Ubiquitous Fine-Grained Computer Vision

Shu Kong

Department of Computer Science, UC Irvine

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

Outline

1. Problem definition
2. Instantiation
3. Challenge
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Problem Definition

1. Problem definition
2. Instantiation
3. Challenge and philosophy
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Problem Definition

Fine-grained

- marginally different or **subtle**

Problem Definition

Fine-grained

- marginally different or **subtle**
- involving great attention to **detail** (Oxford dictionary)

Problem Definition

Fine-grained

- marginally different or **subtle**
- involving great attention to **detail** (Oxford dictionary)

- The devil is in the details!
- ...and **everywhere!**

Problem Definition

Fine-grained

- marginally different or **subtle**
- involving great attention to **detail** (Oxford dictionary)

- The devil is in the details!
- ...and **everywhere! -- ubiquitous**

Problem Definition

Fine-grained computer vision

Problem Definition

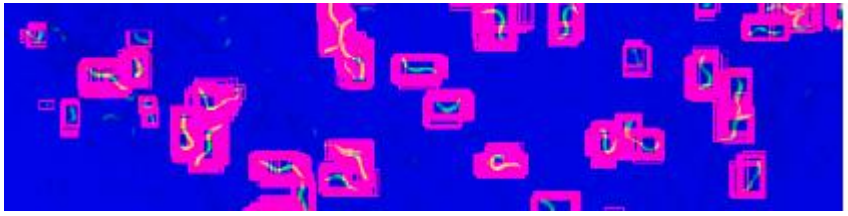
Fine-grained computer vision

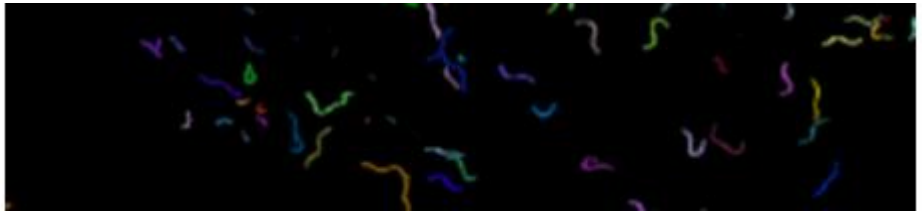
- distinguish subordinate categories within an entry-level category

Problem Definition

Fine-grained computer vision

- distinguish subordinate categories within an entry-level category
- detection -> instance segmentation





Outline

1. Problem definition
2. Instantiation
3. Challenge
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Instantiation -- classification

previously, generic classification -- car vs. bird

S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Instantiation -- classification

now, fine-grained car model classification

S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Instantiation -- classification

now, fine-grained bird species classification

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

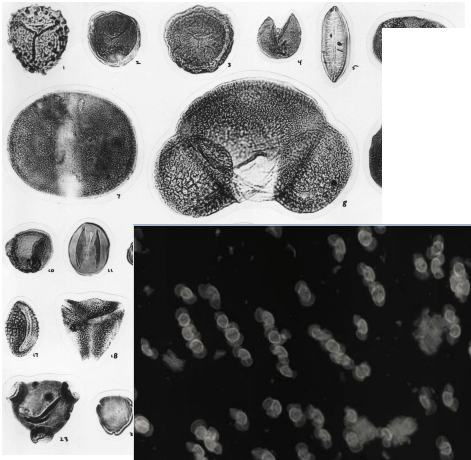
Instantiation -- identification

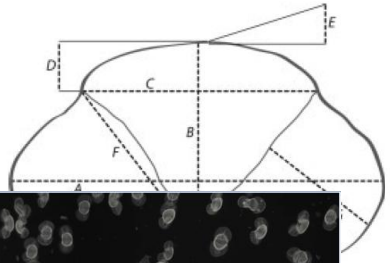
in phytology

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

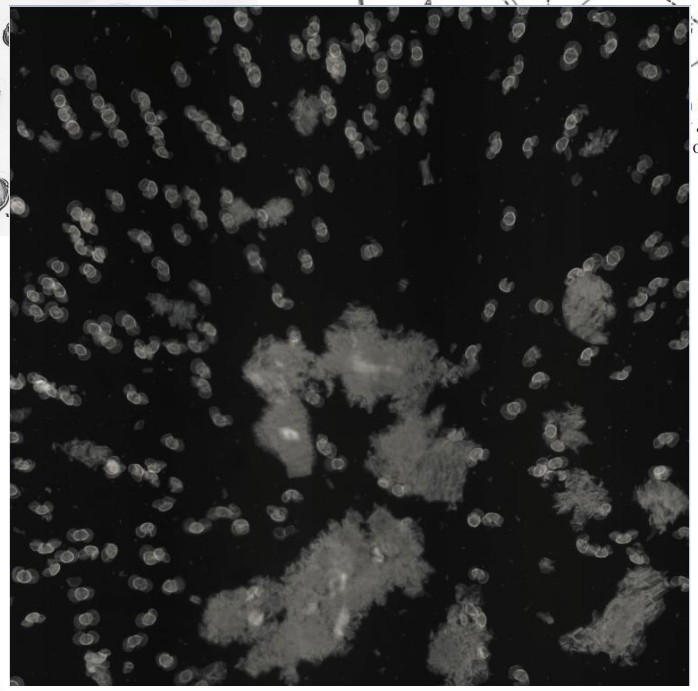
Instantiation -- identification

previously, in phytology, identifying by eye





from specimens
al., 2002). Grain
depth of saccus
(*F*) and saccus

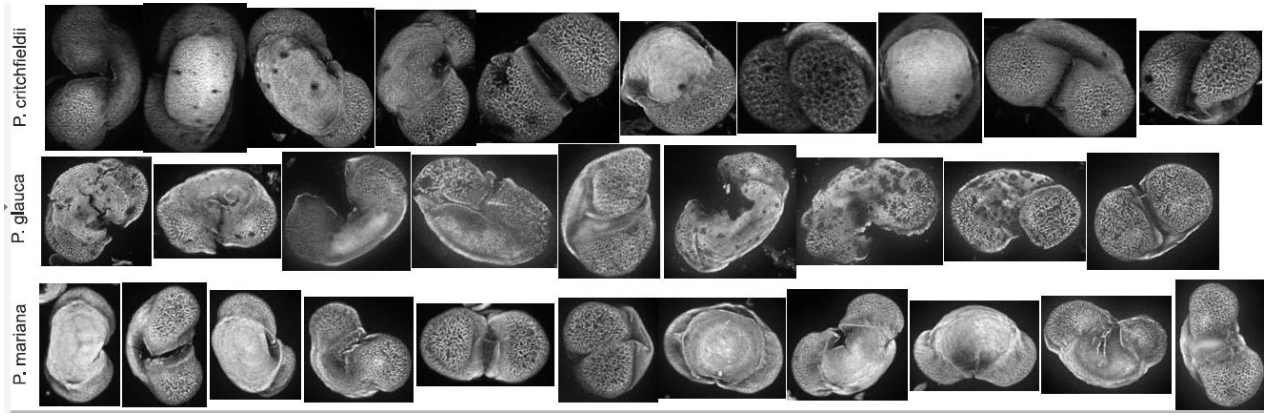


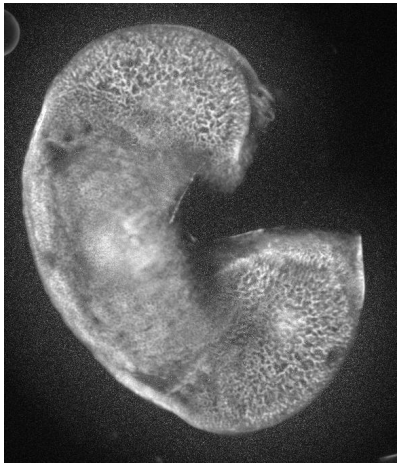
S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

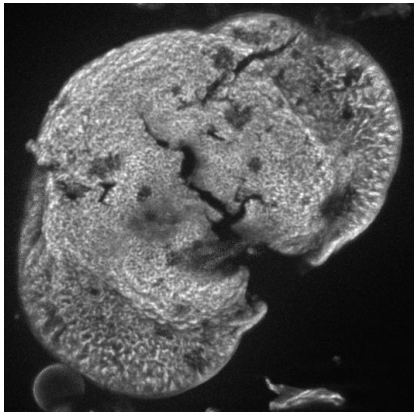
Instantiation -- identification

now, automatically, accurately identifying species-level pollen and matching fossilized pollen grains with modern reference

modern pollen grain from glauca







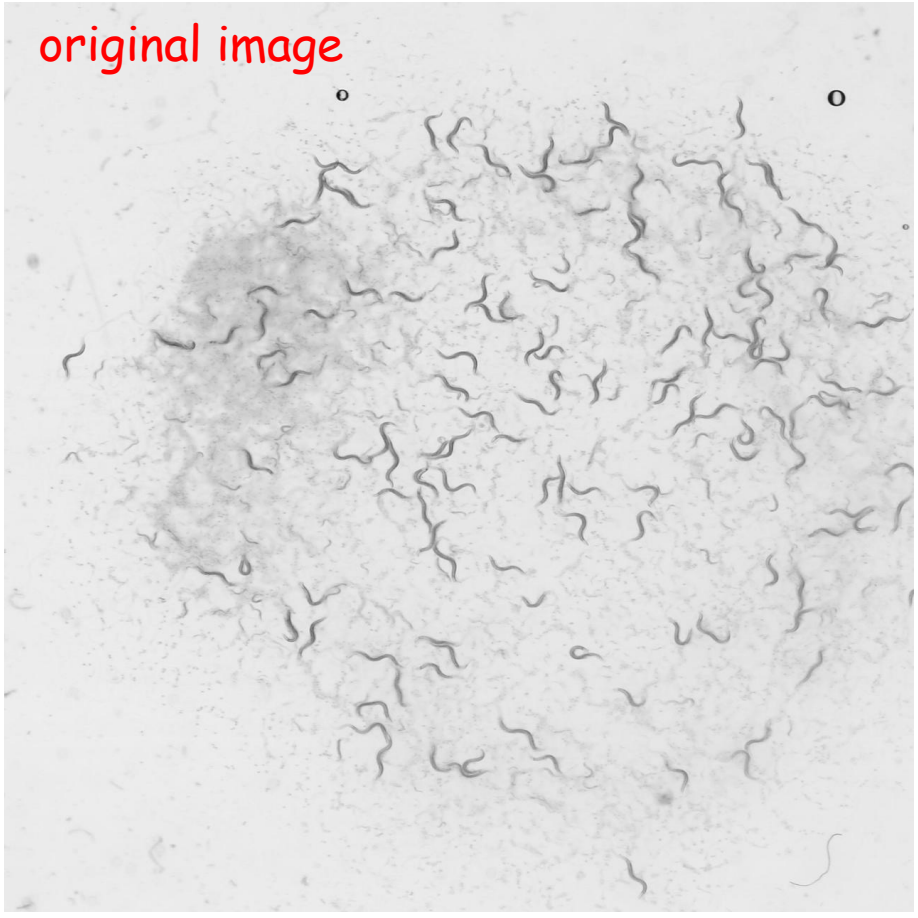
fossil pollen pollen grain from glauca

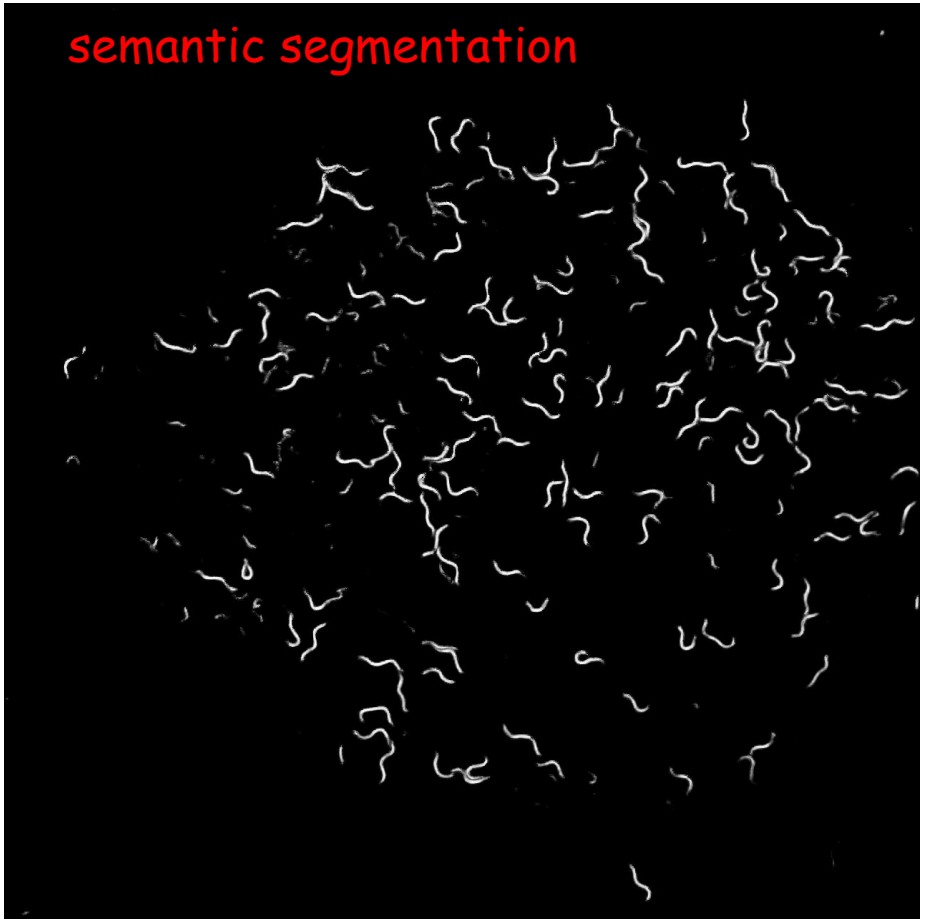
S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Instantiation -- segmentation

previously, in biology, semantic segmentation

e.g. binary label for biological data of *C. elegans*



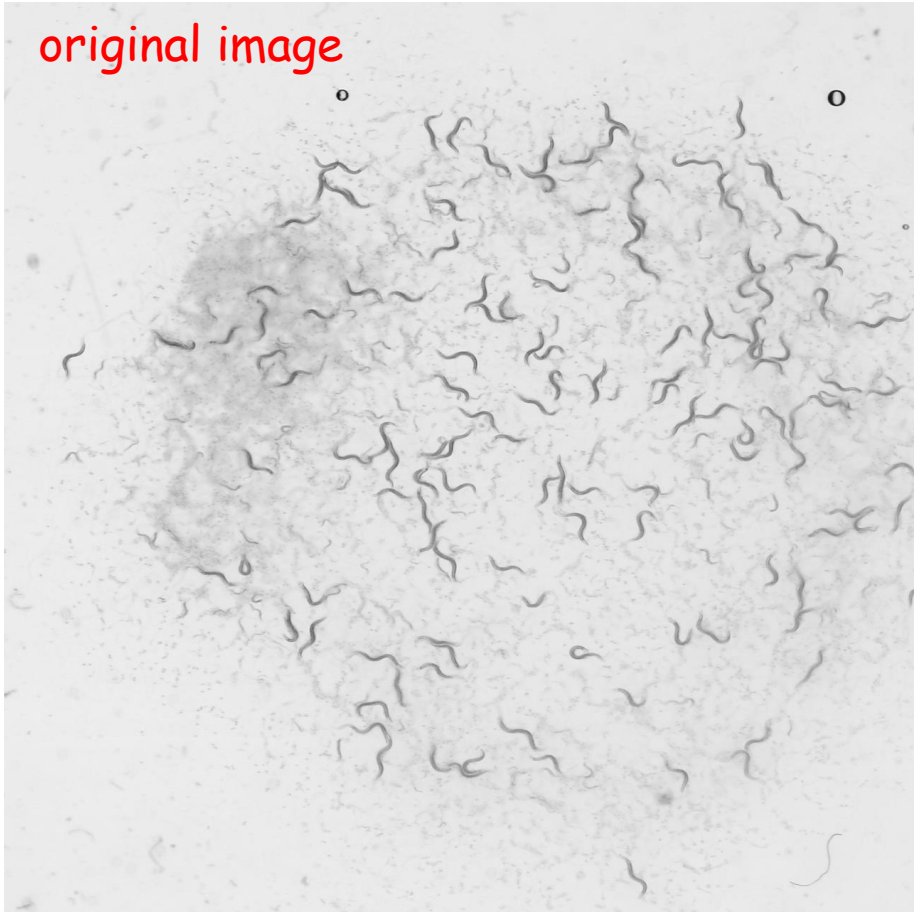


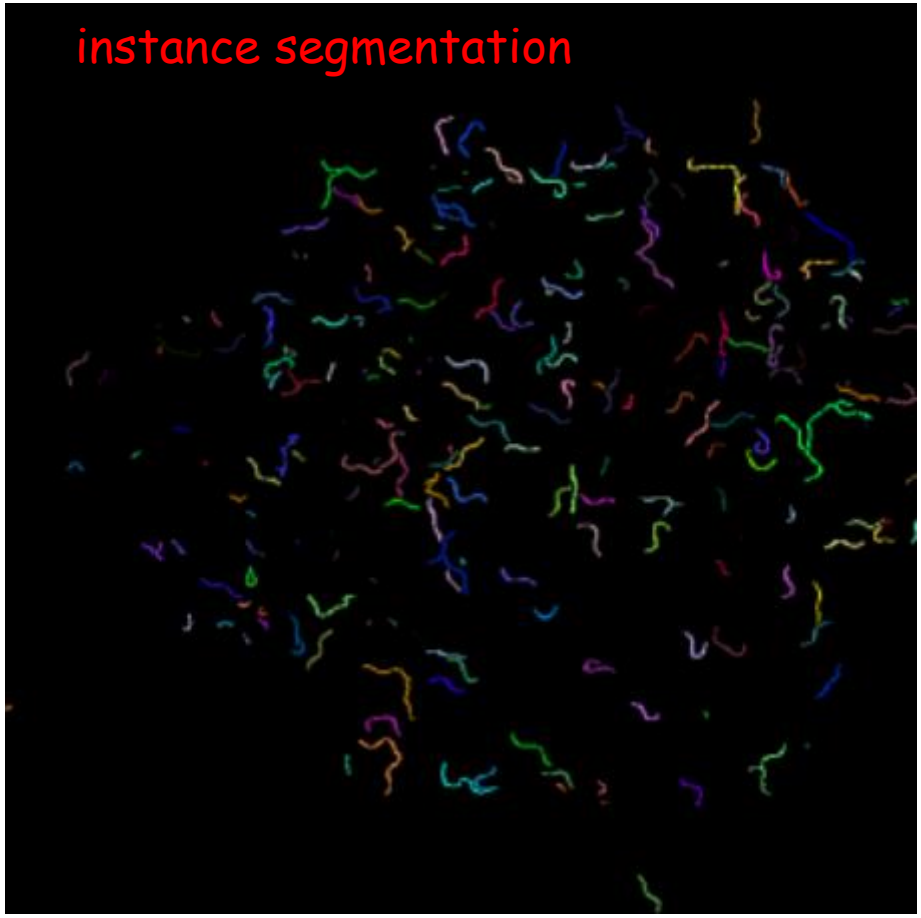
S. Kong, "Automated Biological Image Analysis using Computer Vision and Machine Learning", Janelia workshop, 2016

