# CS381V Paper Presentation

Chun-Chen Kuo

# Selective Search for Object Recognition

# Outline

- Problem statement
- Technical details
- Evaluation
- Extensions

# Problem Statement

# Image Classification

- Input: training set $(I_i, c_i)$ and test set $I$
- Output: the class of the test images and the confidence scores

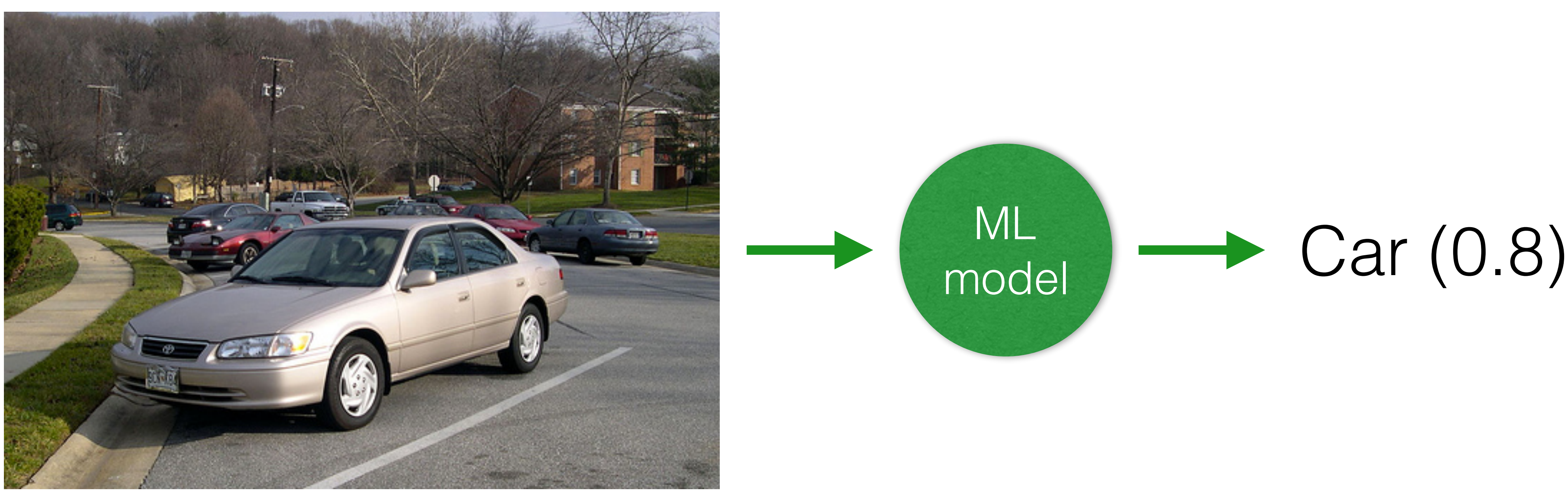



image from ImageNet

# Object Detection

- Input: training set $(I_i, c_i, y1_i, x1_i, y2_i, x2_i)$
  and a test set $I$

- Output: all objects in the test images and
  their **bounding boxes**

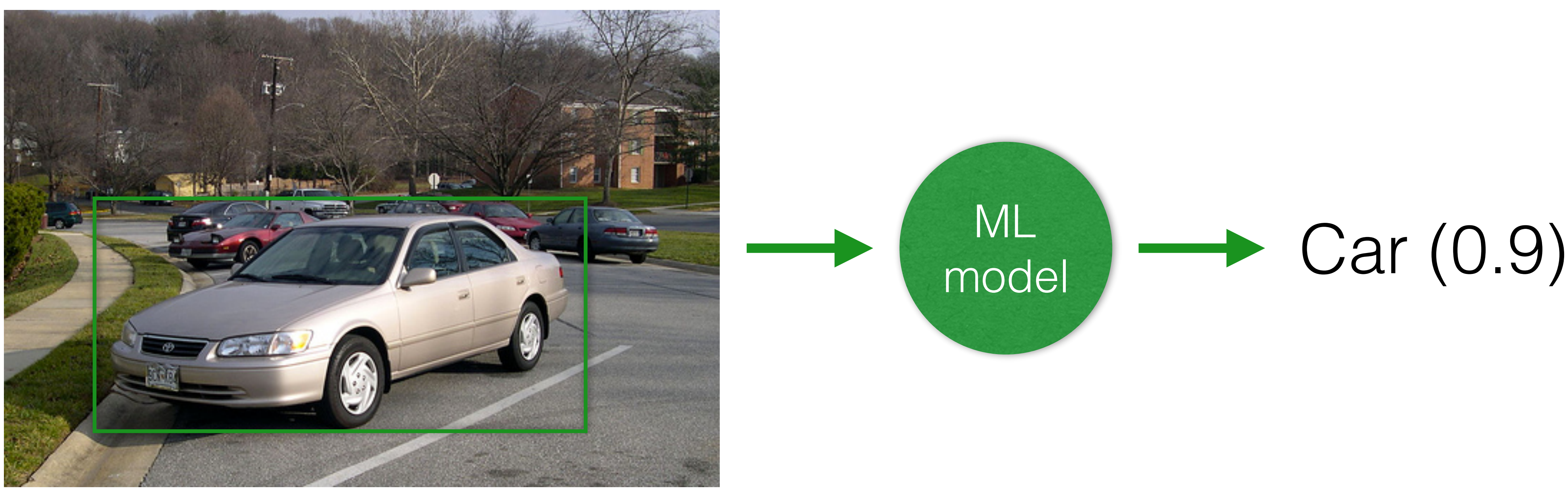



image from ImageNet

- How to turn object detection problem to image classification problem?

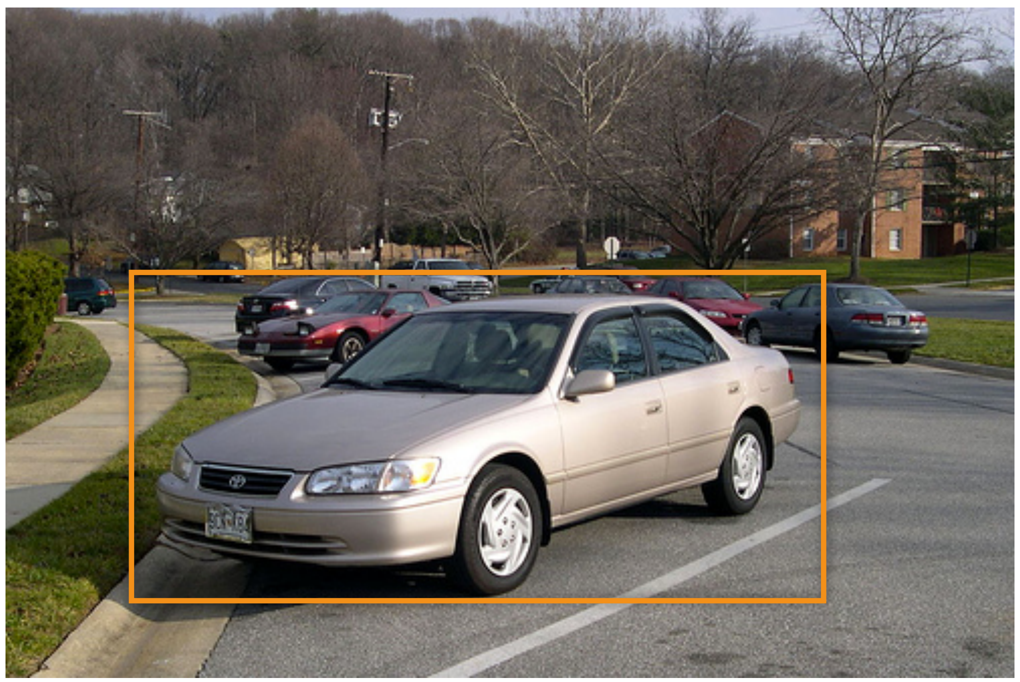
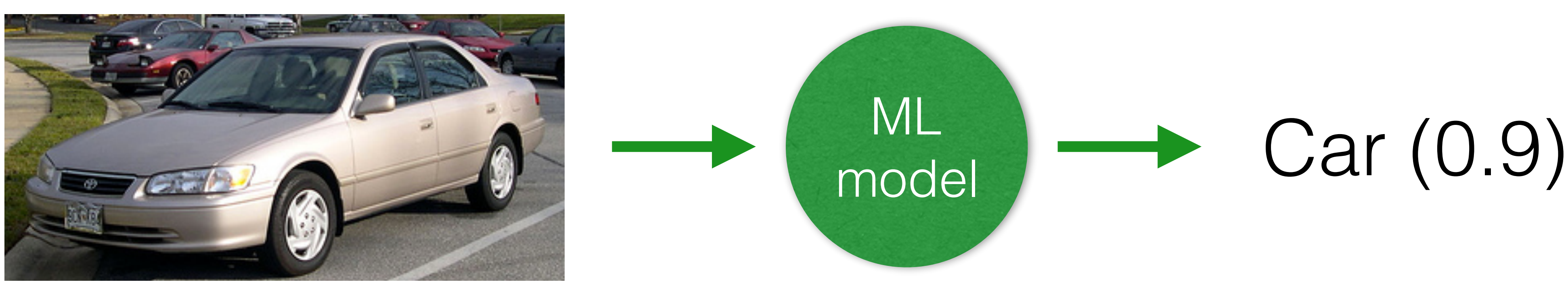


