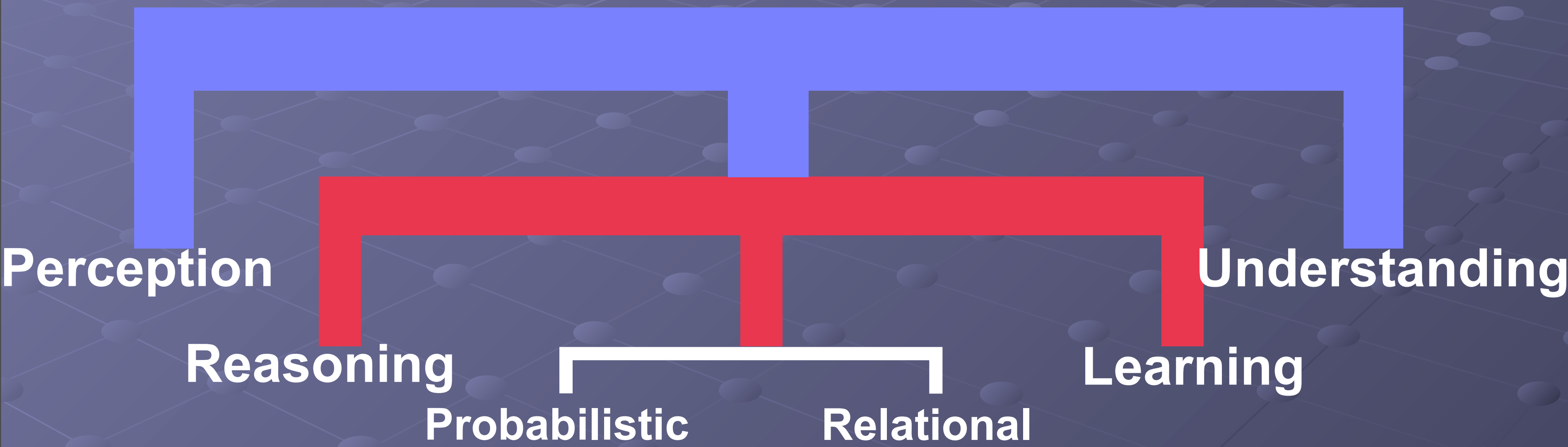
Rich Probabilistic Models for Holistic Scene Understanding

Daphne Koller
Stanford University

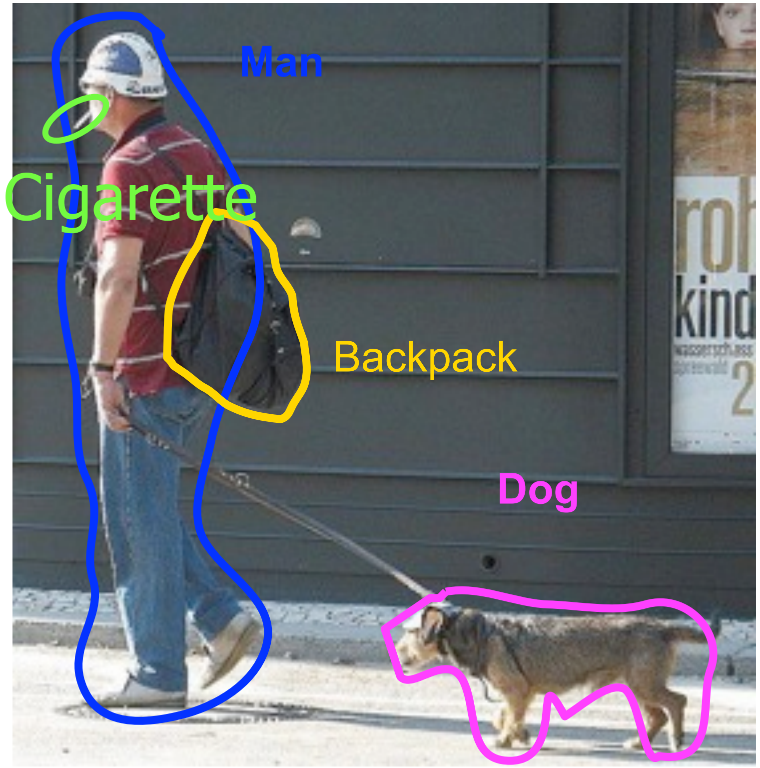
IJCAI 2011

A Tale of Three Bridges*

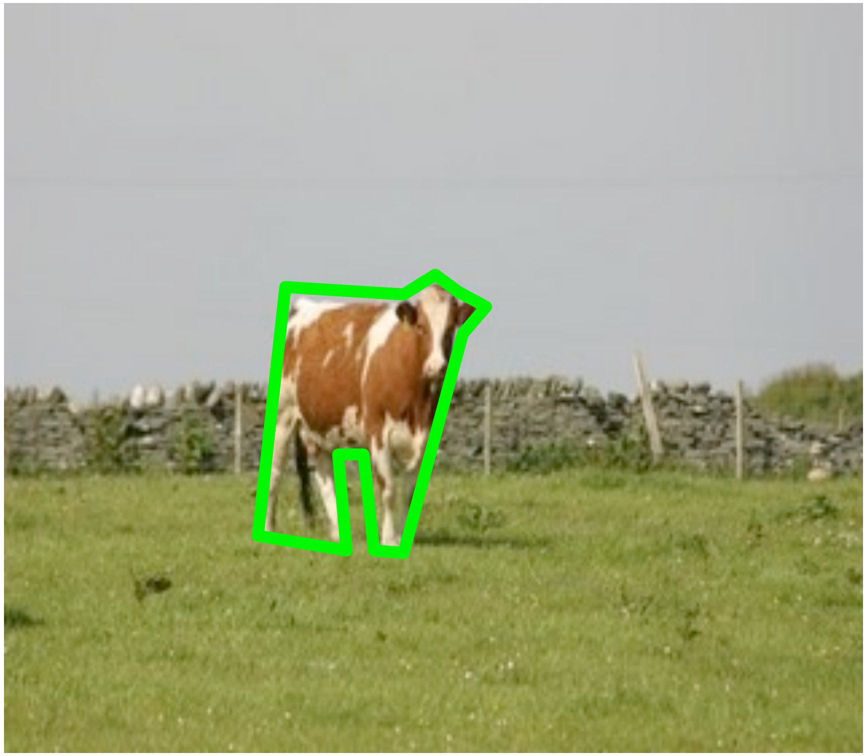


* Final slide, IJCAI 2001 Computers and Thought talk 8/7/01

From Perception to Understanding



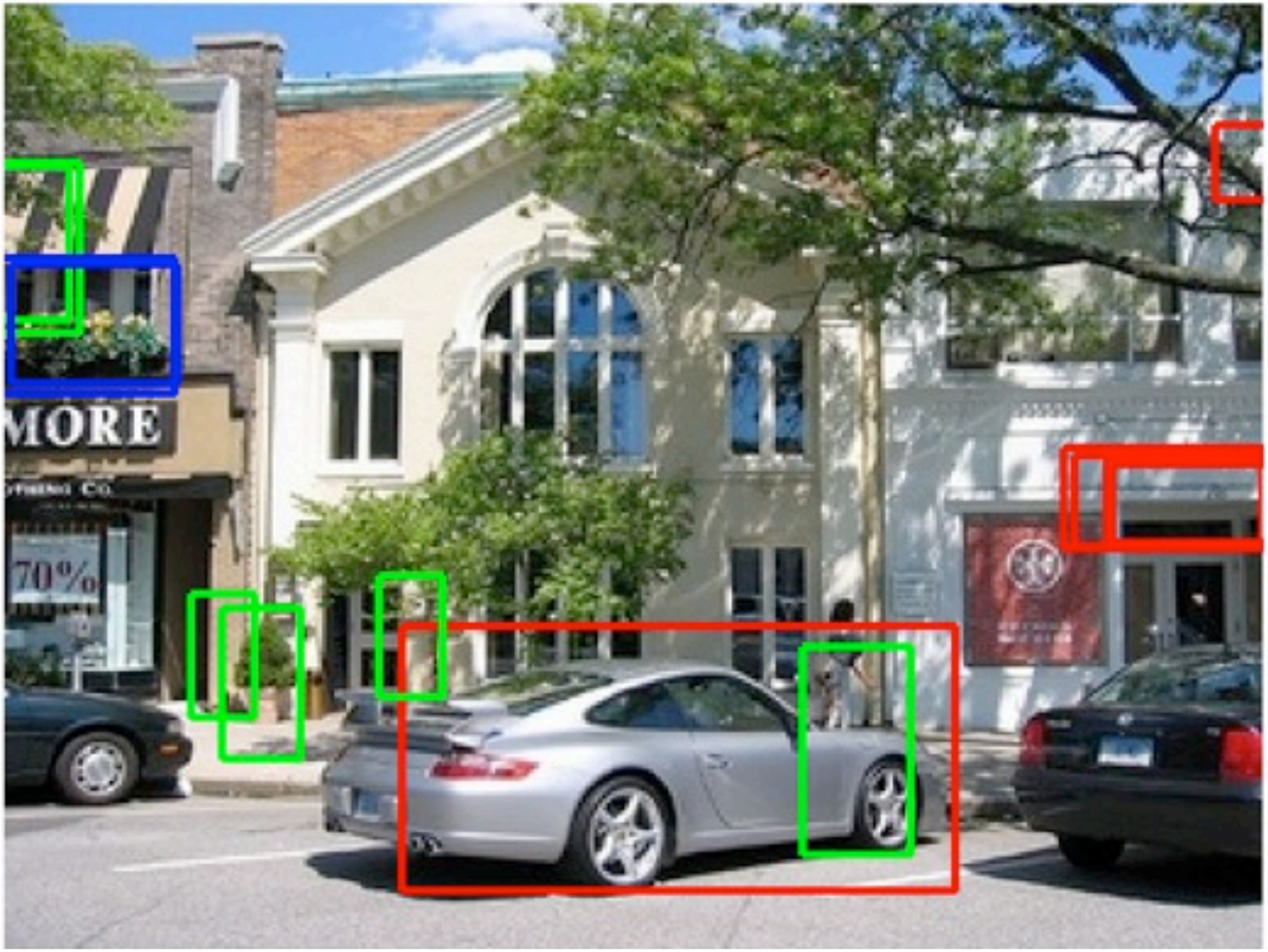
"man wearing a backpack,
smoking a cigarette,
walking a dog"



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

Object Detection

Basic Object Detection

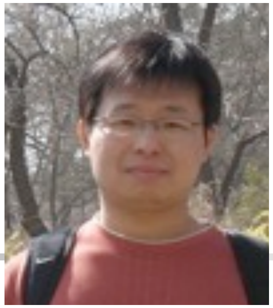


car   person   motorcycle

Outline

- Holistic scene models
  - Indoor scenes
  - Outdoor scenes
- Self-paced learning for latent variables

Outline



Huayan Wang

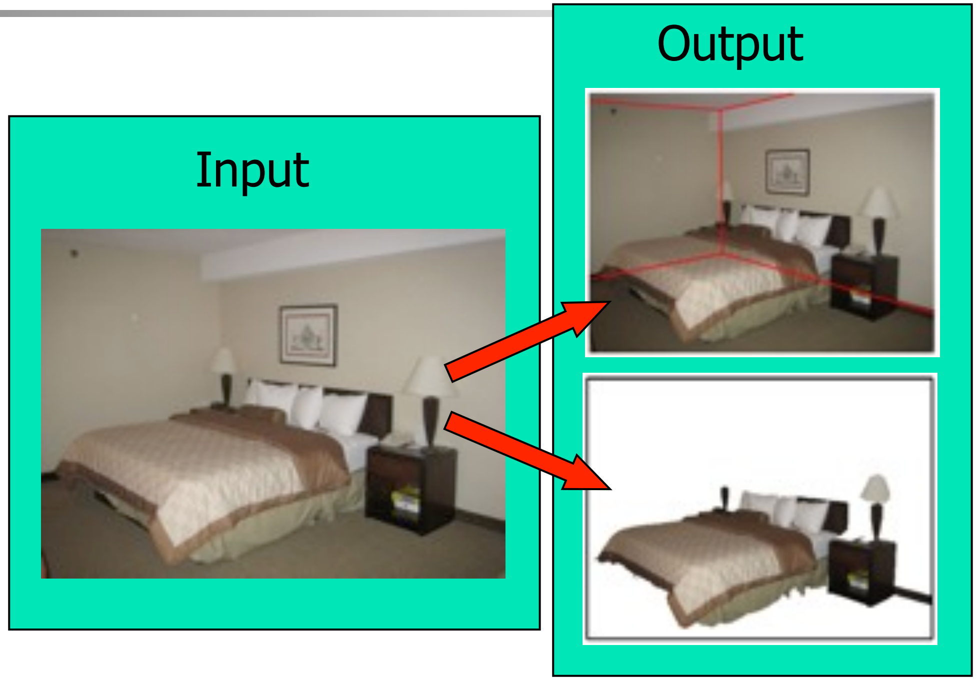
Stephen Gould

- **Holistic scene models**
  - **Indoor scenes**
  - Outdoor scenes

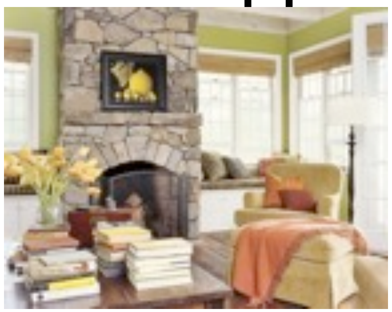
- Self-paced learning for latent variables

Indoor Scene Reconstruction

- **Goal**: Recover
  - Global geometry
  - Furniture layout



- **Challenge:** Clutters occlude boundaries and obscure the appearance of major faces



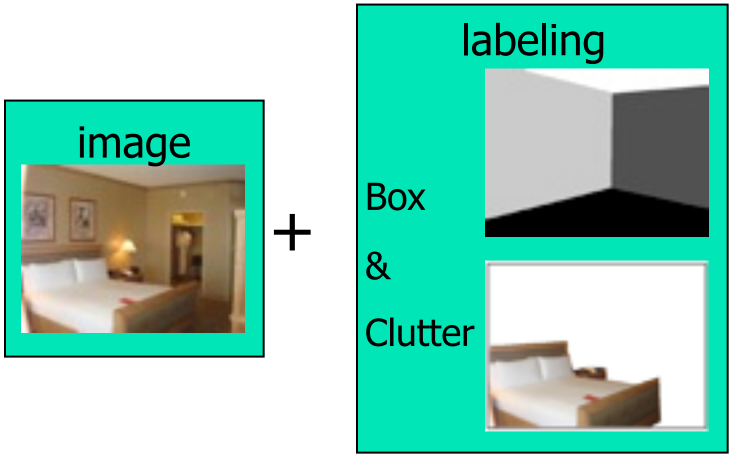
[Wang, Gould, Koller ECCV 2010]

Learning with Clutter

Supervised learning

Hedau et al ICCV 2009

- Training data:

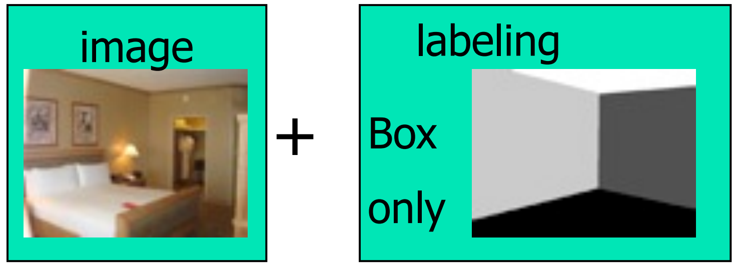


- Approach:
  - Estimate "box"
  - Supervised classification of surface labels

Latent variables

Our approach (ECCV 2010)

- Training data:



- Approach:

  Model clutter layout as latent variables

  Max-margin learning of joint model of clutter and "box"

Energy Function

learned weights

image

box parameters

$$E_{\mathbf{w}}(\mathbf{x},\mathbf{y},\mathbf{h}) = \mathbf{w}^{T}\Psi(\mathbf{x},\mathbf{y},\mathbf{h})$$

features: color, texture, perspective, boundary, …

latent variables (binary mask) specify clutter layout

$$\mathbf{y}^*,\mathbf{h}^* = \operatorname{argmax}_{\mathbf{y}\in Y,\ \mathbf{h}\in H} E_{\mathbf{w}}(\mathbf{x},\mathbf{y},\mathbf{h})$$

[Wang, Gould, Koller ECCV 2010]

Latent Variables are Tricky

Inferred box

Preferred imputation makes most of the room clutter

[Wang, Gould, Koller ECCV 2010]

Grounding Latent Variables

$$E_{\mathbf{w}}(\mathbf{x},\mathbf{y},\mathbf{h}) = \mathbf{w}^T\Psi(\mathbf{x},\mathbf{y},\mathbf{h}) + \underline{E_{\mathbf{0}}(\mathbf{x},\mathbf{y},\mathbf{h})}$$

$$\underline{\alpha_1} E_1(\mathbf{x},\mathbf{y},\mathbf{h}) + \underline{\alpha_2} E_2(\mathbf{y},\mathbf{h})$$

Weights
and fixed

$$\mathbf{y}^*,\mathbf{h}^* = \mathrm{argmax}_{\mathbf{y}\in Y,\, \mathbf{h}\in H}\, E_{\mathbf{w}}(\mathbf{x},\mathbf{y},\mathbf{h})$$

Informed prior about latent variables imposed on the learning process

[Wang, Gould, Koller ECCV 2010]

Effect of Informed Prior

Learning with prior terms

Inferred box | Inferred clutter

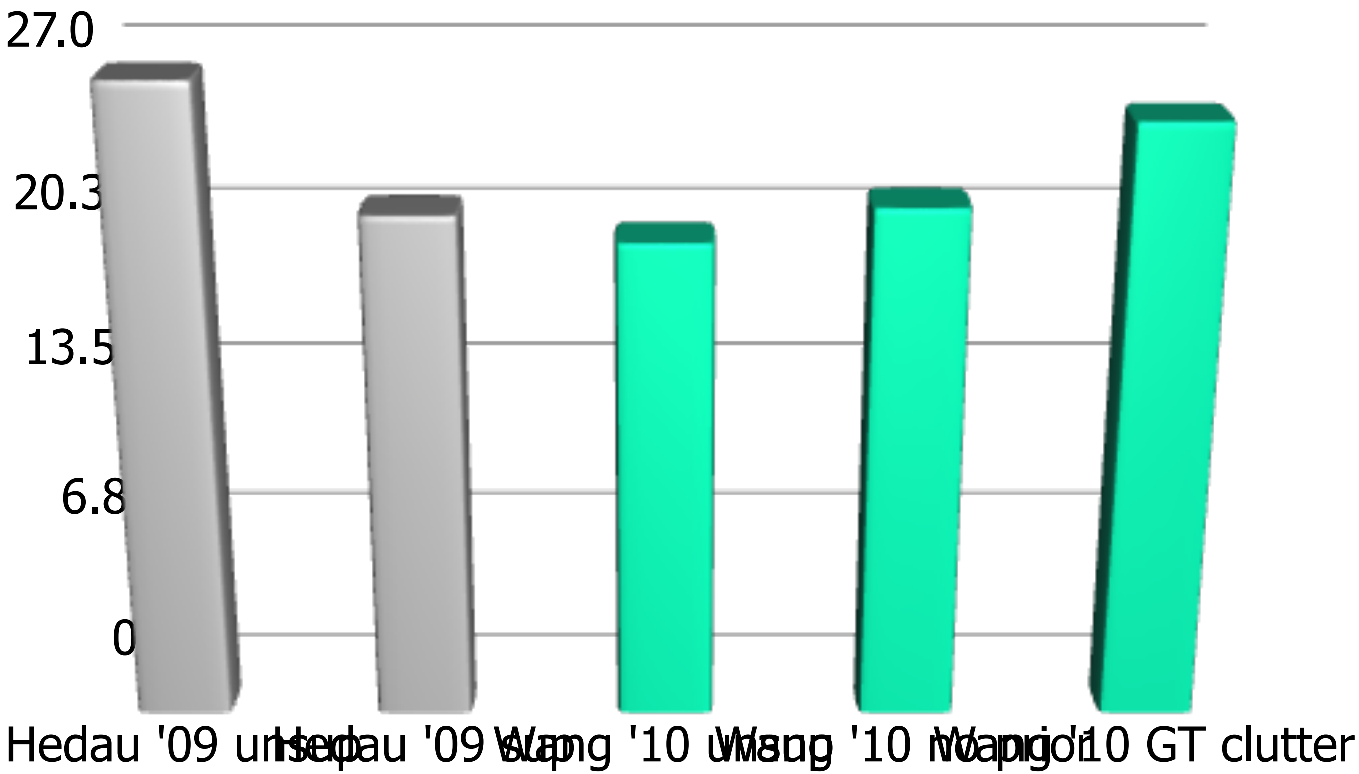
Learning w/o prior terms

Inferred box | Inferred clutter

[Wang, Gould, Koller ECCV 2010]

