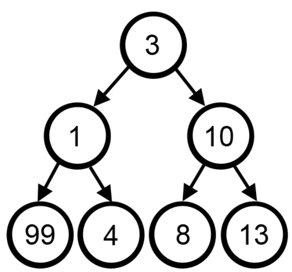
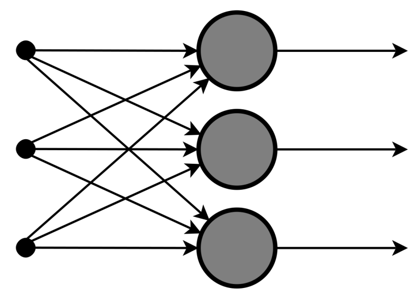
CSE 40171: Artificial Intelligence





Probabilistic Read-Out Layers for Artificial Neural Networks: Combining Bayesian Models with Artificial Neural Networks

Homework #8 is due on 12/11 at 11:59PM

Final Project Deliverable are Due
12/18 at 11:59PM

(See Course Website for Instructions)

Quiz #2 will take place on 12/11 in class. See review checklist on course website.

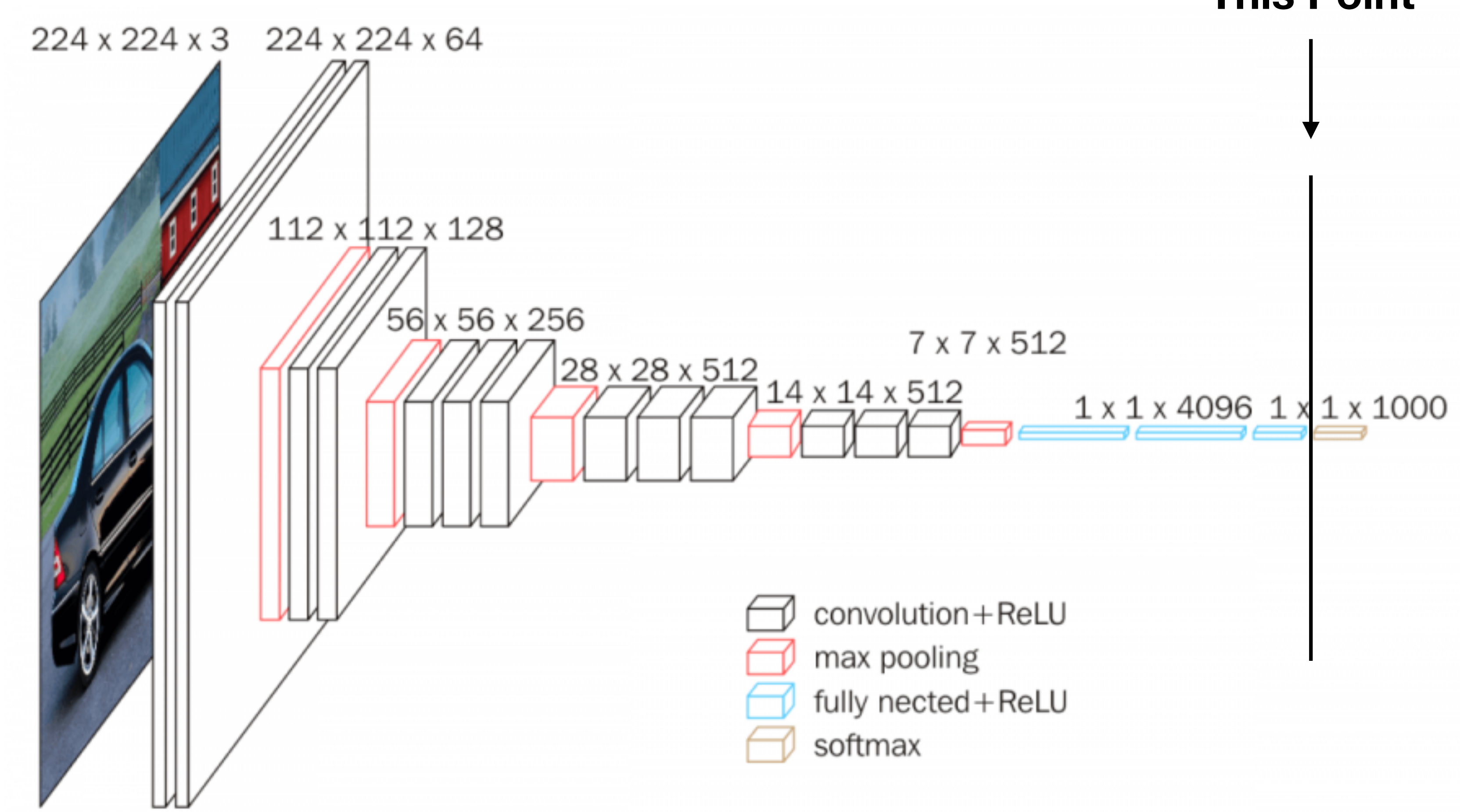
Course Instructor Feedback (CIF)

Deadline: 11:59PM, 12/15/19

How do we deploy Bayes' theorem for decision making?

