Scene Classification

Danna Gurari
University of Colorado Boulder
Fall 2021

https://home.cs.colorado.edu/~DrG/Courses/RecentAdvancesInComputerVision/AboutCourse.html

Review

- Last week:
  - ImageNet Challenge Top Performers
  - Baseline Model: AlexNet
  - VGG
  - ResNet
  - Discussion
- Assignments (Canvas)
  - Reading assignment was due today
  - New reading assignment out later today that is due next week
- Questions?

Scene Classification: Today’s Topics

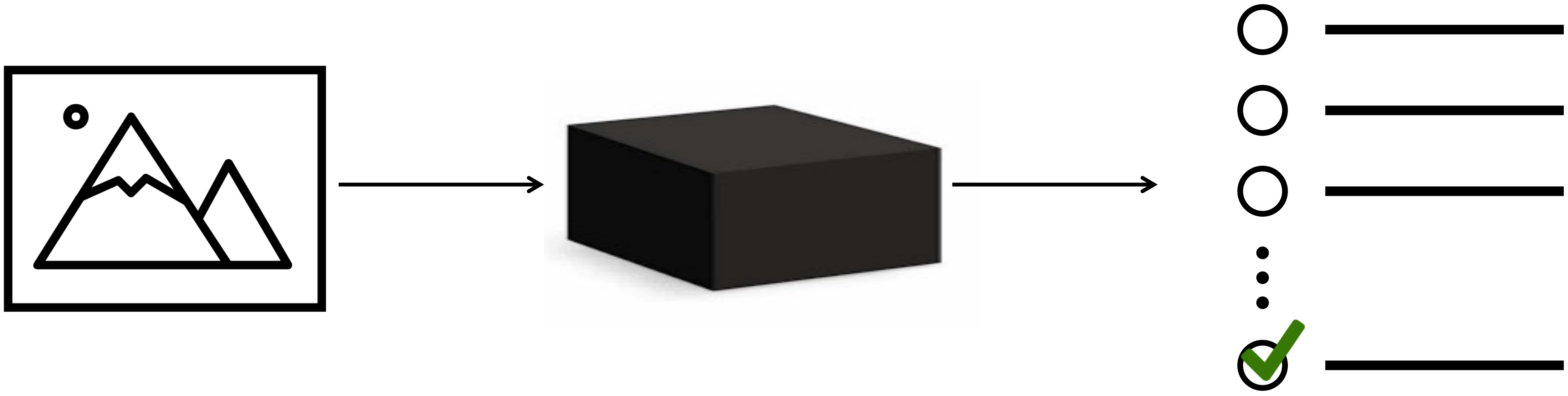
- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

Scene Classification: Today's Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

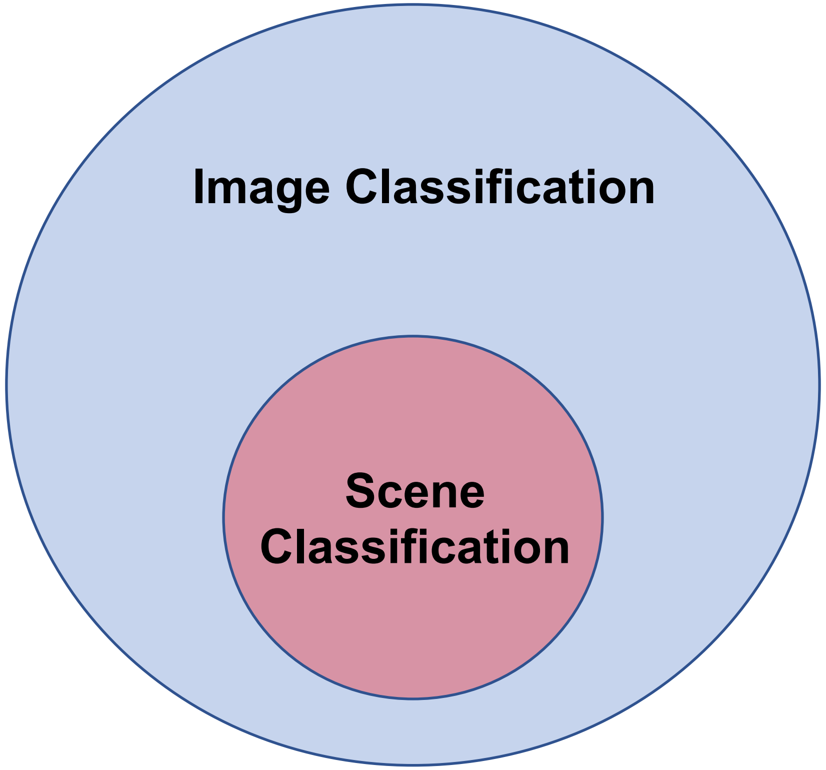
Scene Classification: Image Classification Problem

- Assign an image a label from a set of categories (i.e., multiple choice)



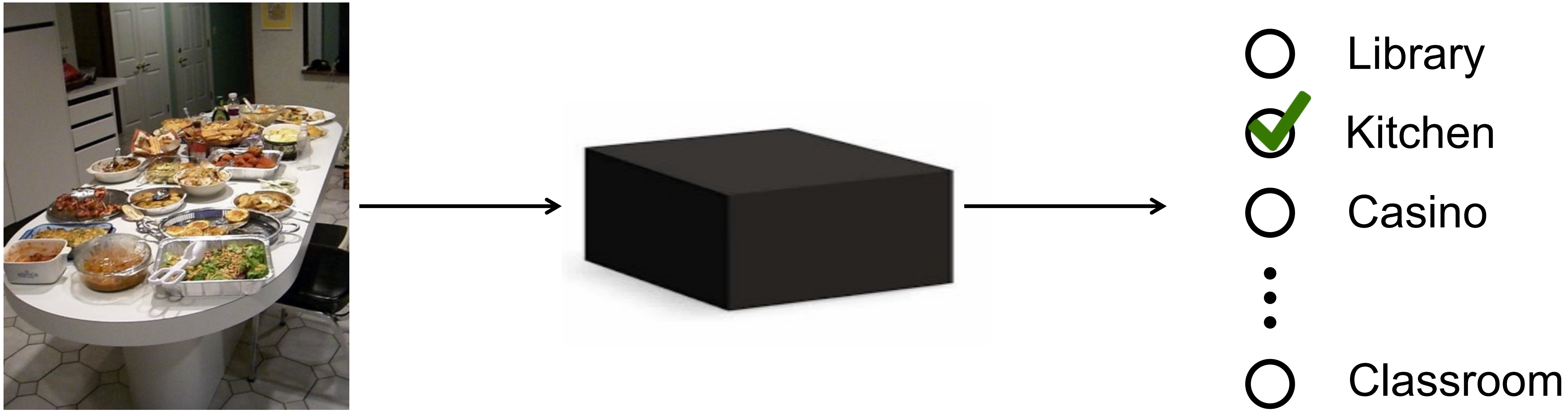
Scene Classification: Image Classification Problem

- Assign an image a label from a set of categories (i.e., multiple choice)



Scene Classification: Image Classification Problem

- Problem: What place is shown in the image?



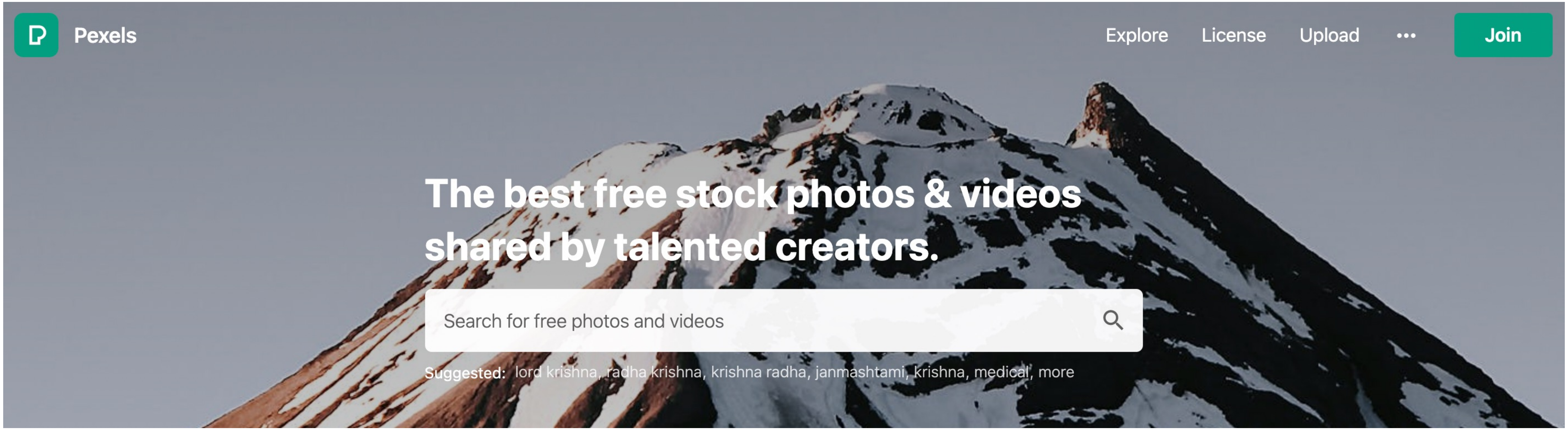
Scene Classification: Today’s Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

