Action Classification: An Integration of Randomization and Discrimination in A Dense Feature Space

Bangpeng Yao, Aditya Khosla, and Li Fei-Fei

Computer Science Department, Stanford University

{bangpeng,aditya86,feifeili}@cs.stanford.edu

Visual Object Classes Challenge 2011 (VOC2011)

Outline

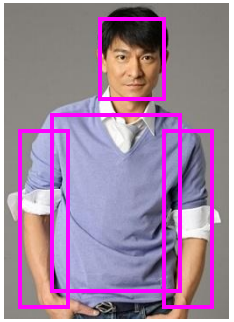
- Action Classification & Intuition
- Our Method
- Our Results
- Conclusion

Outline

- **Action Classification & Intuition**
- Our Method
- Our Results
- Conclusion

Action Classification

Object classification:



Presence of parts and their spatial configurations.

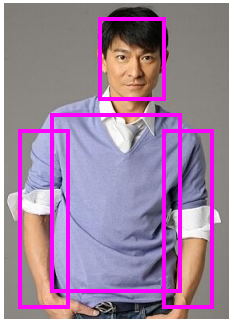
[Lazebnik et al, 2006]
[Fergus et al, 2003]
…

Action Classification



- All images contain humans;

Object classification:



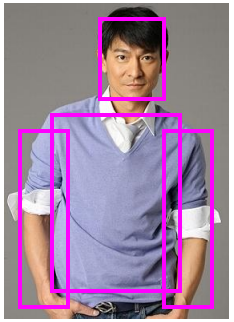
Presence of parts and their spatial configurations.

[Lazebnik et al, 2006]
[Fergus et al, 2003]
…

Action Classification

- All images contain humans;
- Large pose variation;

Object classification:



Presence of parts and their spatial configurations.

[Lazebnik et al, 2006]
[Fergus et al, 2003]
…

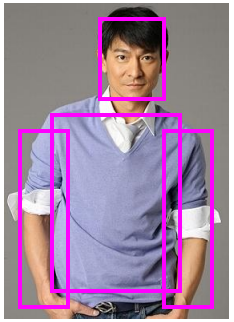
Action Classification



- All images contain humans;
- Large pose variations;
- Objects small or absent;
- Background clutter.

Challenging…

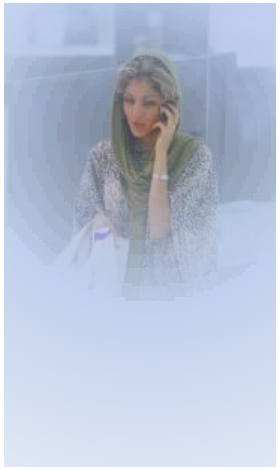
Object classification:

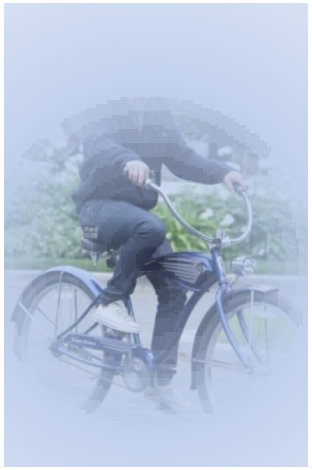


Presence of parts and their spatial configurations.

[Lazebnik et al, 2006]
[Fergus et al, 2003]
…

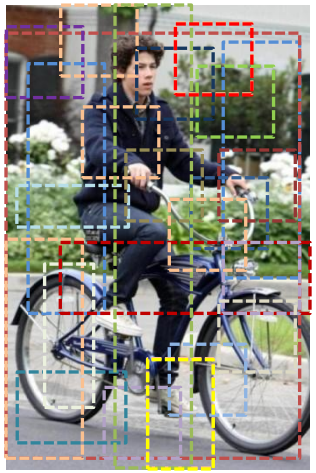
Our Intuition





Focus on image regions that contain the most discriminative information.

Our Intuition


Focus on image regions that contain the most discriminative information.

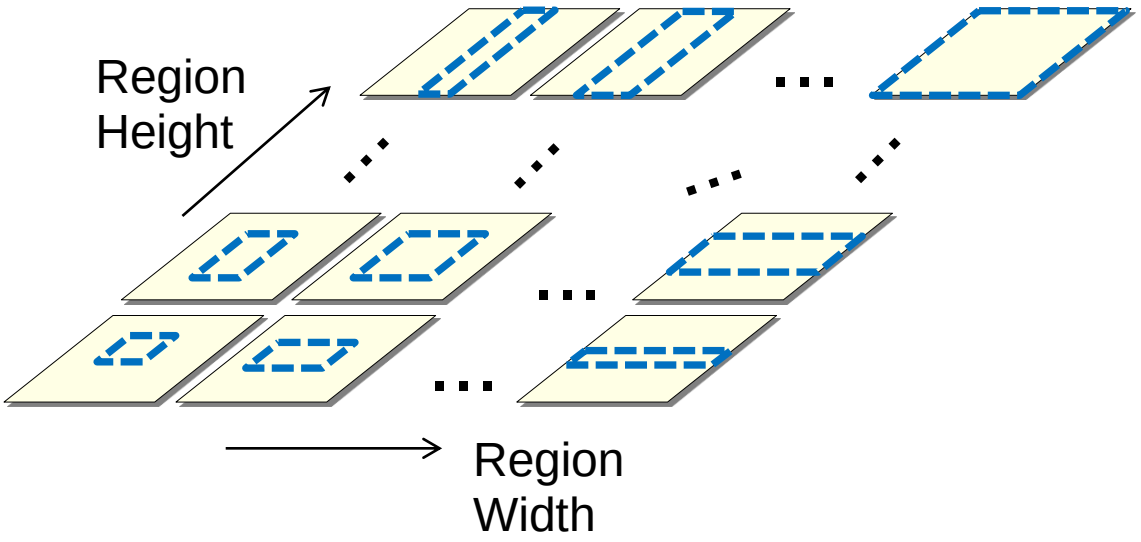
How to represent the features? → **Dense feature space**

