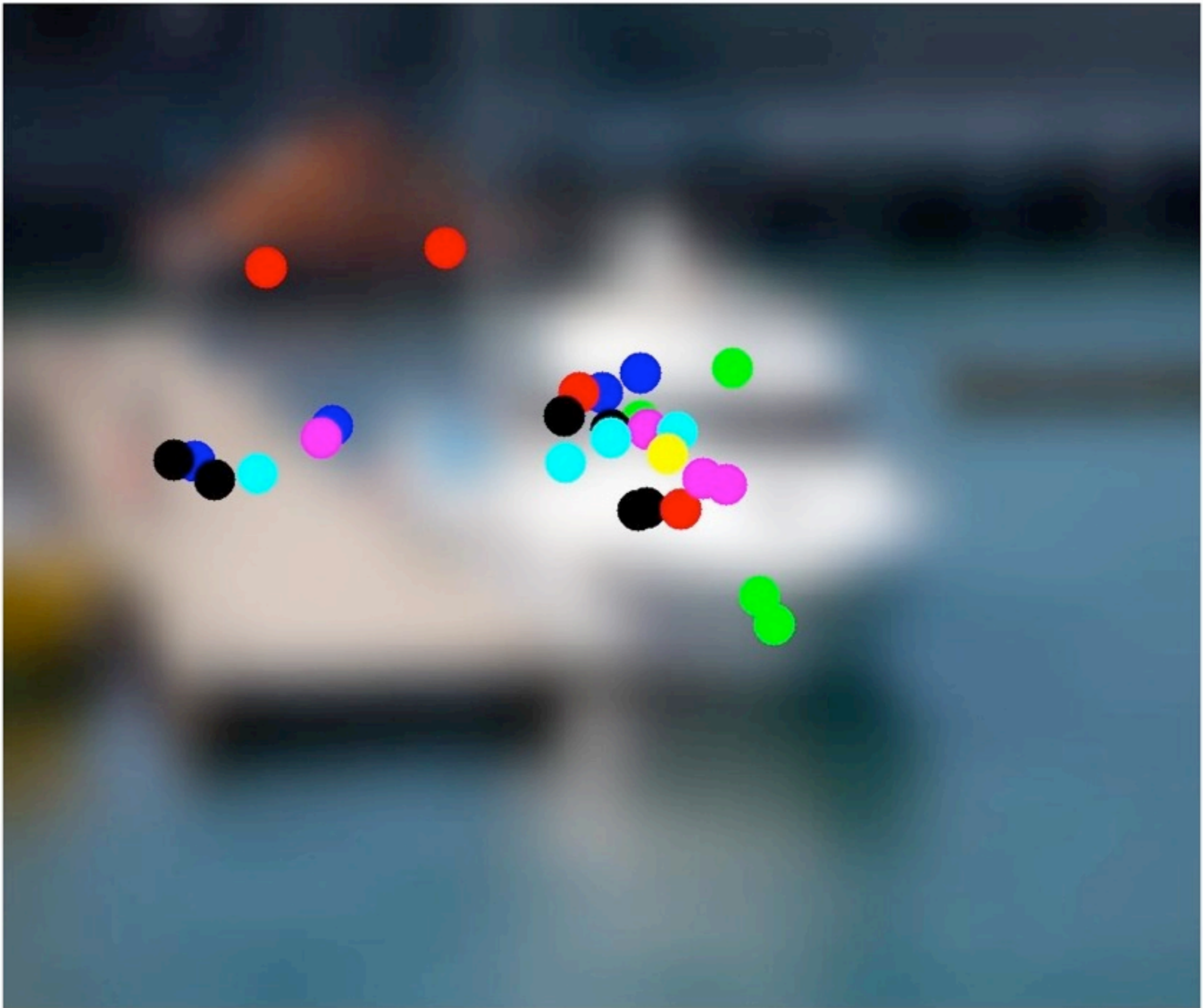
256px



128px



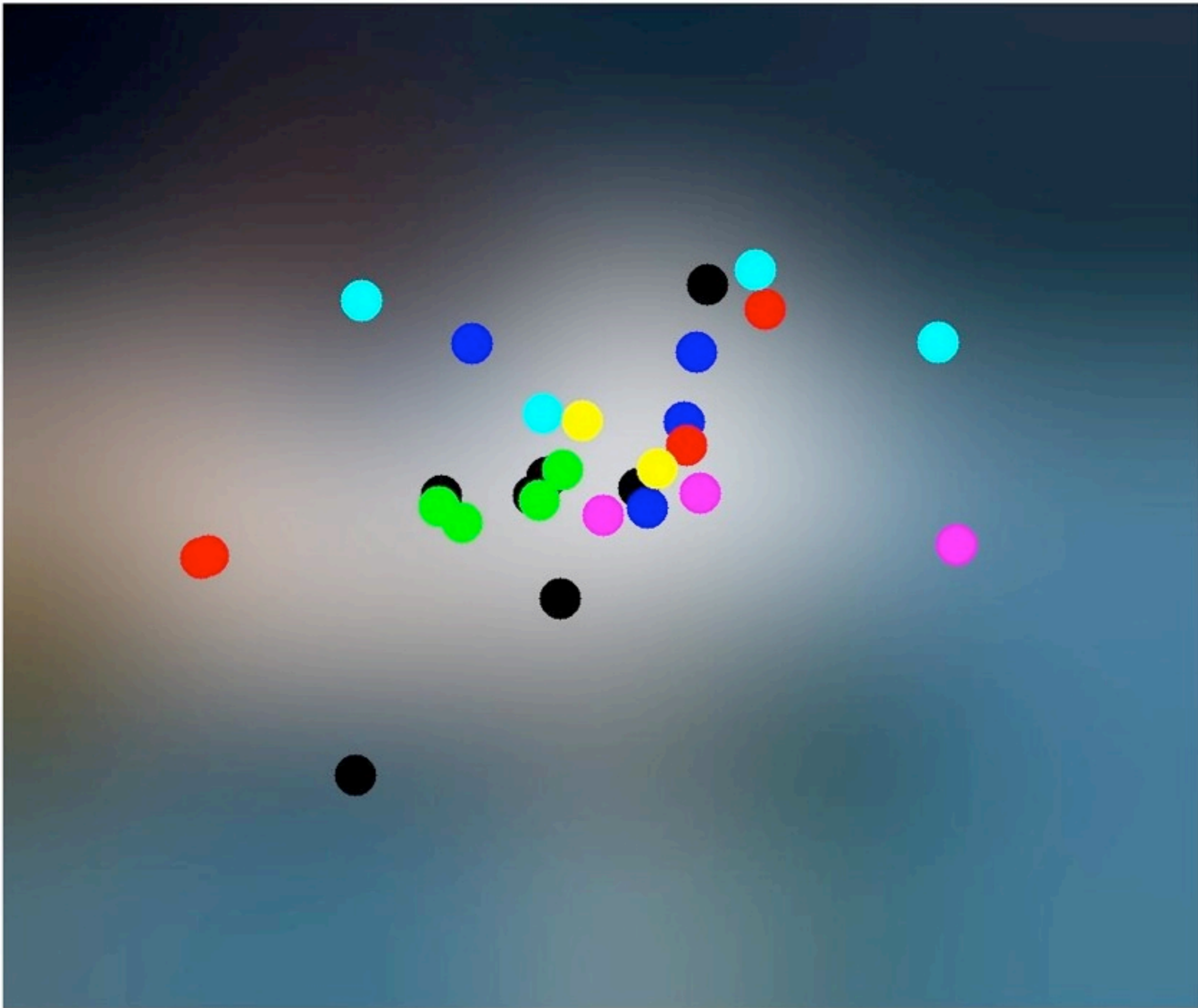
64px



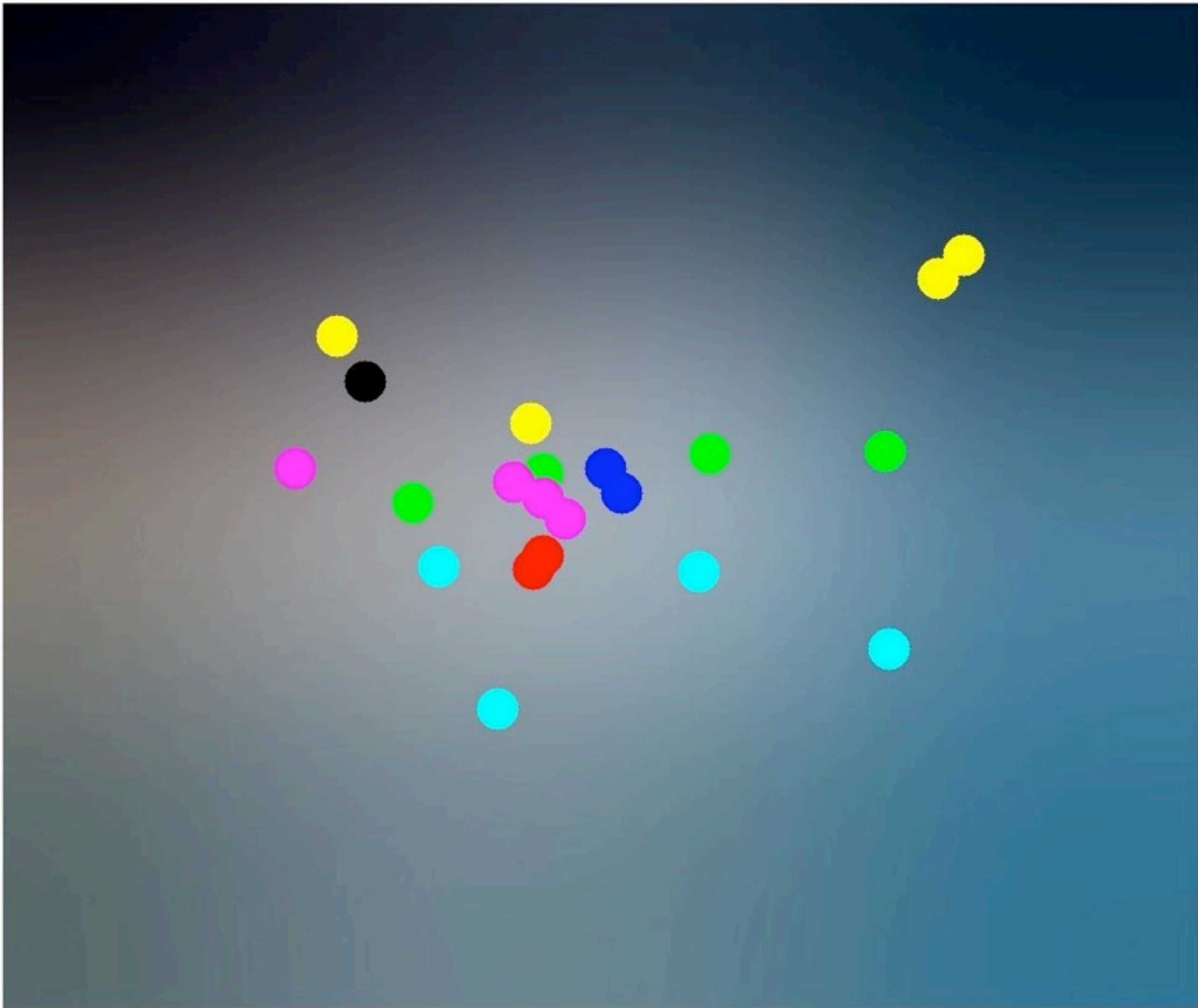
32px



16px

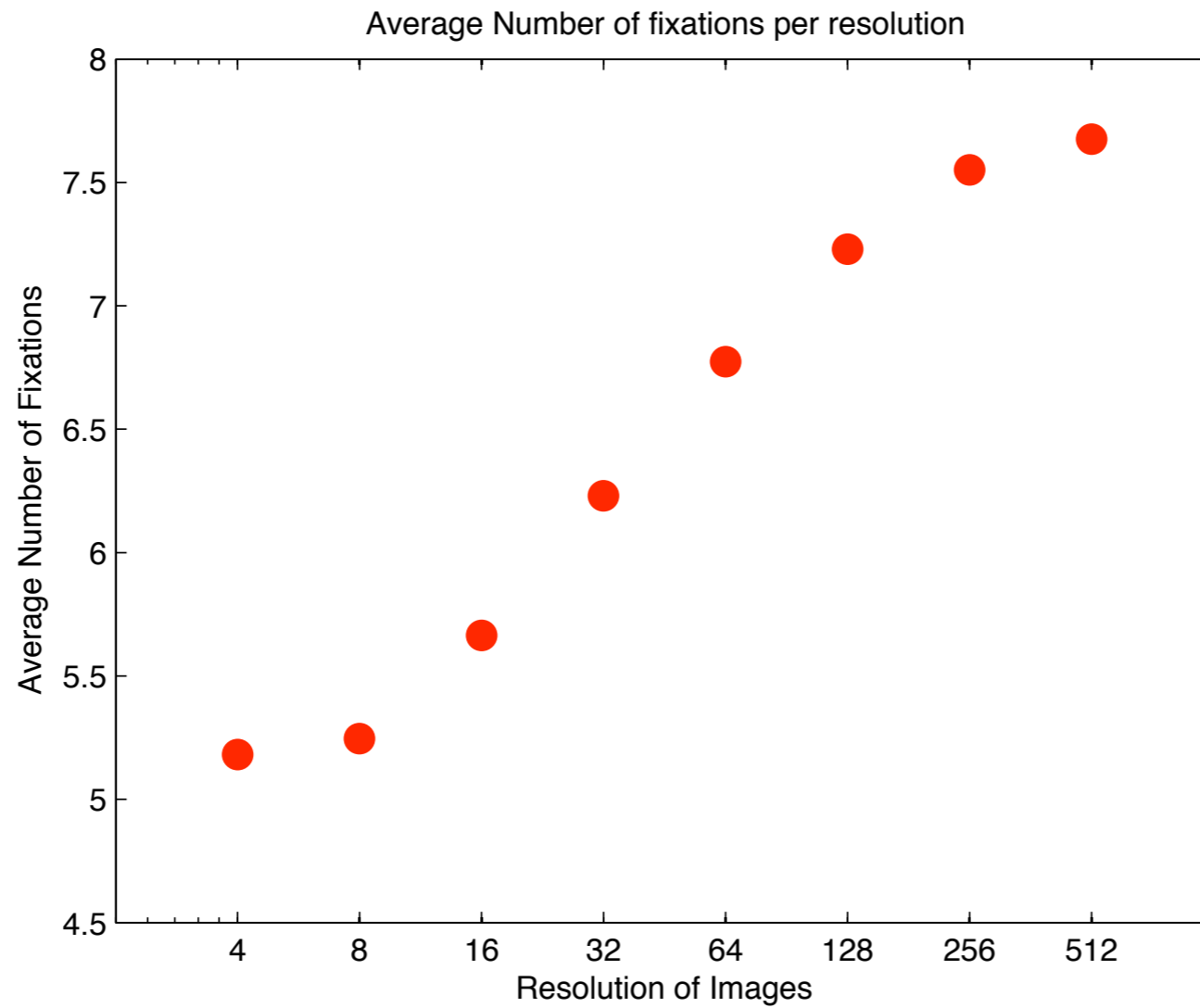


8px

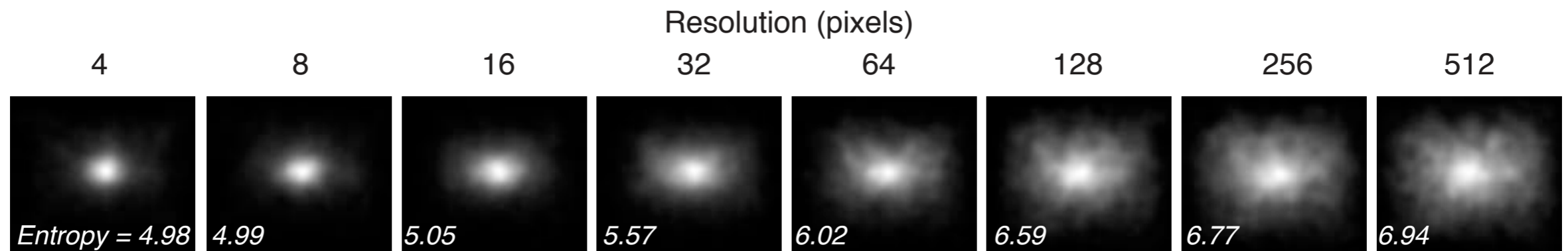


4px

As resolution decreases there are fewer fixations



