Neuro-Inspired Statistical Prior Model for Robust Visual Inference

Qiang Ji
Rensselaer Polytechnic Institute
National Science Foundation

Status of Computer Vision

- CV has been an active area for over 40 years
- Much progress has been made (e.g. visual inspection, medical imaging, and OCR)
- Yet, much remains to be done.   CV still cannot effectively handle many problems that human can easily solve
  - recognize a person from different face angles and under different illuminations
  - segment an object (e.g dog) from an image with cluttered background.
  - recognize human activities from low resolution video.

Why Lack of Significant Progress?

– CV is a highly ill-posed problem

– Image data is uncertain, ambiguous, and often incomplete

– **Lack of effective use of prior knowledge**
  - Human brain encodes prior knowledge about the world.
  - Human uses prior along with sensory data for visual understanding

How to Solve this Challenge?

- Identify the related prior knowledge external to image data from different sources
- Capture and represent such knowledge
- Systematically combine the captured knowledge with image data to perform visual inference and understanding

Types of Prior Knowledge

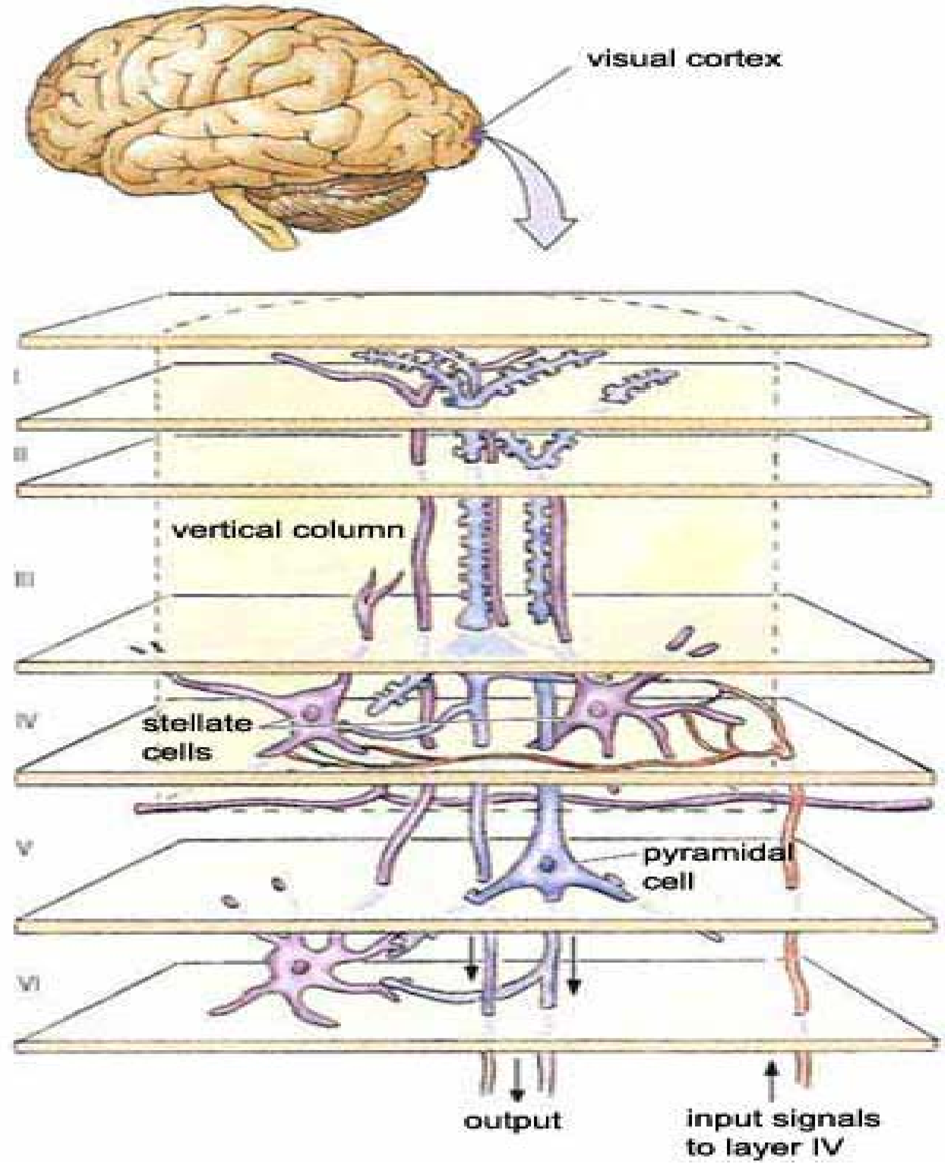
- Permanent-physics, physiology, geometry, anatomy, kinematics, biomechanics etc..
  - Various knowledge, theories or principles that govern the properties and behavior of the objects
  - Tend to be generic, applicable to different objects and different situations.
  - Hard to capture
- Temporary-statistical pattern-based
  - Tend to be object, situation or database specific
  - widely used in CV.
- Exact versus approximate

Visual Cortex

- The primary visual cortex is organized hierarchically and each layer consists of neurons and synapse

- Encoding (learning) is used to identify and store spatial-temporal patterns by establishing connections among neurons at the same level and at next higher level.

