Scenes - Objects

Aayush Bansal

What is a scene?

Slide from David’s presentation in previous class.

How should we represent scenes?

Slide from David’s presentation in previous class.

How do we represent scenes?

Slide from David's presentation in previous class.

And..

We discussed about content, expanse and distance.

And..

We discussed about content, expanse and distance?

BUT we have not looked at objects so far.

Focus of this Class

Are objects important for scene understanding?

Which portions of brain encode information about object content and spatial layout?

Are objects important for scene understanding?

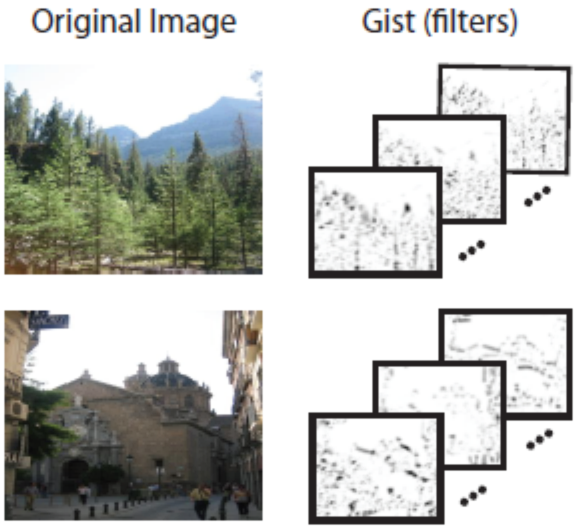
Computer Vision ☐ Human Brain ☐

A computer vision perspective

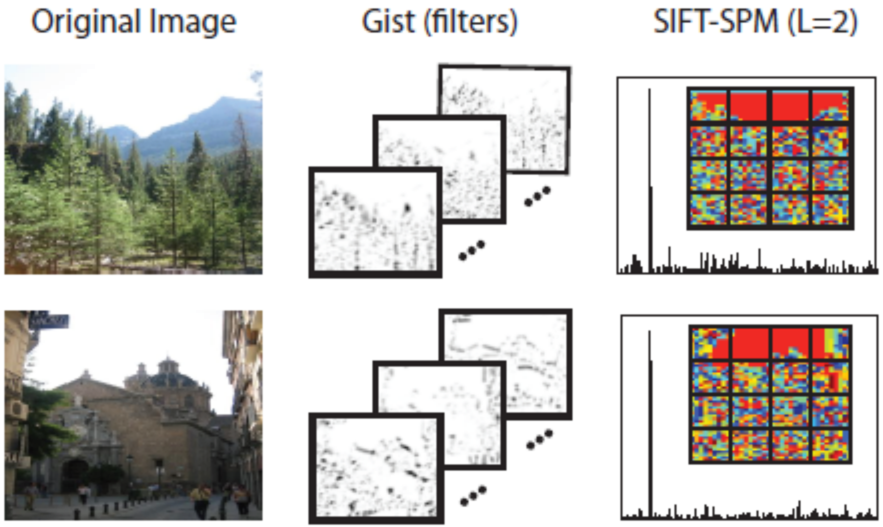
Object Bank: A High-Level Image Representation for Scene Classification & Semantic Feature Sparsification

Li-Jia Li*[1], Hao Su*[1], Eric P. Xing[2], Li Fei-Fei[1]
1 Computer Science Department, Stanford University
2 Machine Learning Department, Carnegie Mellon University

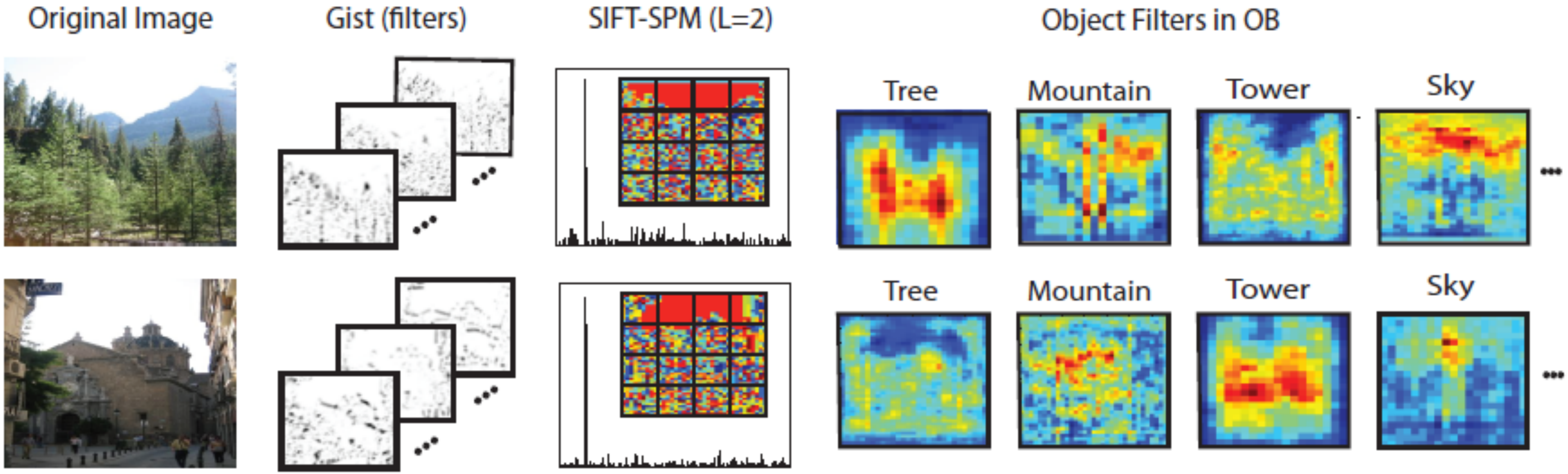
GIST



Spatial Pyramid Matching



What if we use objects?



Objects in Object Bank

1. Objects in ESP, LabelMe, ImageNet and the Flickr Photos were ranked according to their frequencies in each dataset.

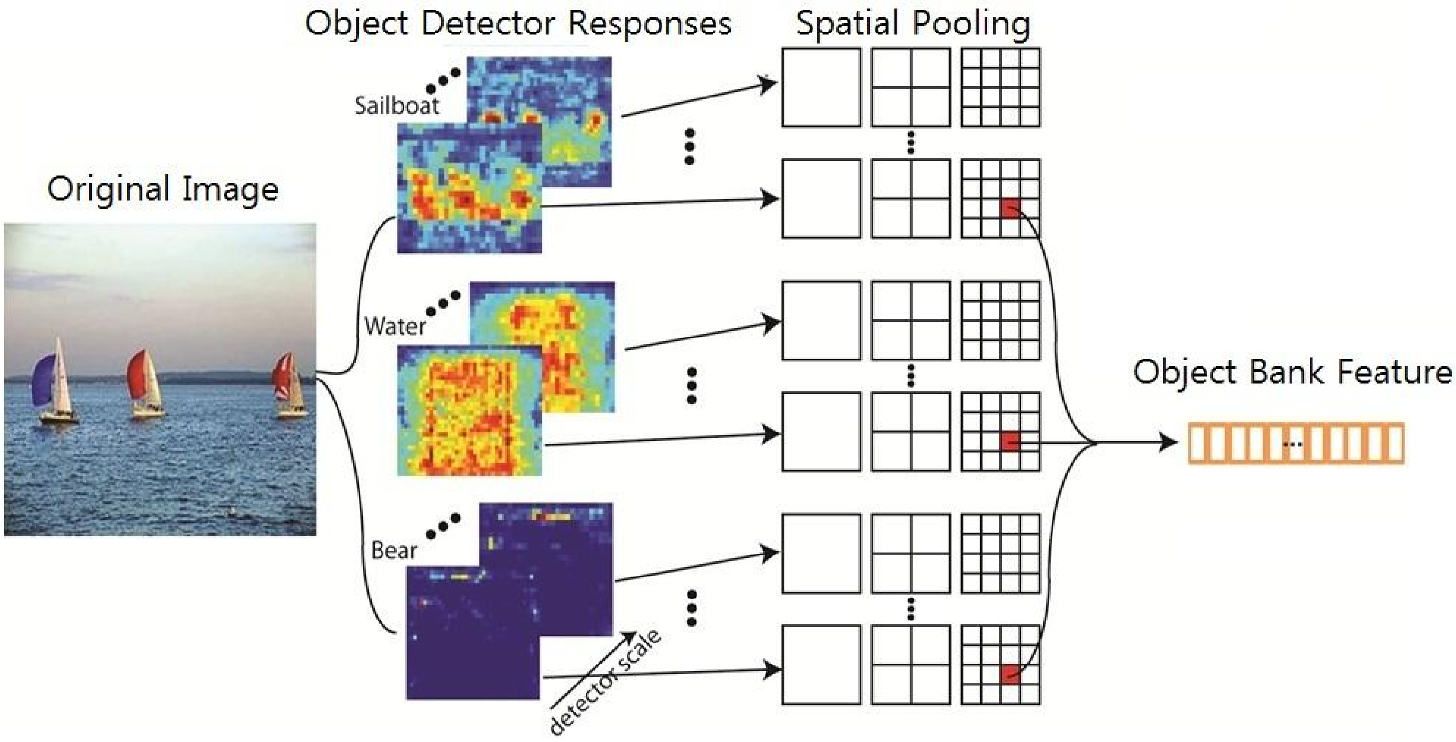
2. The intersection of top 1000 objects from each dataset resulted in 200 objects in Object Bank.

