



Mining Discriminative Adjectives and Prepositions for Natural Scene Recognition

Bangpeng Yao¹, Juan Carlos Niebles^{2,3}, and Li Fei-Fei¹

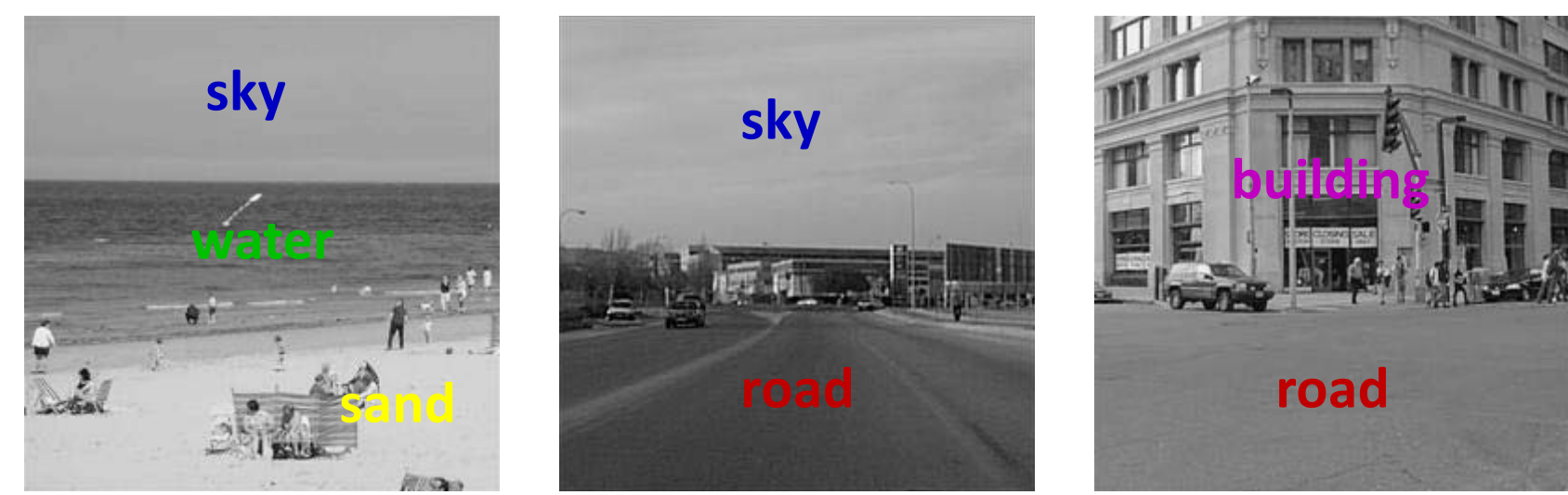
¹Department of Computer Science, Princeton University, NJ 08540, USA ²Department of Electrical Engineering, Princeton University, NJ 08544, USA

³Robotics and Intelligent Systems Group, Universidad del Norte, Colombia



Summary

Natural scene recognition:



Coast: sky above water & water above sand;
Highway: sky above road;
Inside city: building above road.

Our Objective:

- Use information encoded by adjectives (e.g. bigger, smoother) and prepositions (e.g. above, near) for natural scene recognition.

Our Method:

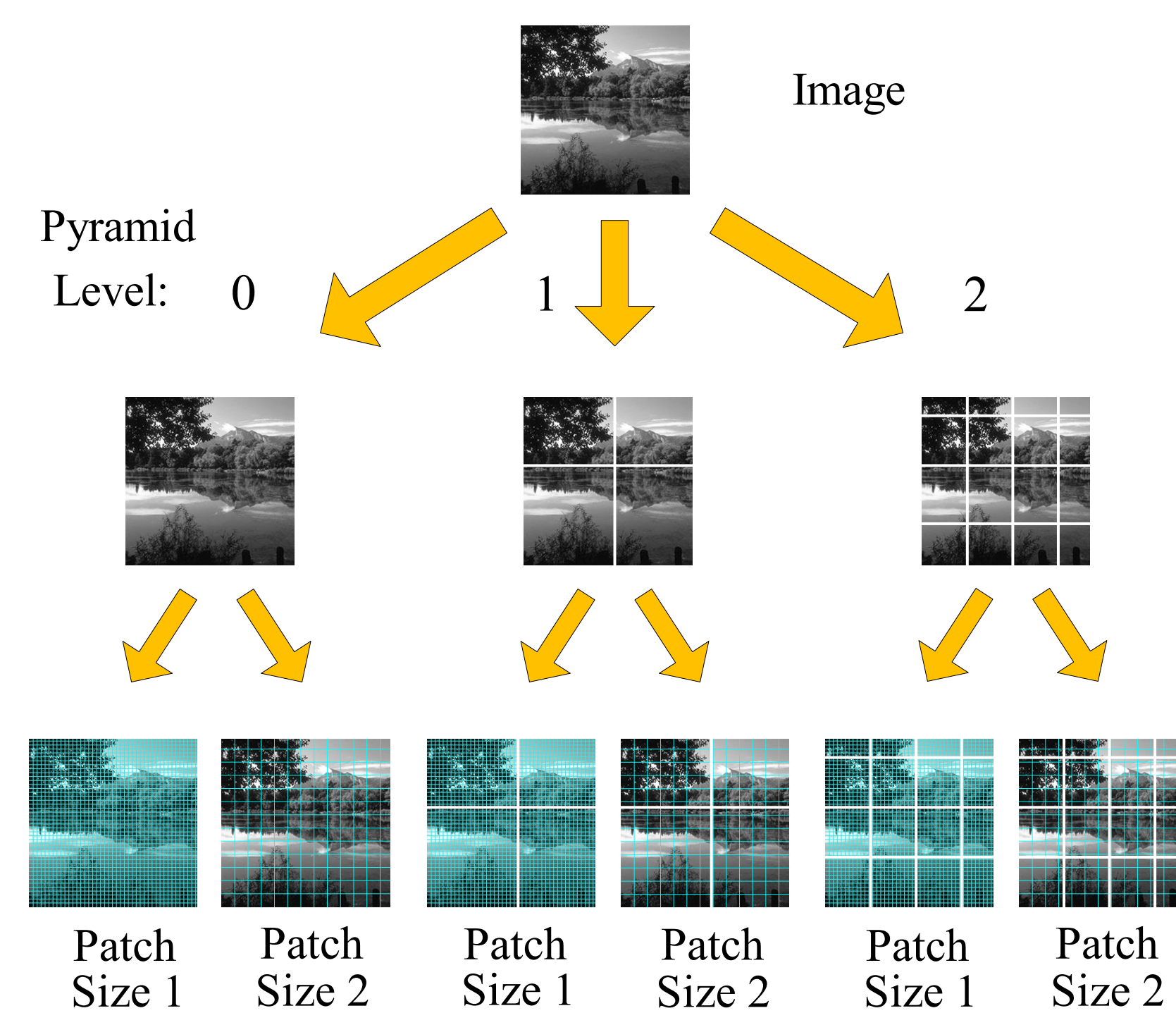
- Define the adjectives and prepositions based on an image representation of multi-scale patches and spatial pyramid;
- An apriori algorithm to mine discriminative adjectives & prepositions.

Image Representation and Adjective & Preposition Definition

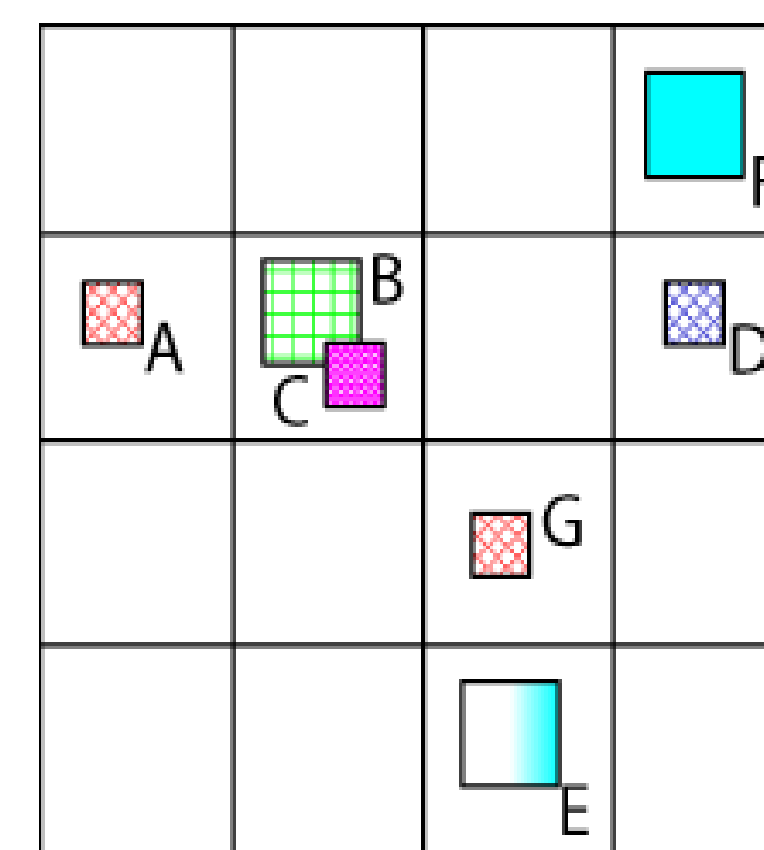
The adjectives and prepositions can be formed by simply considering **particular spatial relationships** among patches with **specific appearances** and **specific size**.