- Visual decoding is performed to combine the visual observations with the captured knowledge via inference



Probabilistic Graphical Models

Probabilistic graphical model (PGM) is a graphical representation of probabilistic relationships among random variables**.**

- Representation
  - Its hierarchical structure allows effectively capturing the prior knowledge at different levels of abstraction.
  - Its probability foundation allows to model, capture, and propagate uncertainties
- Learning (encoding)
  - It offers mechanisms to automatically learn the model structure and parameters to store the knowledge
- Inference (decoding)
  - It provides inference mechanism to systematically combine the prior knowledge with image data to perform effective visual tasks

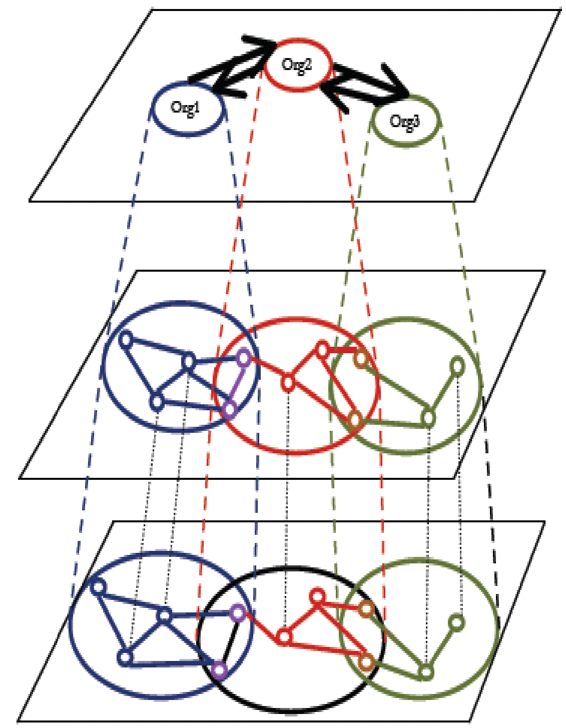
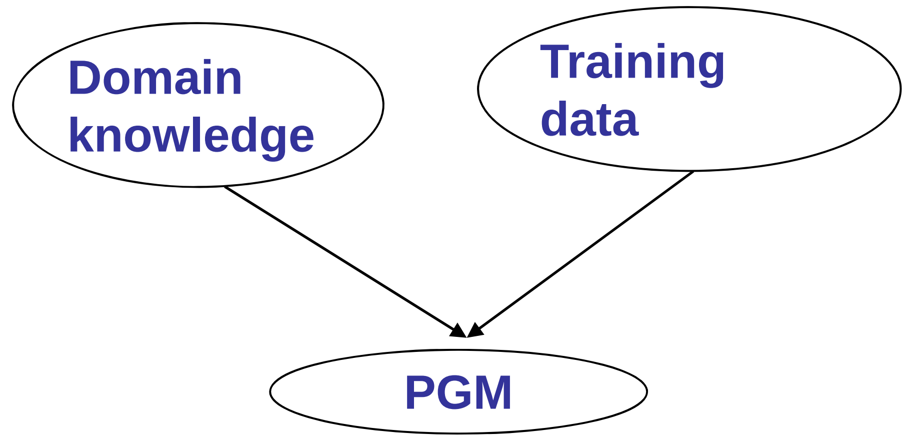


Figure 3. A hierarchical probabilistic graphical model.

Constrained Model Learning and Inference: A hybrid approach

- Combine the domain knowledge and the statistics learned from data through constrained learning:
  - Hard constraints
    - Constrained ML/EM
  - Soft constraints
    - Sample the constraints to produce constraint data
- Inference with constraints



