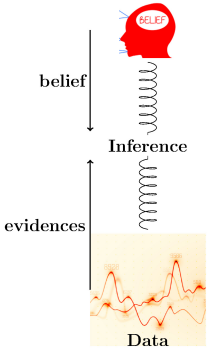
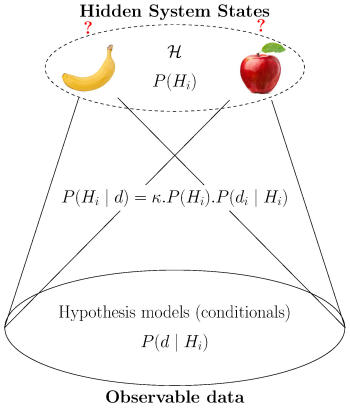
Biological Vision and Applications
Module 04-03: Object recognition

Hiranmay Ghosh

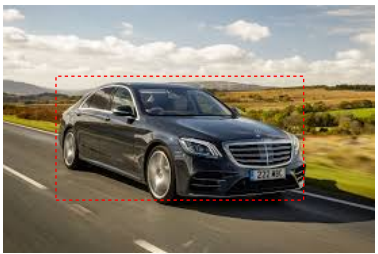
Bayesian Model for object recognition





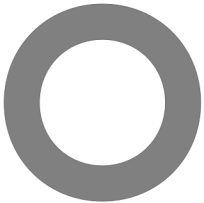
Bayesian Model for object recognition

- $O^* = \text{argmax}_i P(O_i \mid v)$
- when
  - $P(O_i \mid v) = \frac{P(O_i).P(v|O_i)}{P(v)}$
  - O_i = Object hypothesis
  - v = Visual features
- Context contributes to the visual features of the image
  - $v = (v_l, v_c)$ where
    - v_l = Object features
    - v_c = Context features
- In traditional object recognition
  - v_c is minimized
  - $v_l \approx v$



- Can we ignore the context ?

What is the object in this picture ?



Context matters !

- Seeing the whole provides the cues for identifying the parts

A practical example

Context is especially useful for imperfect images

- Context is especially useful for robust interpretation in imperfect images
  - Ambiguous features, blur, occlusion, clutter, etc.

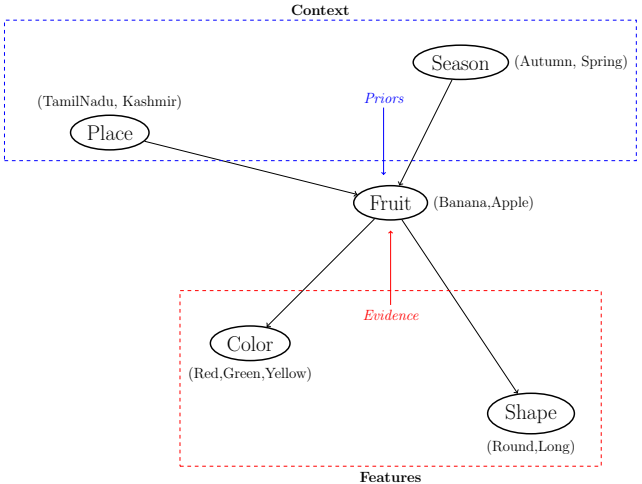
In-context object recognition

We drop the suffix i for convenience

- $P(O \mid v) = k.P(O).P(v \mid O), \qquad [v = (v_l, v_c)]$
- In traditional object recognition $v \approx v_l$
  - $P(O \mid v_l) = k.P(O).P(v_l \mid O)$
- $P(O \mid v_l, v_c) = k'.P(O \mid v_c).P(v_l \mid O, v_c)$ [*Please deduce*]
  - $P(O \mid v_c)$: Prior probability of the object to appear ... in a specific context
  - $P(v_l \mid O, v_c)$: The model of visual feature of an object ... in a specific context
- We can assume, visual features of an object is independent of context: $P(v_l \mid O, v_c) = P(v_l \mid O)$
  - Has some other significance that we shall analyze in a later lesson

Torralba. Contextual Priming for Object Detection (2003)

Programming assignment 2



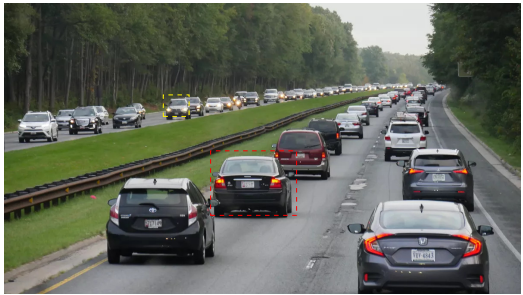
The context (in image)