Experimental Results

Pixel-wise classification error

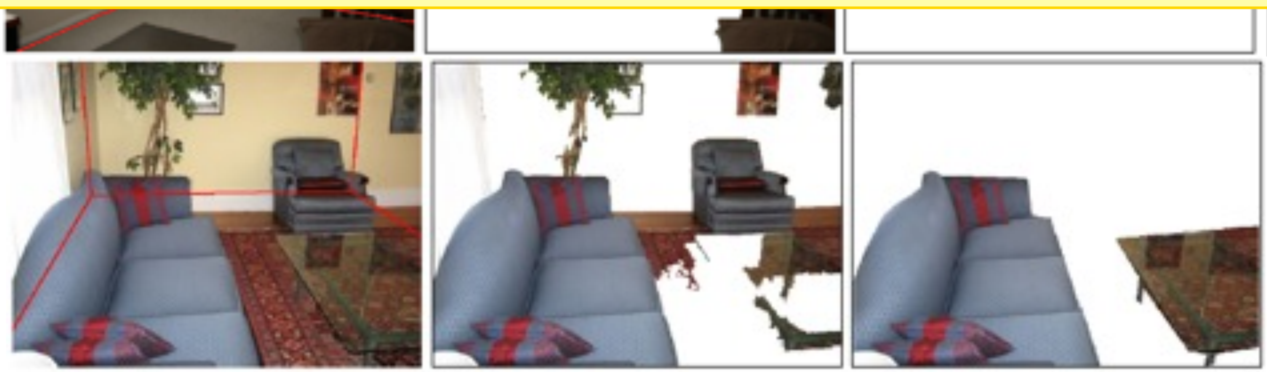


Comparison to labeled clutter

[Wang, Gould, Koller ECCV 2010]

Comparison to labeled clutter

Human labeling for latent variables can be suboptimal



Outline

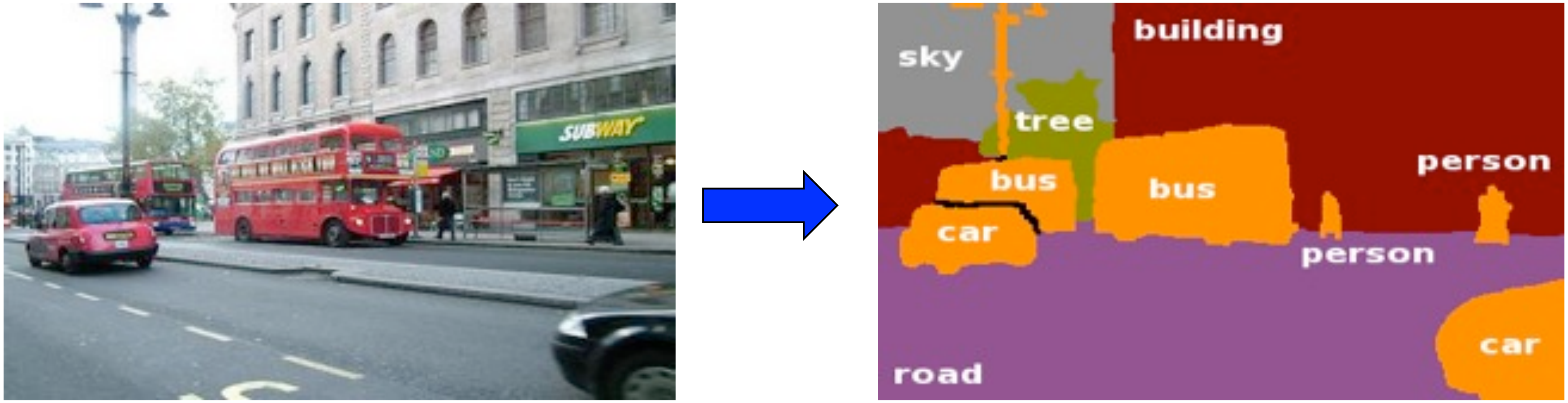
Stephen Gould

Tianshi Gao

Pawan Kumar

- **Holistic scene models**
  - Indoor scenes
  - **Outdoor scenes**
- Self-paced learning for latent variables

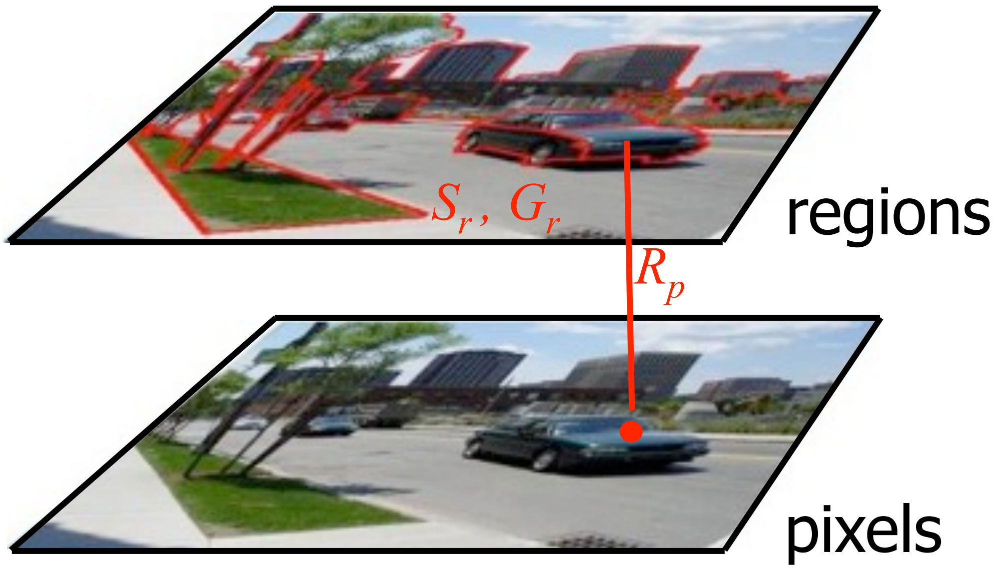
Scene Segmentation



x **y**

$$\mathbf{y}^* = \operatorname{argmin}_{\mathbf{y}} E(\mathbf{x},\mathbf{y};\,\mathbf{w})$$

Region-Based Model



Variables
R_p: pixel-to-region correspondence
A_r: region appearance
S_r: region semantic class
G_r: region geometry
v^{hz}: location of horizon

Model assigns each pixel to a region while respecting global coherence

[Gould, Fulton, Koller ICCV 09; Gould, Gao, Koller NIPS 09]

[Gould, Fulton, Koller ICCV 09]

Region-Based Model

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

$$=$$

$$\psi^{\text{horizon}}(v^{\text{hz}}) \qquad \psi^{\text{region}}(S_r, G_r, v^{\text{hz}}) \qquad \psi^{\text{boundary}}(A_r, A_s)$$

Horizon Term
e.g., vanishing lines

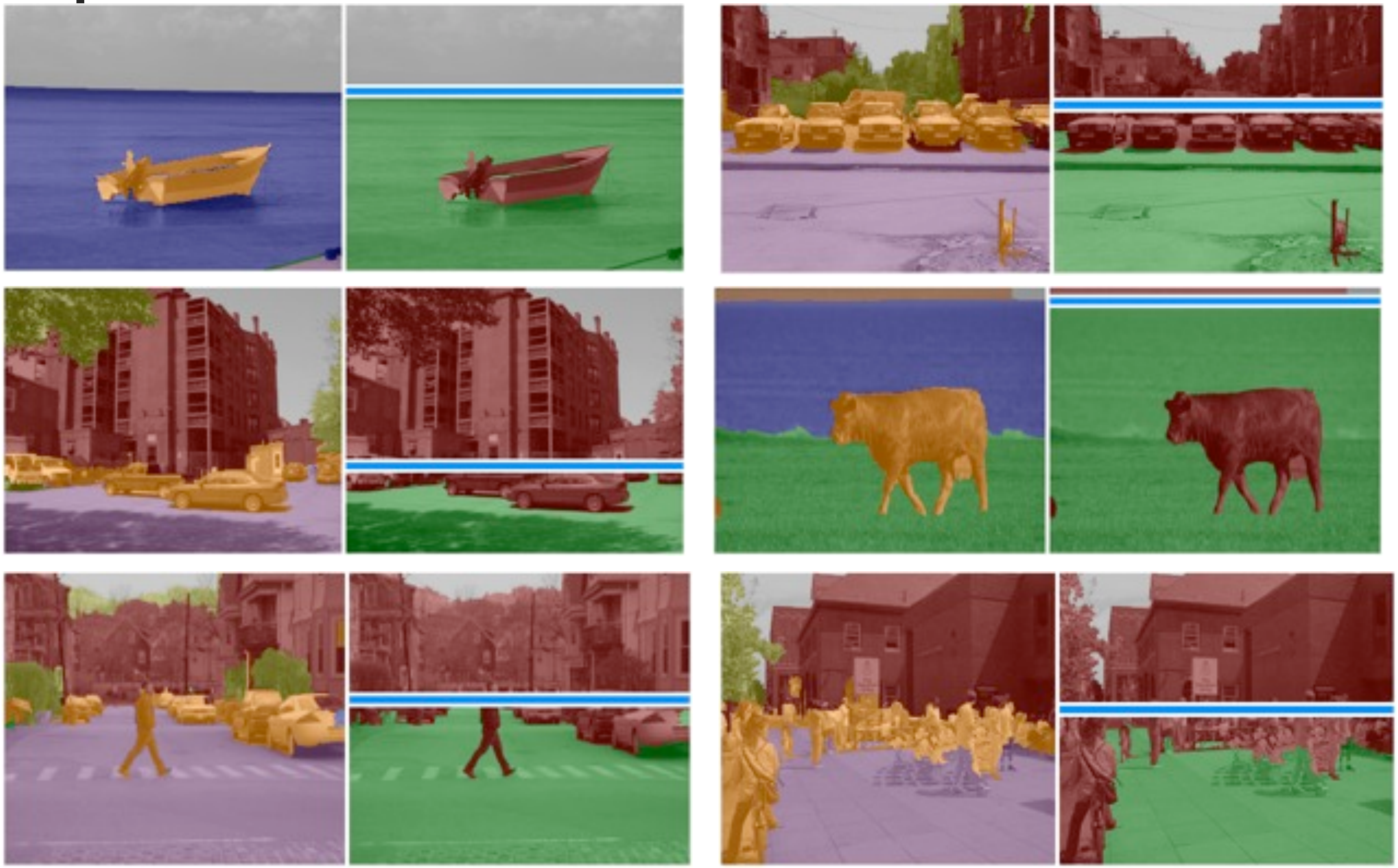
$+$

Region Term
e.g., consistent appearance and location

$+$

Boundary Term
e.g., difference in color/texture between regions

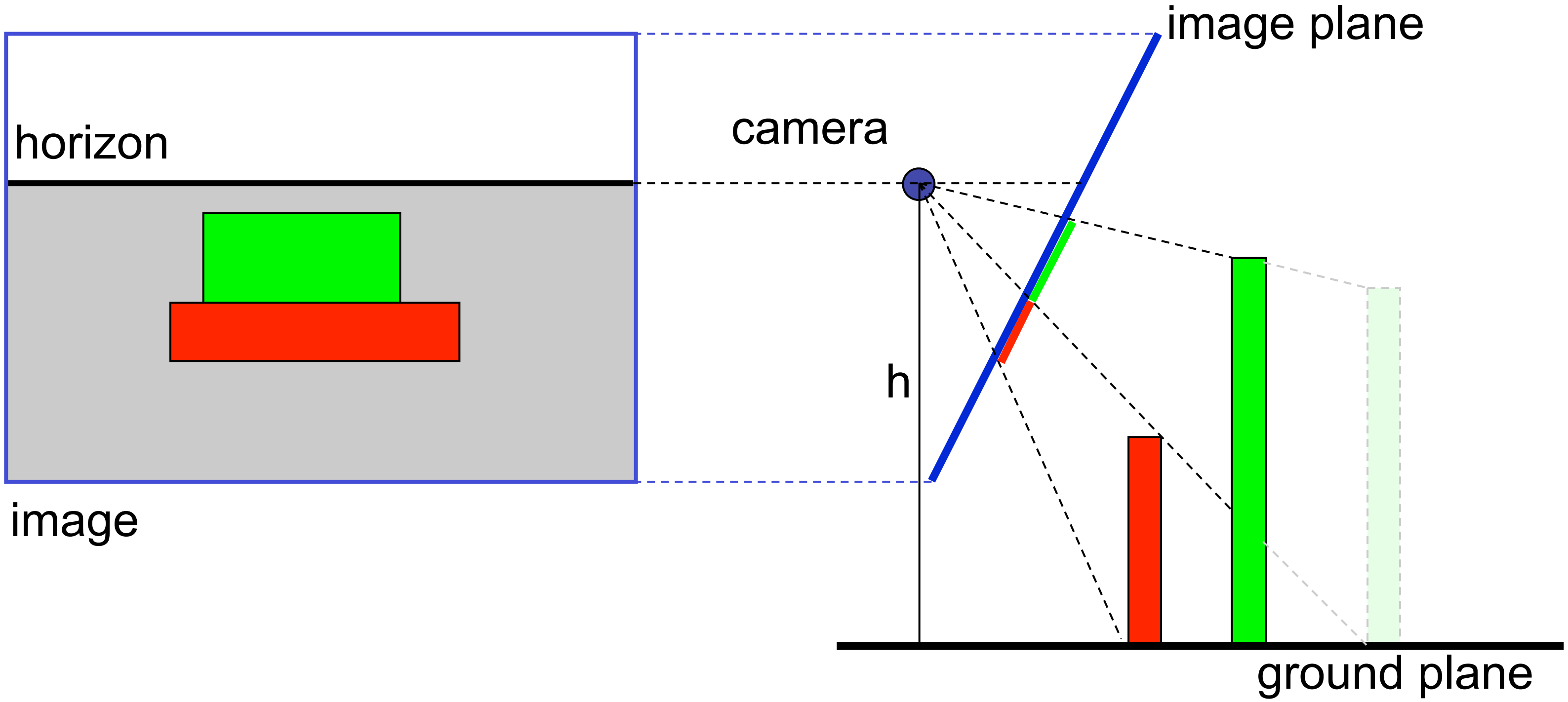
Example Results



[Gould, Fulton, Koller, ICCV 2009]

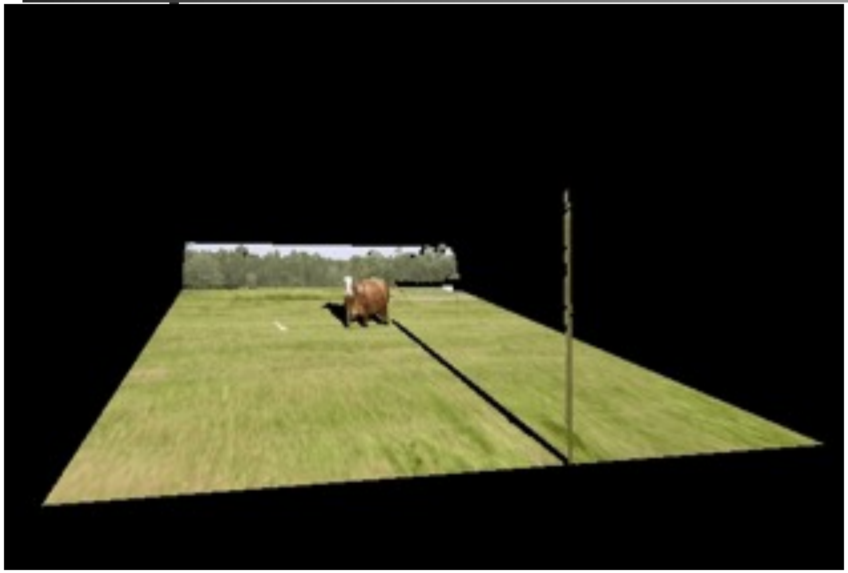
Application: 3d Reconstruction

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



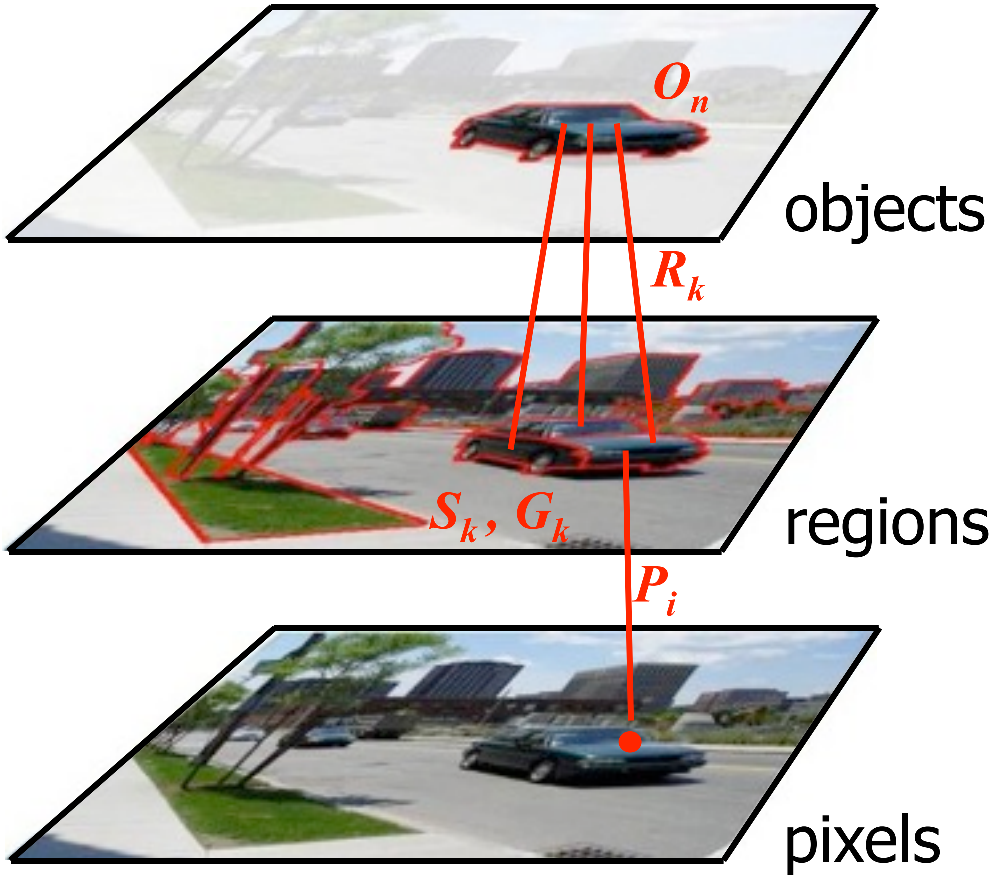
[Gould, Fulton, Koller, ICCV 2009]

Example 3D Reconstructions



[Gould, Gao, Koller NIPS 09]

Object Detection



$\psi^{object}(O_n, v^{hz})$ + $\psi^{context}(O_n, S_k)$

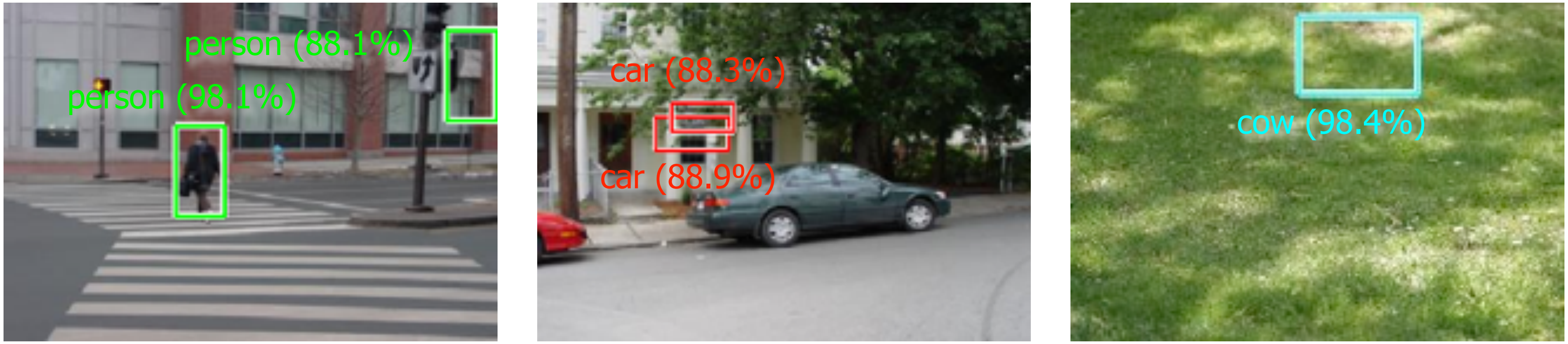
Object Model
e.g. wheel-like appearance in bottom corner

Context Term
e.g., cars on road

[Gould, Gao, Koller NIPS 09]