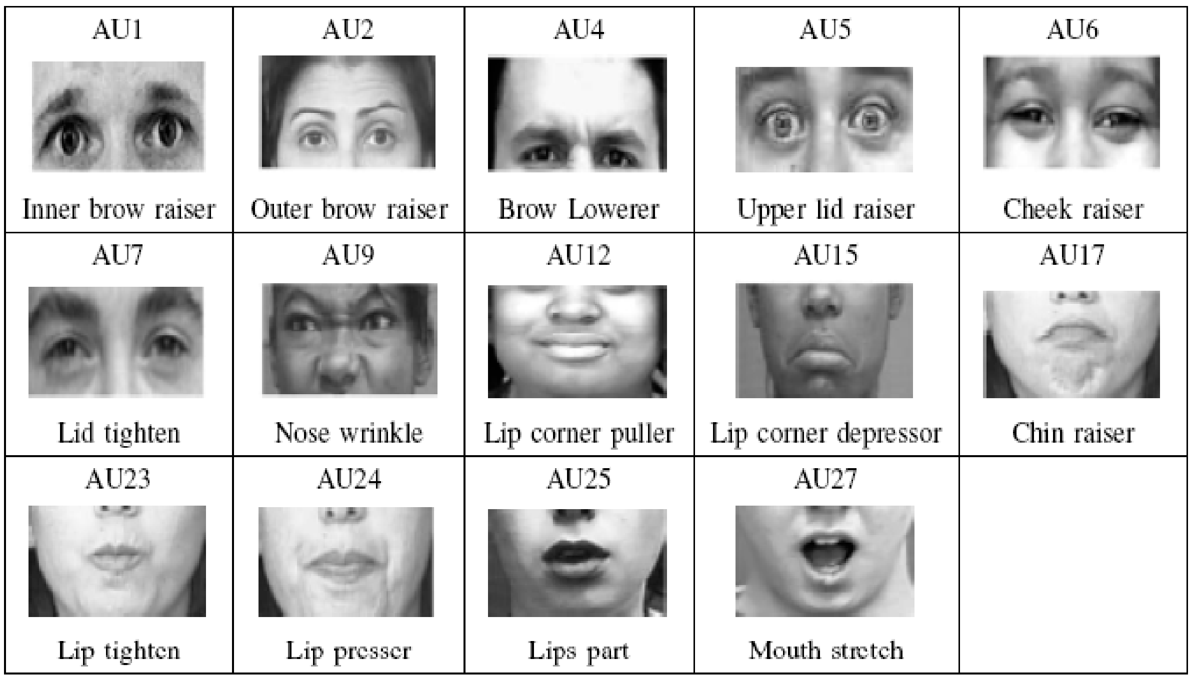
Prior model for Computer Vision Applications

- **Facial action modeling and recognition**
- **Image segmentation**
- Human body tracking
- Human activity modeling and recognition

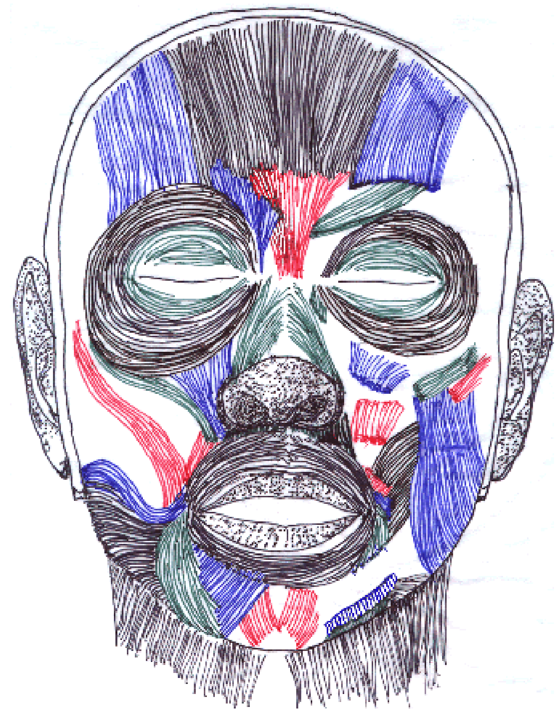
Facial Action Units Recognition

(Tong and Ji, CVPR07, PAMI07, and PAMI 10)

- ➢ Facial Action Units (AUs) capture the non-rigid muscular activities that produce facial appearance changes (defined in Facial Action Coding System)
- Each AU is related to the contraction of a set of facial muscles.
- ➢ A small set of AUs can describe a large number of facial behaviors



(a) A list of AUs and their interpretations



(b) Muscles underlying facial AUs

Existing Methods

- mainly data-driven
- tend to recognize each AU independently
- ignore the fact that often facial actions act in a "synchronized, smooth, symmetrical, and consistent" way to produce a meaningful facial expression

Spatial Relationships among AUs

In a spontaneous facial behavior, there are some spatial relationships among AUs:

- Groups of AUs often appear together to show meaningful expression

Happiness
AU6+12+25

Surprise
AU1+2+5+25+27

Sadness
AU1+4+15+17

Due to the underlying facial anatomy

- Co-occurrence relationships such as AU1 (inner brow raiser) and AU2 (outer brow raiser)
- Mutually exclusive relationships, e.g. AU24 (lip presser) and AU25 (lips apart)

