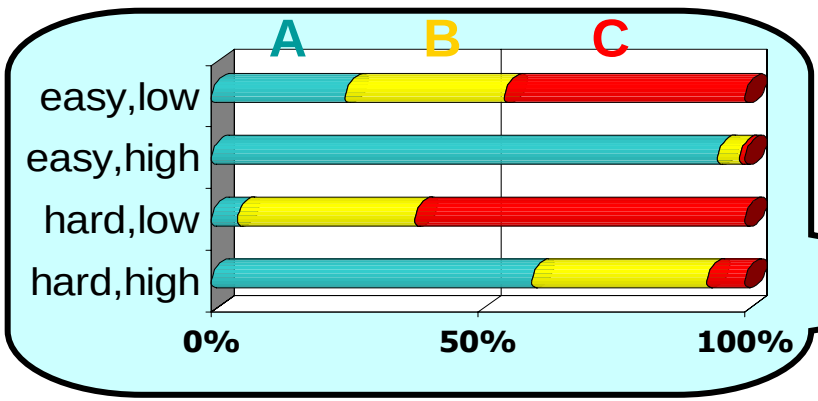
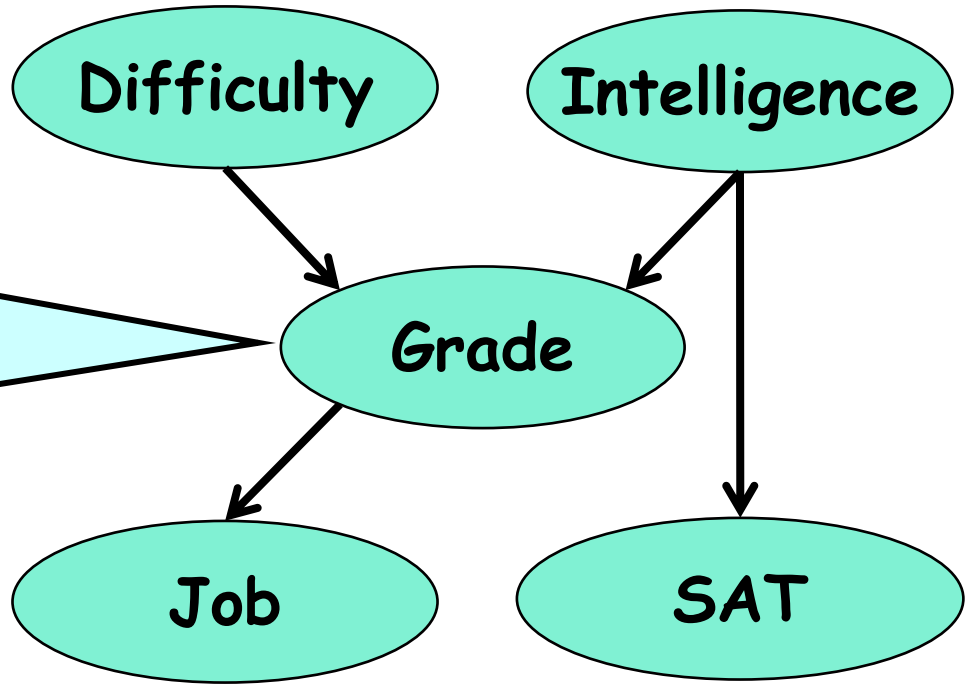
Probabilistic Models of Structured Data

Daphne Koller

Stanford University

Bayesian Networks



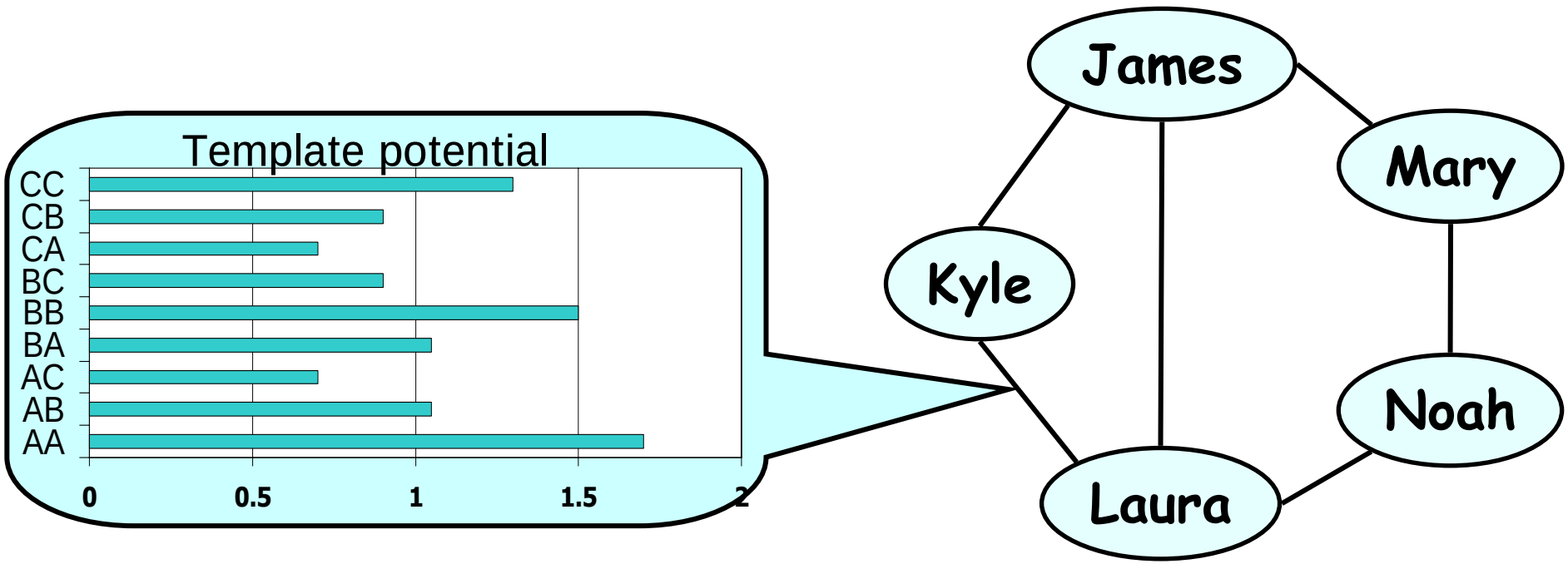


nodes = variables
edges = direct influence

Graph structure encodes independence assumptions:

Job conditionally independent of *Intelligence* given *Grade*

Markov Networks



$$P(J,K,L,M,N) = \frac{1}{Z}$$

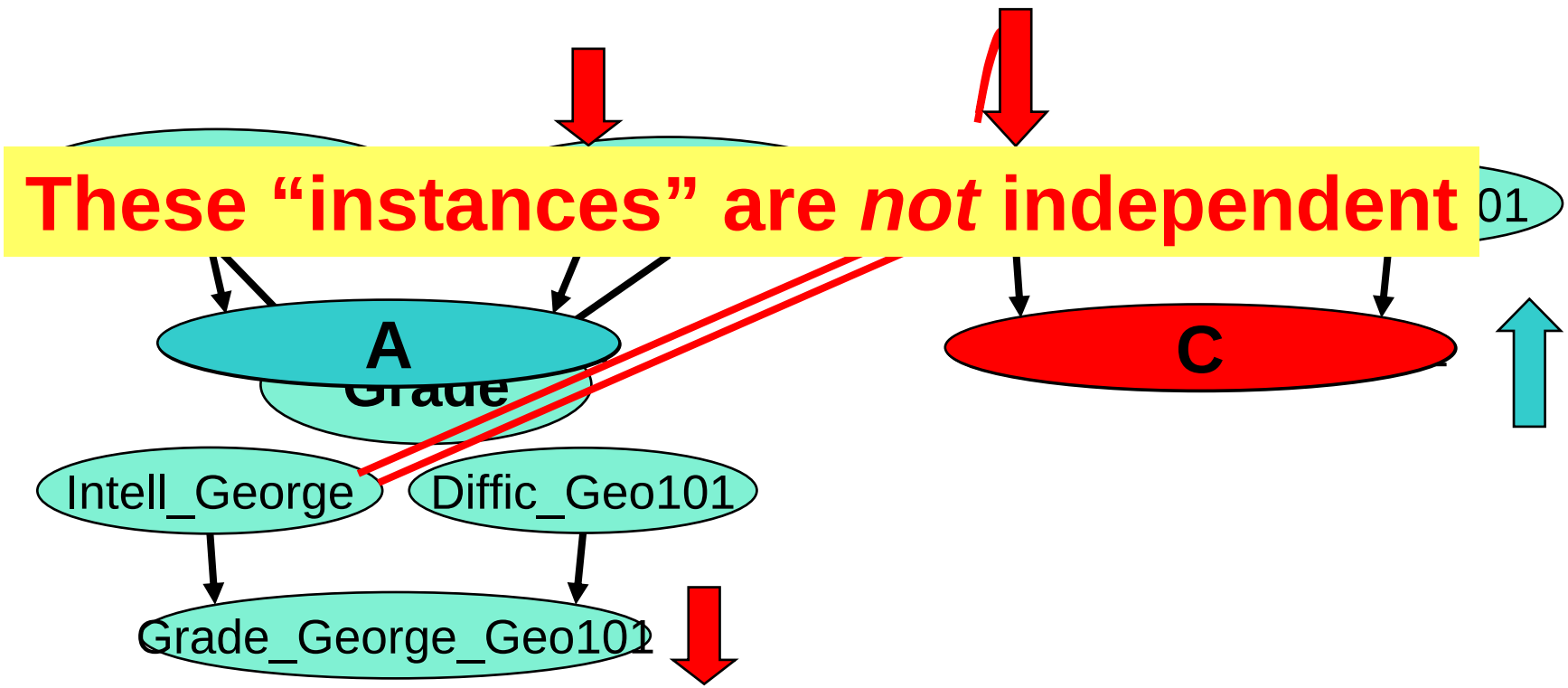
$$\phi(J,K)\phi(J,L)\phi(K,L)\phi(J,M)\phi(M,N)\phi(L,N)$$

The World is Richly Structured

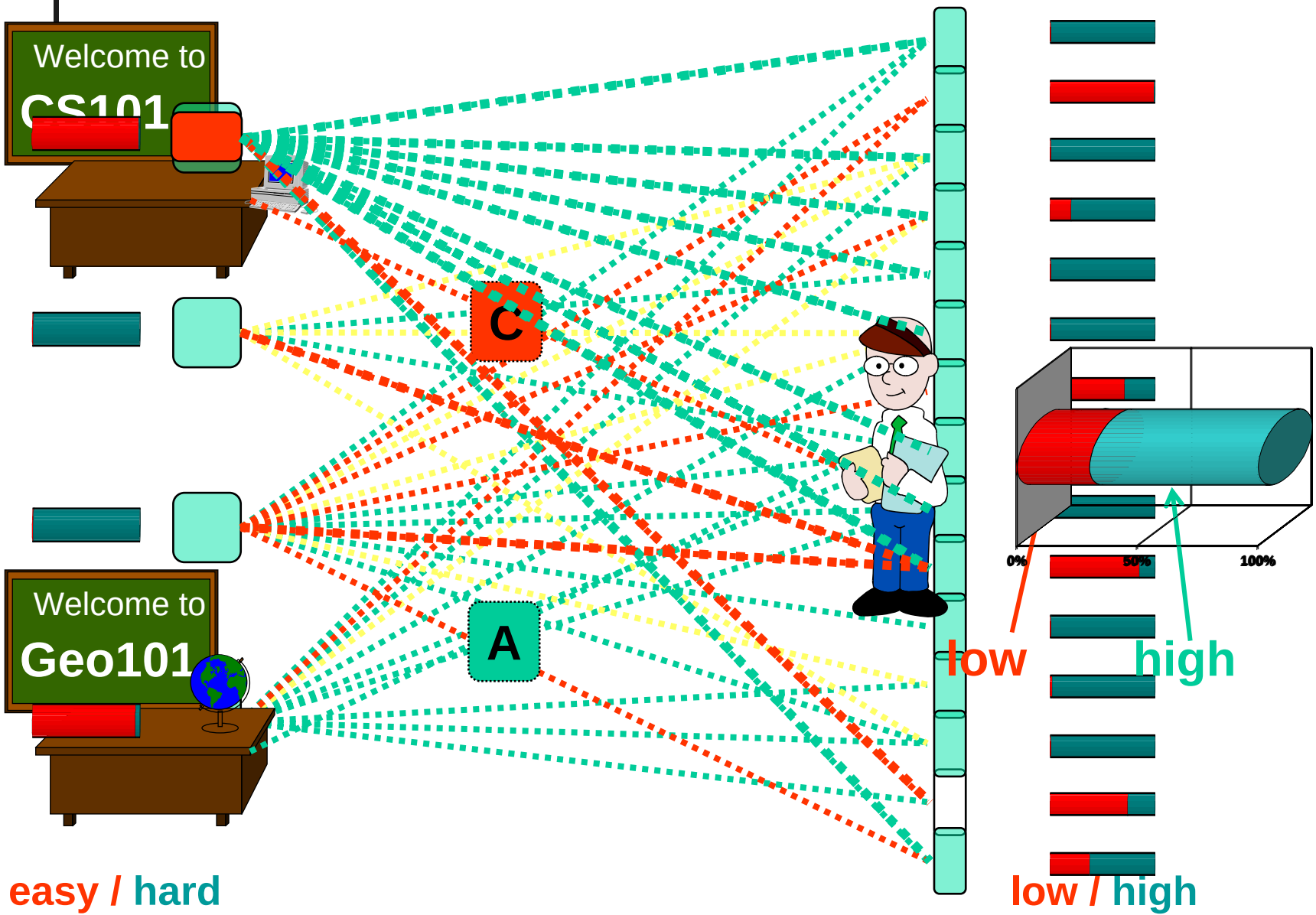
- The web
  - Webpages (& the entities they represent), hyperlinks
- Biological data
  - Genes, proteins, interactions, regulation
- Physical environments
  - People, rooms, objects
- Natural language

Problem

- Bayesian/Markov nets use attribute representation
- Real world has objects, related to each other



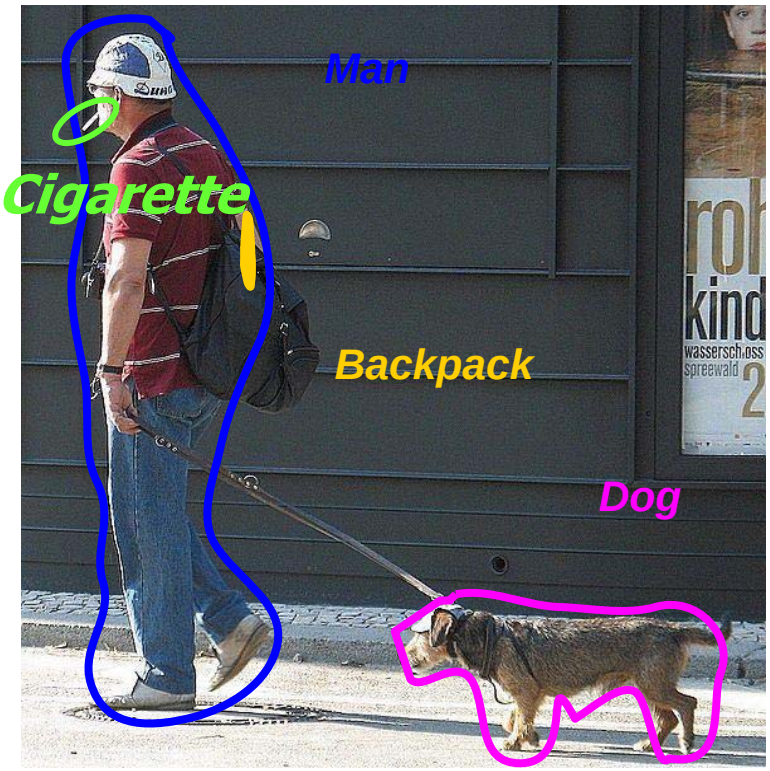
Holistic Reasoning

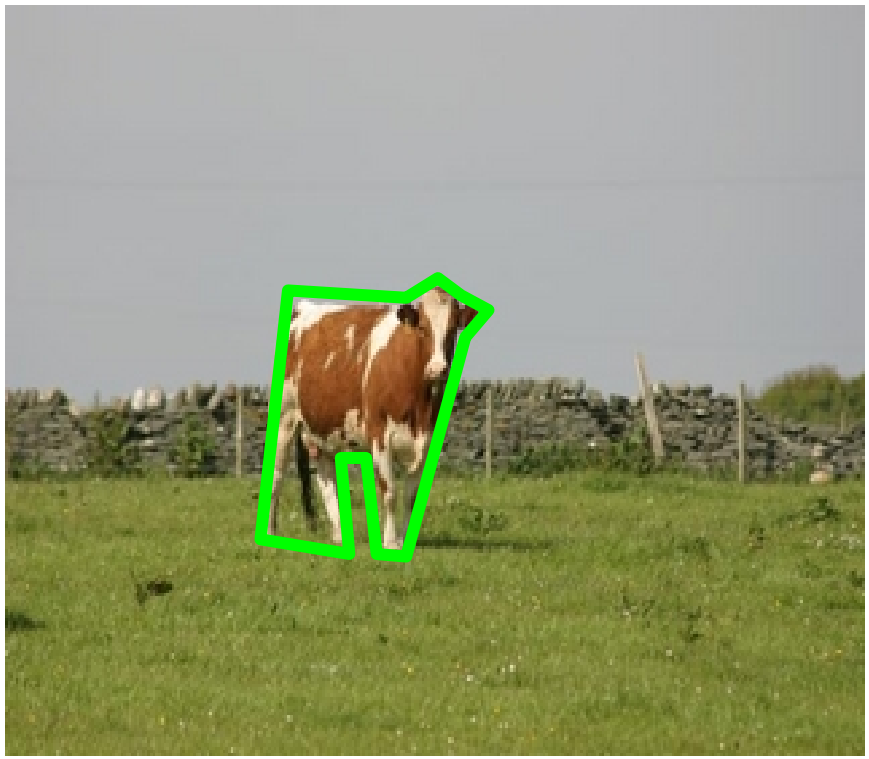


Main Application Domains

- Symbolic understanding of the physical world
- Understanding and reconstructing cellular processes from genomic data

