Lecture 1: Introduction

Course staff

Andrew Owens

Instructor

Xixi Hu

GSI

Name	Office hours time
Andrew Owens	Monday 4:45 - 5:30pm
Xixi Hu	Thursday 4:00 - 4:45pm

Today

- Unsupervised visual learning
- Class logistics
- Discussion

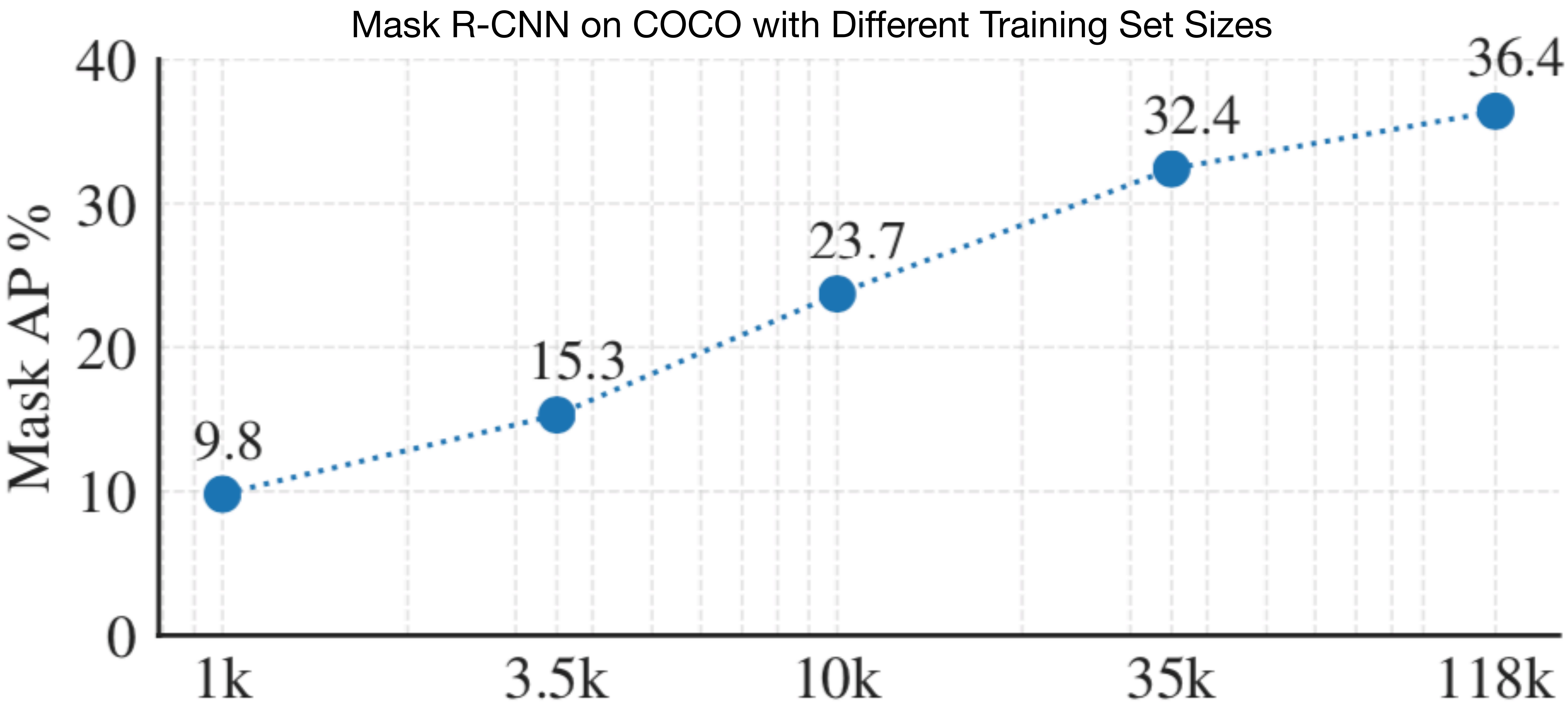
Supervised computer vision

Object recognition [Russakovsky et al., “ImageNet”, 2015]

Object segmentation [Gupta et al., “LVIS”, 2019]

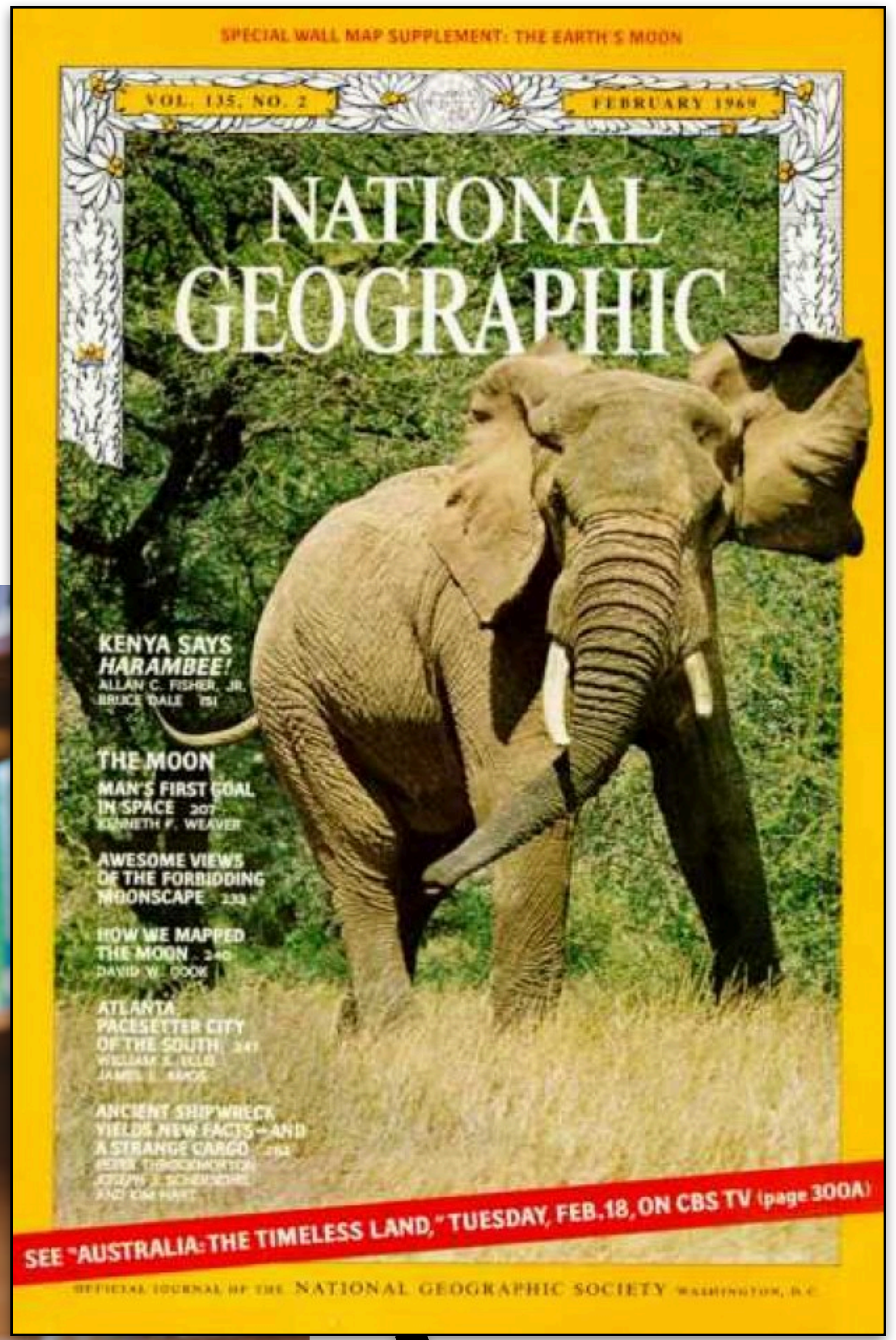
Today's computer vision methods mostly learn about the world using **supervision provided by humans**.