Image representation:

- Multi-scale local patches; -- (size)
 - Codewords of SIFT descriptor to represent each local patch. -- (appearance)
- Spatial pyramid. -- (location)



Examples of adjectives & prepositions: **(Relationship Set - RSet):**

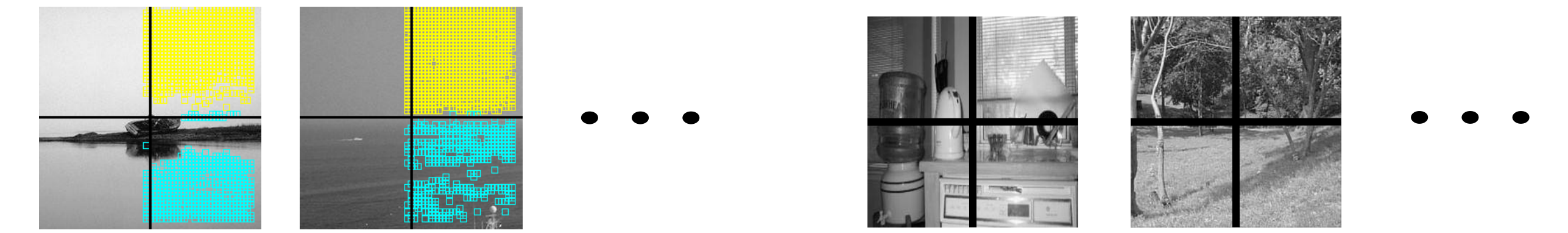


Above (G,E) Larger (B,C)
 Left,Near (A,C) Brighter (D,A)
 Left,Far (A,D) Smoother (F,E)

(A,B,C) – {patches B and C are on the right of A} AND {B is larger than C}

Mining Discriminative Relationship Sets (Rsets)

Discriminative RSet:



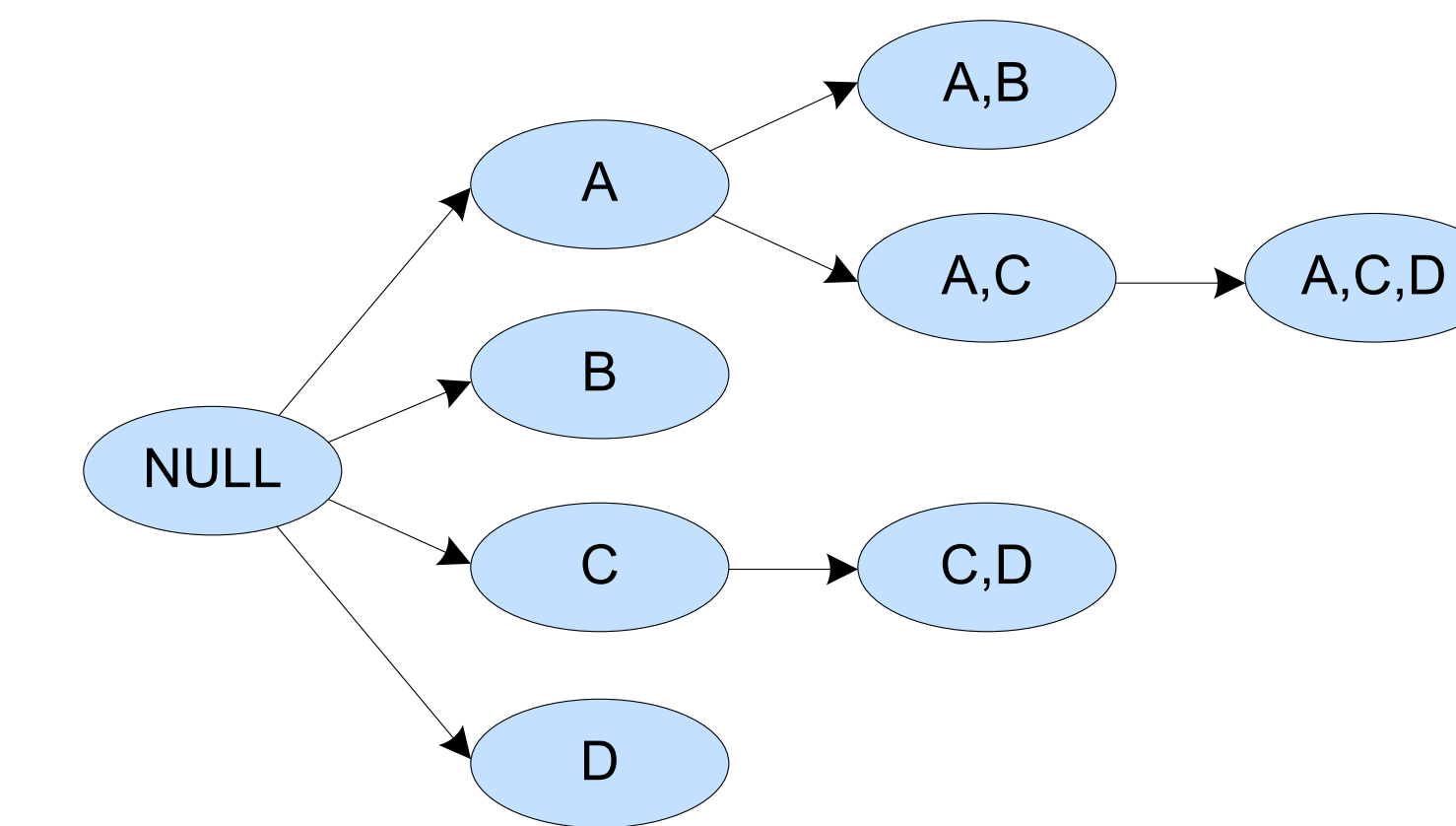
Appear frequently on images of one class. Rarely appear on images of other classes.

No. of occurrence of RSet \mathcal{R} on image i :

$$Supp(\mathcal{R}, c) = \frac{\sum_{c_i=c} s_i}{\sum_{c_i=c} 1} \quad Conf(\mathcal{R}, c) = \frac{Supp(\mathcal{R}, c)}{\text{Avg}_{c' \neq c} Supp(\mathcal{R}, c')}$$

Find all RSets with large $Supp$ & small $Conf$ – **Computationally too expensive.**

Data mining – the Apriori algorithm:



Input: Support threshold T_{Supp} and confidence threshold T_{Conf} .

```

foreach Class do
  Scan all RUnits, select the RUnits with support values larger than  $T_{Supp}$  as 1-RSets;
  for  $p = 2$  to  $P$  do
    Generate candidate  $p$ -RSets based on the selected  $(p-1)$ -RSets;
    Scan all candidate  $p$ -RSets and remove the  $p$ -RSets if the support values are smaller than  $T_{Supp}$ ;
    if The number of  $p$ -RSets < 2 then
      break;
    end
  end
end
Scan all selected RSets, remove the RSets whose confidence values are smaller than  $T_{Conf}$ .
end
    
```

