

# Scene and Attribute Classification

**Danna Gurari**

University of Colorado Boulder  
Fall 2023



# Review

- Last week:
  - ImageNet Challenge Top Performers
  - Baseline Model: AlexNet
  - VGG
  - ResNet
  - Discussion
- Assignments (Canvas)
  - Reading assignment was due earlier today
  - Next reading assignments due next Monday and Wednesday
- Questions?

# Scene & Attribute Classification: Today's Topics

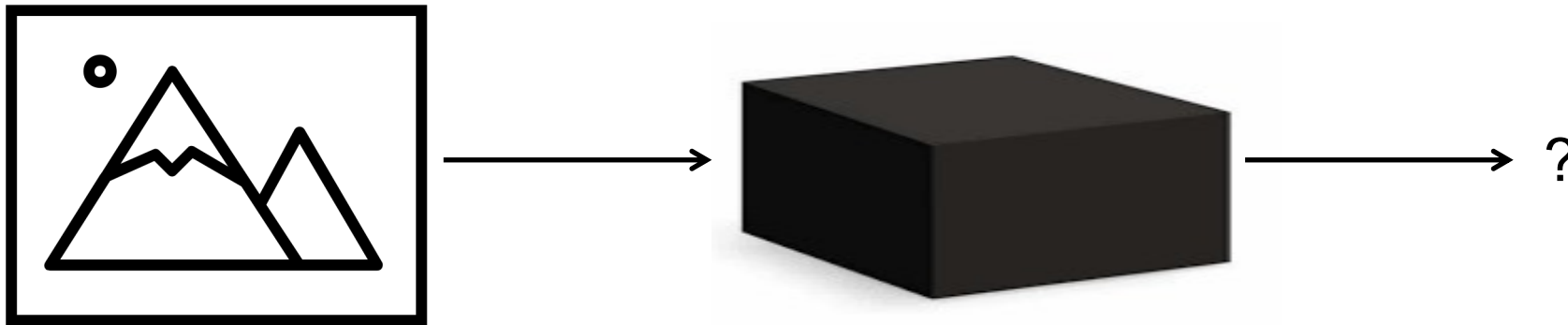
- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

# 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

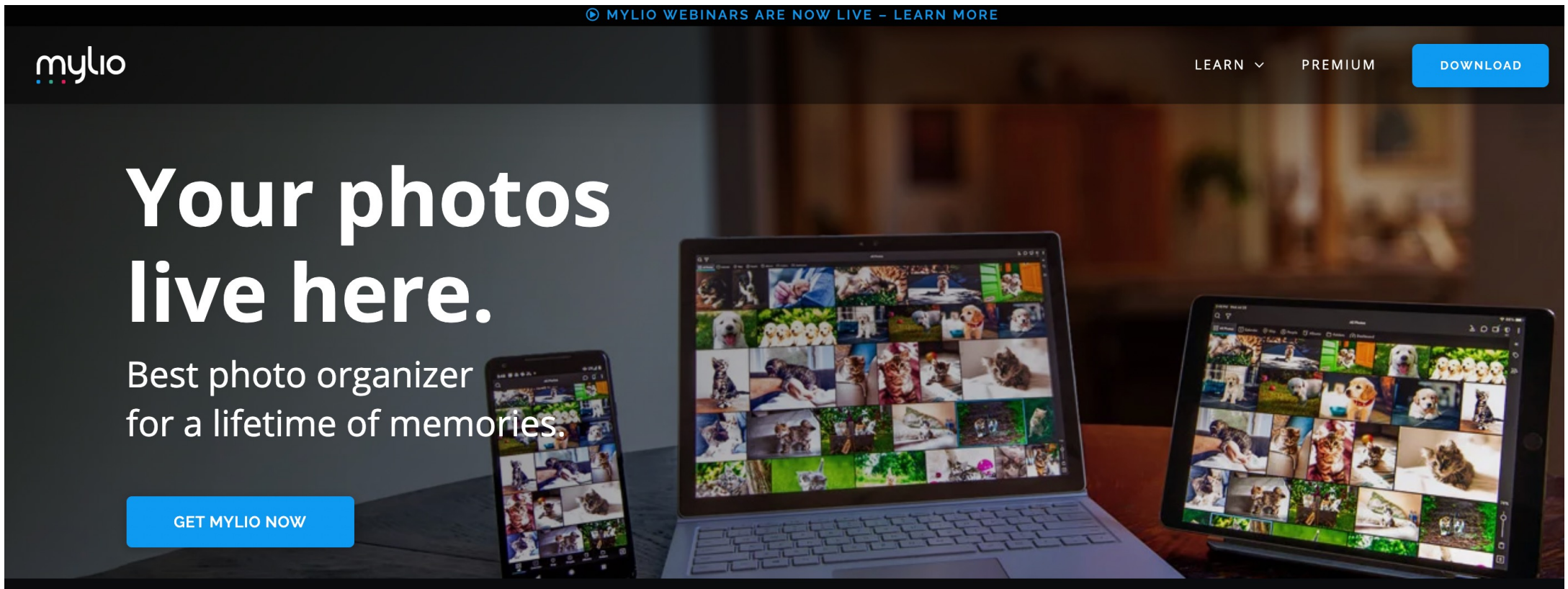
INPUT



OUTPUT



# Application: Photo Organization



© MYLIO WEBINARS ARE NOW LIVE - LEARN MORE

mylio

LEARN ▾ PREMIUM

DOWNLOAD

## Your photos live here.

Best photo organizer for a lifetime of memories.

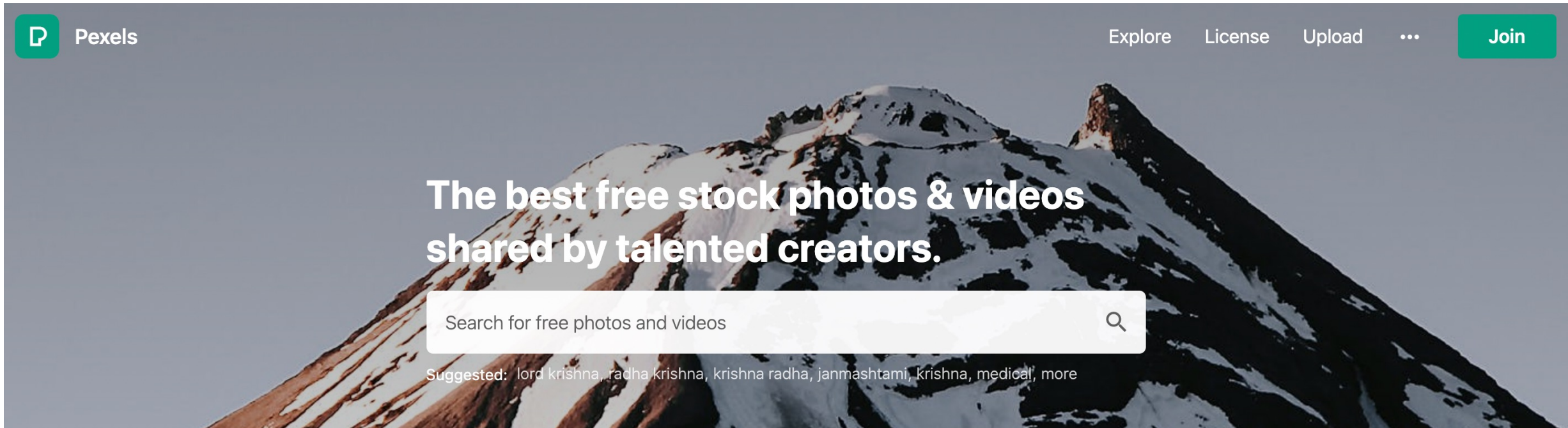
GET MYLIO NOW

The screenshot shows the Mylio website interface. At the top, there is a navigation bar with the Mylio logo on the left, a link to 'LEARN' with a dropdown arrow, a 'PREMIUM' link, and a blue 'DOWNLOAD' button. Below the navigation bar, the main content area features a large headline 'Your photos live here.' and a sub-headline 'Best photo organizer for a lifetime of memories.' A blue button labeled 'GET MYLIO NOW' is positioned below the sub-headline. The background of the main content area is a blurred image of a laptop, a tablet, and a smartphone, all displaying a grid of photos of various animals, primarily dogs and cats, illustrating the application's multi-device synchronization and photo organization capabilities.

Demo: <https://www.youtube.com/watch?v=aBqmWUalnh0>  
(start video at 1:46)

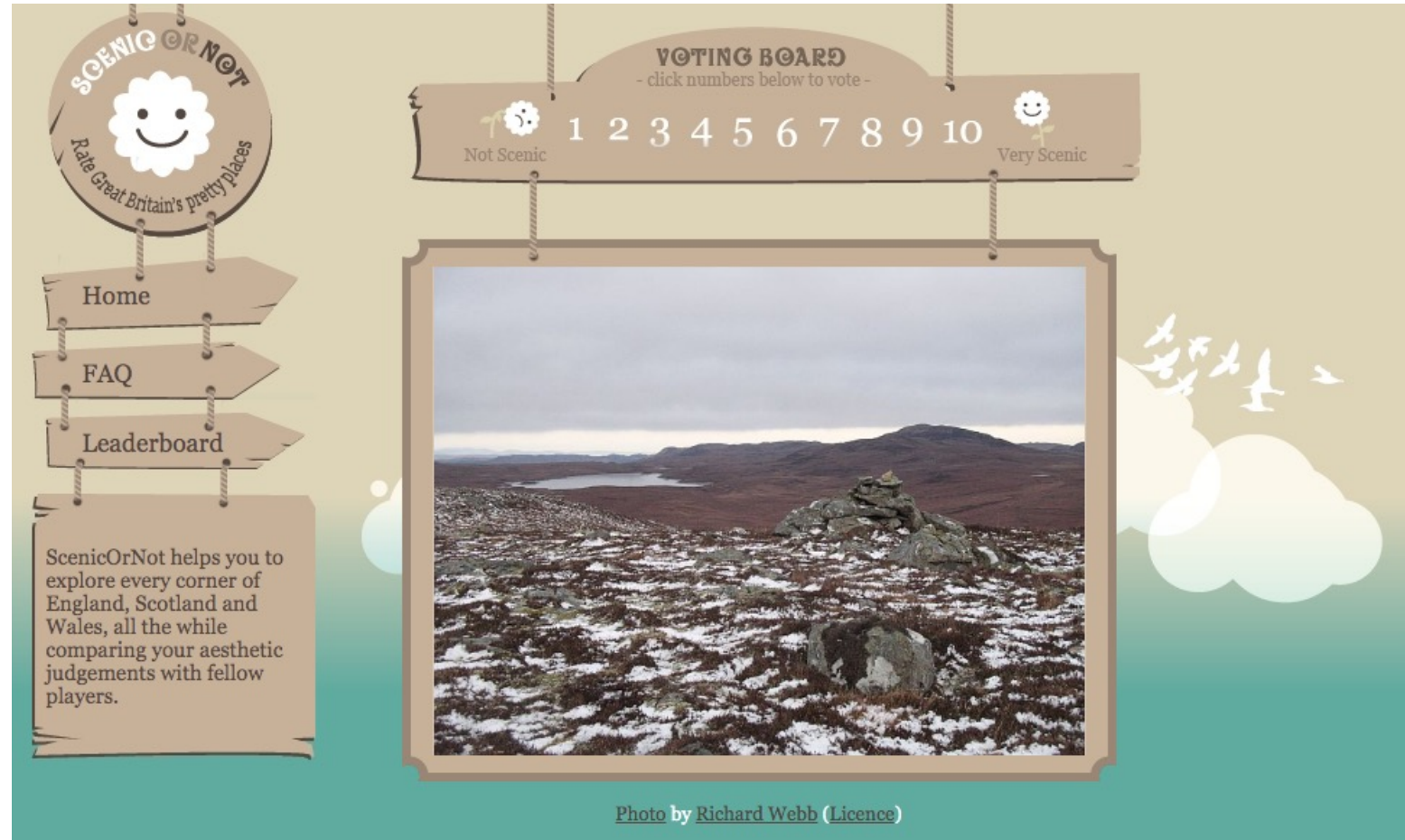


# Application: Image Search



# Application: Urban Planning

People's *well-being* is correlated with *scenic* places



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)

Can you think of any other  
potential applications?