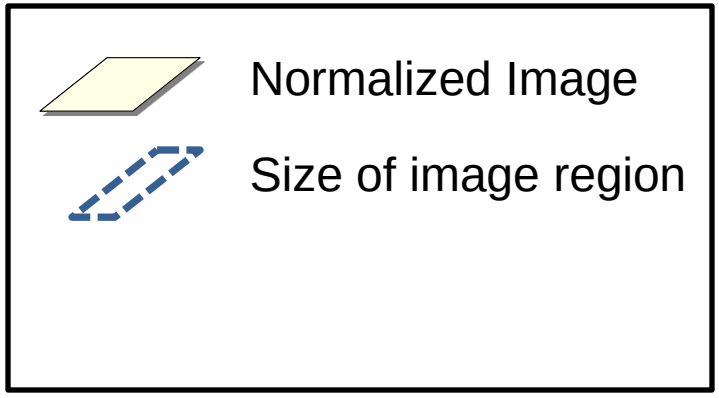
How to explore this feature space? → **Randomization & Discrimination**

Outline

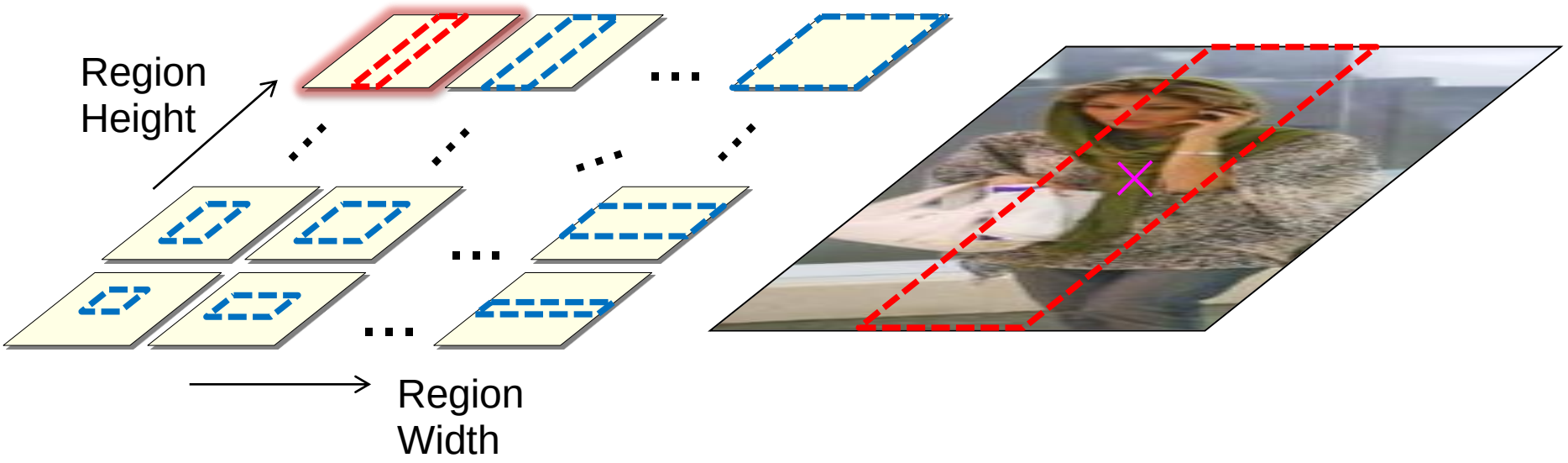
- Action Classification & Intuition
- **Our Method**
- Our Results
- Conclusion