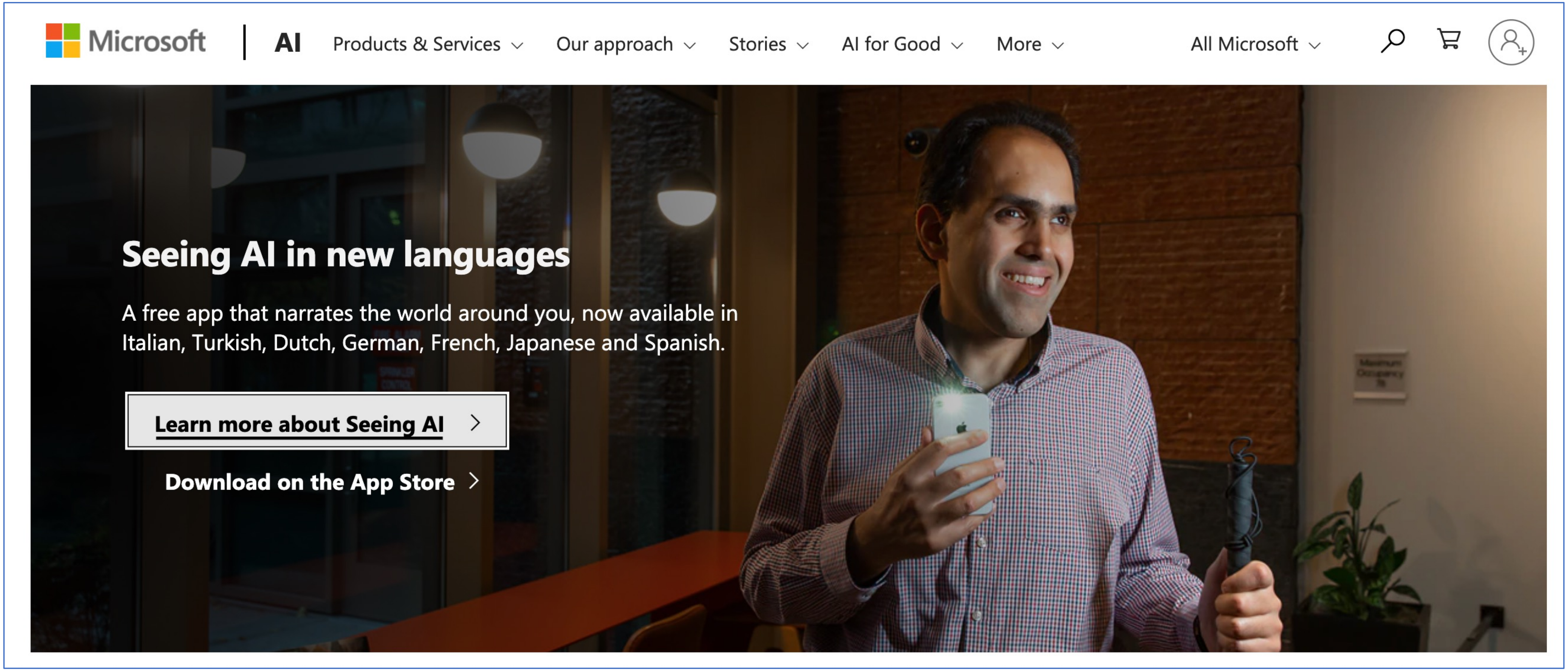
Photo Organization

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

Image Search



Assistive Technology



Seeing AI Demo: https://www.youtube.com/watch?v=R2mC-NUAmMk

Urban Planning

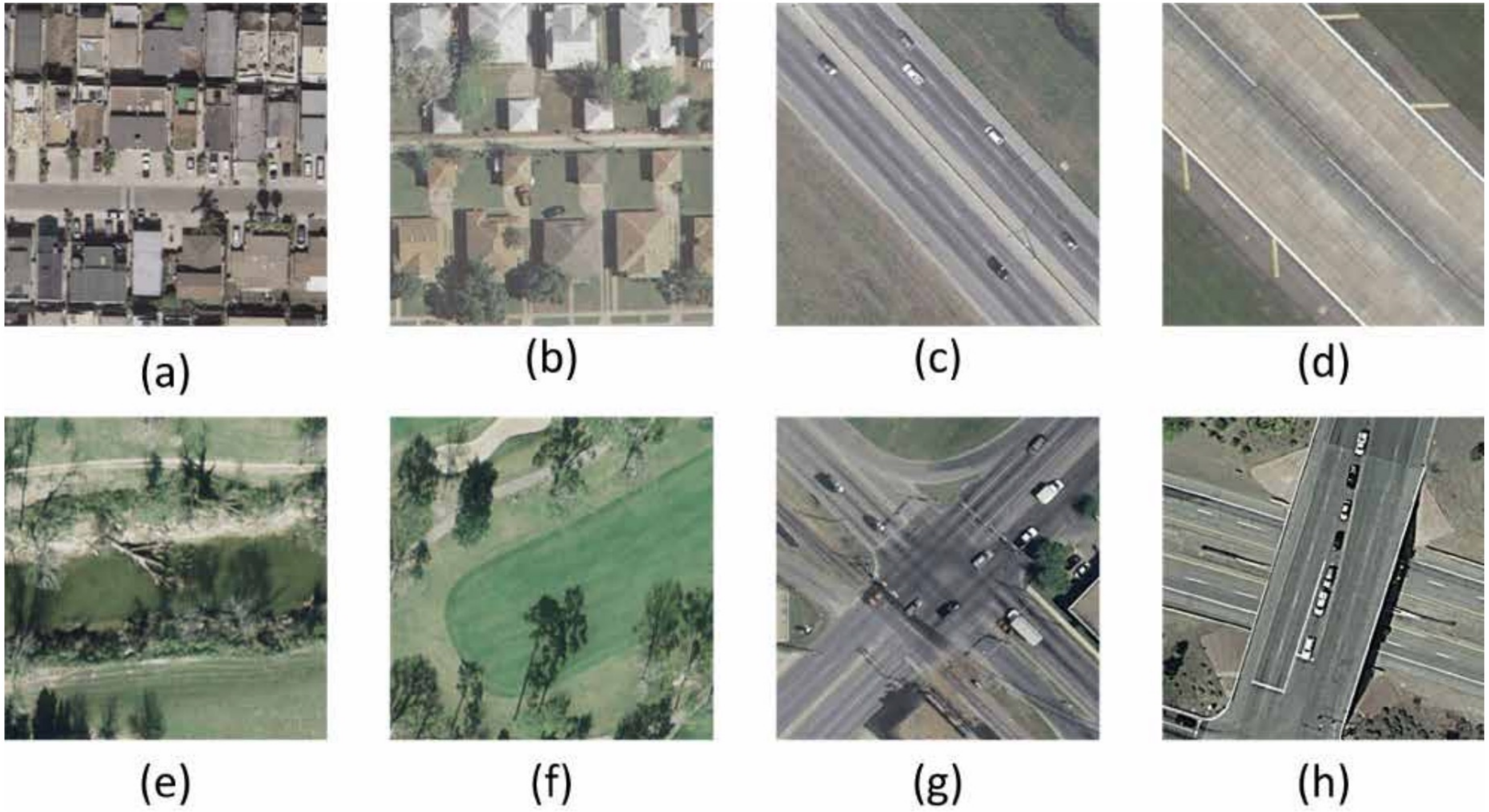
People's *well-being* is correlated with *scenic* places

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

Urban Planning, Natural Hazard Detection, and Environmental Monitoring (via Remote Sensing)



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

Can you think of any other potential applications?

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

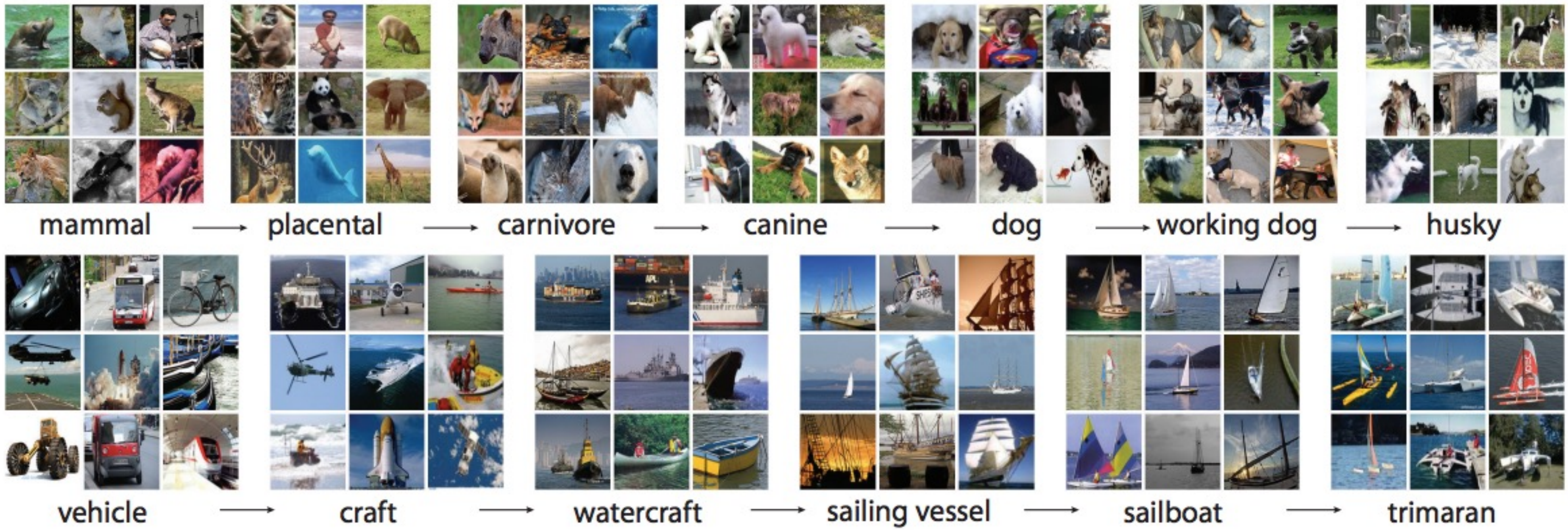
- Object Recognition
  - e.g., What would you expect (or not expect) to find in the scene [now, earlier, later]?
- Activity Recognition/Prediction
  - e.g., What would you expect people to do (or not do) in the scene [now, earlier, later]?

Scene Classification: Today’s Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

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)?

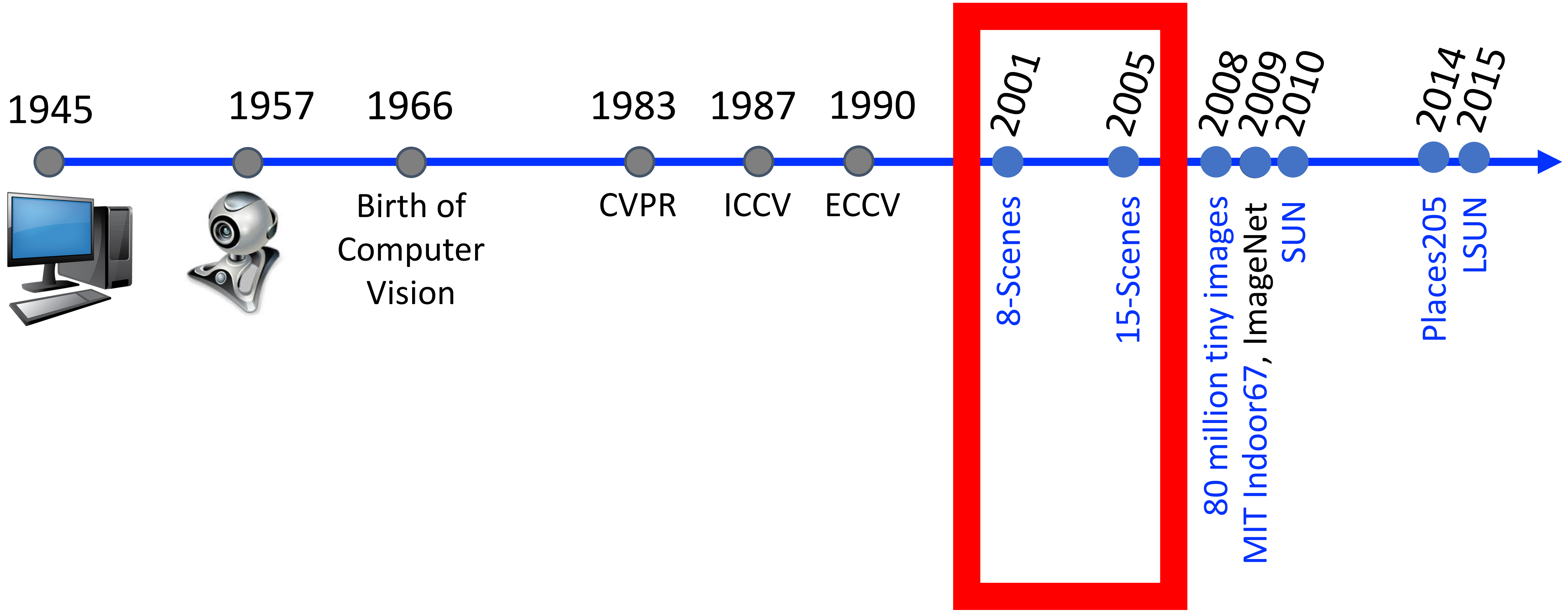
Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009.

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



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.

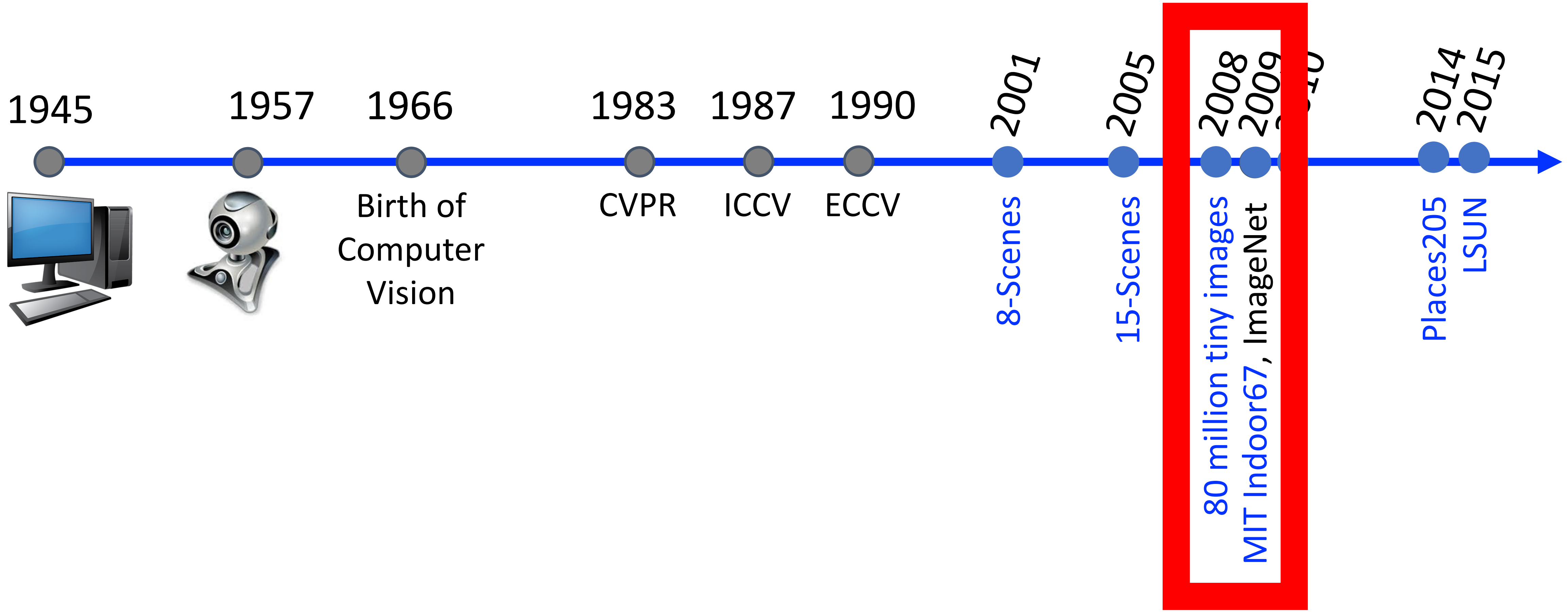
15-Scenes

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.

