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 and philosophy
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)

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 computer vision

- distinguish subordinate categories within an entry-level category

Problem definition

Fine-grained computer vision

- distinguish subordinate categories within an entry-level category
- tasks are like classification, segmentation, specific case studies, etc.

Instantiation -- classification

previously, generic classification -- car vs. bird

Shu Kong, Charless Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Instantiation -- classification

now, fine-grained car model classification

UCIrvine
University of California, Irvine

Shu Kong, Charless 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

Shu Kong, Charless Fowlkes, "Low-rank Bilinear Pooling for Fine-Grained Classification", arXiv:1611.05109, 2016

Instantiation -- identification

previously, in phytology, identifying by eye

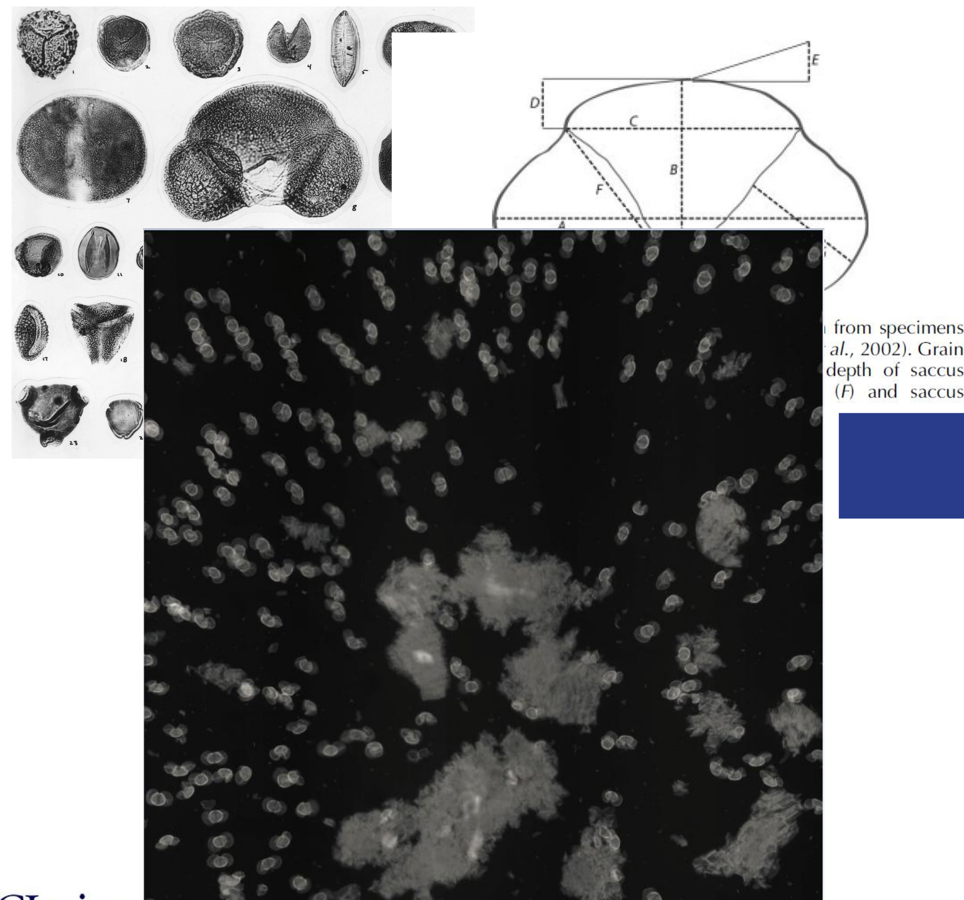


image from Surangi W. Punyasena

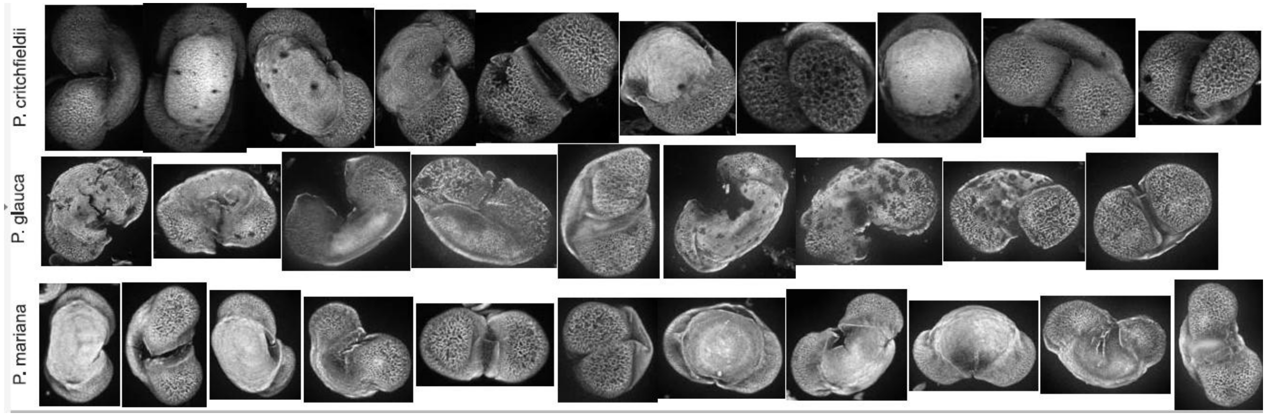
S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

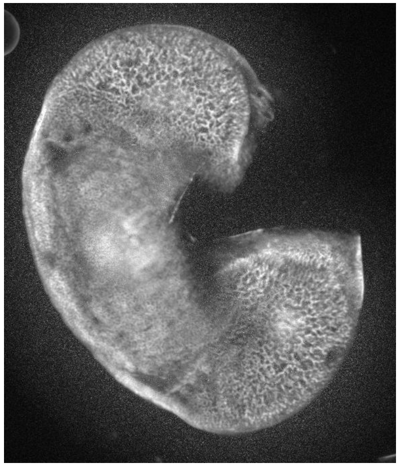
UCIrvine
University of California, Irvine

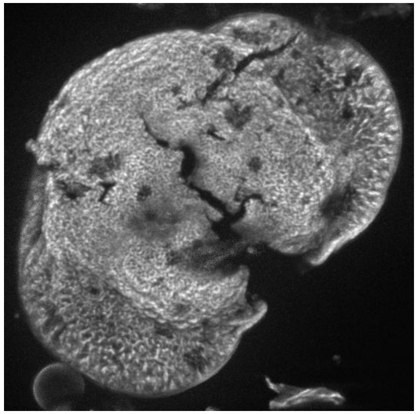
Instantiation -- identification

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

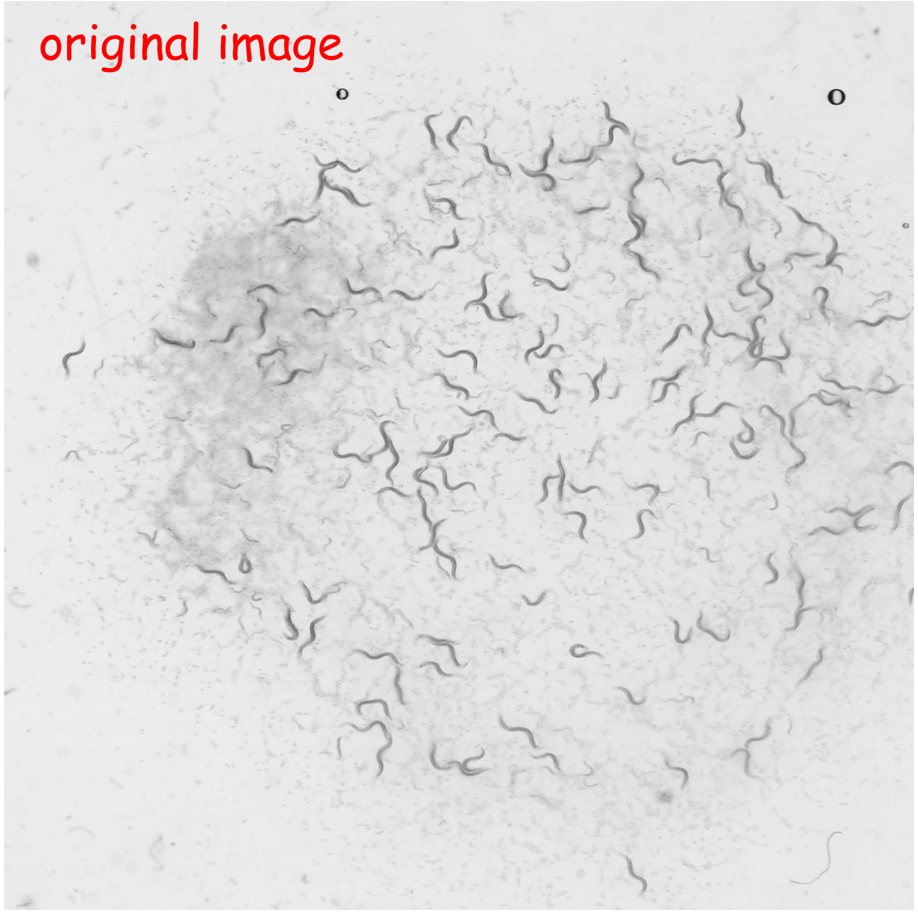
UCIrvine
University of California, Irvine

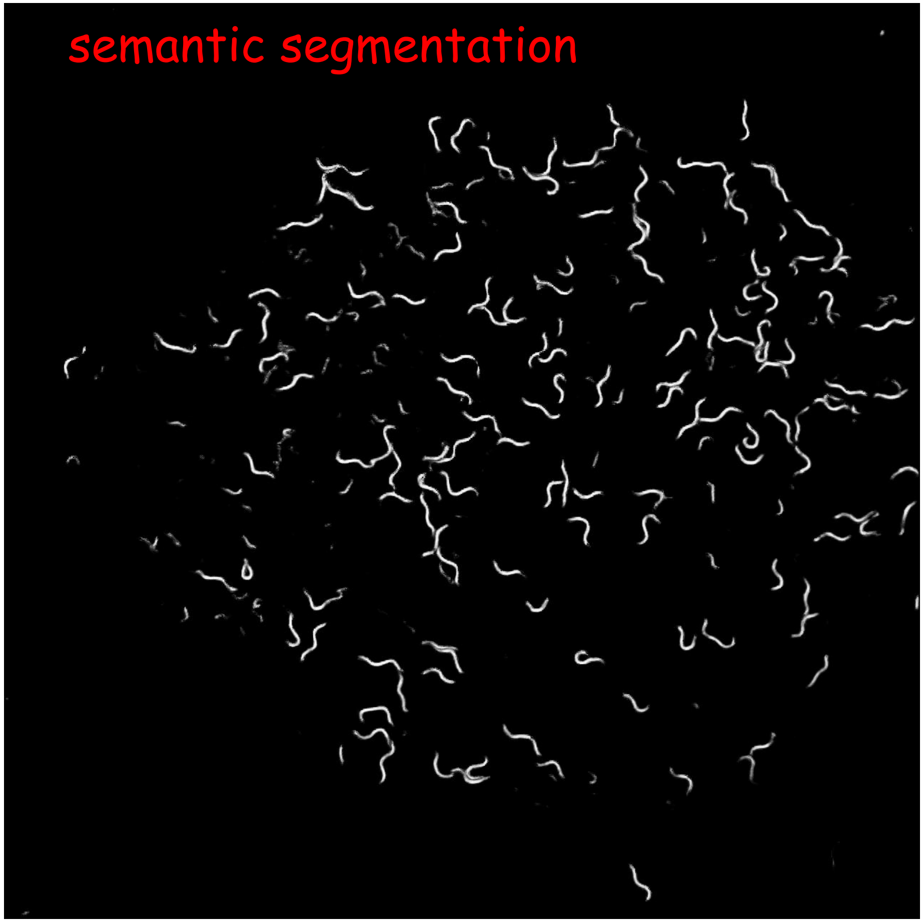
S. Kong, S. Punyasena, C. Fowlkes, "Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification", CVPR CVMI, 2016

Instantiation -- segmentation

previously, in biology, semantic segmentation

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





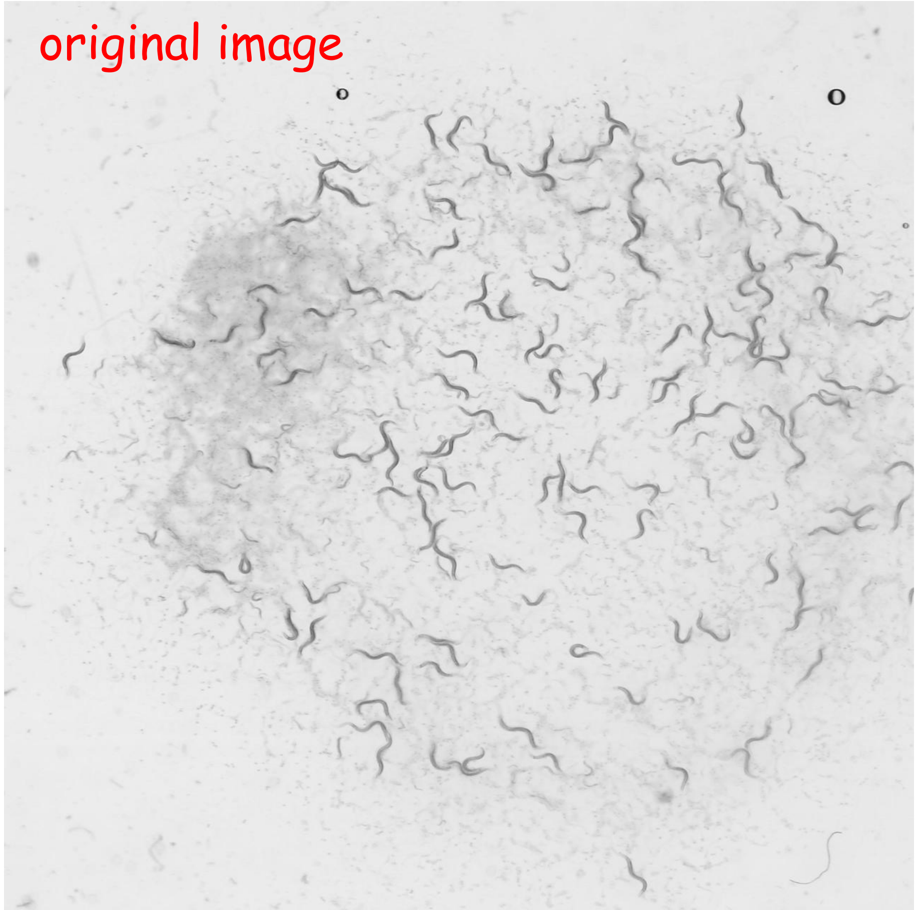
UCIrvine
University of California, Irvine

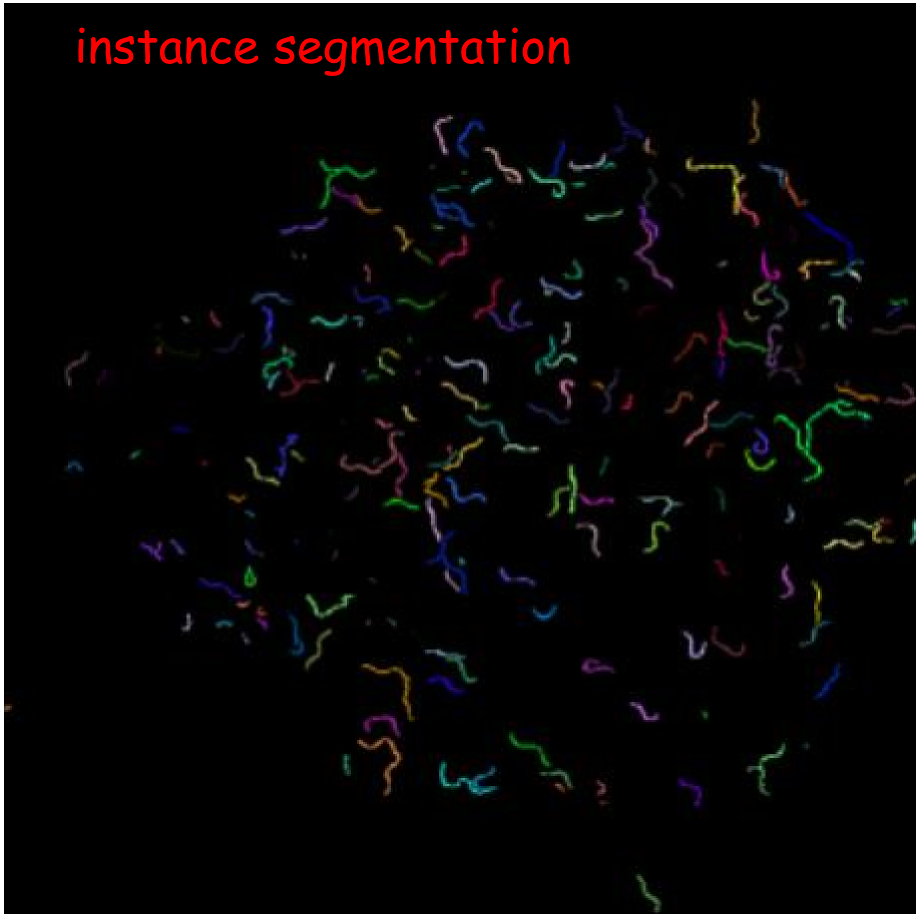
S. Kong, "Automated Biological Image Analysis using Computer Vision and Machine Learning", Janelia workshop, 2016