Long-Term Goal: Scene Understanding



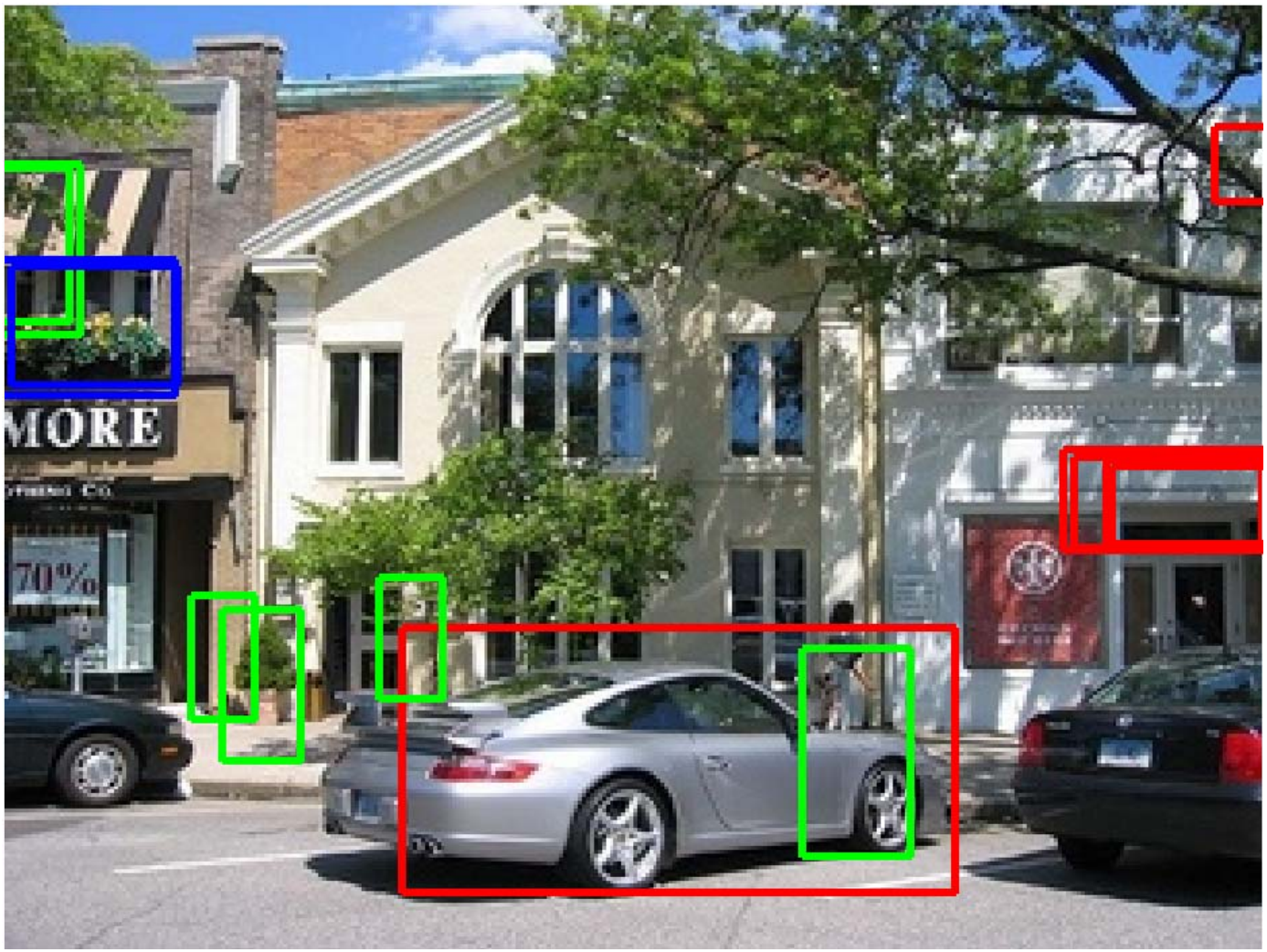


"man wearing a backpack, smoking a cigarette, walking a dog"

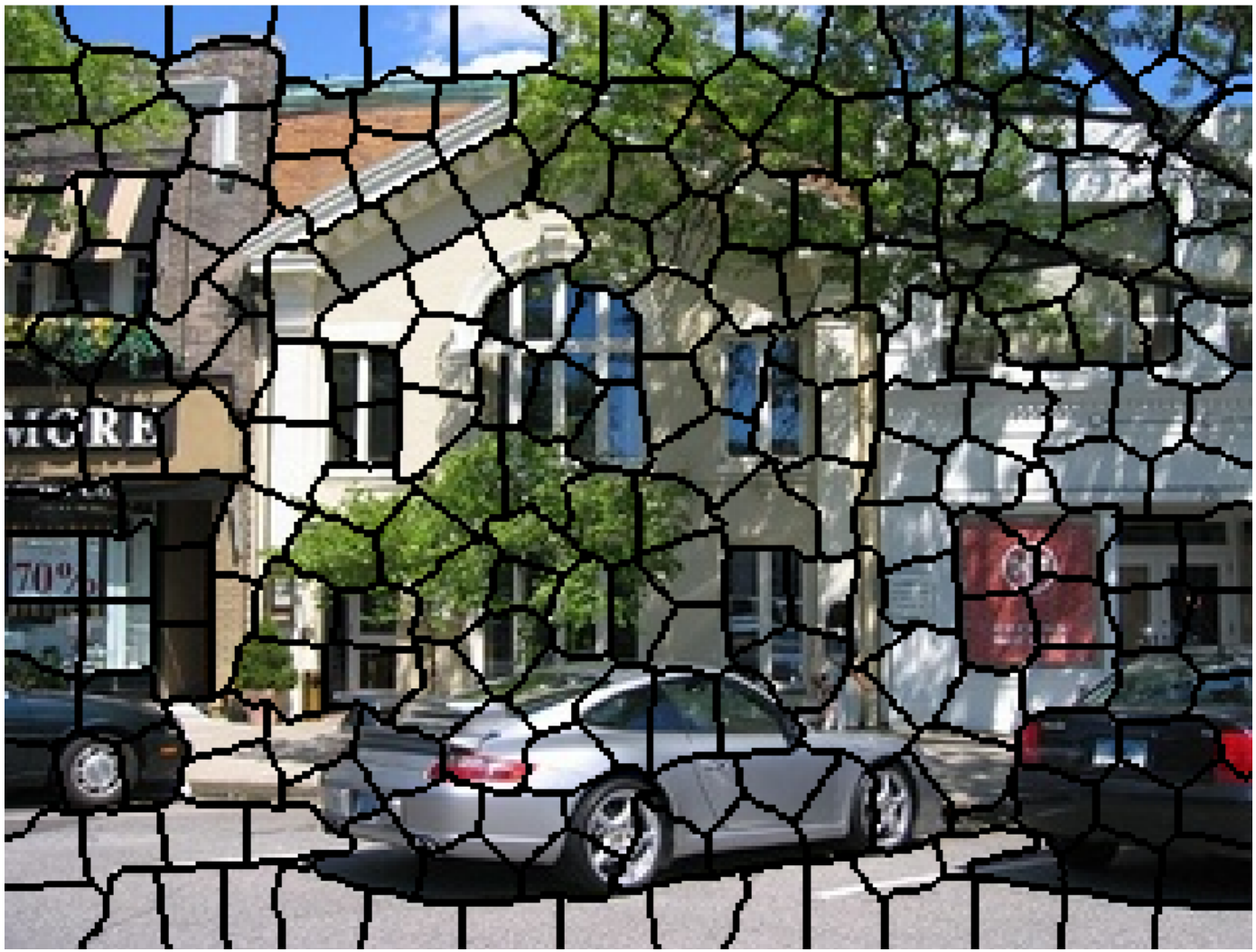
"A cow walking through the grass on a pasture by the sea"

Object Detection

Basic Object Detection

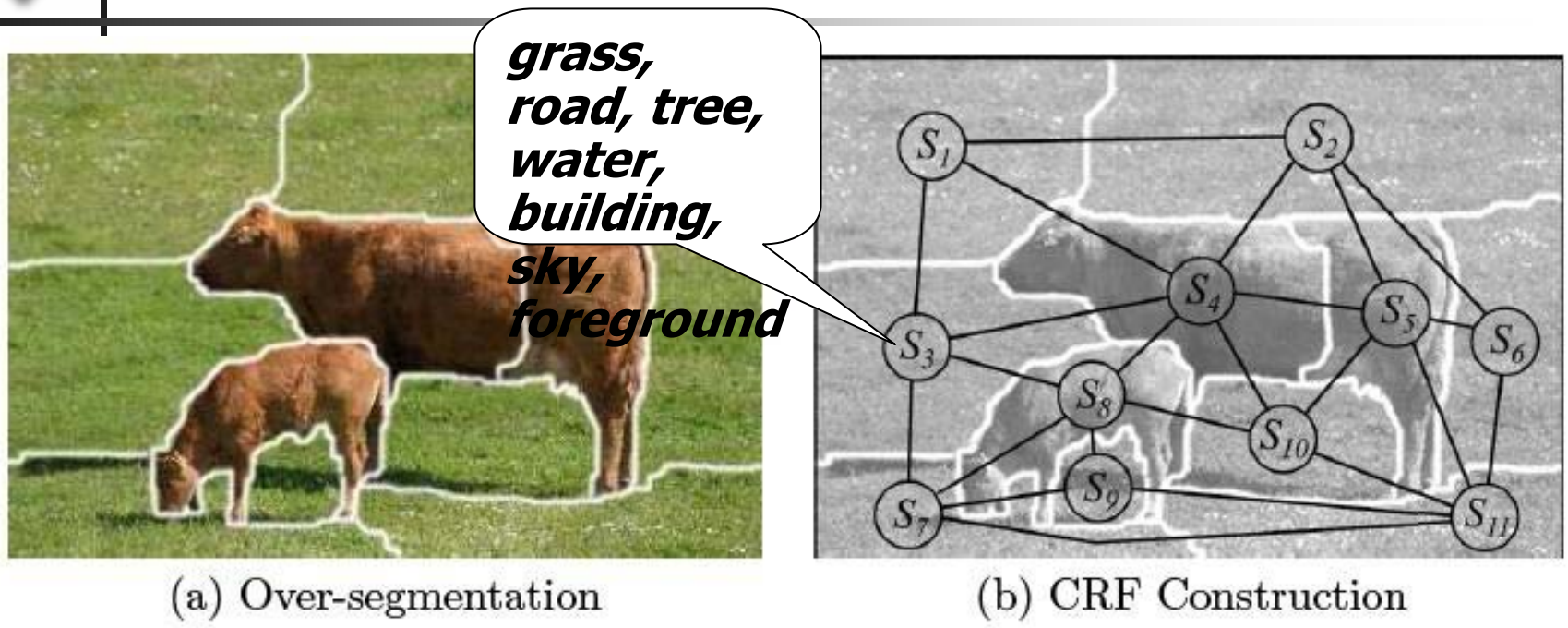


Scene Segmentation



'grass', 'road', 'tree', 'sky', 'water', 'building', 'foreground'

Segmentation CRF



(a) Over-segmentation (b) CRF Construction

$$P(S_1,\ldots,S_N) \propto \prod_i \underbrace{\phi_i(S_i : I)} \prod_{i,j} \underbrace{\phi_{i,j}(S_i, S_j : I)}$$

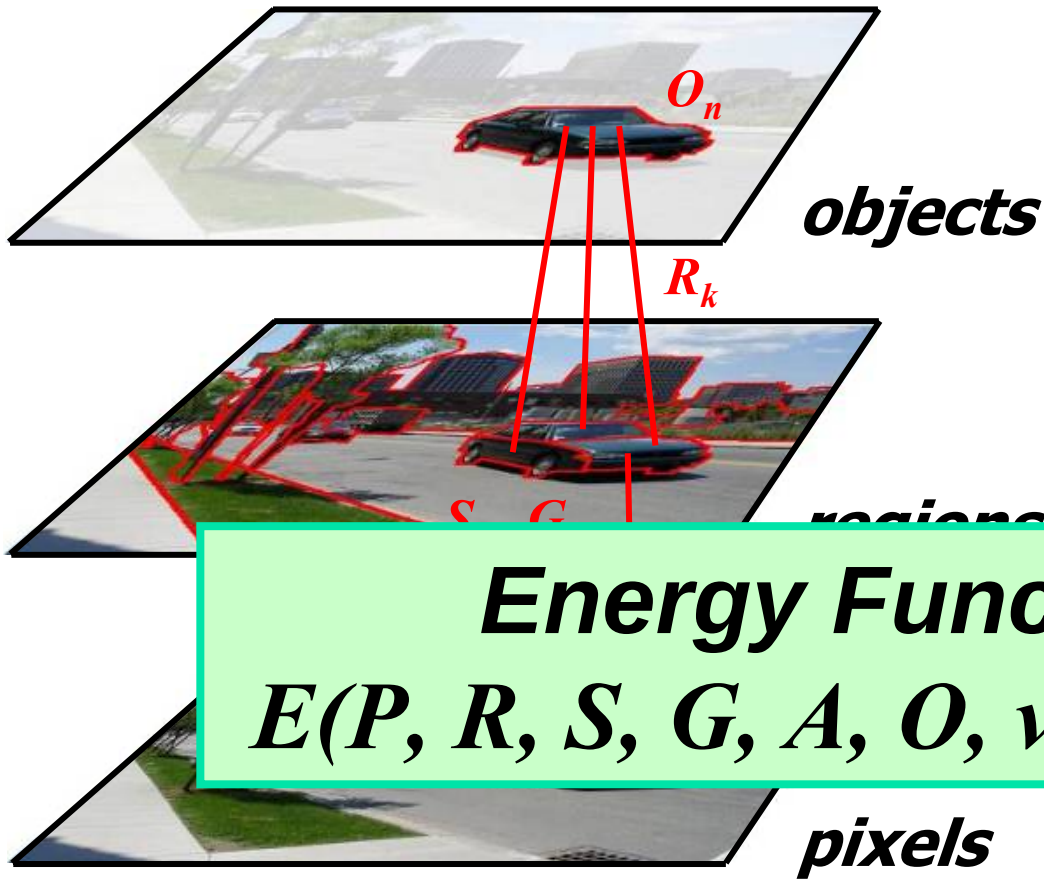
Singleton energy:

Mean R,G,B
Mean H,U,V
Texture Responses
…

Pairwise energy:

Delta R,G,B
Offset Vector
…

Hierarchical Scene Model



Variables

- α_i: pixel appearance
- P_i: pixel-to-region correspondence
- A_k: region appearance
- S_k: region semantic class
- G_k: region geometry
- R_k: region-to-object correspondence
- O_n: object class
- v^{hz}: location of horizon

Energy Function

$$E(P, R, S, G, A, O, v^{hz}, K \mid I, \theta)$$

[Gould, Gao, Koller submitted]

Results: 21-class MSRC

- Validate against state-of-the-art approaches
- Region/pixel class only
- Ground truth labels are approximate
- No geometry information

21 CLASS	Mean
Shotton et al.	*72.2*
Gould et al.	***76.5***
Pixelwise	75.3
Region-based	**75.4**

hand labeled ***image*** ***pixelwise*** ***region-based***

[Gould, Fulton, Koller, ICCV 2009]

High Quality Dataset

- MSRC dataset is limited
  - poorly labeled boundaries
  - many missing pixels (void)
  - no geometry information
- Collected images from MSRC, Hoiem et al., Pascal VOC
- 715 outdoor scenes with high-quality labels
  - region boundaries
  - region class and geometry
  - horizon

[Gould, Fulton, Koller, ICCV 2009]

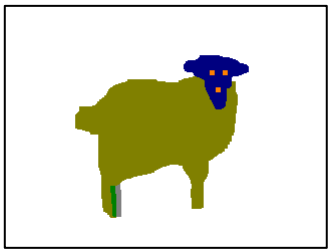
Amazon Mechanical Turk (AMT)

- $0.10 per task (regions, classes, surface types)
- 5-10 minutes per task
- 24-48 hour turn-around time (for 715 images)
- Less than 10% of tasks needed rework

- **Total cost for labels:** under $250 (includes $40 textbook on Adobe Flash)
- **Saving Steve from having to label images:** priceless.

[Gould, Fulton, Koller, ICCV 2009]

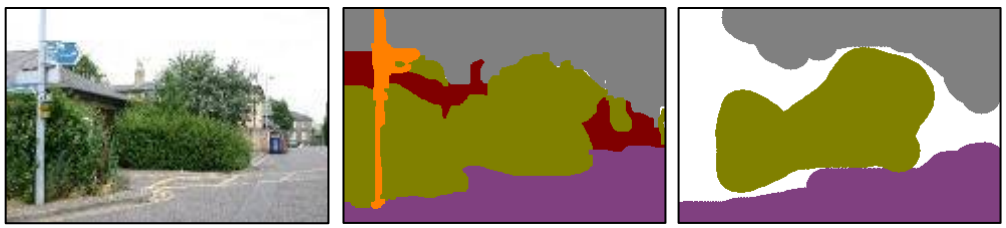
AMT: Label Quality



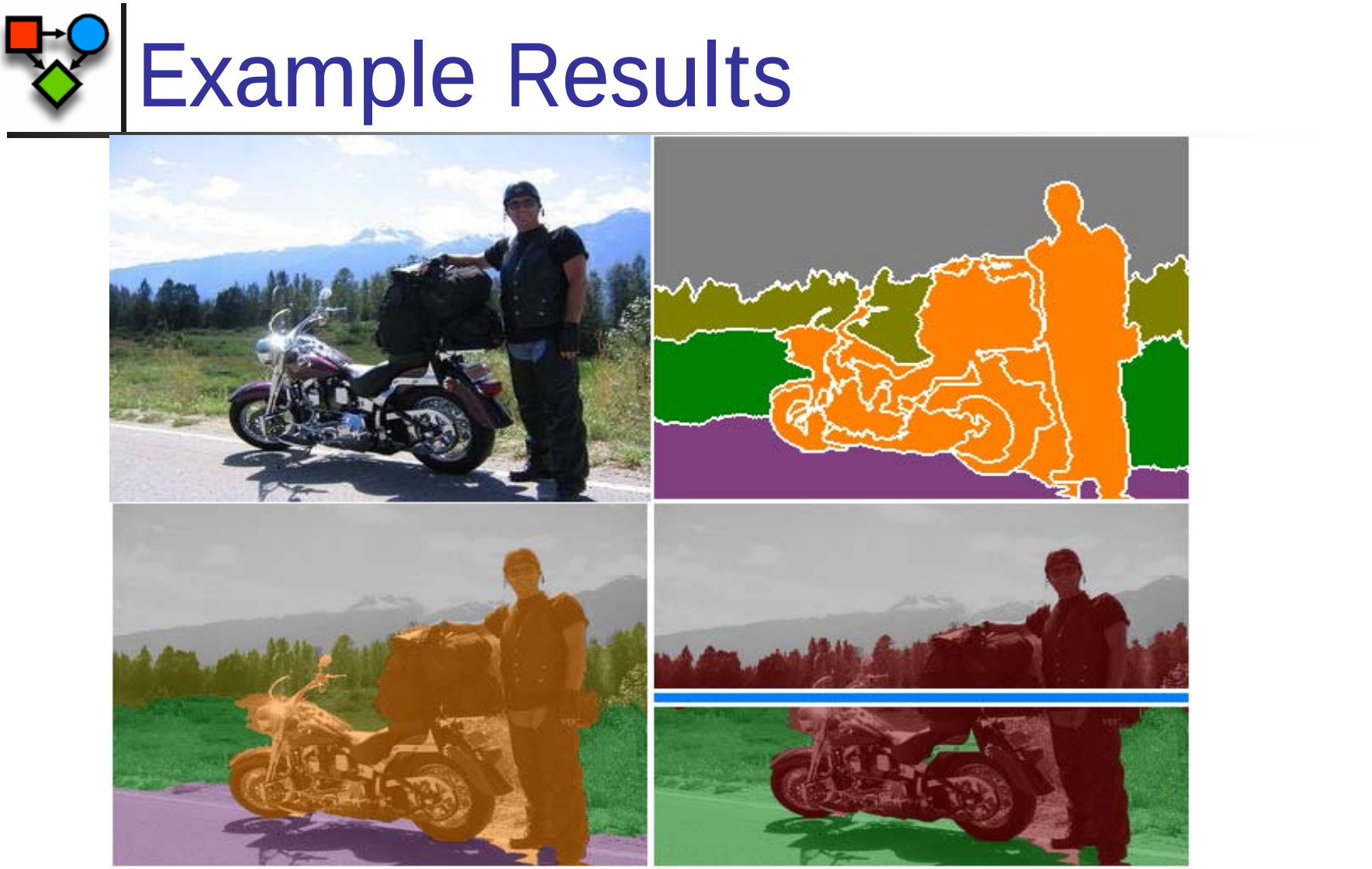
You don't always get what you want

Typical quality (hand labeled)