Instantiation -- segmentation

now, instance segmentation

enabling study of worm population



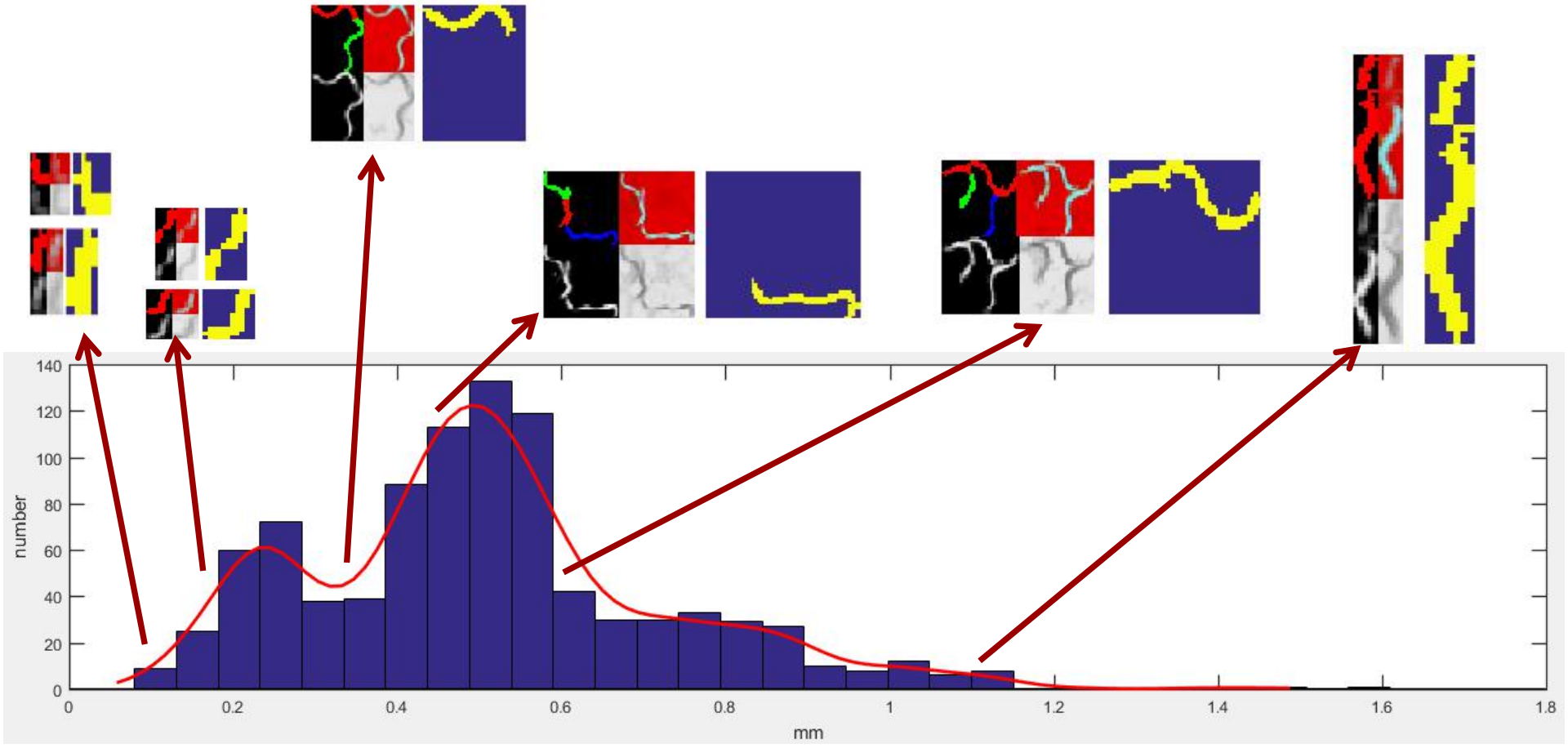


S. Kong, "Automated Biological Image Analysis using Computer Vision and Machine Learning", Janelia workshop, 2016

Instantiation -- segmentation

now, instance segmentation

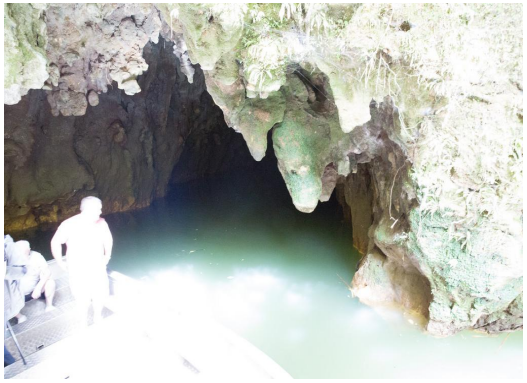
enabling study of worm population

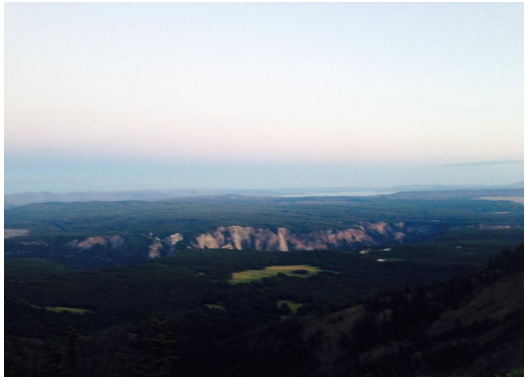


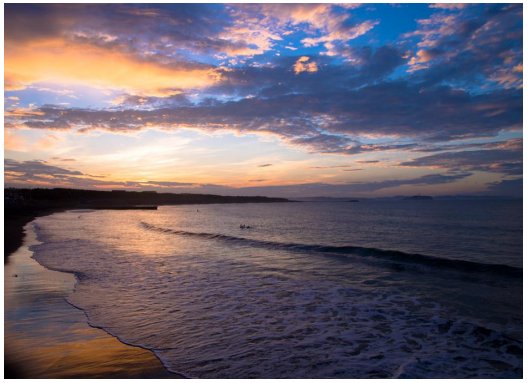
S. Kong, "Automated Biological Image Analysis using Computer Vision and Machine Learning", Janelia workshop, 2016

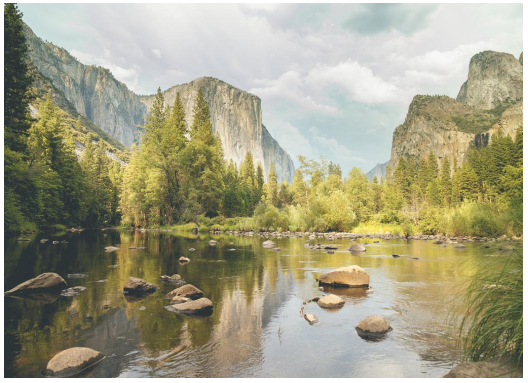
Instantiation -- photo aesthetic ranking

previously, modeling image aesthetics study as binary classification, low- vs. high- aesthetic







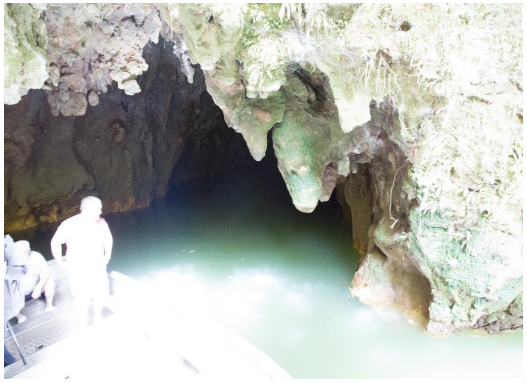
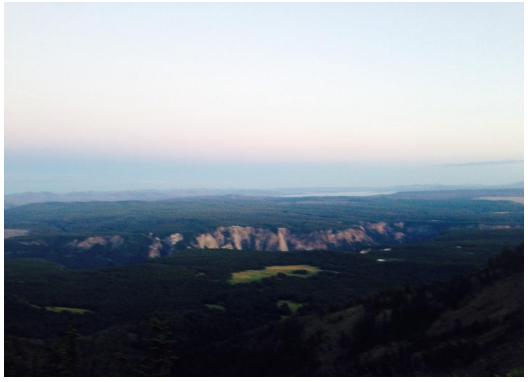


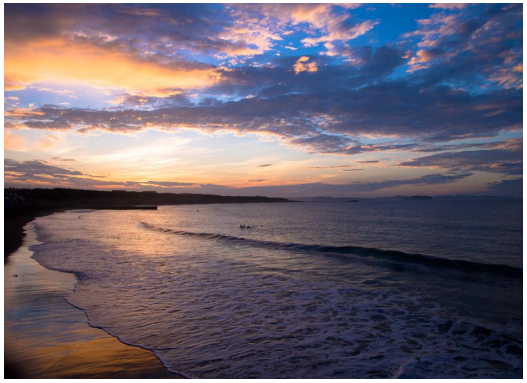
UCIrvine
University of California, Irvine

S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", ECCV, 2016

Instantiation -- photo aesthetic ranking

previously, modeling image aesthetics study as binary classification, low- vs. high- aesthetic

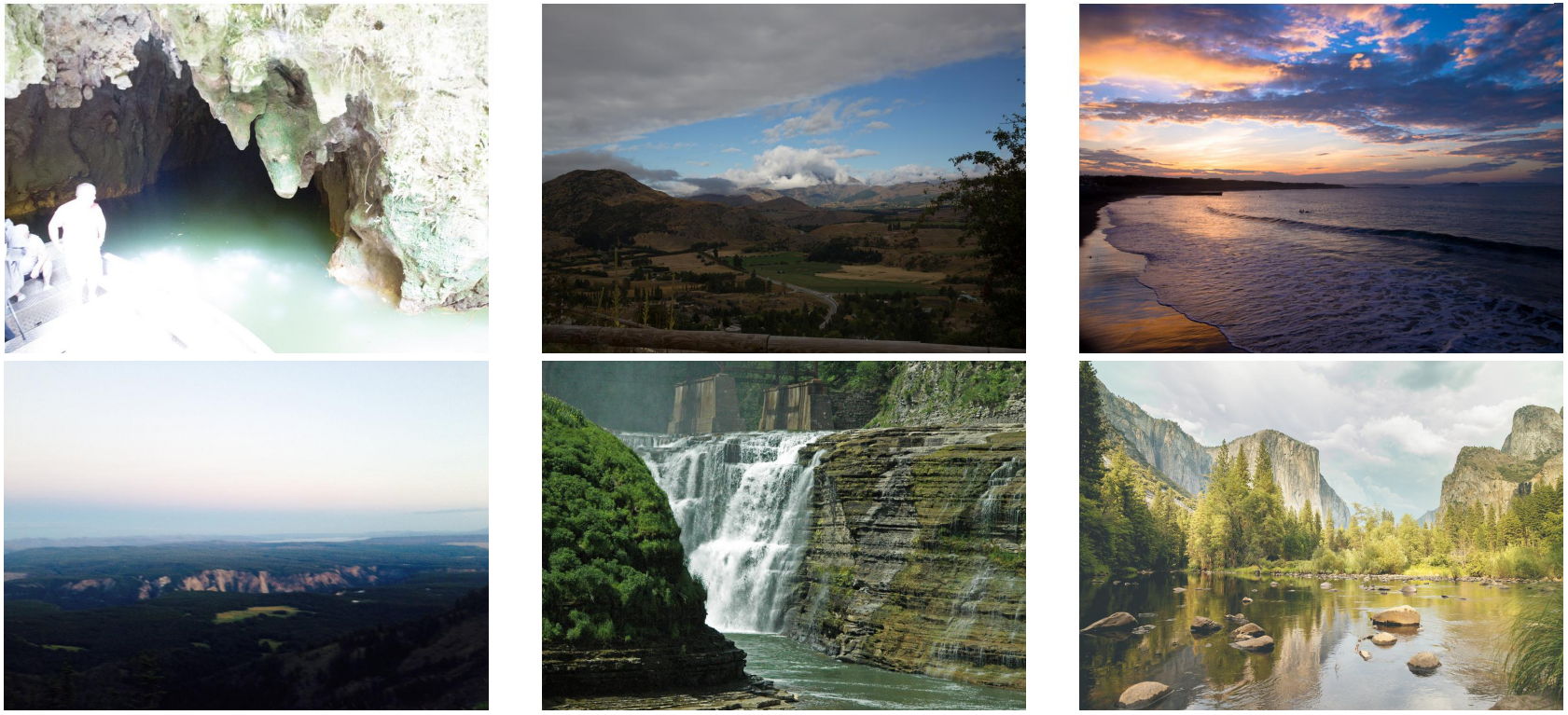




UCIrvine
University of California, Irvine

S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", ECCV, 2016

Instantiation -- photo aesthetic ranking

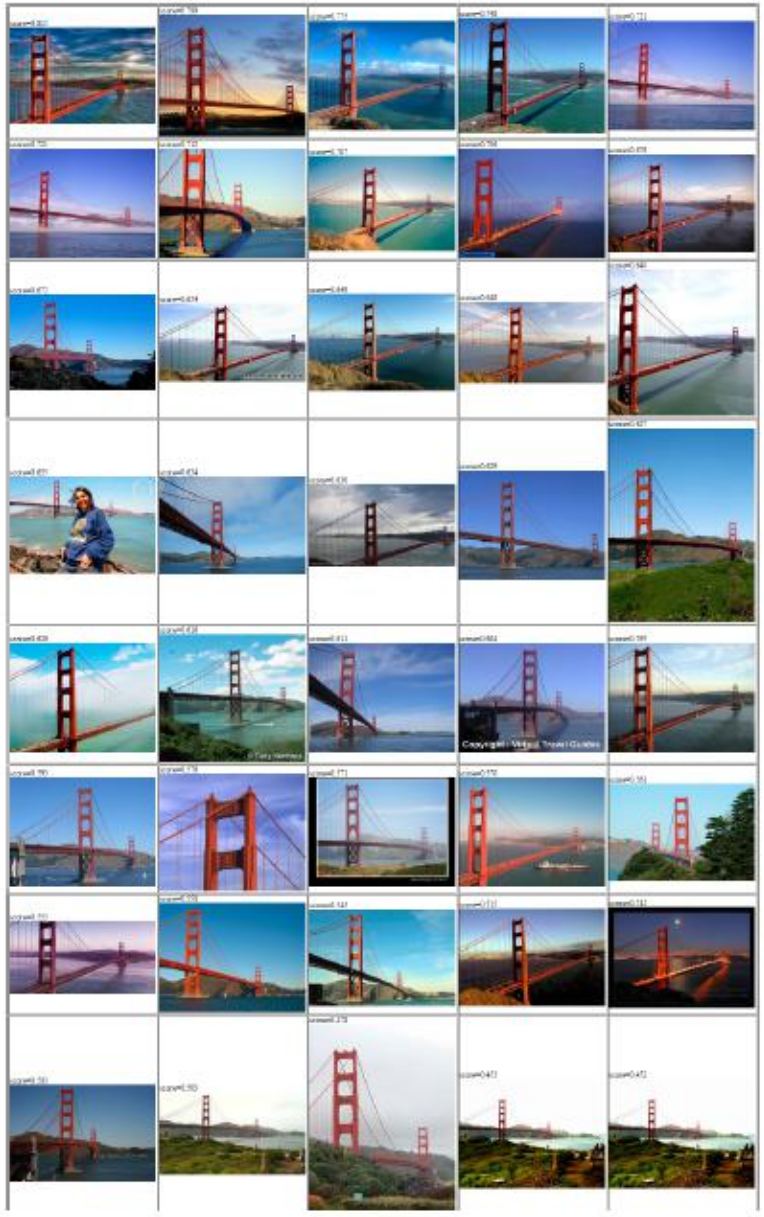
now, fine-grained ranking for personal photo album management



S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", ECCV, 2016

Instantiation -- photo aesthetic ranking

now, fine-grained ranking for personal photo album management



UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", ECCV, 2016

Challenge and philosophy

1. Problem definition
2. Instantiation
3. Challenge
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Challenge and philosophy

- large numbers of categories

Challenge and philosophy

- **large numbers of categories**
  - >14,000 birds

Challenge and philosophy

- **large numbers of categories**
  - >14,000 birds
  - >278,000 butterfly&moth

Challenge and philosophy

- **large numbers of categories**
  - >14,000 birds
  - >278,000 butterfly&moth
  - >941,000 insects

Challenge and philosophy

- large numbers of categories
- **high intra-class vs. low inter-class variance**

Challenge and philosophy

- large numbers of categories
- **high intra-class vs. low inter-class variance**



Challenge and philosophy

- large numbers of categories
- high intra-class vs. low inter-class variance

Caspian Tern

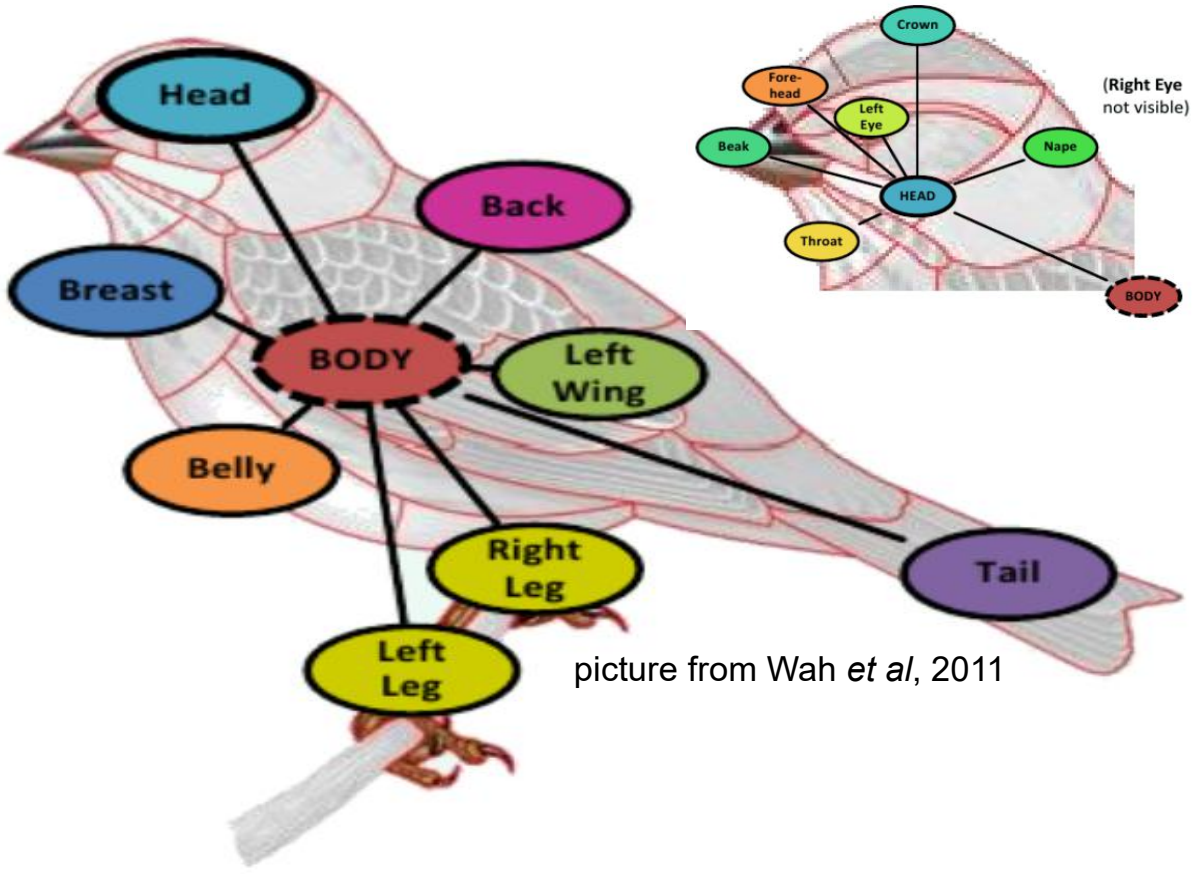
Caspian Tern

Elegant Tern

Challenge and philosophy

- large numbers of
- high intra-class v

Caspian Tern



picture from Wah *et al*, 2011

- **philosophy**
  - finding discriminative parts/keypoints,
  - stacking them and matching for classification

Challenge and philosophy

- large numbers of categories
- high intra-class vs. low inter-class variance
- **expensive to collect and annotate data**
  - lack of training data

Holistic representation based method

1. Problem definition
2. Instantiation
3. Challenge and philosophy
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work
7. Conclusion

Holistic representation based method

recognizing bird species by seeing the photo

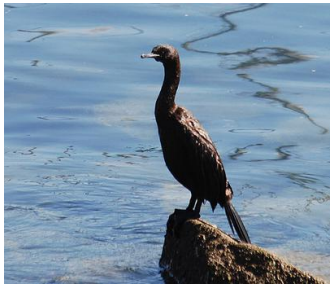
Red_Winged_Blackbird

Brandt_Cormorant

Acadian_Flycatcher

Yellow_Headed_Blackbird

Pelagic_Cormorant



Yellow_Billed_Cuckoo

Holistic representation based method

recognizing bird species by seeing the photo

In literature, detecting keypoint/parts and stacking them as holistic representation

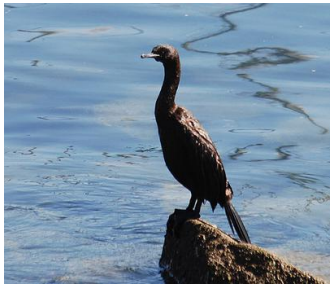
Red_Winged_Blackbird

Brandt_Cormorant

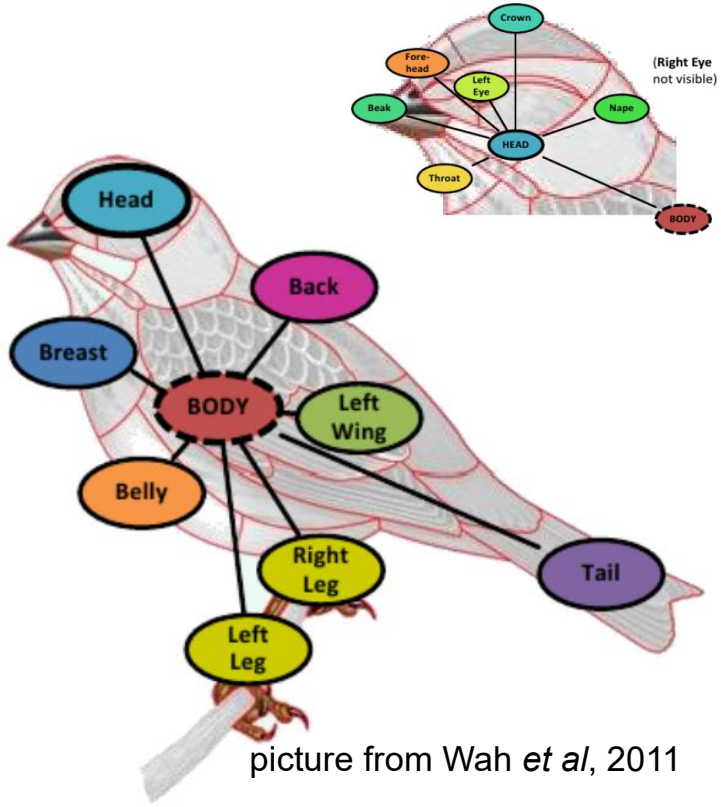
Acadian_Flycatcher

Yellow_Headed_Blackbird

Pelagic_Cormorant



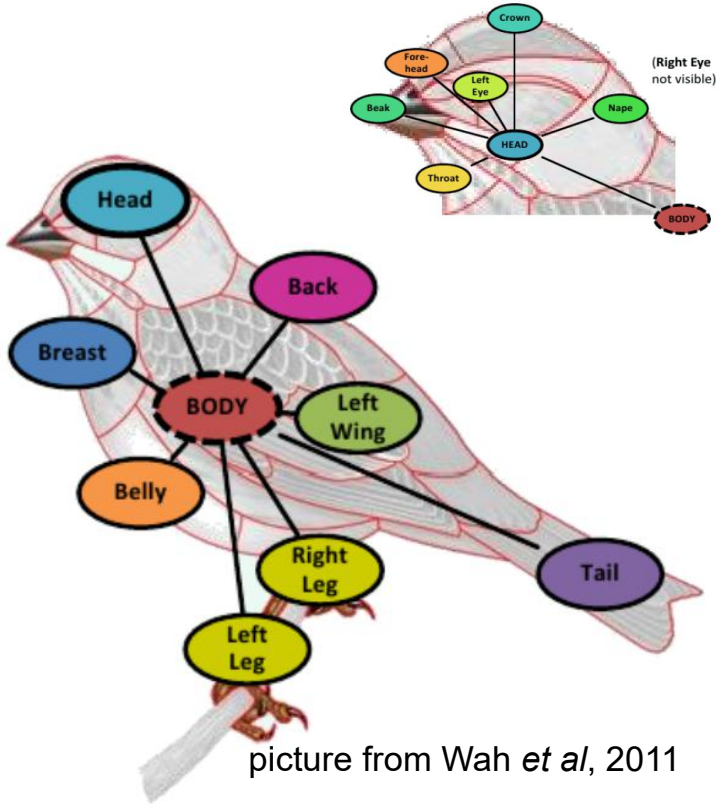
Yellow_Billed_Cuckoo



picture from Wah *et al*, 2011

Holistic representation based method

But, this requires strong-supervised annotation, which is expensive to obtain.



