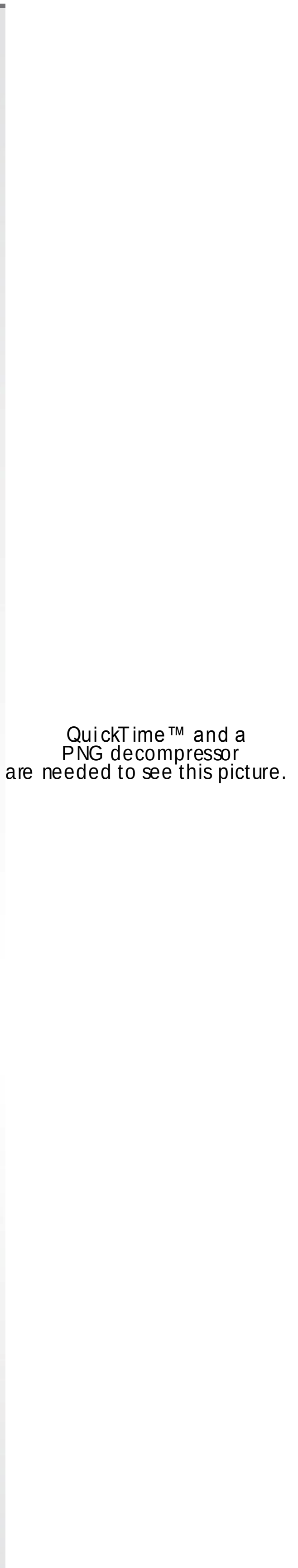
Semantic Scene Segmentation using Random Multinomial Logit

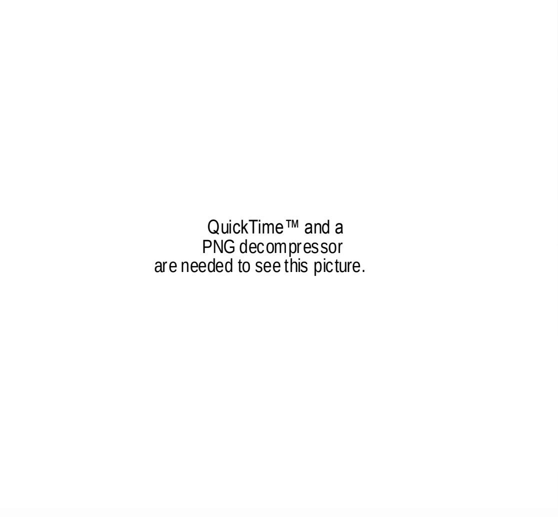
Ananth Ranganathan

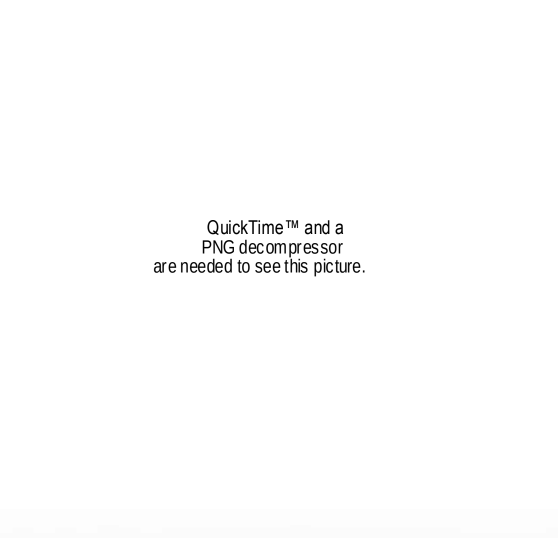
Honda Research Institute, USA

Analysis of Traffic Scenes

- Segment objects of interest in a street scene
- Use in intelligent transportation systems
  - Recognition should be perspective invariant
  - Wide intra-class variability
  - Need to work with video
  - Need to be fast







QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

What this talk is about

- **An Algorithm for Classification:**

  Random Multinomial Logistic Regression

  - Fast
  - Scales better with large intra-class variability, perspective etc
  - Scales well with number of labels
  - Very simple to implement

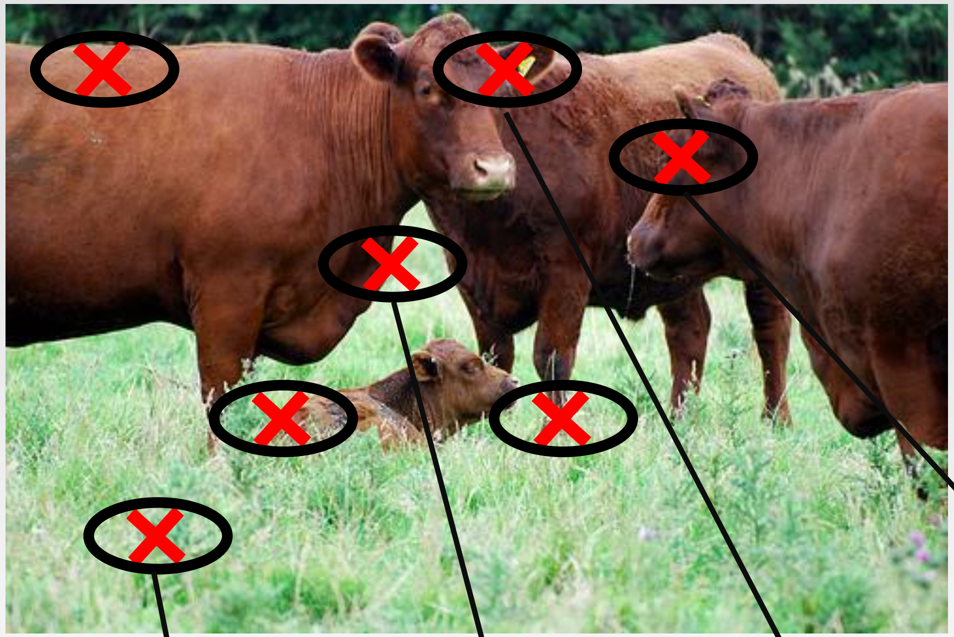
- **A system for Scene Analysis:**

  Segment scenes into constituent object and concept labels

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

Camvid: mi.eng.cam.ac.uk/research/projects/VideoRec

Multinomial Logistic Regression



$$\log p(y=i) \;\square\, \text{®}_0 + \text{®}_1 \,)_1 + \text{®}_2 \,)_2 + \text{®}_3)_3 + \text{®}_4)_4$$

Simple linear model for log-probability

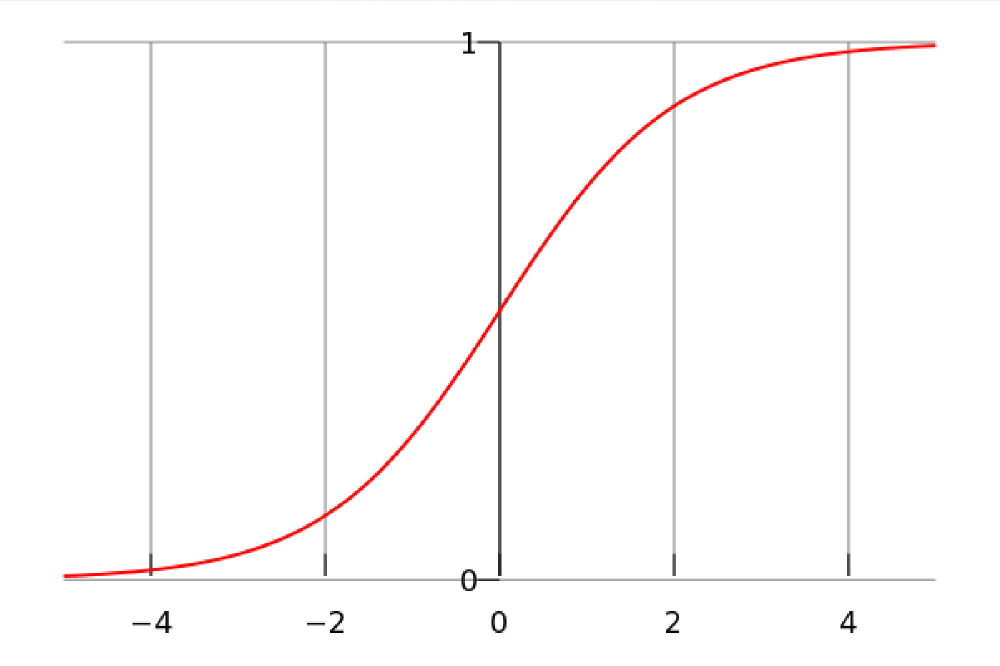
$$\log p(y) \;\square\; \begin{bmatrix} \text{®}_{10} & \text{®}_{11} & \text{®}_{1N} \\ \text{®}_{20} & \text{®}_{21} & \text{®}_{2N} \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} \begin{bmatrix} 1 \\ )_1 \\ )_2 \\ . \\ . \\ )_N \end{bmatrix}$$

Parameters

Multinomial Logistic Regression

log p(y=i) □®0 + ®1)1+ ®2)2 + ®3)3 + ®4)4

$$p(y = i|\beta_i, \Phi) = \frac{\exp \beta_i.\Phi}{1 + \sum_j \exp \beta_j.\Phi}$$



- Supervised learning for β using non-linear least squares
  - -L-BFGS used in this work
  - -Also gives variances of coefficient estimates
- MAP learning with L2-regularization
  - -avoids overfitting and large parameter values

Multinomial Logistic Regression

The Good

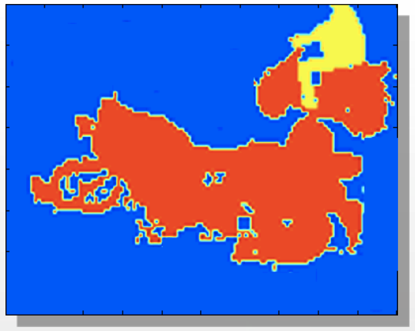
- Fast predictions at runtime
  - -Scales well with number of classes
  - -Labeling probability is available
- Model is stable w.r.t slight changes in training set
- Used widely in biology, sociology, machine learning

The Bad

- Variance of coefficients increases with number of features
- Not suited for large feature spaces
- Sensitive to noise in training data
- Training with large datasets is slow

... and the Beautiful

Random Multinomial Logit!



