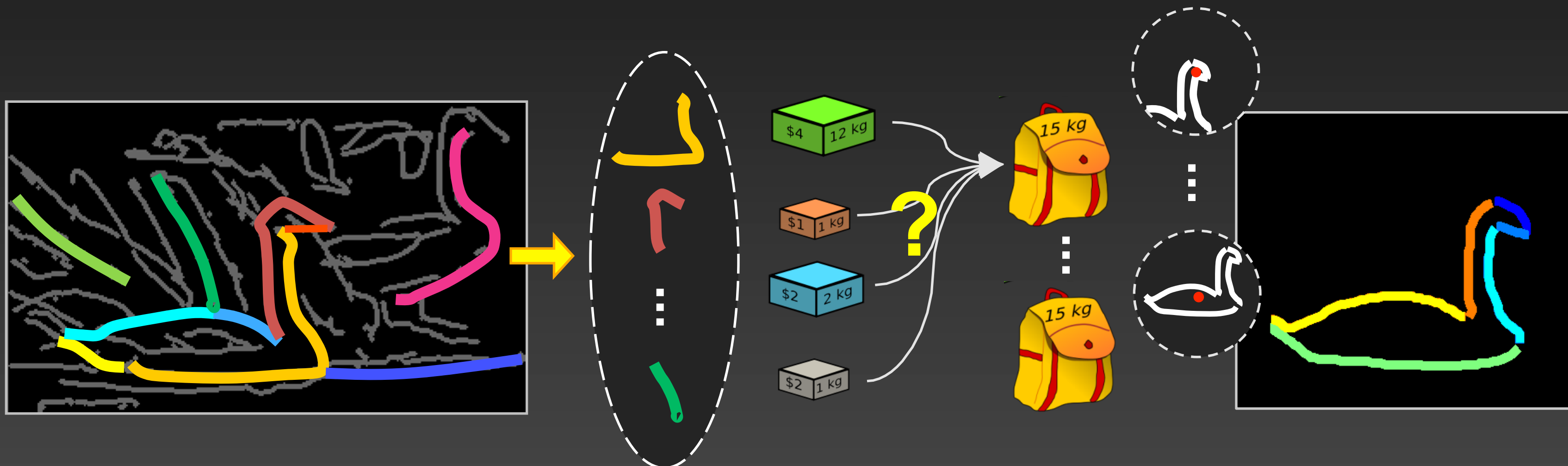
Shape Packing with Contours and Segments



Jianbo Shi

Joint work with Praveen Srinivasan, Qihui Zhu

Penn

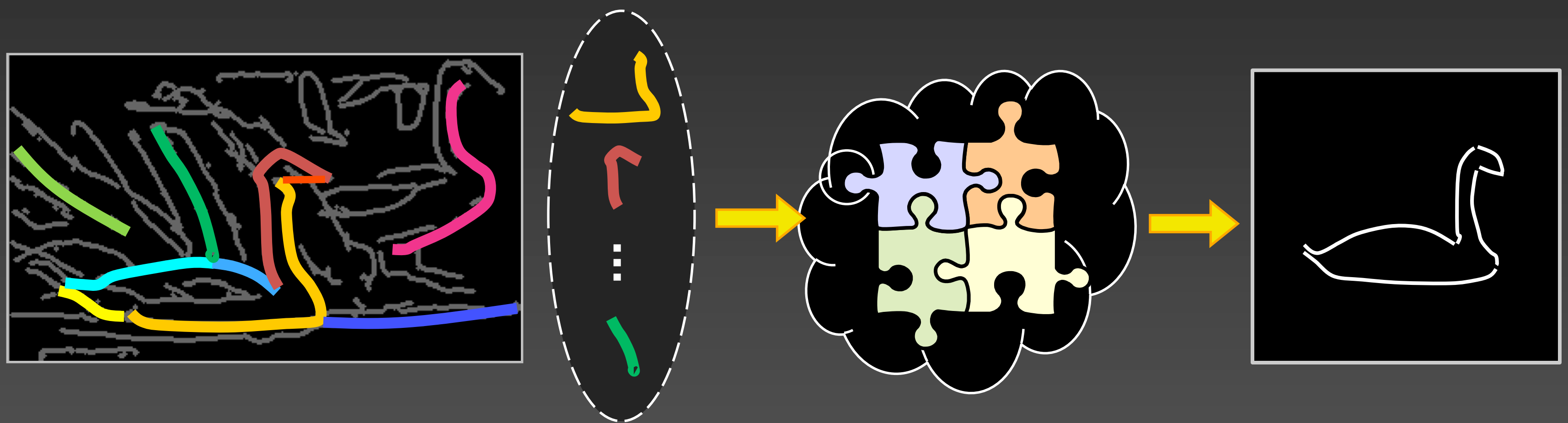
Whole is different from sum of its parts

Shape perception = Many-to-one Packing

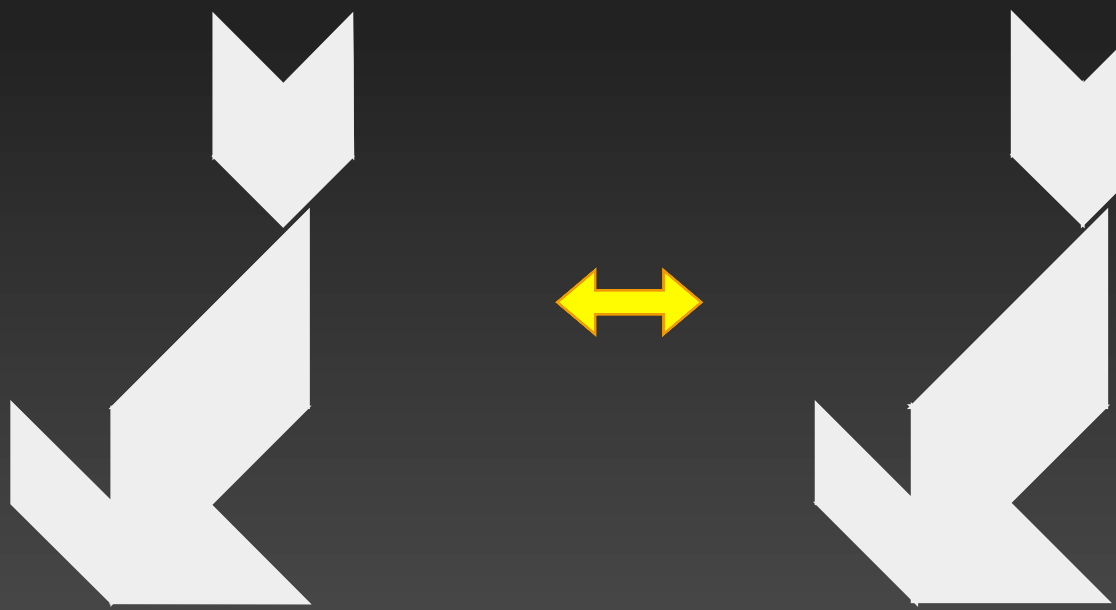
Think of an image made of contours, or segments…

Q: Can we **pack** a set of contours into a recognizable shape?

A: Similar to a **jigsaw puzzle.**

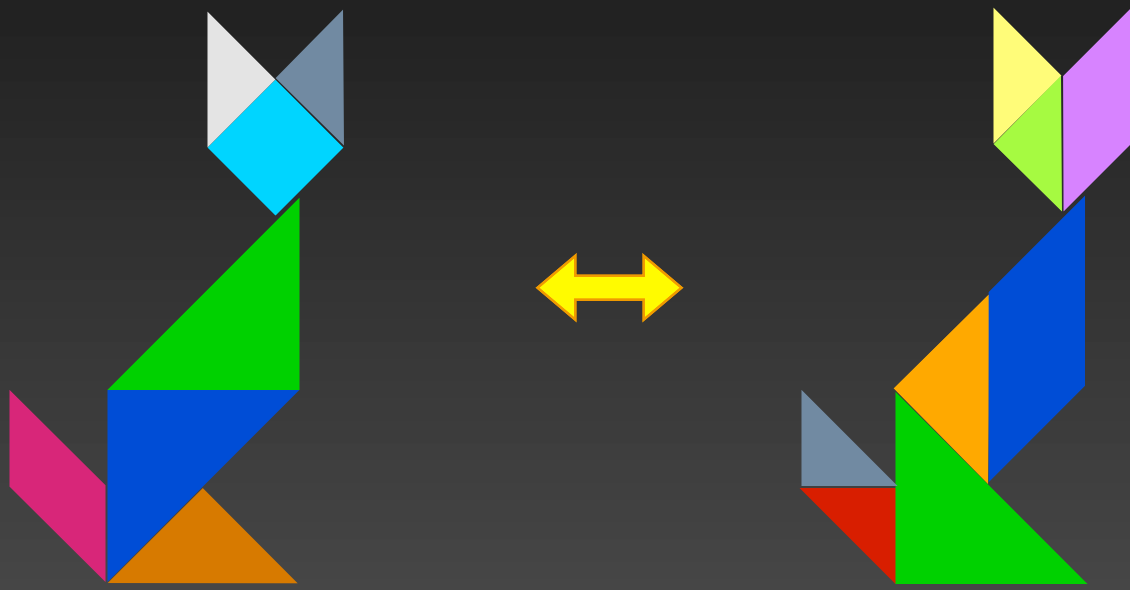


Jigsaw Puzzle



Same shape

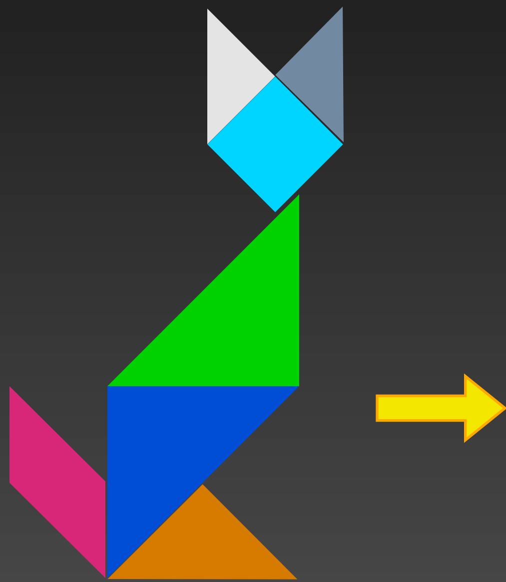
Challenge 1: Fragmentations



No **one-to-one** correspondences of jigsaws!

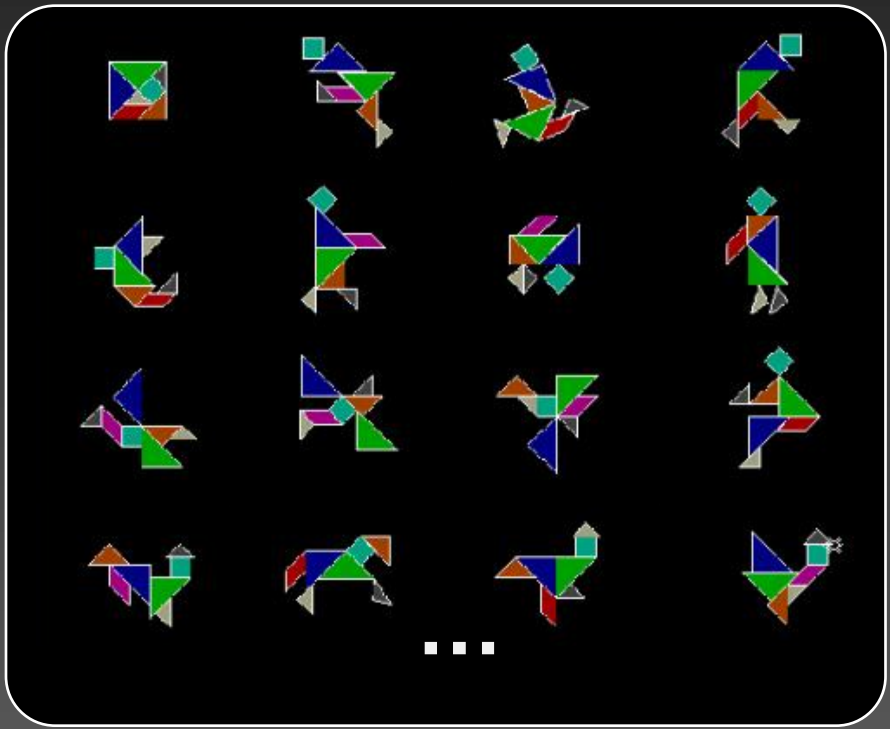
Tangram (*Qi Qiao Ban*), a Chinese jigsaw puzzle game originated in 17th century

Challenge: Exponential Search

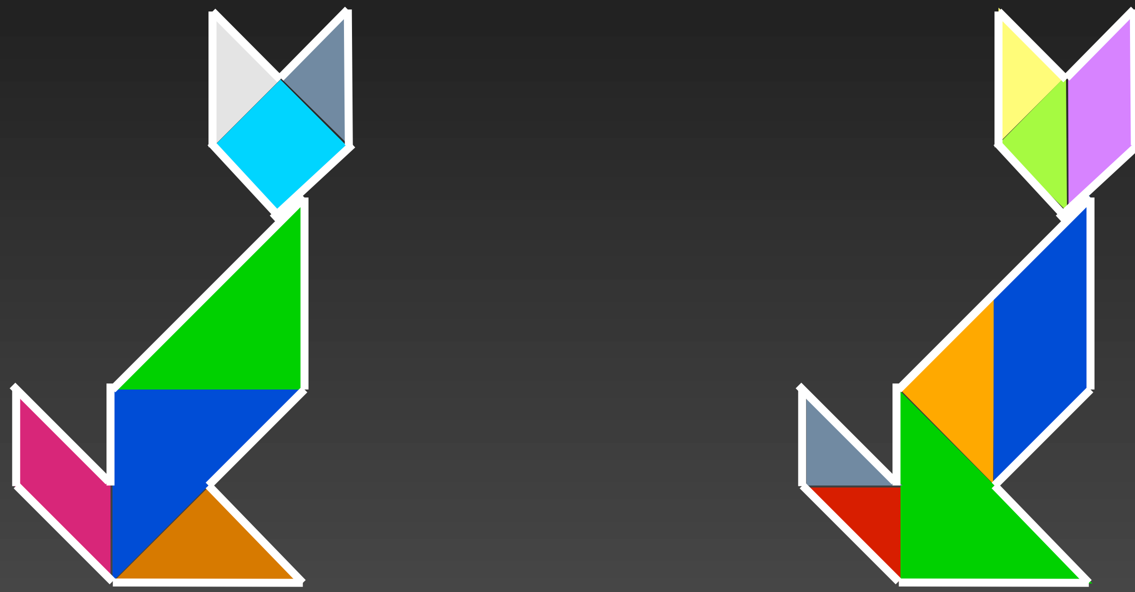


combinations can be exponential!