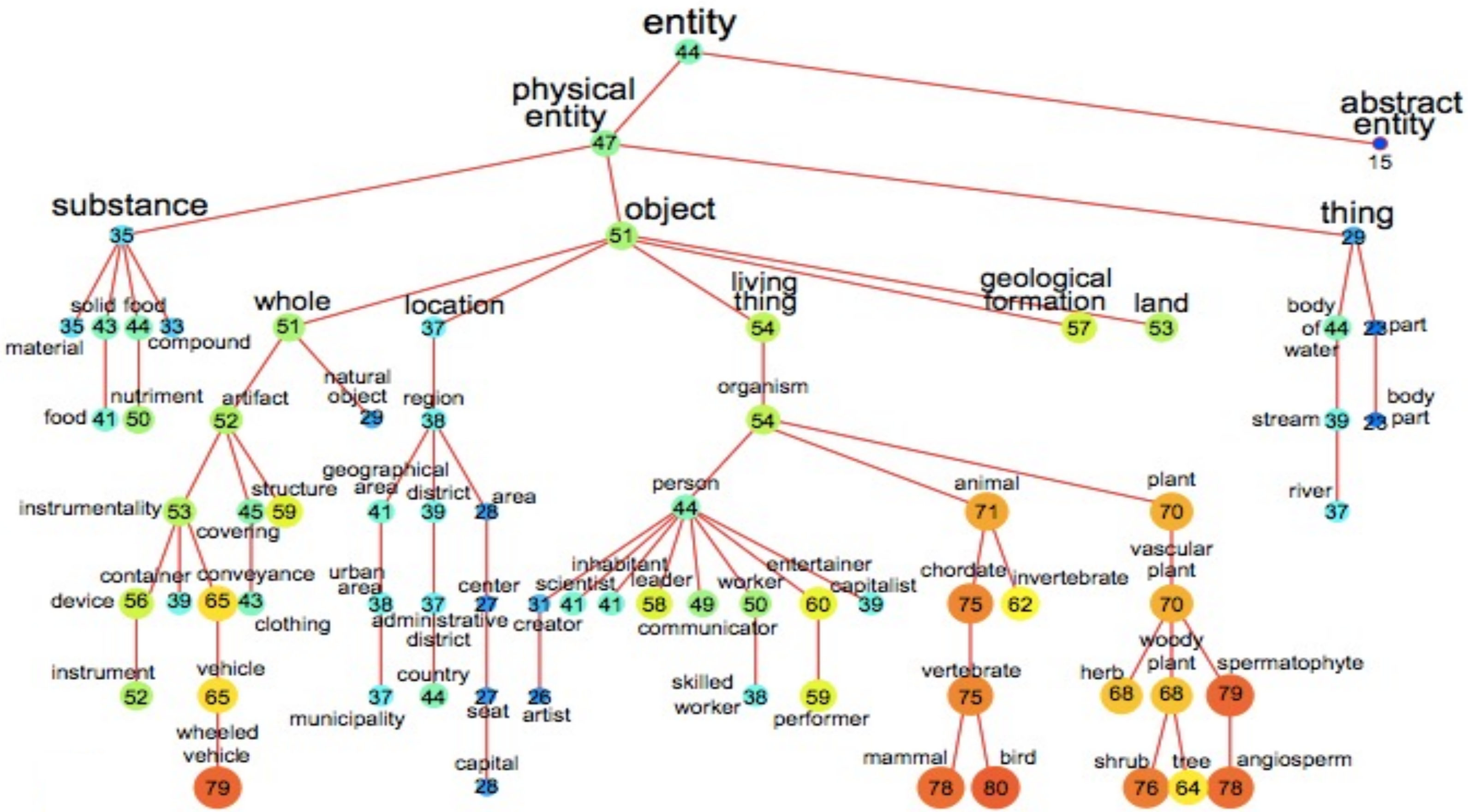
Scene Classification Datasets



80 Million Tiny Images

1. Category Selection

75,000 non-abstract nouns from WordNet



Antonio Torralba et al. 80 Million Tiny Images: A Large Dataset for Non-parametric Object and Scene Recognition. PAMI 2008.

80 Million Tiny Images

1. Category Selection

75,000 non-abstract nouns from WordNet

→

2. Image Collection

Images downloaded for 8 months from 7 online image search engines to 32x32 resolution

(Adapted from slides by Antonio Torralba)

Antonio Torralba et al. 80 Million Tiny Images: A Large Dataset for Non-parametric Object and Scene Recognition. PAMI 2008.

80 Million Tiny Images

1. Category Selection

75,000 non-abstract nouns from WordNet

2. Image Collection

Images downloaded for 8 months from 7 online
ir o

32x32 resolution

Why "tiny" images?

Antonio Torralba et al. 80 Million Tiny Images: A Large Dataset for Non-parametric Object and Scene Recognition. PAMI 2008.

80 Million Tiny Images

Why "tiny" images?

Idea: What resolution does a human need to recognize a scene?

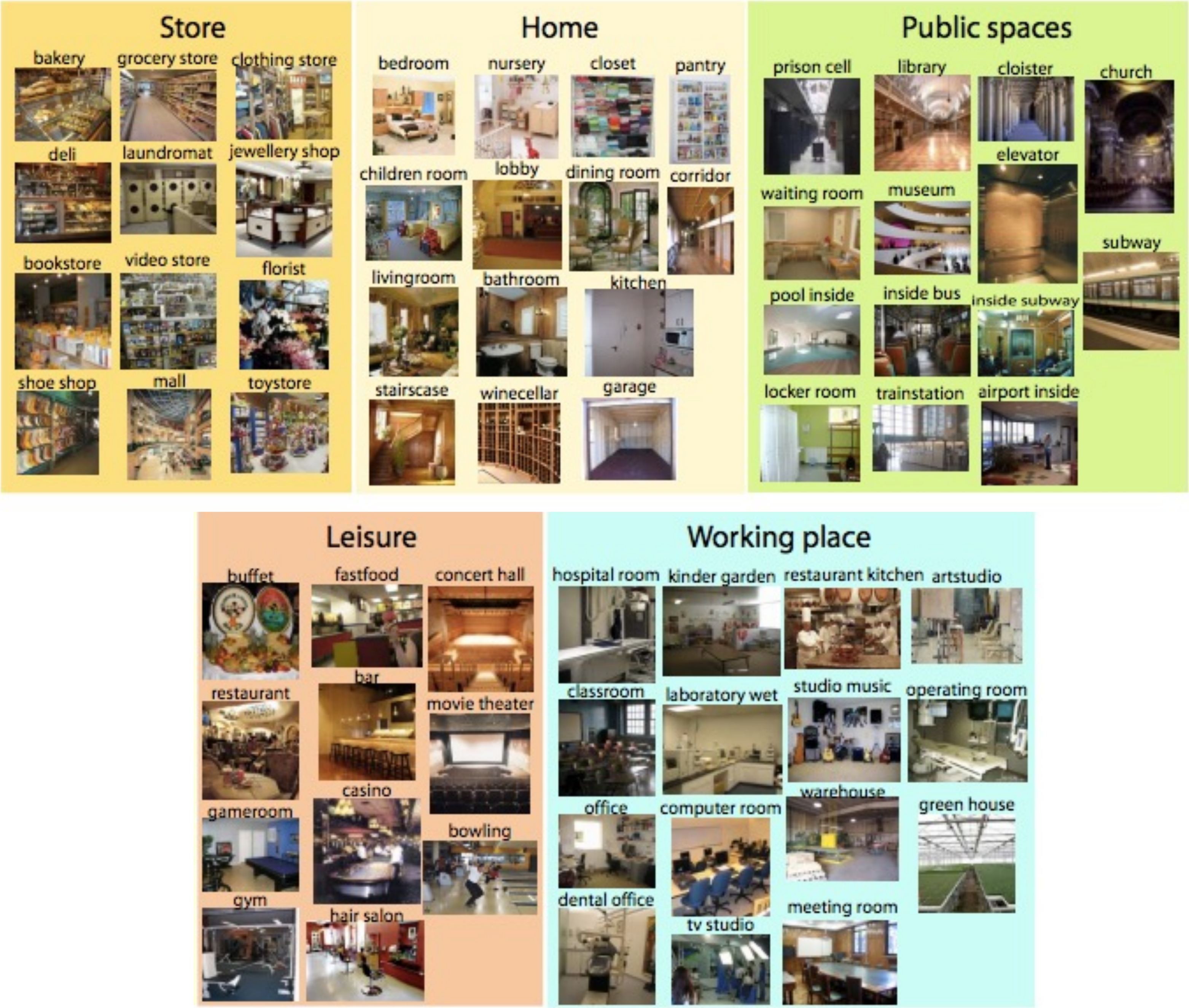
Study:
- 6 participants
- 585 color images
- Classify as 1 of 15 scene categories
- Images presented at 5 possible resolutions (8^2, 16^2, $\boxed{32^2}$, 64^2, 256^2)

Antonio Torralba et al. 80 Million Tiny Images: A Large Dataset for Non-parametric Object and Scene Recognition. PAMI 2008.

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 sites, and 1 vision dataset

