Selective Search for Object Recognition

Uijlings et al. (IJCV 2013)

Some figures are from http://vision.stanford.edu/teaching/cs231b_spring1415/slides/ssearch_schuyler.pdf

Object Recognition

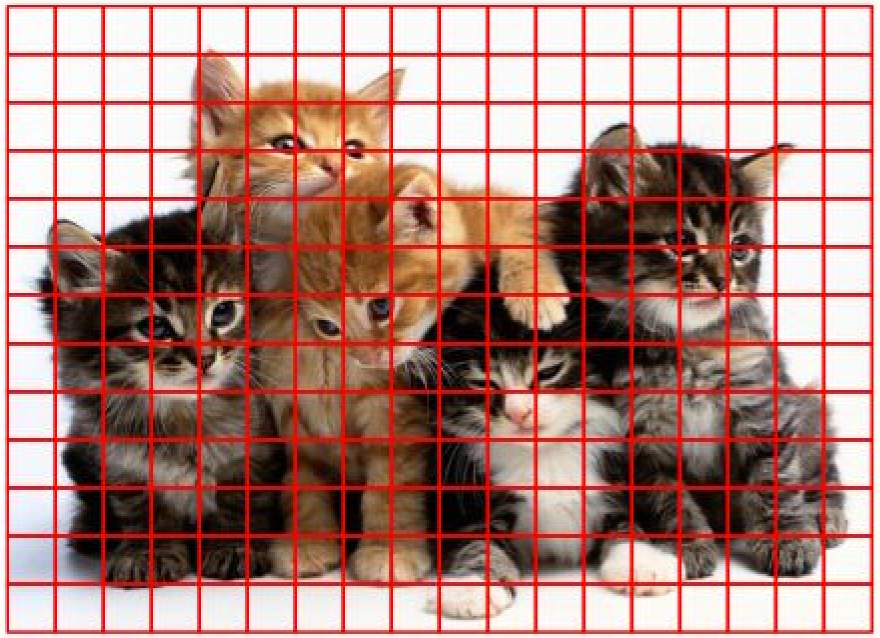
Goal:

Find Object and Recognize it

Contribution

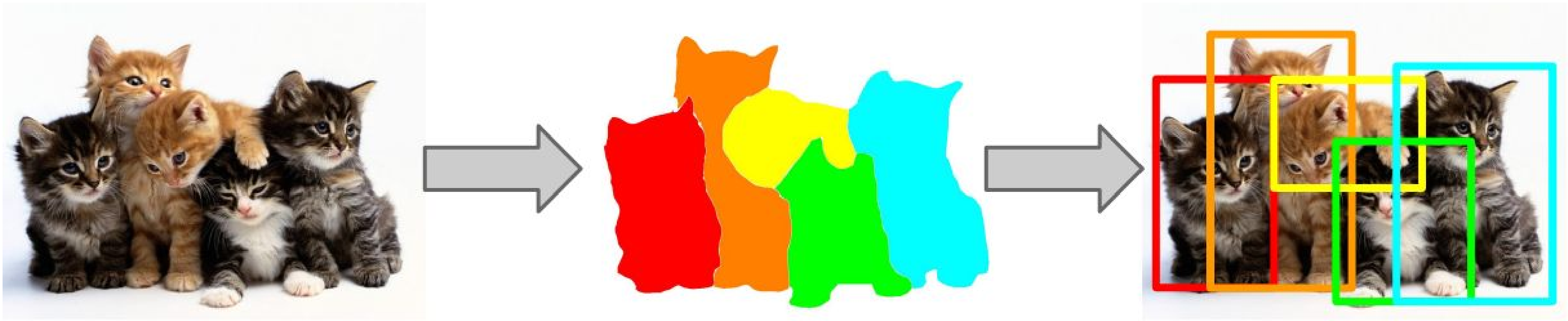
Exhaustive Search

- Exhaustively grid search all possible locations
- Very Slow!! (Imagine you need to process many images)



