

CNN wrapup
and
Visual attributes

Thurs April 26
Kristen Grauman
UT Austin

Last time

- Evaluation
 - Scoring an object detector
 - Scoring a multi-class recognition system
- Spatial pyramid match kernel

- (Deep) Neural networks

Today

- Convolutional neural networks
- Attributes

Learning a Hierarchy of Feature Extractors

- Each layer of hierarchy extracts features from output of previous layer
- All the way from pixels → classifier
- Layers have the (nearly) same structure

```

    graph LR
      Input[Image/video] --> L1[Layer 1]
      L1 --> L2[Layer 2]
      L2 --> L3[Layer 3]
      L3 --> Output[Labels]
    
```

- Train all layers jointly

Slide: Rob Fergus

Significant recent impact on the field

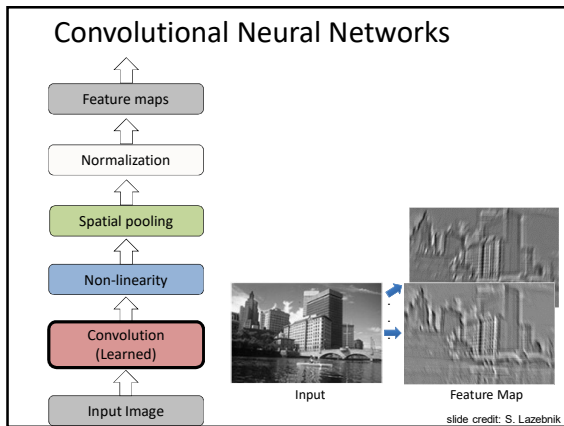
Year	Error (%)
1	25
2	16
3	11
4	7
5	4
6	3

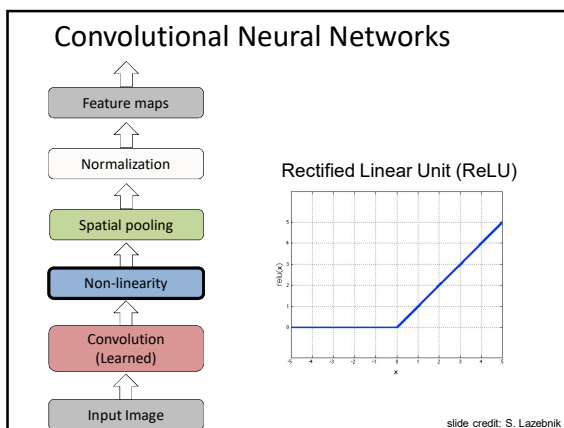
Slide credit: Dinesh Jayaraman

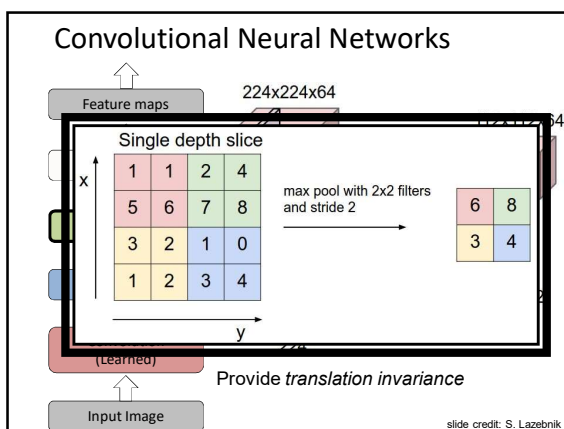
Convolutional Neural Networks (CNN, ConvNet, DCN)

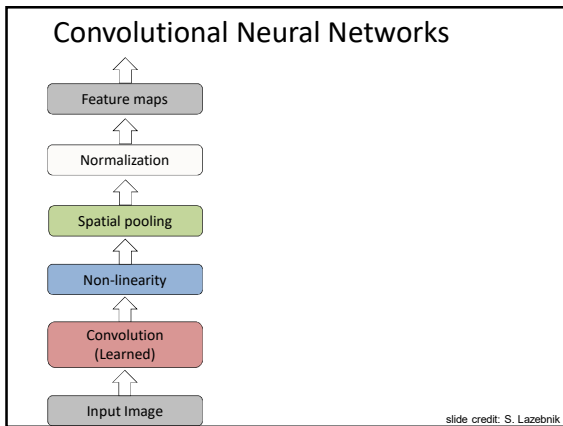
- CNN = a multi-layer neural network with
 - **Local** connectivity:
 - Neurons in a layer are only connected to a small region of the layer before it
 - **Share** weight parameters across spatial positions:
 - Learning shift-invariant filter kernels