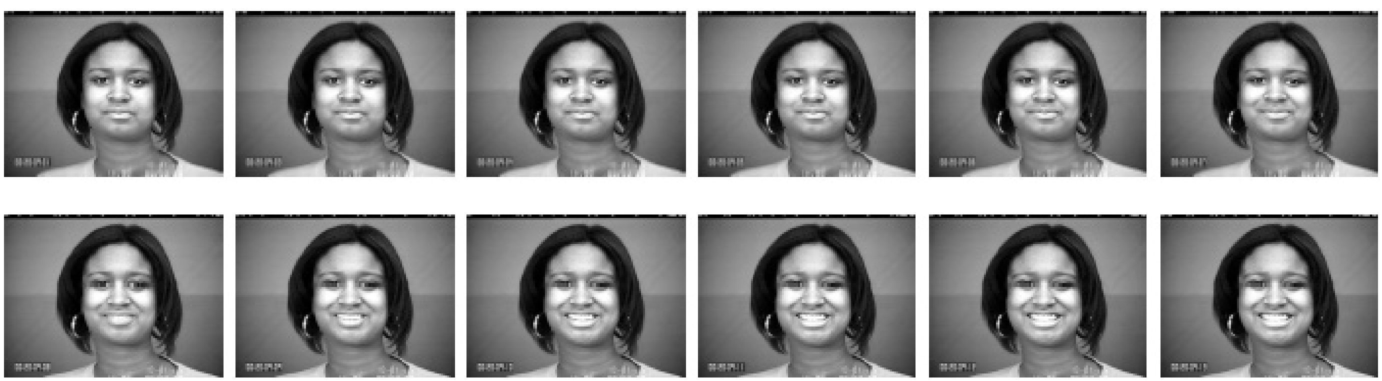
[Tong et al. ’06 CVPR, Tong et al. ’07 PAMI]

Dynamic Relationships among AUs

Dynamic characteristics are crucial for interpreting spontaneous facial behavior. They include:

- Self development of each AU
- Dynamic dependencies between the consecutive occurrences of certain AUs

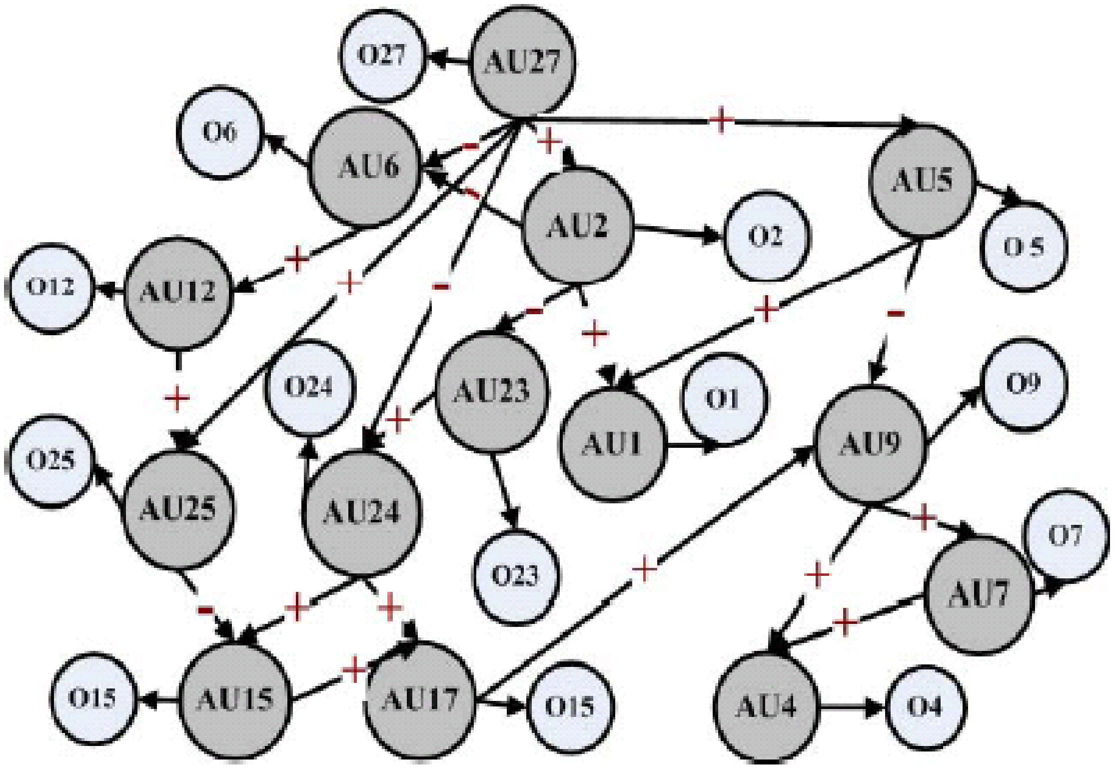
For example, in a smile, AU12 (lip corner raiser) is followed by AU6 (cheek raiser) and AU25 (lips apart).