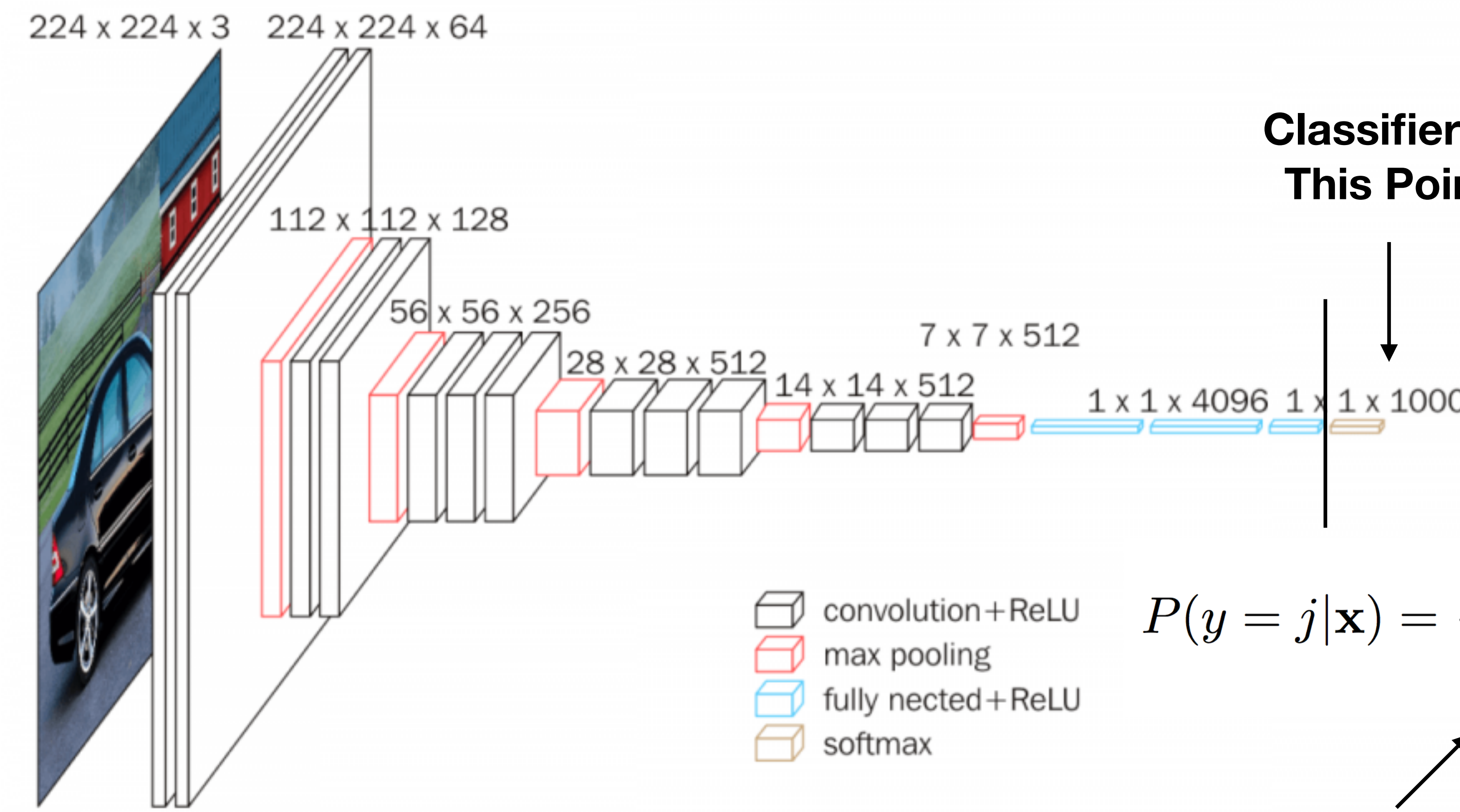
Features

Feature Extractor Until This Point



Simonyan and Zisserman 2015

Decision Making

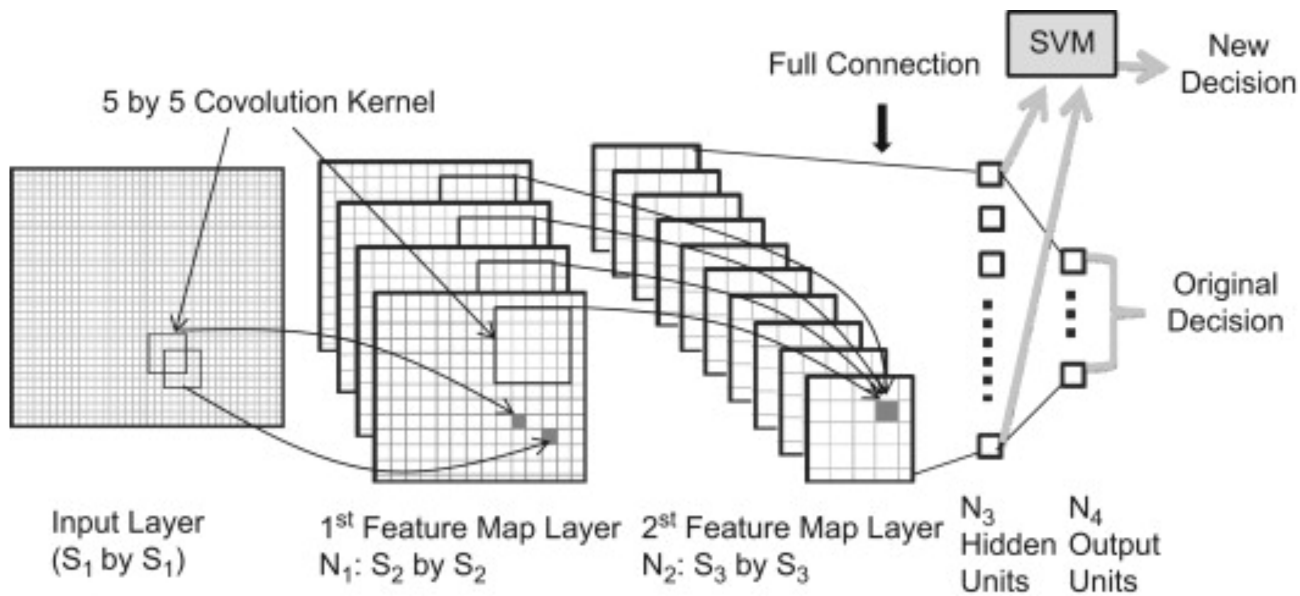


Classifier at This Point

$$P(y = j|\mathbf{x}) = \frac{e^{\mathbf{v_j}(\mathbf{x})}}{\sum_{i=1}^{N} e^{\mathbf{v_i}(\mathbf{x})}}$$

Sum over all of the classes

Decision Making Machines



Niu and Suen Pattern Recognition 2012