RML Result

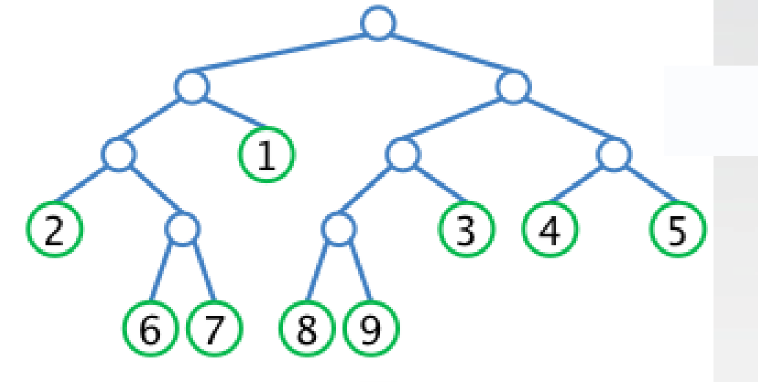
Random Multinomial Logit

QuickTime™ and a
PNG decompressor
are needed to see this picture.

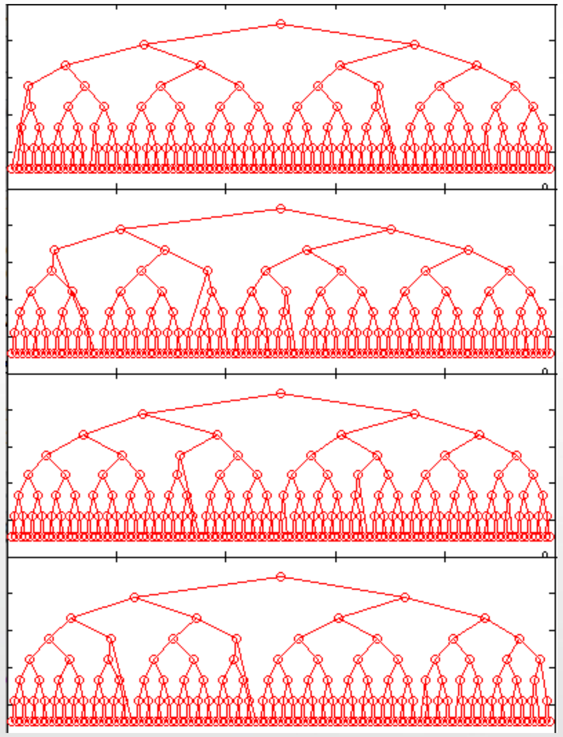
A. Prinzie, D. Van den Poel, “Random forests for multiclass classification: Random Multinomial Logit”, Expert Systems with Applications, 34(3), 2008.

Random Multinomial Logit

Basic idea similar to Random Forests of Decision trees



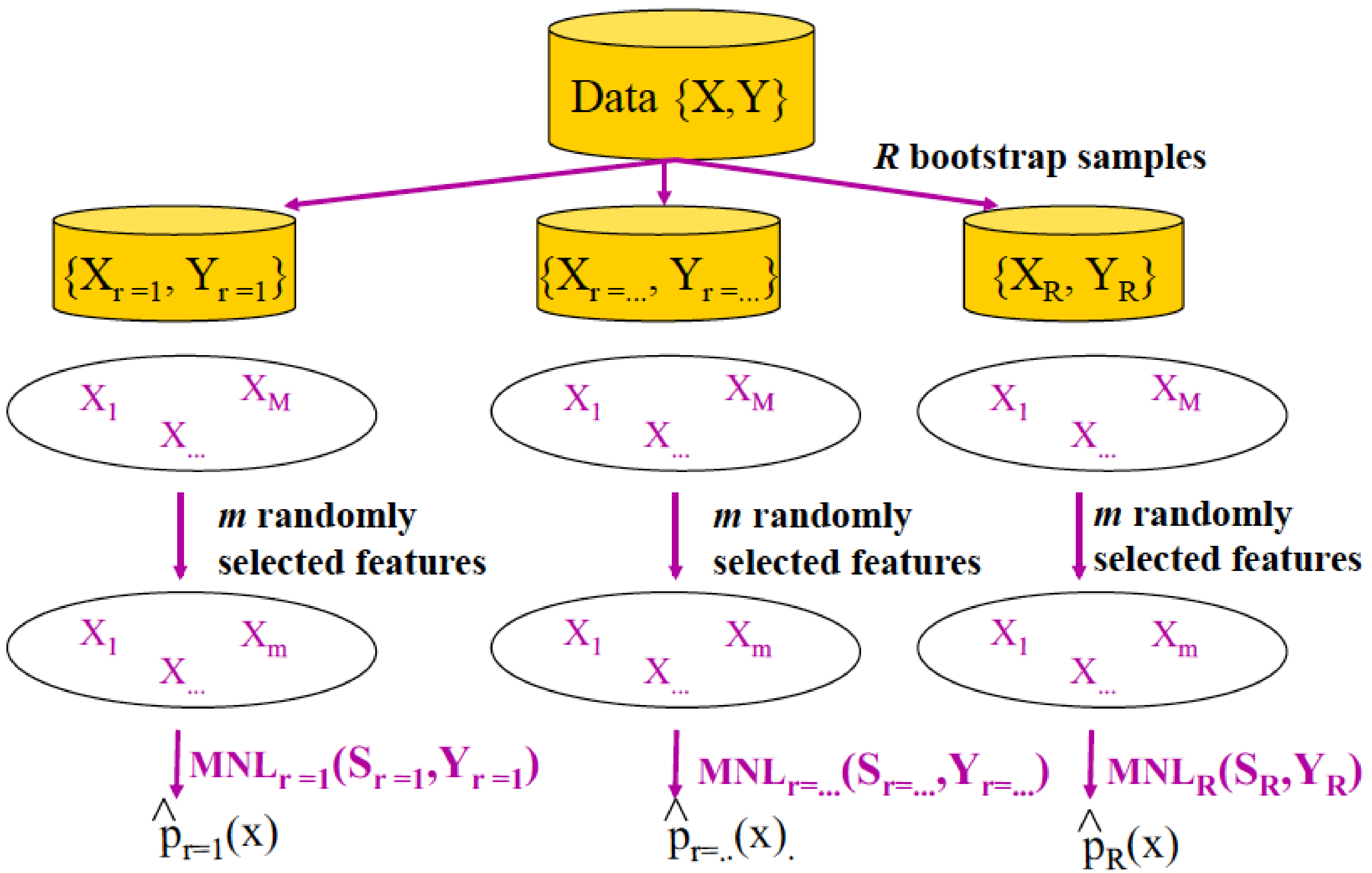
high variance, overfitting, sensitive to noise, unsuitable for large feature spaces



Randomly generated trees

Result obtained by averaging

Random Multinomial Logistic Regression



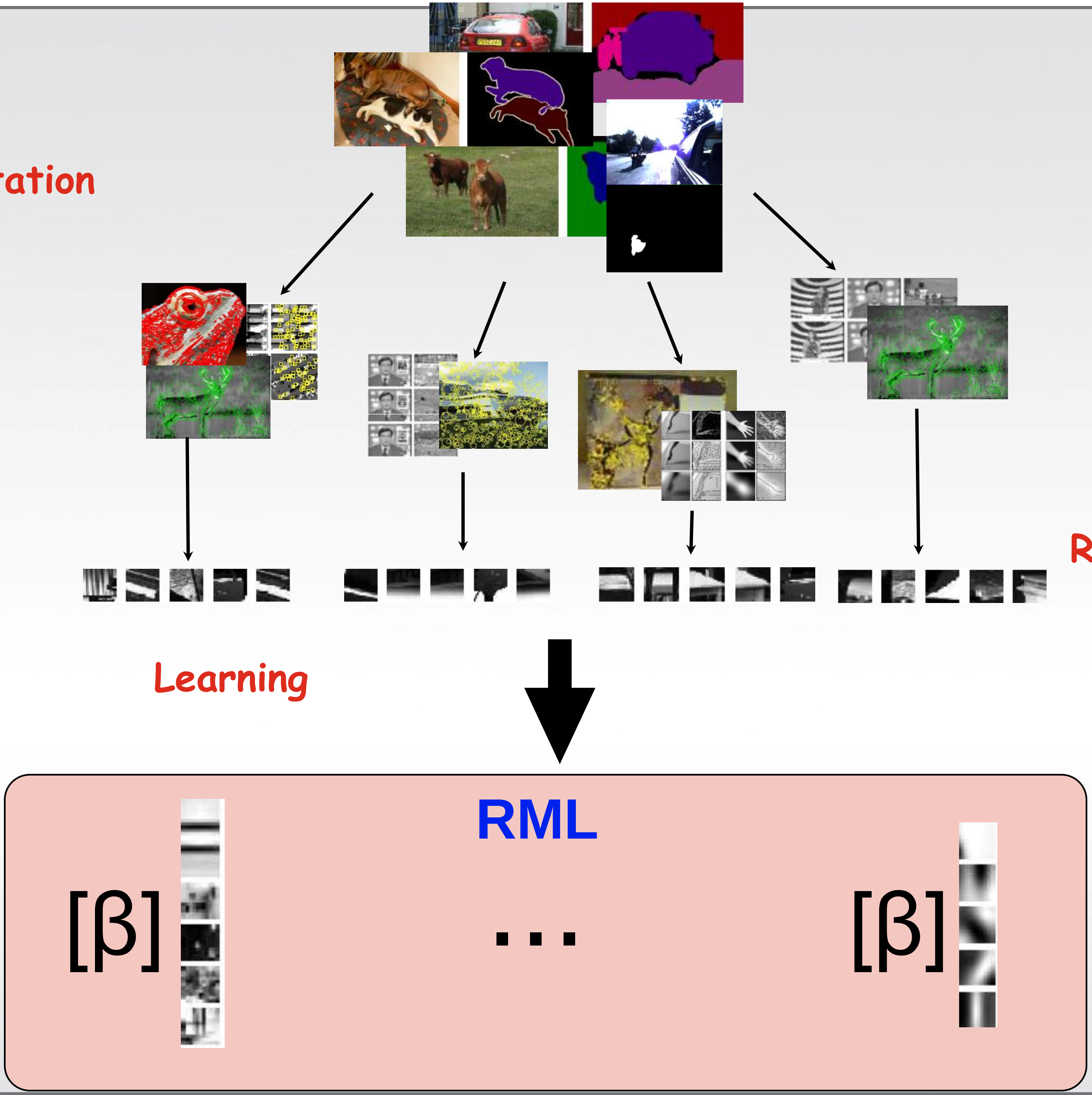
From Prinzie & Van den Poel, 2008

- Use multiple Multinomial Logit models each trained from a different subset of training data
- Randomly selected feature set
- Final prediction is simply the average
- Model bias and prediction variance are both reduced
- Robust to noise and can work in large feature spaces