Instantiation -- segmentation

now, instance segmentation

enabling study of worm population



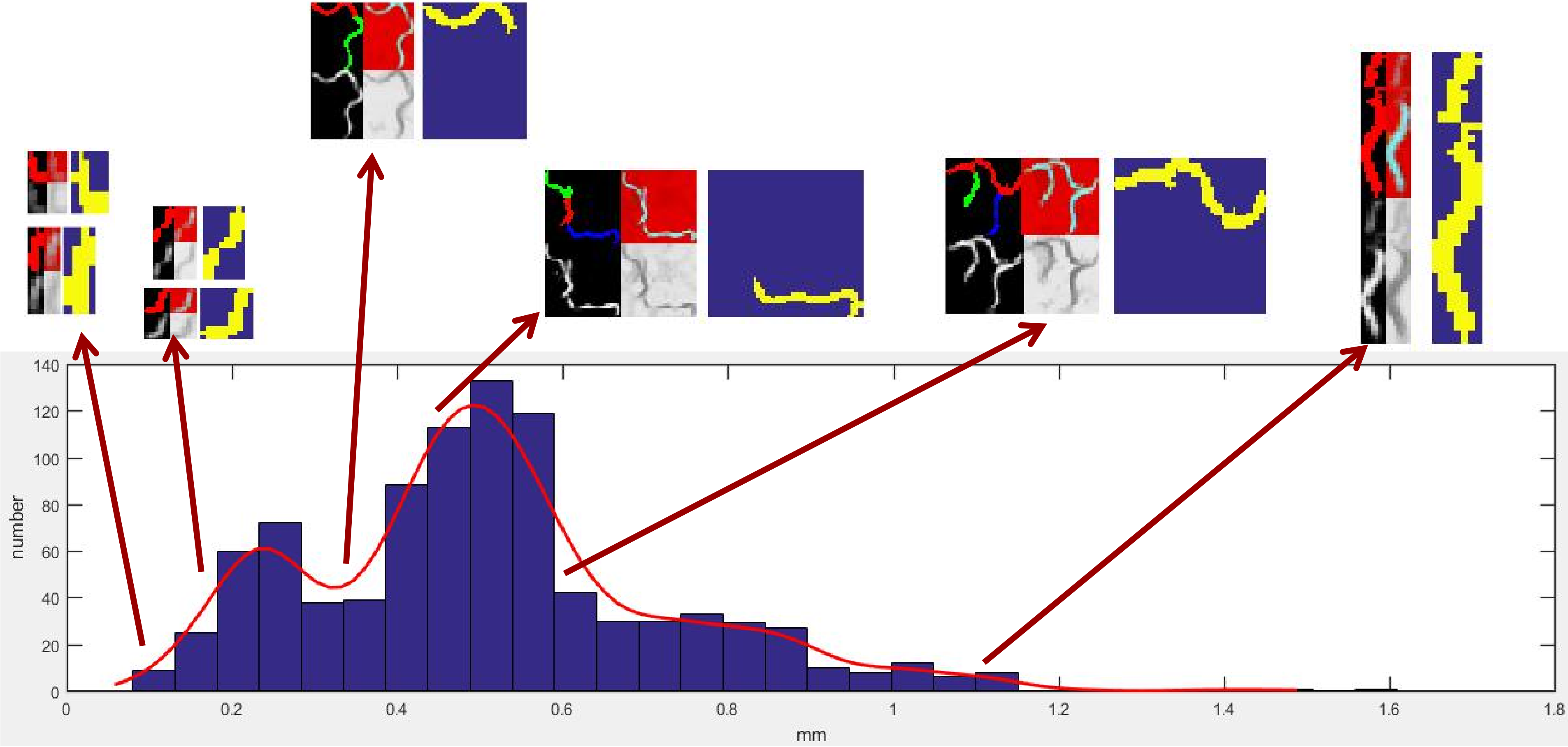


S. Kong, "Automated Biological Image Analysis using Computer Vision and Machine Learning", Janelia workshop, 2016

Instantiation -- segmentation

now, instance segmentation

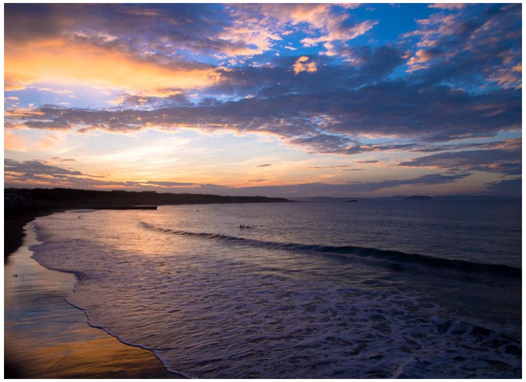
enabling study of worm population



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



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

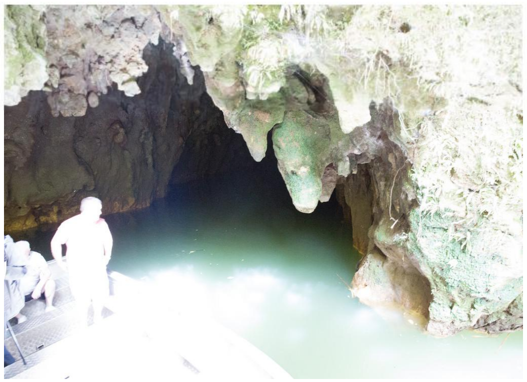
previously, modeling image aesthetics study as binary classification, low- vs. high- aesthetic

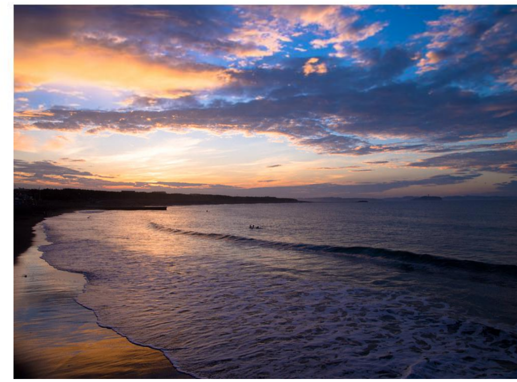
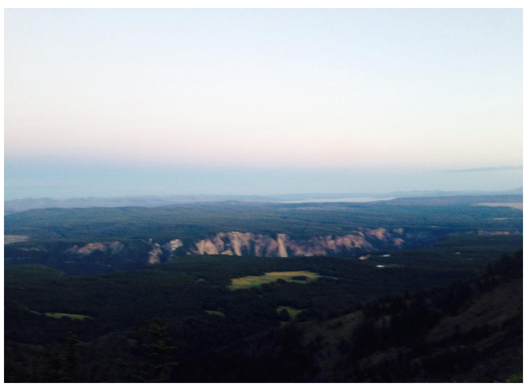
UCIrvine
University of California, Irvine

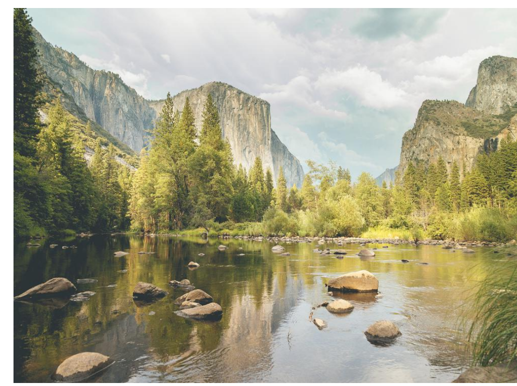
S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation", ECCV, 2016

Instantiation -- photo aesthetic ranking

now, fine-grained ranking for personal photo album management



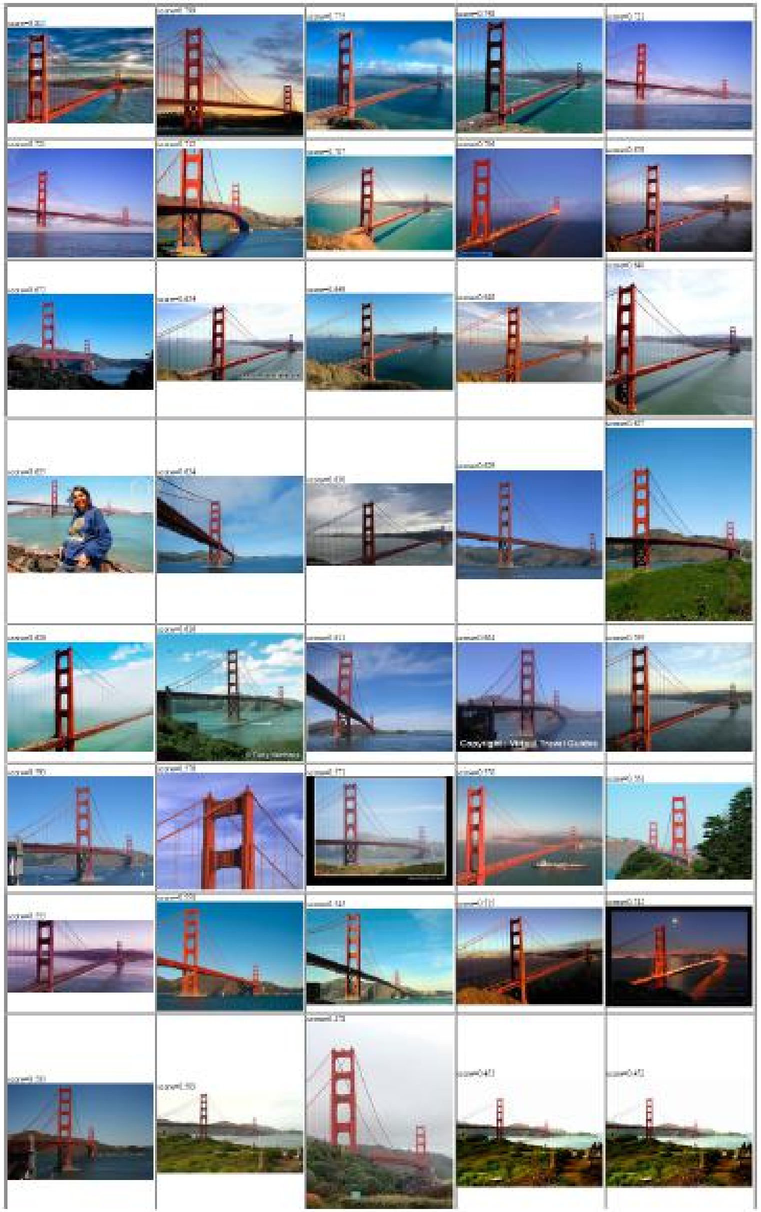


UCIrvine
University of California, Irvine

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



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 and philosophy
4. Fine-grained classification with holistic representation
5. Fine-grained identification by matching local patches
6. Future work and conclusion

Challenge and philosophy

- **lack of training data**
  - costly data collection and annotation

Challenge and philosophy

- lack of training data
  - costly data collection and annotation
- **large numbers of categories**

Challenge and philosophy

- lack of training data
  - costly data collection and annotation
- **large numbers of categories**
  - >14,000 birds
  - >278,000 butterfly&moth
  - >941,000 insects

Challenge and philosophy

- **lack of training data**
  - costly data collection and annotation
- **large numbers of categories**
- **high intra-class vs. low inter-class variance**

Challenge and philosophy

- **lack of training data**
  - costly data collection and annotation
- **large numbers of categories**
- **high intra-class vs. low inter-class variance**

Caspian Tern

Caspian Tern

Elegant Tern

Challenge and philosophy

- **lack of training data**
  - costly data collection and annotation
- **large numbers of categories**
- **high intra-class vs. low inter-class variance**
- **philosophy**
  - **finding discriminative parts, and matching them effectively**

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

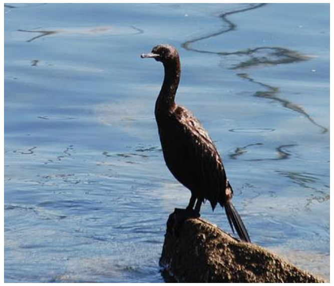
Red_Winged_Blackbird

Brandt_Cormorant

Acadian_Flycatcher

Yellow_Headed_Blackbird

Pelagic_Cormorant



Yellow_Billed_Cuckoo

Holistic representation based method

recognizing bird species by seeing the photo

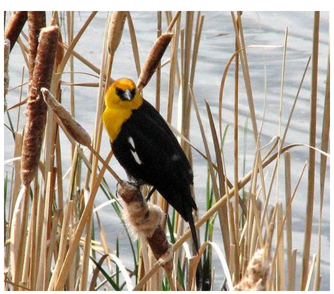
In literature, detecting keypoint/parts and stacking them as holistic representation

Red_Winged_Blackbird

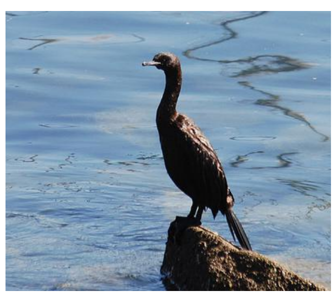
Brandt_Cormorant

Acadian_Flycatcher

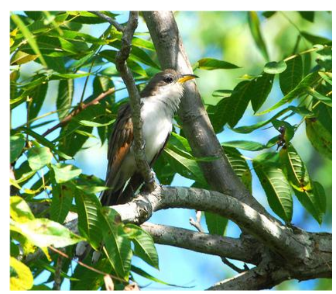
Yellow_Headed_Blackbird

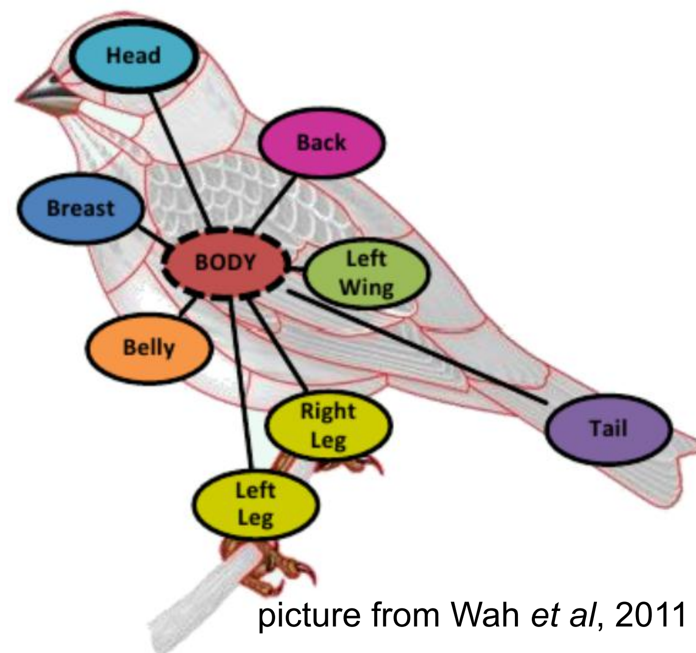


Pelagic_Cormorant



Yellow_Billed_Cuckoo

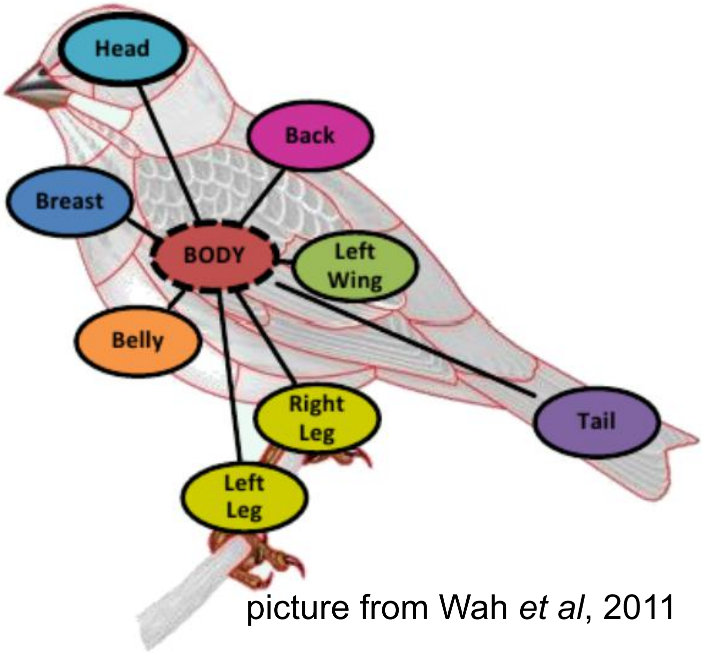




picture from Wah *et al*, 2011

Holistic representation based method

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



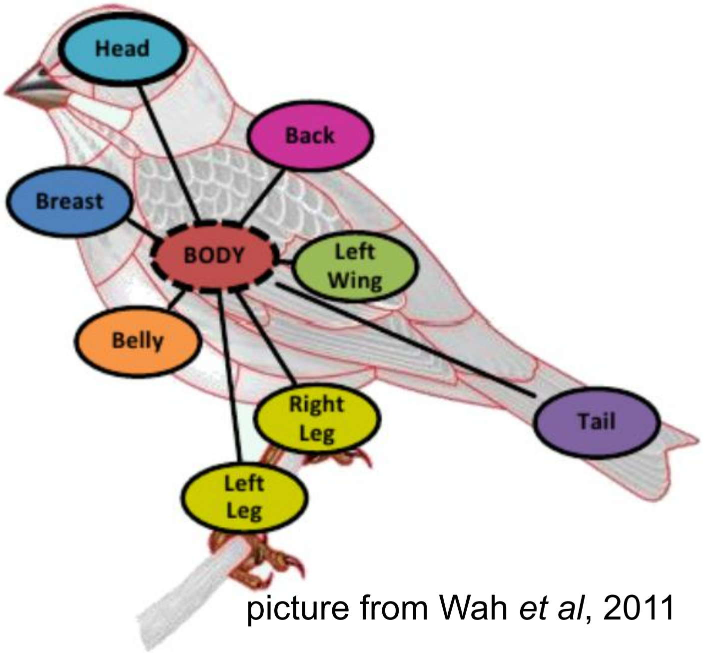
picture from Wah *et al*, 2011

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/masks.