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



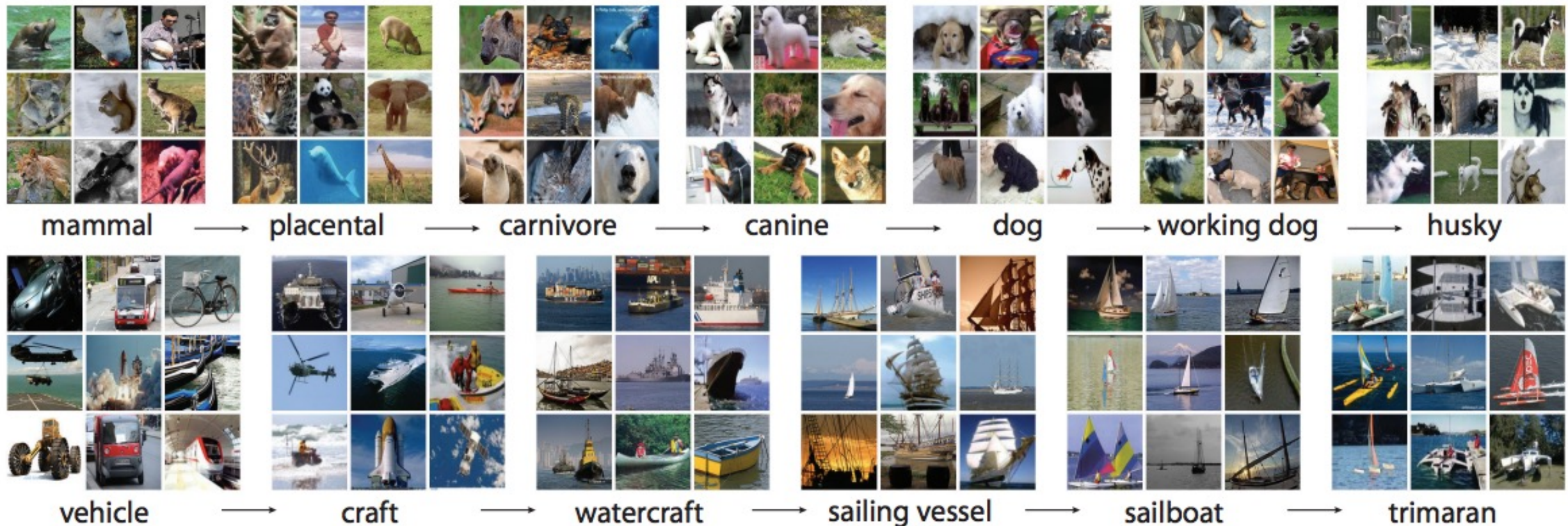
- Object Recognition
  - e.g., What would you expect (or not expect) to find in the scene [now, earlier, later]?
- Activity Recognition/Prediction
  - e.g., What would you expect people to do (or not do) in the scene [now, earlier, later]?

# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

# Motivation for Scene Classification Datasets

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



# Motivation for Scene Classification Datasets

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



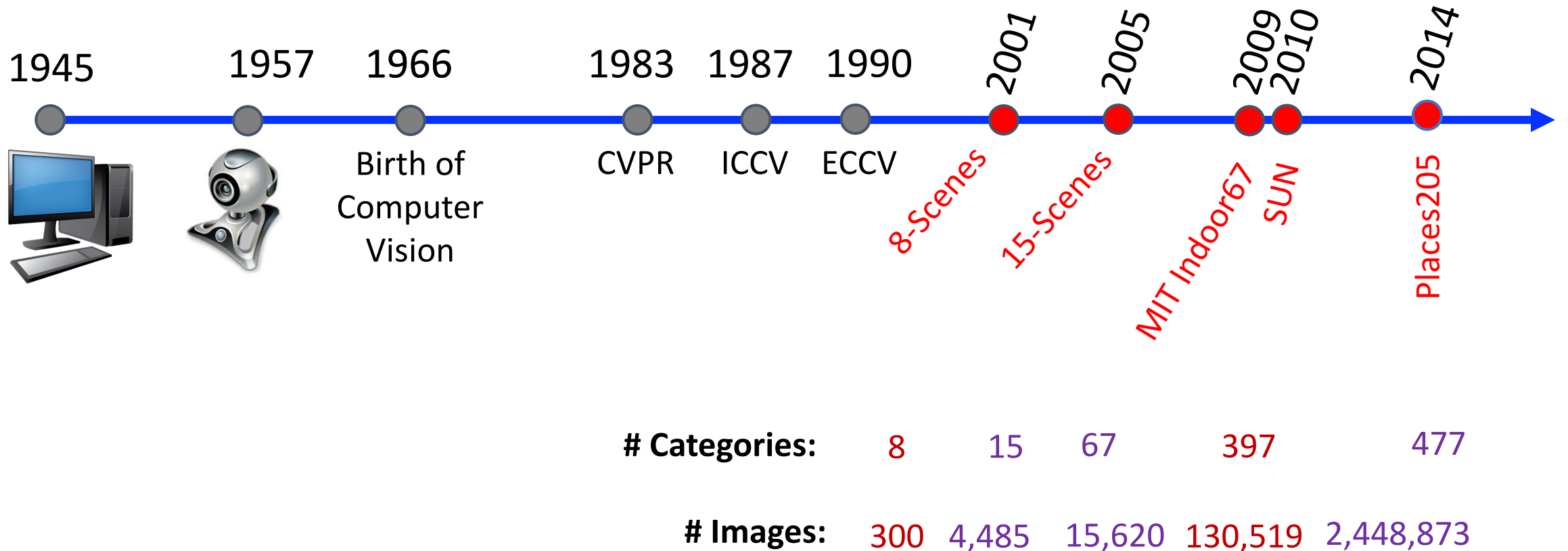


# Motivation for Scene Classification Datasets

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



# Scene Classification Datasets



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

Coast



Fields



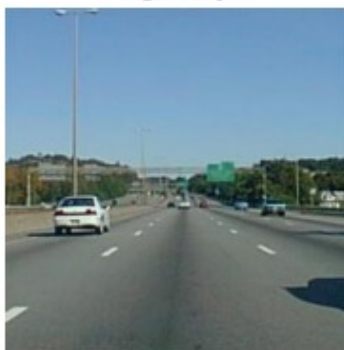
Forests



Mountains



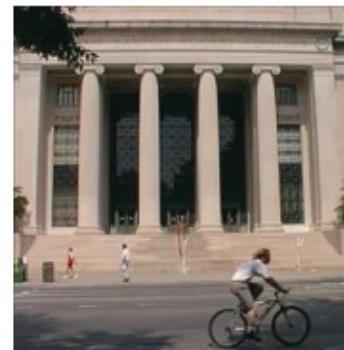
Highways



Streets



Inside City



Skyscrapers



# 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



# 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



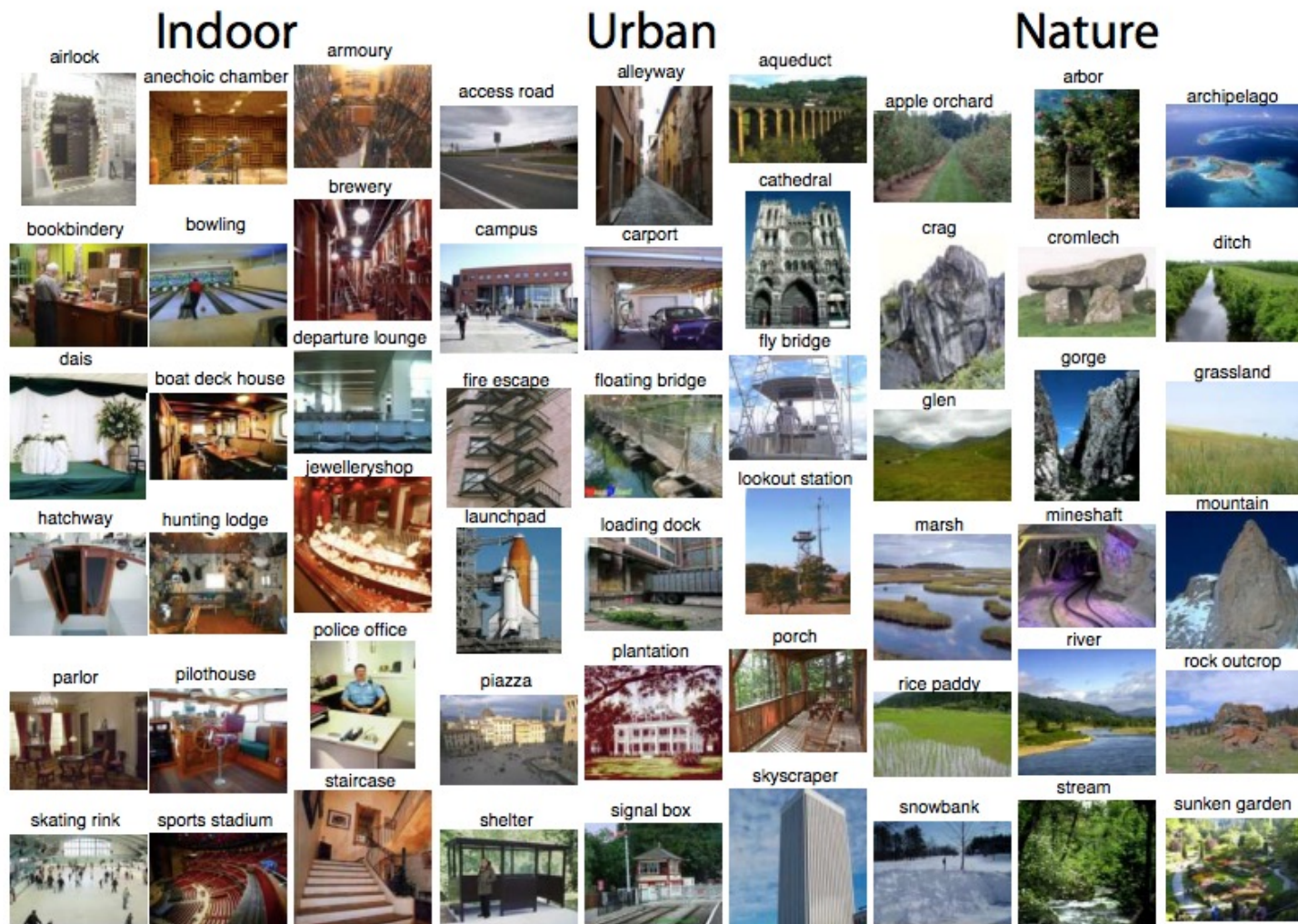
# SUN

## 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



# 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



# 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)

# 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

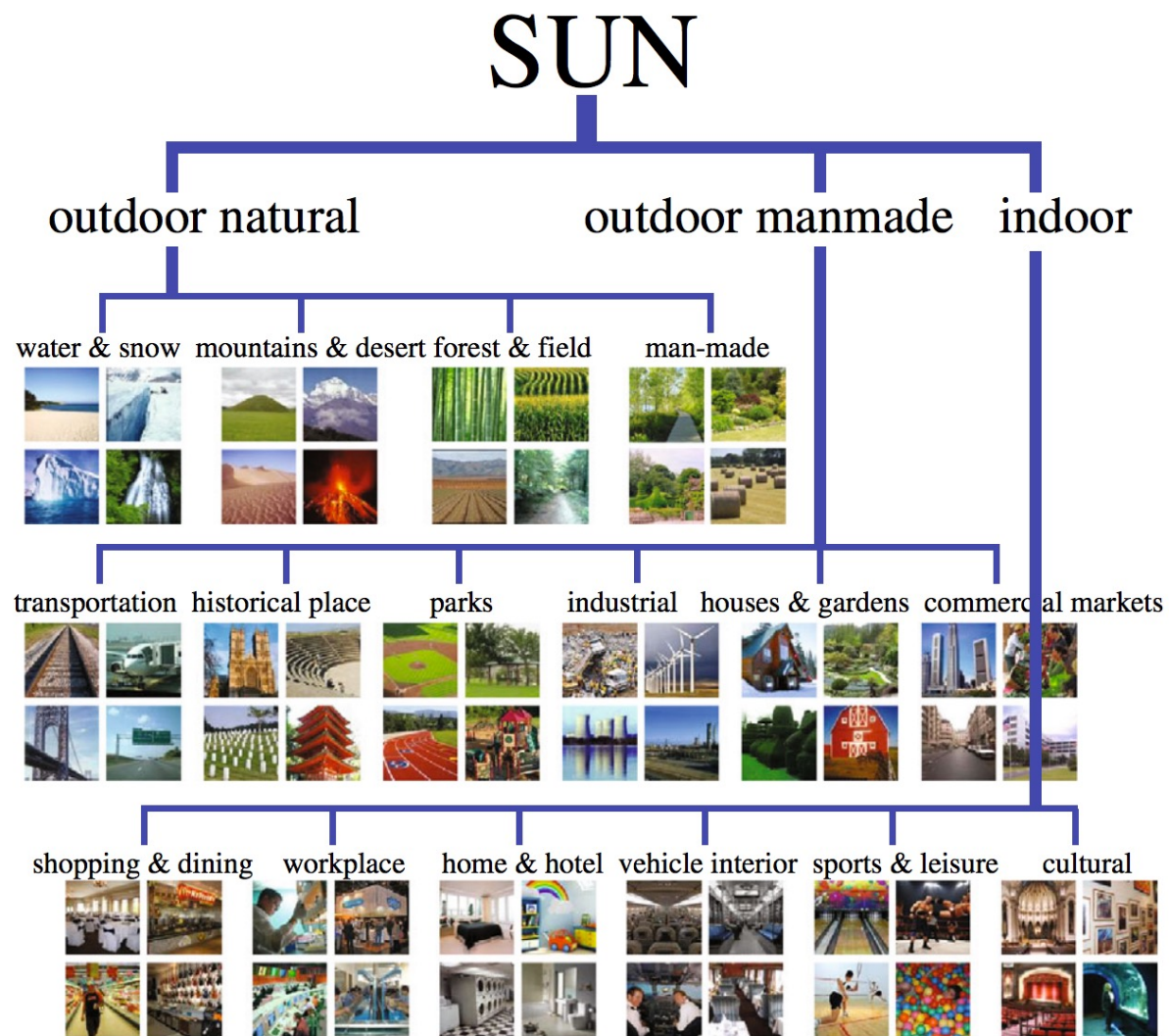
- 9 in-house people reviewed & discarded irrelevant images
- Result is 130,519 images spanning 397 categories with >99 images per category



# Places205

## 1. Category Selection

Same taxonomy as SUN



# 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

The logo for Bing, featuring the word "bing" in a blue, lowercase, sans-serif font with a small orange dot above the letter 'i'.The logo for Google Image Search, featuring the word "Google" in its multi-colored font and "Image Search" in a smaller blue font below it.The logo for Flickr, featuring the word "flickr" in a blue, lowercase, sans-serif font with "GAMMA" in a smaller grey font above the "r" and a pink dot above the "i".