Anatomic Constraints

- Positive and negative causal influences
  - Mouth stretch increases the chance of lips apart; it decreases the chance of cheek raiser and lip presser.
  - Cheek raiser and lid compressor increases the chance of lip corner puller.
  - Outer brow raiser increases the chance of inner brow raiser.
  - Upper lid raiser increases the chance of inner brow raiser and decreases the chance of nose wrinkler.
  - Lip tightener increases the chance of lip presser.
  - Lip presser increases the chance of lip corner depressor and chin raiser.

Positive and Negative Influences



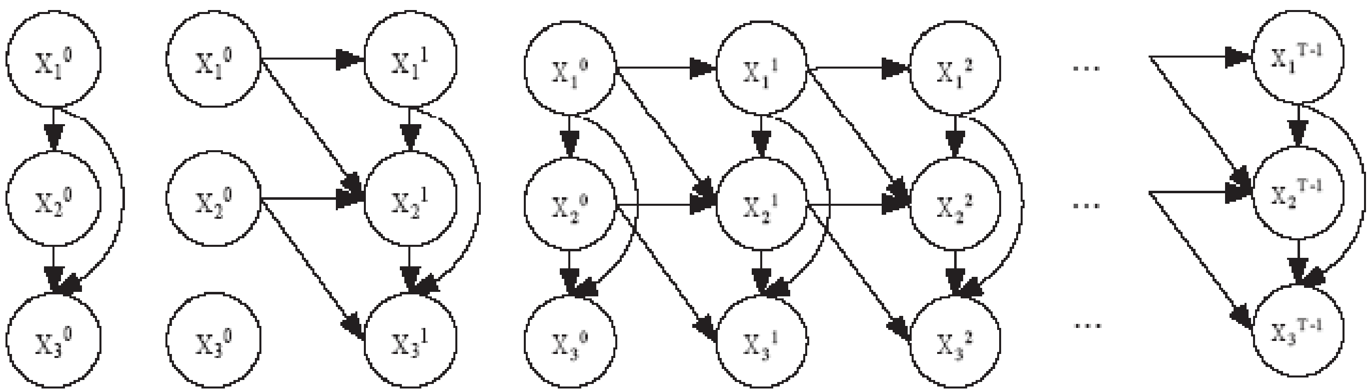
For an AU_i with positive influence by its parent node AU_j $P(AU_i = 1 | AU_j = 1) > P(AU_i = 1 | AU_j = 0)$

For an AU_i with negative influence by its parent node Au_j $P(AU_i = 1 | AU_j = 1) < P(AU_i = 1 | AU_j = 0)$

Dynamic Bayesian Network

DBN is a directed acyclic graph (DAG) consists of two parts:

- Static network $B_0 = (G_0, \theta_0)$ models the static relationships among variables
- Transition network $B_{\rightarrow} = (G_{\rightarrow}, \theta_{\rightarrow})$ models the dynamic relationships among variables