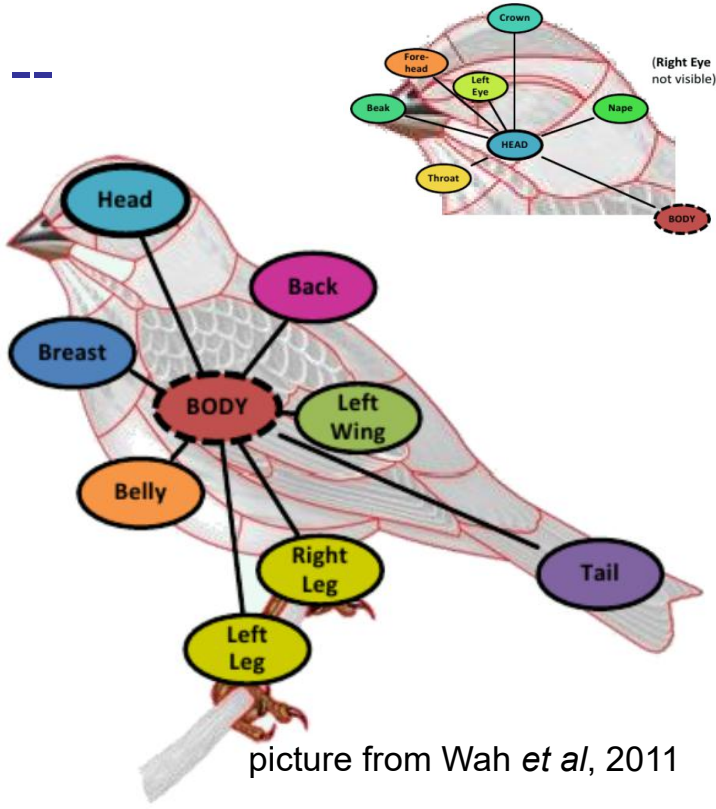
picture from Wah *et al*, 2011

Holistic representation based method

But, this requires strong-supervised annotation, which is expensive to obtain.

Preferably in weakly supervised manner --

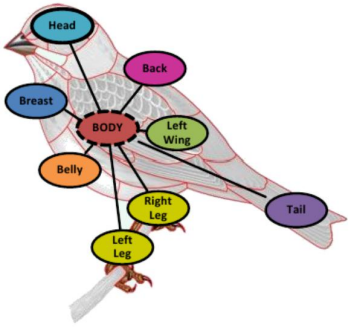
- solely based on category labels
- without any part annotation.



picture from Wah *et al*, 2011

Holistic representation based method

One method for this is called **bilinear pooling**

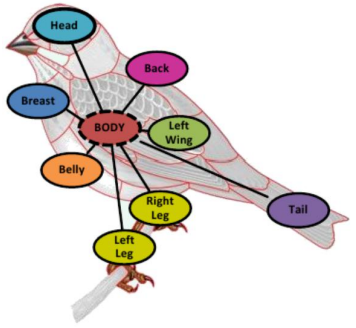


Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

One method for this is called bilinear pooling

compute second-order statistics of local features, and average them as a single holistic representation



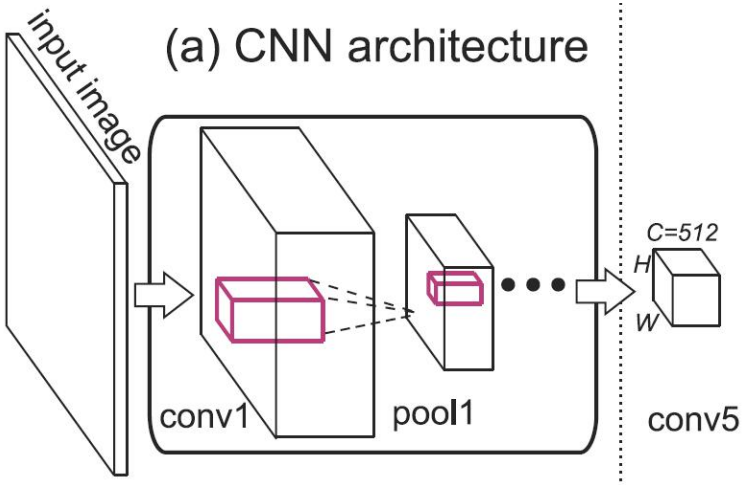
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

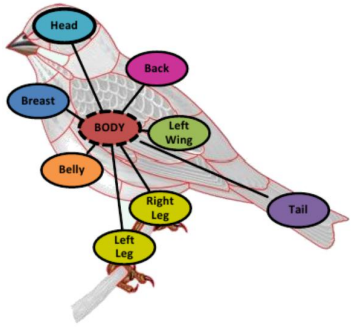
Holistic representation based method

One method for this is called bilinear pooling

compute second-order statistics of local features, and average them as a single holistic representation

The local features can be activations at a hidden layer of a convolutional neural network (CNN)



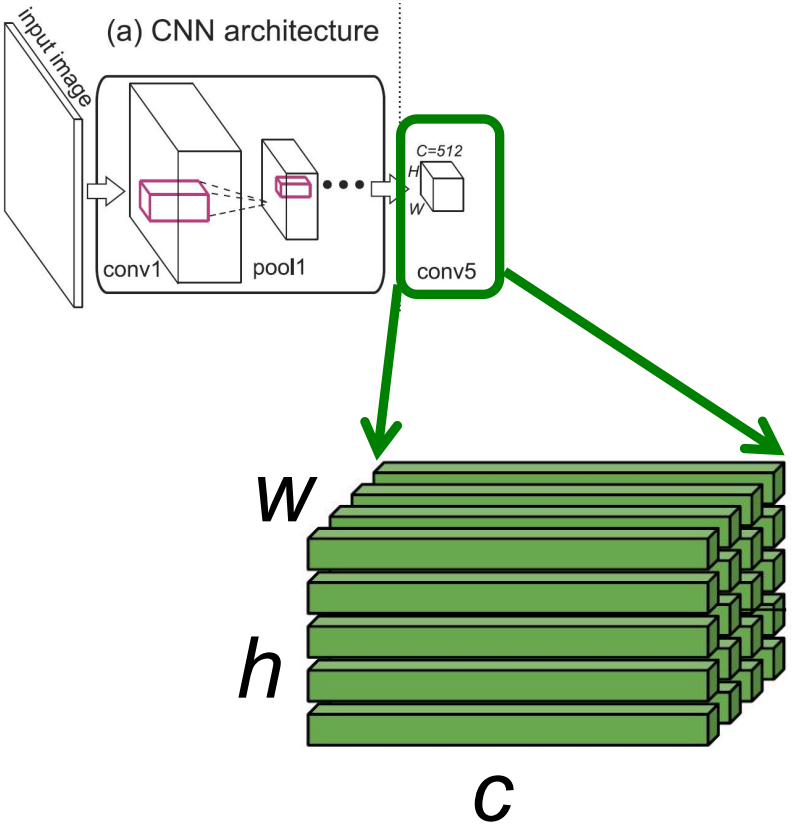


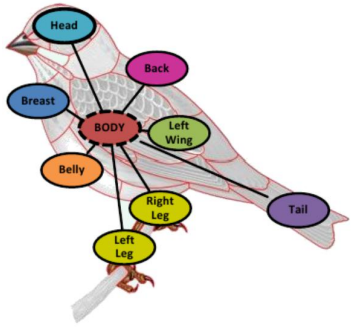
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

Bilinear Pooling

$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$





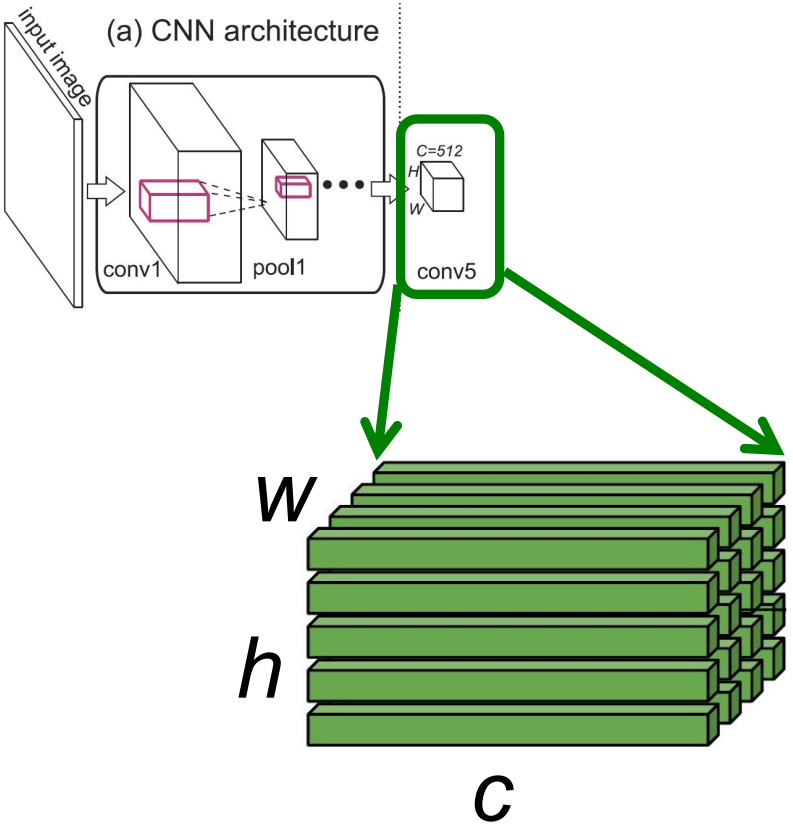
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

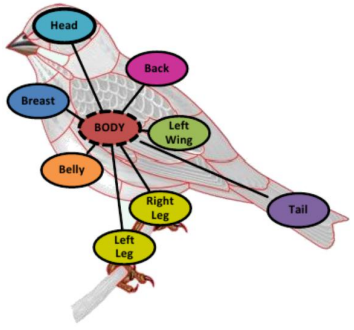
Holistic representation based method

Bilinear Pooling

$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

$$\mathbf{x}_i \in \mathbb{R}^c \quad i \in [1, hw]$$





Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

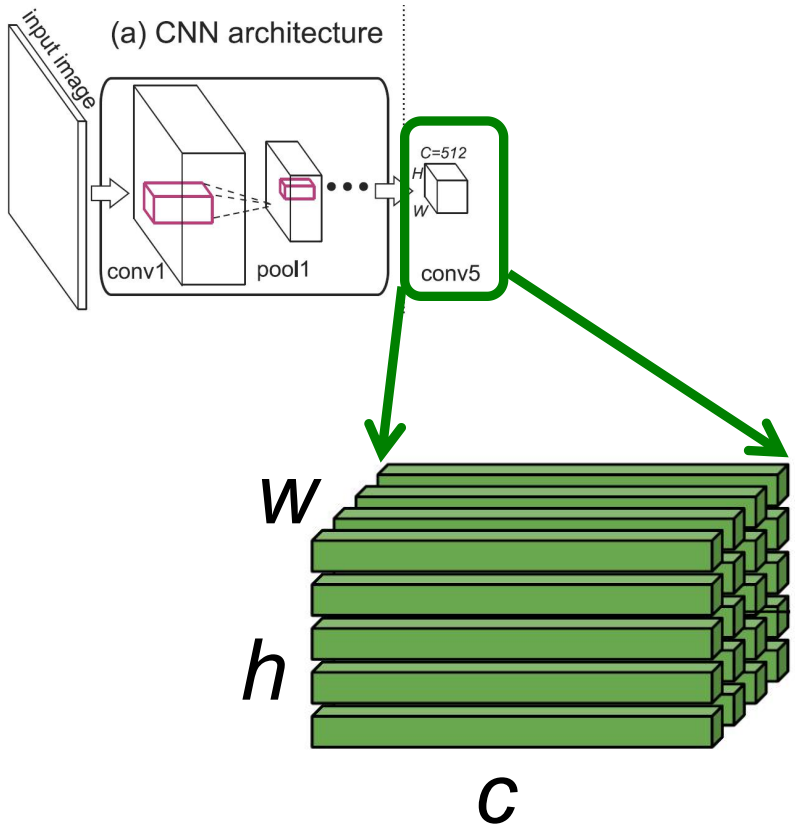
Holistic representation based method

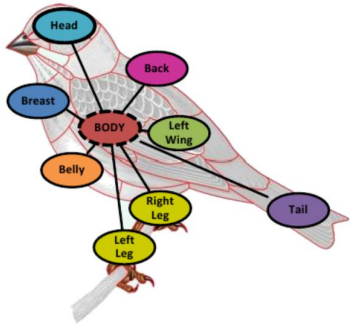
Bilinear Pooling

$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

$$\mathbf{x}_i \in \mathbb{R}^c \quad i \in [1, hw]$$

$$\mathbf{X} \in \mathbb{R}^{c \times hw}$$





Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

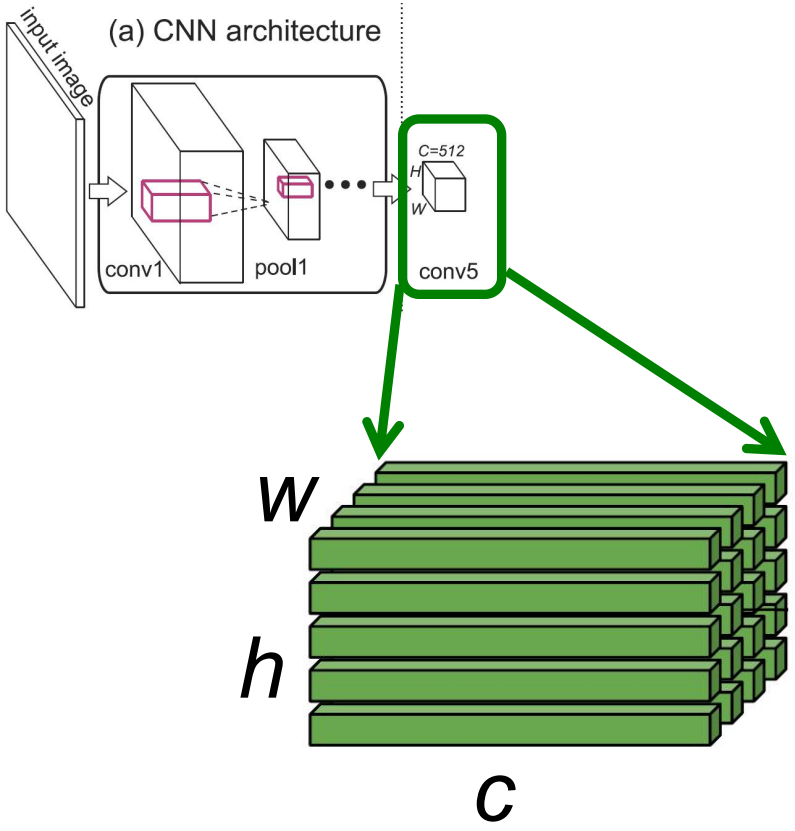
Bilinear Pooling

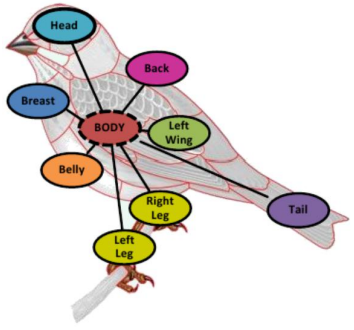
$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

$$\mathbf{x}_i \in \mathbb{R}^c \quad i \in [1, hw]$$

$$\mathbf{X} \in \mathbb{R}^{c \times hw}$$

$$\mathbf{X}\mathbf{X}^T = \sum_{i=1}^{hw} \mathbf{x}_i \mathbf{x}_i^T$$





Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

Bilinear Pooling

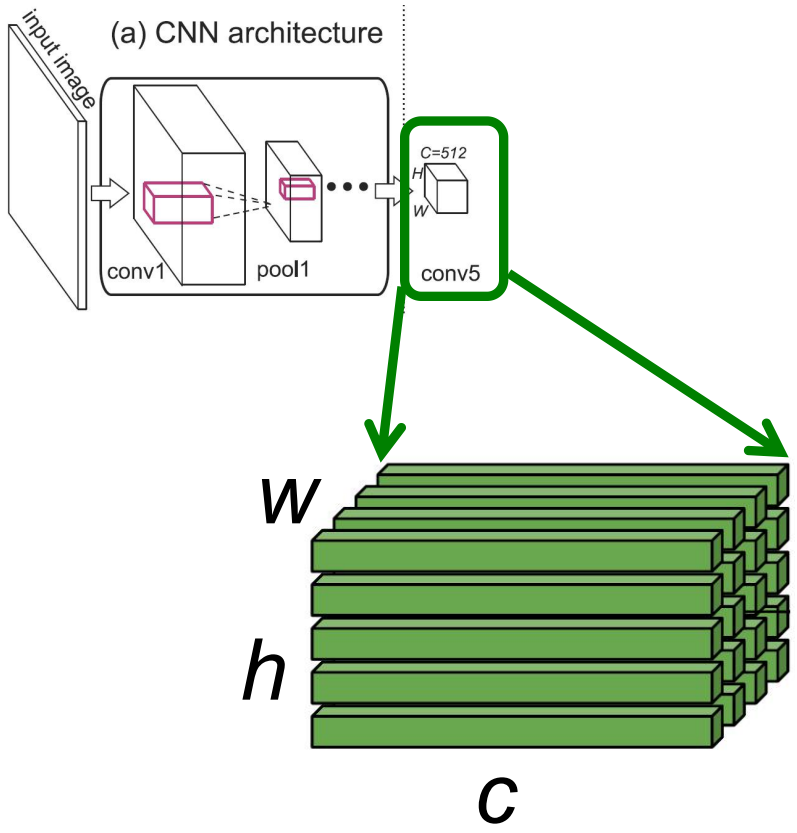
$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