Even with these read-out options, why are we lacking good decision making?

Softmax, SVM, and other decision machines cannot express structure in their outputs

Most read-outs are not probabilistic (or weakly calibrated to produce probability-like outputs)

Under-appreciated problem: much of a network's performance is in the classifier, not the feature extraction layers

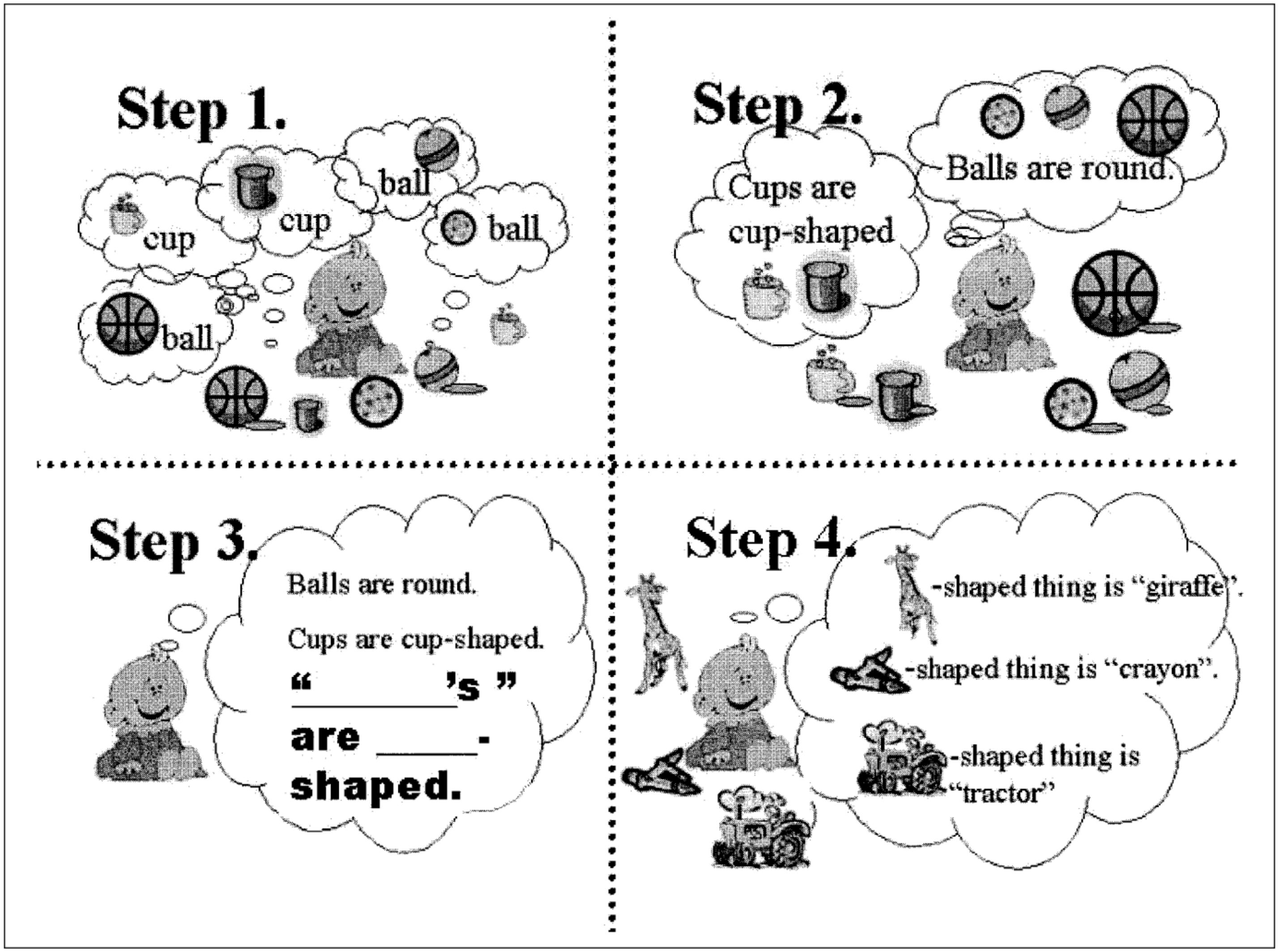
Largest Constraint: Generalization

“These are goats”

