Scene and Attribute Classification

Danna Gurari
University of Colorado Boulder
Fall 2024

https://dannagurari.colorado.edu/course/recent-advances-in-computer-vision-fall-2024/

Review

- Last lecture: Vision Transformers
  - Motivation
  - ViT architecture
  - ViT training
  - Guidance for student-led lectures

- Assignments (Canvas)
  - Reading assignment was due earlier today
  - Next reading assignments due next Wednesday and Monday
  - Project proposal due in 1.5 weeks

- Questions?

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Scene & Attribute Classification: Today’s Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Image Classification: General Problem

- Given an image, indicate what [fill-in-the-blanks] are in the image

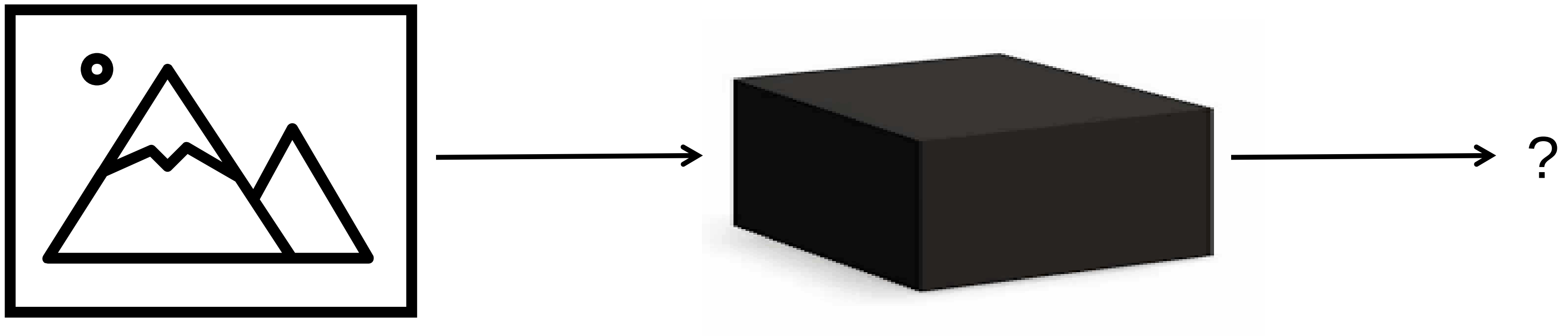
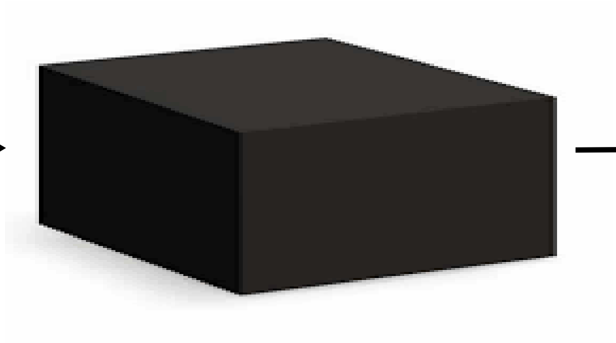


Image Classification: Recall Object Recognition

- Given an image, indicate what objects are in the image

INPUT

OUTPUT

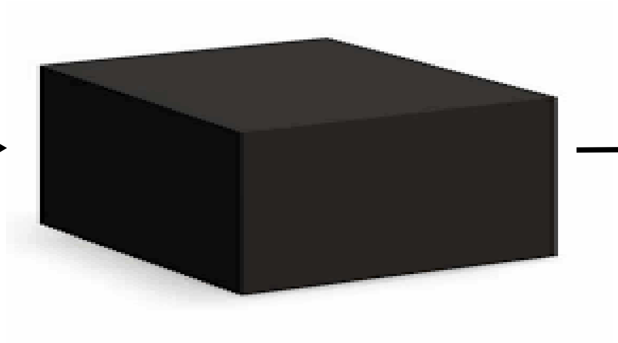


Sunflower

Image Classification: Scene Classification

- Given an image, indicate what scenes are in the image

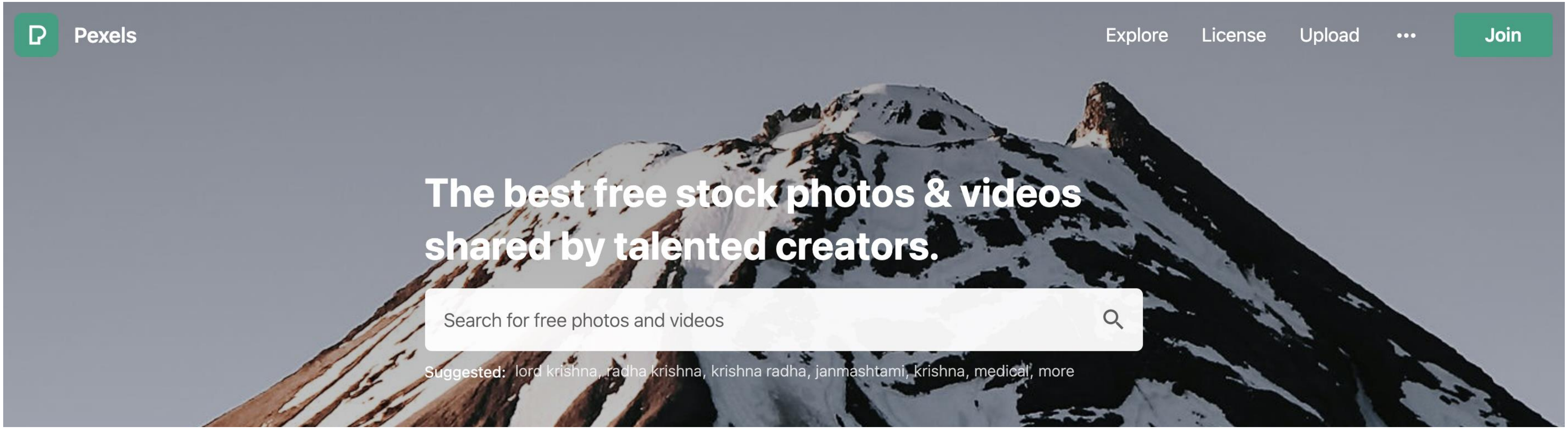
INPUT



OUTPUT

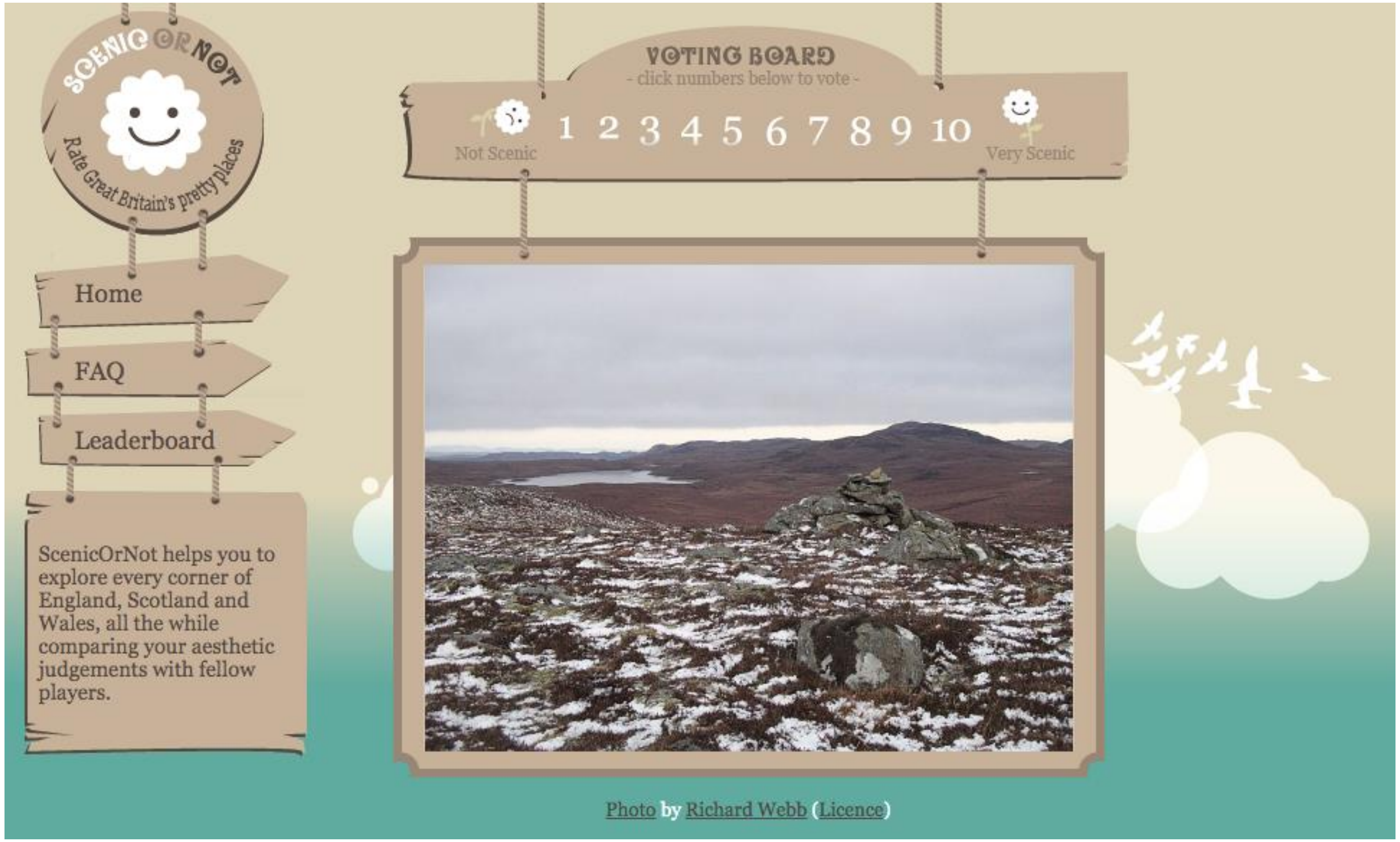
Kitchen

Application: Image Search



Application: Urban Planning

Analyzing correlation of human well-being with scene types

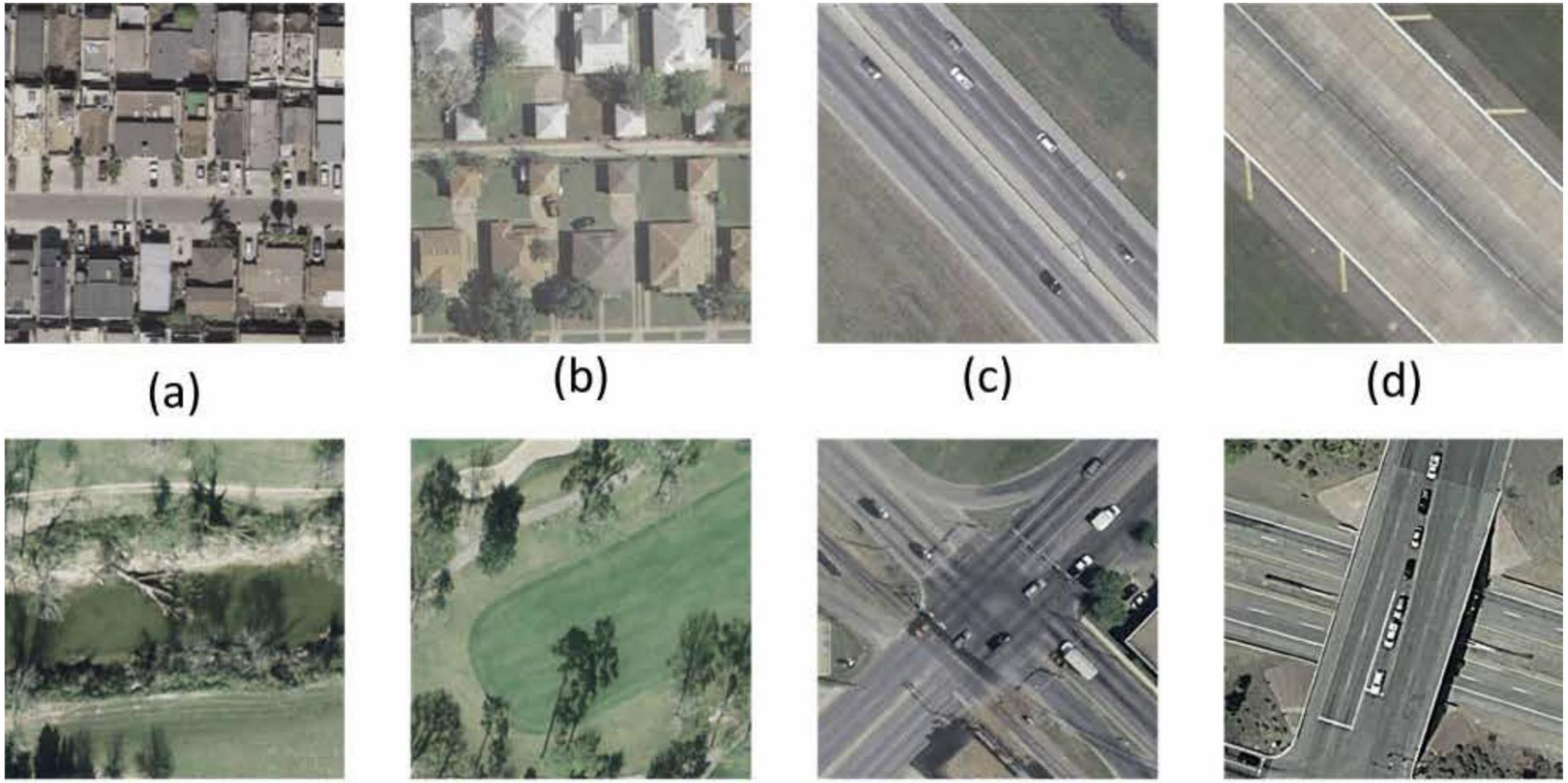


Dataset: http://scenicornot.datasciencelab.co.uk/

Chanuki Illushka Seresinhe et al. Happiness is greater in more scenic locations. *Scientific reports,* 2019.

https://www.economist.com/science-and-technology/2017/07/20/computer-analysis-of-what-is-scenic-may-help-town-planners

Application: Natural Hazard Detection and Environmental Monitoring (via Remote Sensing)



(a) (b) (c) (d)
(e) (f) (g) (h)

Gong Cheng, Junwei Han, and Xiaoqiang Lu. *Proceedings of the IEEE* 2017

What Other Vision Tasks/Applications Can Scene Classification Can Help With?