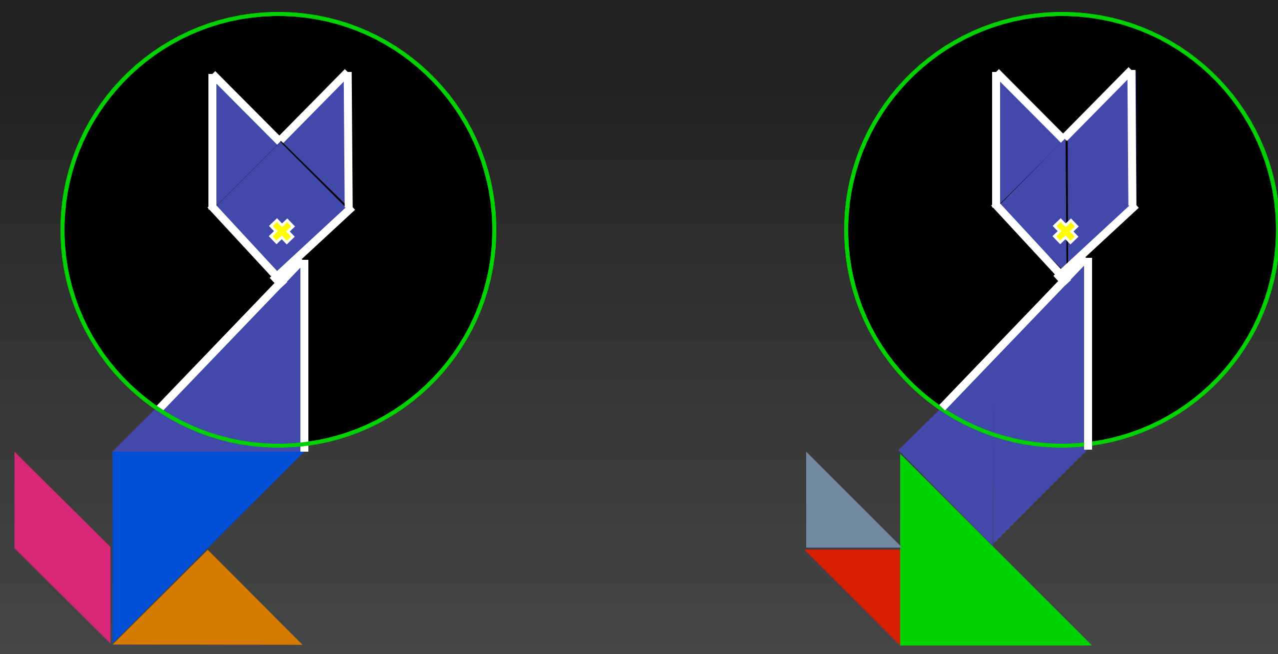
7 pcs → over 1600 shapes…



How Can We Match Shape As A Whole?

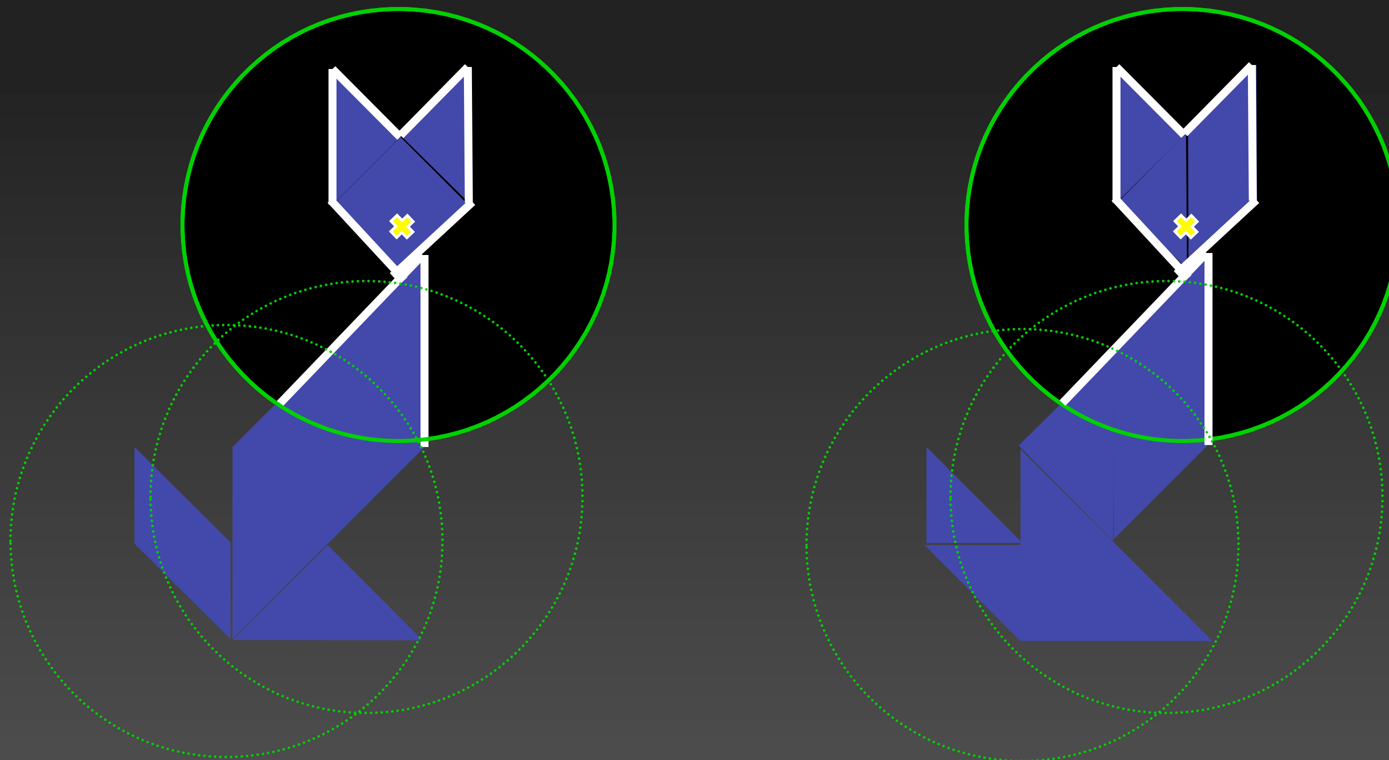


Holistic Shape Matching



Pack all jigsaws within a large neighbourhood at some control points…

Holistic Shape Matching

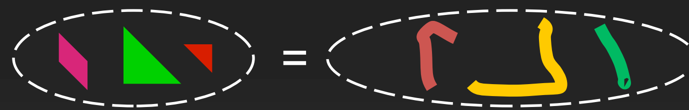


With more control points: same shape regardless of fragmentation!

Many-to-one Contour Matching

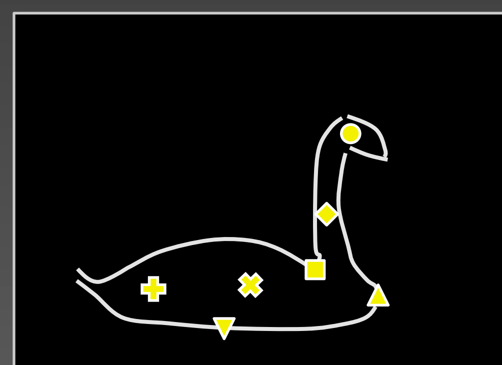
Back to the contour matching problem:

jigsaws = contours



- **Holistic shape matching between contour sets centered at control points**
- **Control point correspondence**

Image contours



Model contours

Context Selective Shape Features

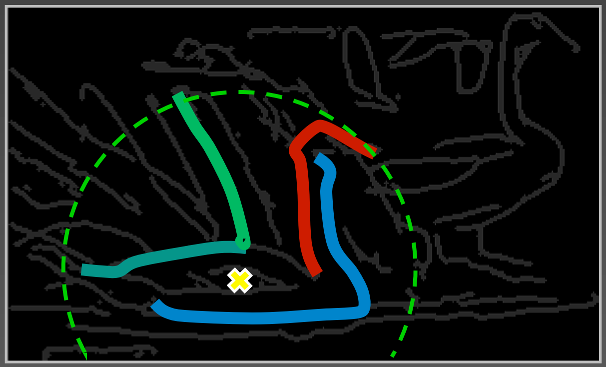
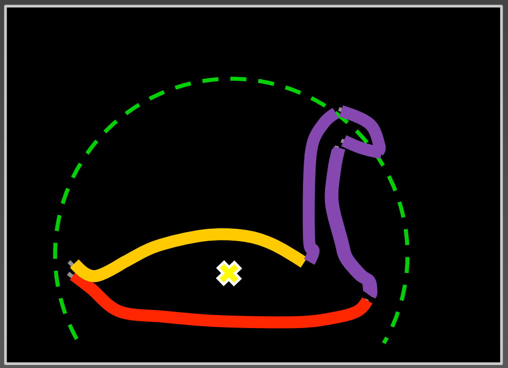


Image contours



Model contours

Context Selective Shape Features

Shape contexts

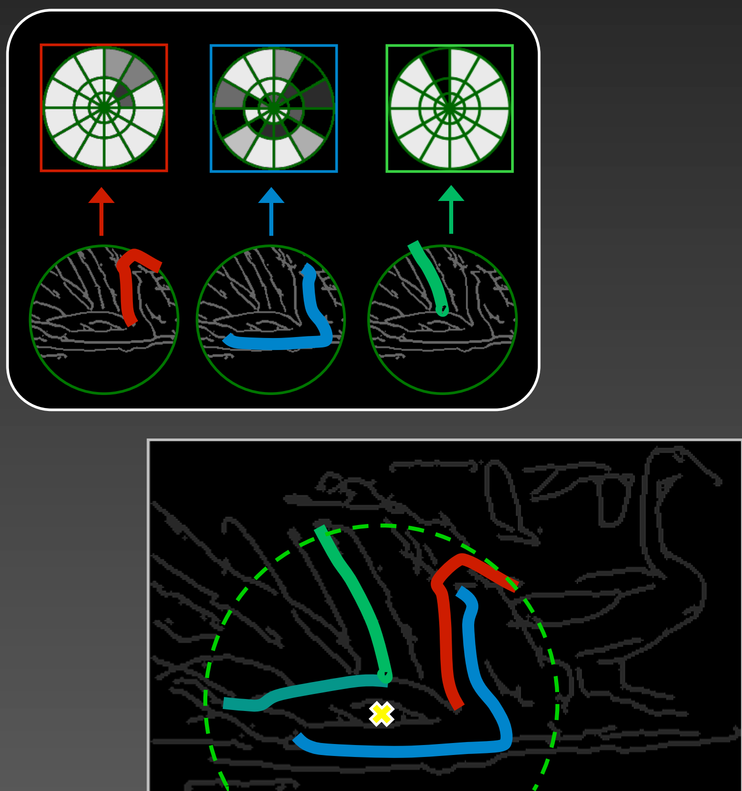
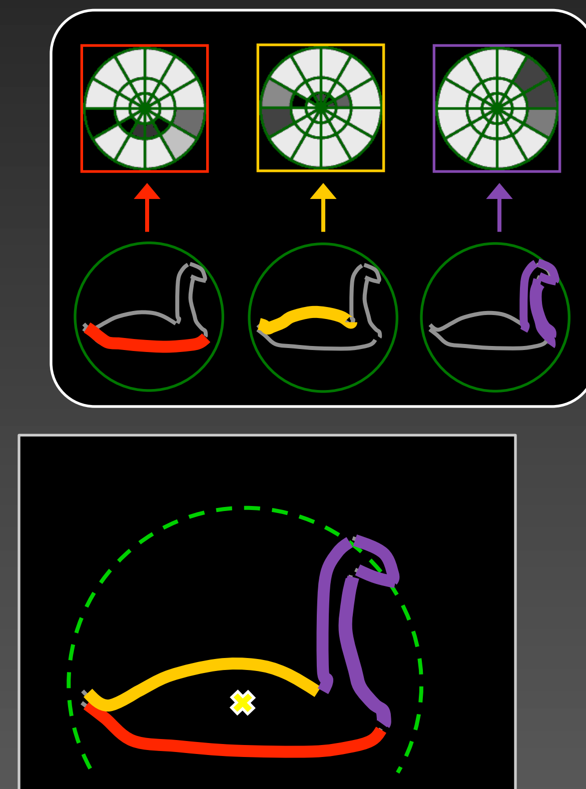


