Intelligent Perception

S. M. Ali Eslami

December 2016

THE UNIVERSITY of EDINBURGH

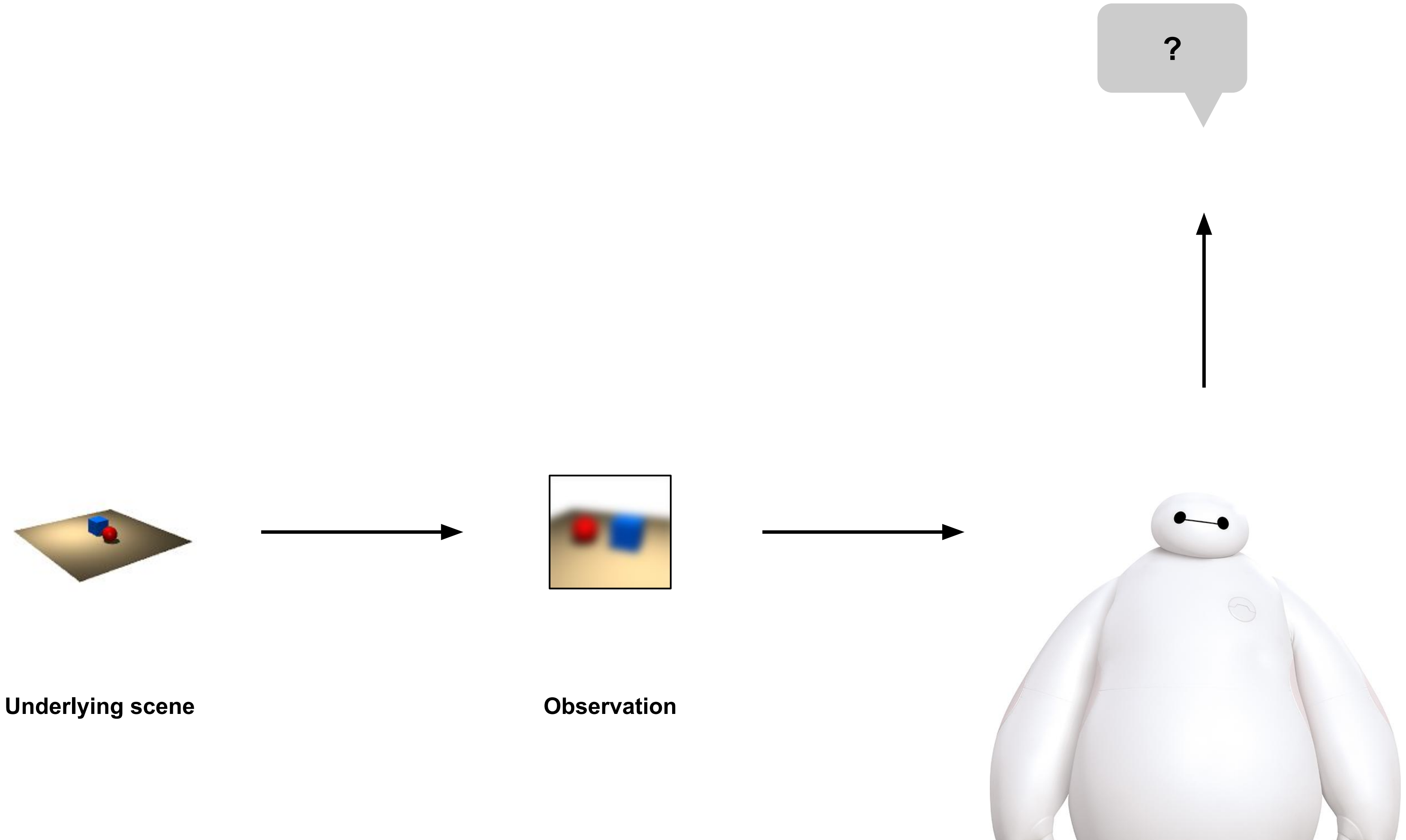
Rare

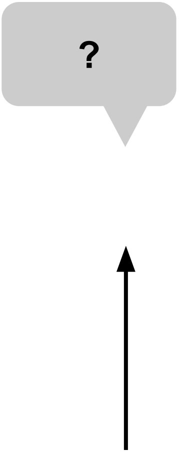
UNIVERSITY OF OXFORD

Microsoft Research

Google DeepMind

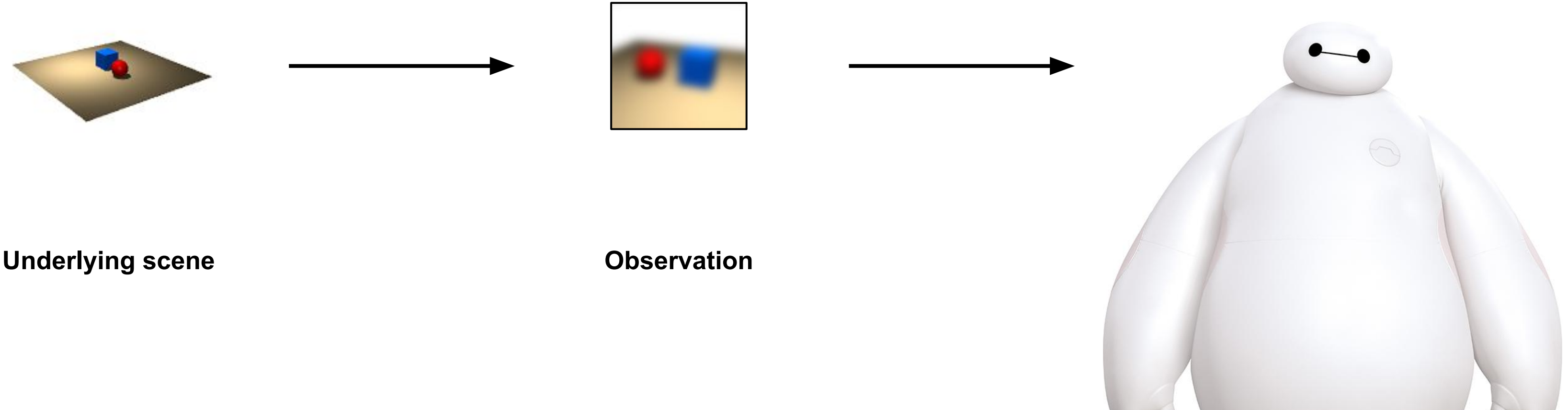

INSTITUTE
CAUTION


?
Underlying scene
Observation

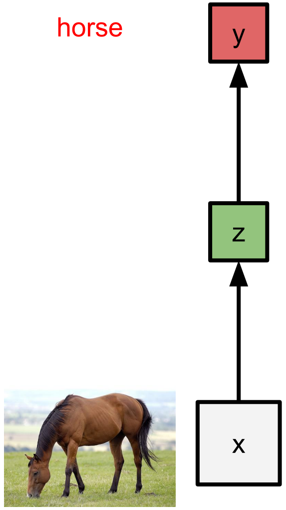


1. How should the scene be represented?

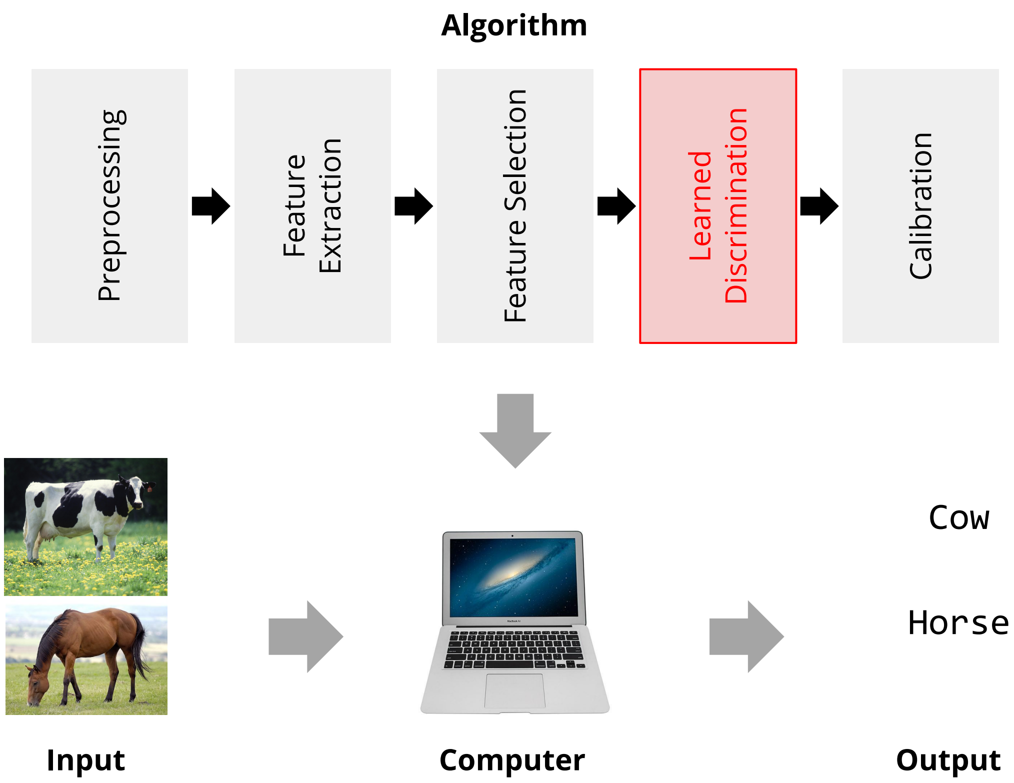
2. How should the representation be computed?

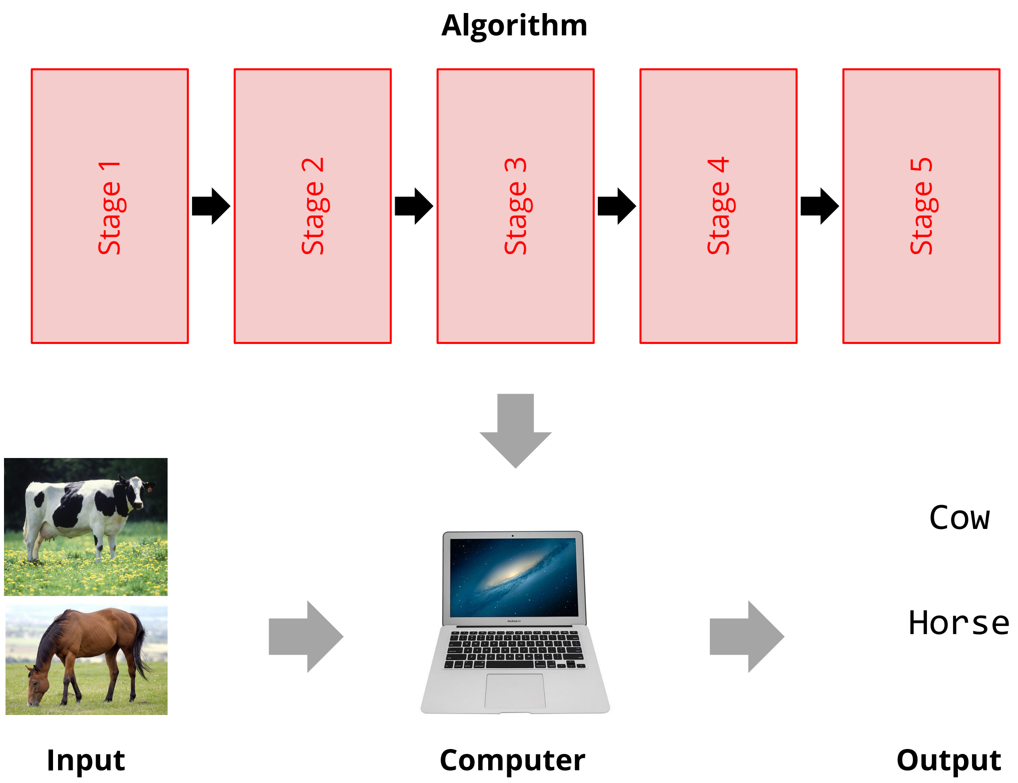


Learning paradigms



Supervised Learning


Algorithm
Preprocessing
Feature Extraction
Feature Selection
Learned Discrimination
Calibration
Cow
Horse
Input
Computer
Output