As resolution decreases fixations are more centered



Average fixation map for 168 natural images at each resolution

We measure how well fixations of one image predict fixations of another image

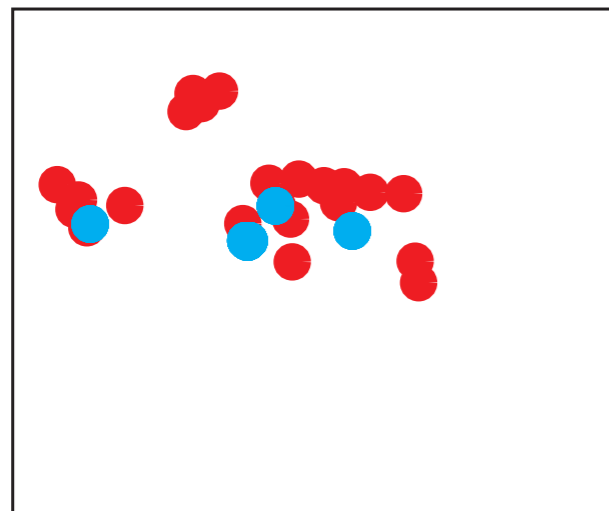
Image 1 (64px)



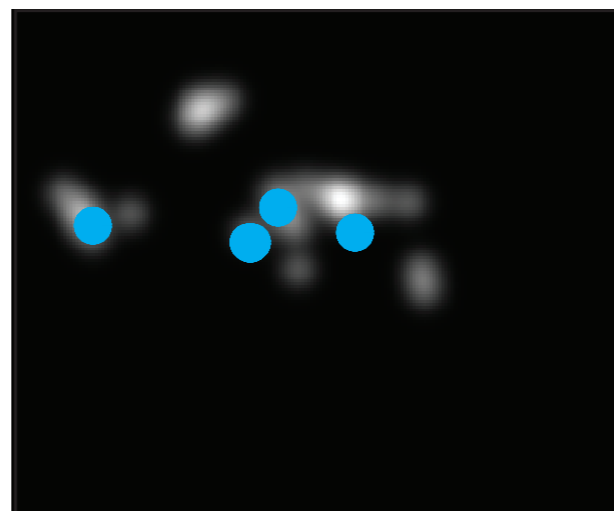
Image 2 (512px)