# 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

# Places205

## User interface: Instructions

### 1. Task Design

#### Instructions:



#### Interface:



### Examples

**Is this a cliff scene?**

**Definition:** high, steep or overhanging face of rock.

**Task**

For each of the **810** images, answer yes or no to the above question. Only answer **Yes** to **real photos**. Always answer **No** to **cartoon, drawing, CG rendering**, or real photos with a **large text overlay** on the photo. Here are some examples:

No Single Object	No Text Overlay	No Drawing	No Screenshot	No Graphics	No Bad Photo
Not Only Logo	No Magazine/Newspaper	No	No	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes

# Places205

User interface: Task

# Tasks left

## 1. Task Design

Instructions:



Interface:



Instruction **Is this a cliff scene?** Submit (790 images left)

Definition: a high, steep or overhanging face of rock.

Current Task: press a key on keyboard

Completed Tasks

No



Yes



Next Tasks

No





# Places205

## 1. Task Design

**Instructions:**

**Start** **Is this a cliff scene?**  
Definition: a high, steep or overhanging face of rock.

**Task**  
For each of the **#10** images, answer yes or no to the above question. Only answer **Yes** to real photos. Always answer **No** to cartoons, drawings, CG rendering, or real photos with a large text overlay on the photo. Here are some examples:

No Simple Object No Text Overlay No Drawing No Screenshot No Graphics No Bad Photo

Not Only Logo No Magazine/Newspaper No No Yes Yes

**Interface:**

**Instruction** **Is this a cliff scene?** **Submit (78) Images Left**  
Definition: a high, steep or overhanging face of rock.

**Yes**

**No** **No** **No**

## 2. Crowdsourcing Platform

amazon mechanical turk™  
Artificial Artificial Intelligence


# Places205

## 1. Task Design

**Instructions:**



**Interface:**



## 2. Crowdsourcing Platform



## 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"

# Places205 Summary

## 1. Category Selection

Same taxonomy as SUN

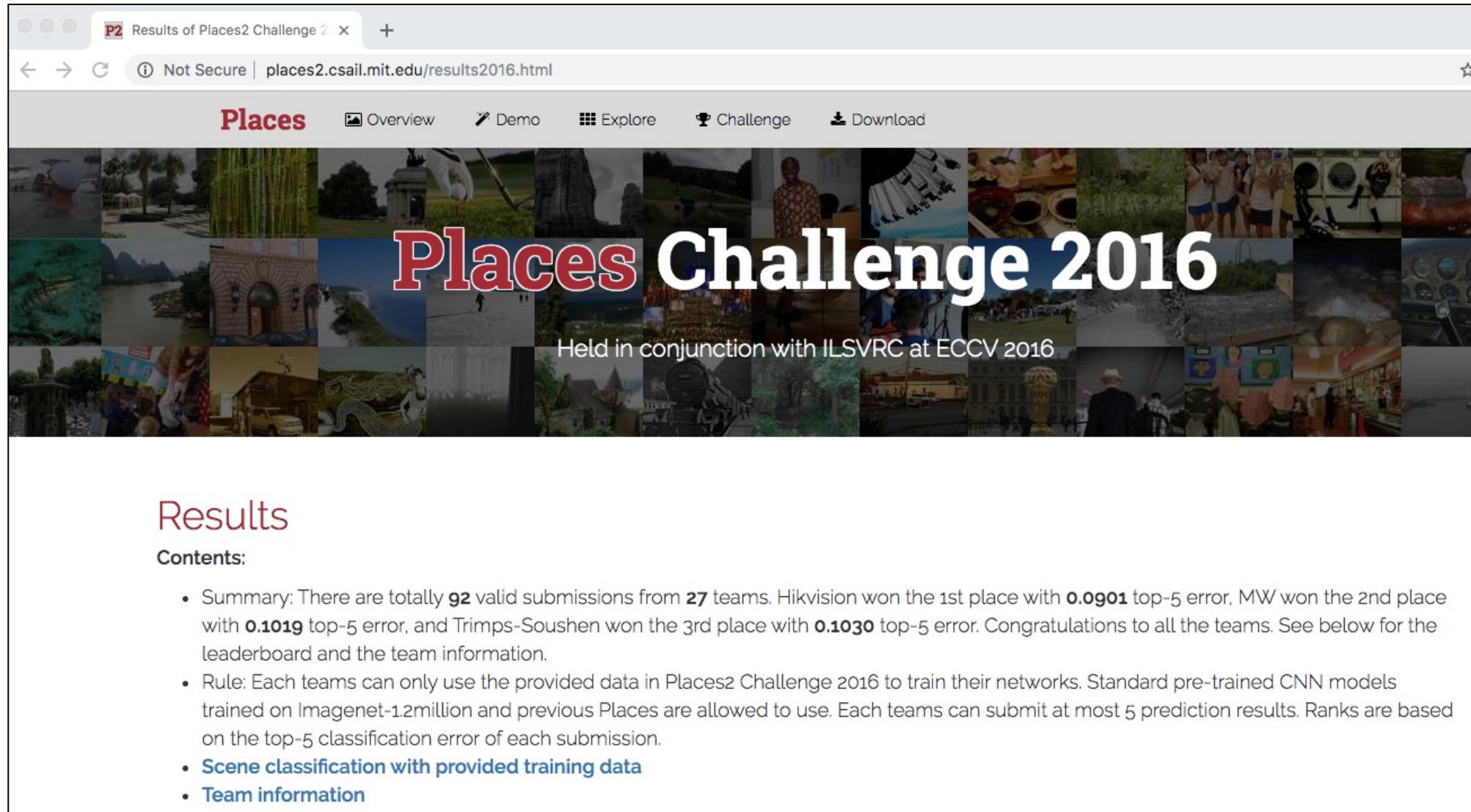
## 2. Image Collection

- 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

Steel drum



**Output:**  
Scale  
T-shirt  
Steel drum  
Drumstick  
Mud turtle



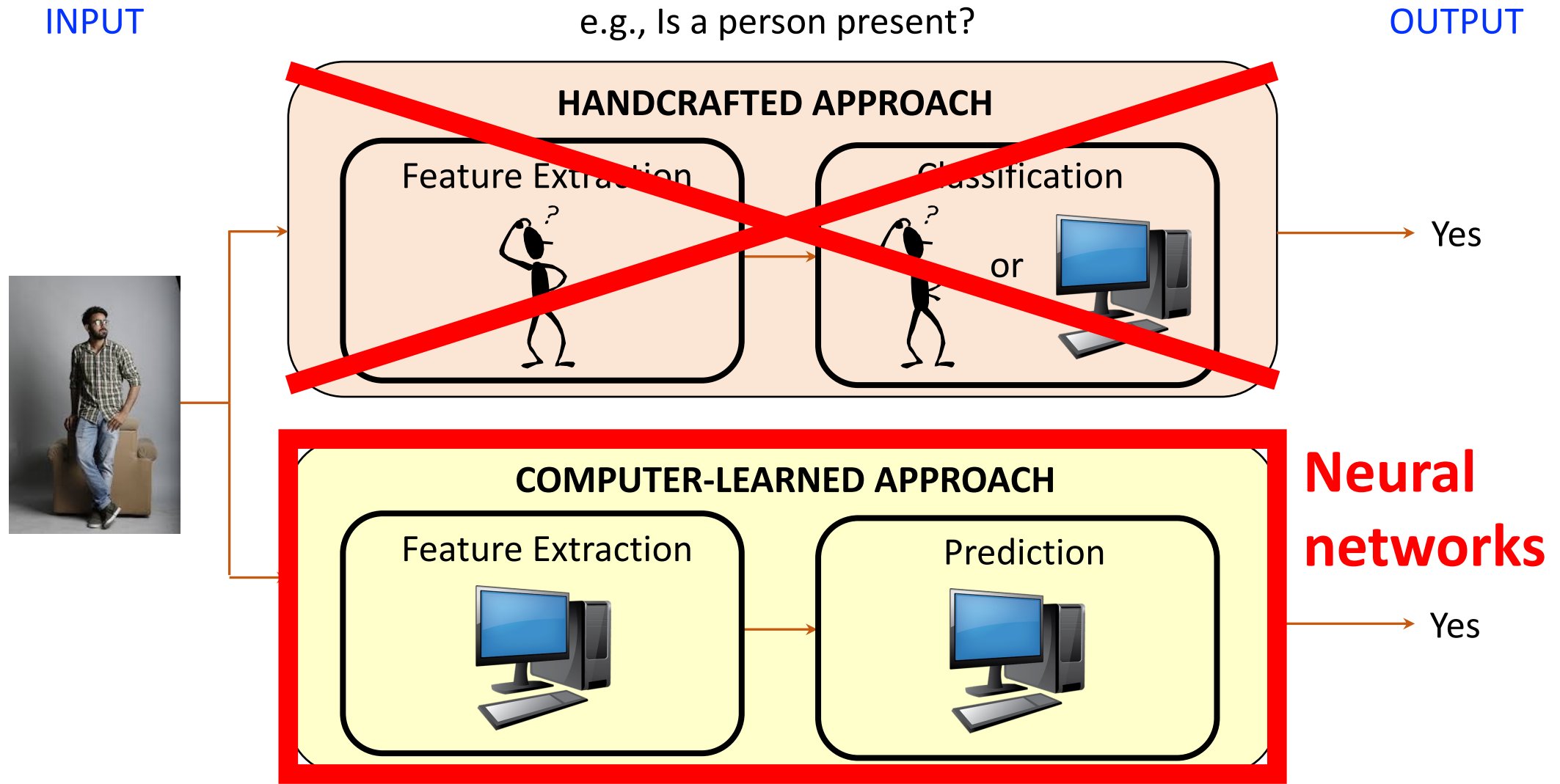
**Output:**  
Scale  
T-shirt  
Giant panda  
Drumstick  
Mud turtle