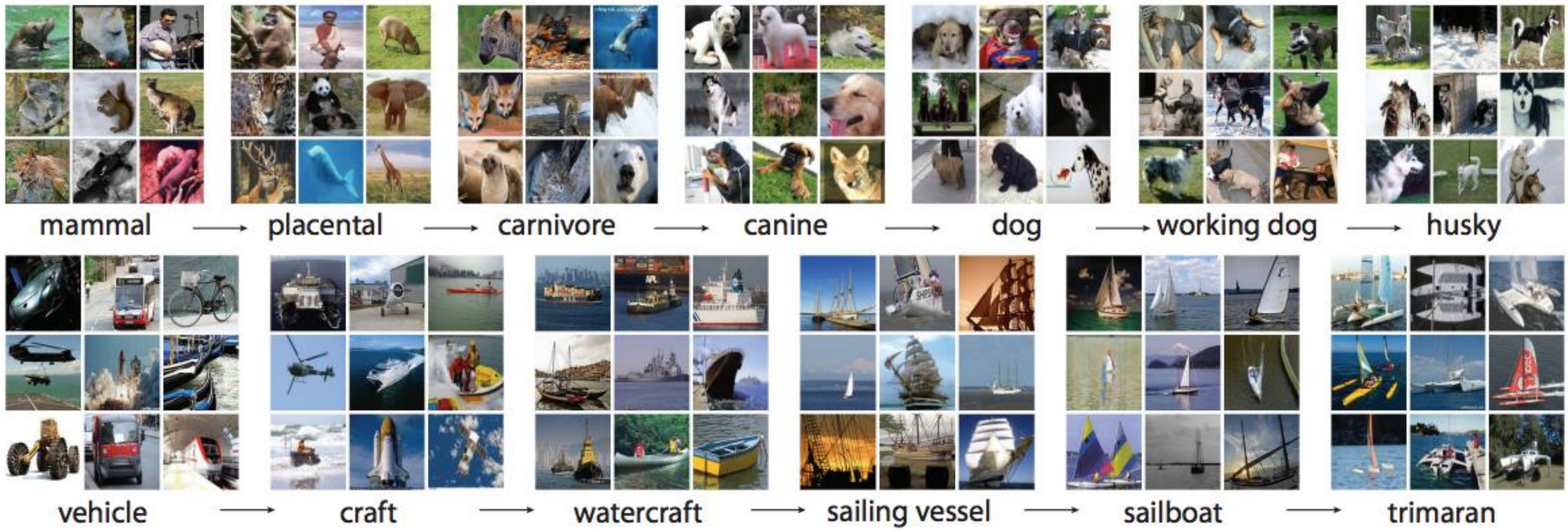
- Object Recognition
  - e.g., What would you expect (or not expect) to find in the scene [now, earlier, later]?
- Activity Recognition/Prediction
  - e.g., What would you expect people to do (or not do) in the scene [now, earlier, later]?

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Motivation for Scene Classification Datasets

What commonality/limitation do you observe for object recognition images (e.g., ImageNet)?



Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009

Motivation for Scene Classification Datasets

What commonality/limitation do you observe for object recognition images (e.g., ImageNet)?

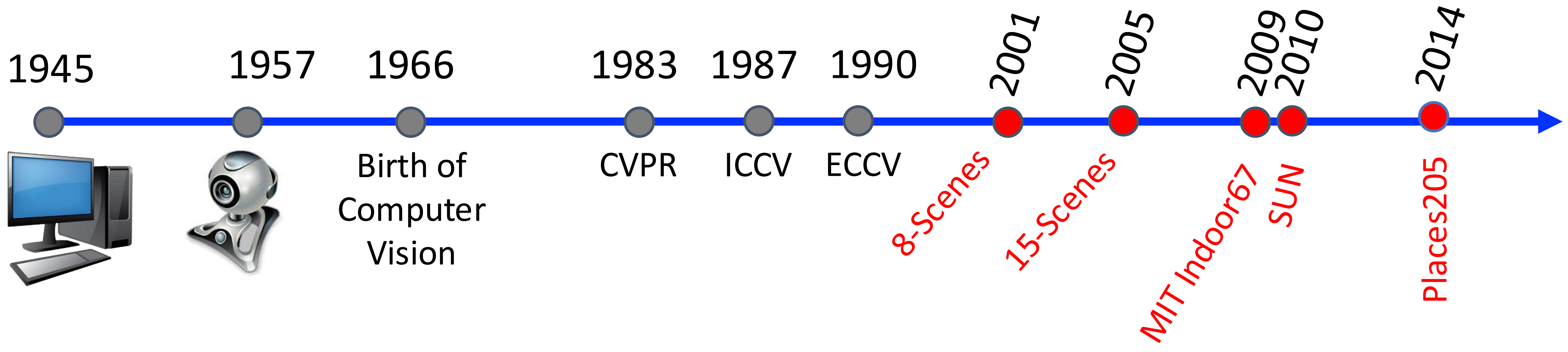
Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009

Motivation for Scene Classification Datasets

Images are **iconic** (i.e., objects are in the center of the images)!

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009

Scene Classification Datasets



# Categories:	8	15	67	397	477
# Images:	300	4,485	15,620	130,519	2,448,873

Trend: build bigger datasets

8-Scenes

Taxonomy Source: unclear

Image Source: COREL stock photo library, personal photographs, Google image search engine

Image Type: 256x256 resolution of roughly even amounts of natural and urban environments

Dataset: https://people.csail.mit.edu/torralba/code/spatialenvelope/

Aude Oliva and Antonio Torralba. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. IJCV 2001

15-Scenes

Taxonomy Source: unclear

Image Source: COREL stock photo library, personal photographs, Google image search engine (contains 8-scenes dataset)

Dataset: https://www.kaggle.com/zaiyankhan/15scene-dataset
Fei Fei Li and Pietro Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. CVPR 2005
Svetlana Labeznik et al. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. CVPR 2005

MIT Indoor67

1. Category Selection

67 categories for 5 domains

Ariadna Quattoni & Antonio Torralba. Recognizing Indoor Scenes. CVPR 2009

MIT Indoor67

1. Category Selection

67 categories for 5 domains

→

2. Image Collection

Images downloaded from 2 image search tools, 1 online photo sharing site, and 1 vision dataset

Ariadna Quattoni & Antonio Torralba. Recognizing Indoor Scenes. CVPR 2009

SUN

1. Category Selection

- From 70,000 categories in "Tiny Images" (WordNet), chose 908 categories describing scenes, places, and environments, excluding:
1) names of specific places (e.g., New York)
2) non-navigable scenes
3) "mature" data

- Extra categories; e.g., mission, jewelry store

Jianxiong Xiao et al. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010

SUN

1. Category Selection

- From 70,000 categories in "Tiny Images" (WordNet), chose 908 categories describing scenes, places, and environments, excluding:
1) names of specific places (e.g., New York)
2) non-navigable scenes
3) "mature" data
- Extra categories; e.g., mission, jewelry store

Category Validation Experiment:

- 7 subjects wrote every 30 minutes the name of the scene category for their location
- All resulting 52 categories were in SUN

Jianxiong Xiao et al. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010

SUN

1. Category Selection

- From 70,000 categories in "Tiny Images" (WordNet), chose 908 categories describing scenes, places, and environments, excluding:
1) names of specific places (e.g., New York)
2) non-navigable scenes
3) "mature" data

- Extra categories; e.g., mission, jewelry store

2. Image Collection

- Downloaded from search engines

- Automatically discarded images that are:
  1) not color
  2) less than 200x200
  3) very blurry or noisy
  4) aerial views
  5) duplicates

(Adapted from slides by Antonio Torralba)

Jianxiong Xiao et al. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010

SUN

1. Category Selection

- From 70,000 categories in "Tiny Images" (WordNet), chose 908 categories describing scenes, places, and environments, excluding:
1) names of specific places (e.g., New York)
2) non-navigable scenes
3) "mature" data
- Extra categories; e.g., mission, jewelry store

→

2. Image Collection

- Downloaded from search engines

- Automatically discarded images that are:
  1) not color
  2) less than 200x200
  3) very blurry or noisy
  4) aerial views
  5) duplicates

→

3. Human Verification

- 9 in-house people reviewed & discarded irrelevant images

- Result is 130,519 images spanning 397 categories with >99 images per category

Jianxiong Xiao et al. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010

SUN

3. Human Verification

- 9 in-house people reviewed & discarded irrelevant images

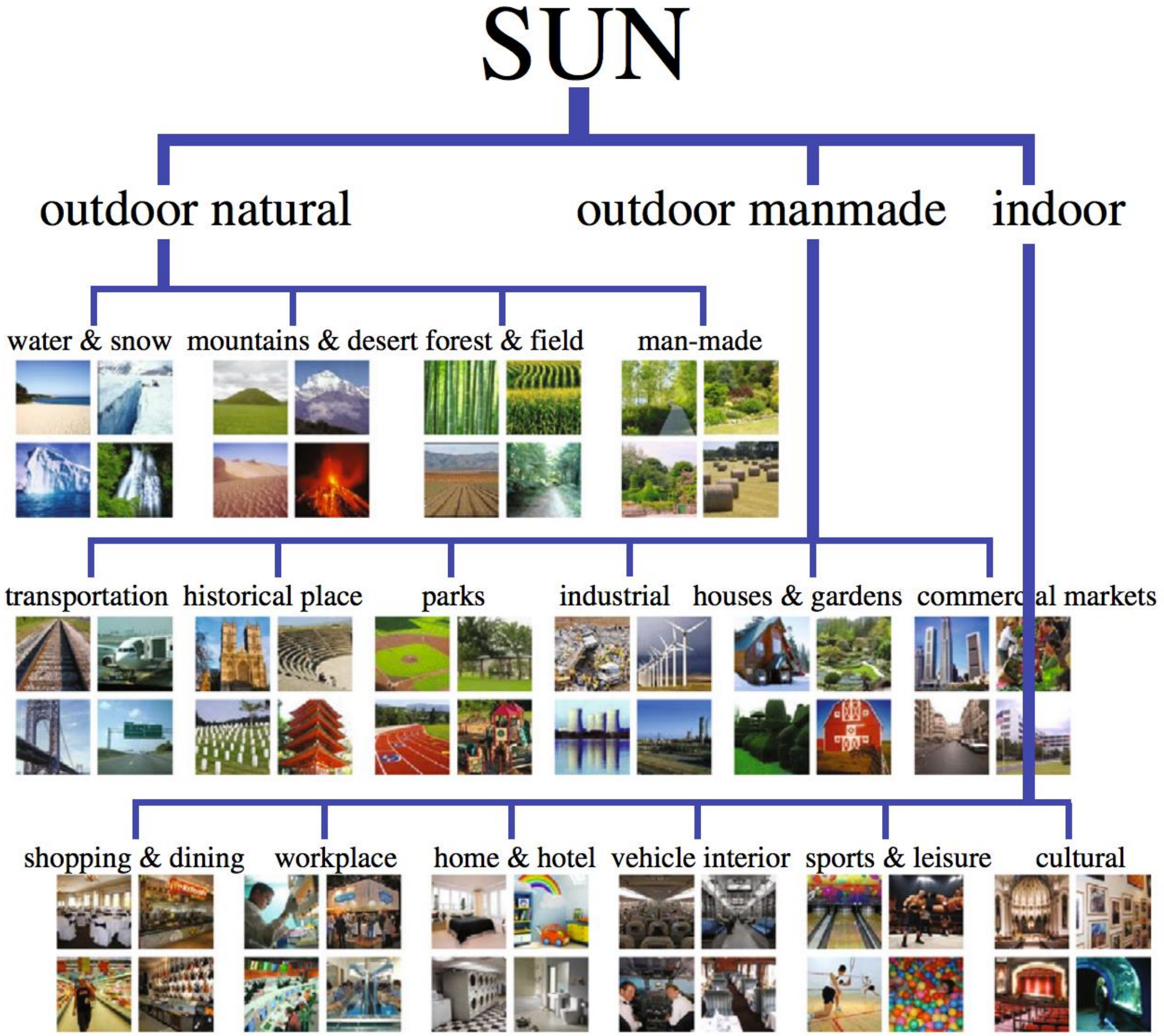
- Result is 130,519 images spanning 397 categories with >99 images per category

Jianxiong Xiao et al. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010

Places205

1. Category Selection

Same taxonomy as SUN



Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Places205

1. Category Selection

Same taxonomy as SUN

2. Image Collection

- Downloaded images from three search engines; query terms were 696 common adjectives (messy, spare, sunny, desolate, etc) with each scene category

- Automatically discarded images that are:
  1) not color
  2) less than 200x200

bing

Google Image Search

flickr GAMMA

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Places205

1. Category Selection

Same taxonomy as SUN

2. Image Collection

- Downloaded images from three search engines; query terms were 696 common adjectives (messy, spare, sunny, desolate, etc) with each scene category

- Automatically discarded images that are:
  1) not color
  2) less than 200x200

3. Human Verification

- AMT crowd workers identified (ir)relevant images for batches of 750 images

- Result is 7,076,580 images spanning 476 categories

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

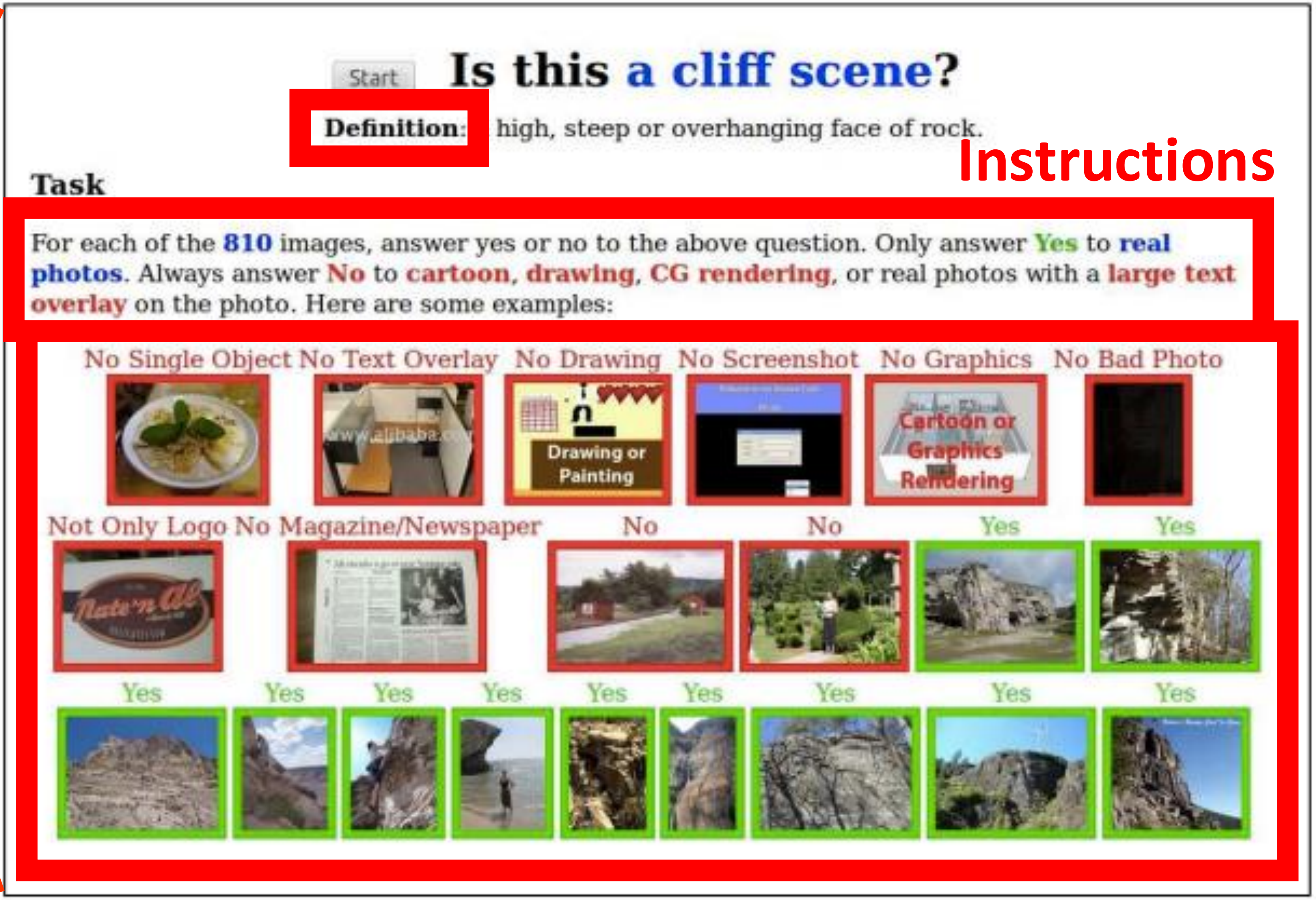
Places205

User interface: Instructions

1. Task Design

Instructions:

Interface:



Examples

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Places205

User interface: Task

1. Task Design

Instructions:

Interface:

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Places205

1. Task Design

Instructions:

Interface:

2. Crowdsourcing Platform

amazon mechanical turk™
Artificial Artificial Intelligence

Places205

1. Task Design

Instructions:

Interface:

2. Crowdsourcing Platform

amazon mechanical turk™
Artificial Artificial Intelligence

3. Quality Control

- Run images through crowd twice with default "yes" and then default "no" answer

- "Honeypot"
  - labelled at least 90% on control set correctly, where it includes 30 known positive and negative labelled images per "HIT"

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Places205 Summary

1. Category Selection

Same taxonomy as SUN

→

2. Image Collection

- Downloaded images from three search engines; query terms were 696 common adjectives (messy, spare, sunny, desolate, etc) with each scene category

- Automatically discarded images that are:
  1) not color
  2) less than 200x200

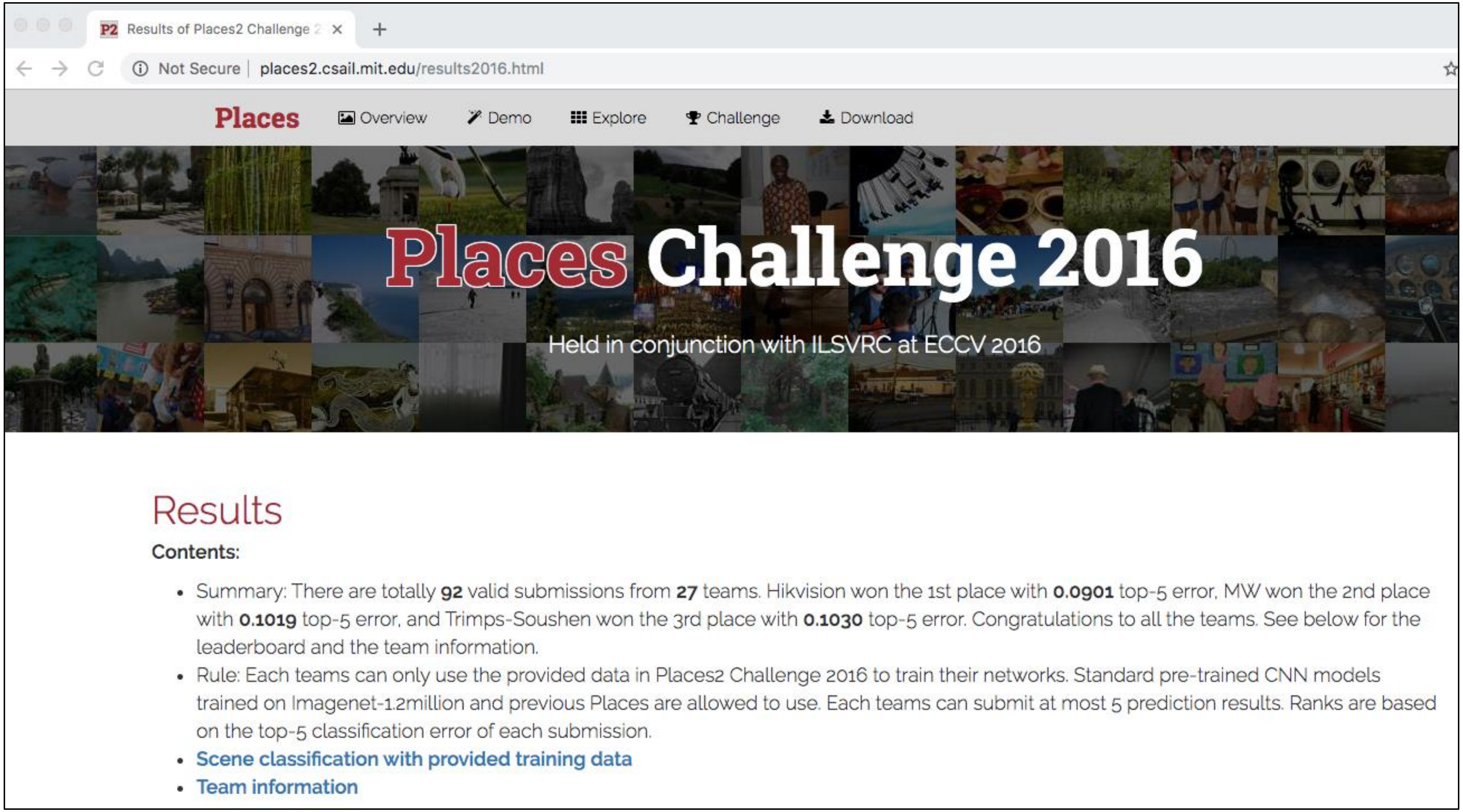
→

3. Human Verification

- AMT crowd workers identified (ir)relevant images for batches of 750 images

- Result is 7,076,580 images spanning 476 categories

Scene Classification: Places Challenge



Results

Contents:

- Summary: There are totally **92** valid submissions from **27** teams. Hikvision won the 1st place with **0.0901** top-5 error, MW won the 2nd place with **0.1019** top-5 error, and Trimps-Soushen won the 3rd place with **0.1030** top-5 error. Congratulations to all the teams. See below for the leaderboard and the team information.
- Rule: Each teams can only use the provided data in Places2 Challenge 2016 to train their networks. Standard pre-trained CNN models trained on Imagenet-1.2million and previous Places are allowed to use. Each teams can submit at most 5 prediction results. Ranks are based on the top-5 classification error of each submission.
- **Scene classification with provided training data**
- **Team information**

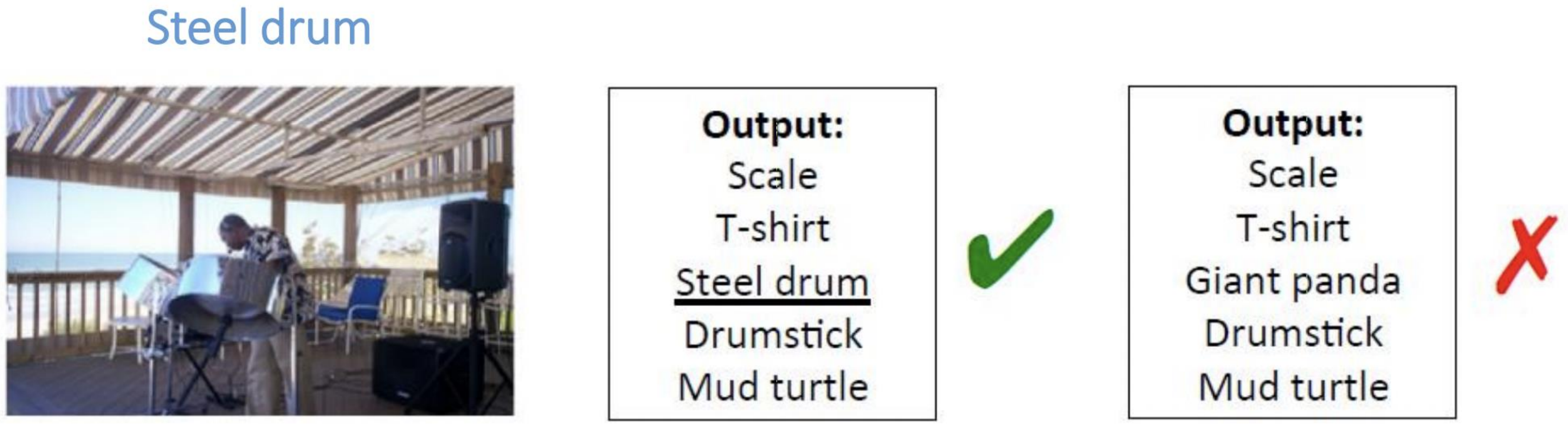
Evaluation: Metric Used for ImageNet

Assumption: 1 ground truth label per image

Error is average over all test images using this rule per image:

* 0 if any predictions match the ground truth
* 1 otherwise

e.g., top 5 error

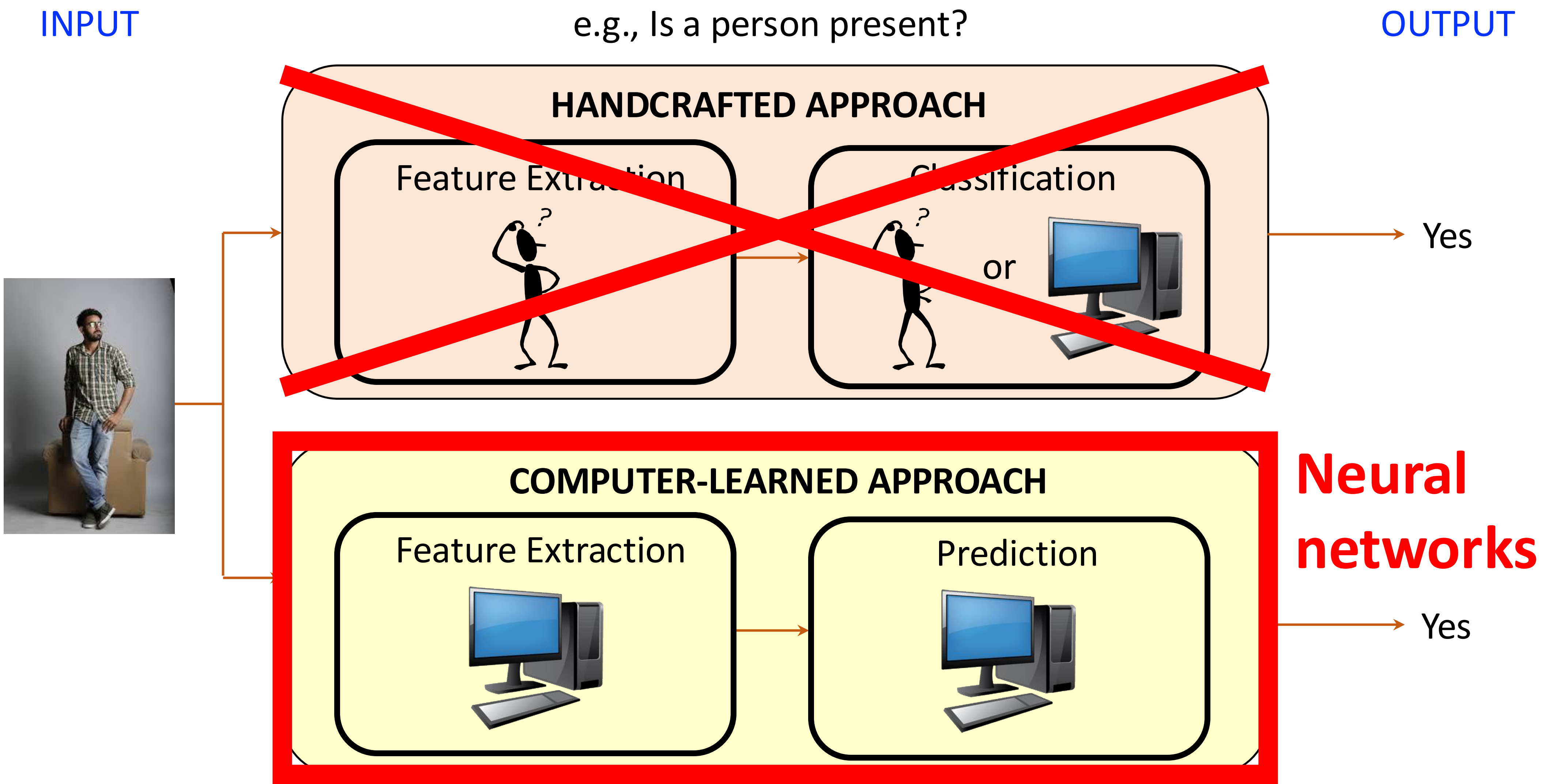


Source: https://image-net.org/static_files/files/ILSVRC2017_overview.pdf

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Recall Computer Vision Revolution: Algorithm Design Shifted from Handcrafted to Computer-Learned Rules



Key Idea: Establish Good “Deep Features”

Key challenge is finding features that generalize across datasets and tasks

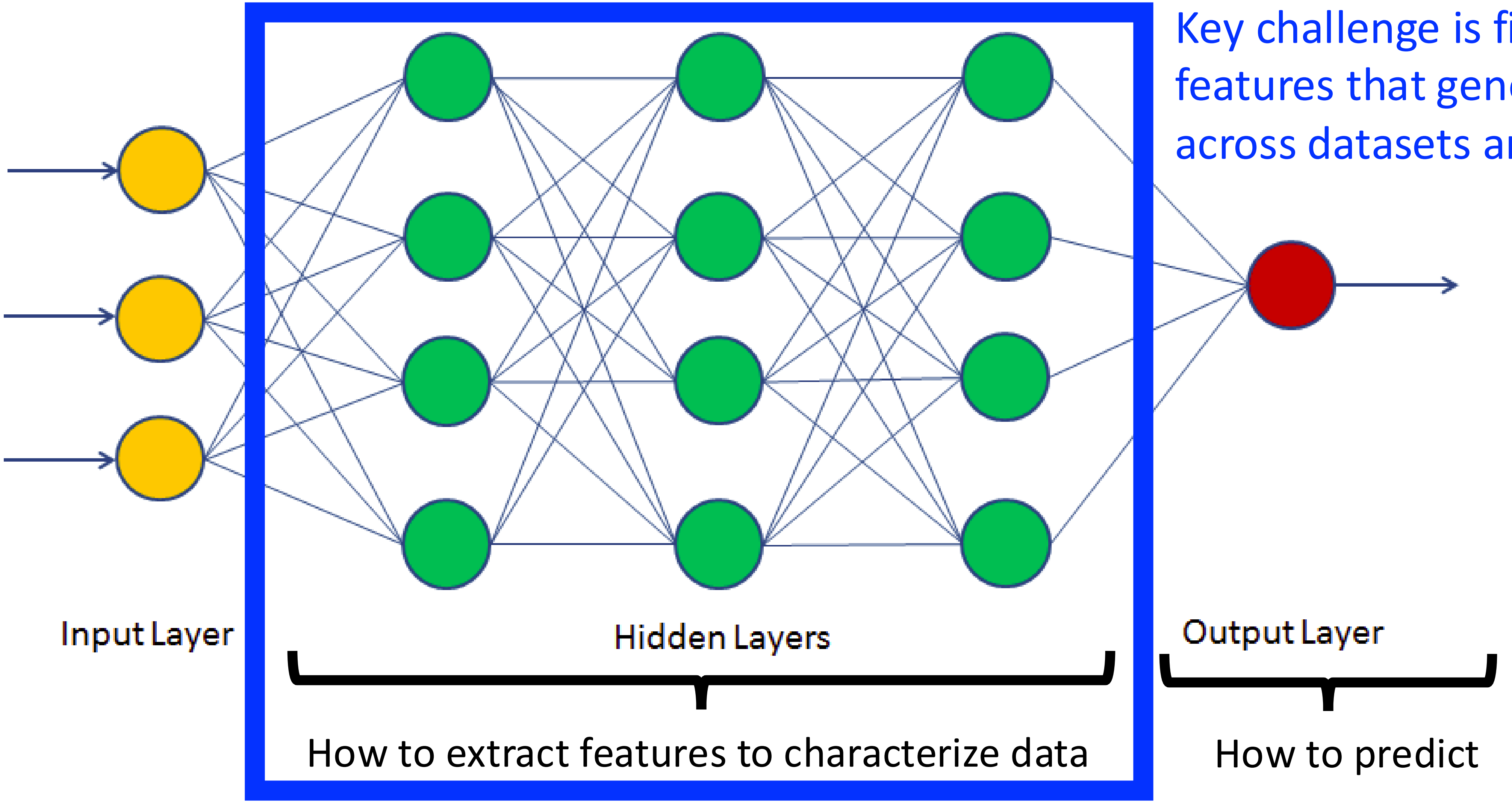


Figure Source: https://www.datacamp.com/community/tutorials/neural-network-models-r

Approach (Step 1): Train AlexNet on a Scenes-Based Dataset

- Prior work: trained on ImageNet (~1.5 million images of objects scraped from search engines)

Deng et al. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009.

