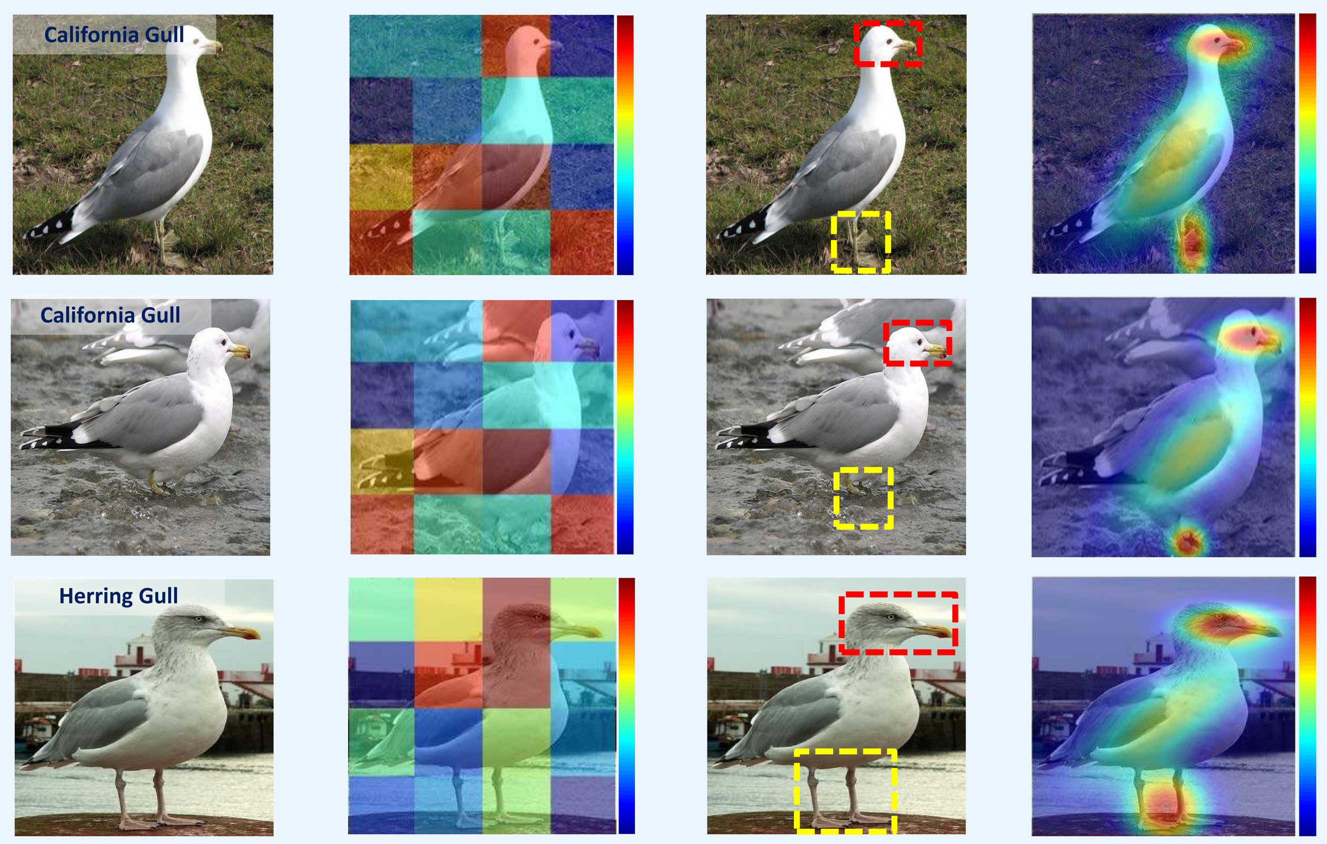
Combining Randomization and Discrimination for Fine-Grained Image Categorization

The Vision Lab
Computer Science Dept.