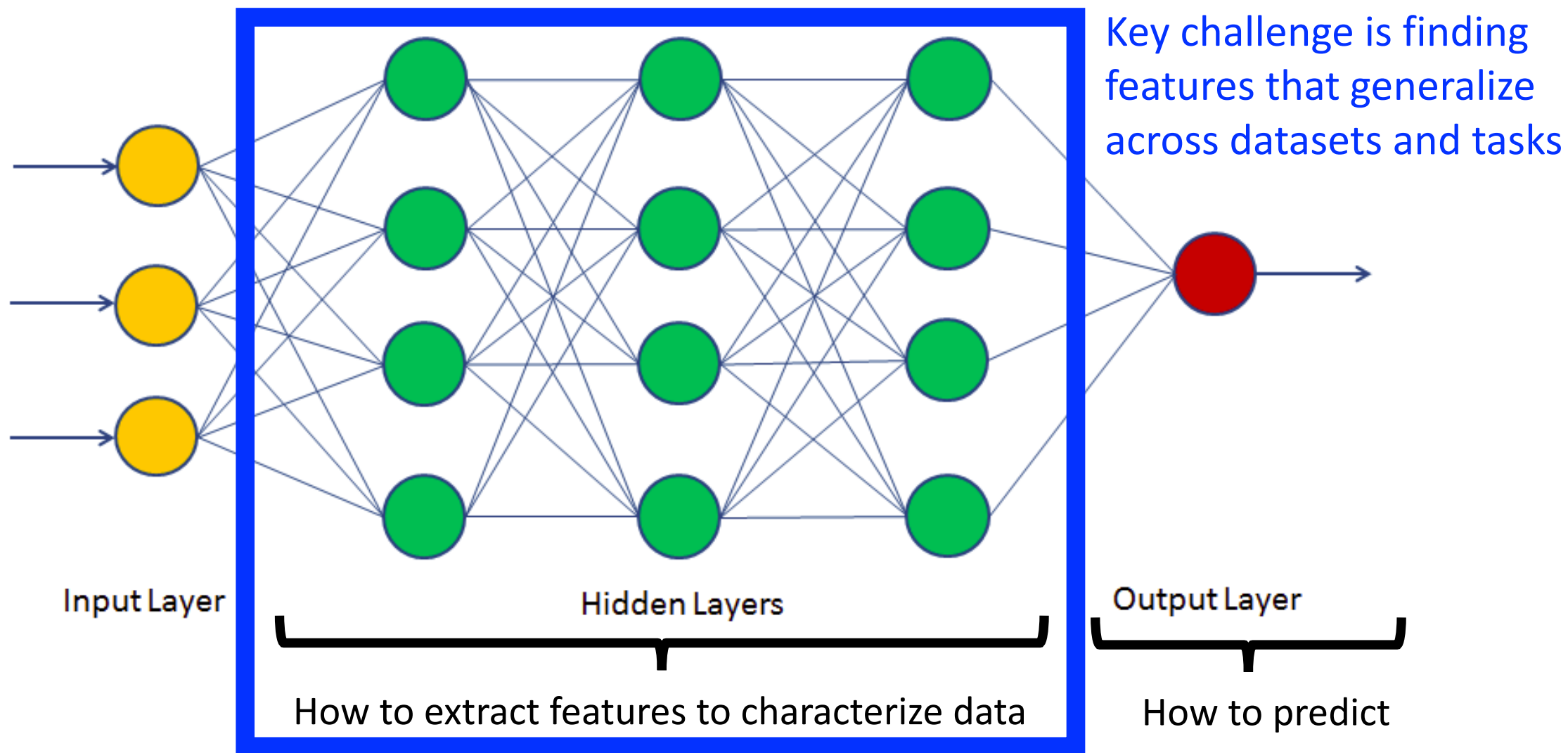
# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

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



# Key Idea: Establish Good “Deep Features”





# 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?

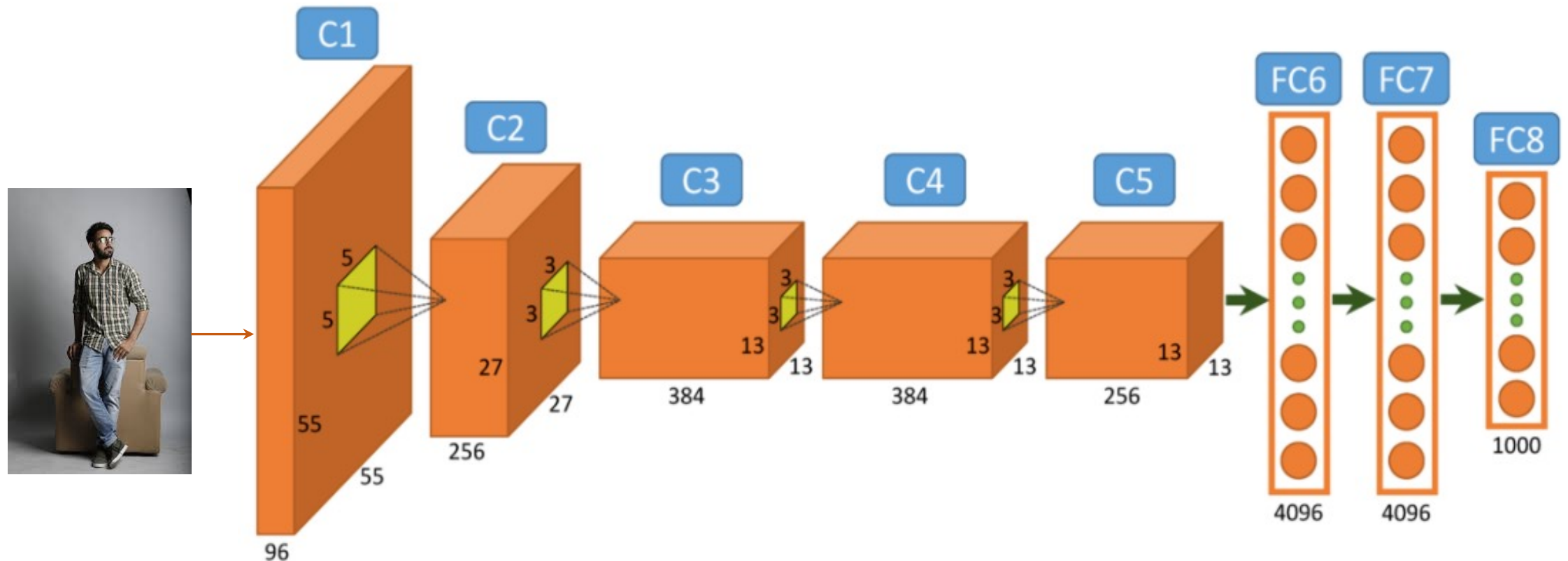


image source: [https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers\\_fig2\\_312303454](https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers_fig2_312303454)

# 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	<b>54.32±0.14</b>	<b>68.24</b>	<b>90.19±0.34</b>	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	<b>87.22±0.92</b>	<b>67.23±0.27</b>

What trends do you see?

# 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	<b>54.32±0.14</b>	<b>68.24</b>	<b>90.19±0.34</b>	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	<b>87.22±0.92</b>	<b>67.23±0.27</b>

State-of-the-art  
performance at the time

# Performance Comparison When Using Features Extracted from Two AlexNet Models

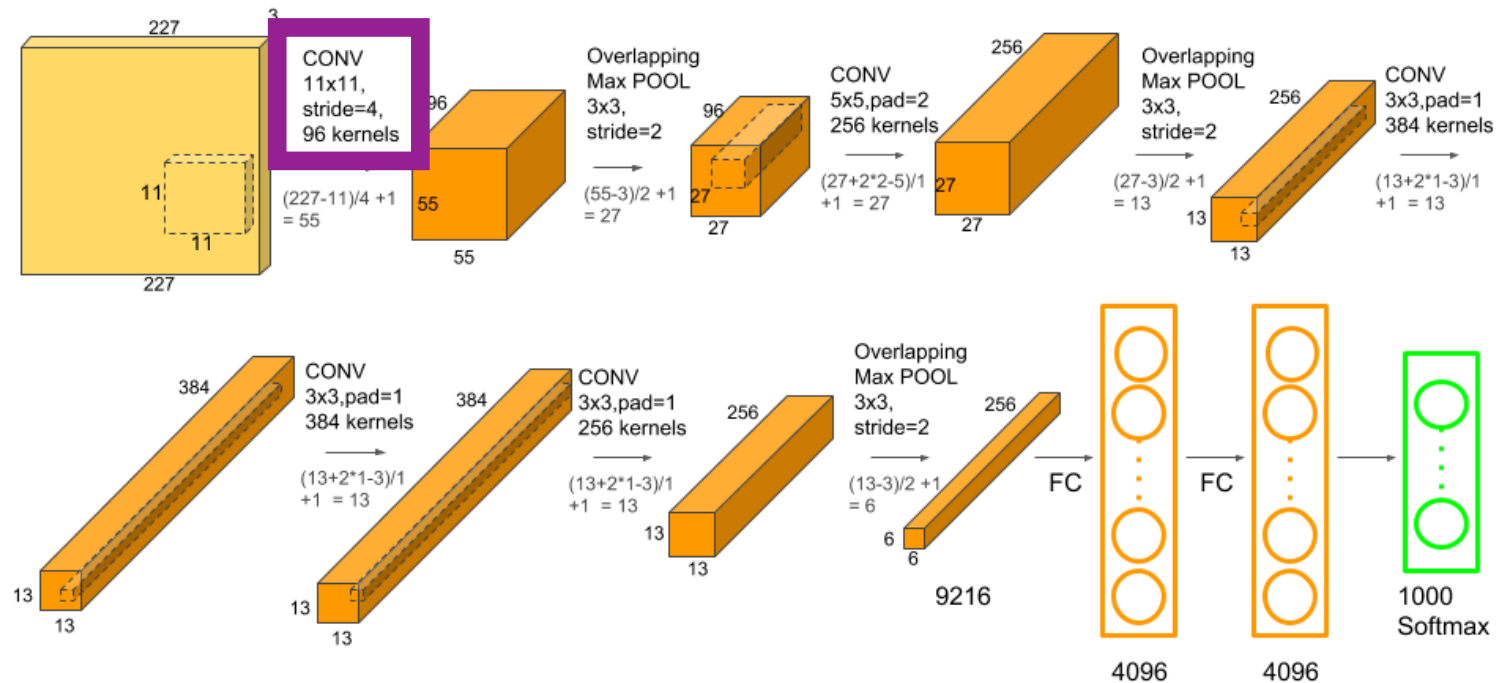
	SUN397	MIT Indoor67	Scene15	Caltech101	Caltech256
Places-CNN feature	<b>54.32±0.14</b>	<b>68.24</b>	<b>90.19±0.34</b>	65.18±0.88	45.59±0.31
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	<b>87.22±0.92</b>	<b>67.23±0.27</b>
Feature from AlexNet trained on both datasets	53.86±0.21	<b>70.80</b>	<b>91.59±0.48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

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

# Comparing Representations Learned When Training AlexNet on Different Datasets

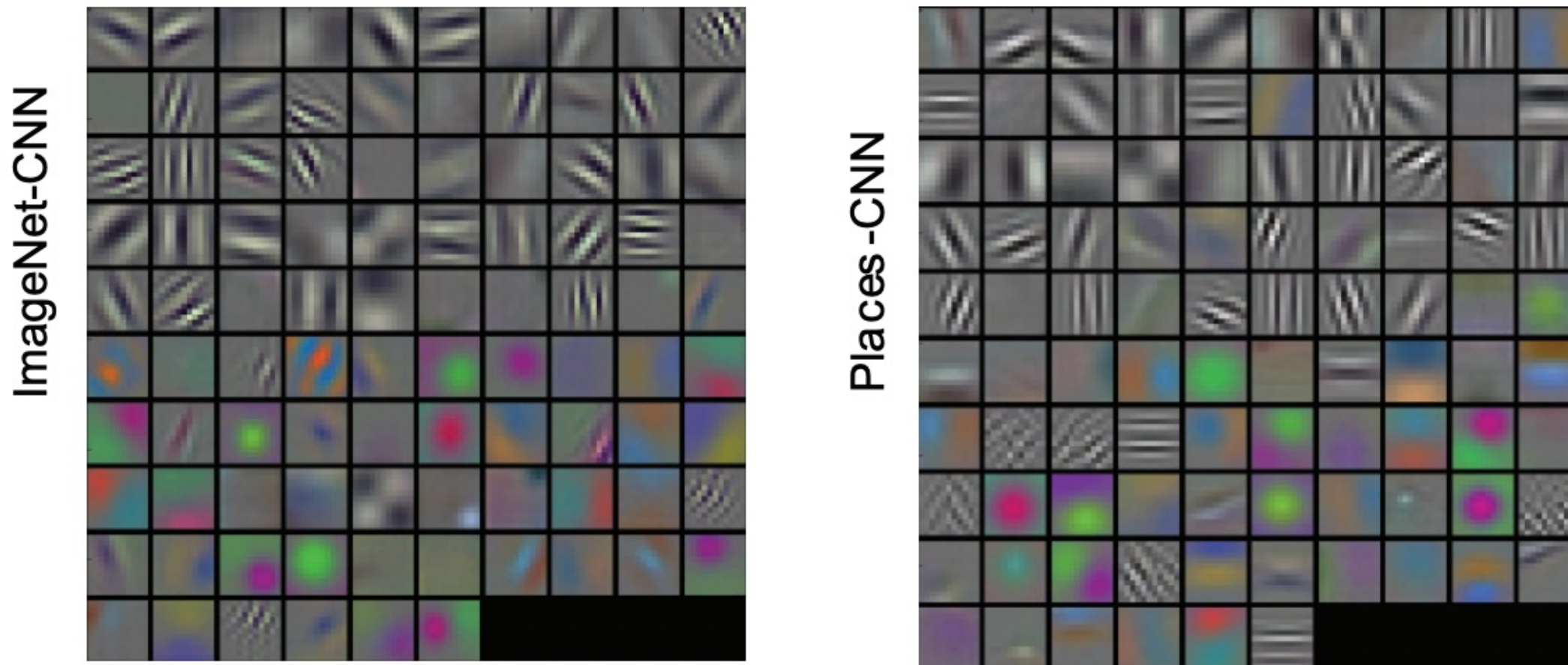
- **Dataset 1:** ImageNet (~1.5 million images of **objects** scraped from search engines)