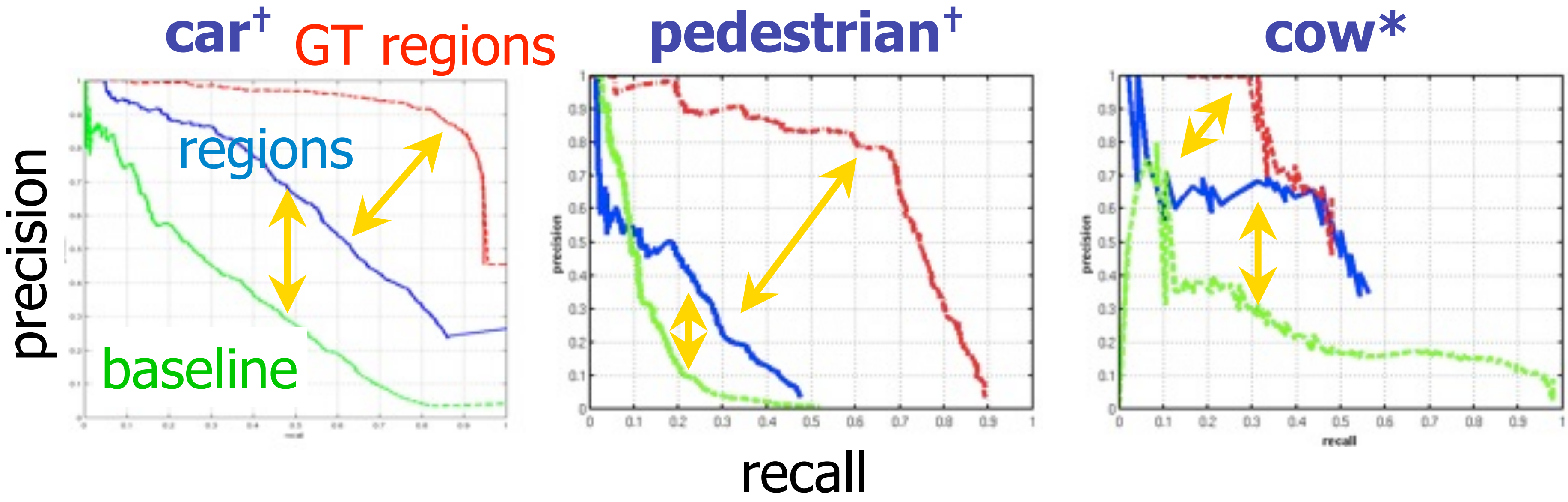
Examples

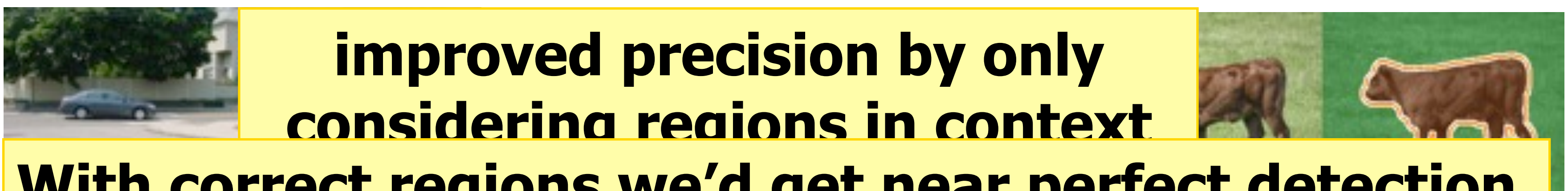
Typical sliding-window detector results (top two detections per image)



Our region-based approach (MAP assignment)

Detection Performance





improved precision by only considering regions in context

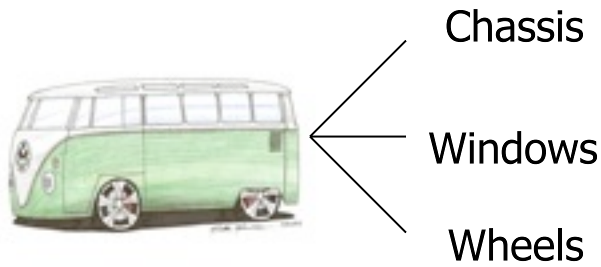
With correct regions we'd get near perfect detection, but region model still has a way to go

†

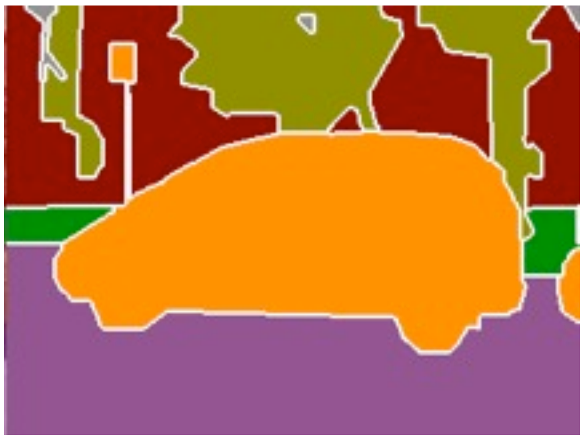
*** run on subset of 21-class MSRC dataset**

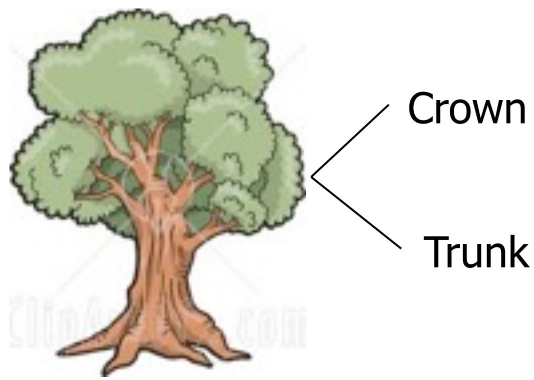
[Gould, Gao, Koller NIPS 09]

Latent Variables Revisited



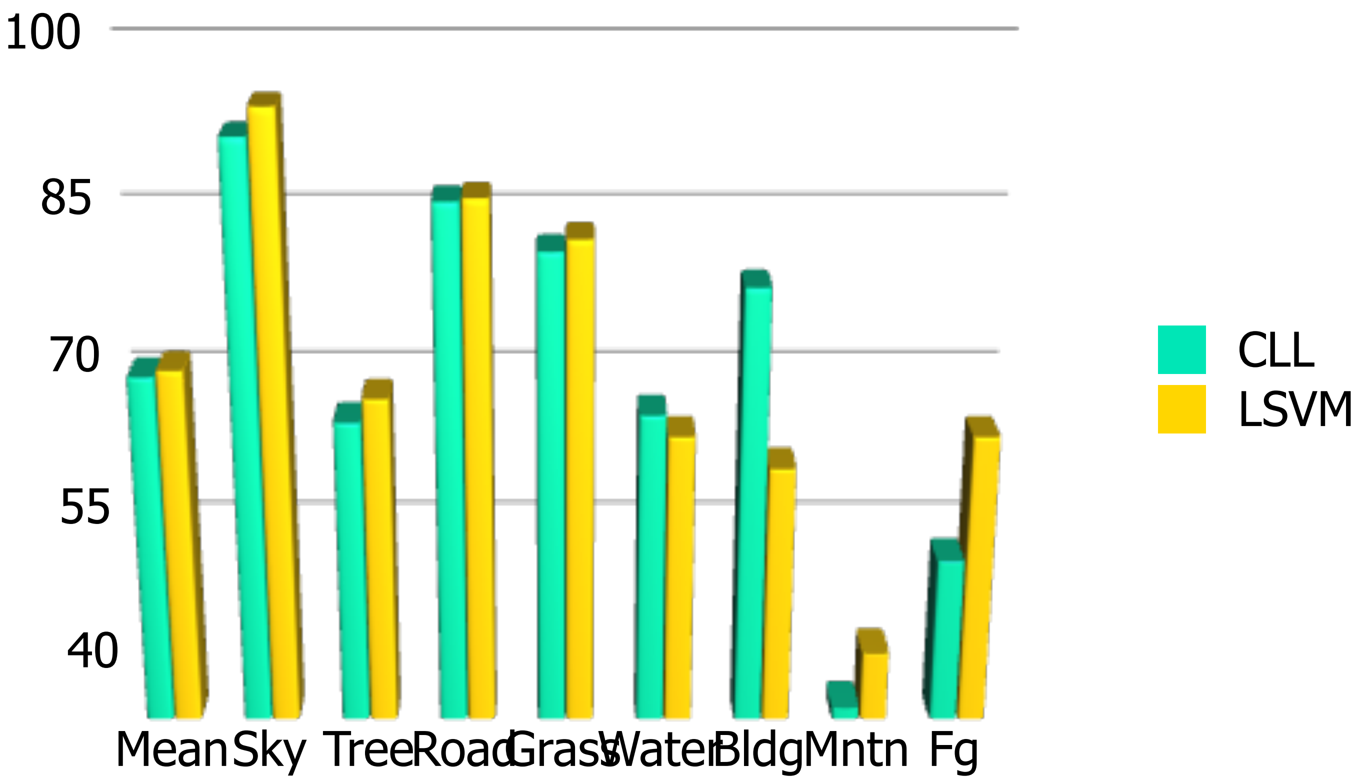
Human specified regions are not the most discriminative





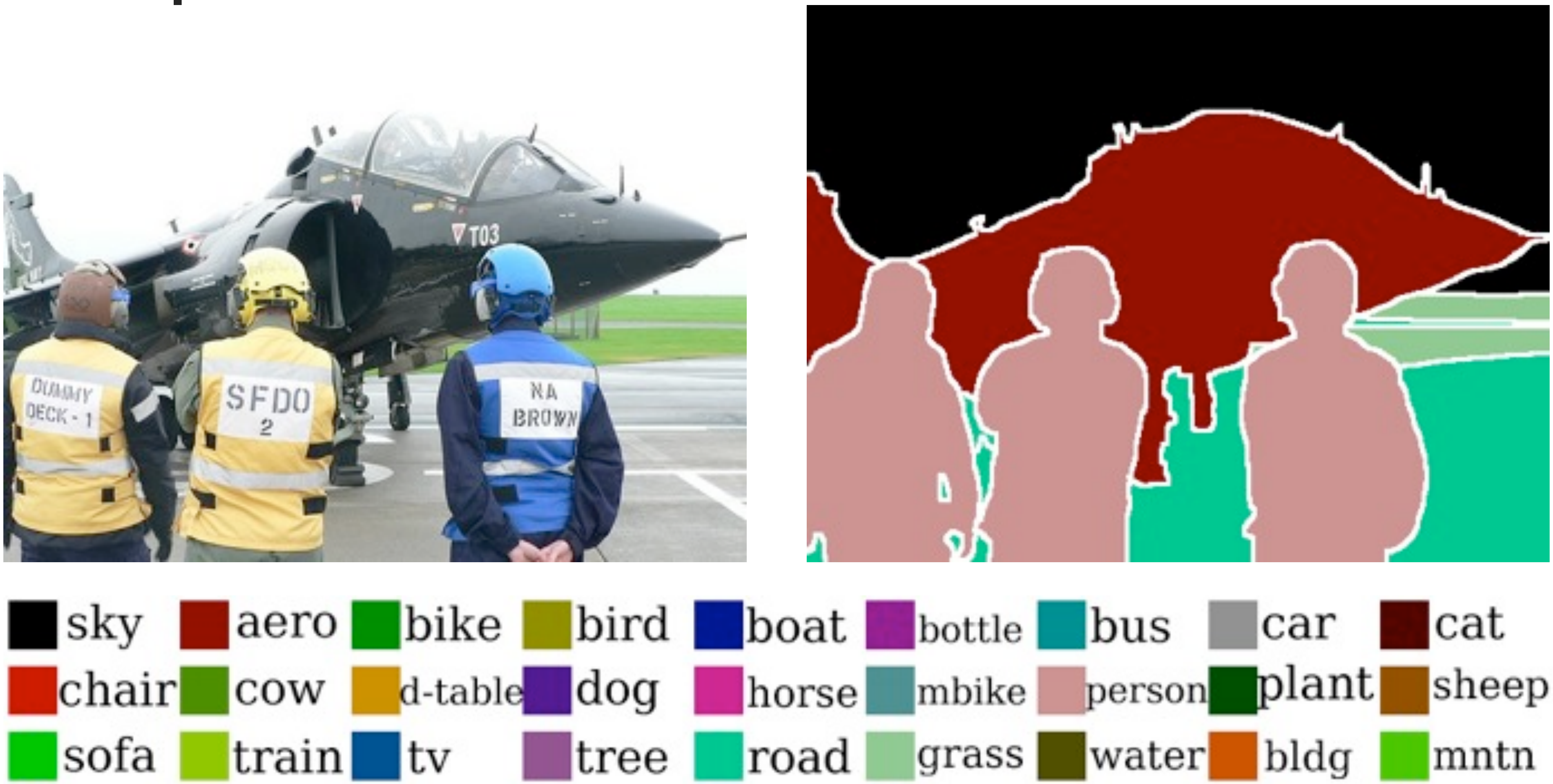
Latent Variables Revisited

Learn with latent variables encoding pixel-to-region assignments



[Kumar, Turki, Preston, Koller ICCV11]

Real Multi-Class Segmentation



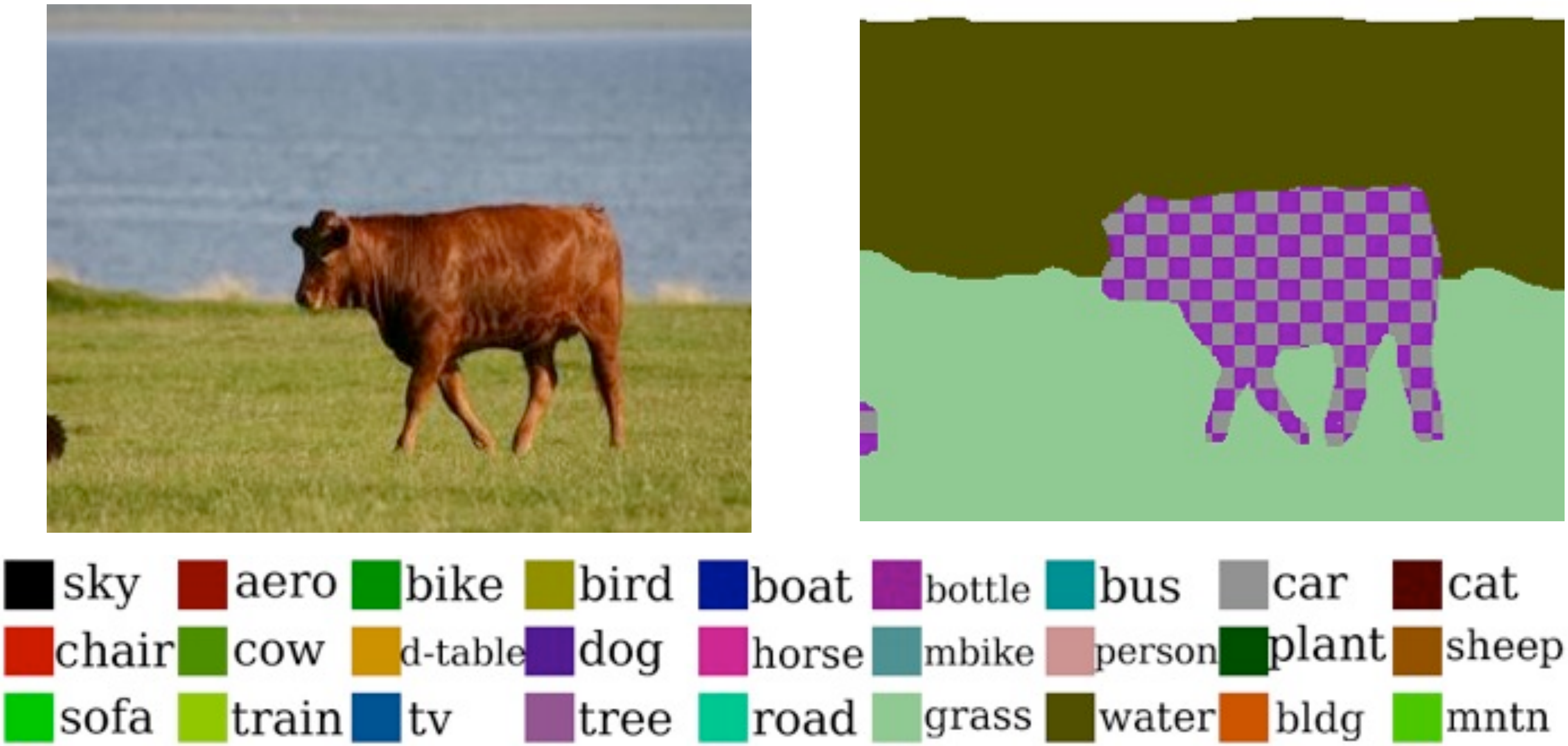
"Fully" Supervised Data

Specific foreground classes, generic background class

PASCAL VOC Segmentation Datasets

“Fully” Supervised Data

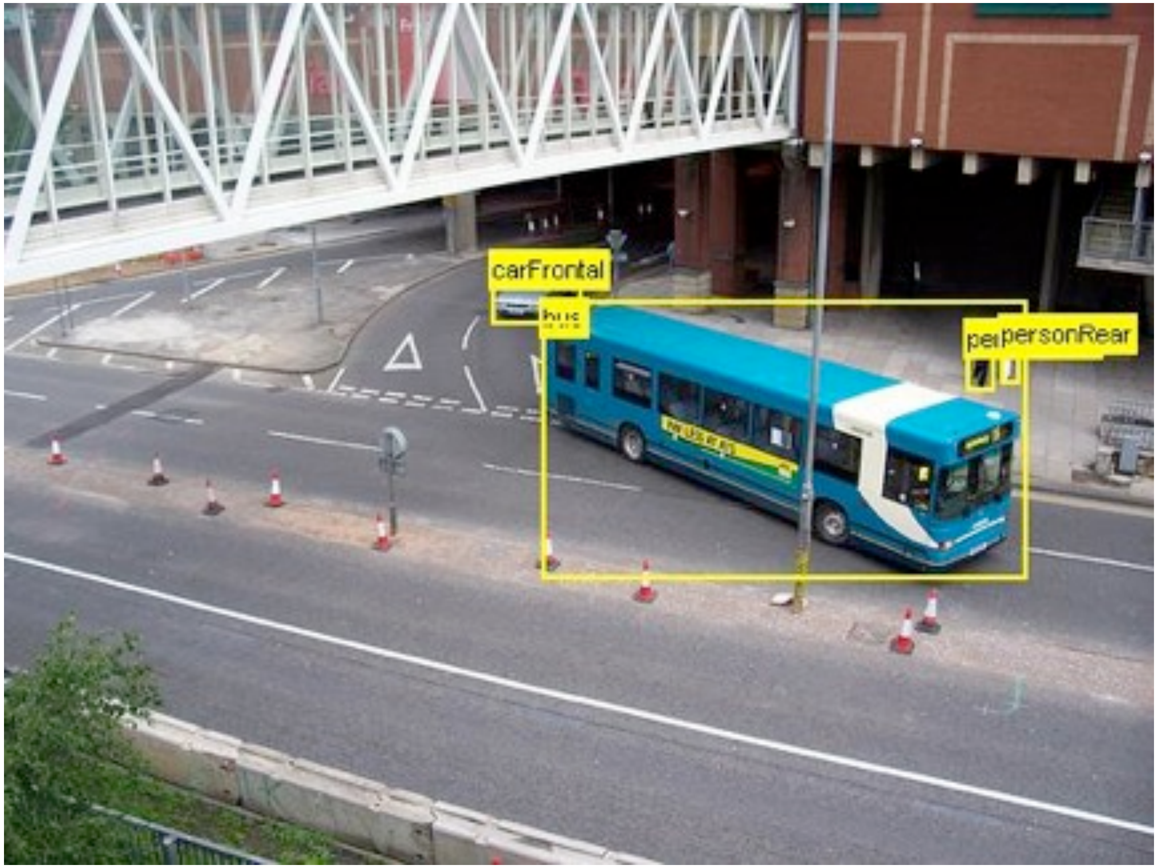
Specific background classes, generic foreground class



Stanford Background Datasets

Weakly Supervised Data

Bounding Boxes for Objects



PASCAL VOC Detection Datasets

Thousands of images

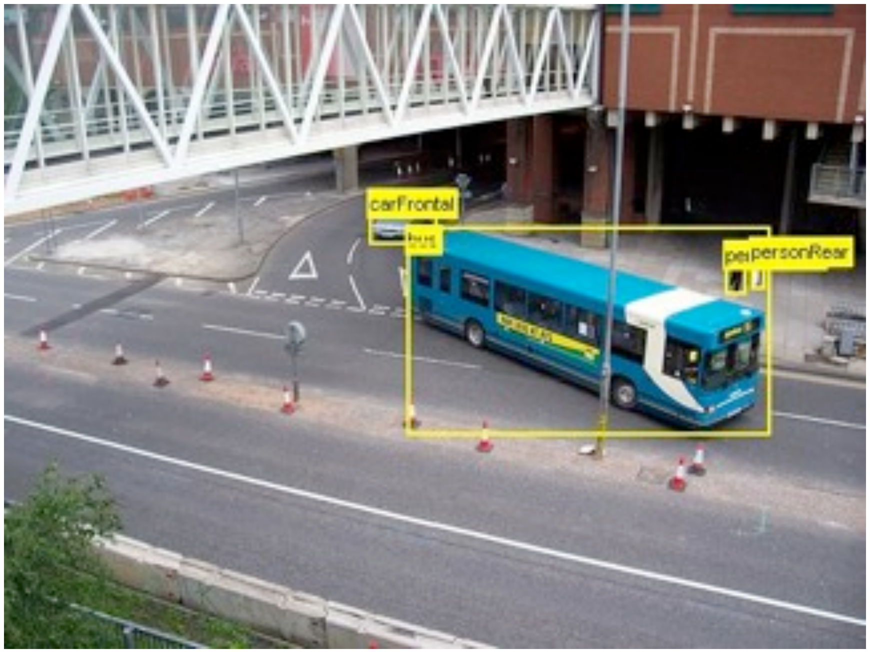
Weakly Supervised Data

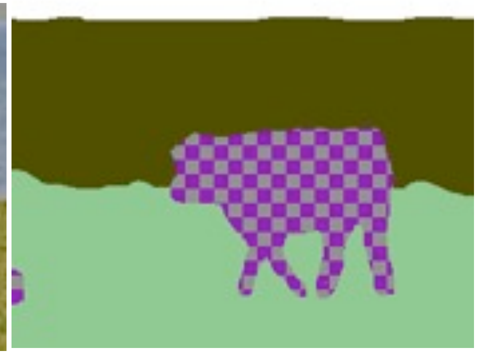
Image-Level Labels

ImageNet, Caltech…

Thousands of images

Diverse Data

Latent Variable Formulation:

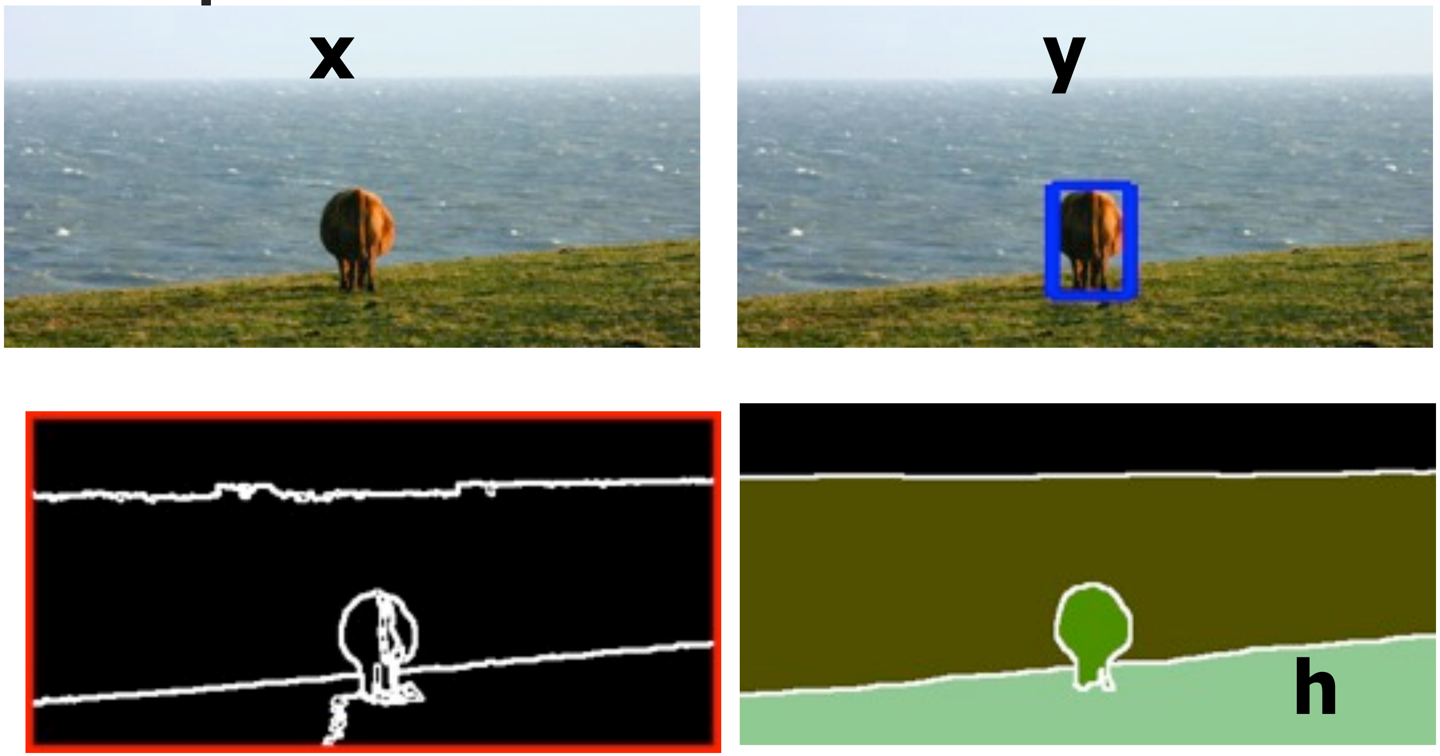
x

y

h

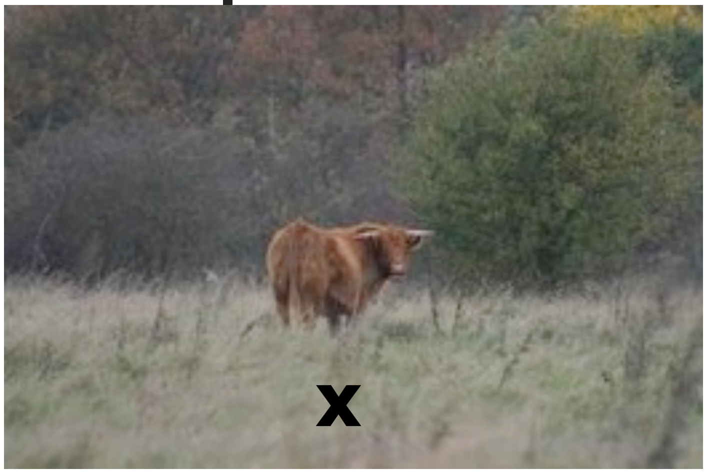
Specific classes must agree with generic classes

Latent Variable Formulation:



Every row & column in bounding box must
contain pixel labeled with bounding-box class

Latent Variable Formulation:



y = "Cow"

Image must contain region labeled with image class

Learning with Diverse Data

Comparison to previous results

New

Sta

Back

d

Dataset

53.1%

Using weakly labeled data provides only marginal improvement

Imputing latent variables is hard and can introduce significant noise

Pas

VOC

Old

24.7%

Classes

Outline

Pawan Kumar

Ben Packer

- Holistic scene models
- **Self-paced learning for latent variables**
  - **Instance selection**
  - Model selection

Max-Margin Training

Multi-class SVM (Crammer & Singer, 2001)

learned weights

Feature vector

$$y^* = \text{argmax}_{y \in Y}\, \mathbf{w}^T \Psi(\mathbf{x}, y)$$

$$\text{Minimize}_{\mathbf{w},\xi}\ \|\mathbf{w}\|^2 + C \textstyle\sum_i \xi_i$$

$$\mathbf{w}^T \Psi(\mathbf{x}_i, y_i) - \mathbf{w}^T \Psi(\mathbf{x}_i, y) \geq 1 - \xi_i \qquad \forall\, y$$

Margin

Slack

Maximize margin between ground truth and all other labels

Structured Max-Margin Training

Taskar, Guestrin, Koller, 2003; Tsochantaridis, Hofmann, Joachims, Altun, 2004

structured output

$$\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y} \in Y} \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$$

$$\text{Minimize}_{\mathbf{w},\xi} \ \|\mathbf{w}\|^2 + C \sum_i \xi_i$$

$$\mathbf{w}^T \Psi(\mathbf{x}_i, \mathbf{y}_i) - \mathbf{w}^T \Psi(\mathbf{x}_i, \mathbf{y})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}) - \xi_i \qquad \forall\, \mathbf{y}$$

Loss-dependent Margin

Exponentially many constraints

Max-Margin Structured Prediction

- Tractable models admit polynomial size formulation [Taskar, Guestrin, Koller, 2003]

- Cutting plane approach [Tsochantaridis et al., 2004]
  - Often requires only MAP inference
    - admits tractable algorithms that avoid computing the partition function
  - For many models, only polynomial # of cutting planes required for "close to optimal" learning

Latent SVM

Felzenswalb, McAllester, Ramanan 2008; Yu, Joachims 2009

$$\mathbf{y}^*, \mathbf{h}^* = \operatorname{argmax}_{\mathbf{y}\in Y,\, \mathbf{h}\in H} E_{\mathbf{w}}(\mathbf{x},\mathbf{y},\mathbf{h})$$

$$\text{Minimize}_{\mathbf{w},\xi}\ ||\mathbf{w}||^2 + C\sum_i \xi_i$$

$$\max_{\mathbf{h}_i} \mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h}_i) - \mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{y},\mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall\ \mathbf{y}, \mathbf{h}$$

Best imputation of **h** consistent with ground truth label is better than any imputation and any other label

CCCP

Felzenszwalb et al., NIPS 2007, Yu et al., ICML 2008

Start with an initial estimate $\mathbf{w}_0$

Impute $\mathbf{h}_i = \mathrm{argmax}_{\mathbf{h}}\ \mathbf{w}_t^T\Psi(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h})$ MAP Inference

Update $\mathbf{w}_{t+1}$ by solving a convex problem

How well can we impute $\mathbf{h}_i$?

$$\mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h}_i) - \mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{y},\mathbf{h}) \leq \Delta(\mathbf{y}_i,\mathbf{y},\mathbf{h}) - \xi_i$$

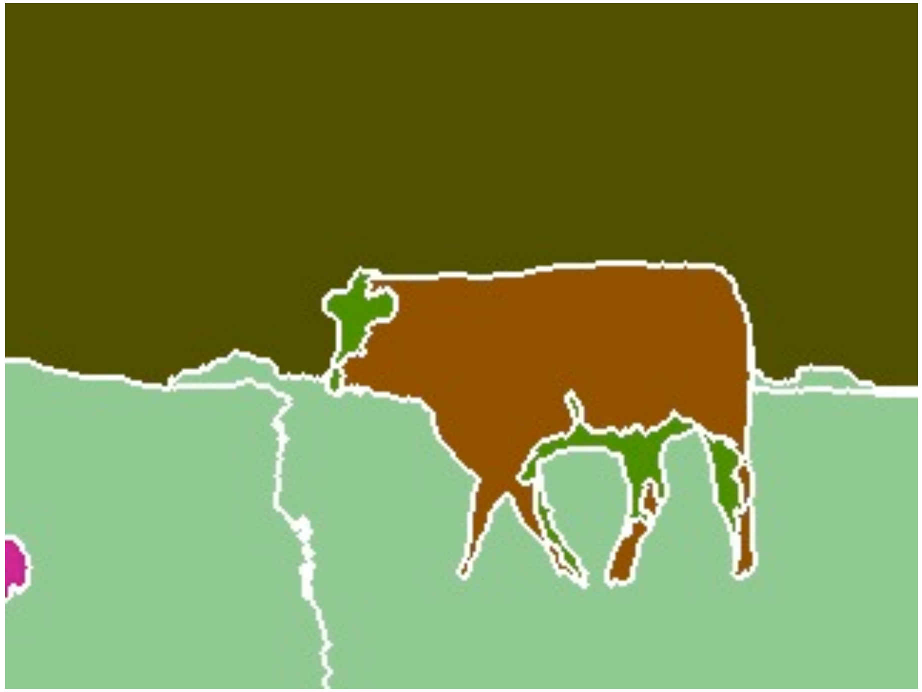
EASY

White sky Green grass Grey road

EASY

White sky Green grass Blue water

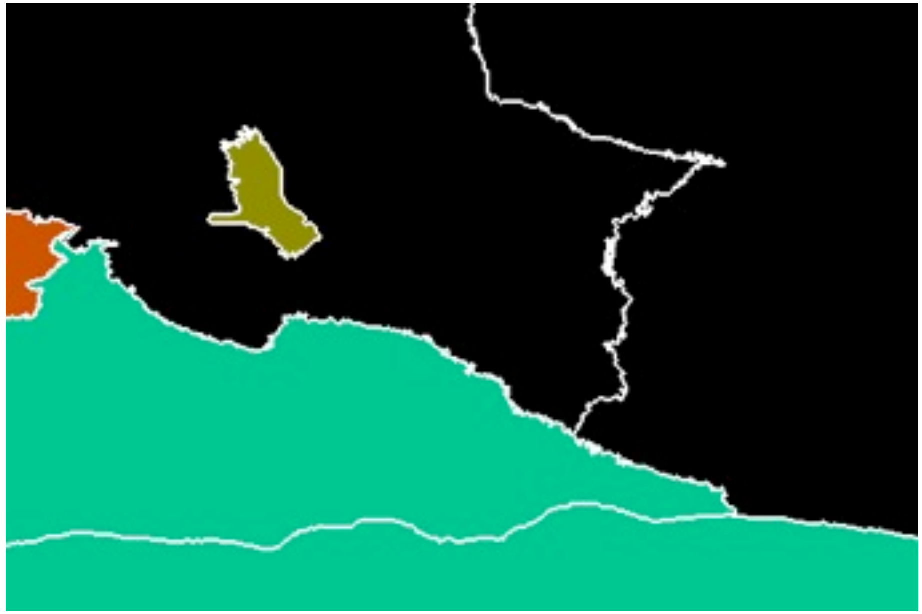
HARD



Cow? Horse? Cat?

HARD

Red Sky?



Black Mountain?

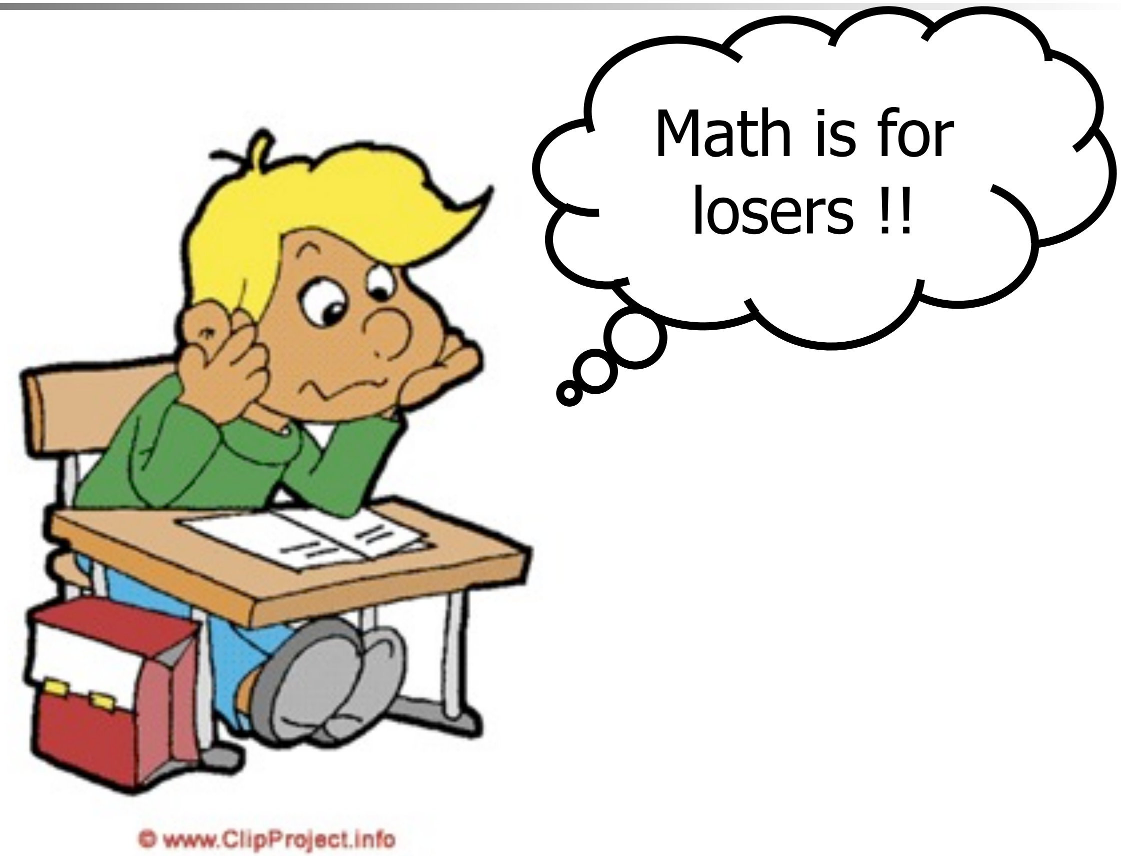
[Kumar, Packer, Koller NIPS 2010]

Inspiration: Human Learning

Real Numbers

Imaginary Numbers

$e^{i\pi}+1 = 0$



FAILURE … BAD LOCAL MINIMUM

[Kumar, Packer, Koller NIPS 2010]

Inspiration: Human Learning

Real Numbers

Imaginary Numbers

$e^{i\pi}+1 = 0$

SUCCESS … GOOD LOCAL MINIMUM

Curriculum Learning: Bengio et al, ICML 2009

Bengio et al, ICML 2009

Curriculum Learning

Start with easy examples, then consider hard ones

Easy vs. **hard???**

Easy for human

≠ Easy for machine

Self-Paced Learning

Easiness is a property of data sets and classifiers, not of isolated instances

Computer should figure out for itself which instances are hard for it right now

Kumar, Packer and Koller, NIPS 2010

Self-Paced Learning

$v_i \in \{0,1\}$

Start with an initial estimate $\mathbf{w}_0$

Update $\boxed{\mathbf{h}_i = \min_{\mathbf{h}} \mathbf{w}_t^T\Psi(\mathbf{x}_i,\mathbf{a}_i,\mathbf{h})}$

v_i $-\sum_i v_i/K$

Update $\mathbf{w}_{t+1}$ by solving a convex problem

$$\boxed{\begin{array}{c}\min \sum_i \xi_i + \lambda\|\mathbf{w}\|^2 \\ \mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{a}_i,\mathbf{h}_i) - \mathbf{w}^T\Psi(\mathbf{x}_i,\mathbf{a},\mathbf{h}) \le \Delta(\mathbf{a}_i,\mathbf{a},\mathbf{h}) - \xi_i\end{array}}$$

$v_i = 1$ for easy examples $v_i = 0$ for hard examples

Biconvex Optimization

Alternate Convex Sea

[Kumar, Packer, Koller NIPS 2010]

Self-Paced Learning

Start with an initial estimate $\mathbf{w}_0$

Update $\boxed{\mathbf{h}_i = \min_{\mathbf{h}} \mathbf{w}_t^\top \Psi(\mathbf{x}_i,\mathbf{a}_i},\mathbf{h})$

As simple as CCCP!!

Update $\mathbf{w}_{t+1}$ by solving a biconvex problem

$$\min \sum_i \xi_i v_i + \lambda||\mathbf{w}||^2 - \sum_i v_i/K$$

$$\mathbf{w}^\top \Psi(\mathbf{x}_i,\mathbf{a}_i,\mathbf{h}_i) - \mathbf{w}^\top \Psi(\mathbf{x}_i,\mathbf{a},\mathbf{h})$$

$$\leq \Delta(\mathbf{a}_i,\mathbf{a},\mathbf{h}) - \xi_i$$

Decrease $K \leftarrow K/\mu$

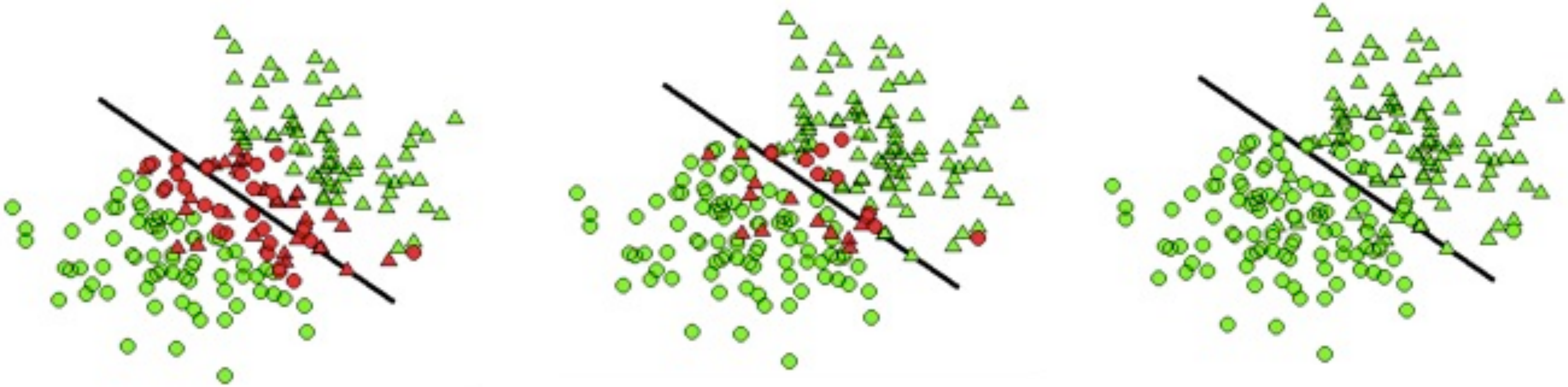
[Kumar, Packer, Koller NIPS 2010]

Self-Paced Learning

$$\min \sum_i \xi_i v_i + \lambda \|w\| - \sum_i v_i / K$$

v_i = 1 (use)

v_i = 0 (don't use)



Large K Medium K Small K

[Kumar, Packer, Koller NIPS 2010]

Simple Example: Object Detection

Input **x** - Image

Output **y** $\in$ Y

Latent **h** - Box

Δ - 0/1 Loss

Y = {"Bison", "Deer", "Elephant", "Giraffe", "Llama", "Rhino"}

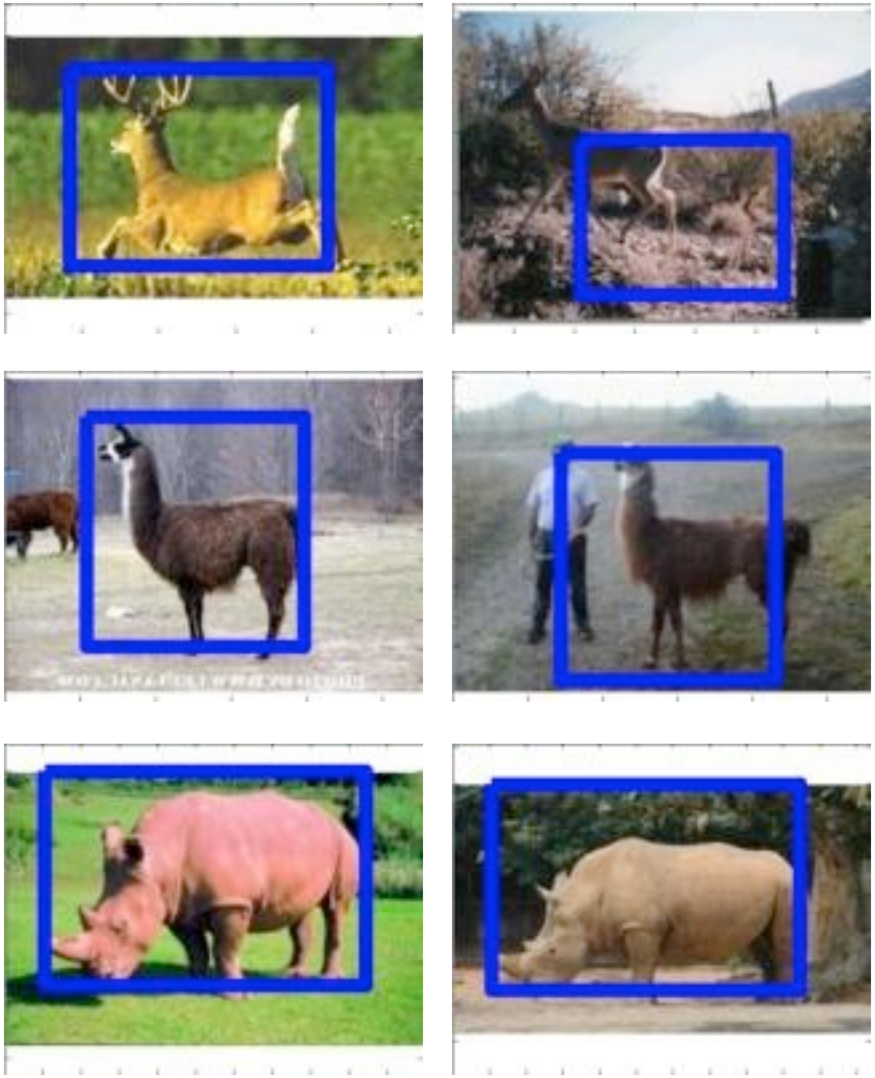
Feature $\Psi(\mathbf{x},\mathbf{y},\mathbf{h})$ – Standard HOG

$$(\mathbf{y}^*,\mathbf{h}^*) = \max_{\mathbf{y}\in Y,\mathbf{h}\in H} \mathbf{w}^T\Psi(\mathbf{x},\mathbf{y},\mathbf{h})$$