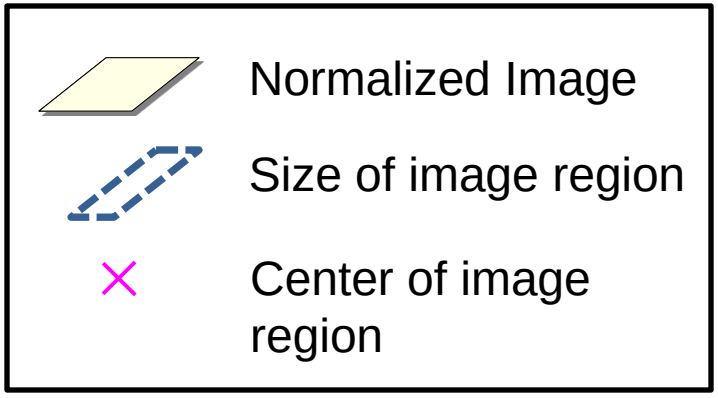
Dense Feature Space



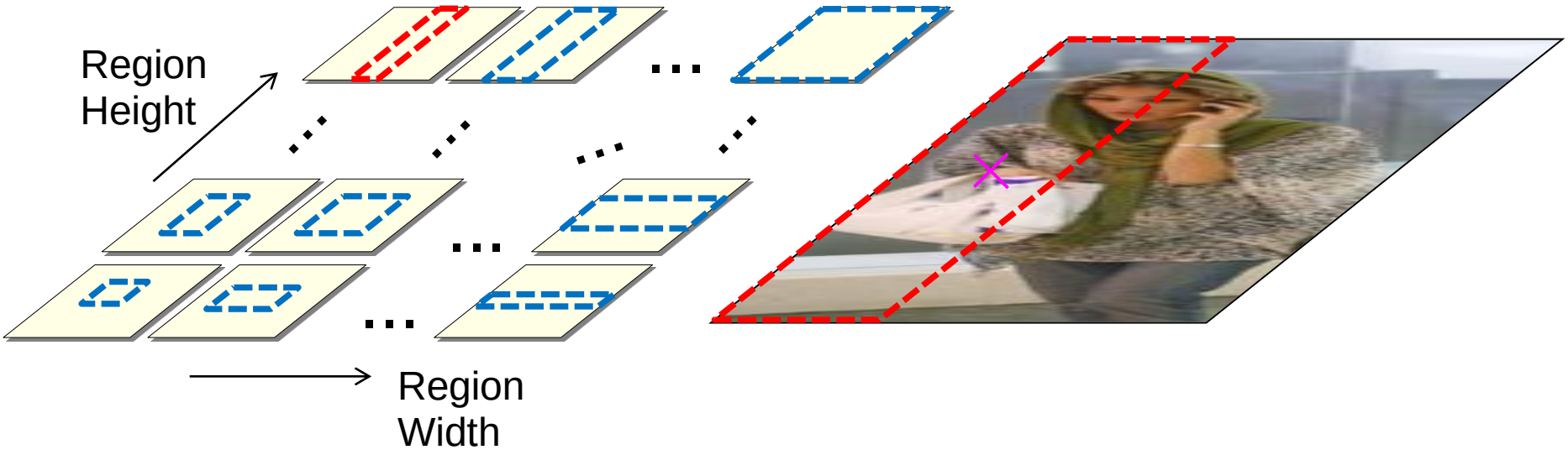


Dense Feature Space





Dense Feature Space

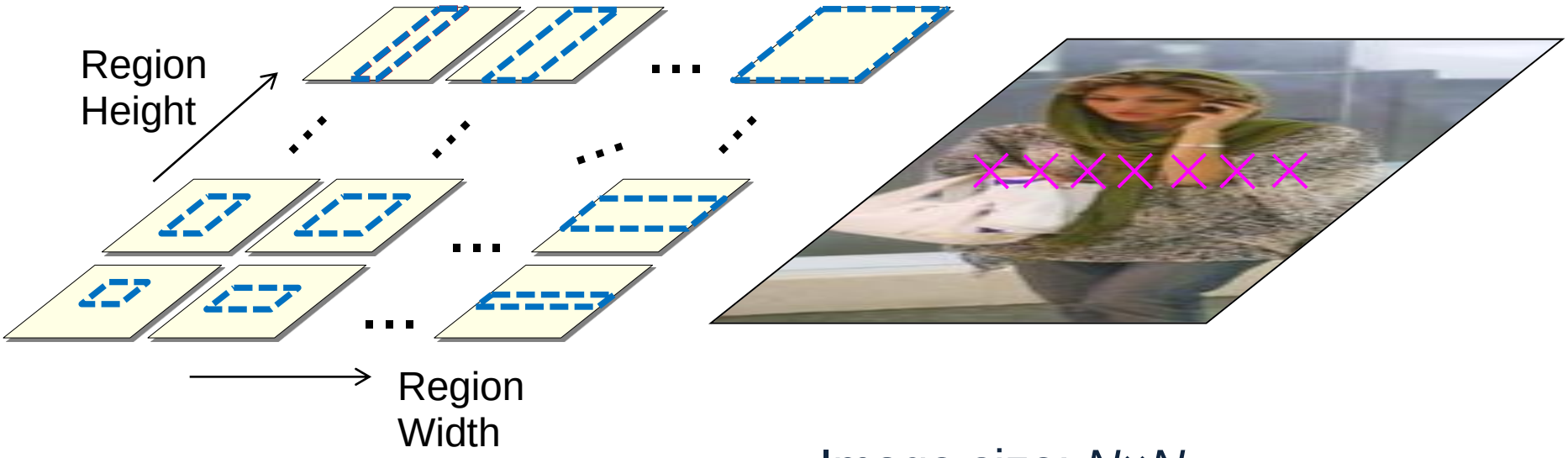


Normalized Image

Size of image region

Center of image region

Dense Feature Space



Normalized Image

Size of image region

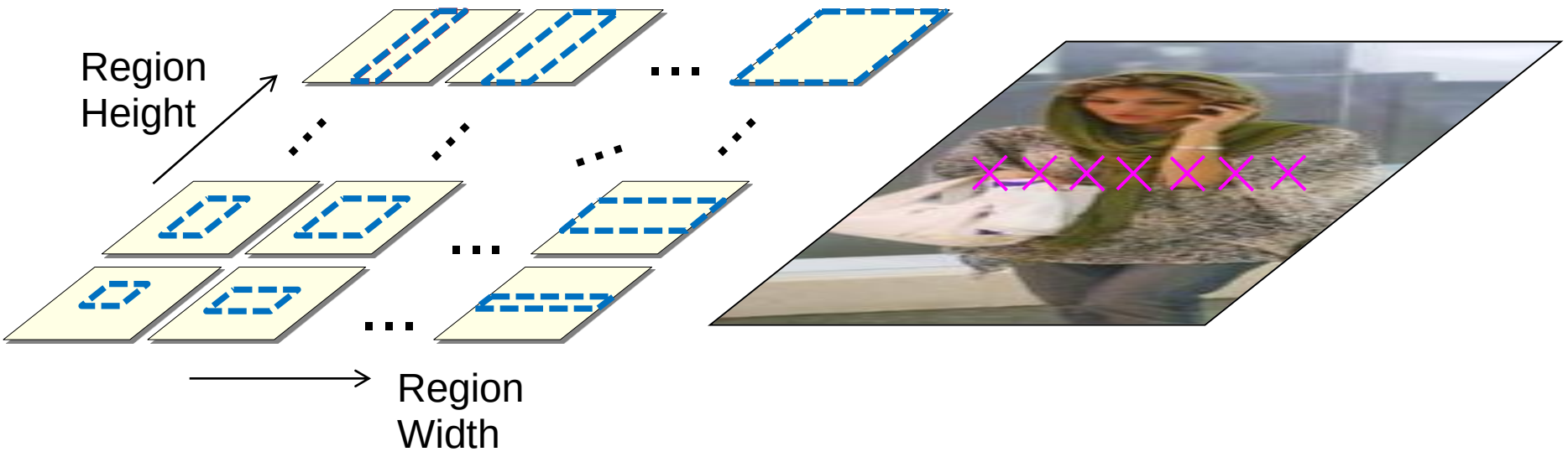
Center of image region

Image size: $N \times N$

Image regions: $O(N^6)$

How can we identify the **discriminative** regions **efficiently** and **effectively**?