Detectors

1. Pedro’s latent SVM object detector for most of blobby objects such as tables, humans, cars etc.

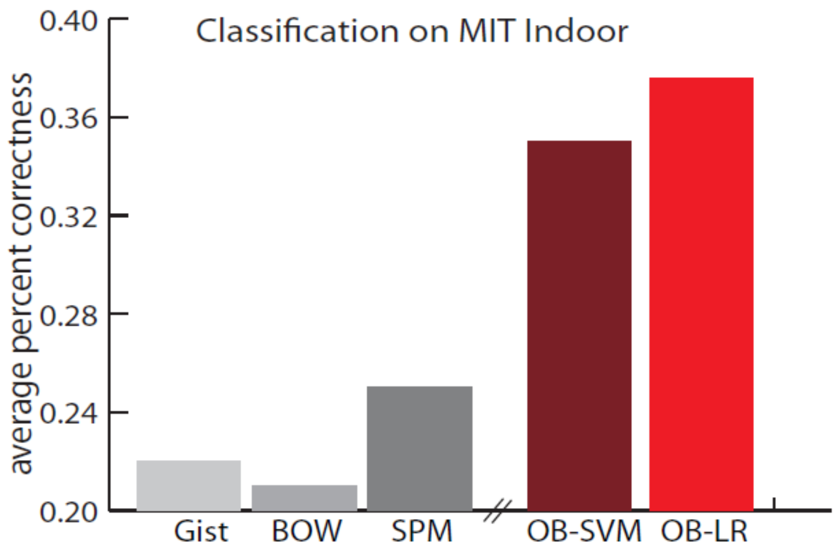
2. Derek’s texture classifier for more texture- and material-based objects such as sky, road, sand etc.

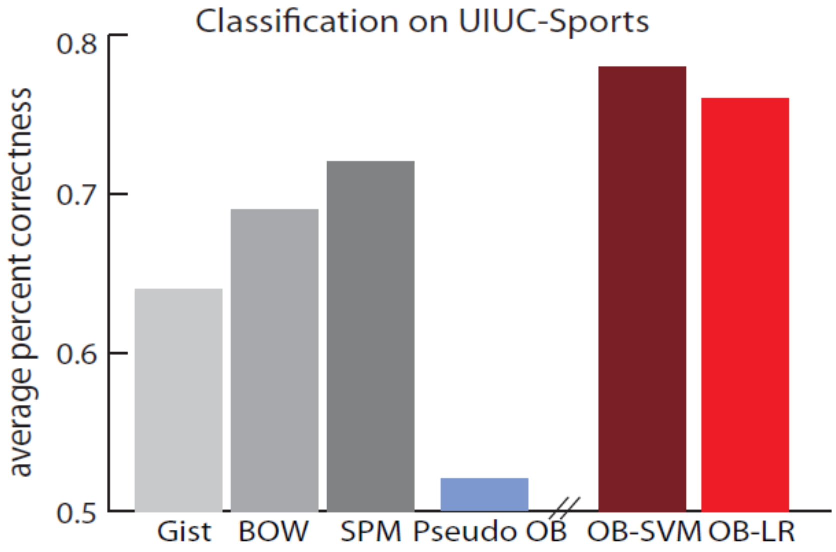
How does object bank approach work?



Li et.al. NIPS 2010

Evaluation





Li et.al. NIPS 2010

Are objects important for scene understanding?

Computer Vision ☑ Human Brain ☐

A neuroscience perspective

Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex

Dustin E. Stansbury,[1] Thomas Naselaris,[2,4] and Jack L. Gallant[1,2,3,*]
[1]Vision Science Group
[2]Helen Wills Neuroscience Institute
[3]Department of Psychology
University of California, Berkeley, CA 94720, USA

Inferring Scenes

Beach

Office

Living Room

Inferring Scenes

Probably '***objects***' helped us in inference..

Beach

<u>Objects</u> -
sky,
water waves,
sand etc.

Office

<u>Objects</u> -
table, chair,
monitor,
keyboard etc.

Living Room

<u>Objects</u> -
sofa,
table etc.

What comes to your mind

when you hear following words -

1. Beach
2. Kitchen
3. Office
4. Living Room

Probably objects!

1. Beach

   Sky, water waves, palm tree, sand, people etc.

2. Kitchen

   Stove, utensils, refrigerator, oven etc.

3. Office

   Table, chair, computer, books etc.

4. Living Room

   Sofa, table etc.

Probably objects!

These observations intute that humans use knowledge about how objects co-occur in the natural world.

sand,
people etc.

oven etc.

books etc.

But

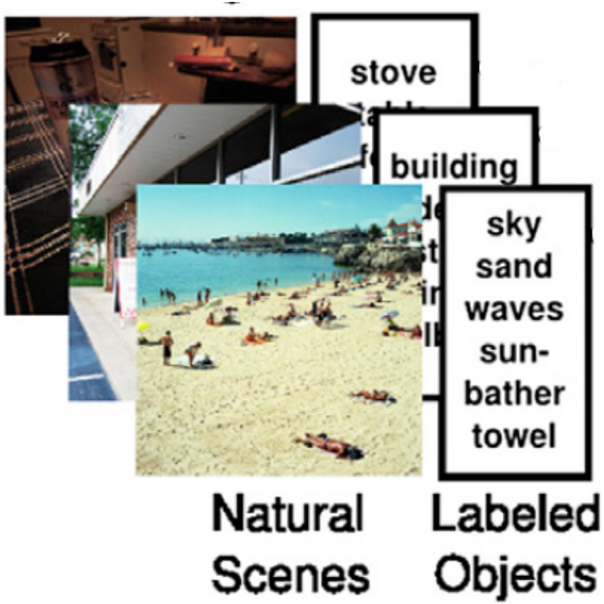
Can we define scene categories in terms of object co-occurrences themselves?

But

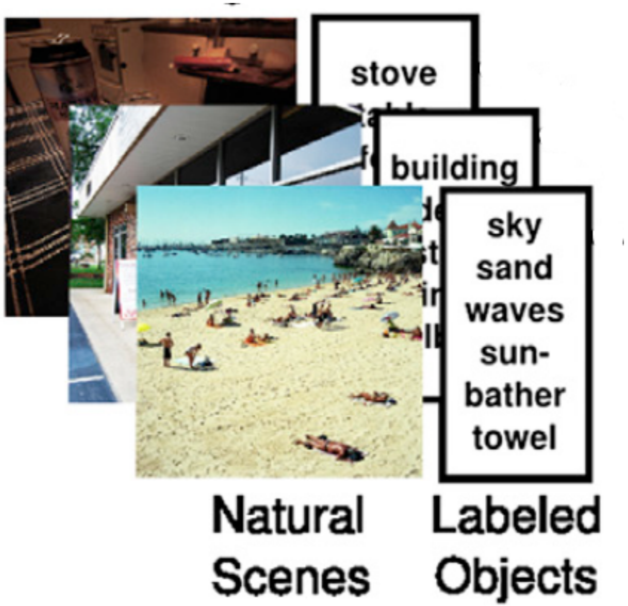
Can we define scene categories in terms of object co-occurrences themselves?

Does human brain represent scene categories in this manner?

For Example: Given Natural Scenes & Labeled Objects



For Example: Given Natural Scenes & Labeled Objects



Can we learn the objects which co-occur?

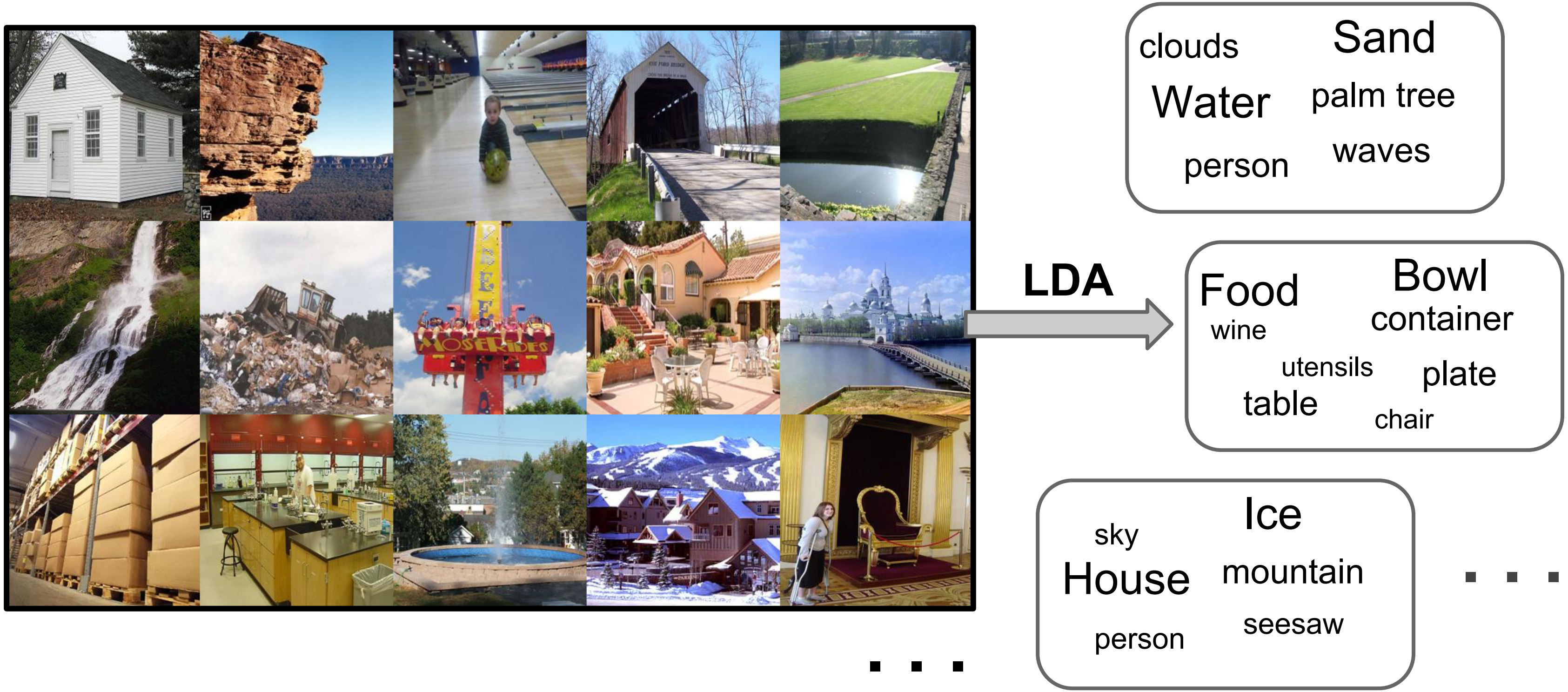
Latent Dirichlet Allocation (LDA)

LDA = topic-modeling

- learns an underlying set of scene categories that capture the co-occurrence of objects in database.