- Proposal: train on Places (~2.5 million images of scenes scraped from search engines)

Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NeurIPS 2014.

Approach (Step 2): Train SVM classifiers Using Deep Features Extracted from FC7 Layer

- What is the dimensionality of the fc7 feature?

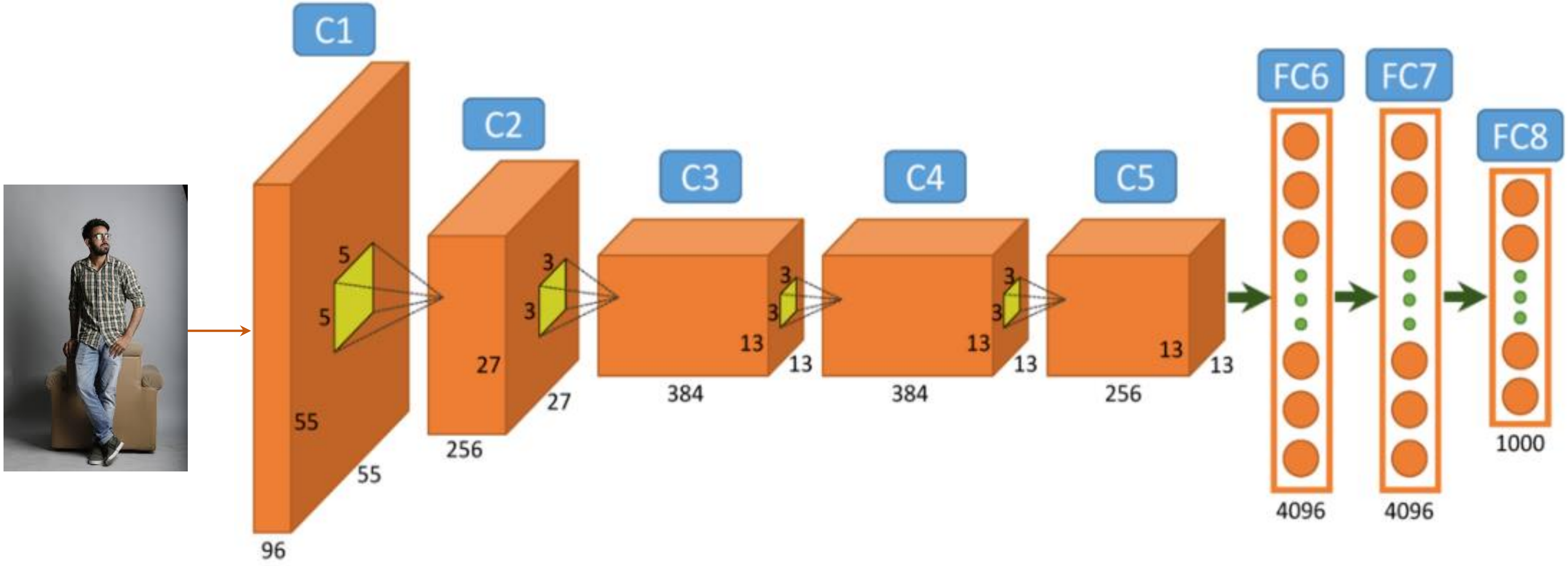


Image Source: https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers_fig2_312303454

Performance Comparison When Using Features Extracted from Two AlexNet Models

Scene classification datasets | Object recognition datasets

	SUN397	MIT Indoor67	Scene15	Caltech101	Caltech256
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	**87.22±0.92**	**67.23±0.27**

What trends do you see?

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Performance Comparison When Using Features Extracted from Two AlexNet Models

Places training data better for scene classification datasets!

ImageNet training data better for object recognition datasets!

	SUN397	MIT Indoor67	Scene15	Caltech101	Caltech256
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	**87.22±0.92**	**67.23±0.27**

State-of-the-art performance at the time

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Performance Comparison When Using Features Extracted from Two AlexNet Models

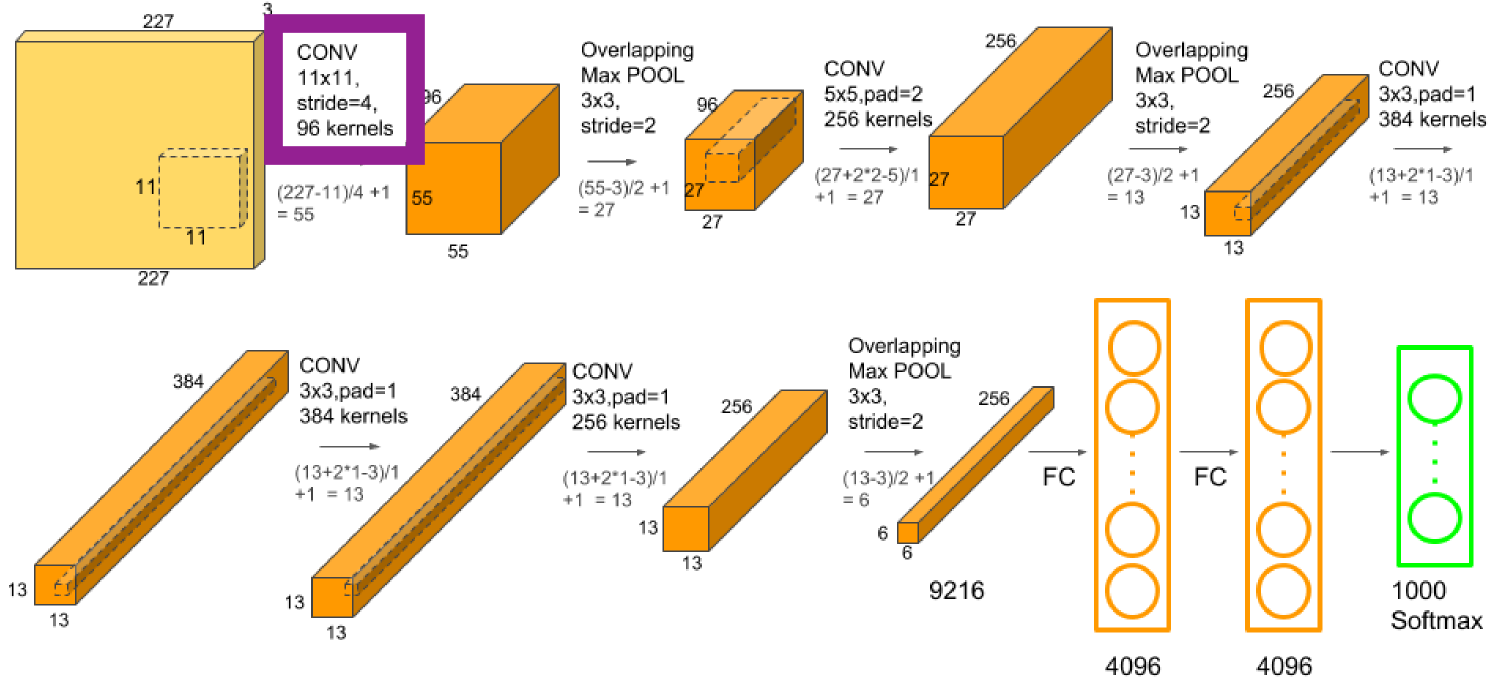
	SUN397	MIT Indoor67	Scene15	Caltech101	Caltech256
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	**87.22±0.92**	**67.23±0.27**
Feature from AlexNet trained on both datasets	53.86±0.21	**70.80**	**91.59±0.48**	84.79±0.66	65.06±0.25

Using MORE training data can diminish the benefit of the deep features; Why?

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Comparing Representations Learned When Training AlexNet on Different Datasets

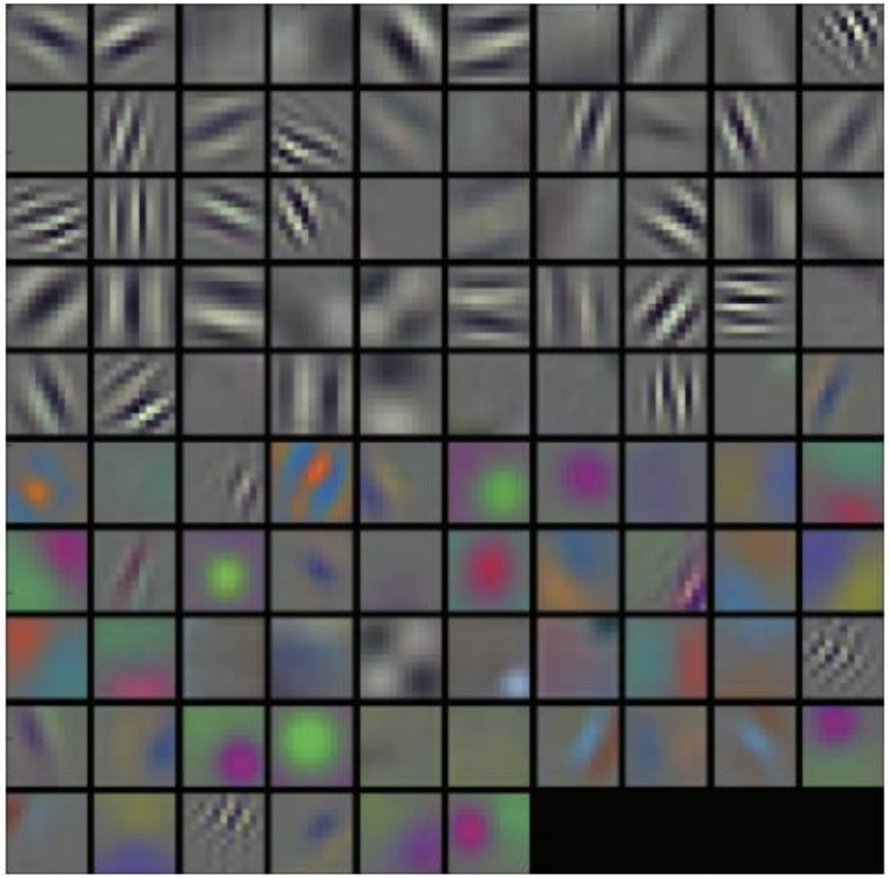
- Dataset 1: ImageNet (~1.5 million images of objects scraped from search engines)
- Dataset 2: Places (~2.5 million images of scenes scraped from search engines)



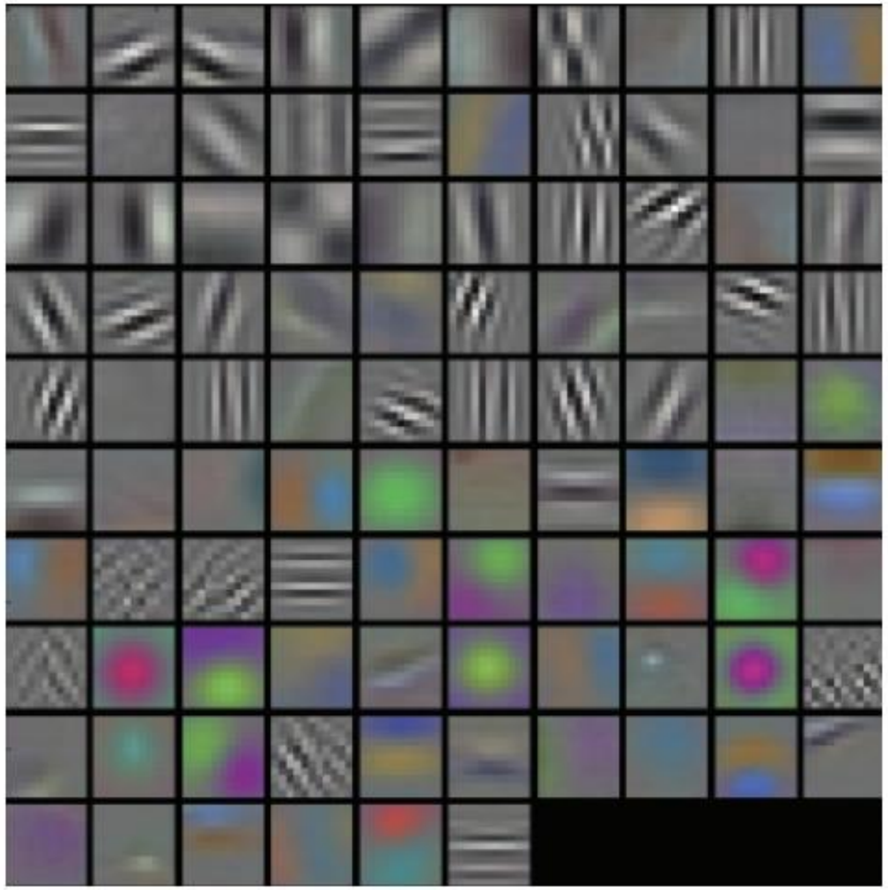
Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