B_0 $B_{\rightarrow}$ Unrolled DBN

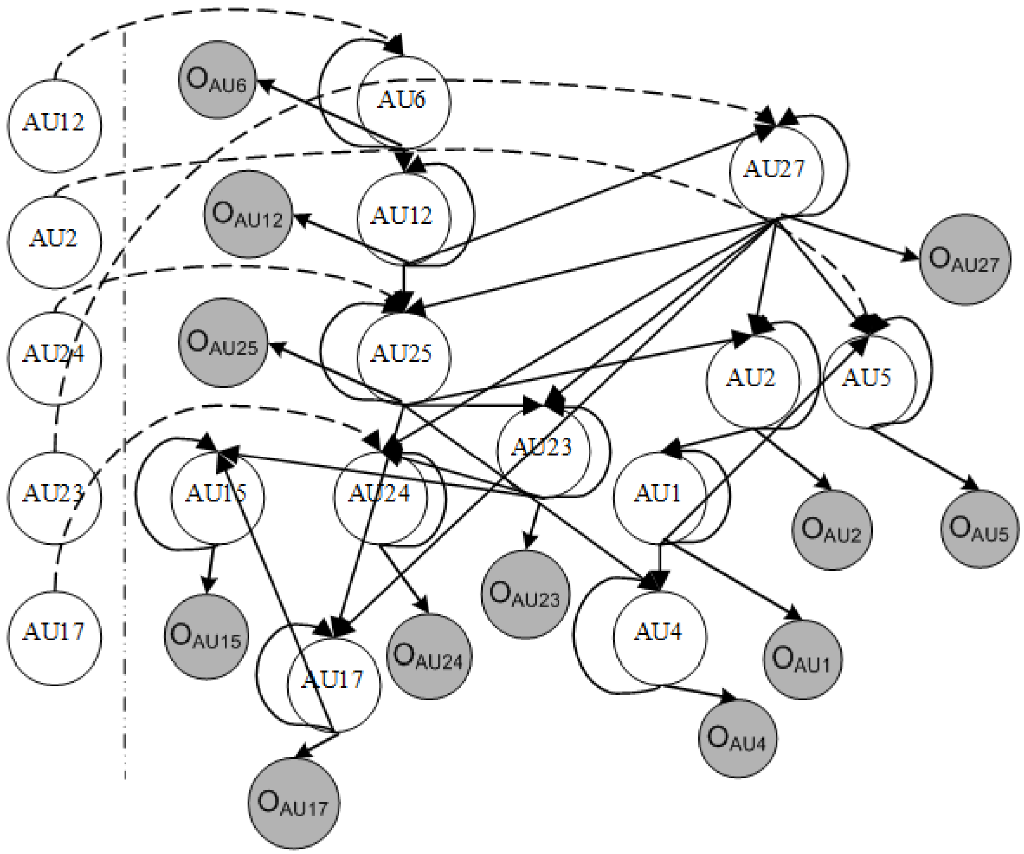
- DBN learning: constrained learning from both the training data and the anatomic constraints (CVPR08, ECCV08)

Constrained DBN Learning

- Learning the DBN model with analytic solution by KKT conditions **(Tong and Ji, CVPR08)**

- Learning the DBN model with iterative constrained optimization (Campos, Tong and Ji, ECCV08)

The Learnt DBN for AU Relationship Modeling



- Solid line: spatial relationship among AUs
- Self-arrow: temporal evolution of a single AU
- Dashed line from time *t-1* to time *t*: temporal relationship between two different AUs

t-1

t

$$AU^{*}_{1..N} = \arg\max_{AU_{1..N}} P(AU_{1..N} \mid O_{AU_{1..N}})$$

AU Recognition Experimental Results

Compared with the AdaBoost,

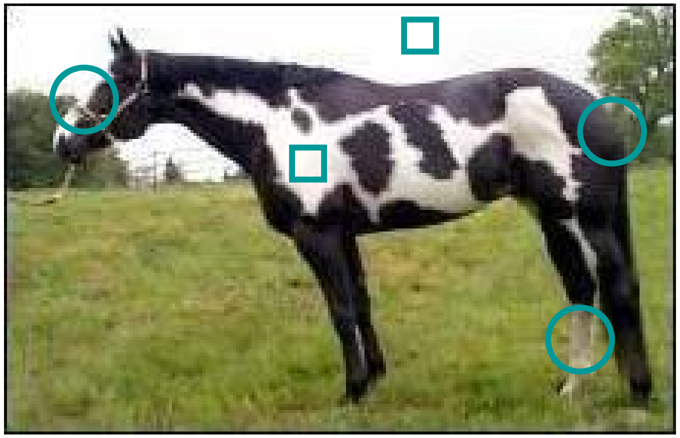
- ✓ The overall average recognition performance improves
  - Average false negative error decreases from 44% to 24.3%
  - Average false positive error decreases from 8.58% to 5.3%

- ✓ Significantly improves the recognition performances of difficult AUs
  - False negative rate of AU23 (lip tighten) decreases from 94.4% to 25.9%, with a moderate increase in false positive error rate from 3.6% to 5.8%
  - False negative error of AU12 (lip corner puller) decreases from 53% to 37.8%, and its false positive error decreases from 32% to 12.8%

A Hybrid Framework for Image Segmentation (PAMI, Zhang and Ji, 2010)

- Image segmentation aims to partition an image into constituent regions of interest.

- Challenges in image segmentation:
  - appearance changes
  - illumination changes
  - low contrast edges
  - noises
  - occlusion
  - cluttering



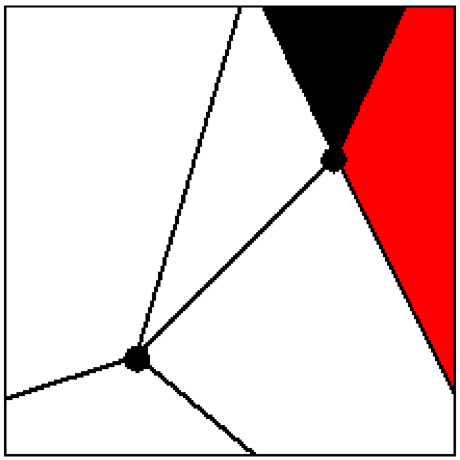
- Existing methods are mostly data-driven
- Incorporation of prior knowledge help solve the problems. A flexible framework is needed to integrate different prior knowledge with image measurements.