Algorithm
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5
Cow
Horse
Input
Computer
Output

Introduction

Deep Supervised Learning

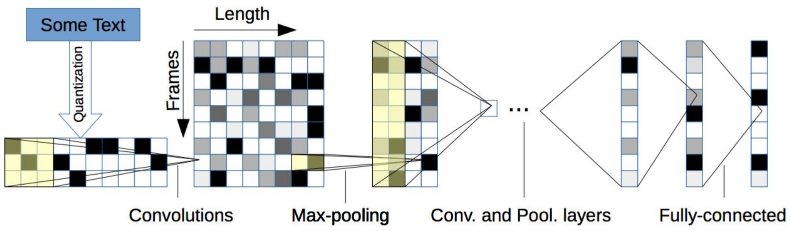
- Optimize directly for the end loss
- End-to-end training, no engineered inputs
- With enough data, learn a big non-linear function
- Supervised labeling is often enough for transferrable representations
- Large labeled dataset + big / deep neural network + GPUs

Clarifai (2014)

Introduction

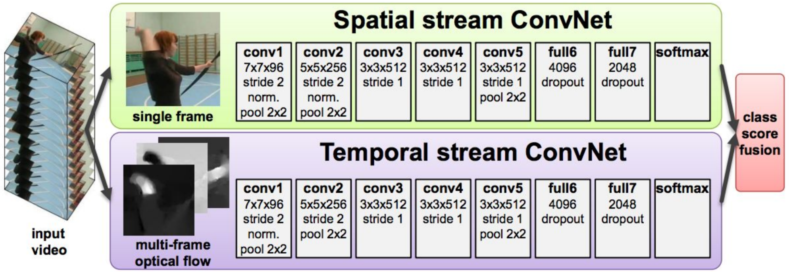
Deep Supervised Learning

Text Classification



Zhang et al. (2015)

Video Classification



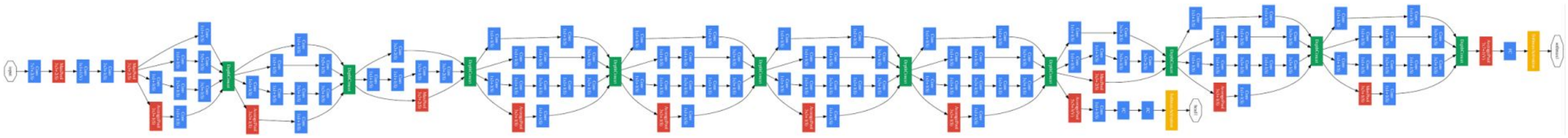
Simonyan et al. (2014)

Introduction

Deep Supervised Learning

- Innovation continues
  - Inception (Szegedy et al., 2015)
  - Residual connections (He et al., 2015)
  - Batchnorm (Ioffe et al., 2015)

- Performance is continuously improving

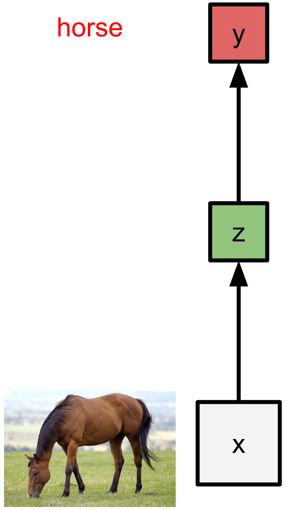


Szegedy et al., (2015)

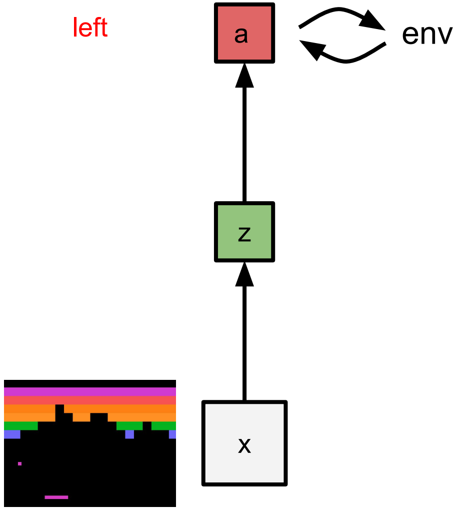
Where does the data come from?

What is the correct representation?

Learning paradigms



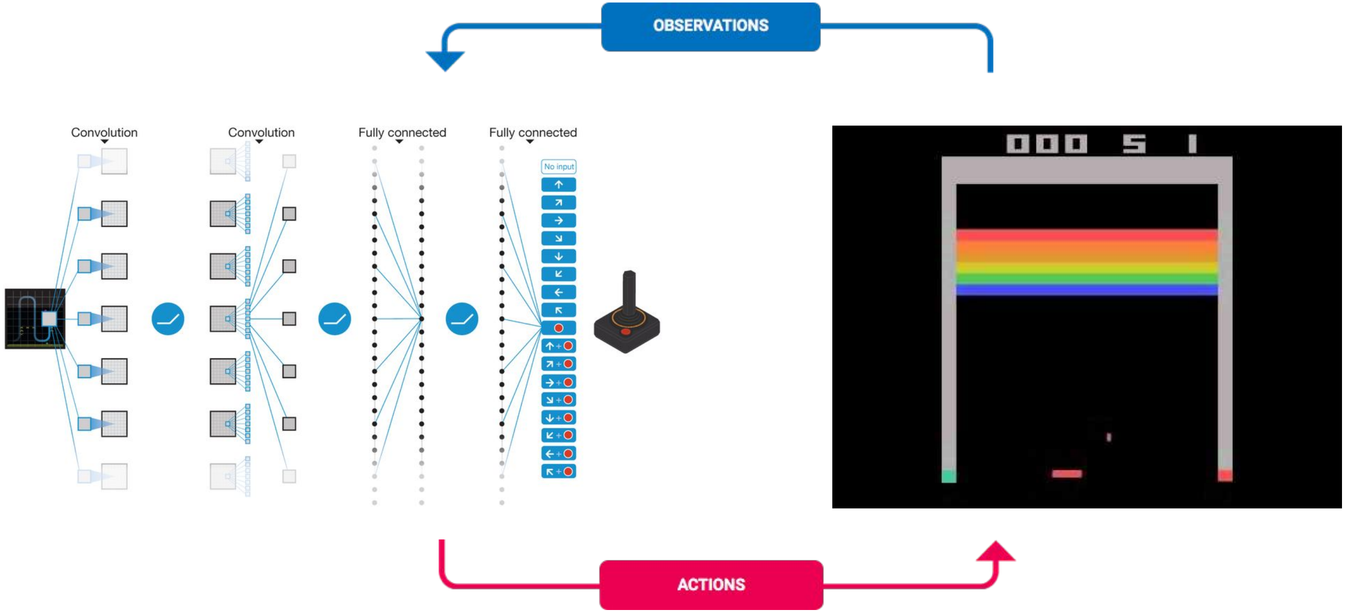
Supervised Learning



Reinforcement Learning

Human-level control in ATARI

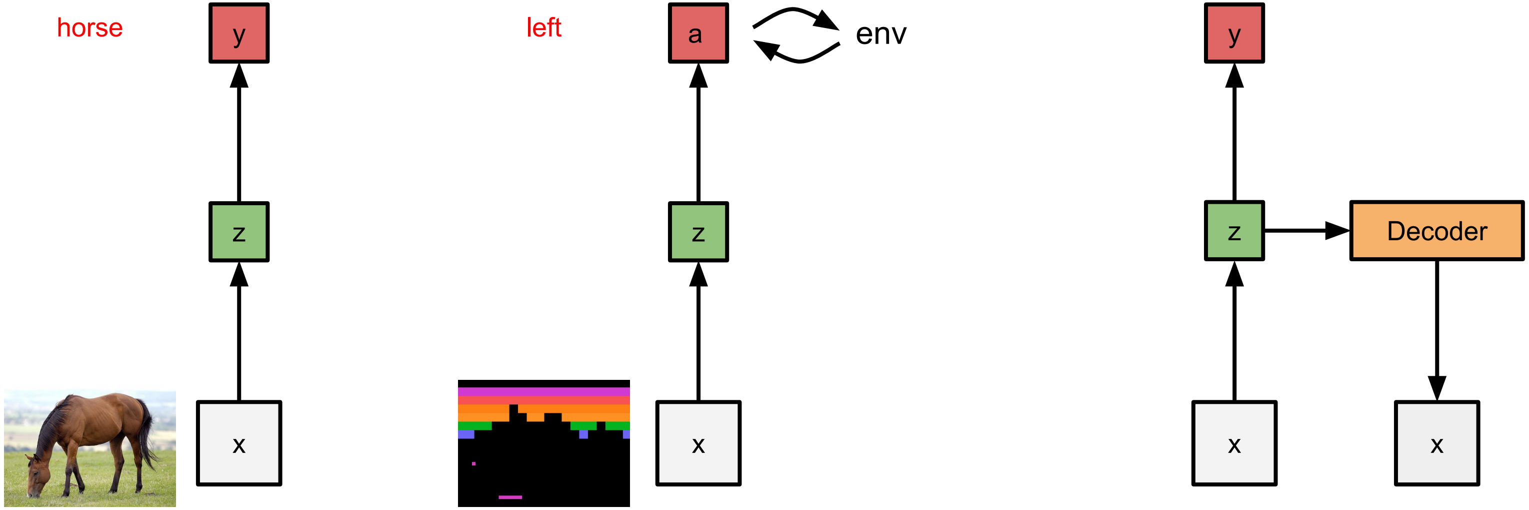
End-to-end reinforcement learning



Mnih et al. (2015)

How much experience do we really need?

