Western Science

Artificial Intelligence II

Part 2: Lecture 8

Yalda Mohsenzadeh

Representation Learning

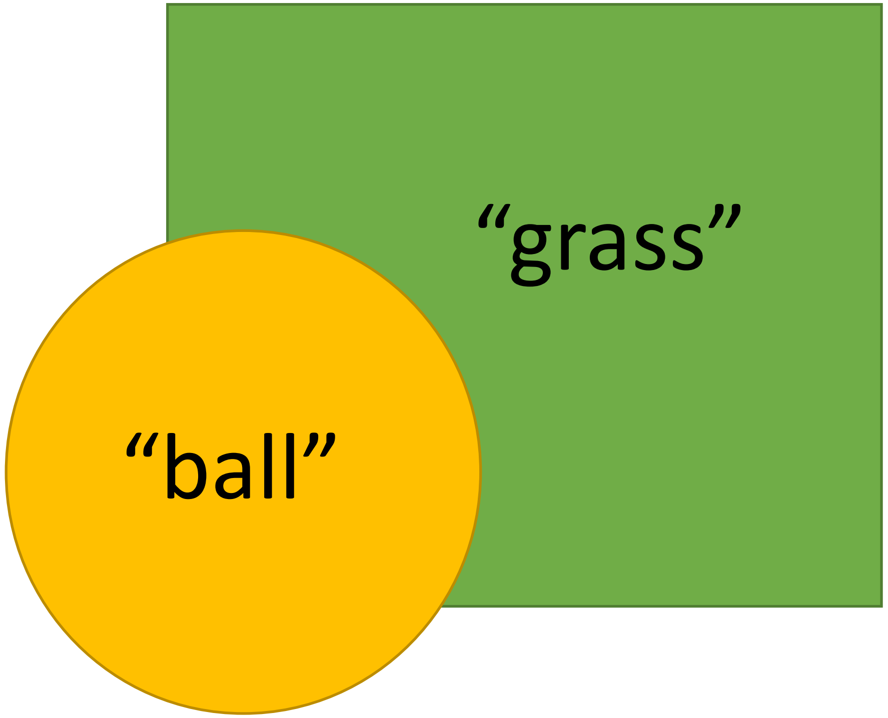
- Representations in the brain
- What is learned by a deep net?
- Transfer learning and fine-tuning
- Unsupervised and self-supervised learning
- Anomaly detection

"I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have "327"? No. I have sky, house, and trees."

— Max Wertheimer, 1923

Representation Learning

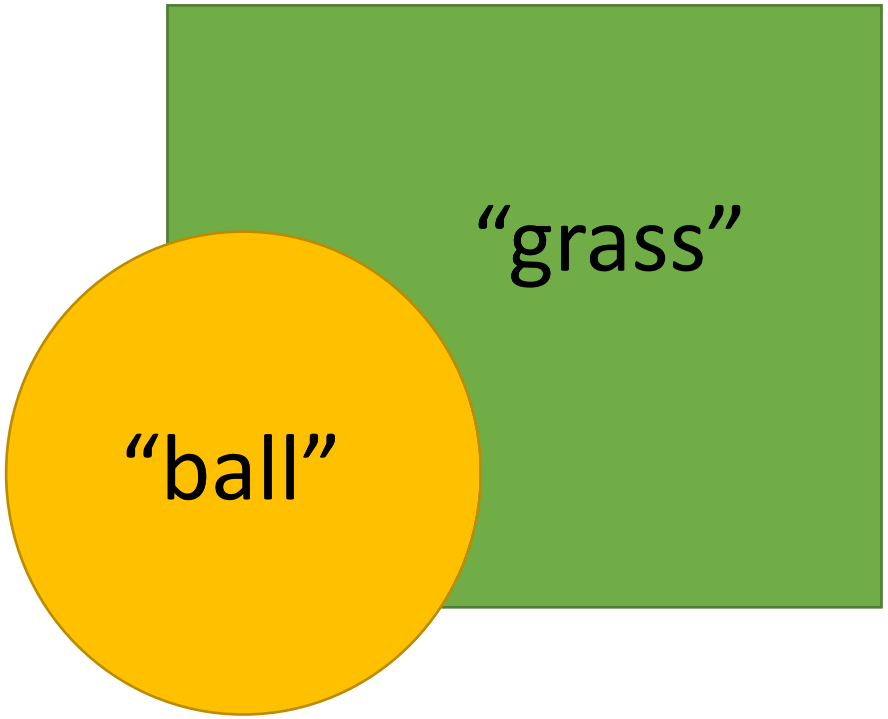
Image



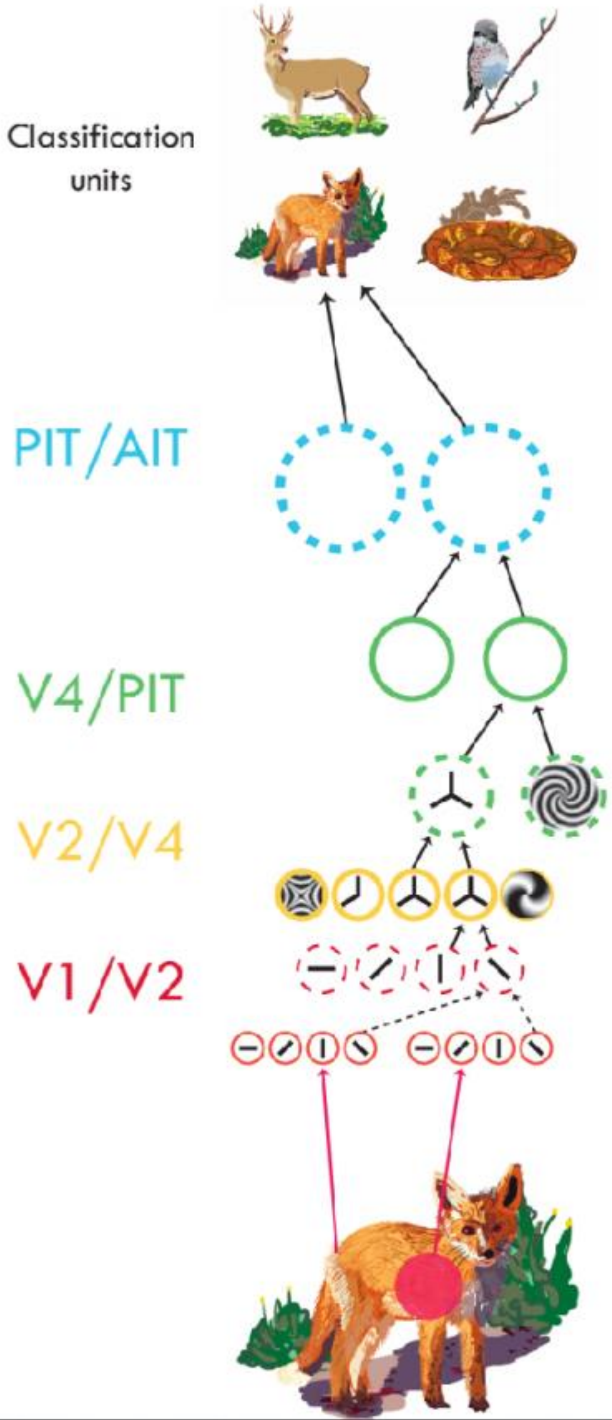
Compact Mental Representation

Representation

- Good representations are:
  - Compact (minimal)
  - Explanatory (sufficient)
  - Disentangled (independent factors)
  - Interpretable
  - Make subsequent problem solving easy