Few-shot Learning

Tenenbaum et al. Science 2011

Learning Inductive Biases



Smith et al. Psych. Science 2002

Unsupervised Hierarchical Categorization

Tenenbaum et al. Science 2011

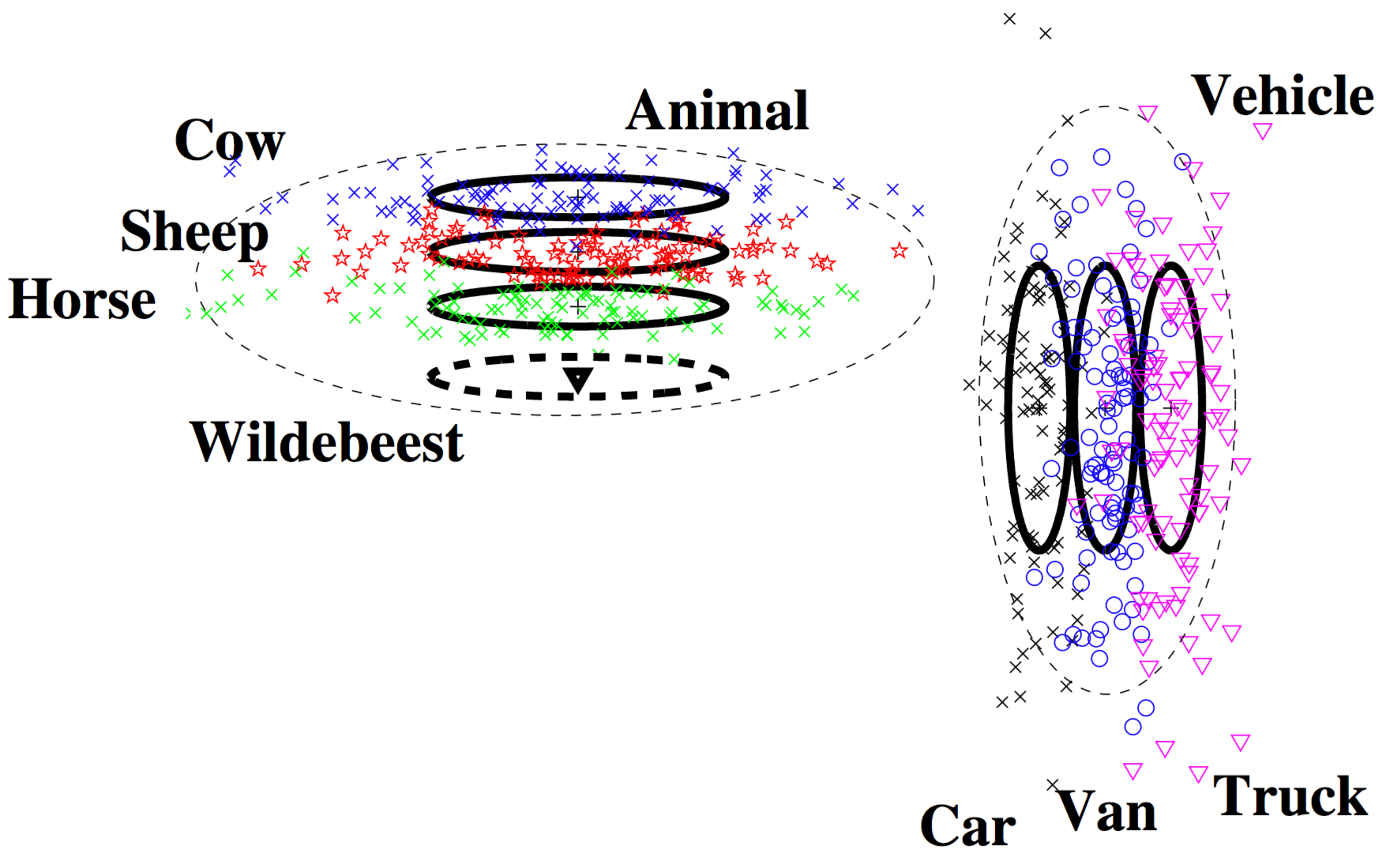
Salakhutdinov et al. Workshop on Unsupervised and Transfer Learning 2012