image from ImageNet



- How many sub-regions should we test and how do we generate them?

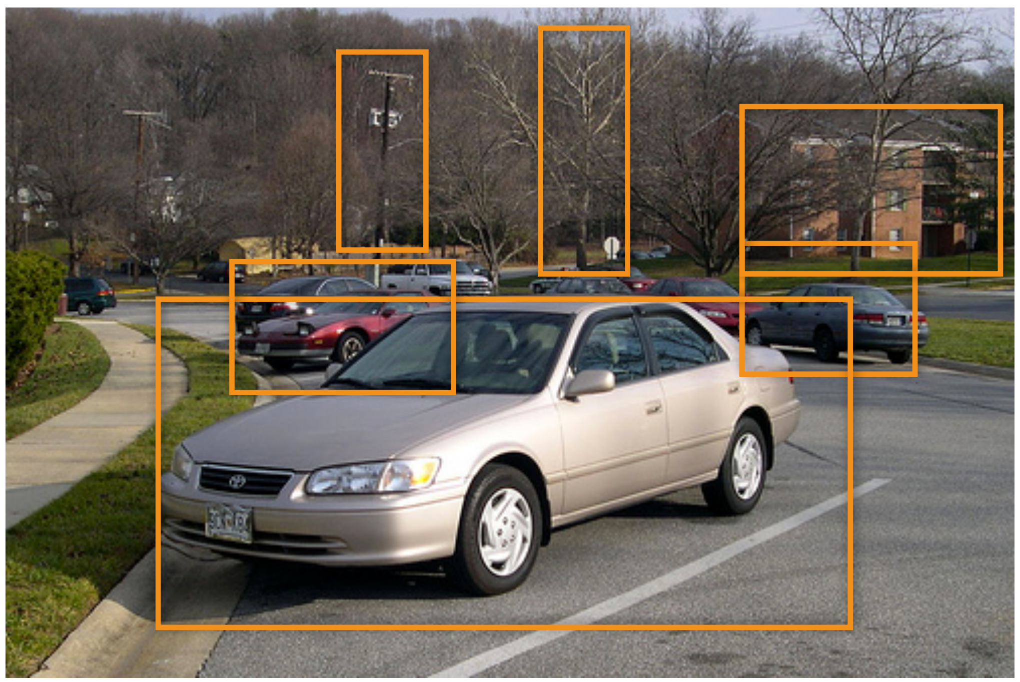
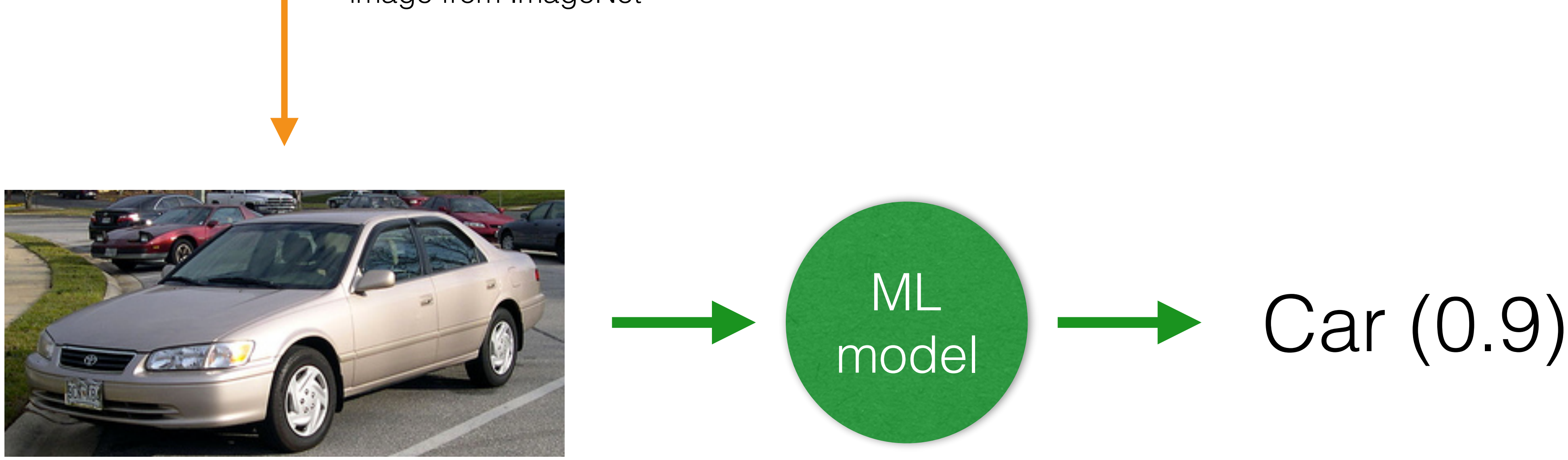


image from ImageNet



# Exhaustive Search

- Generate all possible windows
- Complexity: $\Theta(\underbrace{w \times h}_{\text{all size}} \times \underbrace{w \times h}_{\text{all location}}) = \Theta(w^2 h^2)$

