Very many sources of variability

Image
Sources of image variability

Scene type
Street scene

Scene geometry
### Sources of image variability

<table>
<thead>
<tr>
<th>Scene type</th>
<th>Street scene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sky</td>
<td>Bicycle</td>
</tr>
<tr>
<td>Building × 3</td>
<td>Car × 5</td>
</tr>
<tr>
<td>Road</td>
<td>Bench</td>
</tr>
<tr>
<td>Sky</td>
<td>Sidewalk</td>
</tr>
<tr>
<td>Tree × 3</td>
<td></td>
</tr>
<tr>
<td>Person × 4</td>
<td></td>
</tr>
<tr>
<td>Bicycle</td>
<td></td>
</tr>
<tr>
<td>Car × 5</td>
<td></td>
</tr>
<tr>
<td>Bench</td>
<td></td>
</tr>
</tbody>
</table>

**Scene geometry**

* Object classes:
  - Street scene: Sky, Building × 3, Road, Sidewalk, Tree × 3, Person × 4, Bicycle, Car × 5, Bench, Bollard
## Sources of image variability

<table>
<thead>
<tr>
<th>Scene type</th>
<th>Scene geometry</th>
<th>Object classes</th>
<th>Object position</th>
<th>Object orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Street scene</td>
<td>Sky</td>
<td>Bicycle</td>
<td>Car</td>
<td>Bollard</td>
</tr>
<tr>
<td></td>
<td>Building×3</td>
<td>Tree×3</td>
<td>Person×4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Road</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Street scene

- Sky
- Building×3
- Road
- Sidewalk
- Tree×3
- Person×4
- Bicycle
- Car×5
- Bench
- Bollard
Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
Sources of image variability

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance

Illumination
Shadows

Image from LabelMe [Russell et al, 2008]
Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows

Image from LabelMe [Russell et al, 2008]
Sources of image variability

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows
Motion blur
Camera effects
Generative models

Generative approach
- select some sources of variability as latent variables
- design a tractable generative model $P(\text{Image}, \text{Latent})$ of the imaging process
- apply to unlabelled data

Sources of variability which are not explicitly represented must be:
- **fixed** by restricting the data set/imaging environment or
- **captured by noise** or outlier process
otherwise, the inference of other variables will be distorted.
<table>
<thead>
<tr>
<th>Comparison</th>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td>Complicated</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>Slow</td>
</tr>
</tbody>
</table>
Generative example I

Layered Sprites [Jojic & Frey, 2001]

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

Composition:
- Object appearance
- Object shape
- Position
- Translated appearance
- Translated shape
- Image

Objects 1..N
Generative example I

Input video

Flexible Sprites [Jojic & Frey, 2001]
<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>-</td>
<td>Complicated</td>
</tr>
<tr>
<td></td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>Often limited domain</td>
</tr>
</tbody>
</table>
Sources of variability which are not explicitly represented must be:

- **fixed** by restricting the data set/imaging environment or
- **thoroughly explored** in the training set

Otherwise, the inference of other variables will be distorted.

*All winning systems in Pascal VOC 2009 were discriminative.*
<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>Fast</td>
<td></td>
</tr>
<tr>
<td>Robust to model mismatch</td>
<td>Complicated</td>
</tr>
<tr>
<td>Needs labelled data</td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>Often limited domain</td>
</tr>
</tbody>
</table>
Discriminative example I

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

[Viola & Jones, 2001]

Sliding window

Features $f$ Classifier Face?
<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>Fast</td>
<td></td>
</tr>
<tr>
<td>Robust to model mismatch</td>
<td></td>
</tr>
</tbody>
</table>

- Needs labelled data
- Ignores helpful cues
- Inaccurate if too few labelled examples

- Complicated
- Slow
- Often limited domain
Discriminative example II

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows
Motion blur
Camera effects