- defines each scene category as a list of probabilities that are assigned to each of the object labels within an available vocabulary.

Recovering intrinsic categorical structure of natural scenes



All the visible objects were labelled in library of 4116 natural scene images.

Examples

1	2	3	. . .	N
table	**animal**	**desk**		**food**
sofa	**ocean**	**chair**	. . .	**container**
wall	**fish**	monitor		**bowl**
floor	water	wall		**table**
decoration	mammal	book		beverage
window	seal	keyboard		plate
ceiling	coral	floor		wine
lamp	boulder	.		utensils
.	.	.		.
.	.	.		.
.	.			.

Examples

Hand - labeled afterwards

"Living Room"	"Aquatic"	"Office"	. . .	"Dining"
table	**animal**	**desk**		**food**
sofa	**ocean**	**chair**	. . .	**container**
wall	**fish**	monitor		**bowl**
floor	water	wall		**table**
decoration	mammal	book		beverage
window	seal	keyboard		plate
ceiling	coral	floor		wine
lamp	boulder	⋮		utensils
⋮	⋮			⋮

Output from LDA

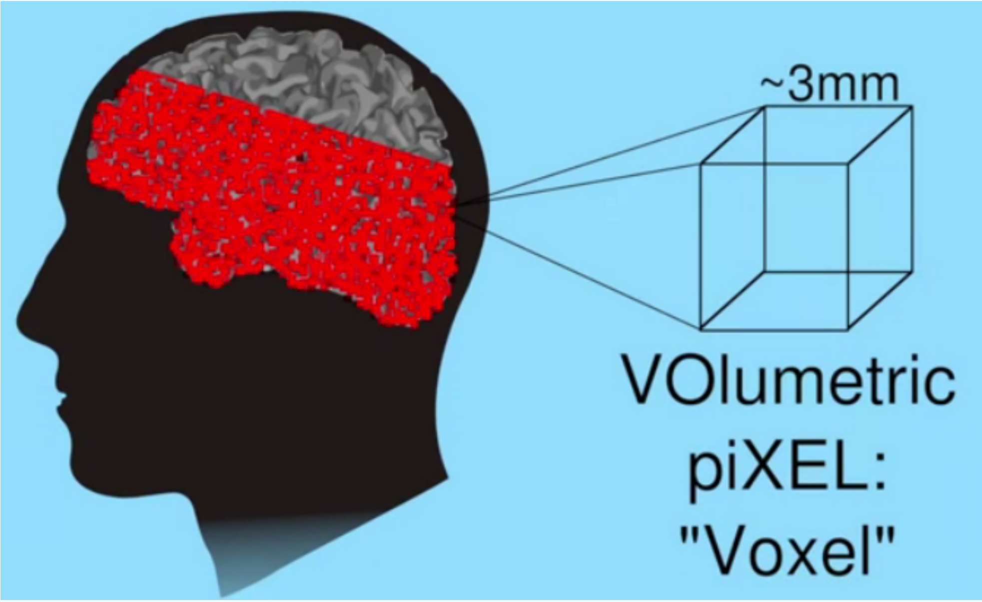
Human Studies

fMRI data

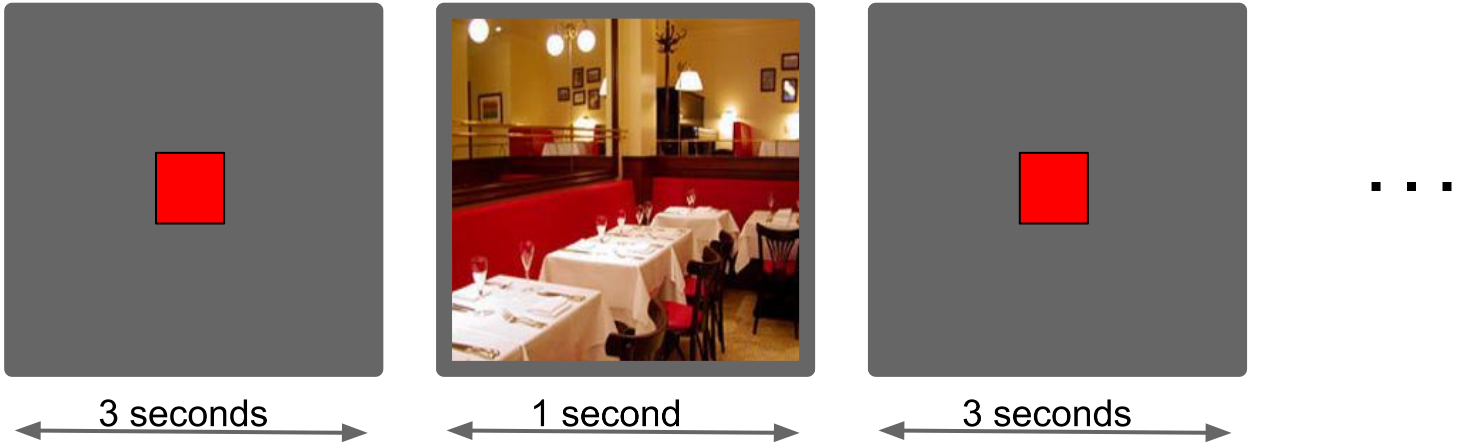
4 human subjects viewed 1,260 images for the experiment.

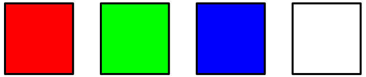
The voxels used were nearly 3 mm in side.

Approx. 20, 000 of these voxels were studied in visual cortex area in each subject



Stimuli





Fixation square were randomly permuted at 3 Hz in color to facilitate fixation.

1,260 stimulus scenes in the estimation set were sampled from the learning database.

Can we use scene-category probabilities to predict voxel responses?

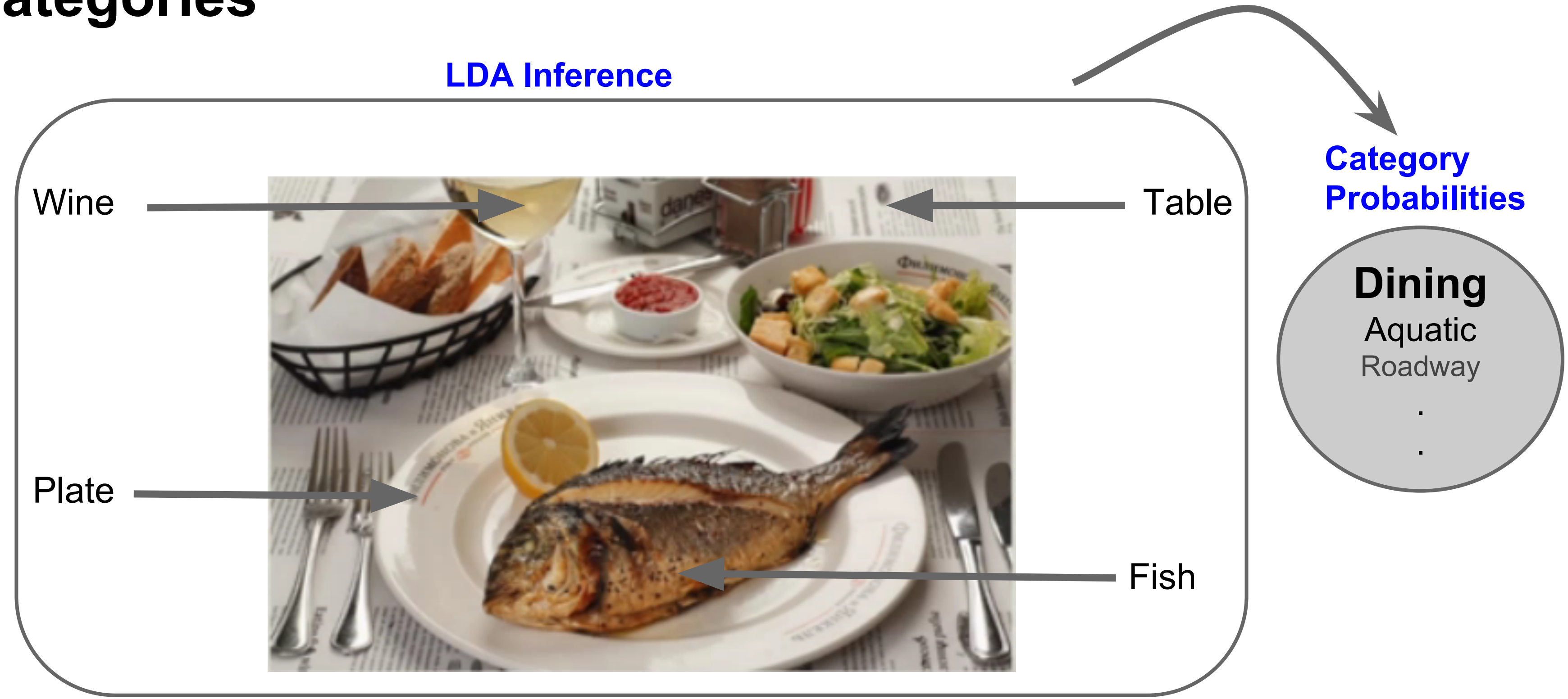
Voxelwise Encoding Models Based on Learned Scene Categories

Stimulus Scene

Voxelwise Encoding Models Based on Learned Scene Categories

Labelled Objects

Voxelwise Encoding Models Based on Learned Scene Categories



Voxelwise Encoding Models Based on Learned Scene Categories

LDA

fMRI data

Category Probabilities

Dining
Aquatic
Roadway
.
.

x1
x2
x3
x4
...
xN

Voxelwise Encoding Models Based on Learned Scene Categories

LDA

fMRI data

Category Probabilities

Dining
Aquatic
Roadway
.
.

X

x1
x2
x3
x4
...
xN

For each voxel,
learn weight '**w**' s.t. **X** can be mapped to **Y**.