Introduced Hierarchical Nonparametric Bayesian Models for handcrafted features

Campero et al. CogSci 2017

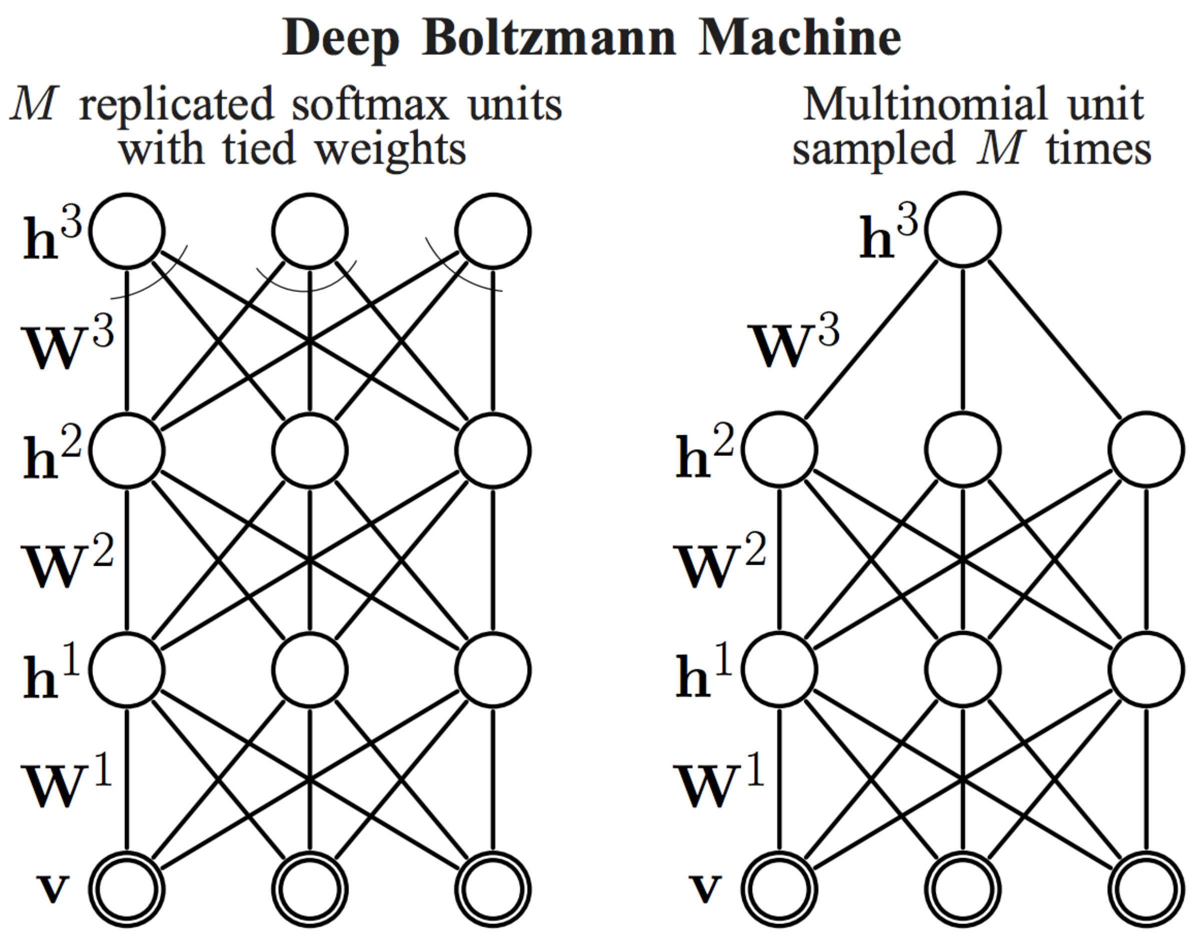
Extends the work of Salakhutdinov et al. for use with features from deep learning