RML Training

Feature computation

Sampling



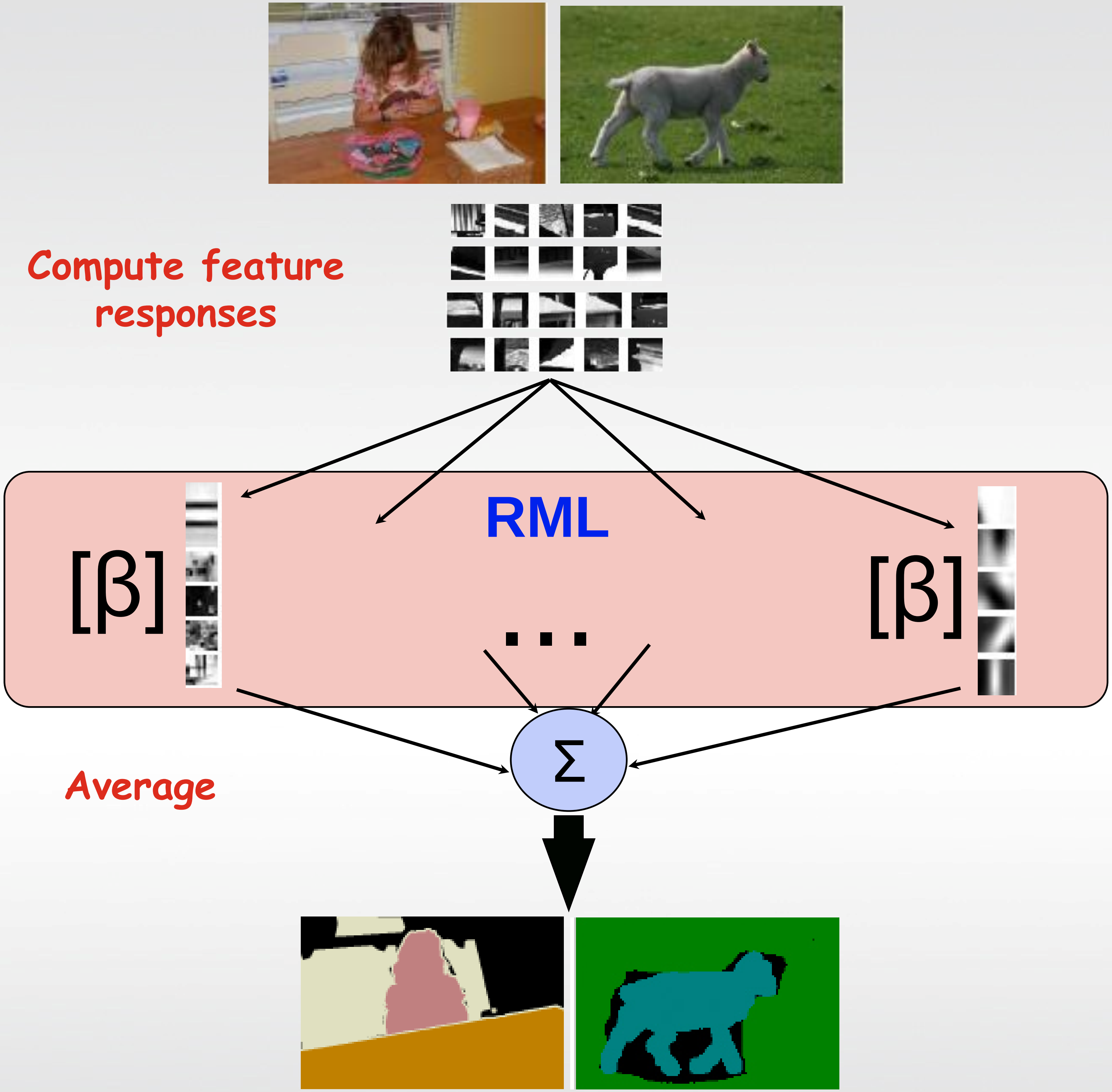
Random feature selection

Learning

RML

[β] … [β]

Classification using RML



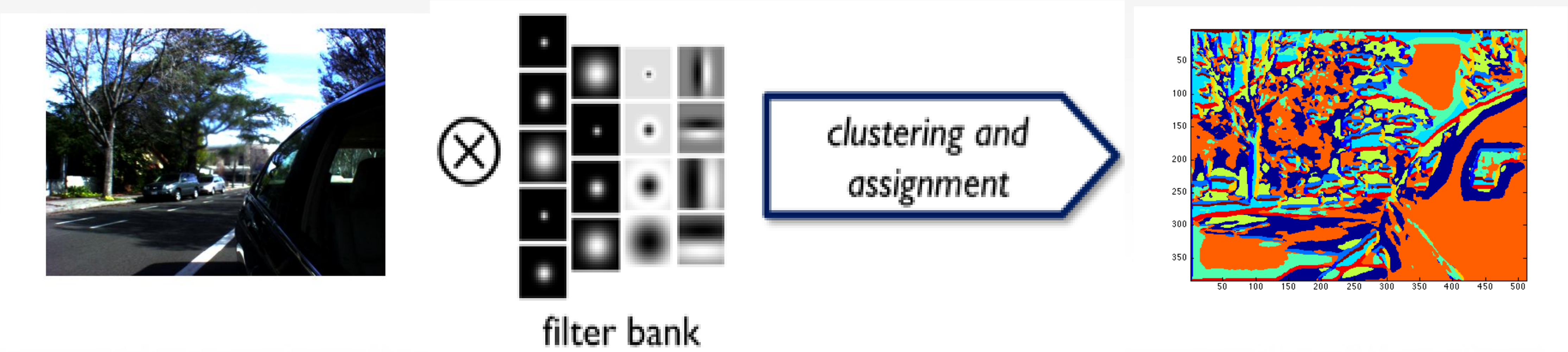
Compute feature responses

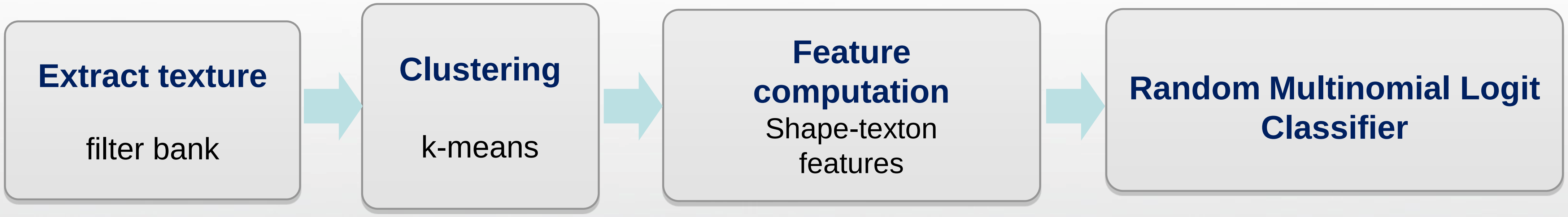
Get model outputs

Average

Scene Segmentation using RML

- RML used as texture-based classifier
- Texture space is discretized into *Textons*
- Leung-Malik filter bank to compute texture
  - -17 filters Gaussian, DoG, LoG
- Shape-texton features used as input to RML