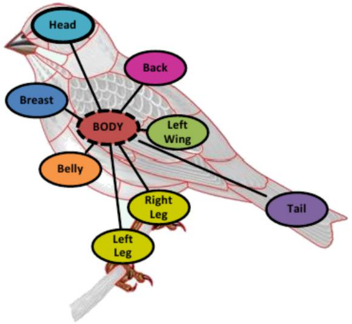


picture from Wah *et al*, 2011

UCIrvine
University of California, Irvine

Holistic representation based method

One method for this is called bilinear pooling

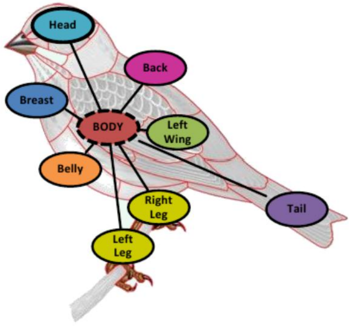


Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

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



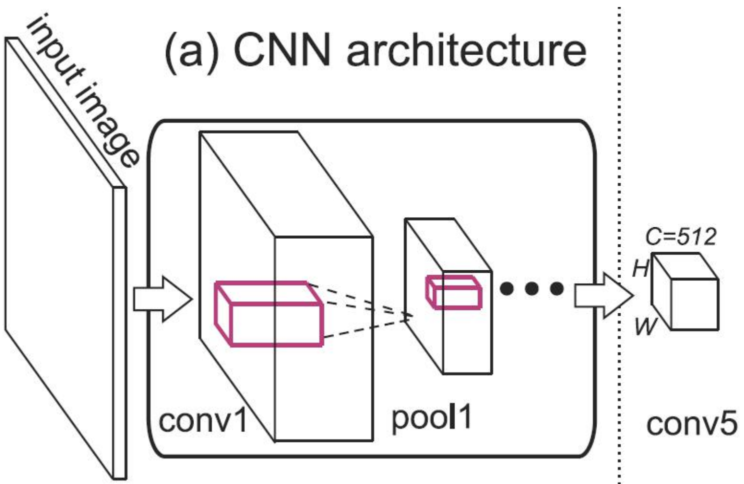
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

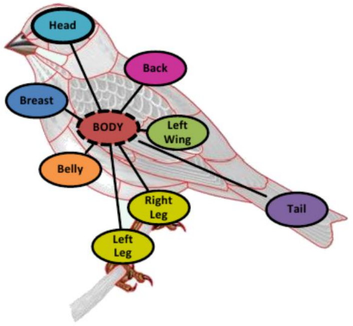
Holistic representation based method

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 hidden layers of a convolutional neural network (CNN)





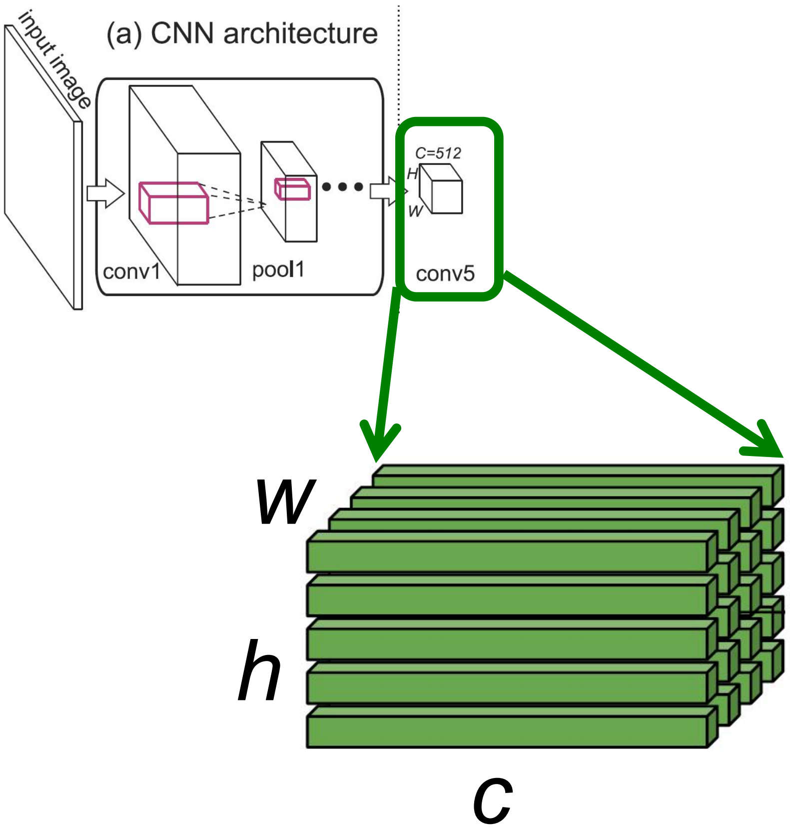
UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

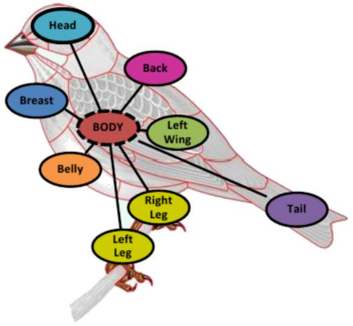
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

Holistic representation based method

Bilinear Pooling

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





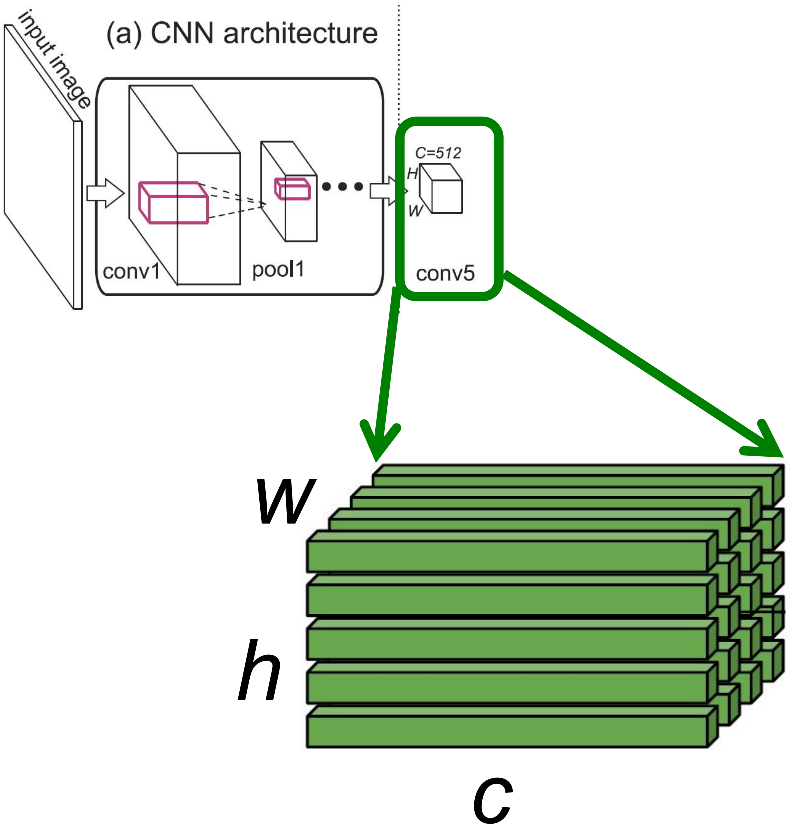
Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

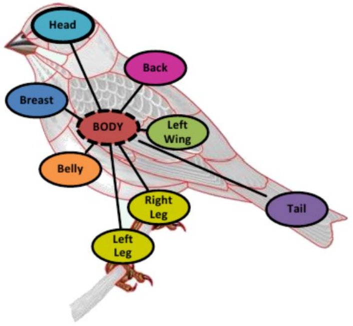
Holistic representation based method

Bilinear Pooling

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

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





UCIrvine
University of California, Irvine

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

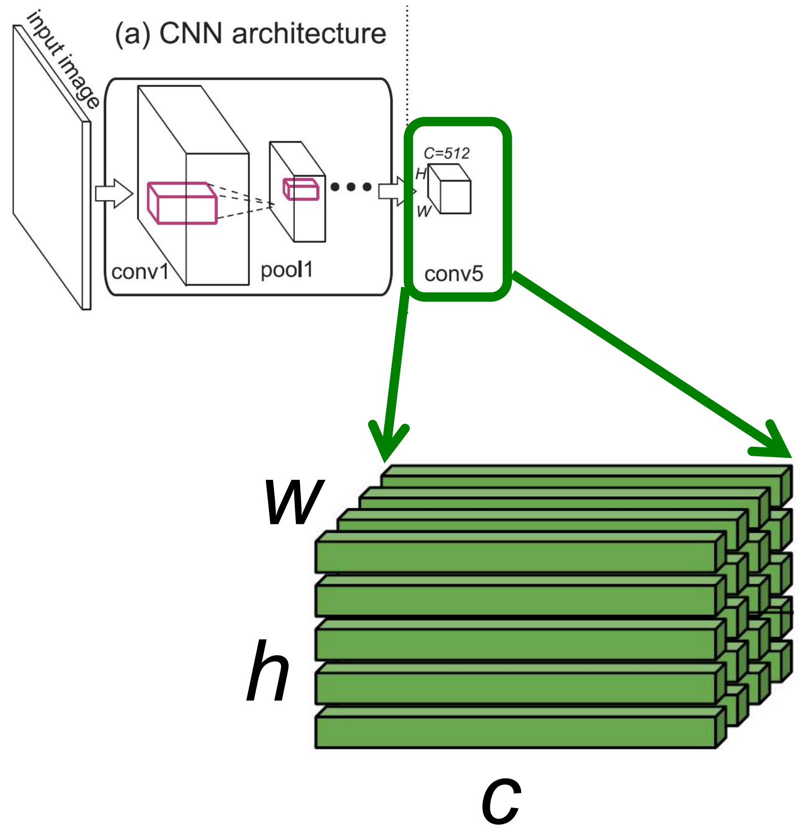
Holistic representation based method

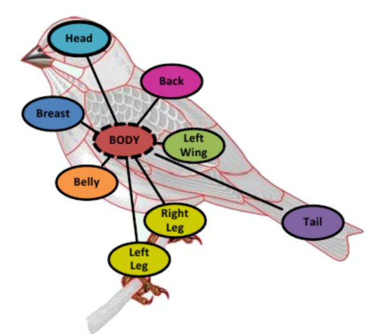
Bilinear Pooling

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

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

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





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

Holistic representation based method

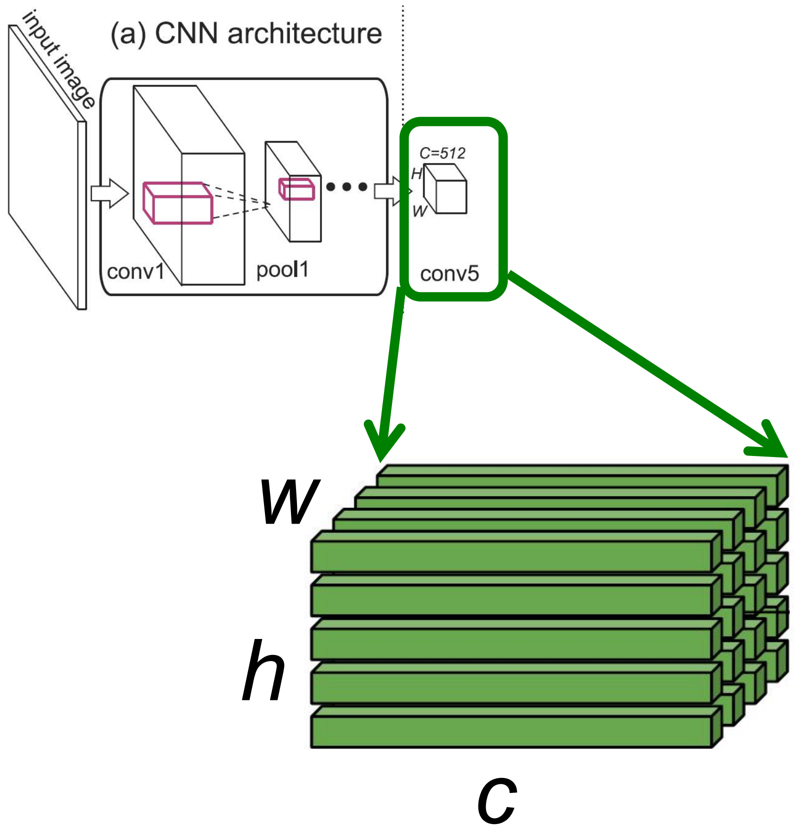
Bilinear Pooling

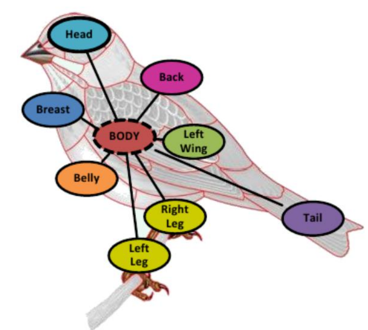
$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

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

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

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





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

Holistic representation based method

Bilinear Pooling

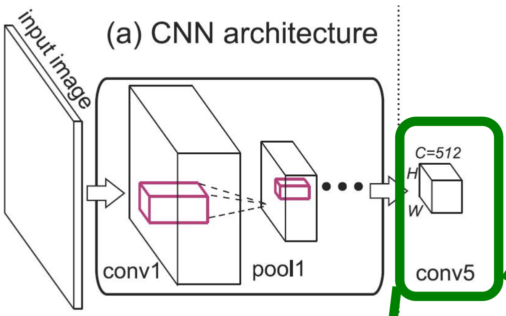
$$\mathcal{X} \in \mathbb{R}^{h \times w \times c}$$

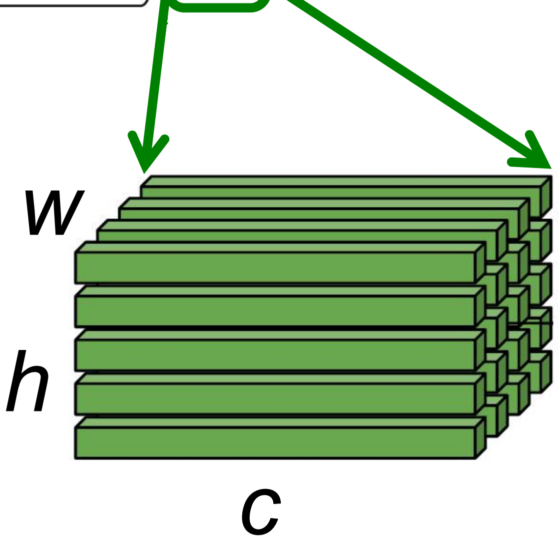
$$\mathbf{x}_i \in \mathbb{R}^c \quad i \in [1, hw]$$

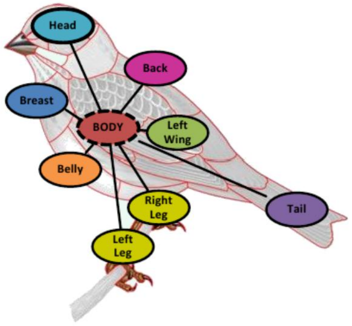
$$\mathbf{X} \in \mathbb{R}^{c \times hw}$$

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

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



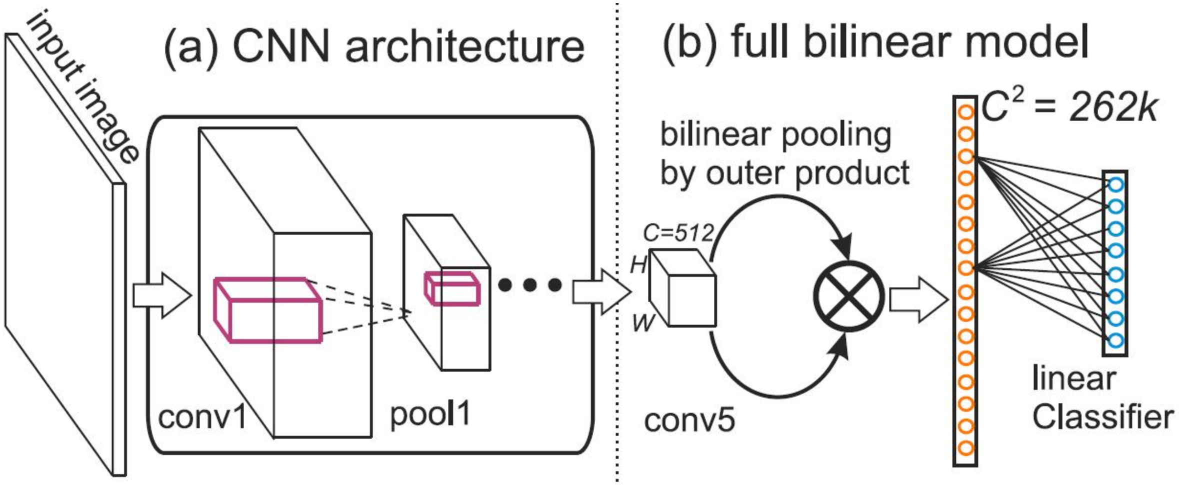


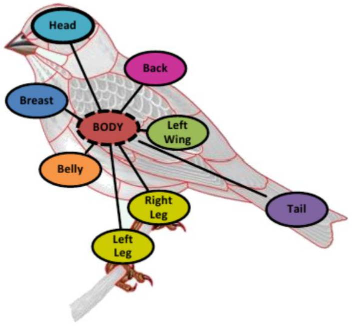


Lin et al., Bilinear CNN models for fine-grained visual recognition, ICCV, 2015

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

Low-rank Bilinear Pooling

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

Holistic representation based method

Low-rank Bilinear Pooling

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

linear SVM

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

Holistic representation based method

Low-rank Bilinear Pooling

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

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

Low-rank Bilinear Pooling

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

linear SVM

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

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

UCIrvine
University of California, Irvine

Holistic representation based method

Low-rank Bilinear Pooling

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

linear SVM

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

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

linear SVM in matrix

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

Holistic representation based method

Low-rank Bilinear Pooling

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

linear SVM

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

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

linear SVM in matrix

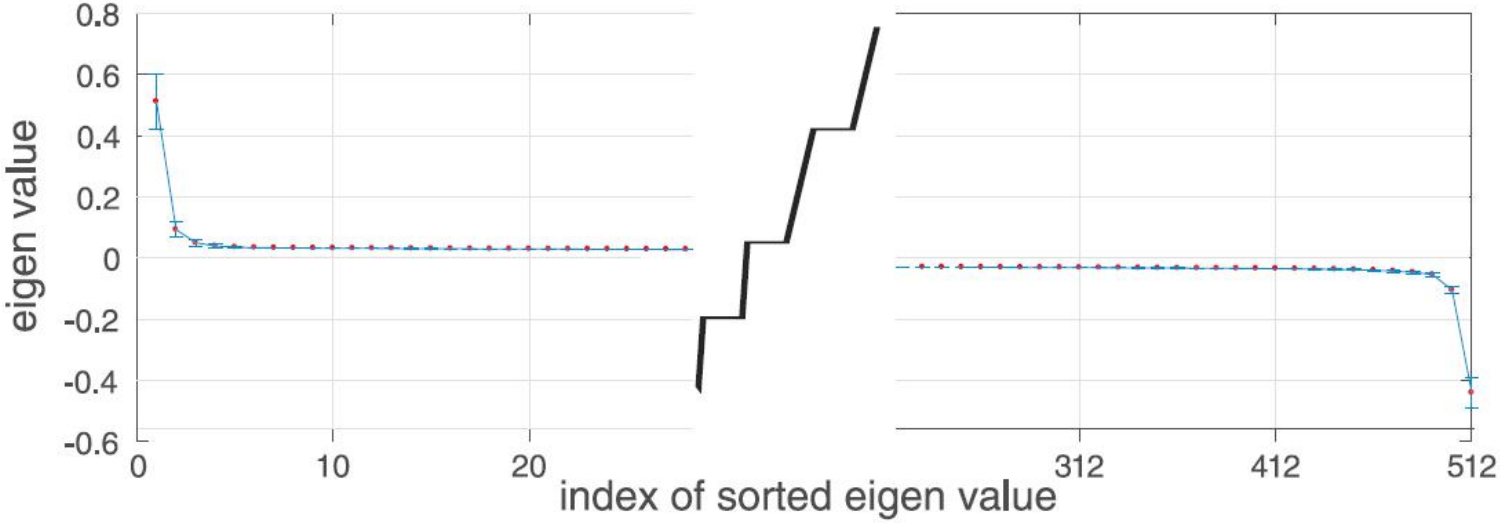
$$\max(0, 1 - y_i \mathrm{tr}(\mathbf{W}^T\mathbf{X}\ \mathbf{X}^T) + b)$$