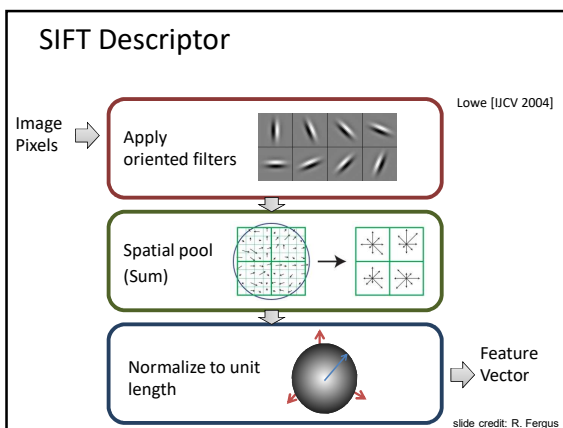
Jia-Bin Huang and Derek Hoiem, UIUC Image credit: A. Karpathy

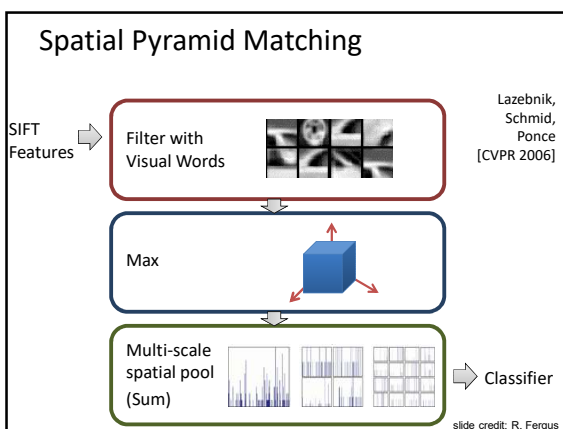






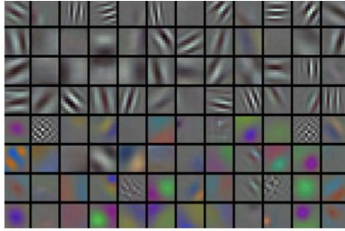






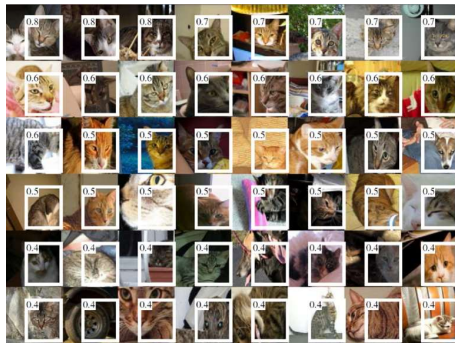
Visualizing what was learned

- What do the learned filters look like?



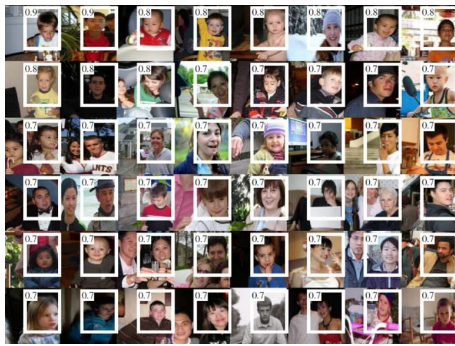
Typical first layer filters

Individual Neuron Activation




Jia-Bin Huang and Derek Hoiem, UIUC. RCNN [Girshick et al. CVPR 2014]