Dense Feature Space



Normalized Image

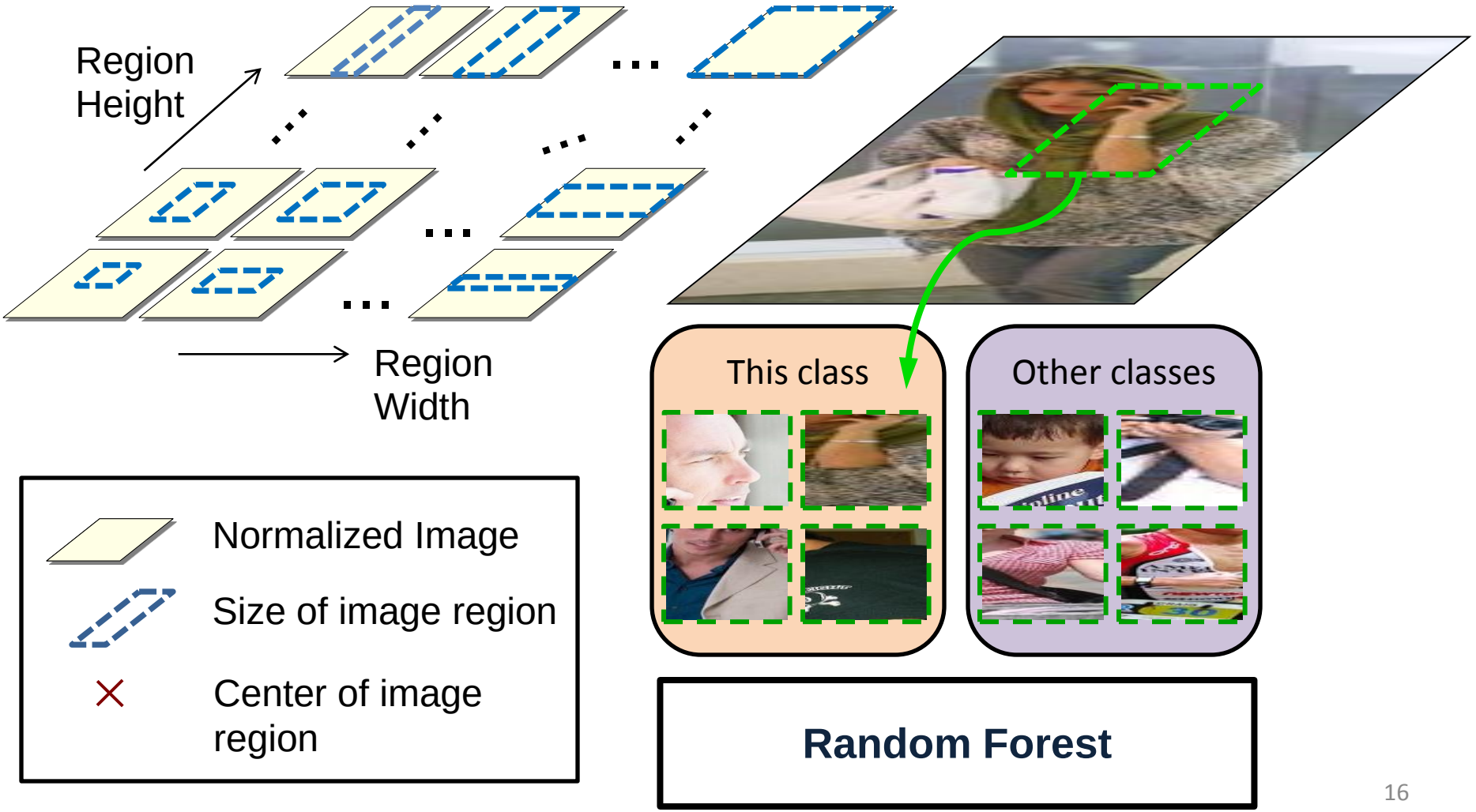
Size of image region

Center of image region

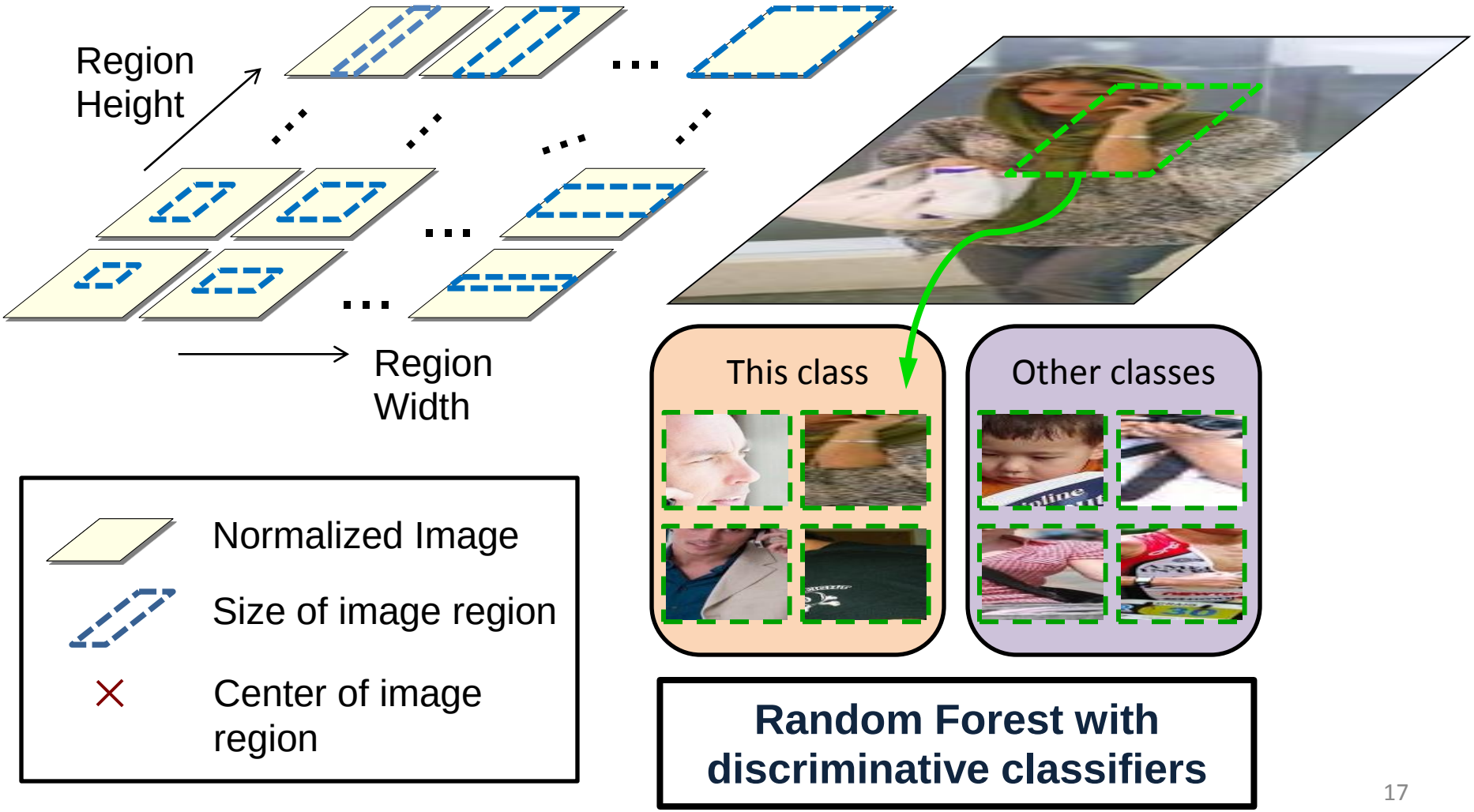
Apply **randomization** to sample a **subset of image patches**