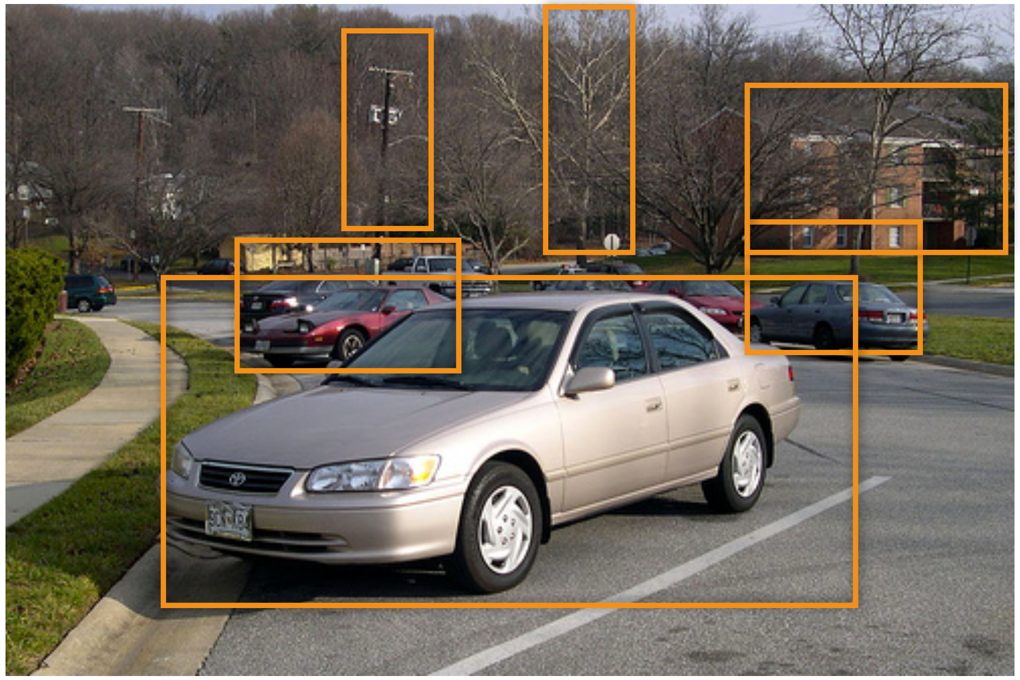


image from ImageNet

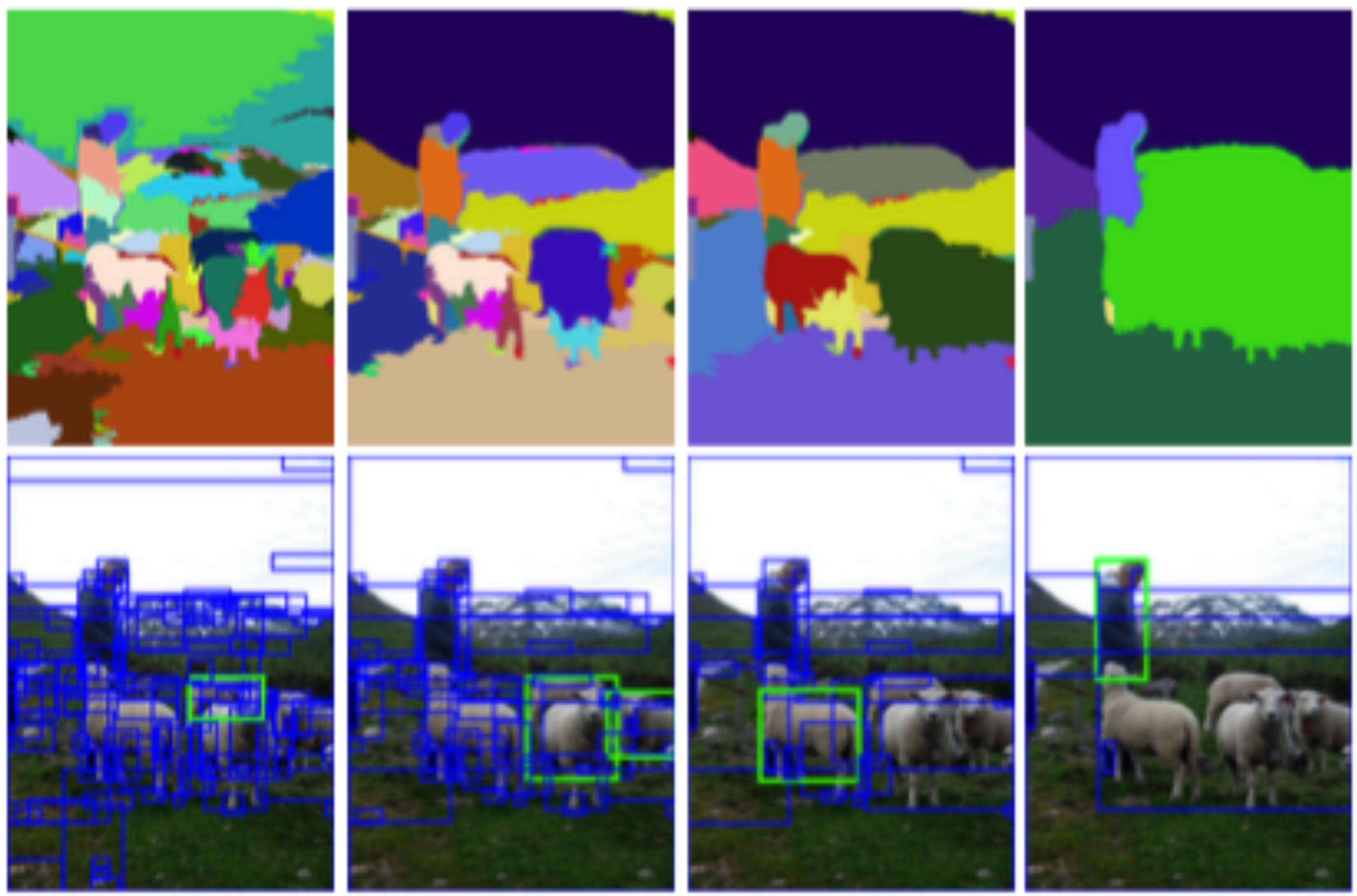
# Selective Search

- Reduce the number of hypotheses while keep recall high
- **Select** some high quality hypotheses, which are subset of all possible hypotheses

# Technical Details

# Intuition

- Explore image **structure** and group regions from small scale to high scale (**hierarchical** grouping)



Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Algorithm

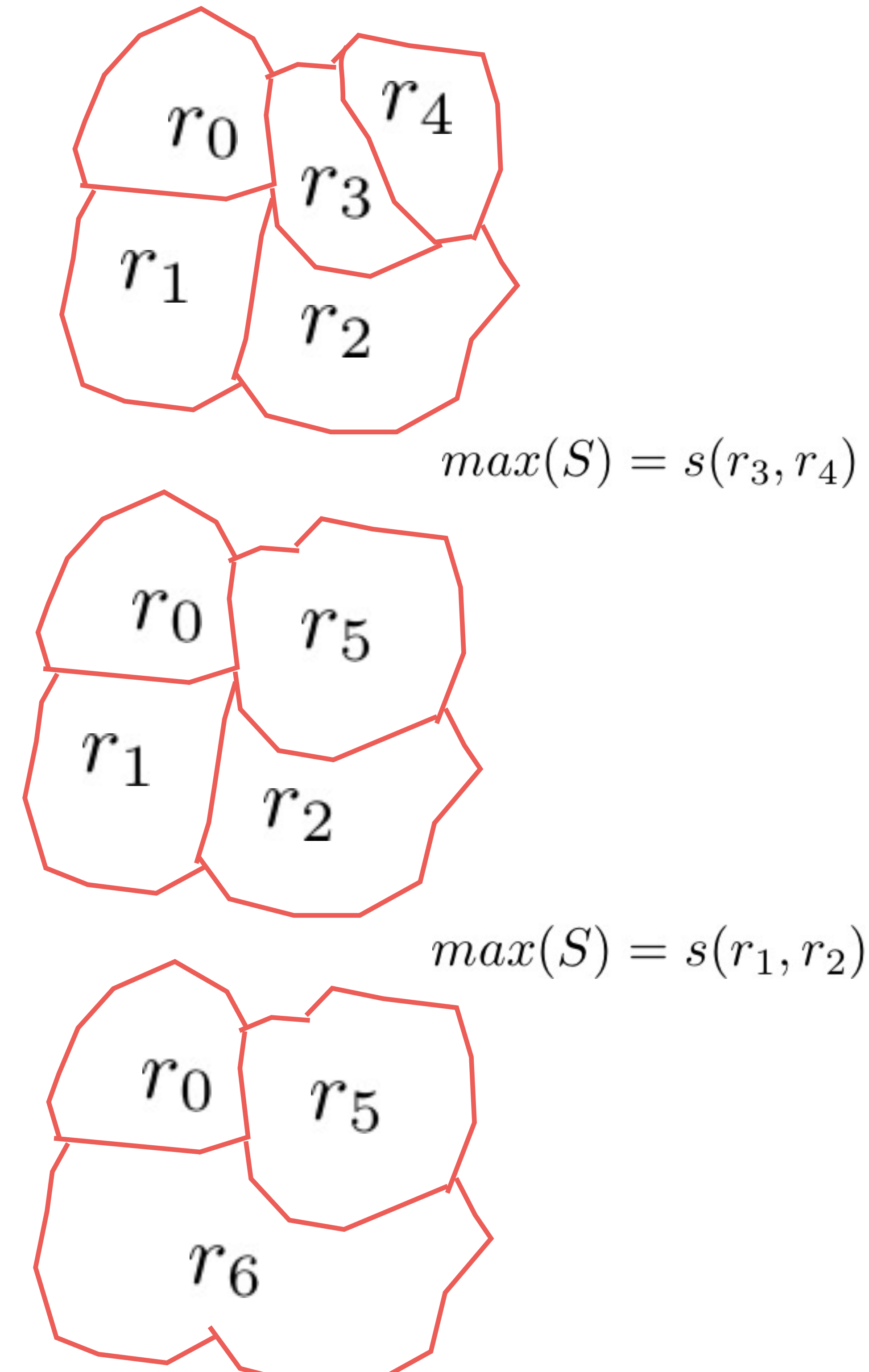
**Algorithm 1:** Hierarchical Grouping Algorithm

**Input**: (colour) image
**Output**: Set of object location hypotheses $L$

Obtain initial regions $R = \{r_1, \cdots, r_n\}$ using [13]
Initialise similarity set $S = \emptyset$
**foreach** *Neighbouring region pair* $(r_i, r_j)$ **do**
  Calculate similarity $s(r_i, r_j)$
  $S = S \cup s(r_i, r_j)$

**while** $S \neq \emptyset$ **do**
  Get highest similarity $s(r_i, r_j) = \max(S)$
  Merge corresponding regions $r_t = r_i \cup r_j$
  Remove similarities regarding $r_i : S = S \setminus s(r_i, r_*)$
  Remove similarities regarding $r_j : S = S \setminus s(r_*, r_j)$
  Calculate similarity set $S_t$ between $r_t$ and its neighbours
  $S = S \cup S_t$
  $R = R \cup r_t$

Extract object location boxes $L$ from all regions in $R$




Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function

---

**Algorithm 1:** Hierarchical Grouping Algorithm

---

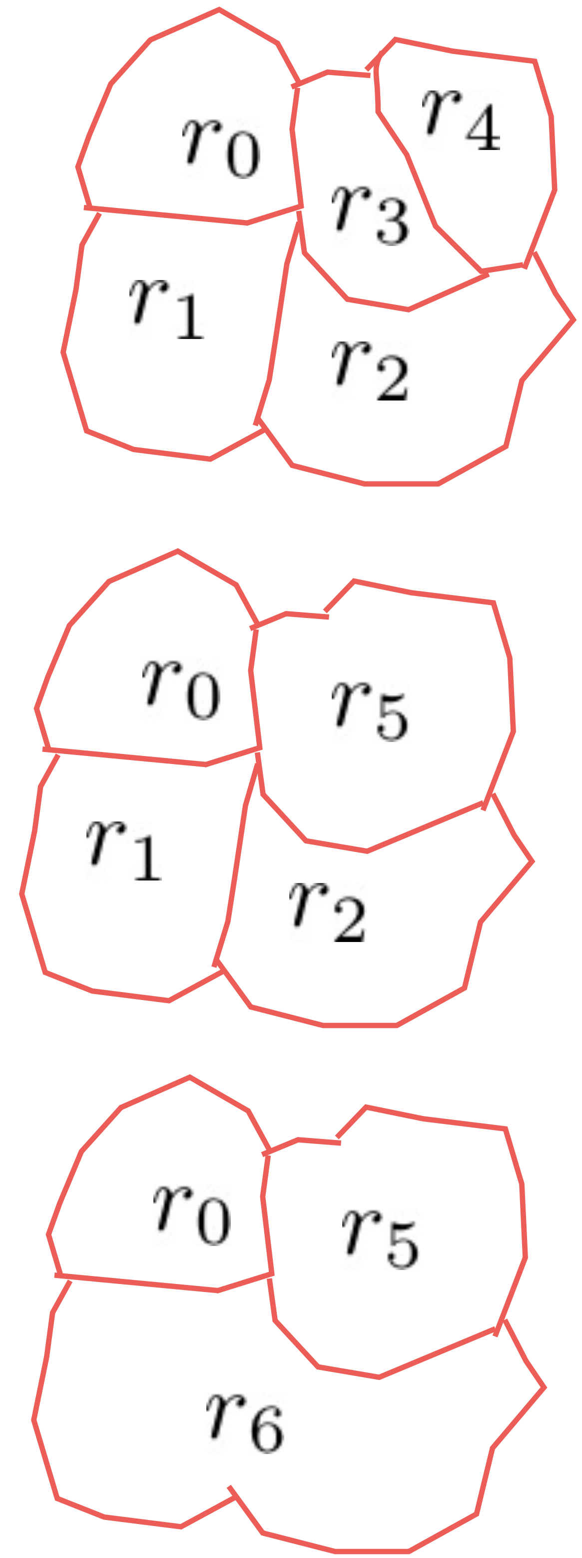
**Input**: (colour) image
**Output**: Set of object location hypotheses $L$

Obtain initial regions $R = \{r_1, \cdots, r_n\}$ using [13]
Initialise similarity set $S = \emptyset$
**foreach** *Neighbouring region pair* $(r_i, r_j)$ **do**
    Calculate similarity $s(r_i, r_j)$
    $S = S \cup s(r_i, r_j)$

**while** $S \neq \emptyset$ **do**
    Get highest similarity $s(r_i, r_j) = \max(S)$
    Merge corresponding regions $r_t = r_i \cup r_j$
    Remove similarities regarding $r_i : S = S \setminus s(r_i, r_*)$
    Remove similarities regarding $r_j : S = S \setminus s(r_*, r_j)$
    Calculate similarity set $S_t$ between $r_t$ and its neighbours
    $S = S \cup S_t$
    $R = R \cup r_t$