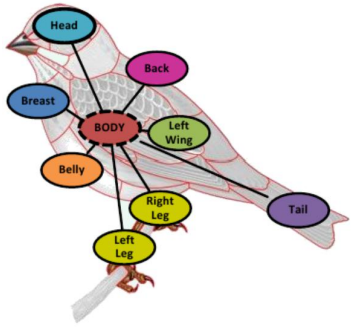
$$\mathbf{x}_i \in \mathbb{R}^c \quad i \in [1, hw]$$

$$\mathbf{X} \in \mathbb{R}^{c \times hw}$$

$$\mathbf{X}\mathbf{X}^T = \sum_{i=1}^{hw} \mathbf{x}_i \mathbf{x}_i^T$$

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

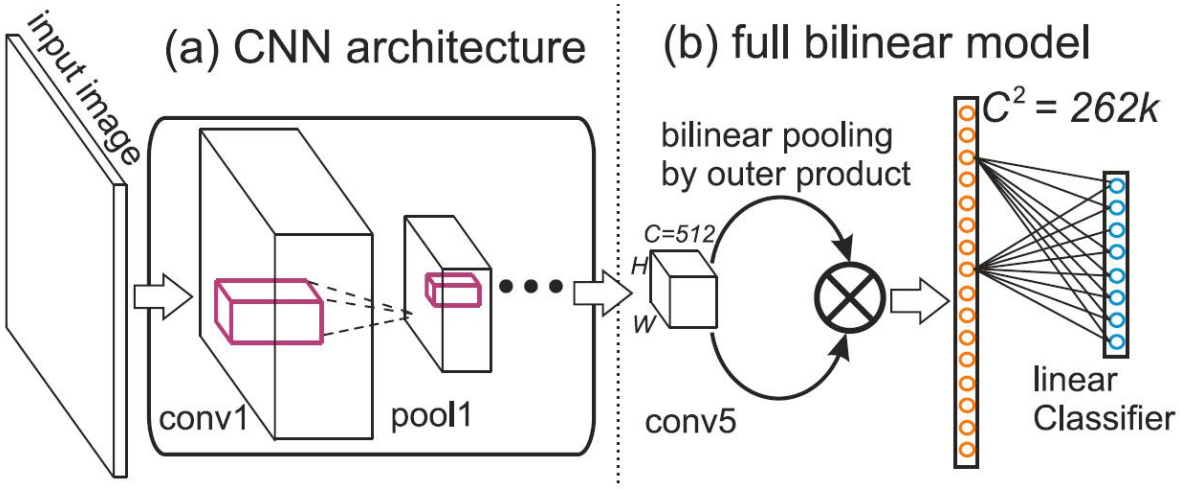


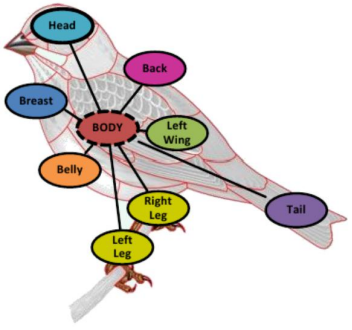


Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

Bilinear Pooling CNN -- training in an end-to-end manner

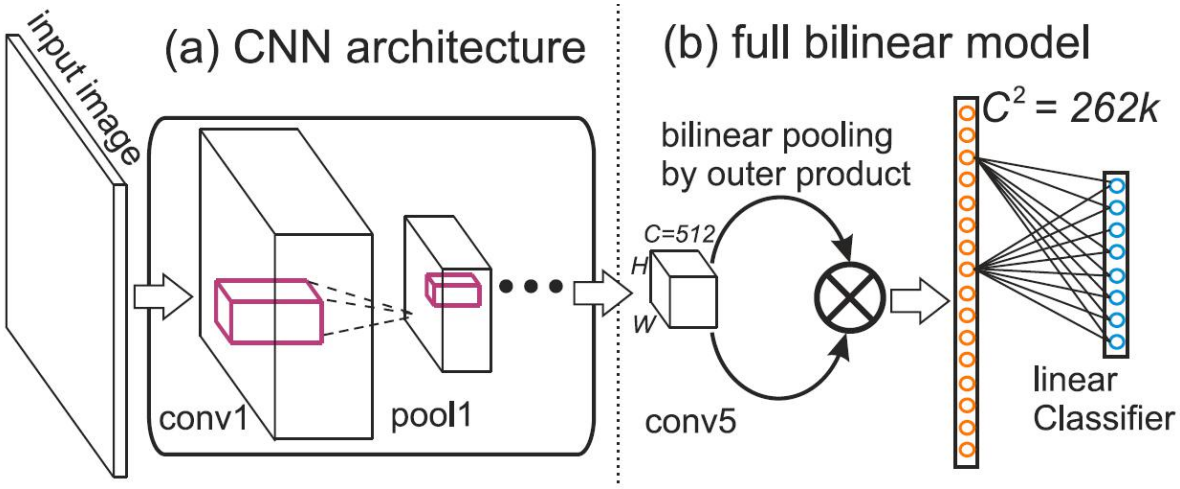




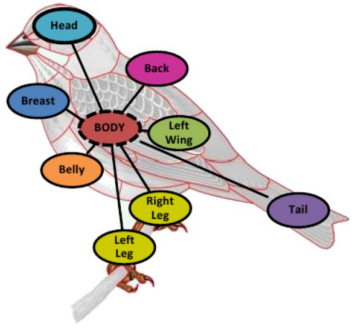
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

Bilinear Pooling CNN -- training in an end-to-end manner



good, but high dim and large model size



Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM

$$\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$$

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM

$$\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$$

$$\mathbf{w}^T vec(\mathbf{X}\mathbf{X}^T) \Leftrightarrow tr(\mathbf{W}^T \mathbf{X}\mathbf{X}^T)$$

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$
2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$
2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

Theorem 1 *Let* $\mathbf{w}^* \in \mathbb{R}^{c^2}$ *be the optimal solution of the linear SVM in Equation* 1 *over bilinear features, then* $\mathbf{W}^* = mat(\mathbf{w}^*) \in \mathbb{R}^{c \times c}$ *is the optimal solution in Equation* 2. *Moreover,* $\mathbf{W}^* = \mathbf{W}^{*T}$.

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$
2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

$$\mathbf{w}^* = \sum_{y_i=1} \alpha_i \mathbf{z}_i - \sum_{y_i=-1} \alpha_i \mathbf{z}_i$$

$$\mathbf{W}^* = \sum_{y_i=1} \alpha_i \mathbf{X}_i \mathbf{X}_i^T - \sum_{y_i=-1} \alpha_i \mathbf{X}_i \mathbf{X}_i^T$$

where $\alpha_i \geq 0, \forall i = 1, \ldots, N$

Holistic representation based method

$$\mathbf{z} = vec(\mathbf{X}\mathbf{X}^T) \in \mathbb{R}^{c^2}$$

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$

2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

$$\begin{aligned}\mathbf{W}^* &= \mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^T = \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T + \mathbf{\Psi}_-\mathbf{\Sigma}_-\mathbf{\Psi}_-^T \\ &= \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T - \mathbf{\Psi}_-|\mathbf{\Sigma}_-|\mathbf{\Psi}_-^T \\ &= \mathbf{U}_+\mathbf{U}_+^T - \mathbf{U}_-\mathbf{U}_-^T\end{aligned}$$

$$\mathbf{w}^* = \sum_{y_i=1} \alpha_i \mathbf{z}_i - \sum_{y_i=-1} \alpha_i \mathbf{z}_i$$

$$\mathbf{W}^* = \sum_{y_i=1} \alpha_i \mathbf{X}_i \mathbf{X}_i^T - \sum_{y_i=-1} \alpha_i \mathbf{X}_i \mathbf{X}_i^T$$

where $\alpha_i \geq 0, \forall i = 1, \ldots, N$

Holistic representation based method

When bilinear SVM meets bilinear feature

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$

2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

$$\begin{aligned}\mathbf{W}^* =& \mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^T = \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T + \mathbf{\Psi}_-\mathbf{\Sigma}_-\mathbf{\Psi}_-^T \\ =& \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T - \mathbf{\Psi}_-|\mathbf{\Sigma}_-|\mathbf{\Psi}_-^T \\ =& \mathbf{U}_+\mathbf{U}_+^T - \mathbf{U}_-\mathbf{U}_-^T\end{aligned}$$

$$\max(0, 1 - y_i\{\|\mathbf{U}_+^T\mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T\mathbf{X}_i\|_F^2\} + b)$$

$$\max(0, 1 - y_i\{\mathrm{tr}(\mathbf{U}_+\mathbf{U}_+^T\mathbf{X}_i\mathbf{X}_i^T) - \mathrm{tr}(\mathbf{U}_-\mathbf{U}_-^T\mathbf{X}_i\mathbf{X}_i^T)\} + b)$$

Holistic representation based method

maximum Frobenius margin

$$\max(0, 1 - y_i\{\|\mathbf{U}_+^T\mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T\mathbf{X}_i\|_F^2\} + b)$$

$$\max(0, 1 - y_i\{\mathsf{tr}(\mathbf{U}_+\mathbf{U}_+^T\mathbf{X}_i\mathbf{X}_i^T) - \mathsf{tr}(\mathbf{U}_-\mathbf{U}_-^T\mathbf{X}_i\mathbf{X}_i^T)\} + b)$$

Holistic representation based method

maximum Frobenius margin

no need to compute bilinear features when testing

$$\max(0, 1 - y_i\{\|\mathbf{U}_+^T\mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T\mathbf{X}_i\|_F^2\} + b)$$

$$\max(0, 1 - y_i\{\mathsf{tr}(\mathbf{U}_+\mathbf{U}_+^T\mathbf{X}_i\mathbf{X}_i^T) - \mathsf{tr}(\mathbf{U}_-\mathbf{U}_-^T\mathbf{X}_i\mathbf{X}_i^T)\} + b)$$

Holistic representation based method

When bilinear SVM meets bilinear feature

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$

2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

$$
\begin{aligned}
\mathbf{W}^* &= \mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^T = \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T + \mathbf{\Psi}_-\mathbf{\Sigma}_-\mathbf{\Psi}_-^T \\
&= \mathbf{\Psi}_+\mathbf{\Sigma}_+\mathbf{\Psi}_+^T - \mathbf{\Psi}_-|\mathbf{\Sigma}_-|\mathbf{\Psi}_-^T \\
&= \mathbf{U}_+\mathbf{U}_+^T - \mathbf{U}_-\mathbf{U}_-^T
\end{aligned}
$$

$$\max(0, 1 - y_i\{\|\mathbf{U}_+^T\mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T\mathbf{X}_i\|_F^2\} + b)$$

$$\max(0, 1 - y_i\{\mathrm{tr}(\mathbf{U}_+\mathbf{U}_+^T\mathbf{X}_i\mathbf{X}_i^T) - \mathrm{tr}(\mathbf{U}_-\mathbf{U}_-^T\mathbf{X}_i\mathbf{X}_i^T)\} + b)$$

Holistic representation based method

When bilinear SVM meets bilinear feature

1. linear SVM $\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$

2. linear SVM in matrix $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$

3. rank-r linear SVM $\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}_r^T \mathbf{X}\ \mathbf{X}^T) + b)$

$$\max(0, 1 - y_i \{ \|\mathbf{U}_+^T \mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T \mathbf{X}_i\|_F^2 \} + b)$$

$$\max(0, 1 - y_i \{ \mathrm{tr}(\mathbf{U}_+ \mathbf{U}_+^T \mathbf{X}_i \mathbf{X}_i^T) - \mathrm{tr}(\mathbf{U}_- \mathbf{U}_-^T \mathbf{X}_i \mathbf{X}_i^T) \} + b)$$

Holistic representation based method

When bilinear SVM meets bilinear feature

1. linear SVM	$\max(0, 1 - y_i \mathbf{w}^T \mathbf{z}_i + b)$
2. linear SVM in matrix	$\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T \mathbf{X}_i \mathbf{X}_i^T) + b)$
3. rank-r linear SVM	$\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}_r^T \mathbf{X}\ \mathbf{X}^T) + b)$

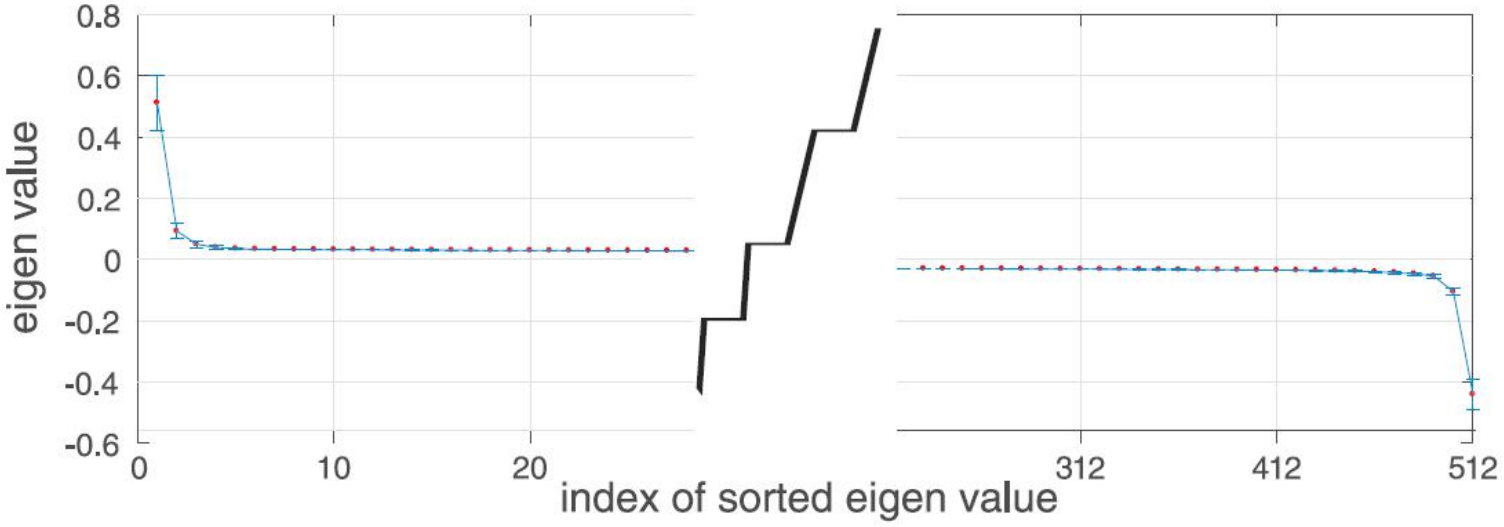
This reduces degrees of freedom of learning parameters

$$\max(0, 1 - y_i \{ \|\mathbf{U}_+^T \mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T \mathbf{X}_i\|_F^2 \} + b)$$

$$\max(0, 1 - y_i \{ \mathrm{tr}(\mathbf{U}_+ \mathbf{U}_+^T \mathbf{X}_i \mathbf{X}_i^T) - \mathrm{tr}(\mathbf{U}_- \mathbf{U}_-^T \mathbf{X}_i \mathbf{X}_i^T) \} + b)$$

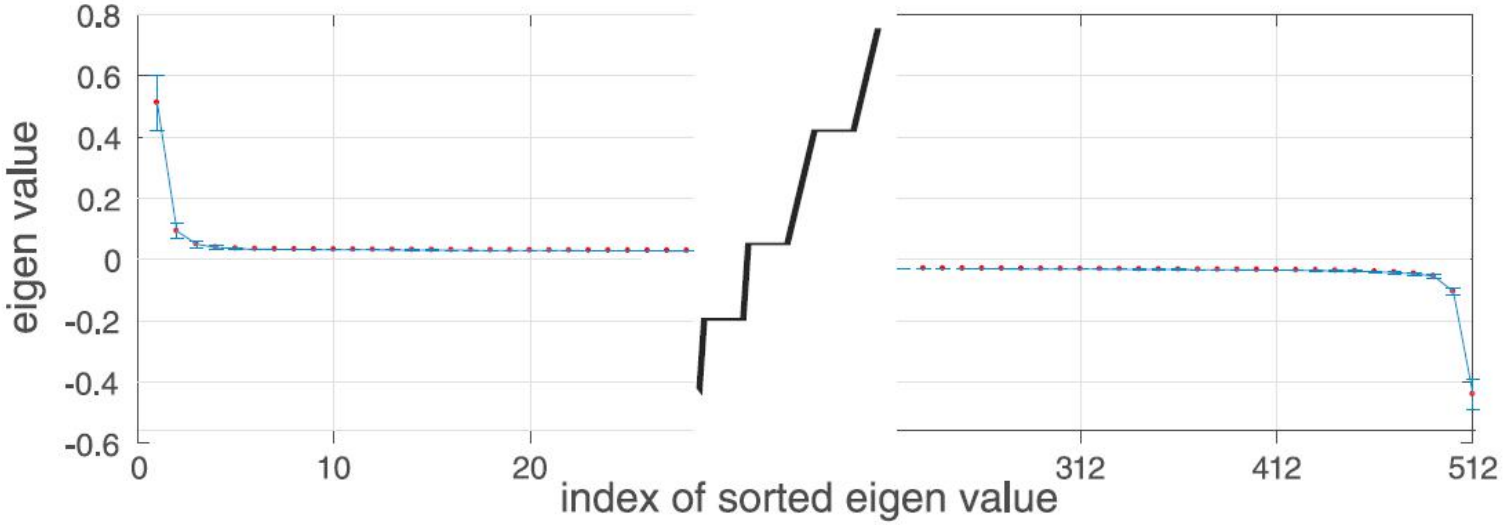
Holistic representation based method

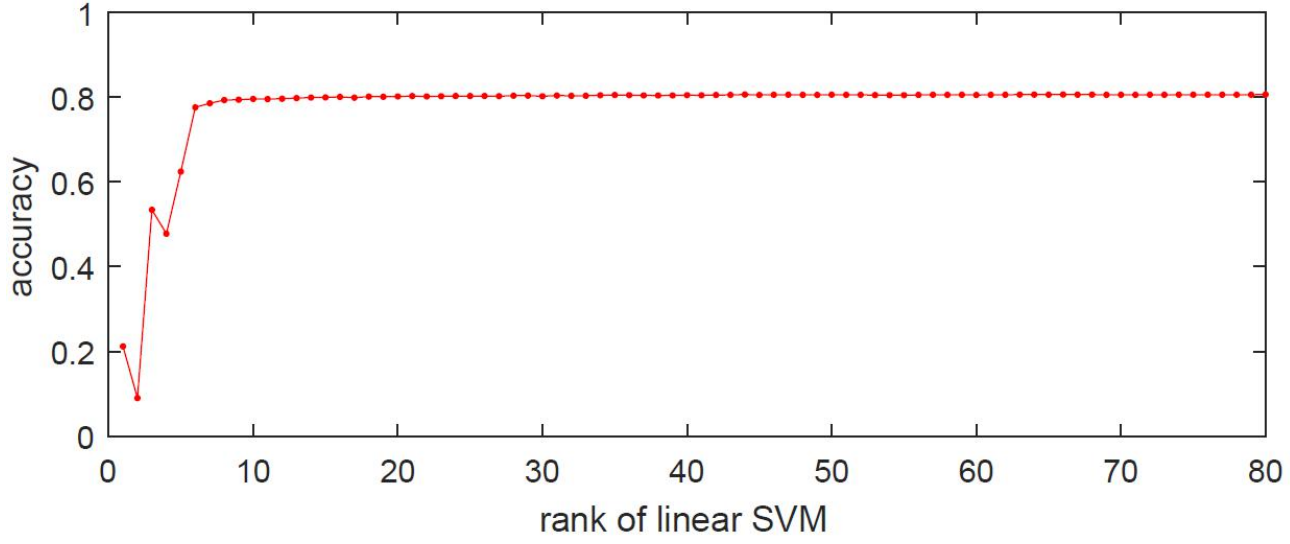
Low-rank SVM



Holistic representation based method

Low-rank SVM





Holistic representation based method

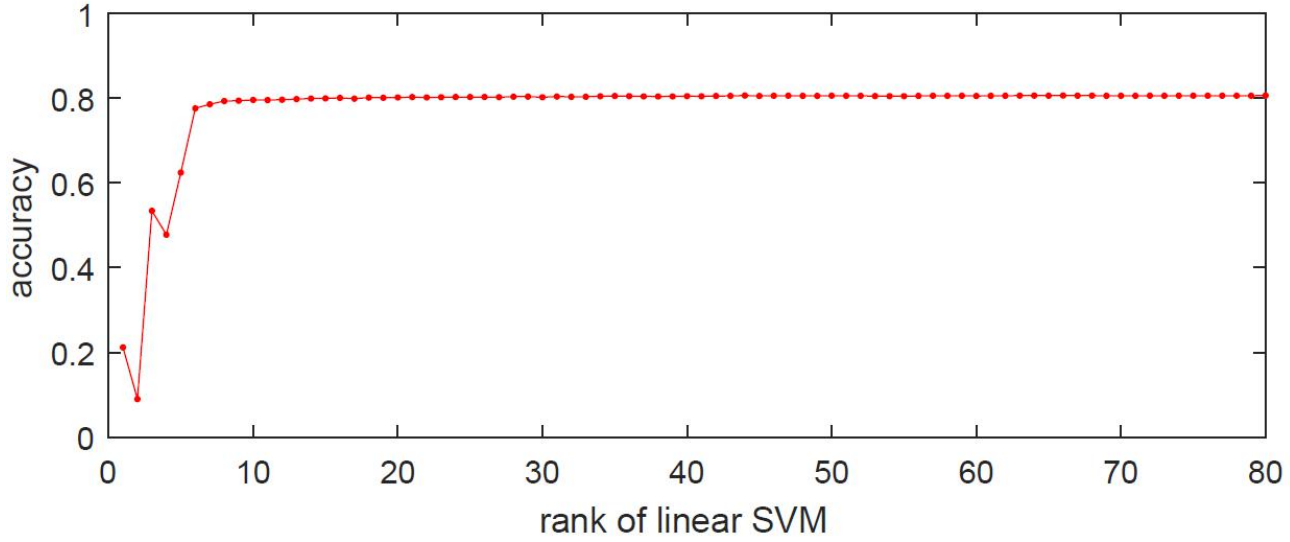
Low-rank SVM

200 classes, then param

size is reduced

from **200*512*512**

to **200*512*8**



Holistic representation based method

classifier co-decomposition -- learning a common factor and class-specific parameters of smaller size

$$\min_{\mathbf{V}_k,\mathbf{P}} \sum_{k=1}^{K} \|\mathbf{U}_k - \mathbf{P}\mathbf{V}_k\|_F^2$$

$$\mathbf{U}_k = [\mathbf{U}_{+k}, \mathbf{U}_{-k}] \in \mathbb{R}^{c\times r}$$

$$\mathbf{P} \in \mathbb{R}^{c\times m}$$

$$\mathbf{V}_k \in \mathbb{R}^{m\times r}$$

$$m < c$$

Holistic representation based method

classifier co-decomposition -- learning a common factor and class-specific parameters of smaller size