rank-*r* SVM

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

Holistic representation based method

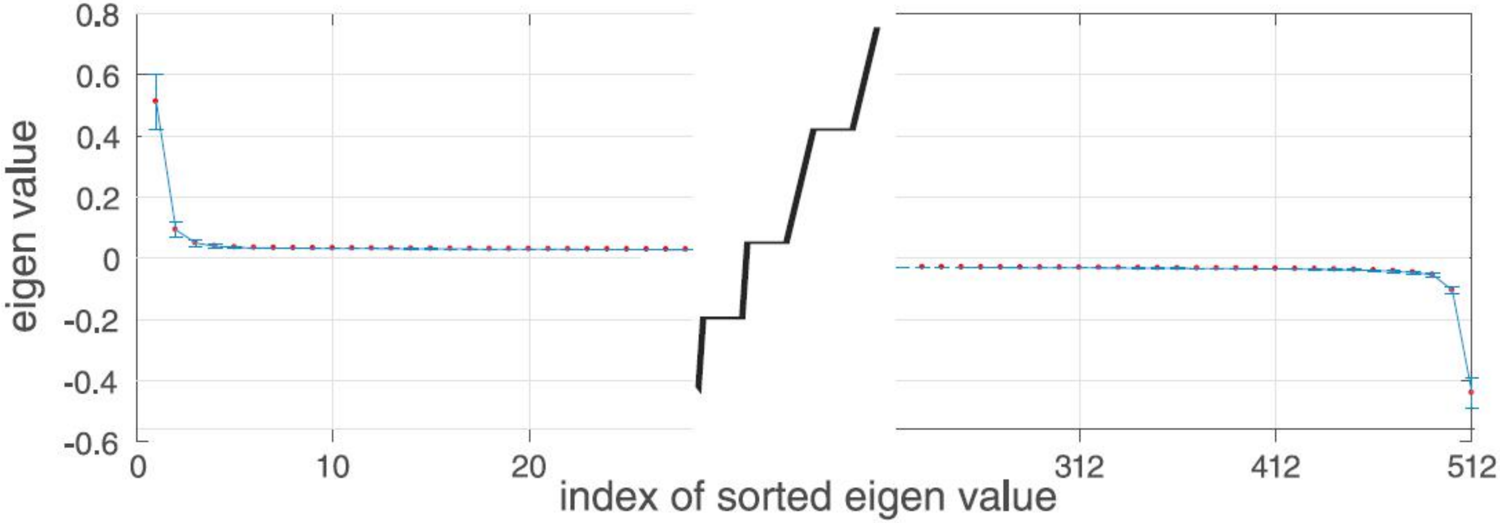
Low-rank SVM

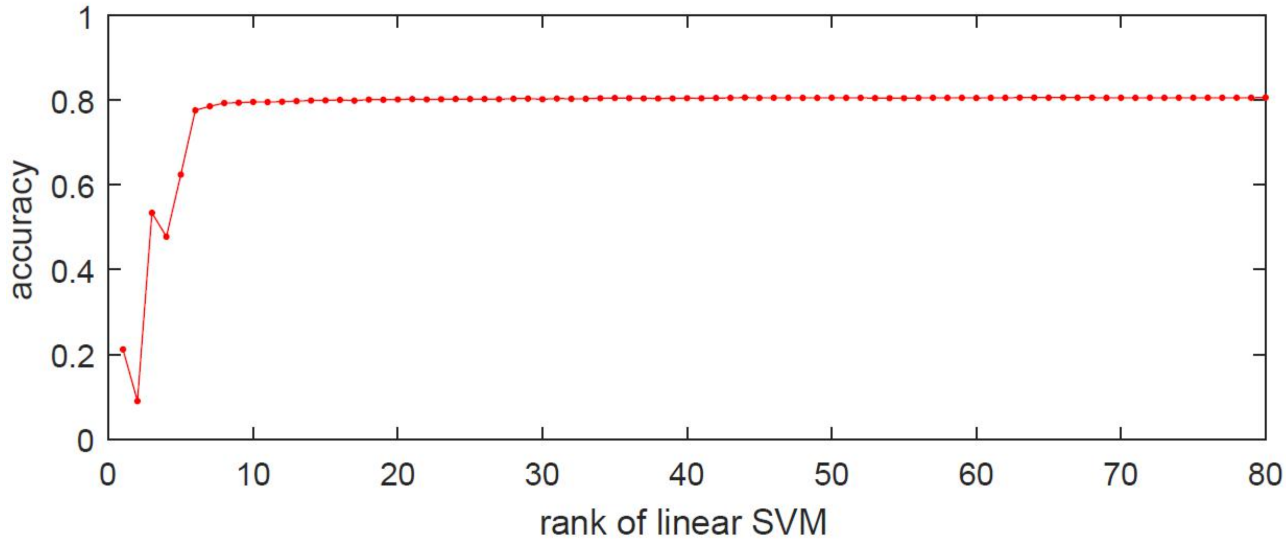


UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

Holistic representation based method

Low-rank SVM





UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

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

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

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

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

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

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

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

$$\text{where} \quad \alpha_i \geq 0, \forall i = 1, \ldots, N$$

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

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

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

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

$$\text{where} \quad \alpha_i \geq 0, \forall i = 1, \ldots, N$$

UCIrvine
University of California, Irvine

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

$$\boxed{\begin{array}{c} \mathbf{w}^T vec(\mathbf{X}\mathbf{X}^T) \Leftrightarrow tr(\mathbf{W}^T \mathbf{X}\mathbf{X}^T) \Leftrightarrow tr(\mathbf{U}\mathbf{U}^T \mathbf{X}\ \mathbf{X}^T) \\ \Updownarrow \\ \|\mathbf{U}^T \mathbf{X}\|_F^2 \Leftrightarrow tr(\mathbf{U}^T \mathbf{X}\mathbf{X}^T \mathbf{U}) \end{array}}$$

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

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

$$\text{where} \quad \alpha_i \geq 0, \forall i = 1, \ldots, N$$

UCIrvine
University of California, Irvine

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

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

$$\Updownarrow$$

$$\|\mathbf{U}^T \mathbf{X}\|_F^2 \Leftrightarrow tr(\mathbf{U}^T \mathbf{X}\mathbf{X}^T \mathbf{U})$$

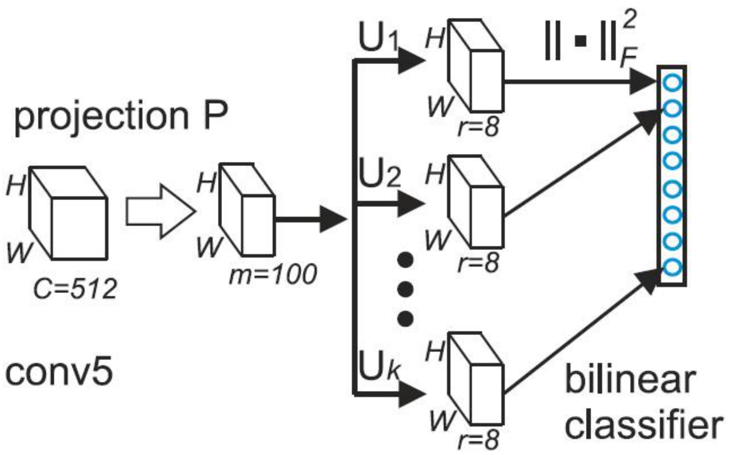
$$\boxed{\begin{array}{l}\max(0, 1 - y_i \{\|\mathbf{U}_+^T \mathbf{X}_i\|_F^2 - \|\mathbf{U}_-^T \mathbf{X}_i\|_F^2\} + b) \\ \max(0, 1 - y_i \{\mathrm{tr}(\mathbf{U}_+ \mathbf{U}_+^T \mathbf{X}_i \mathbf{X}_i^T) - \mathrm{tr}(\mathbf{U}_- \mathbf{U}_-^T \mathbf{X}_i \mathbf{X}_i^T)\} + b)\end{array}}$$

Holistic representation based method

When bilinear SVM meets bilinear feature

maximum Frobenius norm

(d) our model (**LRBP-I**)



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

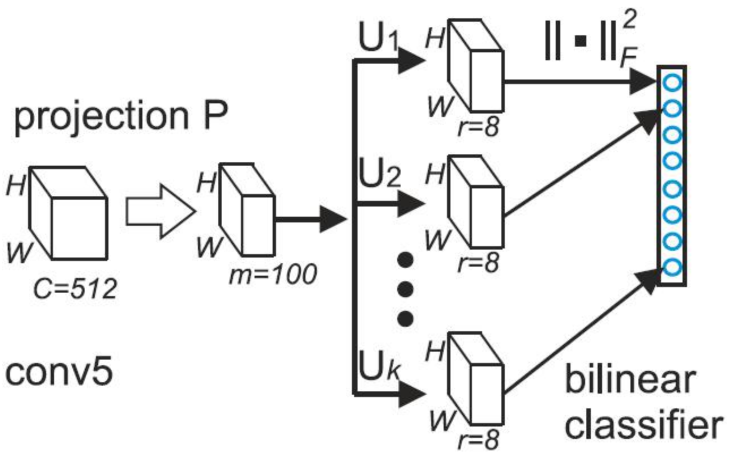
Holistic representation based method

When bilinear SVM meets bilinear feature

maximum Frobenius norm

no need to compute bilinear features when testing

(d) our model (**LRBP-I**)



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

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

UCIrvine
University of California, Irvine

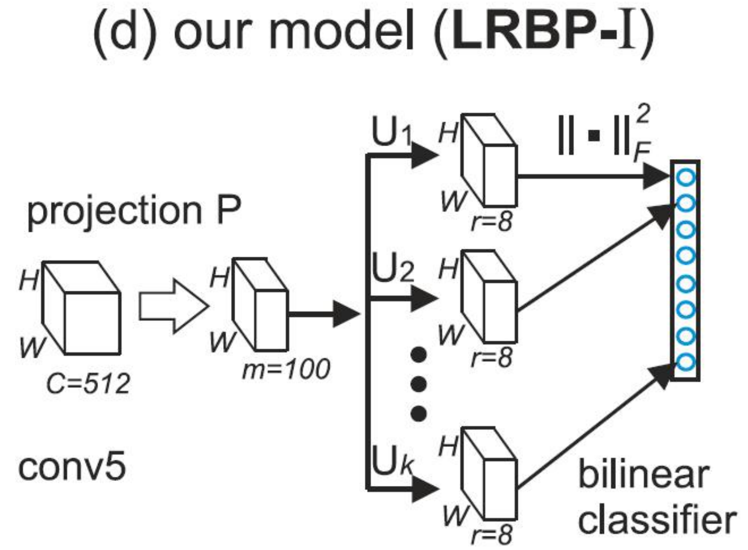
Holistic representation based method

When bilinear SVM meets bilinear feature

maximum Frobenius norm

no need to compute bilinear features when testing

200 classes, then param size is reduced from **200*512*512** to **200*512*8**



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

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

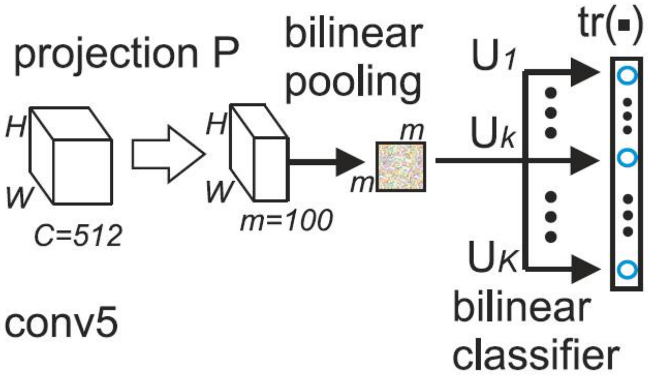
UCIrvine
University of California, Irvine

Holistic representation based method

explicitly computing bilinear features

more efficient useful when hw>m

our model (**LRBP-**II)



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* $\boxed{11}$ *spans the subspace of the singular vectors corresponding of the largest* m *singular values of* $[\mathbf{U}_1, \ldots, \mathbf{U}_K]$.

Holistic representation based method

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

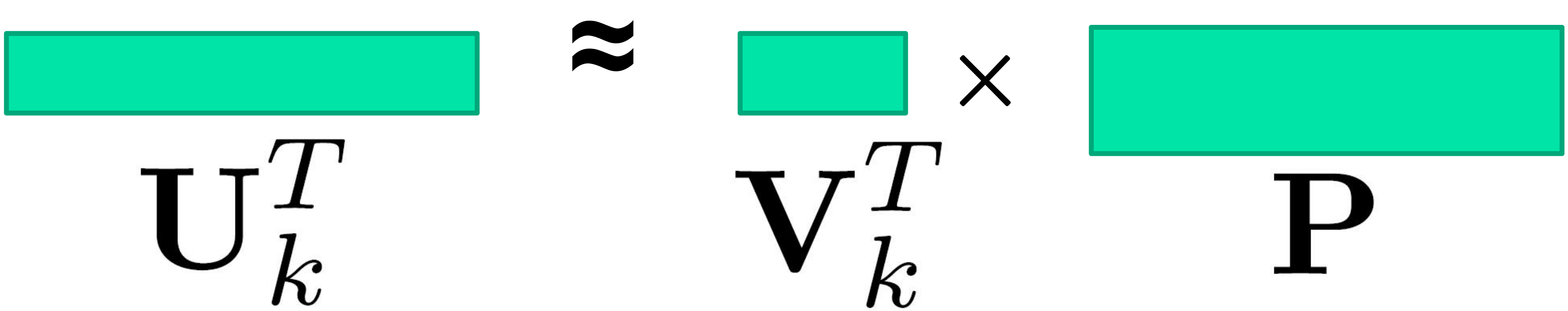
$$\min_{\mathbf{V}_k, \mathbf{P}} \sum_{k=1}^{K} \|\mathbf{U}_k - \mathbf{P}\mathbf{V}_k\|_F^2$$