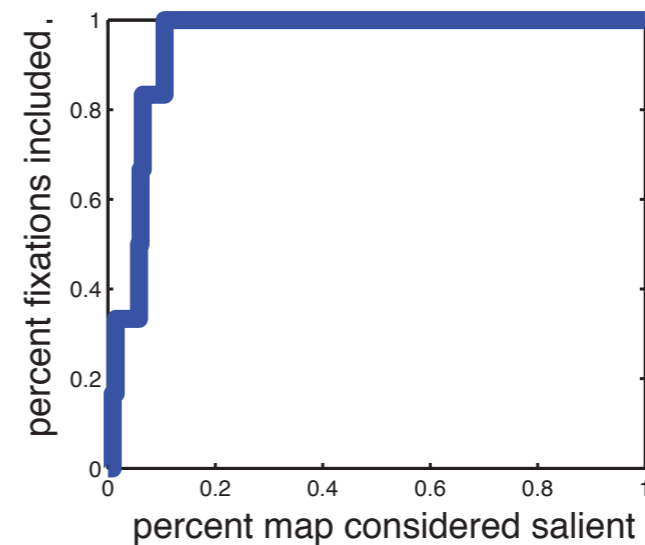
Fixations from both imgs



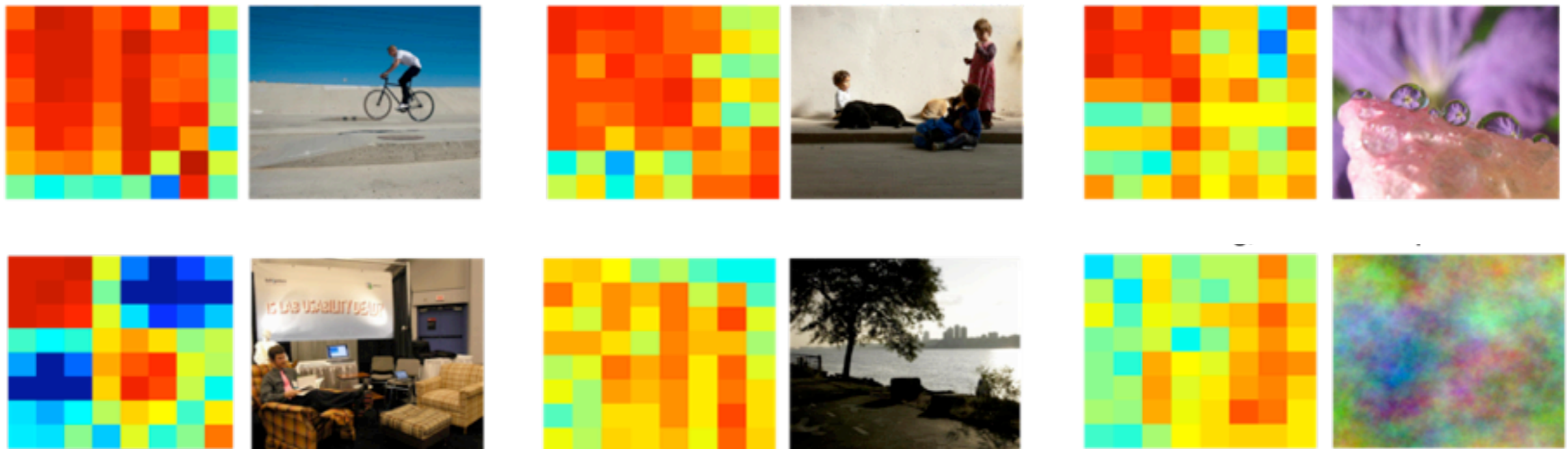
Fixation Map of Image 1



Area under ROC curve: 0.947



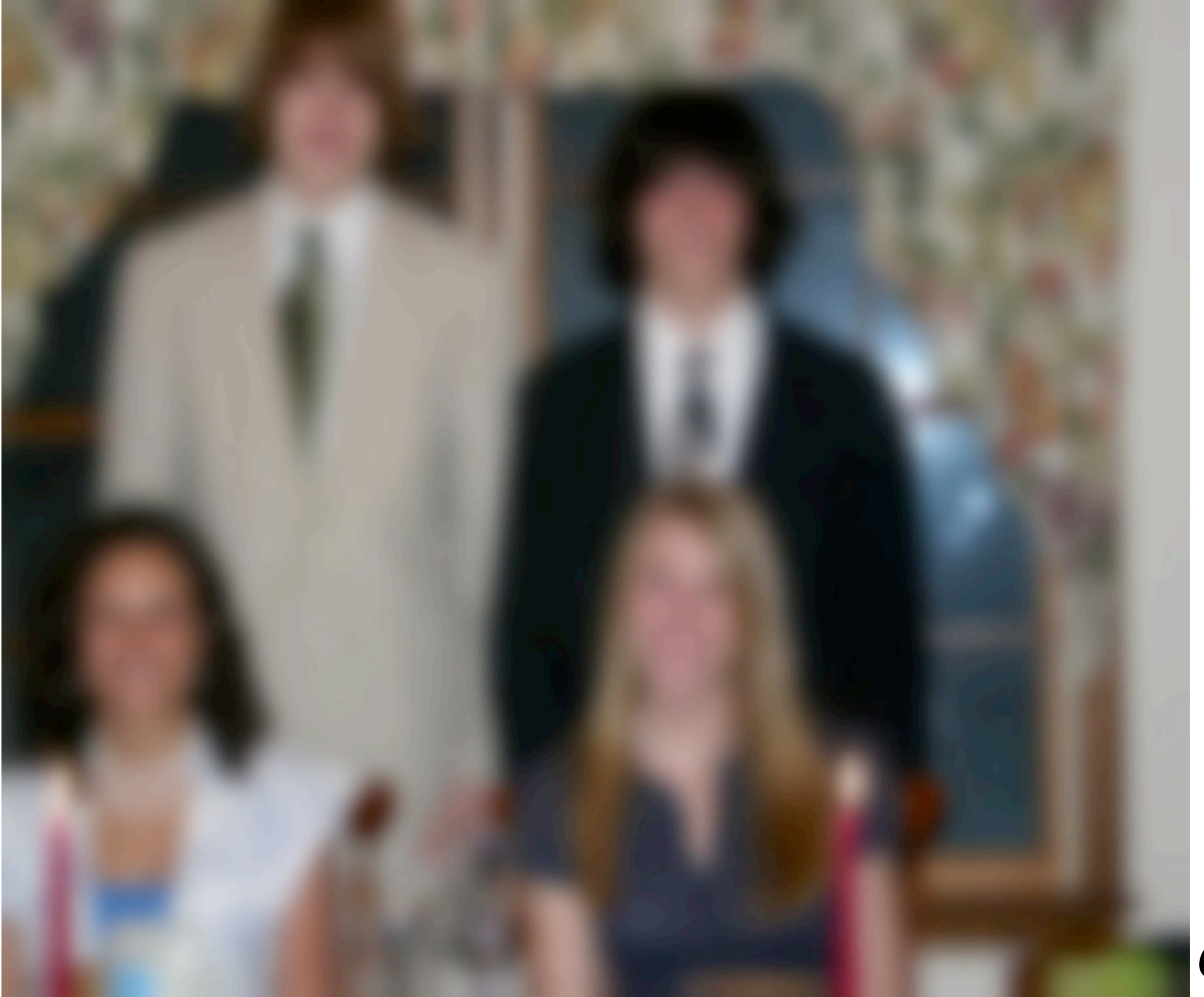
Fixation consistency depends on image complexity



Clap when you are pretty sure you know what is in the following image





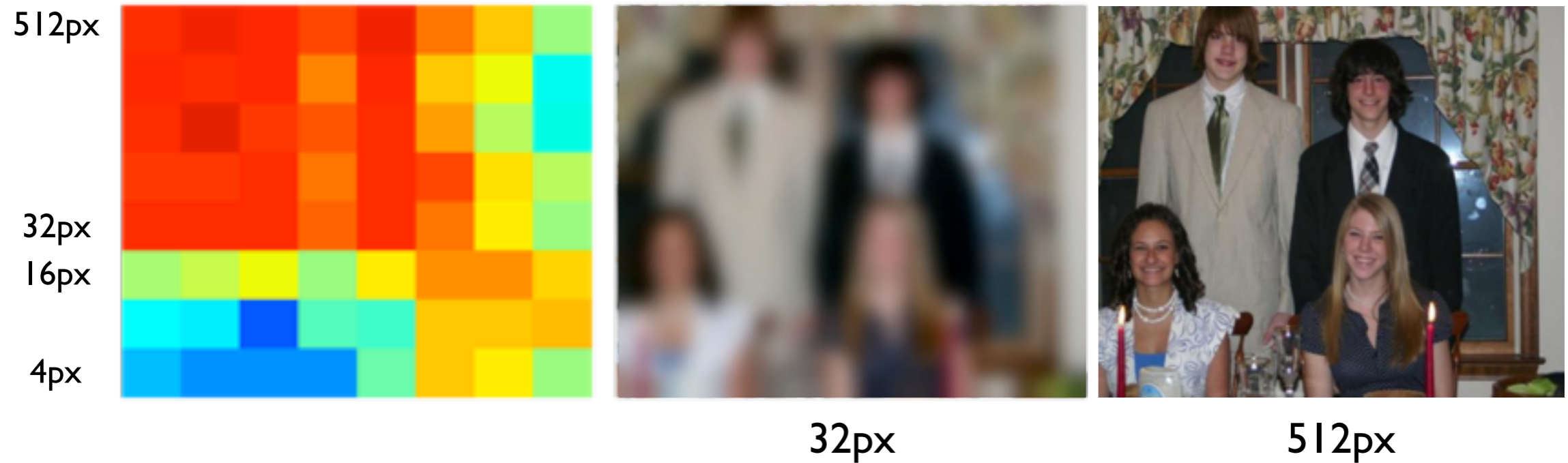








Hypothesis: fixations become consistent when people understand the image



Hypothesis: fixations consistent on blurrier images because resolution eye can resolve in periphery is same



high-res image with simulated visual falloff



low-res image has same cycles/deg in periphery

Main takeaways

- Fixations on images of $\sim 64\text{px}$ (2cpd) and above provide very good predictions of fixations on high-res images
- Depending on your application, you can use images of lower resolution... $>512\text{px}$ not necessary.
 - depends on image complexity and
 - number of fixations you want to model