Random Forest

Dense Feature Space



Dense Feature Space



Generalization of Random Forest

• Generalization error of a Random Forest (Breiman, 2001):

$$\rho \frac{(1 - s^2)}{s^2}$$

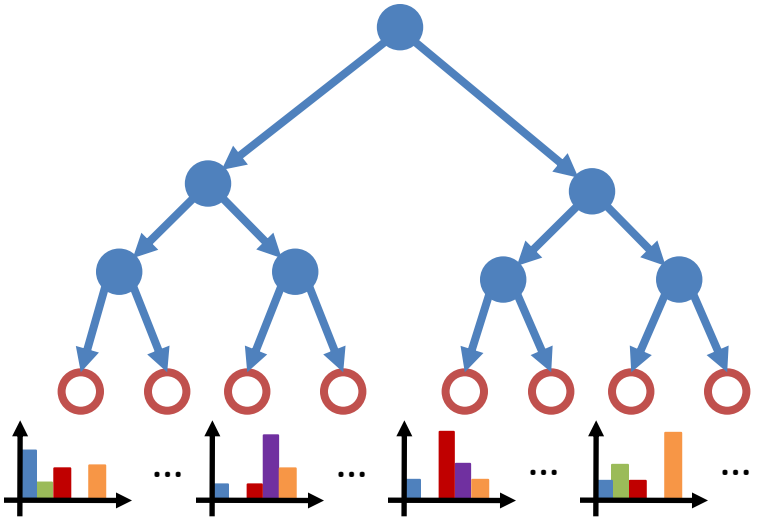
s : strength of the decision trees

ρ : correlation between decision trees

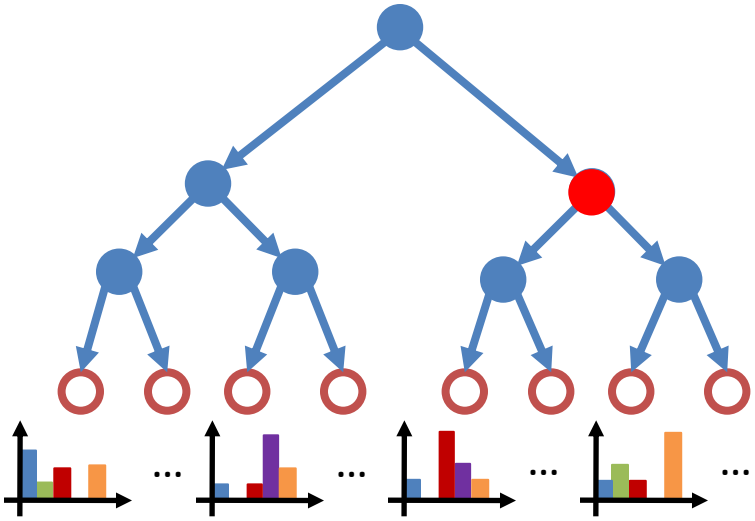
- Discriminative classifiers ➡ s increases
- Dense feature space ➡ ρ decreases