Segmentation

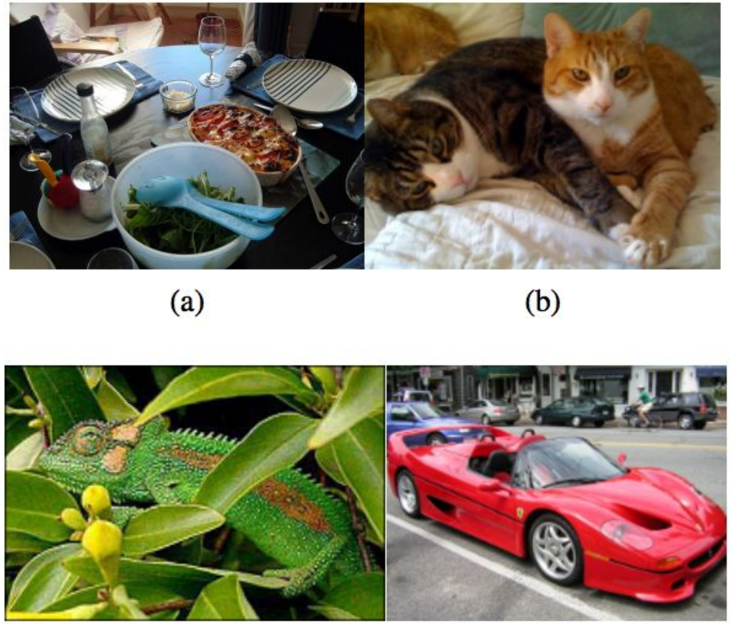
- Run detection before recognition
- Many existing segmentation algorithms



Difficulties of Segmentation

- No single golden criteria for segmentation

- Scale
- Color
- Texture
- Enclosure



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

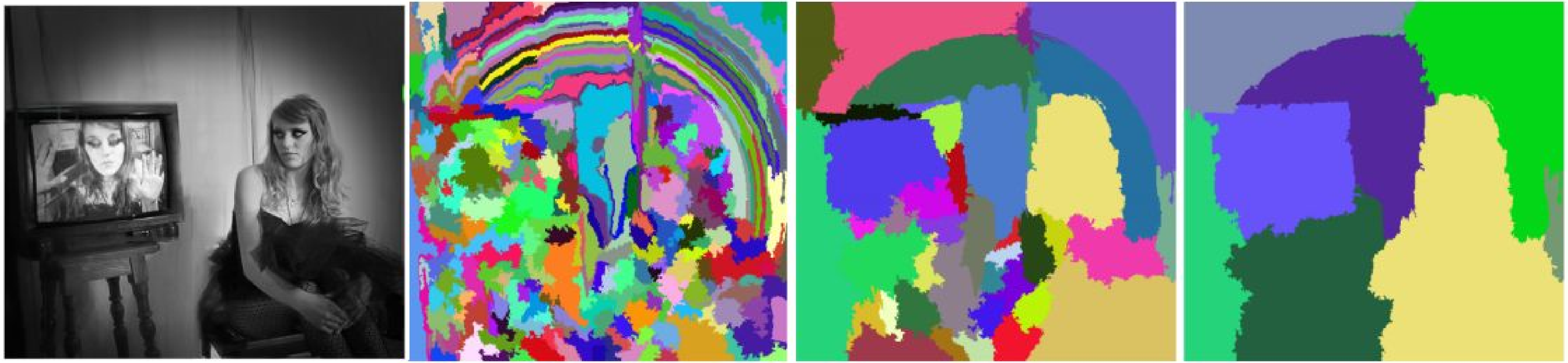
Selective Search

Goals:

- **Capture all scales** - How could we know the size of object?
- **Diversifications** - Different criteria for segmentation
- **Fast to compute**

Hierarchical Segmentation

- Apply existing algorithms to find sub-segmentations
  - **Small segmentations**
- Recursively combine small segmentations into big segmentations
  - **Big segmentations**