Comparison with MSRC labels



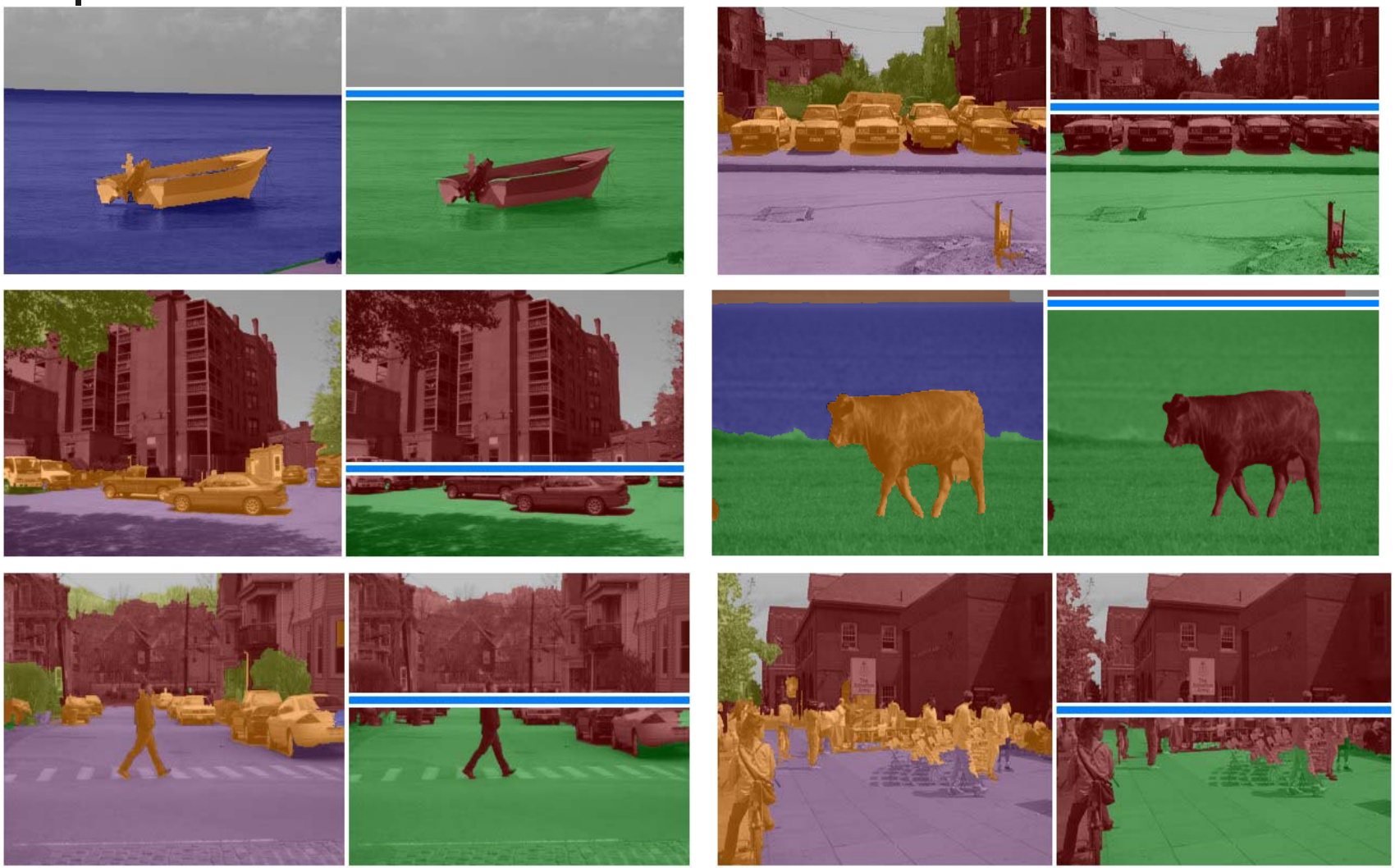
[Gould, Fulton, Koller, ICCV 2009]



Example Results

[Gould, Fulton, Koller, ICCV 2009]

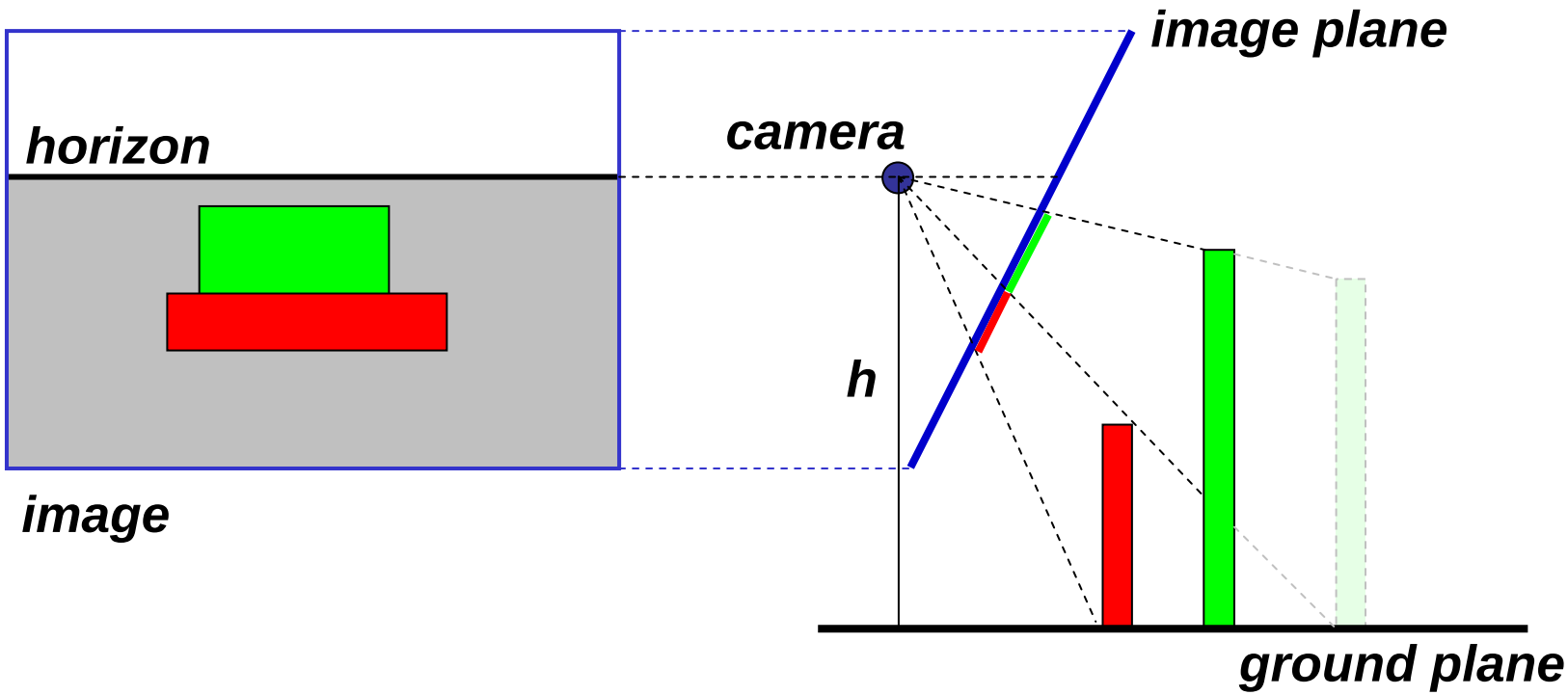
More Example Results



[Gould, Fulton, Koller, ICCV 2009]

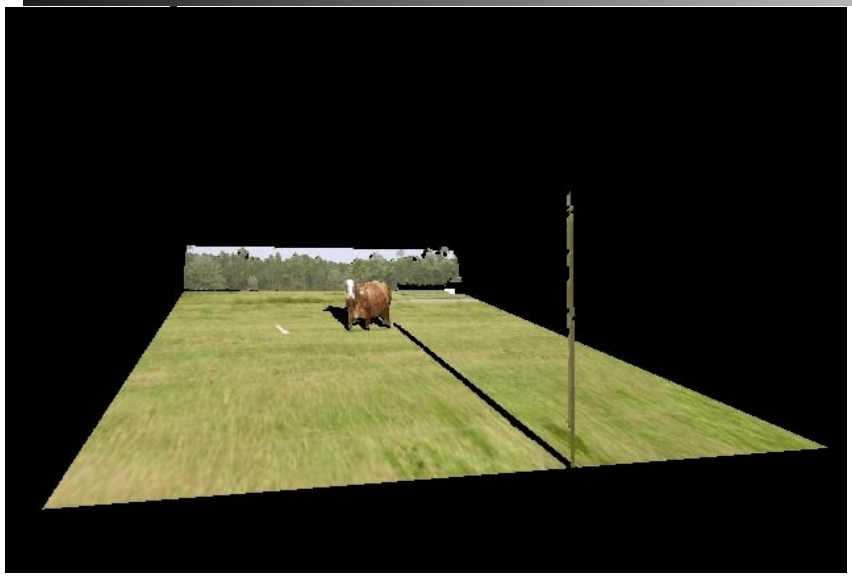
Application: 3d Reconstruction

- Estimate camera tilt from location of horizon
- Predict region 3D position using ray projected through camera plane

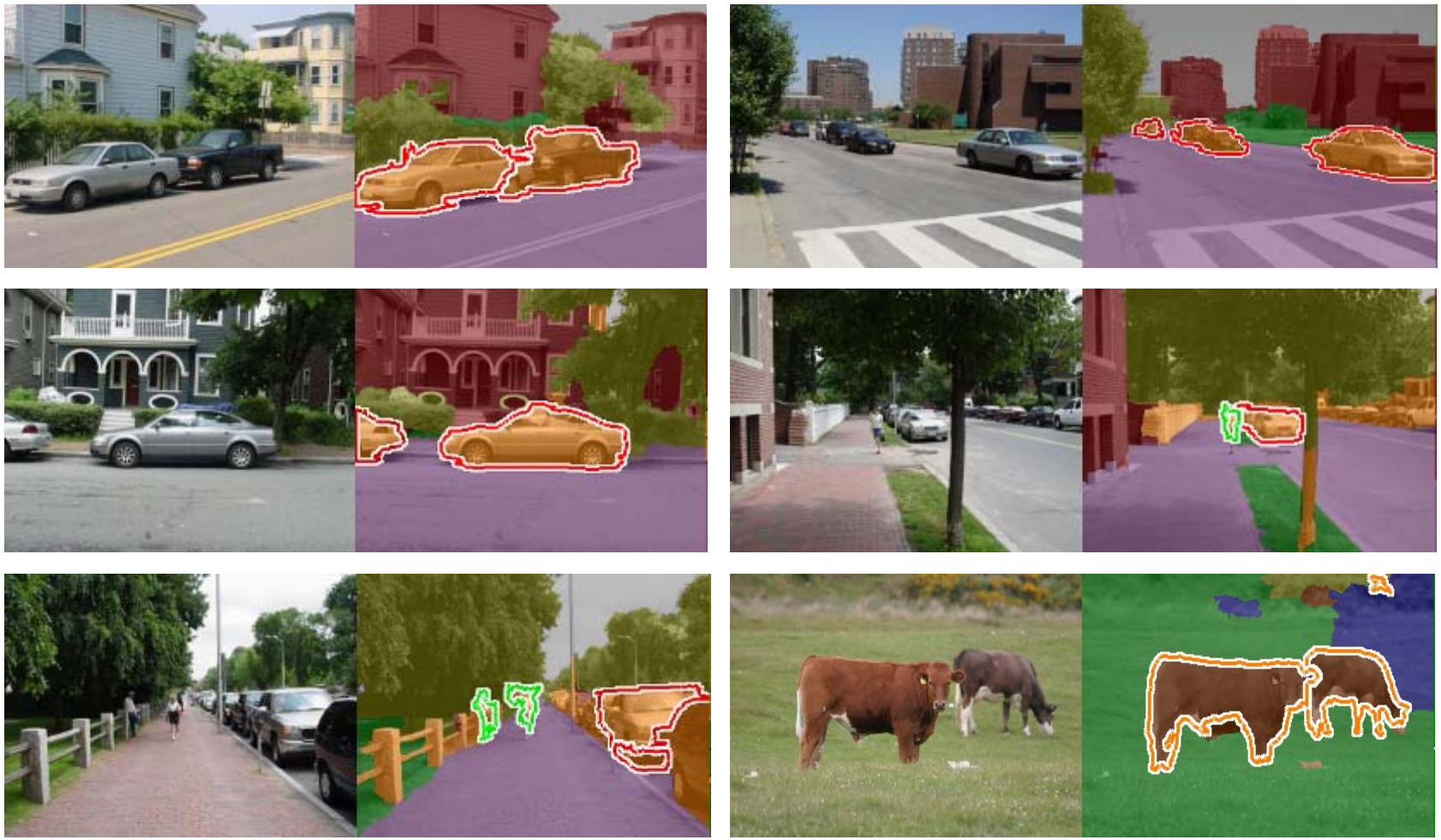


[Gould, Fulton, Koller, ICCV 2009]

Example 3D Reconstructions

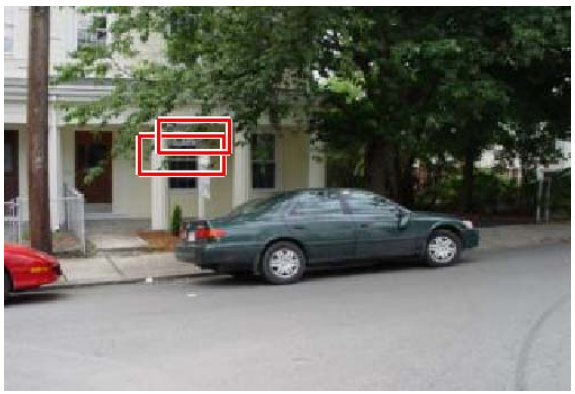


Object Detection Examples



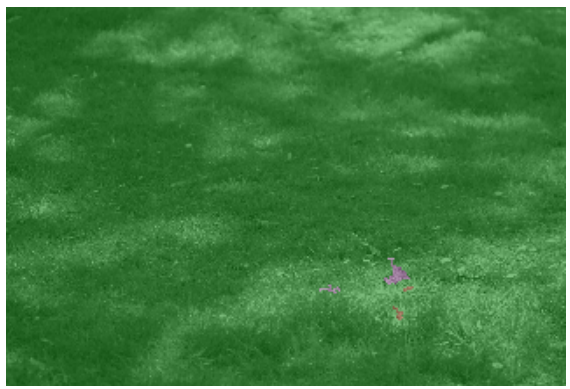
Sliding-Window Failures

Sliding-window detector top results

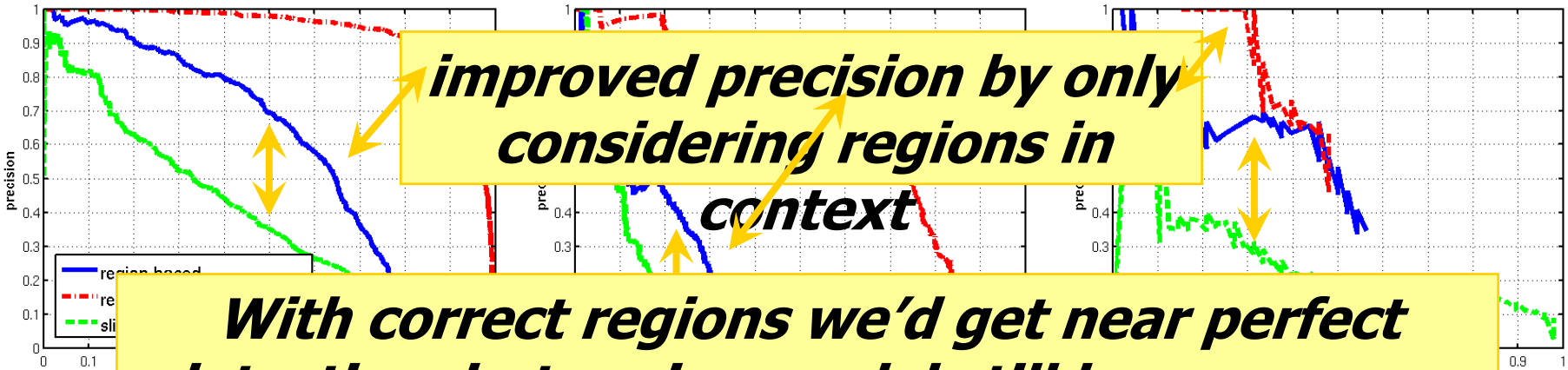


Our region-based object detector results





Object Detection

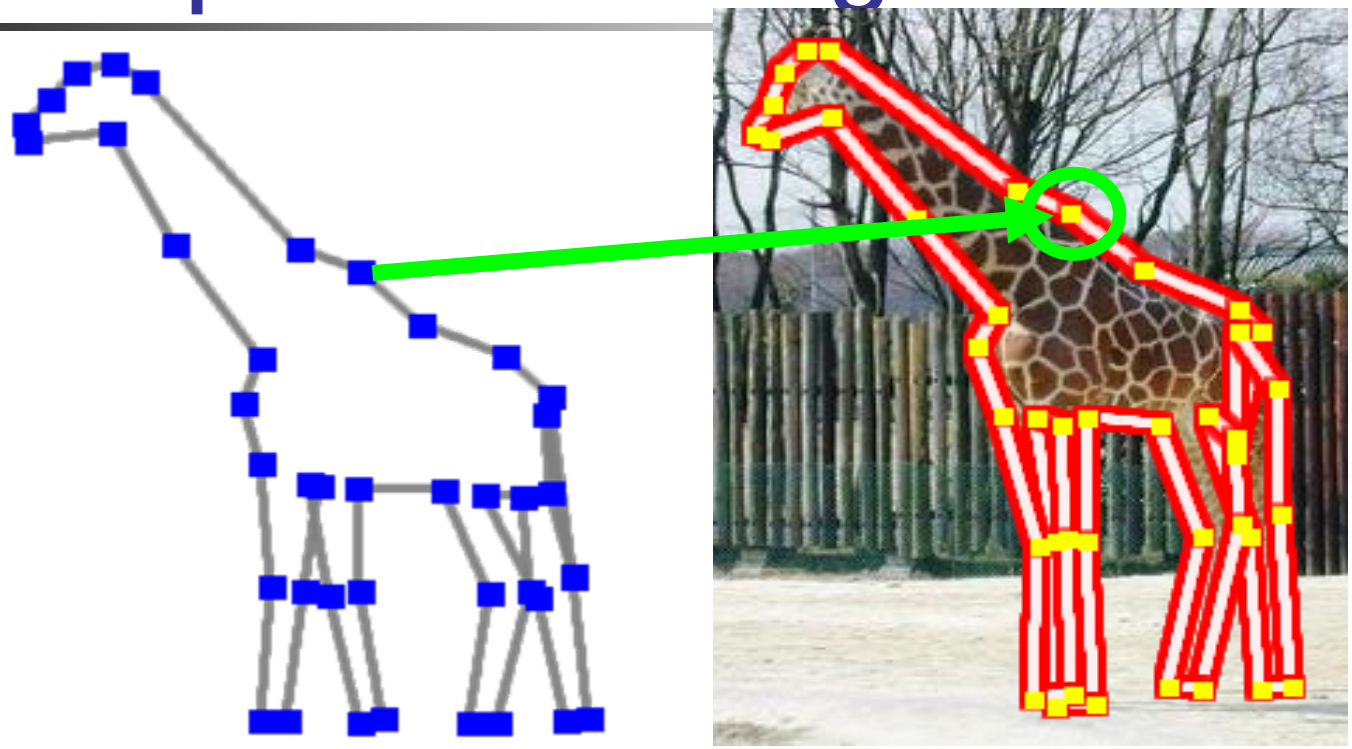


† run on Street Scene dataset

* run on subset of 21-class MSRC data

[Gould, Gao, Koller submitted]