LabelMe

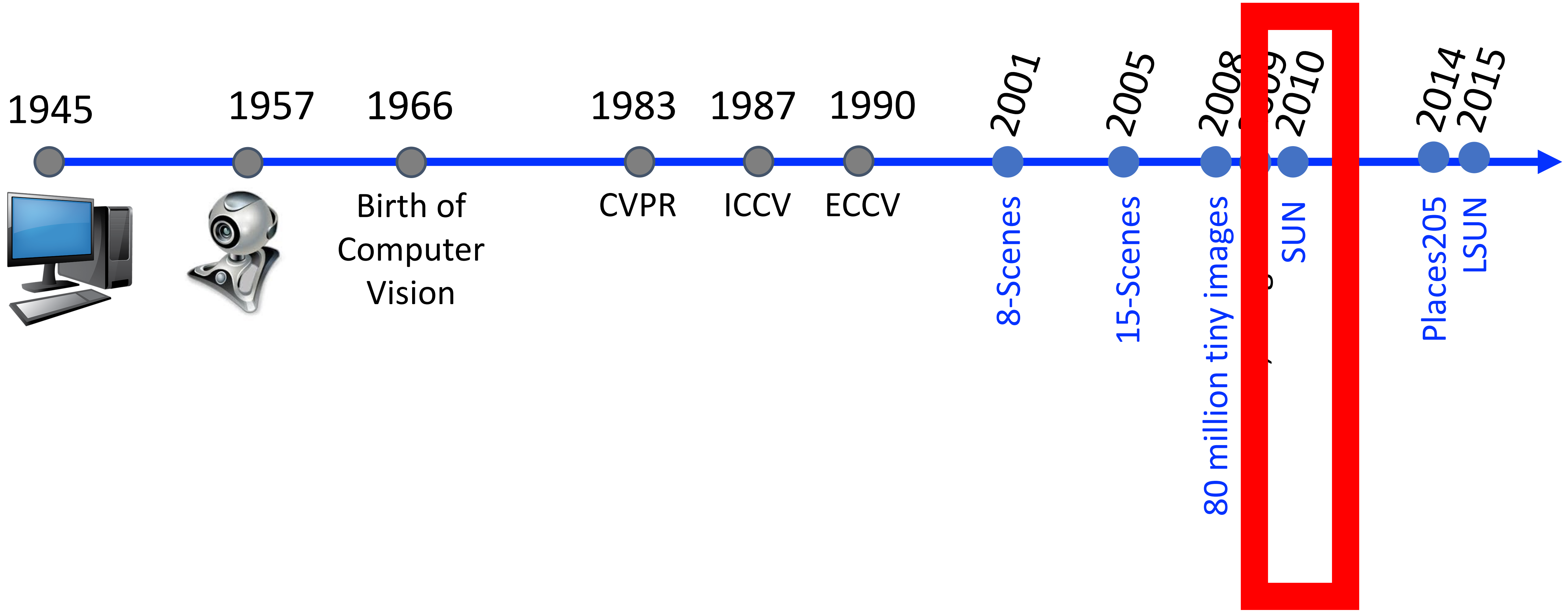
Google Image Search

altavista

flickr GAMMA

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

Scene Classification Datasets



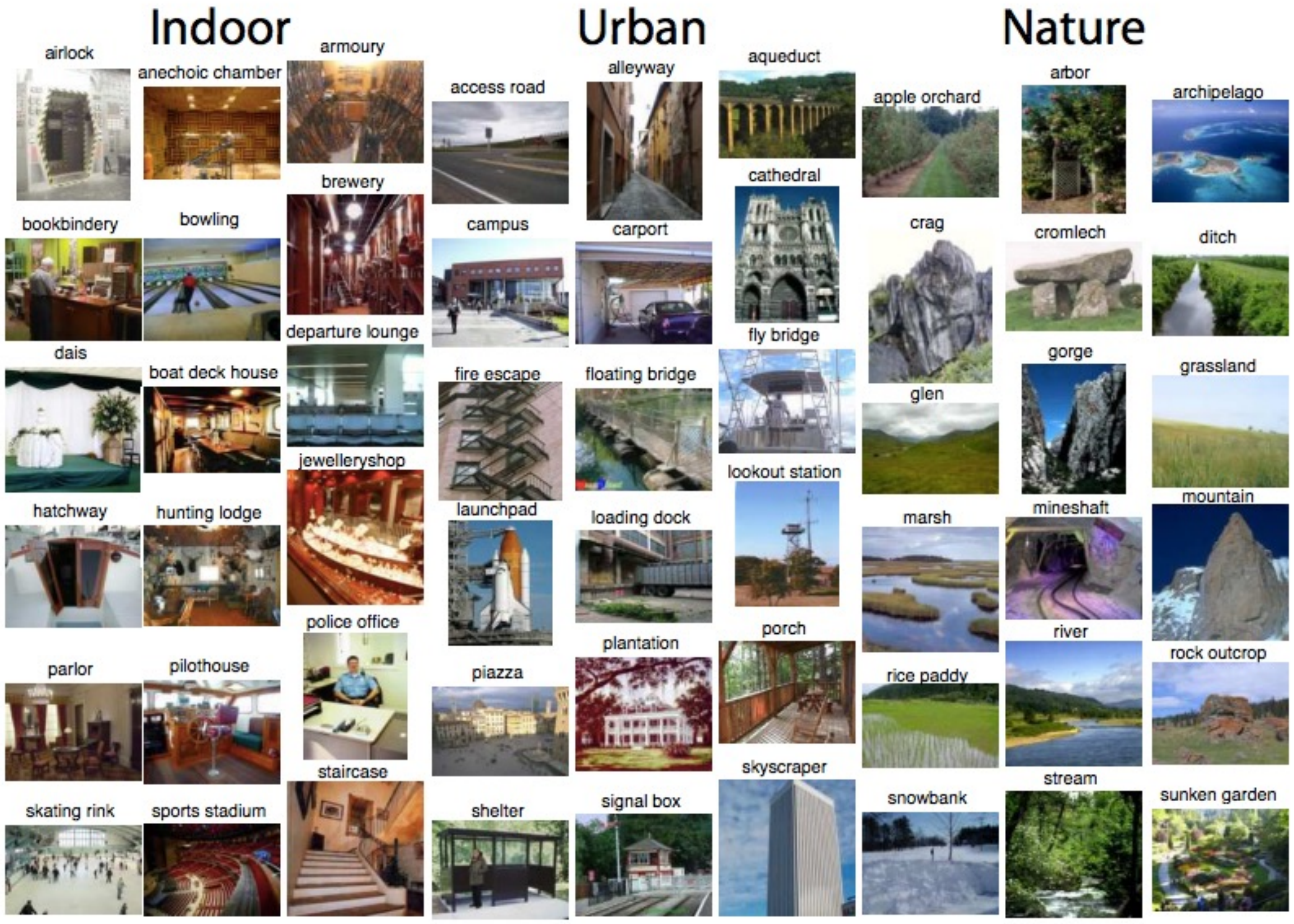
New Typical Process for Creating a Dataset

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



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.

New Typical Process for Creating a Dataset

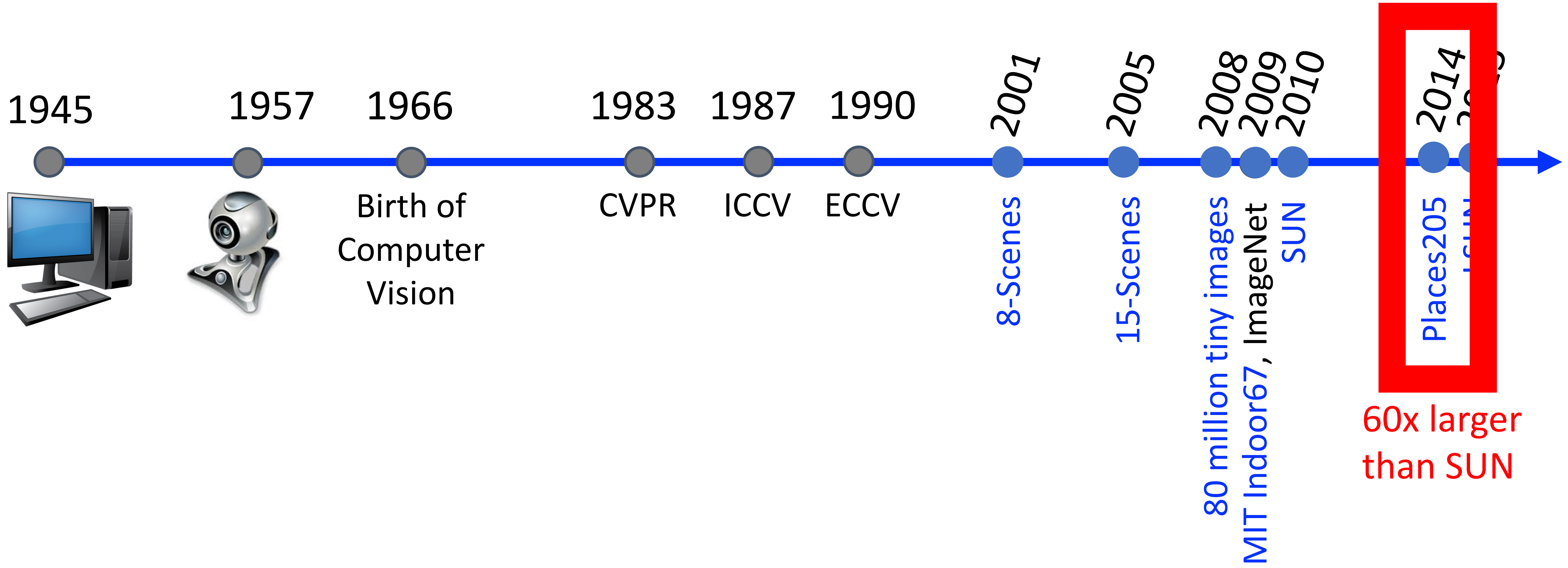
1. Category Selection

2. Image Collection

3. Human Verification

Limitation: relies on a few in-house people to verify images

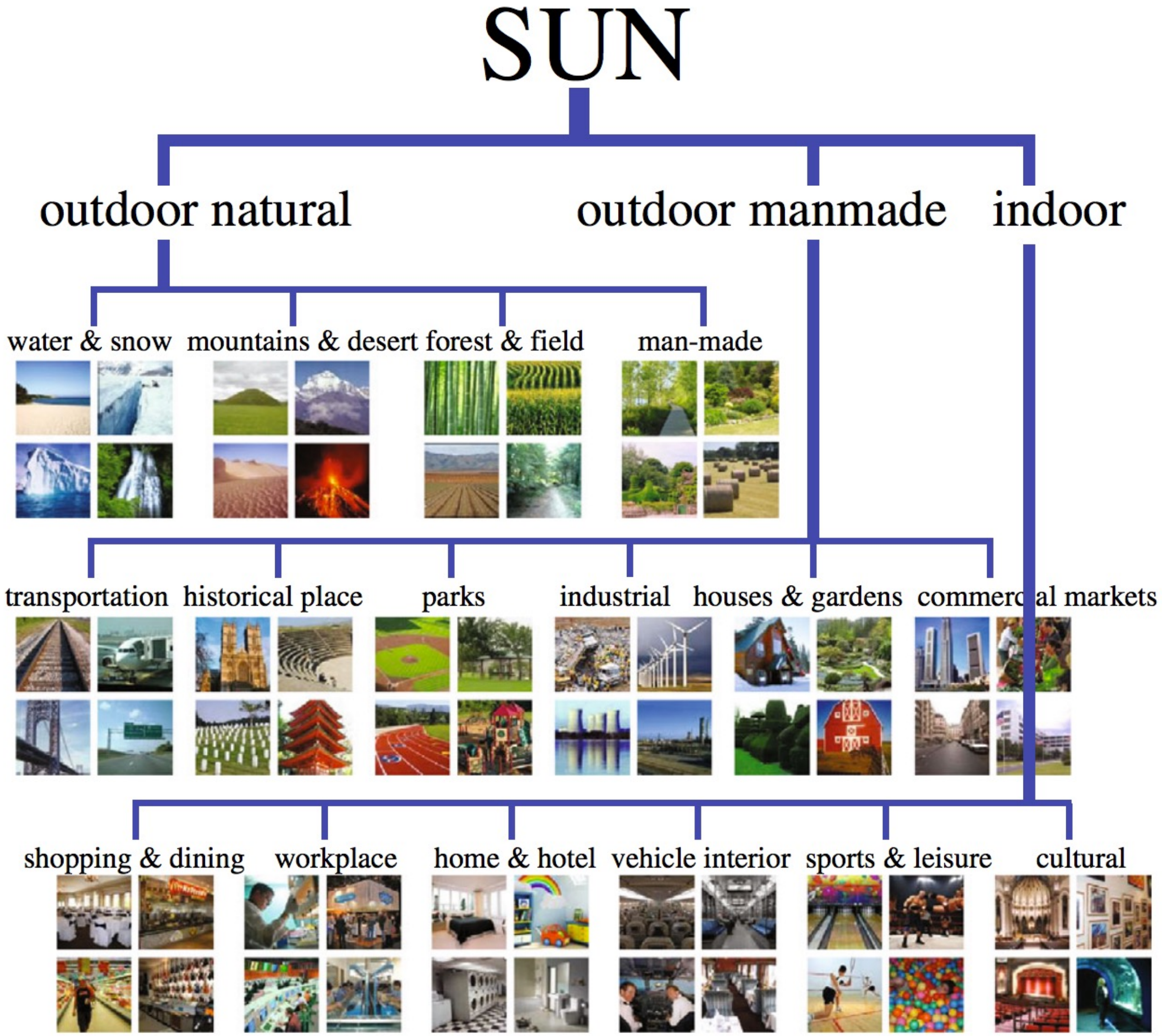
Scene Classification Datasets



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

1. Task Design

Instructions:

Interface:

User interface: Instructions

Examples

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

Places205

1. Task Design

Instructions:

Interface:

User interface: Task

Tasks left

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

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

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

→

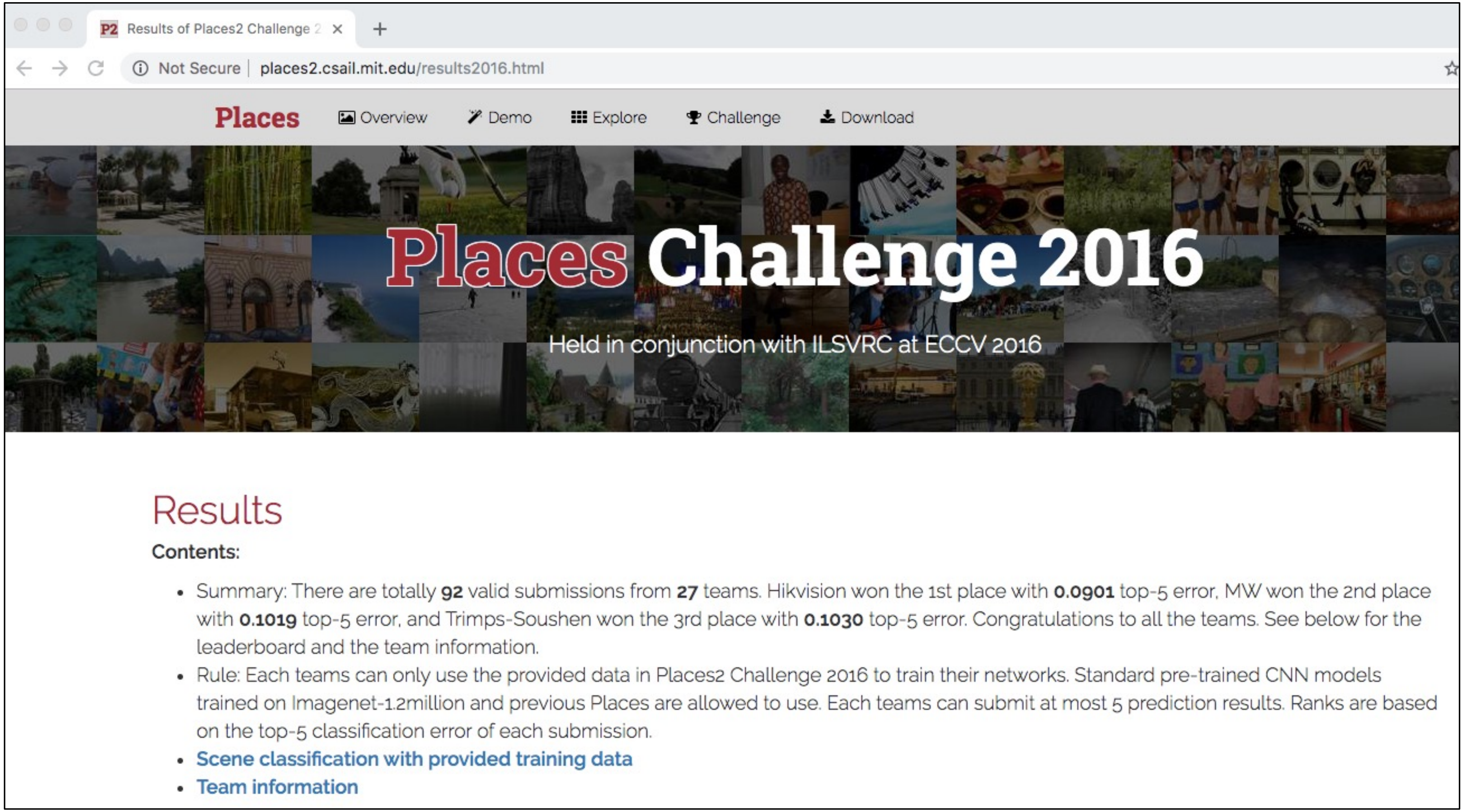
3. Human Verification

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

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

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

Scene Classification: Places Challenge


Results of Places2 Challenge 2

Not Secure | places2.csail.mit.edu/results2016.html

Places | Overview | Demo | Explore | Challenge | Download

Places Challenge 2016

Held in conjunction with ILSVRC at ECCV 2016

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**

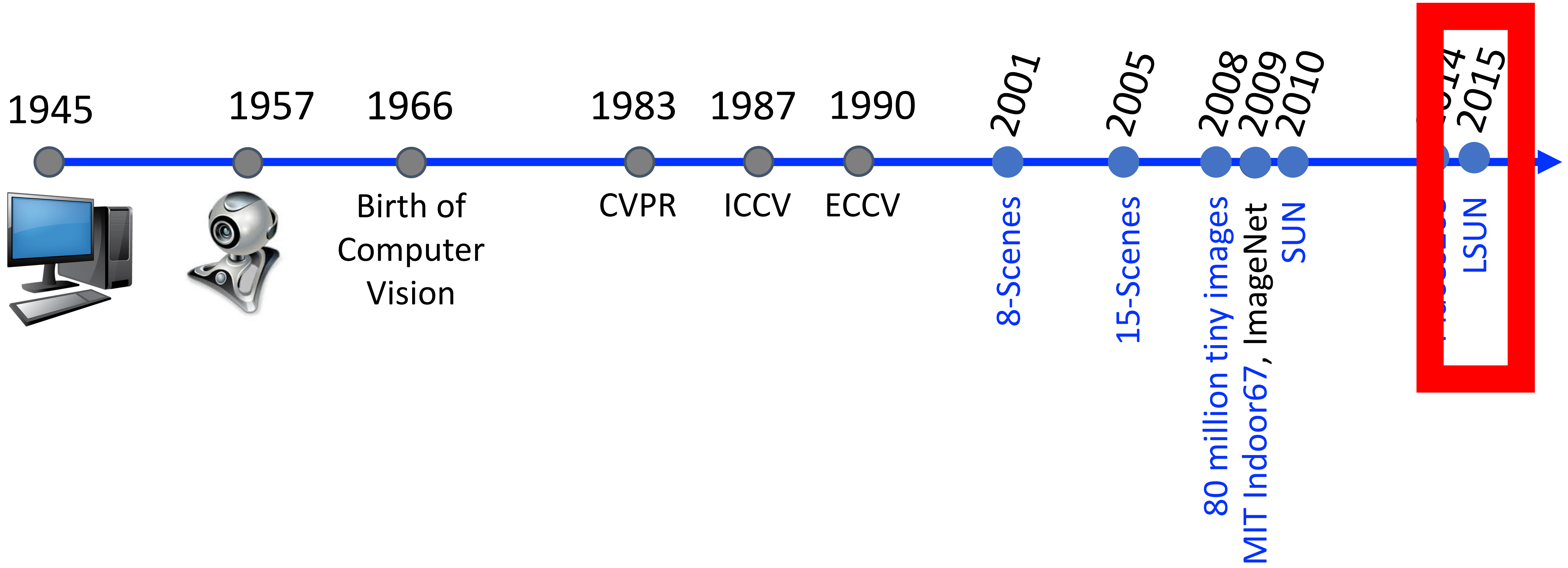
What Are Limitations of the Dataset?

- "… the authors did not translate their image queries like ImageNet did. By not translating these queries, the authors may have missed out on a good portion of images that represent non-eurocentric scenes; a model deployed using Places could be less accurate in classifying scenes in non English speaking countries."

- "It is doubtful how scenes can be categorized by adjectives, which can be quite subjective… For instance, in Figure 1 there are images of teenage bedrooms, romantic bedrooms, and so on. Are there really clear definitions in these categories? What if a teenager grows up and still uses the same room without remodeling - what category does that room fall in, then? Not to mention how subjective it is to define a room romantic. I find some of the sample images spooky in that category, and I would definitely not mark them as romantic."

- "… if an image contains the sea, the coast, and the mountains behind the coast, how should we categorize this image? In my opinion, scenes can be more than just a place or a category..."

Scene Classification Datasets



LSUN

1. Category Selection

10 scene categories from SUN

Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

LSUN

1. Category Selection

10 scene categories from SUN

2. Image Collection

- Downloaded images from Google Images; query terms were 696 common adjectives (messy, spare, sunny, desolate, etc) with each scene category for all 3-day time spans since 2009

- Automatically discarded images that are < 256 x 256

Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

LSUN

1. Category Selection

10 scene categories from SUN

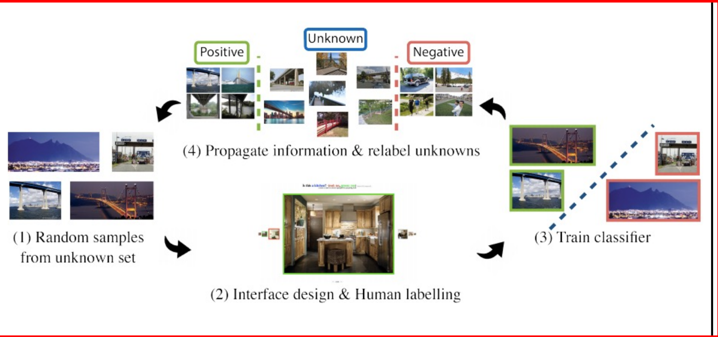
2. Image Collection

- Downloaded images from Google Images; query terms were 696 common adjectives (messy, spare, sunny, desolate, etc) with each scene category for all 3-day time spans since 2009

- Automatically discarded images that are < 256 x 256

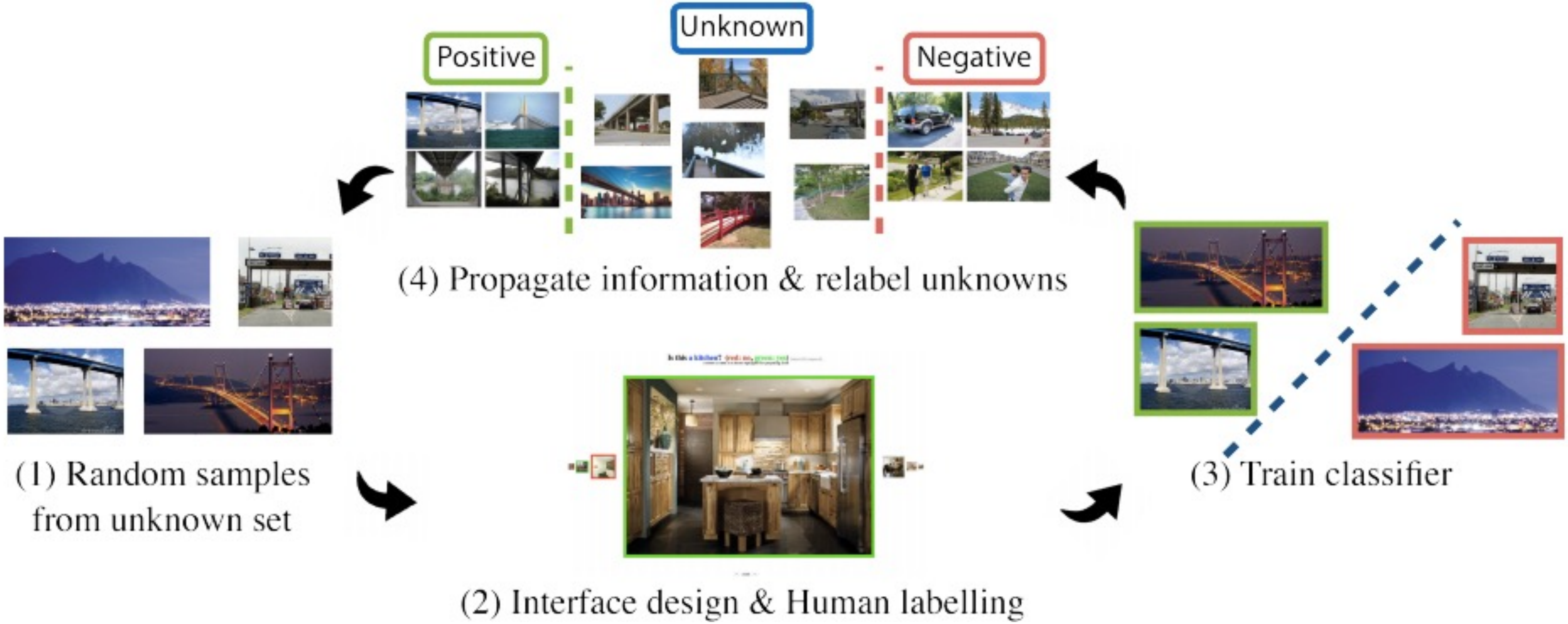
3. **Label** Verification

- Human in the loop



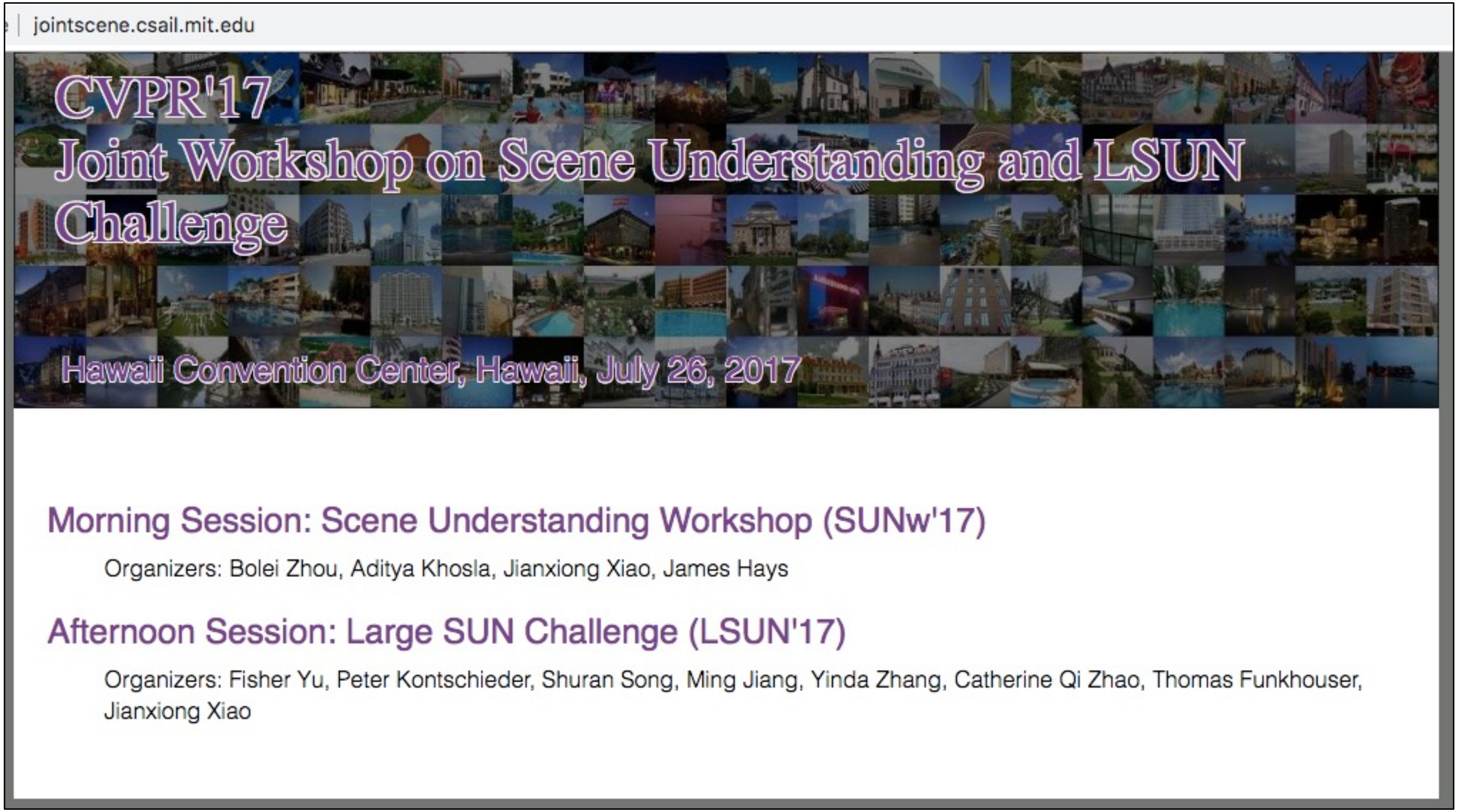
Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