Image I

Part map h

LayoutCRF

CVPR'06
with
Jamie Shotton
Discriminative example II

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

Image I

Part map $h$

LayoutCRF

CVPR'06

with

Jamie Shotton
<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>Fast</td>
<td>Complicated</td>
</tr>
<tr>
<td>Robust to model mismatch</td>
<td>Slow</td>
</tr>
<tr>
<td><strong>Need not be ‘black box’</strong></td>
<td>Often limited domain</td>
</tr>
<tr>
<td>Needs labelled data</td>
<td></td>
</tr>
<tr>
<td>Ignores helpful cues</td>
<td></td>
</tr>
<tr>
<td>Inaccurate if too few labelled examples</td>
<td></td>
</tr>
<tr>
<td>Hard to diagnose flaws</td>
<td></td>
</tr>
</tbody>
</table>
Generative example II

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows
Motion blur
Camera effects

Class shape
Object shape
Background appearance
Image
Position
Object appearance

Duplicated for each image
Generative example II

Unsupervised segmentation

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows
Motion blur
Camera effects

Accuracy: 93%
(cf. 94-95% supervised)
<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>Fast</td>
<td>Data can be unlabelled</td>
</tr>
<tr>
<td>Robust to model mismatch</td>
<td>Easier to diagnose flaws</td>
</tr>
<tr>
<td>Need not be ‘black box’</td>
<td>Complicated</td>
</tr>
<tr>
<td></td>
<td>Slow</td>
</tr>
<tr>
<td>Needs labelled data</td>
<td>Often limited domain</td>
</tr>
<tr>
<td>Ignores helpful cues</td>
<td></td>
</tr>
<tr>
<td>Inaccurate if too few</td>
<td></td>
</tr>
<tr>
<td>labelled examples</td>
<td></td>
</tr>
<tr>
<td>Hard to diagnose flaws</td>
<td></td>
</tr>
</tbody>
</table>
Generative example III

Unsupervised learning

Layered pictorial structures
[Pawan Kumar et al., 2005]
### Comparison

<table>
<thead>
<tr>
<th>Discriminative</th>
<th>Generative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Accurate (in domain)</td>
</tr>
<tr>
<td>Fast</td>
<td>Data can be unlabelled</td>
</tr>
<tr>
<td>Robust to model mismatch</td>
<td>Easier to diagnose flaws</td>
</tr>
<tr>
<td>Need not be ‘black box’</td>
<td>Uses all available cues</td>
</tr>
<tr>
<td>Needs labelled data</td>
<td>Can be used for many different applications</td>
</tr>
<tr>
<td>Ignores helpful cues</td>
<td></td>
</tr>
<tr>
<td>Inaccurate if too few labelled examples</td>
<td></td>
</tr>
<tr>
<td>Hard to diagnose flaws</td>
<td></td>
</tr>
</tbody>
</table>
A third way...

Hybrid approach

- select some sources of variability as latent variables
- design a generative model \( P(\text{Image}, \text{Latent}) \)
- apply to unlabelled (or partially labelled) data

- develop **hybrid** inference method using both models

- design a discriminative model (classifier) to predict latent variables given image, \( P(\text{Latent} \mid \text{Image}) \)
- train on data set with *manually annotated* latent variables
Hybrid Belief Propagation

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

Duplicated for each image

CVPR'07 with Julia Lasserre and Anitha Kannan

Location map $L$ (pos./shape/occl.)
Hybrid Belief Propagation

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows
Motion blur
Camera effects

Generative likelihood $P(I_i | I_i)$

Classifier prediction $P(I_i | I)$

Hybrid likelihood
Hybrid Belief Propagation

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

Increasing size of classifier training set (from 1 to 20 images)

Averaged over 10 test images.
The goal of machine vision

- Model **all sources** of image variability jointly:
  - Scene type
  - Scene geometry
  - Object classes
  - Object position
  - Object orientation
  - Object shape
  - Depth/occlusions
  - Object appearance
  - Illumination
  - Shadows
  - Motion blur
  - Camera effects

- Possible using hybrid methods?
Deep Segmentation Networks

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects

Watch this space…
Thank you