We still need *lots* of labeled examples



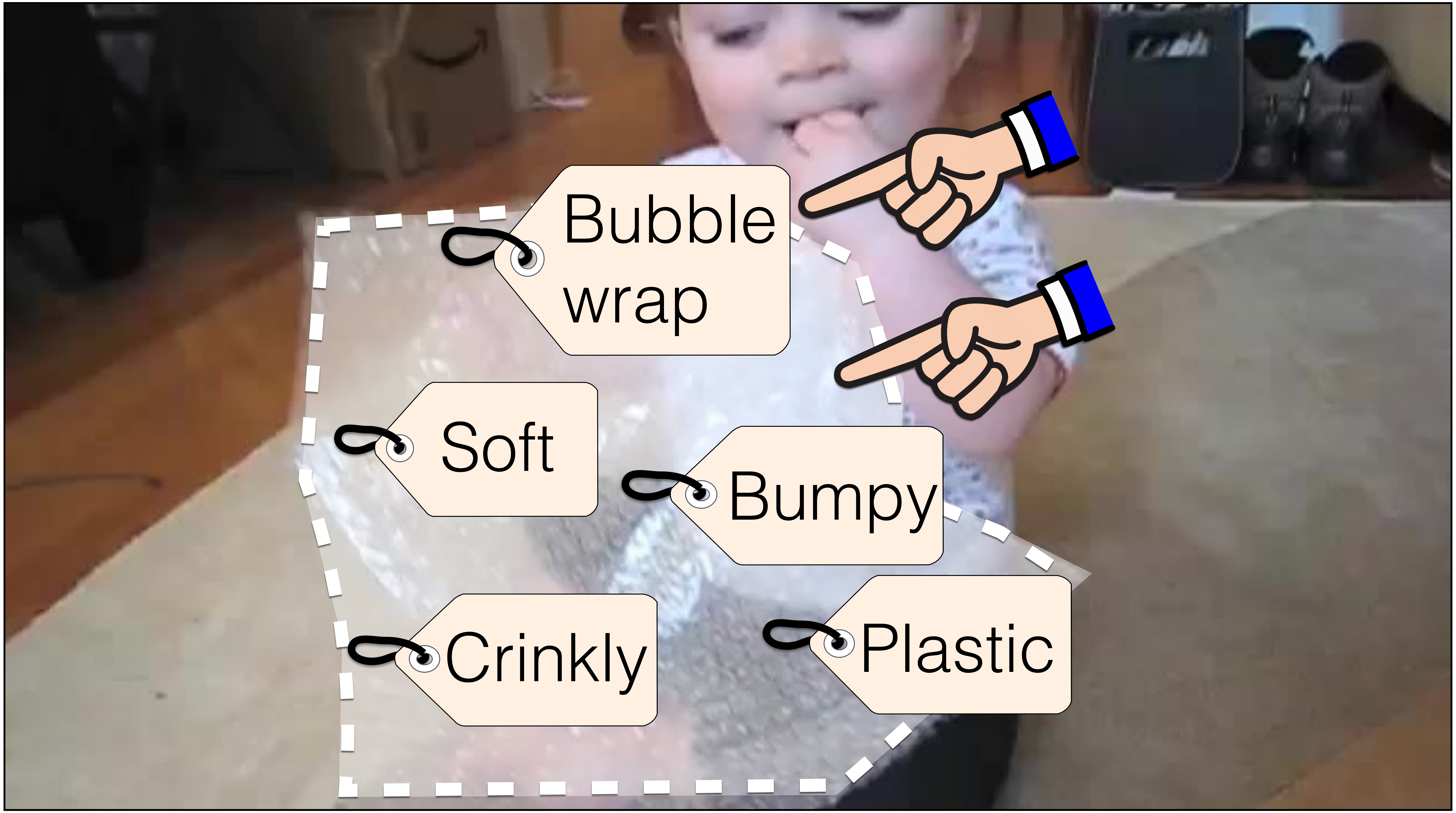
Object detection accuracy vs. dataset size

Image source: R. Girshick



We want methods that can generalize with *as little intervention* from humans as possible.