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

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

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

$$m < c$$



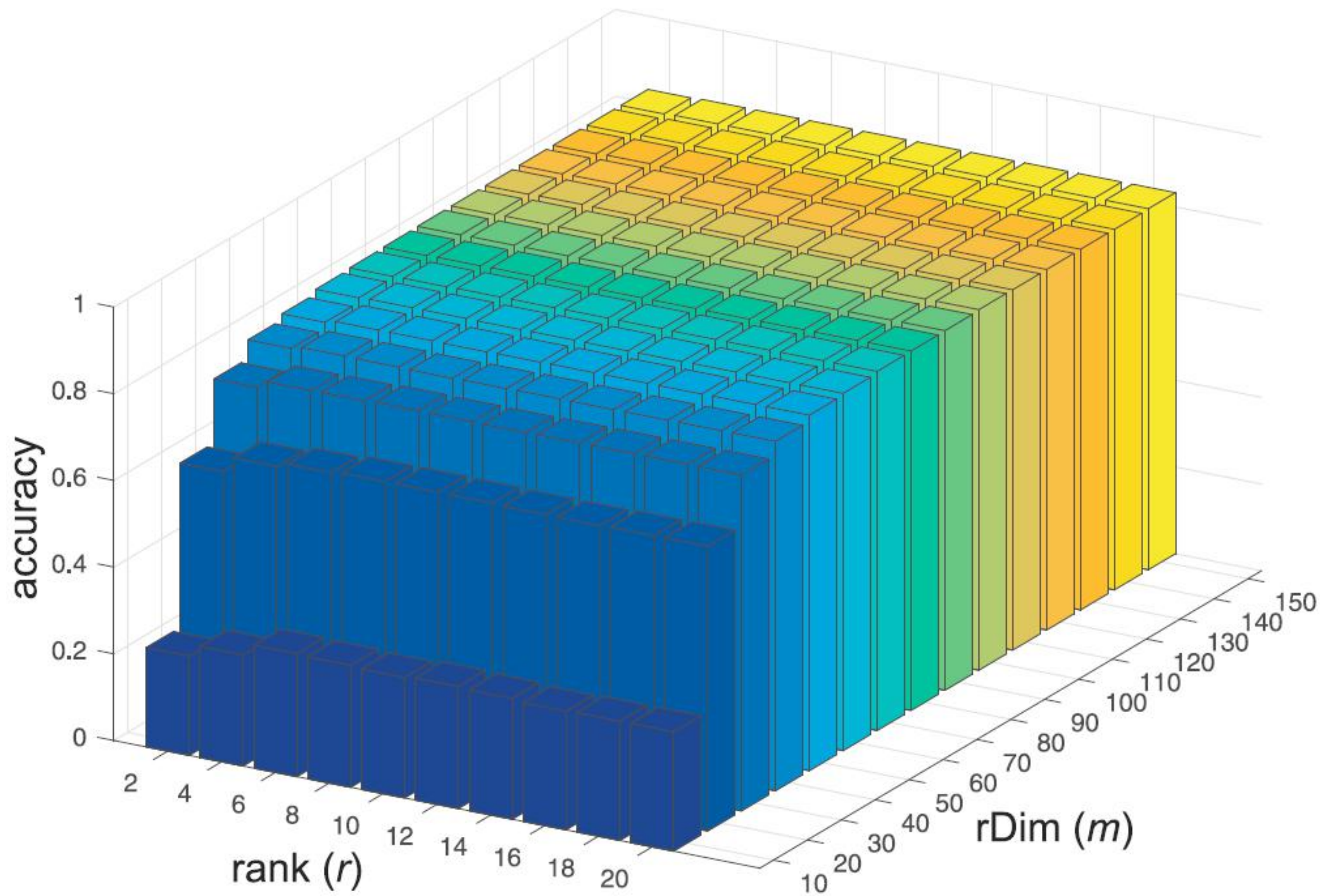
Holistic representation based method

Studying the two hyperparameters

- low dimension *m*
- low *rank r*

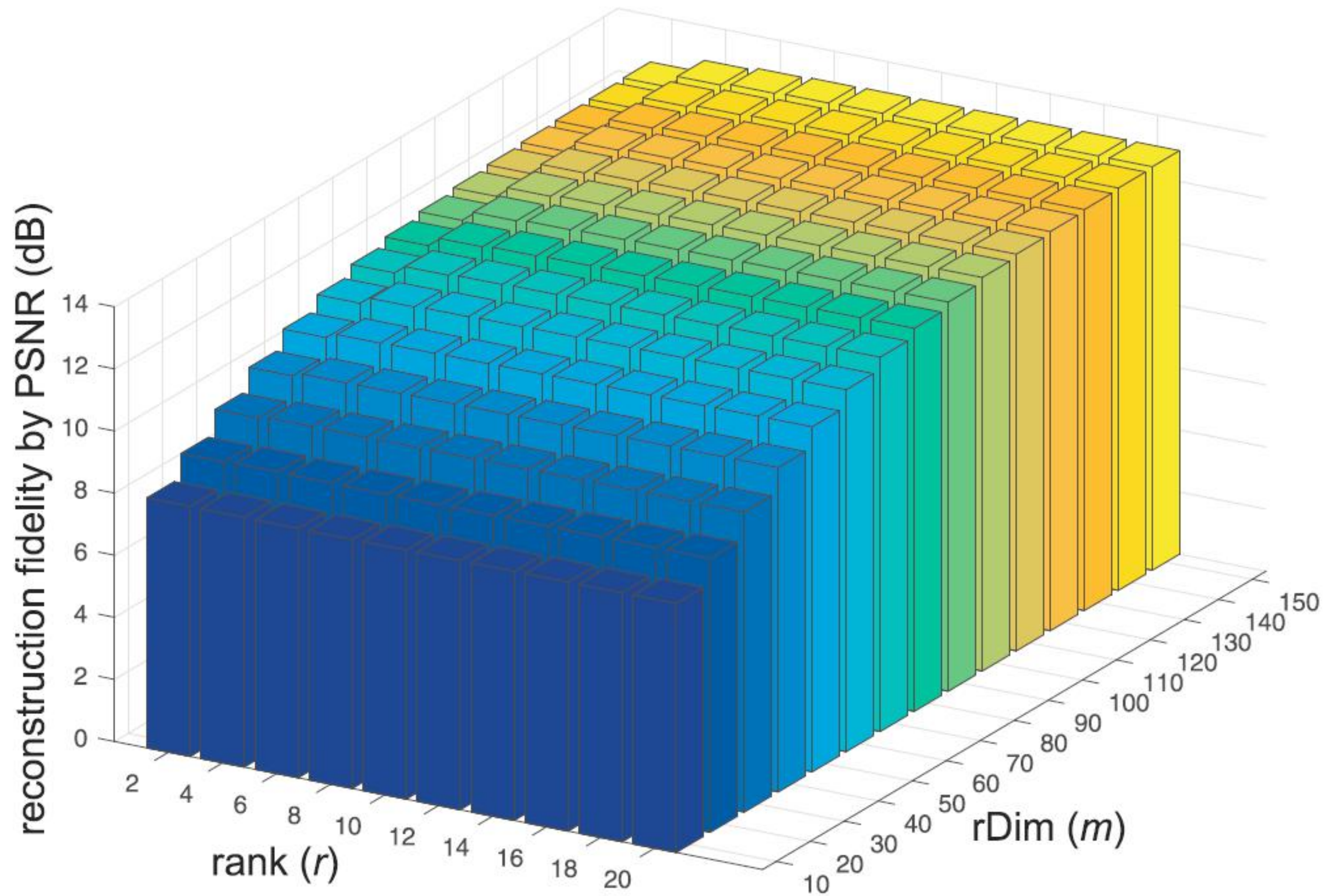
Holistic representation based method

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



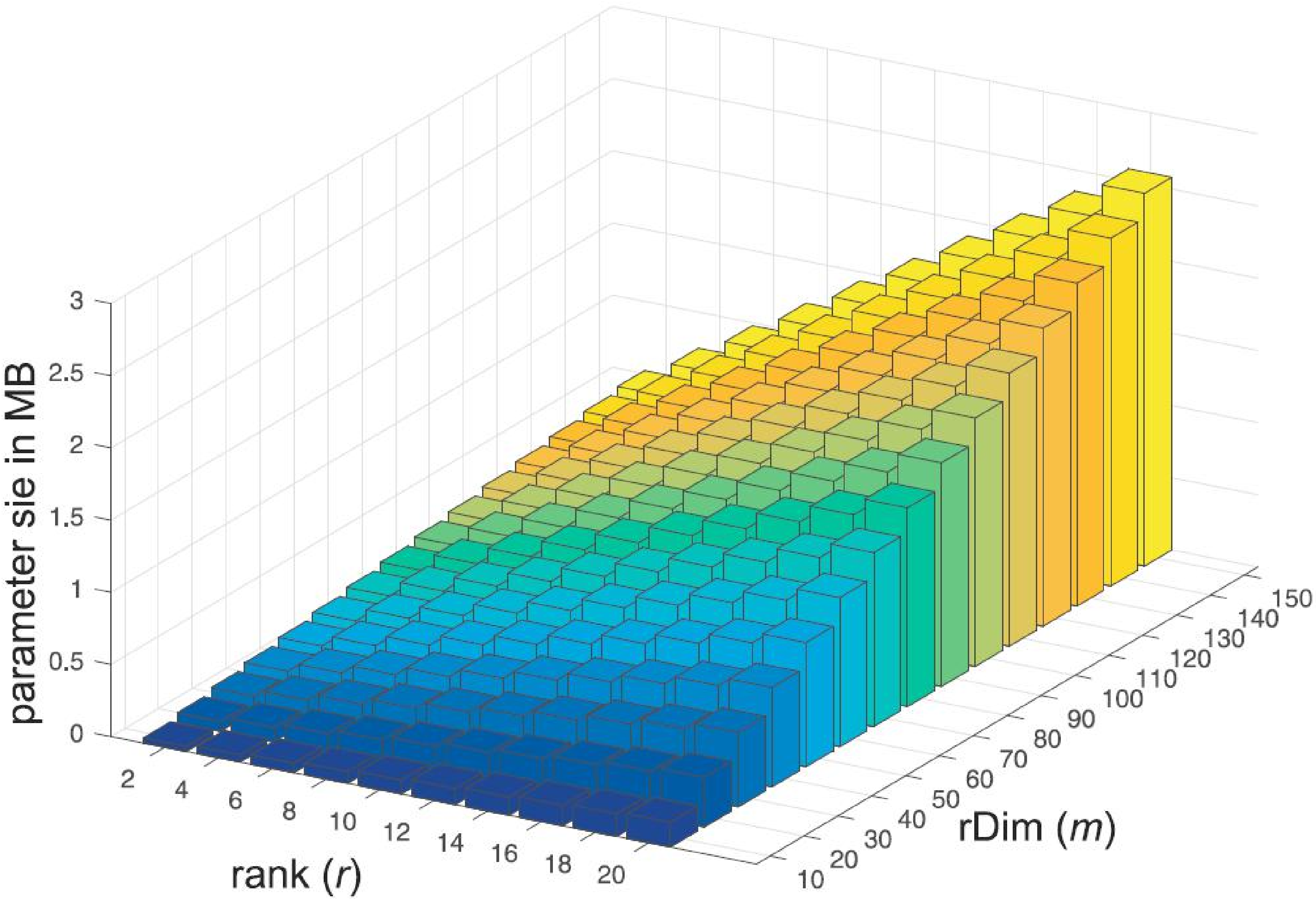
Holistic representation based method

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



Holistic representation based method

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



UCIrvine
University of California, Irvine

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

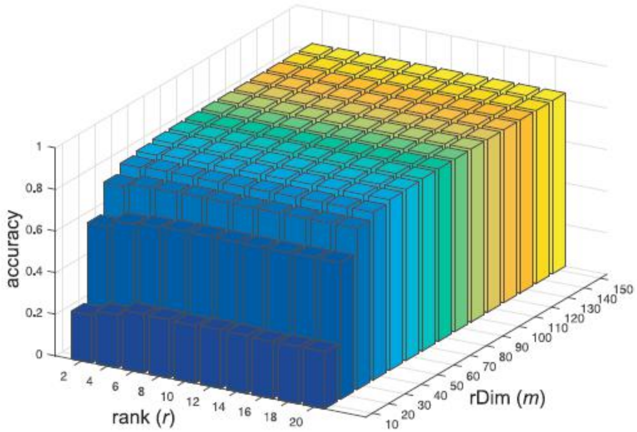


Figure 5: Classification accuracy on CUB-200 dataset [31] vs. reduced dimension (m) and rank (r).

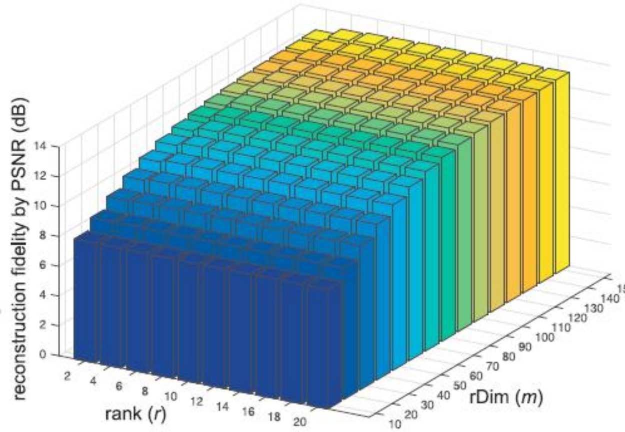


Figure 6: Reconstruction fidelity of classifier parameters measured by peak signal-to-noise ratio versus reduced dimension (m) and rank (r).

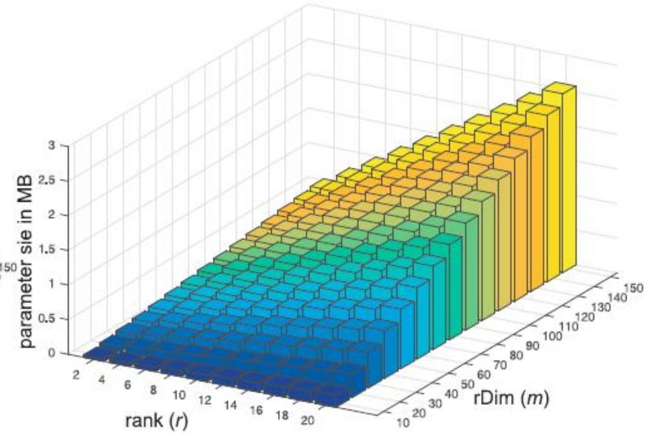


Figure 7: The learned parameter size versus reduced dimension (m) and rank (r).

UCIrvine
University of California, Irvine

Holistic representation based method

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

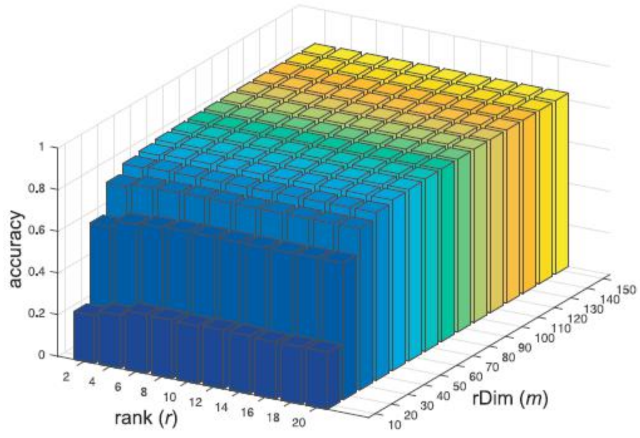


Figure 5: Classification accuracy on CUB-200 dataset [31] vs. reduced dimension (m) and rank (r).

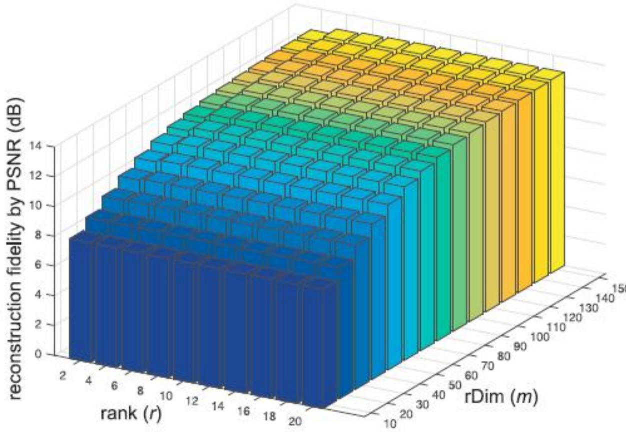


Figure 6: Reconstruction fidelity of classifier parameters measured by peak signal-to-noise ratio versus reduced dimension (m) and rank (r).

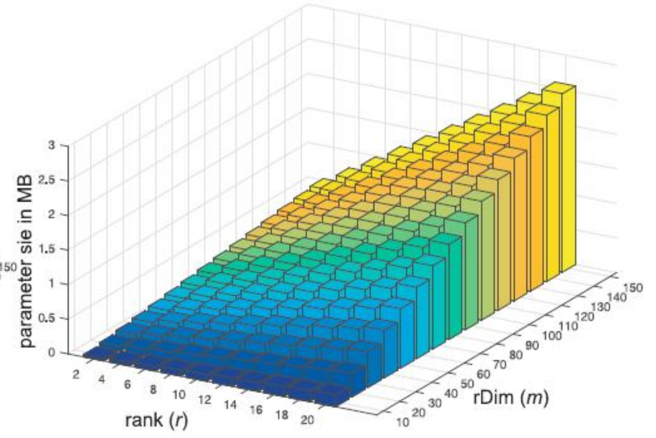


Figure 7: The learned parameter size versus reduced dimension (m) and rank (r).

if 200 classes, then param size is reduced

from 200*512*512 (~5.2 x 10e7 single precision)

to (200*8*100+100*512) (~2.1 x 10e5 single precision)

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

Holistic representation based method

Details on the complexity

	Full Bilinear	Random Maclaurin	Tensor Sketch	LRBP-I	LRBP-II
Feature Dim	c^2 [262K]	d [10K]	d [10K]	mhw [78K]	m^2 [10K]
Feature computation	$O(hwc^2)$	$O(hwcd)$	$O(hw(c + d\log d))$	$O(hwmc)$	$O(hwmc + hwm^2)$
Classification comp.	$O(Kc^2)$	$O(Kd)$	$O(Kd)$	$O(Krmhw)$	$O(Krm^2)$
Feature Param	0	$2cd$ [40MB]	$2c$ [4KB]	cm [200KB]	cm [200KB]
Classifier Param	Kc^2 [KMB]	Kd [K·32KB]	Kd [K·32KB]	Krm [K·3KB]	Krm [K·3KB]
Total ($K = 200$)	Kc^2 [200MB]	$2cd + Kd$ [48MB]	$2c + Kd$ [8MB]	$cm + Krm$ [0.8MB]	$cm + Krm$ [0.8MB]

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

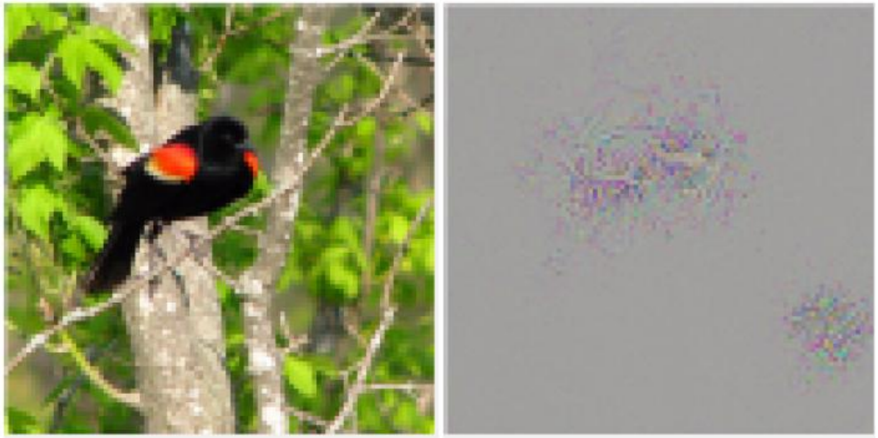
Holistic representation based method

Qualitative evaluation for understanding the model

Holistic representation based method

Qualitative evaluation for understanding the model

- gradient map --- backpropogating error to input image

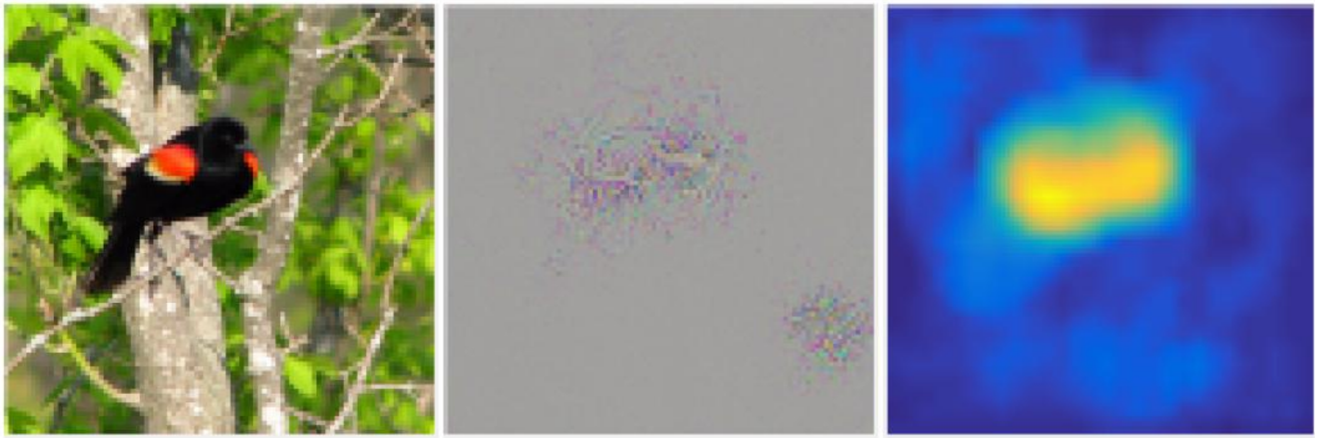


UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

Holistic representation based method

Qualitative evaluation for understanding the model

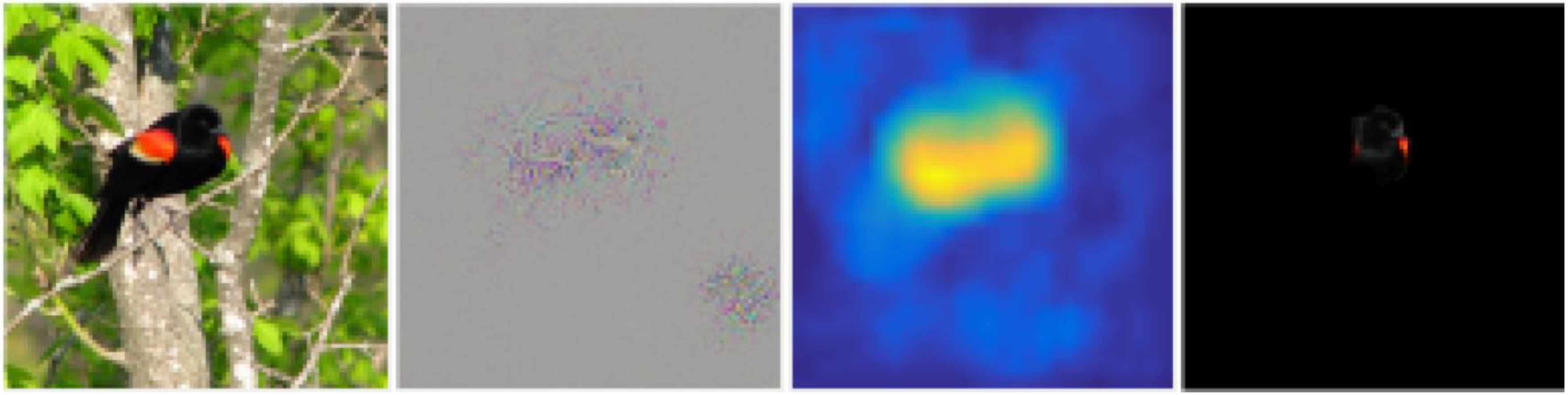
- gradient map --- backpropogating error to input image
- average activation map