Learning a Class-Specific Similarity Metric from One Example



Salakhutdinov et al. Workshop on Unsupervised and Transfer Learning 2012

Complex feature spaces with a hierarchical semantic structure

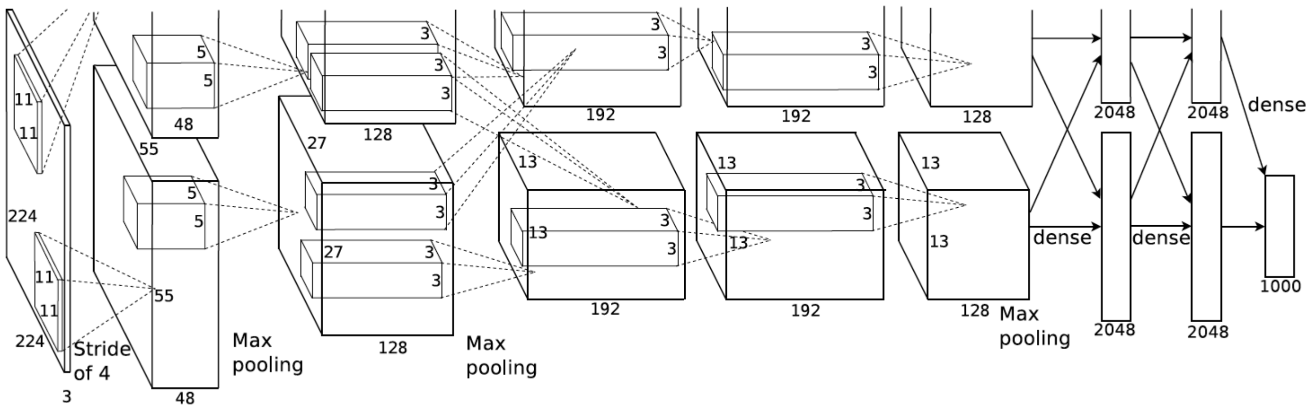


Motivation: Generative Model

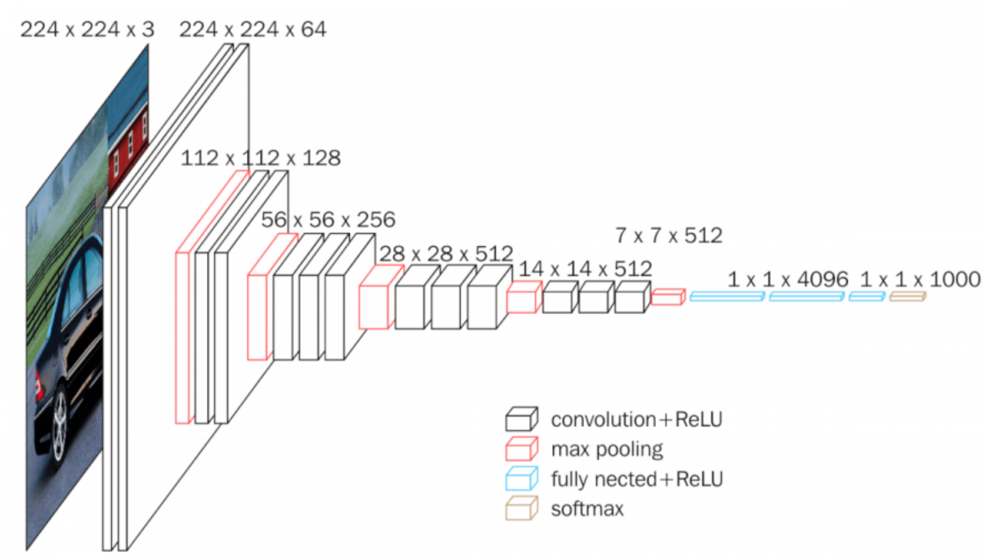
Salakhutdinov et al. 2013

Modern Deep Network Features

AlexNet: Krizhevsky et al. 2012



VGG-16: Simonyan and Zisserman 2015



Generative Semantic Organization

Step 1: Extract features from a chosen DNN

Step 2: Hierarchical Bayesian Model's parameters are inferred by approximating the posterior via Markov Chain Monte Carlo methods

Generative Semantic Organization

Assume two-level hierarchy where N observed inputs are partitioned into C basic-level categories

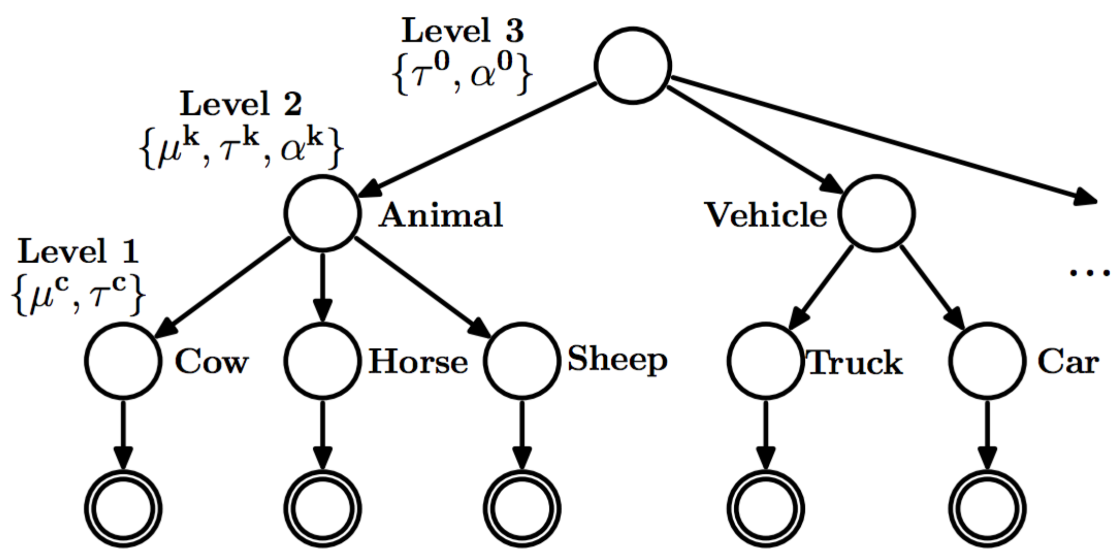
These categories are in turn partitioned into K supercategories

The distributions over features of each basic-level category are assumed to be multivariate Gaussian with a category specific mean M_c and with precision terms τ_d^c that are assumed to be independent

Conjugate Normal-Gamma prior over $\{\mu_c, \tau_c\}$; determined by the supercategory specific level-2 parameters μ_k, τ_k, α_k.

μ_k and τ_k constitute the expected values of the lower-level parameters and α_k controls the variability of τ_c around its mean

For the conjugate priors over the level-2 parameters, assume Normal, Exponential and Inverse-Gamma distributions that are further shaped by parameters α_0 and γ_0