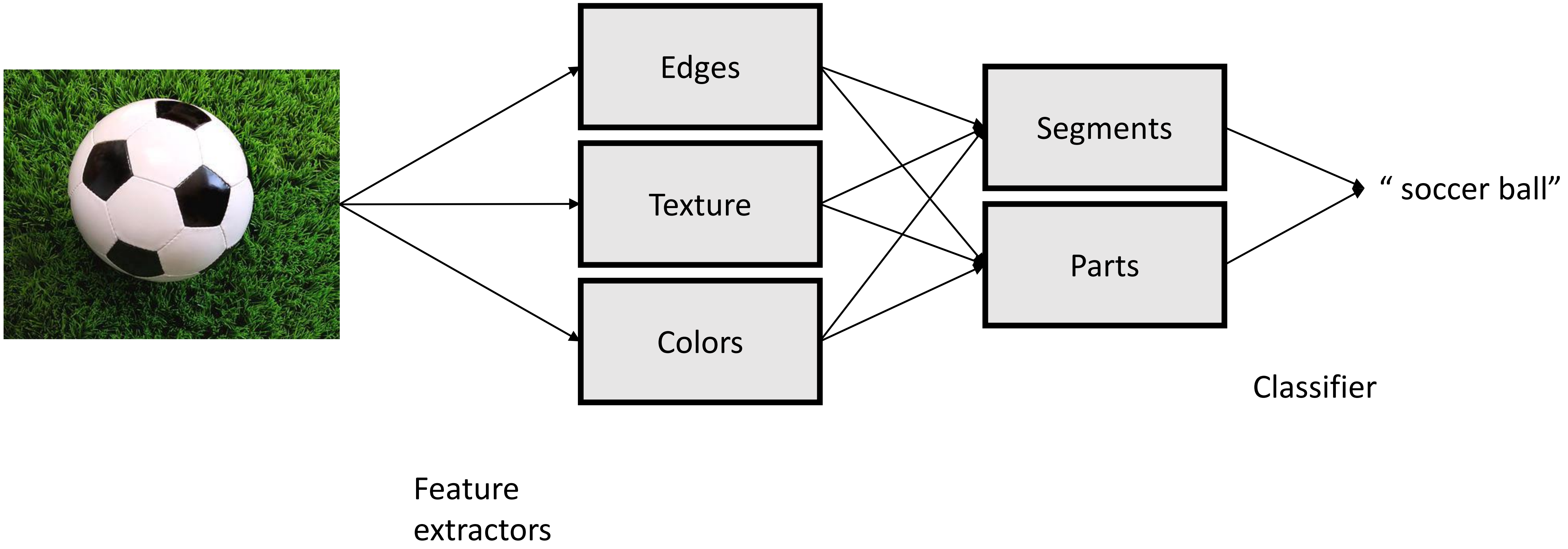


Bengio 2013

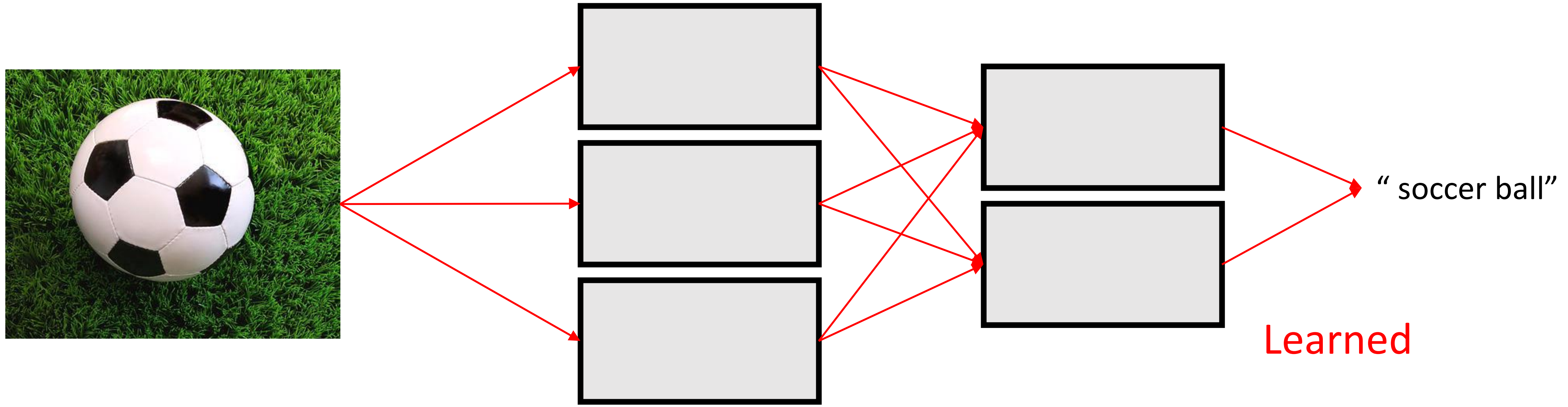


Serre, 2014

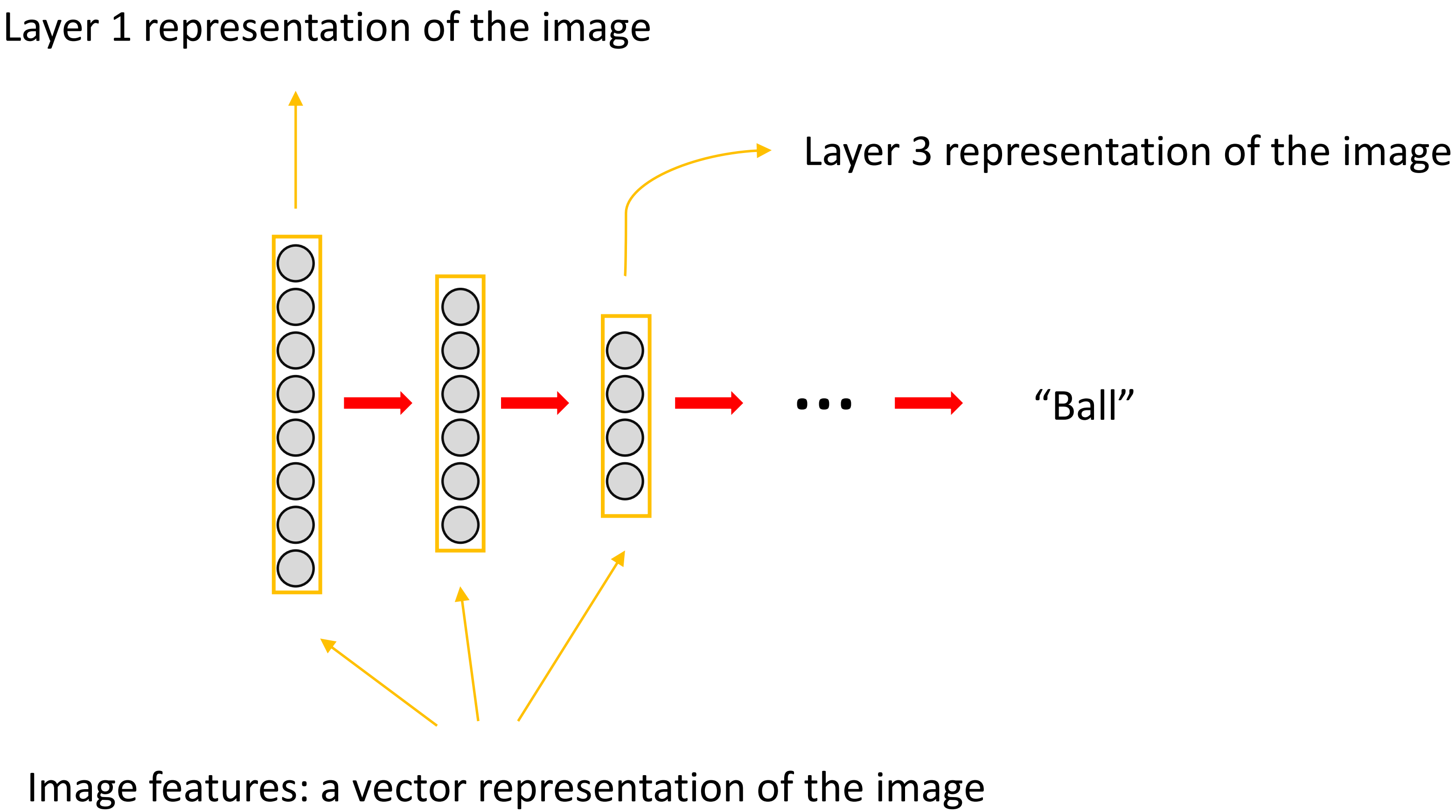
Classic Object Recognition



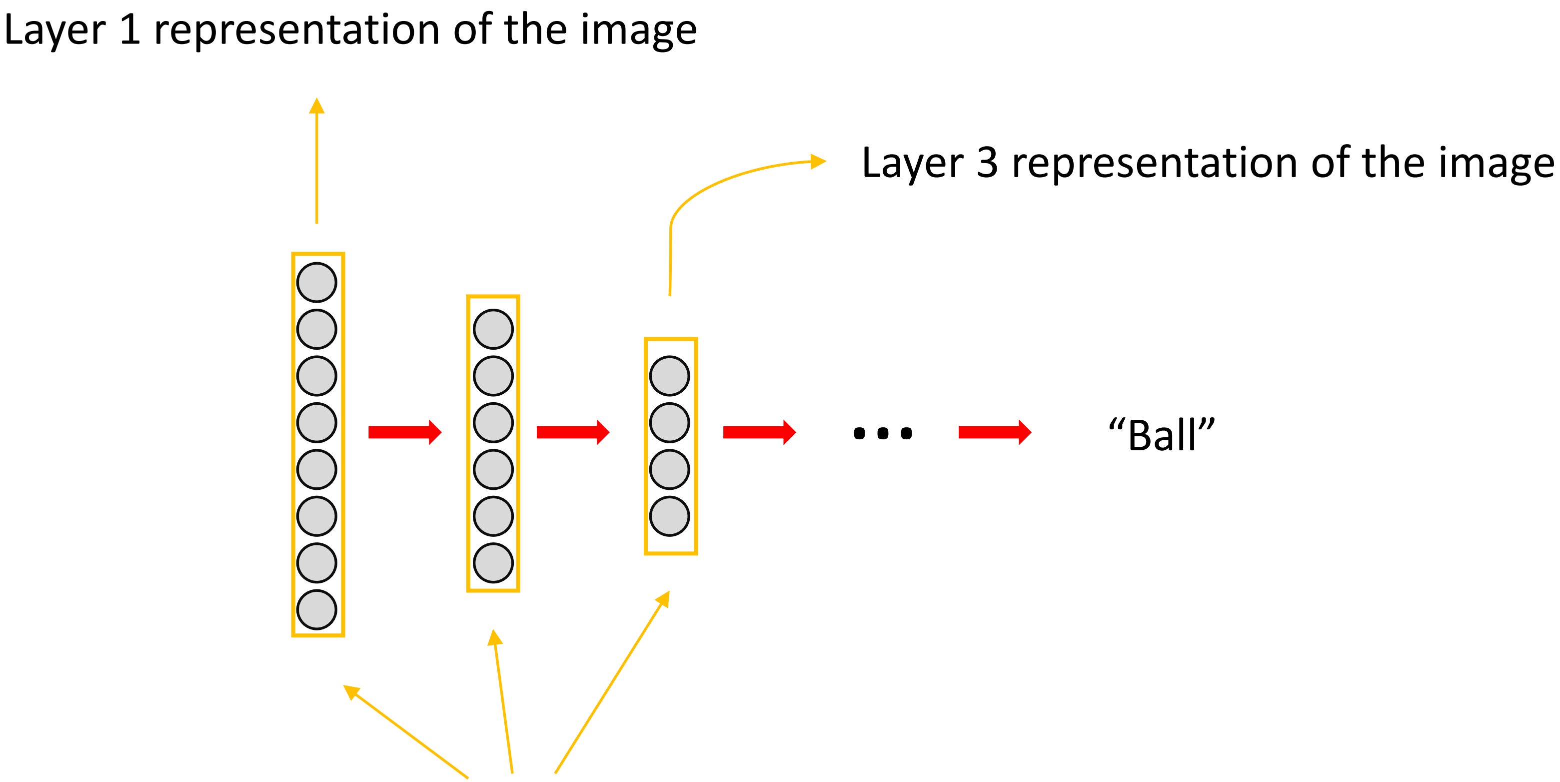
Deep Learning



Internal Representations in Deep Nets



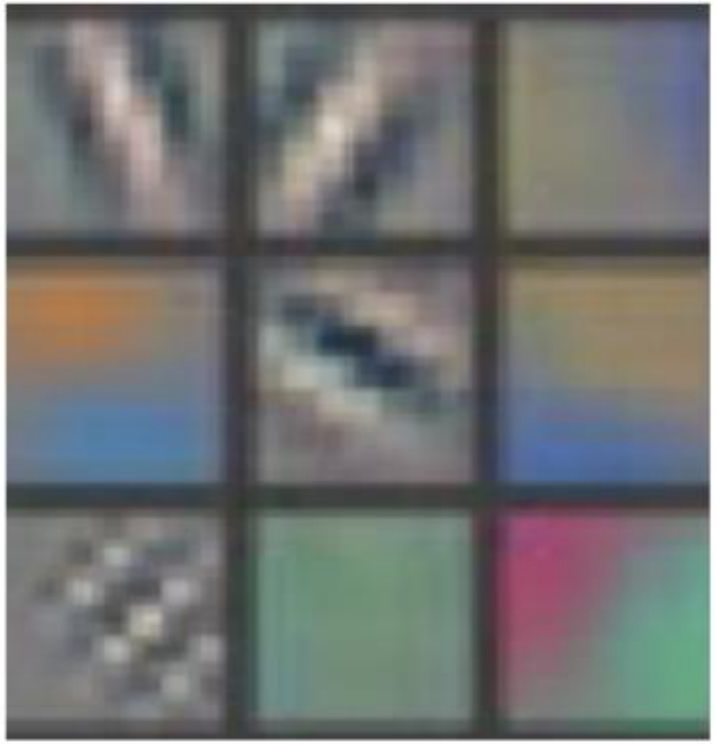
Internal Representations in Deep Nets



A DNN is a multiscale, hierarchical representation of data

Understanding CNNs

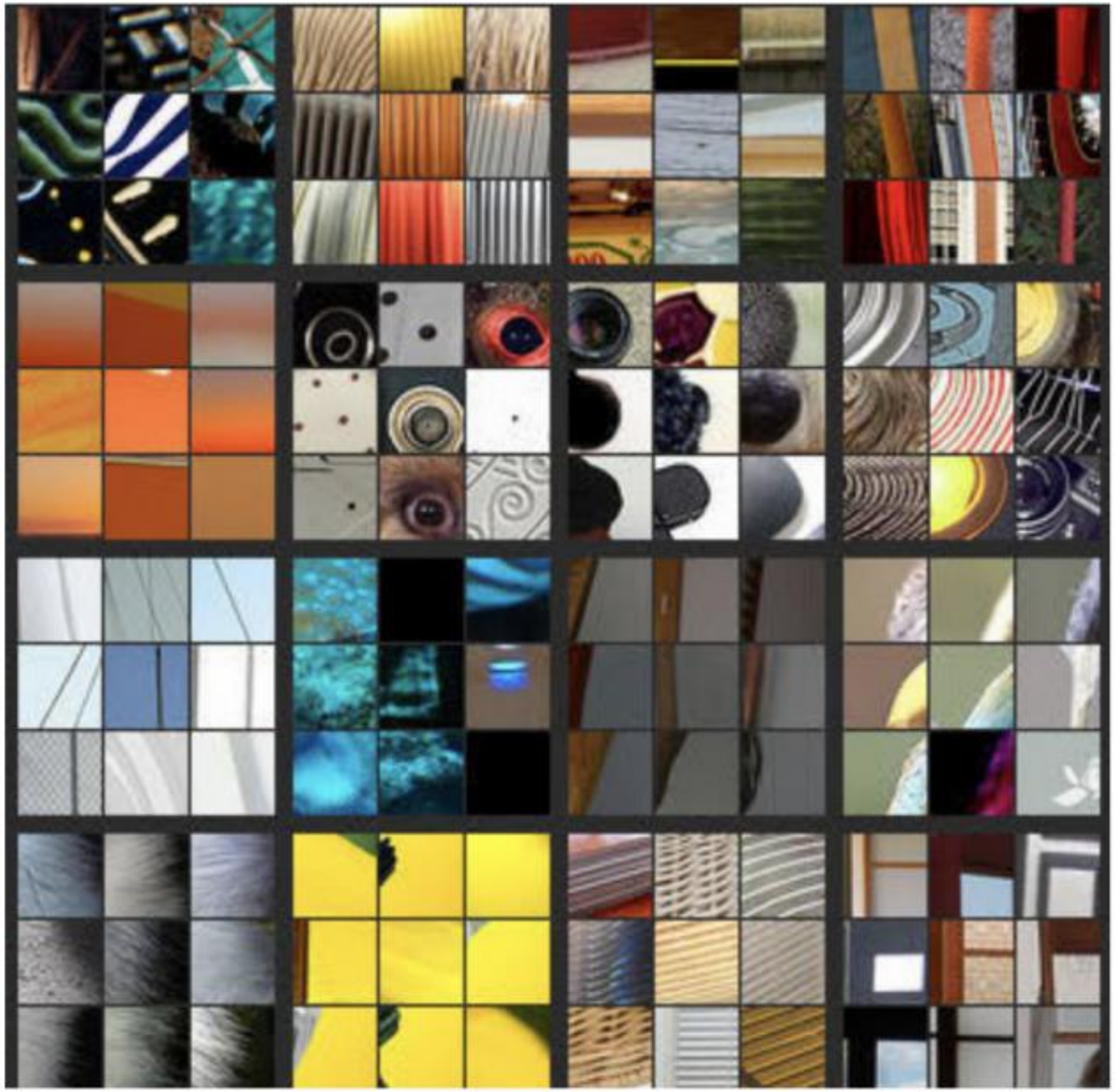
Layer1: Gabor-like filters



Zeiler and Fergus, 2014

Image patches that strongly activates units in Layer 1