Supervised learning

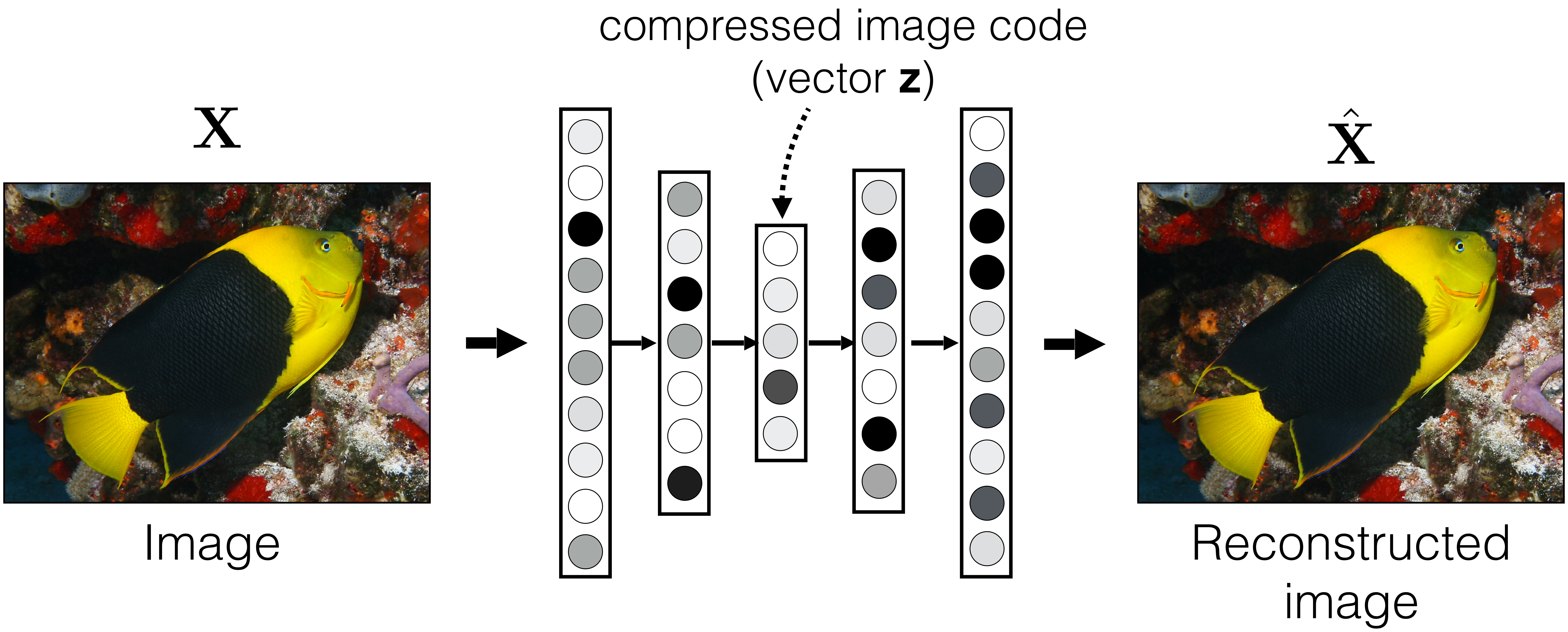


Supervised learning

Unsupervised learning

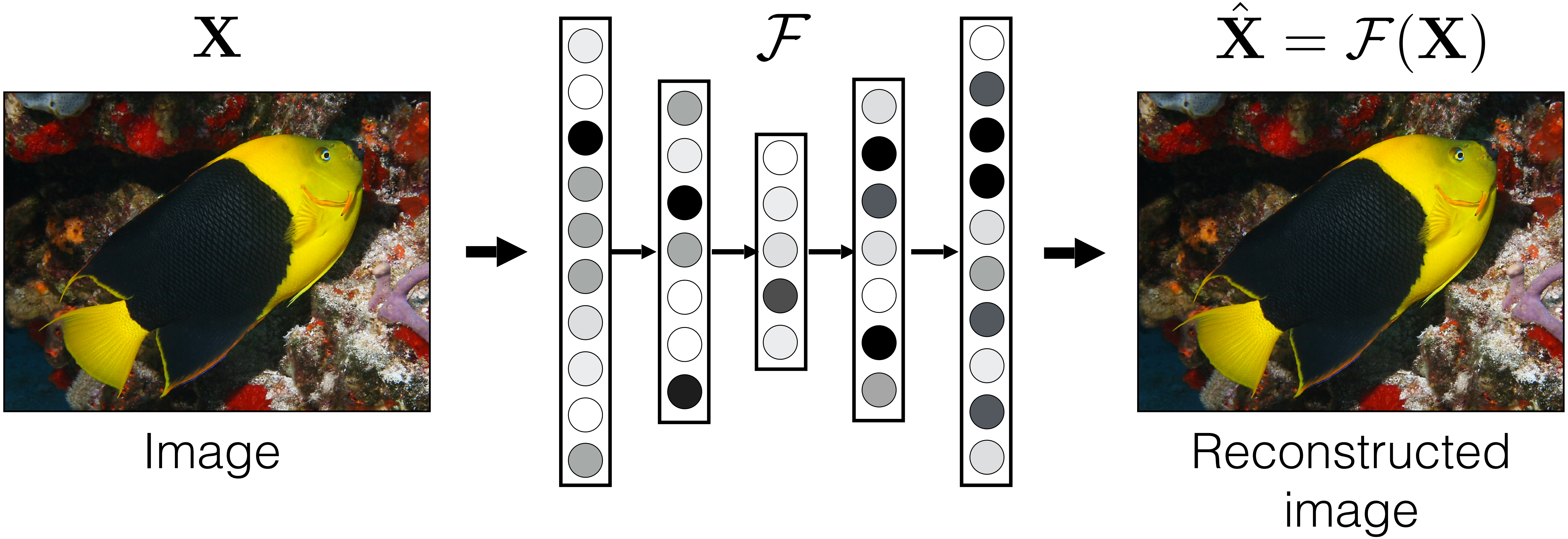
a.k.a. self-supervised learning

Example: autoencoder



[e.g., Hinton & Salakhutdinov, Science 2006]