Extract texture

filter bank

Clustering

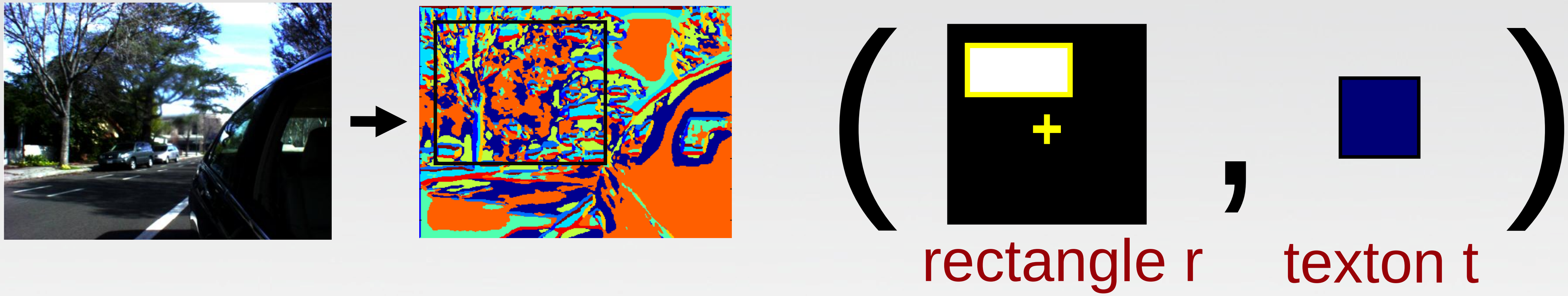
k-means

Feature computation

Shape-texton features

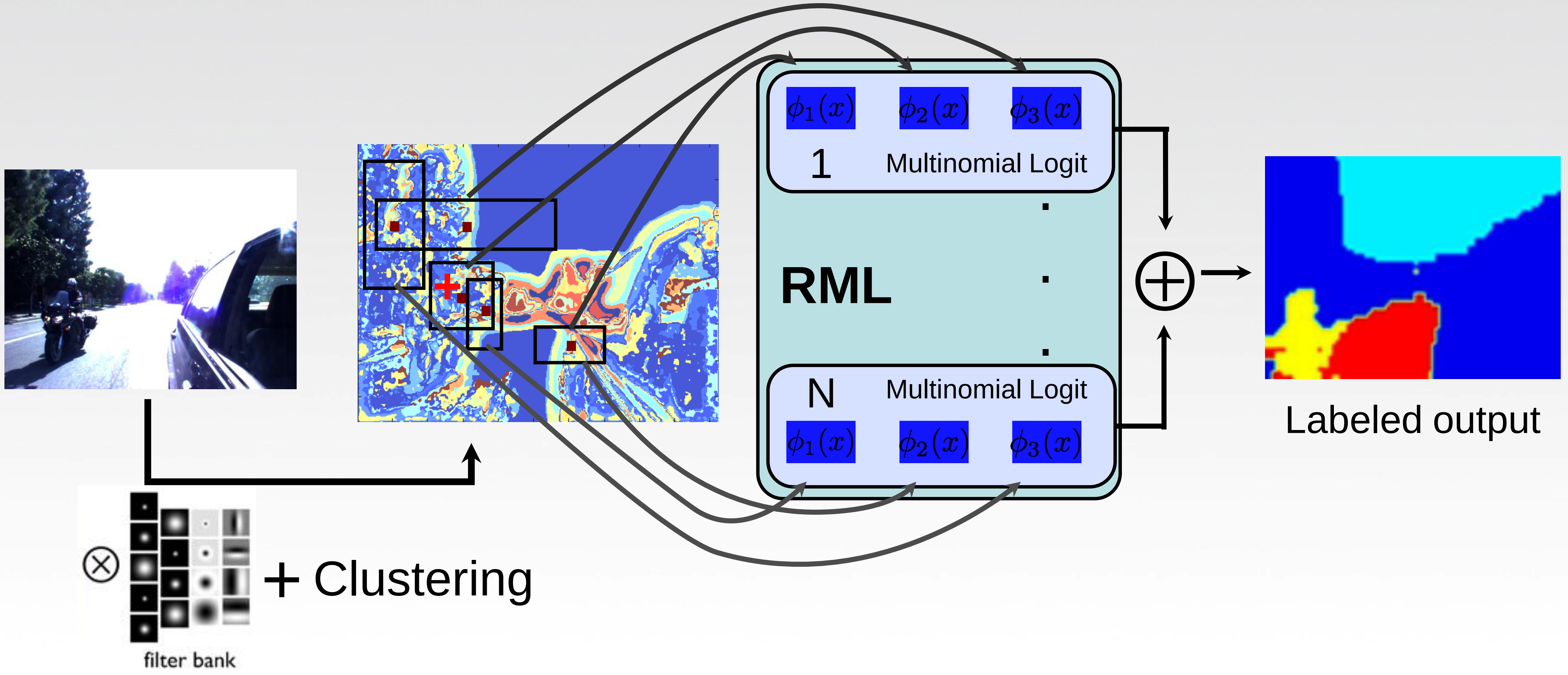
Random Multinomial Logit Classifier

Shape-texton Features



- Introduced in (Shotton et al., ECCV 2006)
- Features computed on texton-mapped images
- Each feature comprises a rectangular region r and a texton t
- The feature measures the proportion of the texton t inside the rectangle r, and is applied to each pixel of the image
- Size of rectangle r and the texton t are generated randomly
- Fast computation using integral images
- Layout and context is captured by rectangular regions

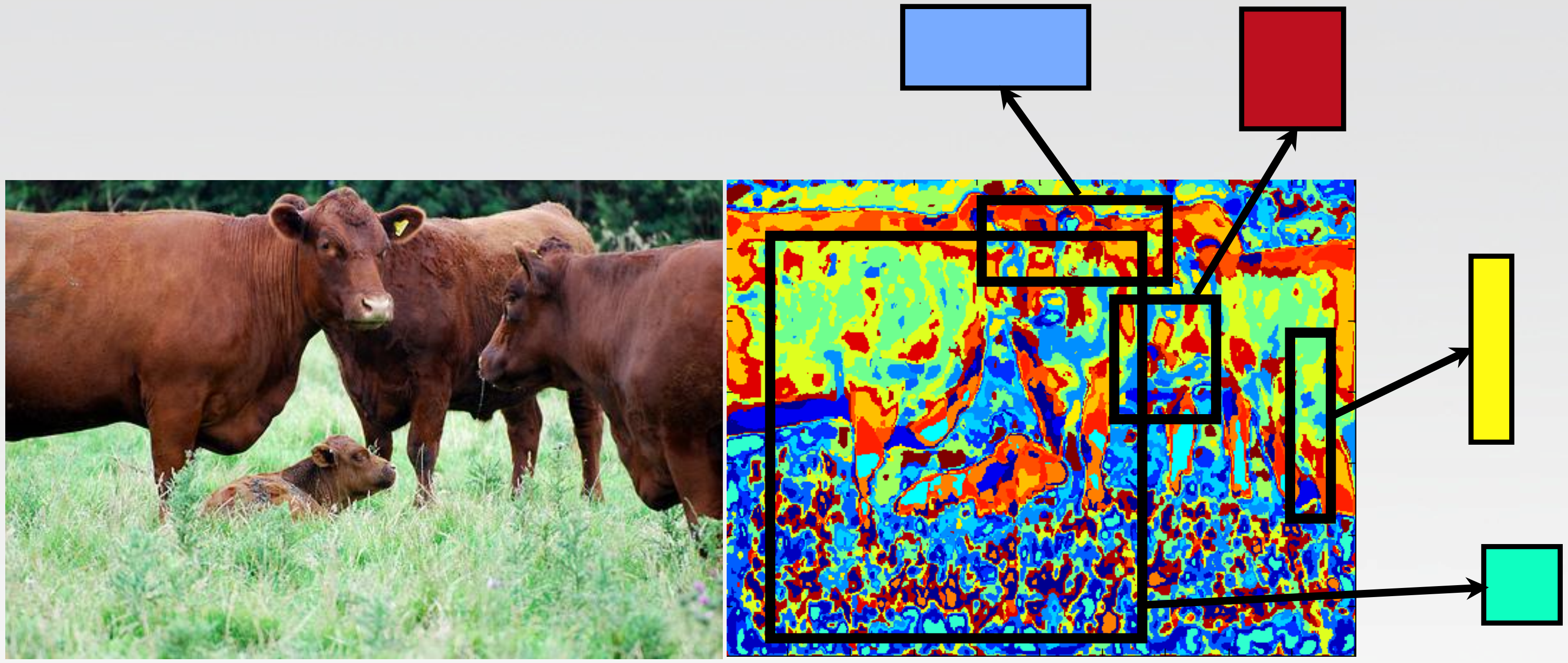
RML for Scene Segmentation

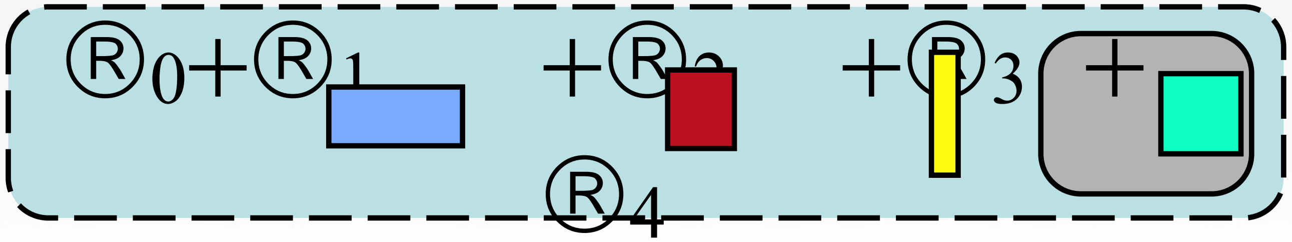


Caveat

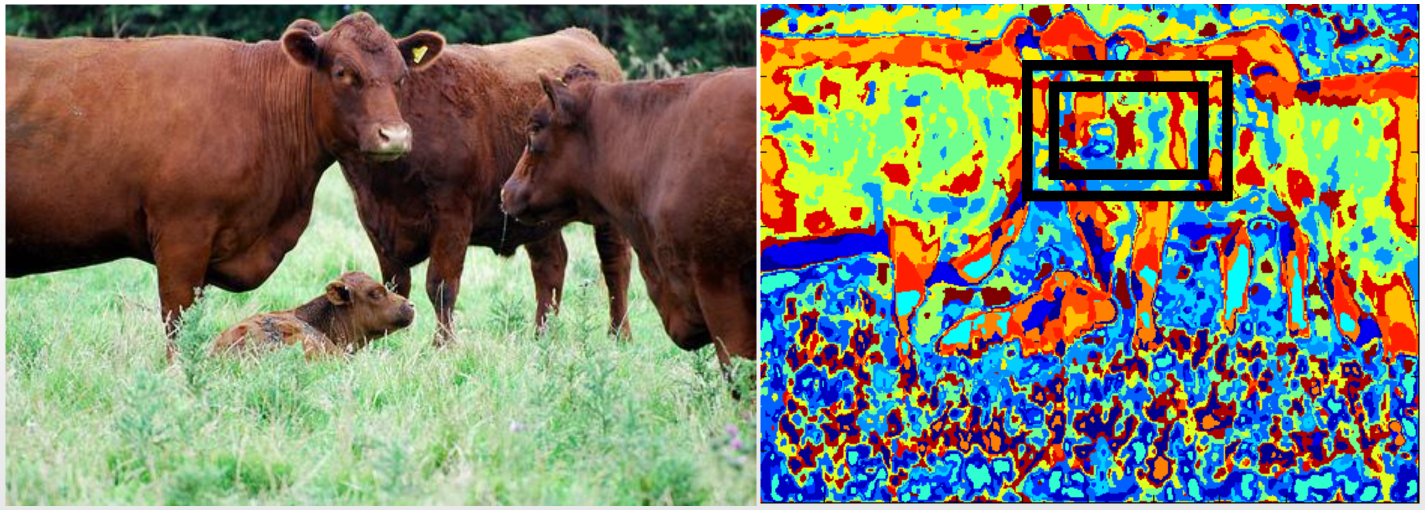
- Feature space is huge
- Many features are useless
- Large number of models will be required to get good results