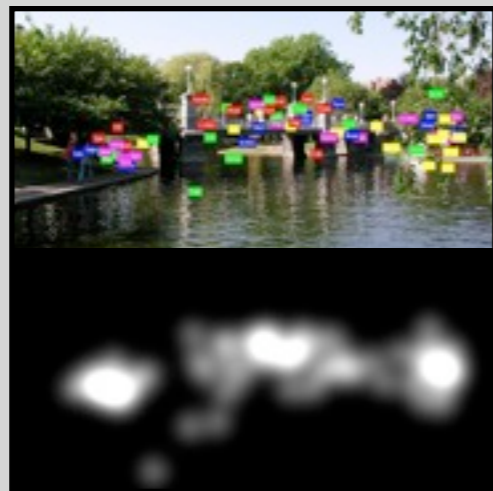
In the thesis we also show

- Human fixations are equally consistent with each other at all resolutions above 16px
- Prediction and consistency performance is higher for easy images than for hard images
- There is very little consistency among fixations on noisy images

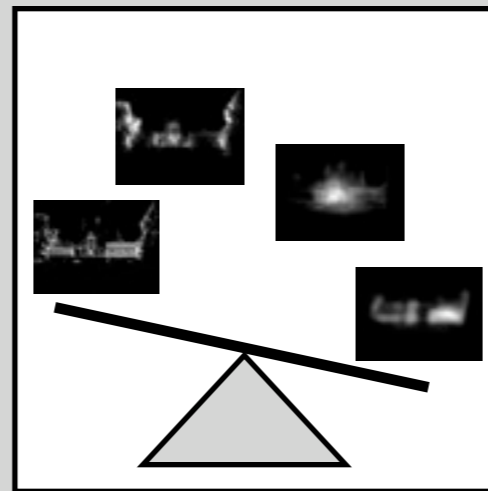


Contributions

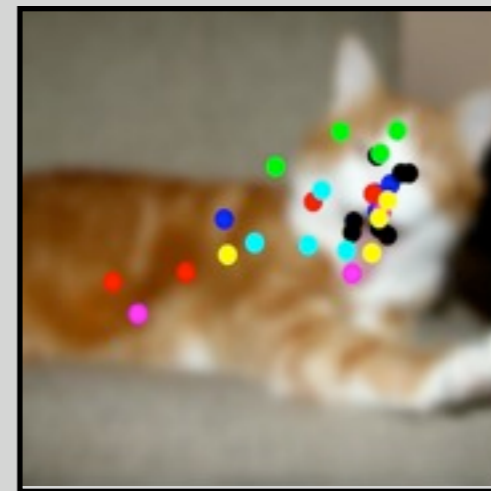
Introduction



Learning a model



Benchmarking
models of saliency



Fixations on low-
resolution images

Conclusion

Conclusions

- Humans are best predictors of where others will look
- Many features correlate with where people look. Using them *together* gets more power.
- Saw surprisingly high performance of the center model.
High weight on center feature.
- Face, person, object detectors useful but need to be used with other features
- High frequency features might not be as important. Blurrier models better, and fixations are in the same locations on blurry images
- Computational systems can gain efficiency by using low resolution images

Open questions

- **Better ways of integrating top-down information**
- **Better features**
(Devi Parikh show features more important than amount of training data).
- **Better data sets**
(Are they good enough, diverse enough? What biases do they have? Torralba and Efros 2011 and Pinto 2008 show data sets often biased).
- **Better understanding of human vision**
(Understand effects of image retargeting, image warping, image compression and degradation, image size)

Acknowledgments

- Fredo Durand and Antonio Torralba
- Aude Oliva and Bill Freeman
- Krista Ehinger, Barbara Hidalgo-Sotelo, Nicolas Pinto
Yann Le Tallec, Sylvain Paris for fruitful conversations
and help with the work and eye tracking experiments
- Eye tracking subjects
- MIT Graphics Group
- Tom Buelher and Bryt Bradley
- Friends
- Family and Yann

image images visual models fixations human attention resolution information look system many computational room benchmark

mechanisms bottom-up areas visual model features ref fixations people saliency task factors movements

research thought James experiments researchers Attentional JustCarpenter1980

quote top-down hypothesis cite author high

different data shortcite et provide subject relevant scene

interest models begin Treisman198097

eye human resolution end saliency

center called information Many content able low

understanding interact room benchmark figure selective

current one robots particular difficult

make recorded

outlined available

return related regions inhibition mechanism also several next vision developed

complexity databases picture show three

section interesting location amount

way around shown feature focus order need measure

naive still

mechanisms

bottom-up

emph

weights brain textbf help consistency map

al applications

given gets

first looking

Yarbus orientation trackers fovea center object chapter computational pixel

observer's movement easy notice Visitor

Repin's followed surrounding

move

certain

far processed

viewing spotlight item

often time

begin

stimuli bias

end

salient

Yarbus

observer's

movement

Repin's

move

certain

far processed

viewing spotlight item

often time

begin

stimuli bias

end

salient

Yarbus

observer's

movement

Repin's

move

certain

far processed

viewing spotlight item

often time

begin

stimuli bias

end

salient

Do males and females look at different locations?

Gender Differences for Specific Body Regions When Looking at Men and Women,”

Johannes Hewig, Ralf H. Trippe, Holger Hecht, Thomas Straube and Wolfgang H.R. Miltner,
Journal of Nonverbal Behavior, vol. 32, no. 2, June 2008

Do males and females look at different locations?

Gender Differences for Specific Body Regions When Looking at Men and Women,”

Johannes Hewig, Ralf H. Trippe, Holger Hecht, Thomas Straube and Wolfgang H.R. Miltner,
Journal of Nonverbal Behavior, vol. 32, no. 2, June 2008

“Participants were exposed to 30 pictures
of 15 male and 15 female models in casual clothing.

Do males and females look at different locations?

Gender Differences for Specific Body Regions When Looking at Men and Women,”

Johannes Hewig, Ralf H. Trippe, Holger Hecht, Thomas Straube and Wolfgang H.R. Miltner,
Journal of Nonverbal Behavior, vol. 32, no. 2, June 2008

“Participants were exposed to 30 pictures of 15 male and 15 female models in casual clothing.

The results show that both male and female observers **primarily gaze at people’s face.**

Do males and females look at different locations?

Gender Differences for Specific Body Regions When Looking at Men and Women,”

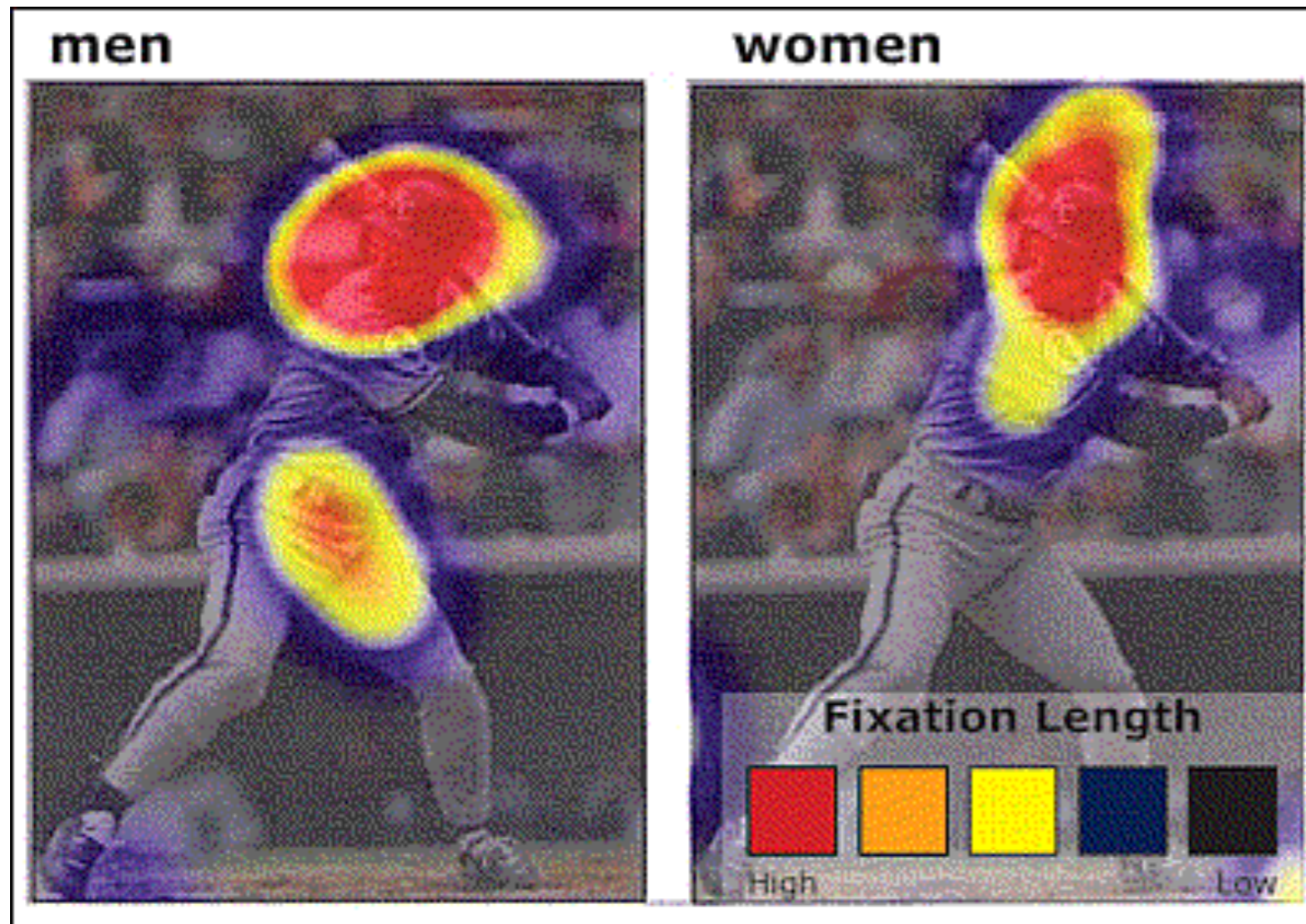
Johannes Hewig, Ralf H. Trippe, Holger Hecht, Thomas Straube and Wolfgang H.R. Miltner,
Journal of Nonverbal Behavior, vol. 32, no. 2, June 2008

“Participants were exposed to 30 pictures of 15 male and 15 female models in casual clothing.