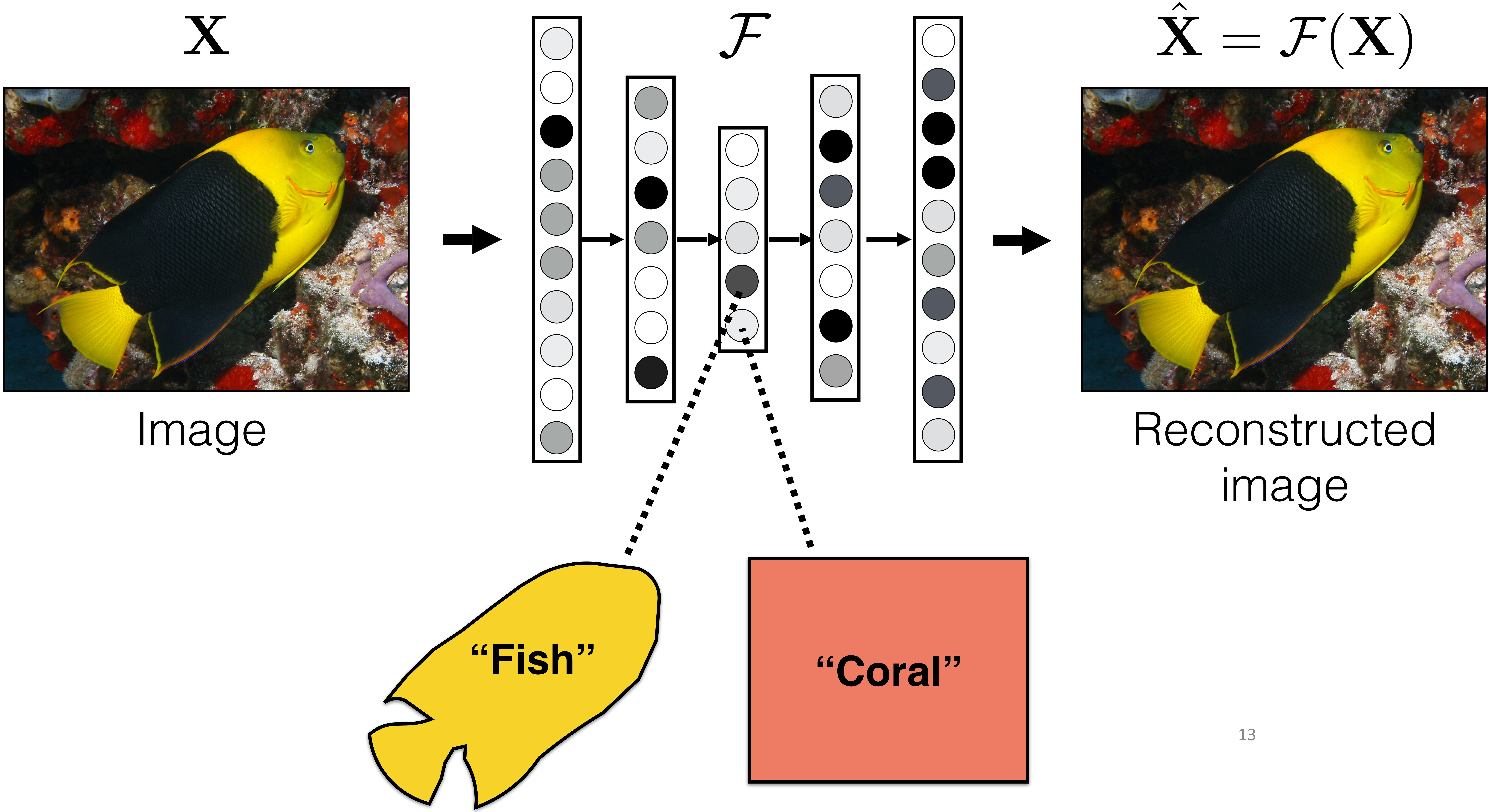
Source: Isola, Freeman, Torralba



Through this training process, our model learns to code for objects and other “useful” image structures.

This method is **unsupervised**. It requires no labeled data.

Source: Isola, Freeman, Torralba



Source: Isola, Freeman, Torralba

Unsupervised learning

a.k.a. self-supervised learning

Unsupervised learning

a.k.a. self-supervised learning

+ Less work and (potentially) money for labeling
+ Avoid the need for human judgement calls
+ No need for a human “in the loop” to update training

- Learning methods are still being developed. In some cases there are good methods, but not “one size fits all”
- Requires lots of *unlabeled* data and compute (potentially)

Questions?