Holistic representation based method

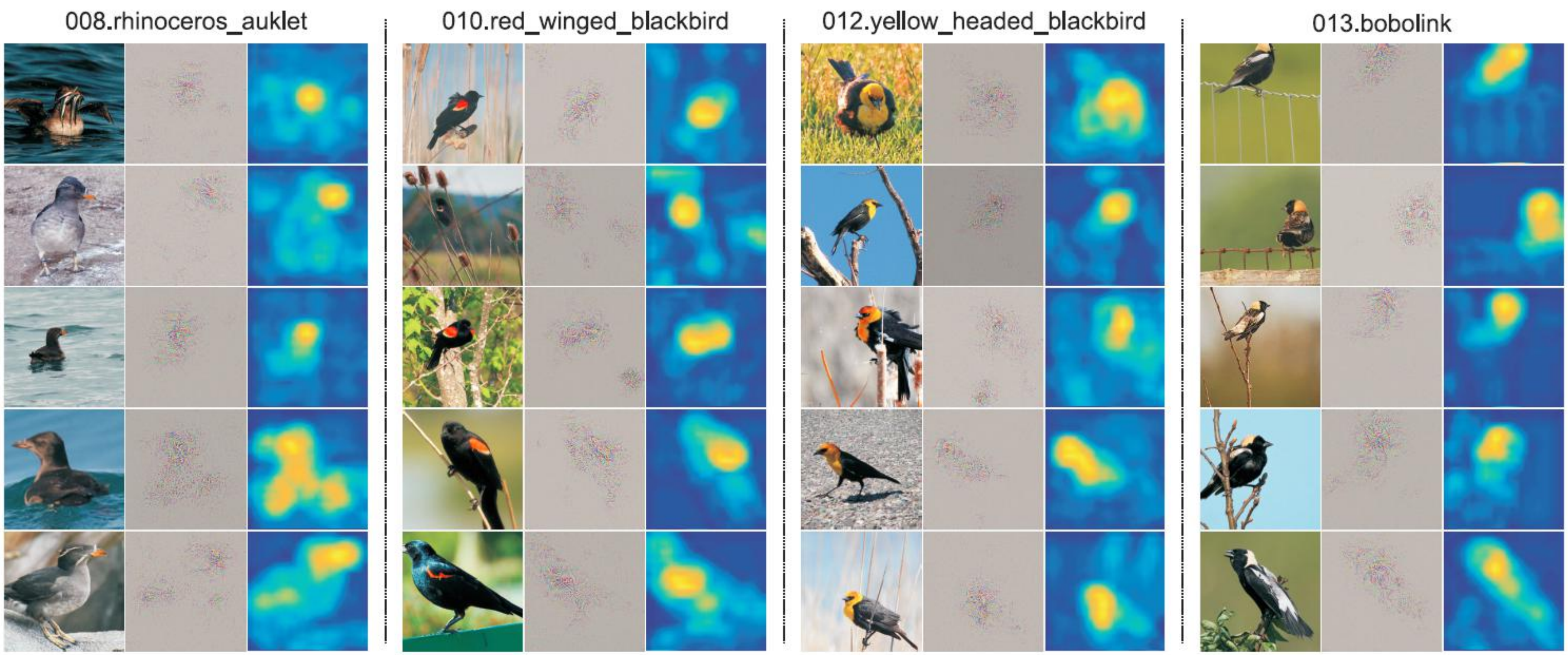
Qualitative evaluation for understanding the model

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



Holistic representation based method

Qualitative evaluation for understanding the model



Holistic representation based method

Conclusion

Holistic representation based method

Conclusion

1. a more compact and powerful model by coupling bilinear classifier and bilinear feature for fine-grained classification

Holistic representation based method

Conclusion

1. a more compact and powerful model by coupling bilinear classifier and bilinear feature for fine-grained classification
2. a new direction for a weakly supervised visual learning

Holistic representation based method

Conclusion

1. a more compact and powerful model by coupling bilinear classifier and bilinear feature for fine-grained classification
2. a new direction for a weakly supervised visual learning
3. useful for learning interpretable attentions

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

UCIrvine
University of California, Irvine

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

Skilled experts trained for years have to identify by eye

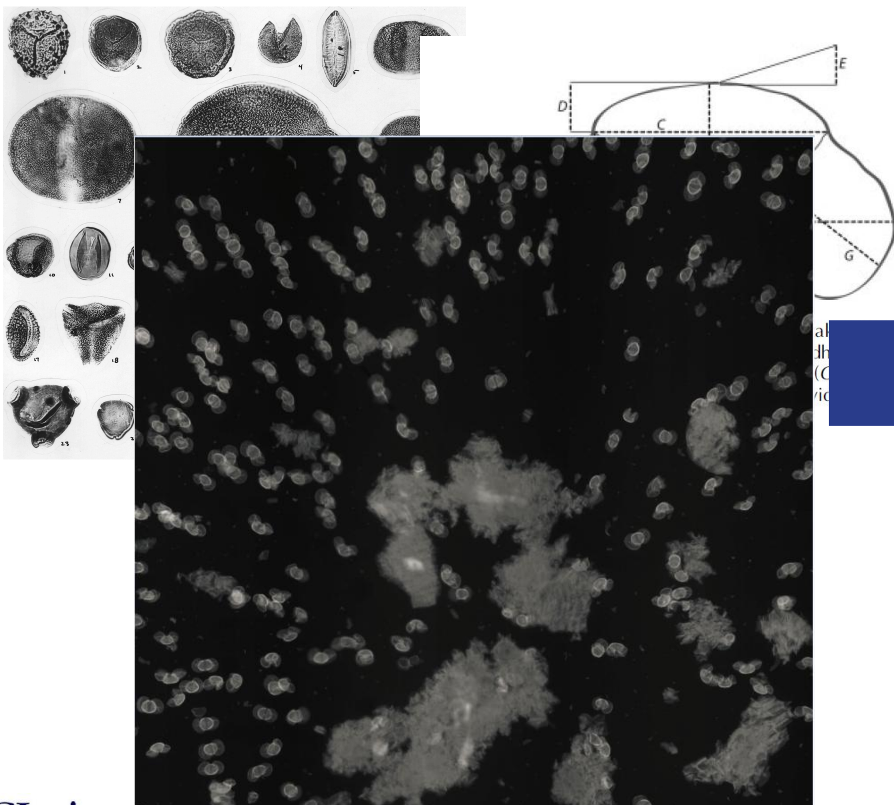
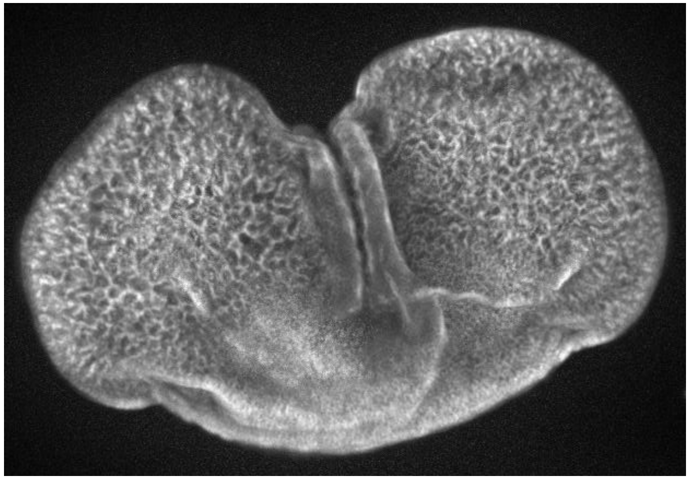


image from Surangi W. Punyasena

Why do we care about identifying pollen?

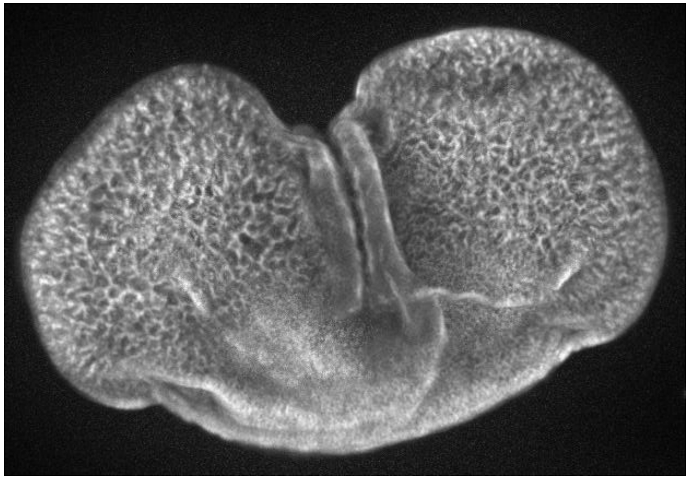
- Pollen grains are ubiquitous and well preserved in the fossil record



UCIrvine
University of California, Irvine

Why do we care about identifying pollen?

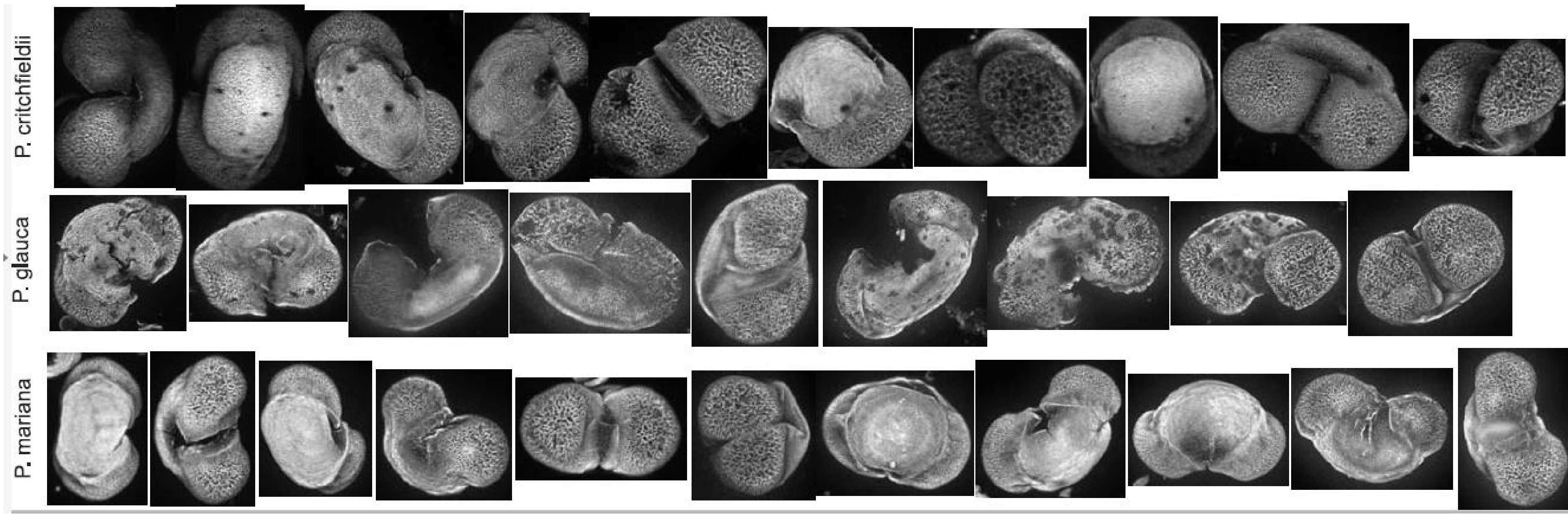
- Pollen grains are ubiquitous and well preserved in the fossil record
- Identification of pollen samples allows for analysis of plant biodiversity and evolution, understanding history of long-term climate change, etc...



UCIrvine
University of California, Irvine

Patch-match based method

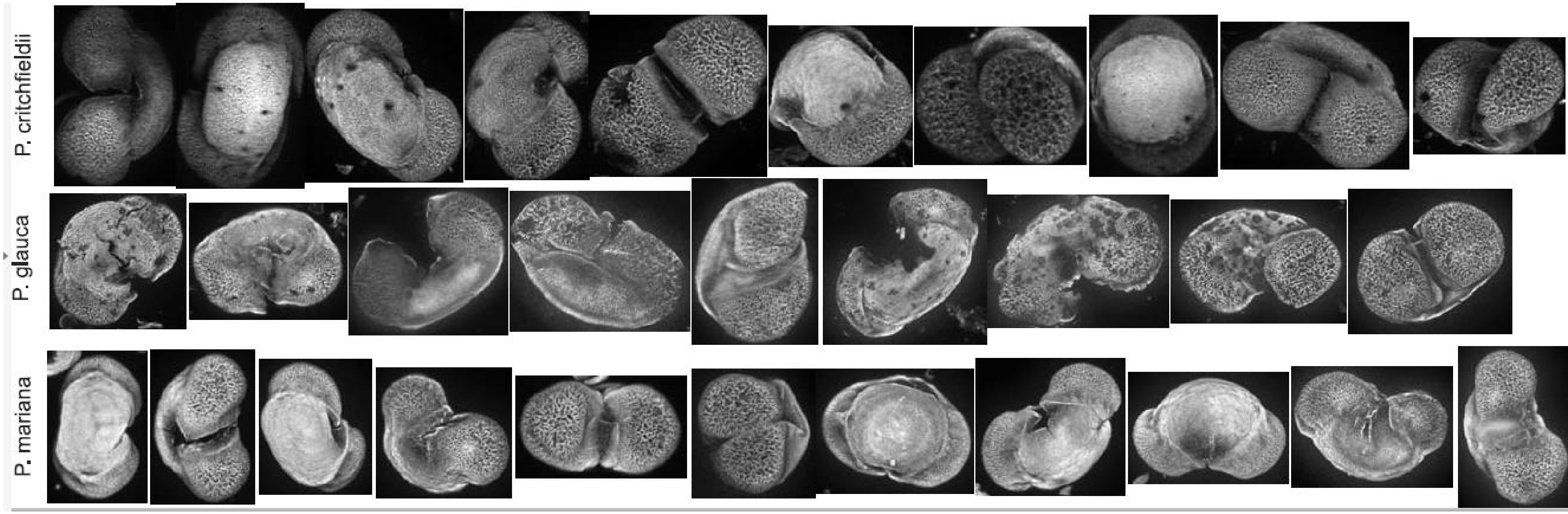
A specific dataset for this exploration



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?

1. It is expensive to collect and annotate data.

Quantitative Result on Fossil Pollen

Why not holistic representation?

1. It is expensive to collect and annotate data.

2. There are not enough training data using 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

1. It is expensive to collect and annotate data.

2. There are not enough training data using 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

1. It is expensive to collect and annotate data.

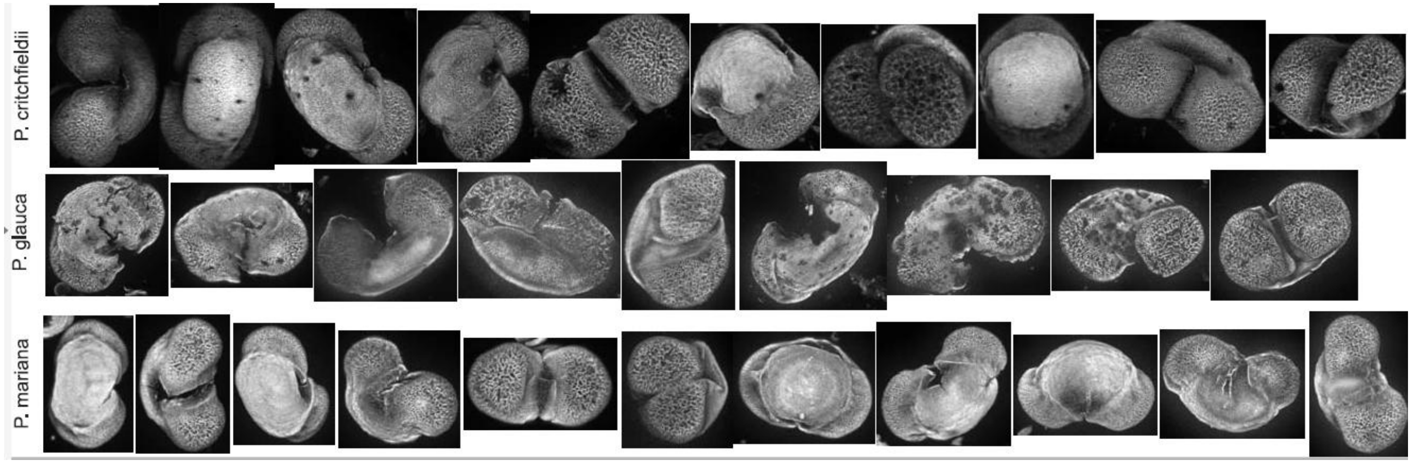
2. There are not enough training data using holistic representation.