The results show that both male and female observers **primarily gaze at people’s face.**

Only after this initial face-scan, men look significantly earlier and longer at women’s breasts, while women look earlier at men’s legs.”

Do males and females look at different locations?



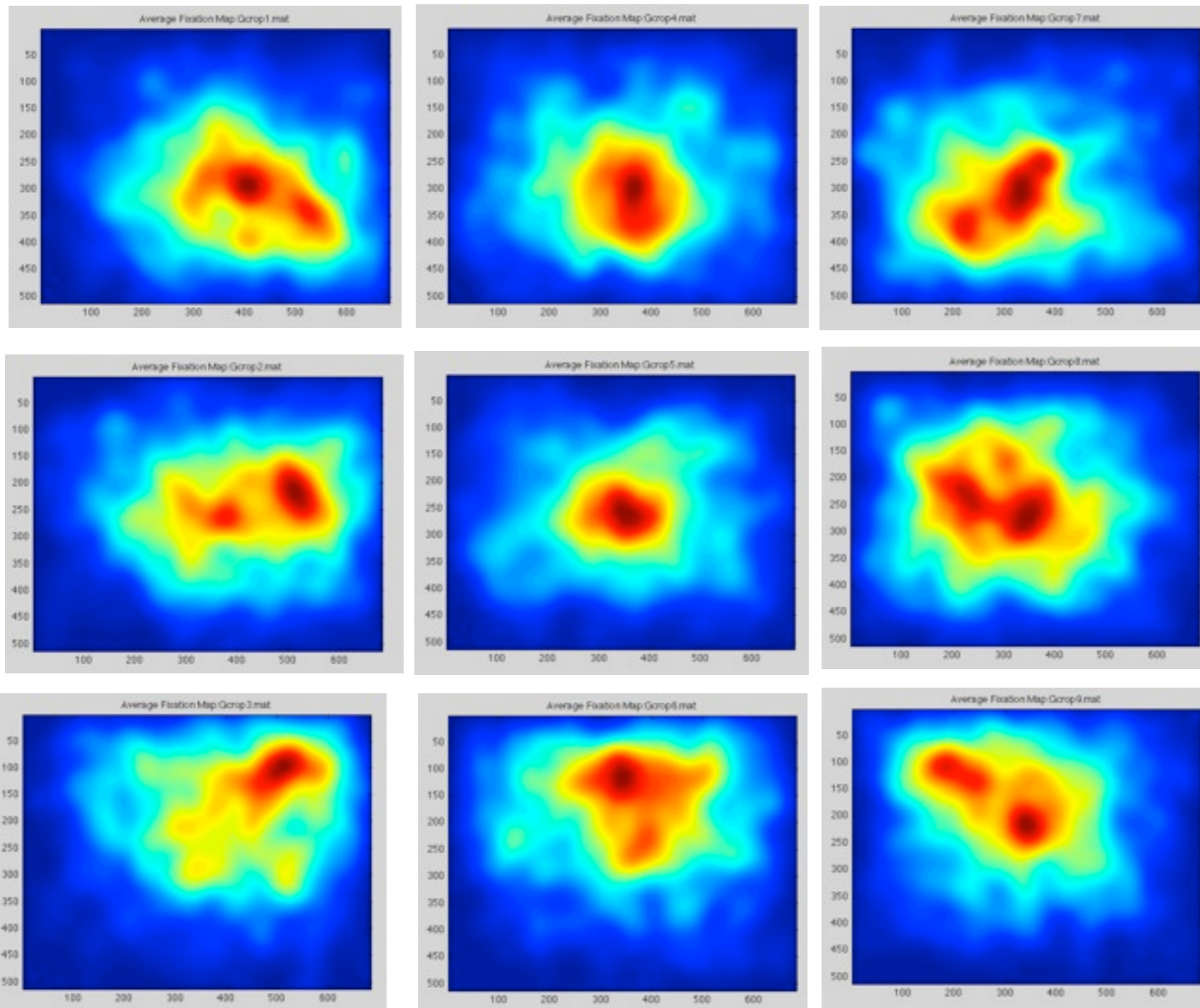
“Men tend to focus on private anatomy as well as the face. For the women, the face is the only place they viewed”.

Online Journalism Review
(2007)

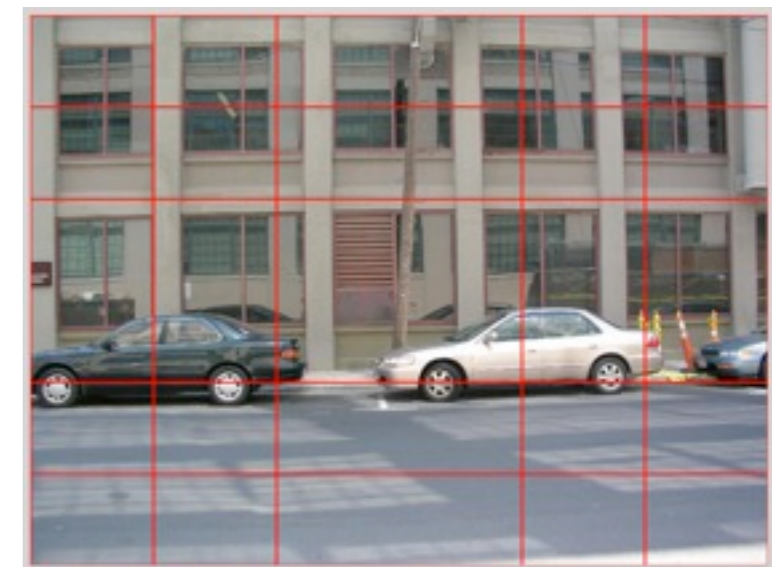
[Eyetracking points the way to effective news article design](http://www.ojr.org/ojr/stories/070312ruel/)

<http://www.ojr.org/ojr/stories/070312ruel/>

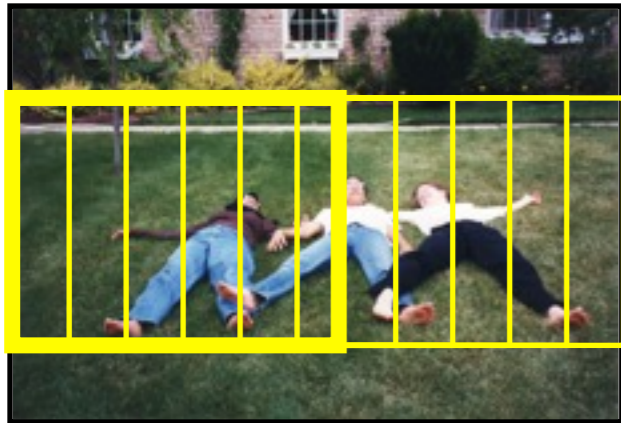
Does cropping an image affect the center bias of fixations?



Crop Reference

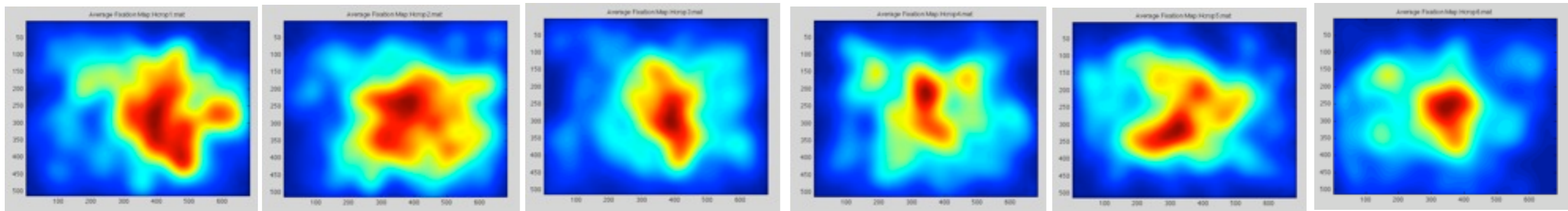


Does cropping an image affect the center bias of fixations?



Crop Reference

All crops have some center bias, but in addition...



fixations lean
to the right

fixations lean
to the left

Is there less center bias if you do not put the cross hair in the center?