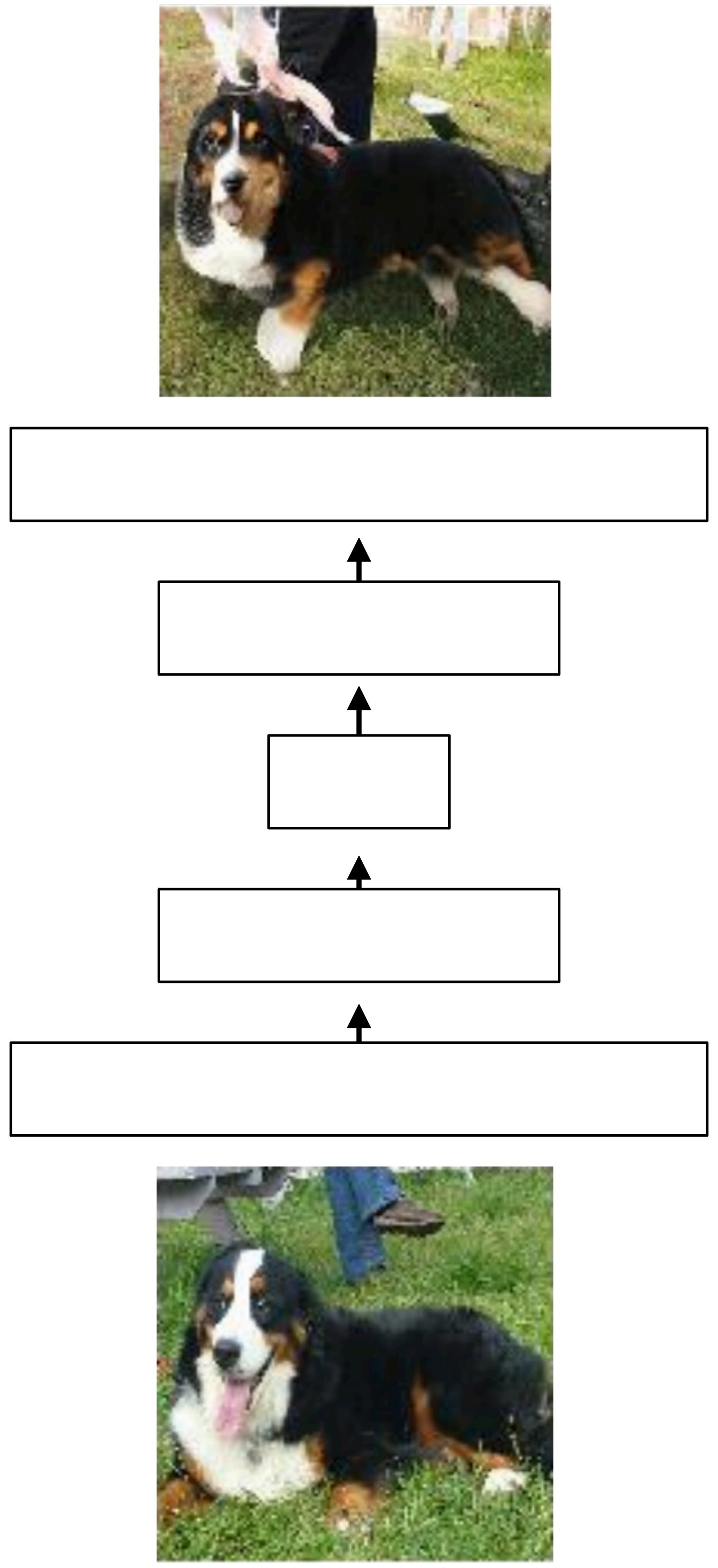
Class topics

1. Overview of recent unsupervised learning methods
2. Applications of unsupervised learning in vision
3. New directions in unsupervised learning

Part #1: deep generative models

Learn about the world by learning to reconstruct it

			Generative models
Lec. 2	Mon, Jan 25	**Energy-based models**	• Song & Kingma: How to Train Your Energy-Based Models • Grathwohl et al: Your classifier is secretly an energy based model and you should treat it like one
Lec. 3	Wed, Jan 27	**Variational autoencoders**	• Razavi, van den Oord, Vinyals: Generating Diverse High-Fidelity Images with VQ-VAE-2 • Vahdat & Kautz, NVAE: A Deep Hierarchical Variational Autoencoder • Doersch, Tutorial on Variational Autoencoders (optional)
Lec. 4	Mon, Feb 1	**Normalizing flows**	• Dinh et al.: Density Estimation using Real NVP • Kingma & Dhariwal, Glow: Generative flow with invertible 1x1 convolutions
Lec. 5	Wed, Feb 3	**GANs**	• Donahue & Simonyan: Large Scale Adversarial Representation Learning • Bau et al:, Rewriting a deep generative model
Lec. 6	Mon, Feb 8	**Autoregressive models**	• van den Oord et al: Pixel-CNN • Chen et al.: Generative pretraining from pixels

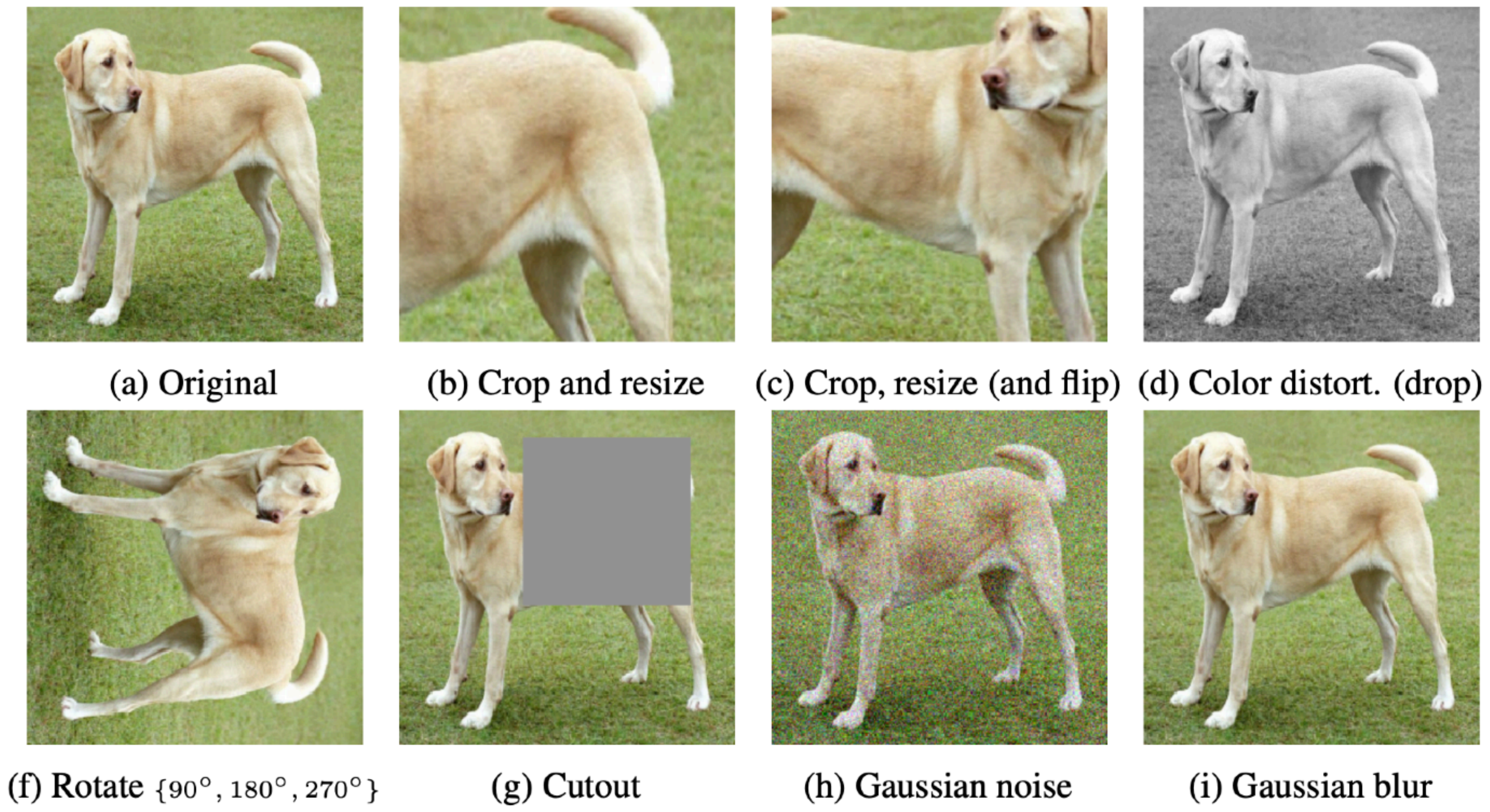


[Donahue & Simonyan, "BigBiGAN", 2019]

Part #2: discriminative unsupervised learning

Learn a representation by solving specially designed discriminative learning tasks.