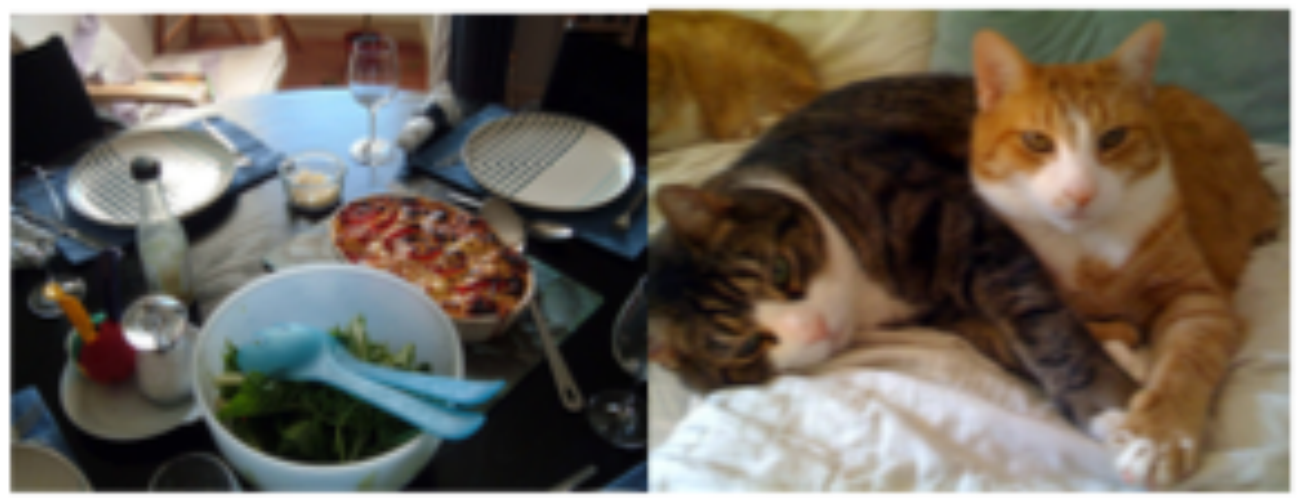
Extract object location boxes $L$ from all regions in $R$

---




Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function



(a) (b) Color

(c) Texture (d) Part

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function

- Color? Texture? Part?
- No single strategy to group regions
- Need to **diversify** by using **complementary** similarity measures

# Similarity Function

- Color similarity: $s_{colour}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k)$
- Normalized color histogram with 25 bins: $C_i = \{c_i^1, \cdots, c_i^n\}$
- Propagate through the hierarchy:

$$C_t = \frac{\text{size}(r_i) \times C_i + \text{size}(r_j) \times C_j}{\text{size}(r_i) + \text{size}(\text{r}_\text{j})}$$

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function

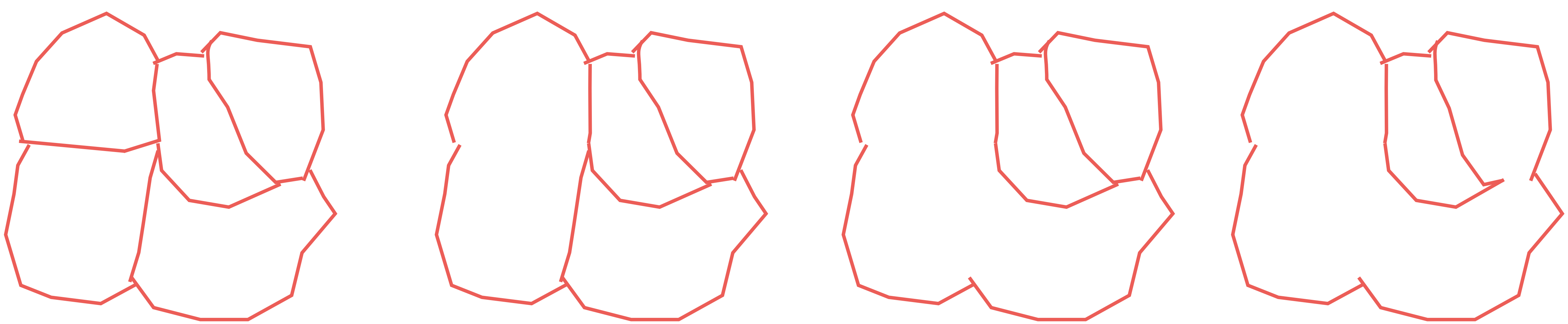
- Texture similarity: $s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k)$
- Take Gaussian derivatives in 8 orientations, and extract histogram with bin size=10: $T_i = \{t_i^1, \cdots, t_i^n\}$

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function

- Size similarity: $s_{size}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(r_j)}{\text{size}(im)}$
- Merge small regions first

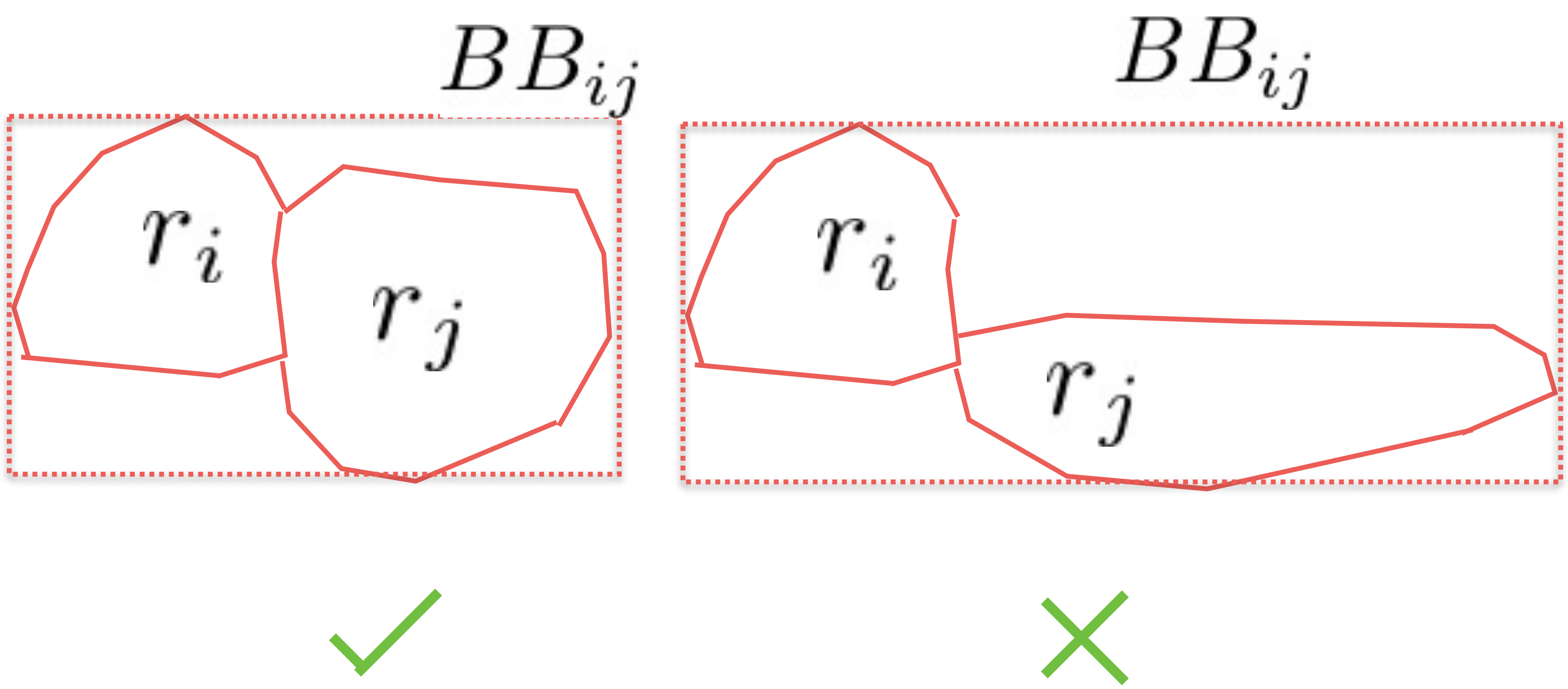
Prevent a big region eating small regions