Image patches that strongly activates units in Layer 2



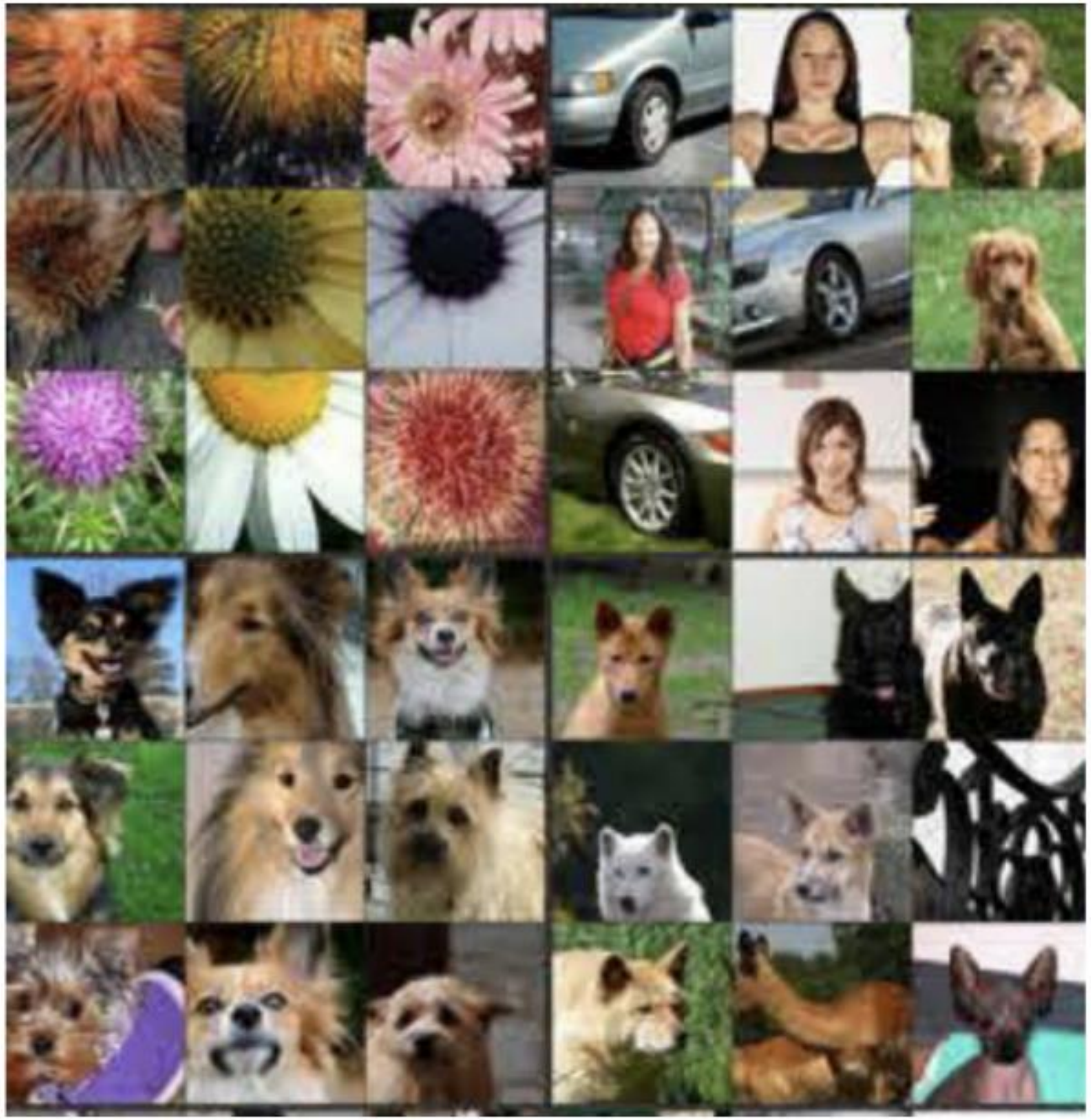
Zeiler and Fergus, 2014

Image patches that strongly activates units in Layer 3

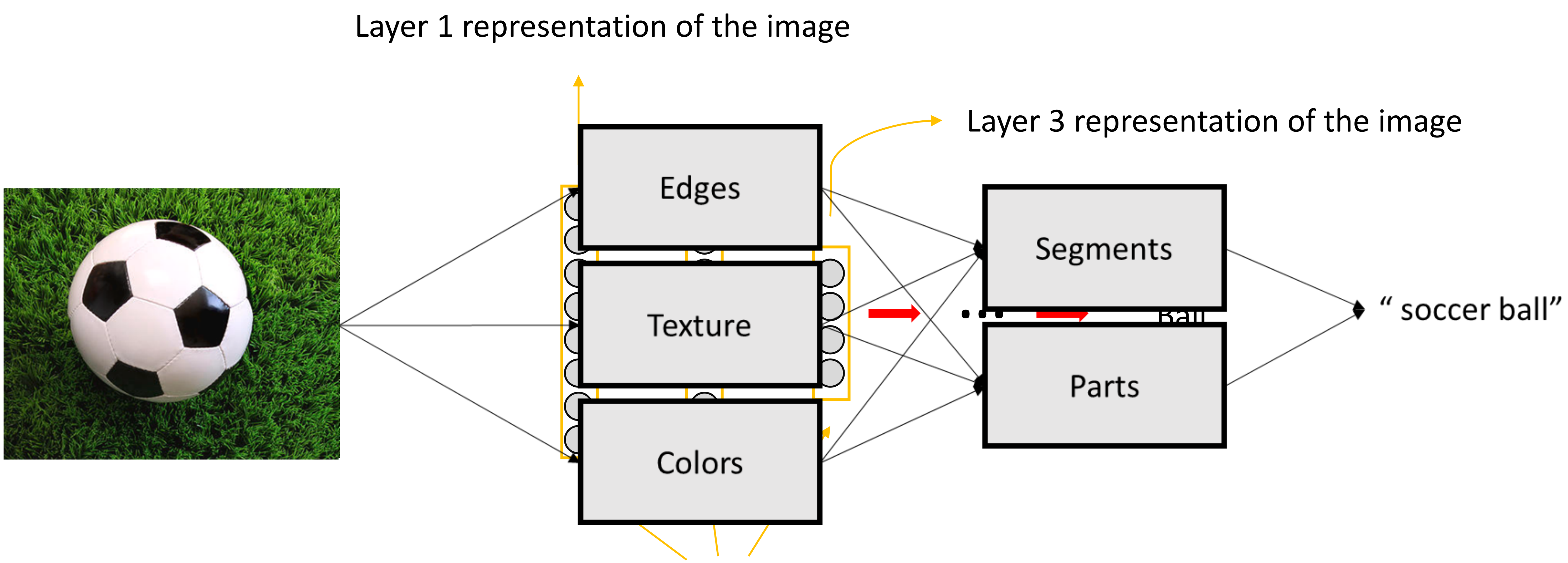
Zeiler and Fergus, 2014

Image patches that strongly activates units in Layer 4

Image patches that strongly activates units in Layer 5

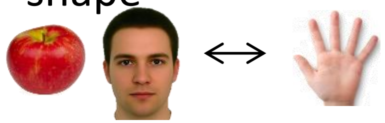


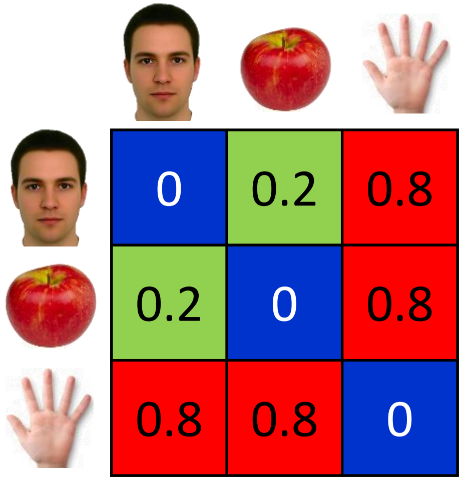
Internal Representations in Deep Nets



Represent image as a vector of neural activations
Perhaps representing a vector of detected texture patterns and object parts