Discriminative methods for unsupervised learning			
Lec. 7	Wed, Feb 10	**Contrastive learning**	• He et al.: Momentum Contrast for Unsupervised Visual Representation Learning • Grill et al, Bootstrap your own latent: A new approach to self-supervised Learning • Chen et al: A Simple Framework for Contrastive Learning of Visual Representations • van den Oord et al: Representation Learning with Contrastive Predictive Coding (op • Tian et al: Contrastive Multiview Coding (optional)
Lec. 8	Mon, Feb 15	**Pretext tasks**	• Piergiovanni et al: Evolving Losses for Unsupervised Video Representation Learnin • Are Labels Necessary for Neural Architecture Search?

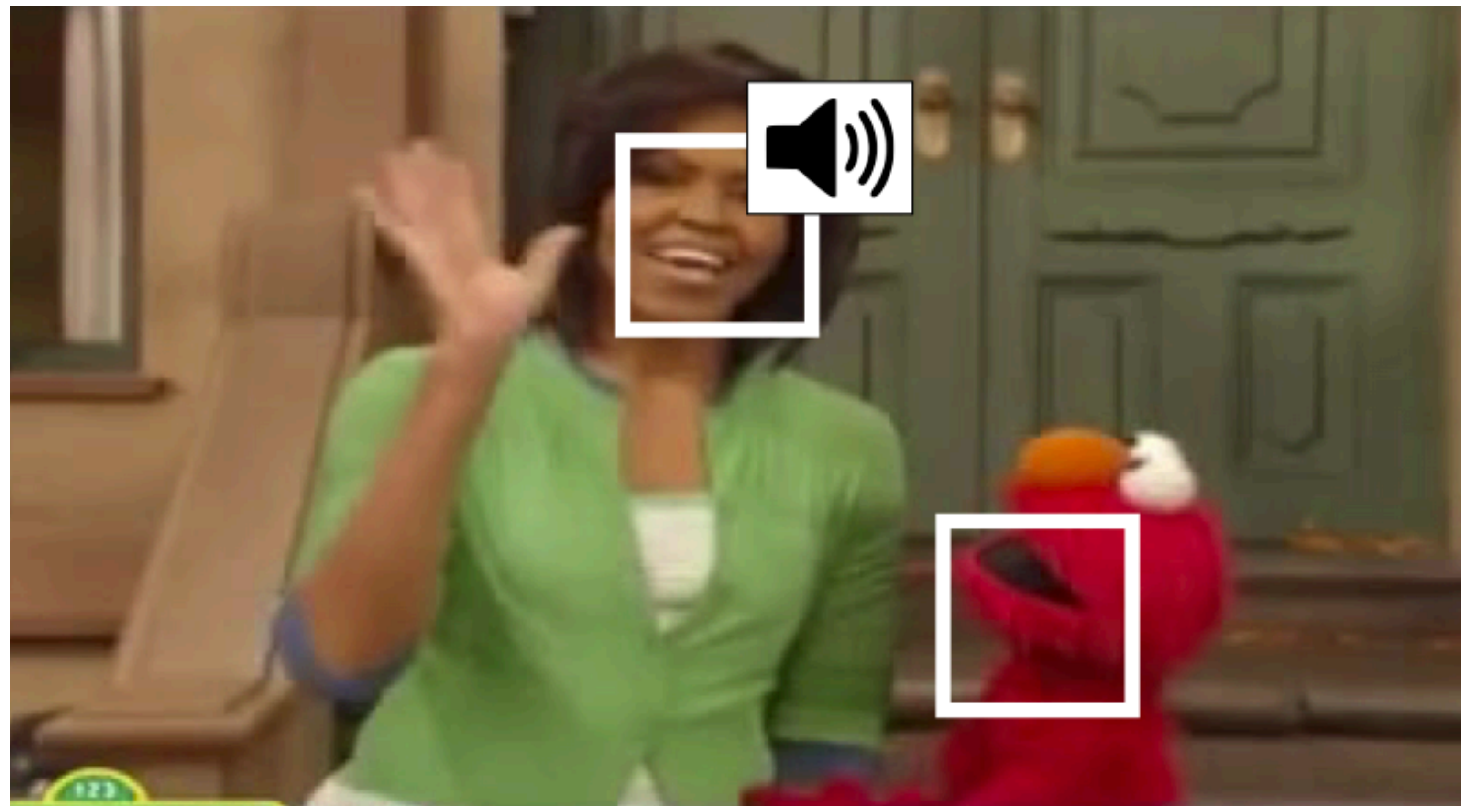


[Chen et al., “SimCLR”, 2019]

Part #3: learning from other modalities

Learn things from non-visual modalities that would be hard to learn from vision alone.

Learning from non-visual signals			
Lec. 9	Wed, Feb 17	**Language & vision**	• Desai & Johnson: VirTex: Learning Visual Representations from Textual Annotations • Radford et al: Learning Transferable Visual Models From Natural Language Supervision
Lec. 10	Mon, Feb 22	**Sound & vision**	• Afouras et al.: Self-Supervised Learning Of Audio-Visual Objects From Video • Asano, Patrick et al.: Labelling unlabelled videos from scratch with multi-modal self-supervision