Experiments

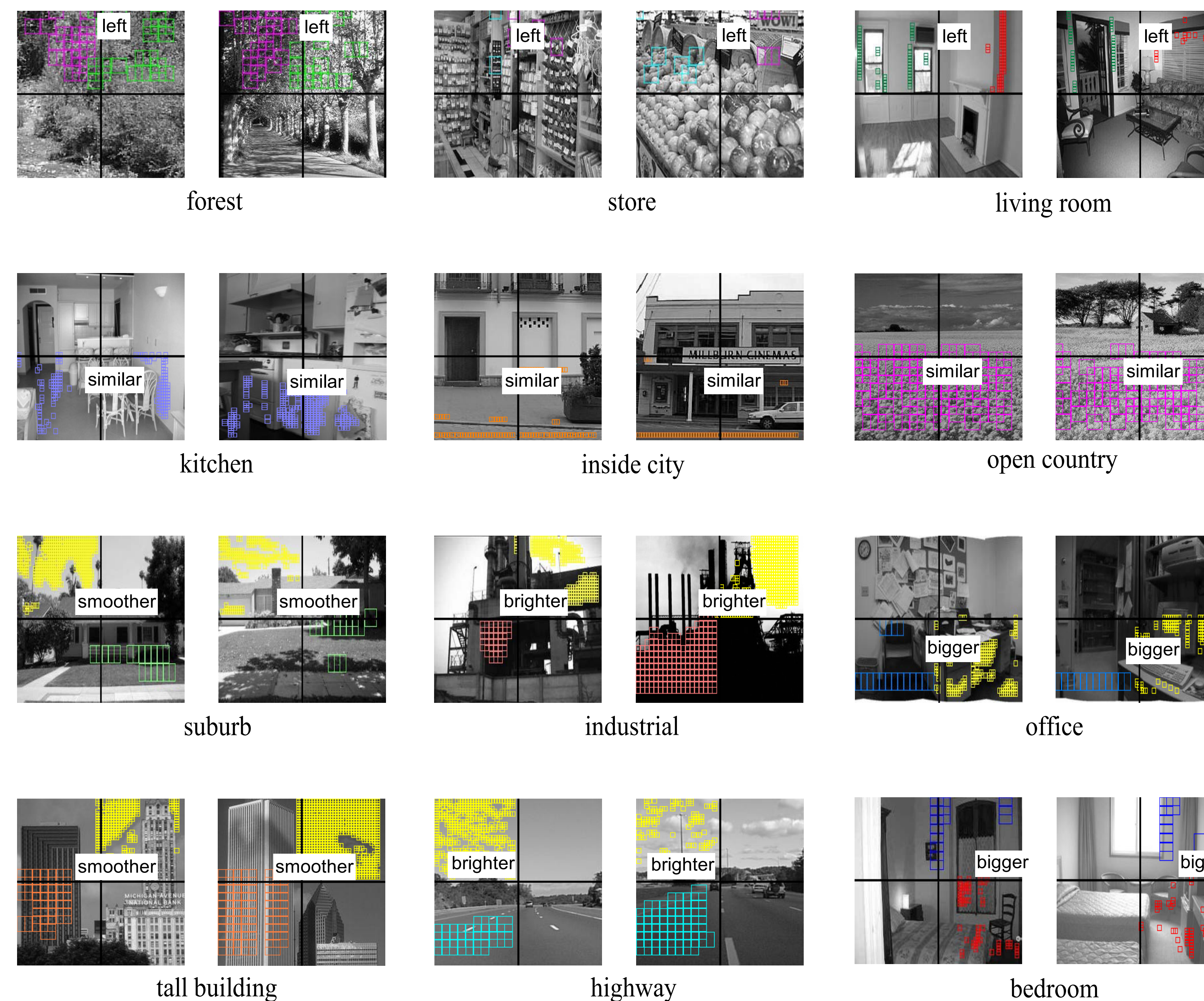
Classifier:

- Input:
- Feature vector:
- Discriminative RSets:
- Classifier: SVM

Dataset:

- 15 classes of natural scenes;
- [Lazebnik et al., CVPR 2006].

bedroom	.68	.01	.04	.06	.14	.01	.03	.01	.01	.01	.01	.01	.01	.01	.01
suburb	.98	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
industrial	.01	.01	.72	.02	.05	.01	.03	.01	.01	.01	.01	.01	.01	.01	.01
kitchen	.06	.03	.75	.06	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
living room	.16	.01	.03	.11	.60	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
coast	.01	.01	.01	.01	.01	.78	.05	.02	.12	.01	.01	.01	.01	.01	.01
forest	.01	.01	.01	.01	.01	.01	.98	.01	.01	.01	.01	.01	.01	.01	.01
highway	.01	.01	.01	.01	.01	.01	.01	.91	.01	.01	.01	.01	.01	.01	.01
inside city	.02	.01	.01	.01	.01	.01	.01	.01	.87	.01	.01	.01	.01	.01	.01
mountain	.01	.01	.01	.01	.02	.01	.01	.01	.01	.91	.03	.01	.01	.01	.01
open country	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.12	.03	.05	.07	.71
street	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.95	.01	.01	.01
tall building	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.89	.01	.01
office	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.96	.01
store	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.79



Discussions

Limitations

- Lacks of automatic semantic understanding of the discriminative adjectives & prepositions.

Future Work:

- Further reduce computational complexity;
- Automatically link the mined relationships to semantic descriptions in natural language.

Reference

Bangpeng Yao, Juan Carlos Niebles, and Li Fei-Fei. "Mining Discriminative Adjectives and Prepositions for Natural Scene Recognition." In *Proceedings of the 1st International Workshop on Visual Scene Understanding*, 2009.

<http://vision.cs.princeton.edu/>