Y

Voxelwise Encoding Models Based on Learned Scene Categories

LDA

fMRI data

Category Probabilities

Dining
Aquatic
Roadway
.
.

X

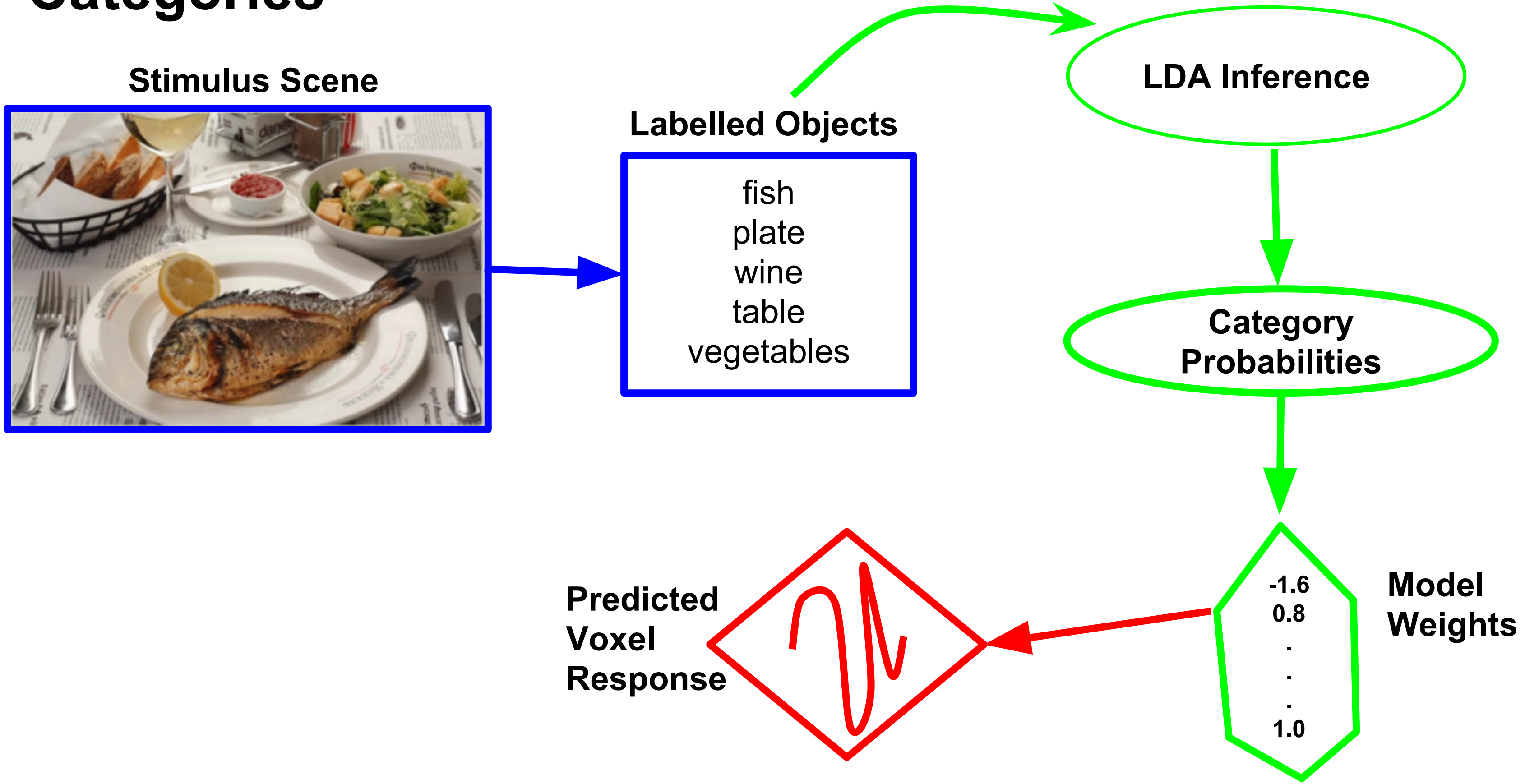
x1
x2
x3
x4
...
xN

For each voxel,
learn weight '**w**' s.t. **X** can be mapped to **Y**.

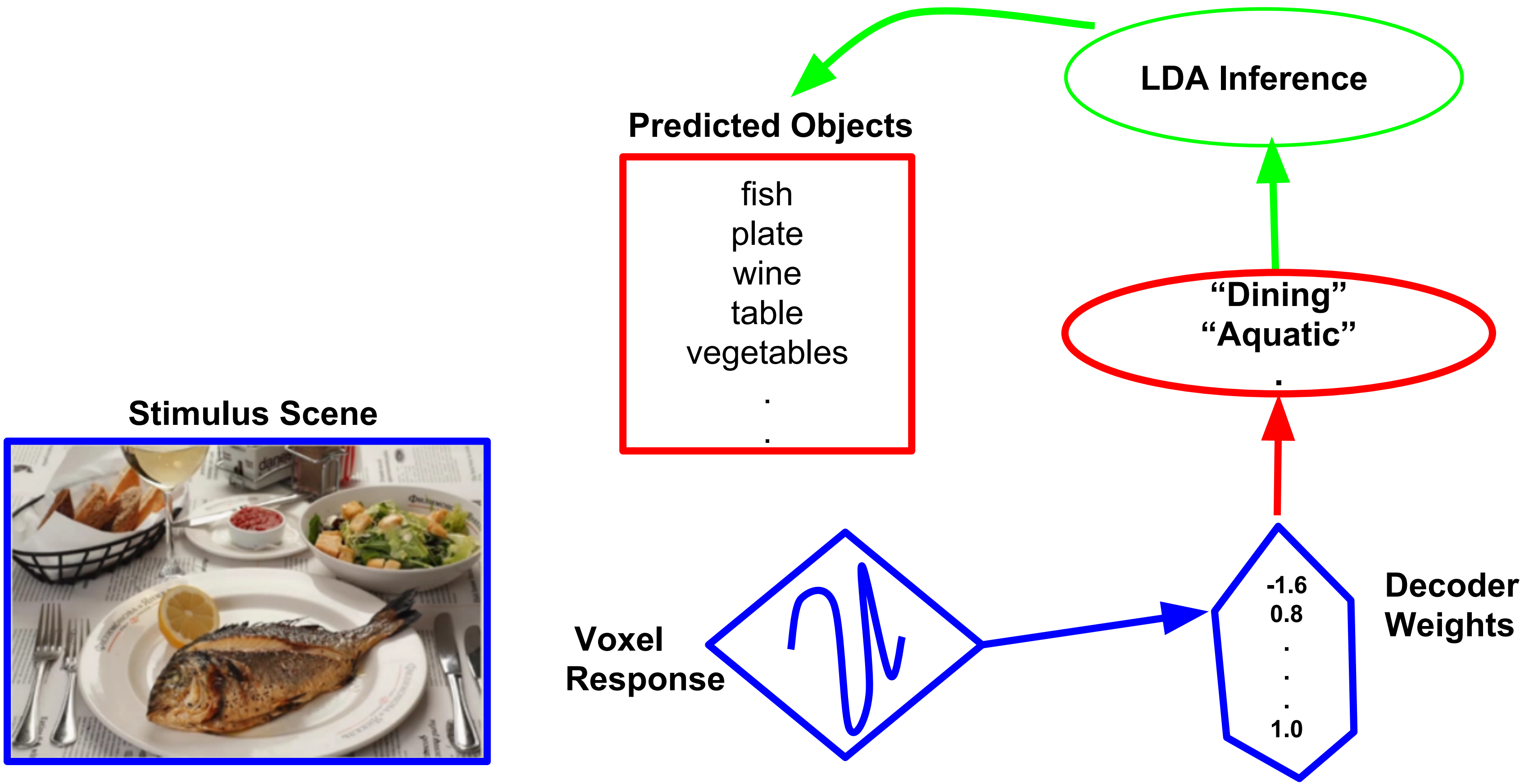
Model weights were estimated using regularized linear regression applied independently for each subject and vote.

Y

Voxelwise Encoding Models Based on Learned Scene Categories



Similarly we can have a Decoding Model by reversing..