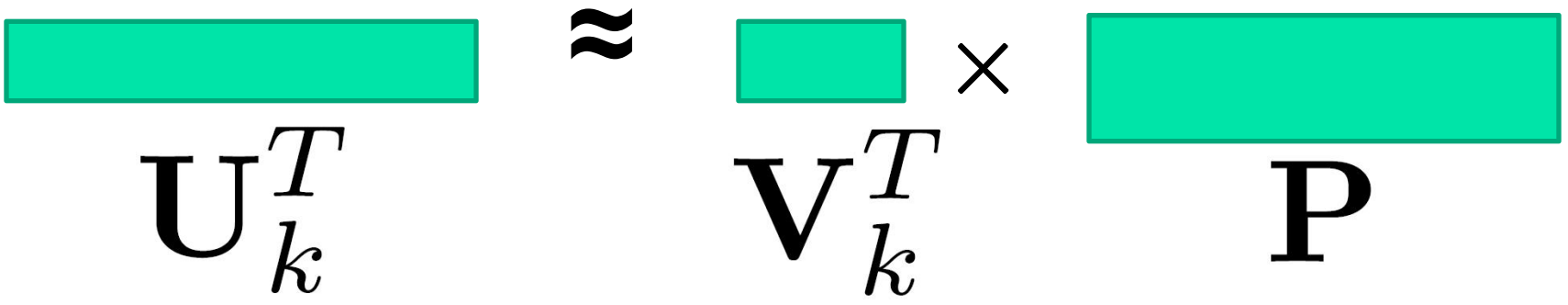
$$\min_{\mathbf{V}_k,\mathbf{P}} \sum_{k=1}^{K} \|\mathbf{U}_k - \mathbf{P}\mathbf{V}_k\|_F^2$$

$$\mathbf{U}_k = [\mathbf{U}_{+k}, \mathbf{U}_{-k}] \in \mathbb{R}^{c\times r}$$

$$\mathbf{P} \in \mathbb{R}^{c\times m}$$

$$\mathbf{V}_k \in \mathbb{R}^{m\times r}$$

$$m < c$$



Holistic representation based method

classifier co-decomposition -- learning a common factor and class-specific parameters of smaller size

$$\min_{\mathbf{V}_k,\mathbf{P}} \sum_{k=1}^{K} \|\mathbf{U}_k - \mathbf{P}\mathbf{V}_k\|_F^2$$

$$\mathbf{U}_k = [\mathbf{U}_{+k}, \mathbf{U}_{-k}] \in \mathbb{R}^{c\times r}$$

$$\mathbf{P} \in \mathbb{R}^{c\times m}$$

$$\mathbf{V}_k \in \mathbb{R}^{m\times r}$$

$$m < c$$

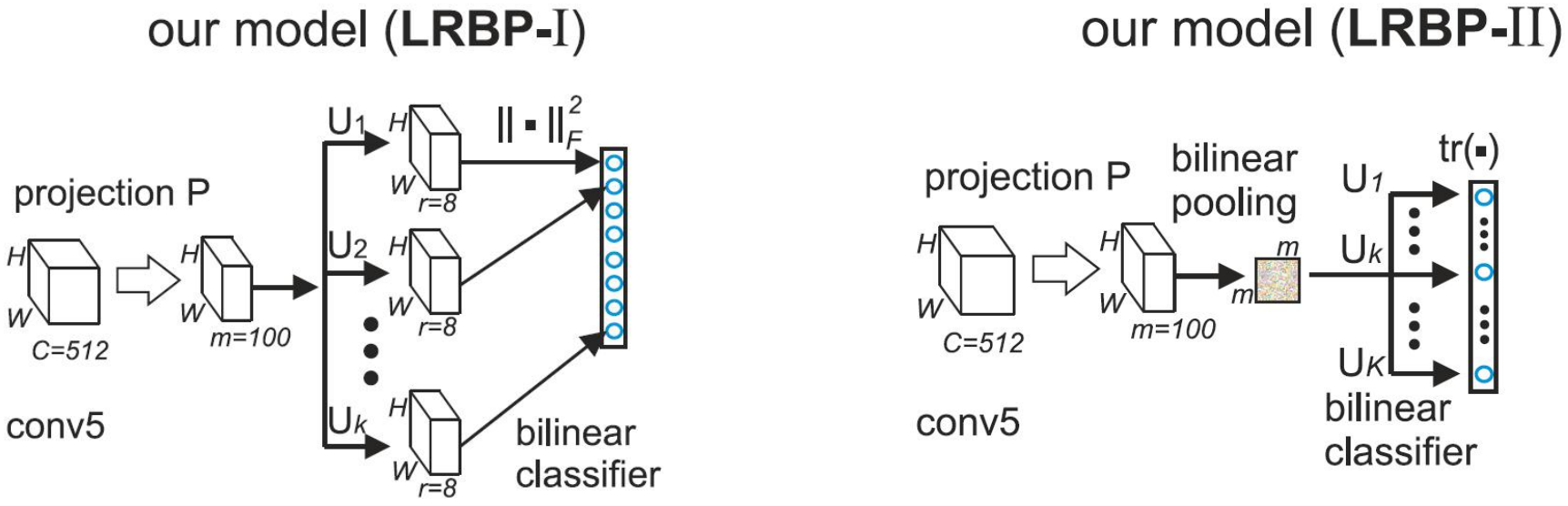
Theorem 2 *The optimal solution of* $\mathbf{P}$ *to Equation 11 spans the subspace of the singular vectors corresponding to the largest* m *singular values of* $[\mathbf{U}_1, \ldots, \mathbf{U}_K]$.

Holistic representation based method

building one convolutional layer for P

Holistic representation based method

building one convolutional layer for P

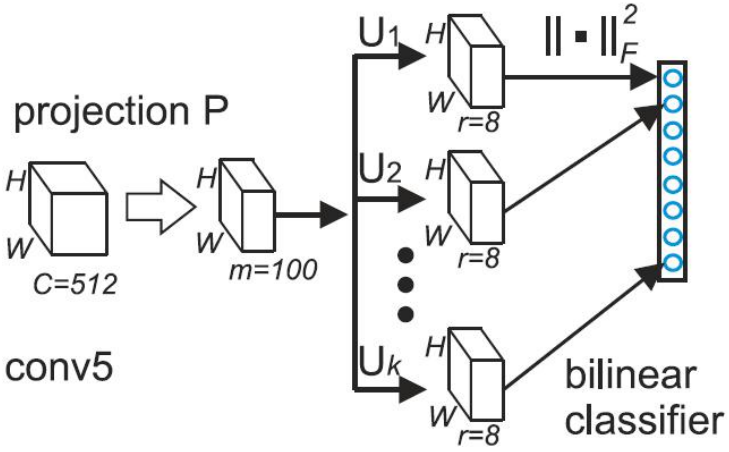


Holistic representation based method

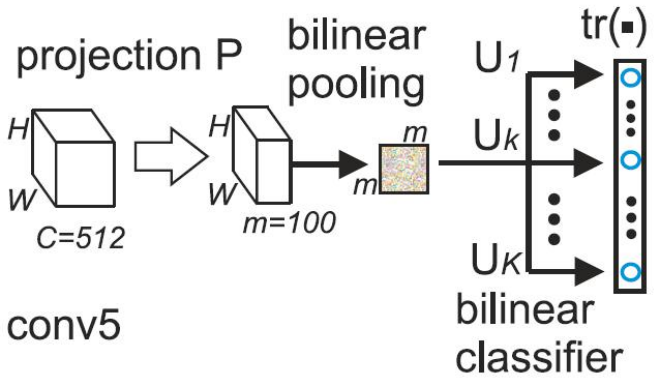
Studying the two hyperparameters -- *m* and *r*

- low dimension *m* determined by P
- low *rank r* for classifier parameters

our model (**LRBP-I**)



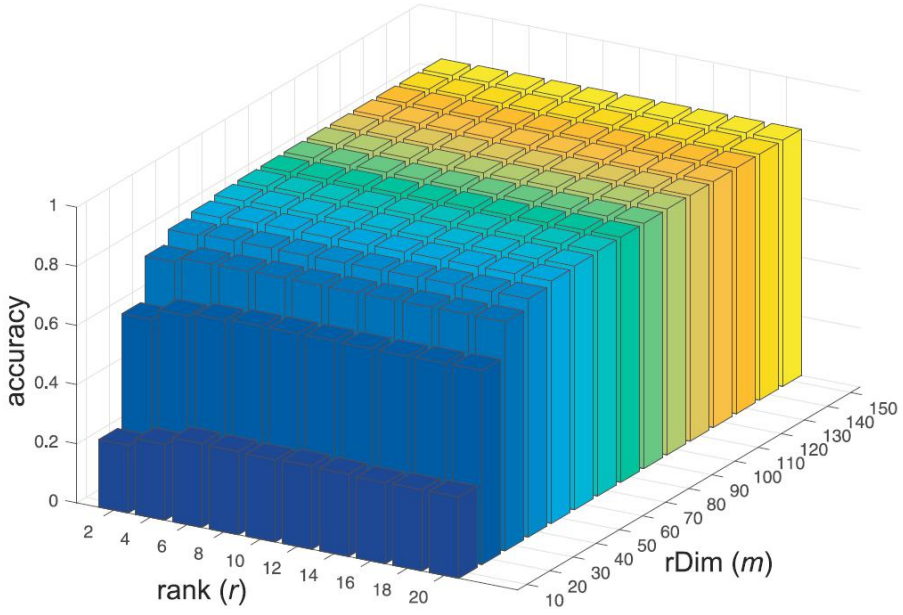
our model (**LRBP-II**)



S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

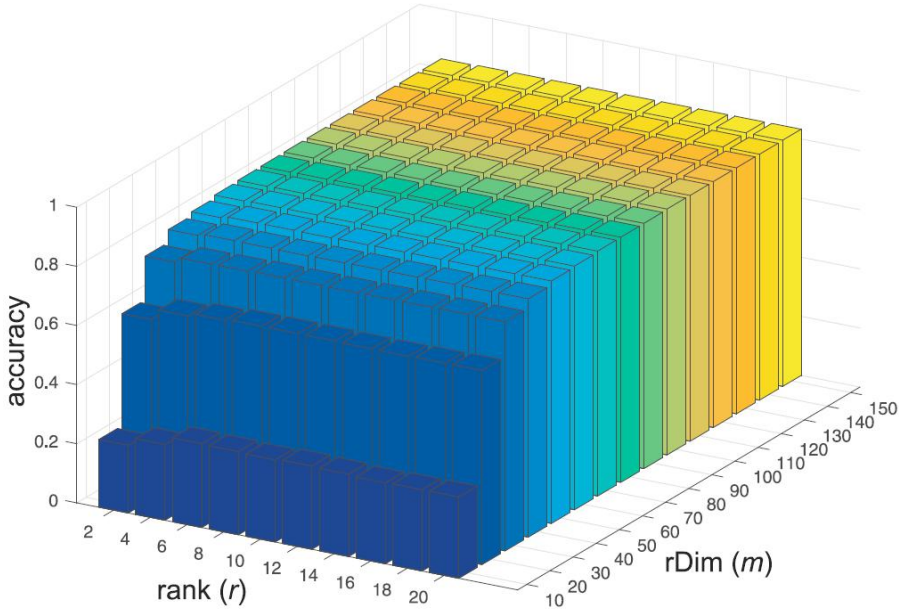
Holistic representation based method

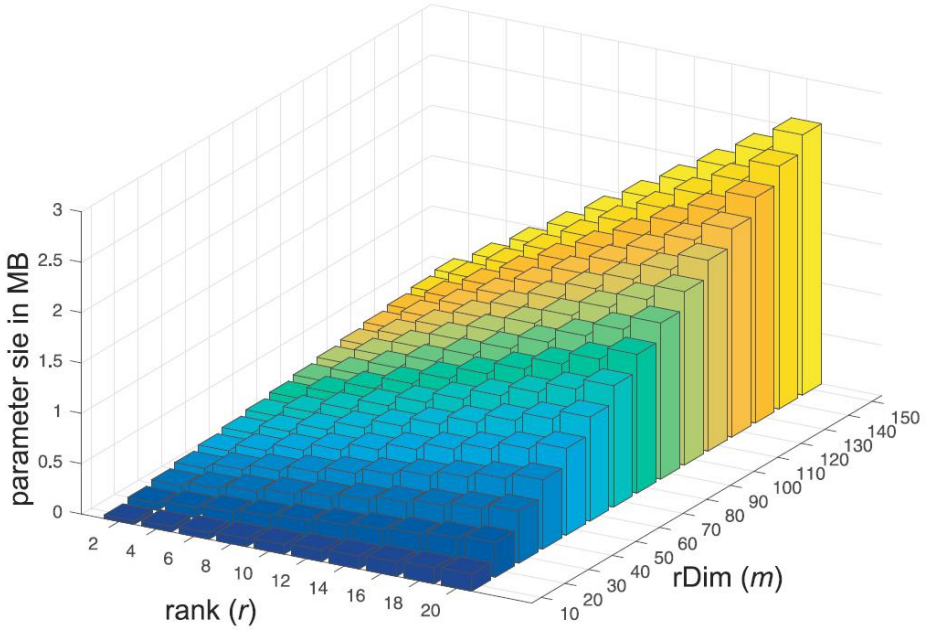
Studying the two hyperparameters -- *m* and *r*



Holistic representation based method

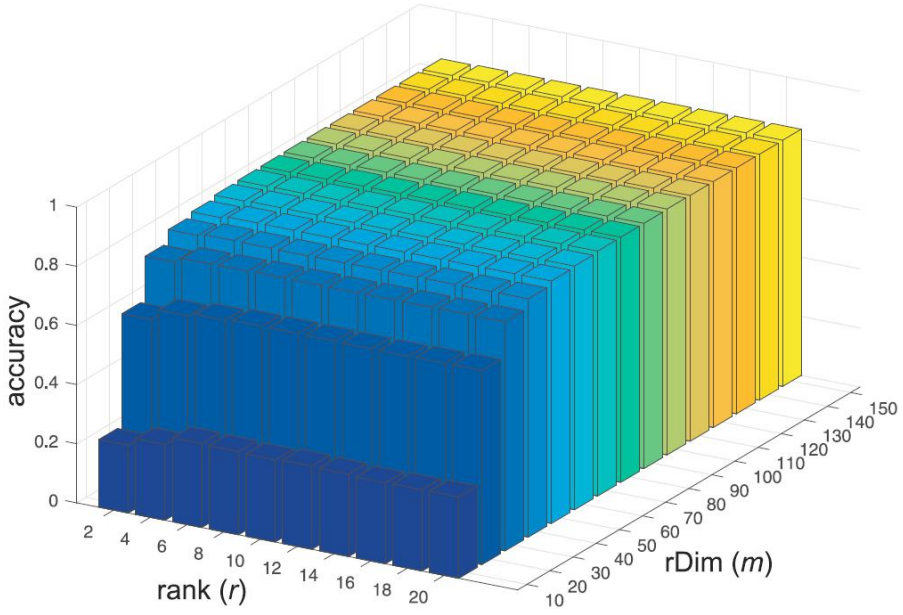
Studying the two hyperparameters -- *m* and *r*

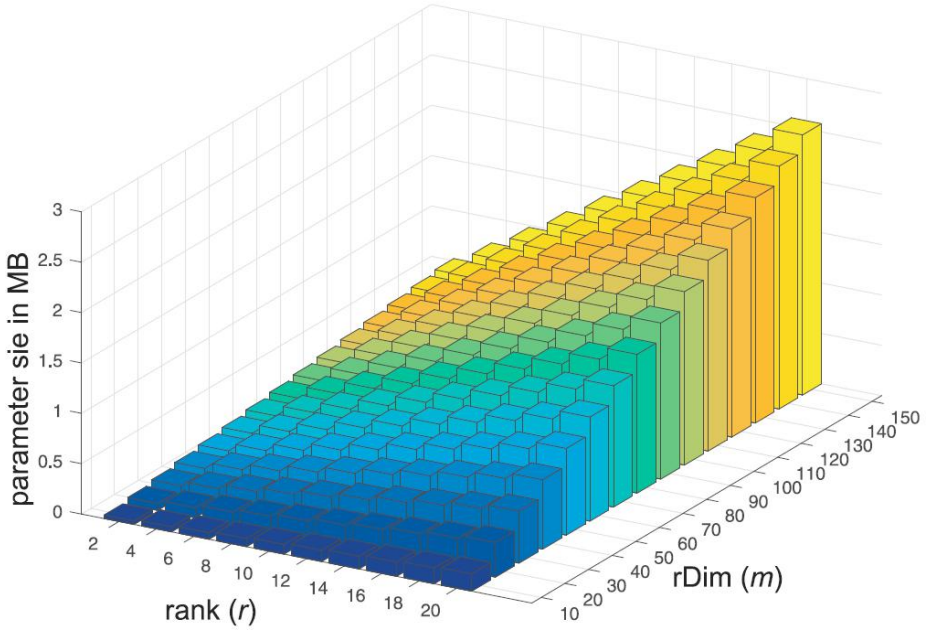




Holistic representation based method

Studying the two hyperparameters -- *m* and *r*





if 200 classes, then param size is reduced

from 200*512*512 (~52.4 x 10e6 single, 200MB)

to (200*8*100+100*512) (~0.21 x 10e6 single, 0.8MB)

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Holistic representation based method

Quantitative evaluation on benchmark datasets

Table 3: Summary statistics of datasets.

	# train img.	# test img.	# class
CUB [31]	5994	5794	200
DTD [4]	1880	3760	47
Car [17]	8144	8041	196
Airplane [21]	6667	3333	100

Holistic representation based method

Quantitative evaluation on benchmark datasets

Table 3: Summary statistics of datasets.