Individual Neuron Activation



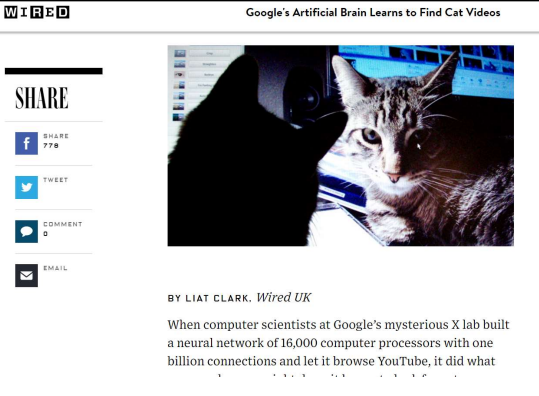
Jia-Bin Huang and Derek Hoiem, UIUC. RCNN [Girshick et al. CVPR 2014]

Individual Neuron Activation



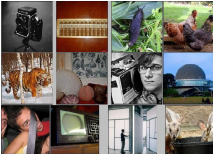

Jia-Bin Huang and Derek Hoiem, UIUC: RCNN [Girshick et al. CVPR 2014]

WIRED Google's Artificial Brain Learns to Find Cat Videos



<https://www.wired.com/2012/06/google-x-neural-network/>

Application: ImageNet



- ~14 million labeled images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk

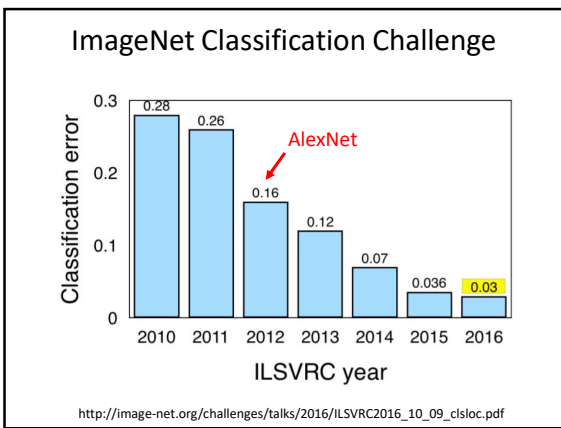
[Deng et al. CVPR 2009]

<https://sites.google.com/site/deeplearningcvpr2014> Slide: R. Fergus

AlexNet

- Similar framework to LeCun'98 but:
 - Bigger model (7 hidden layers, 650,000 units, 60,000,000 params)
 - More data (10^6 vs. 10^3 images)
 - GPU implementation (50x speedup over CPU)
 - Trained on two GPUs for a week

A. Krizhevsky, I. Sutskever, and G. Hinton,
 ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
 Jia-Bin Huang and Derek Hoiem, UIUC



Industry Deployment

- Used in Facebook, Google, Microsoft
- Image Recognition, Speech Recognition,
- Fast at test time

Taigman et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR'14

Slide: R. Fergus

Recap so far

- Neural networks / multi-layer perceptrons
 - View of neural networks as learning hierarchy of features
- Convolutional neural networks
 - Architecture of network accounts for image structure
 - “End-to-end” recognition from pixels
 - Together with big (labeled) data and lots of computation → major success on benchmarks, image classification and beyond

Beyond classification

- Detection
- Segmentation
- Regression
- Pose estimation
- Matching patches
- Synthesis

and many more...

Jia-Bin Huang and Derek Hoiem, UIUC

R-CNN: Regions with CNN features

- Trained on ImageNet classification
- Finetune CNN on PASCAL

R-CNN: *Regions with CNN features*

1. Input image 2. Extract region proposals (~2k) 3. Compute CNN features 4. Classify regions

Jia-Bin Huang and Derek Hoiem, UIUC

R-CNN [Girshick et al. CVPR 2014]

CNN for Regression



DeepPose [Toshev and Szegedy CVPR 2014]

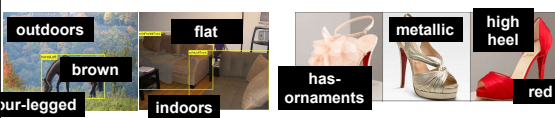
Jia-Bin Huang and Derek Hoiem, UIUC

Today

- Convolutional neural networks
- Attributes

What are visual attributes?

- Mid-level semantic properties shared by objects
- Human-understandable *and* machine-detectable



- Material, Appearance, Function/affordance, Parts...
- Adjectives
- Statements *about* visual concepts


[Oliva et al. 2001, Ferrari & Zisserman 2007, Kumar et al. 2008, Farhadi et al. 2009, Lampert et al. 2009, Endres et al. 2010, Wang & Mori 2010, Berg et al. 2010, Branson et al. 2010, Parikh & Grauman 2011, ...]

