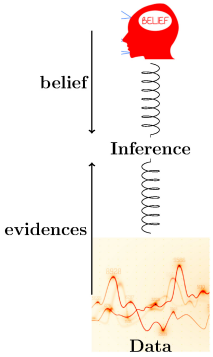
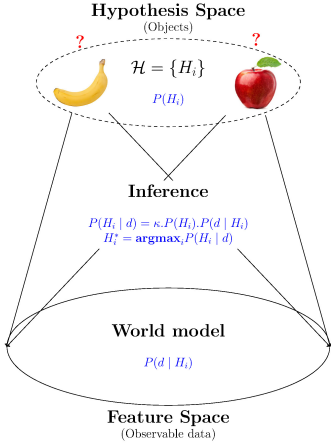
Biological Vision and Applications
Module 04-03: Object Recognition

Hiranmay Ghosh

Bayesian Model for object recognition





Bayesian Model for object recognition

$O^* = \mathbf{argmax}_i P(O_i \mid v)$

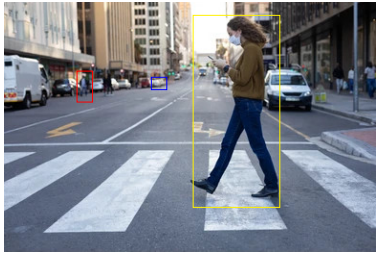
where

$P(O_i \mid v) = \frac{P(O_i).P(v|O_i)}{P(v)} = k.P(O_i).P(v \mid O_i)$

O_i: Object hypothesis

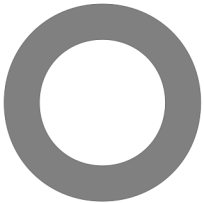
v: Observed visual features

- 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 ?

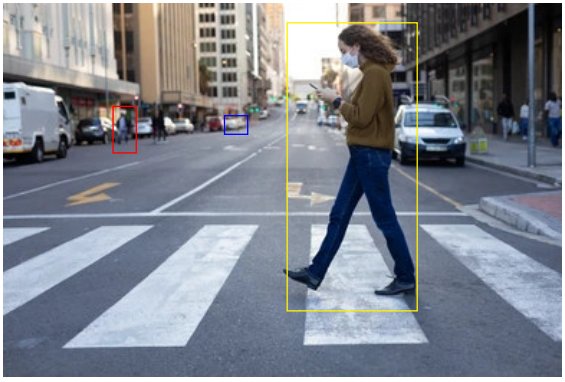
What is this object ?



Context matters !

- Seeing the whole provides the cues for identifying the parts

Another example



Context is especially useful for robust interpretation in imperfect images

- Ambiguous features, blur, occlusion, clutter, etc.

In-context object recognition

$P(O_i \mid v) = k.P(O_i).P(v \mid O_i), \quad v = (v_l, v_c)$

$P(O_i \mid v_l, v_c) = P((O_i \mid v_c) \mid v_l) = \frac{P(O_i|v_c).P(v_l|O_i,v_c)}{P(v_l)}$
$= \kappa.P(O_i \mid v_c).P(v_l \mid O_i, v_c)$

$P(O_i \mid v_c)$: Prob of O_i to appear in a specific context v_c
$P(v_l \mid O_i, v_c)$: Visual model for O_i in the same context v_c

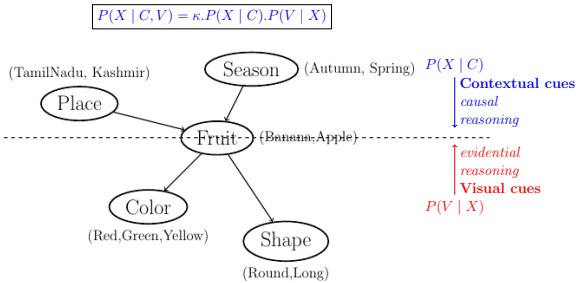
We can assume, visual features of an object is independent of context (how?)

$P(v_l \mid O_i, v_c) = P(v_l \mid O_i)$

$P(O_i \mid v_l, v_c) = \kappa.P(O_i \mid v_c).P(v_l \mid O_i)$

Example of context based reasoning

Context can be of different kinds – maybe external to the image



How do you decide if a photo is taken in the day or in the night ?