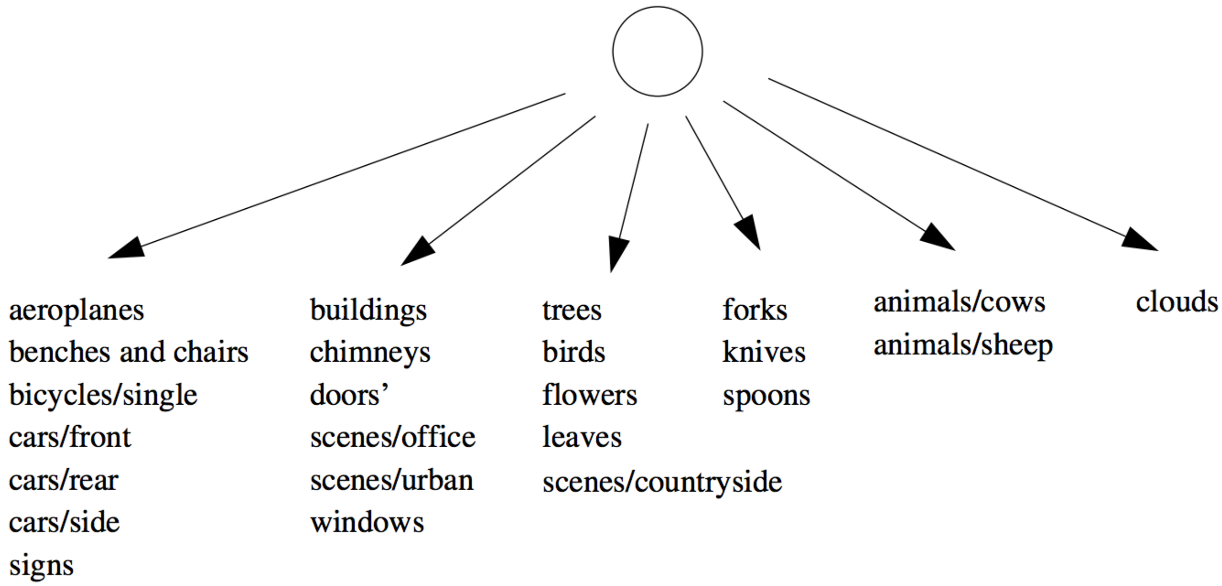
Bayesian Inference

Given a set of observations, the model iteratively performs Bayesian inference by alternating between sampling the parameters and inferring category assignments

When learning distributions at each step of the iteration, supercategory membership is fixed and the parameters are sampled from posteriors that are analytically computed using the conjugate priors

Supercategory membership for each category is learned by fixing the current parameters and the rest of the hierarchical structure. Assignment to any of the existing supercategories or to a newly created one.

MSR Cambridge Dataset



http://research.microsoft.com/en-us/projects/objectclassrecognition/

Baseline: “texture-of-textures” (handcrafted) features

Model	Category: Cow				Category: Flower				Average			
	1 ex	2 ex	4 ex	20 ex	1 ex	2 ex	4 ex	20 ex	1 ex	2 ex	4 ex	20 ex
HB	0.77	0.81	0.84	0.89	0.71	0.75	0.78	0.81	0.76	0.80	0.84	0.87
HB-Flat	0.62	0.69	0.80	0.89	0.59	0.64	0.75	0.81	0.65	0.71	0.78	0.87
HB-Var	0.61	0.73	0.83	0.89	0.60	0.68	0.77	0.81	0.64	0.74	0.81	0.87
Euclidean	0.59	0.61	0.63	0.66	0.55	0.59	0.61	0.64	0.63	0.66	0.69	0.71
Oracle	0.83	0.84	0.87	0.89	0.77	0.79	0.80	0.81	0.82	0.84	0.86	0.87
MLE	0.58	0.64	0.78	0.89	0.55	0.62	0.72	0.81	0.62	0.67	0.77	0.87

Salakhutdinov et al. Workshop on Unsupervised and Transfer Learning 2012

AUROC on the MSR dataset in the one-shot learning task

	# Examples from Withheld Class							
	Alexnet				VGG			
	1ex	2ex	4ex	20ex	1ex	2ex	4ex	20ex
Oracle	.99	1	1	1				
HB-Full	.91	.96	.98	.99	.92	.97	.98	.99
One Supercategory	.87	.94	.97	.99	.88	.95	.98	.99
NearestN	.84	.86	.87	.90	.89	.90	.92	.95
T of T*	.76	.80	.84	.87				