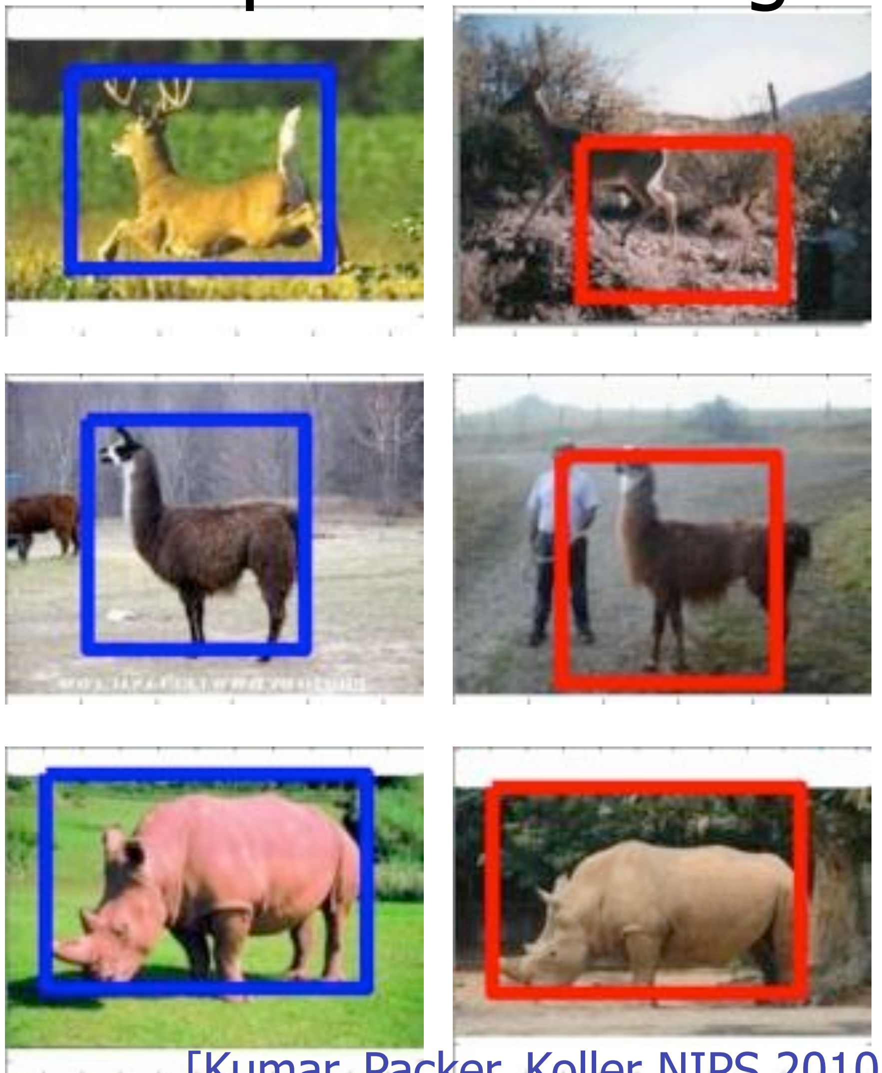
$v_i = 1$ (used) $v_i = 0$ (not used)

Imputation – Iteration 1

CCCP



Self-paced learning

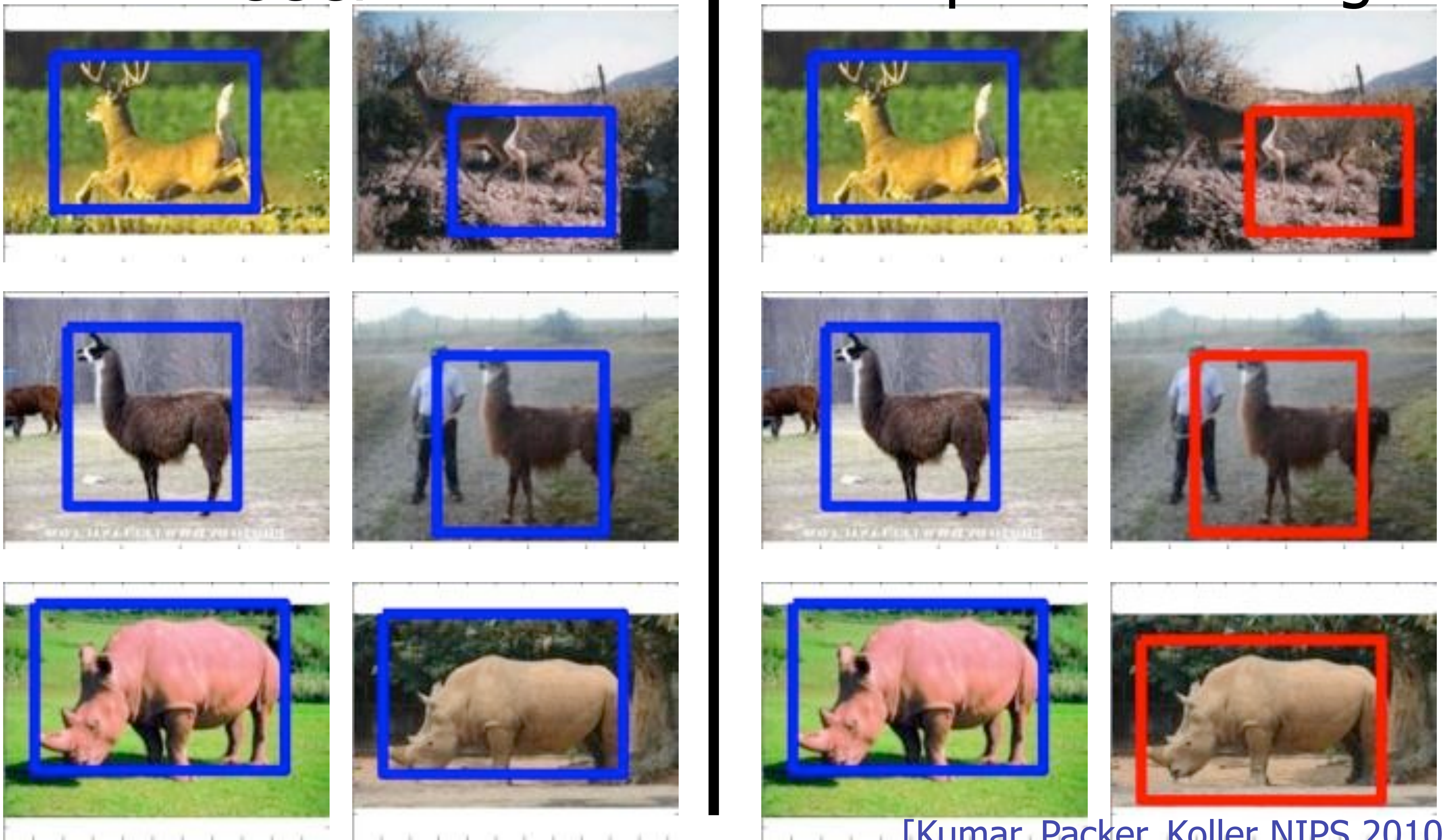


[Kumar, Packer, Koller NIPS 2010]

$v_i = 1$ (used) $v_i = 0$ (not used)

Imputation – Iteration 5

CCCP | Self-paced learning

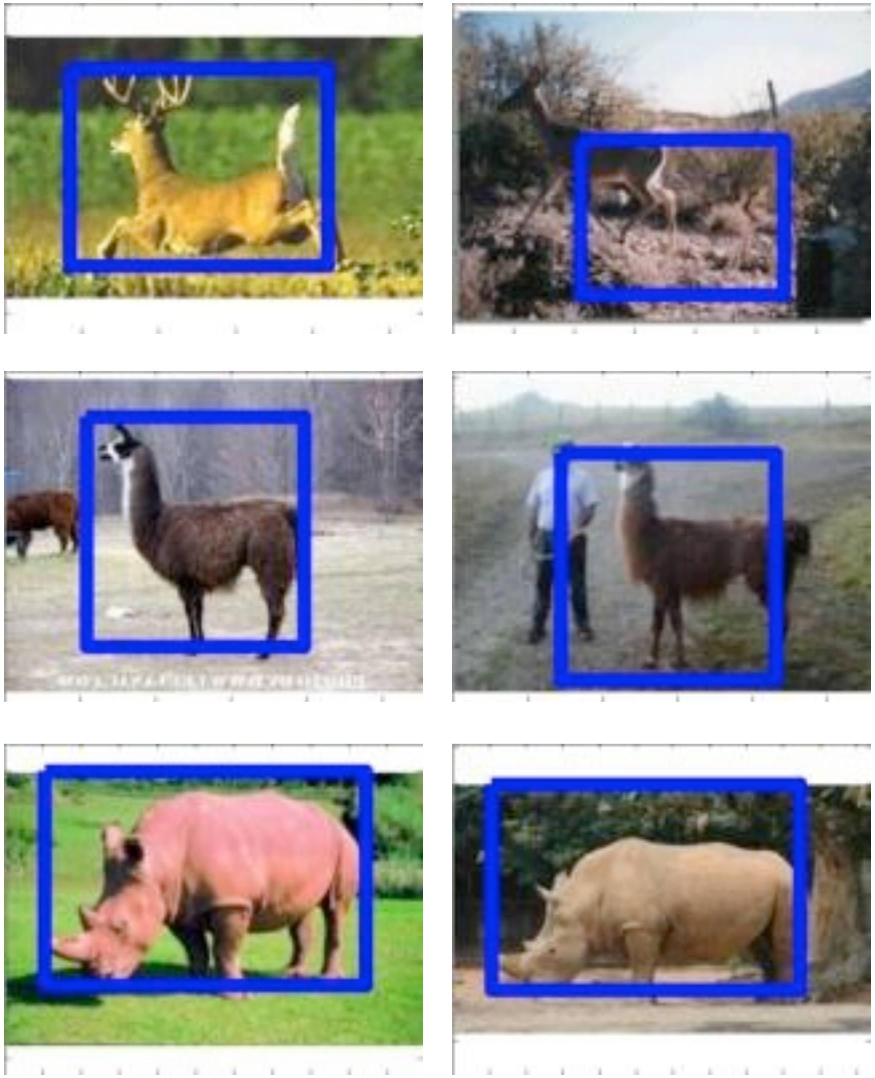


[Kumar, Packer, Koller NIPS 2010]

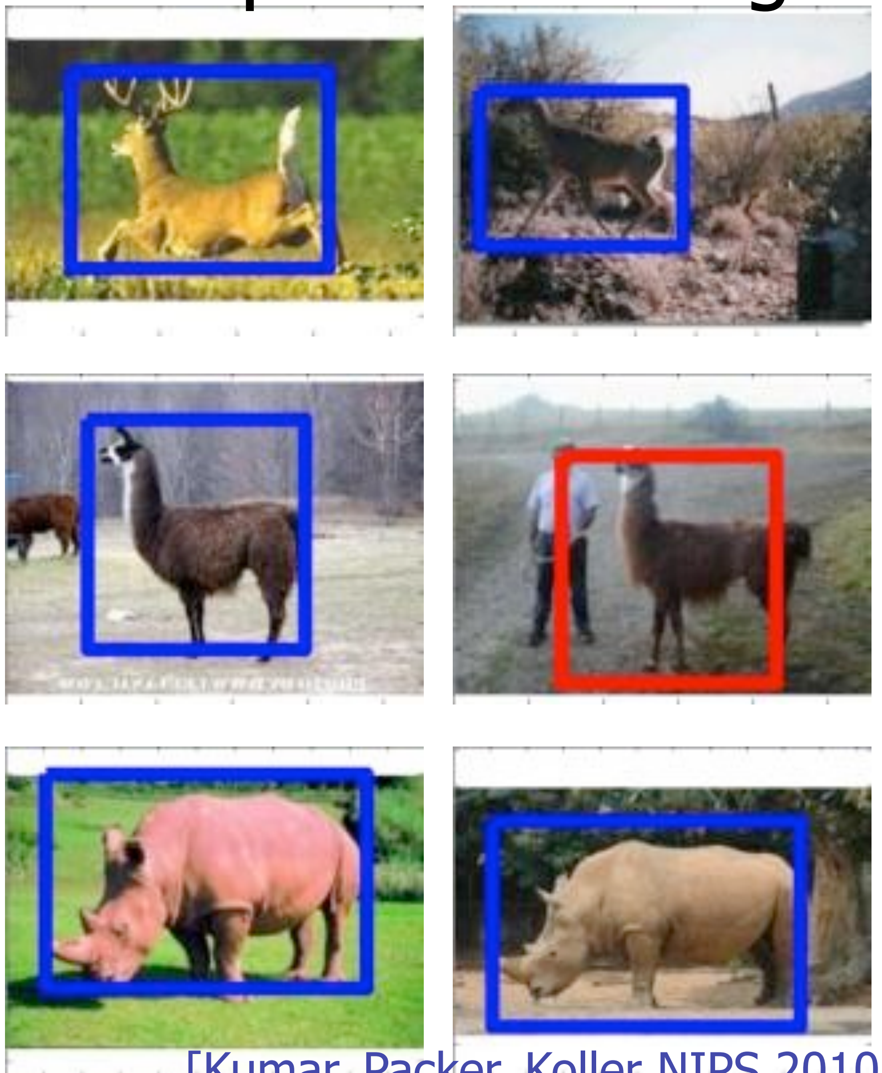
$v_i = 1$ (used) $v_i = 0$ (not used)

Imputation – Iteration 9

CCCP



Self-paced learning

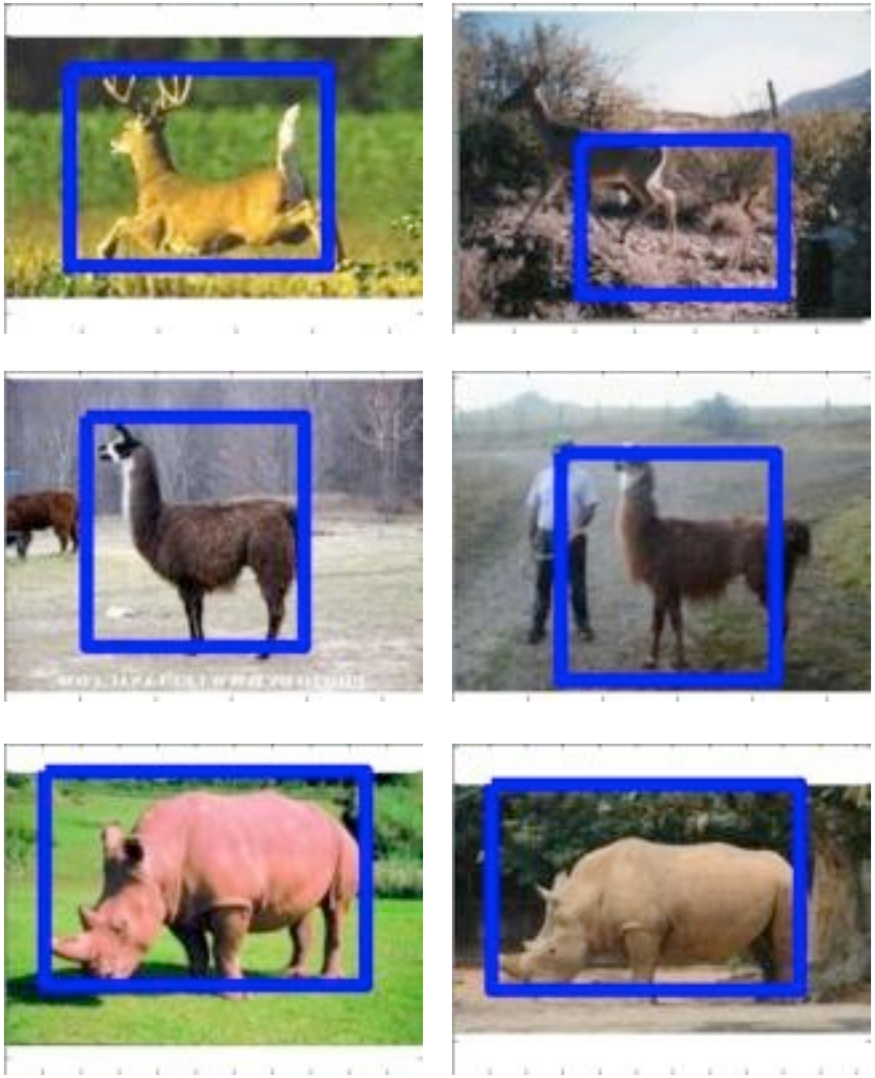


[Kumar, Packer, Koller NIPS 2010]

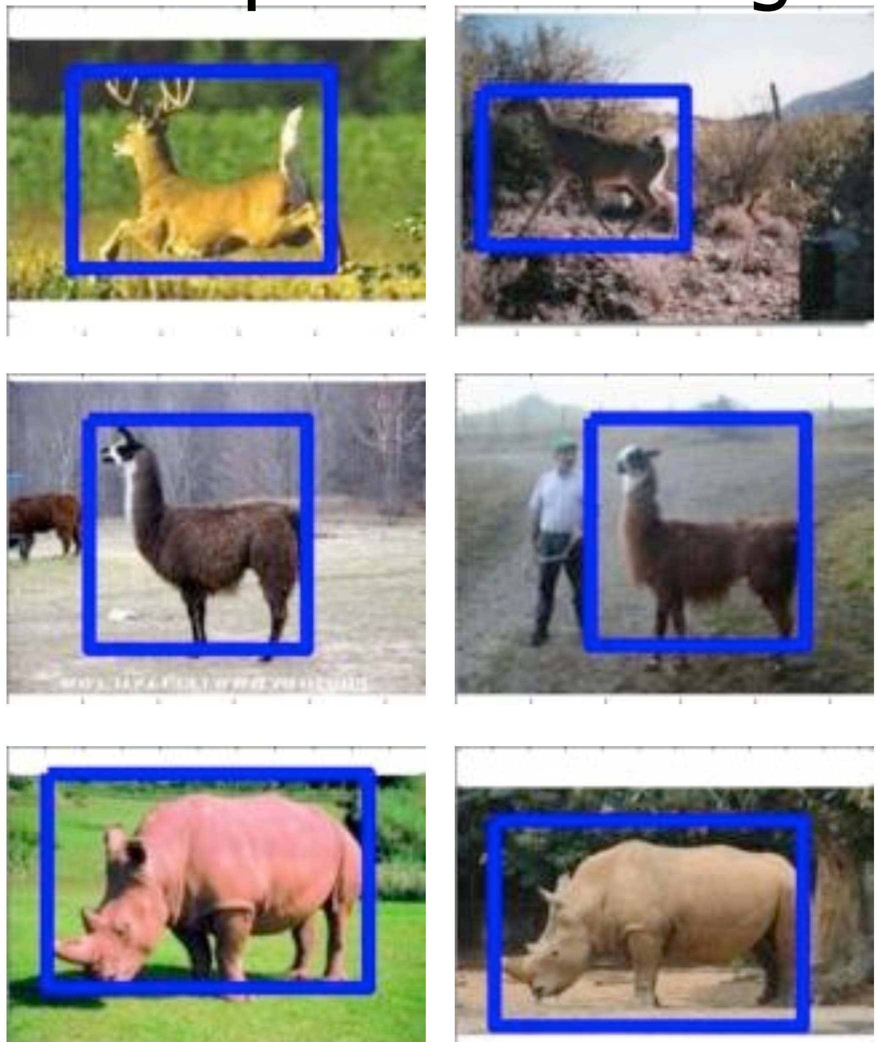
$v_i = 1$ (used) $v_i = 0$ (not used)