Decoding Novel Scenes

Stimulus Scene

Harbor and Skyline Scene

Decoding Novel Scenes

Stimulus Scene

Harbor and Skyline Scene

Predicted Category Probabilities

Urban/Street
Boatway

Decoding Novel Scenes

Stimulus Scene

Harbor and Skyline Scene

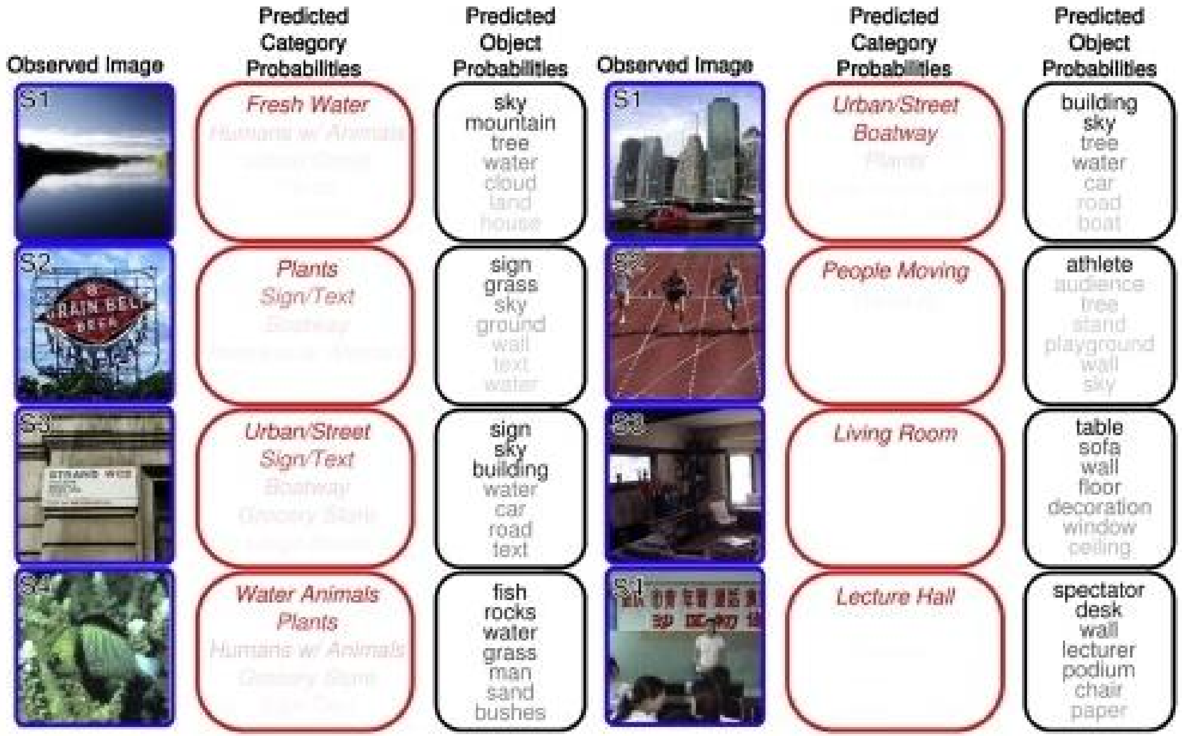
Predicted Category Probabilities

Urban/Street
Boatway

Predicted Object Probabilities

building
sky
tree
water
car
road

More Examples



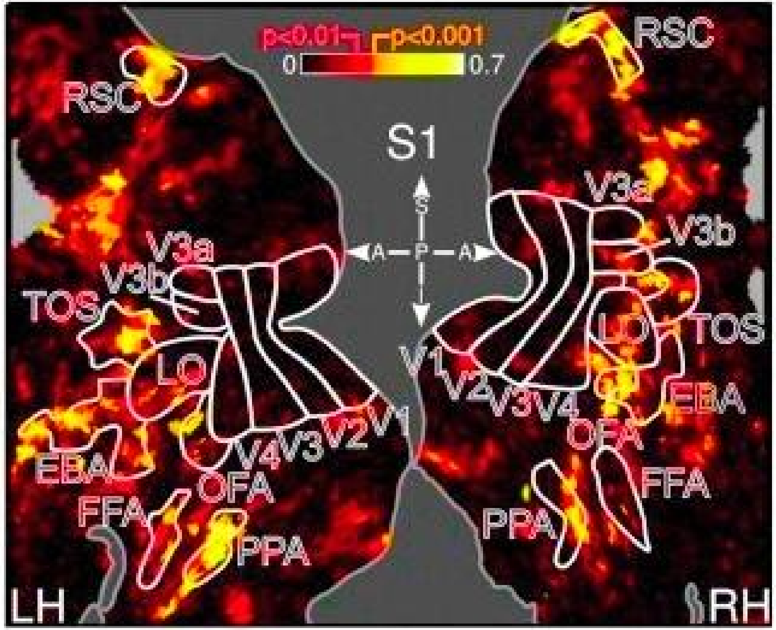
Stansbury et.al. Neuron 2013

Now we know Encoding Models

Now we know Encoding Models

We need to see the performance of these encoding models..

Encoding model performance



1. Gray indicate areas outside of the scan boundary.

2. Bright locations indicate voxels that are accurately predicted by the corresponding encoding model.

3. ROIs identified in separate retinotopy and functional localizer experiments are outlined in white.

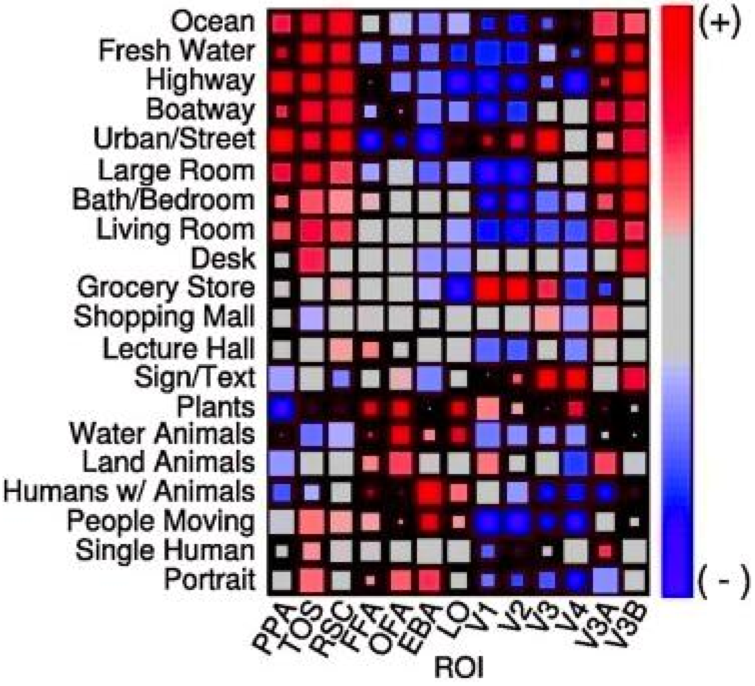
Take Home Message: The data shows that the encoding models accurately predict responses of voxels located in many ROIs with anterior visual cortex.

Stansbury et.al. Neuron 2013

Question

Can selectivity in these regions be explained in terms of the categories learned from the natural scene object statistics?

Average Encoding Model Weights



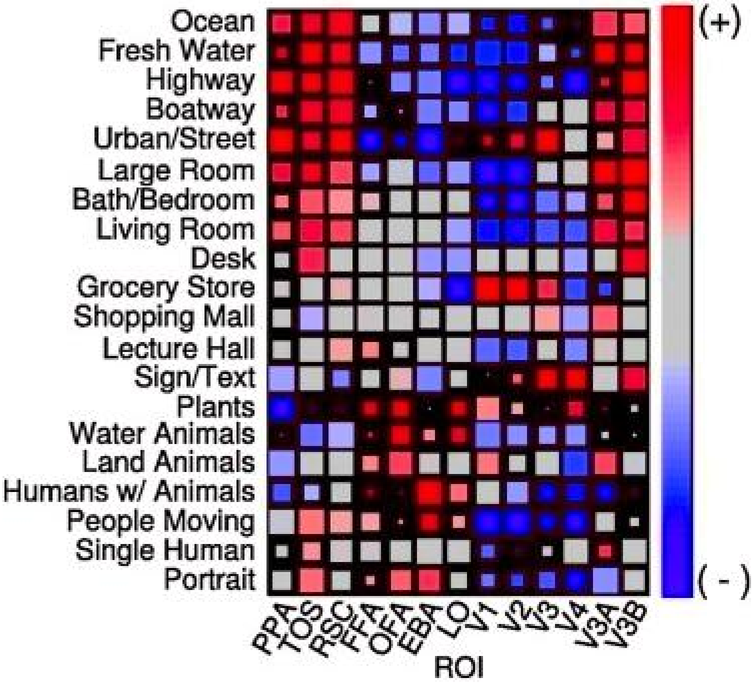
Stansbury et.al. Neuron 2013

Scene Category Selectivity Examples

[Epstein and Kanwisher, 1998] -
PPA is selective for presence of buildings.

LDA Algorithm -
Images containing buildings are most likely to belong to the "Urban/Street" category.

Average Encoding Model Weights



Stansbury et.al. Neuron 2013

Scene Category Selectivity Examples

[Gauthier et al., 2000] -
OFA is selective for presence of human faces.

LDA Algorithm -
Images containing faces are most likely to belong to the "Portrait" category.

Conclusions

1. Categories that capture co-occurrence statistics are consistent with their intuitive interpretations of natural scenes.

2. Voxelwise encoding models based on these categories accurately predict visually evoked BOLD activity across much of anterior visual cortex.

Are objects important for scene understanding?

Computer Vision ☑ Human Brain ☑

Previous Class:

1. PPA encodes spatial layout.
2. Spatial Layout is most important for scenes.

This Class:

Objects co-occurrences define scenes….

Previous Class:

1. PPA encodes spatial layout.
2. Spatial Layout is most important for scenes.

This Class:

Objects co-occurrences define scenes….

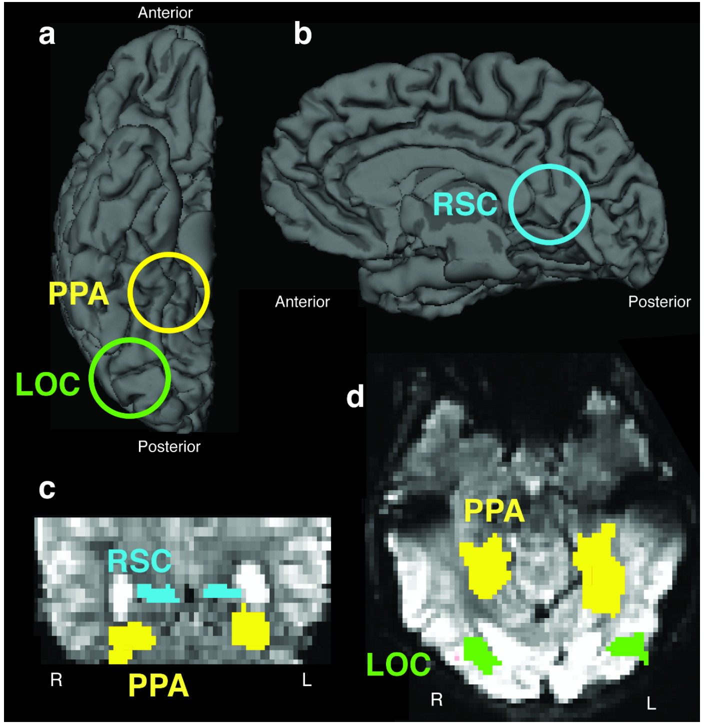
Probably BOTH HAPPEN

Questions

Is object content and spatial layout information stored in different regions? If yes, how are they connected?

What brain regions should we look at?