$P(O \mid v_c)$: v_c = visual feature of the context

- $P(O \mid v_l, v_c) = k.P(O \mid v_c).P(v_l \mid O, v_c)$
- Let O not represent just an object class
  - Modeling the visual features with just the class information is too crude
  - Let $O = (o, x, \sigma)$ where
    - o: object class
    - x: location in image
    - σ: appearance (scale, orientation, etc.)
  - $P(O \mid v_c)$ represents an object of a class to appear in a specific location in an image with a certain appearance
- $P(O \mid v_c) = P(o, x, \sigma \mid v_c) = P(\sigma \mid o, x, v_c).P(x \mid o, v_c).P(o \mid v_c)$

In-context object recognition

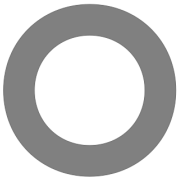
Significance of the decomposition

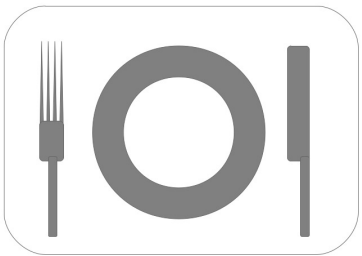


- $P(o \mid v_c)$: Probability of an object class to appear in a context
- $P(x \mid o, v_c)$: Probability of the location where an object class appears in a context
- $P(\sigma \mid o, x, v_c)$: Probability of the appearance of an an object class when it appears in a certain location in an image

- The prior probabilities are determined by the context

How do characterize a context





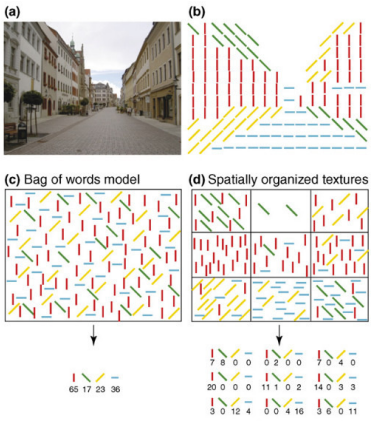
- Plate is recognized by it's context
- Other objects in the scene creates the context
  - Fork, knife, table-mat
- How do you recognize those objects?
  - A chicken-and-egg problem?

Can we see “forest before the trees” ?

Do the scenes have some distinctive features?

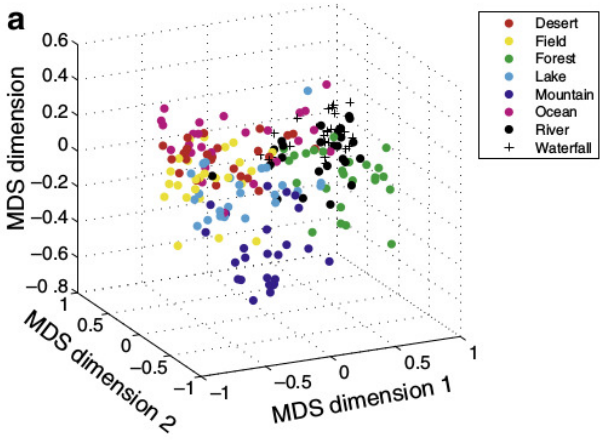
Spatial envelop representation

A holistic representation of a scene layout



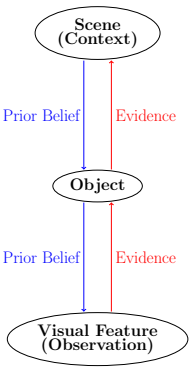
- The edges in a scene constitutes a definite pattern
  - Statistical pattern characterizes a scene
- Recall natural scene statistics
- Happens in early (pre-attentive) vision – fast
- Two types of feature descriptors
  - Global statistics
  - Local statistics
- We skip the detailed mathematical formulation

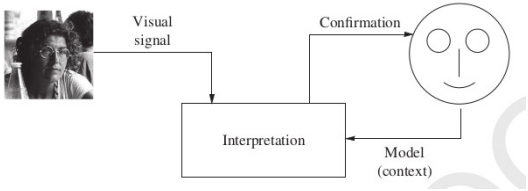
Distinguishing scene classes with spatial envelop representation



Oliva & Torralba. Modeling the Shape of the Scene: ...

Vision as a synthesis of top-down and bottom-up process





- Object recognition is a combination of two processes
  - Top-down: Prior belief (scene context)
  - Bottom-up: Evidence (observation of feature)
- The face model and the face image mutually reinforce belief in each other
- The process is hierarchical

On hypothesis space

Spot the pug

- There are thousands of objects we are familiar with
  - Makes the hypothesis space very large
- Only the hypotheses endorsed by context are analyzed
  - Difficult to detect things at unexpected places

Quiz 04-03

End of Module 04-03