Algorithms

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

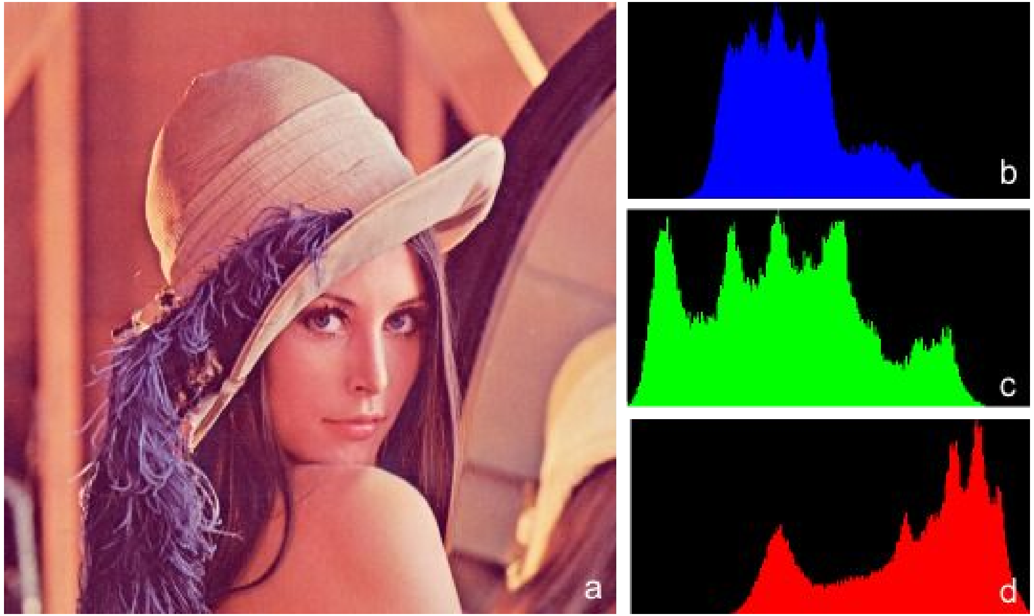
Diversification

- We already capture the scale, how to model different criteria into the algorithm?
- What criteria for combining the segmentations?

Colour Similarity

- Normalized and Histogram Intersection

$$s_{colour}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k)$$



Texture Similarity

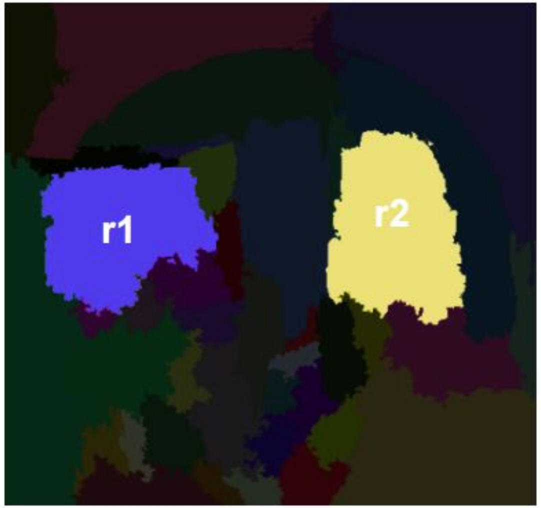
- Extract derivatives in 8 directions for 3 channels
- 10 bins for each, 240 bins in total
- Normalized and Histogram Intersection

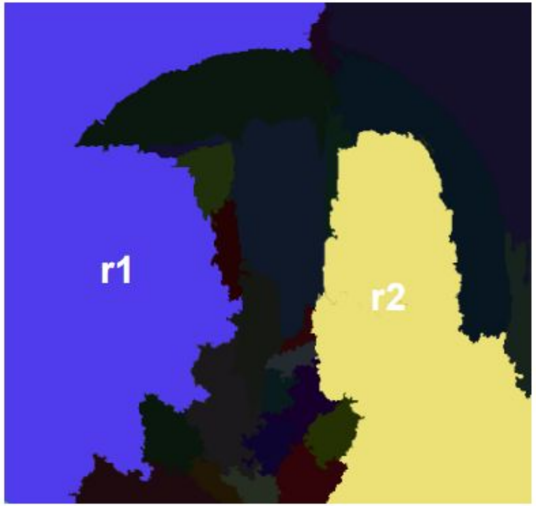
$$s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k).$$

Size Similarity

- We hope to merge two small region into a large segmentation

$$s_{size}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(\text{r}_\text{j})}{\text{size}(im)}$$

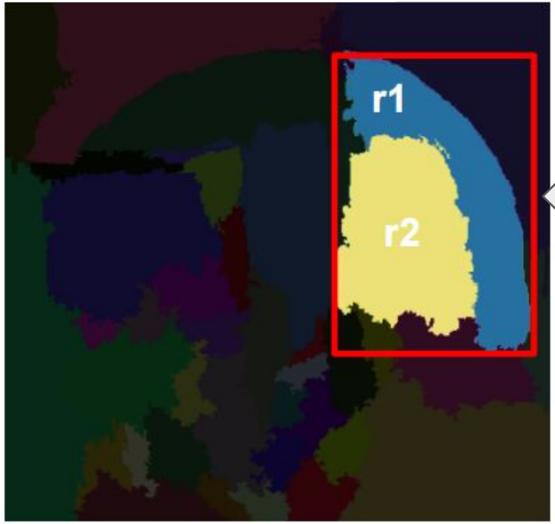




Shape Compatibility

- Whether two segmentations **fit** each other?