Imputation – Iteration 13

CCCP



Self-paced learning



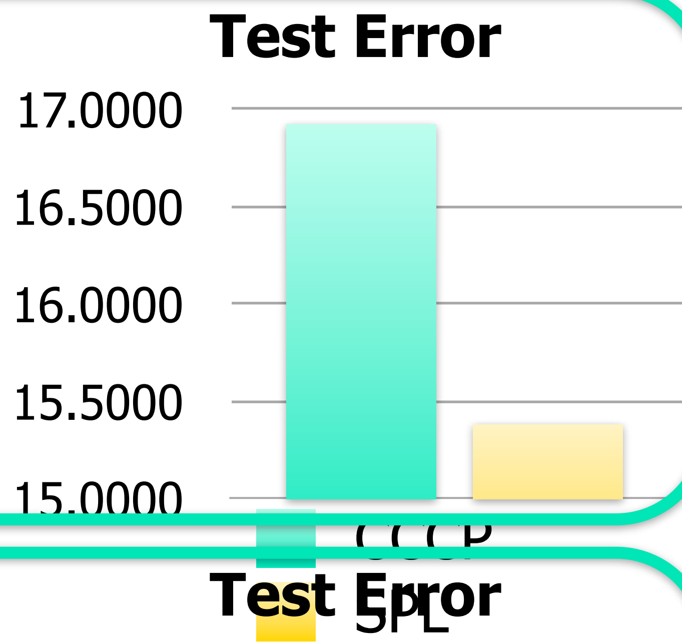
[Kumar, Packer, Koller NIPS 2010]

[Kumar, Packer, Koller NIPS 2010]

Self-Paced Learning

Object detection

y = "Deer"



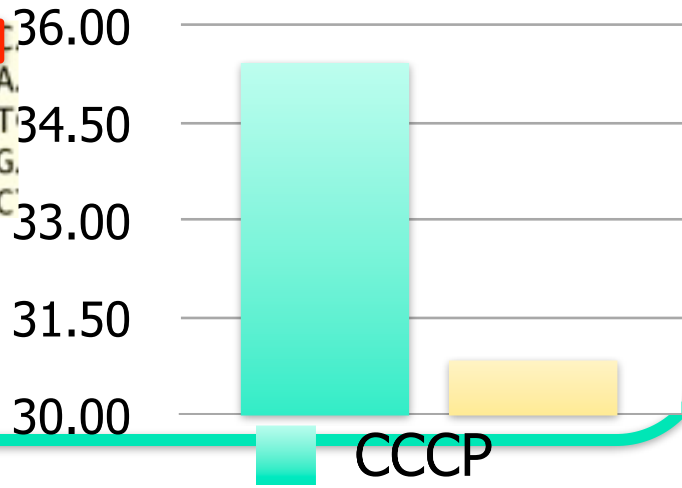
DNA motif finding

x

AGACCTAACCACACATTGCTAATTAG
GAACAATGTAAATTATTGAAAGGGCTA
TGATTAATTAATCCCATGGTTCGCTCT
GCTCTTAATTAATGAACCCGCTCTTGG
ATCTGGTCGTCTAATTAGCTACGGTAC

y = bind/no bind

h = Motif position



Self-Paced Learning

Object detection PASCAL VOC 2007

y = "Car/No Car"

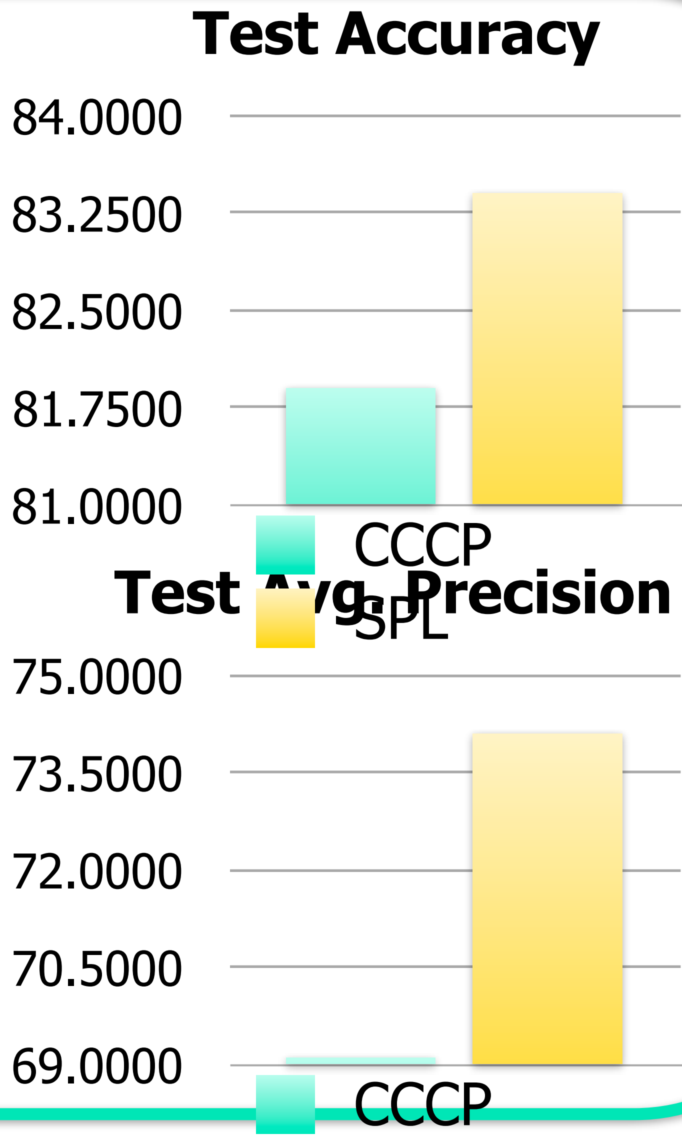
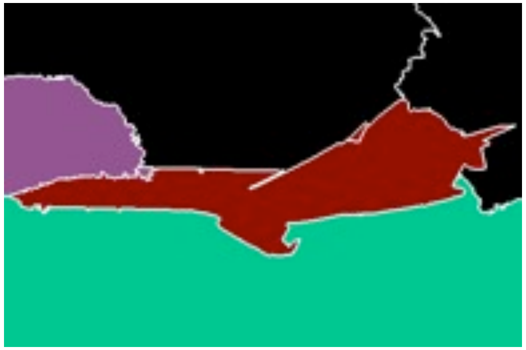


Image Segmentation Revisited

Input

CLL



SPL

SPL can make good use of weak annotations

Difference (SPL-CCCP)

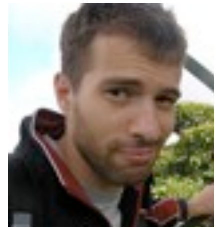
CCCP	SPL
24.7%	28.8%

CCCP	SPL
53.8%	55.3%

Classes

Outline

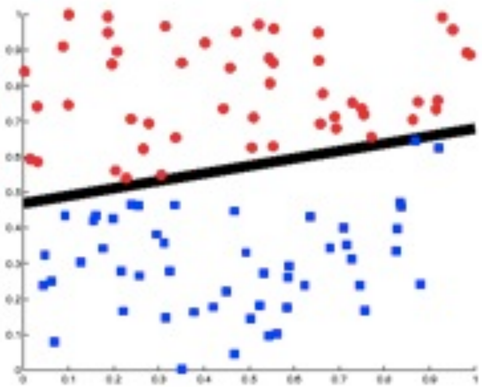
Pawan Kumar



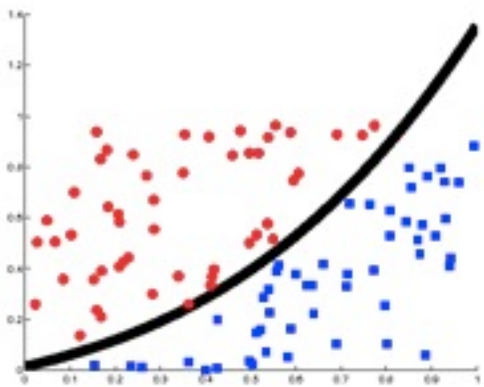
Ben Packer

- Holistic scene models
- **Self-paced learning for latent variables**
  - Instance selection
  - **Model selection**

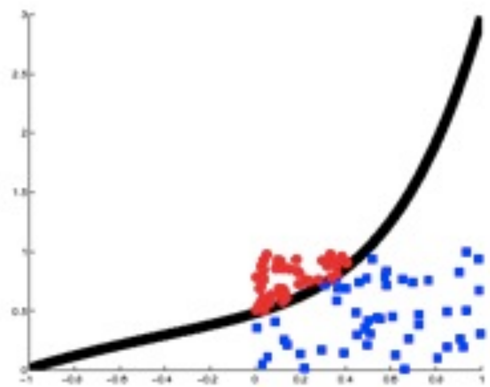
Model Selection



Linear



Cubic

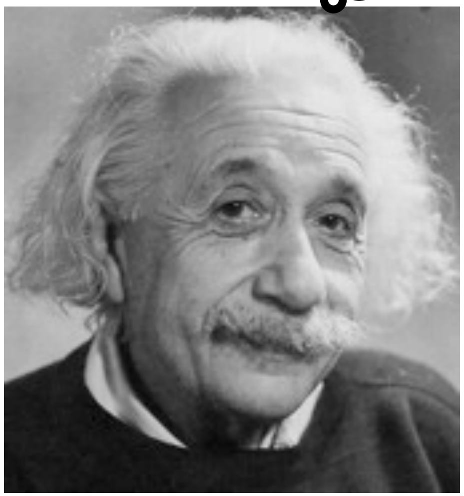


Quintic

Which kernel should I use?

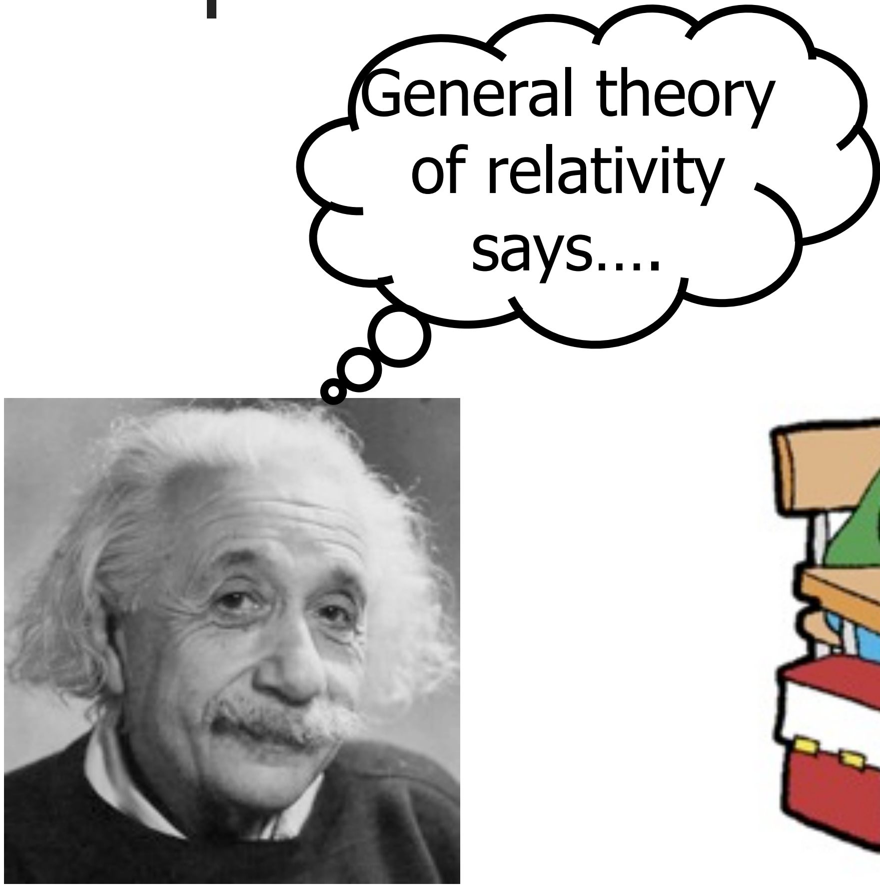
Human learning revisited

General theory
of relativity
says....

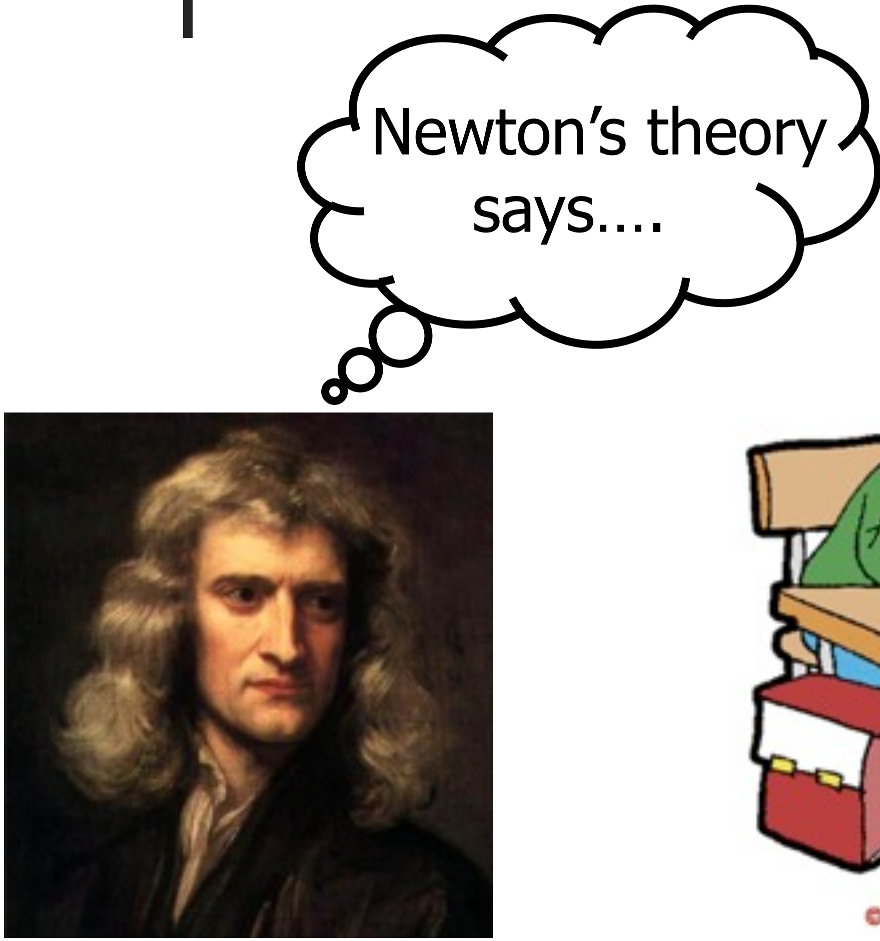


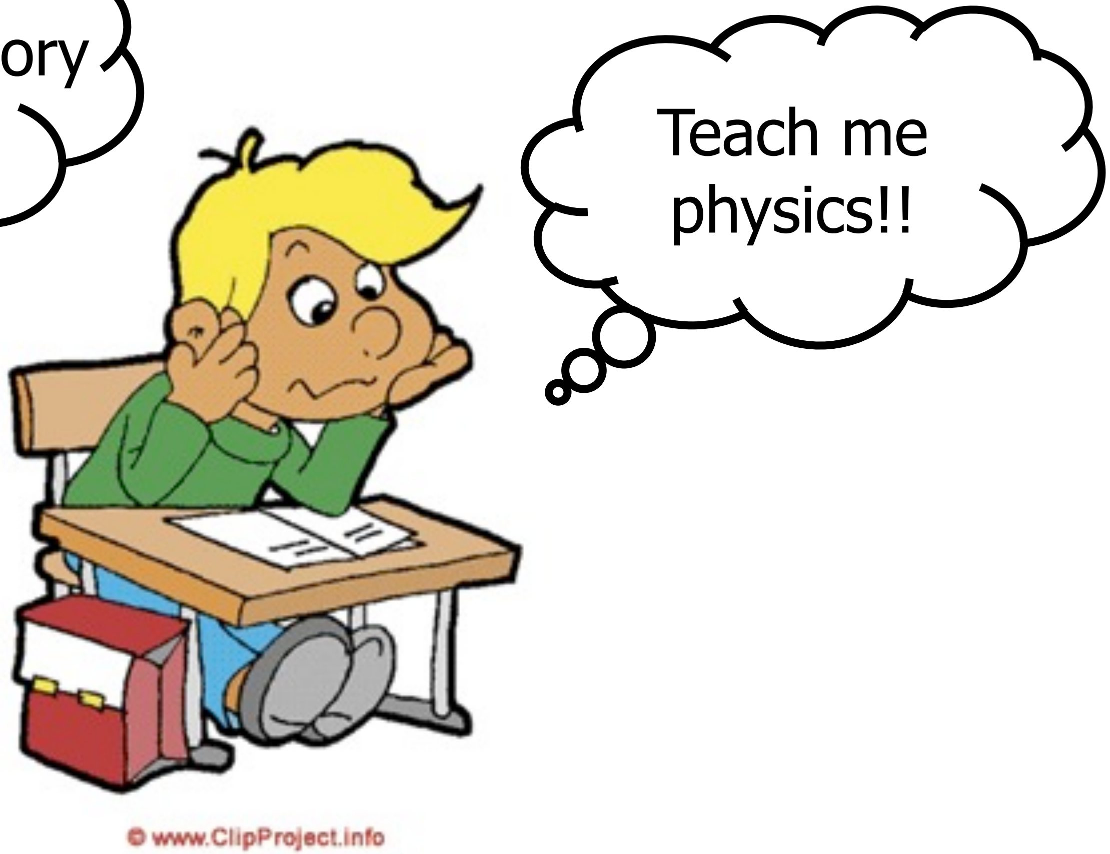
Teach me
physics!!

Human learning revisited



Human learning revisited

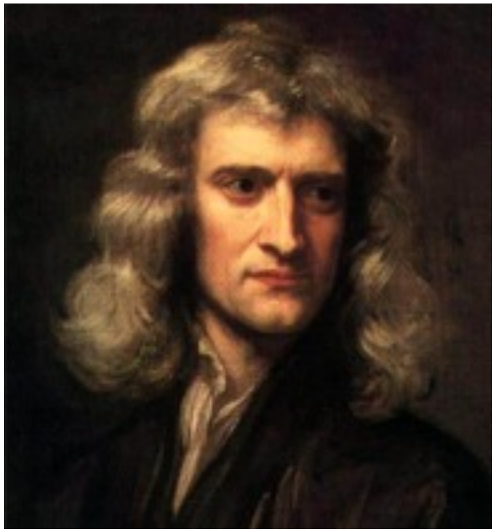




Human learning revisited

Newton's theory says….

Got it!



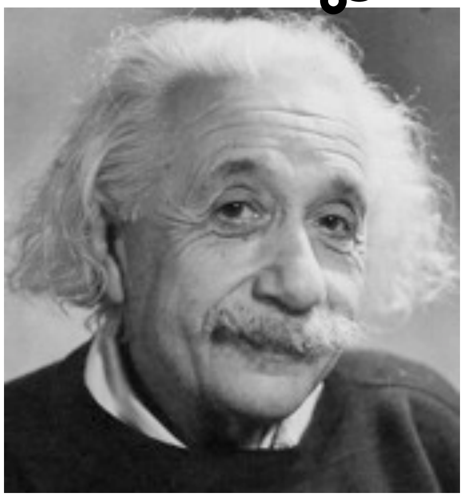
Human learning revisited

Special theory of relativity says….

Got it!

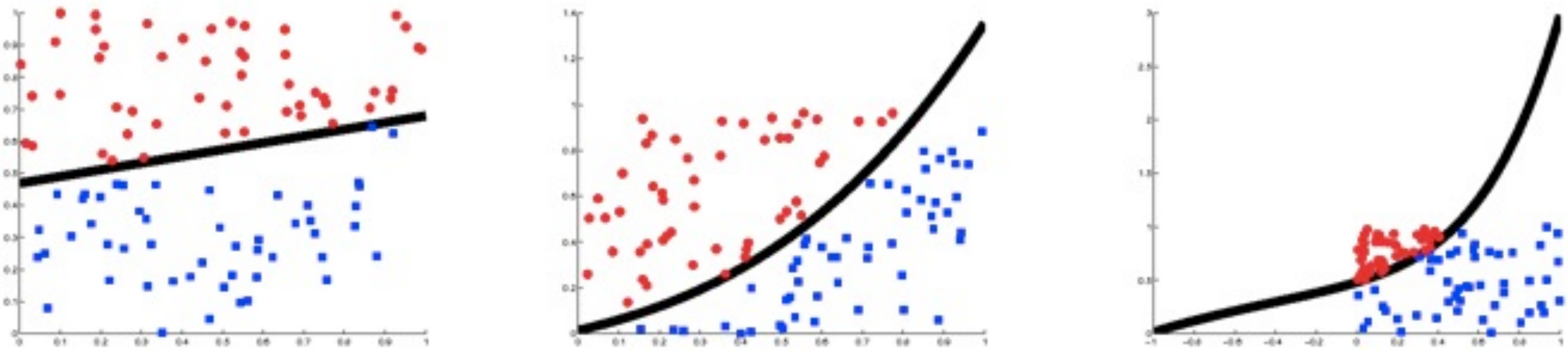
Human learning revisited

General theory
of relativity
says….



Got it!

Multiple Kernel Learning



Linear Cubic Quintic

$$K = \Sigma_i a_i K_i$$

Kernel weights $a_i \geq 0$

$$\boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x},\mathbf{y},\mathbf{h}) = \begin{pmatrix} \sqrt{a_1}\Psi_1(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \sqrt{a_2}\Psi_2(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \dots \end{pmatrix}$$

* Bach, Lanckriet, Jordan, ICML 2004

Multiple Kernel Learning*

$$\text{Minimize}_{\mathbf{w},\mathbf{a},\xi}\ ||\mathbf{w}||^2 + C\sum_i \xi_i$$

$$\max_{\mathbf{h}_i} \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h}_i) - \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i,\mathbf{y},\mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall\ \mathbf{y}, \mathbf{h}$$

Minimizing ξ_i encourages most complex kernel !!