- **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

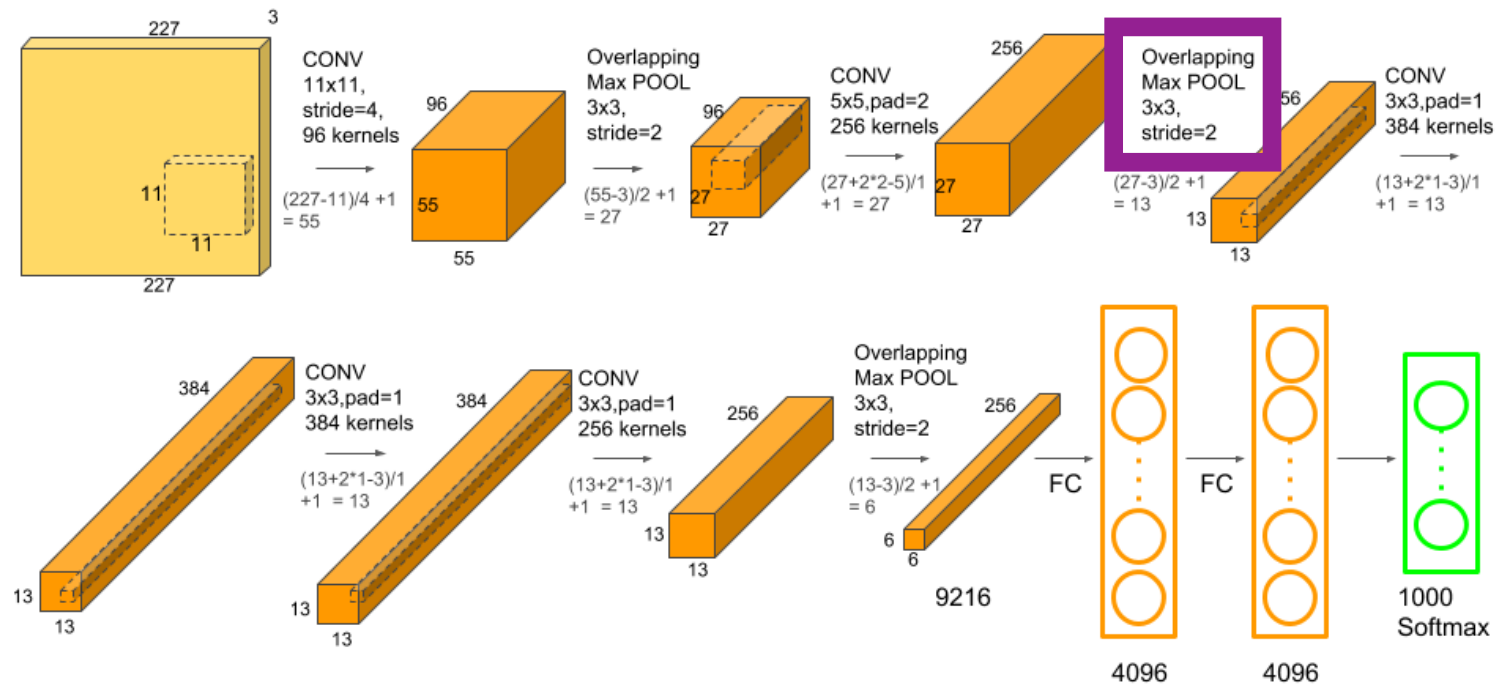


Do filters learned from the different datasets look similar or different?

# 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)



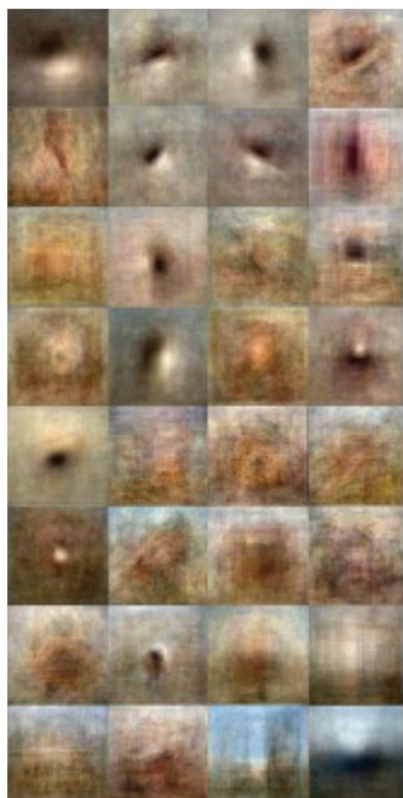
Source: <https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png>



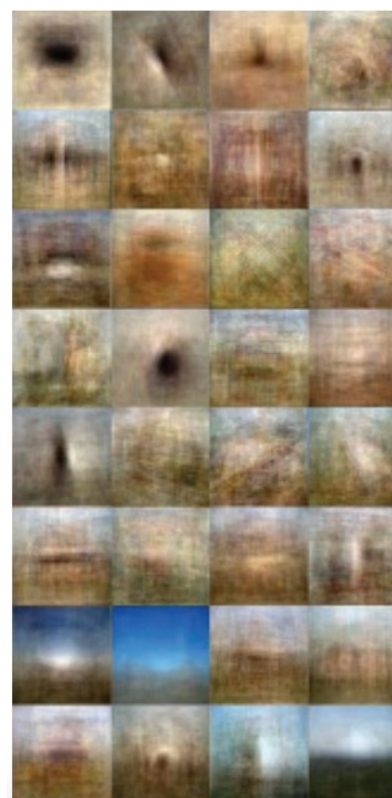
# Comparing Representations Learned When Training AlexNet on Different Datasets

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)

ImageNet-CNN



Places -CNN

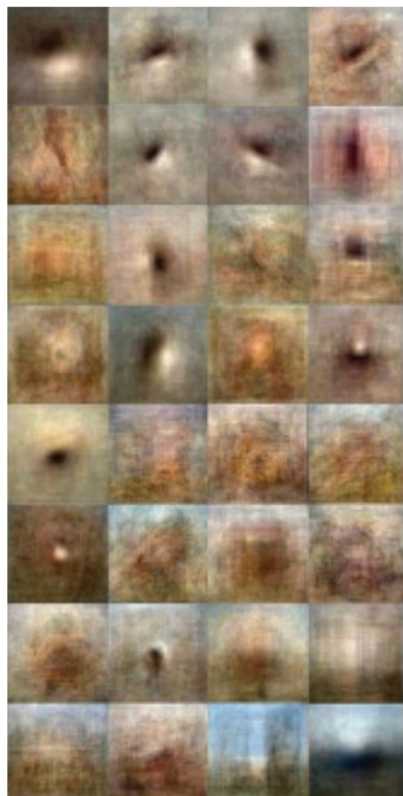


Do the representations from the different datasets appear to be similar or different?

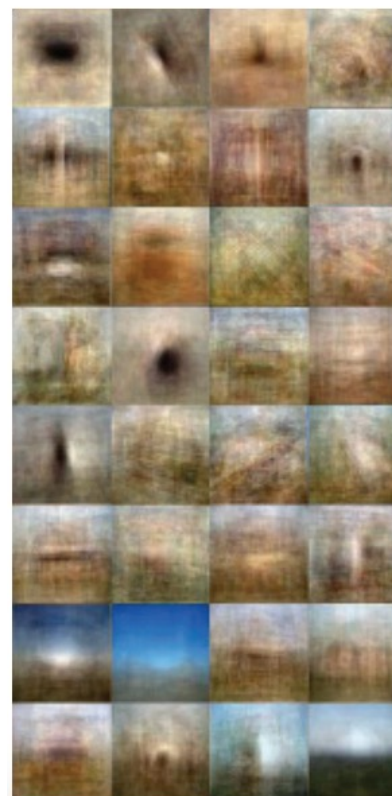
# Comparing Representations Learned When Training AlexNet on Different Datasets