Visual context

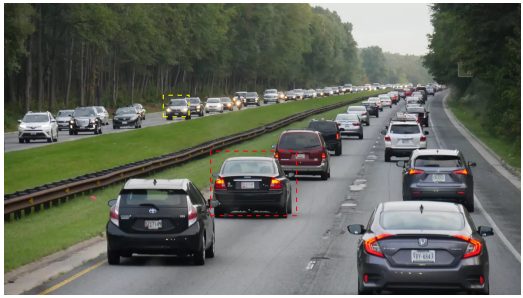
$P(O \mid v_c)$: v_c = visual feature of the context

- O_i is the manifestation of an object instance in a certain location of a scene
  - ... not just an object class
- Let $O_i = (o_i, x_i, \sigma_i)$ where
  - o_i: object class
  - x_i: location in image
  - σ_i: appearance (scale, orientation, etc.)
- $P(O_i \mid v_c)$ represents an object of a class to appear in a specific location in an image with a certain appearance

$$P(O_i \mid v_c) = P(o_i, x_i, \sigma_i \mid v_c) = P(\sigma_i \mid o_i, x_i, v_c).P(x_i \mid o_i, v_c).P(o_i \mid v_c)$$

In-context object recognition

Significance of the decomposition

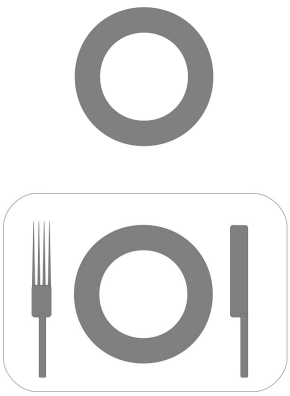


- $P(o_i \mid v_c)$: Probability of an object class to appear in a context
- $P(x_i \mid o_i, v_c)$: Probability of the location where an object class appears in a context
- $P(\sigma_i \mid o_i, x_i, v_c)$: Probability of the appearance of an an object class when it appears in a certain location in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

$$P(O_i \mid v_l, v_c) = \kappa . \underbrace{P(\sigma_i \mid o_i, x_i, v_c).P(x_i \mid o_i, v_c).P(o_i \mid v_c)}_{context}.\underbrace{P(v_l \mid O_i)}_{evidence}$$

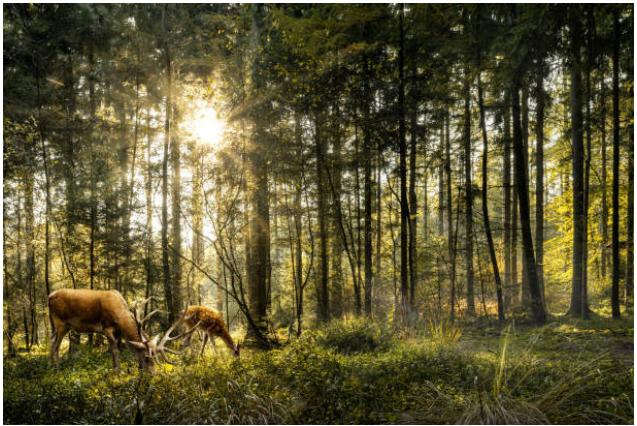
How do we characterize a context ?

What is v_c ?



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

Recap: Forest before the trees ?



Do the “scenes” have some distinctive features?

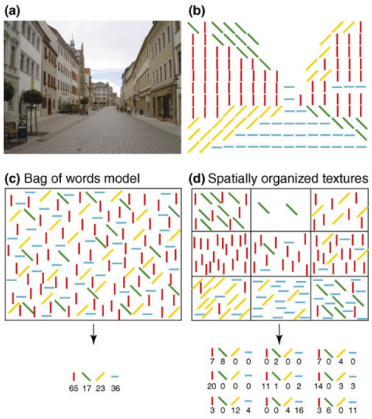





Spatial envelop representation

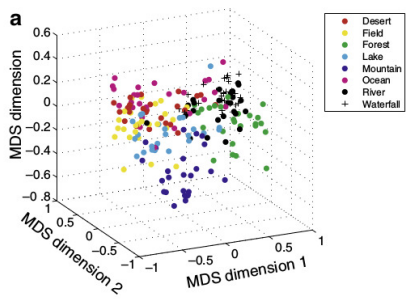
A holistic representation of a scene layout

- The edges in a scene constitutes a definite pattern
  - Statistical pattern characterizes a scene
- Recall natural scene statistics
  - Happens in early (pre-attentive) vision
- Computational Model:
  - Global and local statistics
  - Abstract image features



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

Distinguishing scene classes with spatial envelop representation



The three MDS axes represent three abstract features of an image: openness, ruggedness and expansion.

Quiz

Quiz 04-03

End of Module 04-03