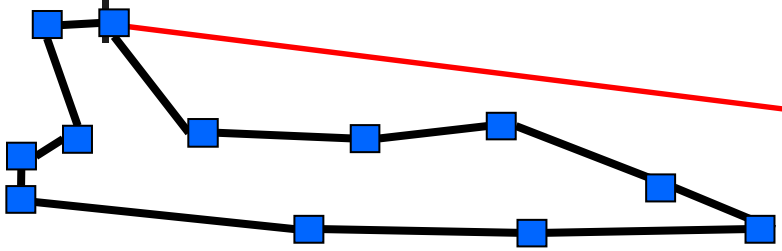
Shape-Based Segmentation



- Set of shape landmarks
- Shape defined by connecting piecewise-linear contour
- Semantic outlining = assignment L of landmarks to pixels

[Heitz, Elidan, Packer, Koller, NIPS-08b]

The LOOPS Model



L – assignment of model landmarks to pixels

$$P(L \mid O=1, I, w, \mu, \Sigma) =$$

pairwise image features

$$\frac{1}{Z} \underbrace{P_{\text{SHAPE}}(L; \mu, \Sigma)} \prod_{i,j} \exp\left\{ w_{i,j} \overbrace{F_{i,j}(l_i, l_j; I)} \right\}$$

Outlining = MAP Inference over L

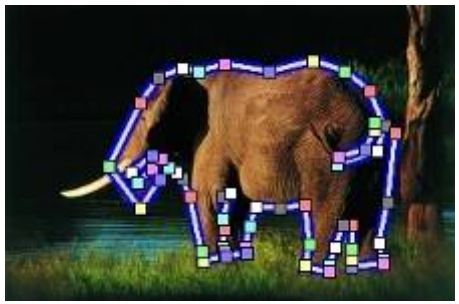
[Heitz, Elidan, Packer, Koller, NIPS-08b]

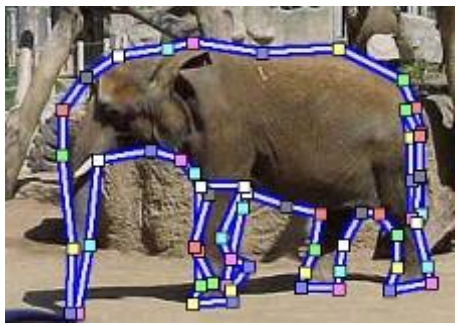
landmark detectors

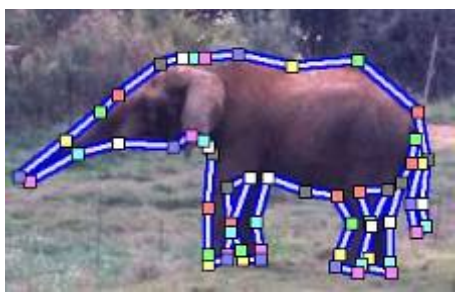
Learning the Shape Model

[UAI 2008]

Training Set:

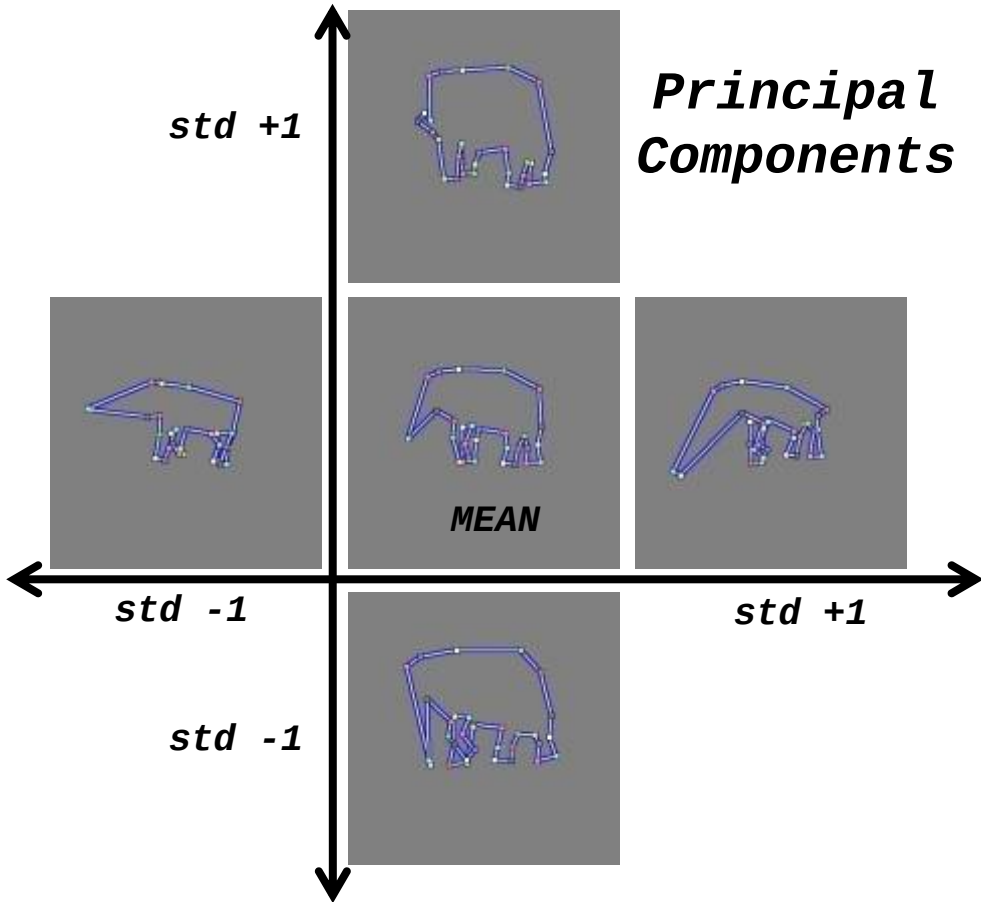






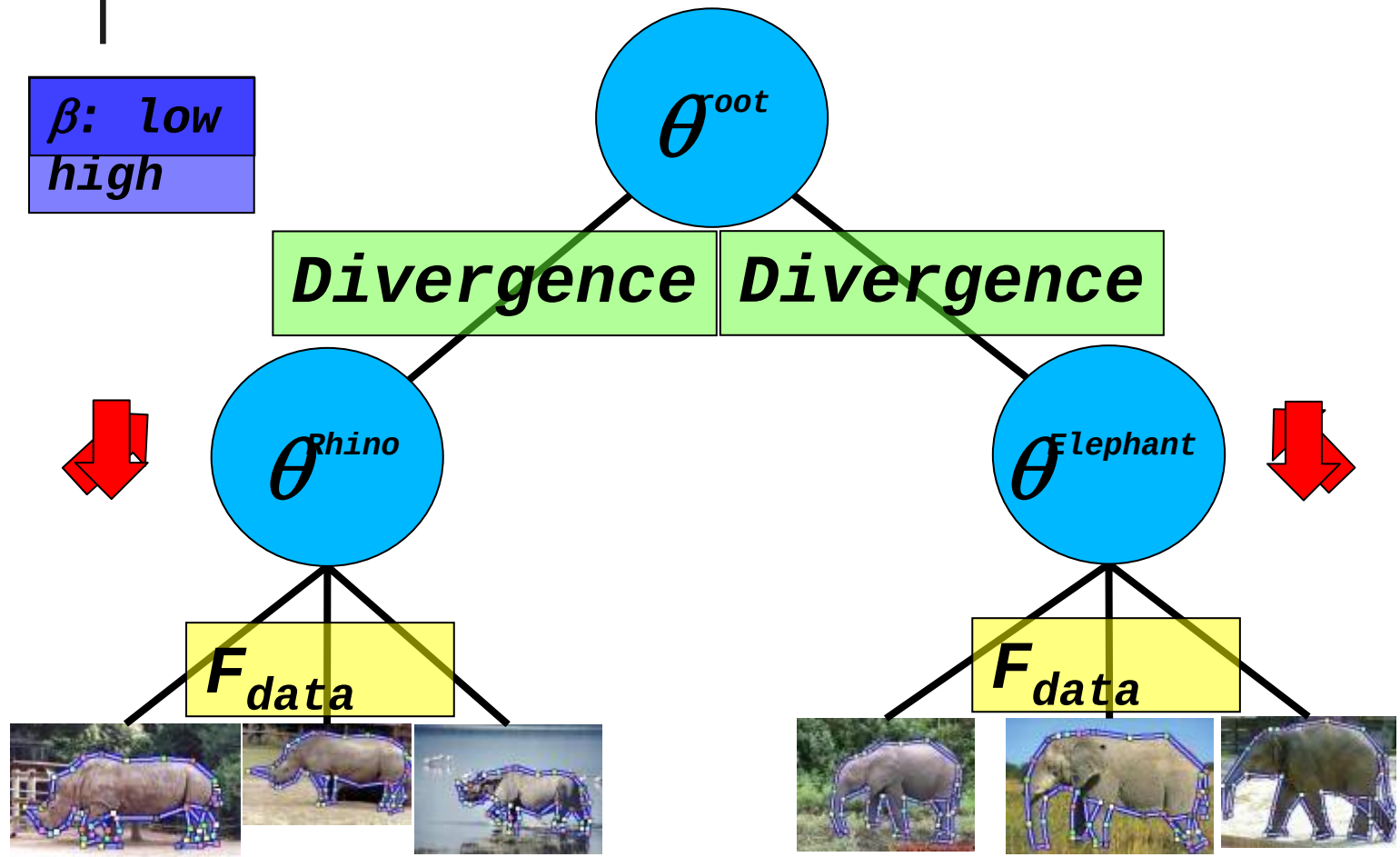
Problem:

With few instances, learned models aren't robust



Undirected Probabilistic Model

[UAI 2008]

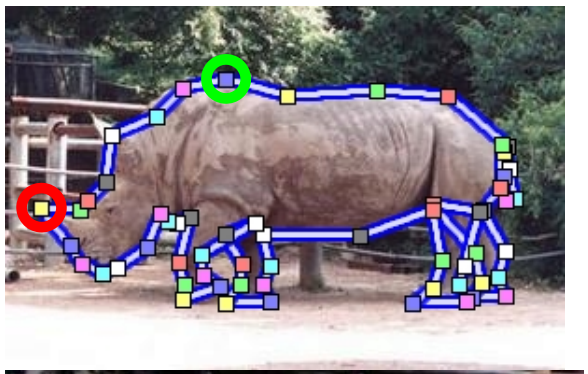


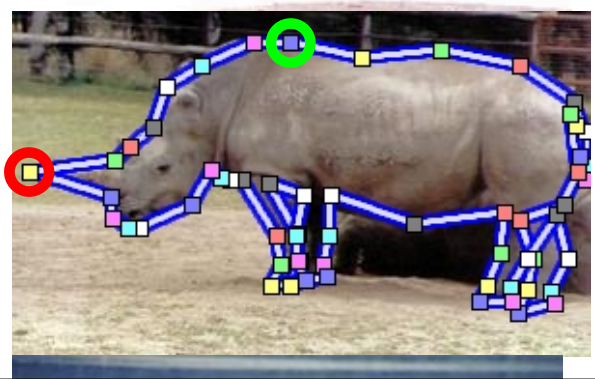
F_{data}:
Encourage parameters to explain data

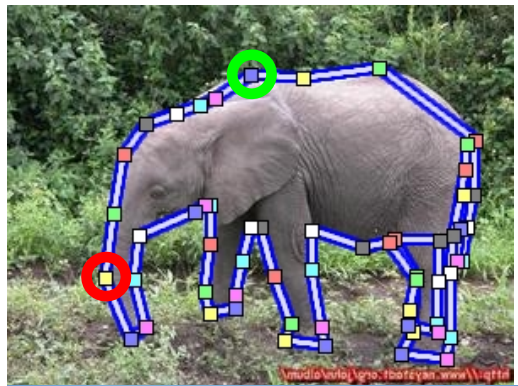
Divergence:
Encourage parameters to be similar to parents

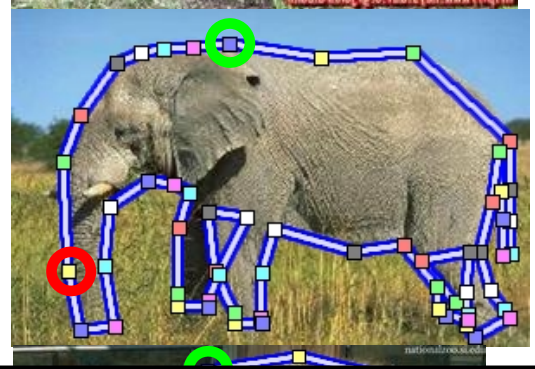
Degrees of Transfer

[UAI 2008]





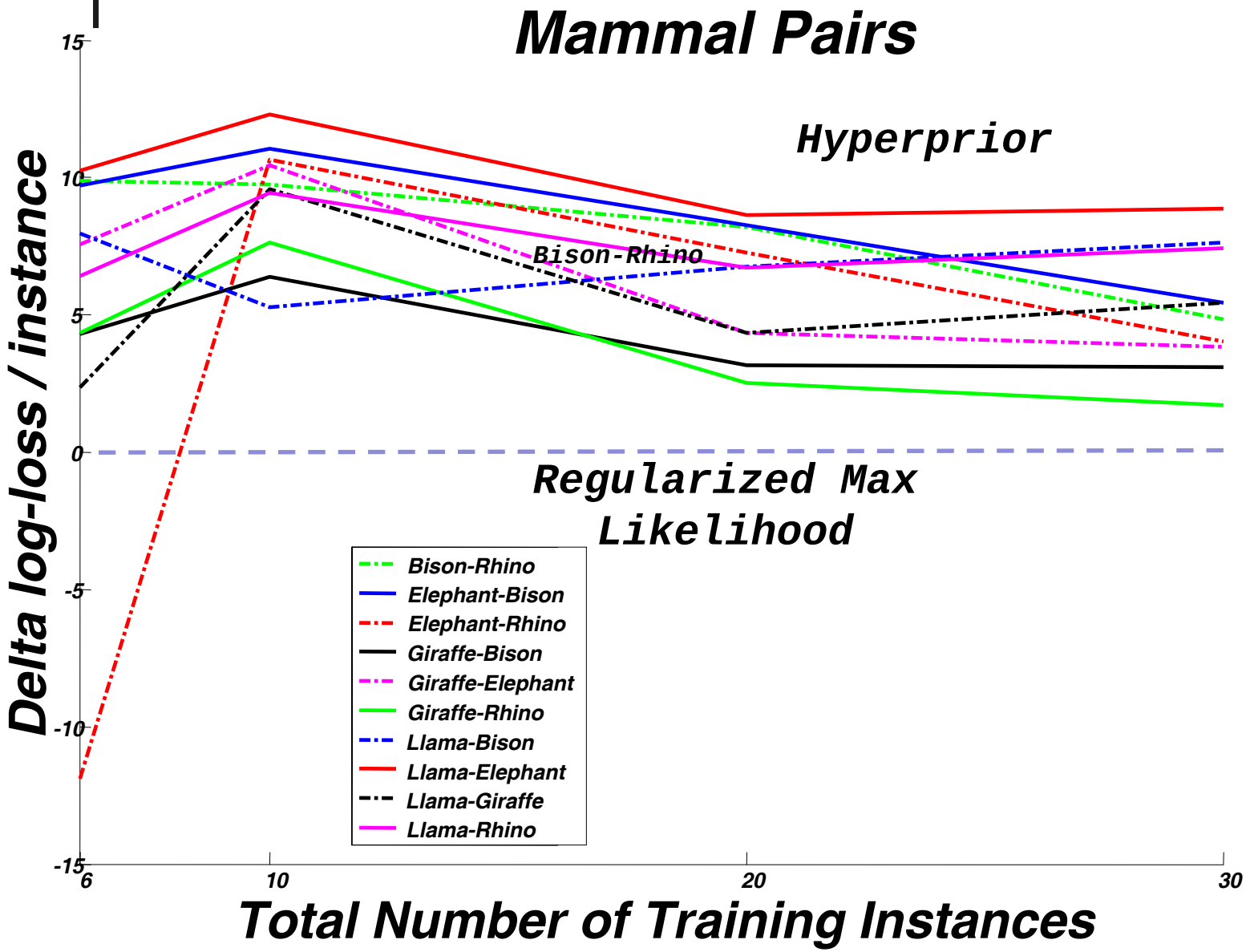




Not all parameters deserve equal sharing

[UAI 2008]

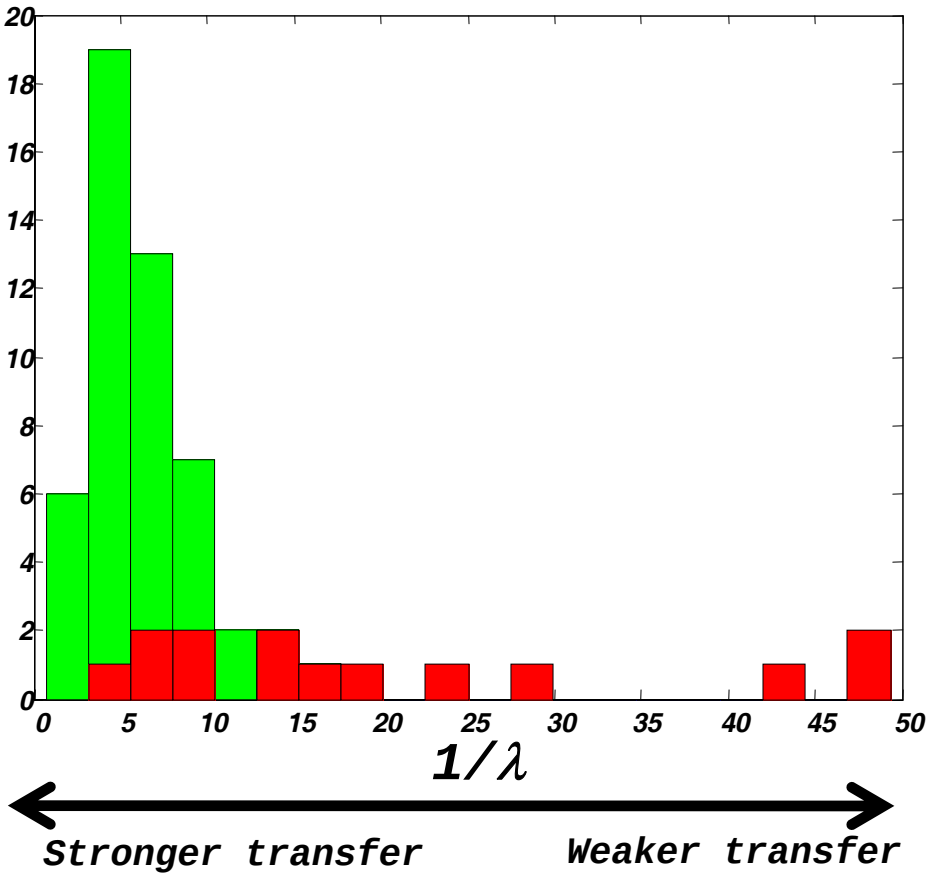
Do Degrees of Transfer Help?