Learning paradigms

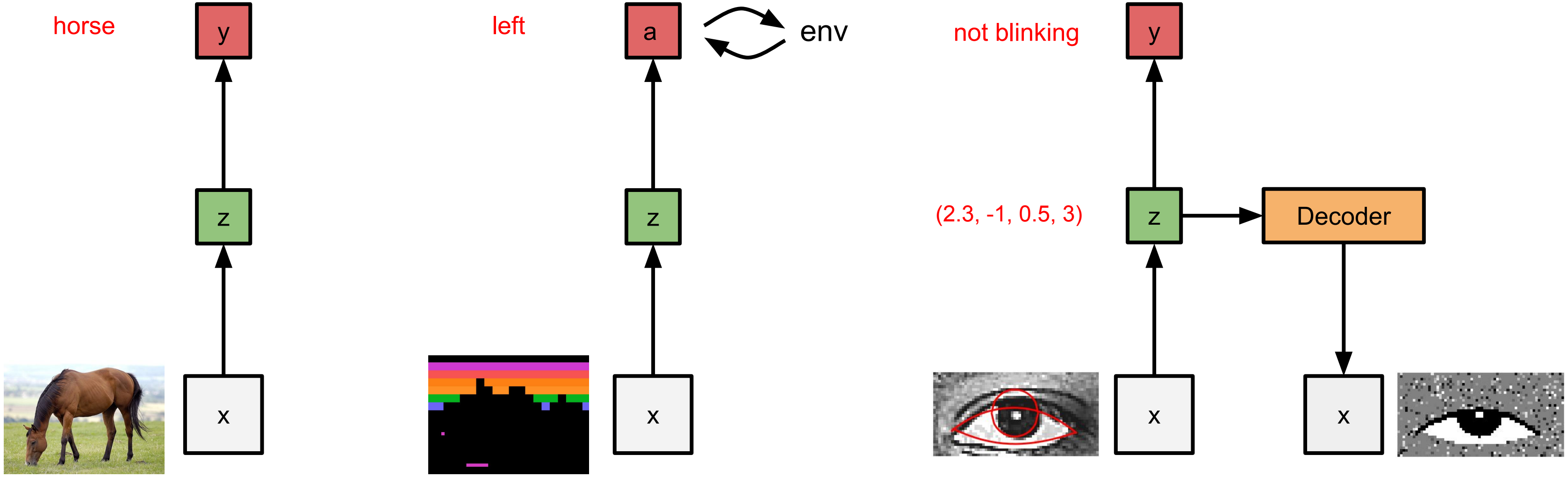


Supervised Learning

Reinforcement Learning

Generative Modelling

Learning paradigms

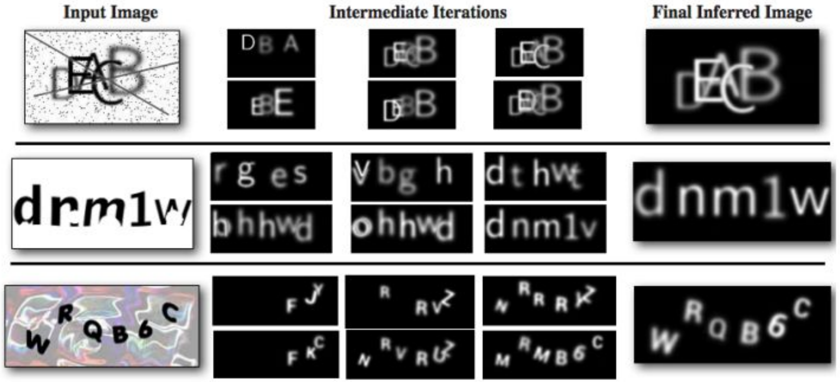


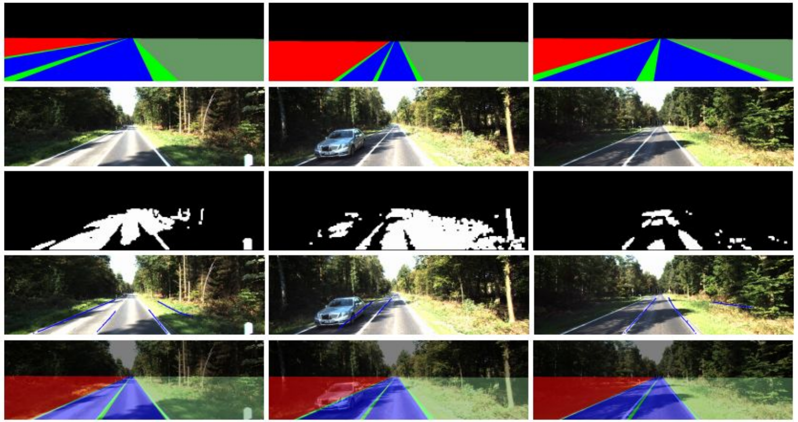
Supervised Learning

Reinforcement Learning

Generative Modelling

Highly structured





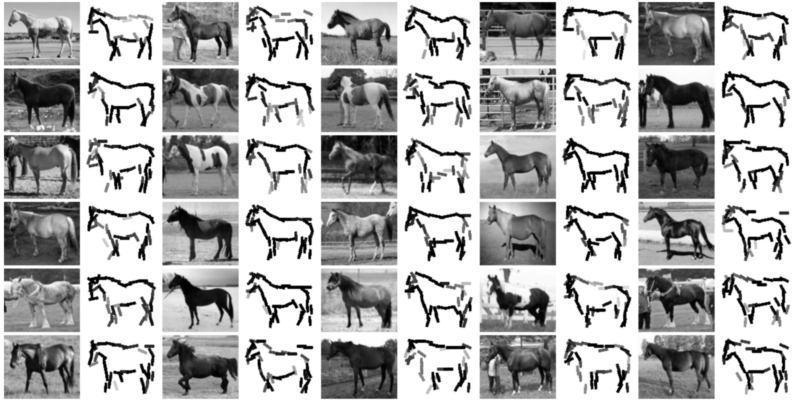
General Purpose Graphics Programming

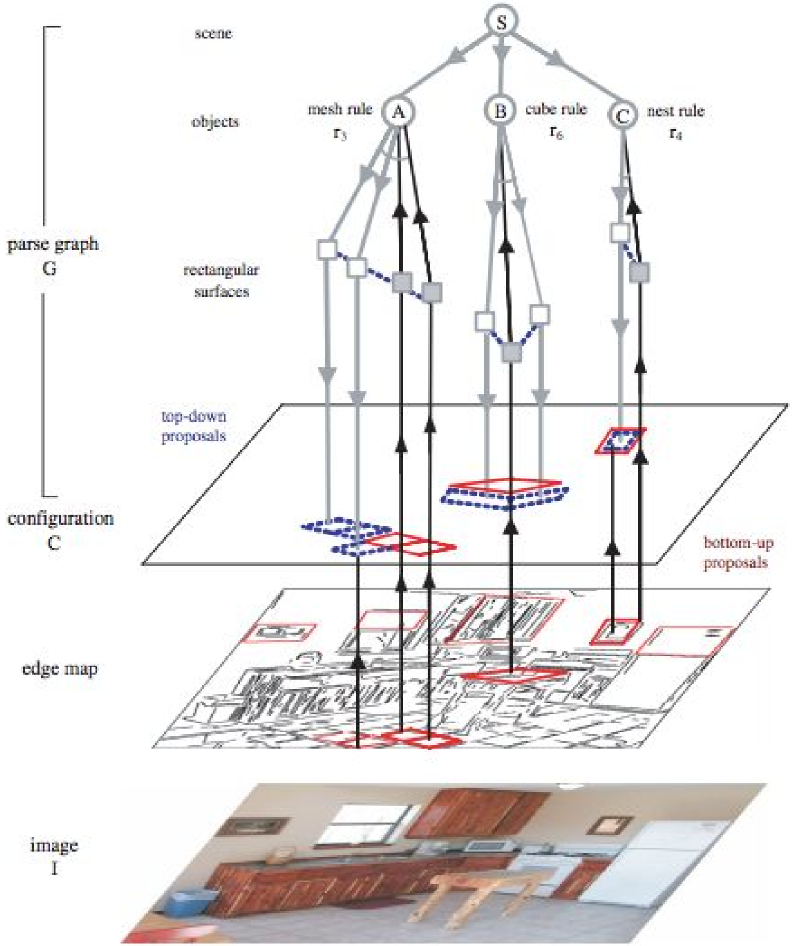
Vikash Mansinghka, Tejas D. Kulkarni, Yura N. Perov, and Joshua B. Tenenbaum (2013)

Partially structured

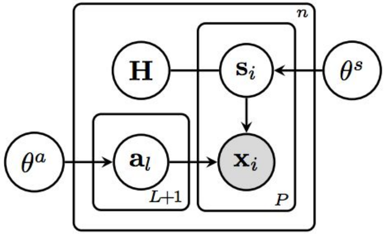
A Stochastic Grammar of Images

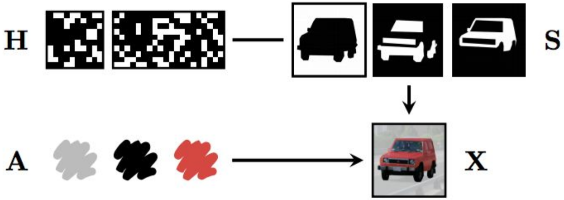
Song-Chun Zhu and David Mumford (2007)





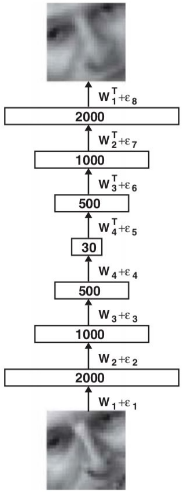
Partially structured

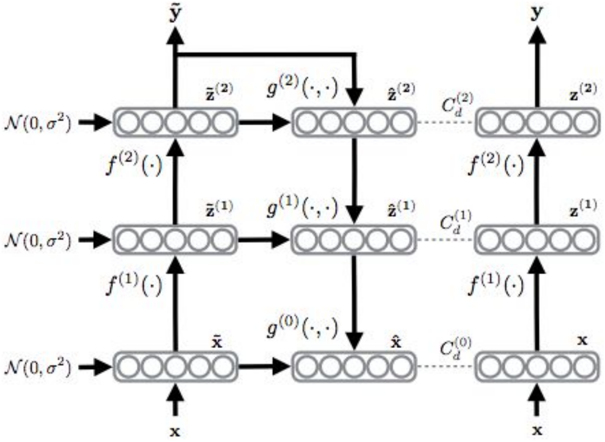


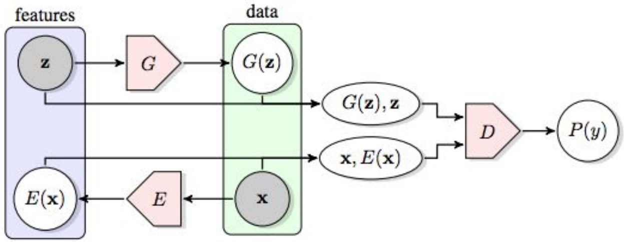


S. M. Ali Eslami and Christopher K. I. Williams (2012)

Fully unstructured







Geoffrey Hinton (2006)

Antti Rasmus et al. (2016)

Jeff Donahue et al. (2016)

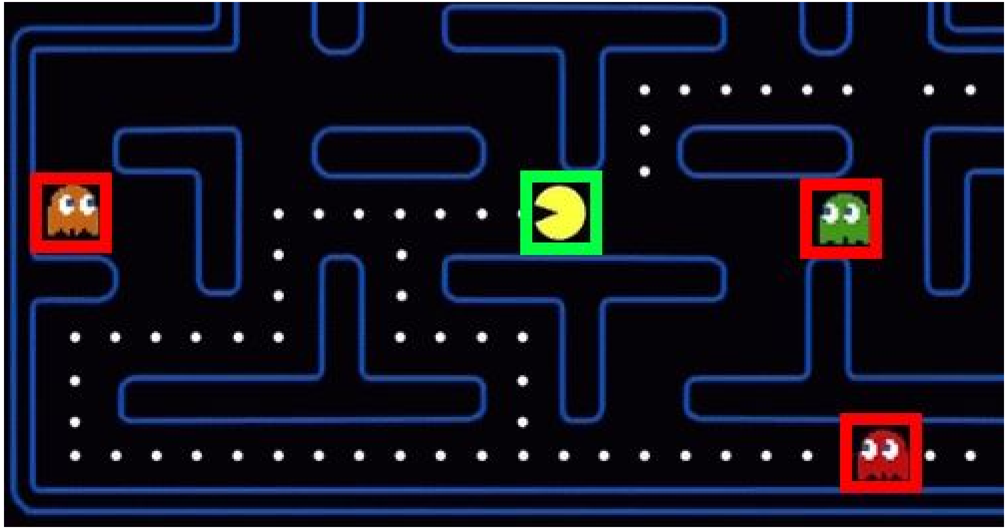
Attend, Infer, Repeat: Fast Scene Understanding with Generative Models

S. M. Ali Eslami, Nicolas Heess, Theophane Weber, Yuval Tassa, Koray Kavukcuoglu, Geoffrey Hinton
Neural Information Processing Systems (NIPS), 2016

Motivation

To obtain **object-based** representations

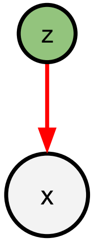
To learn from orders-of-magnitude **less data**