Representational Similarity Analysis





Representational Dissimilarity Matrix (RDM)

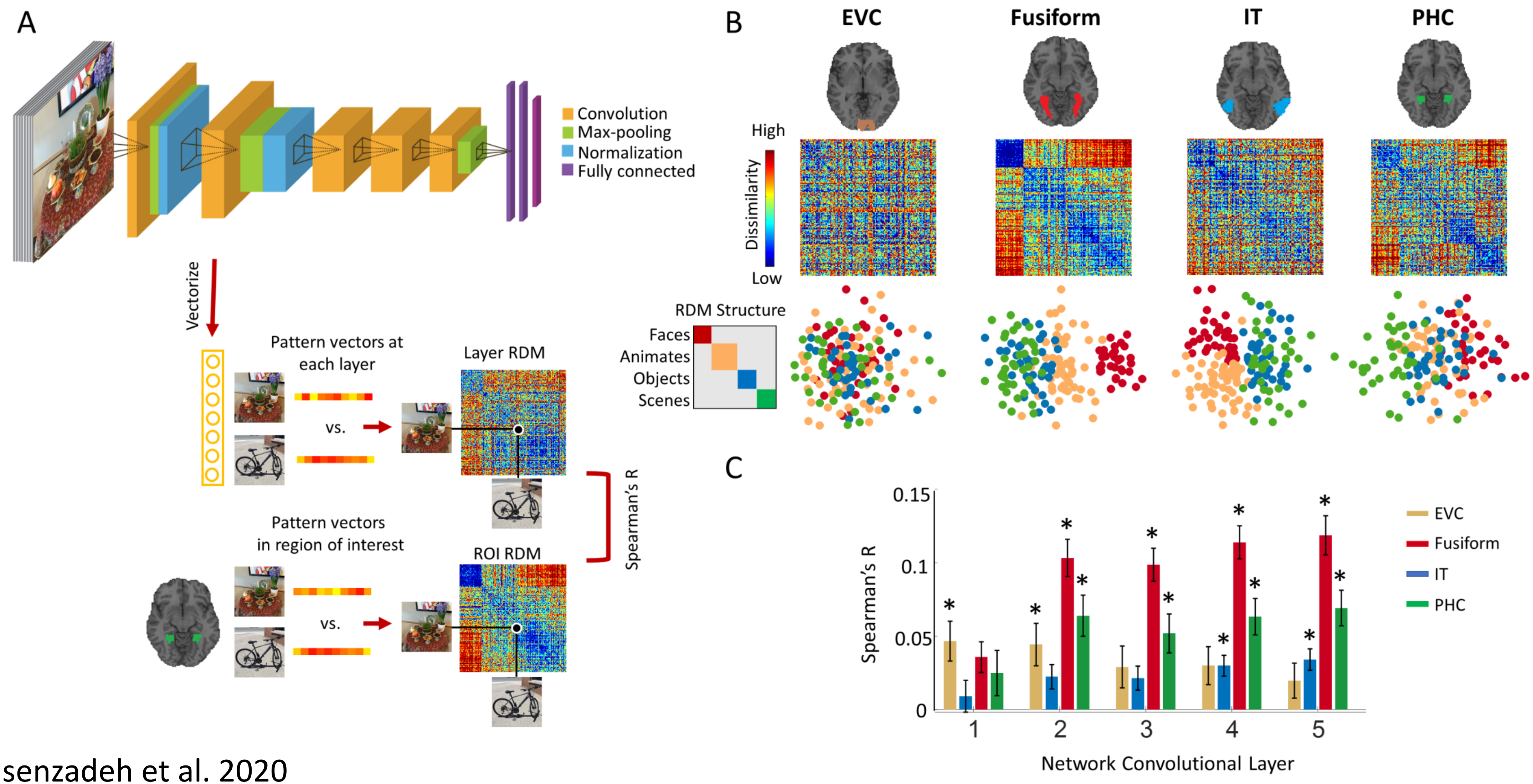
$$||\mathbf{h}_i - \mathbf{h}_j||$$

Neural activation vectors

Kriegesorte et al. 2008

Relating the Brain and CNN representations

- Deep nets organize visual information similarly to how our brains do!



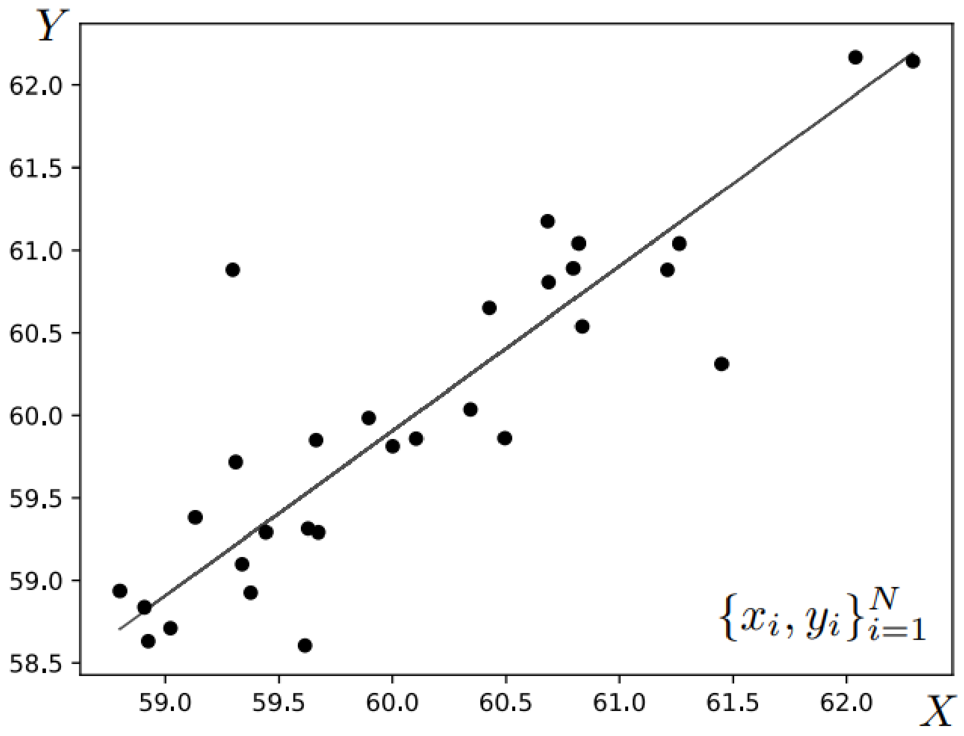
Mohsenzadeh et al. 2020

Transfer Learning

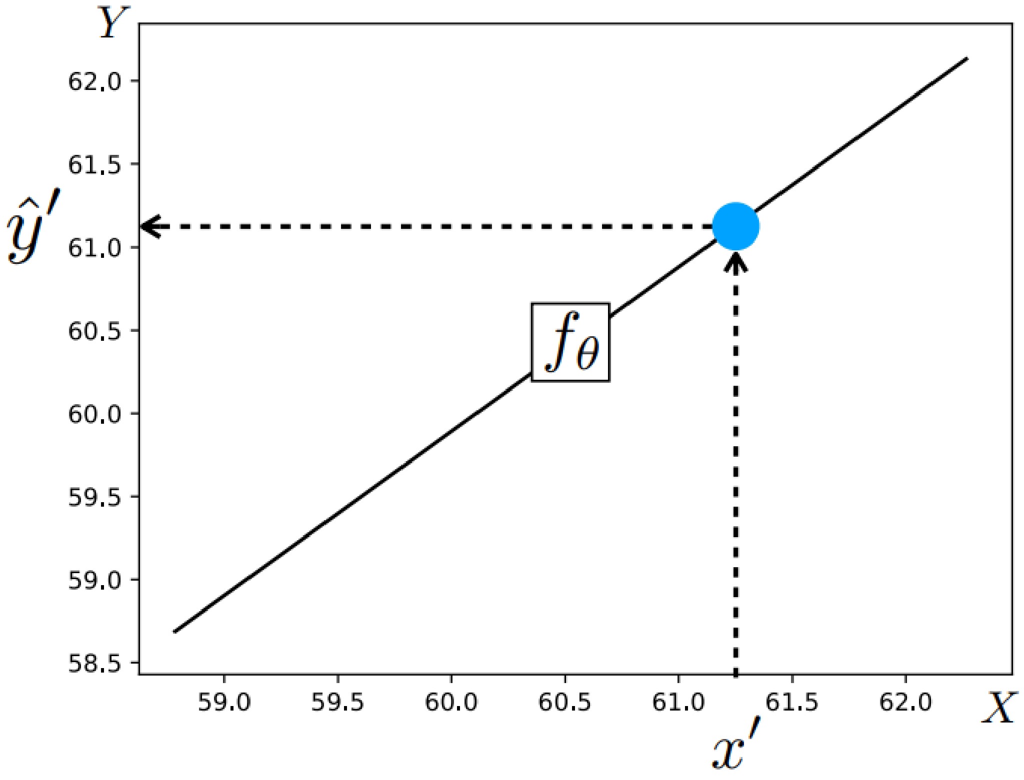
A good representation is one that makes a subsequent learning task easier

- Deep Learning, Goodfellow et al. 2016

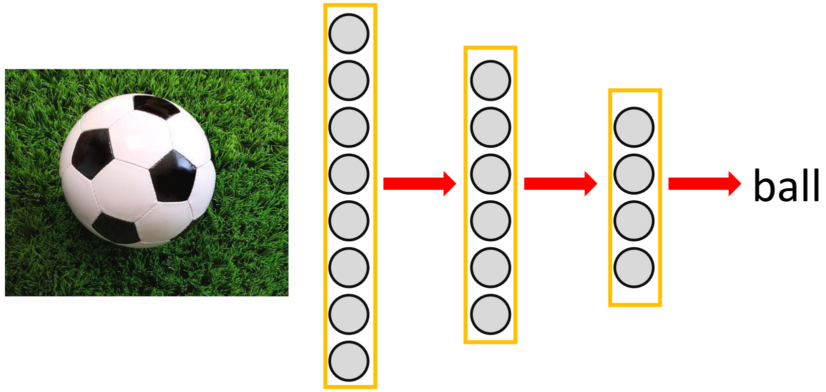
Train


Y
62.0
61.5
61.0
60.5
60.0
59.5
59.0
58.5
59.0
59.5
60.0
60.5
61.0
61.5
62.0
X
$\{x_i, y_i\}_{i=1}^{N}$