	# train img.	# test img.	# class
CUB [31]	5994	5794	200
DTD [4]	1880	3760	47
Car [17]	8144	8041	196
Airplane [21]	6667	3333	100

	FC-VGG16	Fisher	Full Bilinear	Random Maclaurin	Tensor Sketch	LRBP (Ours)
CUB [31]	70.40	74.7	84.01	83.86	84.00	**84.21**
DTD [4]	59.89	65.53	64.96	65.57	64.51	**65.80**
Car [17]	76.80	85.70	91.18	89.54	90.19	**90.92**
Airplane [21]	74.10	77.60	87.09	87.10	87.18	**87.31**
param. size (CUB)	67MB	50MB	200MB	48MB	8MB	0.8MB

S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

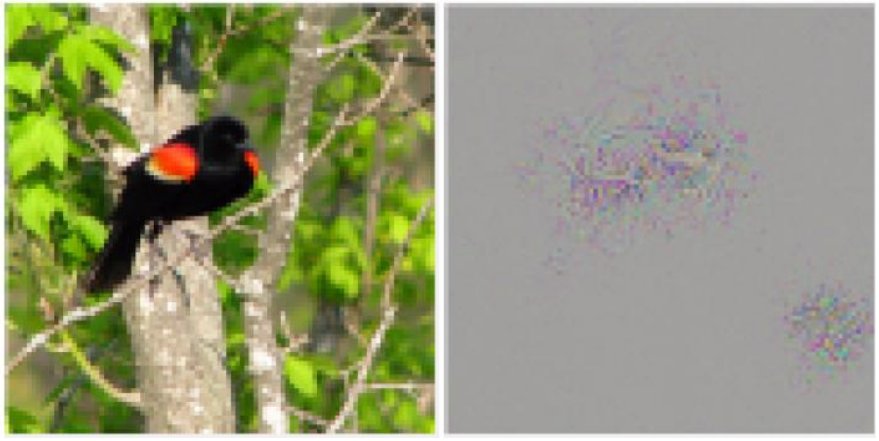
Holistic representation based method

Qualitative evaluation for understanding the model

Holistic representation based method

Qualitative evaluation for understanding the model

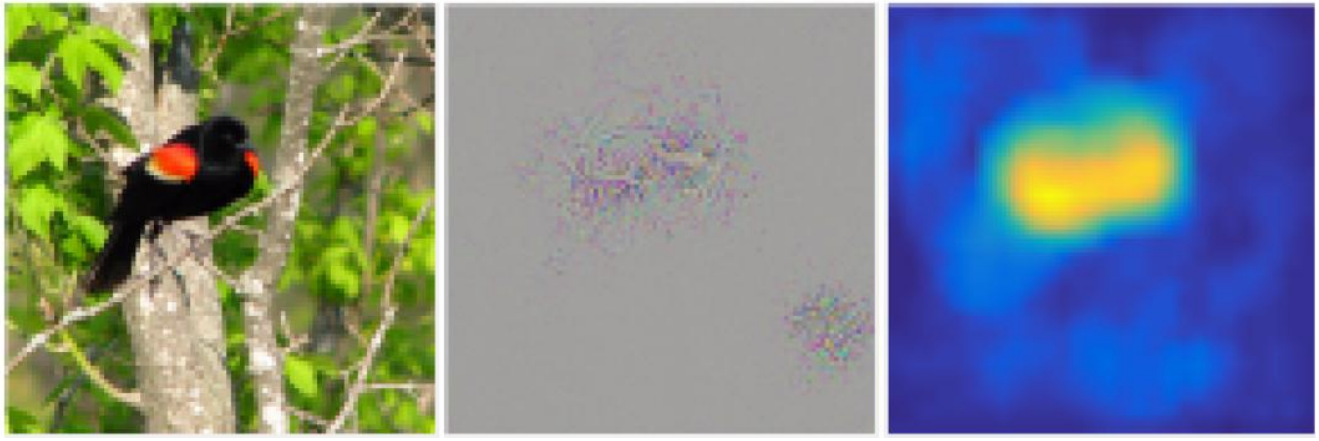
- gradient map --- backpropogating error to input image



Holistic representation based method

Qualitative evaluation for understanding the model

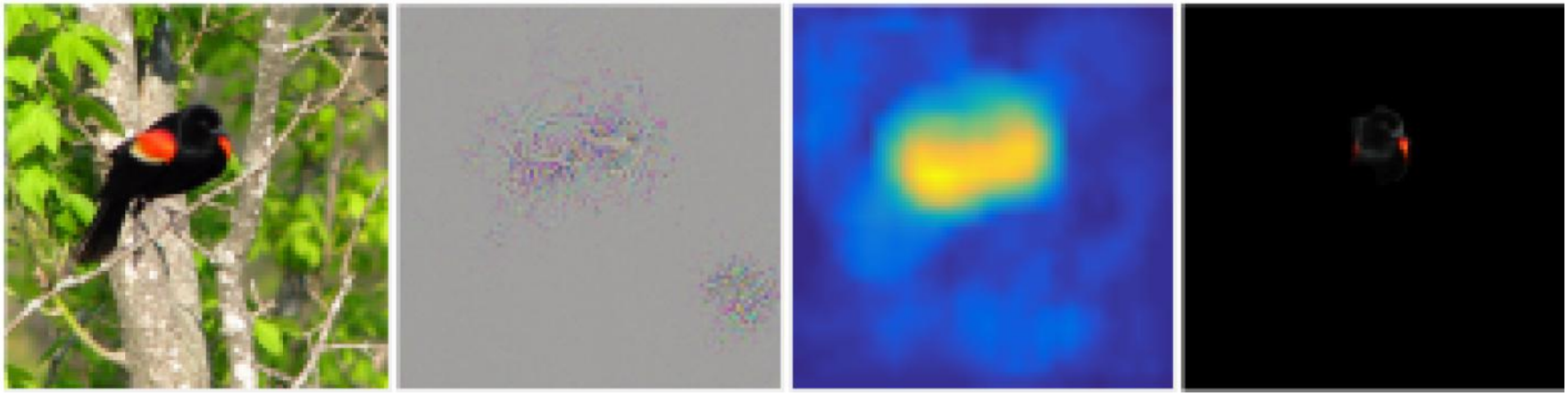
- gradient map --- backpropogating error to input image
- average activation maps



Holistic representation based method

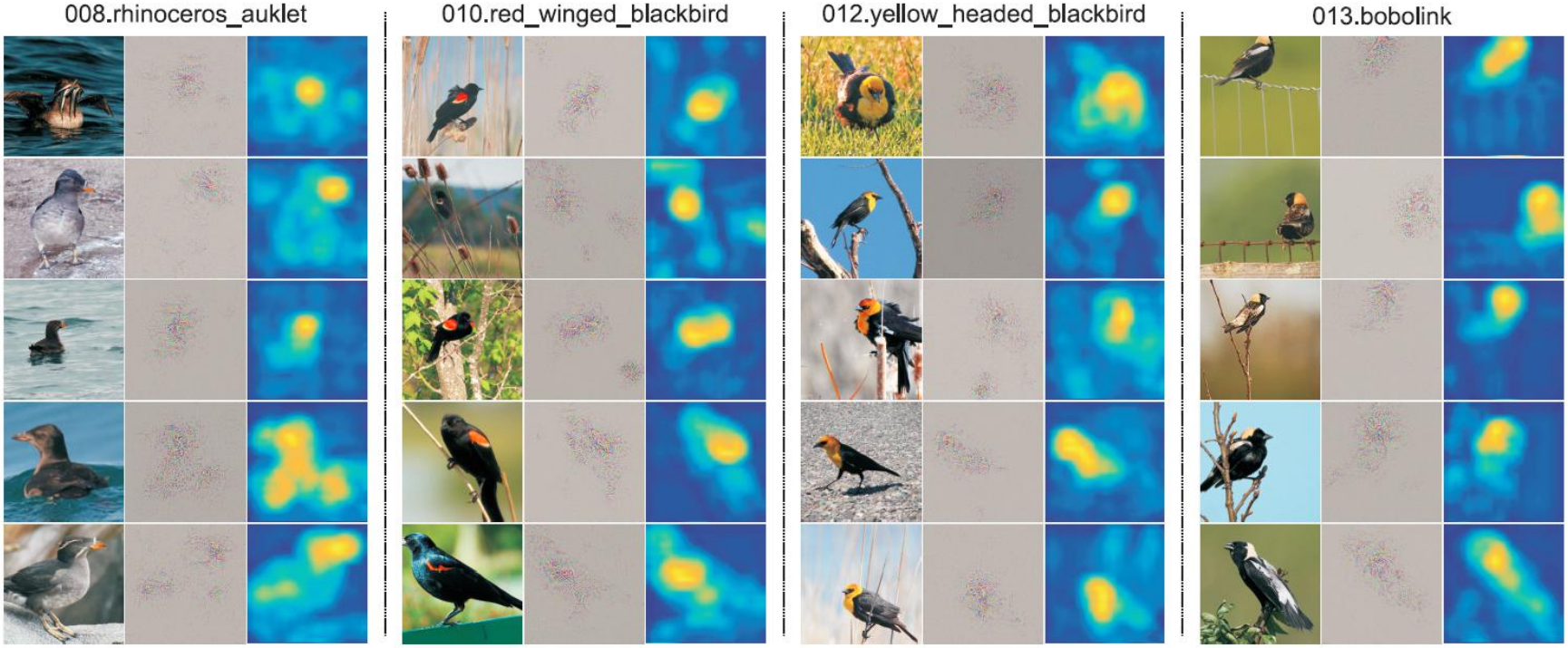
Qualitative evaluation for understanding the model

- gradient map --- backpropogating error to input image
- average activation map
- simplifying input image by removing superpixels



Holistic representation based method

Qualitative evaluation for understanding the model



S. Kong, C. Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Patch-match based method

1. Problem definition
2. Instantiation
3. Challenge and philosophy
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Patch-match based method

patch-match based approach for pollen grain identification

Patch-match based method

patch-match based approach for pollen grain identification

problem

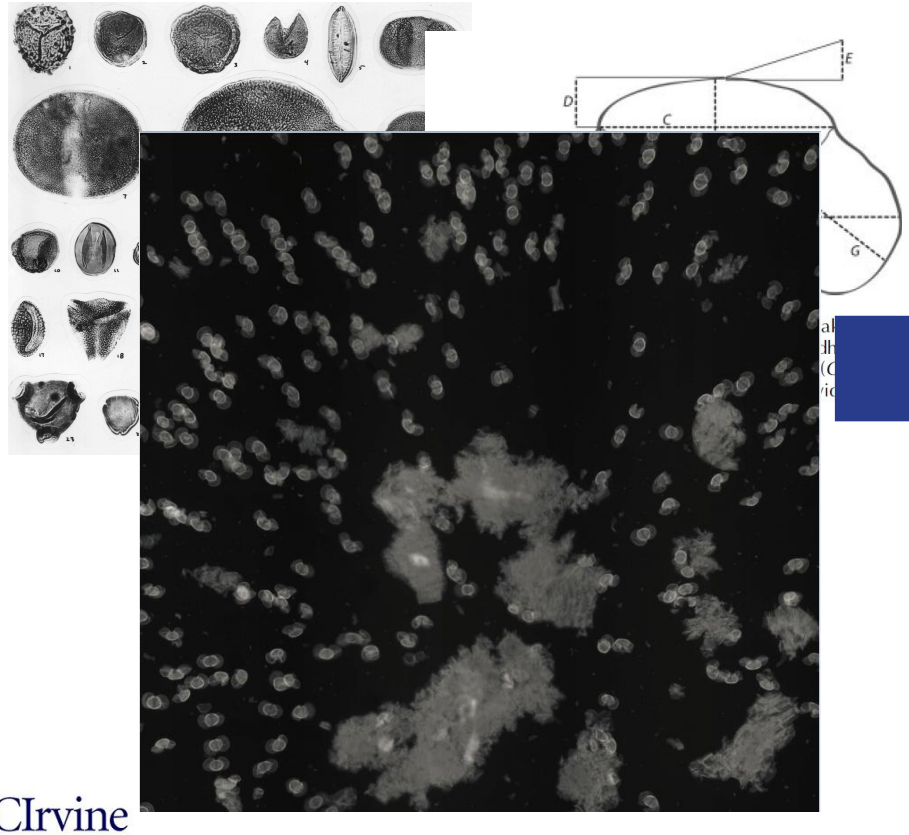
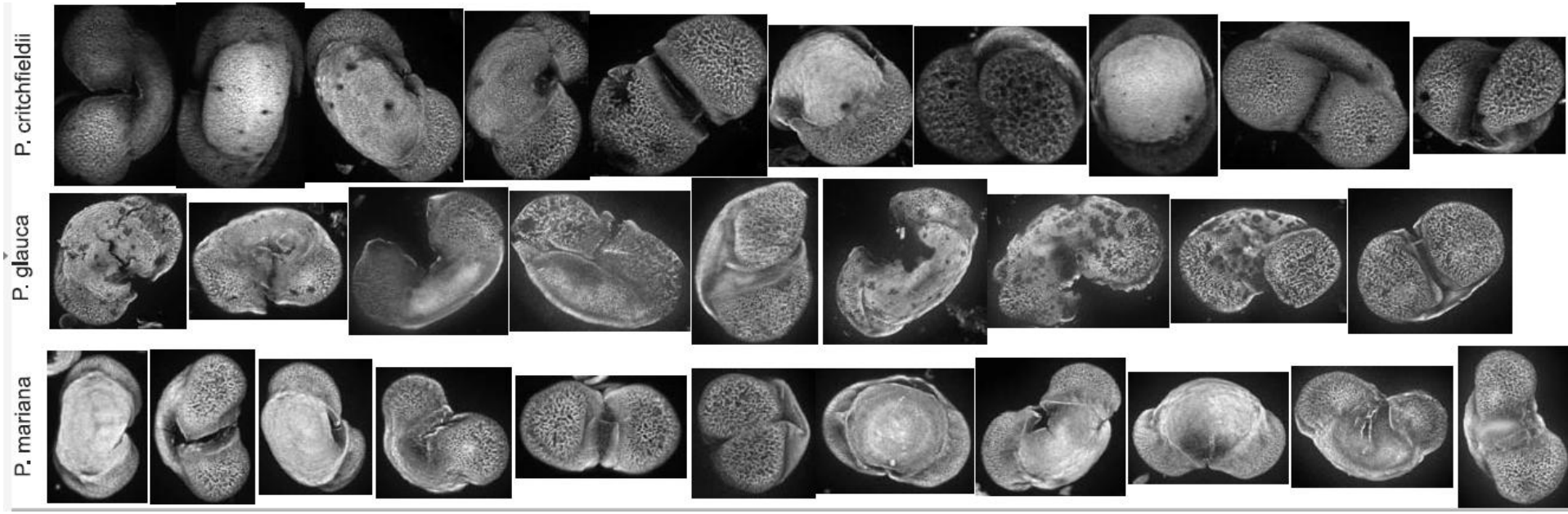


image from Surangi W. Punyasena

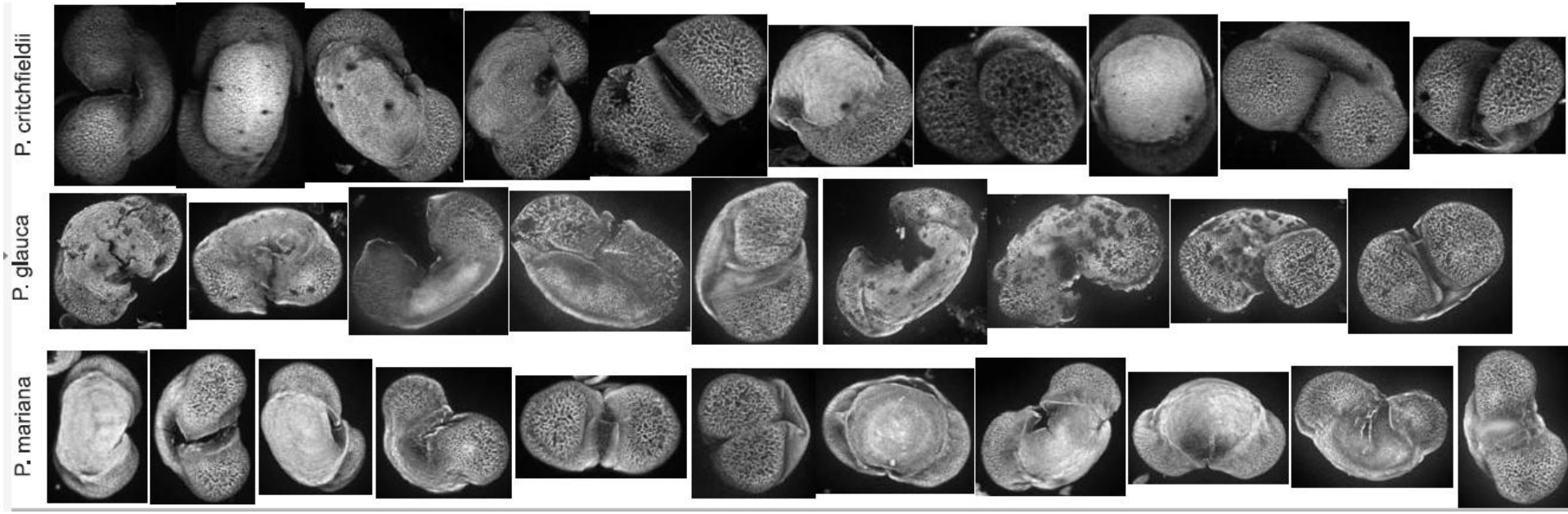
Patch-match based method

A specific dataset for this exploration



Patch-match based method

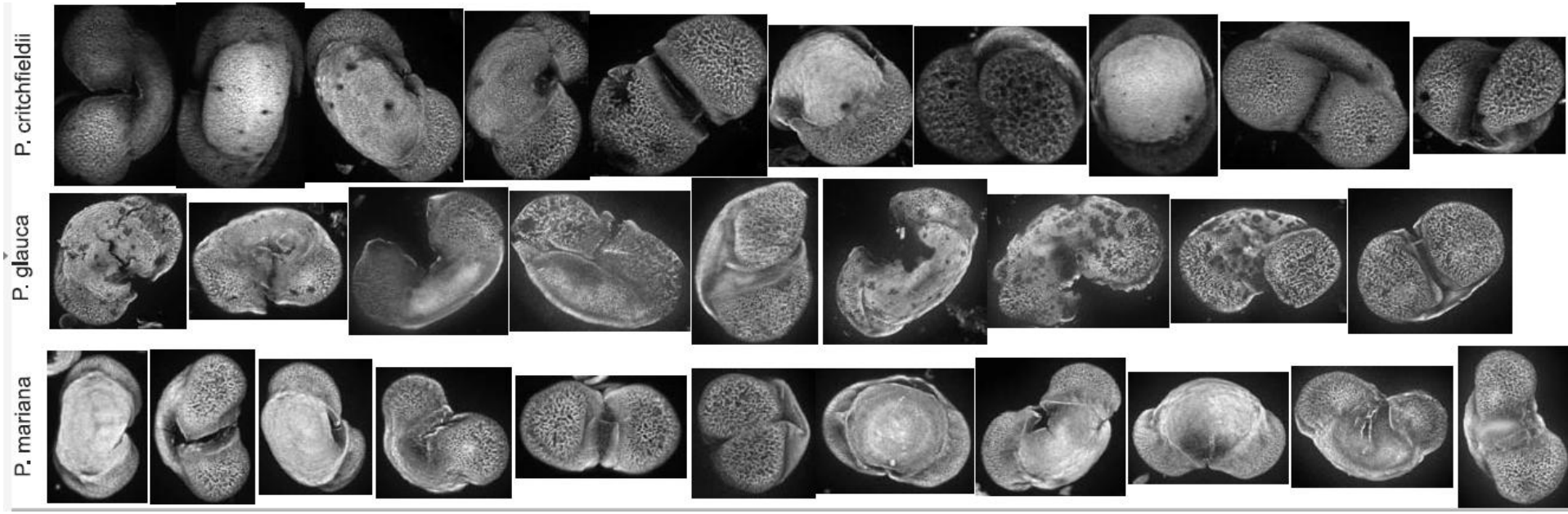
A specific dataset for this exploration



1. arbitrary viewpoint of the pollen grains

Patch-match based method

A specific dataset for this exploration



1. arbitrary viewpoint of the pollen grains
2. Large intra-class and small inter-class variation

Quantitative Result on Fossil Pollen

Why not holistic representation?

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Quantitative Result on Fossil Pollen

Why not holistic representation?

It is expensive to collect and annotate data.

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Quantitative Result on Fossil Pollen

Why not holistic representation?

It is expensive to collect and annotate data.

So there are not enough training data to learn holistic representation.

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Quantitative Result on Fossil Pollen

Why not holistic representation?

Table 1. Statistics of our fossil pollen grain dataset.

	#train	#test	#total
P. critchfieldii	65	43	108
P. glauca	65	355	420
P. mariana	65	287	352
Summary	195	685	880

It is expensive to collect and annotate data.

So there are not enough training data to learn holistic representation.

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Quantitative Result on Fossil Pollen

Why not holistic representation?

Table 1. Statistics of our fossil pollen grain dataset.

	#train	#test	#total
P. critchfieldii	65	43	108
P. glauca	65	355	420
P. mariana	65	287	352
Summary	195	685	880

It is expensive to collect and annotate data.