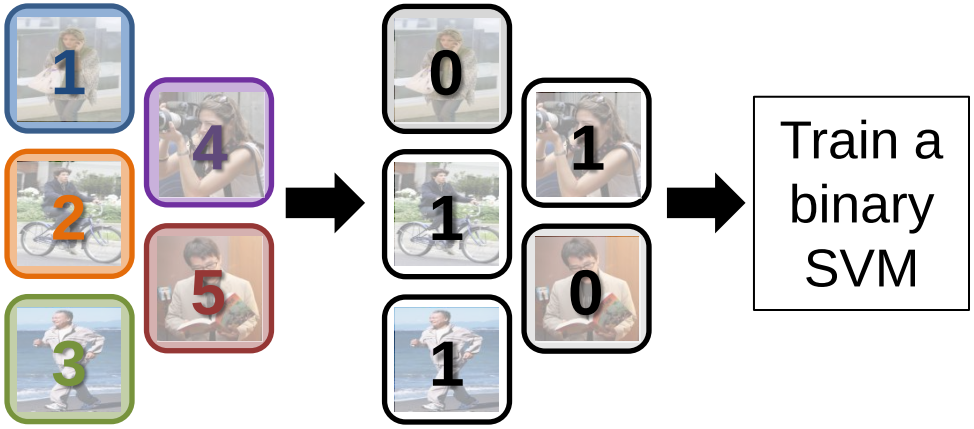
➡ **Better generalization**

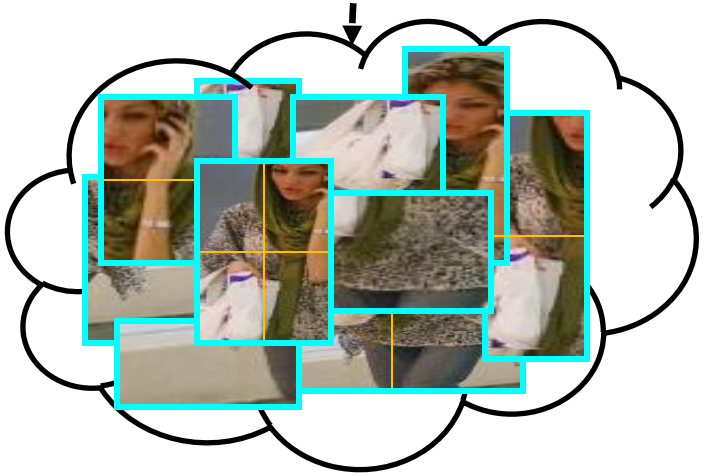
Random Forest with Discriminative Classifiers