Test

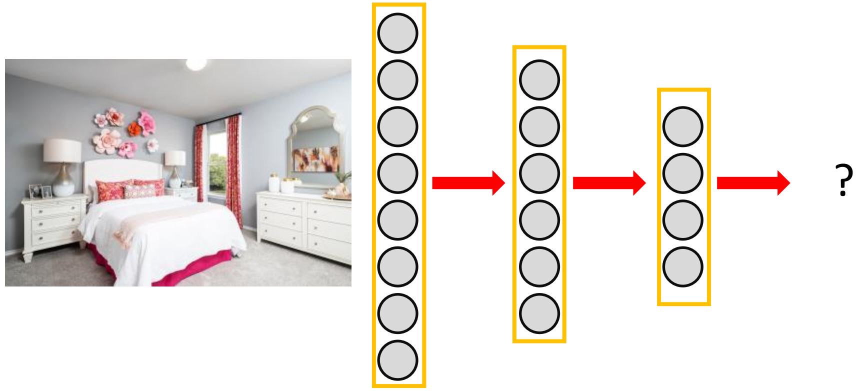

Y
62.0
61.5
61.0
60.5
60.0
59.5
59.0
58.5
$\hat{y}'$
f_θ
59.0
59.5
60.0
60.5
61.0
61.5
62.0
X
x'

Train



object recognition

Test

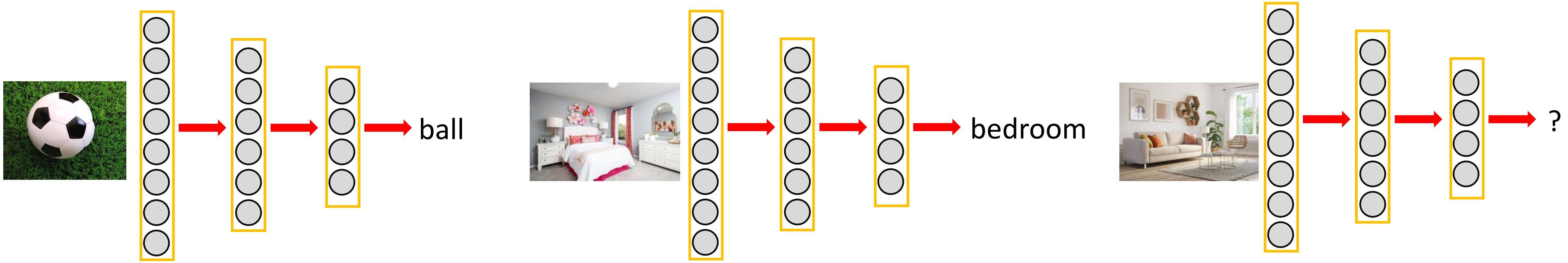


scene recognition

Finetuning

A large object image dataset

A small scene image dataset



Pretraining on object recognition

Finetuning on scene recognition

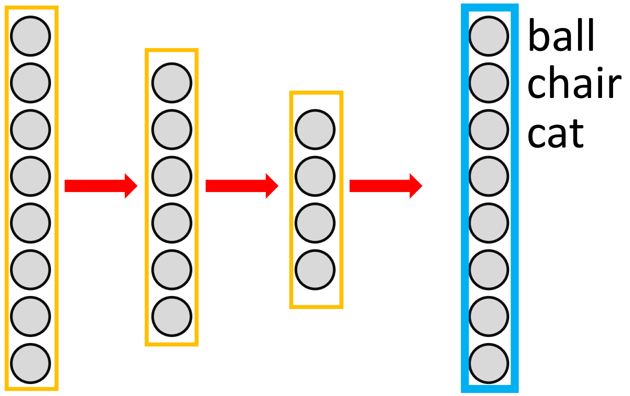
Testing on scene recognition

Finetuning: start with the representations learned on a previous relevant task, and tune it to perform well on the new task

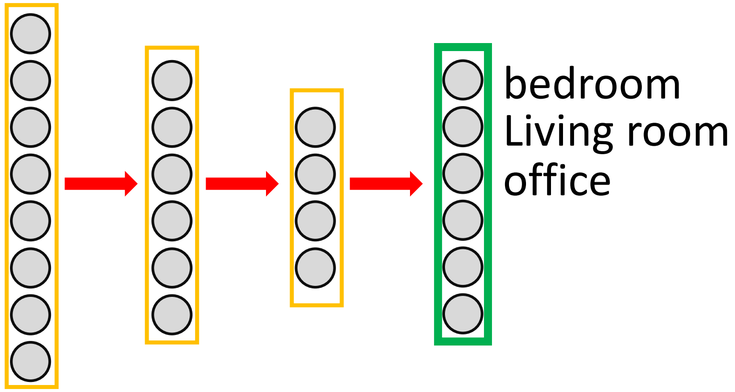
Finetuning

Finetuning: The learned representations is just weight and biases, those are what being transferred!

Pretraining on object recognition



Finetuning on scene recognition



Finetuning

- Pretrain a network on task A (often object recognition), resulting in parameters W and b
- Initialize a second network with some or all of W and b
- Train the second network on task B, resulting in parameters W' and b'

Supervised vs Unsupervised Learning

THE NEXT BIG THING(S) IN UNSUPERVISED MACHINE LEARNING: FIVE LESSONS FROM INFANT LEARNING

Lorijn Zaadnoordijk
Trinity College Institute of Neuroscience
Trinity College Dublin
Dublin, Ireland
L.Zaadnoordijk@tcd.ie

Tarek R. Besold
Neurocat GmbH
Berlin, Germany
tb@neurocat.ai

Rhodri Cusack
Trinity College Institute of Neuroscience
Trinity College Dublin
Dublin, Ireland
CusackRh@tcd.ie

Sep 2020

Supervised Learning

- Hand-curated training data
- Informative
- But
  - Expensive
  - Limited to teacher's knowledge

Infant Learning

- Raw unlabeled training data
- Cheap
- But
  - Noisy
  - Harder to interpret

Learning from examples (supervised learning)

Training data

$$\begin{matrix} \{x_1, y_1\} & & & & \\ \{x_2, y_2\} & \rightarrow & \boxed{\text{Learner}} & \rightarrow & f : X \rightarrow Y \\ \{x_3, y_3\} & & & & \\ \cdots & & & & \end{matrix}$$

$$f^* = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{N} \mathcal{L}(f(x_i), y_i)$$

Learning without examples (unsupervised learning)

Data

$$\begin{array}{c} \{x_1\} \\ \{x_2\} \\ \{x_3\} \\ \cdots \end{array} \quad \rightarrow \quad \boxed{\text{Learner}} \quad \rightarrow \quad ?$$