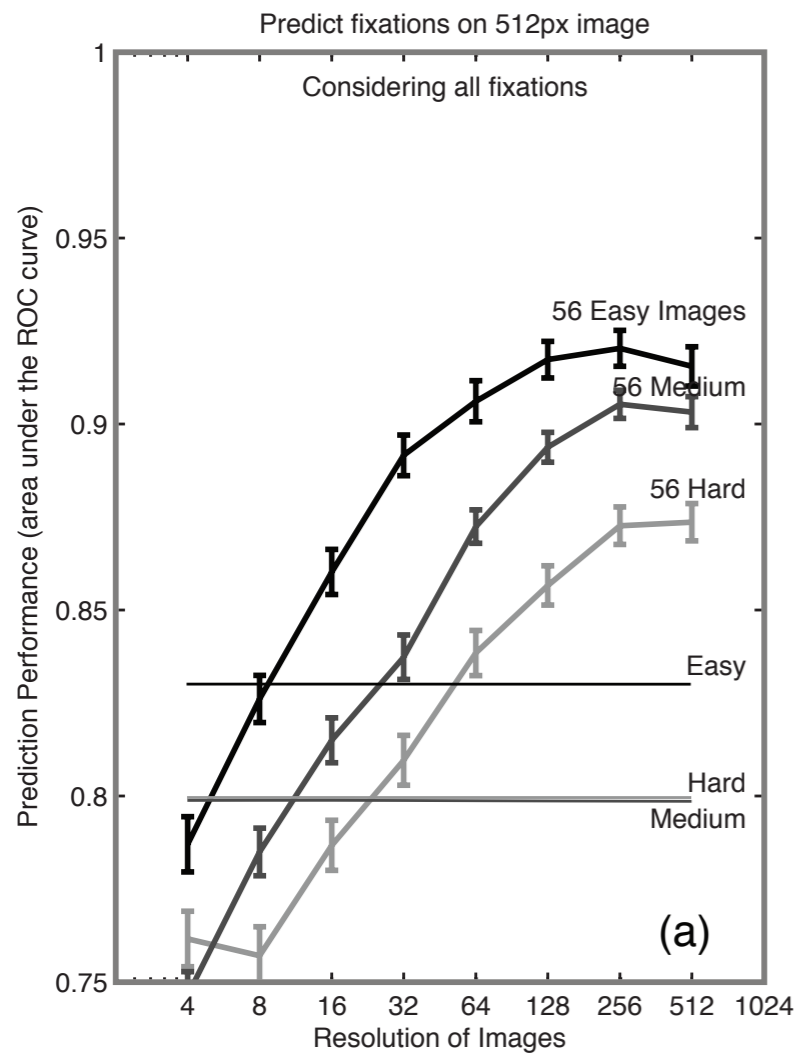
The tendency to look near the center may be reinforced during the eye tracking experiments. Indeed each trial began with a centrally located fixation marker. This marker could be randomly positioned.

However, studies which did not use a central fixation marker (Canosa, Pelz, Mennie, and Peak, 2003) have also shown a central fixation bias.

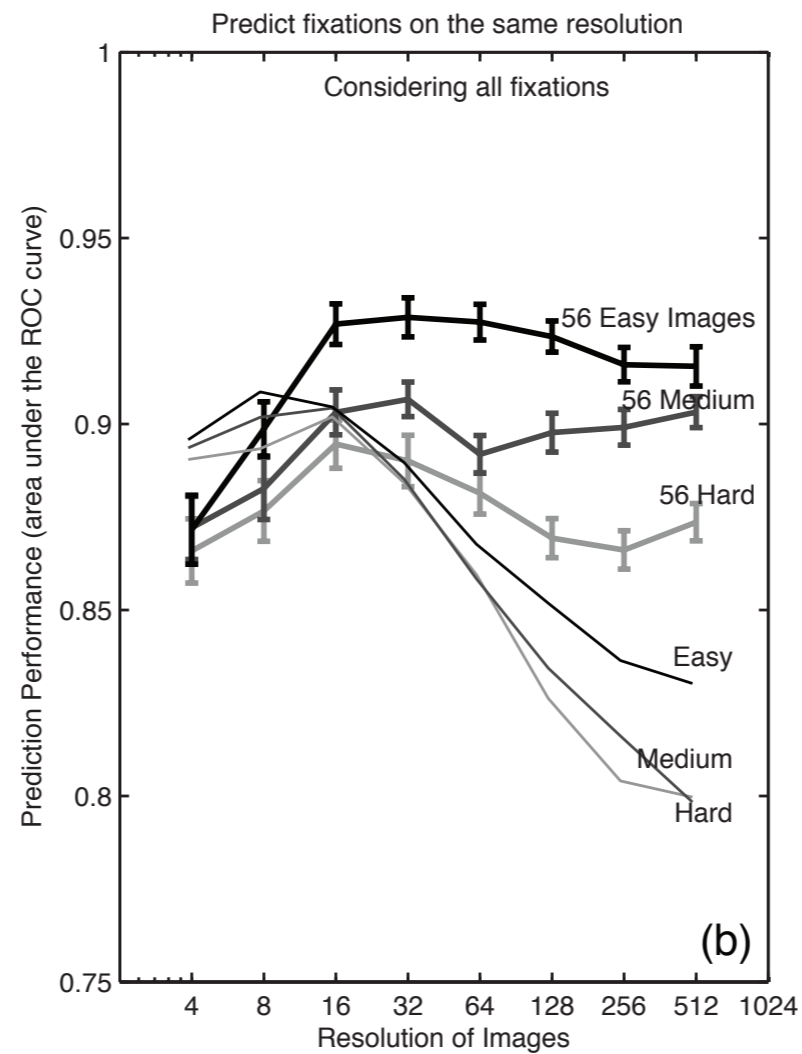
Center bias is also not reduced with increased viewing time as shown by Le Meur et al [2006] who showed images at 2s, 8s and 14s.

Fixation consistency depends on image complexity

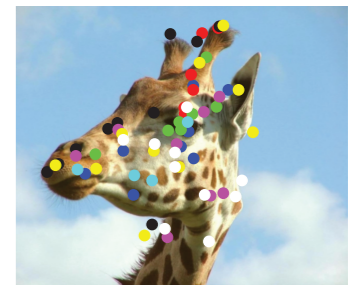
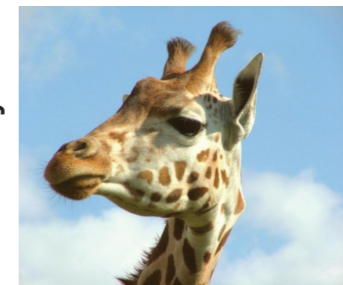
Predict Fixations on High Res Imgs