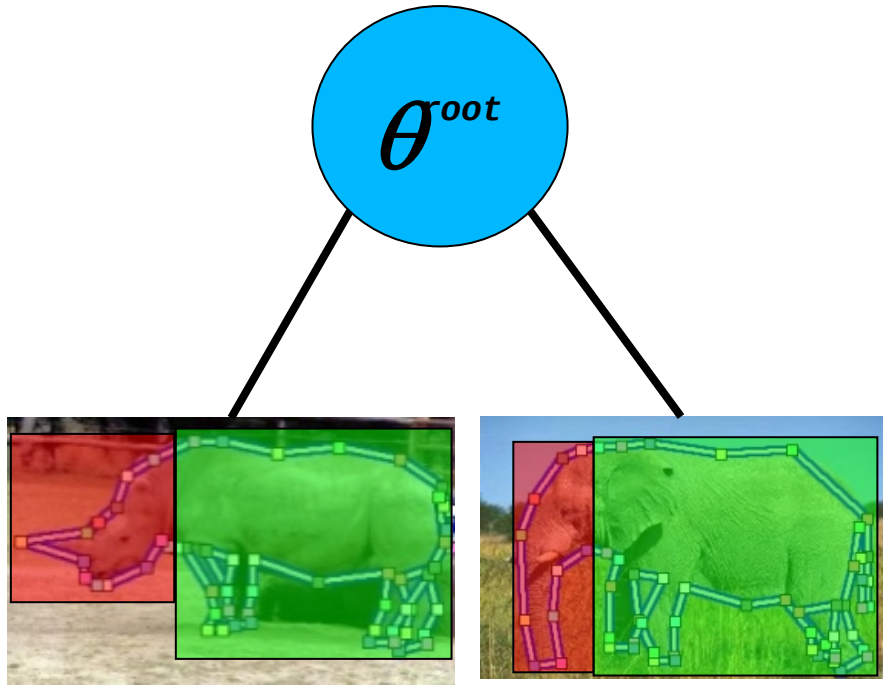


Degrees of Transfer

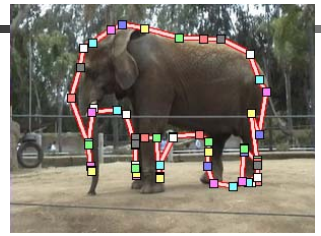
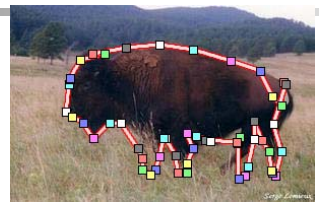
[UAI 2008]

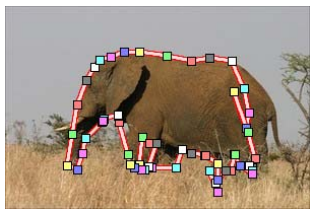
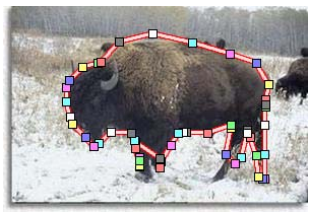
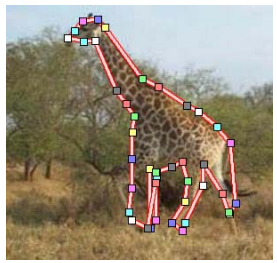
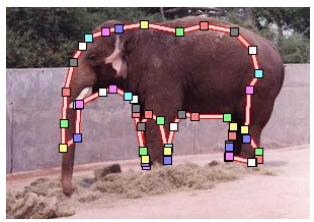
Distribution of DOT coefficients using Hyperprior

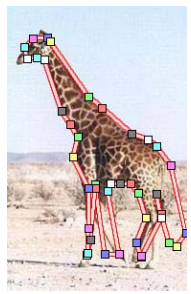
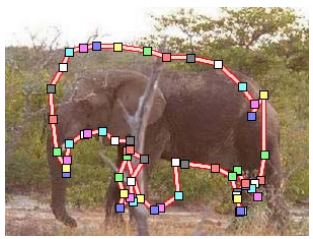
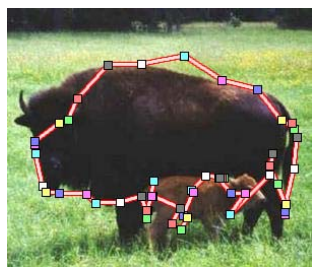
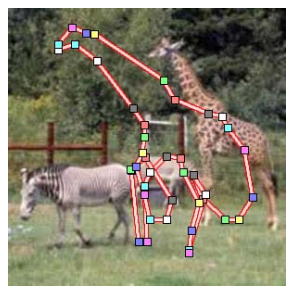




Outlining Results: Mammals

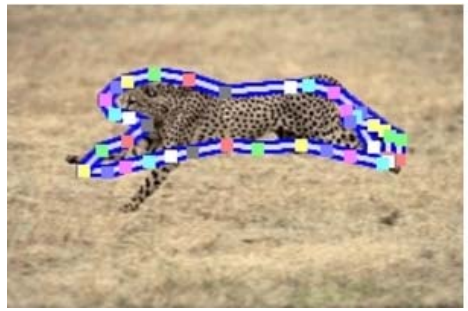





[Heitz, Elidan, Packer, Koller, NIPS-08b]

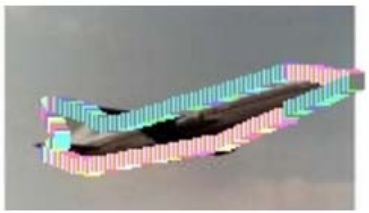
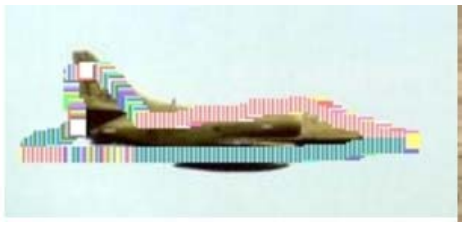
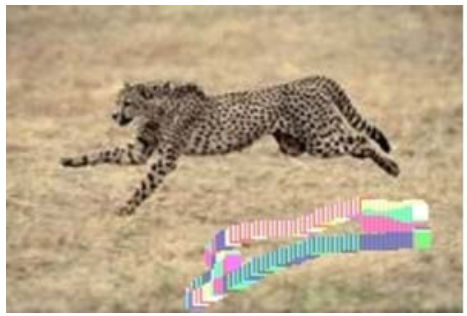
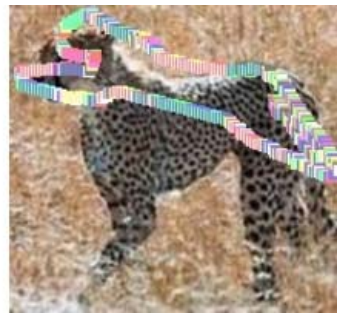
Outlining: Comparison

LOOPS

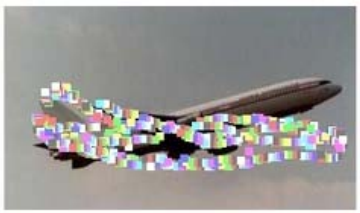





OBJ CUT [Kumar et al., CVPR 05]

kAS [Ferrari et al., CVPR 07]




[Heitz, Elidan, Packer, Koller, NIPS-08b]

Descriptive Classification



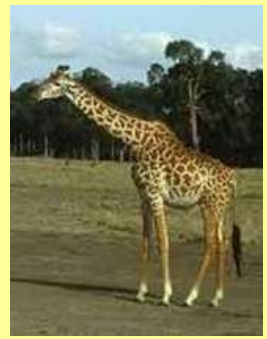
...


Standard Approach

Standard

Feature-based

Classifier

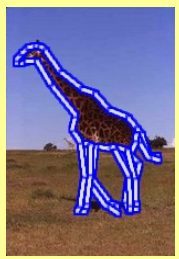


$(X_1, -)$ $(X_2, +)$ $(X_N, -)$

TRAINING DATA

H(x) = +0.6

"UP"



...


LOOPS Approach

LOOPS

X_1 X_2 X_N

LOOPS TRAINING DATA

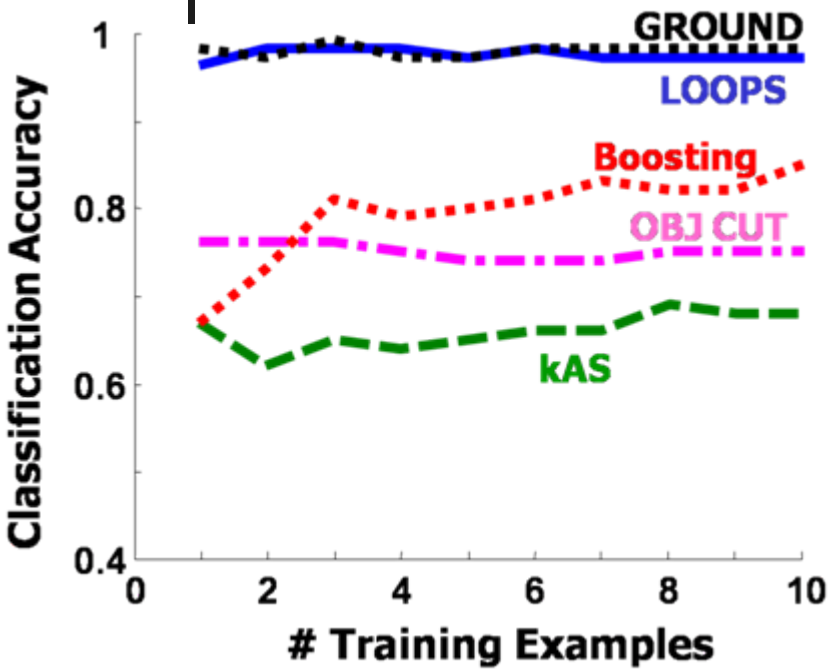
2 << N

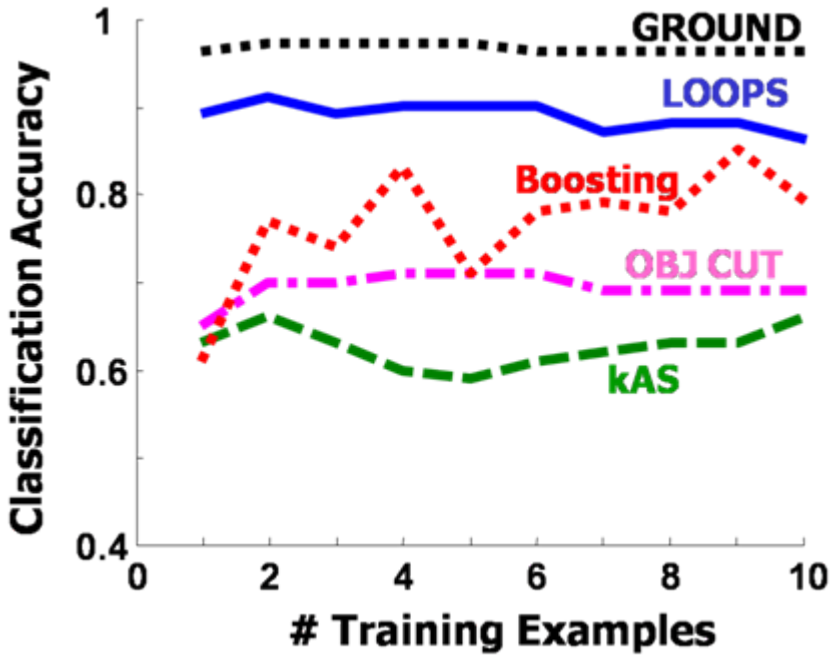
(1,-) ***(2,+)***

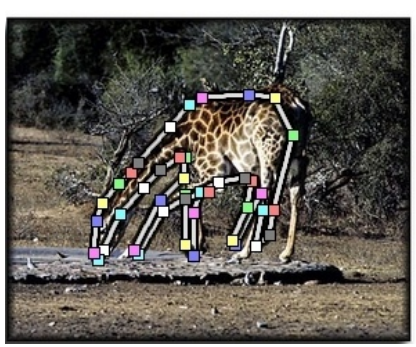
CLASSIFICATION TRAINING DATA

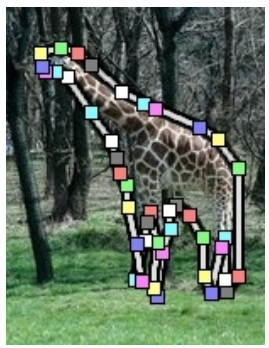
"UP"

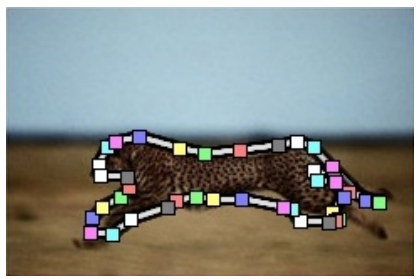
Mammals

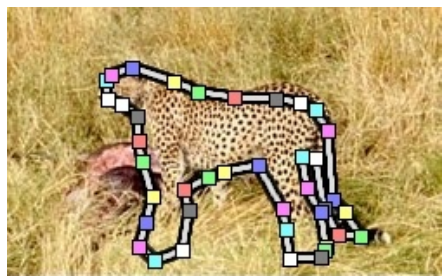

GROUND
LOOPS
Boosting
OBJ CUT
kAS
Classification Accuracy
Training Examples
1
0.8
0.6
0.4
0
2
4
6
8
10


GROUND
LOOPS
Boosting
OBJ CUT
kAS
Classification Accuracy
Training Examples
1
0.8
0.6
0.4
0
2
4
6
8
10