$K = \Sigma_i a_i K_i$

Kernel weights $a_i \geq 0$

$$\boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x},\mathbf{y},\mathbf{h}) = \begin{pmatrix} \sqrt{a_1}\Psi_1(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \sqrt{a_2}\Psi_2(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \ldots \end{pmatrix}$$

* Bach, Lanckriet, Jordan, ICML 2004

Self-Paced Multiple Kernel Learning

$$\text{Minimize}_{\mathbf{w},\mathbf{a},\xi}\ ||\mathbf{w}||^2 + C\sum_i \xi_i + \lambda R(\mathbf{a})$$

$$\max_{\mathbf{h}_i} \mathbf{w}^T \boldsymbol{\varphi}_\mathbf{a}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h}_i) - \mathbf{w}^T \boldsymbol{\varphi}_\mathbf{a}(\mathbf{x}_i,\mathbf{y},\mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall\ \mathbf{y}, \mathbf{h}$$

$R(\mathbf{a}) = \Sigma_j\, r_j\, a_j$ $\qquad r_j$ = Rademacher complexity

$K = \Sigma_i\, a_i K_i$ $\qquad \boldsymbol{\varphi}_\mathbf{a}(\mathbf{x},\mathbf{y},\mathbf{h}) = \begin{pmatrix} \sqrt{a_1}\Psi_1(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \sqrt{a_2}\Psi_2(\mathbf{x},\mathbf{y},\mathbf{h}) \\ \ldots \end{pmatrix}$

Kernel weights $a_i \geq 0$

Self-Paced Multiple Kernel Learning

Start with an initial estimate $\mathbf{w}_0$, $\mathbf{a}_0$

Update $\mathbf{h}_i = \text{argmax}_{\mathbf{h} \in H}\ \mathbf{w}_t^T \boldsymbol{\varphi}_{\mathbf{a}_t}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h})$

Update $\mathbf{w}_{t+1}$ and $\mathbf{a}_{t+1}$ by solving convex problem

$$\text{Minimize}_{\mathbf{w},\mathbf{a},\xi}\ ||\mathbf{w}||^2 + C\sum_i \xi_i + \lambda R(\mathbf{a})$$

$$\mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}_i) - \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i, \mathbf{y}, \mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall\ \mathbf{y}, \mathbf{h}$$

SPMKL Behavior

$$\text{Minimize}_{\mathbf{w},\mathbf{a},\xi} \ ||\mathbf{w}||^2 + C\sum_i \xi_i + \lambda R(\mathbf{a})$$

$$\max_{\mathbf{h}_i} \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i, \mathbf{y}_i, \mathbf{h}_i) - \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i, \mathbf{y}, \mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall \ \mathbf{y}, \mathbf{h}$$

Early iterations:

- $\mathbf{h}_i$ are incorrectly imputed
- ξ_i are large even for complex kernels
- Simple kernels are preferred to minimize R(**a**)

SPMKL Behavior

$$\text{Minimize}_{\mathbf{w},\mathbf{a},\xi}\ ||\mathbf{w}||^2 + C\sum_i \xi_i + \lambda R(\mathbf{a})$$

$$\max_{\mathbf{h_i}} \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{h}_i) - \mathbf{w}^T \boldsymbol{\varphi}_{\mathbf{a}}(\mathbf{x}_i,\mathbf{y},\mathbf{h})$$

$$\geq \Delta(\mathbf{y}_i, \mathbf{y}, \mathbf{h}) - \xi_i \qquad \forall\ \mathbf{y}, \mathbf{h}$$

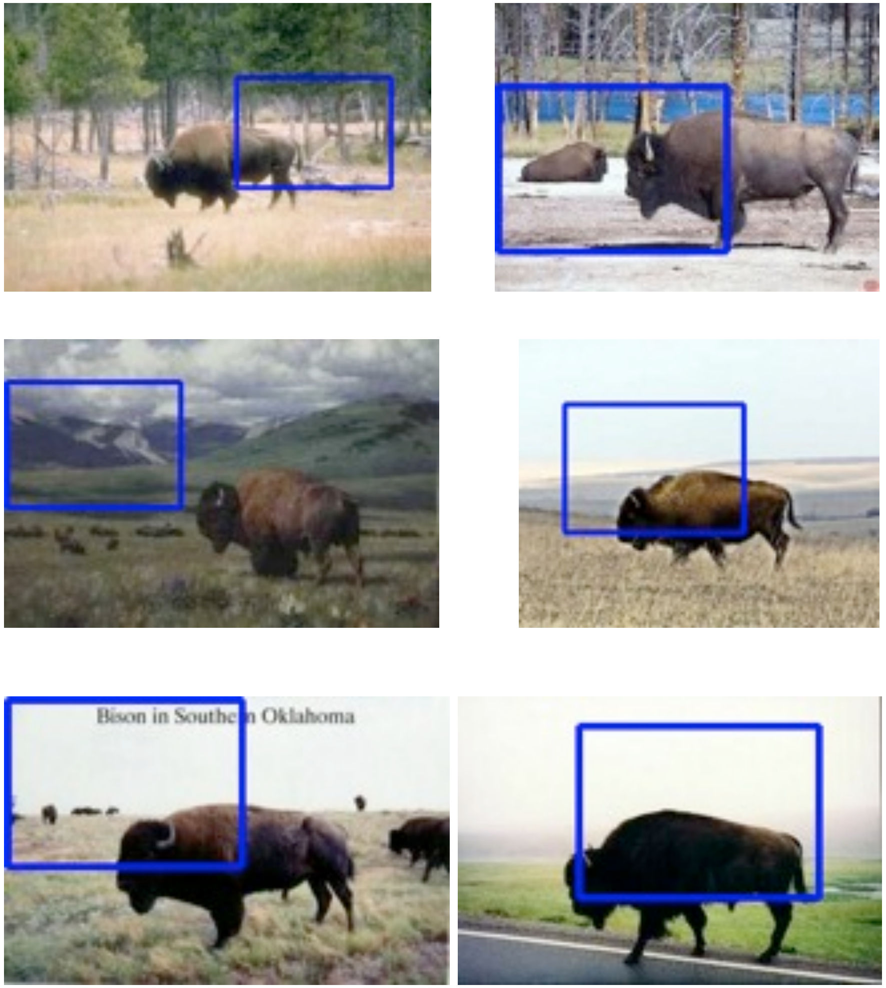
Later iterations:

- $\mathbf{h}_i$ are correctly imputed
- ξ_i is small for complex kernels
- Complex kernels are preferred to minimize $\sum_i \xi_i$