Consistency of Fixations per resolution



Example Images



Easy



Medium



Hard

Fixations

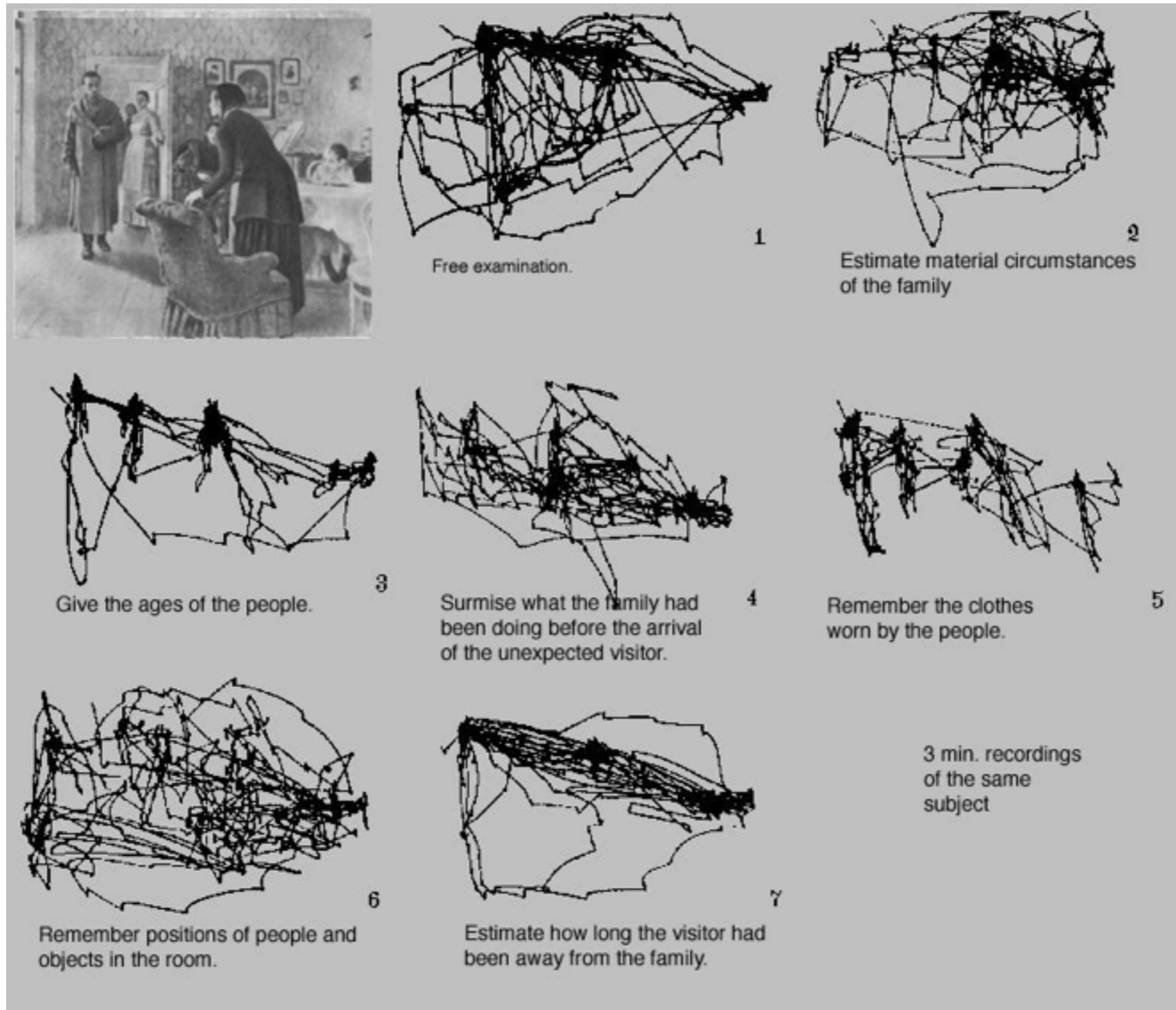
A

Yarbus (1967) was the first to show that *task* influences fixation locations



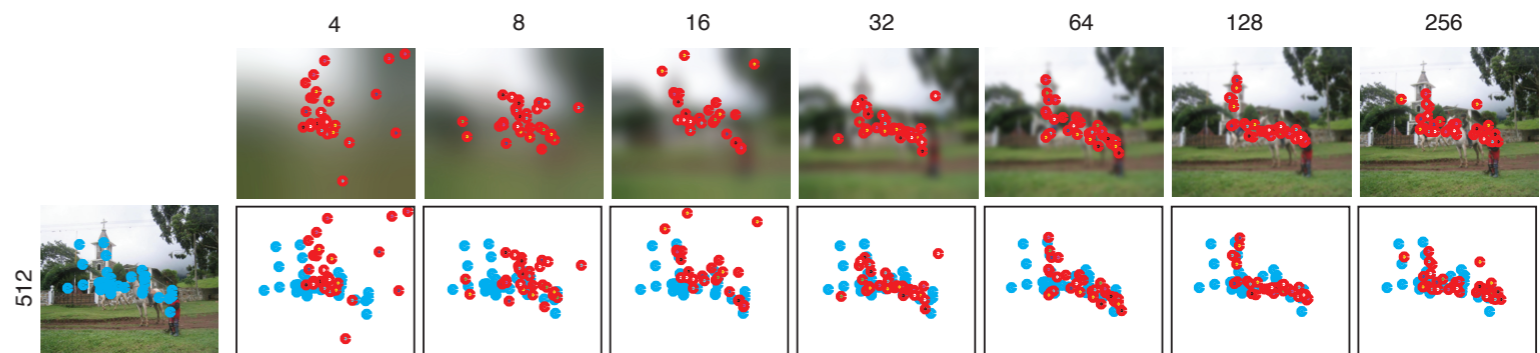
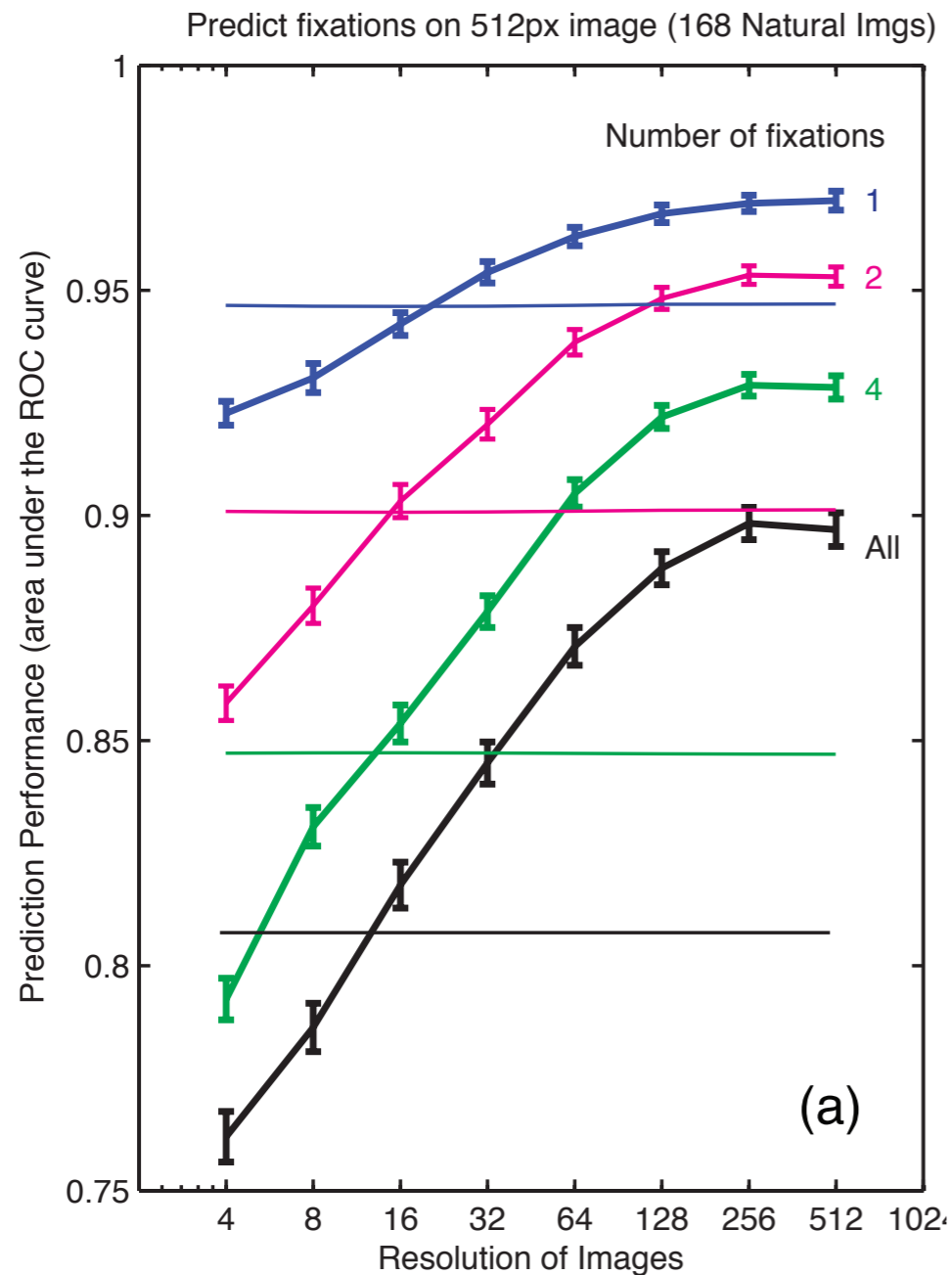
"They did not expect him" by Repin

Yarbus (1967) was the first to show that *task* influences fixation locations



Fixations on low-res images can predict fixations on high-res images

Predict Fixations on High Res Imgs






It is possible to see the fixations online

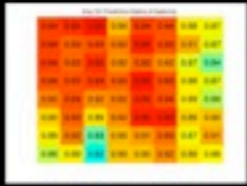
Fixations on Low-Resolution Images
Tilke Judd, Fredo Durand, Antonio Torralba

Here are all the images from our low-resolution experiment, arranged as tiny thumbnails. Click on a thumbnail and the larger image will appear. Press Play to examine the fixation data that was recorded for that image. The related paper is under review.

512
256
128
64
32
16
8
4




Timer: 2.90sec
This is raw data. [\[what?\]](#)
Red vs Yellow dots. [\[what?\]](#)