Representation Learning

Data

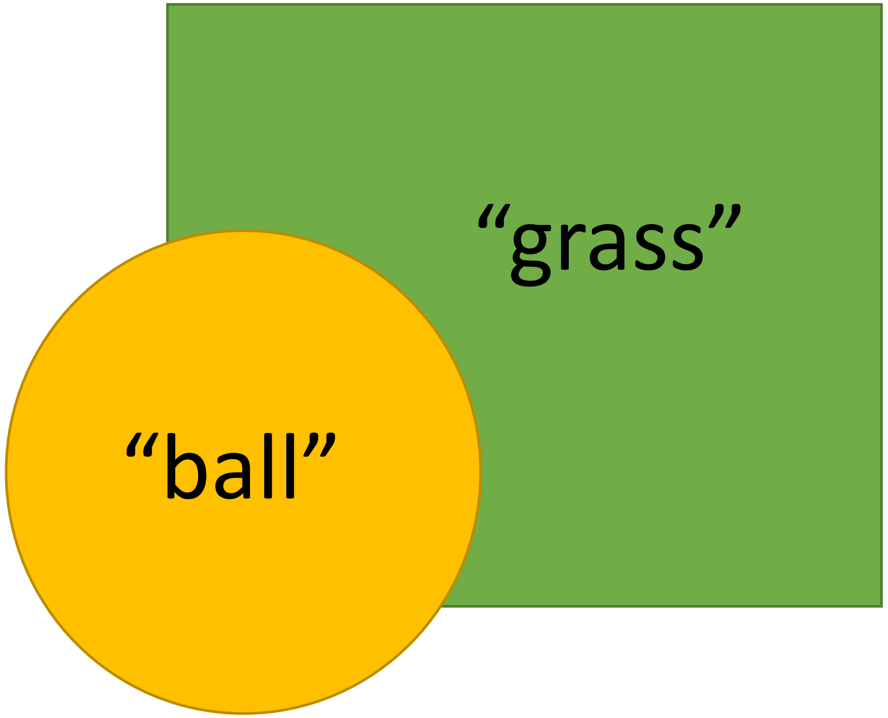
$\{x_1\}$

$\{x_2\}$ $\rightarrow$ Learner $\rightarrow$ Representations

$\{x_3\}$

$\dots$

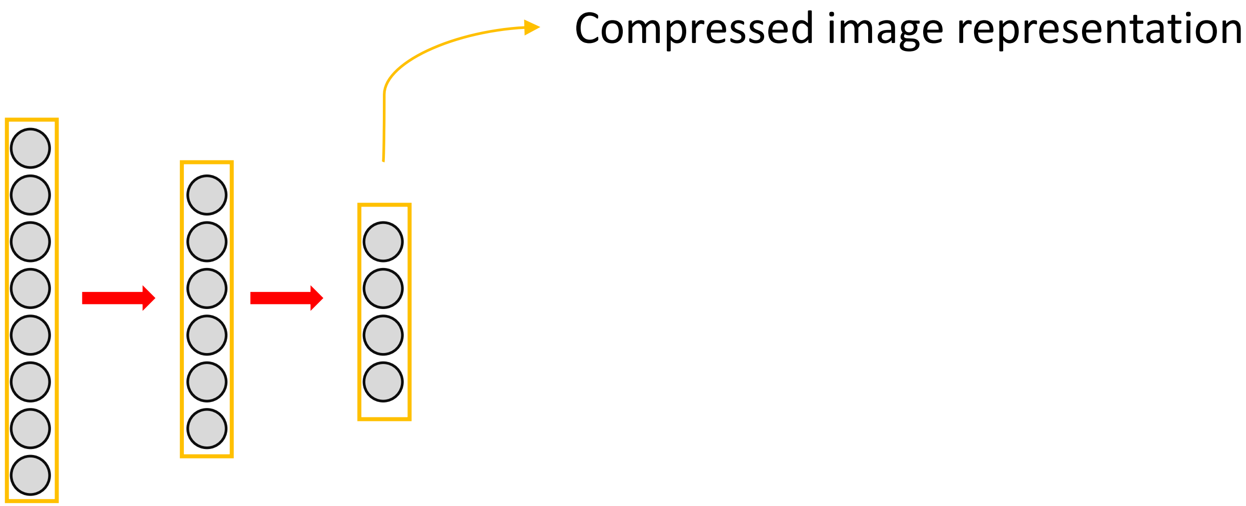
Representation Learning



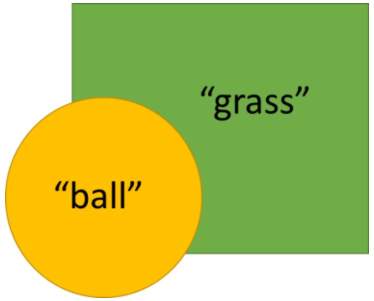
Image

Compact Mental Representation

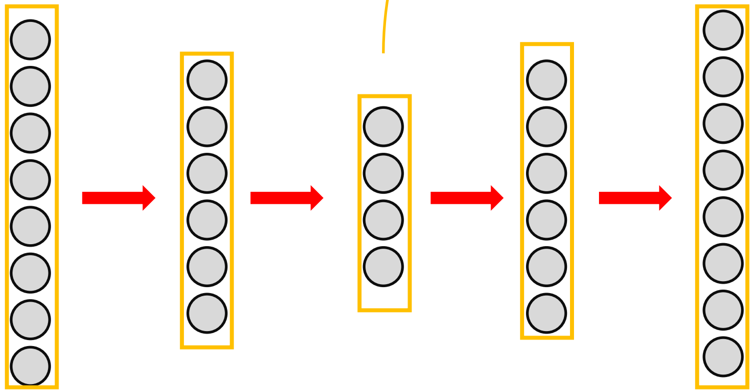
Unsupervised Representation Learning



Unsupervised Representation Learning



Compressed image representation



Image

X

Autoencoder

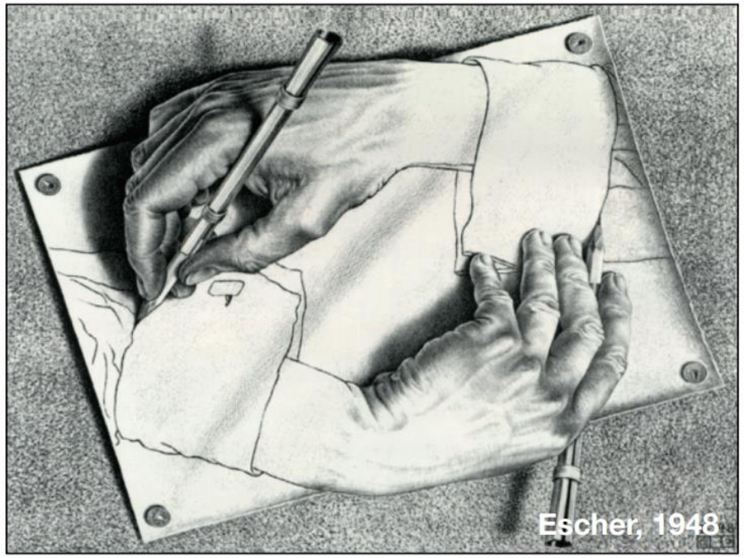
$$argmin_F E_X ||F(X) - X||$$

Reconstructed Image

X' = F(X)

Self-supervised learning

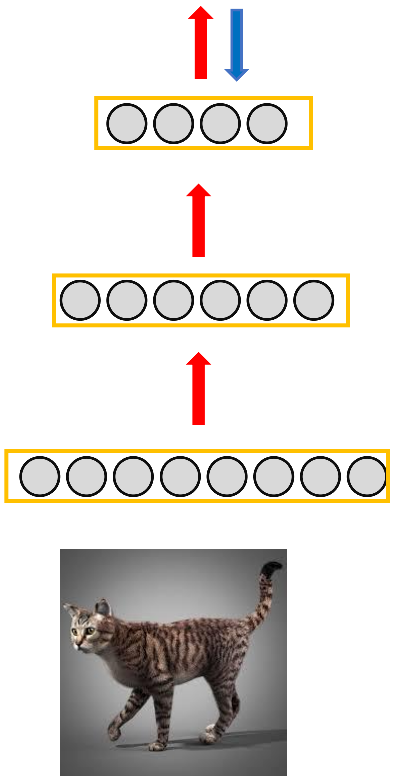
- Trick:
  - Convert “unsupervised” problem into “supervised” empirical risk minimization
  - Do so by cooking up “labels” (prediction targets) from the raw data itself