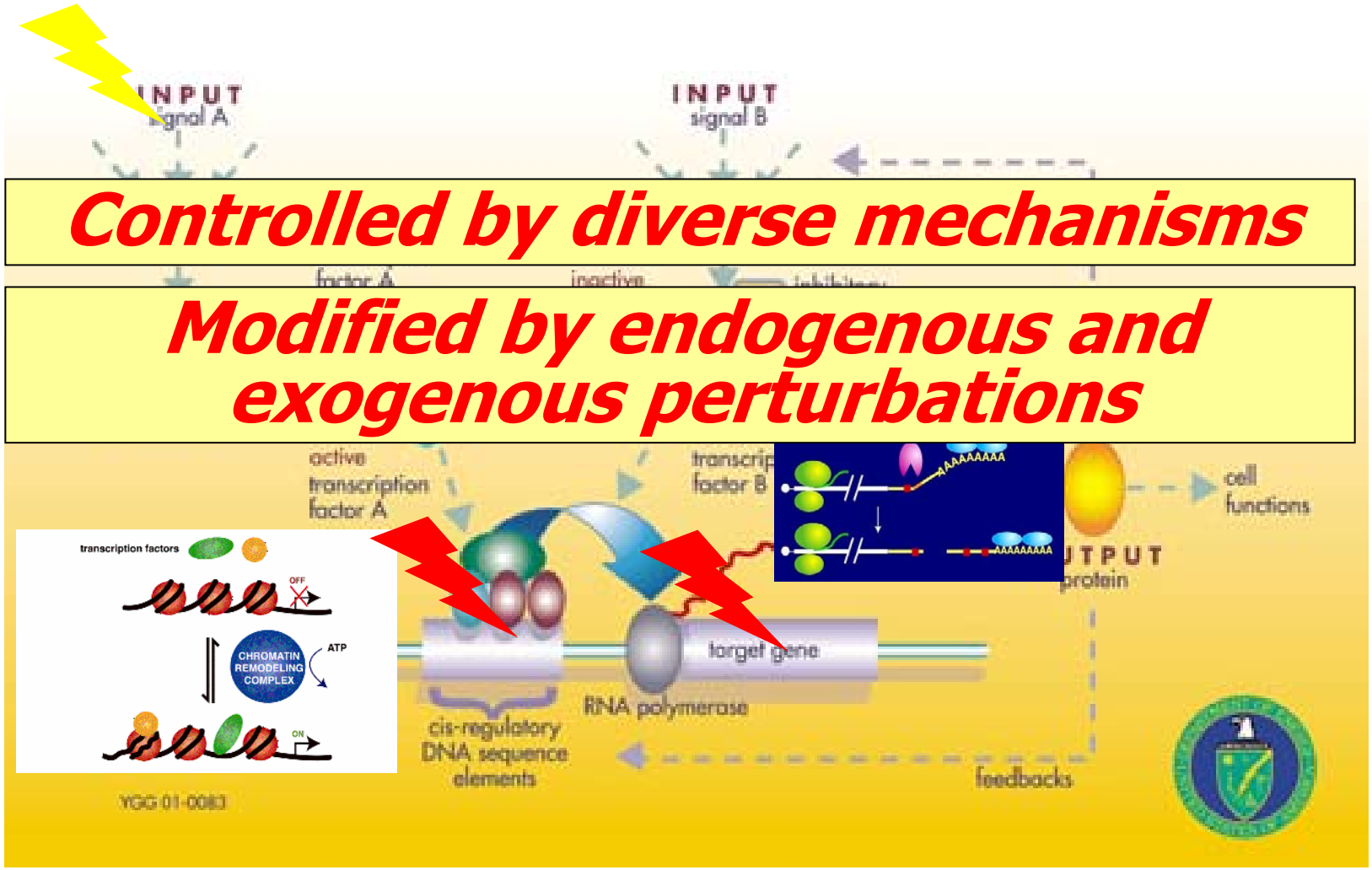




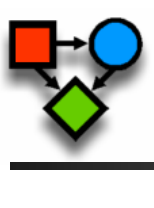




Gene Regulatory Networks



http://en.wikipedia.org/wiki/Gene_regulatory_network



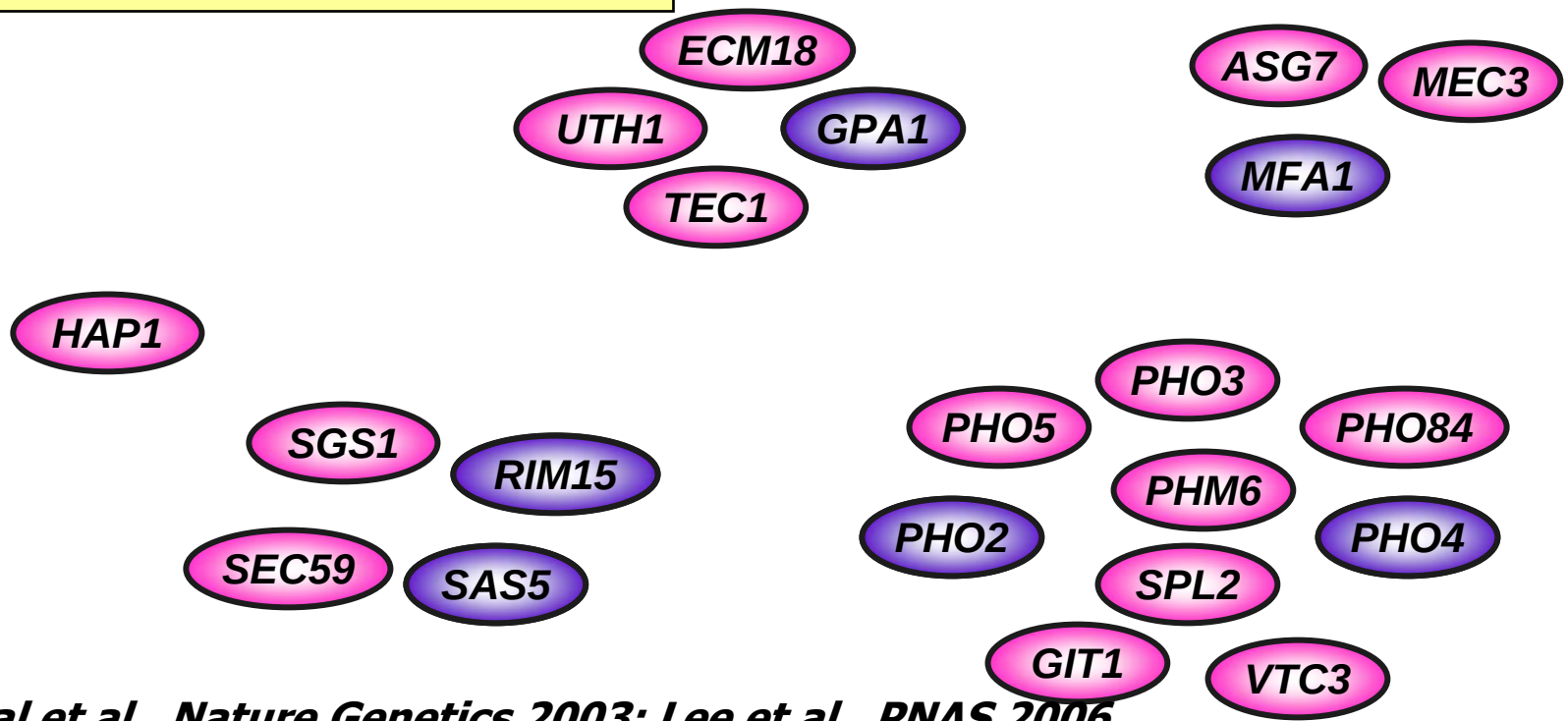
Goals

- Infer **regulatory network and mechanisms** that control gene expression
- Identify **effect of perturbations** on network
- Understand effect of gene regulation on **phenotype**

Regulatory Network I

- mRNA level of regulator can indicate its activity level
- Target expression is predicted by expression of its regulators
- Use expression of regulatory genes as regulators

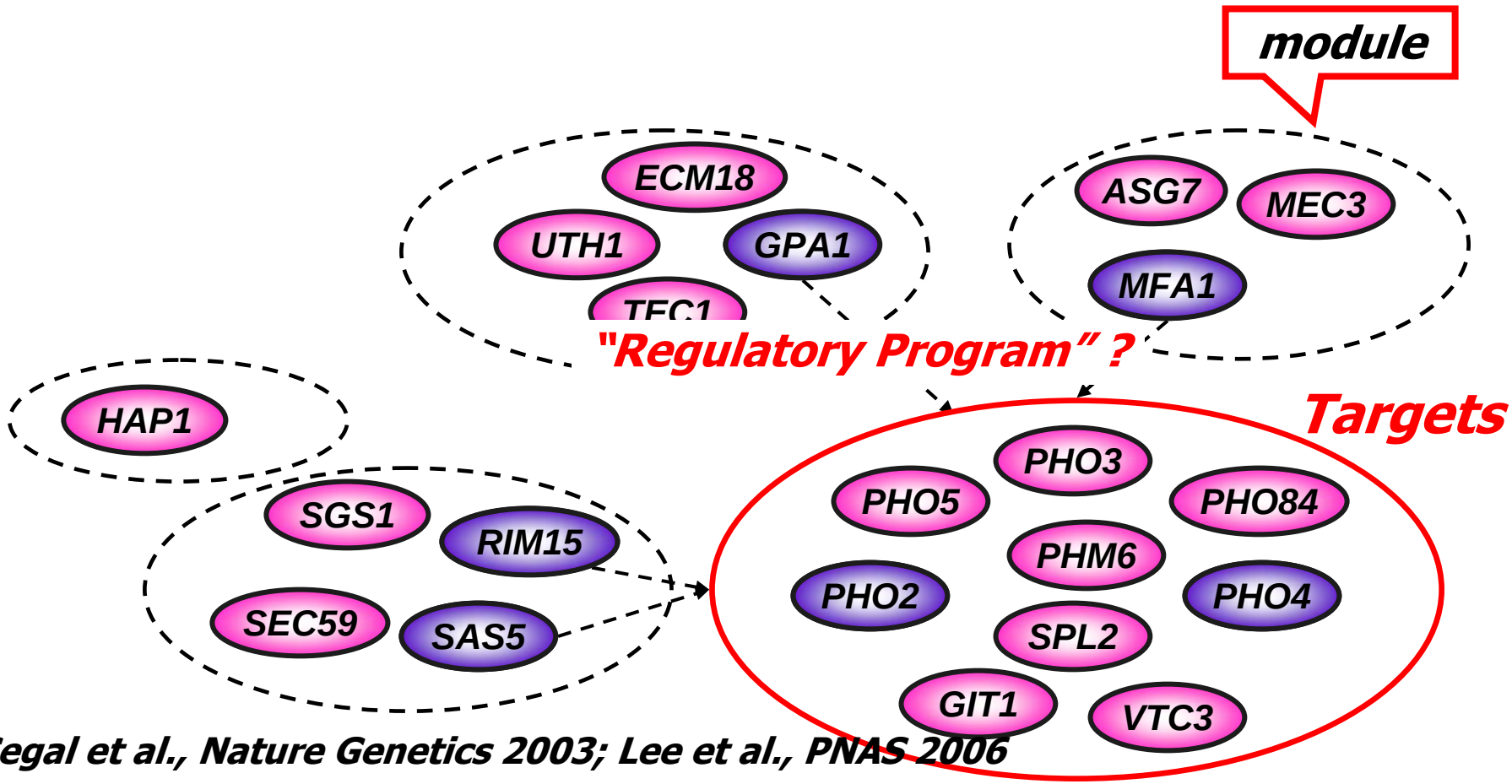
Transcription factors, signal transduction proteins, mRNA processing factors, …



Segal et al., Nature Genetics 2003; Lee et al., PNAS 2006

Regulatory Network II

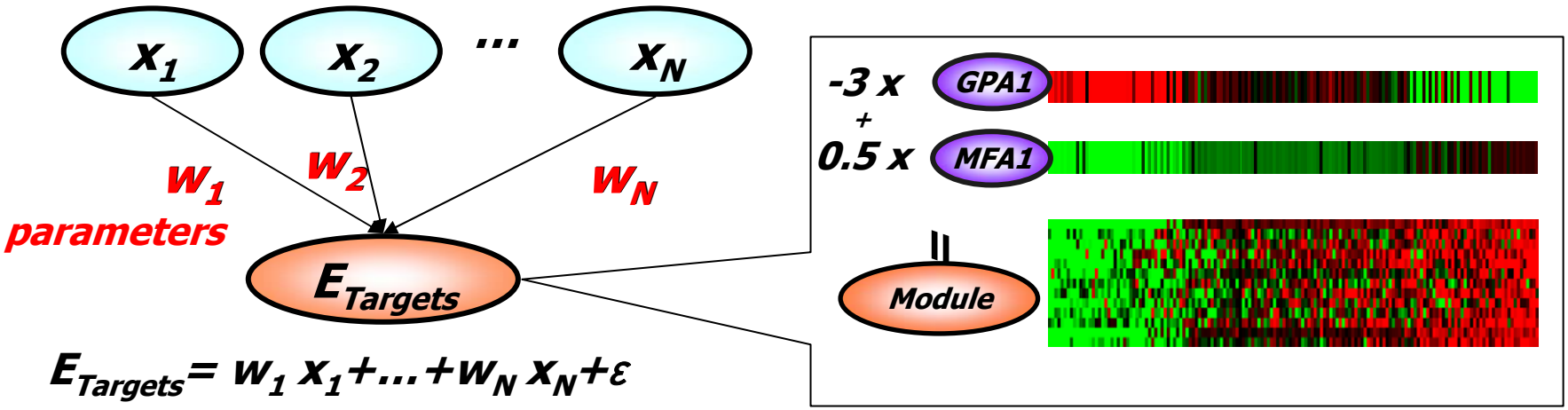
- Co-regulated genes have similar regulation program
- Exploit modularity and predict expression of entire module
- Allows uncovering complex regulatory programs



Segal et al., Nature Genetics 2003; Lee et al., PNAS 2006

Regulation as Linear Regression

$$\text{minimize}_w\ (\Sigma w_i x_i - E_{Targets})^2$$

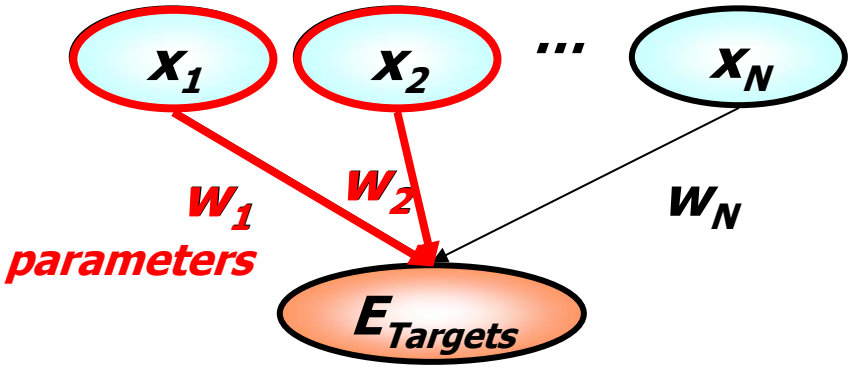


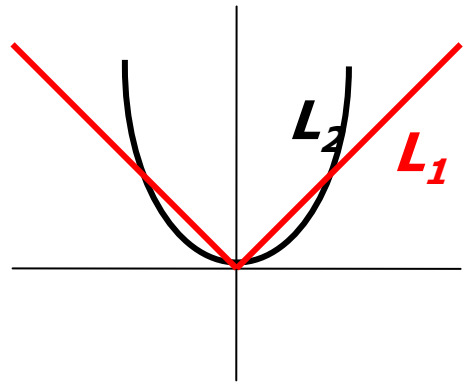
- But we often have hundreds or thousands of regulators
- ... and linear regression gives them all nonzero weight!

Problem: This objective learns too many regulators

Lasso* (L_1) Regression

$$\text{minimize}_{\mathbf{w}} (w_1x_1 + \ldots w_Nx_N - E_{Targets})^2 + \Sigma\, \mathbf{C}\, |w_i|$$



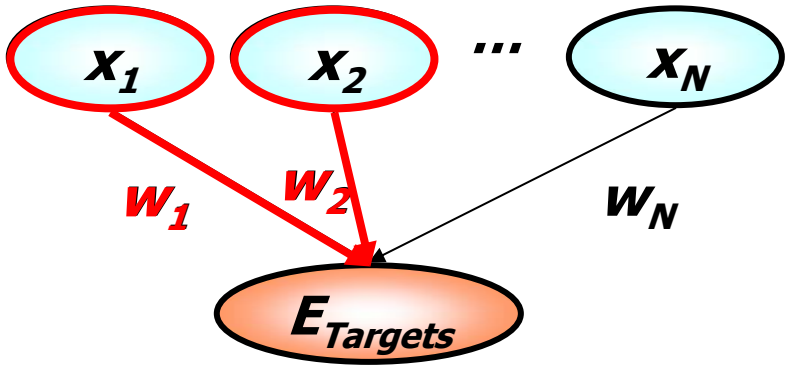


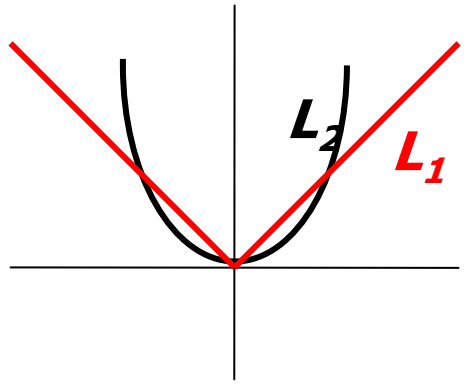
- Induces sparsity in the solution **w** (many w_i's set to zero)
  - Provably selects "right" features when many features are irrelevant
- Convex optimization problem
  - Unique global optimum
  - Efficient optimization
- But, arbitrary choice among correlated regulators

**** Tibshirani, 1996***

Elastic Net* Regression

$$\text{minimize}_{\mathbf{w}} (w_1x_1 + \ldots w_Nx_N - E_{Targets})^2 + \Sigma\, \mathbf{C}\, |w_i| + \Sigma\, \mathbf{D}\, w_i^2$$





- Induces sparsity
- But avoids arbitrary choices among relevant features

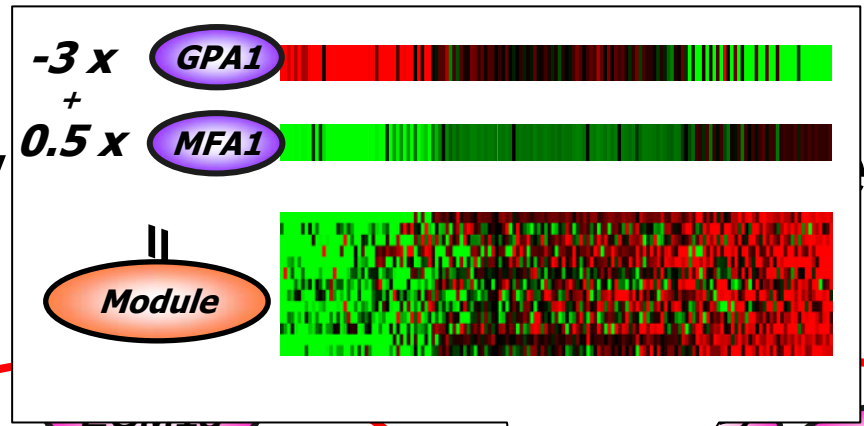
- Convex optimization problem
  - Unique global optimum
  - Efficient optimization algorithms

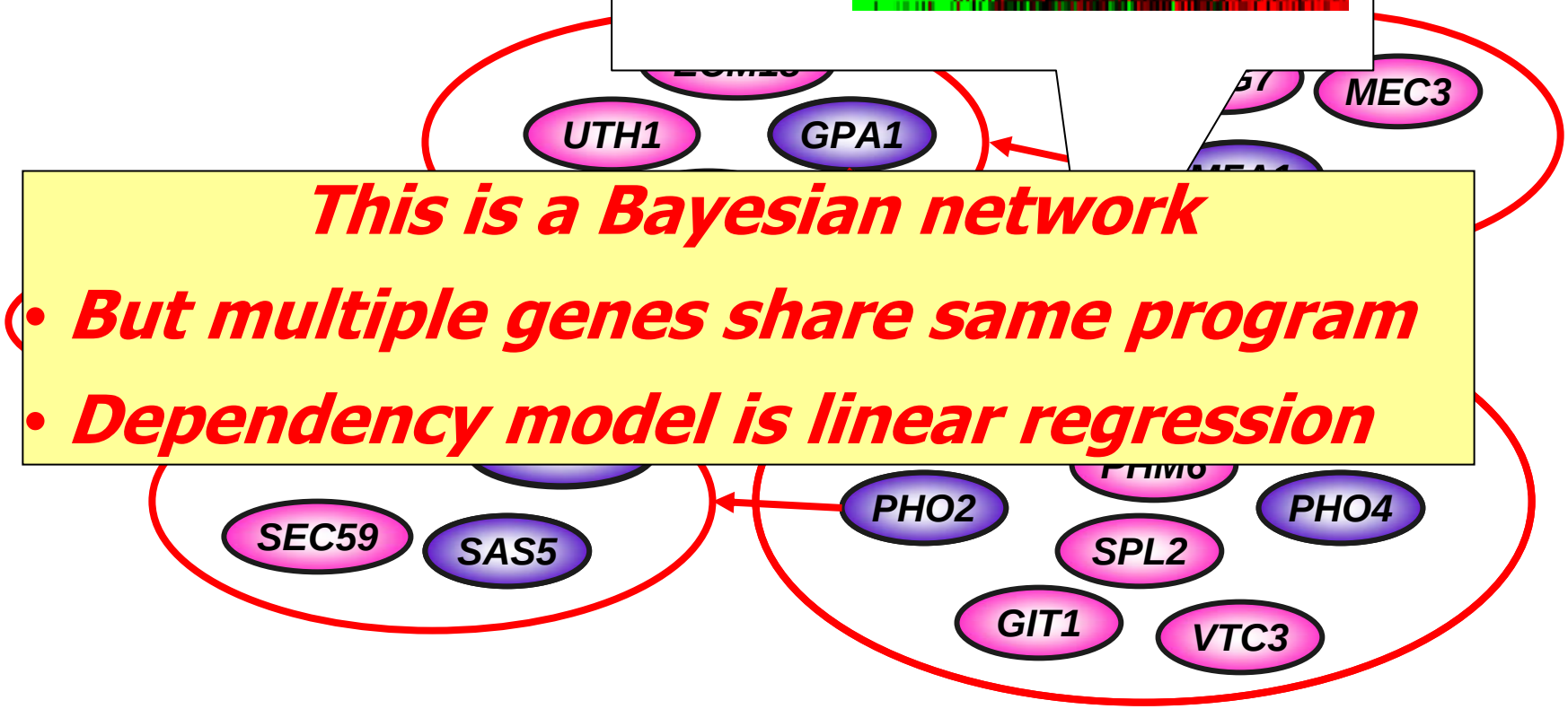
Lee et al., PLOS Genetics 2009

** Zhou & Hastie, 200*

Learning Regulatory Network

- Cluster genes into
- Learn a regulatory





This is a Bayesian network

- ***But multiple genes share same program***
- ***Dependency model is linear regression***

Lee et al., PLoS Genet 20

Genotype → phenotype

Different sequences

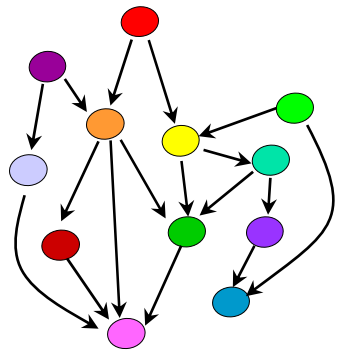
...ACTCGGTTGGCCTAAATTCGGCCCGG...