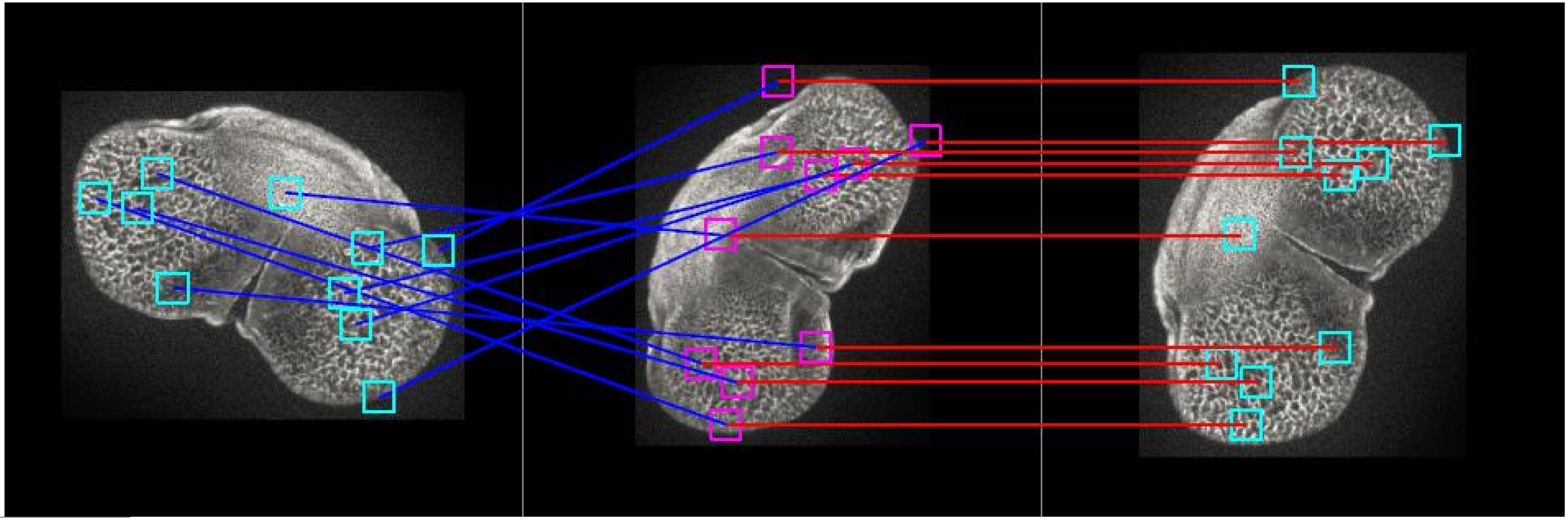
Therefore, it's better to match local patches with geometric constraints.

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

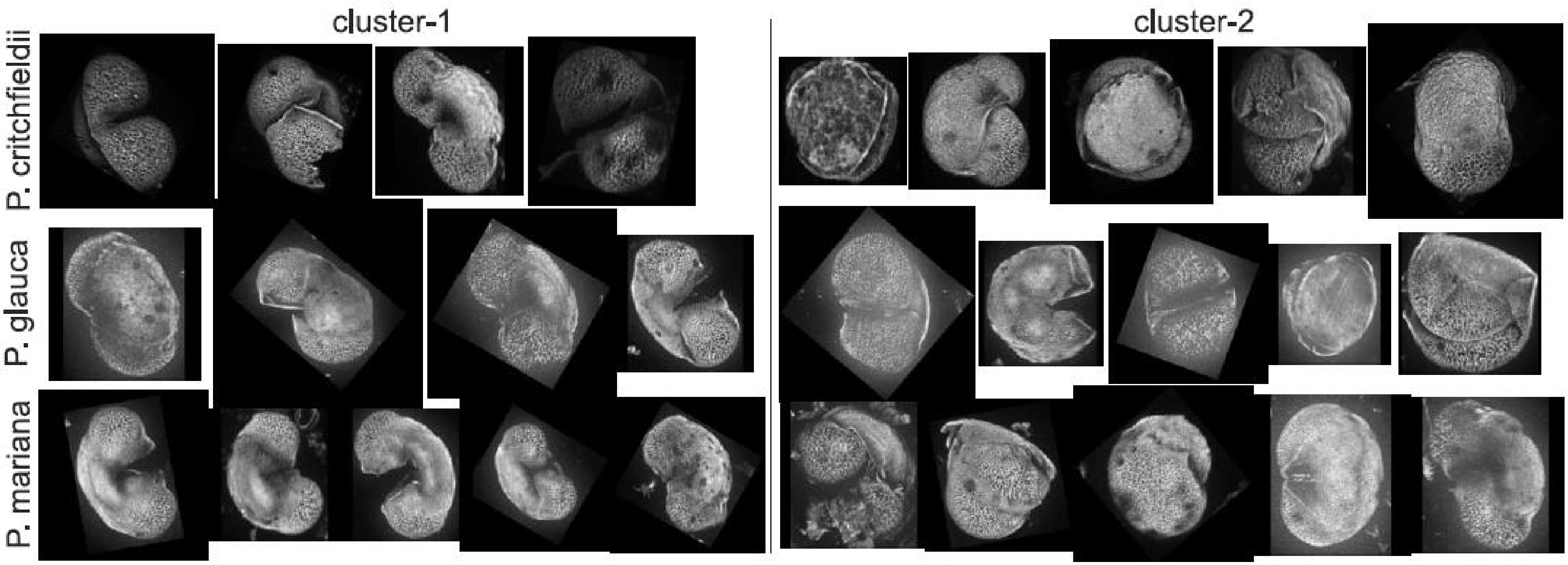
perform *k*-medoids clustering on an affinity graph of training set,

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

in-plate rotation viewpoint calibration

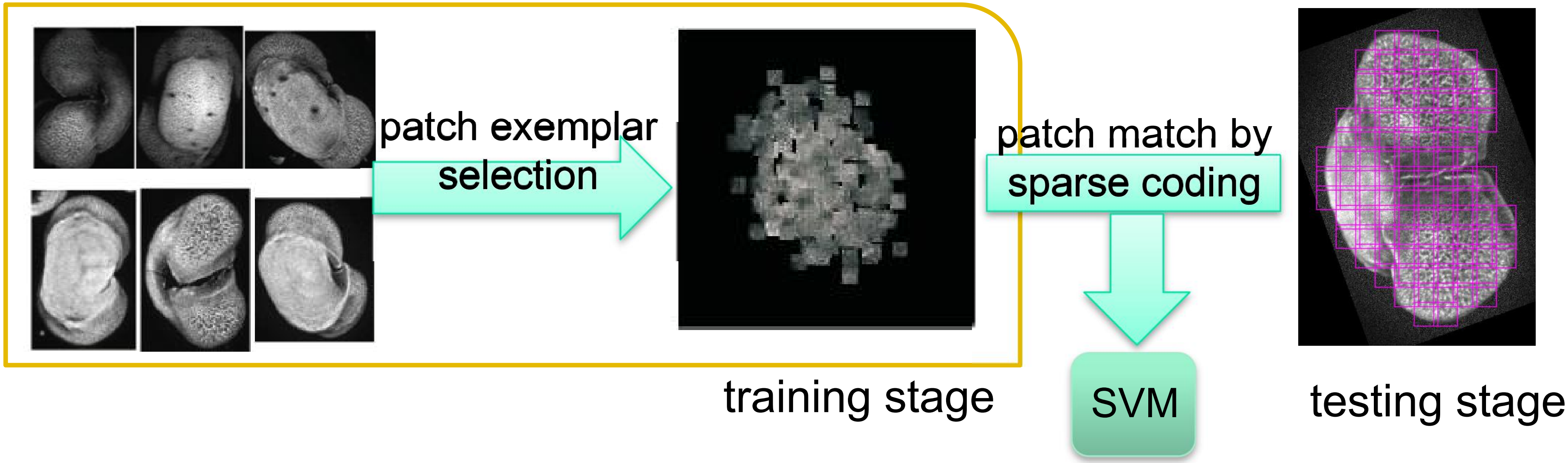
perform *k*-medoids clustering on an affinity graph of training set,

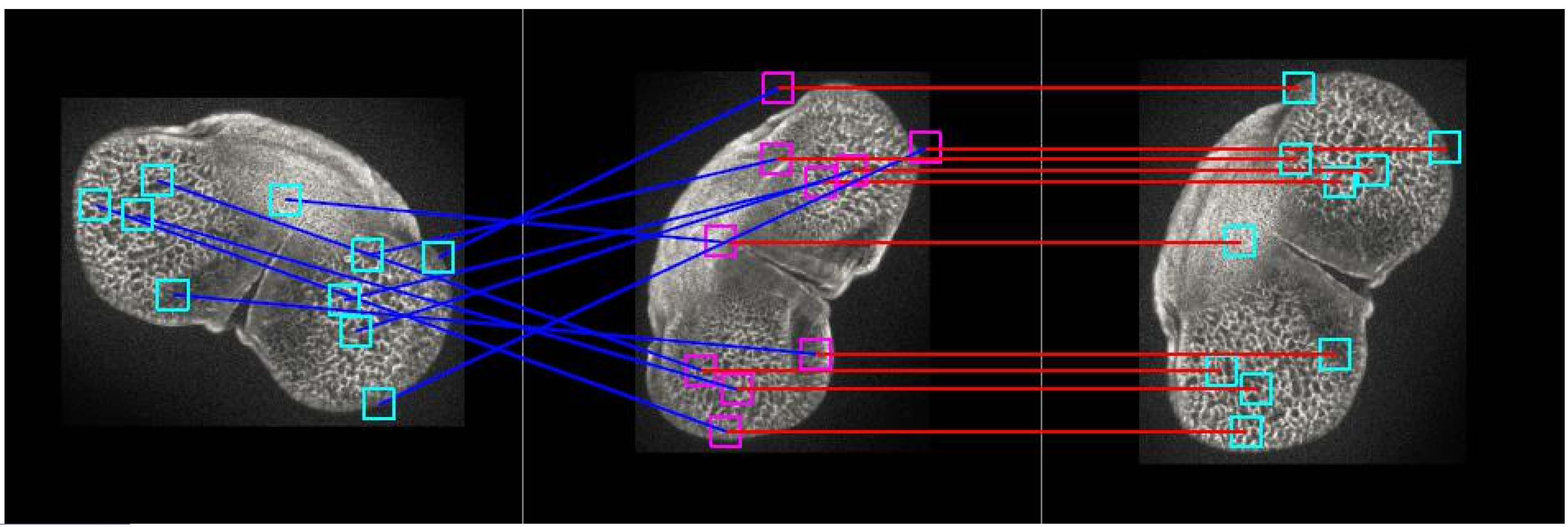
where pairwise similarity is based on Euclidean distance of pollen grain silhouette



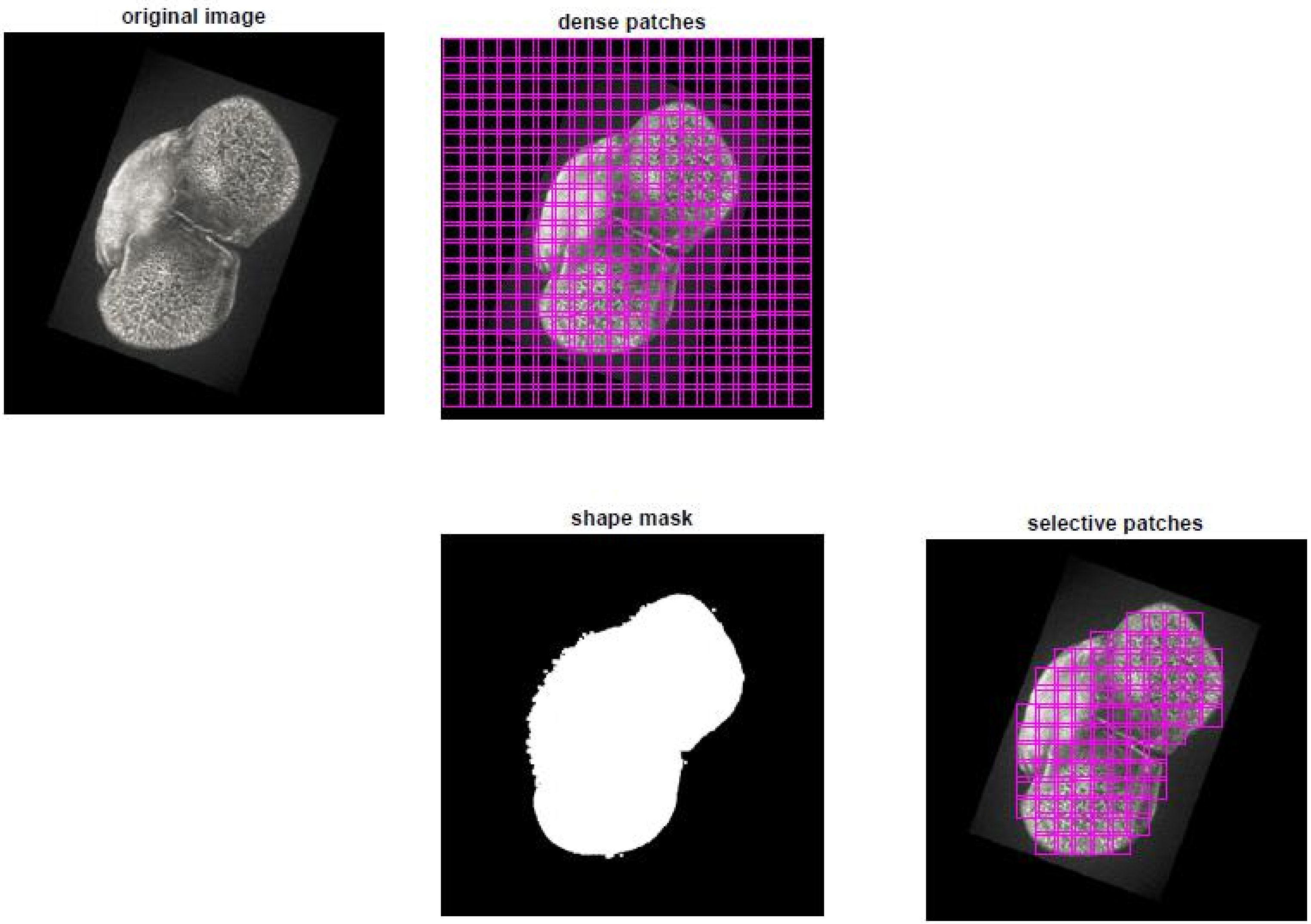
UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

our patch-match based method



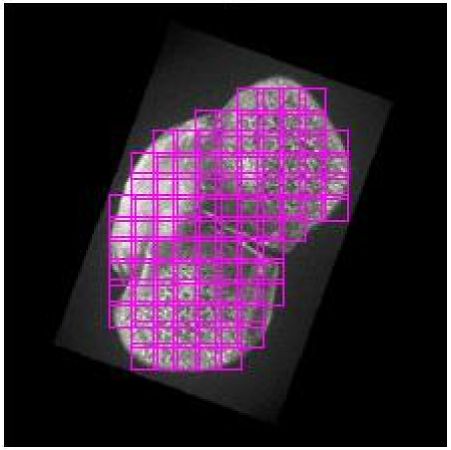


discriminative patch selection



discriminative patch selection

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



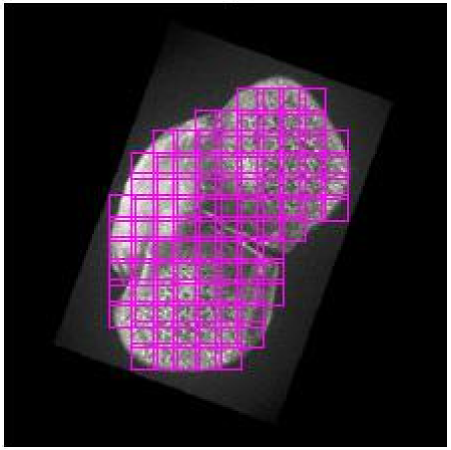
UCIrvine
University of California, Irvine

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

selective patches



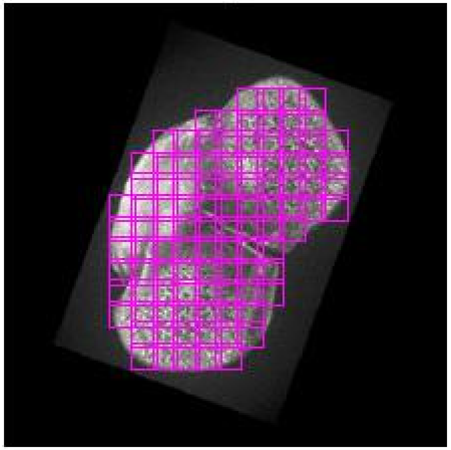
UCIrvine
University of California, Irvine

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

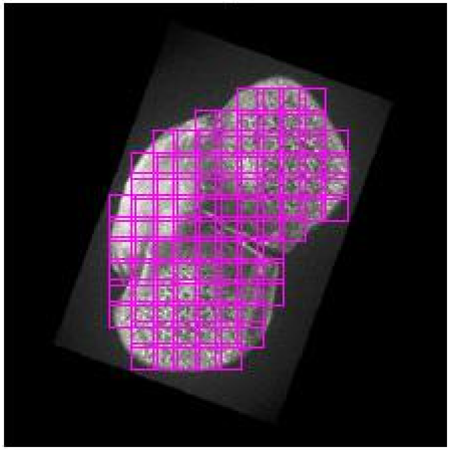
selective patches



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

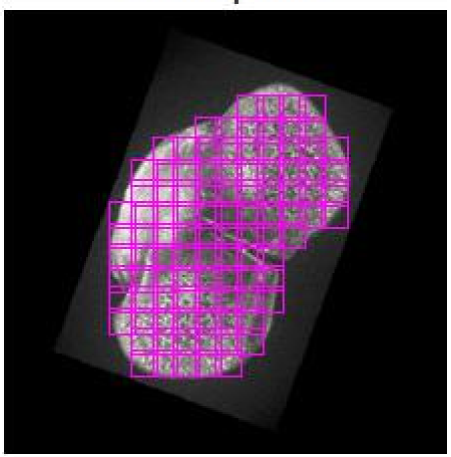


UCIrvine
University of California, Irvine

discriminative patch selection

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

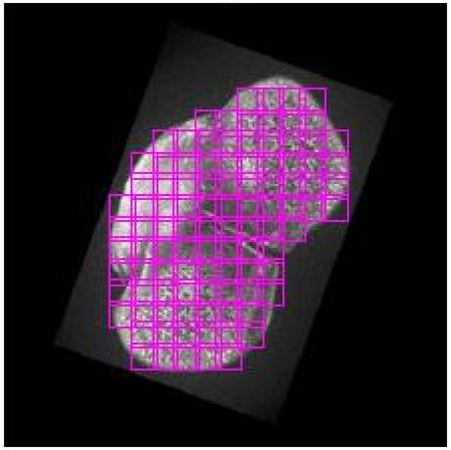
1. representative in feature space
2. spatially distributed in input space
3. discriminative
4. class balance



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



UCIrvine
University of California, Irvine

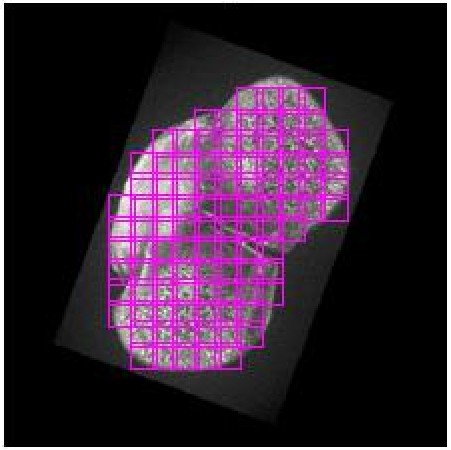
discriminative patch selection

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*

selective patches



example: representational power

Maximizing the following set function is 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 is NP-hard.

