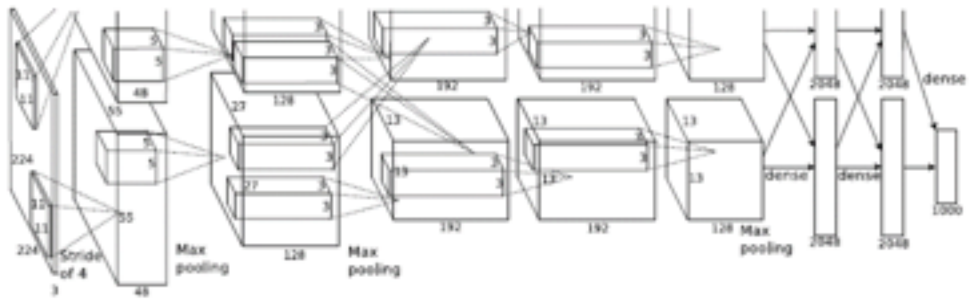
Analysis-by-Synthesis a.k.a Generative Modeling

Tejas D Kulkarni (tejask@mit.edu)

Traditional paradigm for AI research

- Traditional machine learning and pattern recognition has been remarkable successful at questions like: "*what is where*"



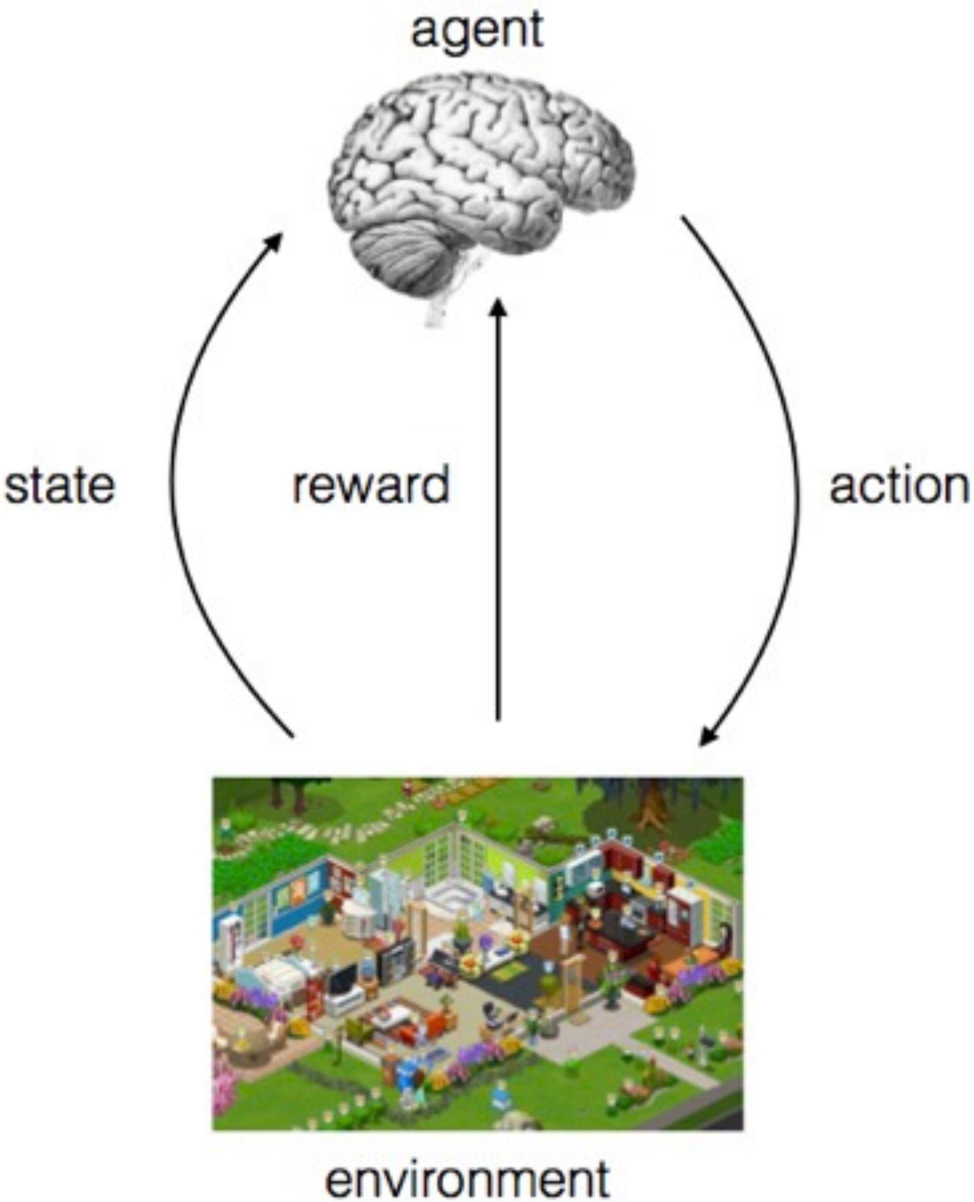
Krizhevsky et. al

- Decades of progress driven by a single experimental paradigm!
  - Train/Test/Validation set split
  - Learn parameters using train set and evaluate on the test set

But infant learning looks like this ...

source: https://www.youtube.com/watch?v=3f3rOz0NzPc

What is the ‘right’ way to think about AI?