LSUN Label Verification with Humans in the Loop



Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

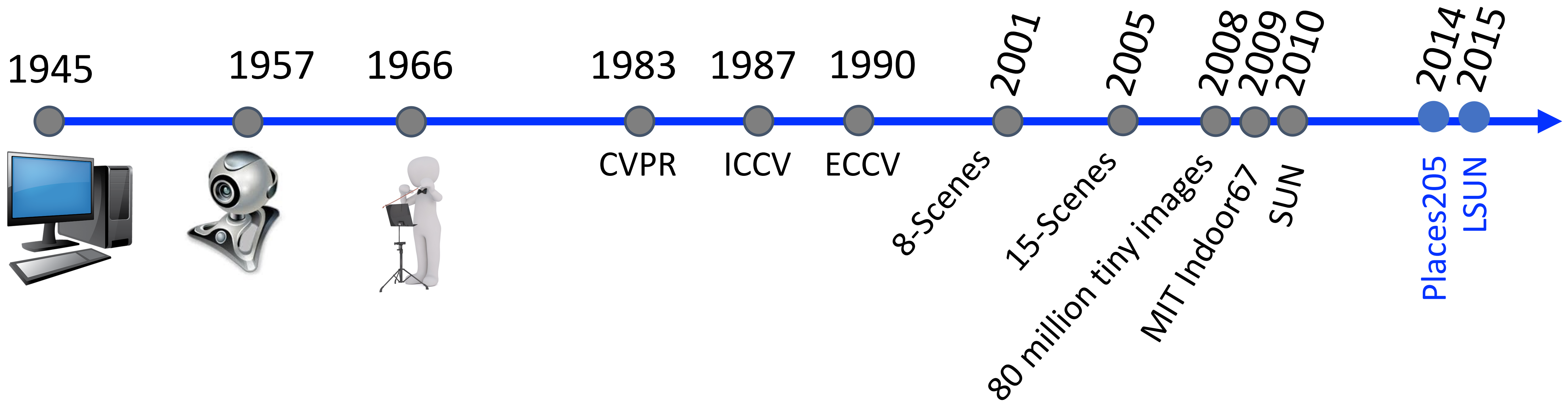
Scene Classification Datasets: LSUN Challenge



What Are Limitations of the Dataset?

- Same limitations as discussed for Places and…
- “It only contains 10 categories, which does not make it very practical as there are many scenes in the real world.”

Scene Classification Datasets



# Categories:	8	15	75,062	67	397	477	10
# Images:	300	4,485	79,302,017	15,620	130,519	2,448,873	10,000,000

Scene Classification: Today's Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

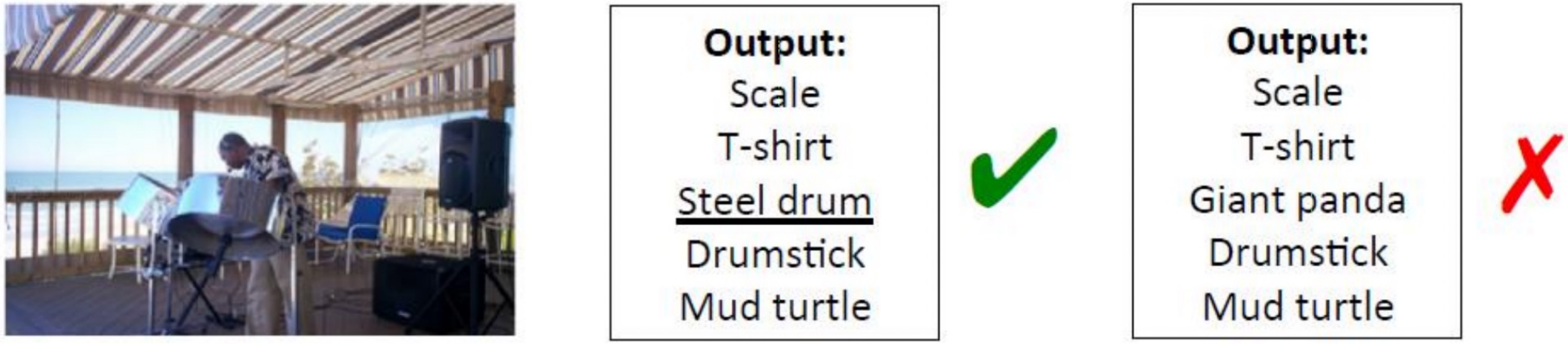
Same Metric As For The ImageNet Challenge

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

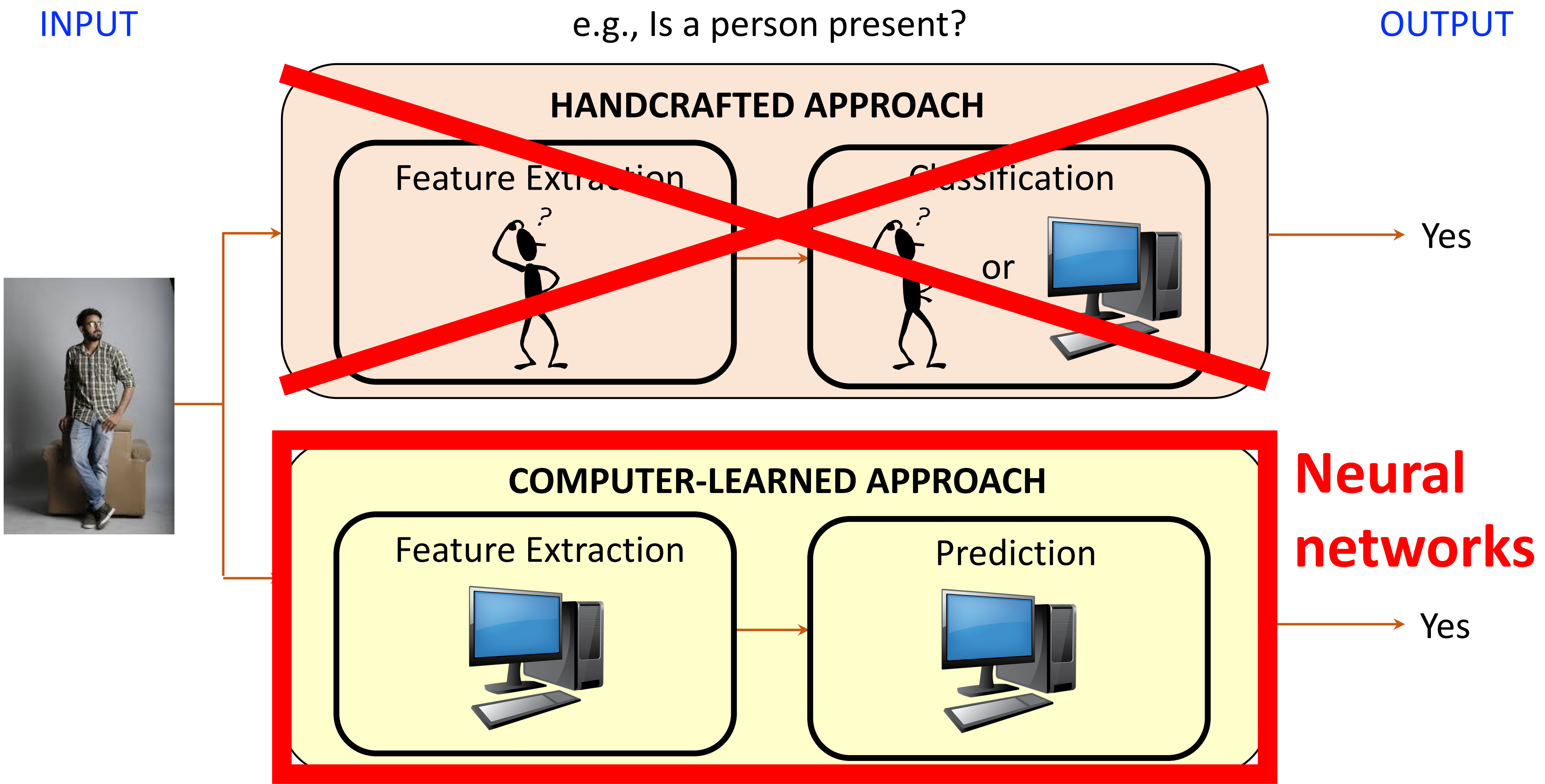


Source: https://image-net.org/static_files/files/ILSVRC2017_overview.pdf

Scene Classification: Today’s Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

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



What Neural Networks Learn

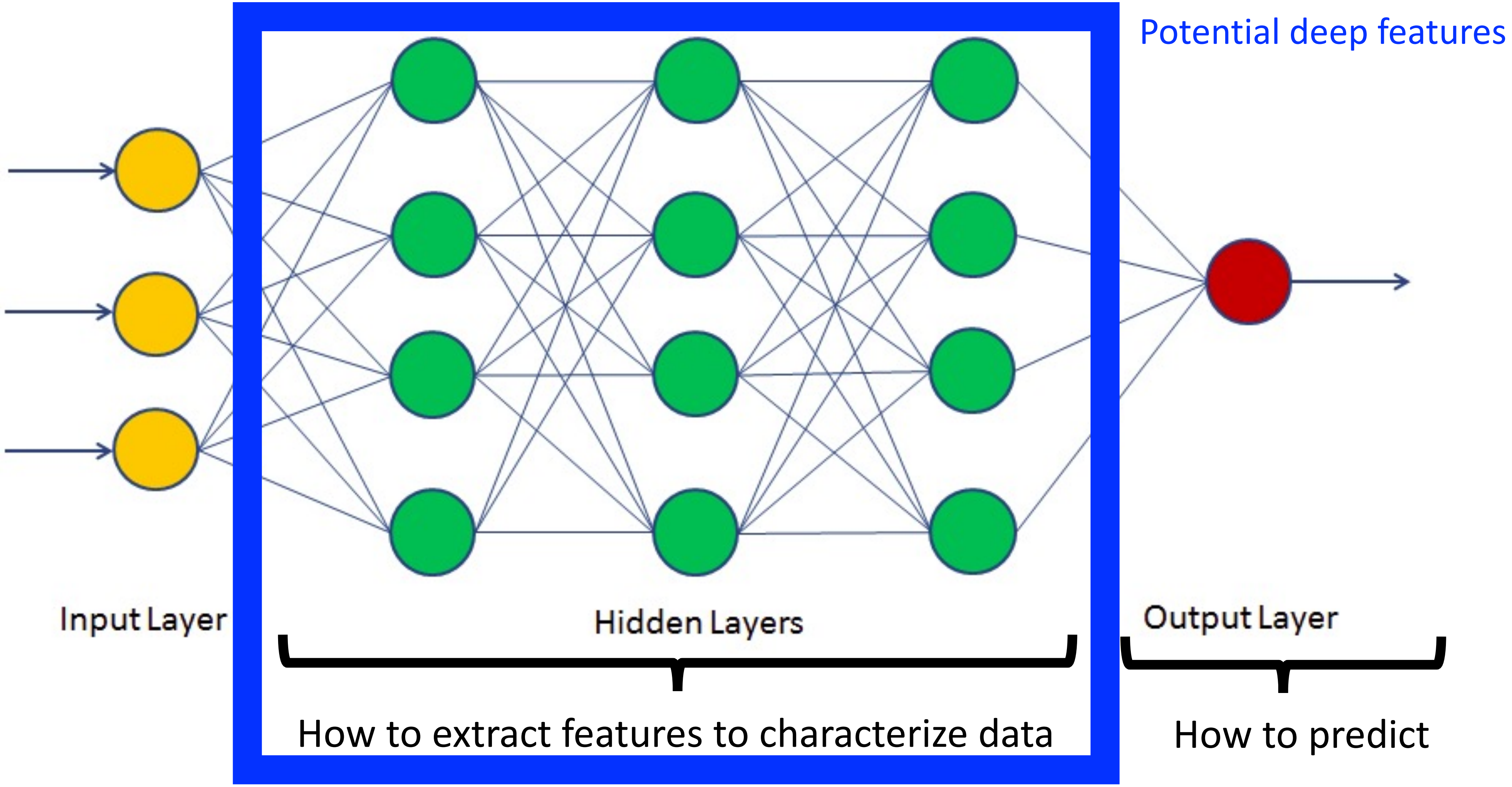


Figure Source: https://www.datacamp.com/community/tutorials/neural-network-models-r

Deep Features: AlexNet

- What is the dimensionality of the fc6 feature?
- What is the dimensionality of the fc7 feature?