Need for Feature Selection





Replace statistically insignificant features



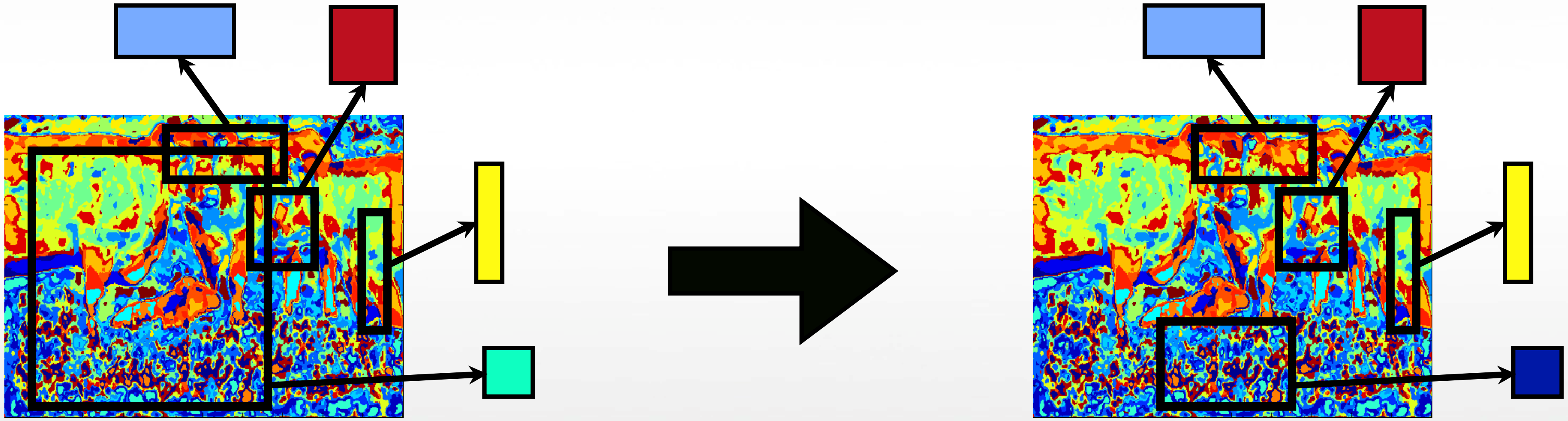
Multi-collinearity

Hard to detect!

Replacing Insignificant features

log p(y=i) □®$_0$ + ®$_1$)1+ ®$_2$)2 + ®$_3$)3 + ®$_4$4

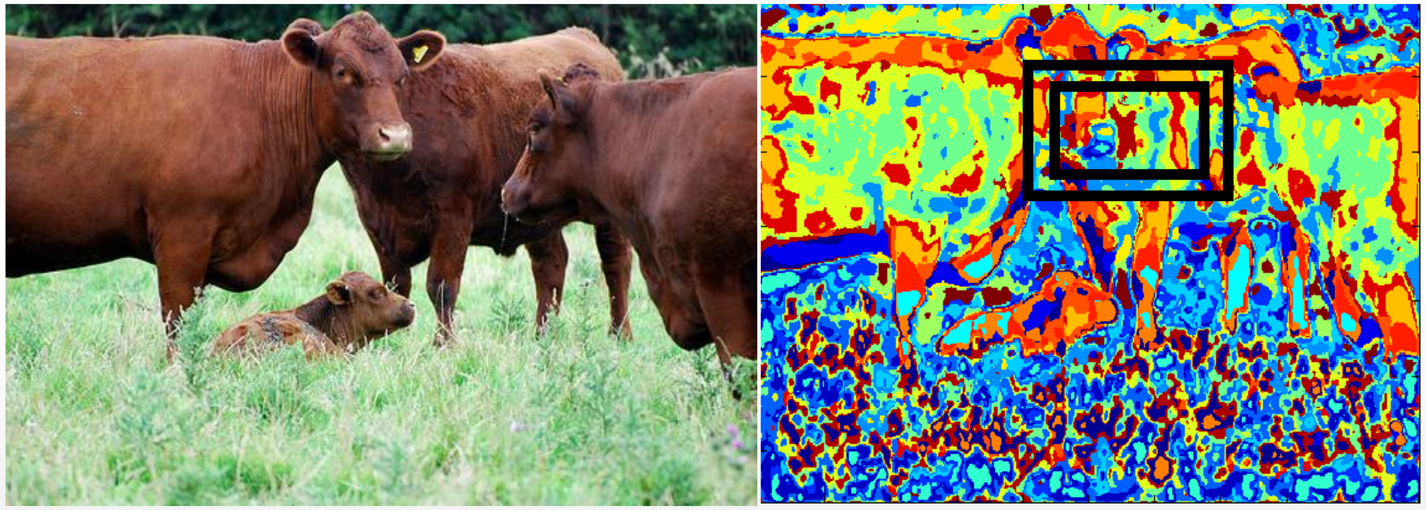
- Statistically insignificant features have “small” β coefficients
- Small - β_i<2*std. dev(β_i)
  - -Variances available from Least-squares learning
- Select new feature randomly to replace the insignificant feature
- Re-learn multinomial logit model



Random Search in Feature Space

- Multi-collinearity is expensive to detect
- Easier to randomly swap features that improve model
- Improvement is quantified by log-likelihood on training data

  -Higher log-likelihood => better feature



Feature Selection Algorithm

- In one round of feature selection do -

- If there are insignificant features
  - replace the feature and re-learn model

Replace insignificant features

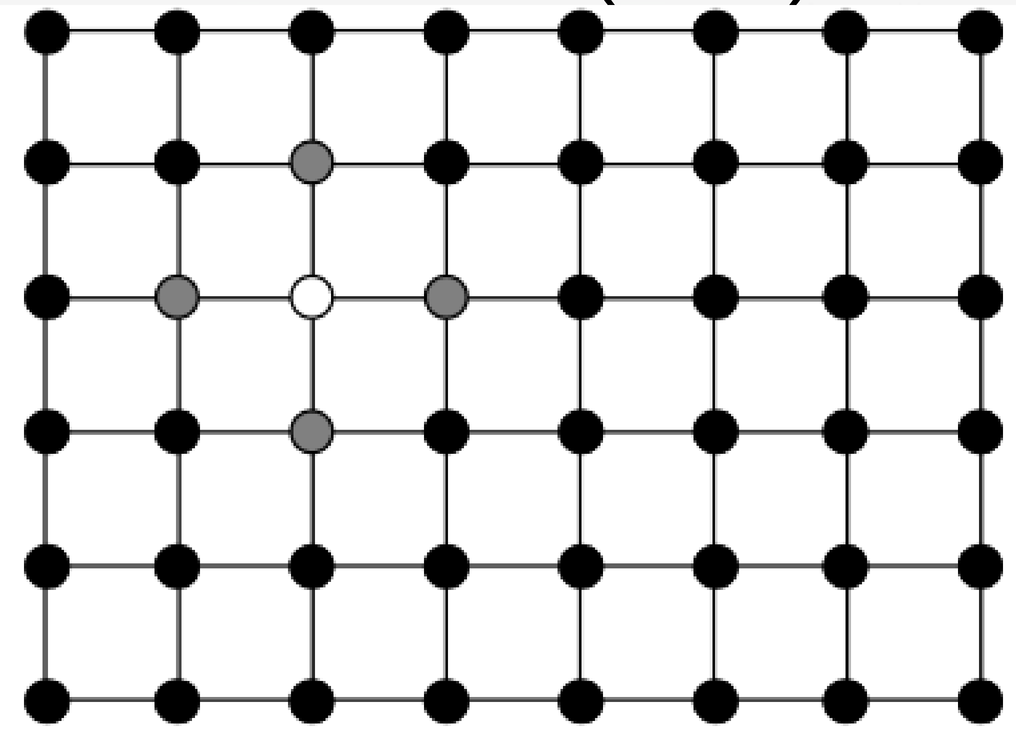
- Else if all features are significant
  - pick a model feature Φ_i at random
  - replace Φ_i with randomly picked feature Φ_j and re-learn model
  - if log-likelihood of new model is greater then keep it, else discard it