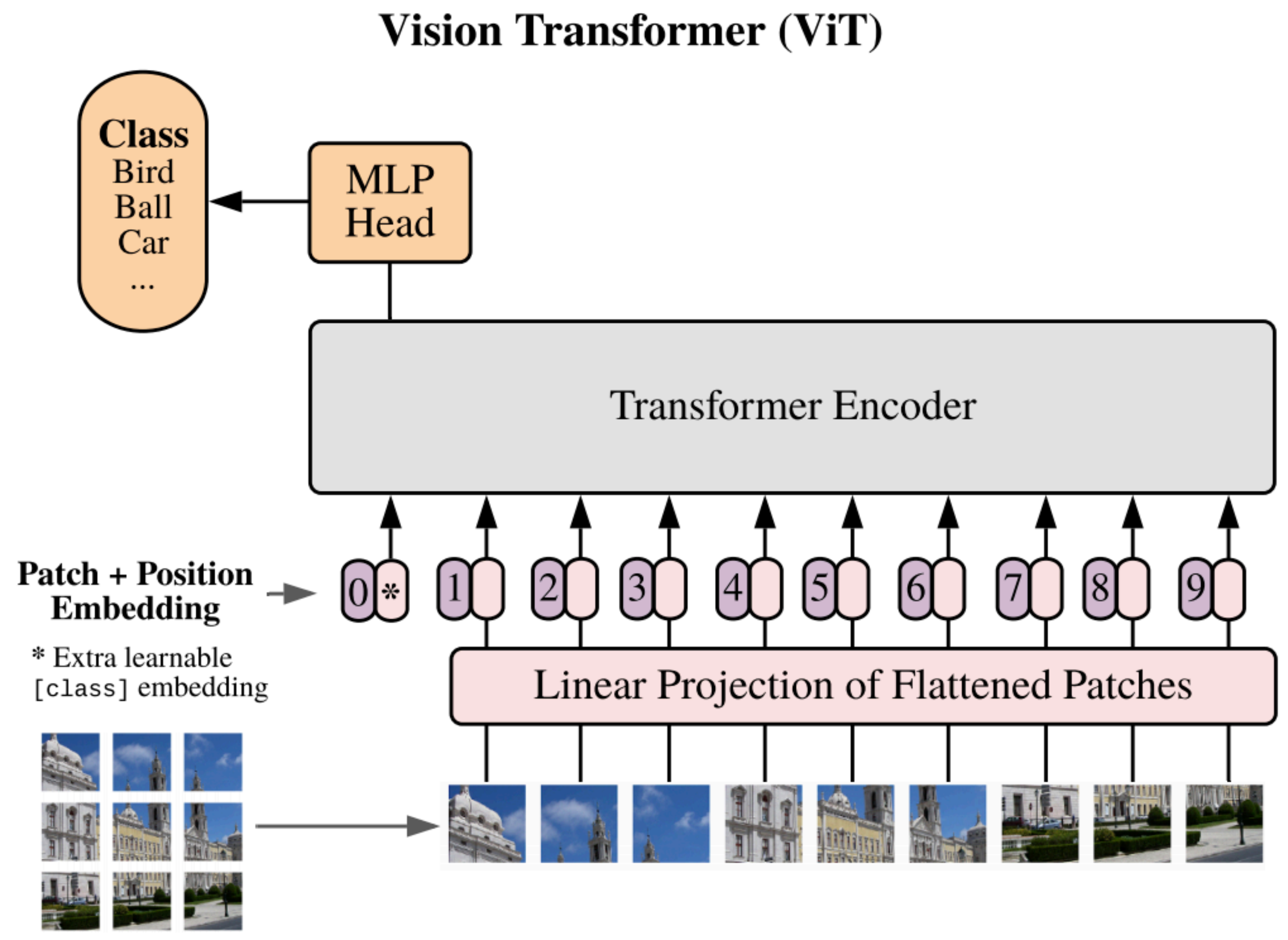


[Afouras et al., 2020]

Part #4: Advances in deep learning

Advances in architectures and optimization that are (or hopefully will) driving progress in unsupervised learning.

Advances in deep learning			
Lec. 14	Wed, Mar 10	**Attention**	• Dosovitskiy et al: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale • Carion et al: End-to-End Object Detection with Transformers
Lec. 15	Mon, Mar 15	**Optimization**	• Ba et al.: Distributed Second-Order Optimization Using Kronecker-Factored Approximations

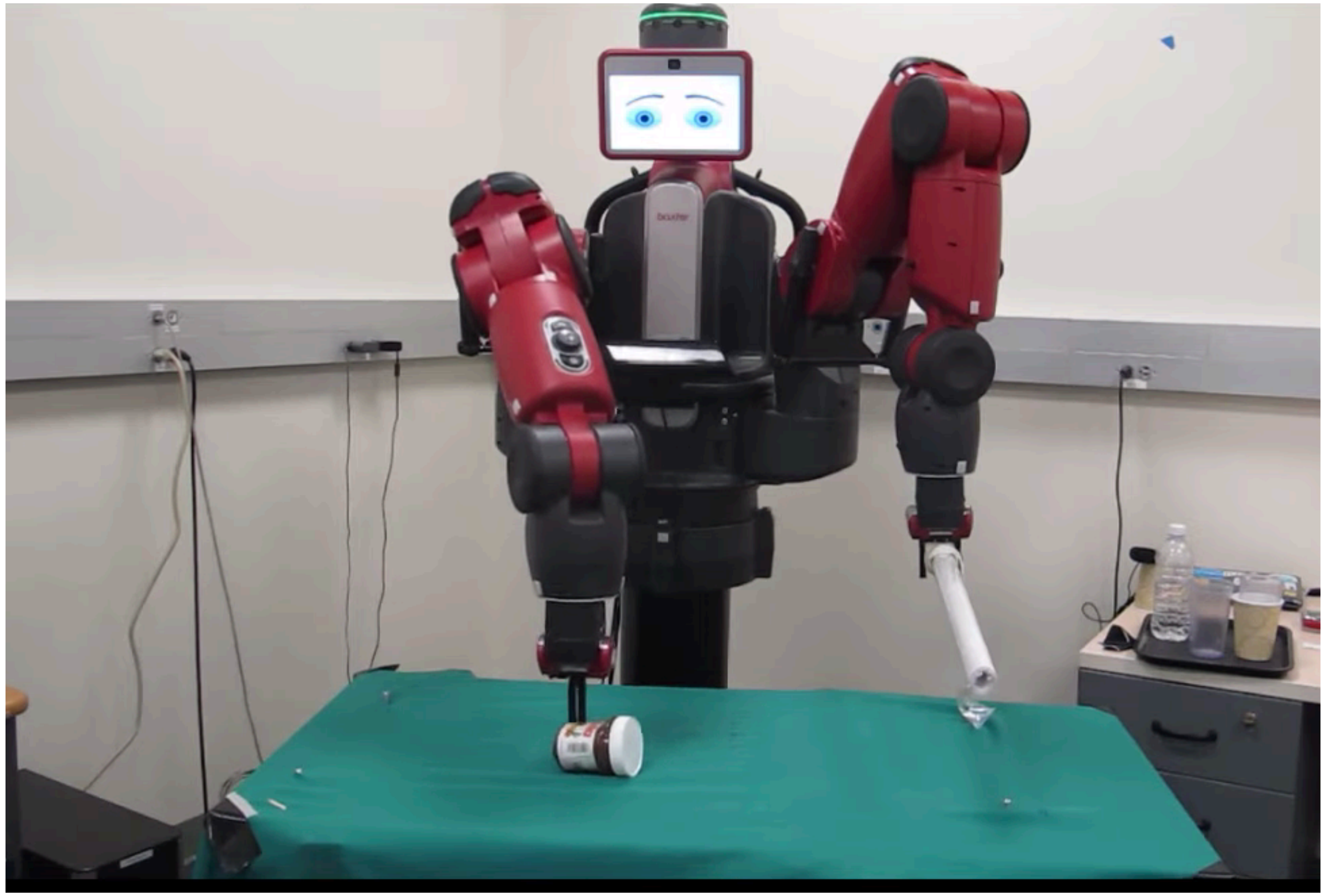


Part #5: Learning with less supervision

Making a little bit of data go a long way.