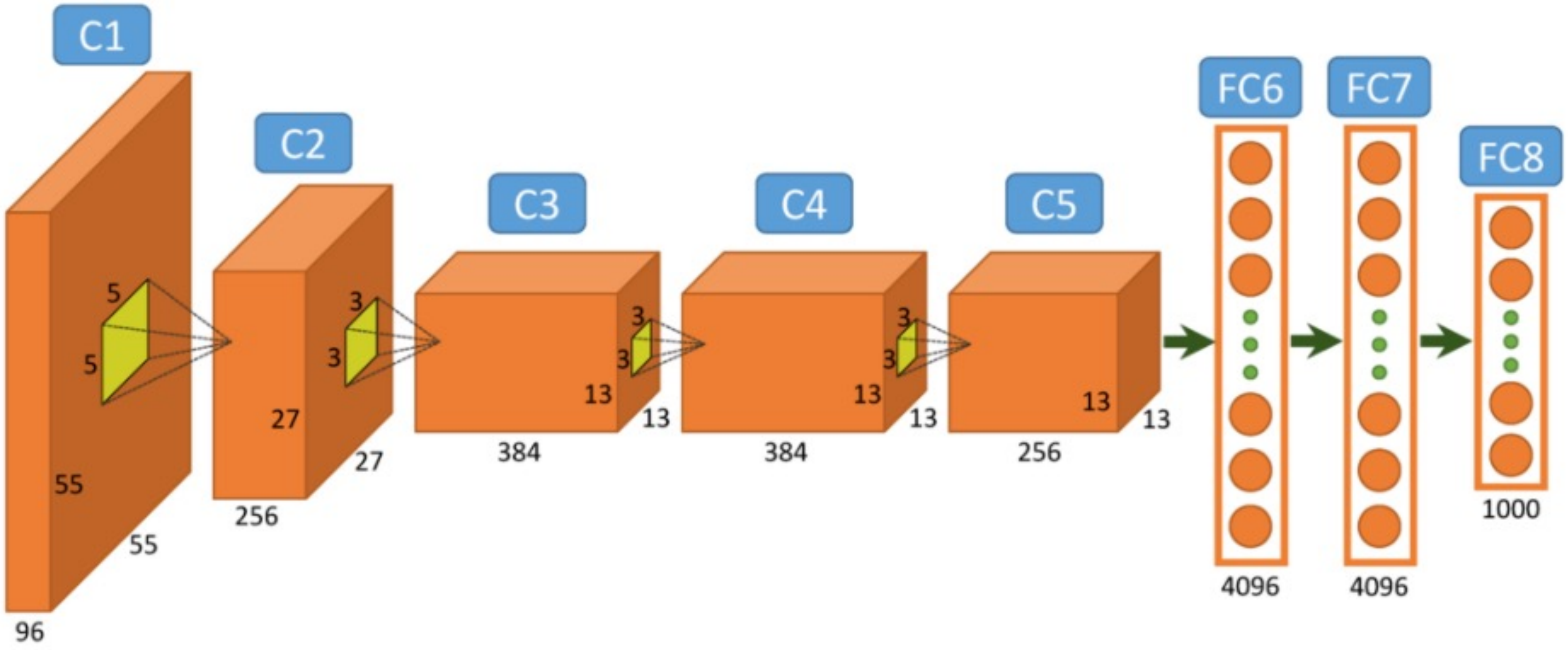


Image Source: https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers_fig2_312303454

Deep Features: e.g., To Use FC7 Layer of AlexNet

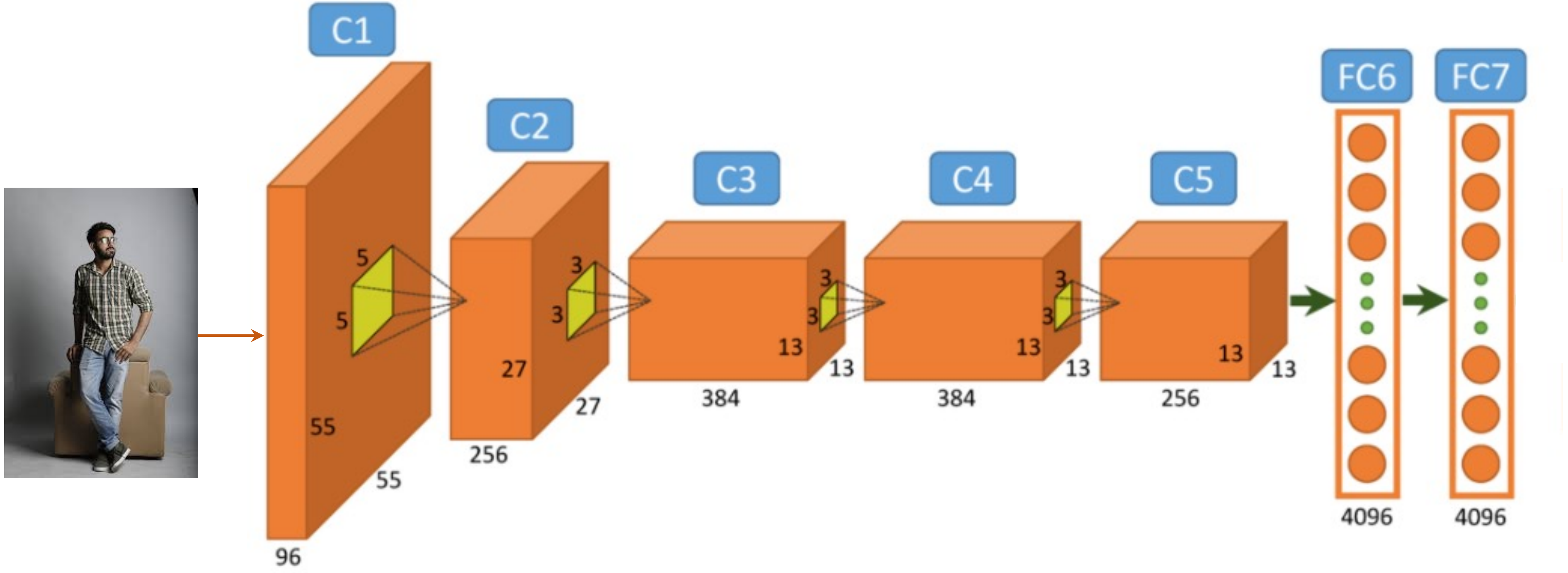


Image Source: https://www.researchgate.net/figure/Architecture-of-Alexnet-From-left-to-right-input-to-output-five-convolutional-layers_fig2_312303454

Deep Features: Which Layer to Use In a Model?

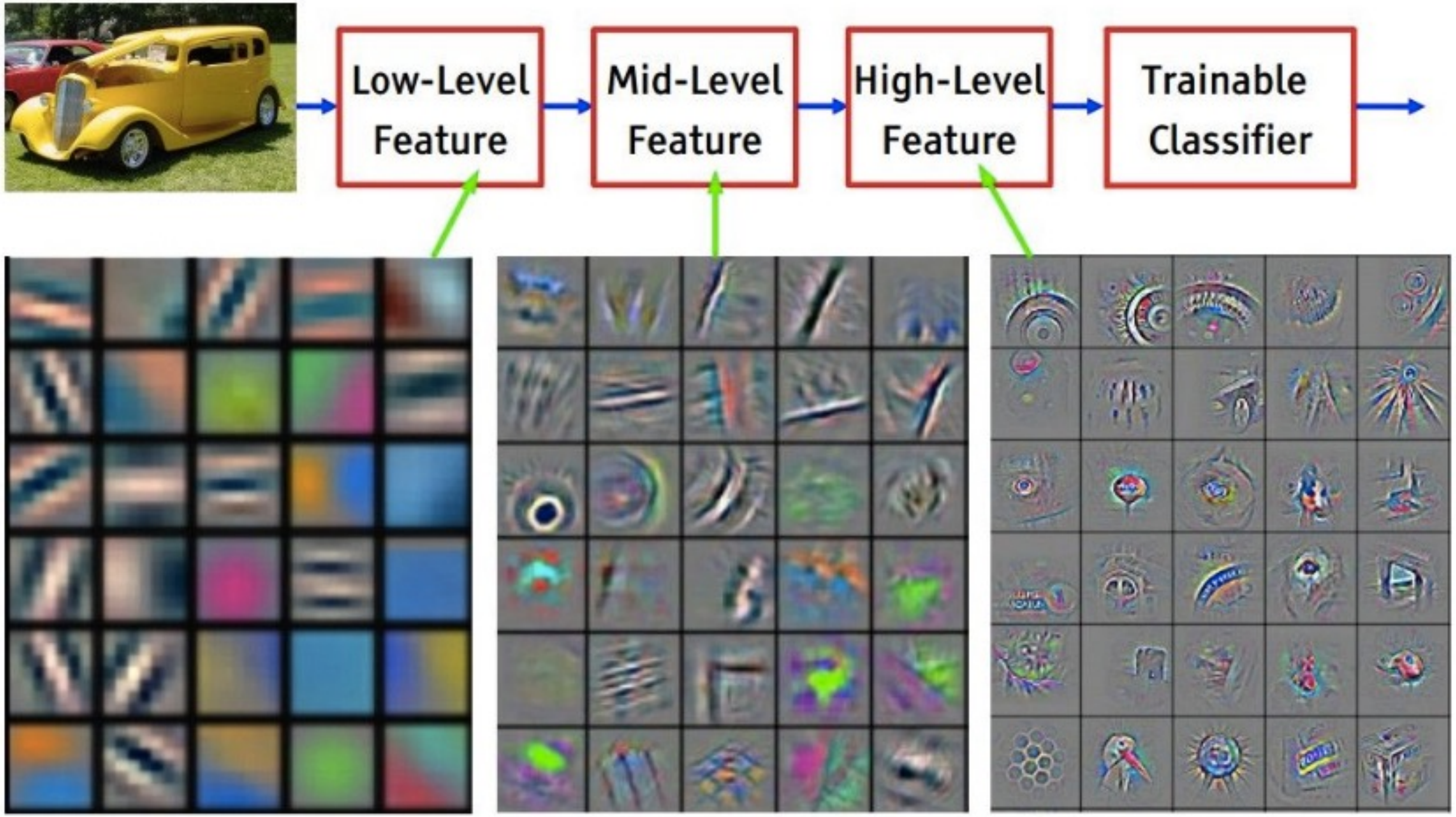


Figure Credit: Yann LeCun

What Neural Networks Learn

A pretrained network can be "fine-tuned" for a different dataset and/or task

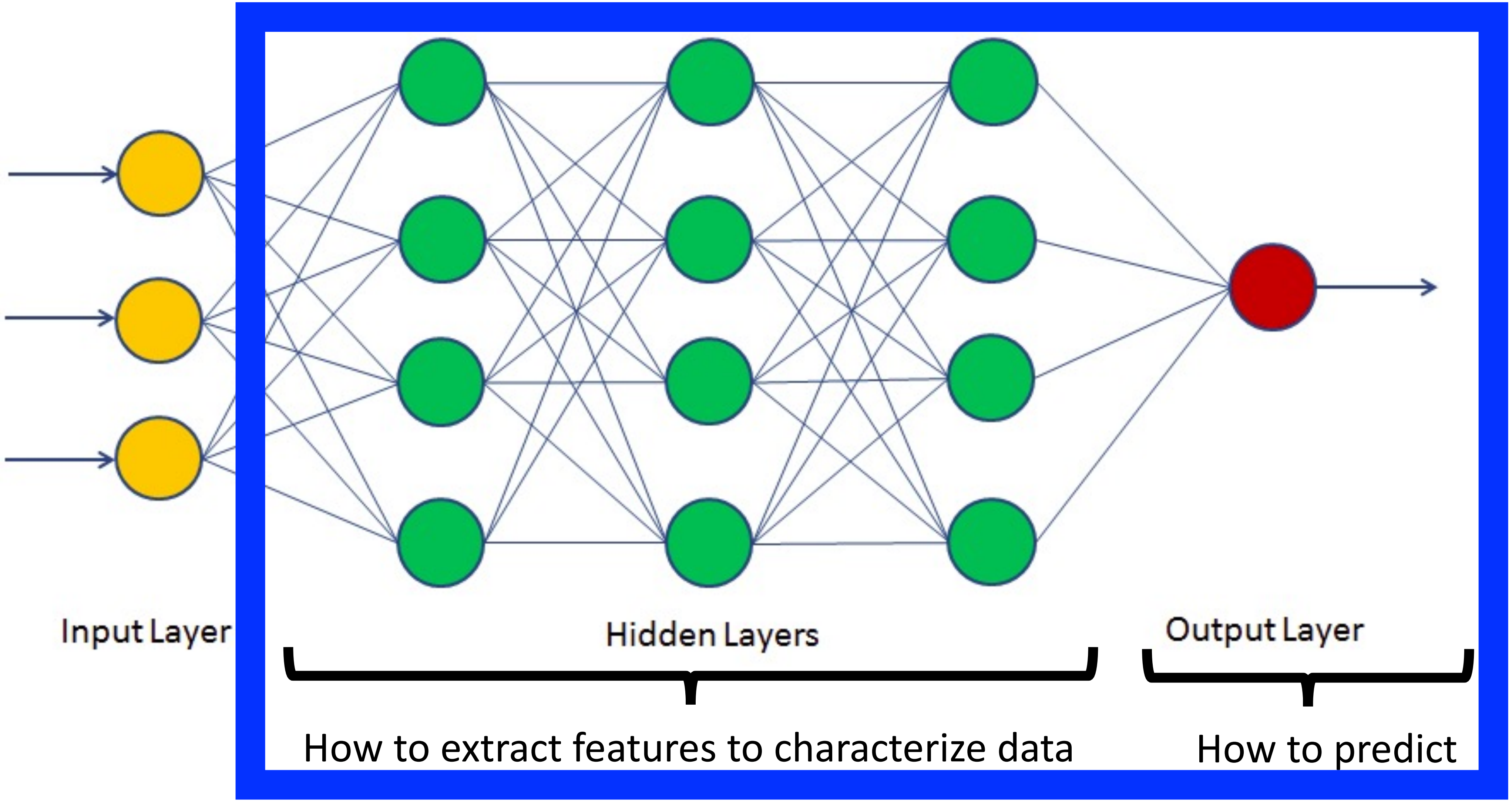


Figure Source: https://www.datacamp.com/community/tutorials/neural-network-models-r