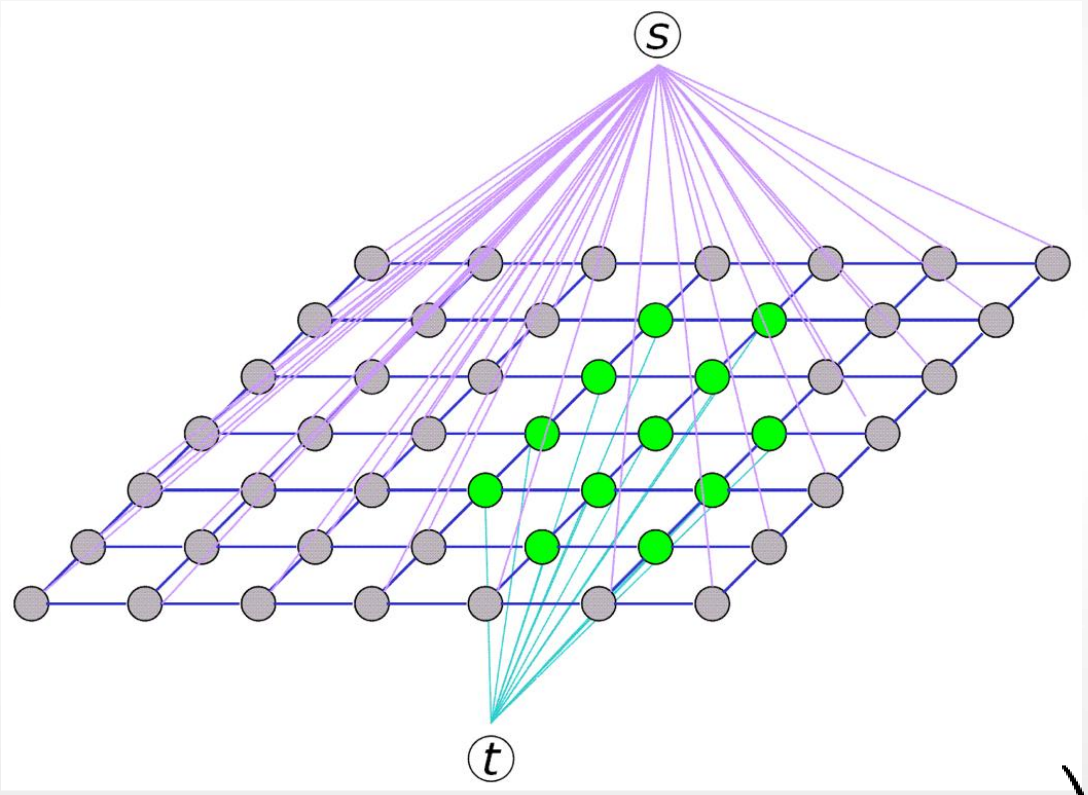
Random feature search

Overall System

- Pixel-wise classification based on texture gives noisy results

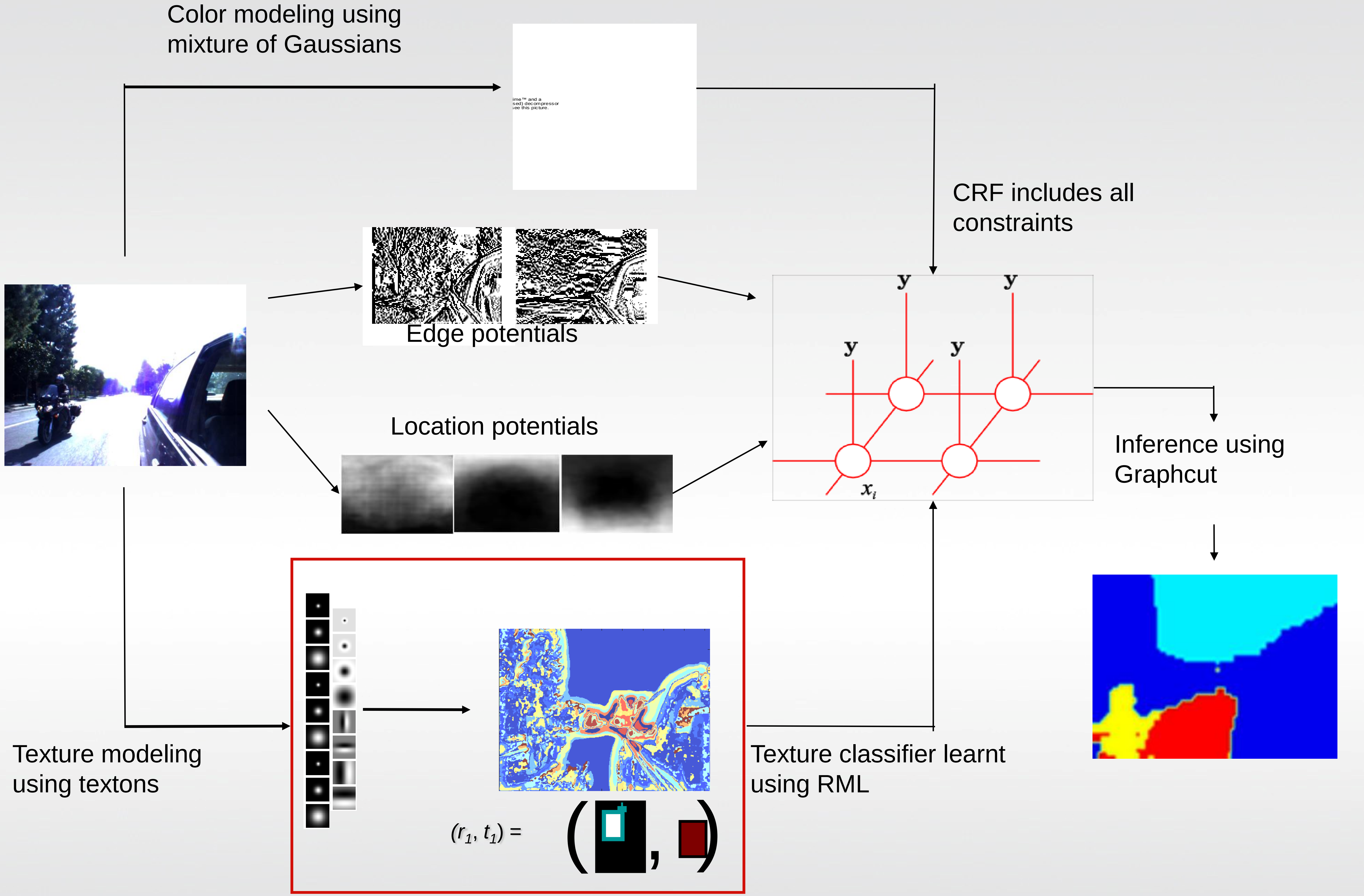
- Include color, location, and edge information in a Conditional Random Field (CRF) - details as in Shotton et al., IJCV 2009





$$p(x) \propto \exp\left(-\sum_i V_{col}(x_i) + V_{tex}(x_i) + V_{loc}(x_i) - \sum_{ij} V_{edge}(x_i, x_j) \right)$$