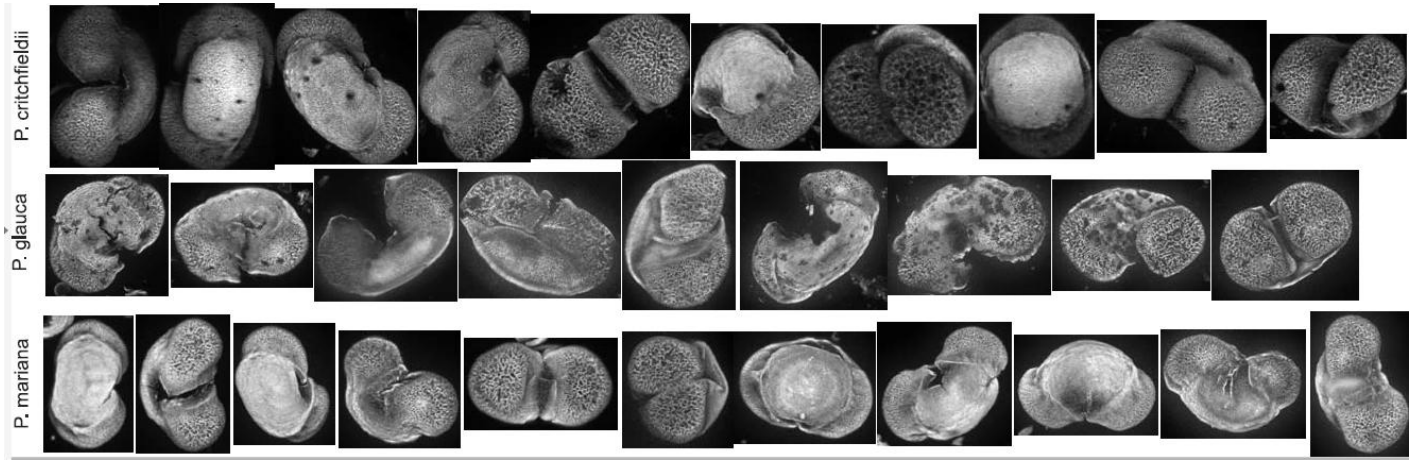
So there are not enough training data to learn holistic representation.

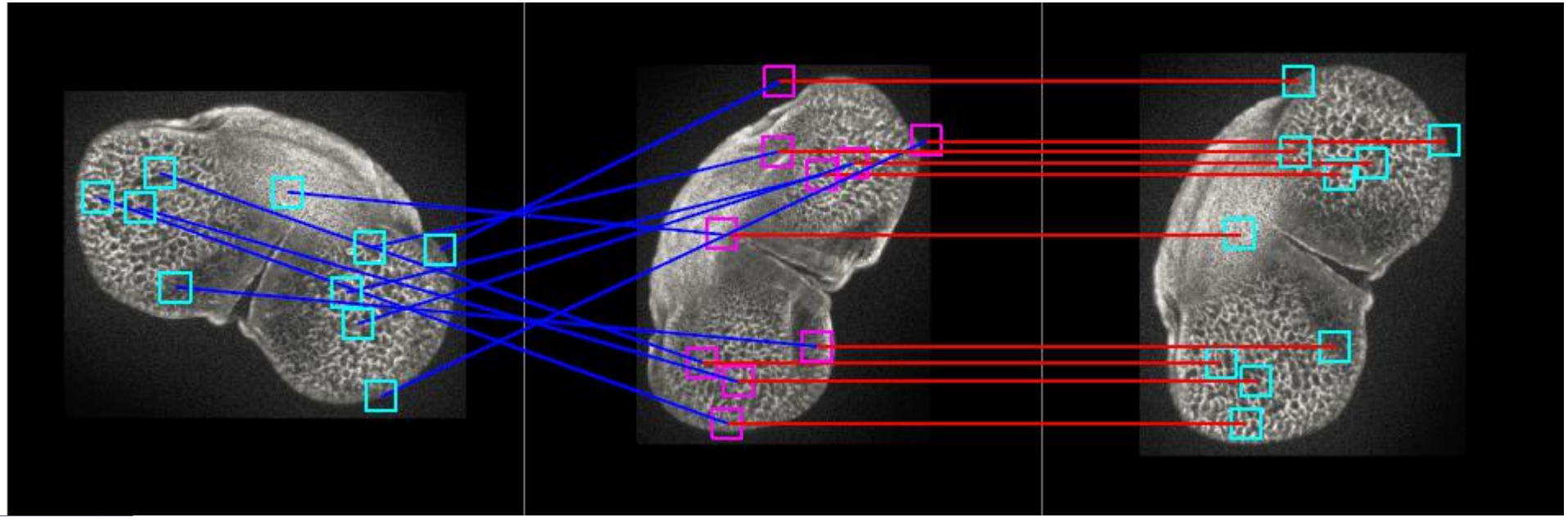
Therefore, it's better to match local patches with geometric constraints.

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

our patch-match based method

The patch-match method needs images to be alligned





in-plate rotation viewpoint calibration

perform *k*-medoids clustering on an affinity graph of training set,

in-plate rotation viewpoint calibration

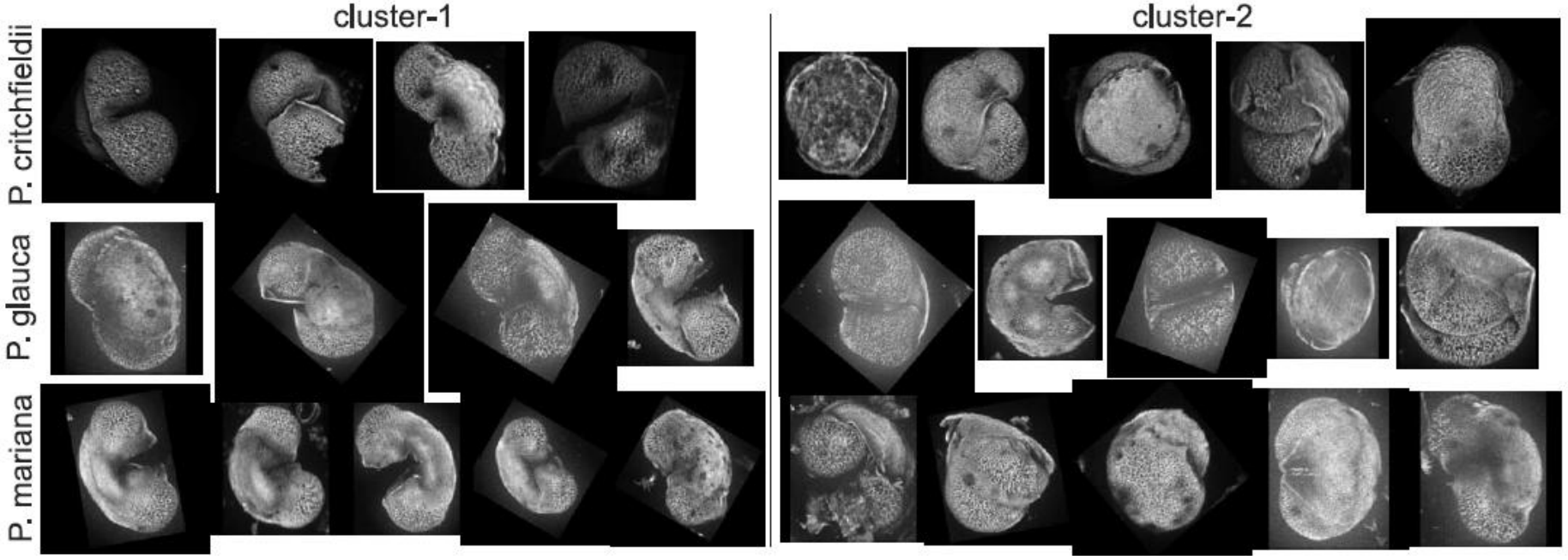
perform *k*-medoids clustering on an affinity graph of training set,

where pairwise similarity is based on Euclidean distance of pollen grain silhouette

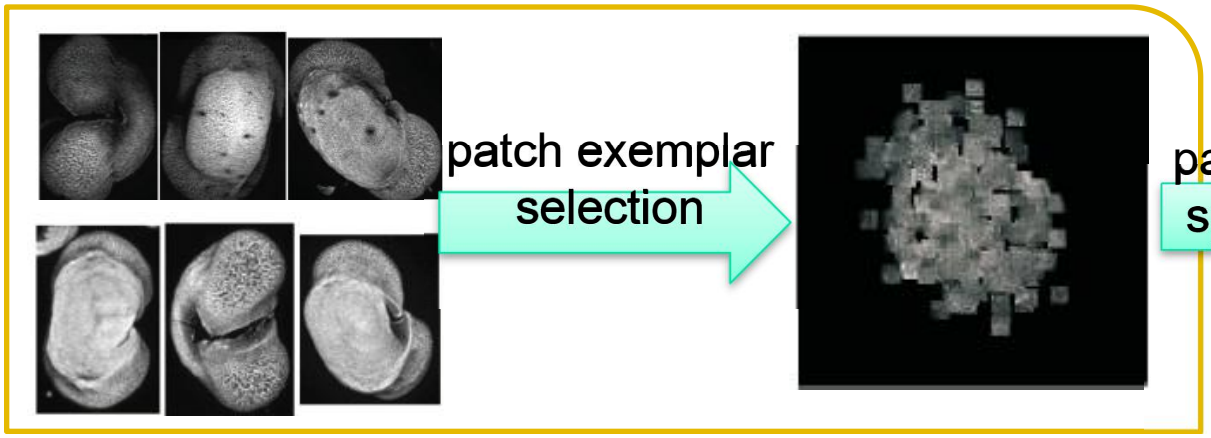
in-plate rotation viewpoint calibration

perform *k*-medoids clustering on an affinity graph of training set,

where pairwise similarity is based on Euclidean distance of pollen grain silhouette



our patch-match based method

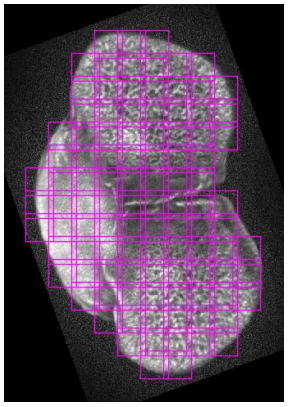


patch exemplar selection

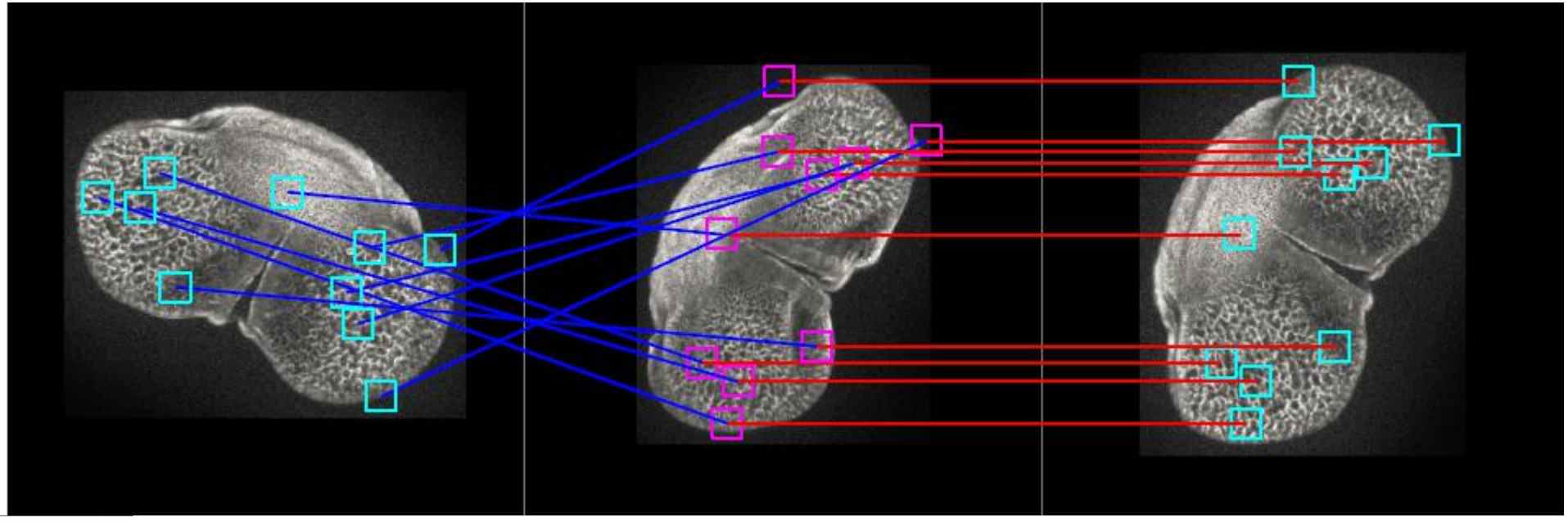
training stage

patch match by sparse coding

SVM

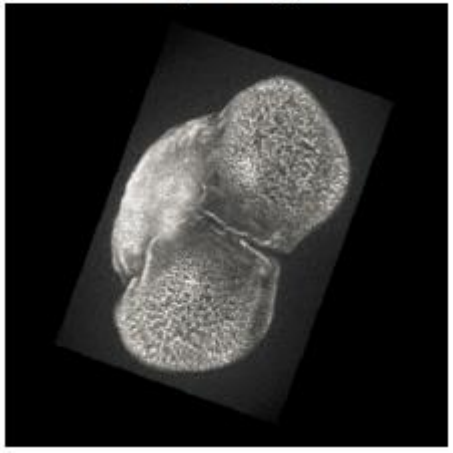


testing stage

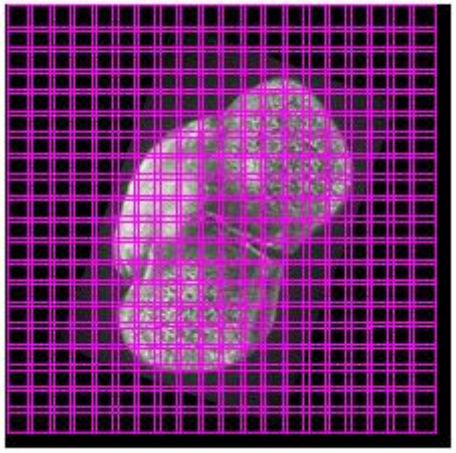


discriminative patch selection

original image

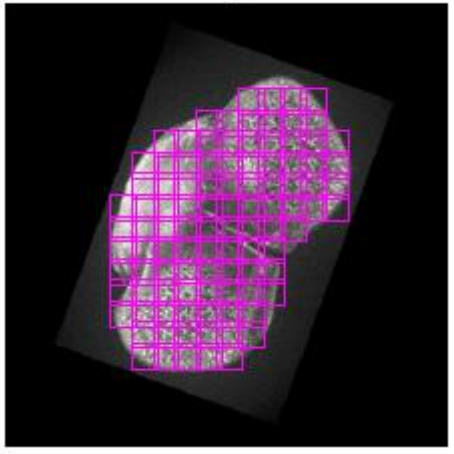


dense patches



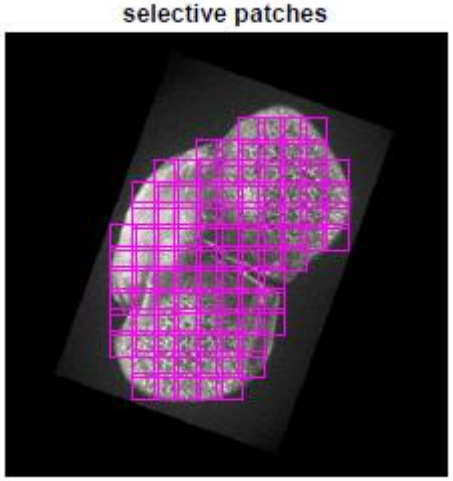
shape mask

selective patches



discriminative patch selection

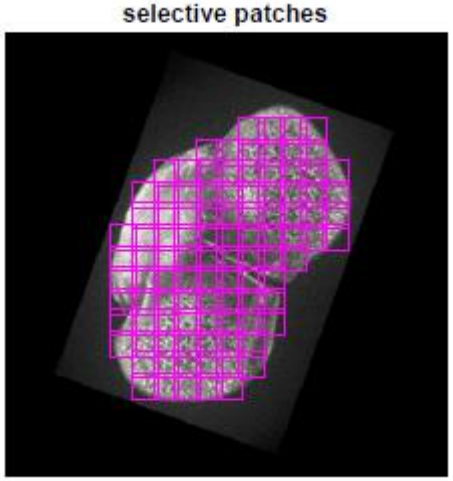
From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have



discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

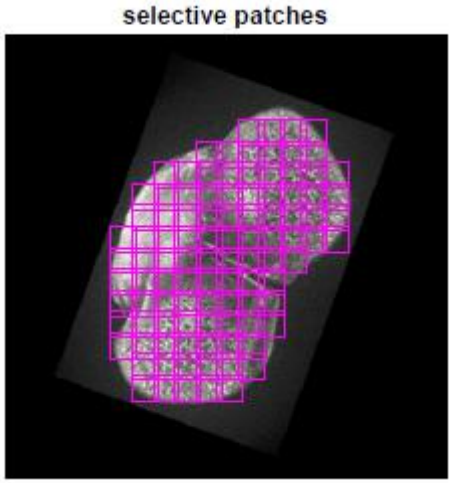
1. representative in feature space



discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

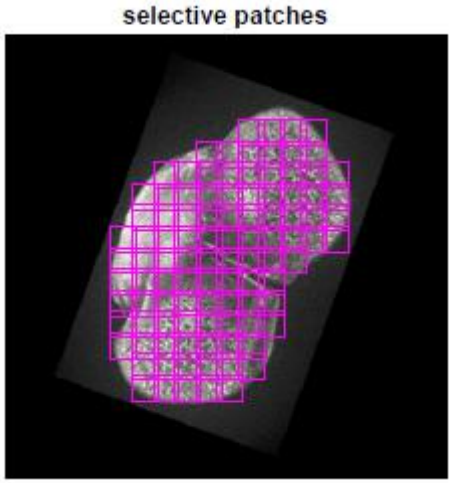
1. representative in feature space
2. spatially distributed in input space



discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

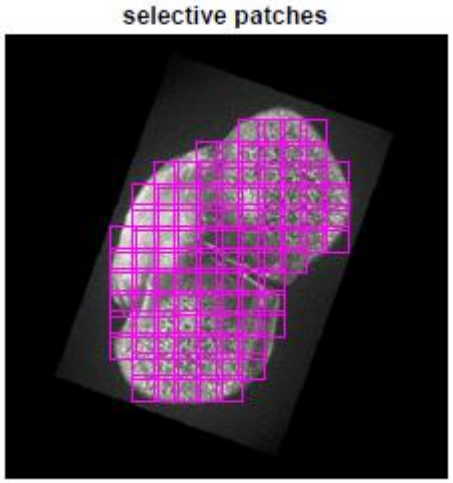
1. representative in feature space
2. spatially distributed in input space
3. discriminative



discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

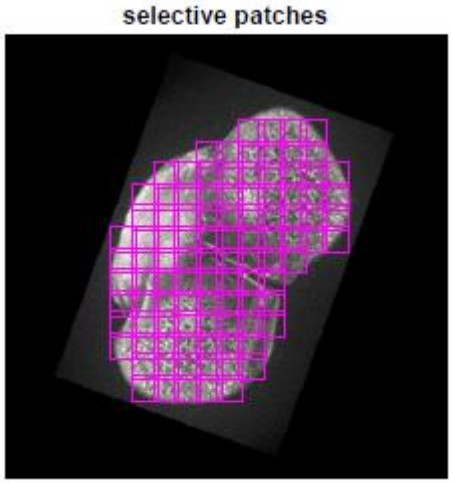
1. representative in feature space
2. spatially distributed in input space
3. discriminative
4. class balance



discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

1. representative in feature space
2. spatially distributed in input space
3. discriminative
4. class balance
5. cluster compactness

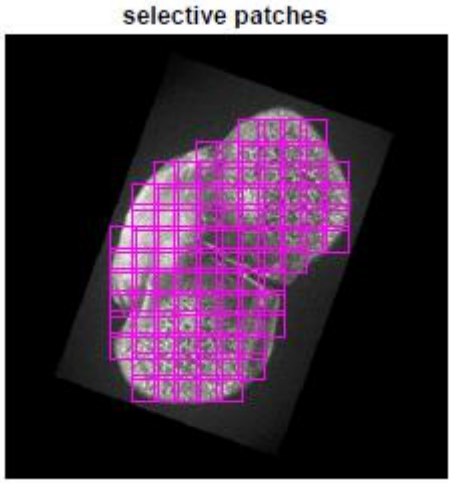


discriminative patch selection

From a finite set of patches, *V*, we'd like to select *M* patches, which should be/have

1. representative in feature space
2. spatially distributed in input space
3. discriminative
4. class balance
5. cluster compactness