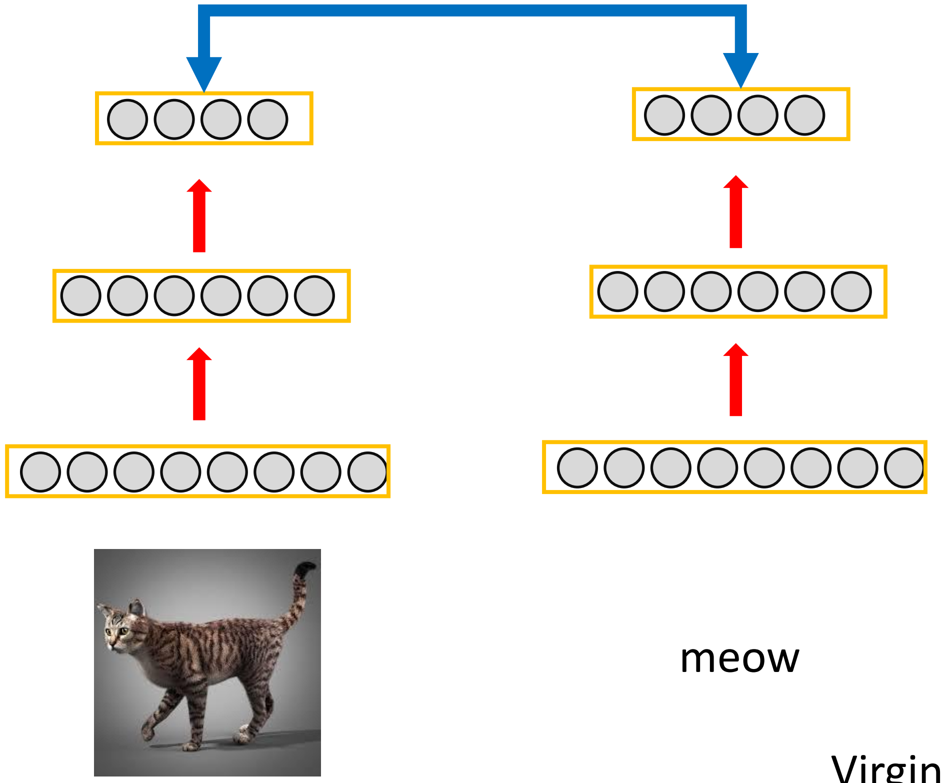
Multisensory self-supervision

Supervised
-labels

cat



Self-Supervised
-Derives label from a
co-occurring input to
another modality



Virginia de Sa. 1994

Multiview self-supervised learning

Data

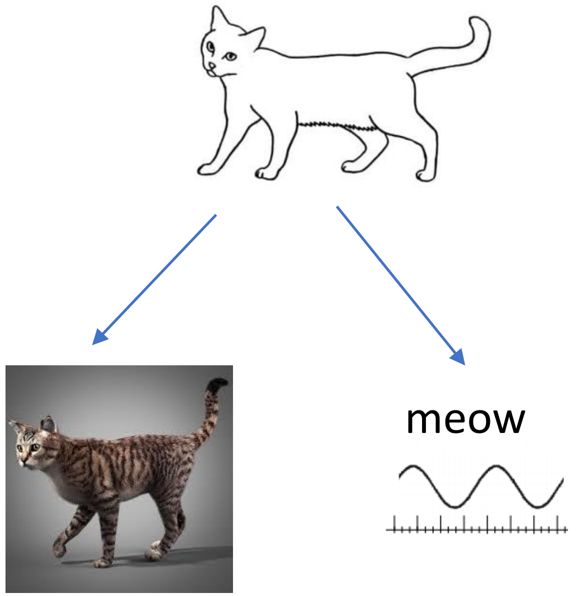
$\{X_i\}_{i=1:N}$

Objective

$$argmin_f \sum_i L(f(g(x_i)) - h(x_i))$$

f

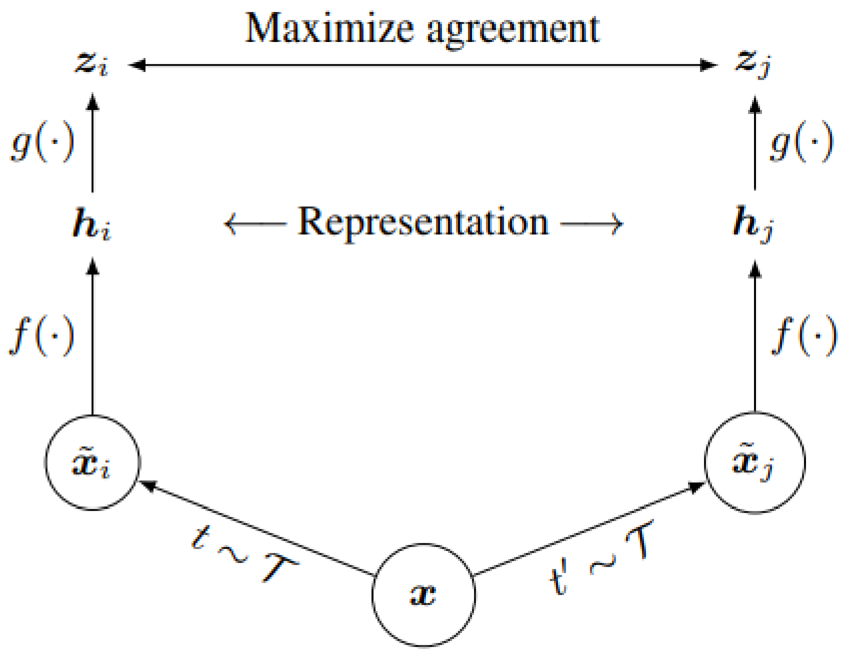
g and h are two views of the data x



Yonglong Tian et al. 2020

Contrastive Self-supervised learning

- composition of data augmentations
- contrastive loss



Ting Chen et al. 2020

Augmentations

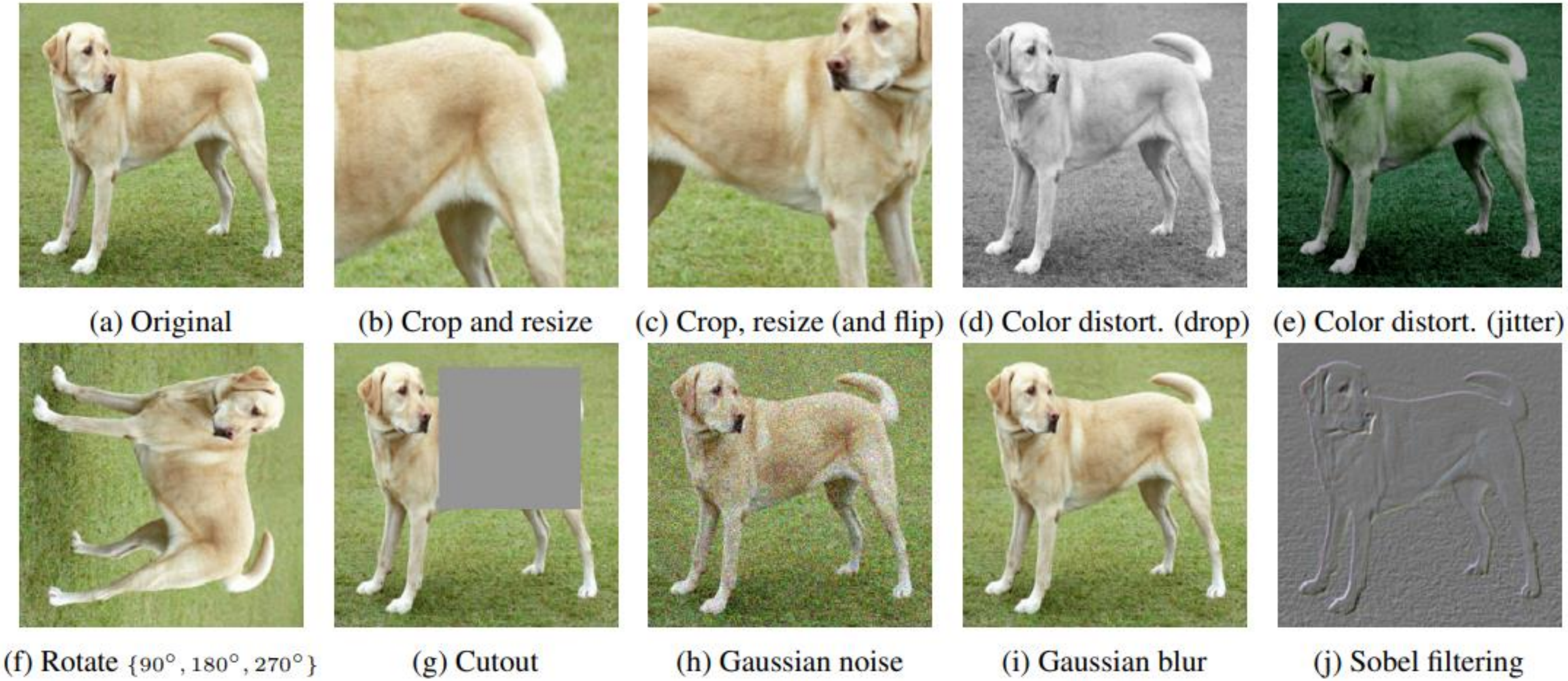


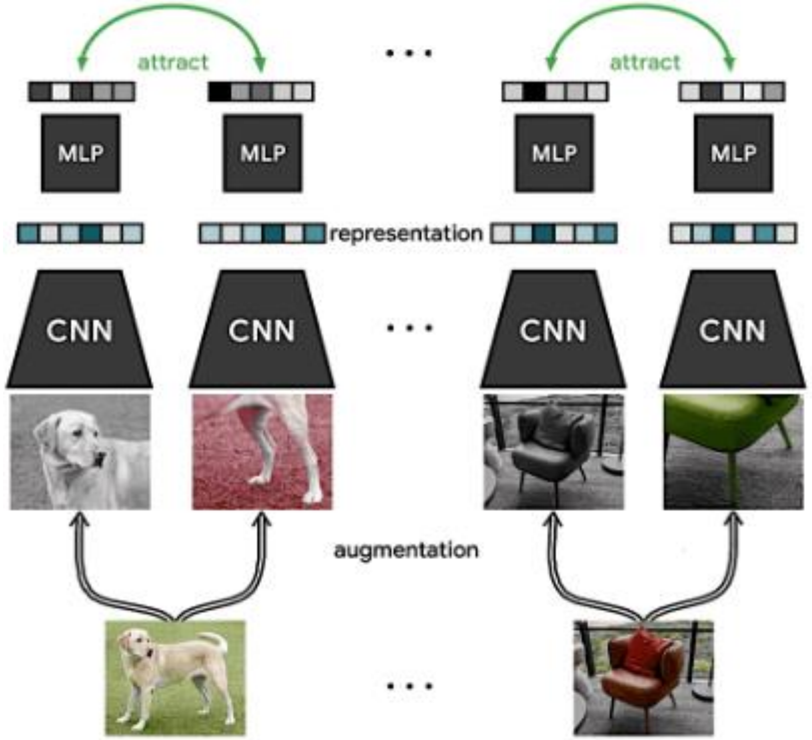
Figure 4. Illustrations of the studied data augmentation operators. Each augmentation can transform data stochastically with some internal parameters (e.g. rotation degree, noise level). Note that we *only* test these operators in ablation, the *augmentation policy used to train our models* only includes *random crop (with flip and resize)*, *color distortion*, and *Gaussian blur*. (Original image cc-by: Von.grzanka)

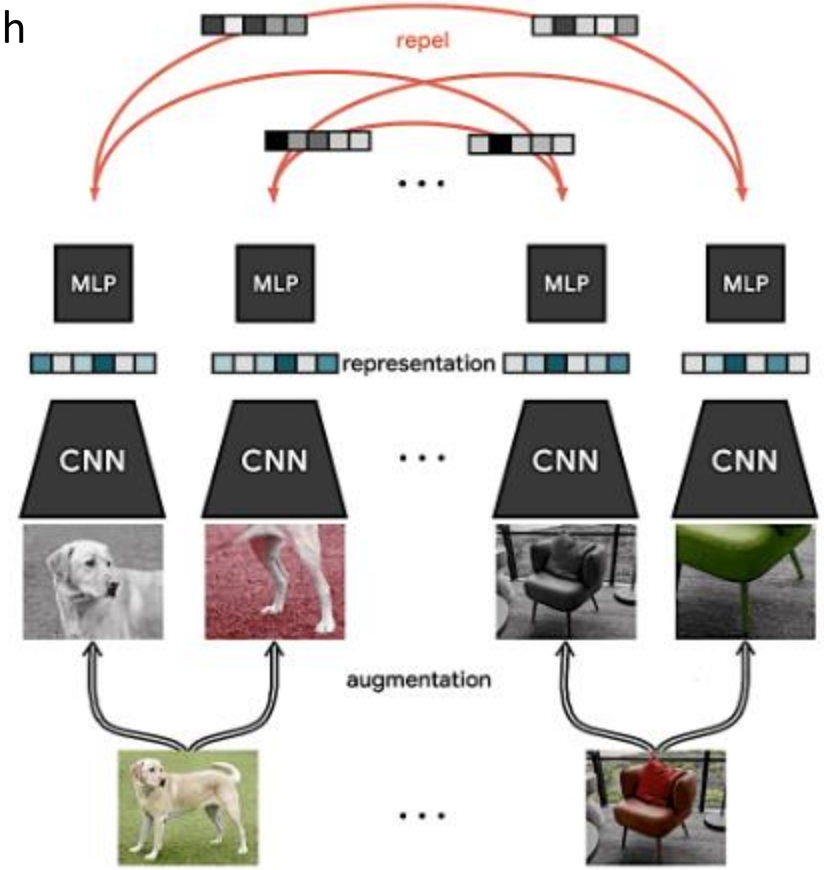
SimCLR

Pair of images originated from the same image

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\overbrace{\boldsymbol{z}_i, \boldsymbol{z}_j})/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\underbrace{\boldsymbol{z}_i, \boldsymbol{z}_k})/\tau)}$$

Pair of images originated from different images within a batch





Ting Chen et al. 2020

CLAR: Contrastive Learning of Auditory Representations

Haider Al-Tahan
Department of Computer Science
The University of Western Ontario
London, Ontario
Canada

Yalda Mohsenzadeh
Department of Computer Science
The University of Western Ontario
London, Ontario
Canada

Abstract

Learning rich visual representations using contrastive self-supervised learning has been extremely successful. However, it is still a major question whether we could use a similar approach to learn superior auditory representations. In this paper, we expand on prior work (SimCLR) to learn better auditory representations. We (1) introduce various data augmentations suitable for auditory data and evaluate their impact on pre-

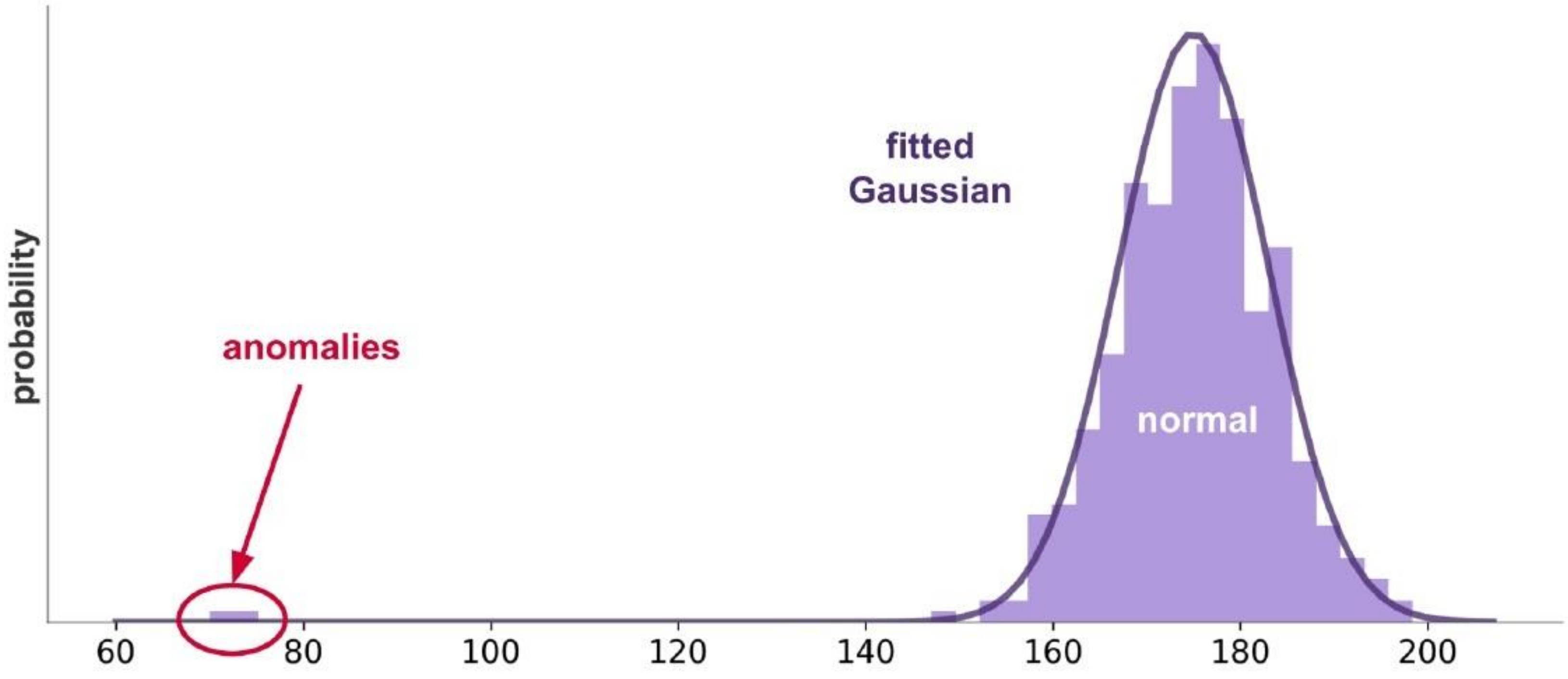
variations in auditory features. Applications of sound understanding range from surveillance (Radhakrishnan et al., 2005) and music classification (Choi et al., 2017; Ibrahim et al., 2020) to audio generation (Engel et al., 2019; Donahue et al., 2019) and deep-fake detection (Mittal et al., 2020).