		Learning with less supervision	
Lec. 16	Mon, Mar 22	**Gradient-based meta-learning**	• Finn et al.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks • Wu, Ren et al.: Understanding Short-Horizon Bias In Stochastic Meta-Optimization • Maclaurin et al: Gradient-based hyperparameter optimization through reversible learning (optional) • Liu et al, DARTS: Differentiable Architecture Search (optional)
Lec. 17	Wed, Mar 24	**More meta-learning**	• Lorraine et al: Optimizing Millions of Hyperparameters by Implicit Differentiation • Nichol et al.: On First-Order Meta-Learning Algorithms (optional)
Lec. 18	Mon, Mar 29	**Semi- and weakly-supervised learning**	• Pham et al: Meta Pseudo Labels • Chen et al: Big Self-Supervised Models are Strong Semi-Supervised Learners • Xie et al: Self-training with Noisy Student improves ImageNet classification (optional)

Part #6: Embodied learning



Images from [Agrawal, Nair et al. 2016] and [Pathak et al, 2017]

What you'll do in this class

	Grade
• Class presentation	25%
• Participation & reviews	10%
• Final project	45%
• One problem set	20%

What this class *isn't*

- It's not an *introduction* to vision or unsupervised learning.
- We're assuming some basic knowledge of these topics (e.g. EECS 442/504)
- The focus is on recent areas of research.
- It's a self-directed class: reading, projects, and presentations in place of homework

Lectures

- This is a discussion-based class, so participation is important
- We’ll take attendance, starting next week
- If you miss more than 2 lectures without an excuse, it will hurt your participation grade
- To supplement lectures, we also have a Piazza forum

Presentations

- This is a seminar-style class, based on **student presentations** of very recent papers.
- These start on week #3. I'll teach until then!
- You'll be part of a 3-person group.
- I encourage you to come to my office hours to discuss the papers and prepare!

What’s in a presentation?

Suggested structure for most classes:

1. Background	24 mins
2. Paper presentation #1	24 mins
3. Paper presentation #2	24 mins
4. Discussion	8 mins

Paper presentations

- Explain the paper to the class
- A *critical* presentation of the paper.
    - Unlike the “test of time” papers you mostly read in intro classes, these will be very new papers.
    - Often they’ll have serious limitations and flaws — perhaps not even acknowledged by the authors!

Paper presentations

- A good presentation should answer:
    1. What problem did the researchers address?
    2. How did they solve it?
    3. How does the idea relate to prior work?
    4. What key experiments did they use to test their ideas?
    5. How successful were they at achieving their goals?

Paper presentations

- No need to present *every* experimental result.
- Ablation experiments are important. What *really* makes the method work?
- Comparisons to state-of-the-art are important, too. How far are they from the best-performing method?
  - Often, well-chosen qualitative results are the best way to convey this.
  - Sometimes the reason they are better is unrelated to the contribution itself, e.g. using a modern architecture.
- Make sure to briefly explain metrics.
- Try to convey the stakes in the experiment (e.g. why was this important?)

Background

- This is meant to be a mini-lecture (or a “tutorial”).
- Bridge the gap between “intro to vision” and “very recent paper”
- Should cover the basics, and important prior work.
- E.g. a generative adversarial networks (GAN) class should have a background section that briefly explains GANs, and gives a quick overview of the previous state-of-the-art.
- More open-ended than the paper presentations.
    - I’m happy to help with this if you drop by my office hours!

Discussion

- This is an optional interactive component of the class (~8 mins)
- Give an opportunity to touch on the larger issues at stake in the papers
- Suggested format: give a prompt, put people into breakout rooms.
- Example prompts for future classes:
  - *Are there things we can only learn about the world by interacting with it?*
  - *Should we really consider natural language to be “unsupervised”, as the authors claim? What are the key differences?*

Groups

- We'll have two sign-up periods
  - One for the first few weeks, another after the drop deadline for the remainder of the course.
- You'll rank the topics, and we'll assign one to you.
- If you have a group of three in mind, that's great. You can sign up jointly.
- You can decide how to split the papers/background up amongst yourselves.
- Your grade: the average score of the group as a whole, and your own presentation.
- Grading is rounded to: $\checkmark^{+}$, $\checkmark$, $\checkmark^{-}$, **0**
  - $\checkmark^{+}$ is only for exceptional work

Paper reviews

- You will write **one paper review per week**
- These should be similar in content to a class presentation
- Aim for half a page
- Submitted to Gradescope *before* lecture, i.e. at 3pm the Monday or Wednesday when the paper is to be presented
- We don't accept late submissions.
- Grading is rounded to: ✓+, ✓, ✓−, **0**
  - Again, ✓+ is only for exceptional work
- We drop the 2 lowest scores

Homework

- This class is mostly about reading and critiquing research.
- However, we'll have a **single** problem set.
- It will cover a few of the "core" ideas in representation learning, e.g. generative models, multi-modal learning.
- Due approximately 2/3 of the way through the semester.
- Uses PyTorch and Google Colab.
- Should be completed individually, though you are allowed to discuss the work with other students

Project

- Can be original work, or reimplementation of existing papers.
- Solo or in groups of up to 4 people (ask permission for more).
- Can be based on your own research (but only if everyone in the group is a collaborator on that research)
- Deliverables:

1. Project proposal (due in early March)
2. Short presentation at the end of class
3. Writeup (due at end of finals period)

Today’s discussion

- Suppose that acquiring labels was *much* cheaper than it is now. Would we still need unsupervised learning? If so, when?

Today’s discussion

- What questions do *you* have about unsupervised learning that you’d like to get answers to this semester?

- Why are *you* interested unsupervised learning?

Next week: generative models

I can chat until 5pm.