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

A more general, well-known problem is the facility location problem, for example optimally placing sensors to monitor temperature.

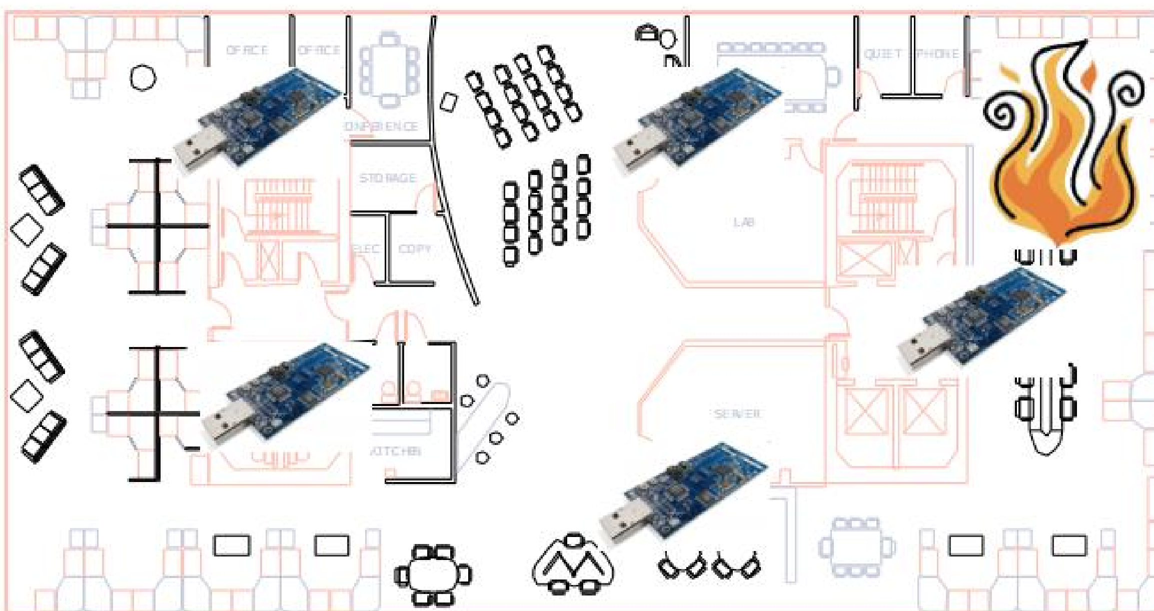


photo credited by Andreas Krause

selected discrminative patches

Identification by patch-match sparse coding

1. Automatic patch exemplar selection (dictionary learning)

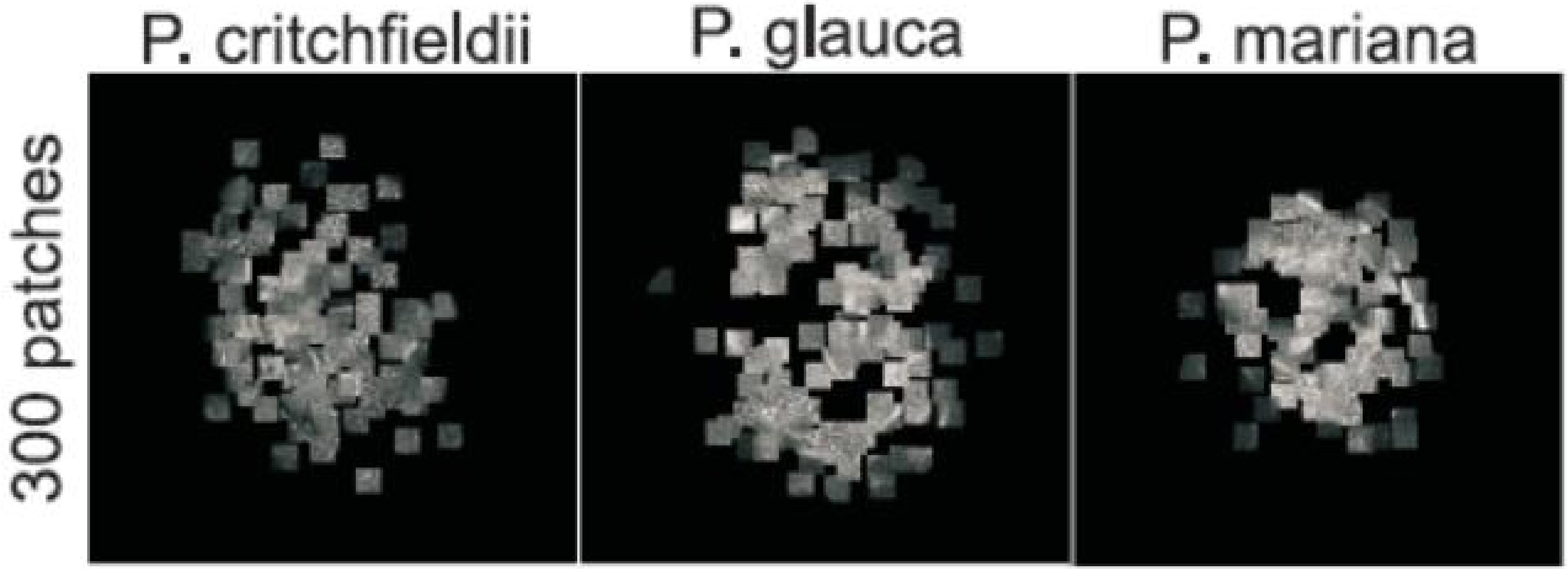
based on discriminative and generative criteria

selected discrminative patches

<u>Identification by patch-match sparse coding</u>

1. Automatic patch exemplar selection (dictionary learning)

based on discriminative and generative criteria



Automatically selected patches

selected discrminative patches

Identification by patch-match sparse coding

1. Automatic patch exemplar selection (dictionary learning)

based on discriminative and generative criteria

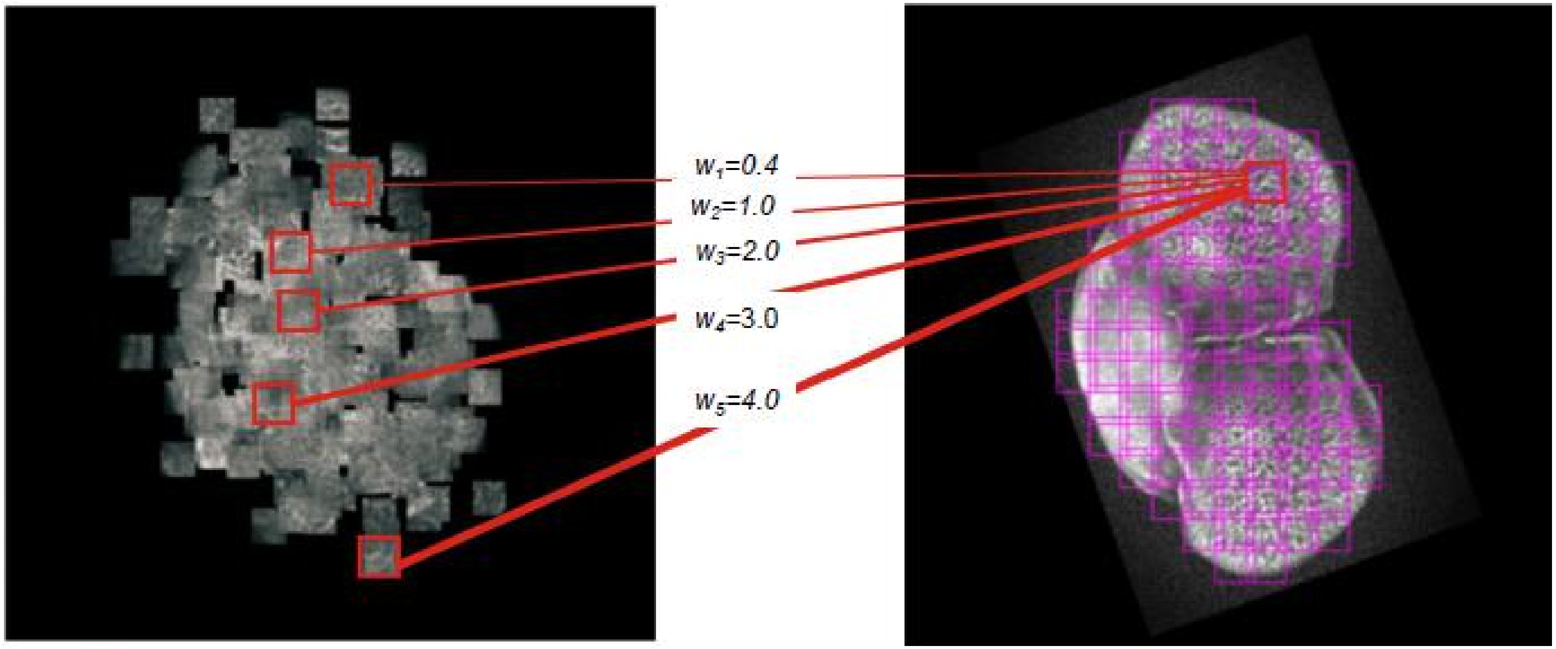
patch-match for identification

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

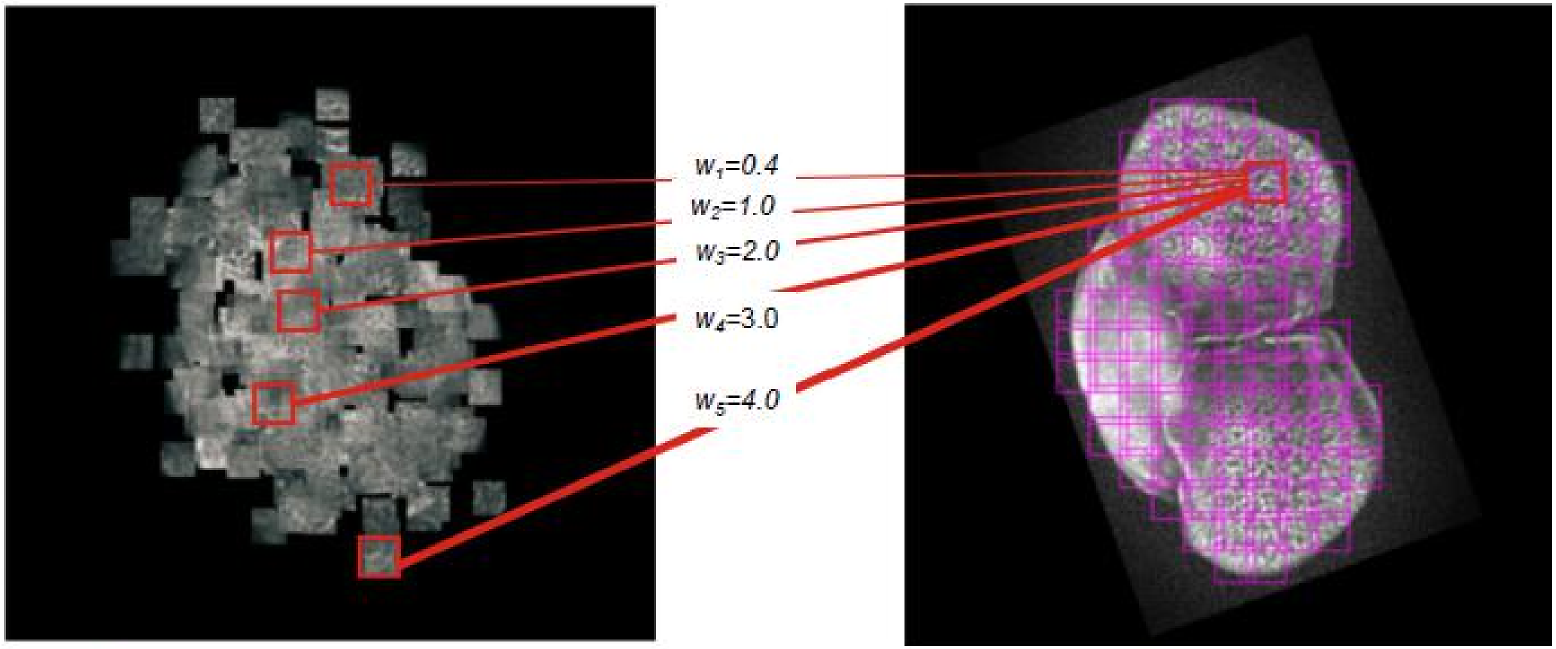


Spatially aware dictionary

Test image

UCIrvine
University of California, Irvine

spatially aware coding (SACO)



Spatially aware dictionary

Test image

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

Spatial weights

Exemplar patches (dictionary)

Test patch

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

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

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

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

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

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

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

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

SACO-I

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

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

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

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

SACO -- Faster Matching

feedforward shrinkage function by transforming dictionary patches into convolutional filters

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

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

SACO-II

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

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!

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**

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

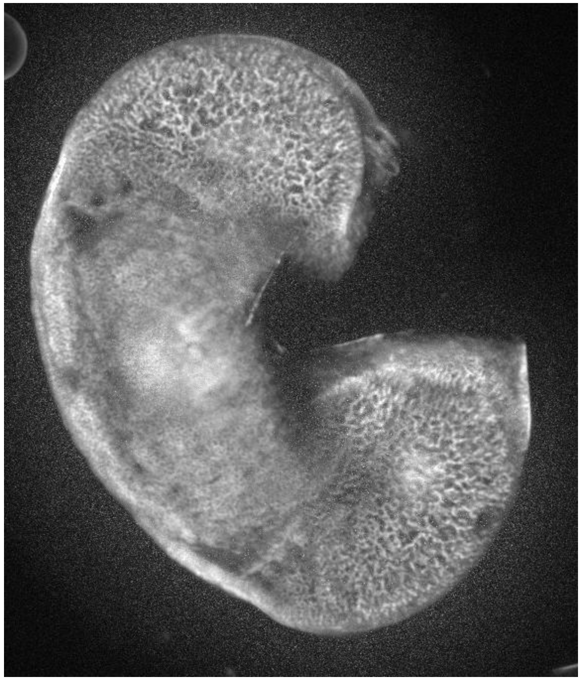
Identifying Fossil Pollen with Modern Reference

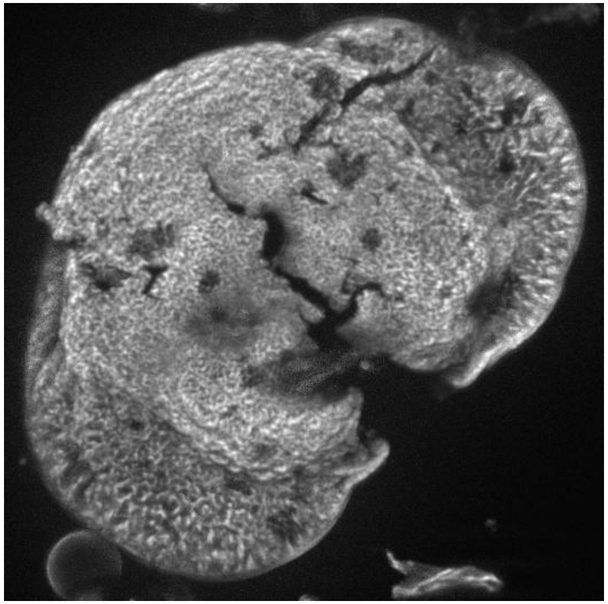
Fossil pollen grains are degraded over time.

using patches from modern pollen reference to identify fossilized ones

modern pollen grain from glauca

fossil pollen pollen grain from glauca

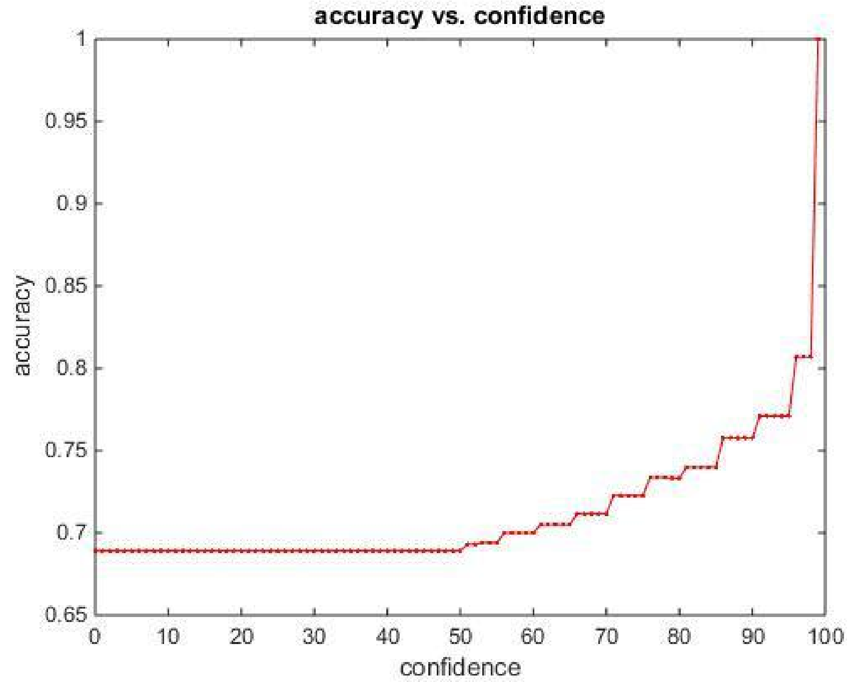




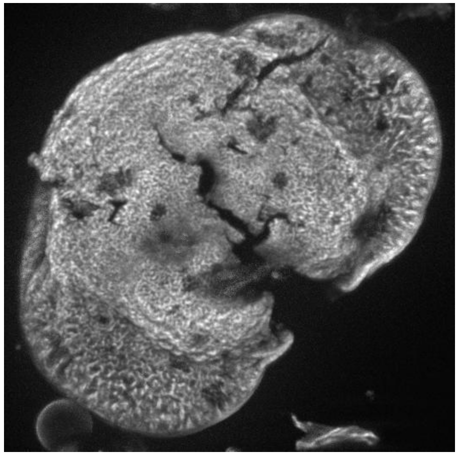
Identifying Fossil Pollen with Modern Reference

- 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



fossil



UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

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

leaving blank

Content after this page is not suitable for people to watch!