Examples: Binary Attributes

Facial properties

"Smiling Asian Men With Glasses"


Found 1344 results for smiling asian men with glasses in 0.220 secs. Displaying results 1 to 48.



Kumar et al. 2008

Examples: Binary Attributes

Object parts and shapes



Farhadi et al. 2009

Examples: Binary Attributes

Shopping descriptors



Berg et al. 2010

Attributes for search and recognition

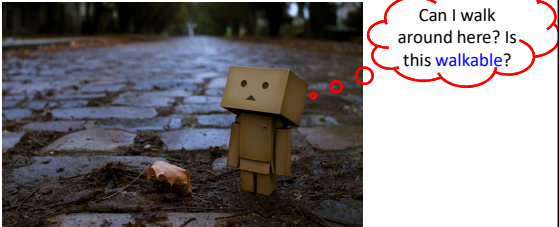
Language-based attributes give human way to

- o Teach novel categories with description
- o Communicate search queries
- o Give feedback in interactive search
- o Assist in interactive recognition

Slide credit: Kristen Grauman

Why attributes?

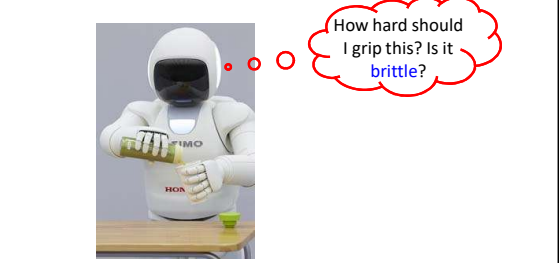
- Why would a robot need to recognize a scene?



Slide credit: Devi Parikh

Why attributes?

- Why would a robot need to recognize an object?



Slide credit: Devi Parikh

Why attributes?

- How do people naturally describe visual concepts?

Slide credit: Devi Parikh

Relative attributes


Idea: represent *visual comparisons* between classes, images, and their properties.

[Parikh & Grauman, ICCV 2011]

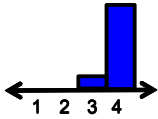
How to teach relative visual concepts?

How much is the person smiling?


How to teach relative visual concepts?



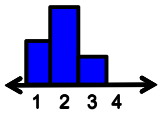
How much is the person smiling?




How to teach relative visual concepts?




How much is the person smiling?



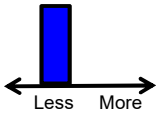
How to teach relative visual concepts?



>



?



Learning relative attributes

For each attribute, use ordered image pairs to train a ranking function:

$w_m^T x_i > w_m^T x_j$
 $\forall (i, j) \in O_m$

[Parikh & Grauman, ICCV 2011; Joachims 2002]

Learning relative attributes

Max-margin learning to rank formulation

$$\min \left(\frac{1}{2} \|w_m^T\|_2^2 + C \left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2 \right) \right)$$

s.t. $w_m^T (x_i - x_j) \geq 1 - \xi_{ij}$

$|w_m^T (x_i - x_j)| \leq \gamma_{ij}$

$\xi_{ij} \geq 0; \gamma_{ij} \geq 0$

Rank margin

Image \rightarrow Relative attribute score

Joachims, KDD 2002

Relating images

Rather than simply label images with their properties,

Not bright

Smiling

Not natural

[Parikh & Grauman, ICCV 2011]

Relating images

Now we can **compare** images by attribute's "strength"

bright →

smiling →

natural →

[Parikh & Grauman, ICCV 2011]

Interactive visual search

- Iteratively **refine** the set of retrieved images based on user feedback on results so far
- Potential to communicate more precisely the desired visual content

Slide credit: Adriana Kovachka

How is interactive search done today?

Keywords + **binary relevance feedback**

- Traditional binary feedback is imprecise
- Coarse communication between user and system

[Rui et al. 1998, Zhou et al. 2003, Tong & Chang 2001, Cox et al. 2000, Ferecatu & Geman 2007, ...]