Bangpeng Yao*, Aditya Khosla* and Li Fei-Fei
{bangpeng,aditya86,feifeili}@cs.stanford.edu
(* - indicates equal contribution)

Stanford University

Motivation



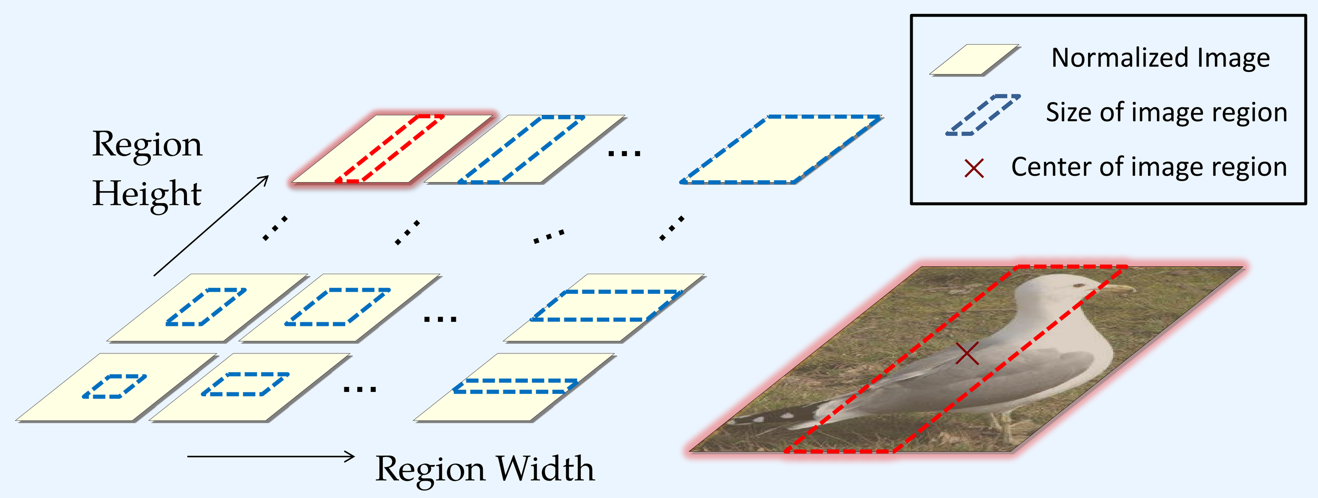
original image | traditional method (SPM) | our intuition: what humans do | **our goal**

Our Work

- **Objective: Finding image regions that contain discriminative information** for fine-grained image categorization.
- **Approach:** A model combining **randomization and discrimination**
  - Dense feature representation;
  - Random forest with discriminative decision trees classifier

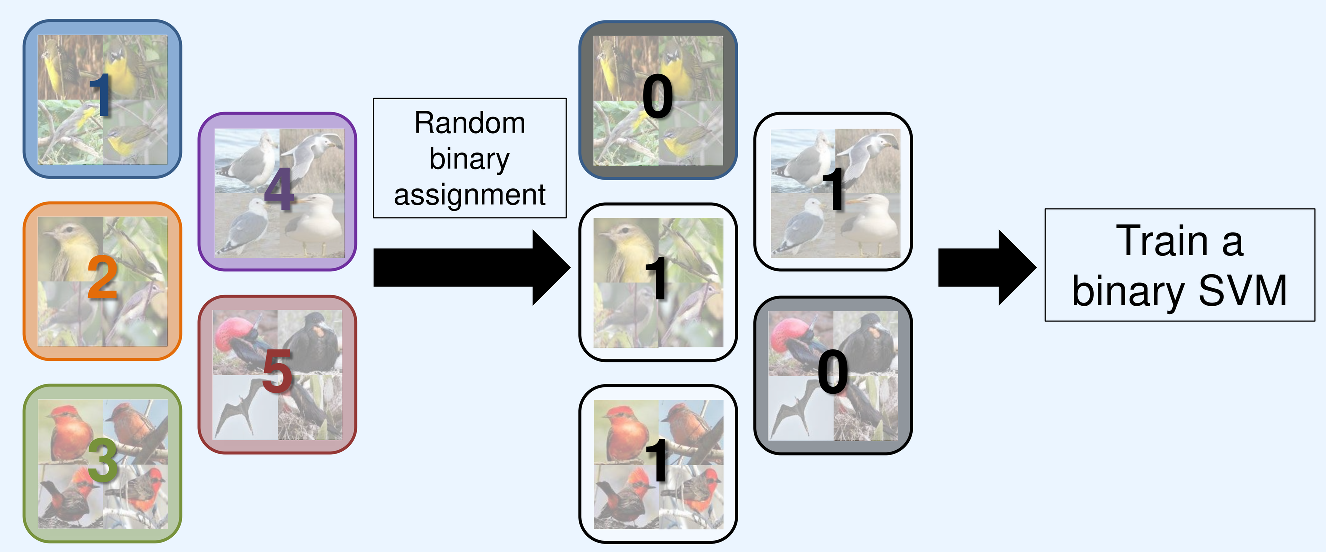
Dense Feature Representation

- Our representation consists of (pairs of) image regions of arbitrary sizes and at arbitrary locations:

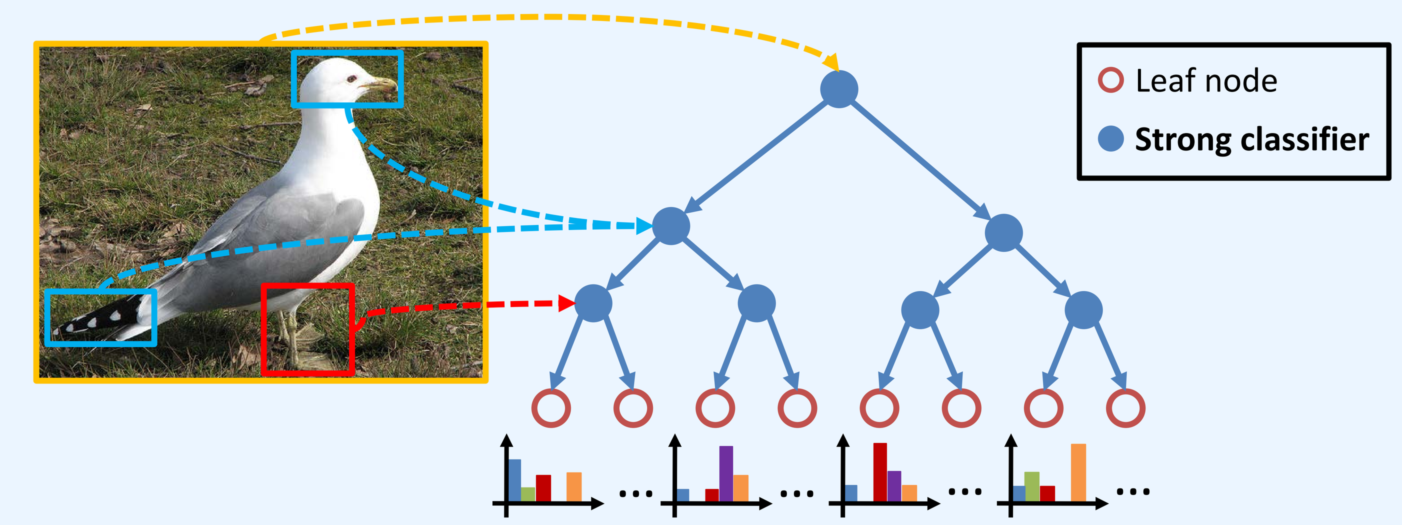


Training a Strong Classifier

- {1, ..., 5} represent original class labels



Random forest with Discriminative Decision Trees



- Learning our random forest classifier:

foreach *tree t* **do**
- Obtain a random set of training examples $\mathcal{D}$;
- `SplitNode` ($\mathcal{D}$);
- **if** *needs to split* **then**
  - i. Randomly sample the candidate (pairs of) image regions;
  - ii. Select the best region to split $\mathcal{D}$ into two sets $\mathcal{D}_1$ and $\mathcal{D}_2$;
  - iii. `SplitNode` ($\mathcal{D}_1$) and `SplitNode` ($\mathcal{D}_2$).
- **else**
  - Return $P_t(c)$ for the current leaf node.
- **end**

end

Features for image regions:
- Grayscale SIFT descriptors for PPMI and PASCAL Action
- ColorSIFT descriptors for Caltech-UCSD Birds-200
- Dense SIFT sampling at multiple scales (8, 12, 16, 24, 30)
- Locality-constrained Linear Coding (LLC) Features

- Select best sample using information gain criterion:

$$\Delta E = -\sum_i \frac{|\mathcal{D}_i|}{|\mathcal{D}|} E(\mathcal{D}_i)$$

$\mathcal{D} = \{\cup_i \mathcal{D}_i\}$: set of all training examples

$E(\mathcal{D}_i)$: entropy of training examples $\mathcal{D}_i$

- Classification of test example:

$$c^* = \arg\max_c \frac{1}{T} \sum_{t=1}^{T} P_{t,l_t}(c)$$

T : number of trees

c^* : class label of test example

$P_{t,l_t}(c)$: probability of test example belonging to class c for tree t

Generalization Error of RF

- Generalization error of a random forest:

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

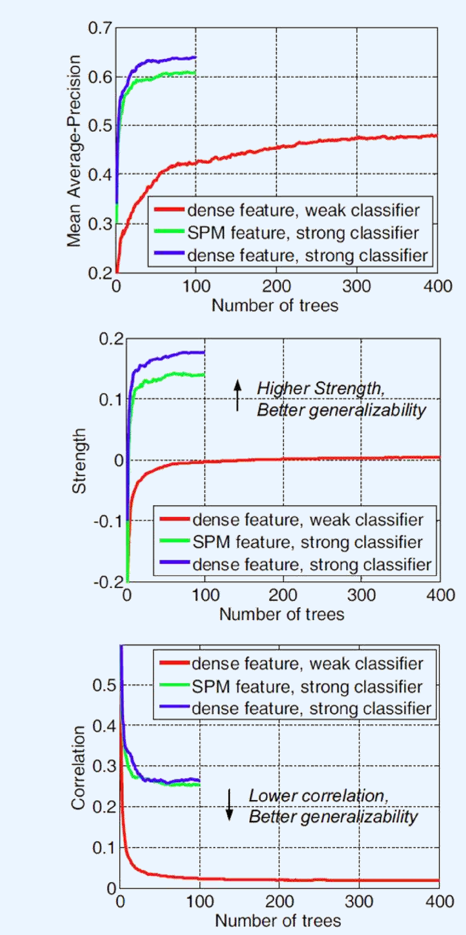
ρ : correlation between decision trees

s : strength of the decision trees

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

➡ Better generalization

Control Experiments

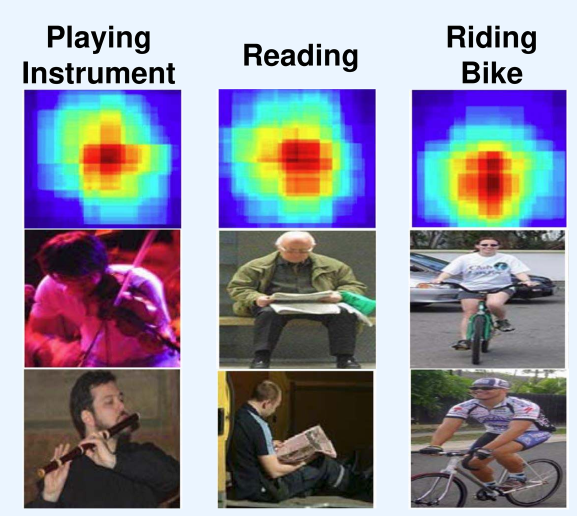


Experiment

PASCAL Action Dataset

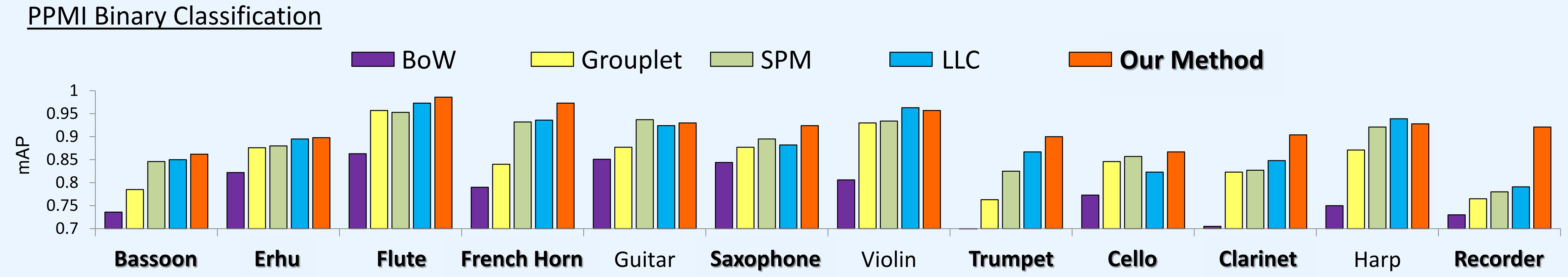
- 9-class classification of human actions (%mAP)