# Similarity Function

- Fill similarity:

$$fill(r_i, r_j) = 1 - \frac{\text{size}(BB_{ij}) - \text{size}(r_i) - \text{size}(r_i)}{\text{size}(im)}$$




Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Similarity Function

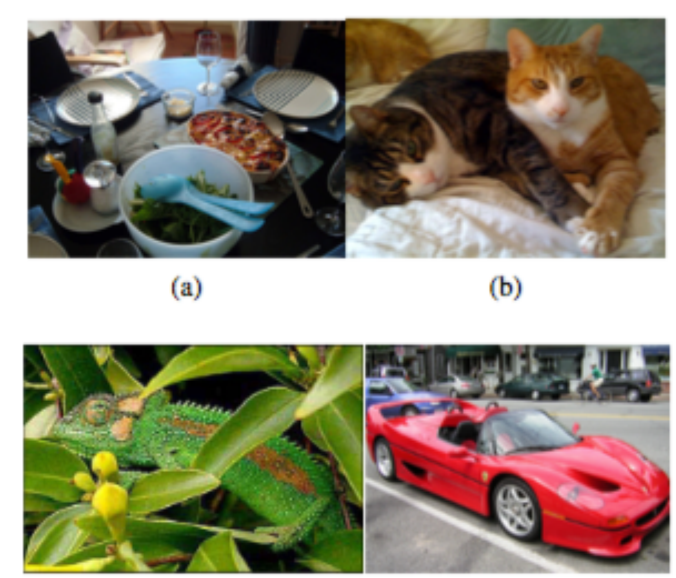
- Combine them:

$$
\begin{aligned}
s(r_i, r_j) \quad = \quad & a_1 s_{colour}(r_i, r_j) + a_2 s_{texture}(r_i, r_j) + \\
& a_3 s_{size}(r_i, r_j) + a_4 s_{fill}(r_i, r_j),
\end{aligned}
$$

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

a=[1,1,1,1] => C+T+S+F

a=[0,1,1,1] => T+S+F



(a) (b) (c) (d)

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Complementary Color Space

- Also diversify in color space

| *colour channels* | R | G | B | I | V | L | a | b | S | r | g | C | H |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Light Intensity | - | - | - | - | - | - | +/- | +/- | + | + | + | + | + |
| Shadows/shading | - | - | - | - | - | - | +/- | +/- | + | + | + | + | + |
| Highlights | - | - | - | - | - | - | - | - | - | - | - | +/- | + |

| *colour spaces* | RGB | I | Lab | rgI | HSV | rgb | C | H |
|---|---|---|---|---|---|---|---|---|
| Light Intensity | - | - | +/- | 2/3 | 2/3 | + | + | + |
| Shadows/shading | - | - | +/- | 2/3 | 2/3 | + | + | + |
| Highlights | - | - | - | - | 1/3 | - | +/- | + |

invariance

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Evaluation

# Metrics

- Average Best Overlap (ABO)

$$\mathrm{ABO} = \frac{1}{|G^c|} \sum_{g_i^c \in G^c} \max_{l_j \in L} \mathrm{Overlap}(g_i^c, l_j)$$

$$\mathrm{Overlap}(g_i^c, l_j) = \frac{\mathrm{area}(g_i^c) \cap \mathrm{area}(\mathrm{l_j})}{\mathrm{area}(g_i^c) \cup \mathrm{area}(\mathrm{l_j})}$$

- Mean Average Best Overlap (MABO)
  mean of ABO over all classes

# Some Examples

(a) Bike: 0.863

(b) Cow: 0.874

(c) Chair: 0.884


(d) Person: 0.882

(e) Plant: 0.873

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Flat v.s Hierarchy

| threshold $k$ in [13] | MABO | # windows |
|---|---|---|
| Flat [13] $k = 50, 150, \cdots, 950$ | 0.659 | 387 |
| Hierarchical (this paper) $k = 50$ | 0.676 | 395 |
| Flat [13] $k = 50, 100, \cdots, 1000$ | 0.673 | 597 |
| Hierarchical (this paper) $k = 50, 100$ | 0.719 | 625 |

Hierarchy is good!

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Diversification Strategies

| Version | Diversification Strategies | MABO | # win | # strategies | time (s) |
|---|---|---|---|---|---|
| Single Strategy | HSV<br>C+T+S+F<br>$k = 100$ | 0.693 | 362 | 1 | 0.71 |
| Selective Search Fast | HSV, Lab<br>C+T+S+F, T+S+F<br>$k = 50, 100$ | 0.799 | 2147 | 8 | 3.79 |
| Selective Search Quality | HSV, Lab, rgI, H, I<br>C+T+S+F, T+S+F, F, S<br>$k = 50, 100, 150, 300$ | 0.878 | 10,108 | 80 | 17.15 |

Diversification is good!

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Compare to Other Methods

| method | recall | MABO | # windows |
|---|---|---|---|
| Arbelaez *et al.* [3] | 0.752 | $0.649 \pm 0.193$ | 418 |
| Alexe *et al.* [2] | 0.944 | $0.694 \pm 0.111$ | 1,853 |
| Harzallah *et al.* [16] | 0.830 | - | 200 per class |
| Carreira and Sminchisescu [4] | 0.879 | $0.770 \pm 0.084$ | 517 |
| Endres and Hoiem [9] | 0.912 | $0.791 \pm 0.082$ | 790 |
| Felzenszwalb *et al.* [12] | 0.933 | $0.829 \pm 0.052$ | 100,352 per class |
| Vedaldi *et al.* [34] | 0.940 | - | 10,000 per class |
| Single Strategy | 0.840 | $0.690 \pm 0.171$ | 289 |
| Selective search "Fast" | 0.980 | $0.804 \pm 0.046$ | 2,134 |
| Selective search "Quality" | 0.991 | $0.879 \pm 0.039$ | 10,097 |

State of the art!

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Contribution and Strength

- Hierarchical grouping and diversification strategies
- Nice trade-off between quality(MABO) and quantity(# window)

# Weakness

- The algorithm for sorting the object hypotheses s.t. the most likely hypothesis comes first

  $r_i^j$: region generated by strategy $j$ in level $i$

  $v_i^j = rand() \times i$

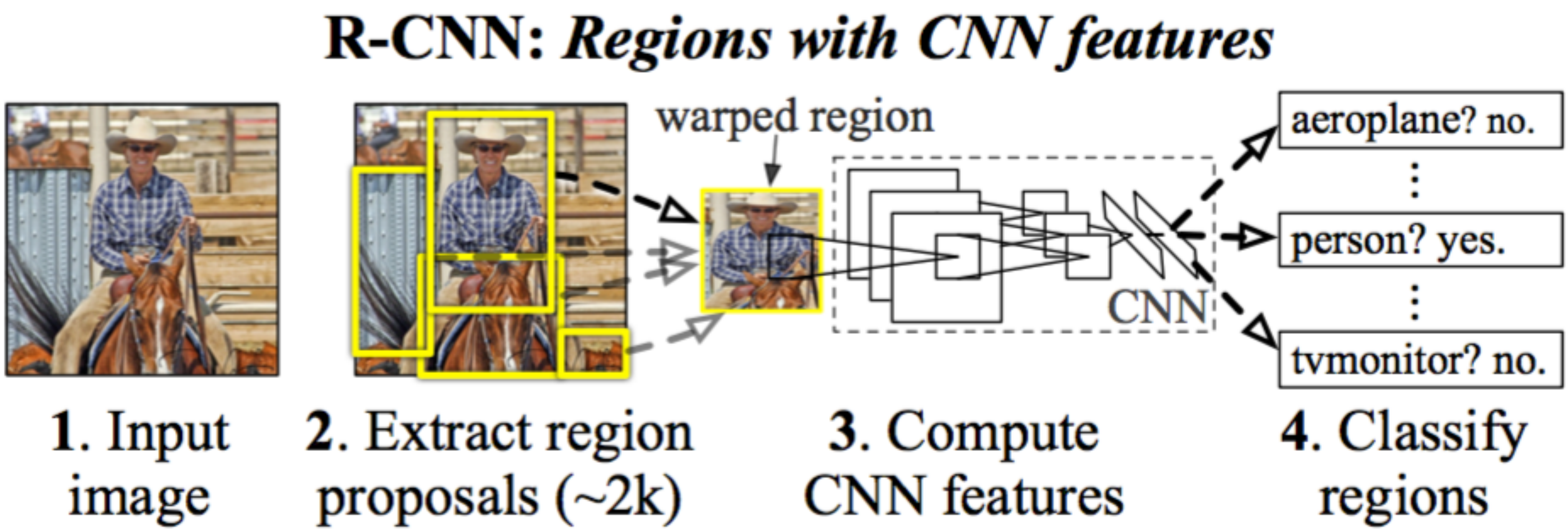
- No evaluation on it?