Fine-Tuning (aka, Transfer Learning)

Use pretrained network as a starting point to train for a different dataset and/or task; e.g.,

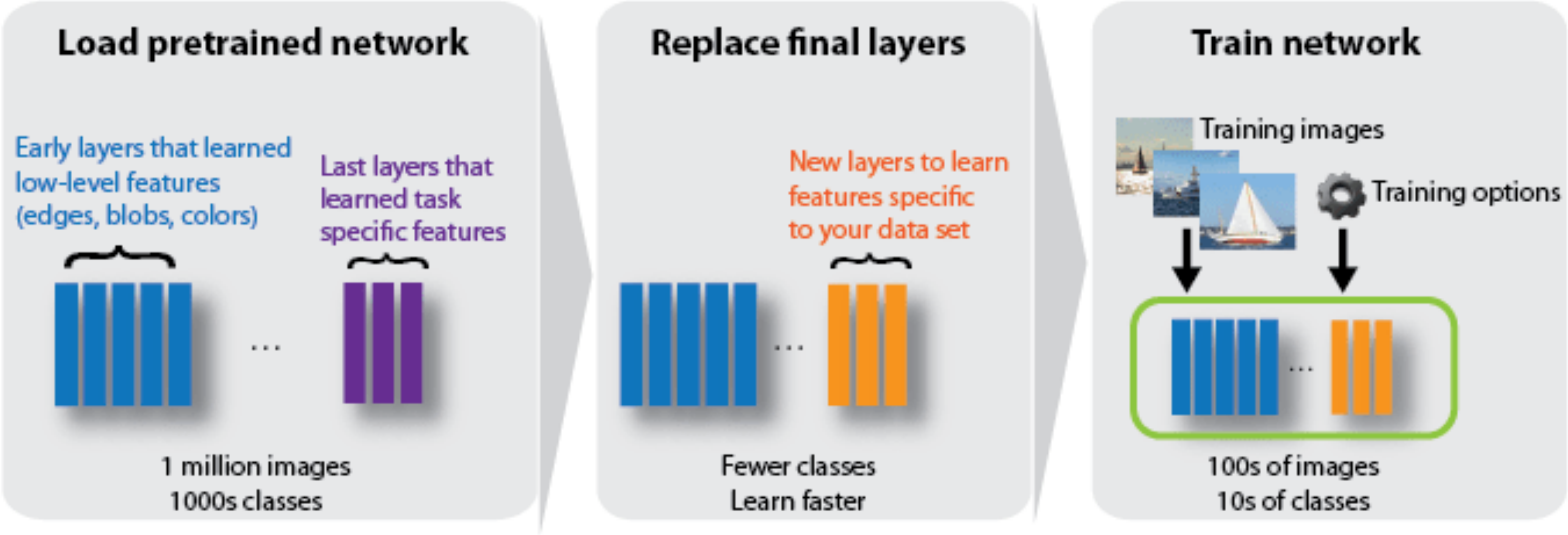


Image Source: https://www.mathworks.com/help/deeplearning/ug/transfer-learning-using-alexnet.html

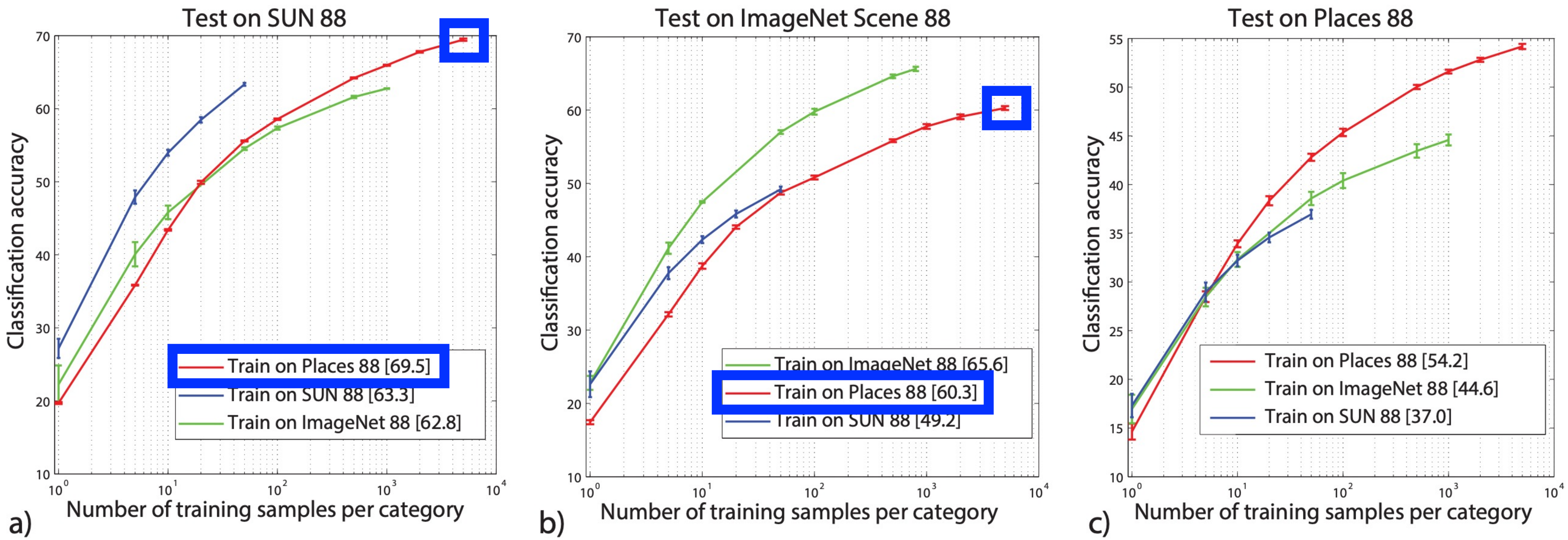
Scene Classification: Today’s Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models

Research Question: Does the Larger Datasets Lead to Better Results from Existing Models?

Research Question: Does the Larger Datasets Lead to Better Results from Existing Models?

- **Model**: AlexNet deep features followed by SVM classifier
- **Experimental Design and Results**: test on 3 different test sets

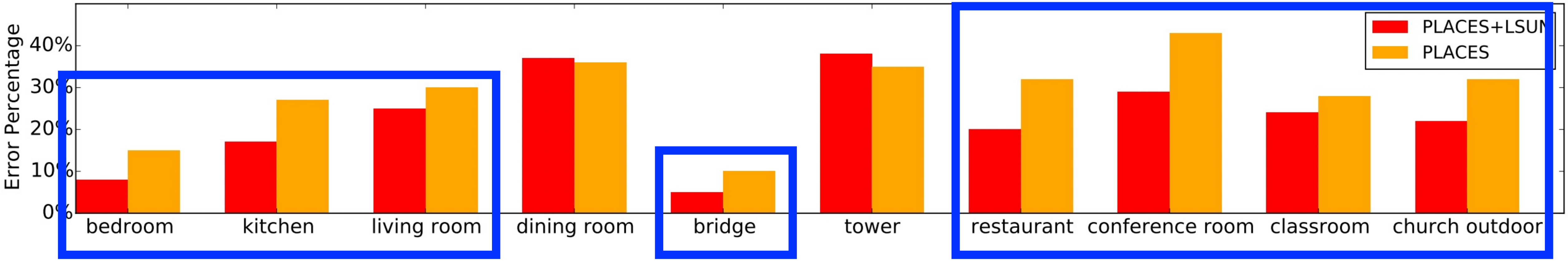


The larger Places dataset leads to better **cross-dataset** performance than existing datasets!

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

Research Question: Does the Larger Datasets Lead to Better Results from Existing Models?

- **Model**: AlexNet architecture
- **Experimental Design and Results**: test on Places dataset

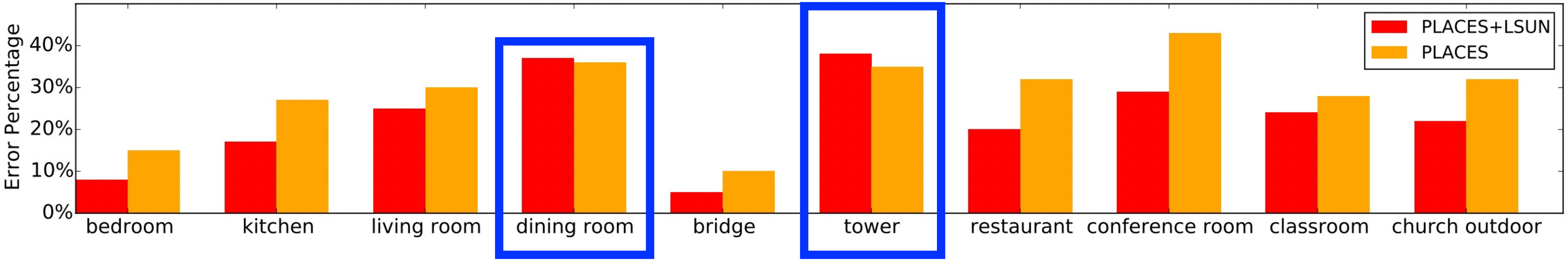


Overall, augmenting the larger LSUN dataset leads to better performance!
(i.e., 22.2% vs 28.6% error)

Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

Research Question: Does the Larger Datasets Lead to Better Results from Existing Models?

- **Model**: AlexNet architecture
- **Experimental Design and Results**: test on Places dataset



Why do you think some categories had worse results when trained with the larger LSUN dataset?

Fisher Yu et al. LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop. arXiv 2015.

Research Question: Which Dataset Leads to Better Deep Features for Image Classification Tasks?

- **Model**: AlexNet deep feature (FC7 layer) followed by SVM classifier
- **Experimental Design and Results**: test on 8 different test sets

	SUN397	MIT Indoor67	Scene15	SUN Attribute	Caltech101	Caltech256	Action40	Event8
Places-CNN feature								
ImageNet-CNN feature								

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

Research Question: Which Dataset Leads to Better Deep Features for Image Classification Tasks?

- **Model**: AlexNet deep feature (FC7 layer) followed by SVM classifier
- **Experimental Design and Results**: test on 8 different test sets

Places better for scene classification datasets!

ImageNet better for object recognition datasets!

	SUN397	MIT Indoor67	Scene15	SUN Attribute	Caltech101	Caltech256	Action40	Event8
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	**91.29**	65.18±0.88	45.59±0.31	42.86±0.25	94.12±0.99
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	89.85	**87.22±0.92**	**67.23±0.27**	**54.92±0.33**	**94.42±0.76**

State-of-art performance

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

Research Question: Which Dataset Leads to Better Deep Features for Image Classification Tasks?

- **Model**: AlexNet deep feature (FC7 layer) followed by SVM classifier
- **Experimental Design and Results**: test on 8 different test sets

	SUN397	MIT Indoor67	Scene15	SUN Attribute	Caltech101	Caltech256	Action40	Event8
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	**91.29**	65.18±0.88	45.59±0.31	42.86±0.25	94.12±0.99
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	89.85	**87.22±0.92**	**67.23±0.27**	**54.92±0.33**	**94.42±0.76**
Hybrid dataset (datasets combined to predict 1183 categories)	53.86±0.21	**70.80**	**91.59±0.48**	**91.56**	84.79±0.66	65.06±0.25	**55.28±0.64**	94.22±0.78

Combining the datasets yields an improvement for half the datasets

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

Research Question: Which Dataset Leads to Better Deep Features for Image Classification Tasks?

- **Model**: AlexNet deep feature (FC7 layer) followed by SVM classifier
- **Experimental Design and Results**: test on 8 different test sets

	SUN397	MIT Indoor67	Scene15	SUN Attribute	Caltech101	Caltech256	Action40	Event8
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	**91.29**	65.18±0.88	45.59±0.31	42.86±0.25	94.12±0.99
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	89.85	**87.22±0.92**	**67.23±0.27**	**54.92±0.33**	**94.42±0.76**
Hybrid dataset (datasets combined to predict 1183 categories)	53.86±0.21	**70.80**	**91.59±0.48**	**91.56**	84.79±0.66	65.06±0.25	**55.28±0.64**	94.22±0.78

What are limitations of what we can learn from these experiments about which deep features to use when?

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

Scene Classification: Today's Topics

- Problem
- Applications
- Evolution of Datasets
- Evaluation Metrics
- Background: Deep Features and Fine-Tuning
- Computer Vision Models


The End