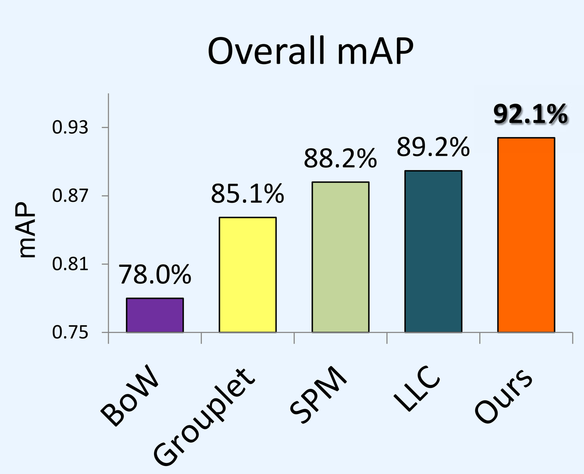
Method	Phoning	Playing instrument	Reading	Riding bike	Riding horse	Running	Taking photo	Using computer	Walking	Overall
CVC-BASE	**56.2**	56.5	34.7	75.1	83.6	86.5	25.4	60.0	69.2	60.8
CVC-SEL	49.8	52.8	34.3	74.2	85.5	85.1	24.9	64.1	72.5	60.4
SURREY-KDA	52.6	53.5	35.9	81.0	89.3	86.5	32.8	59.2	68.6	62.2
UCLEAR-DOSP	47.0	**57.8**	26.9	78.8	89.7	87.3	32.5	60.0	70.1	61.1
UMCO-KSVM	53.5	43.0	32.0	67.9	68.8	83.0	34.1	45.9	60.4	54.3
Our Method	45.0	57.4	**41.5**	**81.8**	**90.5**	**89.5**	**37.9**	**65.0**	**72.7**	**64.6**

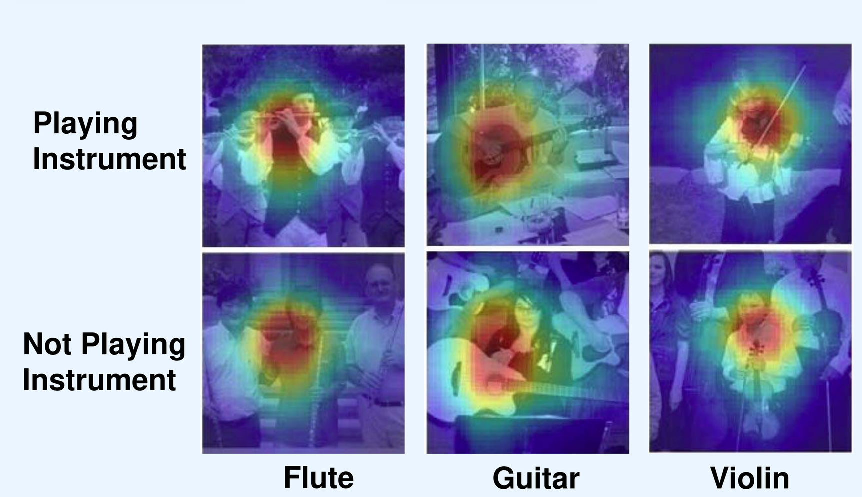


People-Playing-Musical-Instruments (PPMI)