Overall System



Color modeling using mixture of Gaussians

CRF includes all constraints

Edge potentials

Location potentials

Inference using Graphcut

Texture modeling using textons

Texture classifier learnt using RML

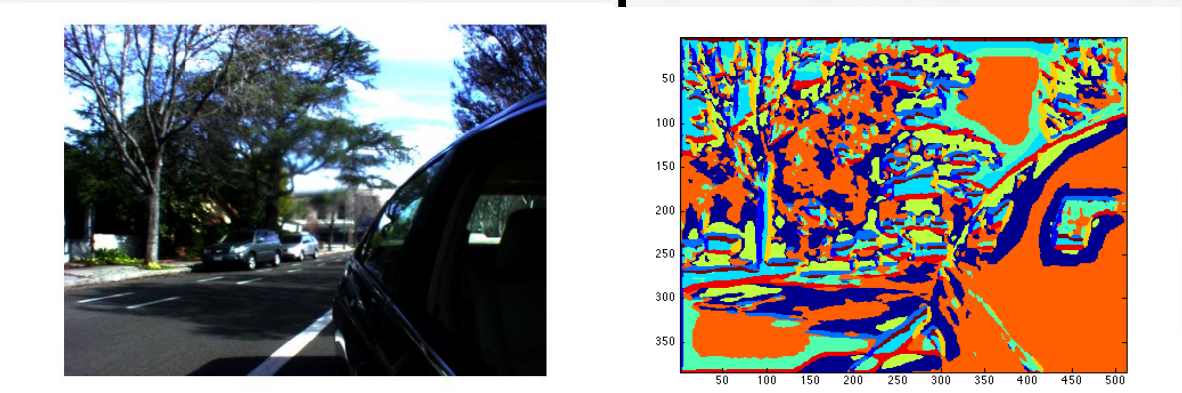
$(r_1, t_1) =$

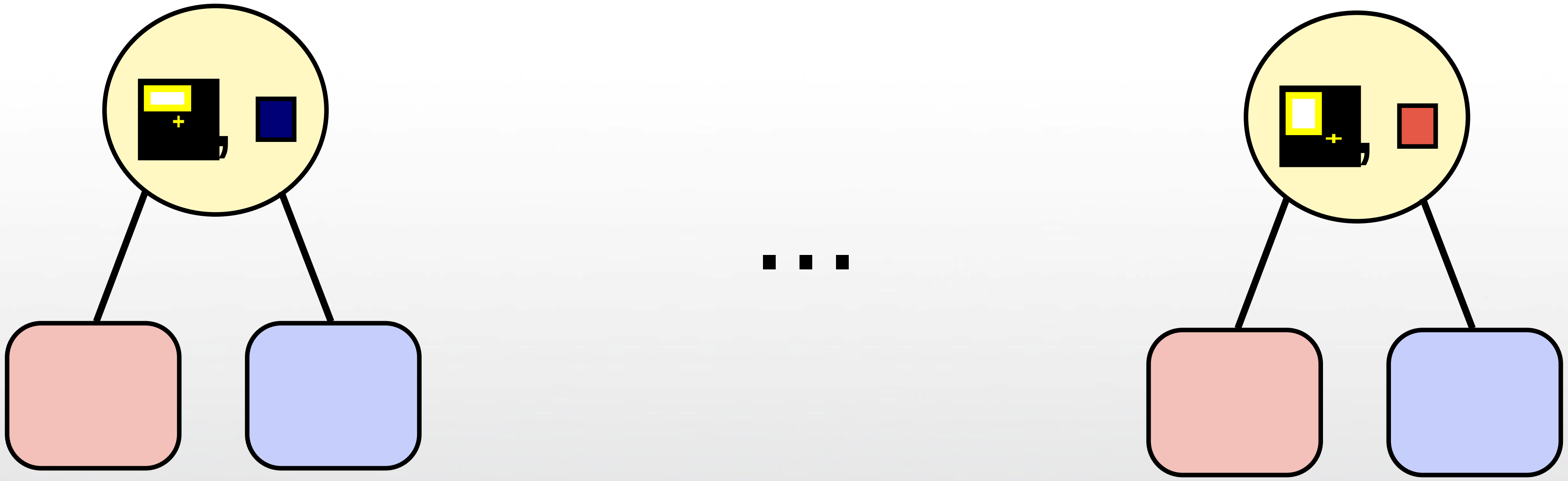
Experiments

- Comparison against random forests and TextonBoost on two datasets

  -implementations in Matlab

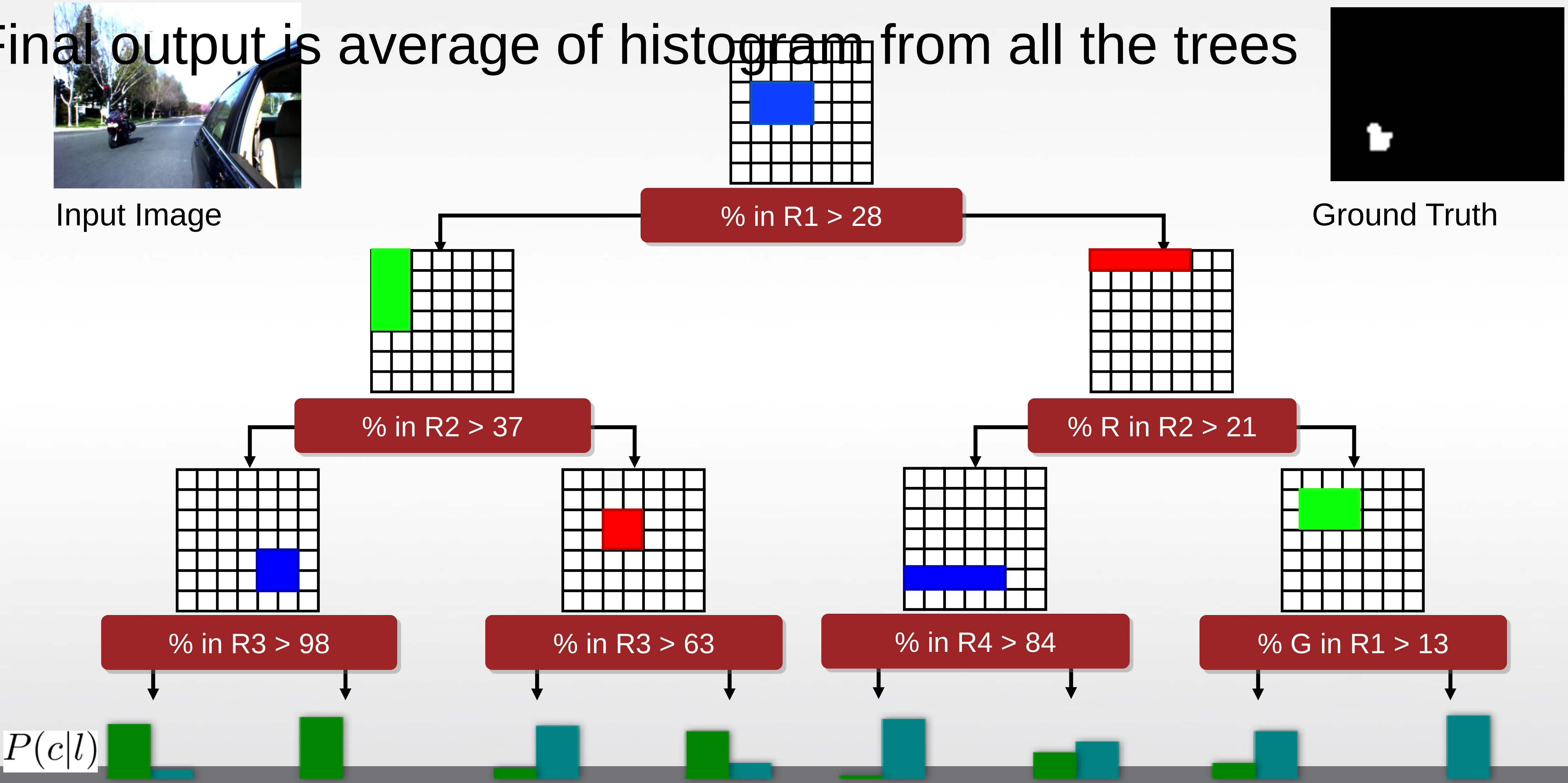
- TextonBoost implemented from (Shotton et al., IJCV 2009)
- Boosting selects decision stumps based on shape-texton features





Random Forests implementation

- Extremely Random Trees (Geurts et al., Machine Learning, April 2006)
- Decision trees have randomly selected shape-texton feature at nodes with random threshold
- Label histogram at each leaf
- Final output is average of histogram from all the trees



Input Image

Ground Truth

% in R1 > 28

% in R2 > 37

% R in R2 > 21

% in R3 > 98

% in R3 > 63

% in R4 > 84

% G in R1 > 13

$P(c|l)$

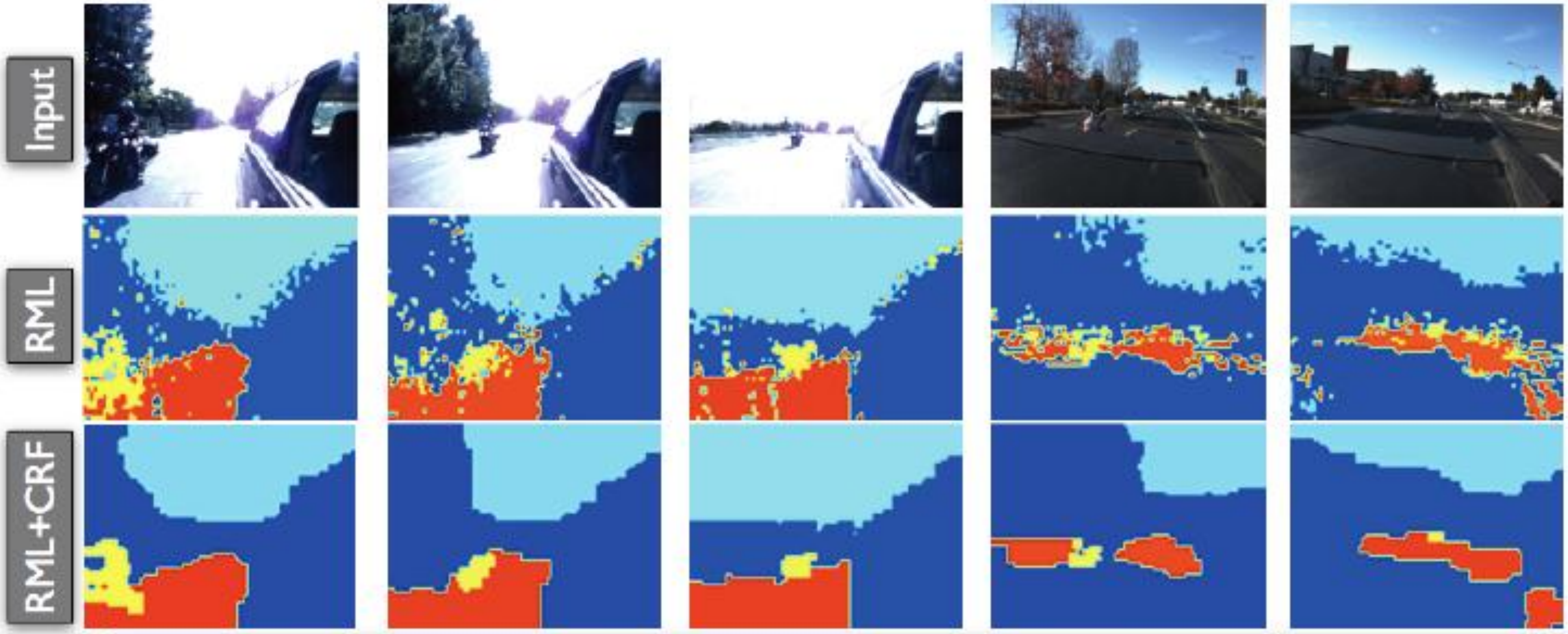
Results - Motorbike video dataset

- Videos from moving vehicles with camera pointed backwards
- 4 categories detected - bike, road, sky, other
- Different types of bikes and road conditions
- 63 labeled frames from 6 sequences, 5800 frames in total
- 15 multinomial logit regressors each with 15 features each

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.