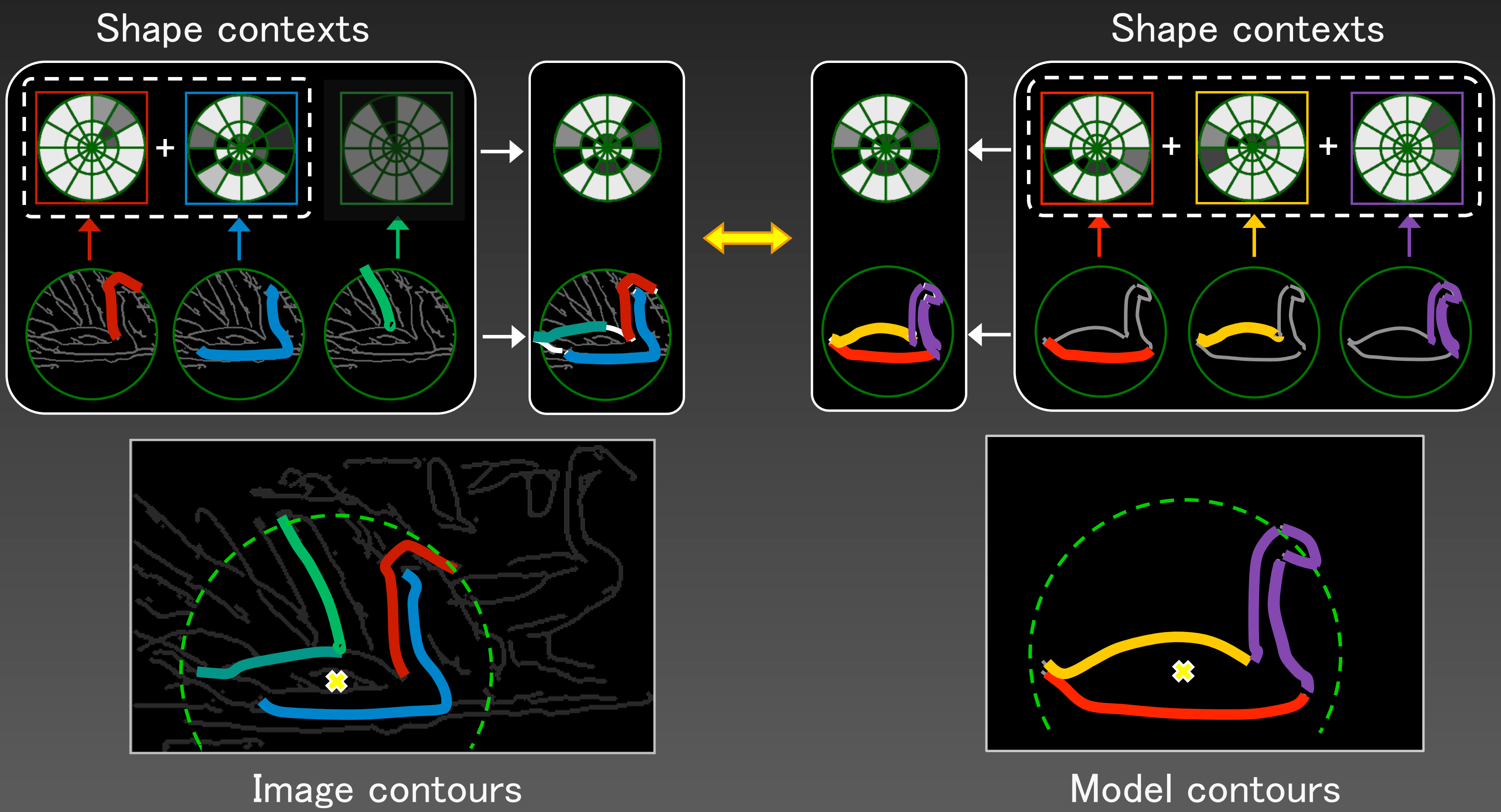
Image contours

Shape contexts



Model contours

Context Selective Shape Features



A compact feature depending on figure/ground contour selection

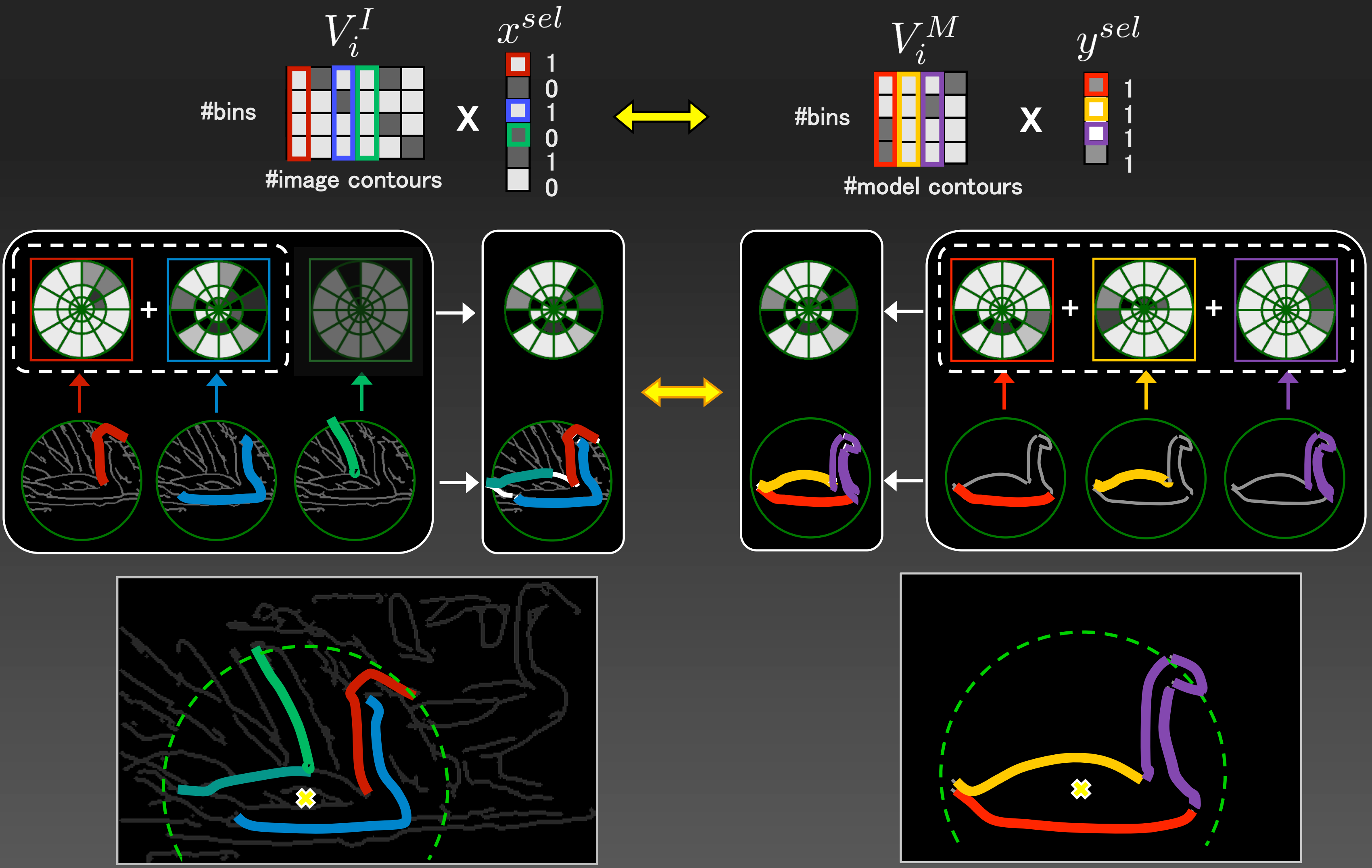


Image contours

Model contours

Shape dissimilarity $D_{ij}(V_i^I \cdot x^{sel}, V_j^M \cdot y^{sel}) = \|V_i^I \cdot x^{sel} - V_j^M \cdot y^{sel}\|_1$

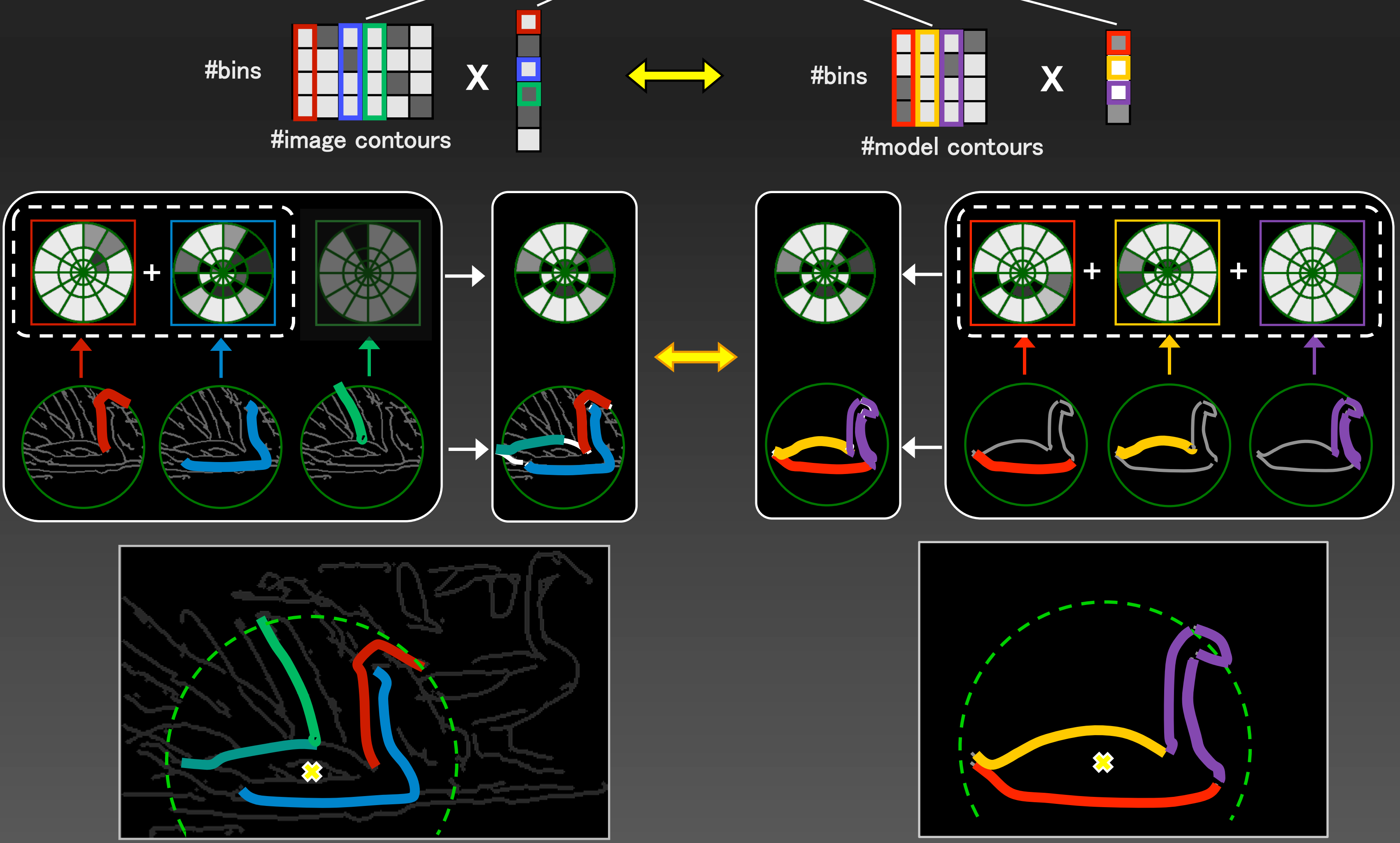
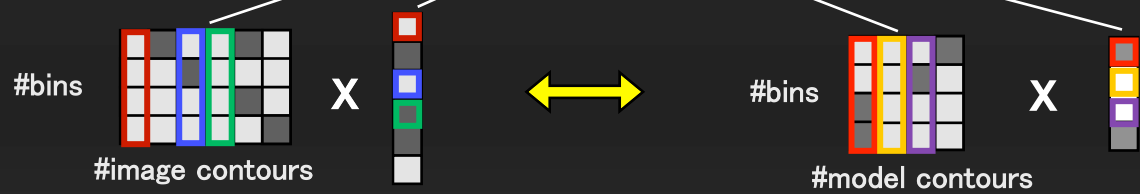


Image contours

Model contours

Shape dissimilarity $D_{ij}(V_i^I \cdot x^{sel}, V_j^M \cdot y^{sel}) = \|V_i^I \cdot x^{sel} - V_j^M \cdot y^{sel}\|_1$

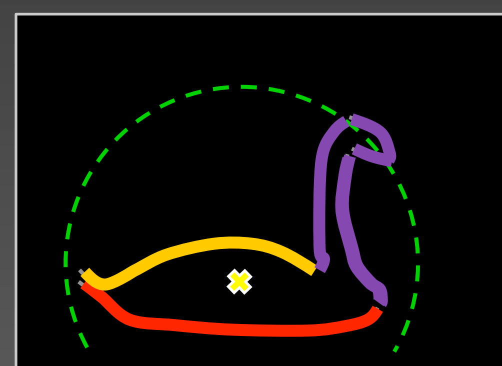


Contour Packing (LP)

$$\min_{x^{sel}} D_{ij} = \|V_i^I \cdot x^{sel} - V_j^M \cdot y^{sel}\|_1$$

$$\text{s.t. } 0 \leq x^{sel} \leq 1$$

Image contours



Model contours

Detection Example

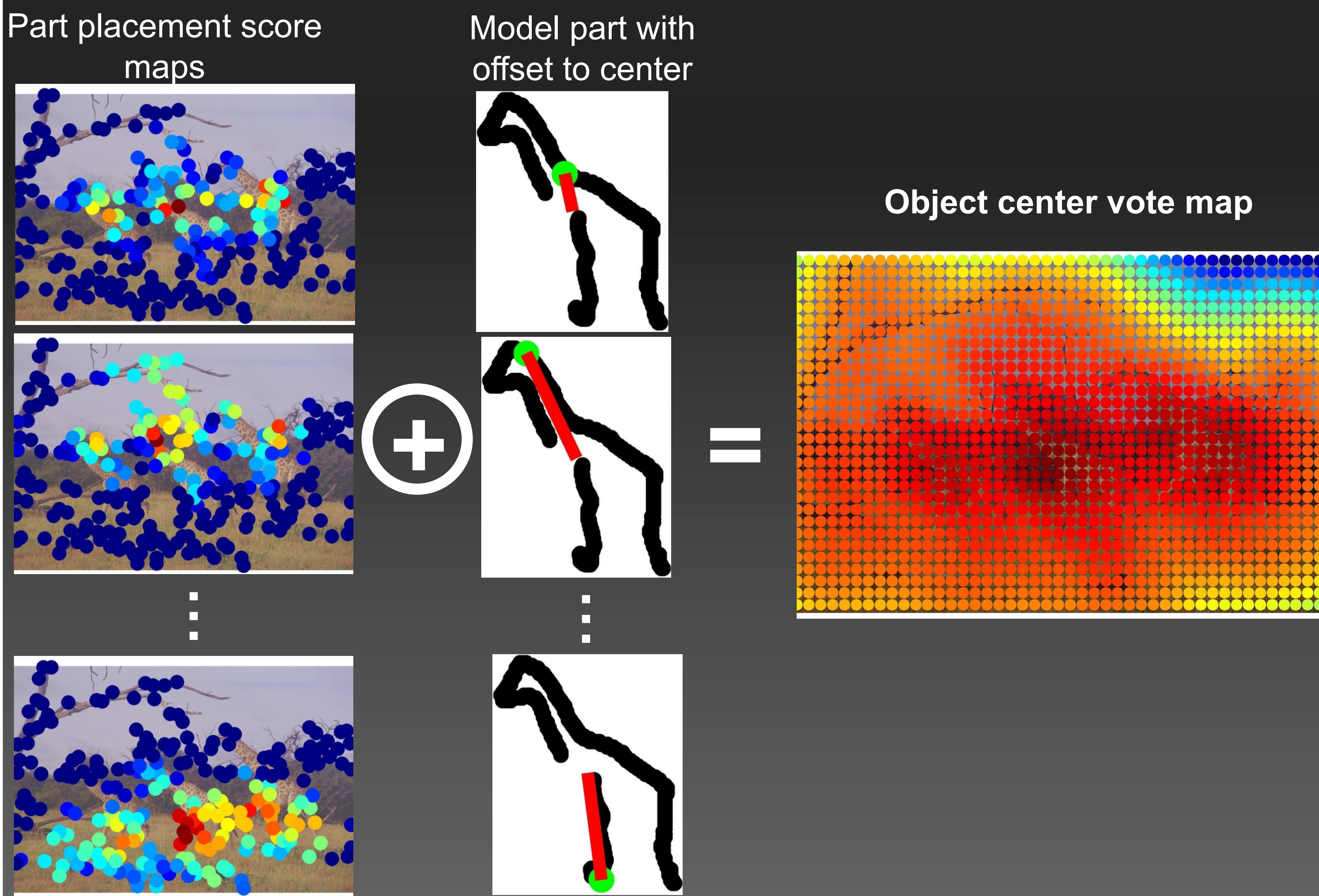
Model parts are matched at multiple locations in image