Structure of Brain



1. LOC lies within visual ventral path.

2. PPA is connected with both the dorsal and ventral stream.

3. RSC is strongly connected with posterior parietal cortex.

Supplementary Material, Harel et.al. Cerebral Cortex 2013

Hypothesis

1. LOC lies in the ventral visual pathway.

Hypothesis

1. LOC lies in the ventral visual pathway.

Since ventral visual pathway contains strong object information but no background, LOC might have more of object information.

Hypothesis

1. LOC lies in the ventral visual pathway.

Since ventral visual pathway contains strong object information but no background, LOC might have more of object information.

2. RSC lies in posterior parietal cortex.

Hypothesis

1. LOC lies in the ventral visual pathway.

Since ventral visual pathway contains strong object information but no background, LOC might have more of object information.

2. RSC lies in posterior parietal cortex.

There is a strong spatial layout information in PPC or dorsal stream. So there can be possibly spatial layout information in RSC.

Hypothesis

1. LOC lies in the ventral visual pathway.

Since ventral visual pathway contains strong object information but no background, LOC might have more of object information.

2. RSC lies in posterior parietal cortex.

There is a strong spatial layout information in PPC or dorsal stream. So there can be possibly spatial layout information in RSC.

3. PPA lies in between dorsal and ventral stream.

Hypothesis

1. LOC lies in the ventral visual pathway.

Since ventral visual pathway contains strong object information but no background, LOC might have more of object information.

2. RSC lies in posterior parietal cortex.

There is a strong spatial layout information in PPC or dorsal stream. So there can be possibly spatial layout information in RSC.

3. PPA lies in between dorsal and ventral stream.

PPA might have both object and spatial layout information.

How to control spatial layout and objects?

Stimuli

Objects

Bed

Crib

Desk

Dresser

Sofa

Stove

Table

Stimuli

Backgrounds

Space Present (Closed)

Space Present (Open)

Space Absent (Gradient)

Stimuli

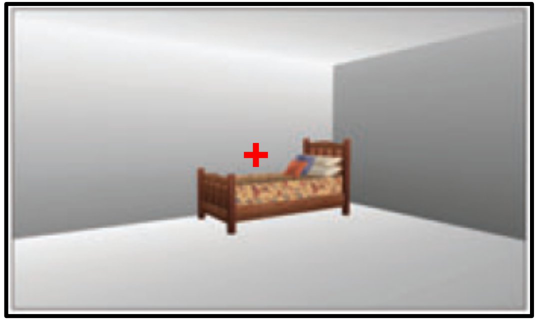
Minimal Scenes

8 objects (7 objects + no object) x 3 backgrounds (open + close + gradient)= 24 scenes

24 scenes x 2 flips x 2 repetitions = 96 trials per run

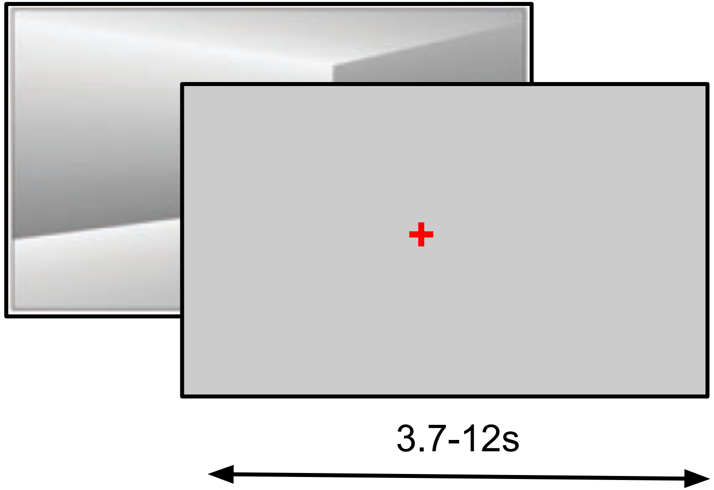
And there are a total of 6 runs..

fMRI Experiment

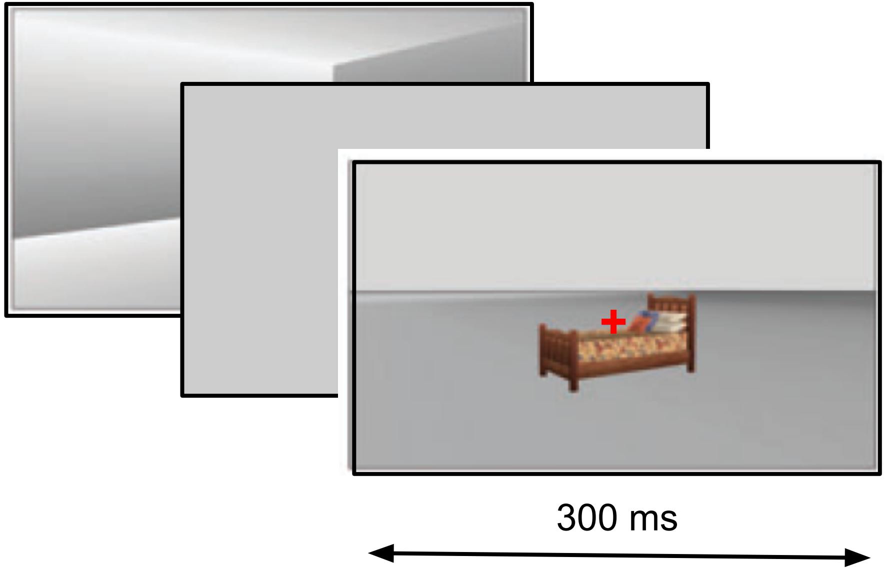


300 ms.

fMRI Experiment



fMRI Experiment



Lets look at activations

Which region has highest differential activation with/without objects?

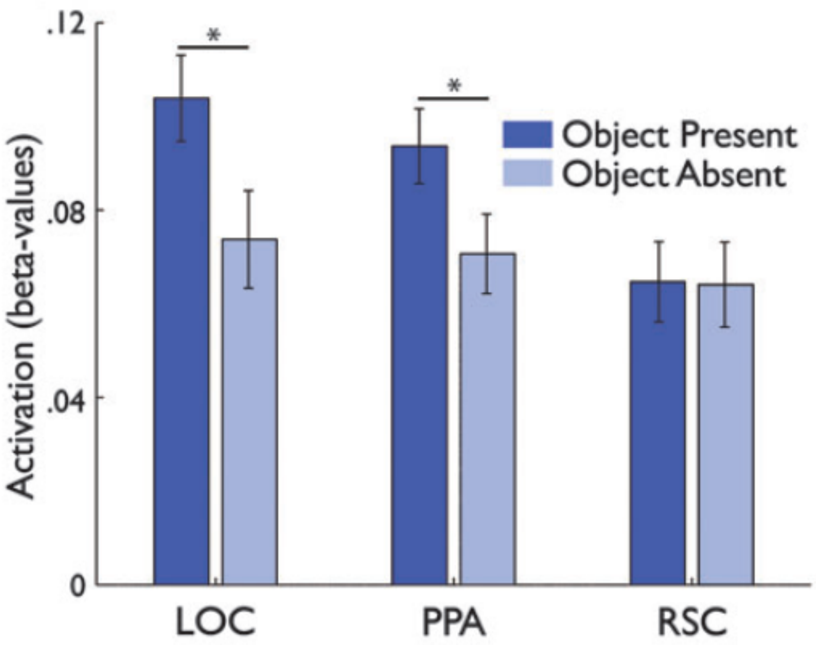
Lets look at activations

Which region has highest differential activation with/without objects?

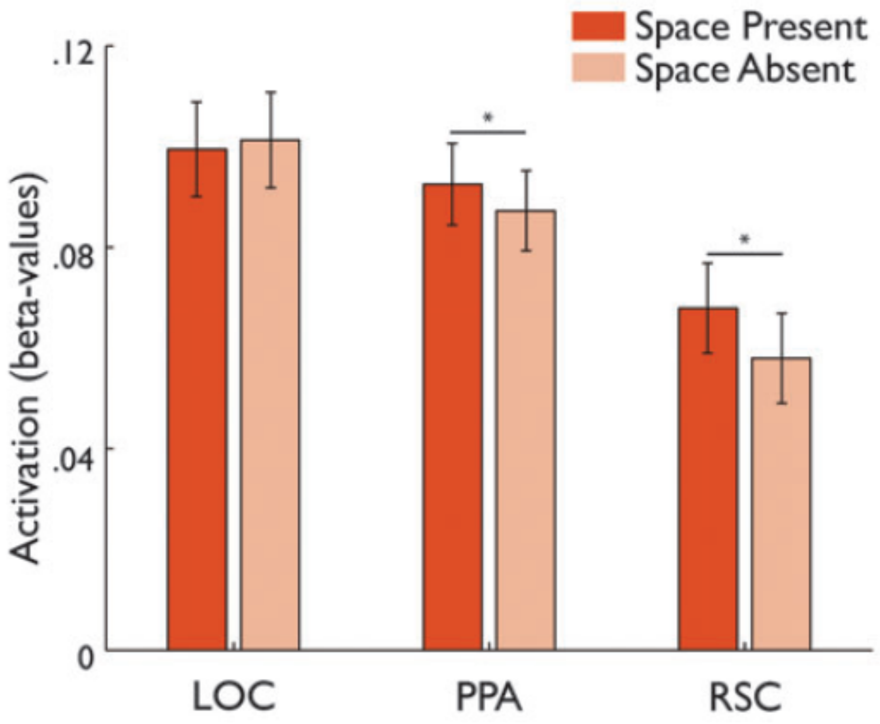
High response when object is present

Low response when object is absent

Object Information across background



Similarly for Scenes



From Activation, we see

1. Object information is prominent in LOC and PPA.

2. Spatial layout information is prominent in RSC and PPA.

But

Can we look just at activations and predict whether scene/background is present or absent?

OR

Can we look at activations and predict object identity? And which region is good at it?

Ideal

Learn SVM on some data and test on held-out

Ideal

Learn SVM on some data and test on held-out

But data is scarce..

Ideal

Learn SVM on some data and test on held-out

But data is scarce..