Idea: Search via comparisons
[Kovashka et al., CVPR 2012]

“Like this... but more ornate”

- Whittle away irrelevant images via comparative feedback on properties of results

WhittleSearch: Relative attribute feedback

Query: “white high-heeled shoes”

Initial top search results

Feedback: “less formal than these”

Feedback: “shinier than these”

Refined top search results

[Kovashka et al., CVPR 2012, IJCV 2015]

WhittleSearch: Relative attribute feedback

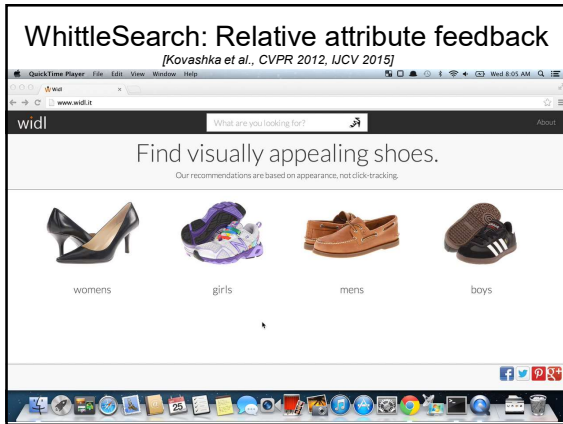
Initial reference images

Feedback: “similar hair style”

Feedback: “broader nose”

Refined top search results

[Kovashka et al., CVPR 2012, IJCV 2015]



Attributes for search and recognition

Attributes give human user way to

- o Teach novel categories with description
- o Communicate search queries
- o Give feedback in interactive search
- o Assist in interactive recognition

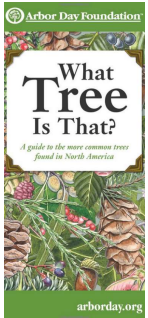
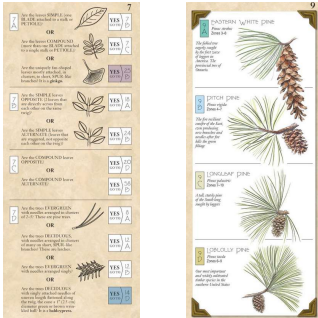
Slide credit: Kristen Grauman

What Plant Species is This?

The photograph shows a dense cluster of bright green, ovate leaves with serrated margins and prominent veins. The leaves are arranged in a somewhat regular pattern on a stem.

Slide credit: Neeraj Kumar




Let's Use a Field Guide

Slide credit: Neeraj Kumar

Categories of Recognition

Basic-Level

↕


Airplane? Chair?
Bottle? ...

Humans
Easy

Computers
Some Success

Slide credit: Steve Branson

Recognition With Humans in the Loop



Computer Vision

Cone-shaped Beak? **yes**


Computer Vision


American Goldfinch? **yes**

- Computers: reduce number of required questions
- Humans: drive up accuracy of vision algorithms

Wah et al., ICCV 2011, Van Horn et al. CVPR 2015 Slide credit: Steve Branson

Example Questions: Localize




 Click on the head
Click on the applicable part in the uploaded image to the left. If the part is not visible in the image, click "Not Visible".

Wah et al., Multi-class Recognition and Part Localization with Humans in the Loop, ICCV 2011 Slide credit: Steve Branson

Example Questions: Name attributes

You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

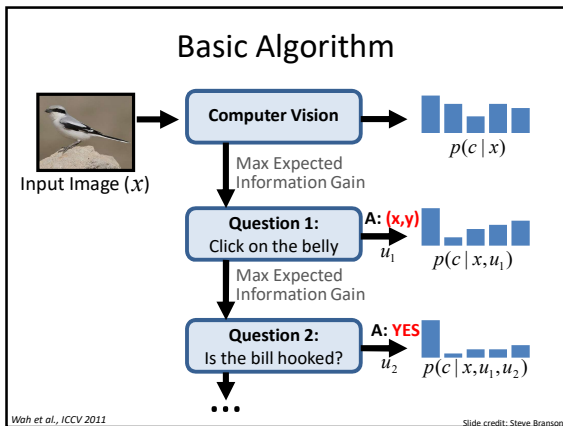


What is the color of the underparts of the bird? 10/20






Select at least one. If the underparts aren't visible, make your best guess, then select "Guessing". If the color is a mixture of two colors, select both (e.g., for blue-green select blue and green). If the underparts have multiple regions or patterns with multiple colors, select all relevant colors (e.g., for yellow with black stripes, select yellow and black).