...ACCCGGTAGGCCTTAATTCGGCCCGG...

:

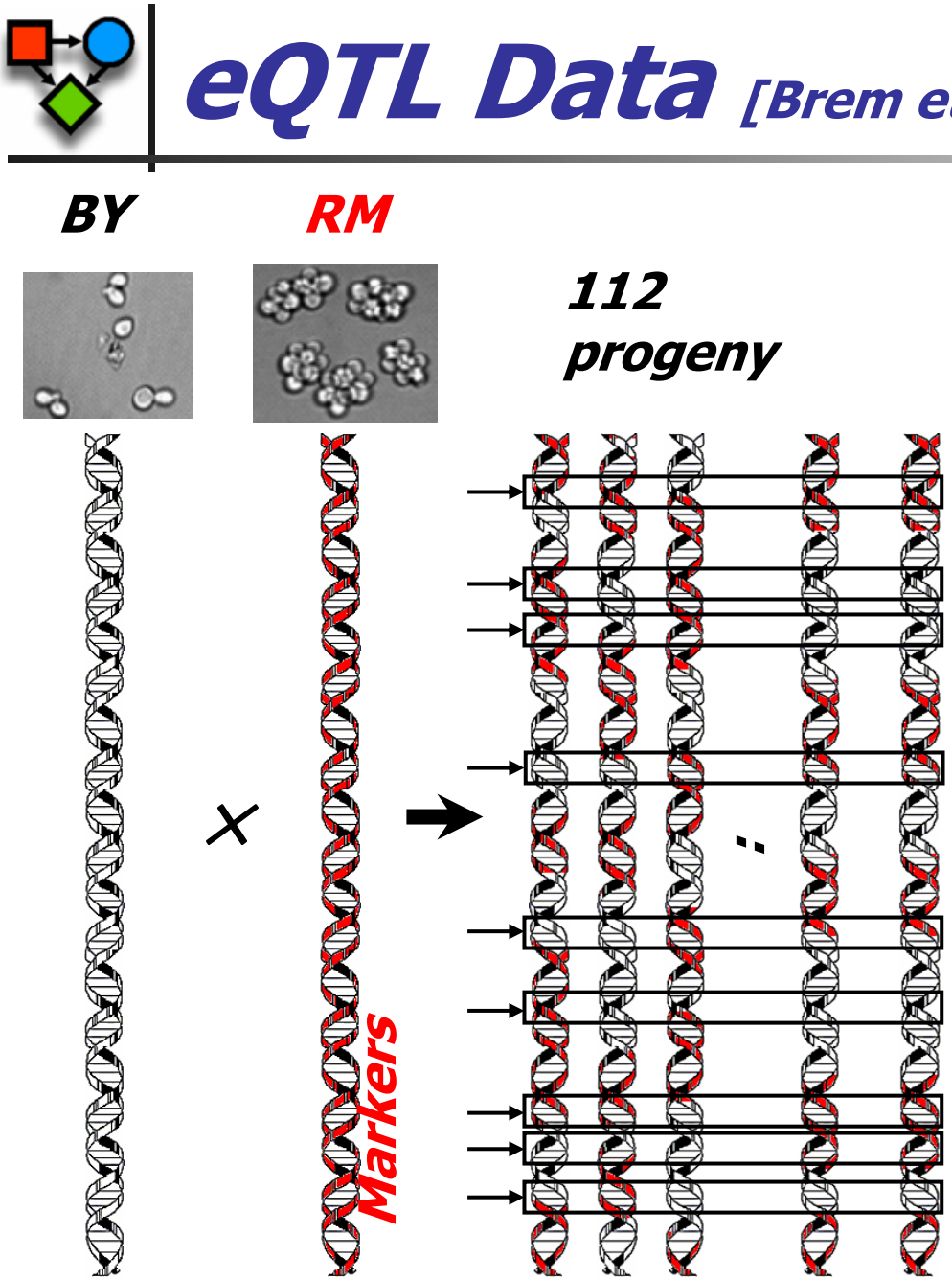
...ACTCGGTAGGCCTATATTCGGCCGGG...

Perturbations to regulatory network



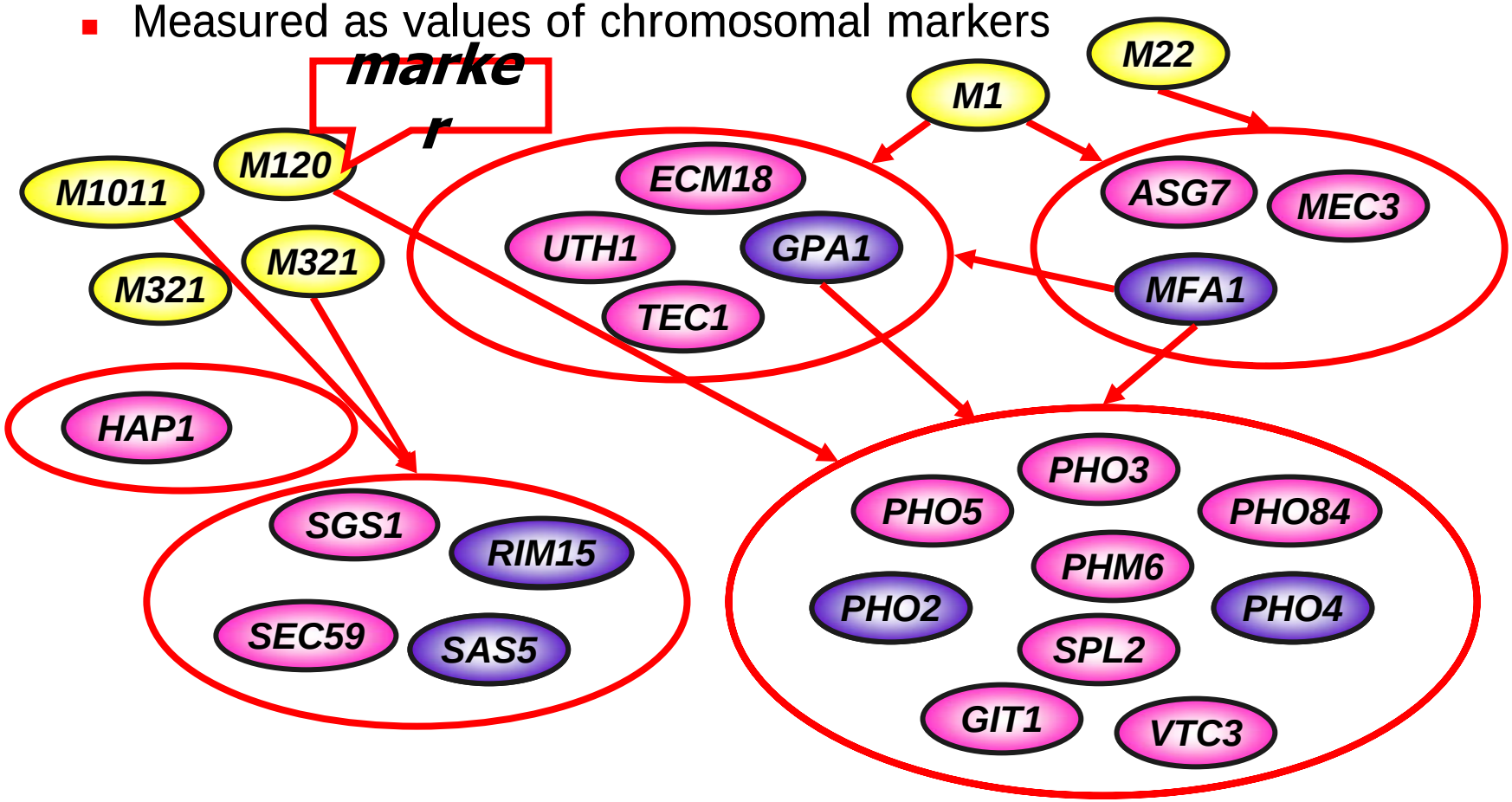
Different phenotypes

eQTL Data [Brem et al. (2002) Science]



LirNet Regulatory network

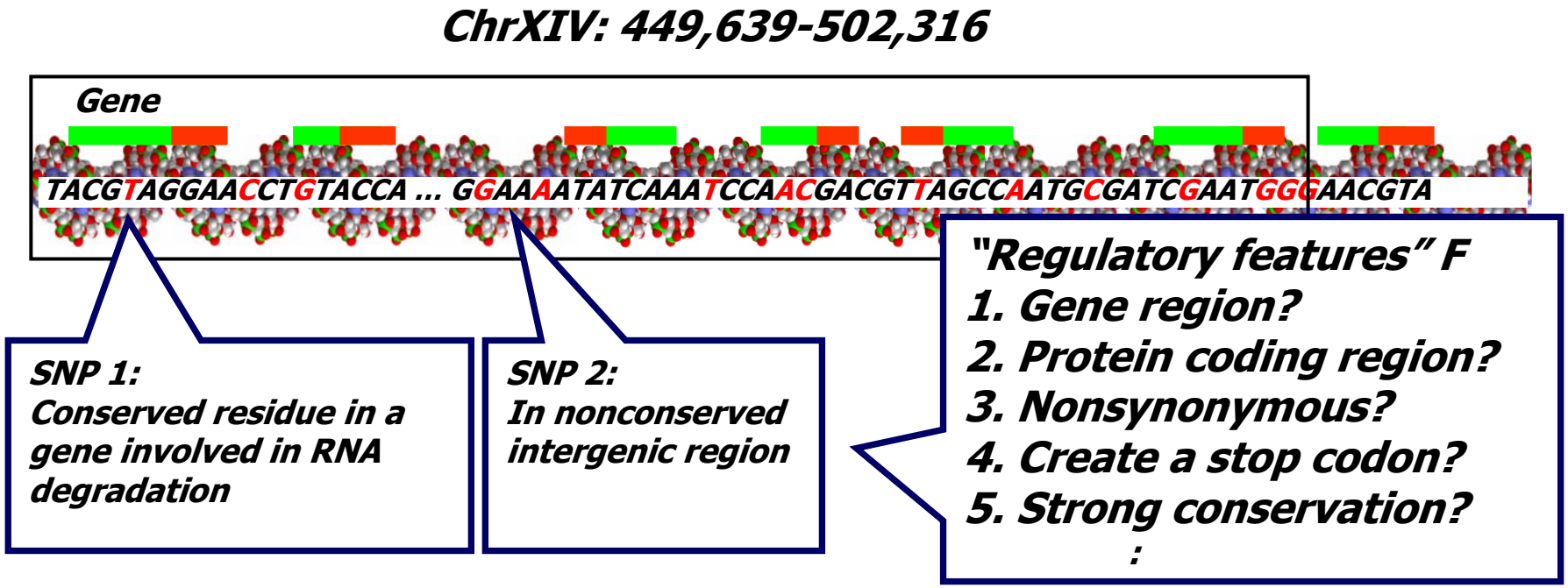
- E-regulators: Activity (expression) of regulatory genes
- G-regulators: Genotype of genes
  - Measured as values of chromosomal markers



Lee et al., PNAS 2006; Lee et al., PLoS Genetics 2009

Motivation

- Not all SNPs are equally likely to be causal.

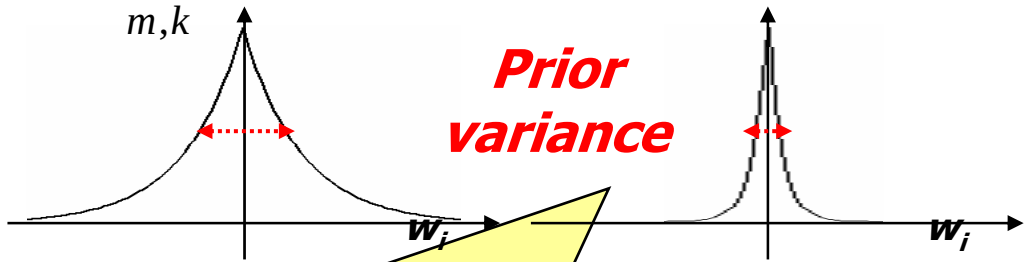


- ***Idea: Prioritize SNPs that have "good" regulatory features***
- ***But how do we weight different features?***

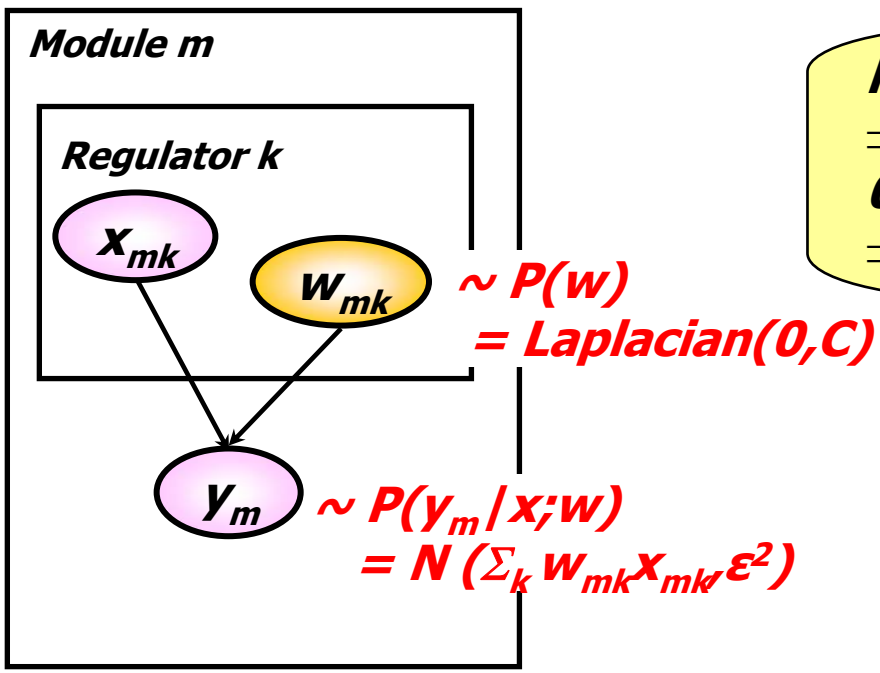
Lee et al., PLOS Genetics 2009

Bayesian L_1-Regularization

$$\sum_{m=1}^{M} \log P(y_m \mid \boldsymbol{x}_m, \boldsymbol{w}_m) - \sum_{m,k} C \mid w_{m,k} \mid$$



higher prior variance
⇒ weight can more easily deviate from 0
⇒ regulator more likely to be selected

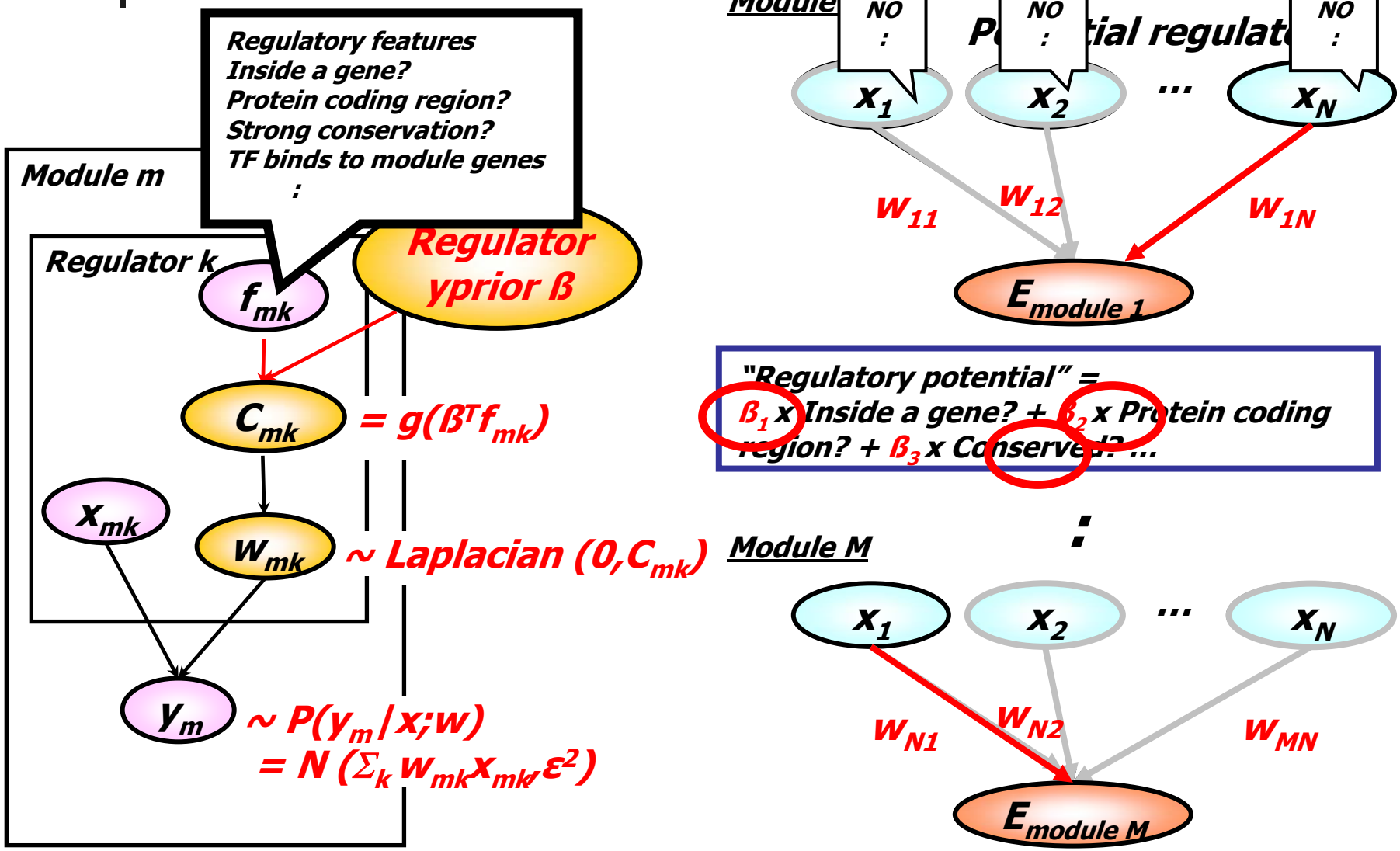


~ P(w)
= Laplacian(0,C)