Comparing Representations Learned When Training AlexNet on Different Datasets

ImageNet-CNN



Places-CNN

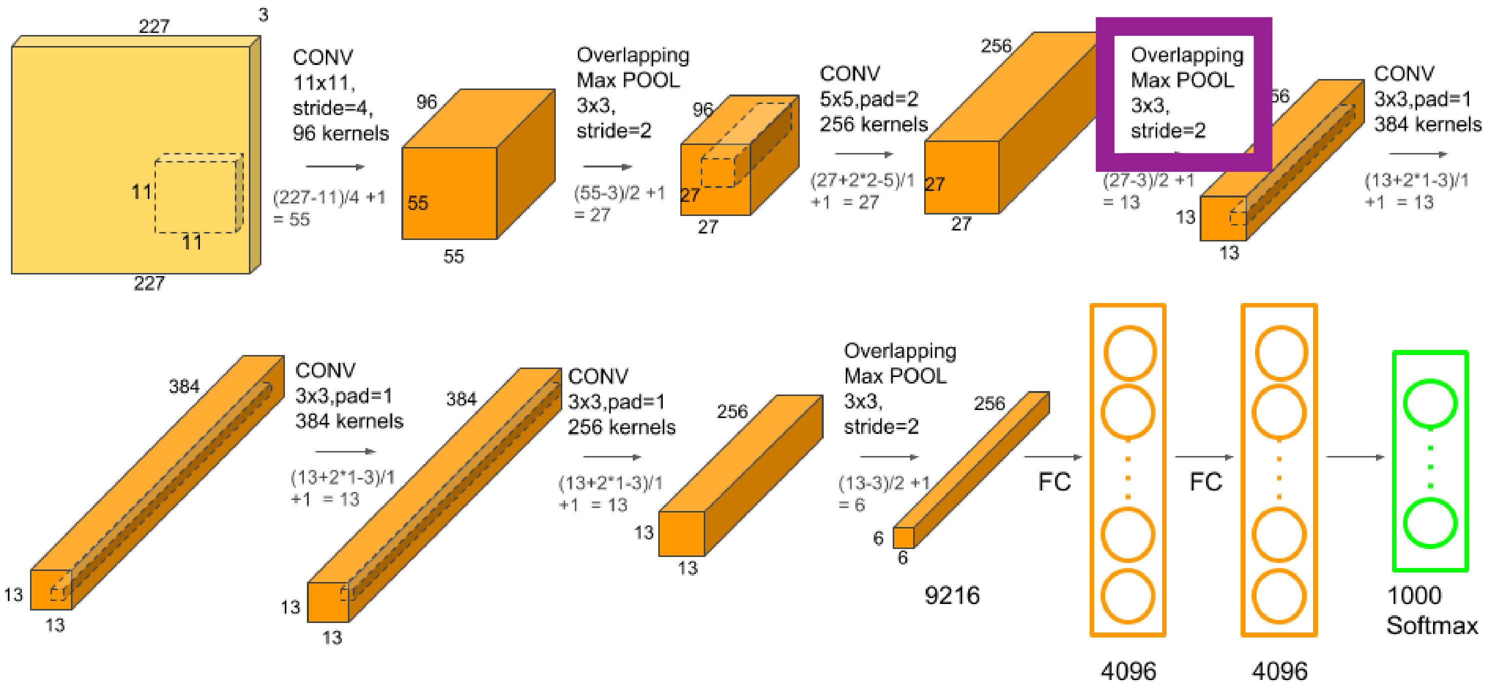


Do filters learned from the different datasets look similar or different?

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Comparing Representations Learned When Training AlexNet on Different Datasets

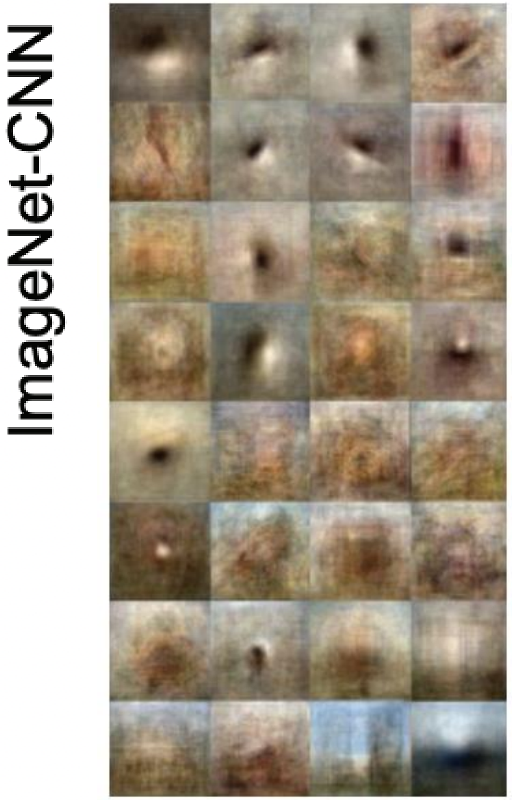
- Dataset 1: ImageNet (~1.5 million images of objects scraped from search engines)
- Dataset 2: Places (~2.5 million images of scenes scraped from search engines)

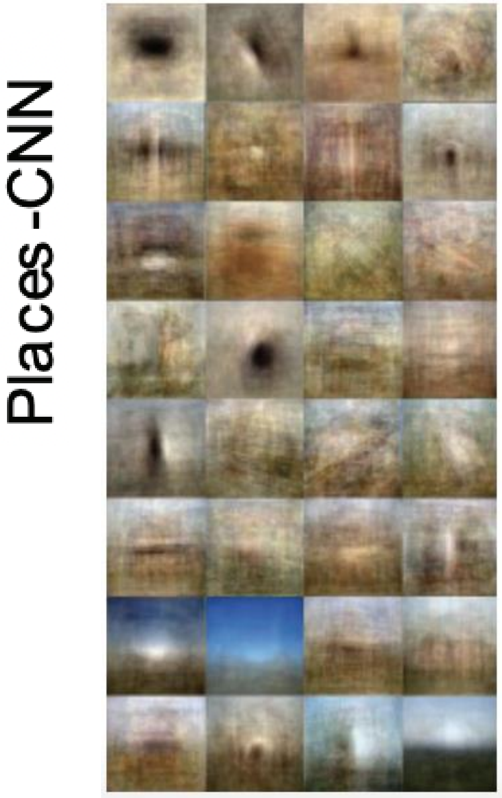


Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

Comparing Representations Learned When Training AlexNet on Different Datasets

Result from singling out different units in the neural networks and then generating the mean image from the 100 images which fire the most (i.e., highest activation scores)



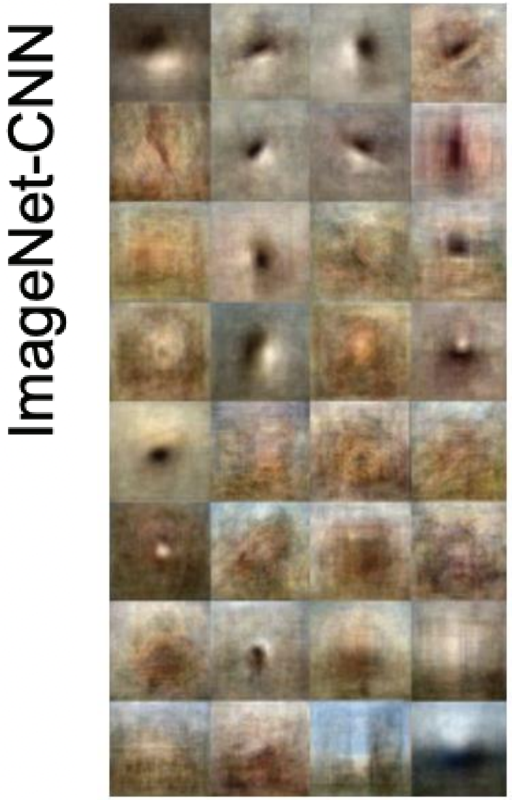


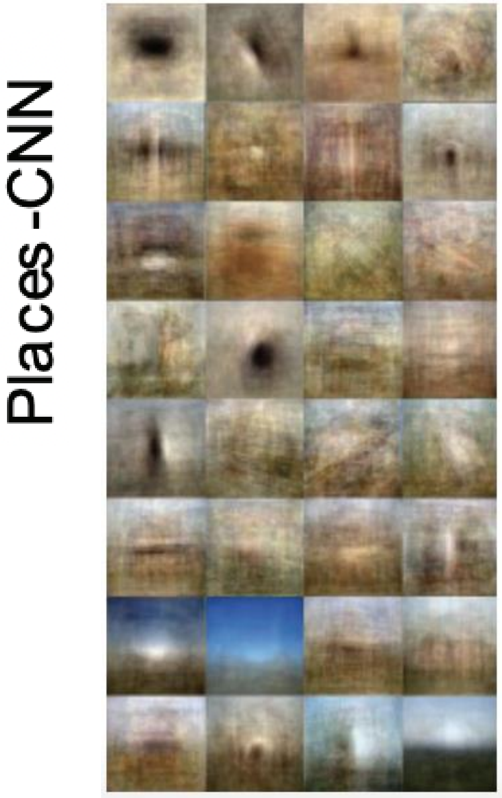
Do the representations from the different datasets appear to be similar or different?

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Comparing Representations Learned When Training AlexNet on Different Datasets

Result from singling out different units in the neural networks and then generating the mean image from the 100 images which fire the most (i.e., highest activation scores)



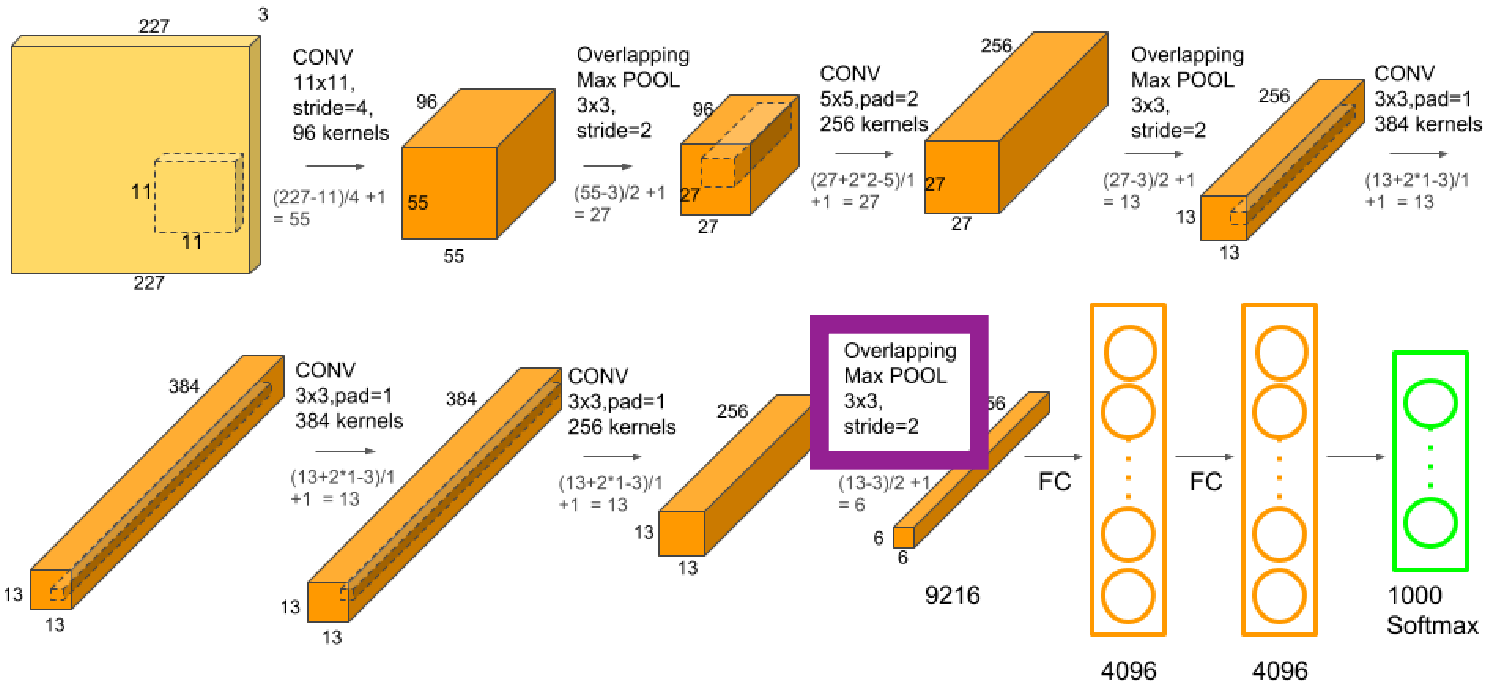


ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Comparing Representations Learned When Training AlexNet on Different Datasets

- Dataset 1: ImageNet (~1.5 million images of objects scraped from search engines)
- Dataset 2: Places (~2.5 million images of scenes scraped from search engines)

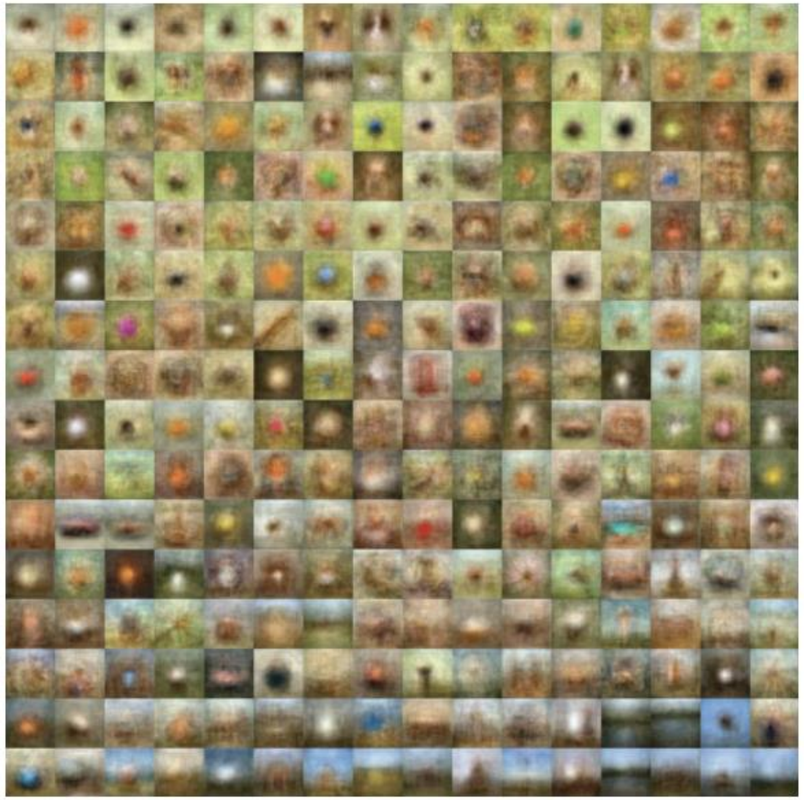


Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

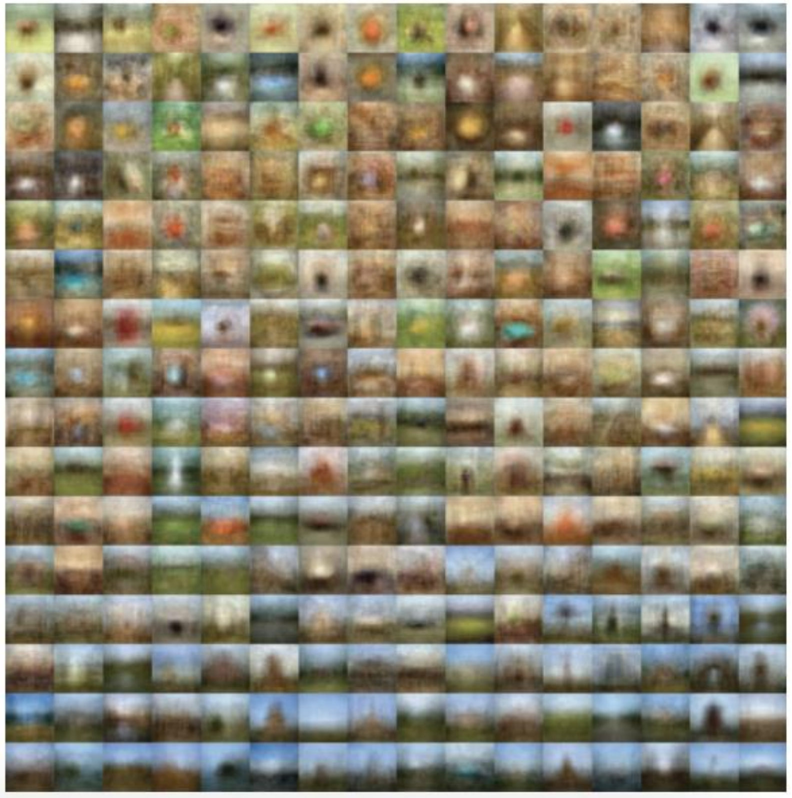
Comparing Representations Learned When Training AlexNet on Different Datasets