Cause

blue brick

Model



Image

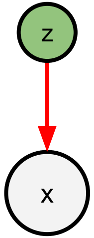
Cause

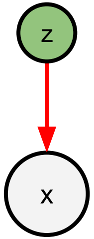
blue brick

pile of bricks

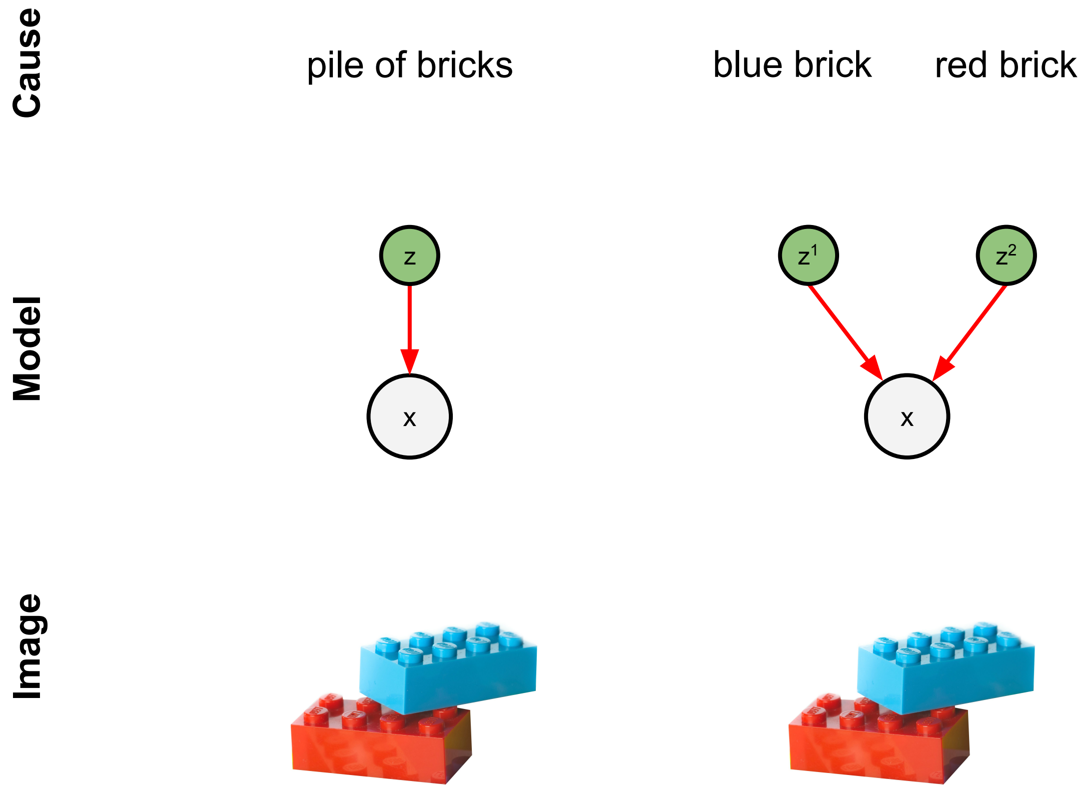
not sufficient for
grasping
counting
transfer
generalisation