Model part

Many-to-one matching at different image locations

Part placement score map

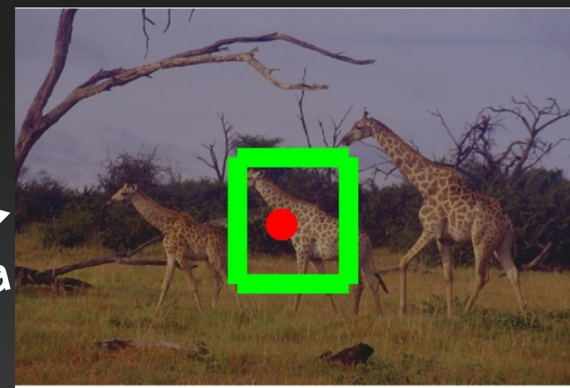
Detection Example



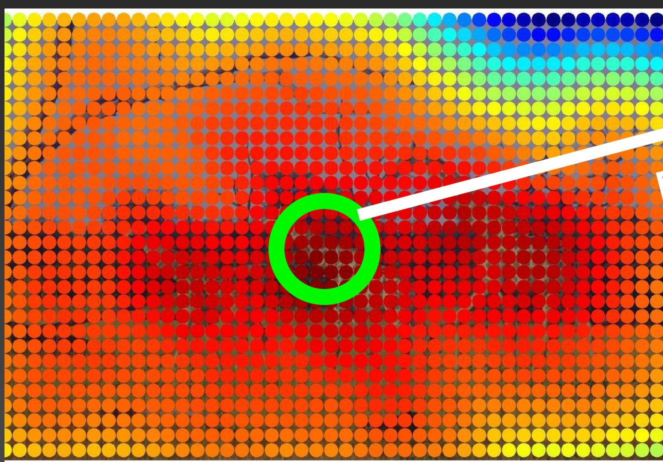
Detection Example

For each detection,

•Estimate bounding box

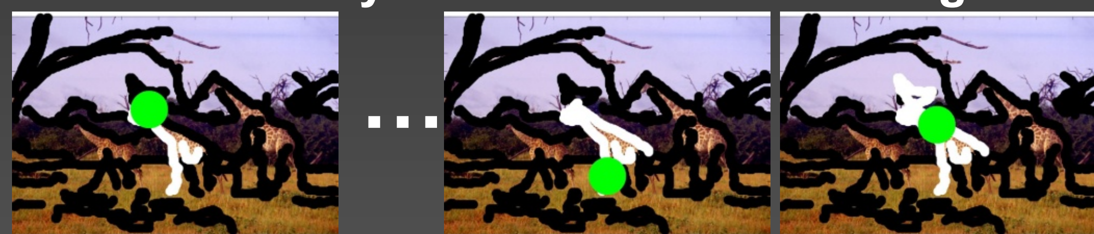


Object center vote map



Extract local maxima

•Traceback of part locations, many-to-one contour matchings



Contour Packing Score $\sum_{ij} U_{ij}^{cor} D_{ij}(V^I \cdot x^{sel}, V^M \cdot y^{sel})$

U_{ij}^{cor} control point correspondence

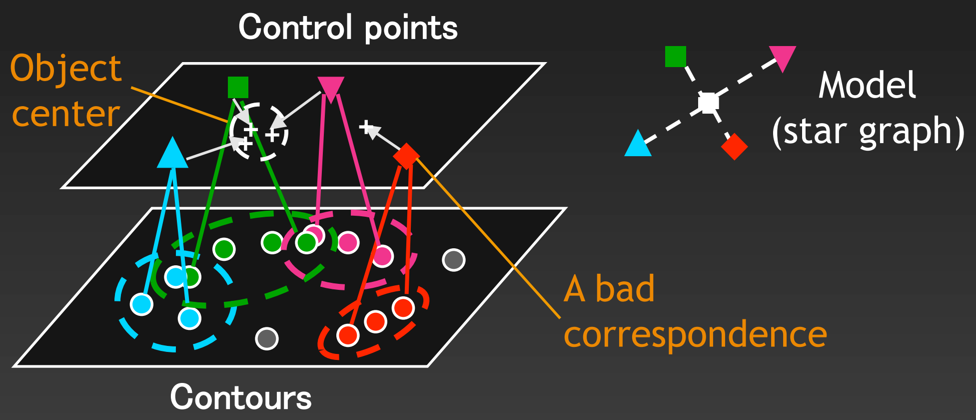
x^{sel} *0/1* indicator of image contours

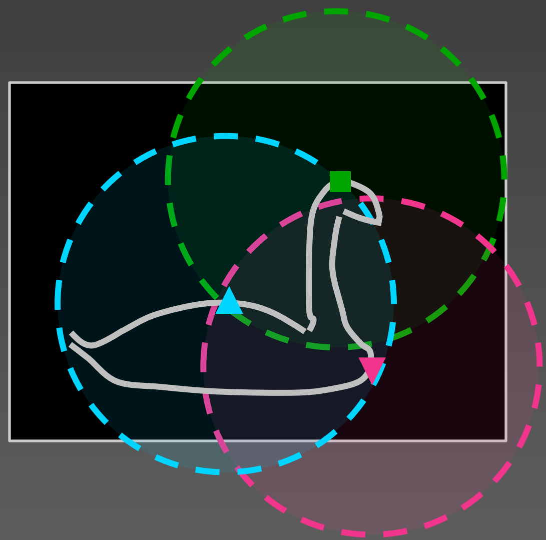
y^{sel} *0/1* indicator of model contours

$$D_{ij} = \text{miss} + \beta \cdot \text{mismatch}$$

miss penalizes matching error;

mismatch encourages maximal packing





Algorithm Overview

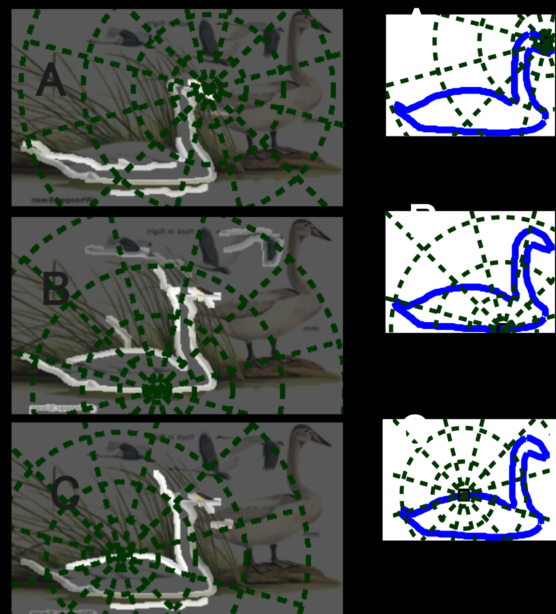
Input image

Run edge detection (Pb)

Contour grouping

Group contours from edgels might be overlapping)

Single point contour selection

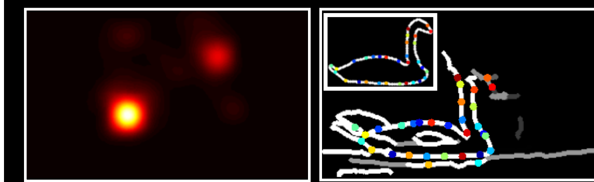


For each control point *i*, find *j* by minimizing D_{ij} (relaxed to an LP):

$$\min_{x^{sel}} D_{ij} = \|V_i^I \cdot x^{sel} - V_j^M \cdot y^{sel}\|_1$$

$$\text{s.t. } 0 \leq x^{sel} \leq 1$$

Voting



Voting map Correspondences

Find consistent U^{cor} by consensus on the center

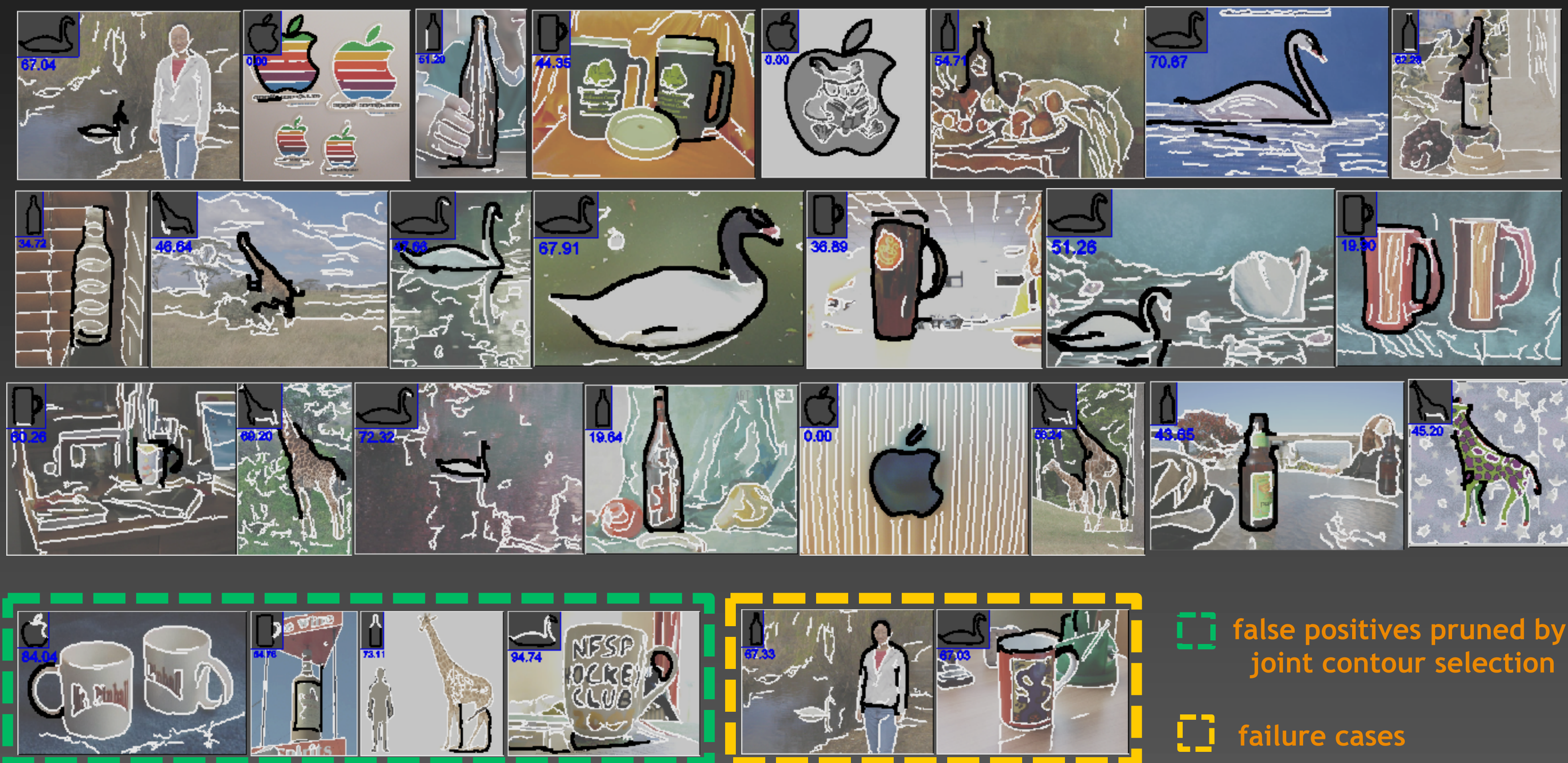
Joint selection

Selected image and model contours

Fixing U^{cor}_{ij}, find a subset of image contours match-ing to a subset of model contours, *i.e.* find x^{sel}, y^{sel} optimizing $\Sigma_{ij} U^{cor}_{ij} D_{ij}(V^I \cdot x^{sel}, V^M \cdot y^{sel})$. Relaxed to an LP as well.

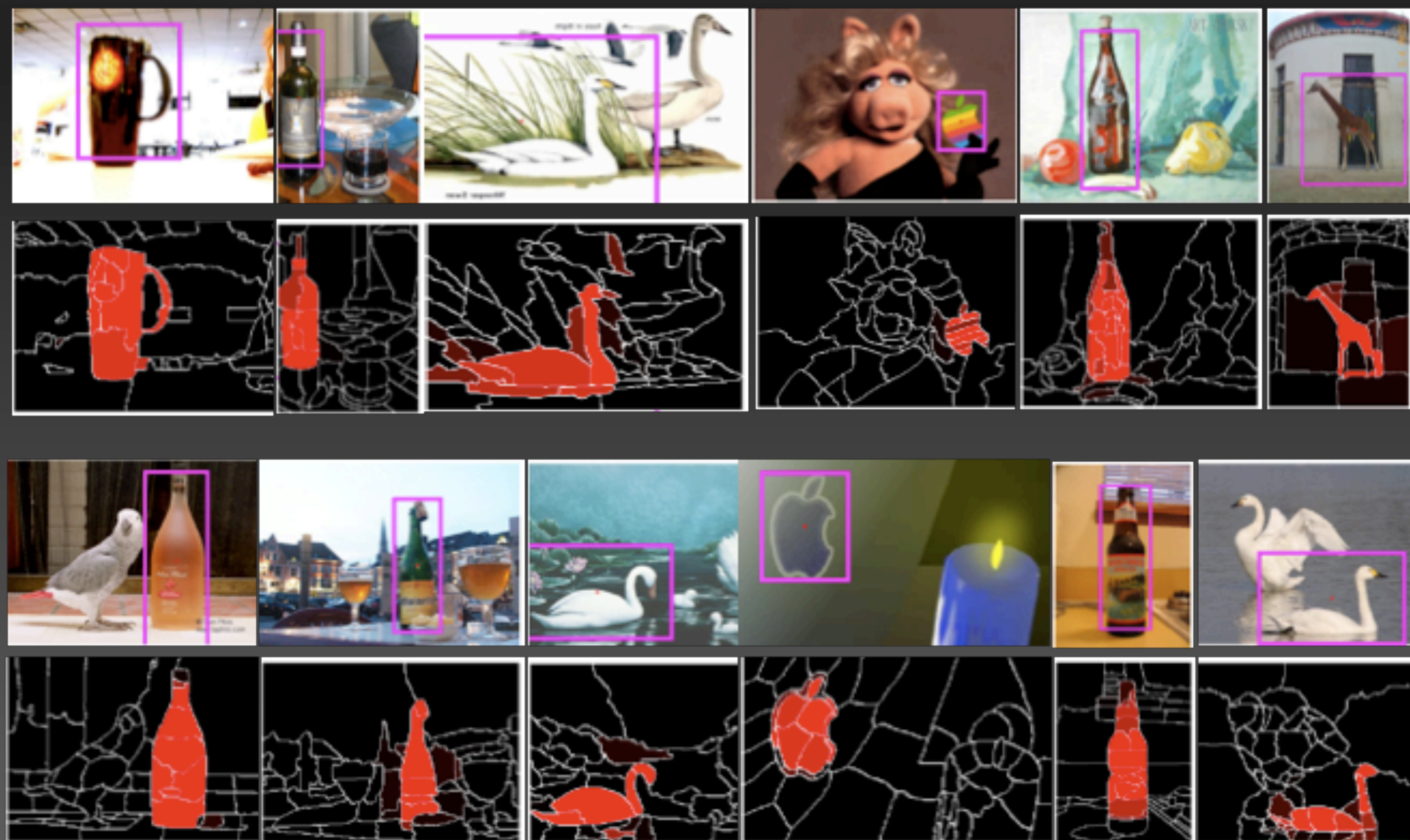
Results on ETHZ

Used only one hand-drawn model per class



Detection results of our method. Model selection is shown on the top-left corner.