Result from generating the mean image from the 100 images which fire the most for a given unit in the neural network (i.e., highest activation scores)

ImageNet-CNN



Places-CNN

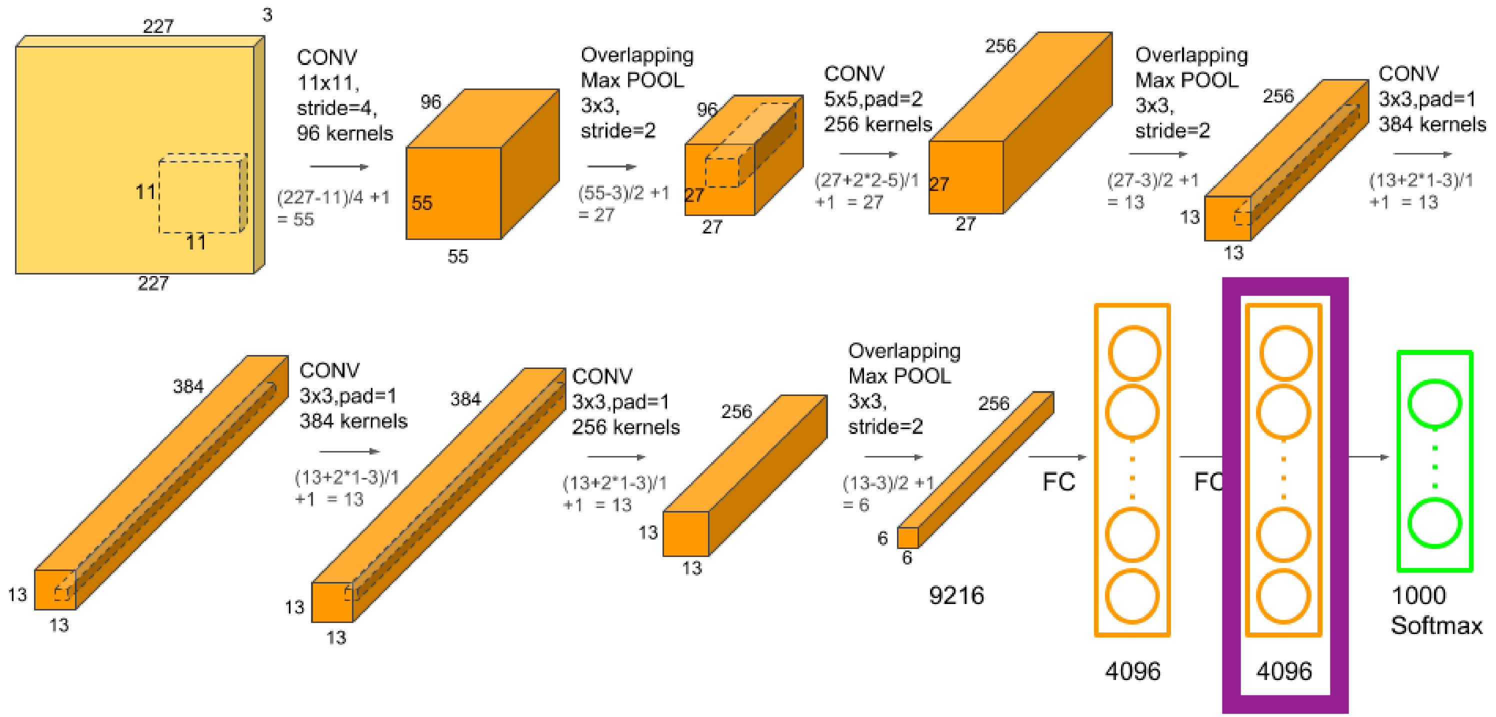


ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

Comparing Representations Learned When Training AlexNet on Different Datasets

- Dataset 1: ImageNet (~1.5 million images of objects scraped from search engines)
- Dataset 2: Places (~2.5 million images of scenes scraped from search engines)

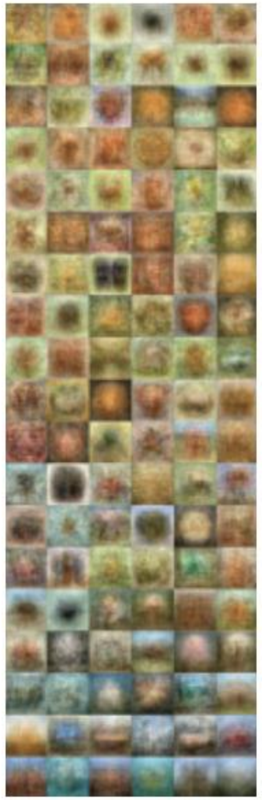


Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

Comparing Representations Learned When Training AlexNet on Different Datasets

Result from generating the mean image from the 100 images which fire the most for a given unit in the neural network (i.e., highest activation scores)

ImageNet-CNN



Places-CNN

ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

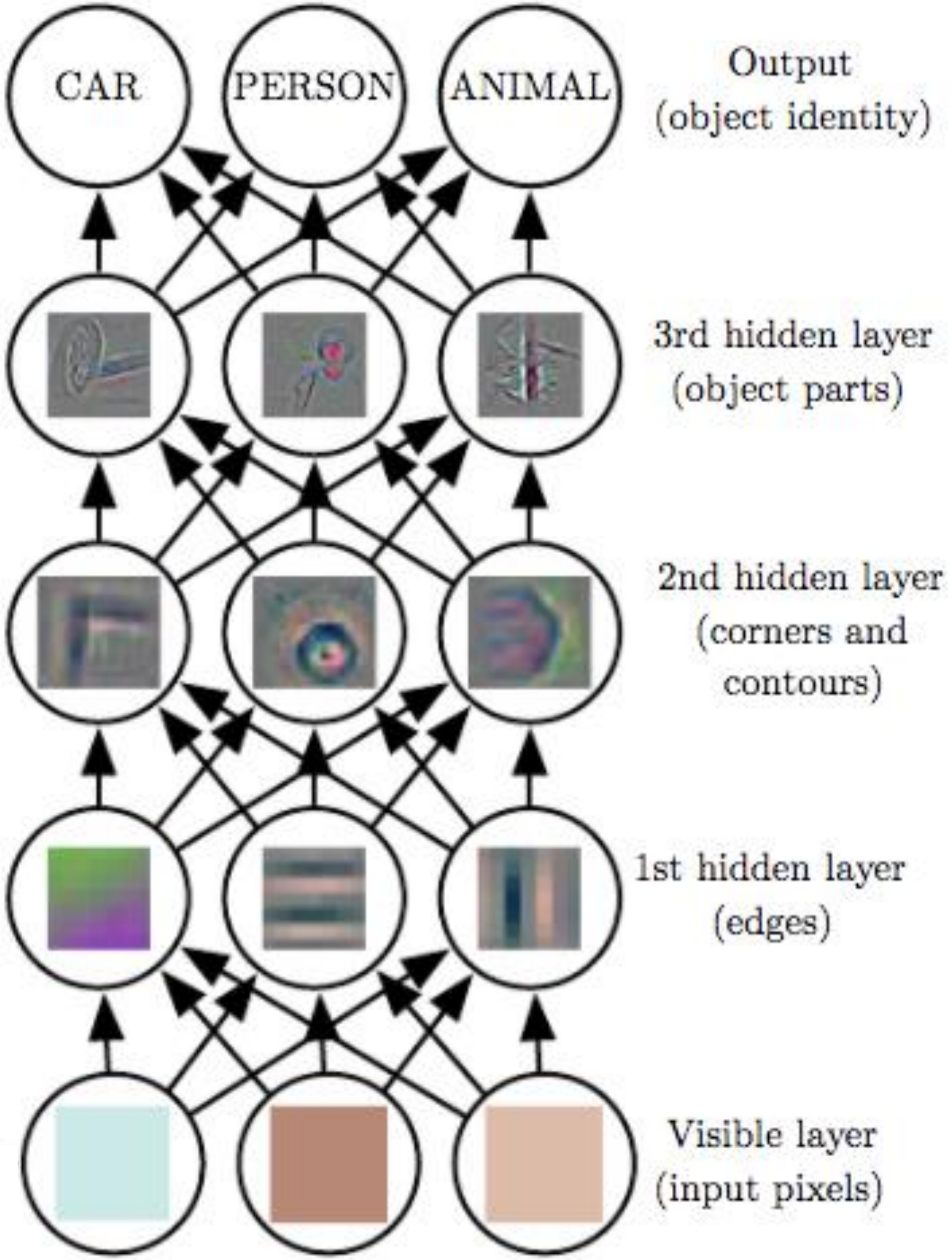
Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014

CNN: Common Representations

High-level feature maps

Mid-level feature maps

Low-level feature maps



Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Summary: Relevant Training Data is Key to Learn Good Deep Features for Downstream Tasks

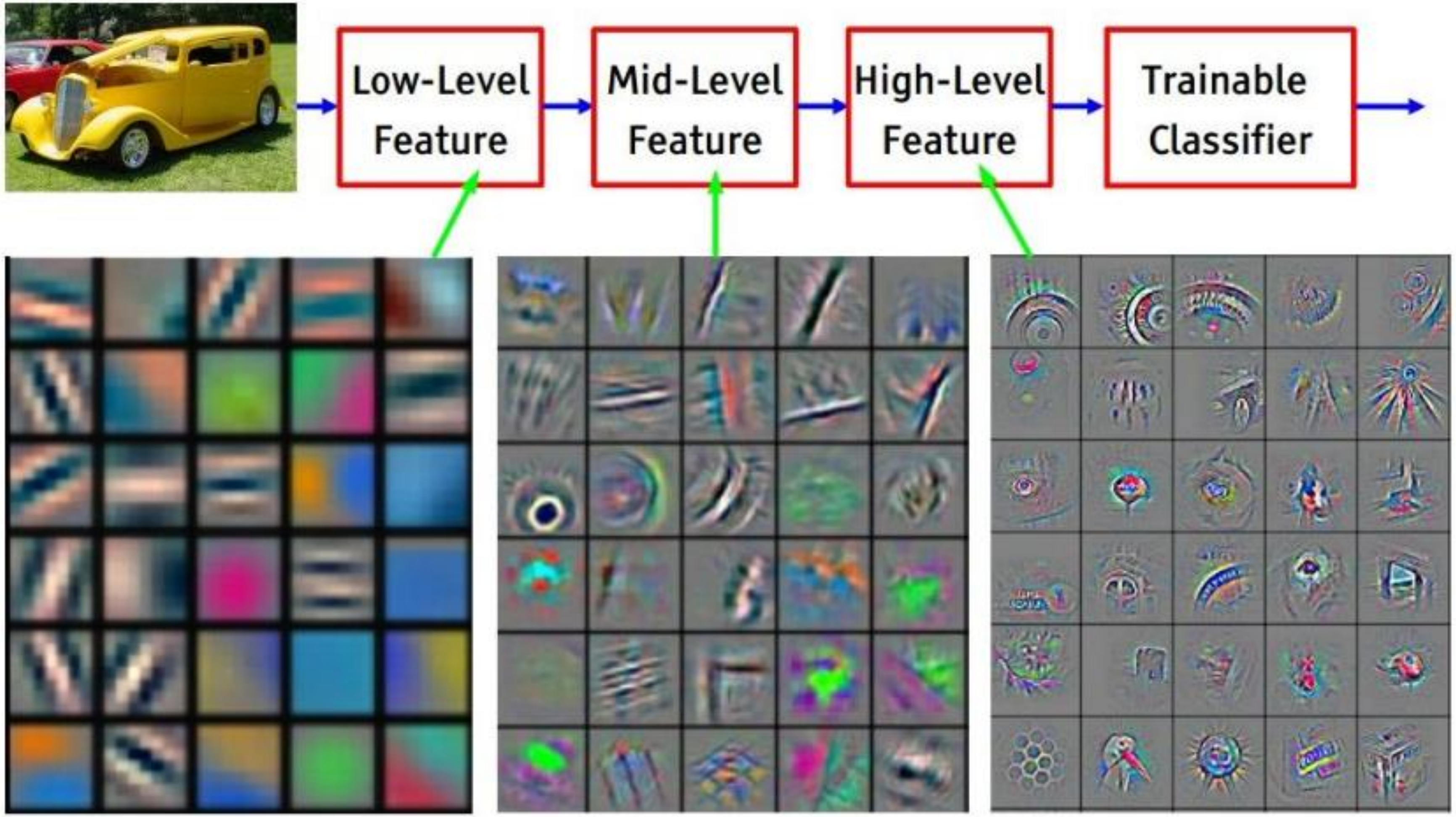


Figure Credit: Yann LeCun

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Attribute Definition

Description

(as opposed to naming)

How would you describe this object?