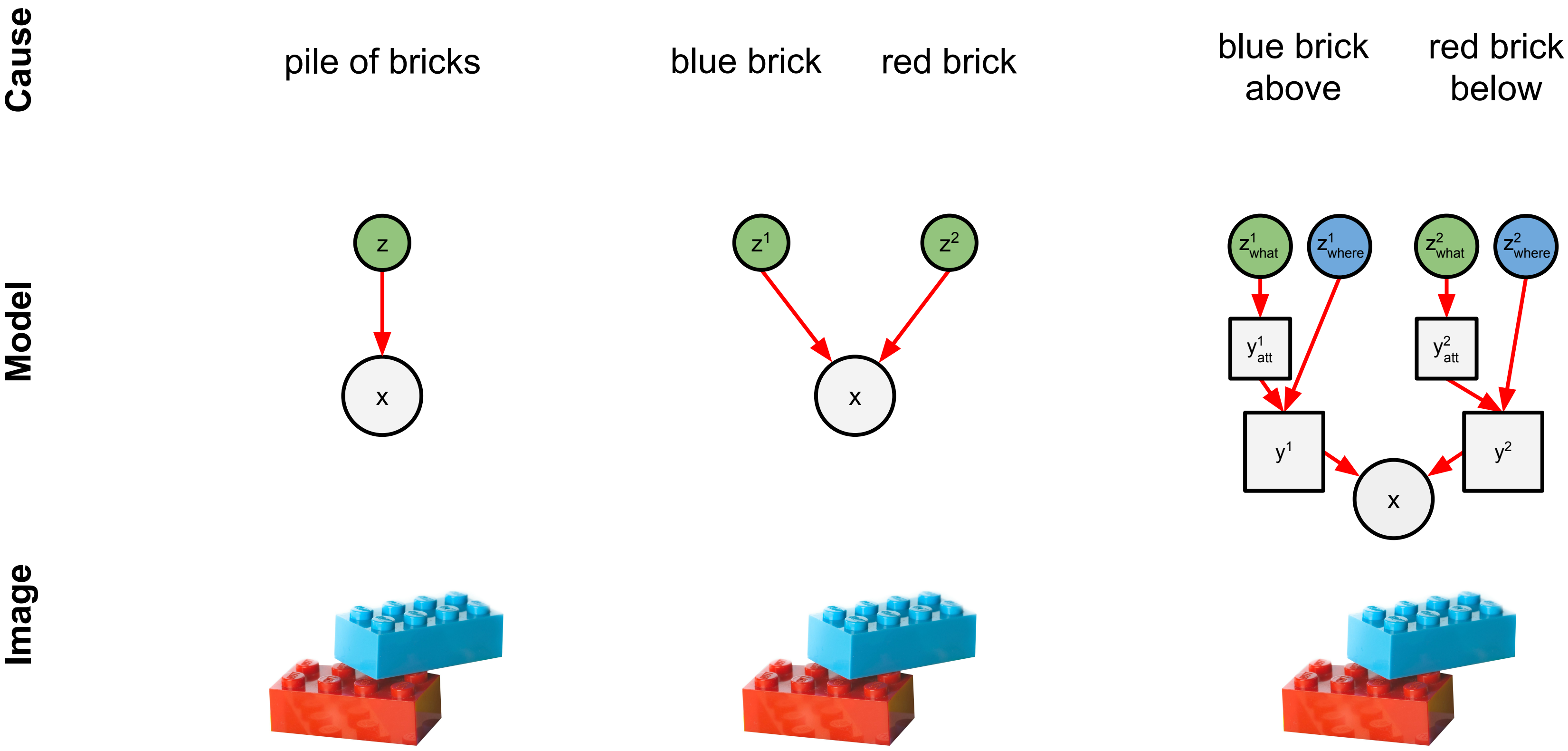
Model



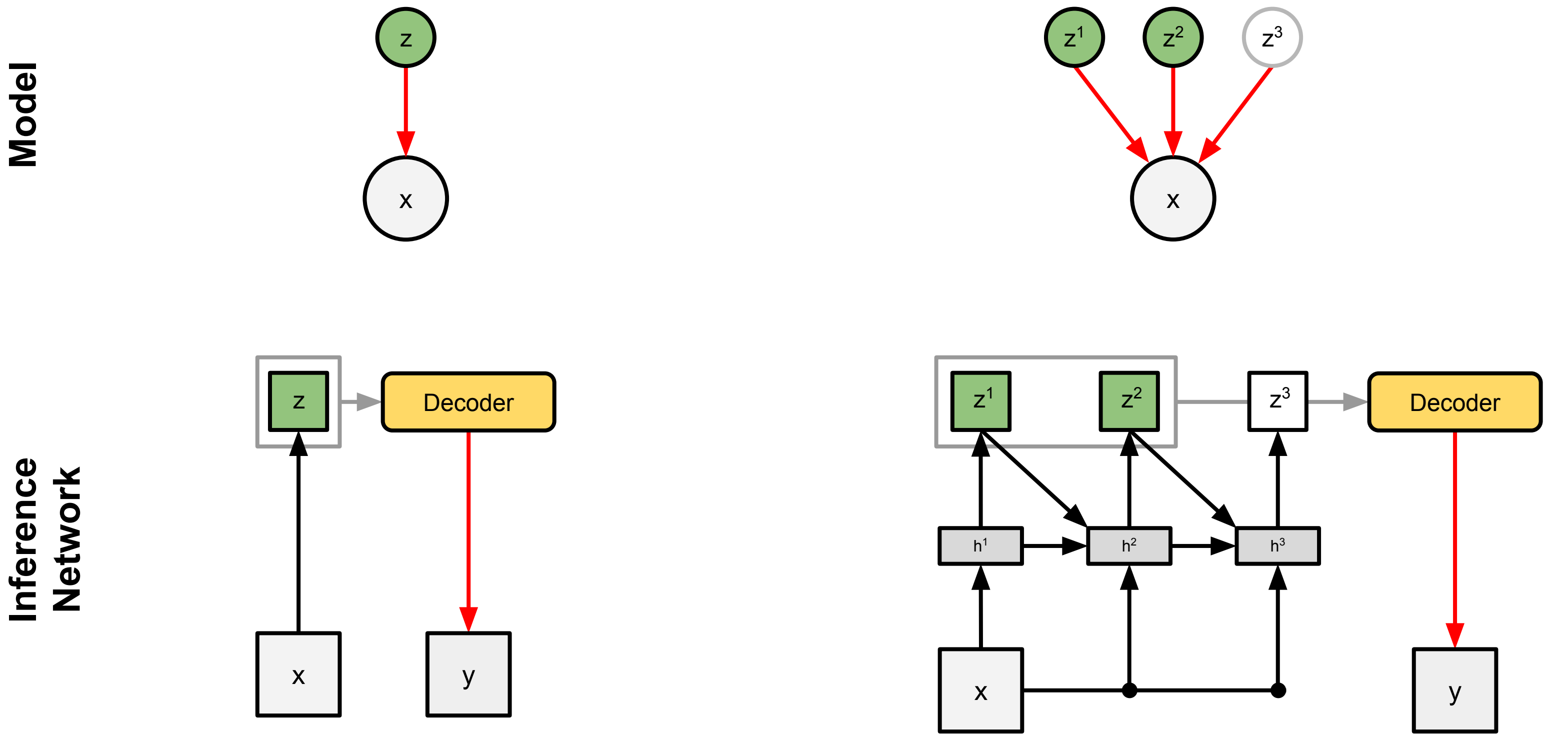


Image

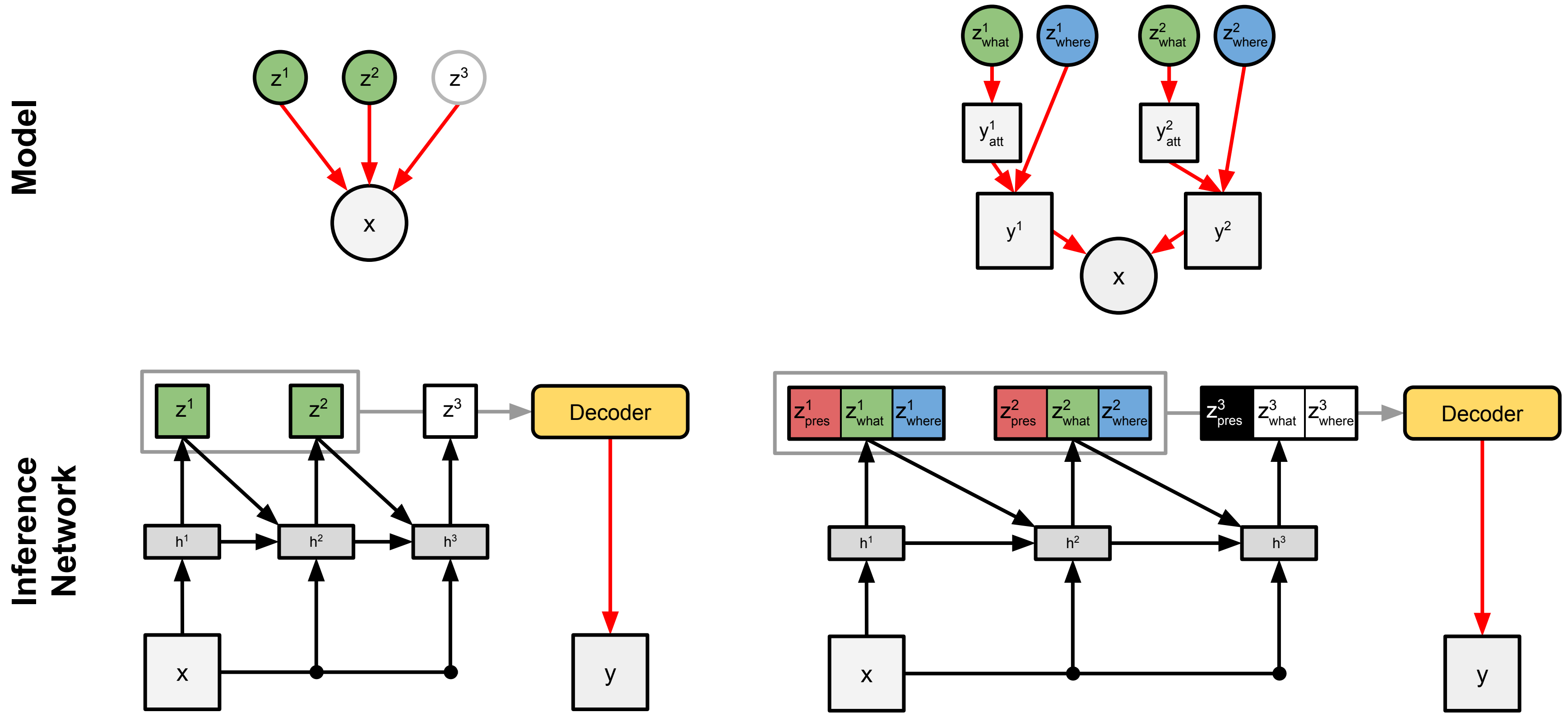

Cause
Model
Image
pile of bricks
blue brick
red brick
z
x
z1
z2
x



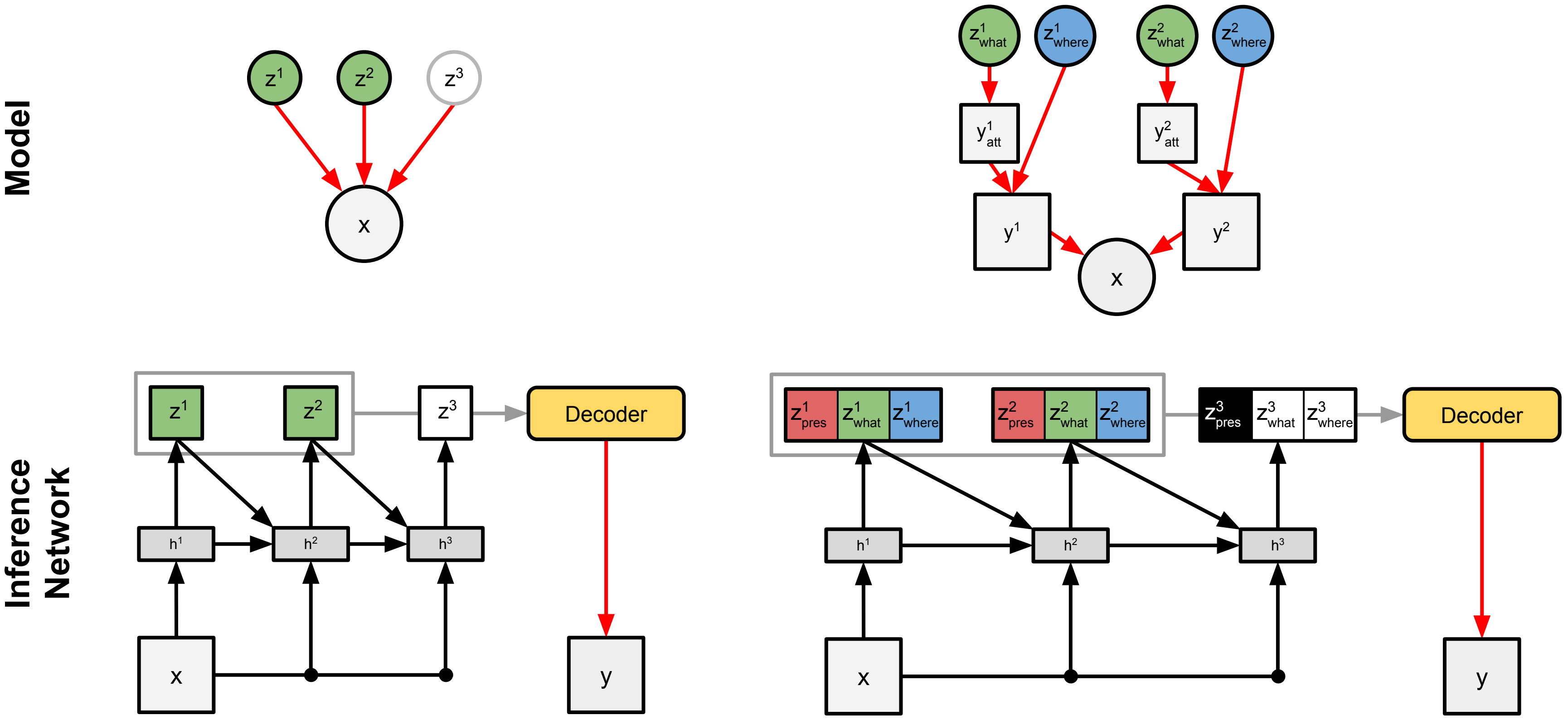
Cause
Model
Image
pile of bricks
blue brick
red brick
blue brick above
red brick below
z
x
z^1
z^2
x
z^1_{what}
z^1_{where}
z^2_{what}
z^2_{where}
y^1_{att}
y^2_{att}
y^1
y^2
x



Model
Inference Network
z
x
z1
z2
z3
x
z
Decoder
x
y
z1
z2
z3
Decoder
h1
h2
h3
x
y

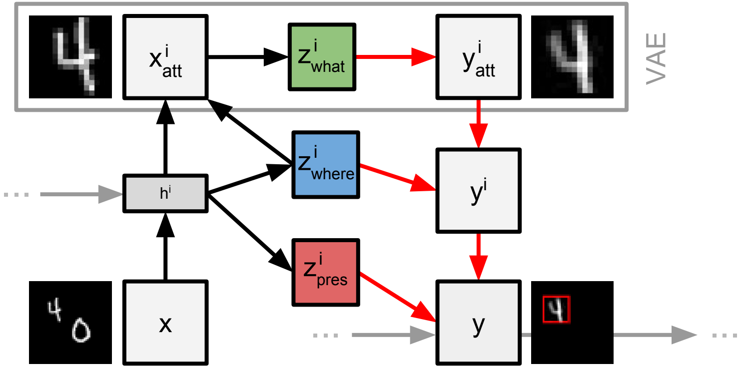


Model
Inference Network
z1
z2
z3
x
z1what
z1where
z2what
z2where
y1att
y2att
y1
y2
h1
h2
h3
Decoder
y
z1pres
z2pres
z3pres
z3what
z3where



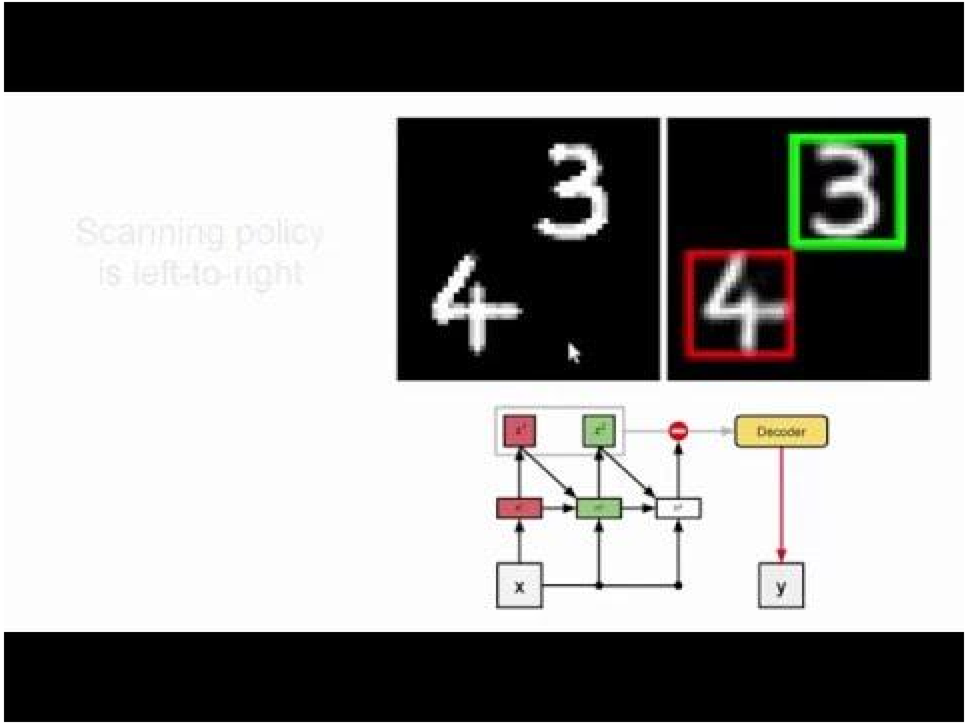
focus on representation
not reconstruction

output is a set
order? count?

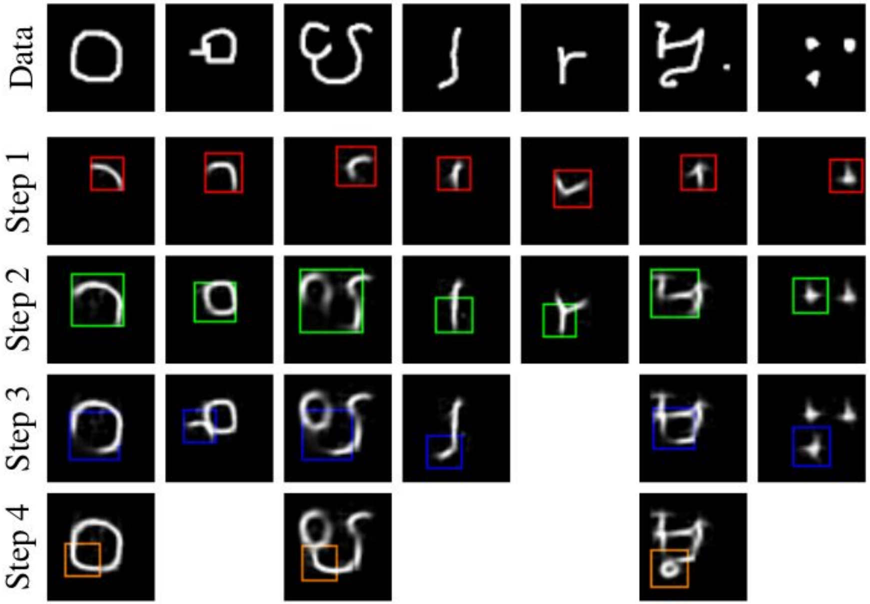


x^i_att
z^i_what
y^i_att
VAE
z^i_where
h^i
y^i
z^i_pres
x
y

Demo reel

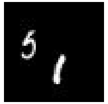


Omniglot

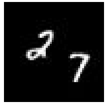


Representational power

Sum?