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



Bonding Box

A mixture Approach

Final Score:

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

Evaluation I - Object Detection

- Average Best Overlap (ABO)

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

Overlap between ground truth and best selected box.

Average of “best overlaps” across all images.

Evaluation I - Object Detection

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

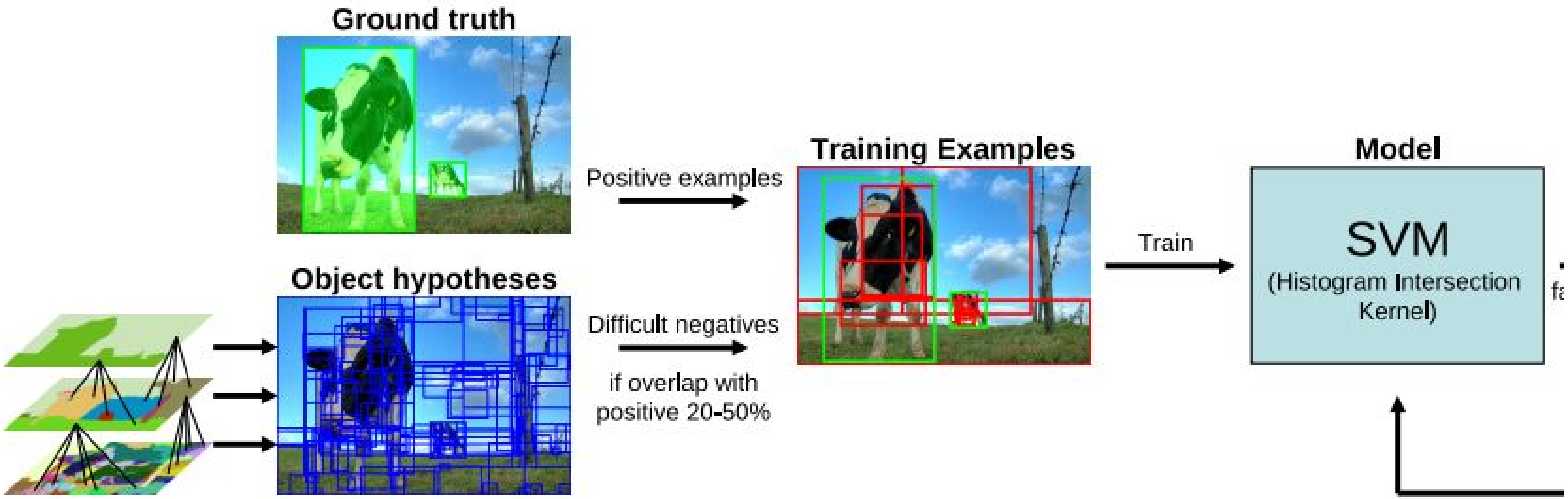
Thresholds	MABO	# box
50	0.676	395
100	0.671	239
150	0.668	168
250	0.647	102
500	0.585	46
1000	0.477	19

Efficiency or Effectiveness?

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

Evaluation II - Object Recognition

Approach: Selective Search + SIFT + SVM



Evaluation II - Object Recognition

PASCAL VOC 2010

System	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv
NLPR	.533	**.553**	**.192**	**.210**	.300	.544	.467	.412	**.200**	.315	.207	.303	.486	.553	.465	.102	.344	.265	**.503**	.403
MIT UCLA [38]	.542	.485	.157	.192	.292	**.555**	.435	.417	.169	.285	.267	.309	.483	.550	.417	.097	.358	.308	.472	.408
NUS	.491	.524	.178	.120	.306	.535	.328	.373	.177	.306	.277	.295	**.519**	**.563**	.442	.096	.148	.279	.495	.384
UoCTTI [12]	.524	.543	.130	.156	**.351**	.542	**.491**	.318	.155	.262	.135	.215	.454	.516	**.475**	.091	.351	.194	.466	.380
This paper	**.562**	.424	.153	.126	.218	.493	.368	**.461**	.129	**.321**	**.300**	**.365**	.435	.529	.329	**.153**	**.411**	**.318**	.470	**.448**

Conclusion

- Hierarchical Segmentation woks
- State-of-the-art algorithm before 2015
- Still many decisions to be made