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



Places -CNN

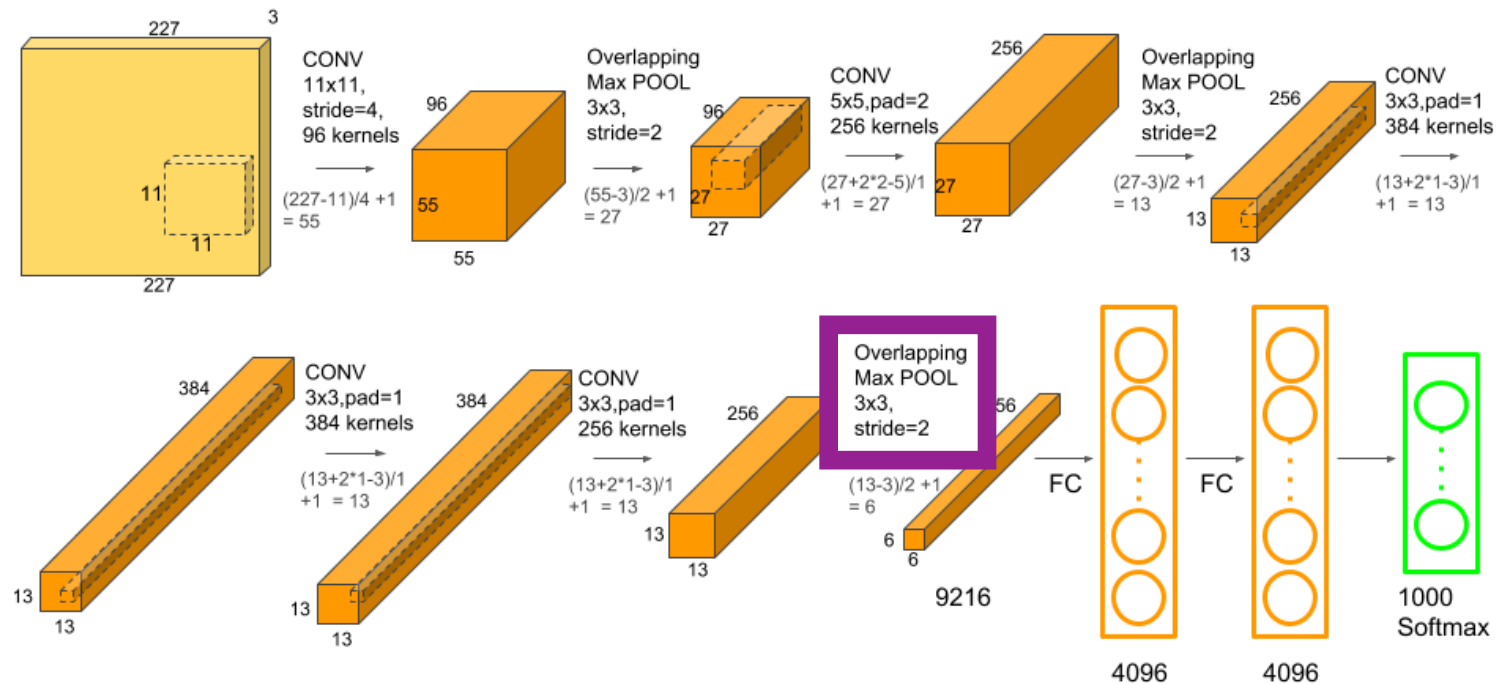


ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

# 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)

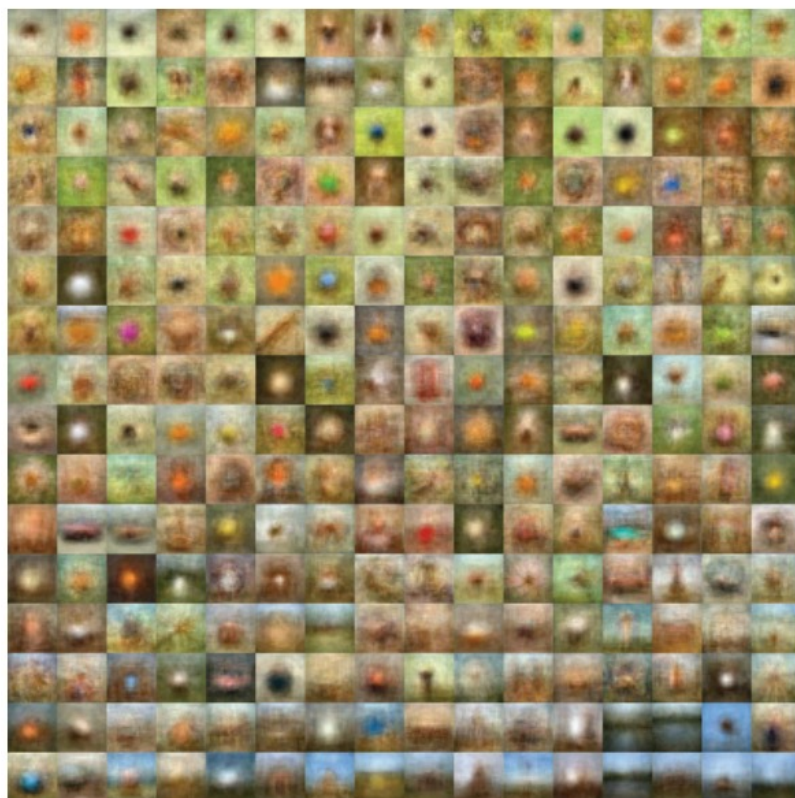


Source: <https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png>

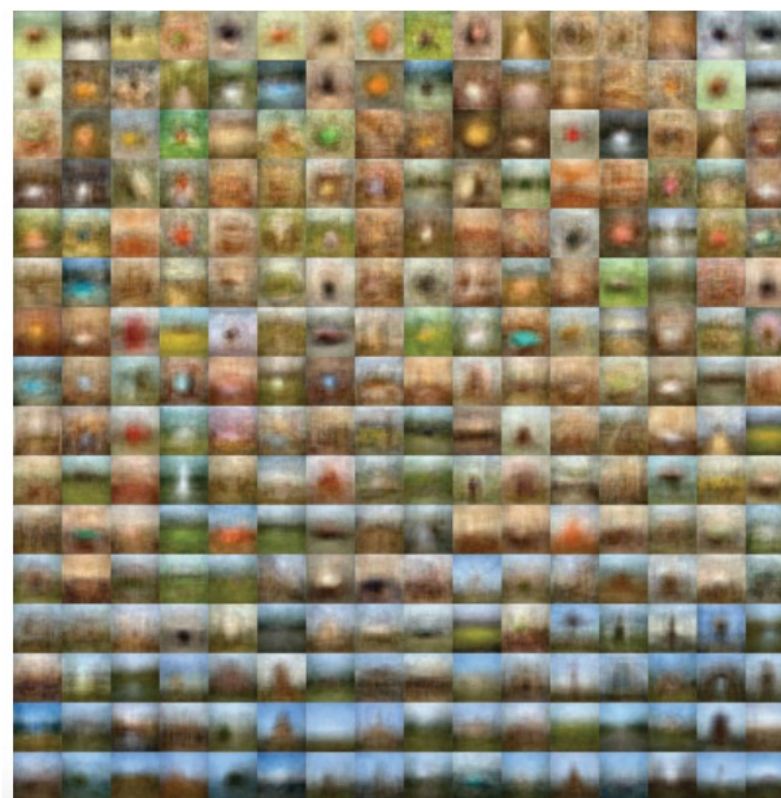
# Comparing Representations Learned When Training AlexNet on Different Datasets

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

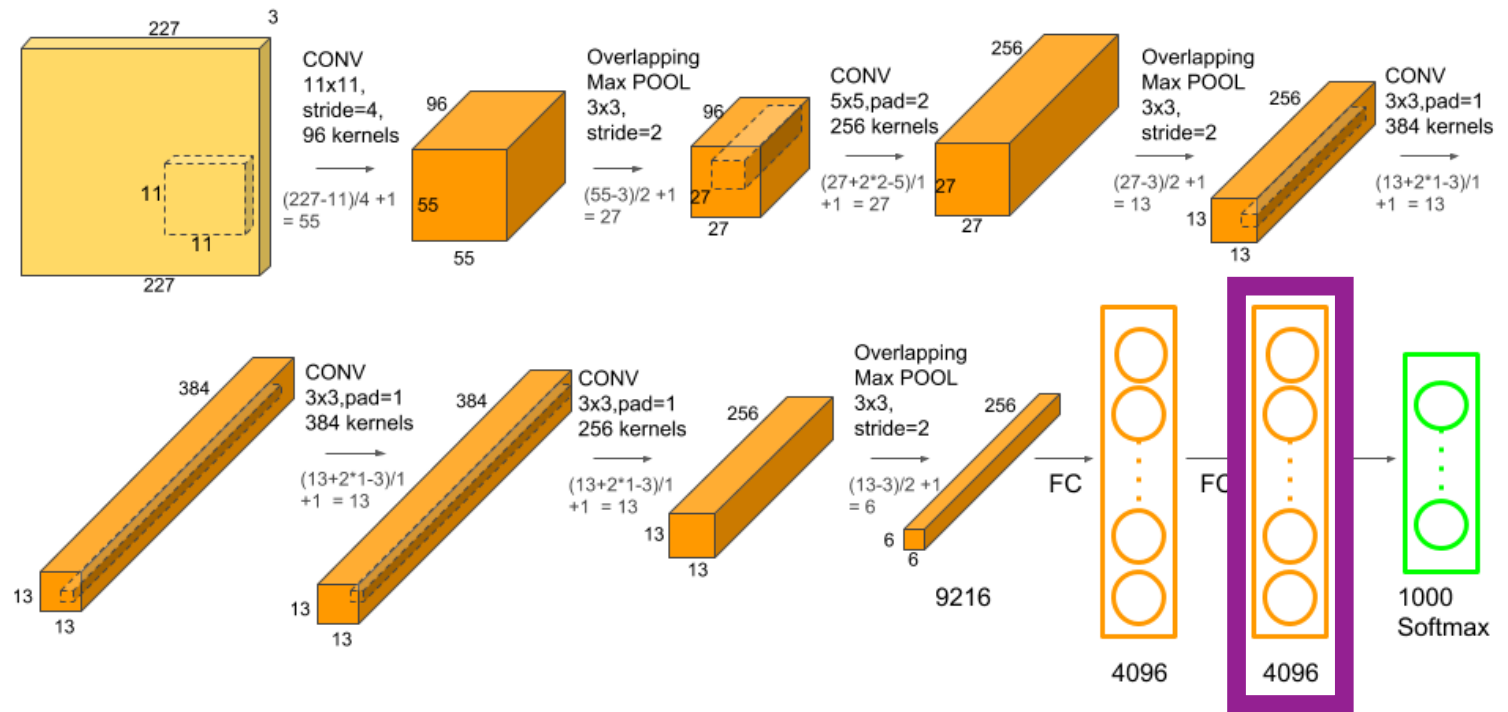


ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

# 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)

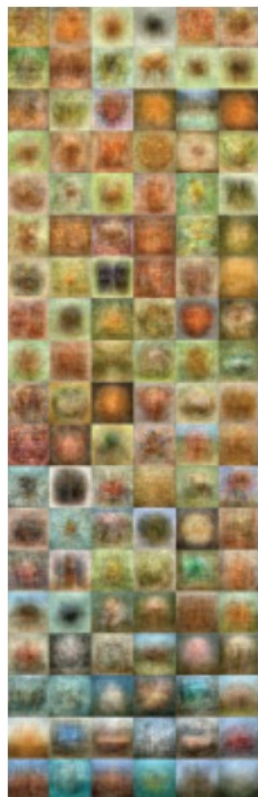


Source: <https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png>

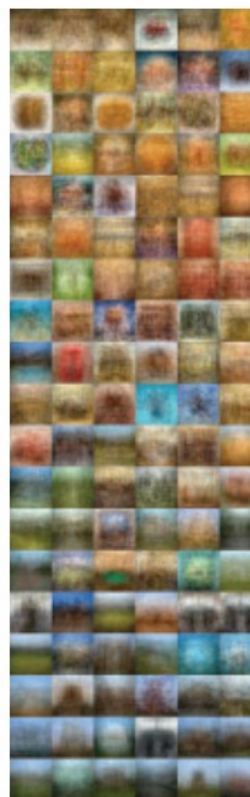
# Comparing Representations Learned When Training AlexNet on Different Datasets