Random Forest with Discriminative Classifiers

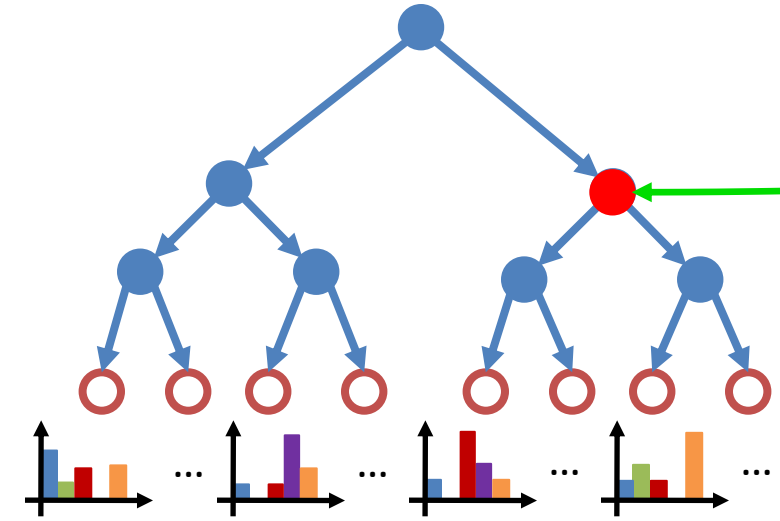




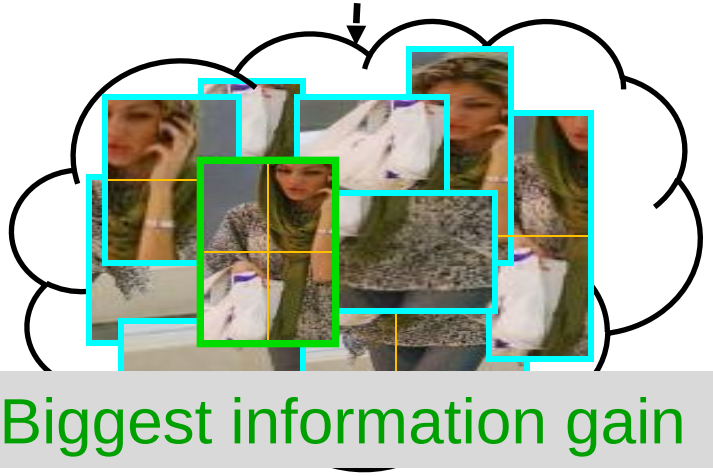


BoW or SPM of SIFT-LLC features

Random Forest with Discriminative Classifiers

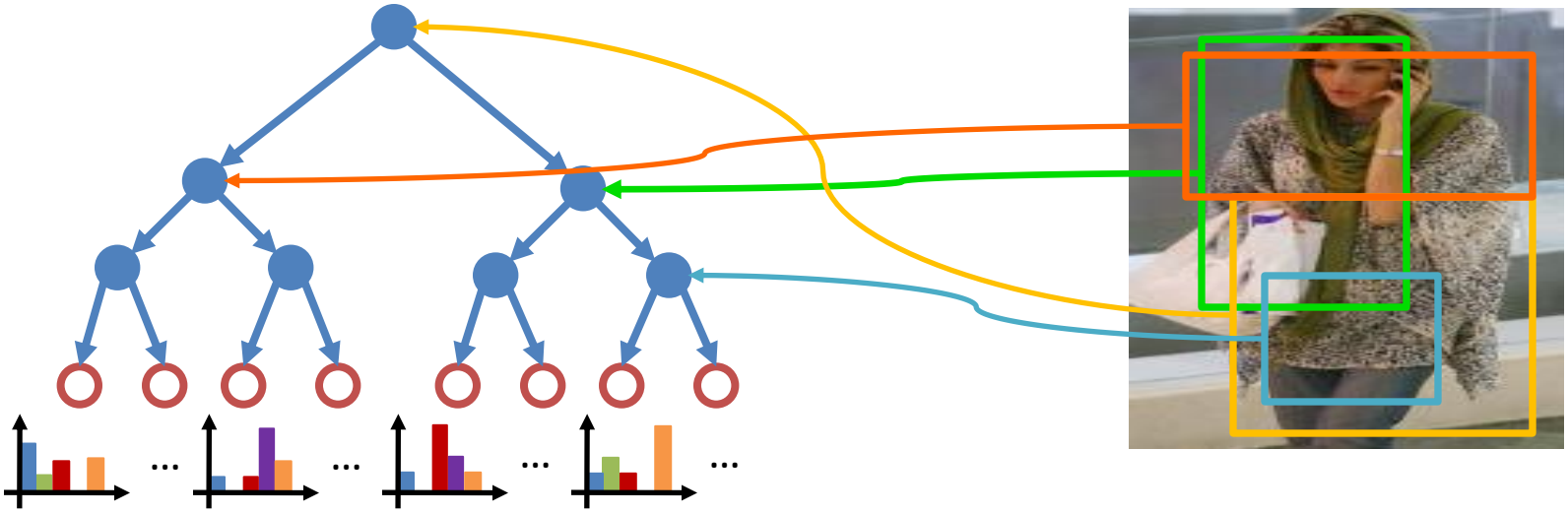


Train a binary SVM

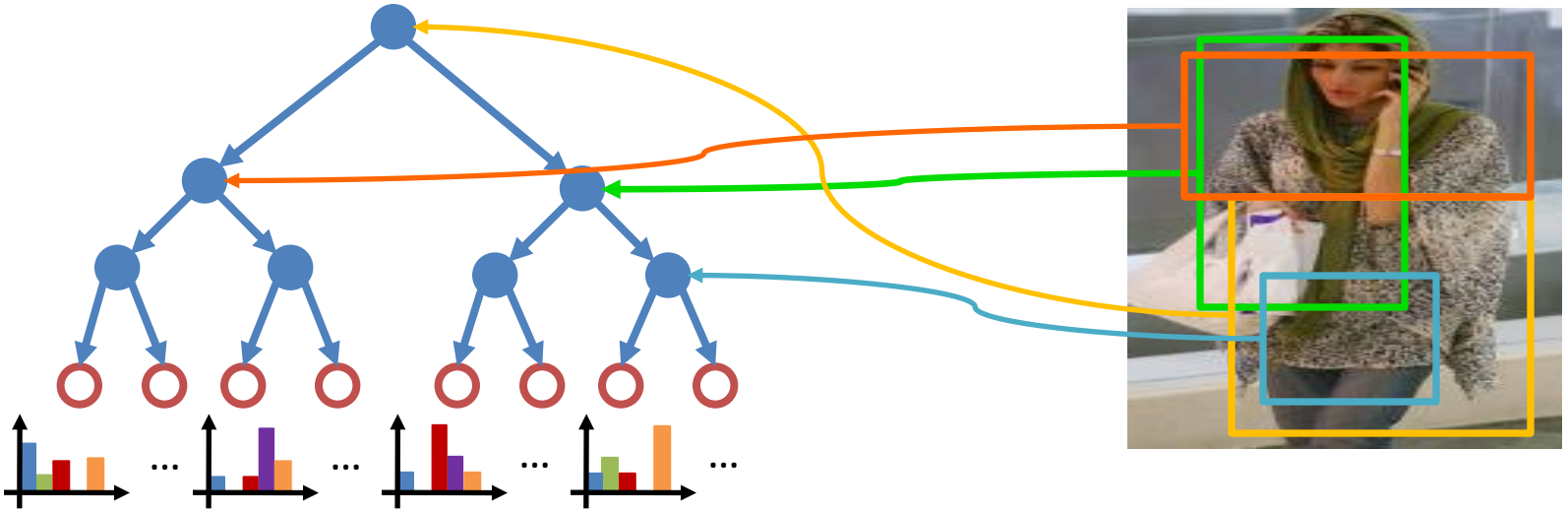


Biggest information gain

Random Forest with Discriminative Classifiers

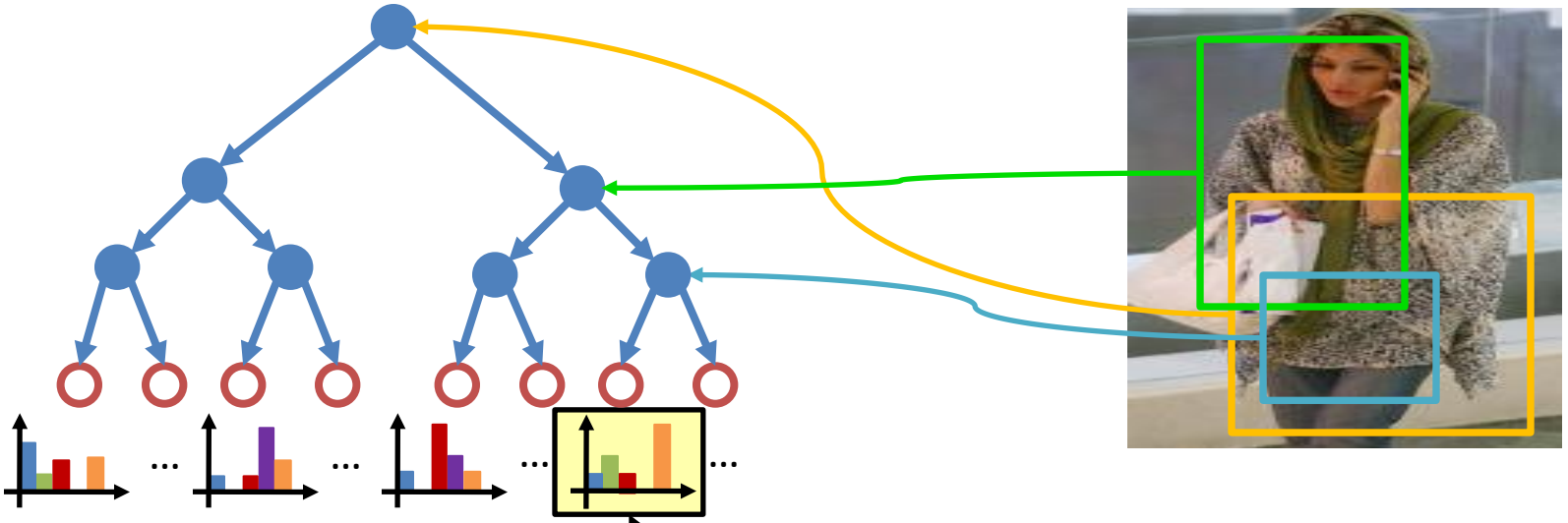


Random Forest with Discriminative Classifiers



- We stop growing the tree if:
  - The maximum depth is reached;
  - There is only one class at the node;