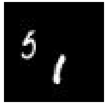


6

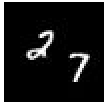


9

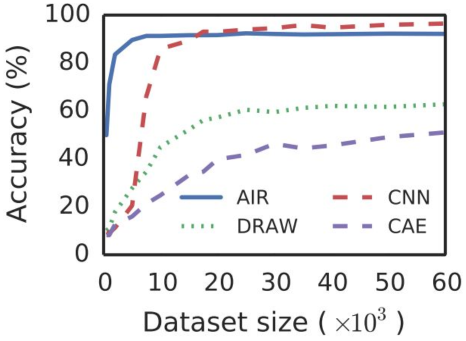
Increasing order?

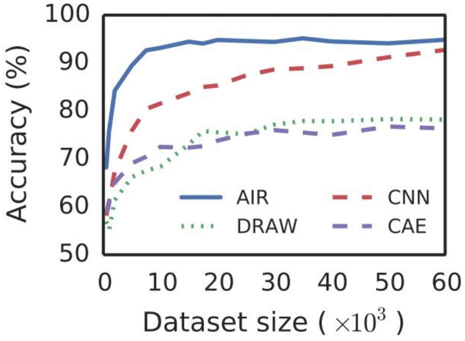


no



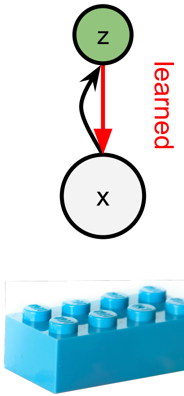
yes





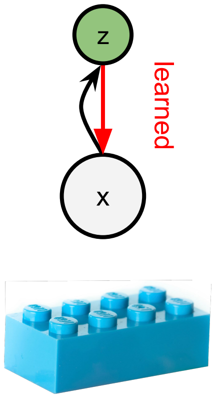
Additional structure

distributed **vector**
that correlates
with blue brick

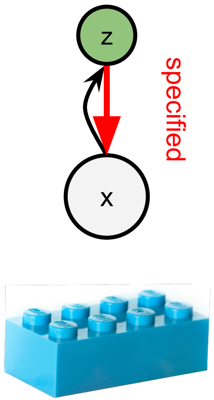


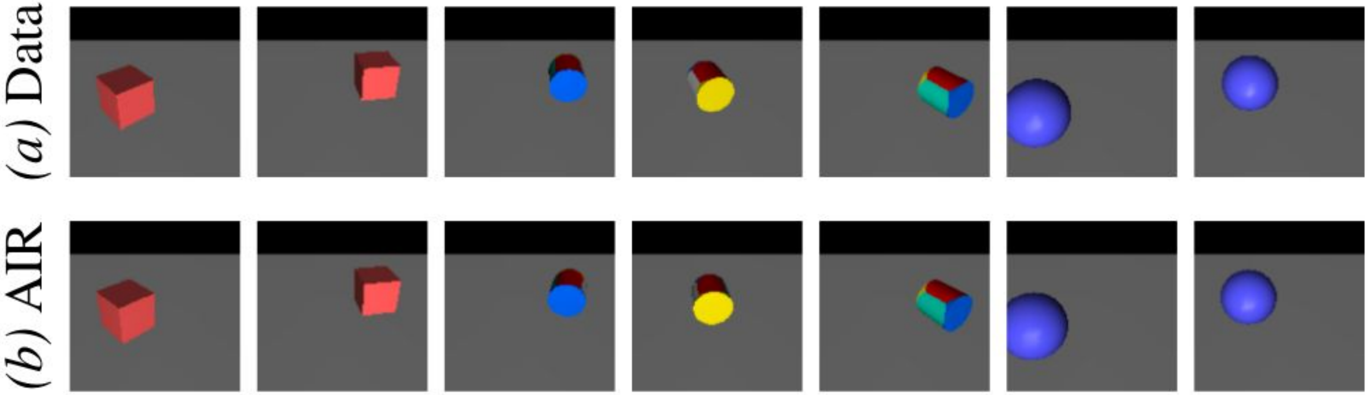
Additional structure

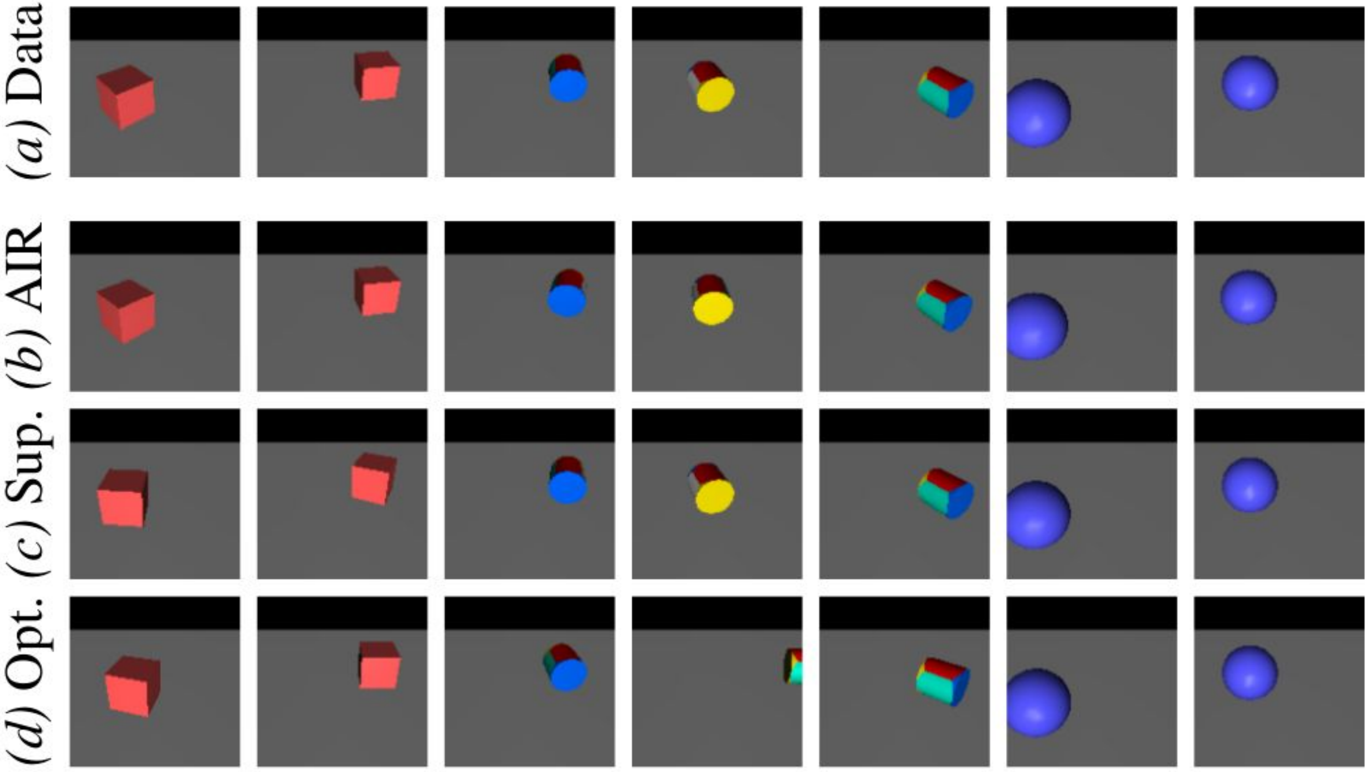
distributed **vector**
that correlates
with blue brick



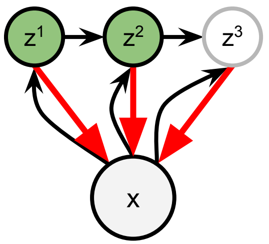
class=**brick**
colour=**blue**
position=**P**
rotation=**R**

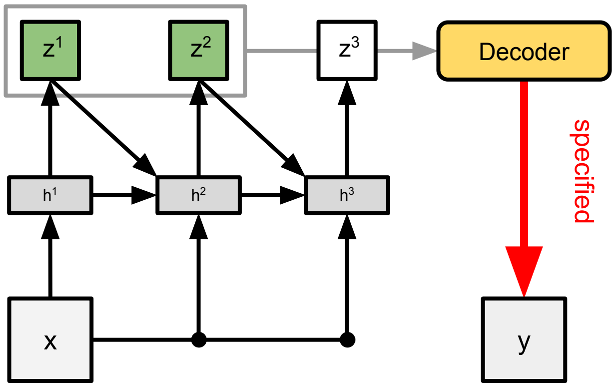



(a) Data
(b) AIR

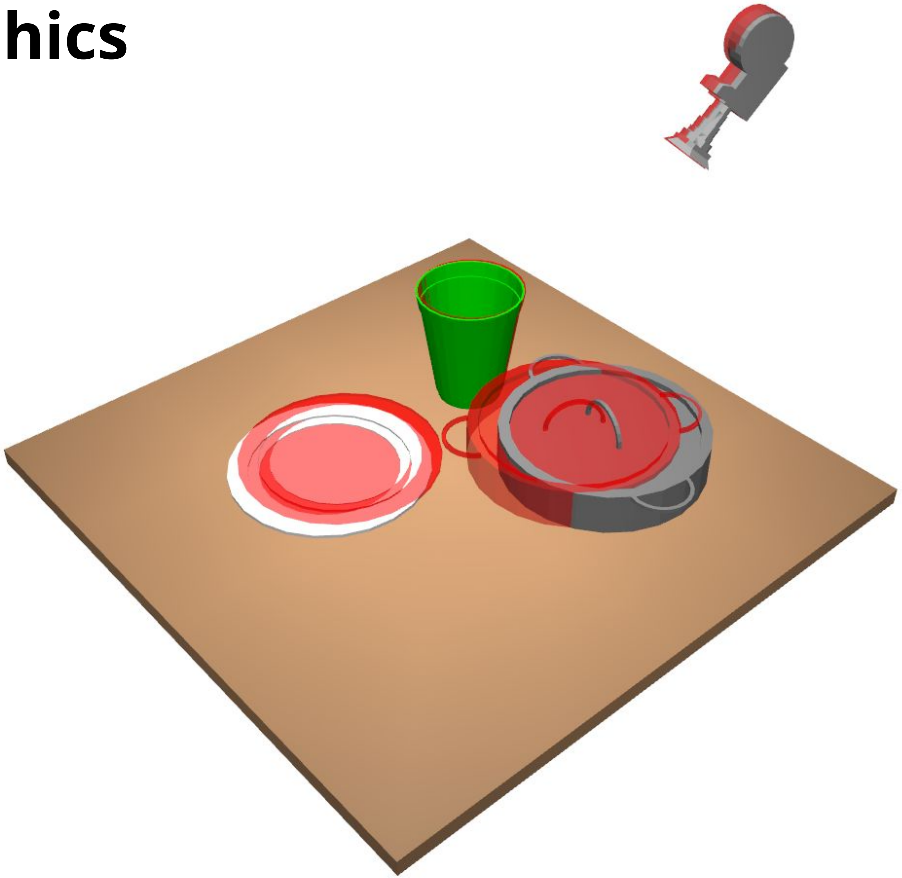

(a) Data
(b) AIR
(c) Sup.
(d) Opt.