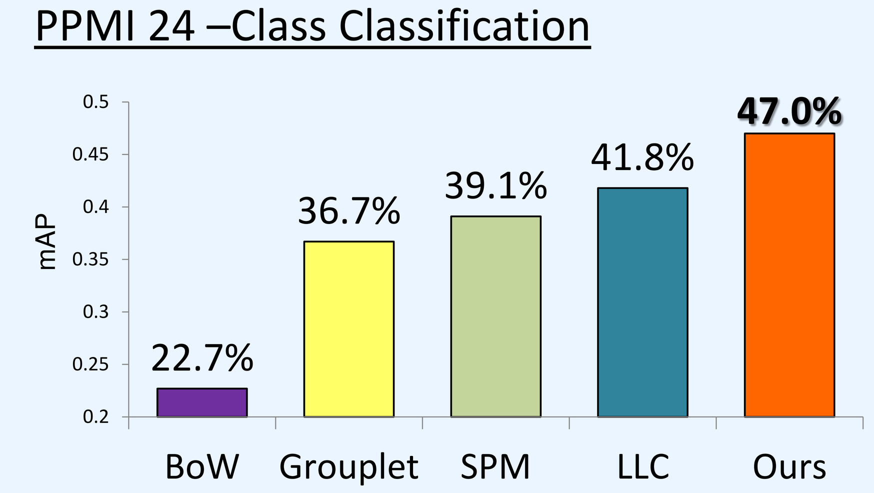
PPMI Binary Classification





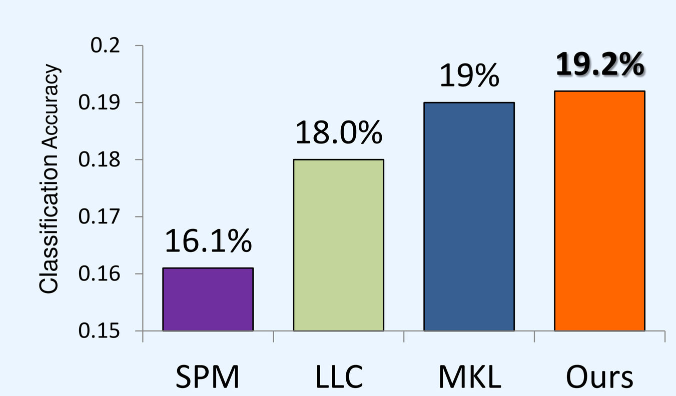


PPMI 24 –Class Classification

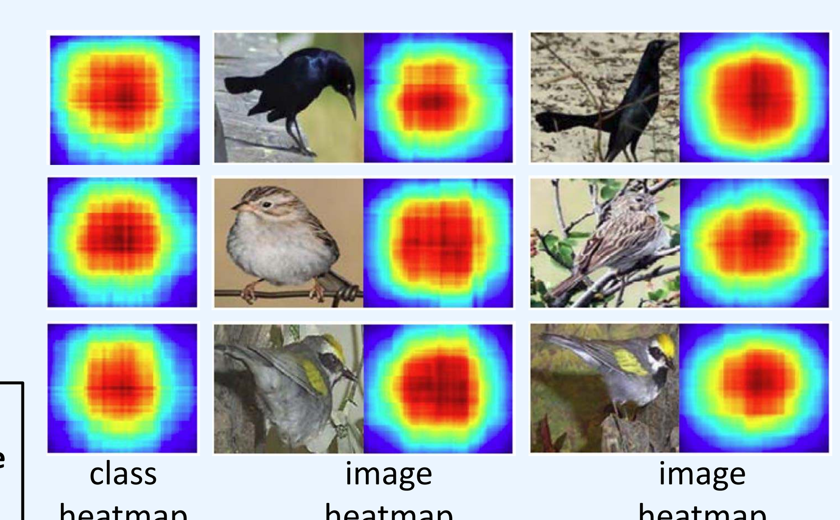


Caltech-UCSD Birds 200

- 200-class classification of 200 bird species from North America

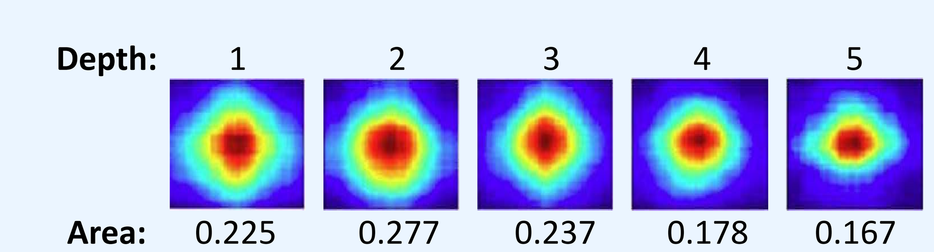


- MKL: Uses **many features** including Gray/ColorSIFT, geometric blur, color histograms, etc.
- Ours: Uses a **single feature** (ColorSIFT)



Coarse-to-fine Learning

- Our method automatically learns a coarse-to-fine region of interest (e.g. shown below for 'playing trumpet' class)
- This is similar to the **human visual system** which is believed to analyze raw input from low to high spatial frequencies or **from large global shapes to smaller local ones**



Future Work:

- Improve speed by exploiting the inherent parallel nature of random forests using GPUs
- Strong classifiers with analytical solution (e.g. LDA)
- Incorporate multiple features

Reference

B. Yao*, A. Khosla* and L. Fei-Fei. "**Combining Randomization and Discrimination for Fine-Grained Image Categorization**." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.