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
White	Black	Grey	Buff	Brown
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Red	Red	Pink	Orange	Yellow
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Green	Olive	Blue	Purple	Shiny / iridescent



Wah et al., ICCV 2011 Slide credit: Steve Branson



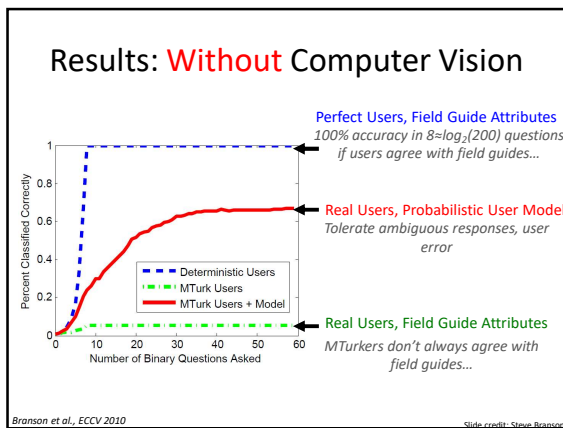
CUB-200-2011 Dataset

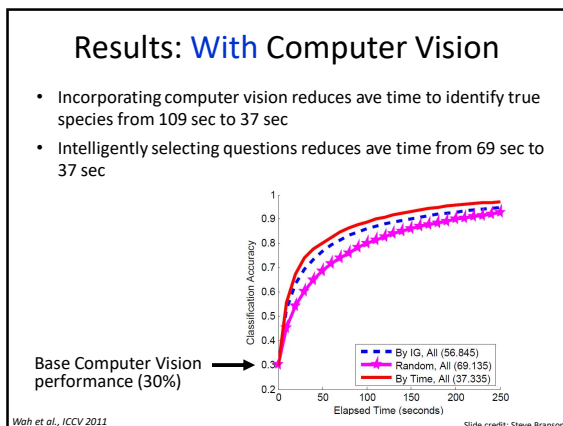






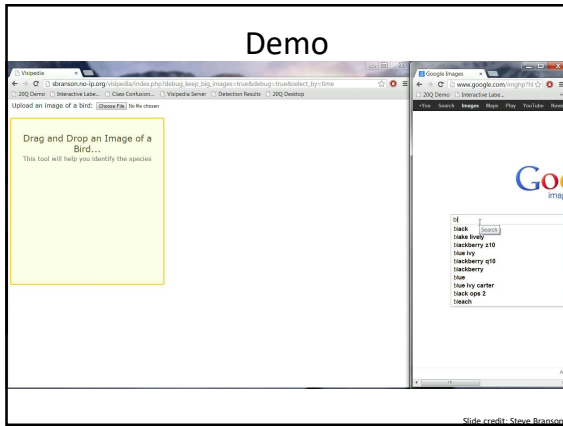
11,877 images, 200 bird species

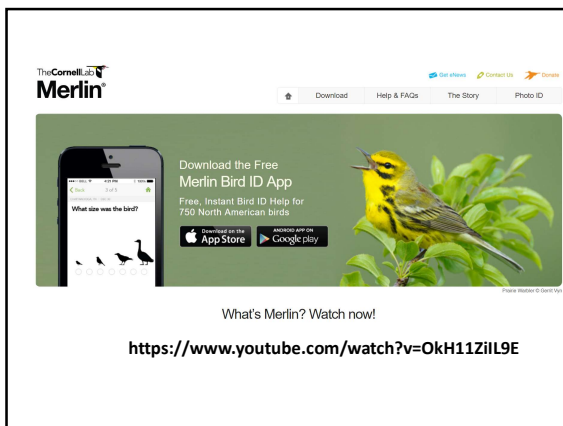



Wah et al., ICCV 2011
Slide credit: Steve Branson









Coming up

- **Tues**
 - Guest lecture by Dr. Suyog Jain
- **Wed**
 - A5 due
- **Thurs**
 - Course wrap-up
 - Applications and frontiers of computer vision