Heterogeneous Relationships between Image Entities

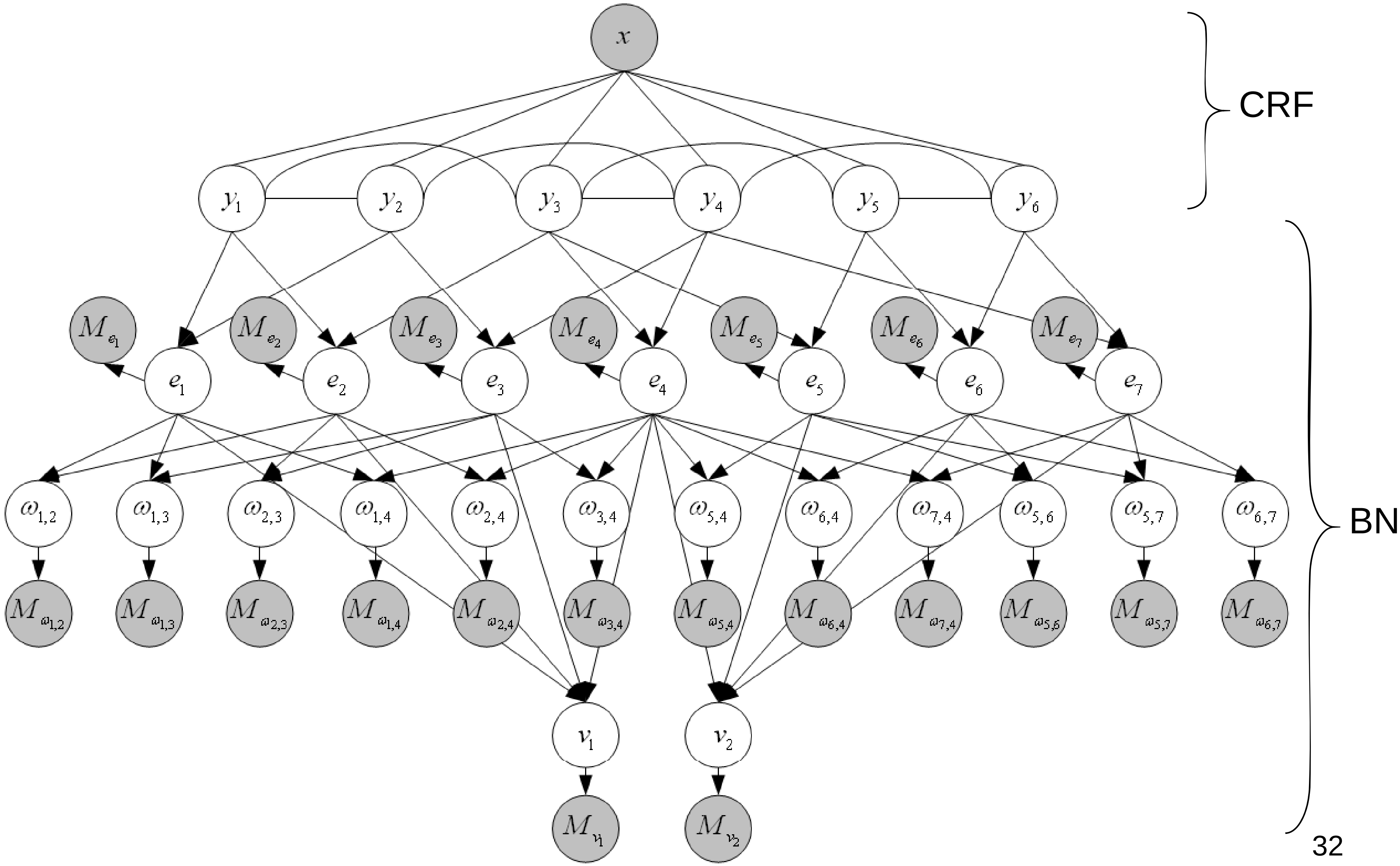
- Image entities: regions (superpixels), edges, vertices.

- Causal relationships:
  - Two regions intersect to form an edge.
  - Multiple edges intersect to form a vertex (or junction).
  - Image entities produce their measurements.