Play Fixation Data

Viewer 1 2 3 4 5 6 7 8 All Uncheck All

















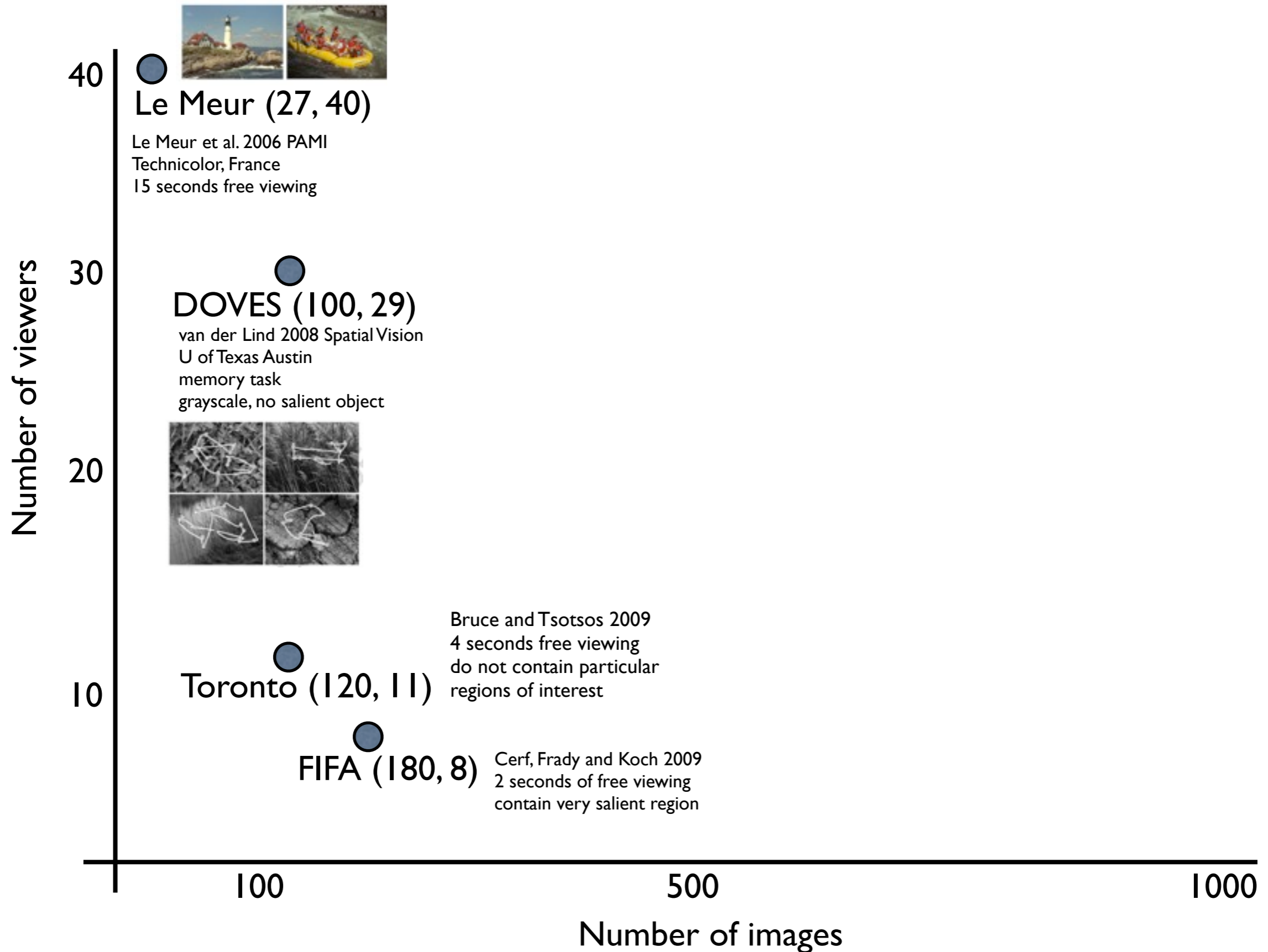




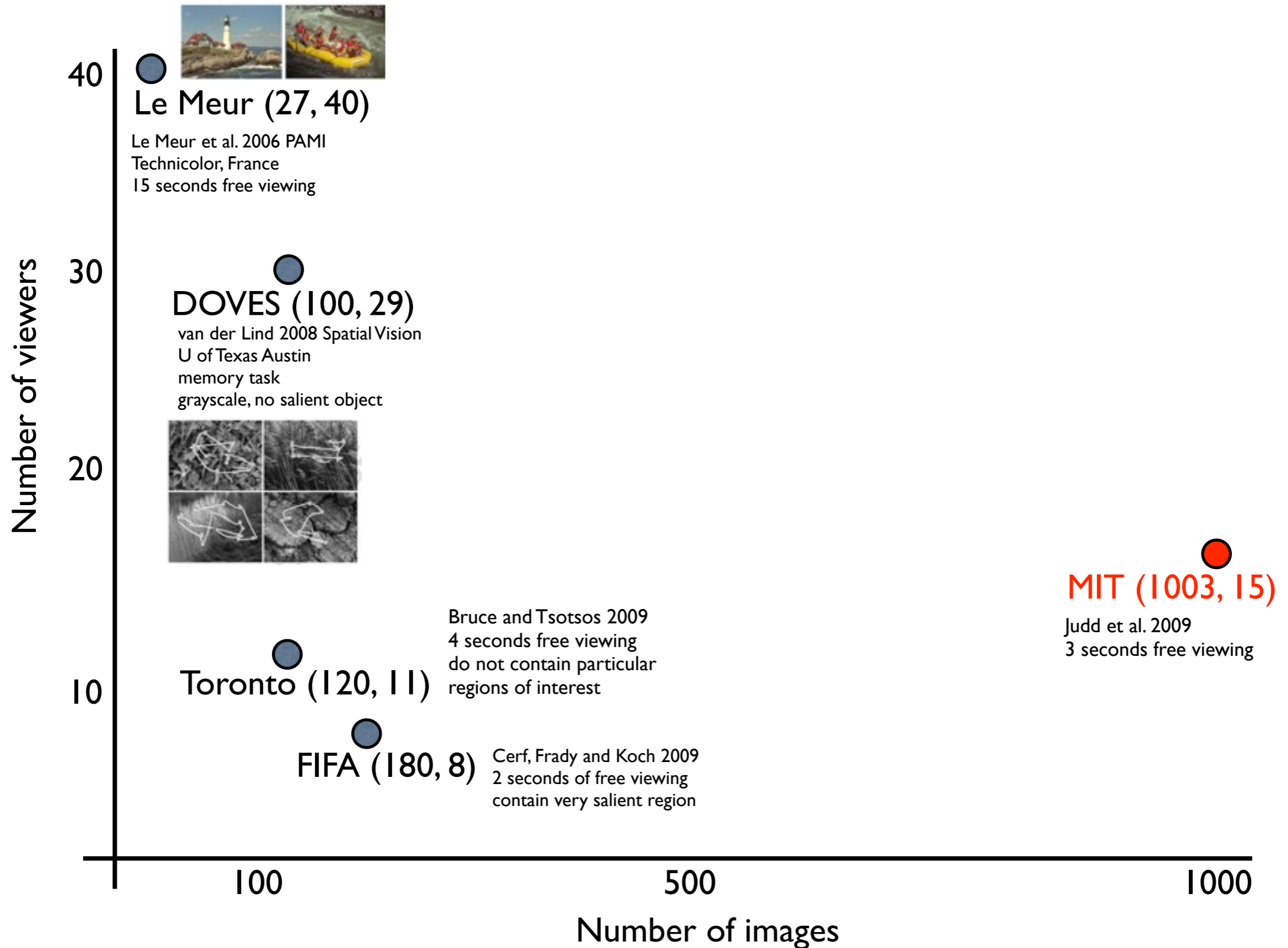




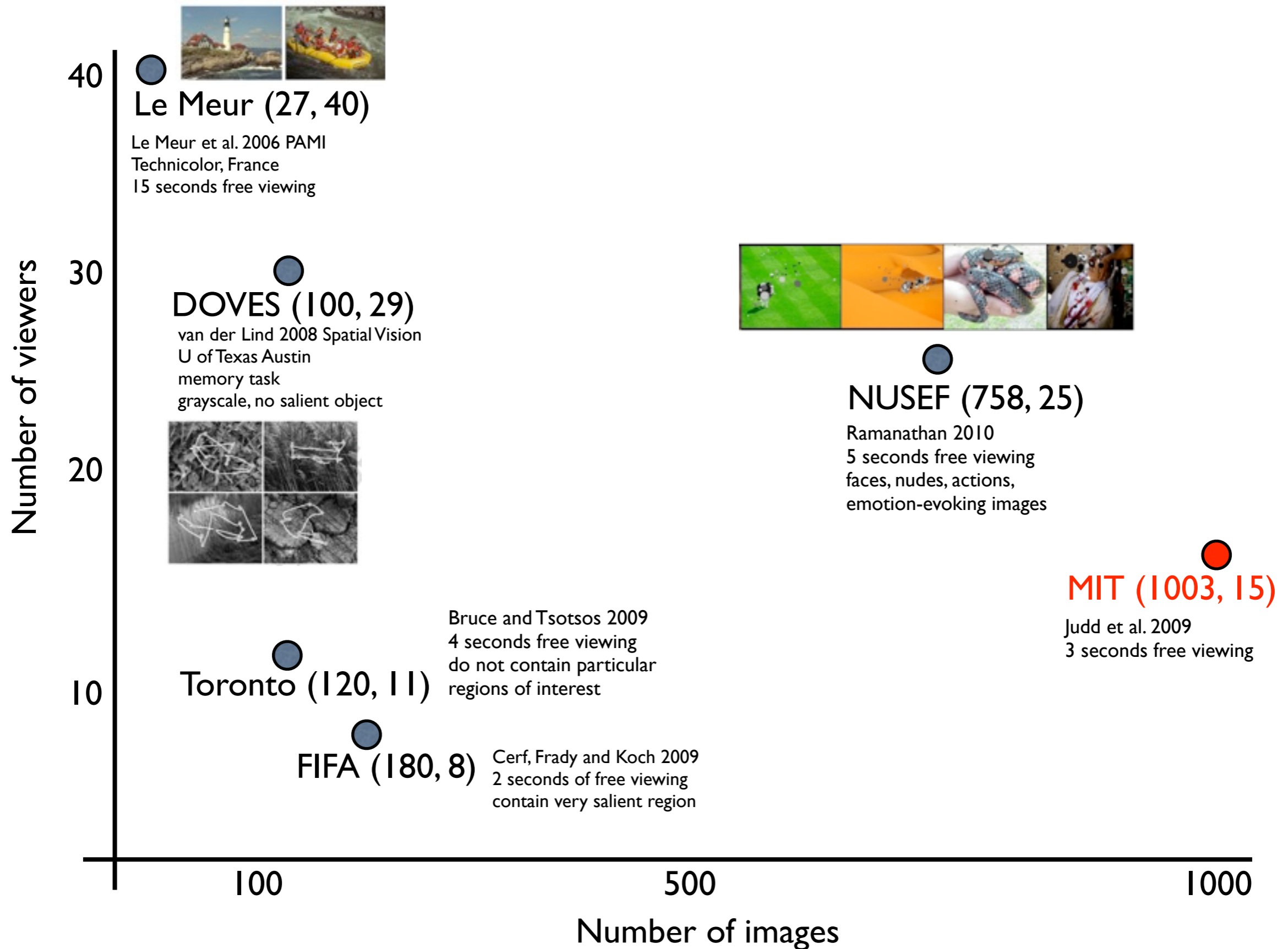
Data sets



Data sets



Data sets



Data sets