Achieving automated auditory perception requires the learning of effective representations. Often prior work derive effective representations through discriminative approaches (Park et al., 2019a; Hershey et al., 2017; Tokozume and Harada, 2017; Guzhov et al., 2020). That is, similar to supervised learning, the model

Deep unsupervised representation learning for Anomaly Detection

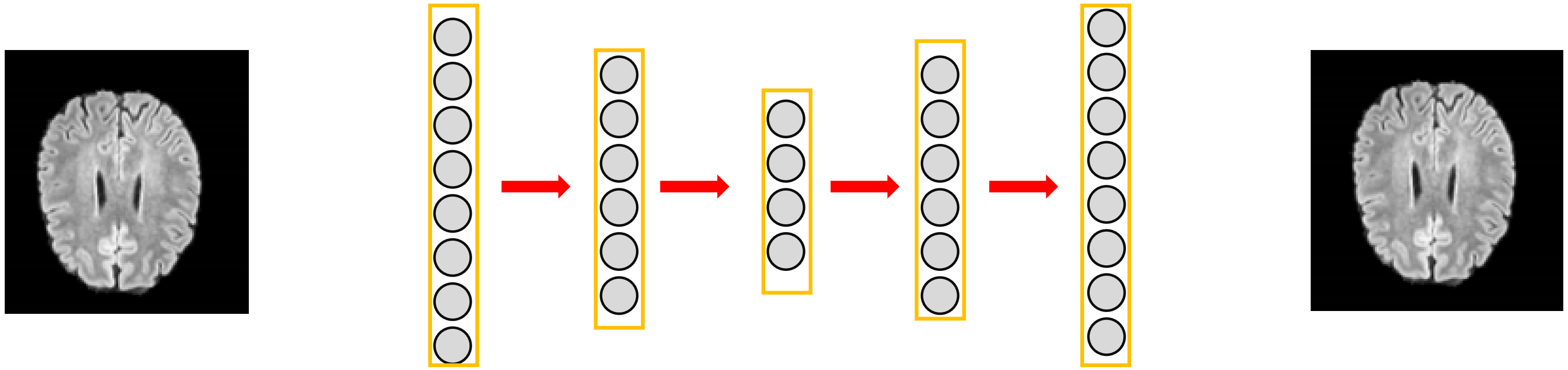
Anomaly Detection?

- Anomaly detection aims at identifying unexpected, abnormal data points given a set of normal data samples only



Example application: Anomaly Detection in Brain MRI Images

- Train an autoencoder to compress and reconstruct the **healthy images**
- **Assumption:** AE trained on only healthy samples cannot properly reconstruct anomalies in pathological data



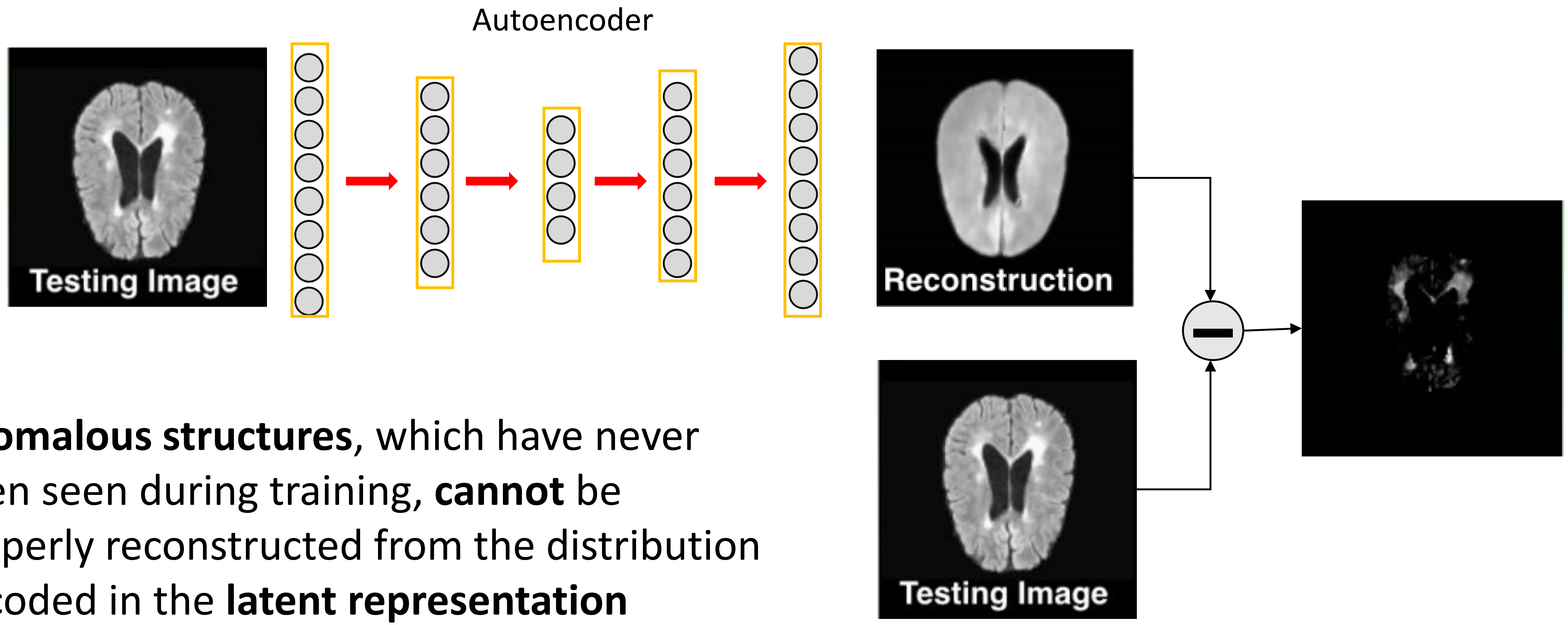
Healthy images

Autoencoder

Reconstructions

Baur et al. 2020

Testing the AE model



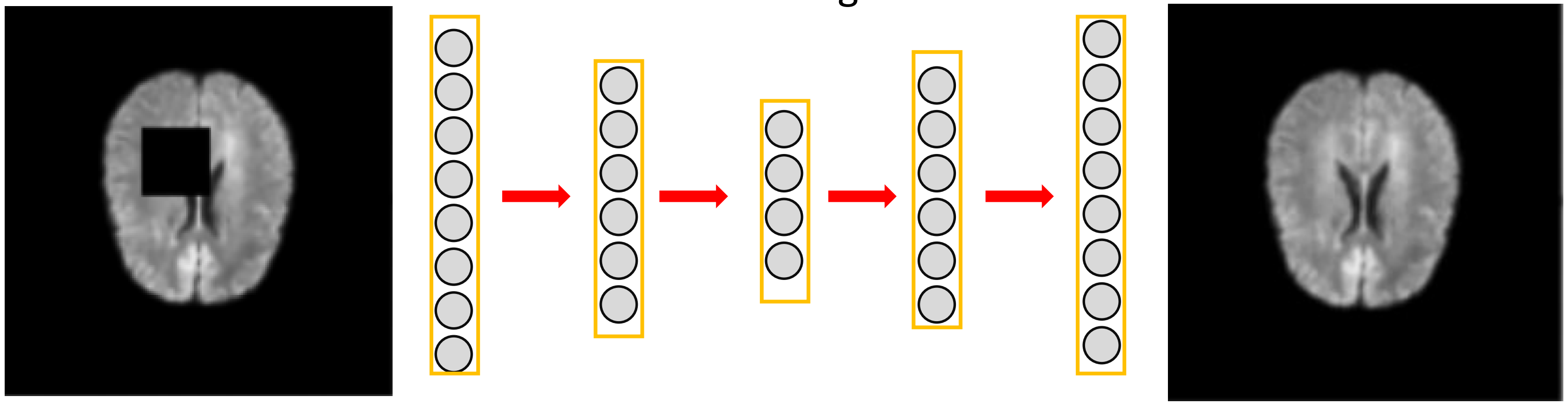
Anomalous structures, which have never been seen during training, **cannot** be properly reconstructed from the distribution encoded in the **latent representation**

Baur et al. 2020

Context Autoencoder (denoising AE)

- AE is trained to recover missing sections in healthy training images

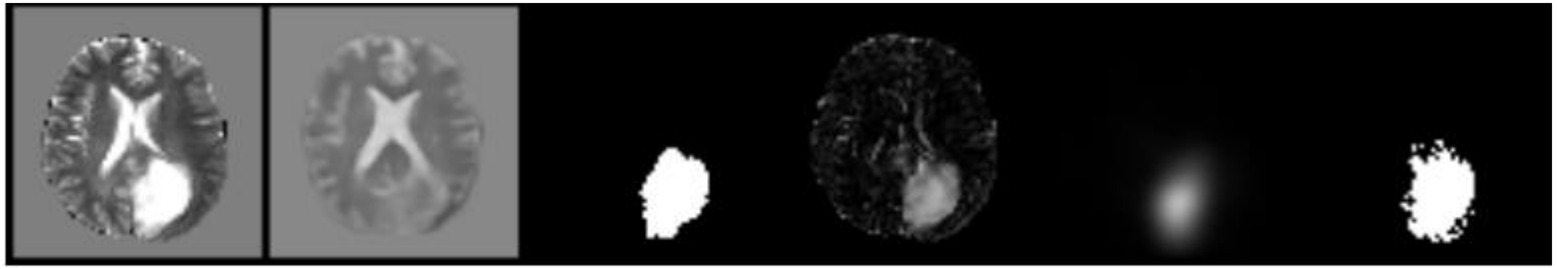
Training



Perturbed/Noisy healthy image

Reconstructed denoised image

Testing



Test image | Reconstructed | Annotated | Error | gradient | segmentation

Zimmerer et al. 2018

Summary

- Deep nets learn representations, just like our brains do
- This is useful because representations transfer — they act as prior knowledge that enables quick learning on new tasks
- Representations can also be learned without labels, which is great since labels are expensive and limiting
- Without labels there are many ways to learn representations. We saw:
    - representations as compressed codes
    - representations that are shared across sensory modalities
    - representations that are predictive of anomality or missing data