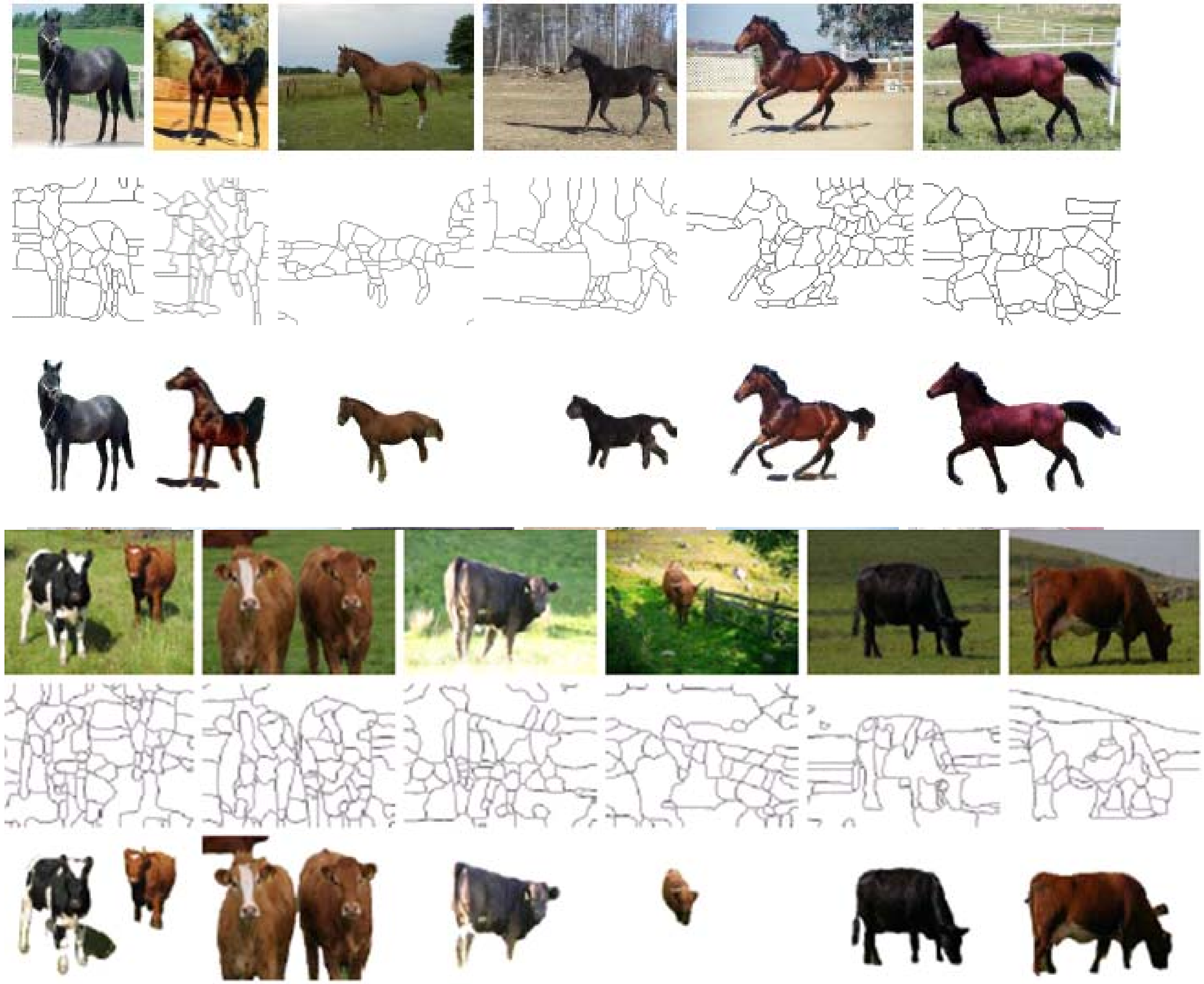
- Mutual dependencies:
  - Spatial correlations between adjacent regions and edges
- Local image constraints: local smoothness constraint and connectivity constraint



A Hybrid Graphical Model for Image Segmentation (zhang& Ji, CVPR09, PAMI10)



Experimental Results on Weizmann and VOC Datasets



Conclusions

- Computer vision is a very ill-posed problem
- Image data alone is not enough to address this problem
- To achieve human level of perception, additional prior knowledge from different sources should be systematically exploited and used to help regularize the problem.
- As a neurologically plausible model, PGM offers a framework that
  - is similar to visual cortex in knowledge representation and processing
  - allows representing knowledge from different sources as well as the uncertainties
  - admits methods for automatic model learning and inference
- Much work remains to be done in PGM model learning in particular structure learning in large model and in semi-supervised and unsupervised learning.