Examples of Region Packing

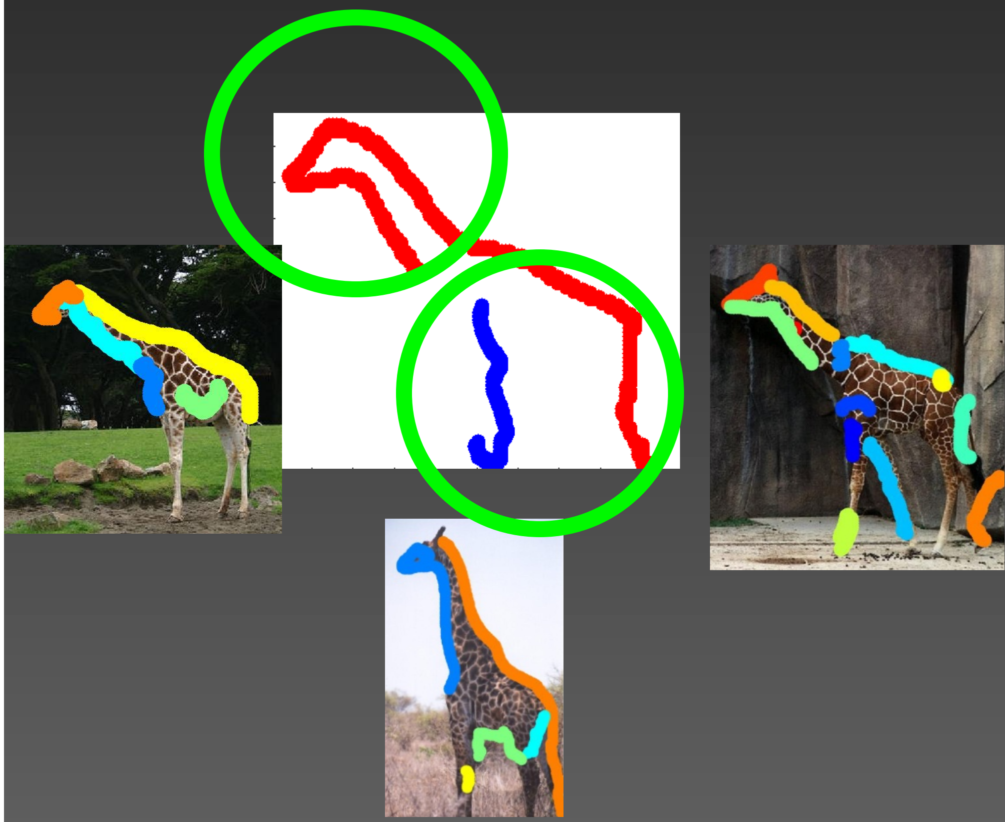


1) Shape packing
with Many-to-one matching

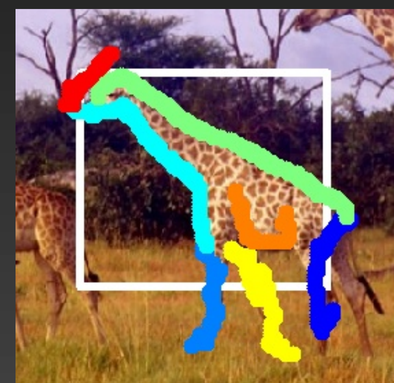
2) Learning this?

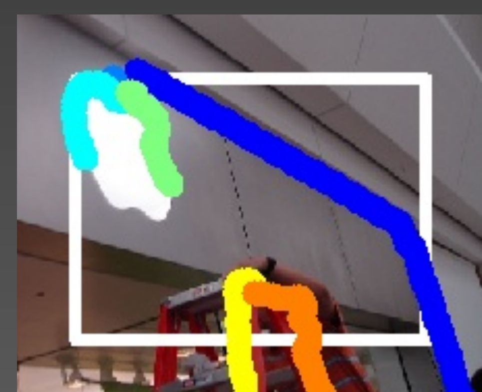
Learn to recognize objects by their shape

- **Describe** positives by commonalities of shape



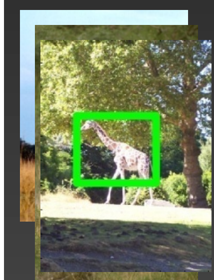
- **Discriminate** negatives by uniqueness of shape



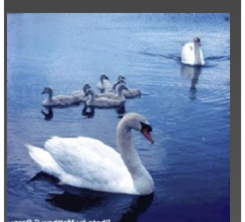


Overview: Describing Shape

Input: images, bounding boxes

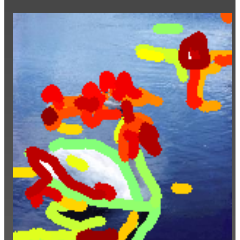


Positives

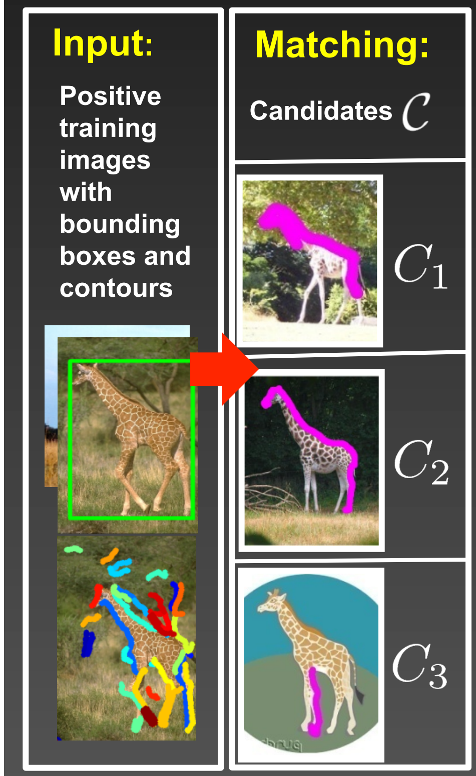


Negatives

Bottom-up contours extracted



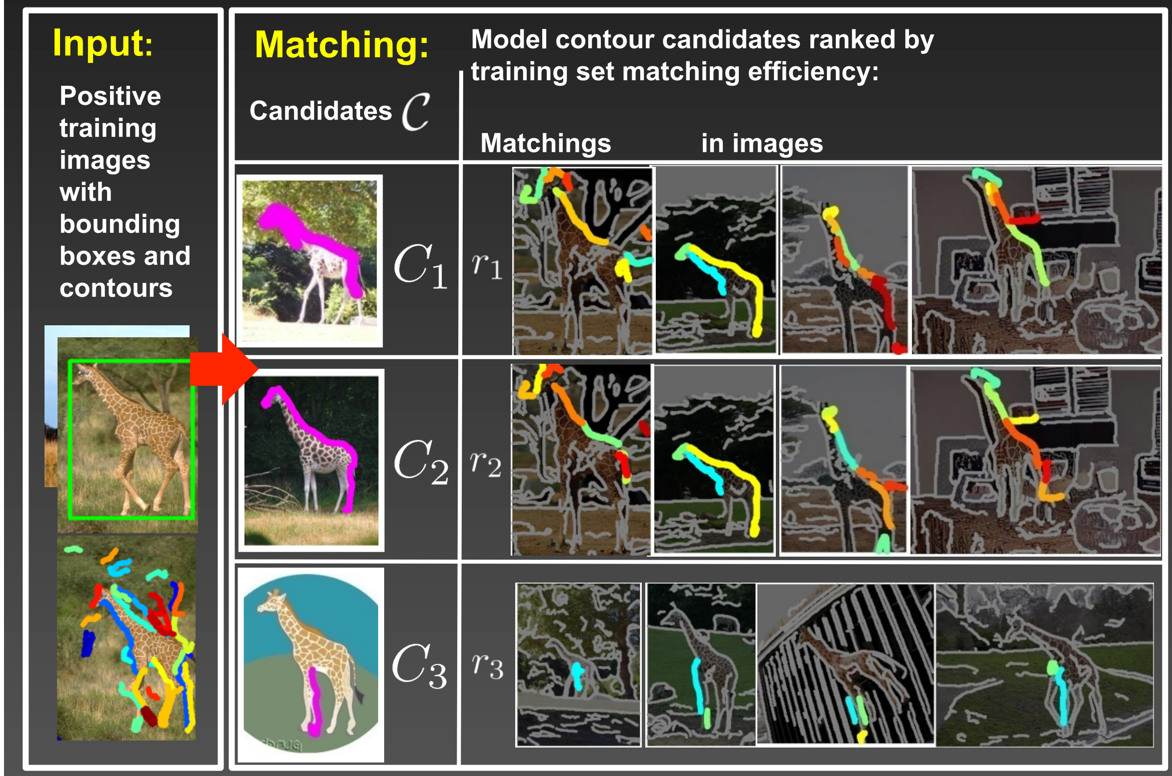
Learning a Descriptive Shape Model



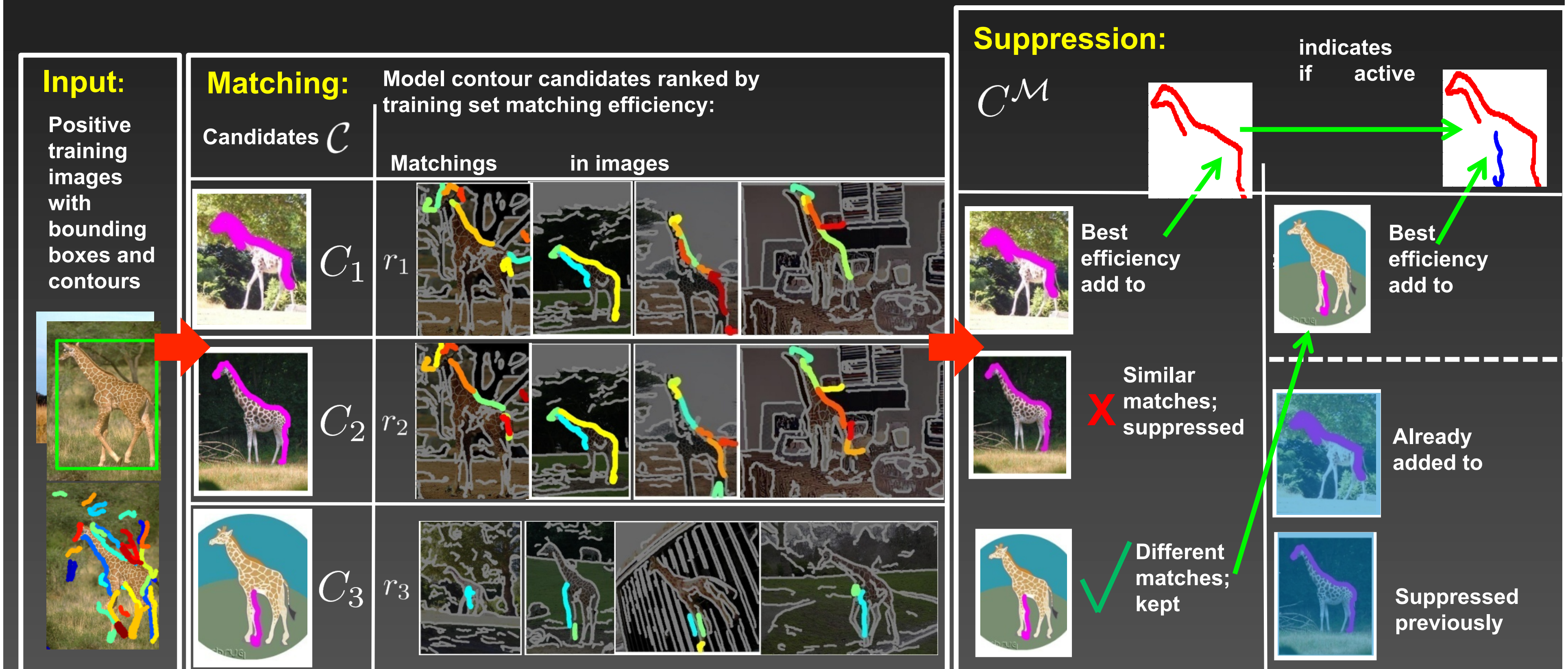
Main idea:

No one lucky giraffe, but
All giraffes are lucky in their own way

Learning a Descriptive Shape Model



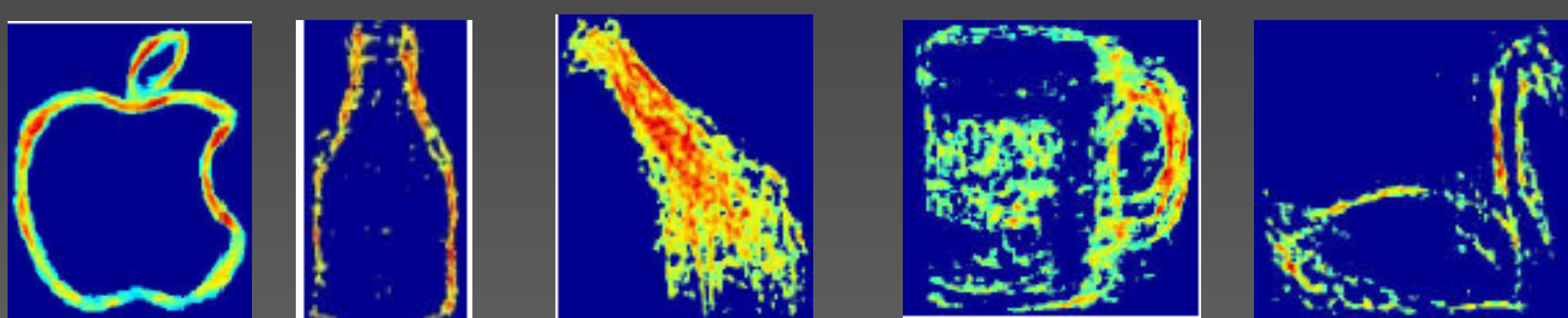
Learning a Descriptive Shape Model



Examples of Learned Model Shapes from ETHZ Dataset

- Models composed of few, non-overlapping contours

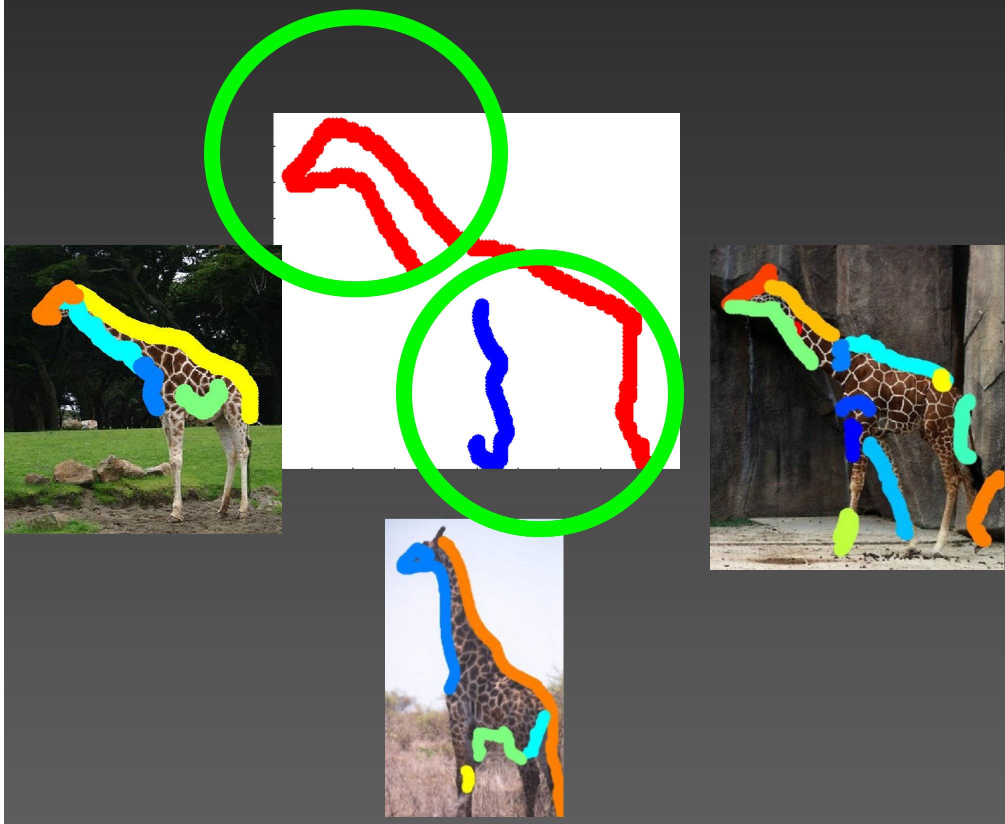
Shape details, e.g. sharp corners, are preserved



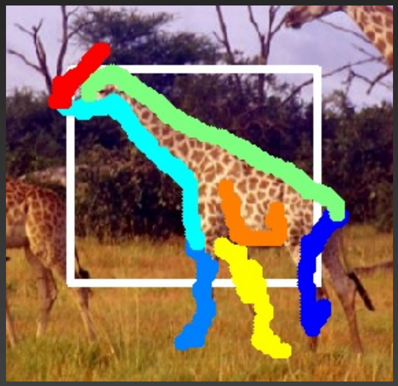
Lee, Grauman: “Shape Discovery from Unlabeled Image Collections” CVPR 2009

How to recognize objects by their shape?

- Describe positives by commonalities of shape

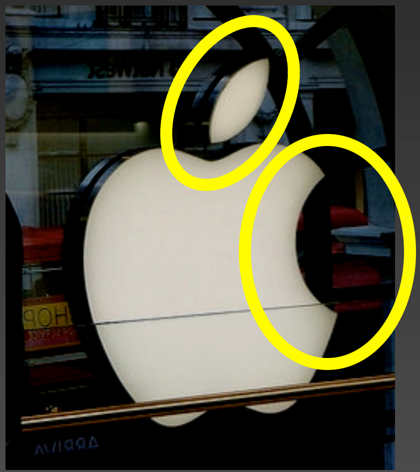


- Discriminate negatives by uniqueness of shape



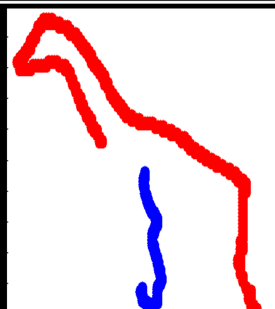
Learning Discriminative Shape Features

- Some object parts are **salient** to recognize
- Should not pack all the parts equally

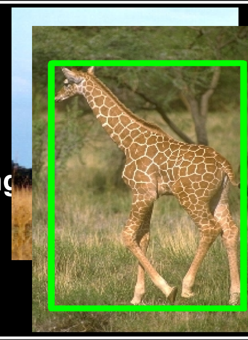


Discriminative Packing

Object outline from model shape learning

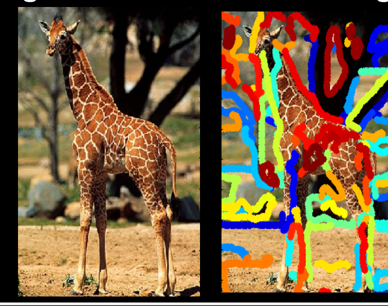


Positive training images with bounding boxes



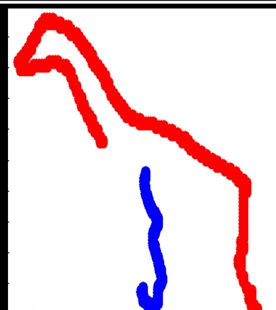
Negative training images

Image contours for all training images

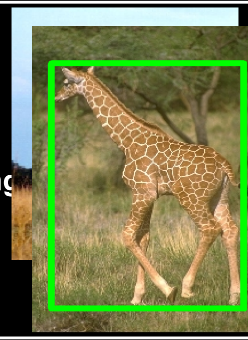


Discriminative Packing

Object outline from model shape learning

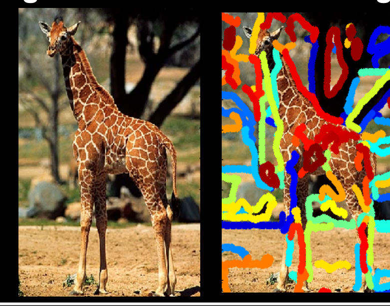


Positive training images with bounding boxes



Negative training images

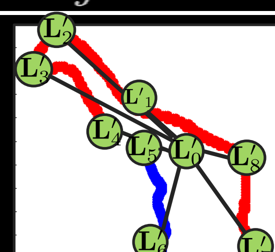
Image contours for all training images



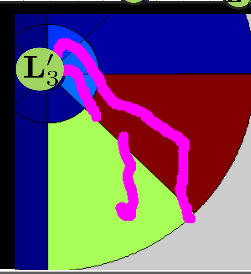
Goal Learn a discriminative packing score function

$$S(R_j, w) = \max_{D|\mathbf{L}_0 = R_j} \mathrm{DetScore}(D)$$

$$\mathrm{DetScore}(D) = \sum_{i=1}^{N} \begin{bmatrix} w_i^{\mathrm{def}} \\ w_i^{\mathrm{app}} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} G(\mathbf{L}_0, \mathbf{L}_i) \\ K(\mathbf{L}_i, \mathbf{x}_i^{\mathrm{sel}}) \end{bmatrix}$$



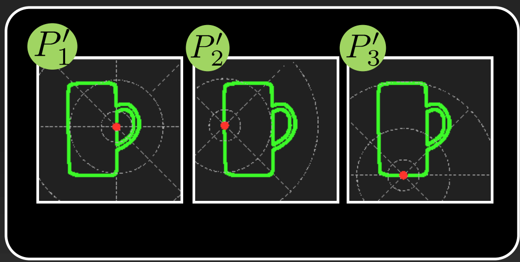
$\{w_i^{def}\}$ weights on spatial relationship of parts



$\{w_i^{app}\}$ weights on model part appearance

Output w

w_i^{app}



Initialized model

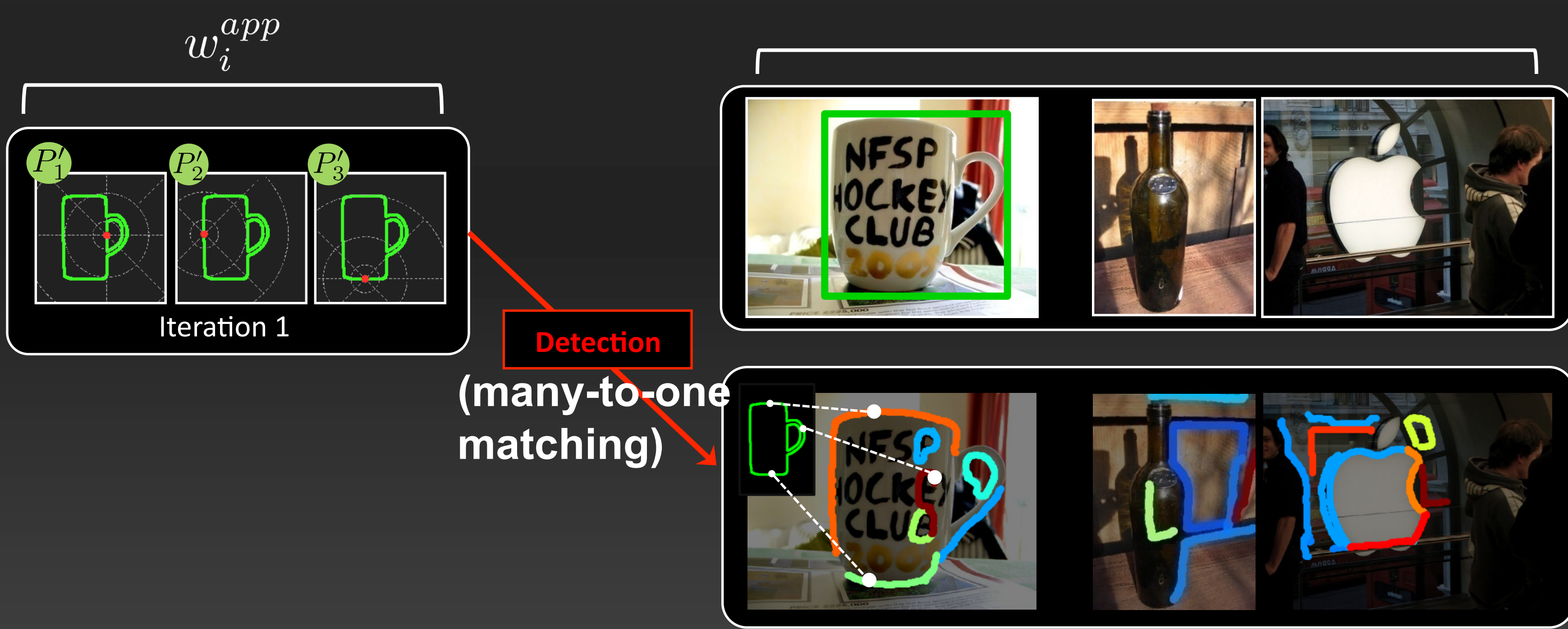
Training images

Discriminative training loss function (latent SVM)

$$\min_{w} C \sum_{j} \max(0, 1 - y_j S(R_j, w)) + \frac{1}{2} w^\top w$$

$$\text{DetScore}(D) = \sum_{i=1}^{N} \begin{bmatrix} w_i^{\text{def}} \\ w_i^{\text{app}} \end{bmatrix}^\top \begin{bmatrix} G(\mathbf{L}_0, \mathbf{L}_i) \\ K(\mathbf{L}_i, \mathbf{x}_i^{\text{sel}}) \end{bmatrix}$$