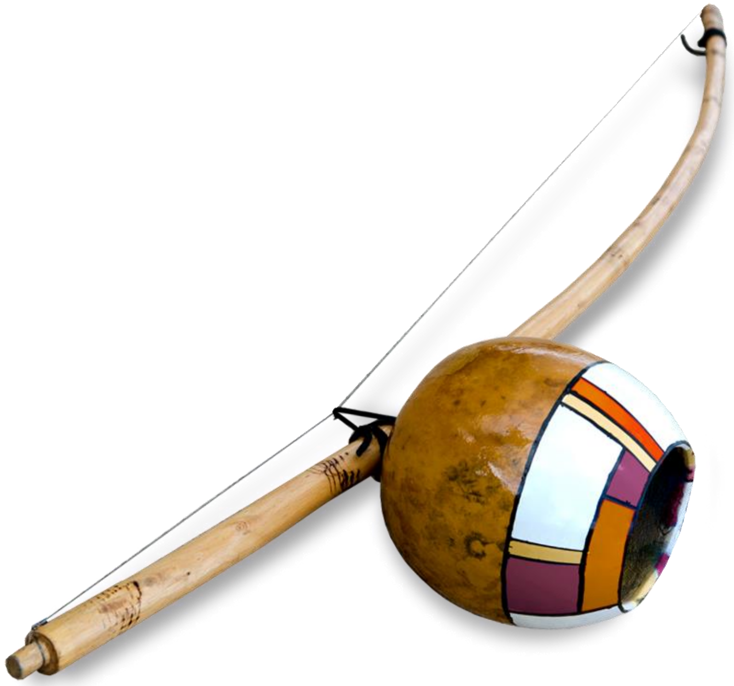
Attribute Definition

* Learning 30,000 objects equates to a person learning ~4.5 objects per day every day for 18 years

* Can be easier to "describe" than to "name" the unknown

Description

(as opposed to naming)



How would you describe this object?

Attribute Definition

Description

(as opposed to naming)

How would you describe this scene?

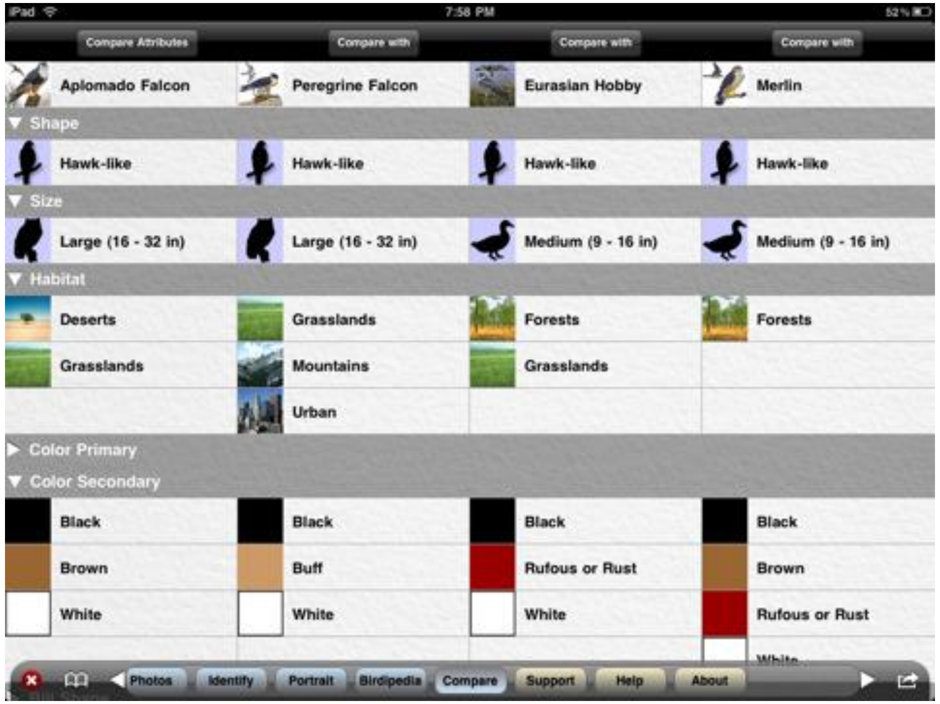
Relative Attributes (Rather Than **Categorical**)

Attributes can have a *spectrum* of strengths; e.g.,

Aron Y and Kristen Grauman. Just Noticeable Differences in Visual Attributes. ICCV 2015

Application: Bird Recognition

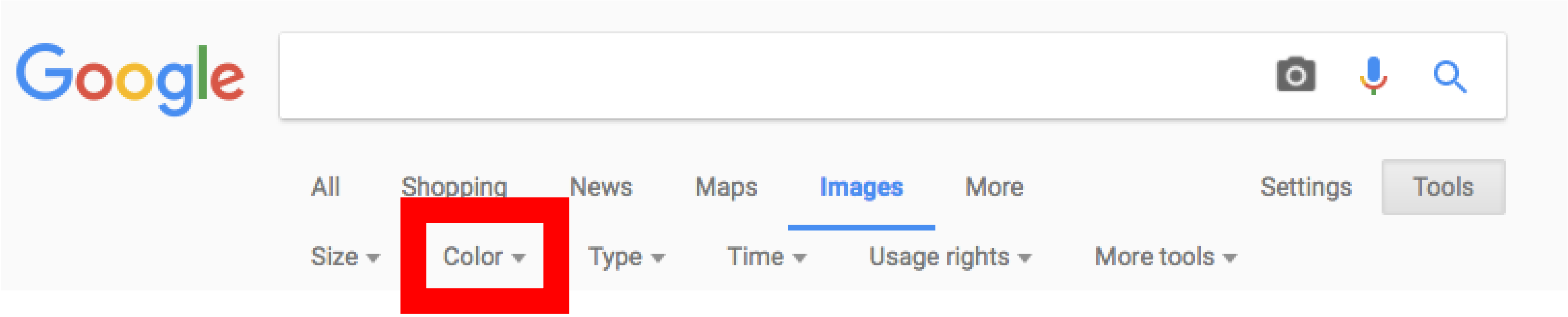
e.g., recognize objects with common knowledge instead of expert knowledge



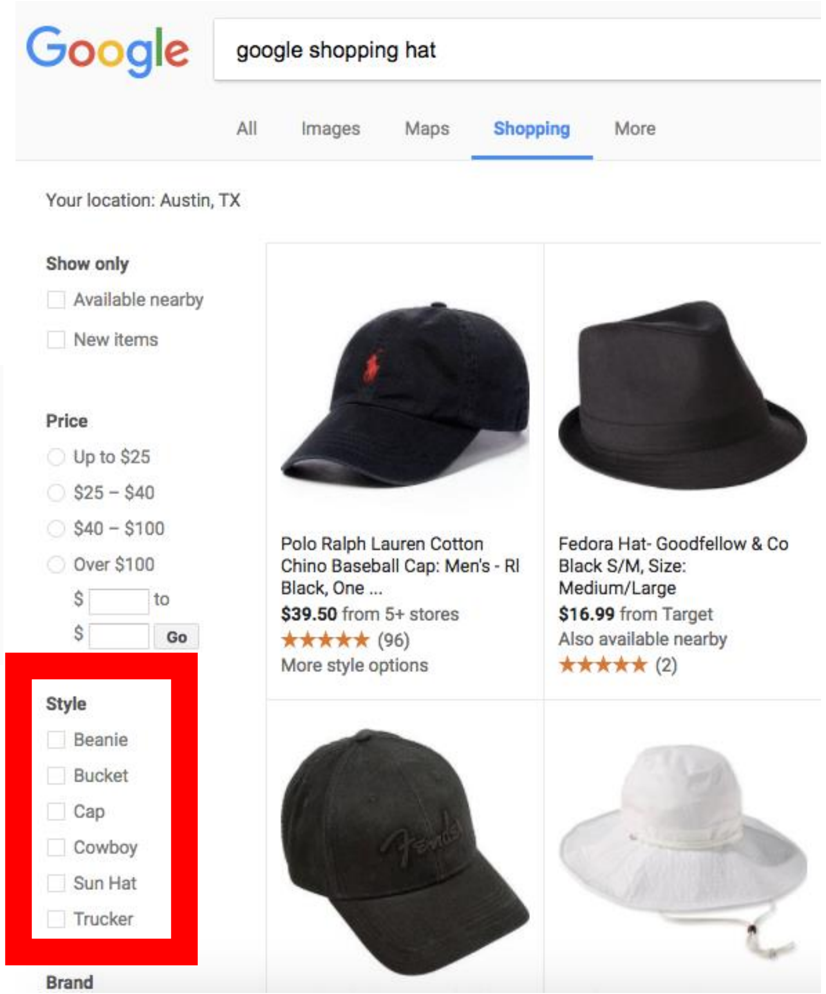
e.g., iBird: describe a bird to learn what type it is

Demo: https://www.youtube.com/watch?v=J1C-Q-z_np0

Application: Expedite Search

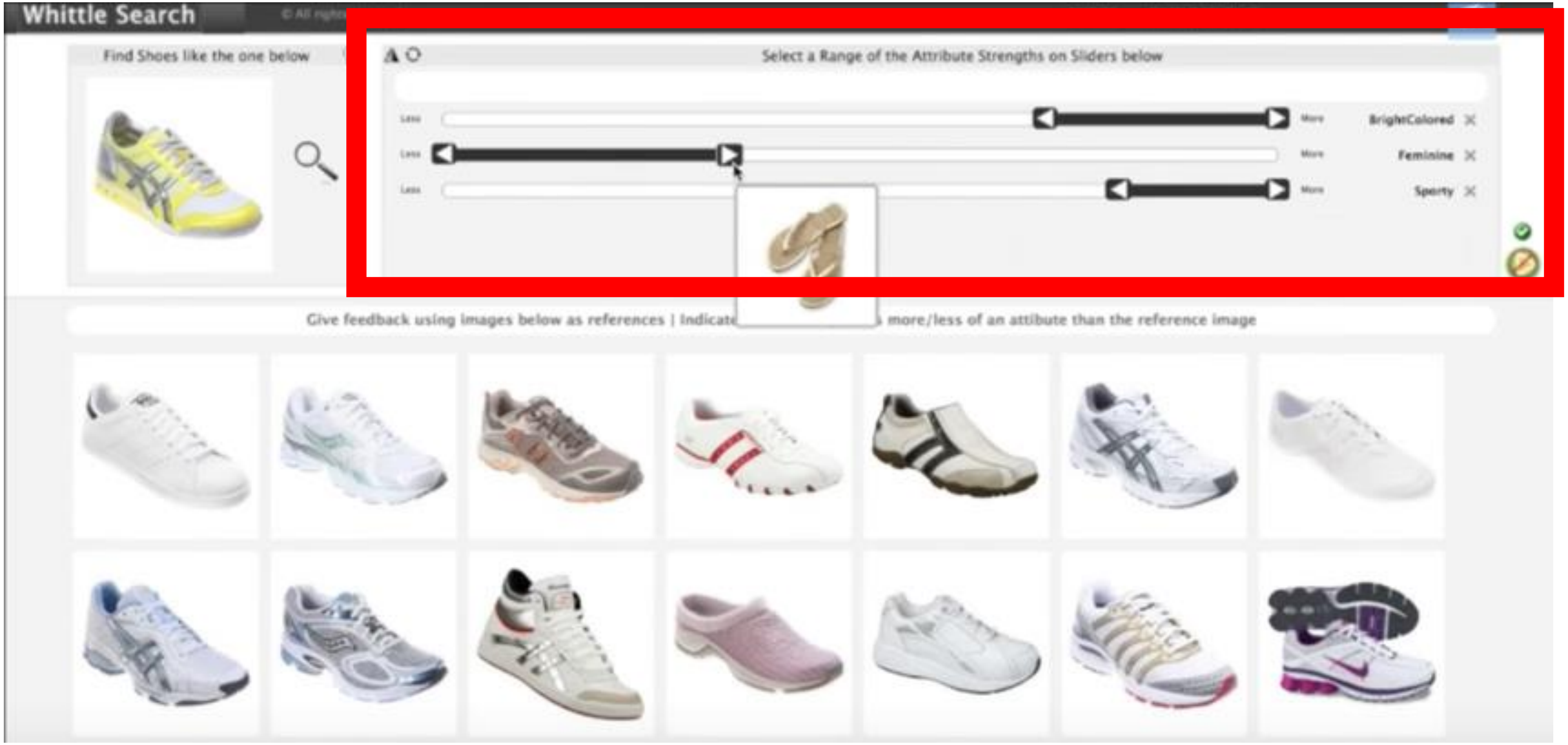


e.g., Image Search



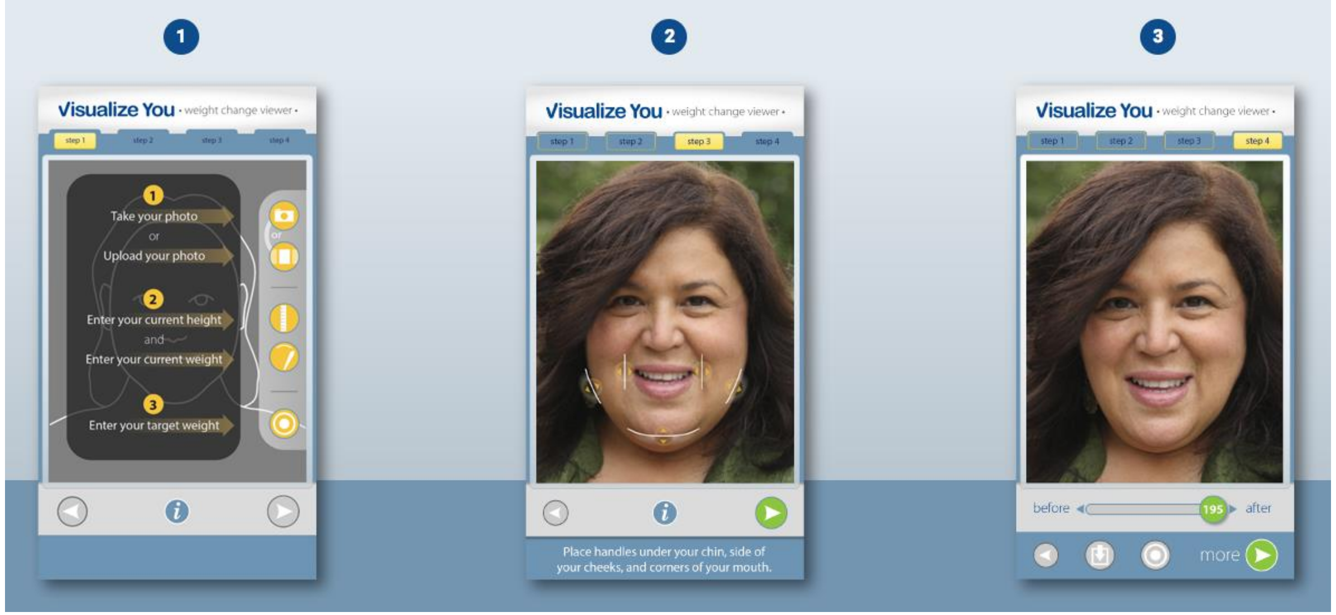
e.g., Clothes Shopping

Application: Shoe Shopping



Demo: https://www.youtube.com/watch?v=3A6YkHn6OU0

Application: Altering Appearance



e.g., simulate weight loss/gain
www.visualizeyourweight.com

e.g., simulate aging and different lifestyles
http://www.mastersingerontology.com/top-25-incredible-age-progression-tools-online.html

Application: Finding Criminals

Please compare the subject in the lower video to the subject in the top video.
For example if the subject in the bottom video is taller than the subjec