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

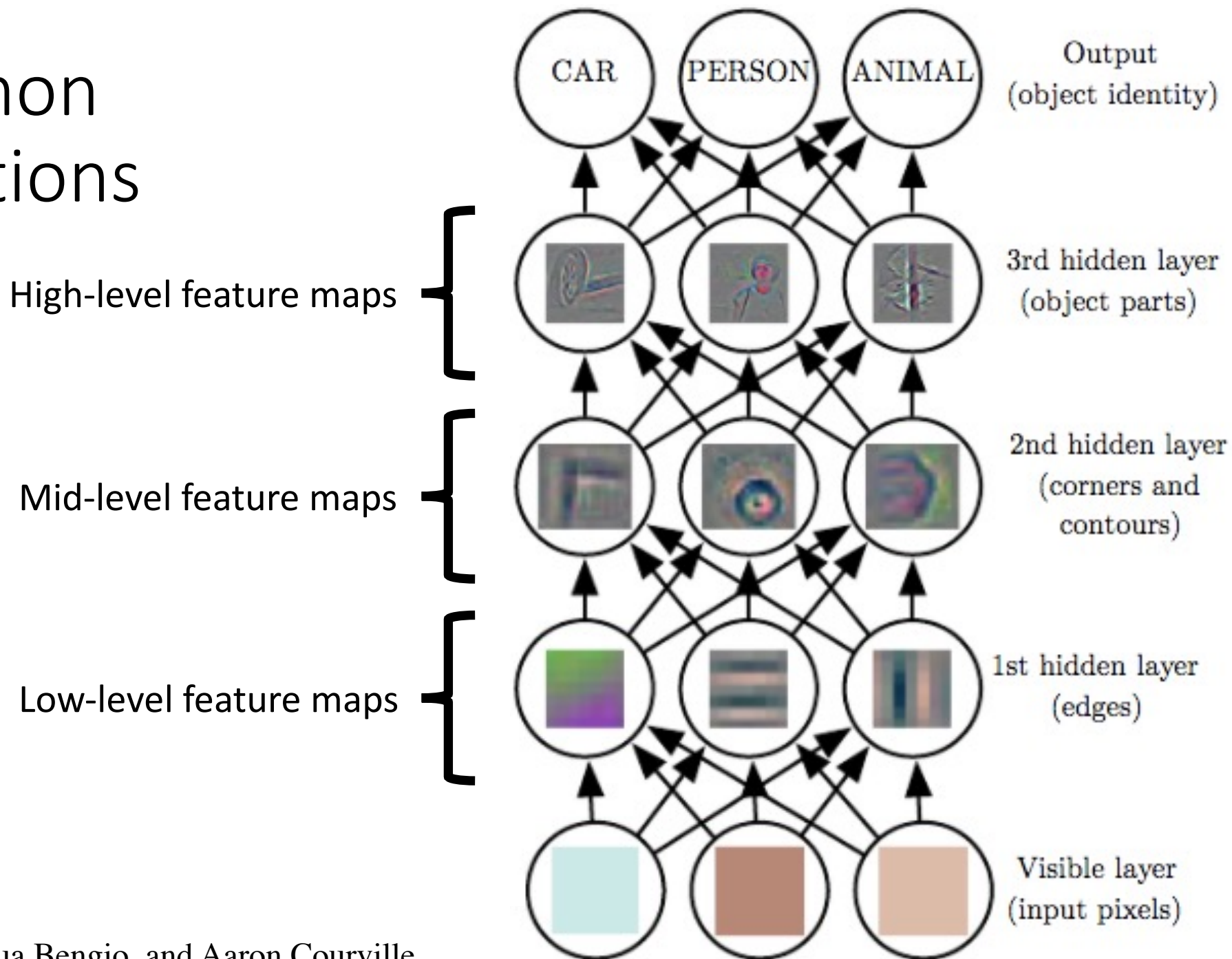


Places -CNN

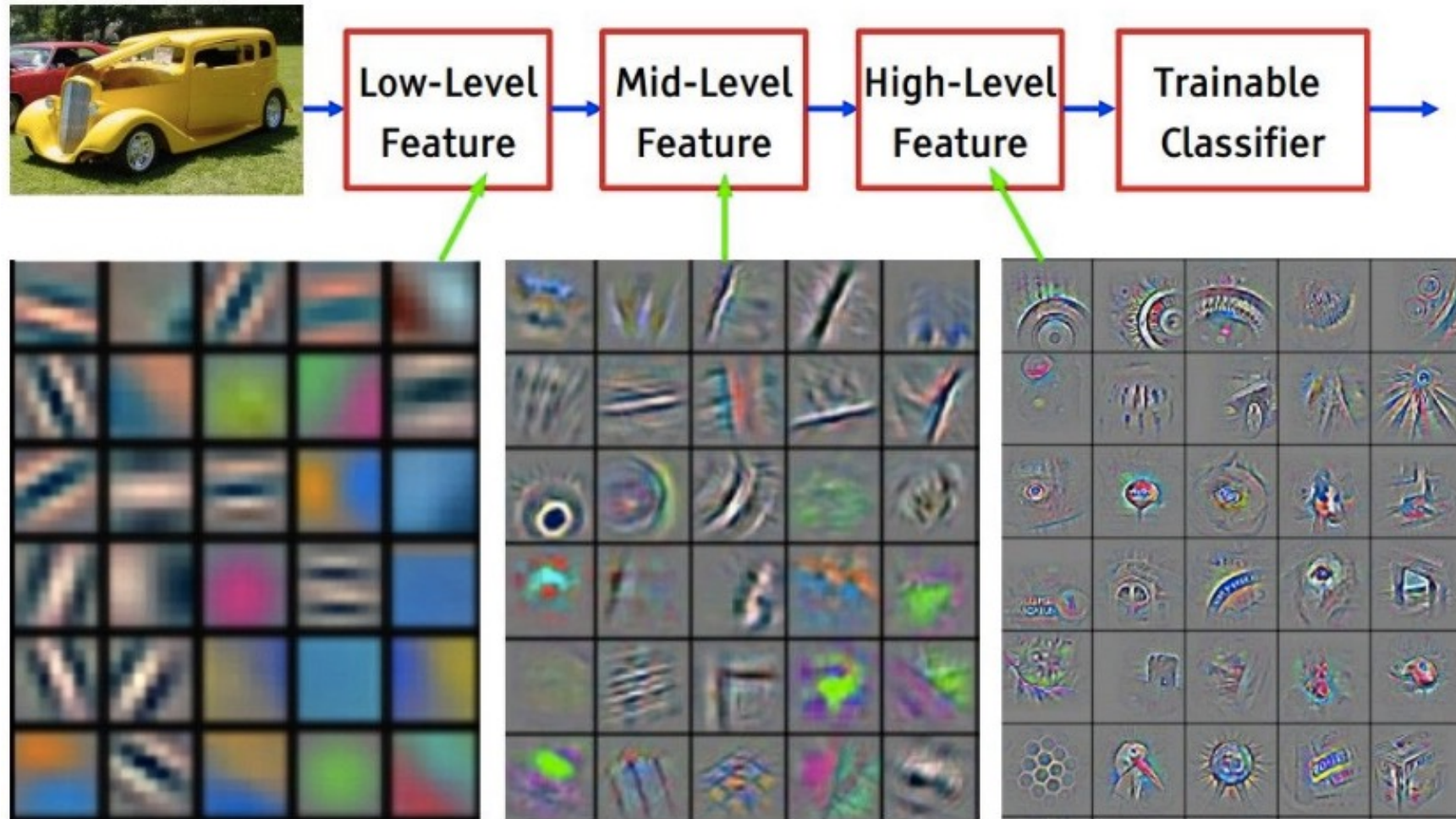


ImageNet-CNN units more often fire on blob-like structures than landscape-like structures

# CNN: Common Representations



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





# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

# Attribute Definition

## Description

(as opposed to naming)



How would you describe this scene?

# Attribute Definition

## Description

(as opposed to naming)



How would you describe this object?

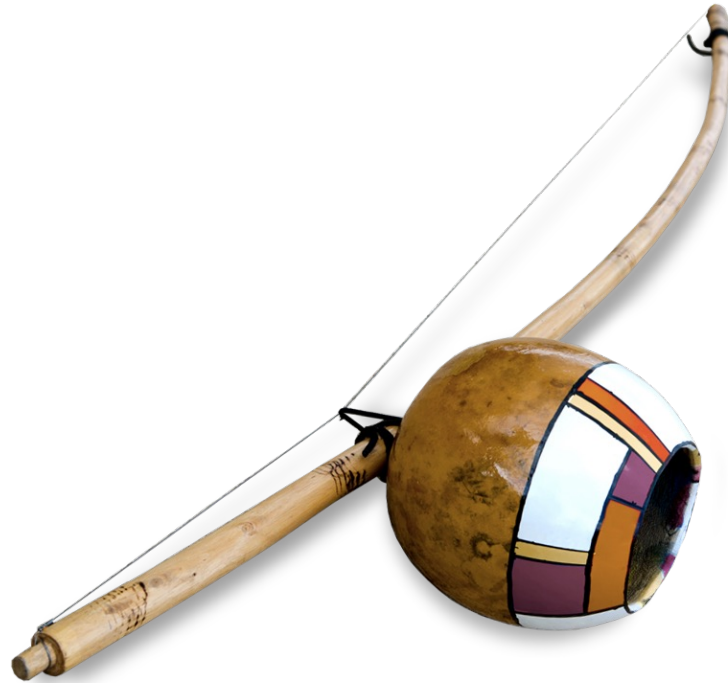
# Attribute Definition

\* 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?

# Relative Attributes (Rather Than Categorical)

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



# Application: Bird Recognition

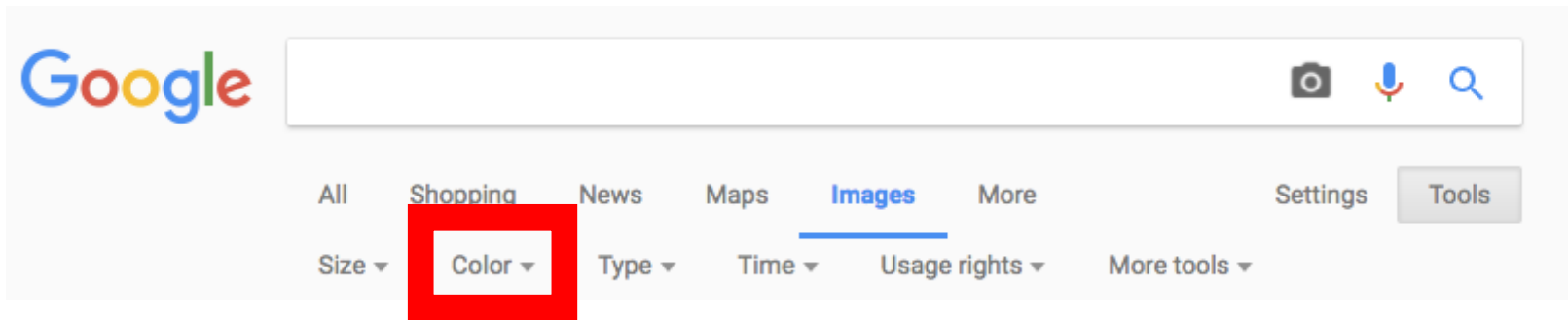
e.g., recognize objects with common knowledge instead of expert knowledge



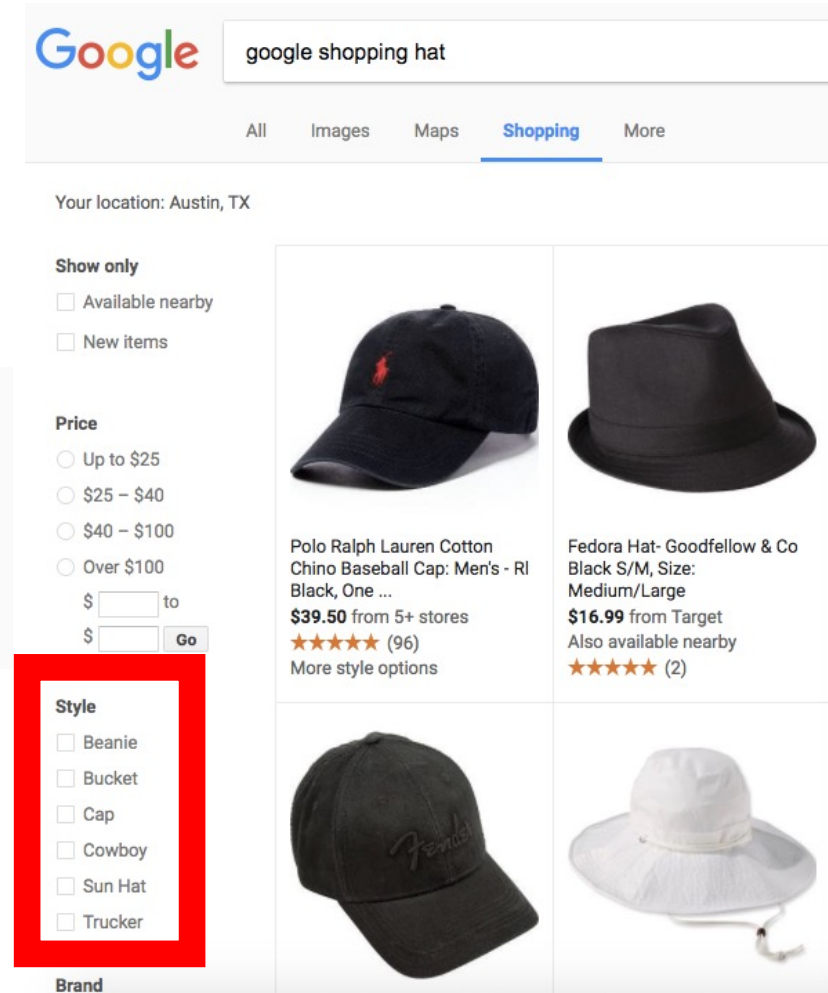
e.g., iBird: describe a bird to learn what type it is

Demo: [https://www.youtube.com/watch?v=J1C-Q-z\\_np0](https://www.youtube.com/watch?v=J1C-Q-z_np0)

# Application: Expedite Search



e.g., Image Search



e.g., Clothes Shopping

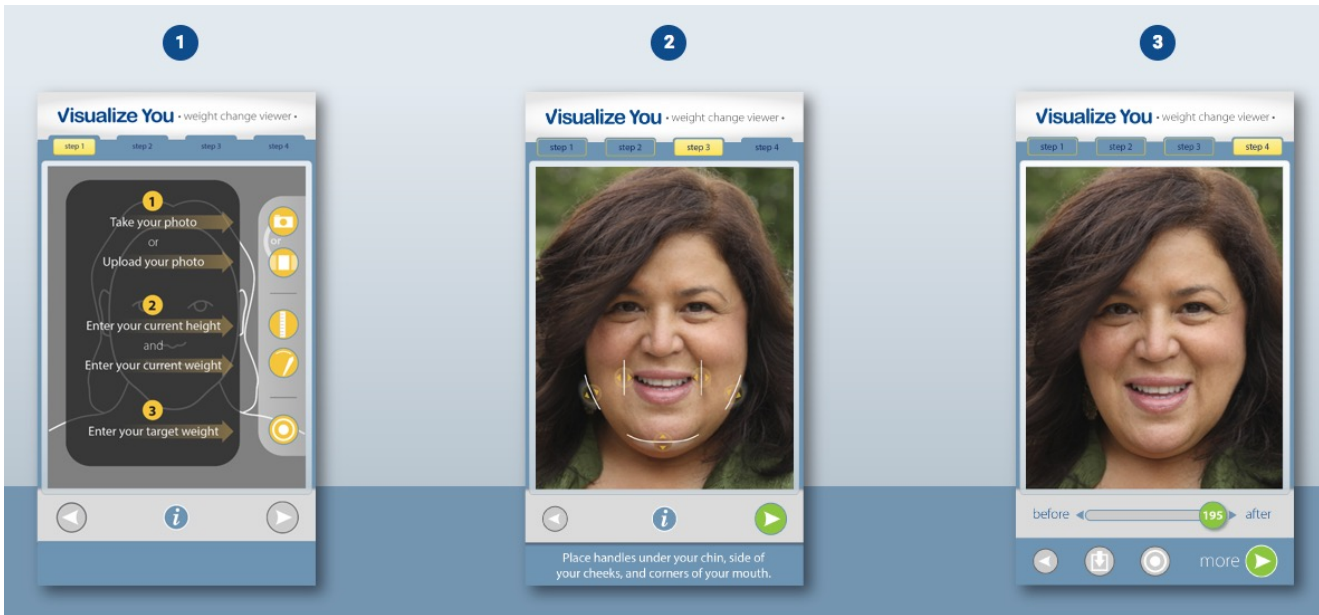
# Application: Shoe Shopping

The screenshot displays the 'Whittle Search' application interface. On the left, a search prompt 'Find Shoes like the one below' is accompanied by an image of a yellow and grey sneaker. The main control area, highlighted with a red border, features a search bar and three sliders for adjusting attribute strengths: 'BrightColored', 'Feminine', and 'Sporty'. Each slider has 'Less' and 'More' labels and a central handle. Below the sliders, a small image of a pink shoe is shown, indicating the current search results. At the bottom, a grid of 14 shoe images is presented for user feedback, with a prompt: 'Give feedback using images below as references | Indicate more/less of an attribute than the reference image'. The interface also includes a copyright notice '© All rights reserved' and a small green checkmark icon.