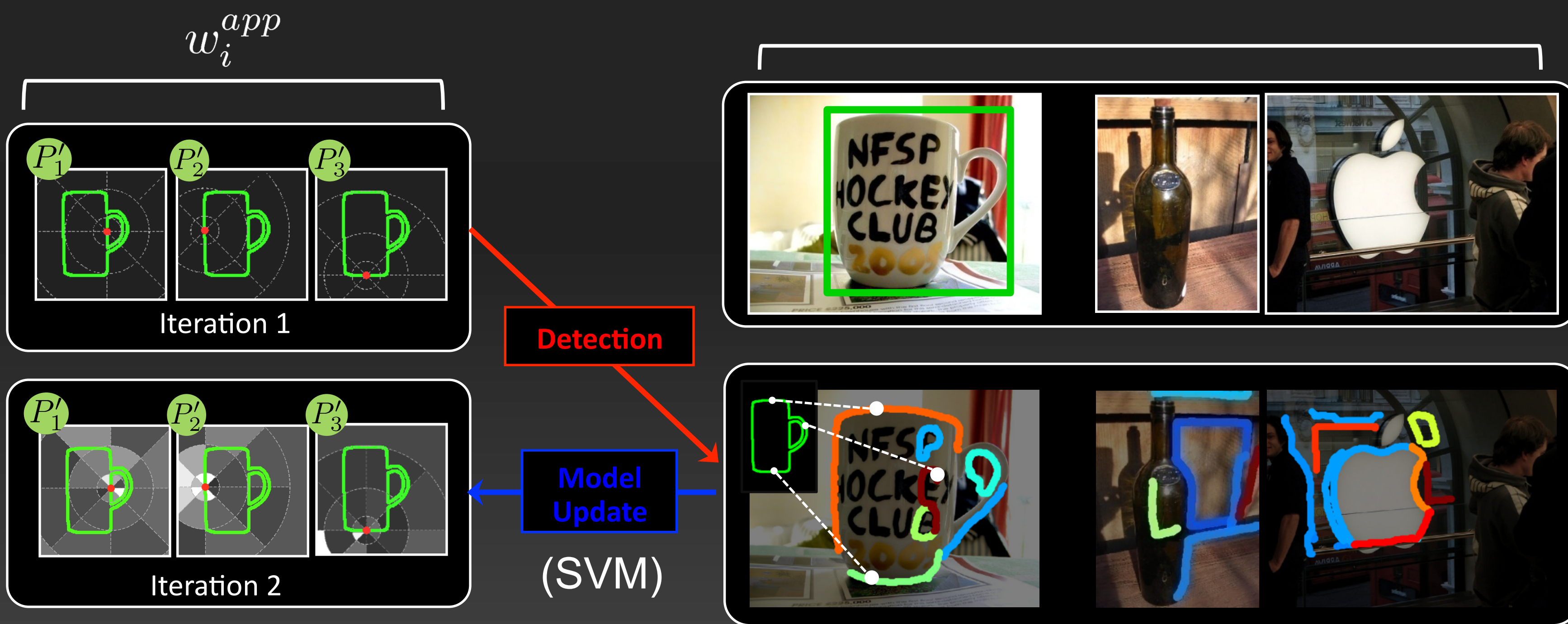
$$S(R_j, w) = \max_{D|\mathbf{L}_0 = R_j} \text{DetScore}(D)$$



$$\min_{w} C \sum_{j} \max(0, 1 - y_j S(R_j, w)) + \frac{1}{2} w^{\top} w$$

$$\text{DetScore}(D) = \sum_{i=1}^{N} \begin{bmatrix} w_i^{\text{def}} \\ w_i^{\text{app}} \end{bmatrix}^{\top} \begin{bmatrix} G(\mathbf{L}_0, \mathbf{L}_i) \\ K(\mathbf{L}_i, \mathbf{x}_i^{\text{sel}}) \end{bmatrix}$$

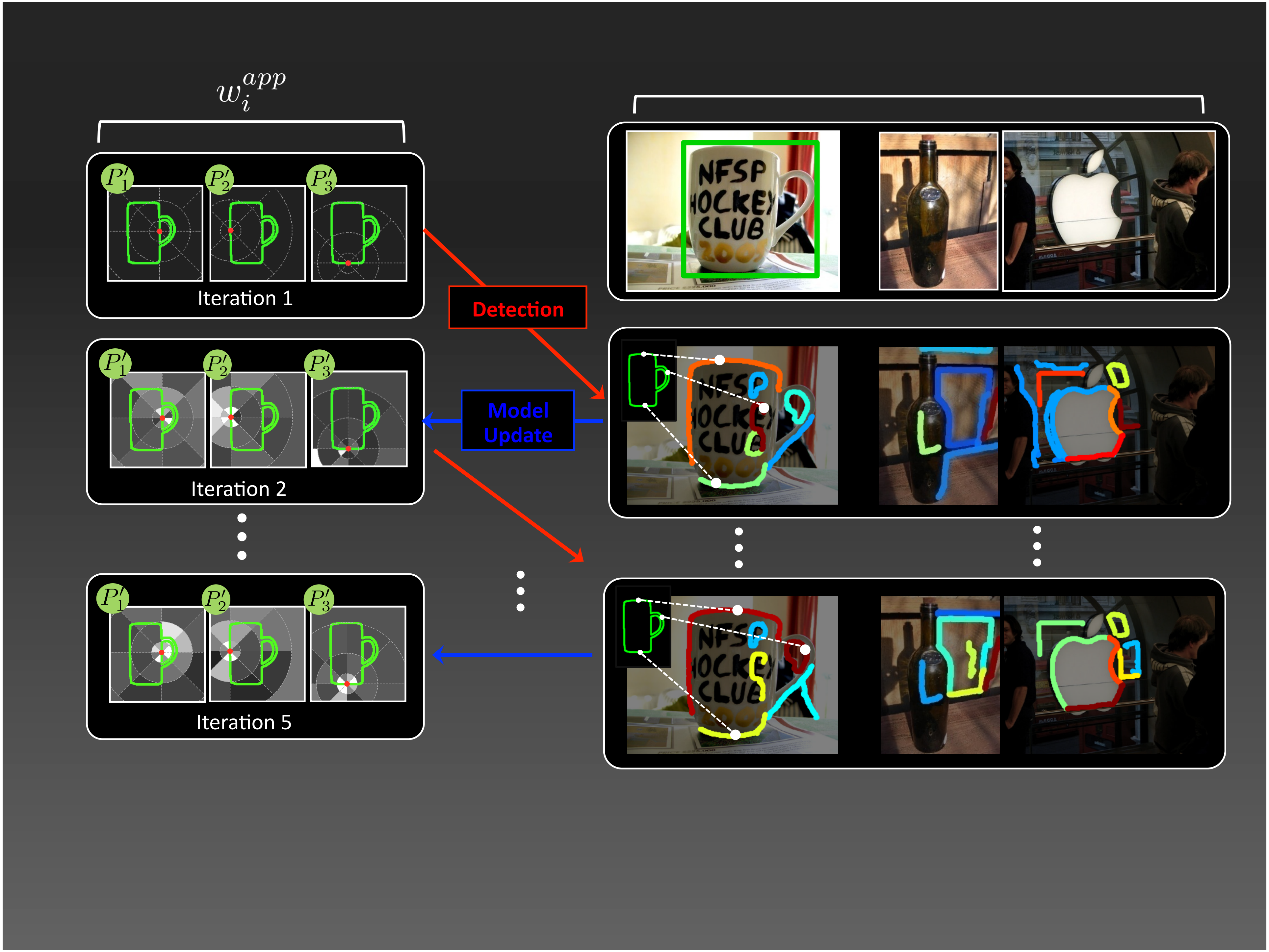
$$S(R_j, w) = \max_{D|\mathbf{L}_0 = R_j} \text{DetScore}(D)$$



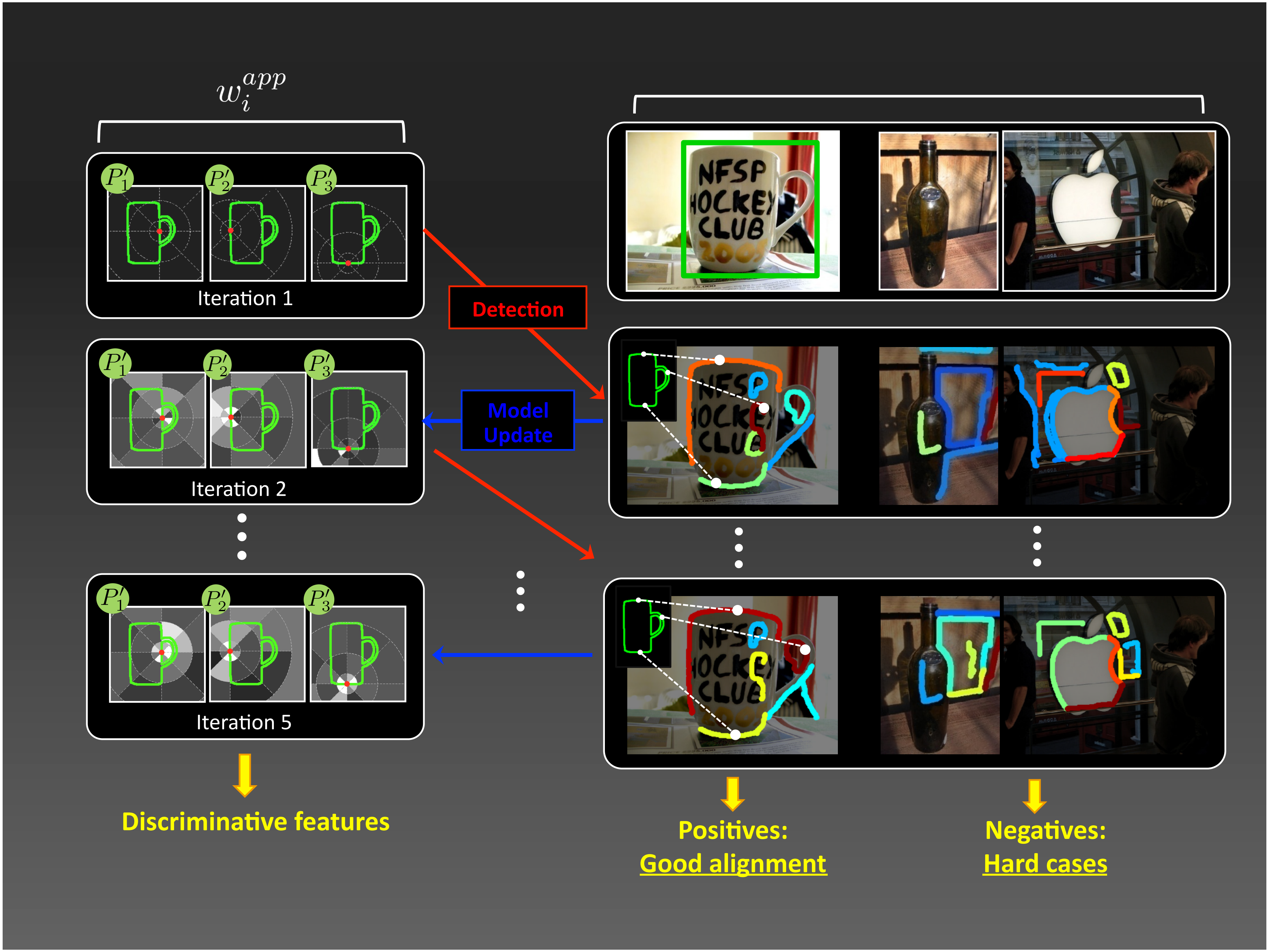
$$\min_{w} C \sum_j \max(0, 1 - y_j S(R_j, w)) + \frac{1}{2} w^\top w$$

$$\mathrm{DetScore}(D) = \sum_{i=1}^{N} \begin{bmatrix} w_i^{\mathrm{def}} \\ w_i^{\mathrm{app}} \end{bmatrix}^\top \begin{bmatrix} G(\mathbf{L}_0, \mathbf{L}_i) \\ K(\mathbf{L}_i, \mathbf{x}_i^{\mathrm{sel}}) \end{bmatrix}$$

$$S(R_j, w) = \max_{D \mid \mathbf{L}_0 = R_j} \mathrm{DetScore}(D)$$

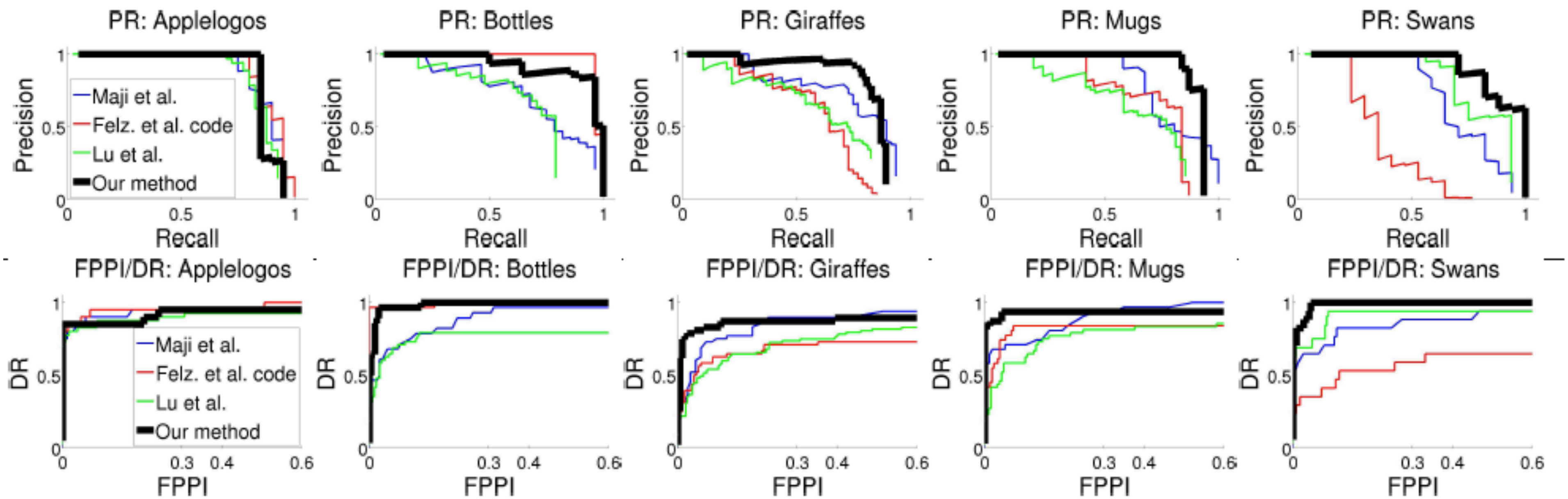


w_i^{app}
P'_1
P'_2
P'_3
Iteration 1
Iteration 2
Iteration 5
Detection
Model Update
NFSP HOCKEY CLUB



w_i^{app}
P'_1
P'_2
P'_3
Iteration 1
Iteration 2
Iteration 5
Detection
Model Update
NFSP HOCKEY CLUB
Discriminative features
Positives: Good alignment
Negatives: Hard cases