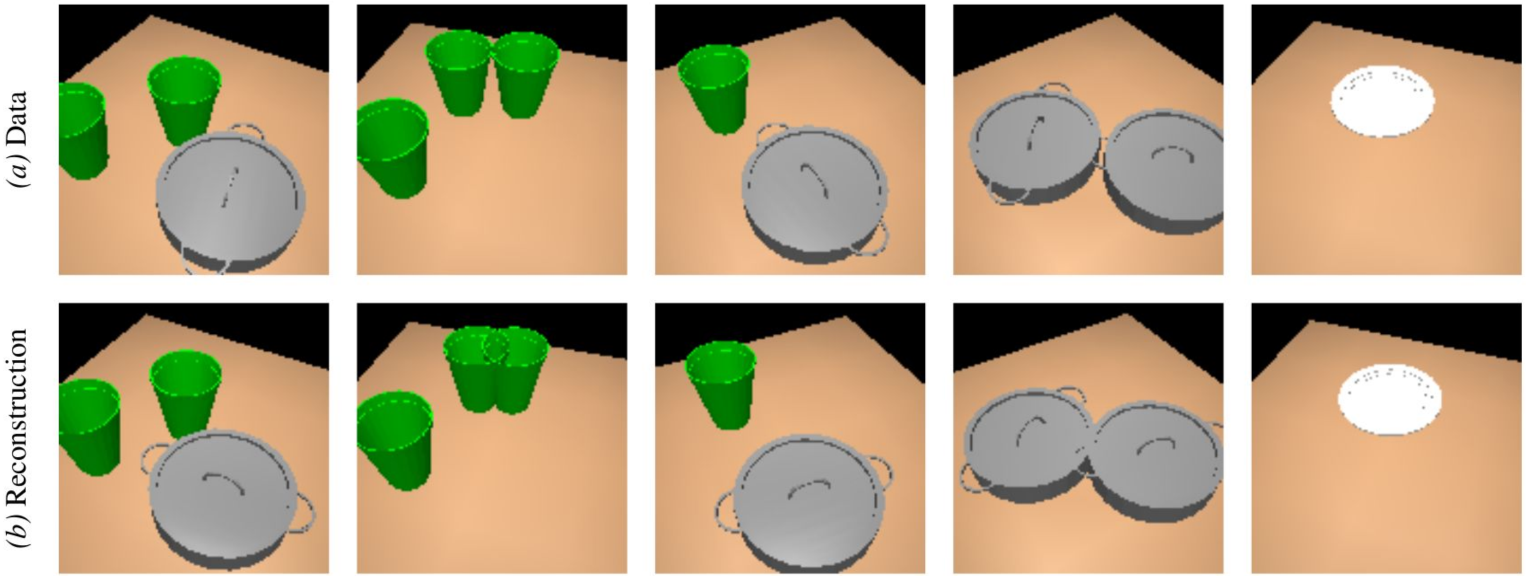
Additional structure





Inverse graphics




(a) Data
(b) Reconstruction

Policy learning

MNIST

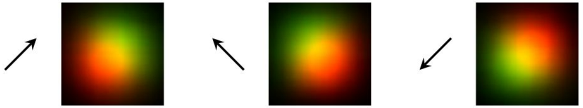
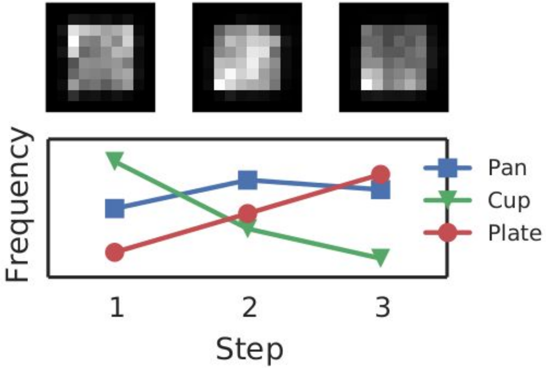


Table-top



Unsupervised Learning of 3D Structure from Images

Danilo Rezende, S. M. Ali Eslami, Shakir Mohamed, Peter Battaglia, Max Jaderberg, Nicolas Heess
Neural Information Processing Systems (NIPS), 2016

Motivation

To recover **3D structure** from **2D images**

To form **stable** representations, regardless of camera position

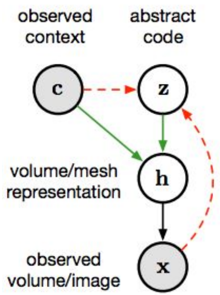
Motivation

To recover **3D structure** from **2D images**

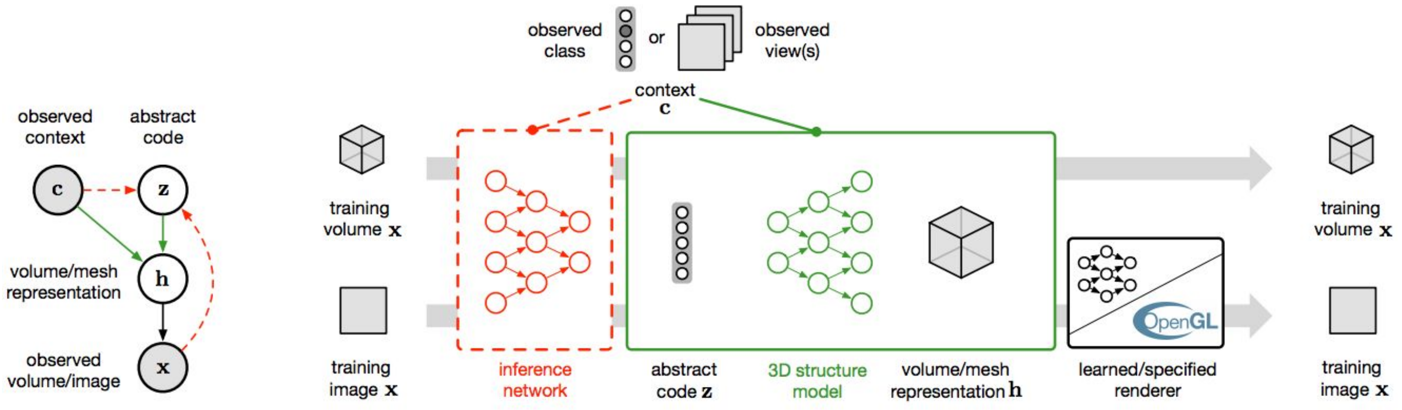
To form **stable** representations, regardless of camera position

- Inherently ill-posed
  - All objects appear under self occlusion, infinite explanations
  - Therefore build statistical models to know what's likely and what's not
- Even with models, inference is intractable
  - Important to capture multi-modal explanations
- How are 3D scenes best represented?
  - Meshes or voxels?
- Where is training data collected from?

Unsupervised Learning of 3D Structure from Images

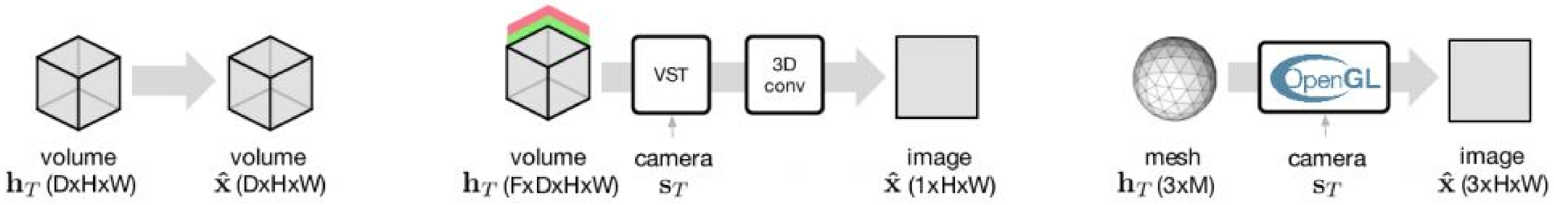


Unsupervised Learning of 3D Structure from Images



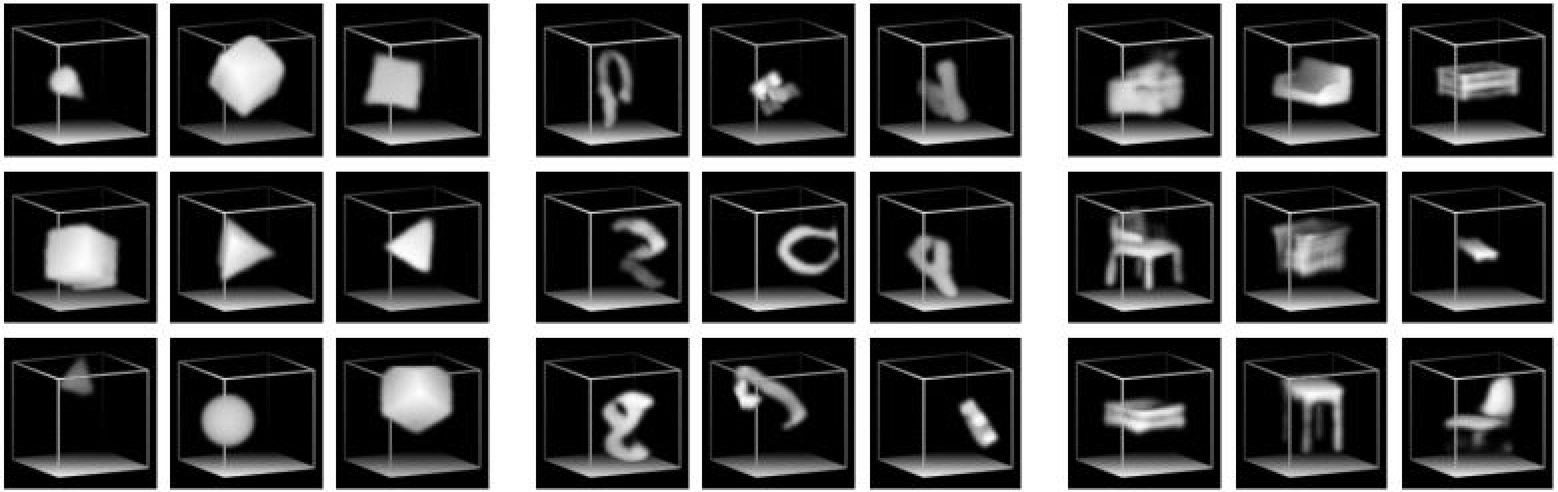
Unsupervised Learning of 3D Structure from Images