Classification With Random Forest



$$\underbrace{c^*}_{\text{Class Label}} = \arg\max_c \underbrace{\frac{1}{T}}_{\text{Number of trees}} \sum_{t=1}^{T} \overbrace{P_{t,l_t}(c)}$$

Outline

- Action Classification & Intuition
- Our Method
- **Our Results**
- Conclusion

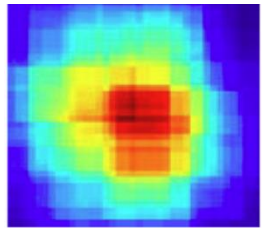
Results on VOC 2011 Action Comp9

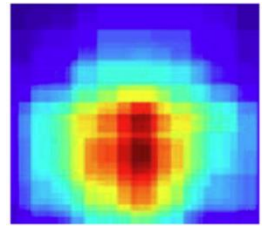
Action	Others' Best	Our Method
Jumping	71.6	66.0
Phoning	50.7	41.0
Playing instrument	77.5	60.0
Reading	37.8	41.5
Riding bike	88.8	90.0
Riding horse	90.2	92.1
Running	87.9	86.6
Taking photo	25.7	28.8
Using computer	58.9	62.0
Walking	59.5	65.9

Our method ranks the first in **six** out of ten classes.

Results on VOC 2011 Action Comp9

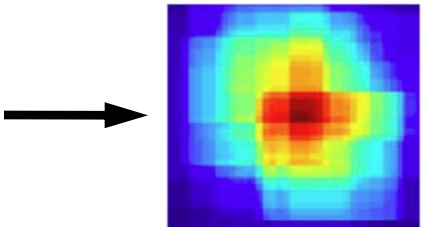
Action	Others' Best	Our Method
Jumping	71.6	66.0
Phoning	50.7	41.0
Playing instrument	77.5	60.0
Reading	37.8	41.5
Riding bike	88.8	90.0
Riding horse	90.2	92.1
Running	87.9	86.6
Taking photo	25.7	28.8
Using computer	58.9	62.0
Walking	59.5	65.9





Results on VOC 2011 Action Comp9

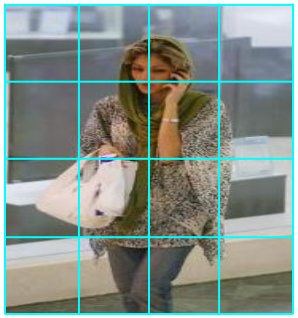
Action	Others' Best	Our Method
Jumping	71.6	66.0
Phoning	50.7	41.0
Playing instrument	77.5	60.0
Reading	37.8	41.5
Riding bike	88.8	90.0
Riding horse	90.2	92.1
Running	87.9	86.6
Taking photo	25.7	28.8
Using computer	58.9	62.0
Walking	59.5	65.9



Generalization Ability of RF

- Dense feature space → Tree correlation decreases
- Discriminative classifiers → Tree strength increases

→ **Better generalization**



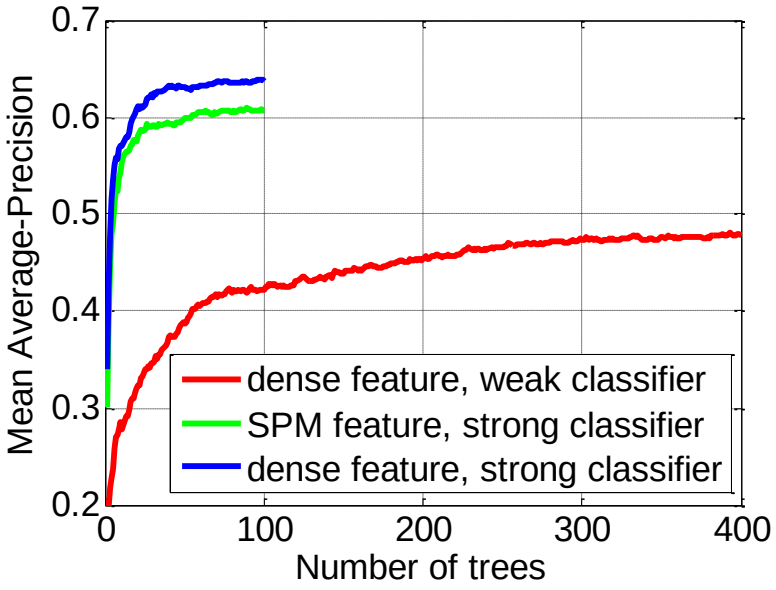
(spatial pyramid) SPM feature **Vs.**

dense feature

Generate feature weights randomly

Train discriminative SVM classifiers

weak classifier **Vs.** strong classifier



(Results on PASCAL VOC 2010)

Outline

- Action Classification & Intuition
- Our Method
- Our Results
- **Conclusion**

Conclusion

- Exploring dense image features can benefit action classification;
- Combining randomization and discrimination is an effective way to explore the dense image representation;
- Achieves very good performance based on only one type of image descriptor;
- Code will be available soon.

PASCAL VOC 2011 Result Comp10

	Others' best	Our method
Jumping	59.5	66.7
Phoning	31.3	41.1
Playing instrument	45.6	60.8
Reading	27.8	42.2
Riding bike	84.4	90.5
Riding horse	88.3	92.2
Running	77.6	86.2
Taking photo	31.0	28.8
Using computer	47.4	63.5
Walking	57.6	64.2

Wednesday 9th November, 12:00-12:30

Acknowledgement

Thanks to Su Hao, Olga Russakovsky, and Carsten Rother.

Reference:

Bangpeng Yao, Aditya Khosla, and Li Fei-Fei. “Combining Randomization and Discrimination for Fine-Grained Image Categorization.” CVPR 2011.