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



# Application: Altering Appearance



e.g., simulate weight loss/gain  
[www.visualizeyourweight.com](http://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 subject

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.,



# Challenges of Attribute Labeling

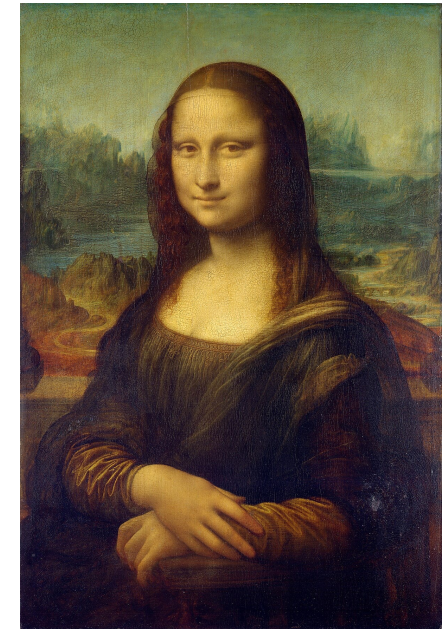
Is this drinkable?



What is the shape of the flag?

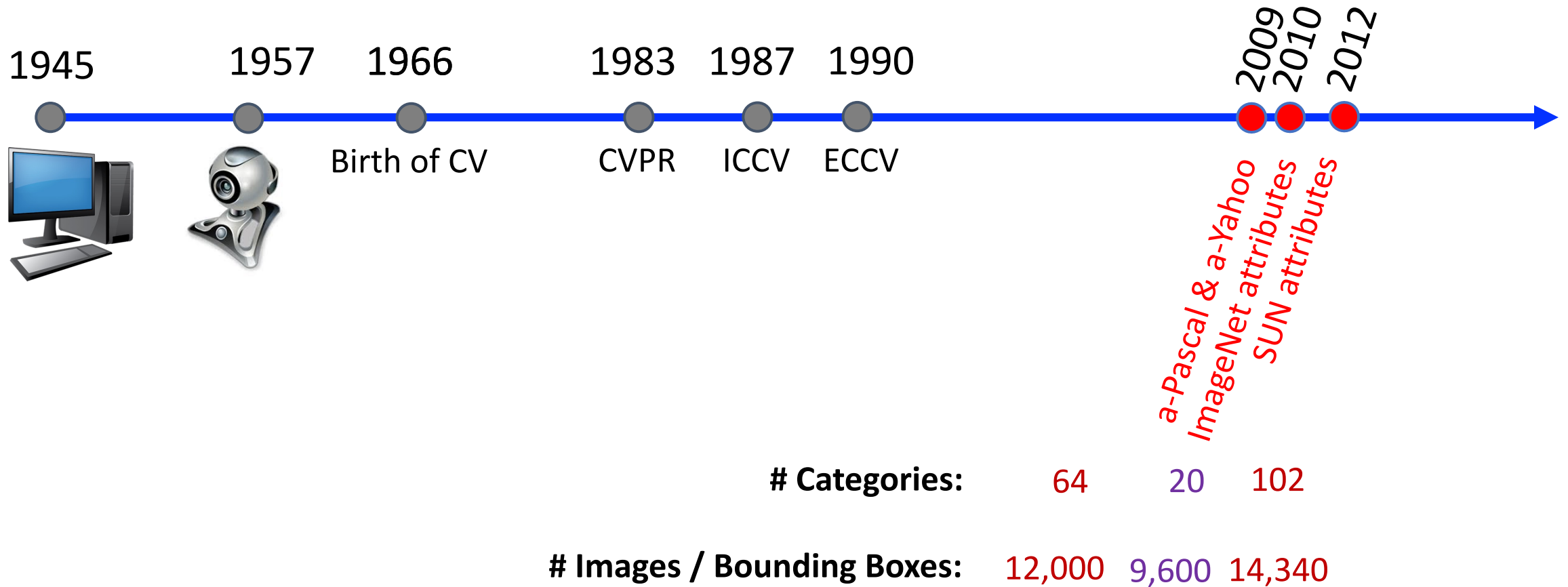


Is this person smiling?



What label to agree on for each task and why?

# Attribute Recognition Datasets



**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



# 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”

# 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



# 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

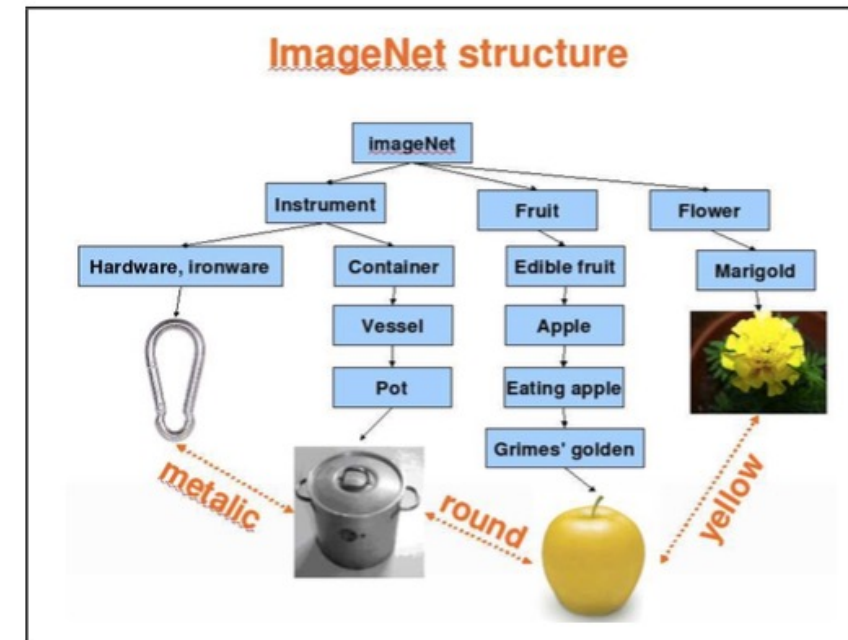
## 1. Image Collection

- 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



# 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

# Dataset: ImageNet Attributes

<p><b>metallic</b></p>	<p>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)</p> 
<p><b>rectangular</b></p>	<p>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)</p> 
<p><b>yellow</b></p>	<p>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)</p> 

# 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

Which attributes distinguish the scenes on the left from the scenes on the right?



rock, warm, barren, natural |

# 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

# Dataset: SUN Attributes

## 1. Task Design

**Instructions:**

**Scene Attribute Labeling** When you mouse over one of the images, a larger version of that image will appear in the box below.

Click on the scenes below that contain the following lighting or material:

**camping** Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.



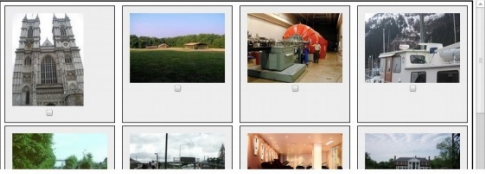
Example Scene Example Scene



These HITs are reviewed before being approved or rejected. [For further instructions Click Here!](#)

This task can be very subjective. If you are not sure about which images should be selected, please **\*SKIP THIS HIT\*** or email us to ask for clarification. There are more HITs with less subjective attributes.

**Interface:**



Images continued down the page ... ↓



# Dataset: SUN Attributes

## 1. Task Design

**Instructions:**

**Scene Attribute Labeling**

Click on the scenes below that contain the following lighting or material:

**camping** *Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.*



Example Scene Example Scene

When you mouse over one of the images, a larger version of that image will appear in the box below.

These HITs are reviewed before being approved or rejected.

[For further instructions Click Here!](#)

This task can be very subjective. If you are not sure about which images should be selected, please **\*SKIP THIS HIT\*** or email us to ask for clarification. There are more HITs with less subjective attributes.

**Interface:**



Images continued down the page ... ↓

**Scene Attribute Labeling**

Click on the scenes below that contain the following lighting or material:

**camping** *Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.*



Example Scene Example Scene

When you mouse over one of the images, a larger version of that image will appear in the box below.



These HITs are reviewed before being approved or rejected.

[For further instructions Click Here!](#)

This task can be very subjective. If you are not sure about which images should be selected, please **\*SKIP THIS HIT\*** or email us to ask for clarification. There are more HITs with less subjective attributes.

# Dataset: SUN Attributes

## 1. Task Design

(grid of 48 images)

**Instructions:**

**Scene Attribute Labeling** When you mouse over one of the images, a larger version of that image will appear in the box below.

Click on the scenes below that contain the following lighting or material:

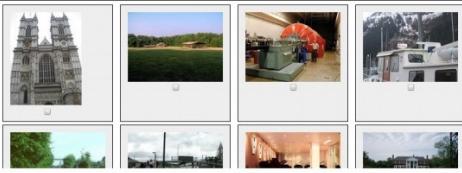
**camping** Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.

*Example Scene* *Example Scene*

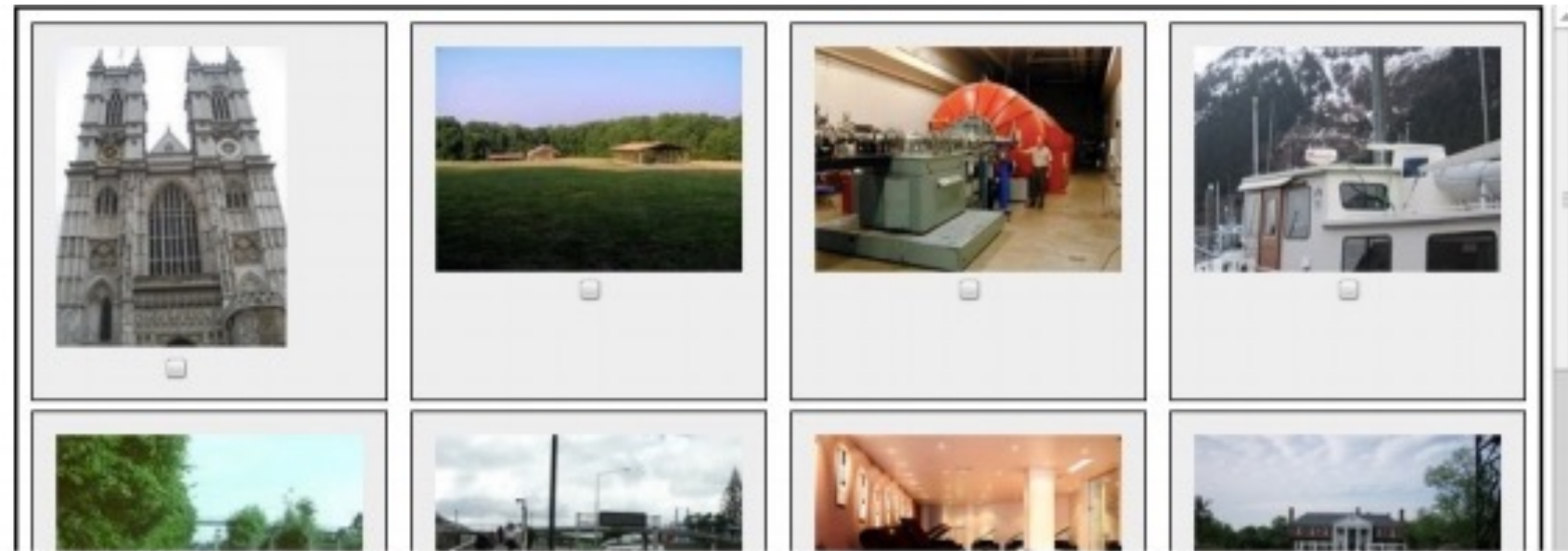
These HITs are reviewed before being approved or rejected. [For further instructions Click Here!](#)

This task can be very subjective. If you are not sure about which images should be selected, please \*SKIP THIS HIT\* or email us to ask for clarification. There are more HITs with less subjective attributes.

**Interface:**



Images continued down the page ... ↓



Images continued down the page ...



# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- **Student-led Lectures**

# Scene & Attribute Classification: Today's Topics

- Scene Classification Problem and Applications
- Scene Classification Datasets and Evaluation Metrics
- Scene Classification Models: Deep Features and Fine-Tuning
- Attribute Classification: Problem, Applications, and Datasets
- Student-led Lectures

The image features a dark gray background with a large, faint, circular glow in the center. A white film strip border, consisting of a series of rectangular sprocket holes, frames the entire scene. In the center of the glow, the words "The End" are written in a white, elegant, cursive script font. The text has a slight drop shadow, giving it a three-dimensional appearance as if it's floating within the scene.

*The End*