Projection operators

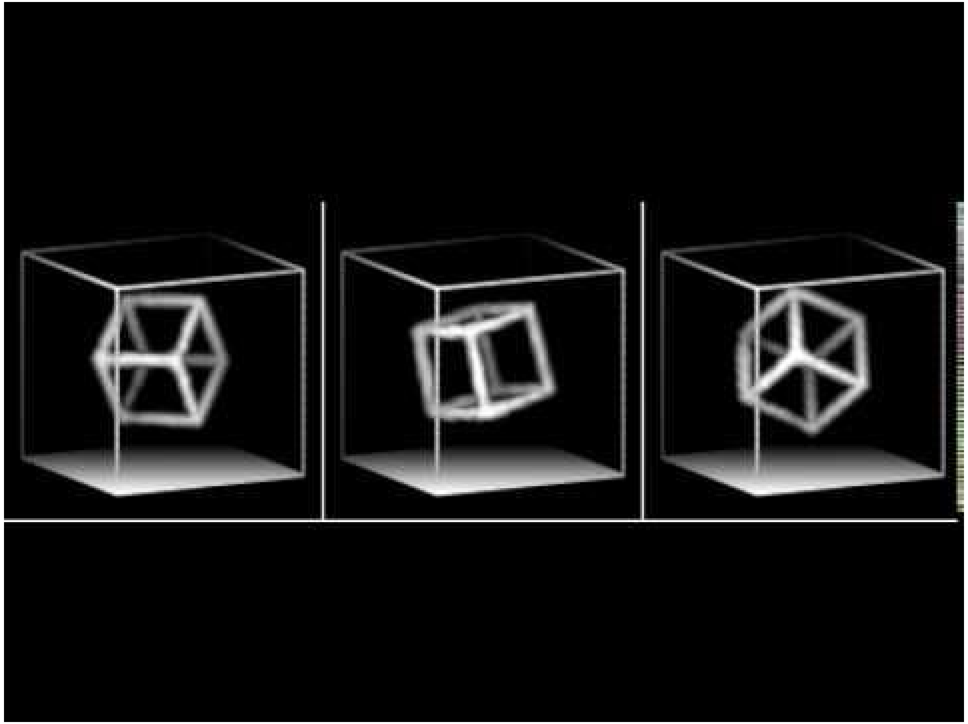


Unsupervised Learning of 3D Structure from Images

Unconditional samples

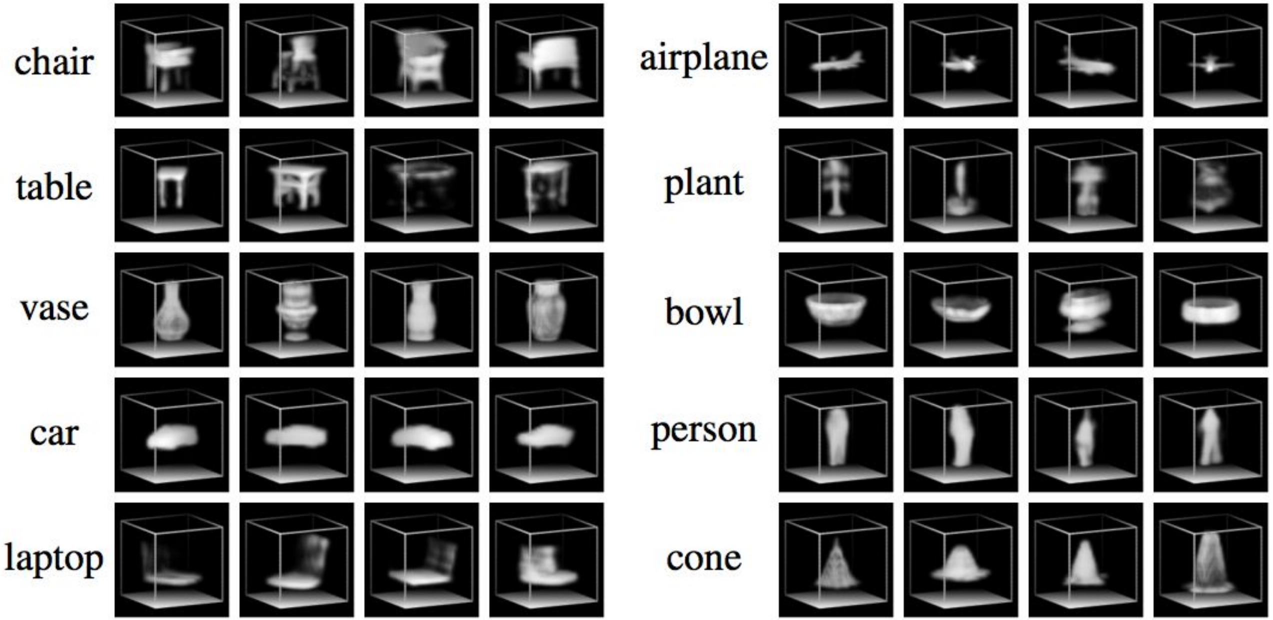


Class-conditional samples



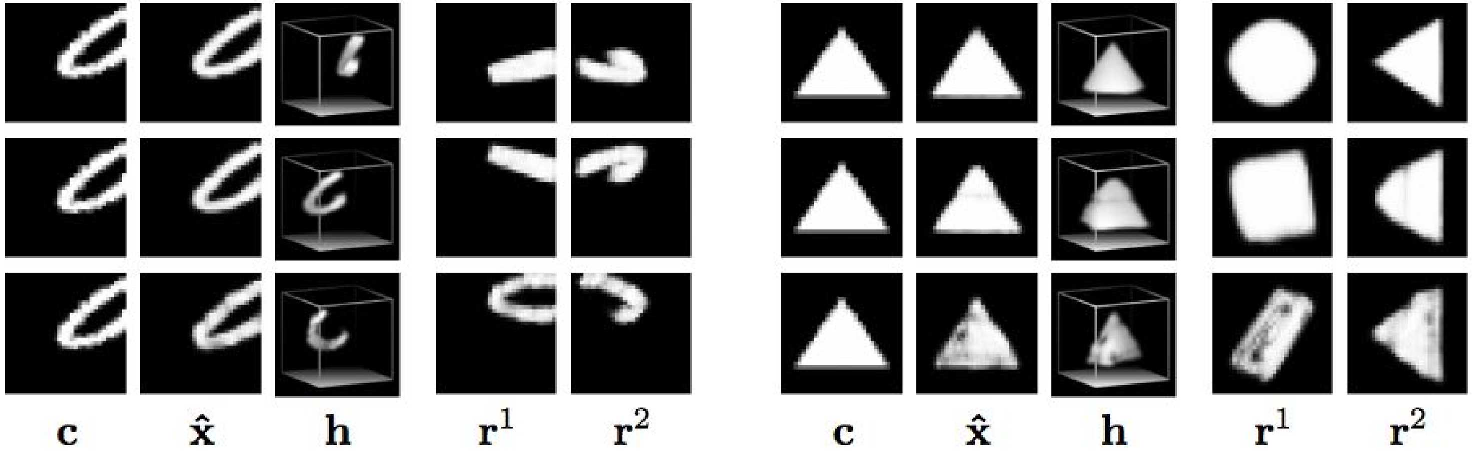
Unsupervised Learning of 3D Structure from Images

Class-conditional samples

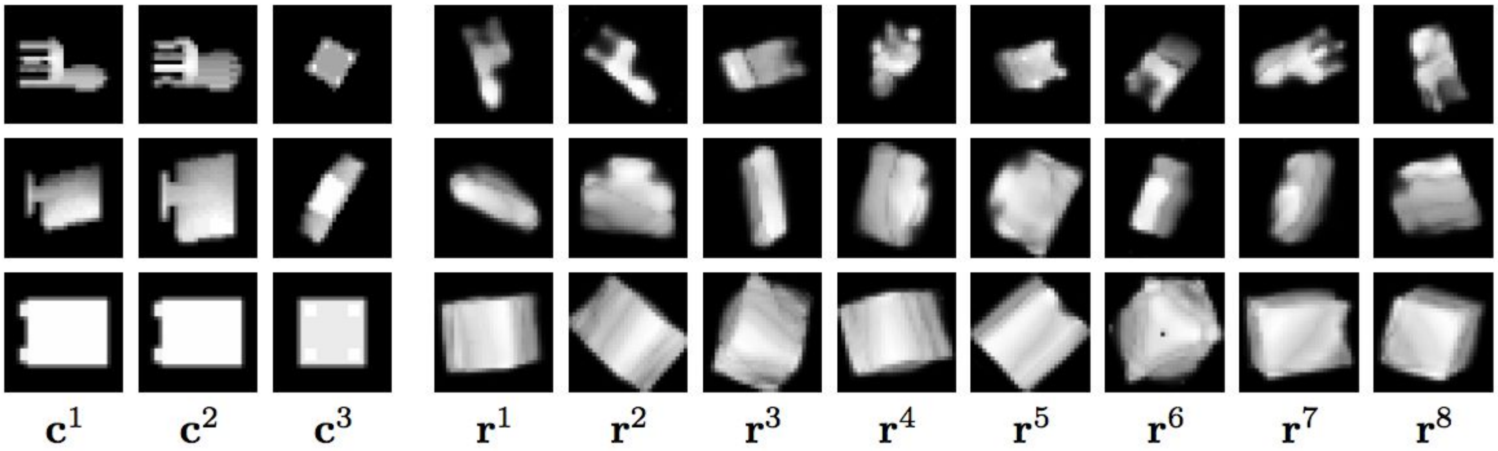


Unsupervised Learning of 3D Structure from Images

Multi-modality of inference

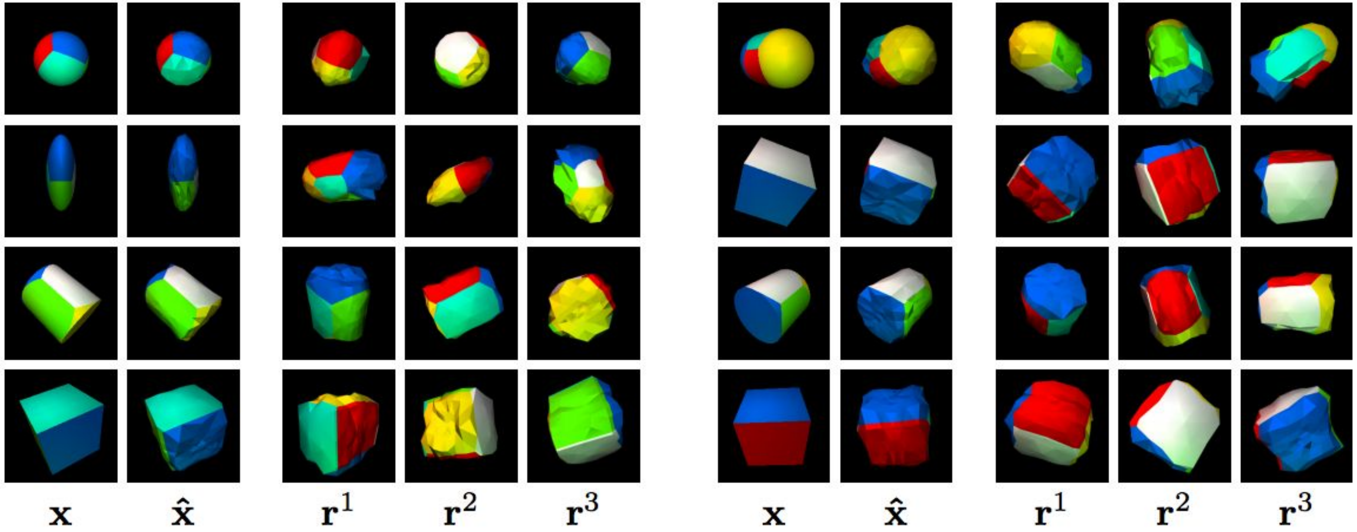


3D structure from multiple 2D images



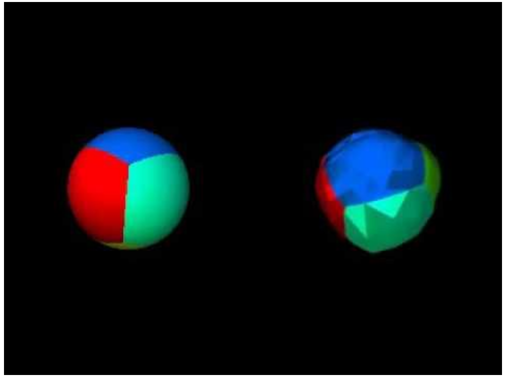
Unsupervised Learning of 3D Structure from Images

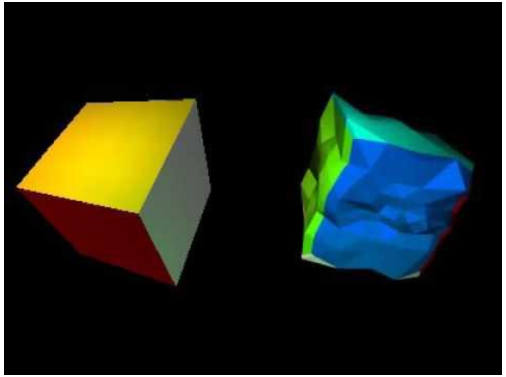
Inferring object meshes

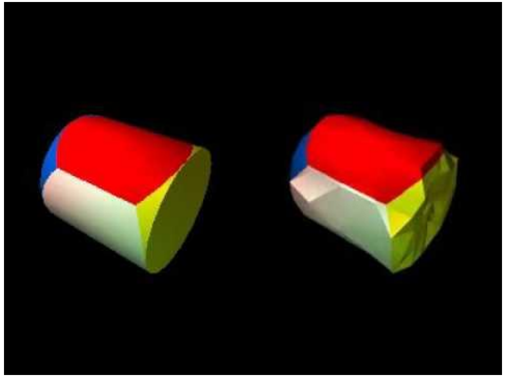


Unsupervised Learning of 3D Structure from Images

Inferring object meshes







Recap

- Deep Supervised Learning
- Deep Reinforcement Learning
- Model-based Methods
- Structured / Unstructured Generative Models

aeslami@google.com
arkitus.com