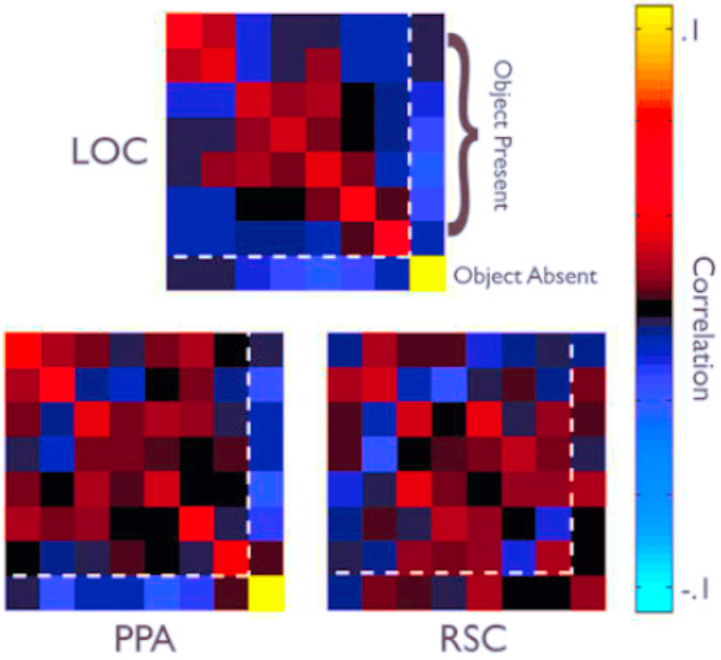
So let us look at correlation differences..

What we want?

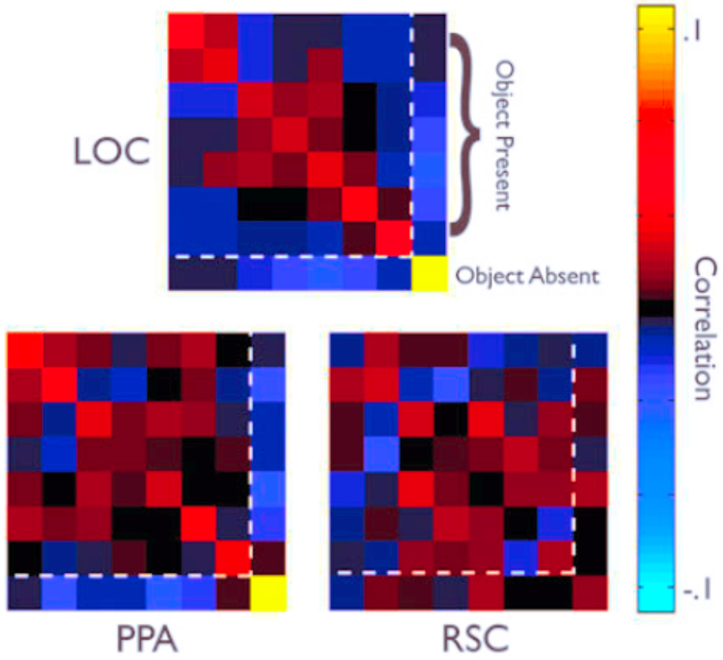
If in a region A, correlation within beds is high as compared to correlation between beds and cupboard, beds and chair.

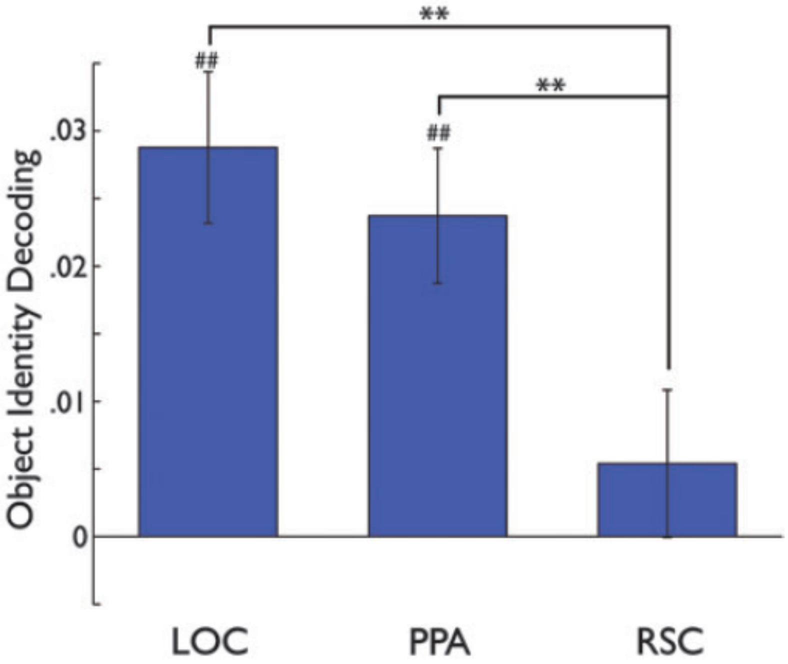
This implies Bed can be decoded using this region.