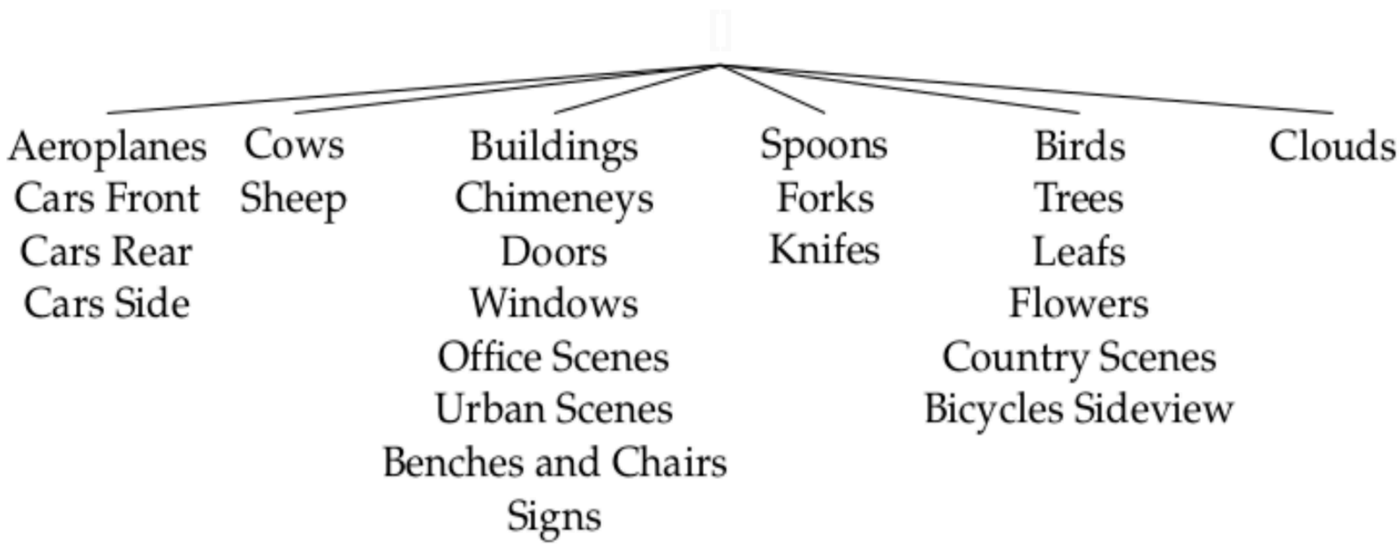
Campero et al. CogSci 2017

AUROC on the MSR dataset with limited training data

	# Examples from Withheld Class							
	Alexnet				VGG			
	1 ex	2 ex	4 ex	20 ex	1 ex	2 ex	4 ex	20 ex
# Training Examples								
1 ex	.87	.87	.88	.89	.90	.90	.90	.92
4 ex	.92	.96	.99	.99	.93	.97	.98	.99
10 ex	.92	.96	.99	.99	.92	.96	.98	.99
18 ex	.92	.95	.98	.99	.91	.96	.98	.99
All examples	.91	.96	.98	.99	.92	.97	.98	.99

Campero et al. CogSci 2017

MSR semantic tree discovered by the Full Model

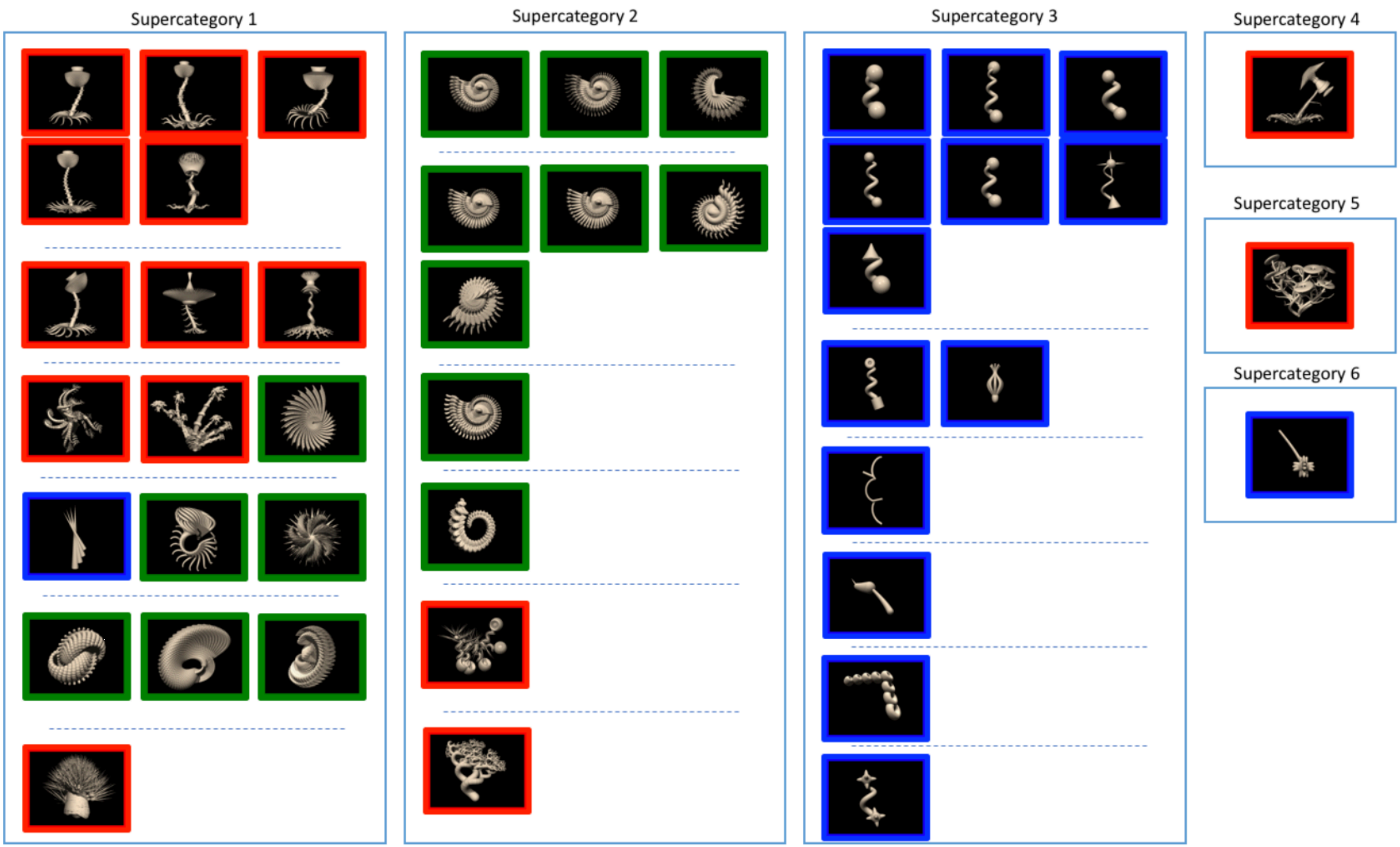


Campero et al. CogSci 2017

Ground-truth Gazoobian Framework

Tenenbaum et al. Science 2011

Model's Inferred Semantic Hierarchy of Gazoobian Objects



Campero et al. CogSci 2017