~ $P(y_m \mid x_i;w)$
= $N(\Sigma_k w_{mk}x_{mk}, \varepsilon^2)$

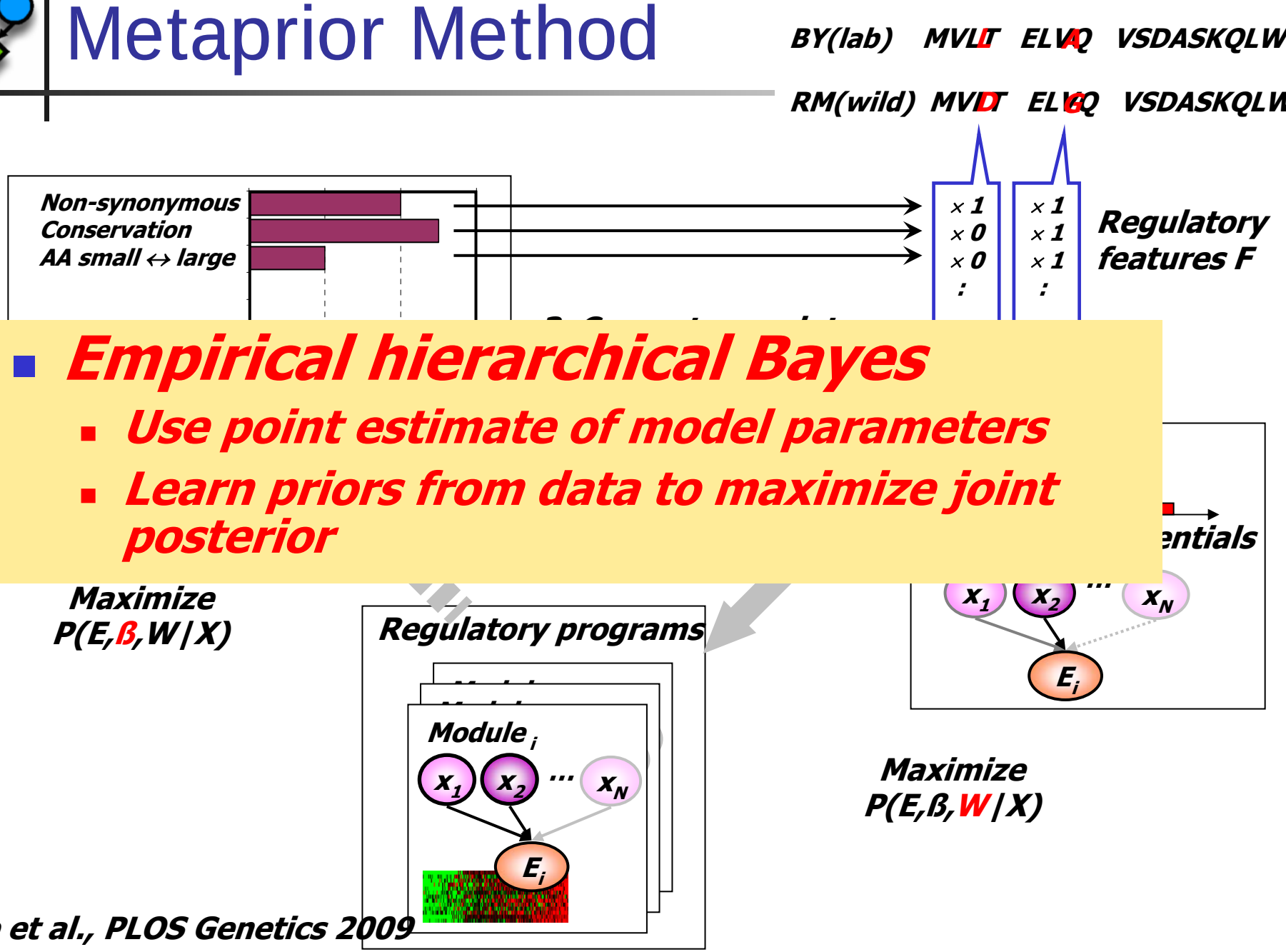
Lee et al., PLOS Genetics 2009

Metaprior Model (Hierarchical Bayes)



Lee et al., PLOS Genetics 2009

Metaprior Method



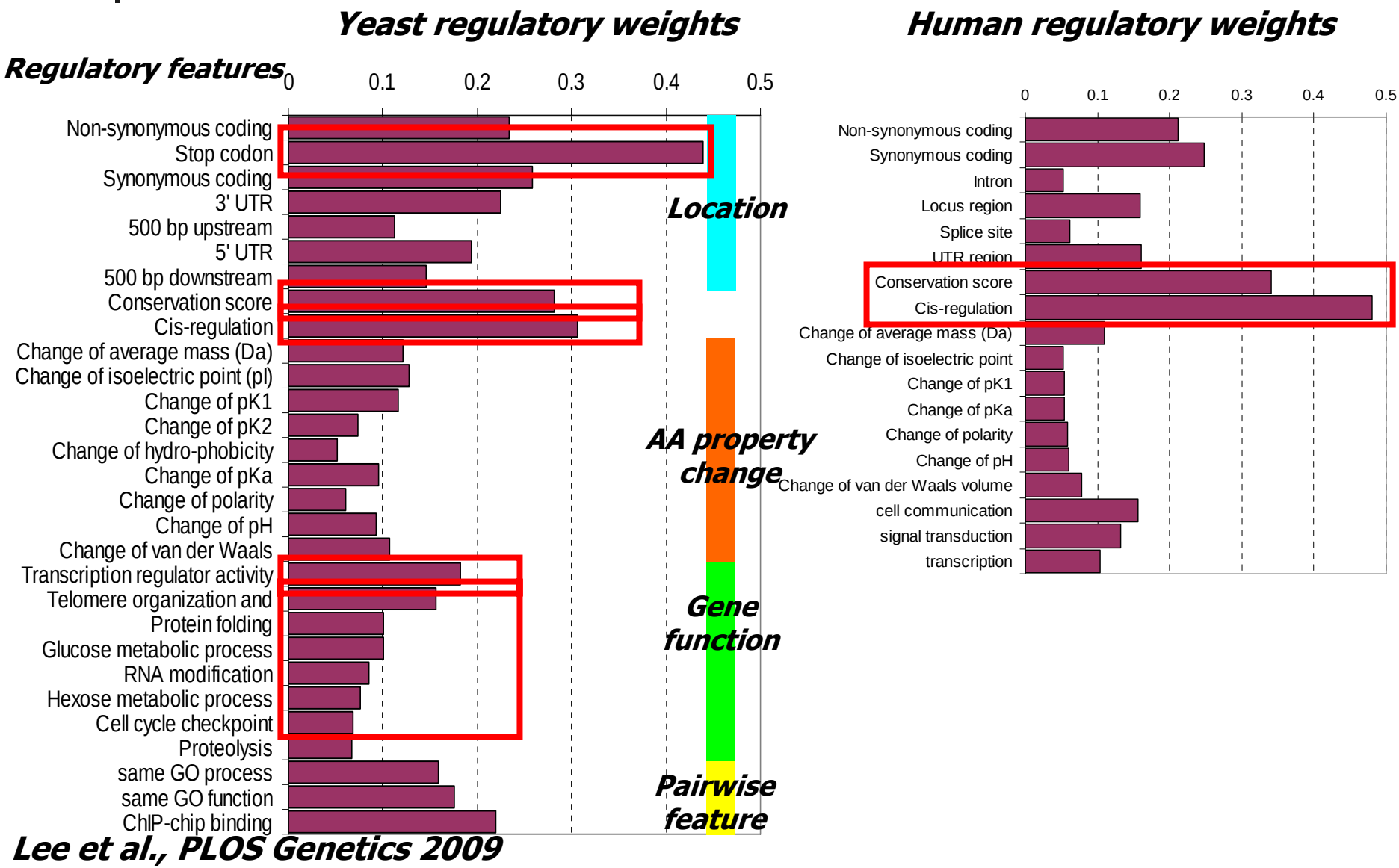
Transfer Learning

- What do regulatory potentials do?
  - They do **not** change selection of “strong” regulators – those where prediction of targets is clear
  - They only help disambiguate between weak ones

- Strong regulators help teach us what to look for in other regulators

Transfer of knowledge between different prediction tasks

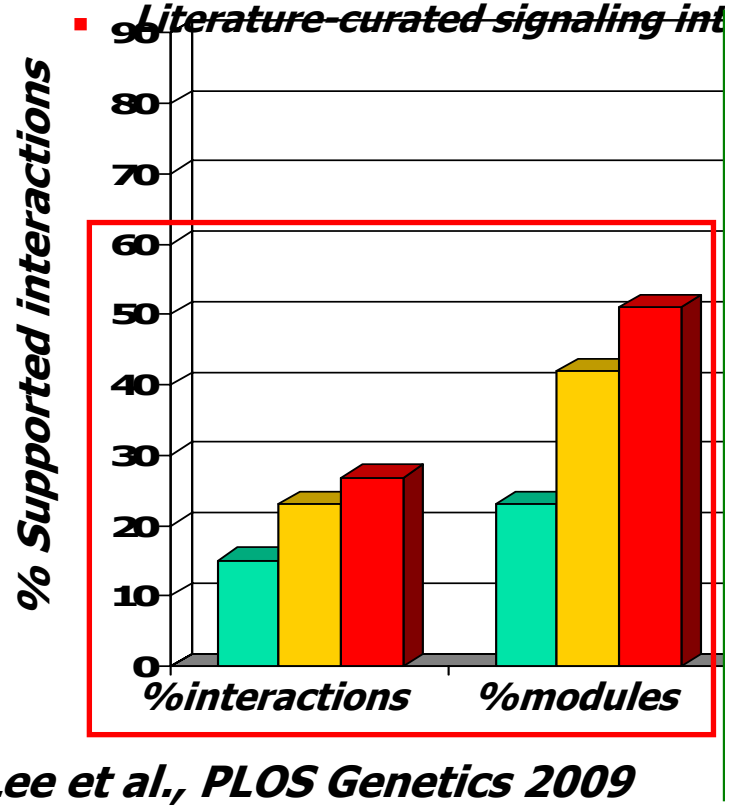
Learned regulatory weights



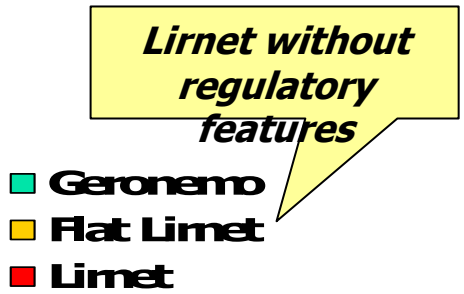
Lee et al., PLOS Genetics 2009

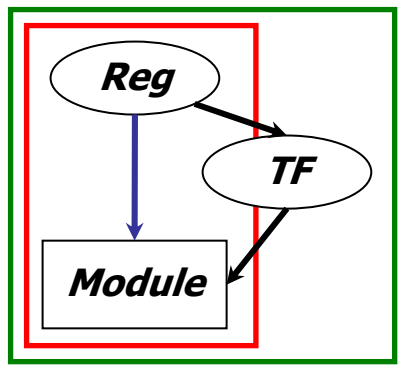
Biological evaluation I

- ***How many predicted interactions have support in other data?***
  - *Deletion/ over-expression microarrays [Hughes et al. 2000; Chua et al. 2006]*
  - *ChIP-chip binding experiments [Harbison et al. 2004]*
  - *Transcription factor binding sites [MacIsaac et al. 2006]*
  - *mRNA binding pull-down experiments [Gerber et al. 2004]*
  - *Literature-curated signaling int*

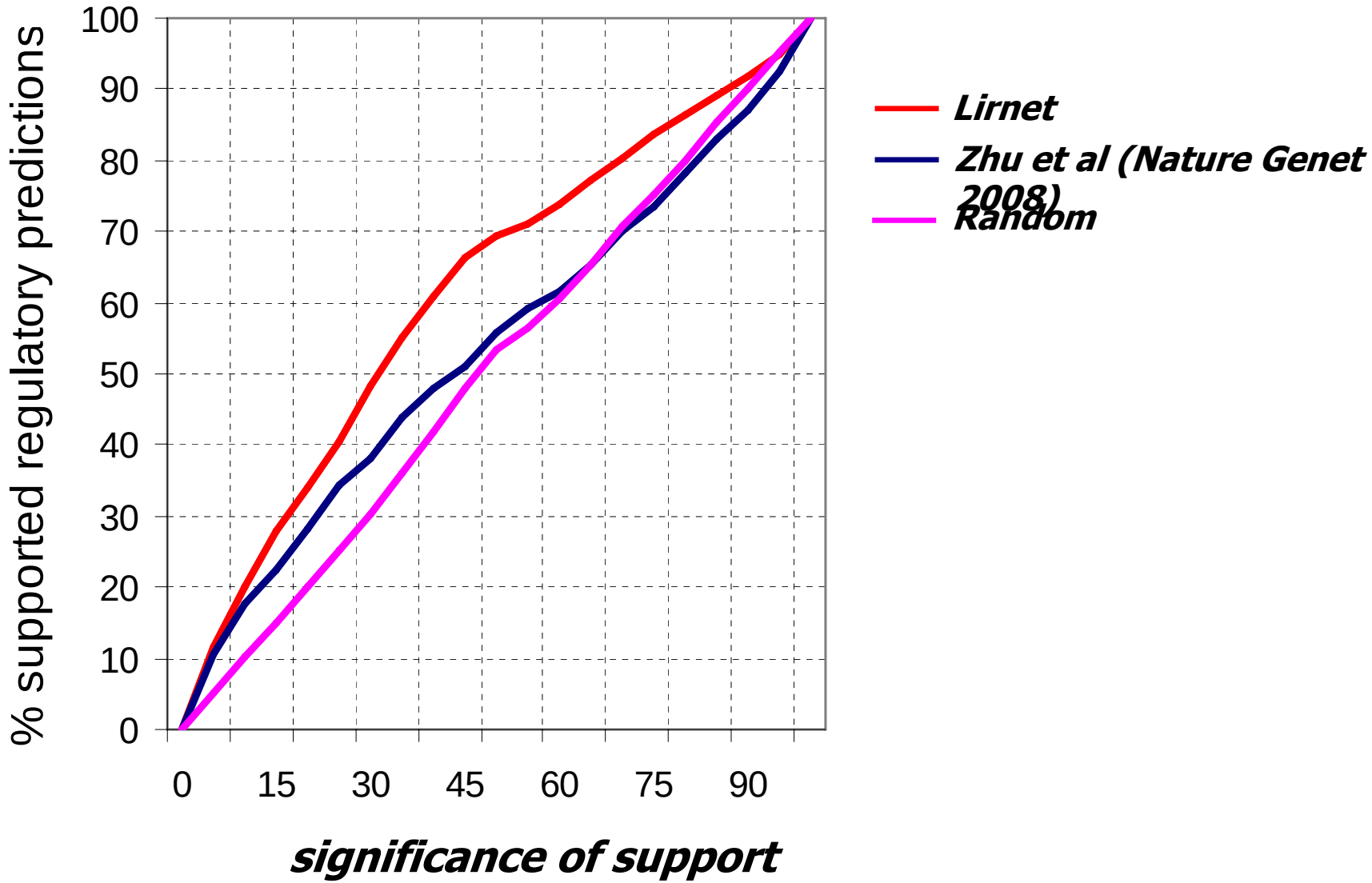


Lee et al., PLOS Genetics 2009





Biological Evaluation II



Lee et al., PLOS Genetics 2009

Pre usal Regula

8 validated regulators in 7 regions

14 validated regulators in 11 regions

- Finding caus regulators for 13 “chromosomal h spots”

Region	Zhu et al [Nat Genet 08]	Lirnet (top 3 are considered)		
1	None	SEC18	RDH54	SPT7
2	TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, **AMN1**	**AMN1**	CNS1	TOS1
3	None	TRS20	ABD1	PRP5
4	**LEU2**, **ILV6**, NFS1, CIT2, MATALPHA1	**LEU2**	PGS1	**ILV6**
5	**MATALPHA1**	**MATALPHA1**	MATALPHA2	RBK1
6	**URA3**	**URA3**	NPP2	PAC2
7	**GPA1**	**STP2**	**GPA1**	NEM1
8	**HAP1**	**HAP1**	NEJ1	GSY2
9	YRF1-4, YRF1-5, YLR464W	**SIR3**	**HMG2**	ECM7
10	None	**ARG81**	TAF13	CAC2
11	SAL1, TOP2	**MKT1**	TOP2	MSK1
12	**PHM7**	**PHM7**	ATG19	BRX1
13	None	ADE2	ORT1	**CAT5**

Lee et al., PLOS Genetics 2009

Current & Future Directions

- Understand mechanism by which individual genotype leads to changes in phenotype
  - Genotype & copy number changes (e.g., in cancer)
  - First step to personalized medicine
- Analysis and reconstruction of cellular pathways
- Understand immune response and how it is affected by aging
- ...

The Computer Science Inside

- **Computational Issues:** Huge graphical models require development of new algorithms
  - Convex optimization methods for learning network structure
  - Learning using MAP inference
  - Using combinatorial optimization within standard inference
- **Statistical issues:** Sparse data in high dimension
  - “Holistic models” to exploit correlations between different labels
  - Transfer learning between related problems
  - New algorithms for feature selection

Thank You!

http://ai.stanford.edu/~koller/

http://dags.stanford.edu/