- Favor large scale but times rand() to prevent over-favor?

# Extension

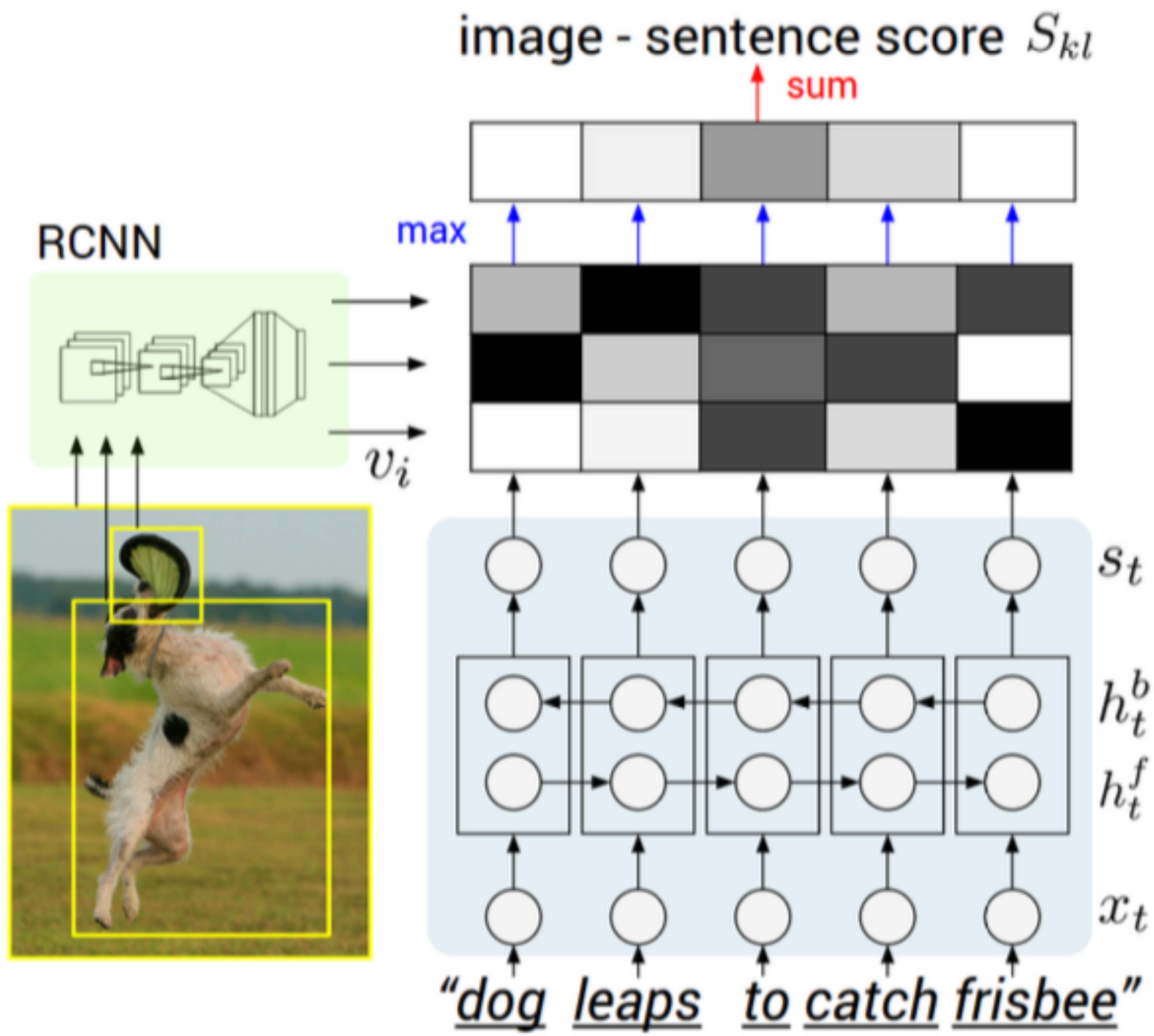
# R-CNN

- Regions with Convolutional Neural Network features




Rich feature hierarchies for accurate object detection and semantic segmentation. R. Girshick et al. CVPR 2013

# Visual-Semantic Alignment




A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3128–3137, 2015.

# Reference

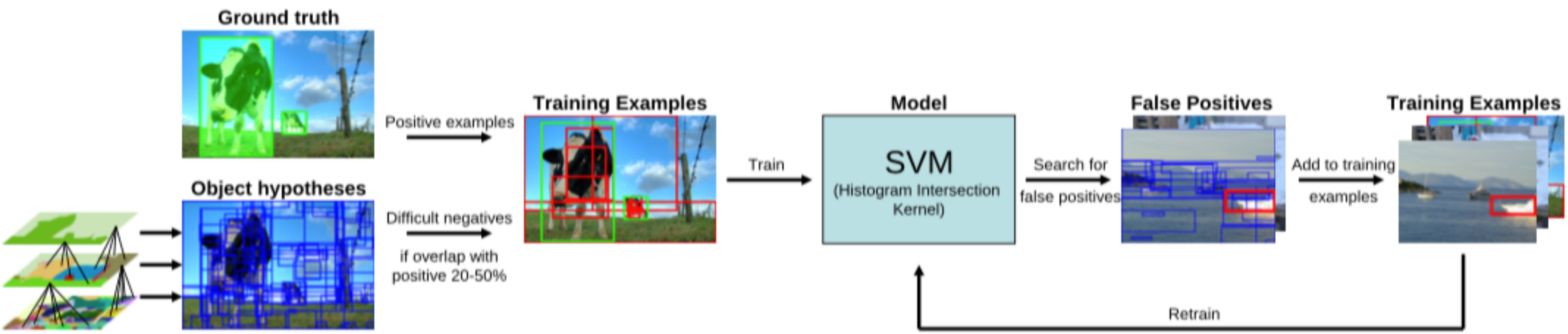

- J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154–171, 2013

- R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

- A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3128–3137, 2015.