Object Identity Decoding

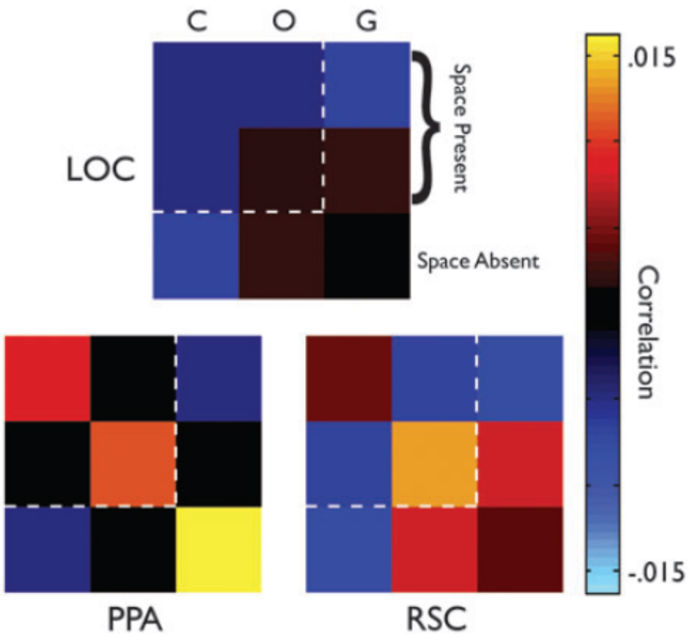


Object Identity Decoding

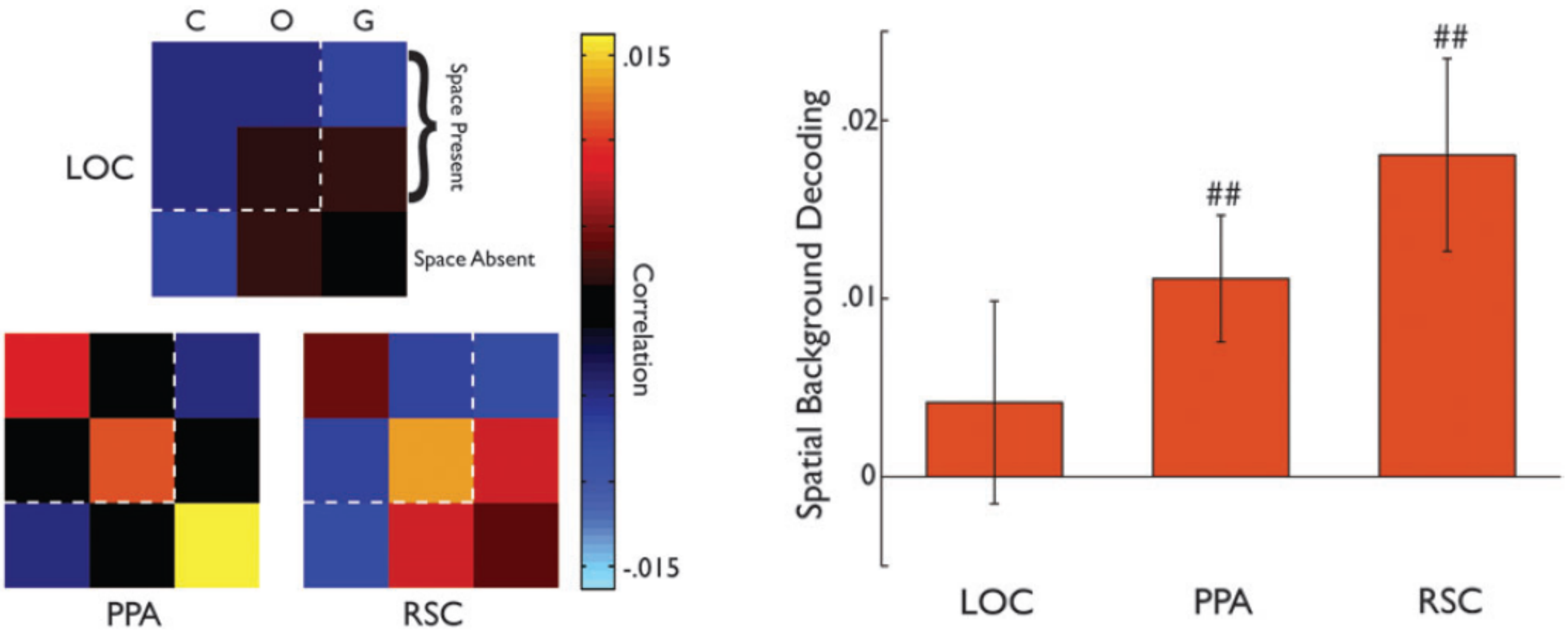




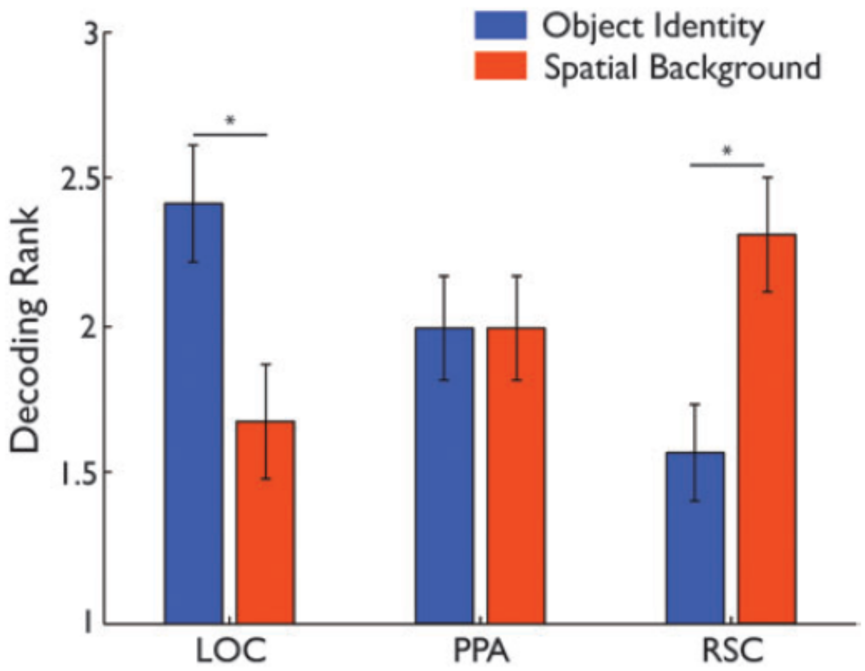
Spatial Background Decoding



Spatial Background Decoding



Combining both Object and Spatial Background



Uptil Now

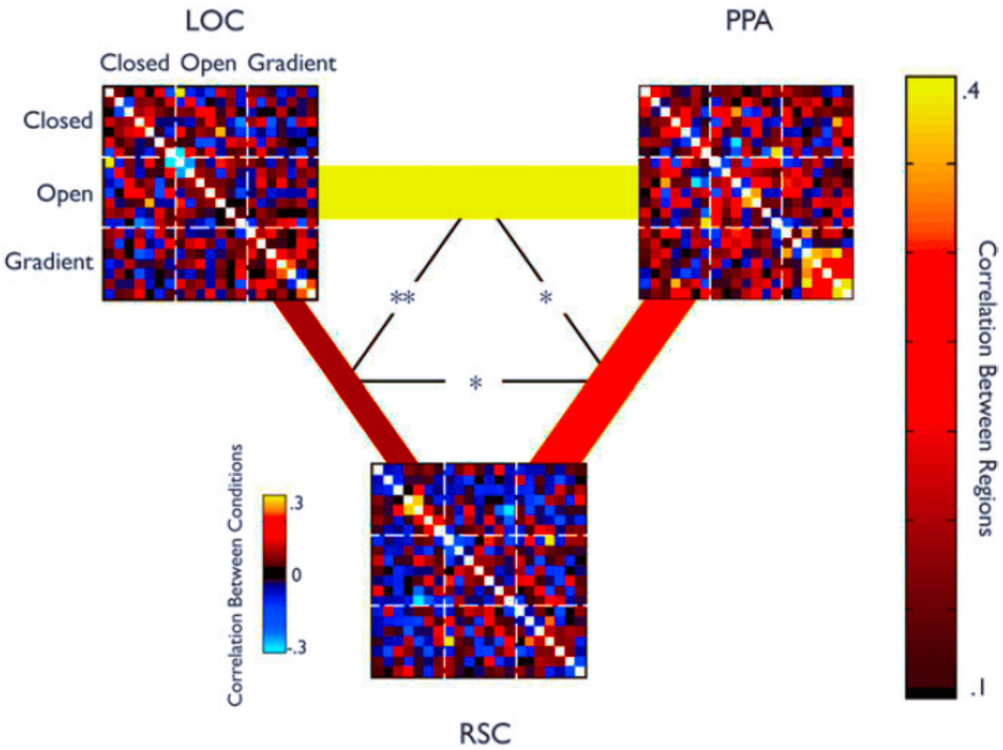
1. The studies suggest that both object and spatial layout are important for scene understanding.

2. Object information is encoded in LOC and PPA, whereas spatial layout information is encoded in RSC and PPA.

Question

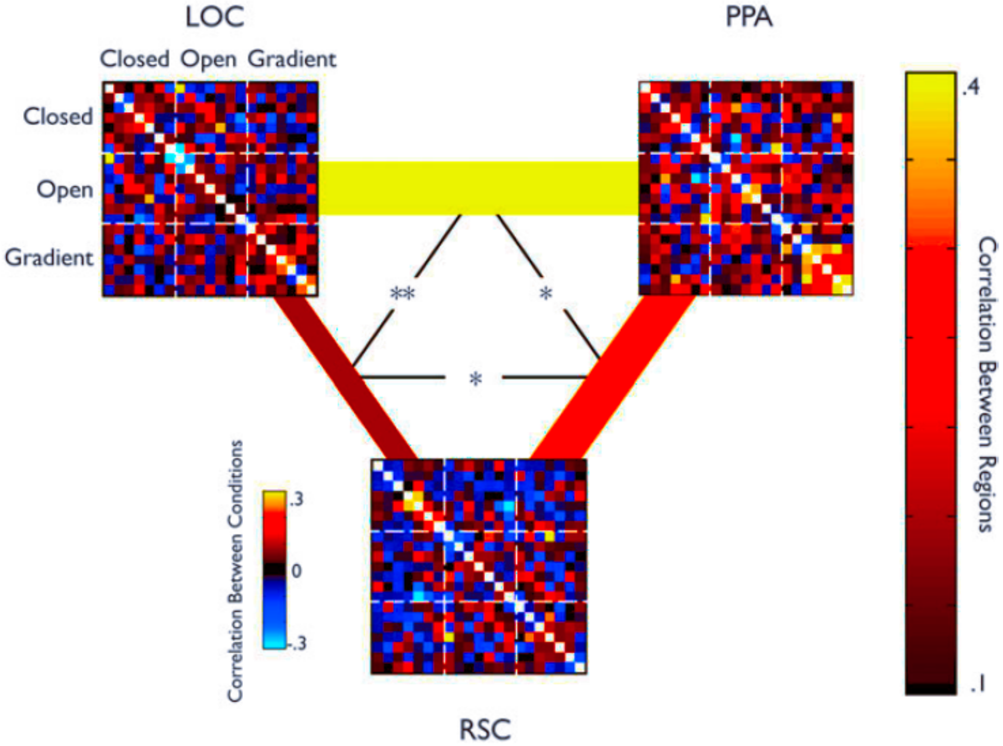
Are these regions (LOC, PPA, and RSC) linked with each other? If Yes, How?

Structure of Representation



Harel et.al. Cerebral Cortex 2013

Structure of Representation



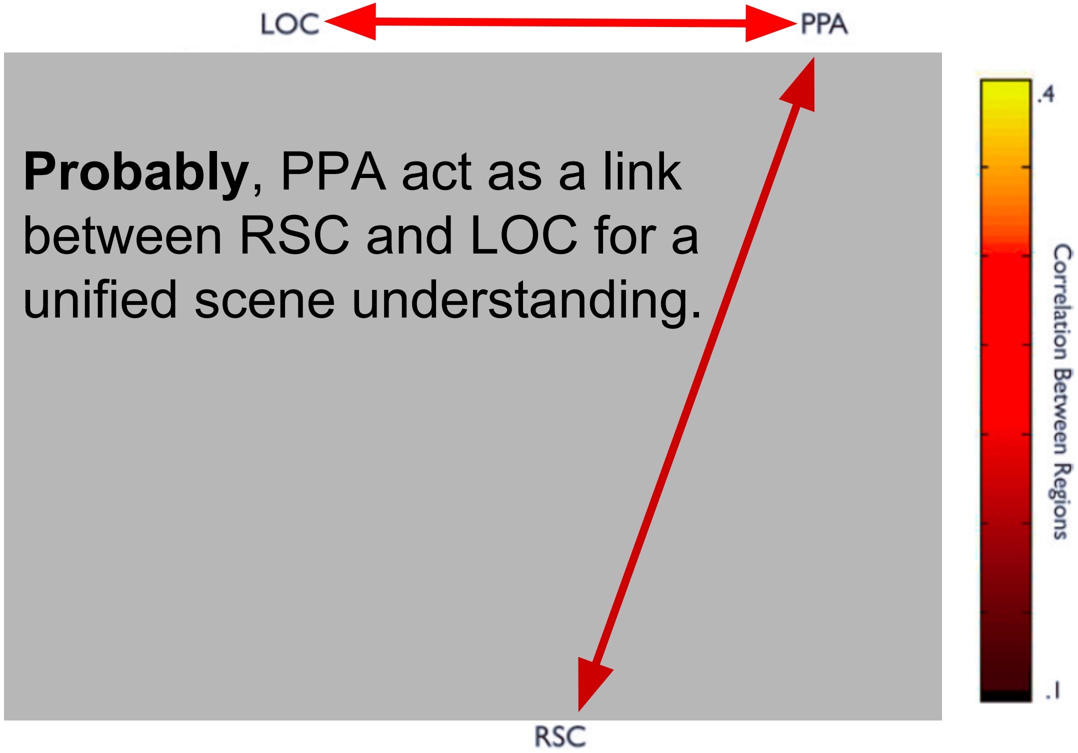
1. Stronger correlations were found in PPA and RSC than between RSC and LOC.

2. LOC was more strongly connected to PPA than RSC.

3. PPA was more strongly connected to LOC than RSC.

Harel et.al. Cerebral Cortex 2013

Structure of Representation



Probably, PPA act as a link between RSC and LOC for a unified scene understanding.

Harel et.al. Cerebral Cortex 2013

Finally, Focus of this Class

Are objects important for scene understanding?

Which portions of brain encode information about object content and spatial layout?

Finally, Focus of this Class

Are objects important for scene understanding?

Yes, Objects seems to be important for scene understanding.

Which portions of brain encode information about object content and spatial layout?

Whereas LOC and PPA encode object information, RSC and PPA encode spatial layout information.

Discussion

There may be bias in results due to objects (furniture) used in dataset.