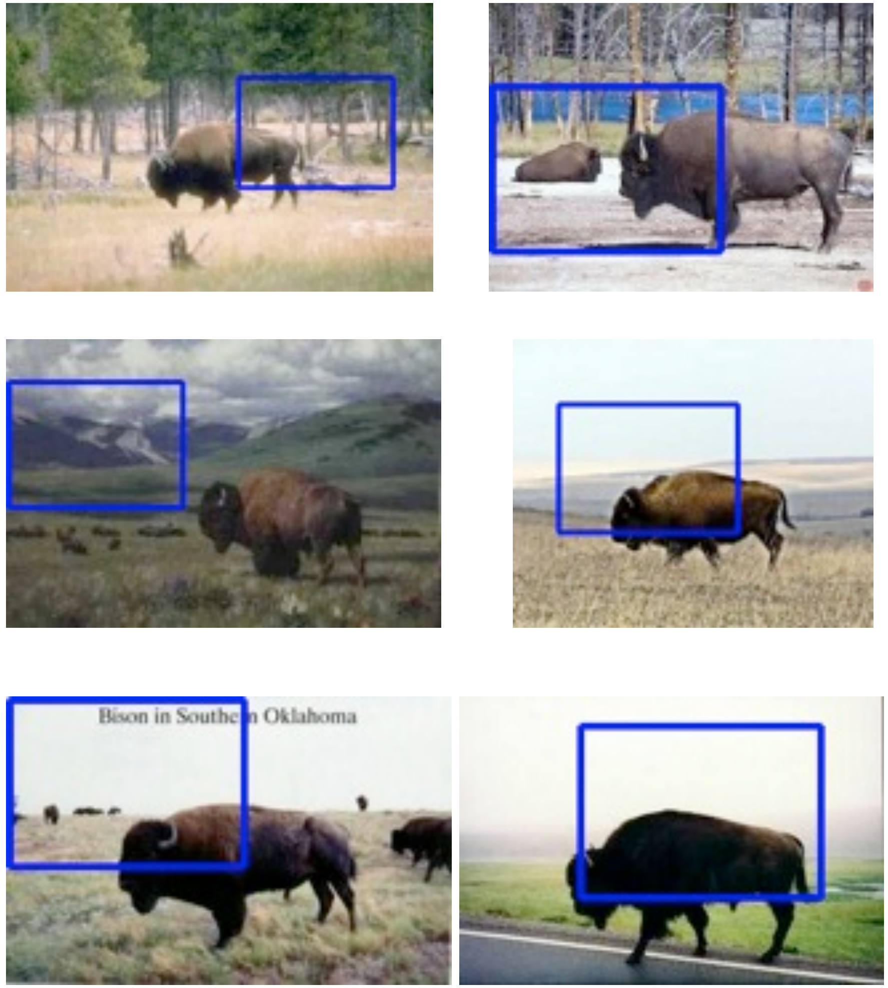
No need to anneal λ

Imputation – Iteration 1

CUBIC

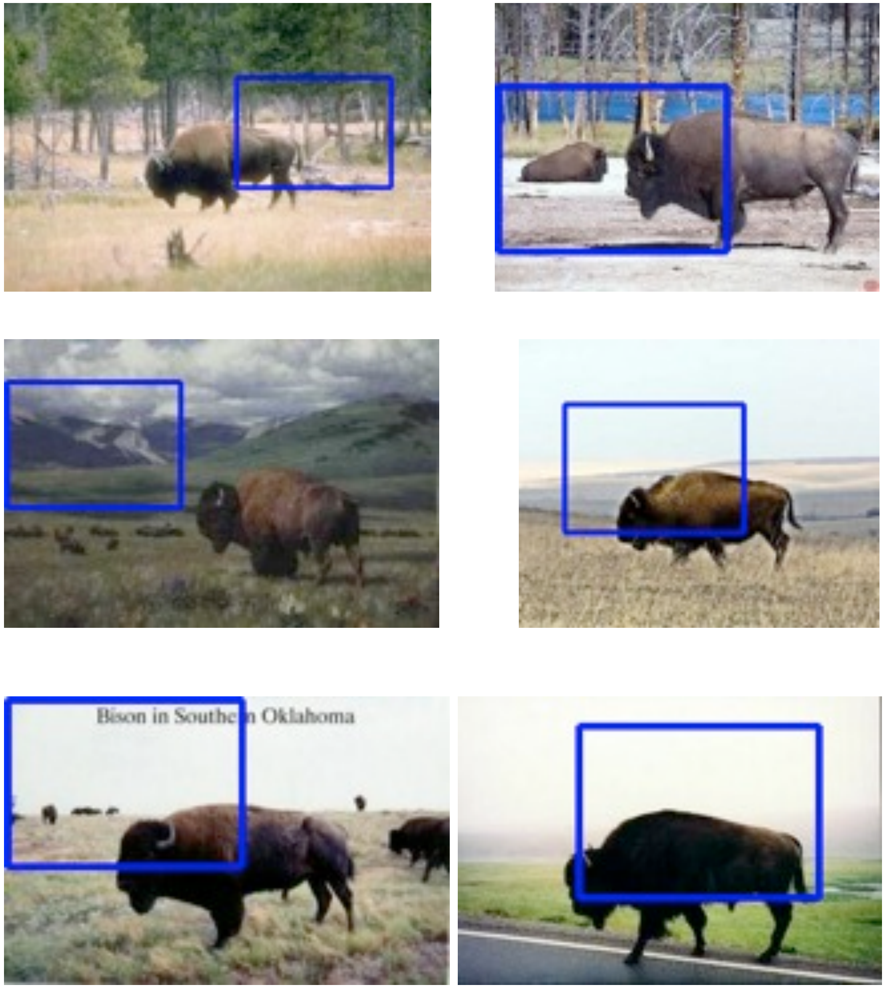


SPMKL

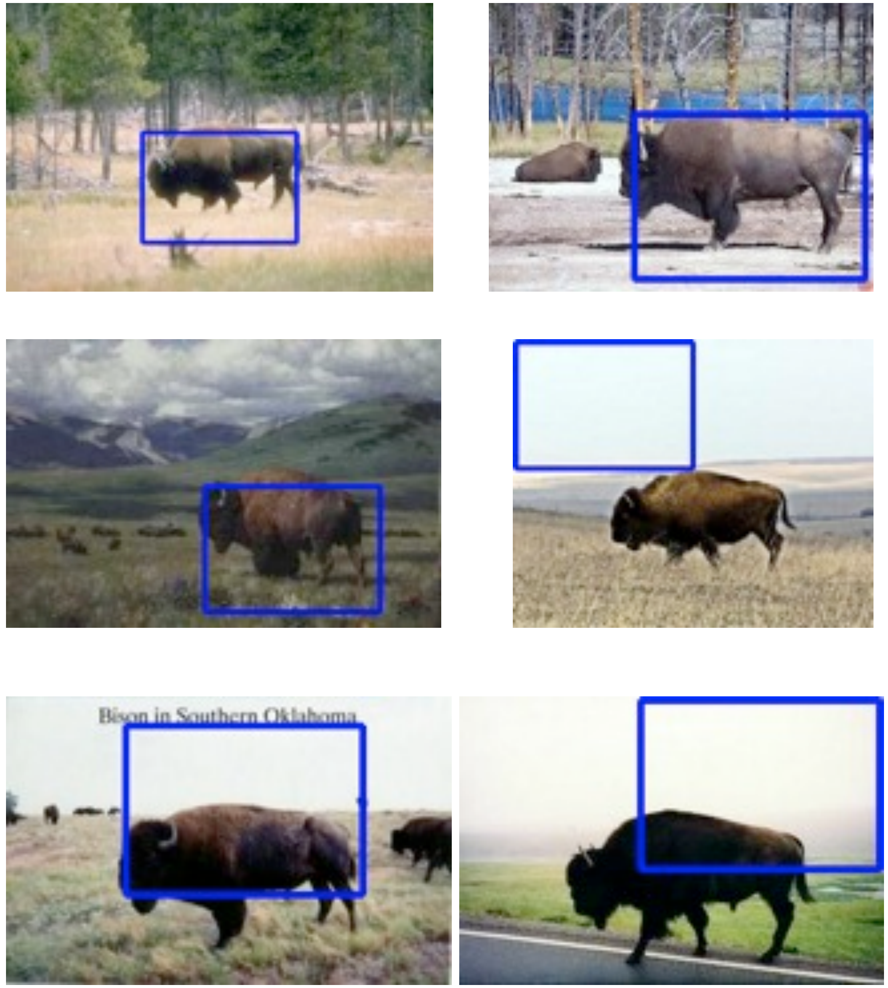


Imputation – Iteration 3

CUBIC

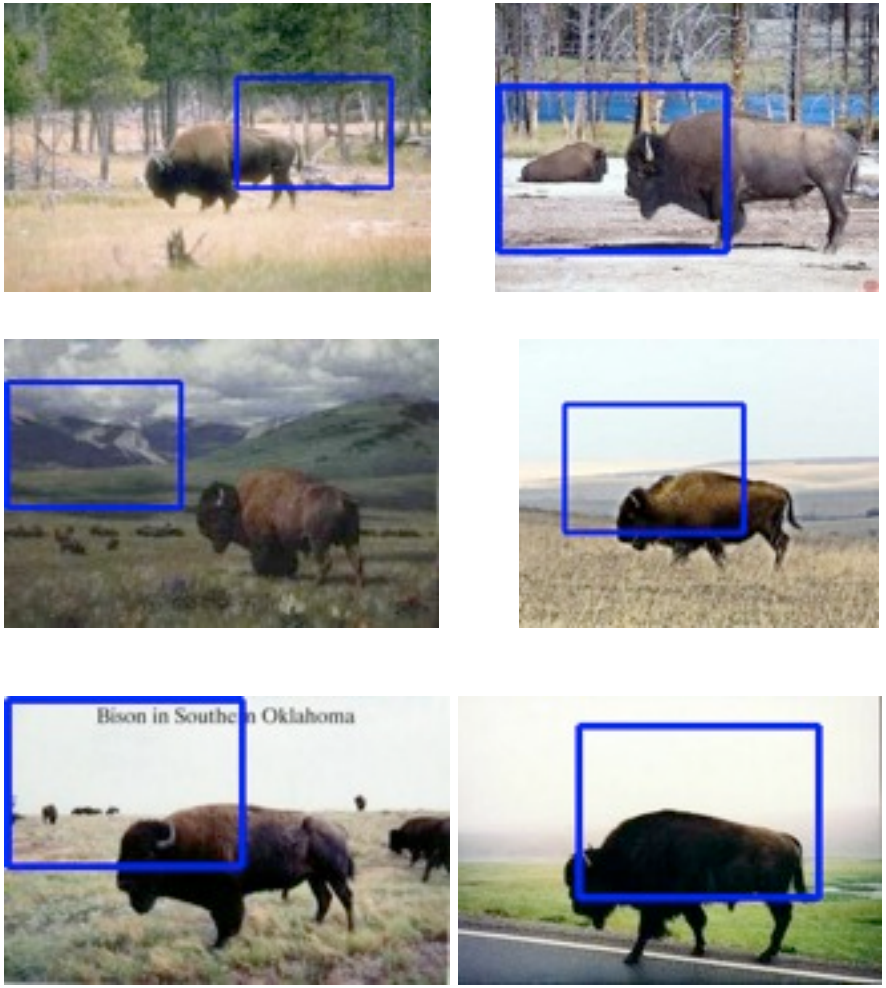


SPMKL

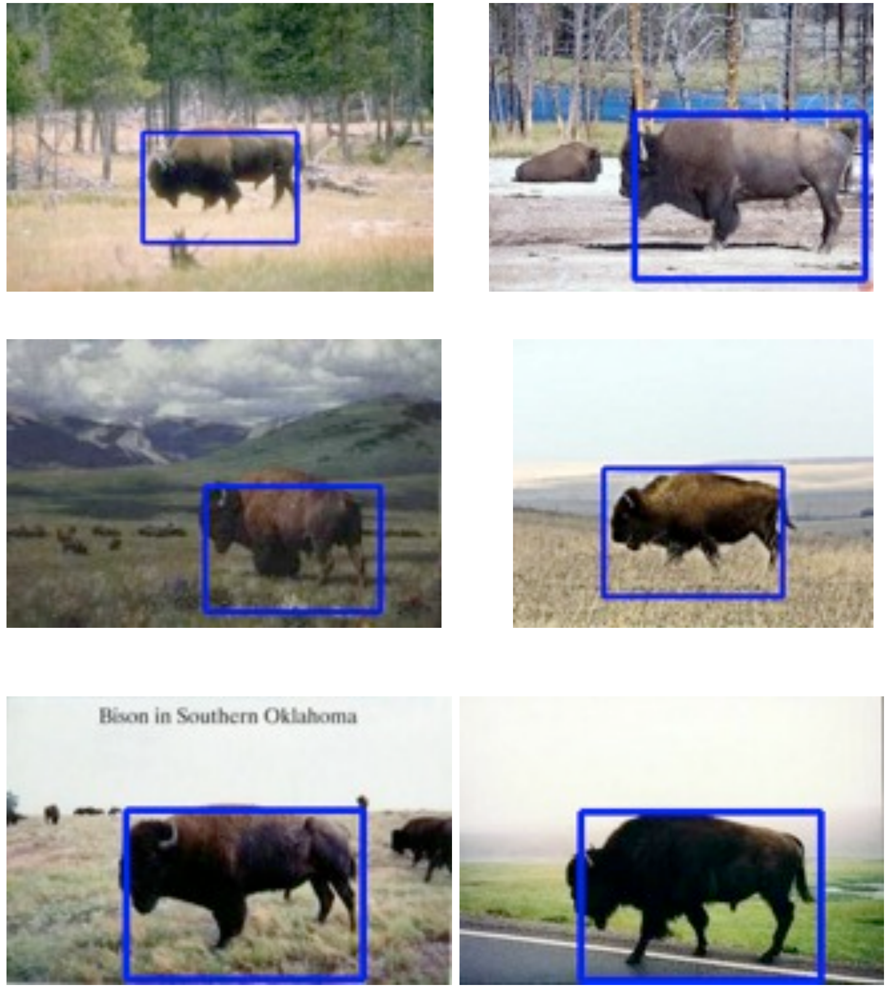


Imputation – Iteration 6

CUBIC

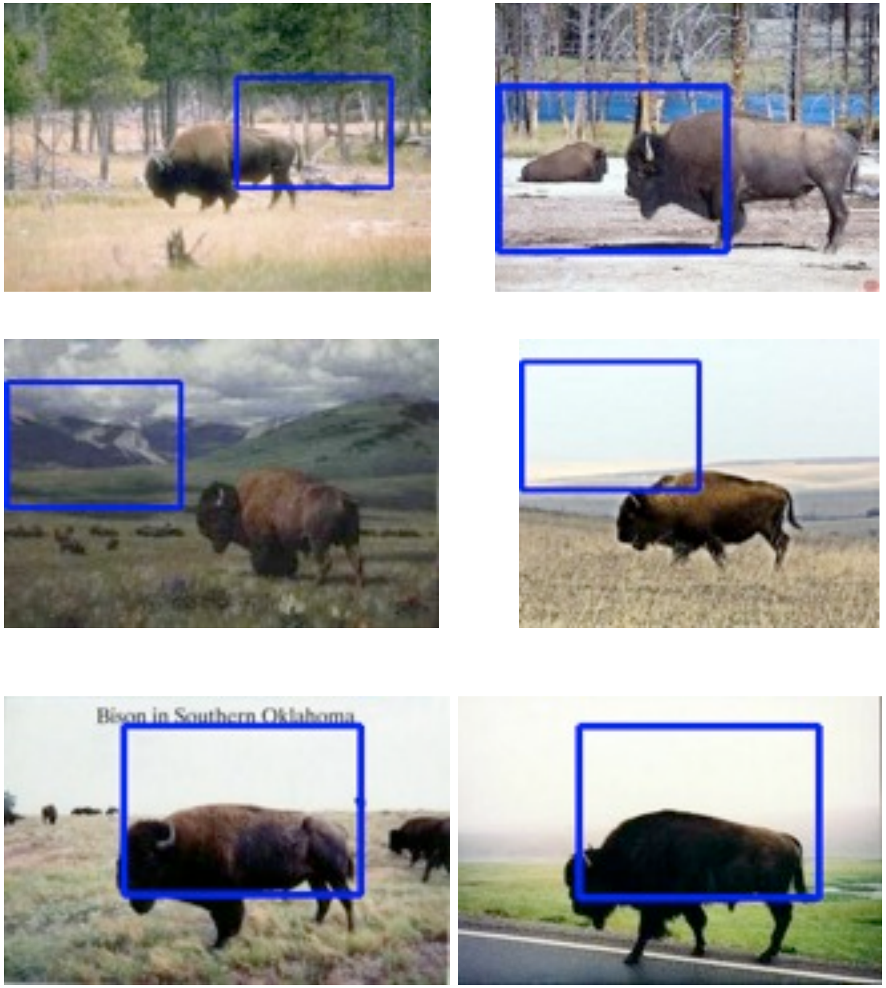


SPMKL

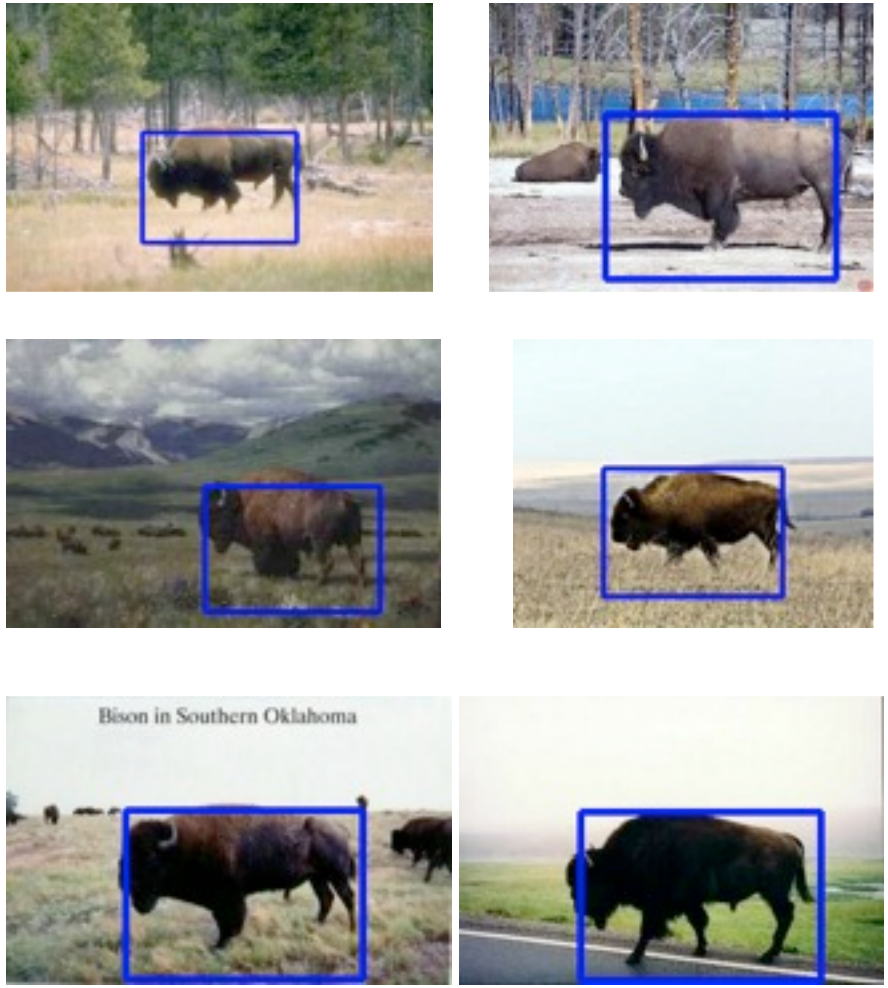


Imputation – Iteration 10

CUBIC

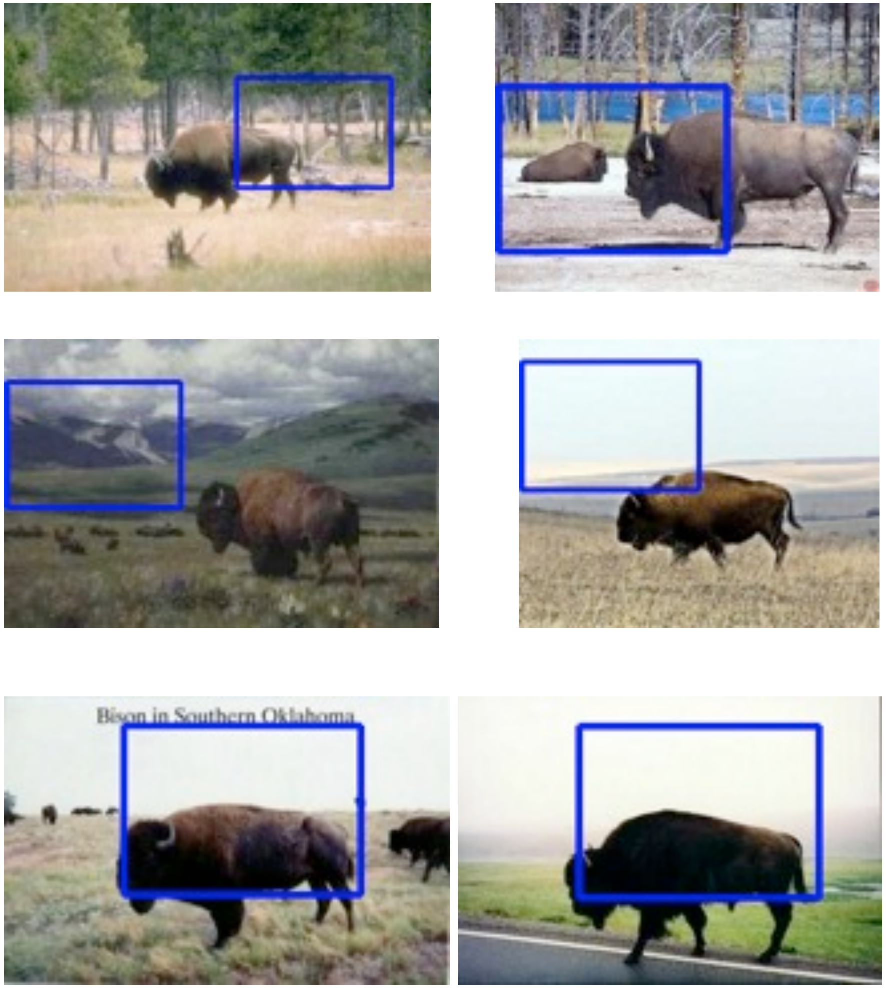


SPMKL

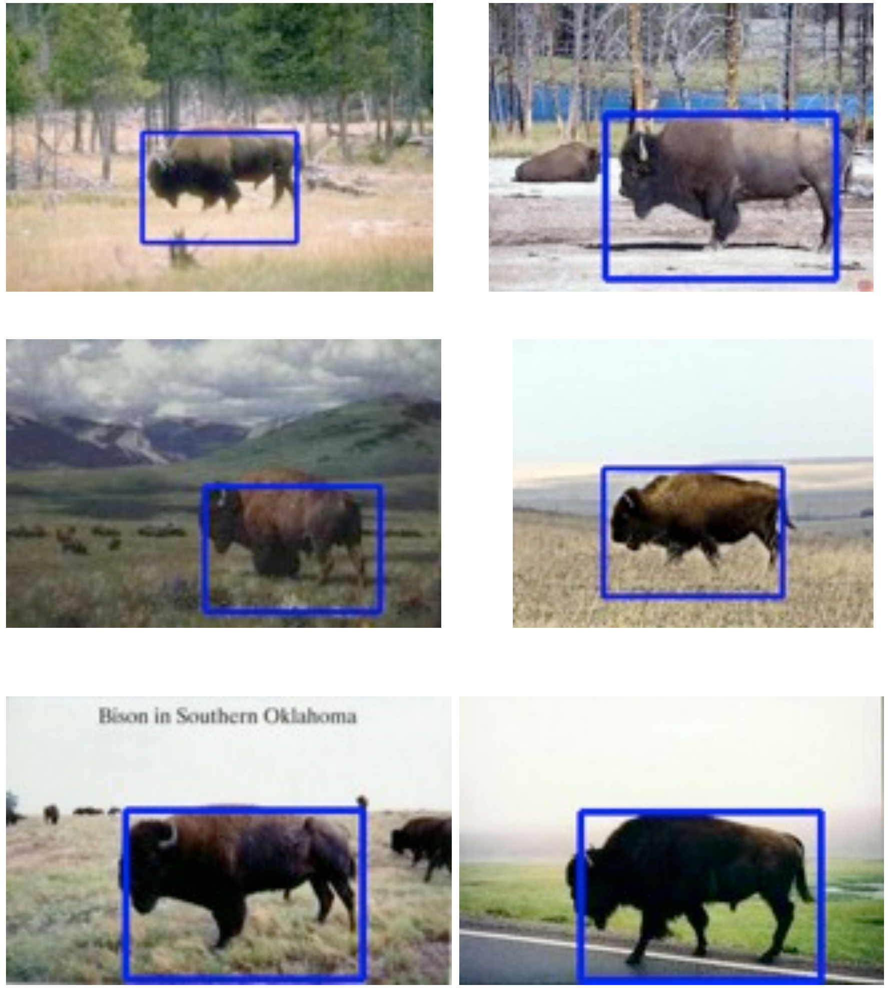


Imputation – At Convergence

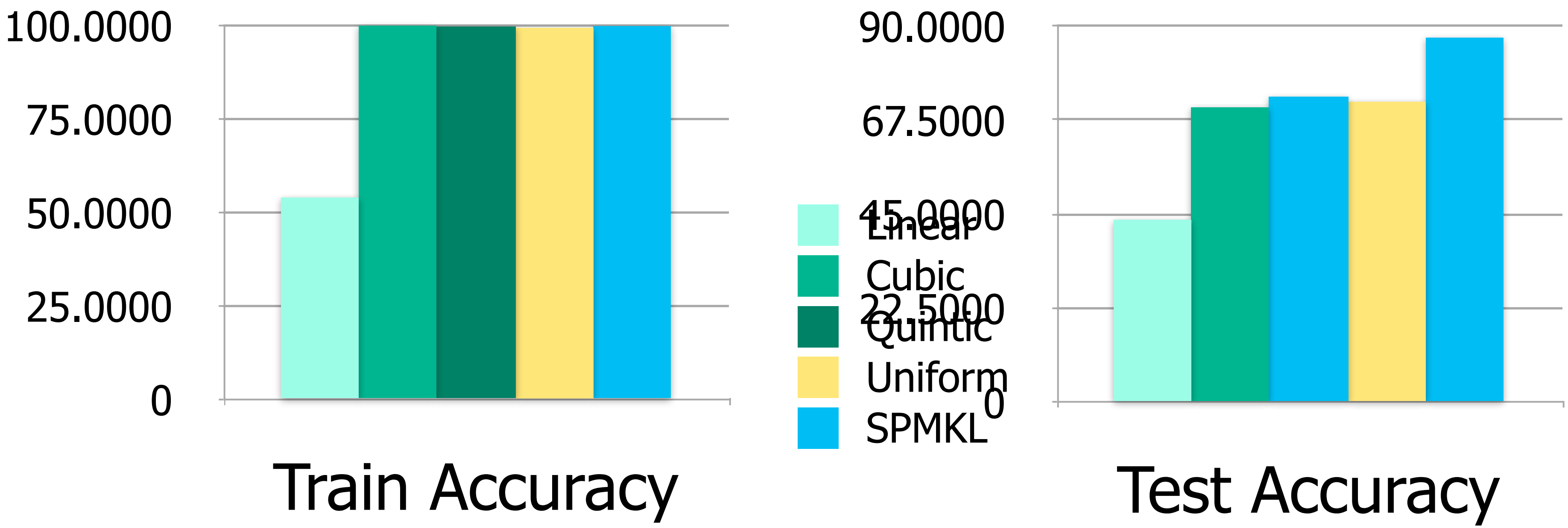
CUBIC



SPMKL



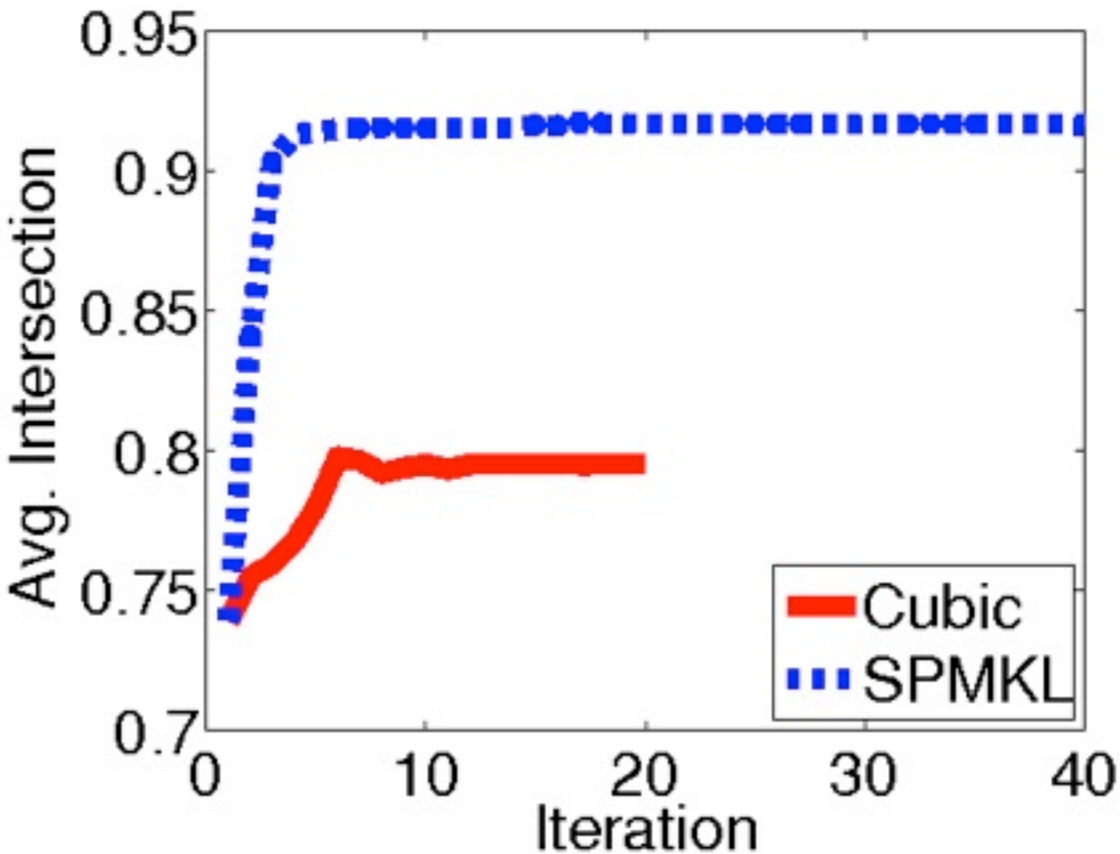
Classification Accuracy



- Linear kernel underfits
- Stronger kernels overfit to noisy imputations and get stuck at local optimum
- SPMKL only uses strong kernels when imputations are accurate, avoiding local optimum

Bounding Box Imputation

$$\text{Score} = \frac{\text{Area of Intersection of A and B}}{\text{Area of A}}$$



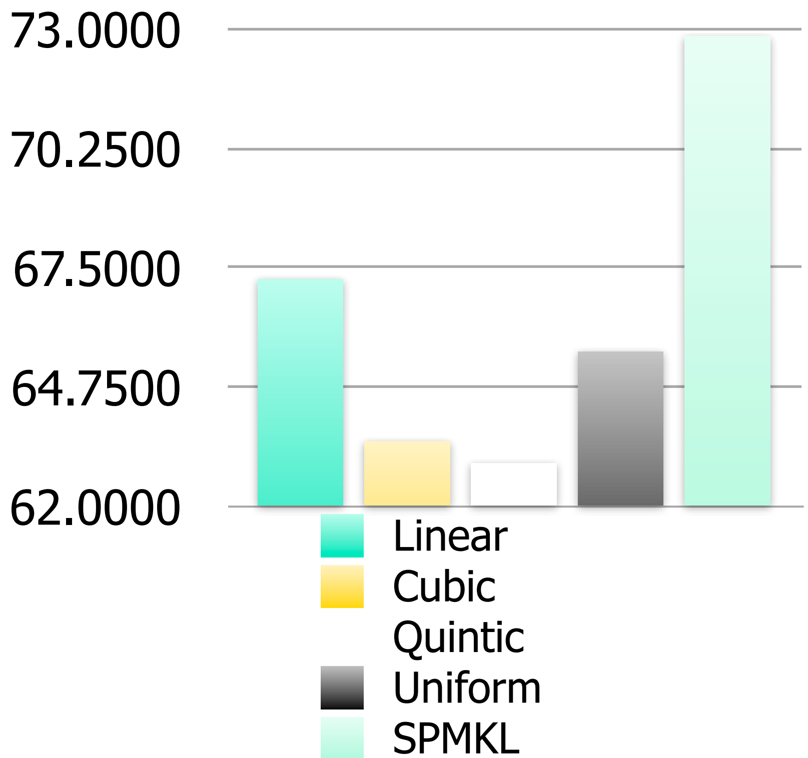
DNA Binding Motif

Input **x**

AGACCTAACCACACATTGCTAATTAGC
GAACAATGTAAATTATTGAAAGGGCTA
TGATTAATTAATCCCATGGTTCGCTCT
GCTCTTAATTAATGAACCCGCTCTTGG
ATCTGGTCGTCTAATTAGCTACGGTAC

y = bind/no bind
h = Motif position

Test Accuracy



Conclusion I

- Pixel-level scene understanding enforces coherent scene interpretation and contextual consistency

- Training data is an issue
  - Pixel-level annotations come in limited amounts
  - Human annotations not always ideal to task

Conclusion II

- We need to make better use of data:
  - Weakly labeled data
  - Diverse data with different levels of annotation
  - Unsupervised data
  - …
- **Latent variables critical**

Conclusion III

- Don’t jump too quickly:
  - Solve the hardest instances
  - Use the richest model

- Let the algorithm gradually adapt to increasing levels of complexity

The Future of Education

January 28, 2011 | 10:33 AM | By Tina Barseghian

KQED | News | Radio | TV | Education | Arts | Food | Science | Community | Support

Future School Day: Self-Paced Learning, Creating, and Collaborating

FILED UNDER: Learning Methods, Tech Tools, individualized learning, Khan Academy, project-based-learning, salman Khan, School Day of the Future

1 Comment Tweet 47 Share 53 Email Post Permalink

Salman Khan has an idea or two about what the future school day should be. In fact, the founder of **Khan Academy** — a series of thousands of YouTube videos that teach everything from calculus to the French Revolution — is working on making it happen as we speak.

It goes something like this:

- Every student working at his or her own pace.
- Students working in groups and helping each other.
- Teachers working one-on-one with students.
- And a school day full of creative, hands-on projects that give kids practical knowledge and experience.

The Future of Machine Learning?

January 28, 2011 | 10:33 AM | By **Tina Barseghian**

KQED | News | Radio | TV | Education | Arts | Food | Science | Community | Support

Future School Day: Self-Paced Learning, Creating, and Collaborating

FILED UNDER: **Learning Methods**, **Tech Tools**, **individualized learning**, **Khan Academy**, **project-based-learning**, **salman Khan**, **School Day of the Future**

1 Comment Tweet 47 Share 53 Email Post Permalink

Salman Khan has an idea or two about what the future school day should be. In fact, the founder of **Khan Academy** — a series of thousands of YouTube videos that teach everything from calculus to the French Revolution — is working on making it happen as we speak.

It goes something like this:

algorithm

- Every ~~student~~ working at his or her own pace.

Kernels

- ~~Students~~ working in groups and helping each other.

algorithms

- Teachers working one-on-one with ~~students~~.

data set diverse, weakly-labeled instances

- And a ~~school day~~ full of ~~creative, hands-on projects~~ that give ~~kids~~ practical knowledge and experience.

algorithms

Acknowledgments

- Holistic scene models
  - **Stephen Gould**
  - Tianshi Gao
  - **Pawan Kumar**
  - Rick Fulton, Haithem Turki, Dan Preston
- Indoor scene models
  - **Huayan Wang**
  - Stephen Gould
- Self-paced learning
  - **Pawan Kumar**
  - **Ben Packer**
  - Kevin Miller, Rafi Witten

National Science Foundation, ONR MURI, Boeing Corp.