We index the selected patches by *A*



example: representational power

representative in feature space

$$\mathcal{F}_R(A) = \sum_{j \in \mathcal{V}} \max_{i \in A} \mathbf{S}_{ij}$$

example: representational power

Maximizing the following set function (NP-hard)

$$\mathcal{F}_R(A) = \sum_{j \in \mathcal{V}} \max_{i \in A} \mathbf{S}_{ij}$$

example: representational power

Maximizing the following set function (NP-hard)

$$\mathcal{F}_R(A) = \sum_{j \in \mathcal{V}} \max_{i \in A} \mathbf{S}_{ij}$$

facility location problem -- optimally placing sensors to monitor temperature

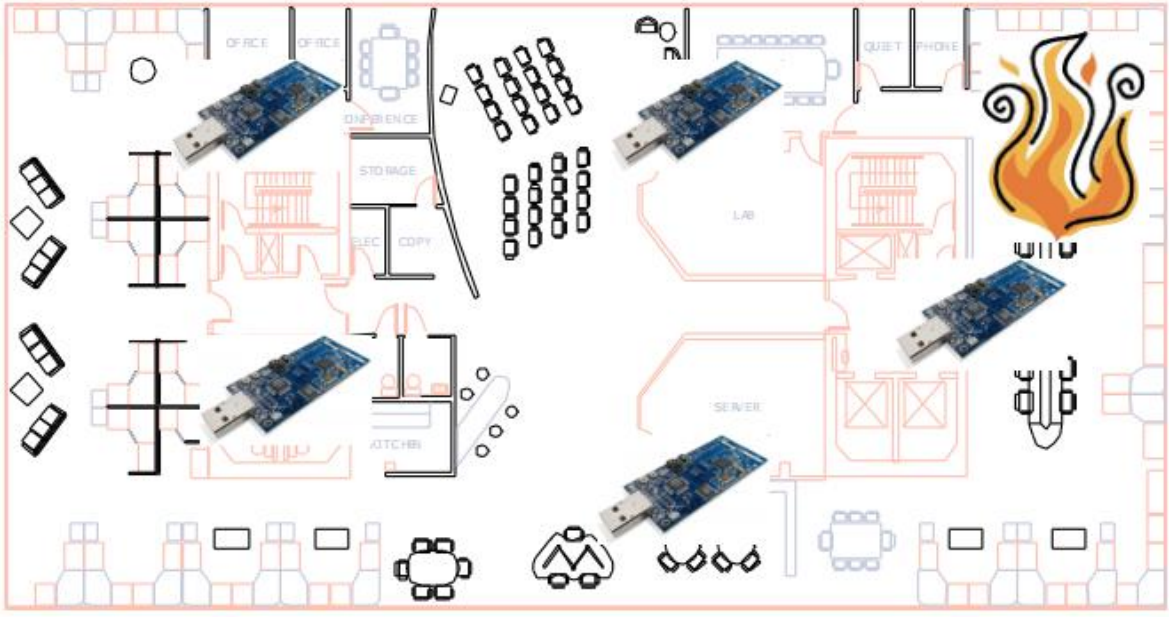


photo credited by Andreas Krause

example: representational power

Maximizing the following set function (NP-hard)

$$\mathcal{F}_R(A) = \sum_{j \in \mathcal{V}} \max_{i \in A} \mathbf{S}_{ij}$$

we can obtain a near optimal solution to this submodular function with a greedy algorithm

photo credited by Andreas Krause

selected discrminative patches

Identification by patch-match sparse coding

1. Automatic patch exemplar selection (dictionary learning)

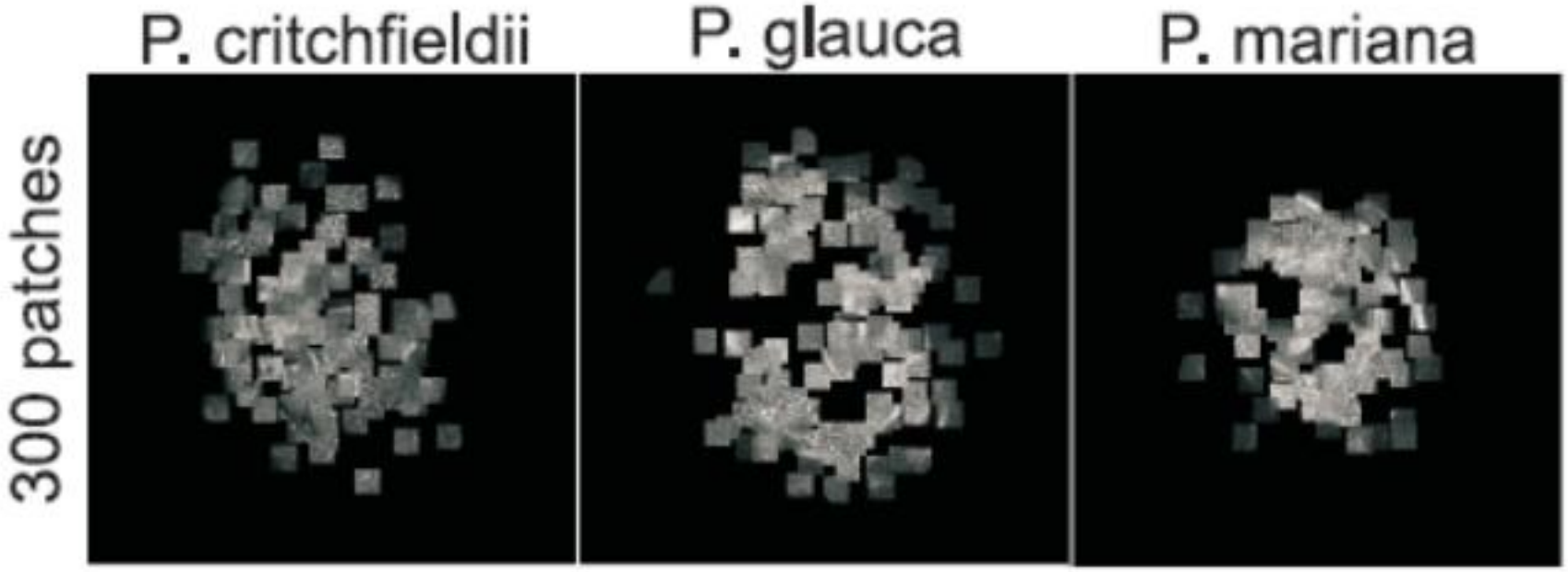
based on discriminative and generative criteria

selected discrminative patches

Identification by patch-match sparse coding

1. Automatic patch exemplar selection (dictionary learning)

based on discriminative and generative criteria



Automatically selected patches

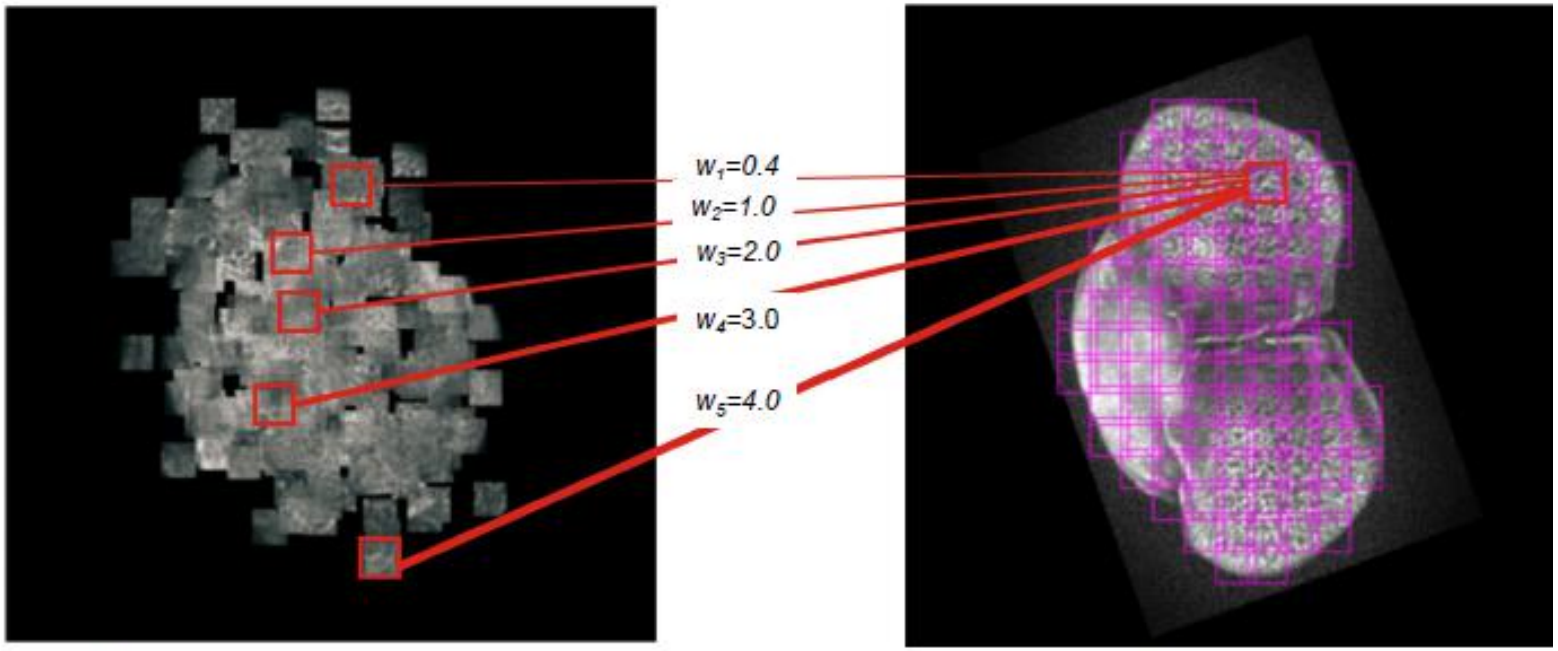
patch-match for identification

Identification by patch-match sparse coding

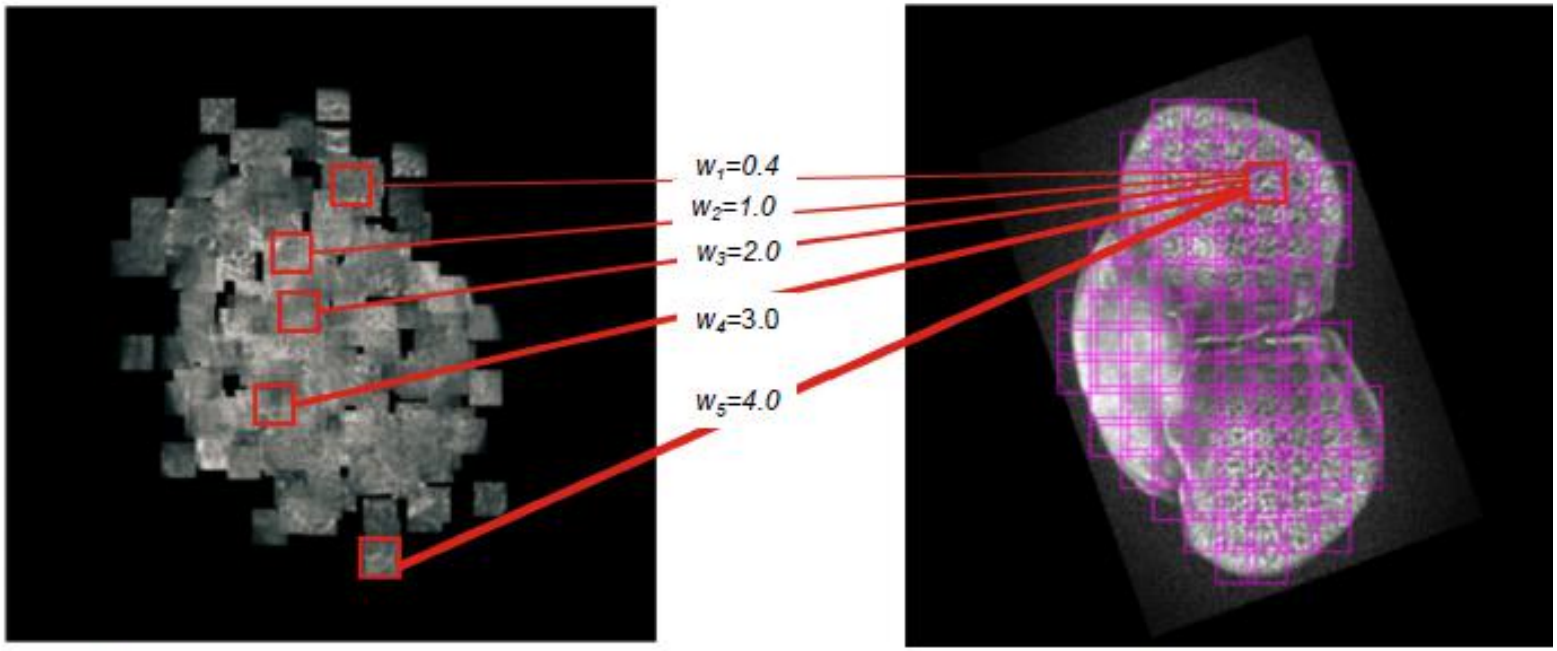
1. Automatic patch exemplar selection (dictionary learning)

2. Spatially-aware sparse coding (SACO)

   - penalize dictionary elements from distant spatial locations



spatially aware coding (SACO)



$$\underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_1$$

Spatial weights

Exemplar patches (dictionary)

Test patch

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

$$\underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_1$$

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

$$\underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_1$$

$$\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 \quad \Rightarrow \quad \|\mathbf{\Omega}\mathbf{x} - \mathbf{a}\|_2^2$$

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

$$\underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_1$$

$$\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 \quad \Rightarrow \quad \|\mathbf{\Omega}\mathbf{x} - \mathbf{a}\|_2^2$$

SACO-I

$$\mathbf{\Omega} \equiv (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T$$

$$\mathbf{u} = \mathbf{\Omega}\mathbf{x}$$

$$a_i^* = \operatorname{sgn}(u_i) \cdot \max(0, |u_i| - \lambda_1 w_i)$$

$$\mathbf{a}^* = [a_1^*, \ldots, a_i^*, \ldots, a_m^*]^T$$

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

$$\underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_2 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_2^2 + \lambda_1 \|\mathbf{a}\|_1$$

$$\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 \quad \Rightarrow \quad \|\mathbf{\Omega}\mathbf{x} - \mathbf{a}\|_2^2$$

SACO-II

$$\begin{aligned}
\mathbf{\Omega} &\equiv (\mathbf{D}^T\mathbf{D} + \lambda_2 \operatorname{diag}(\mathbf{w})^2)^{-1}\mathbf{D}^T \\
\mathbf{u} &= \mathbf{\Omega}\mathbf{x} \\
a_i^* &= \operatorname{sgn}(u_i) \cdot \max(0, |u_i| - \lambda_1) \\
\mathbf{a}^* &= [a_1^*, \ldots, a_i^*, \ldots, a_m^*]^T.
\end{aligned}$$

Quantitative Result on Fossil Pollen

Represent patch using CNN feature extractor (VGG19)

Global average pooling of sparse codes by SACO

linear SVM

SRC	VGG19+SVM	FV+SVM	SACO-I	SACO-II
62.04	65.11	61.46	83.21	86.13

Table 1. Statistics of our fossil pollen grain dataset.

	#train	#test	#total
P. critchfieldii	65	43	108
P. glauca	65	355	420
P. mariana	65	287	352
Summary	195	685	880

Substantially outperforms standard CNN and Fisher-vector based approaches!

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

quantitative result on modern pollen

We apply our approach to modern pollen grain identification.

Our method		*Actual*	
		P. Glauca	*P. Mariana*
Predicted	*P. Glauca*	0.969	0.030
	P. Mariana	0.021	0.980

	Actual	
	P. mariana	*P. glauca*
P. mariana	**0.920**	0.005
P. glauca	0.061	**0.893**

UCIrvine
University of California, Irvine

Surangi W Punyasena, David K Tcheng, Cassandra Wesseln, Pietra G Mueller, Classifying black and white spruce pollen using layered machine learning, New Phytologist, 2012

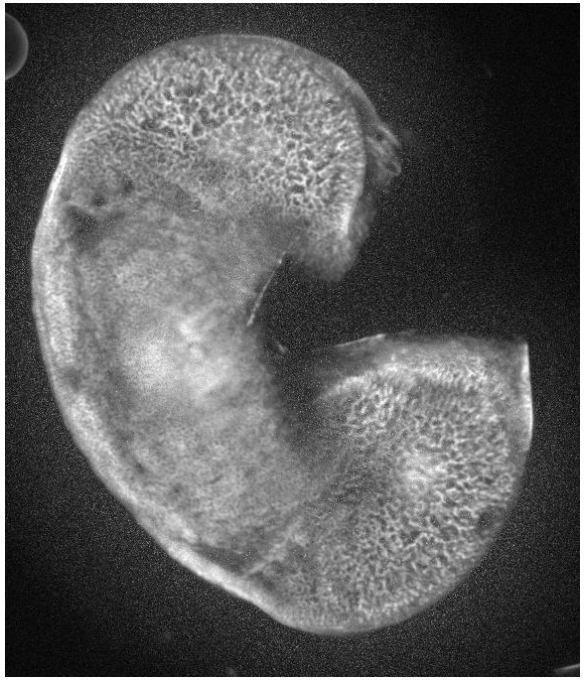
Identifying Fossil Pollen with Modern Reference

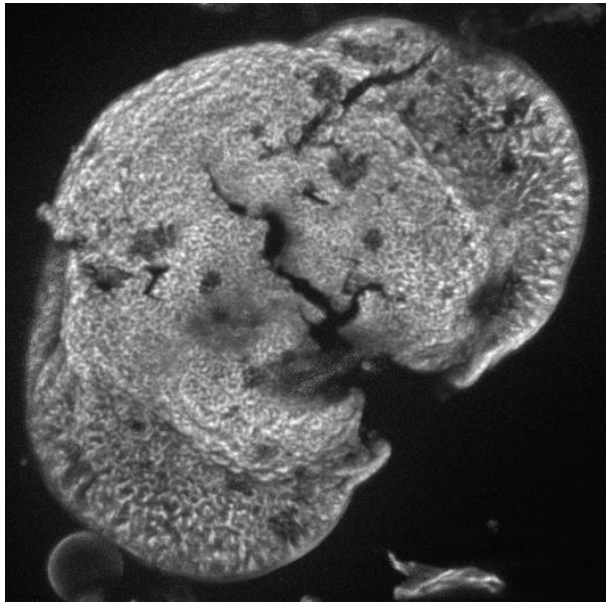
Fossil pollen grains are degraded over time.

using patches from modern pollen reference to identify fossilized ones

modern pollen grain from glauca

fossil pollen pollen grain from glauca



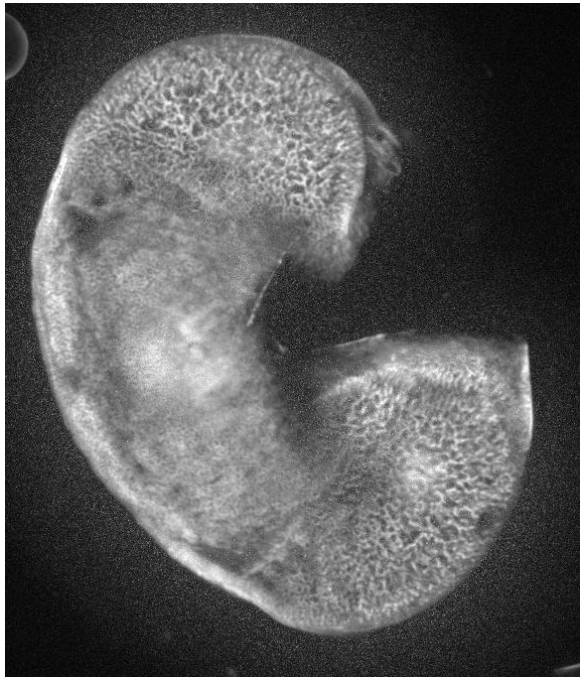


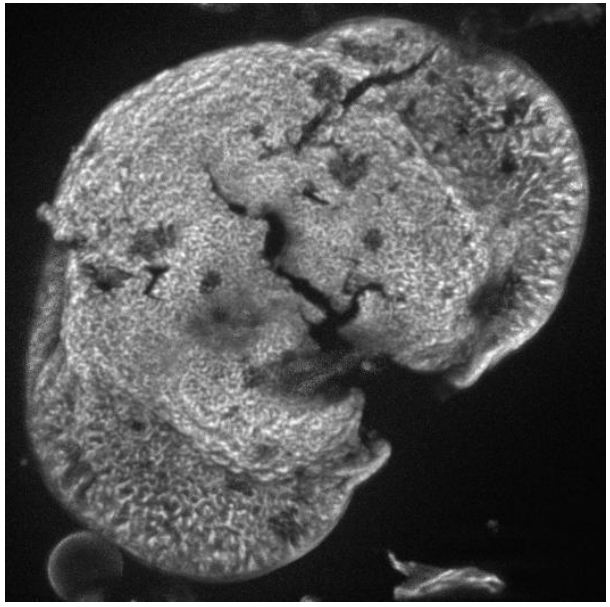
Identifying Fossil Pollen with Modern Reference

- Use our method to select patches from modern pollen grains
- Use the selected modern patches to identify fossil ones
- We achieve **69%** accuracy wrt expert labels.

modern pollen grain from glauca

fossil pollen pollen grain from glauca





Outline

1. Problem definition
2. Instantiation
3. Challenge and philosophy
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Thank you

Question & Answer

Thank you

Question & Answer

Thank you

Question & Answer

Thank you

Question & Answer