Attribute	Annotation	Certainty
Age	Older	100%
Bottom subject is OLDER than the top		
Hair Colour	Same	100%
Subjects have roughly the SAME hair colour		
Hair Length	Longer	100%
Bottom subject has LONGER hair than the top		
Height	Taller	100%
Bottom subject is TALLER than the top		
Figure	Same	100%
Subjects both have roughly the SAME figure		
Neck Length	Same	100%
Subjects have roughly the SAME length neck		
Neck Thickness	Thinner	100%
Bottom subject has a THINNER neck than the top		
Shoulder Shape	Same	100%
Subjects have roughly the SAME shoulder shape		
Chest	Same	100%
Subjects have roughly the SAME size chest		
Arm Length	Longer	100%
Bottom subject has a LONGER arms than the top		

e.g., Biometrics: “the suspect is *taller* than him”

[D. Reid, M. Nixon, IJCB 2011]

Applications: Other

- Recognize new objects with few/no examples; e.g., centaur

- Describe unusual aspects of a familiar object (intra-class variation); e.g.,

Attribute Recognition Datasets

1945 1957 1966 1983 1987 1990 2009 2010 2012

Birth of CV

CVPR ICCV ECCV

a-Pascal & a-Yahoo

ImageNet attributes

SUN attributes

 # Categories: 64 20 102

 # Images / Bounding Boxes: 12,000 9,600 14,340

Trend: build bigger datasets

Datasets: a-Pascal and a-Yahoo

1. Image Collection

- 12,000 VOC 2008 images

- Internet search on Yahoo! for 12 object categories

- Objects are localized in images with bounding boxes

Ali Farhadi, Ian Endres, Derek Hoiem, & David Forsyth. Describing Objects by Their Attributes. CVPR 2009

Datasets: a-Pascal and a-Yahoo

1. Image Collection

- 12,000 VOC 2008 images

- Internet search on Yahoo! for 12 object categories

- Objects are localized in images with bounding boxes

→

2. Category Selection

- 64 attribute categories chosen by authors

1. **Shape attributes**: 2D and 3D properties such as “is 2D boxy”, “is 3D boxy”, “is cylindrical“, etc

2. **Part attributes:** parts that are visible, such as “has head”, “has leg”, “has arm”, “has wheel”, “has wing”, “has window”

3. **Material attributes:** describe what an object is made of, including “has wood”, “is furry”, “has glass”, “is shiny”

Ali Farhadi, Ian Endres, Derek Hoiem, & David Forsyth. Describing Objects by Their Attributes. CVPR 2009

Datasets: a-Pascal and a-Yahoo

1. Image Collection

- 12,000 VOC 2008 images

- Internet search on Yahoo! for 12 object categories

- Objects are localized in images with bounding boxes

→

2. Category Selection

- 64 attribute categories chosen by authors

→

3. Human Labeling

- AMT crowd workers identify presence of each attribute

Ali Farhadi, Ian Endres, Derek Hoiem, & David Forsyth. Describing Objects by Their Attributes. CVPR 2009

Dataset: ImageNet Attributes

1. Image Collection

- Candidate images are all ImageNet images for which objects are localized in images with bounding boxes

- Include images in a "synset" for which the attribute is contained in the synset's name or definition

Dataset: ImageNet Attributes

1. Image Collection

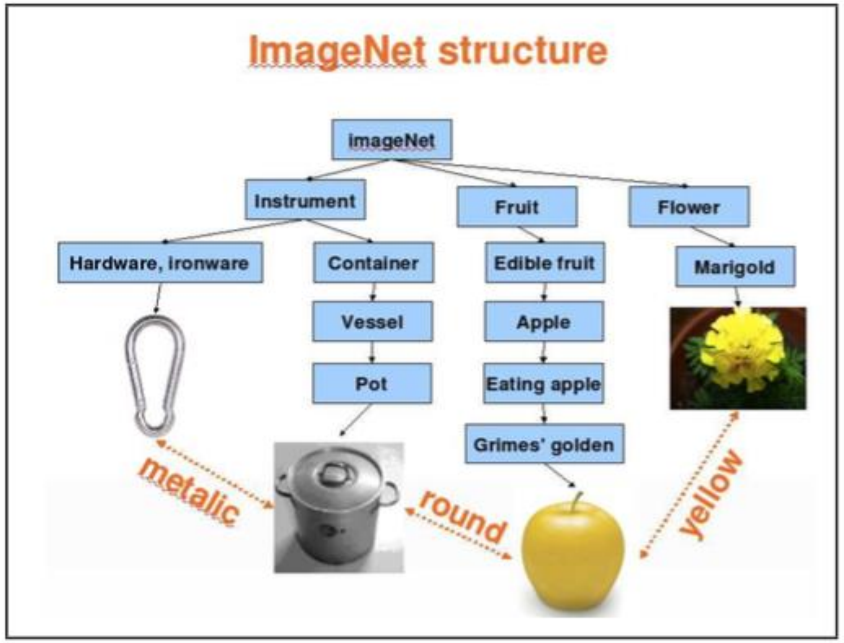
- Candidate images are all ImageNet images for which objects are localized in images with bounding boxes

- Include images in a “synset” for which the attribute is contained in the synset’s name or definition

2. Category Selection

- 20 categories:
  (1) 8 colors
  (2) furry, long, metallic, rectangular, rough, round, shiny, smooth, spotted, square, striped, wet, vegetation, wooden

Aim is to identify *visual* connections between objects



Olga Russakovsky & Li Fei Fei. Attribute Learning in Large-Scale Datasets. ECCV 2010

Dataset: ImageNet Attributes

1. Image Collection

- Candidate images are all ImageNet images for which objects are localized in images with bounding boxes

- Include images in a "synset" for which the attribute is contained in the synset's name or definition

2. Category Selection

- 20 categories:
  (1) 8 colors
  (2) furry, long, metallic, rectangular, rough, round, shiny, smooth, spotted, square, striped, wet, vegetation, wooden

3. Human Labeling

- AMT crowd workers identify presence of each attribute for 106 images per HIT

Olga Russakovsky & Li Fei Fei. Attribute Learning in Large-Scale Datasets. ECCV 2010

Dataset: ImageNet Attributes

metallic	fork (.72), transporter (.56), roller coaster (.49), stick (.41), wheel (.38), police van (.37), keyboard (.34), sail (.31), bridge (.31), building (.28), ski (.25), bowhead (.25) 
rectangular	police van (.90), transporter (.84), cabinet (.61), marimba (.50), window (.44), varietal (.42), flag (.38), bridge (.38), kummel (.31), pot (.29), generic (.28), pool table (.26) 
yellow	egg yolk (1.00), sunflower (.86), omelet (.70), kedgeree (.64), flan (.61), tostada (.48), succotash (.42), pizza (.35), zabaglione (.26), ravigote (.25), curry (.23), casserole (.21) 

Olga Russakovsky & Li Fei Fei. Attribute Learning in Large-Scale Datasets. ECCV 2010

Dataset: SUN Attributes

1. Image Collection

- 20 scenes from each of the 717 SUN scene categories

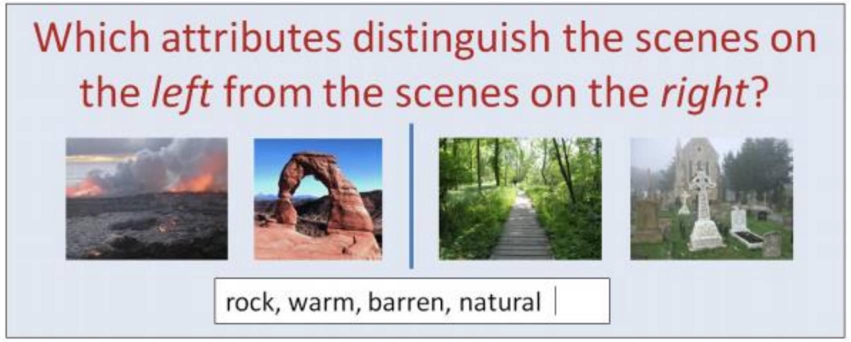
Dataset: SUN Attributes

1. Image Collection

- 20 scenes from each of the 717 SUN scene categories

2. Category Selection

- Discover *attribute types* from image descriptions by AMT workers: material, object & envelope, surface property, affordance, spatial
- Choose *discriminative* attributes offered by AMT workers for the 5 types
- Authors removed and added some categories resulting in 102 categories



Genevieve Patterson & James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Attributes. CVPR 2012

Dataset: SUN Attributes

1. Image Collection

- 20 scenes from each of the 717 SUN scene categories

2. Category Selection

- Discover *attribute types* from image descriptions by AMT workers: material, object & envelope, surface property, affordance, spatial

- Choose *discriminative* attributes offered by AMT workers for the 5 types

- Authors removed and added some categories resulting in 102 categories

3. Human Labeling

- AMT crowd workers identify presence of each attribute for 48 images per HIT

Genevieve Patterson & James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Attributes. CVPR 2012

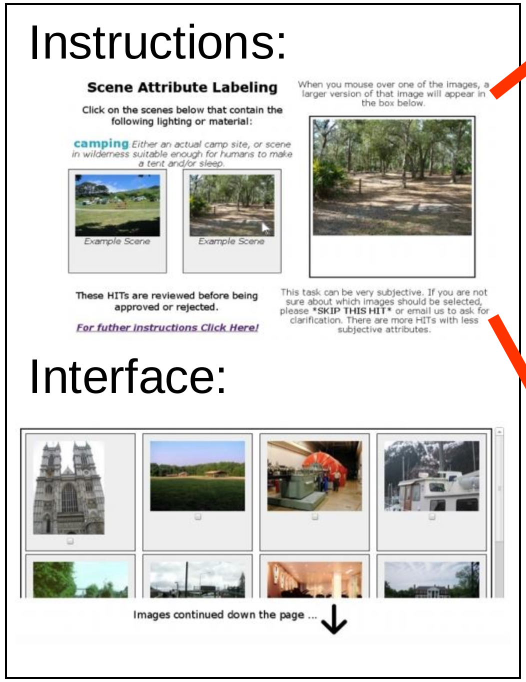
Dataset: SUN Attributes

1. Task Design

Genevieve Patterson & James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Attributes. CVPR 2012

Dataset: SUN Attributes

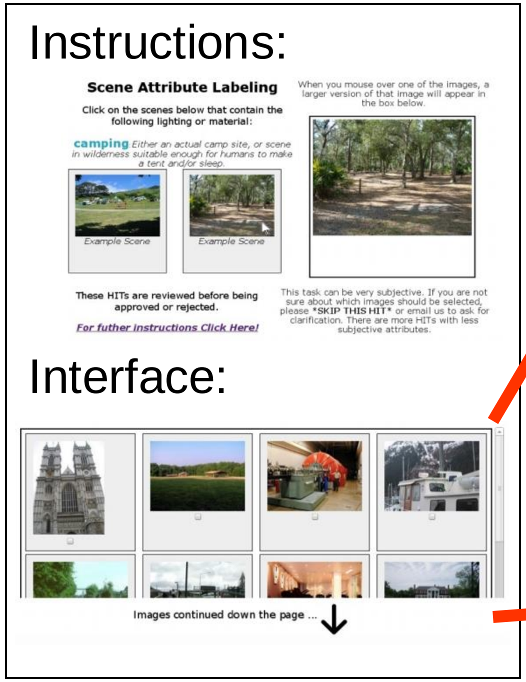
1. Task Design



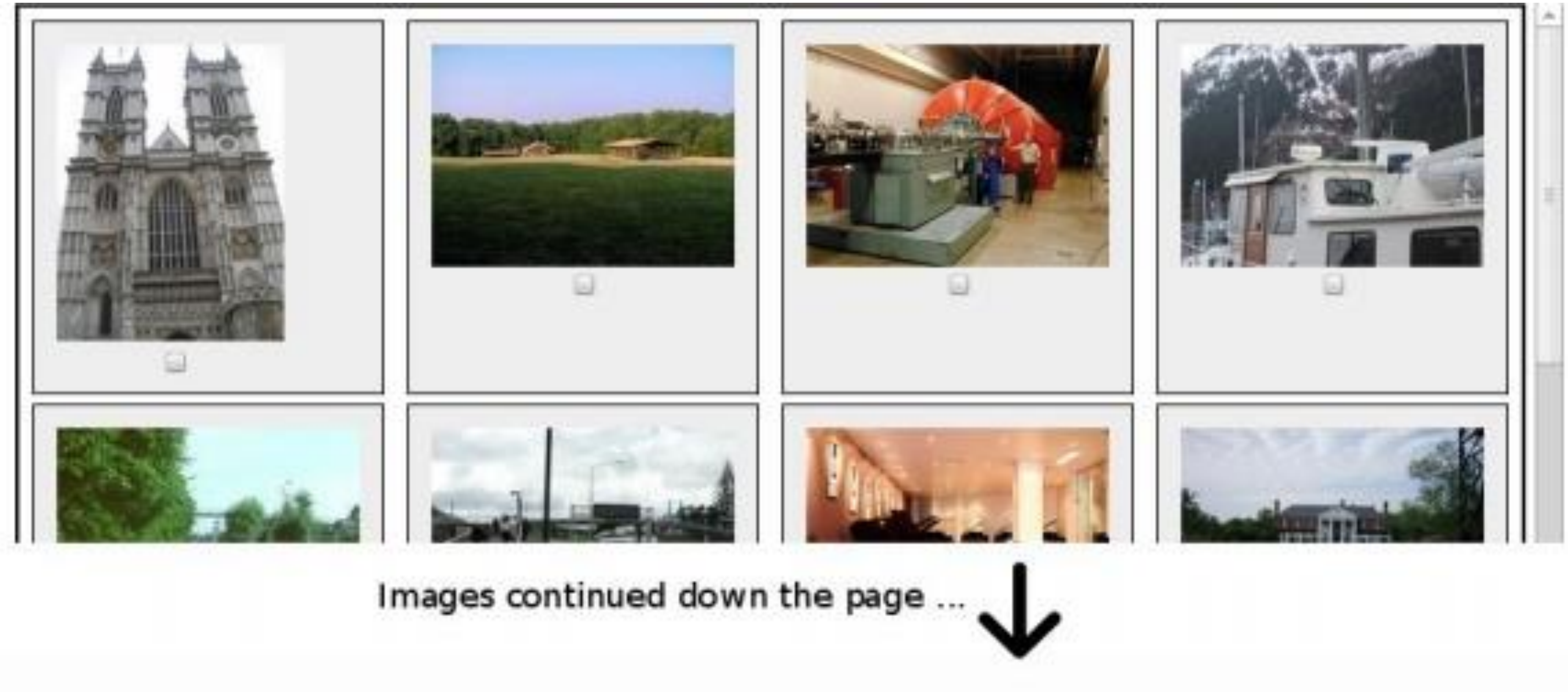
Genevieve Patterson & James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Attributes. CVPR 2012

Dataset: SUN Attributes

1. Task Design



(grid of 48 images)



Genevieve Patterson & James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Attributes. CVPR 2012

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features
- Attribute Classification: Problem, Applications, and Datasets
- Discussion (chosen by YOU ☺)

The End