# Appendix

# Application on Object Detection




Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Diversification Strategies

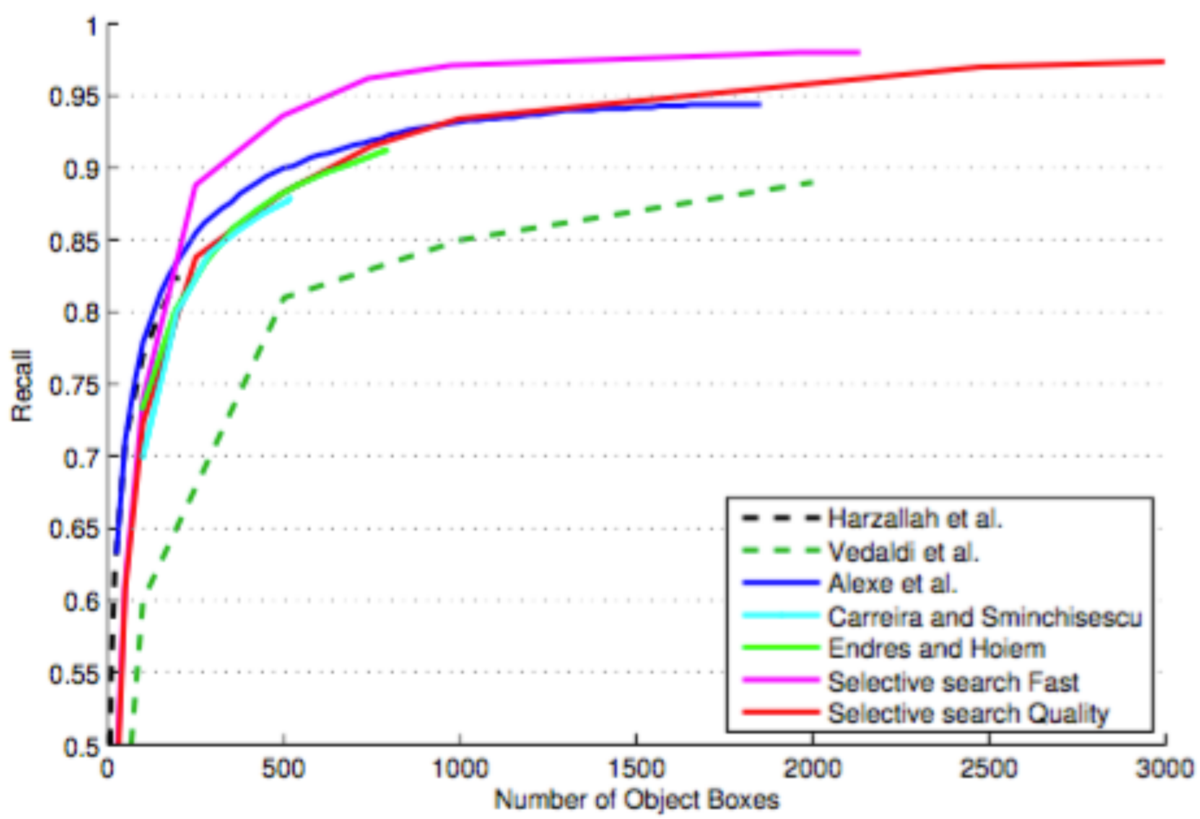
| Similarities | MABO | # box |
|---|---|---|
| C | 0.635 | 356 |
| T | 0.581 | 303 |
| S | 0.640 | 466 |
| F | 0.634 | 449 |
| C+T | 0.635 | 346 |
| C+S | 0.660 | 383 |
| C+F | 0.660 | 389 |
| T+S | 0.650 | 406 |
| T+F | 0.638 | 400 |
| S+F | 0.638 | 449 |
| C+T+S | 0.662 | 377 |
| C+T+F | 0.659 | 381 |
| C+S+F | 0.674 | 401 |
| T+S+F | 0.655 | 427 |
| C+T+S+F | 0.676 | 395 |

| Colours | MABO | # box |
|---|---|---|
| HSV | 0.693 | 463 |
| I | 0.670 | 399 |
| RGB | 0.676 | 395 |
| rgI | 0.693 | 362 |
| Lab | 0.690 | 328 |
| H | 0.644 | 322 |
| rgb | 0.647 | 207 |
| C | 0.615 | 125 |

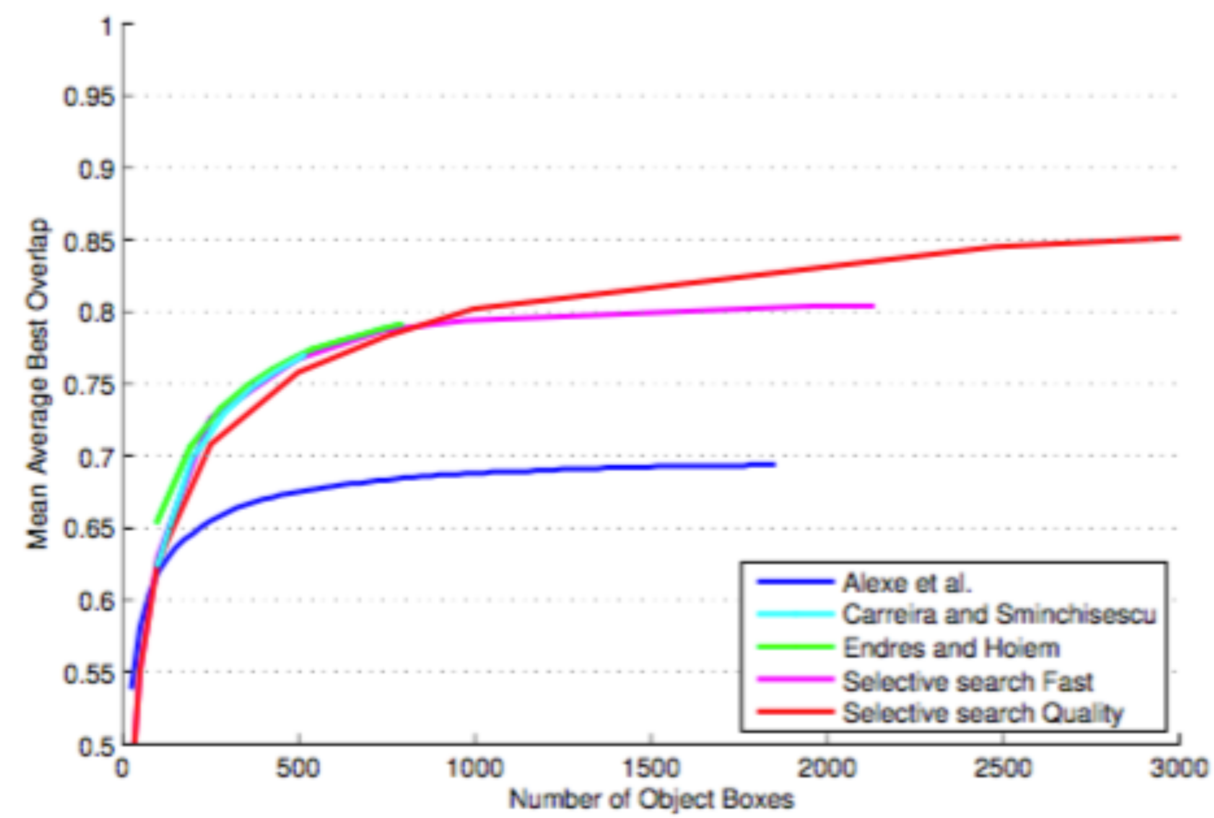
| Thresholds | MABO | # box |
|---|---|---|
| 50 | 0.676 | 395 |
| 100 | 0.671 | 239 |
| 150 | 0.668 | 168 |
| 250 | 0.647 | 102 |
| 500 | 0.585 | 46 |
| 1000 | 0.477 | 19 |

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013

# Trade-off between Quality and Quantity




(a) Trade-off between number of object locations and the Pascal Recall criterion.




(b) Trade-off between number of object locations and the MABO score.

Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013