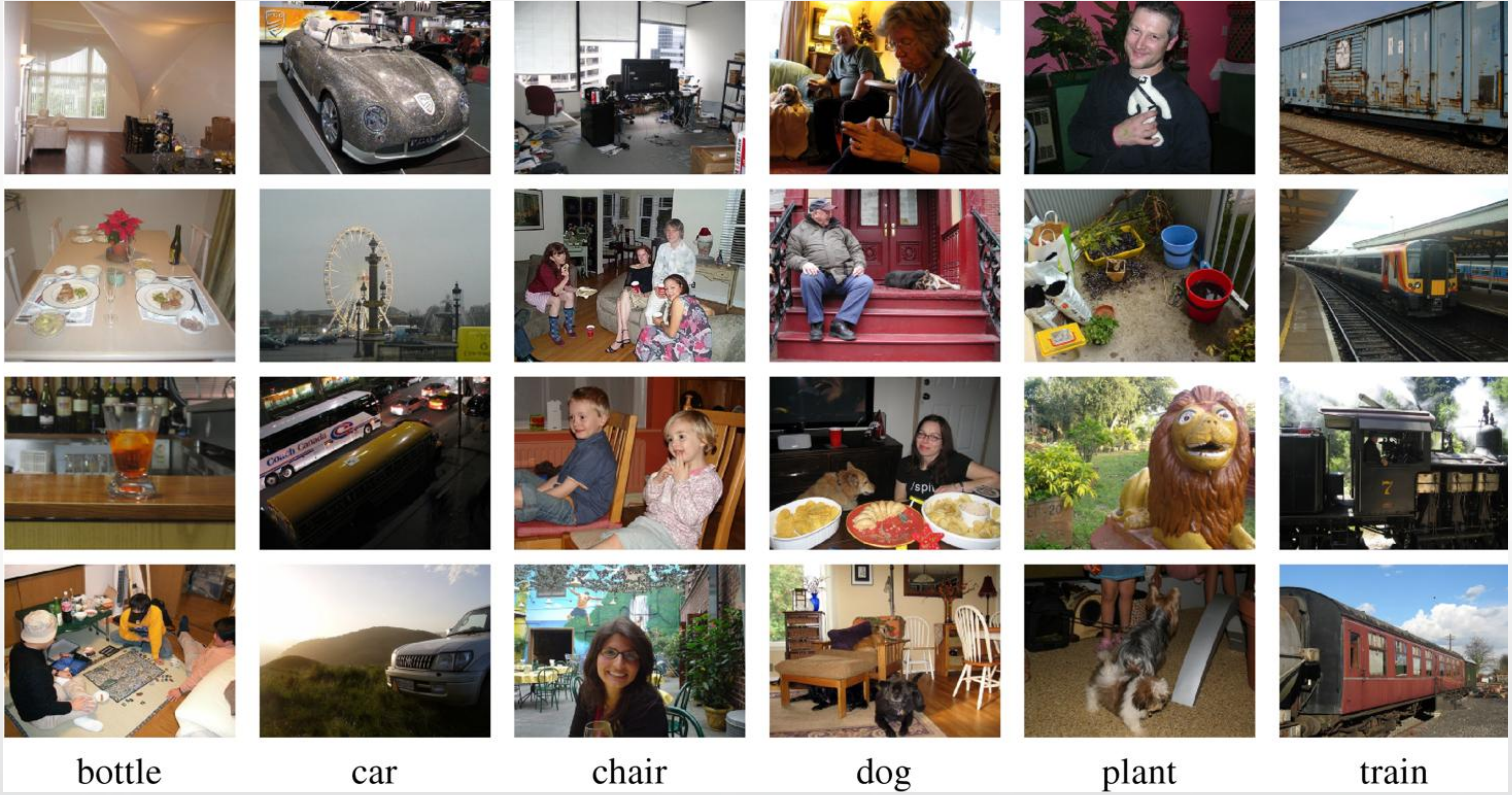
Comparison



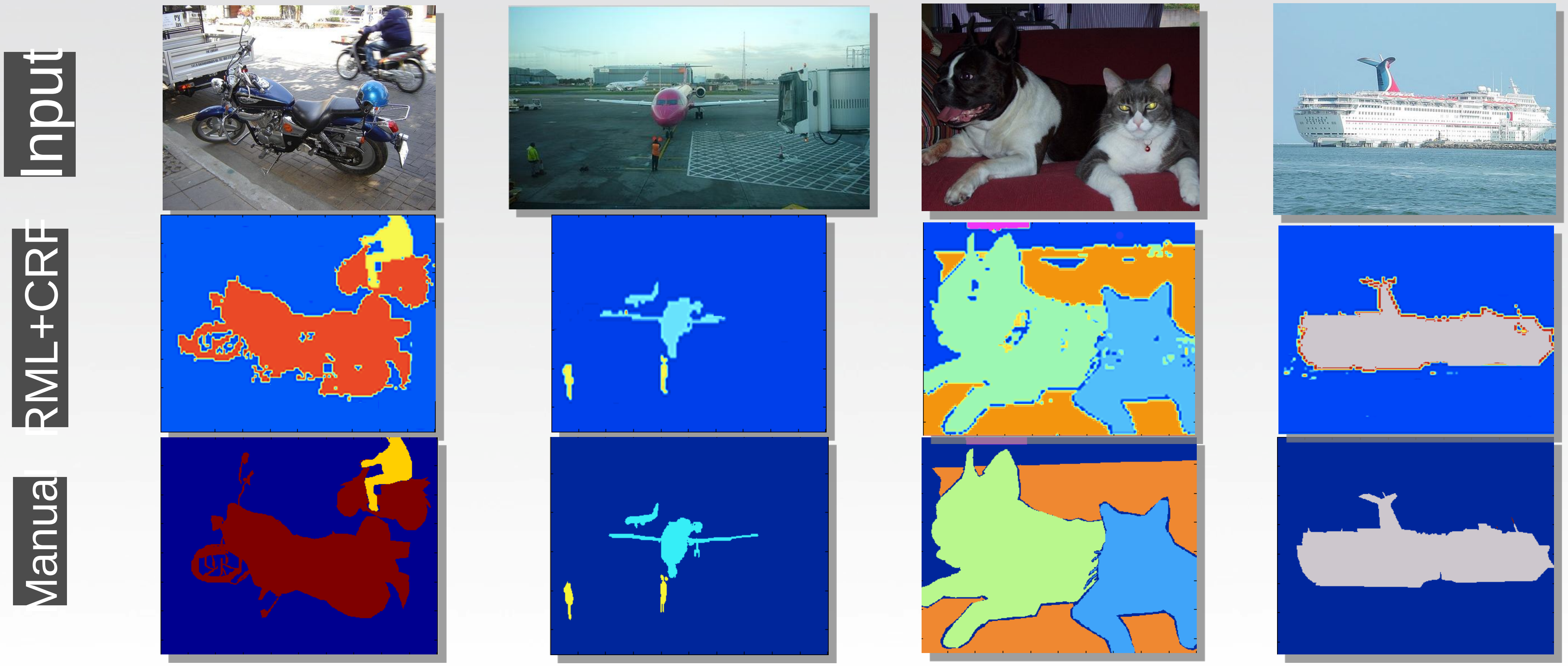
	RML	RML+ Feature selection	RML + Feat. Sel. + CRF	Random Forests	Depth-limited Random Forests	Random Forests + CRF	TextonBoost	TextonBoost + CRF
Overall (%)	73.6	77.1	**82.1**	63.2	49.8	66.7	78.5	81.2
Bike (%)	51.6	53.1	**62.0**	42.7	31.6	45.9	57.8	60.7

VOC 2008 Dataset

- Test dataset for the Pascal VOC 2008 object detection challenge
- 20 classes
- 25 multinomial regressors with 20 features each
- Compared against winner of 2008 challenge (Csurka & Perronin, BMVC 2008)



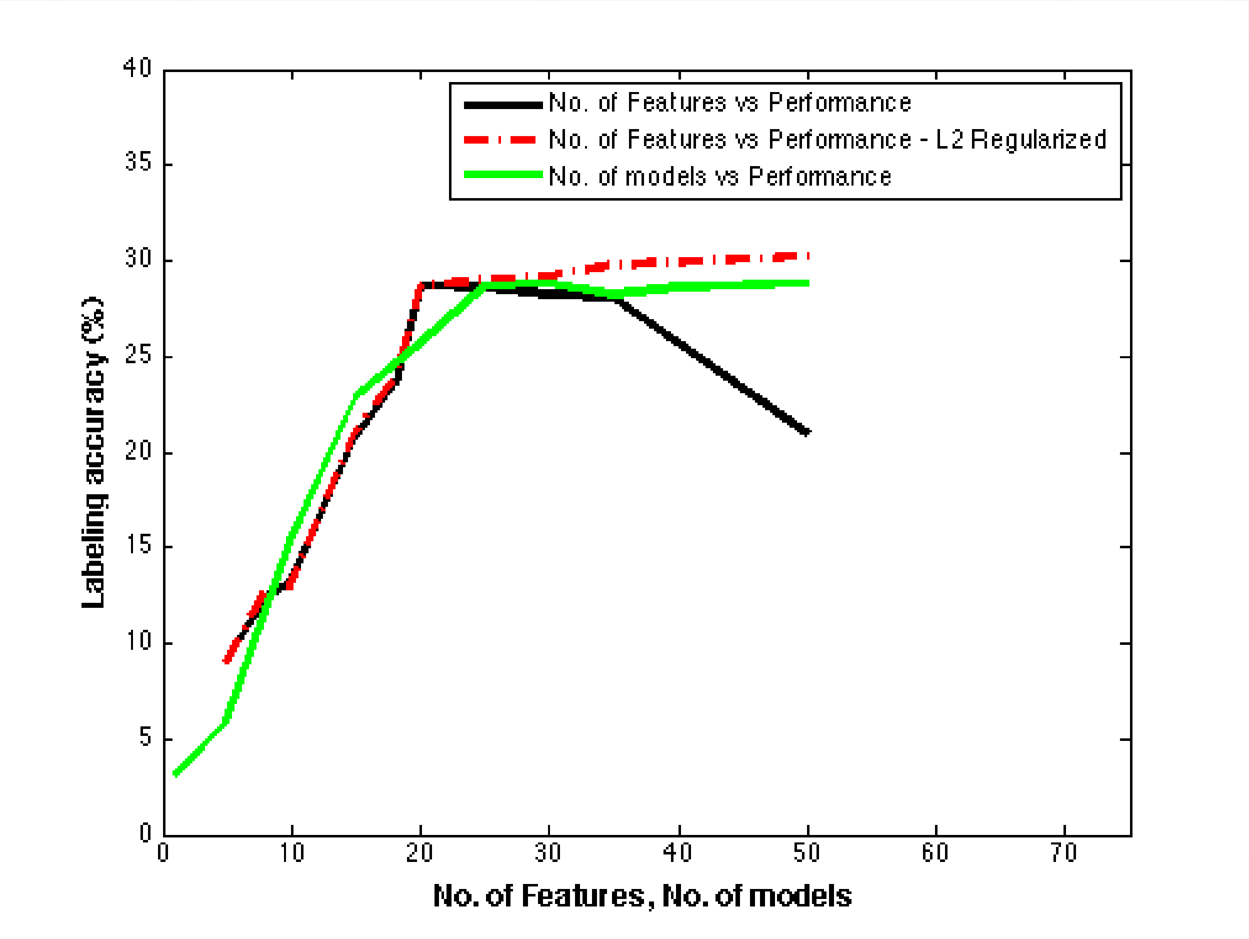
VOC 2008 Dataset



XRCE_Seg	25.8	15.7	19.2	21.6	17.2	27.3	25.5	24.2	7.9	25.4	9.9	17.8	23.3	34.0	28.8	23.2	32.1	14.9	25.9	37.3	75.9
RML+CRF	**31.2**	**20.1**	16.7	**27.2**	**22.6**	**41.2**	**29.3**	22.2	3.2	**36.2**	4.8	12.8	**34.8**	**43.5**	26.0	**23.8**	**39.8**	11.9	**34.1**	**47.7**	73.0

Trends

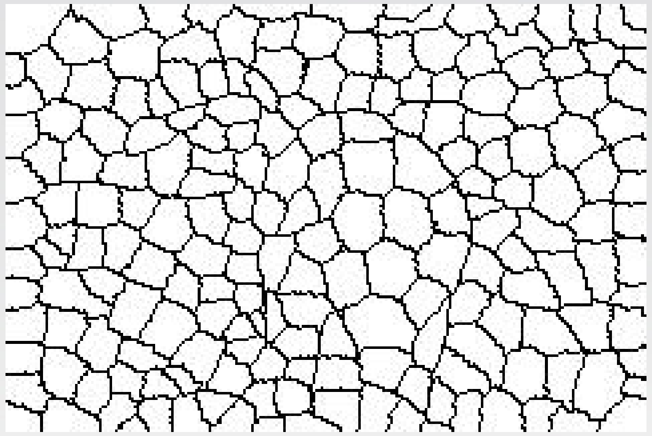
- Performance levels off with increasing number of features
- Regularization is essential with large number of features
- Performance also levels off with number of models



- Runtime on motorbike dataset :
  - RML - 0.12 sec/frame
  - Random forests - 4.1 sec/frame
  - TextonBoost - 6.03 sec/frame

Future

- Sparse multinomial logit
  - L1 regularization
  - No need for feature selection
  - Slow!
- Temporal constraints - optic flow, label tracking
- Superpixels and region statistics
- Shape models etc



TIFF
TIFF
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

The End

QuickTime™ and a
decompressor
are needed to see this picture.

QuickTime™ and a
decompressor
are needed to see this picture.