Quantitative Evaluation



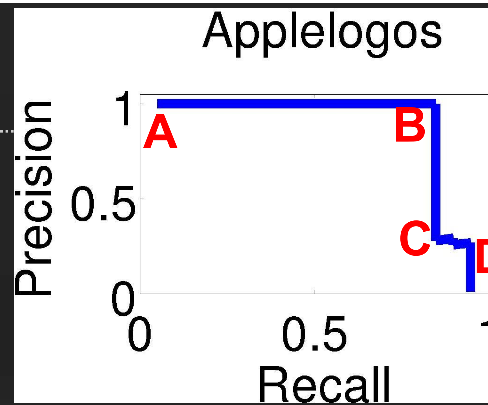
Average PR

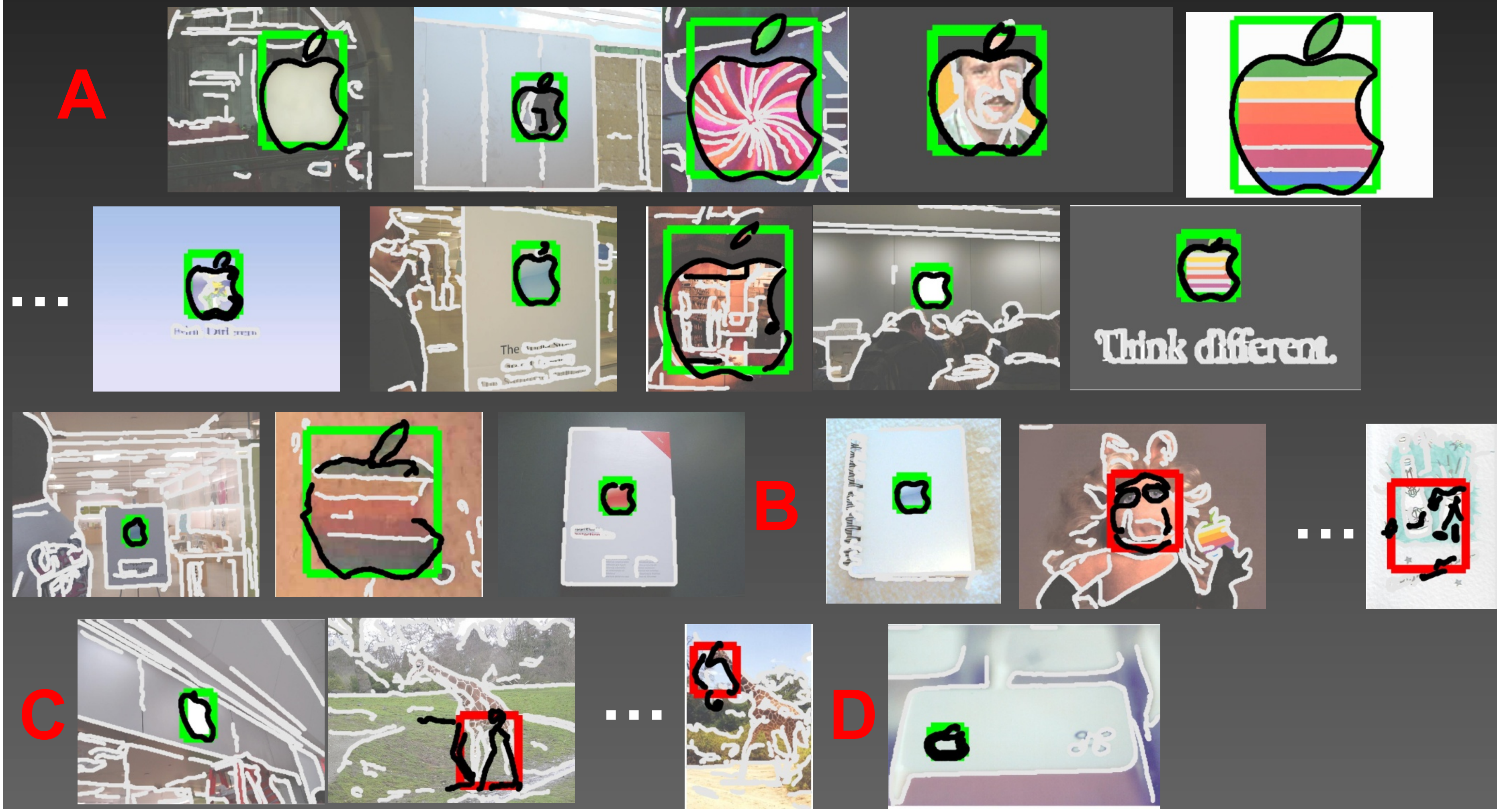
	Applelogos	Bottles	Giraffes	Mugs	Swans	Mean
Our method	0.845	0.916	**0.787**	**0.888**	**0.922**	**0.872**
Maji et al. [12]	0.869	0.724	0.742	0.806	0.716	0.771
Felz. et al. [4] code	**0.891**	**0.950**	0.608	0.721	0.391	0.712
Lu et al. [11]	0.844	0.641	0.617	0.643	0.798	0.709

Srinivasan, Shi: Many-to-one Contour Matching for Describing and Discriminating Object Shape, *CVPR* 2010

Applelogo Detections Ordered by Score

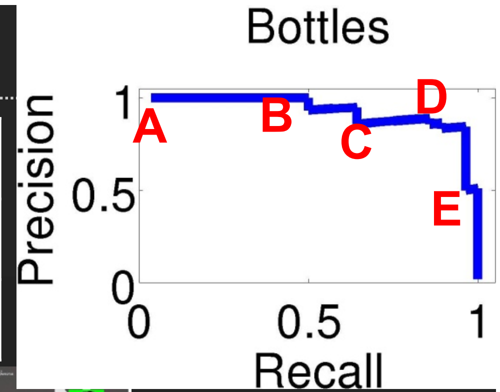
	A	B	C	D	Max Recall
Prec (%)	100	100	29	27	27
Rec (%)	5	85	90	95	95

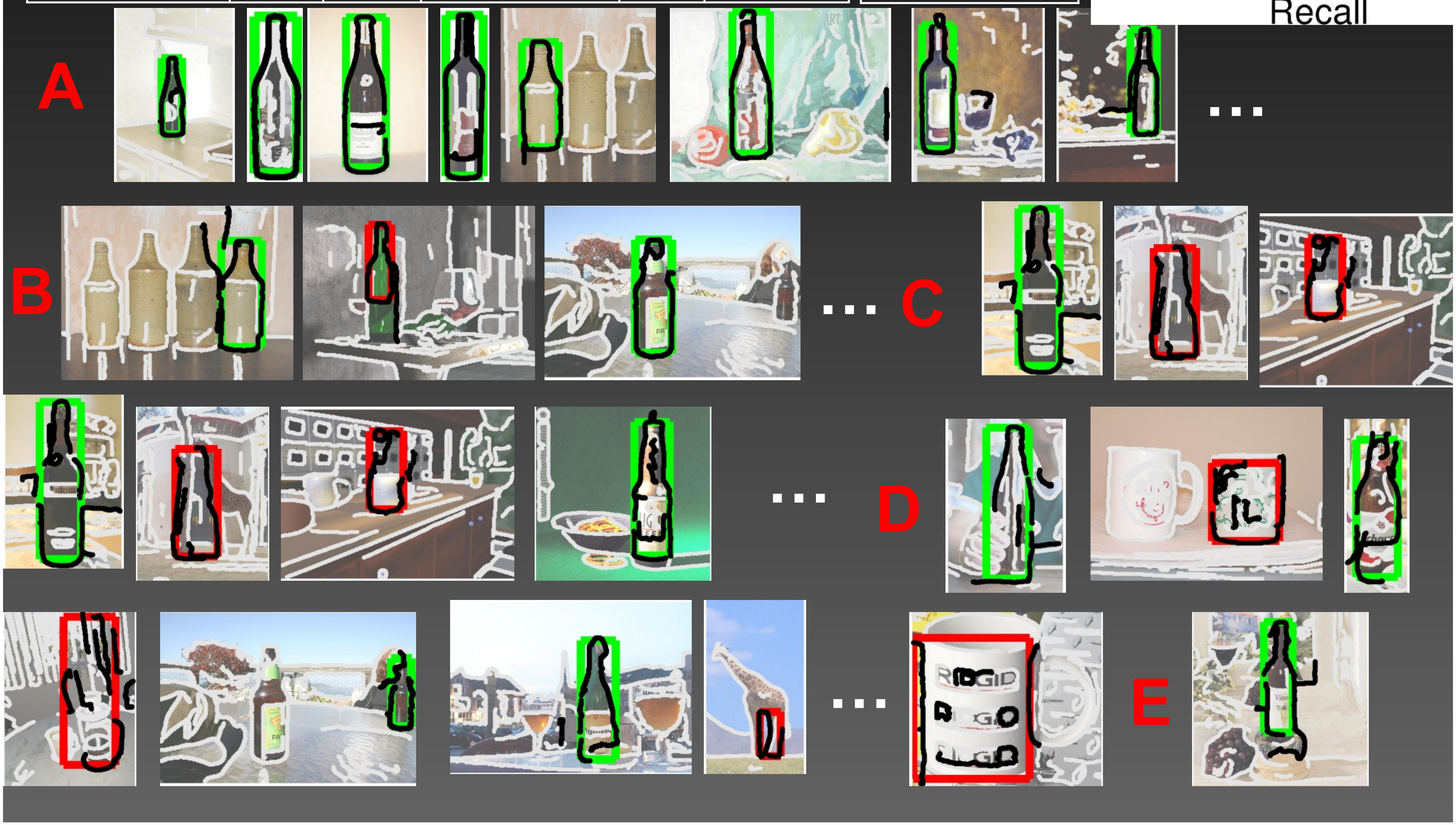




Bottle Detections Ordered by Score

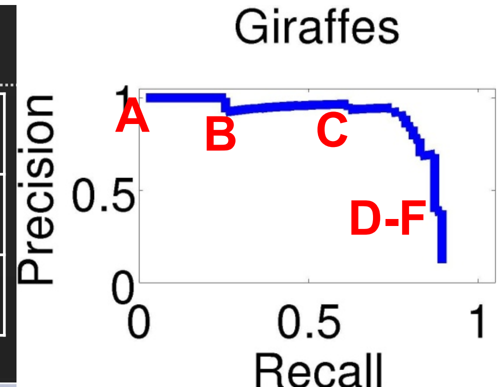
	A	B	C	D	E	Max Recall
Prec (%)	100	100	95	89	51	51
Rec (%)	3.6	50	64	86	100	100





Giraffe Detections Ordered by Score

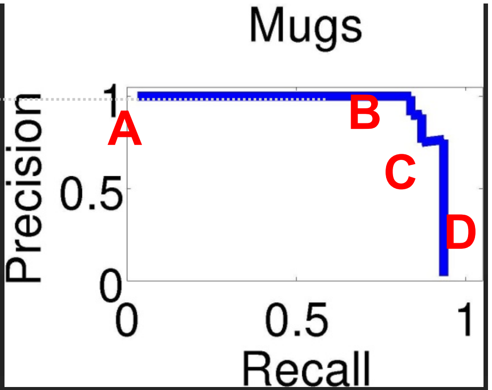
	A	B	C	D	E	F	Max R.
Prec (%)	100	100	97	95	86	68	39
Rec (%)	2.1	26	62	75	79	83	89

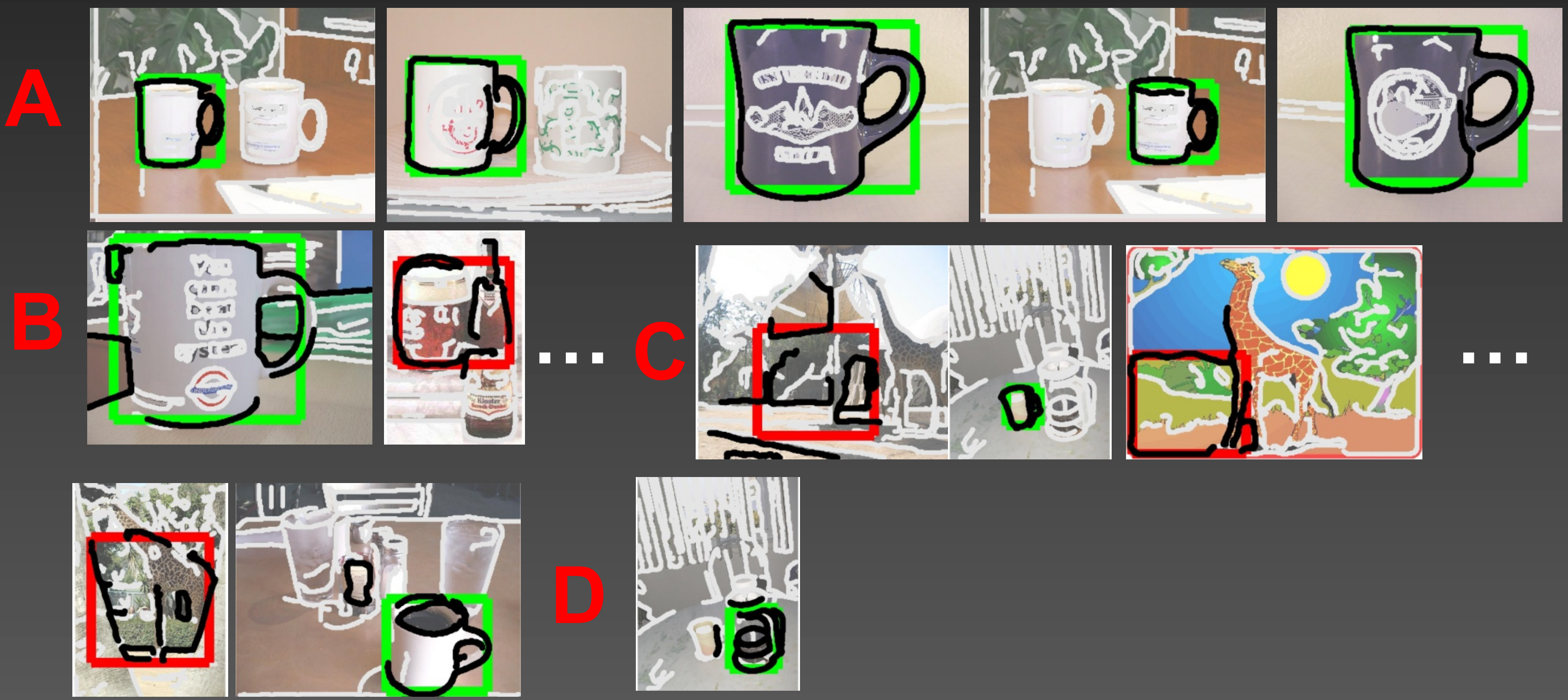


Mug Detections Ordered by Score

	A	B	C	D
Prec (%)	100	100	87	76
Rec (%)	3.2	84	87	94

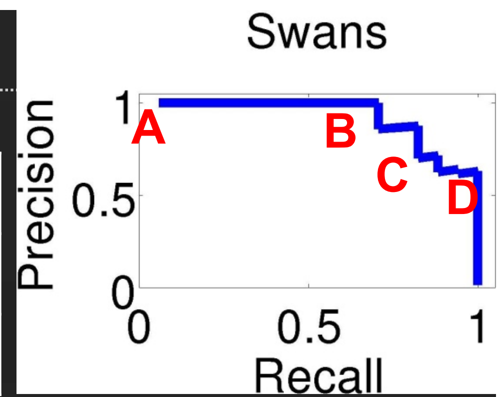
Max Recall
76
94





Swan Detections Ordered by Score

	A	B	C	D	Max Recall
Prec (%)	100	100	71	63	63
Rec (%)	6	71	88	100	100



Shape Packing

1) Global Shape

2) Active feature construction