- Agent learns an internal model of the world (generative model) given sensory states
- Agent uses learnt model to mentally hypothesize plans and pick actions that maximizes expected future rewards
- While exploring, agents pick actions that minimizes prediction error of it’s internal model (i.e. minimizes entropy)

Analysis-by-Synthesis in Perception

Hermann Von Helmholtz

The general rule determining the ideas of vision that are formed whenever an impression is made on the eye, is that such objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism

(1865)

Geoff
Hinton et al

Boltzmann Machines
Helmholtz Machine

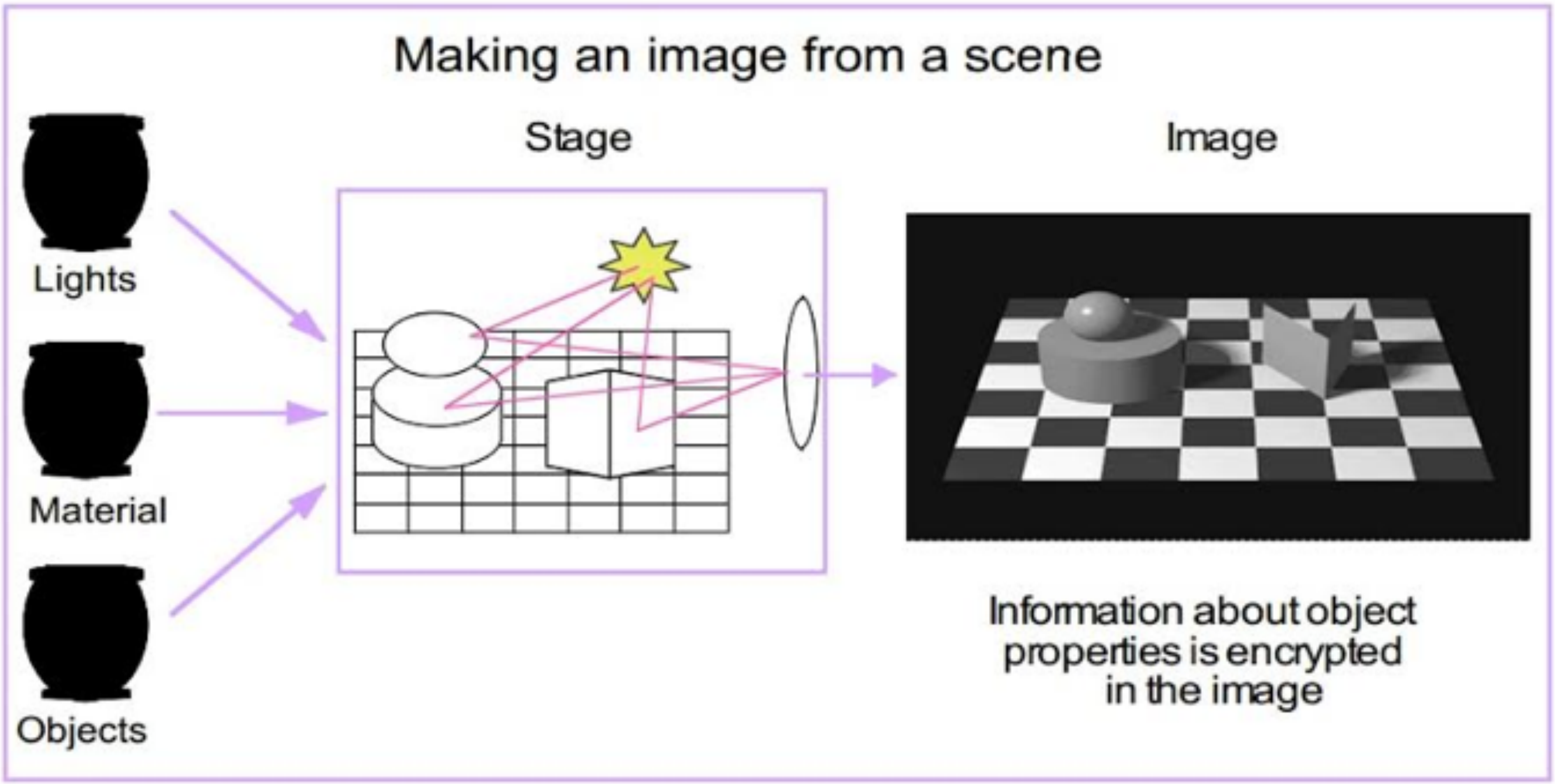
(1885, 1995)

Karl Friston

The free-energy principle says that any self-organizing system that is at equilibrium with its environment must minimize its free energy

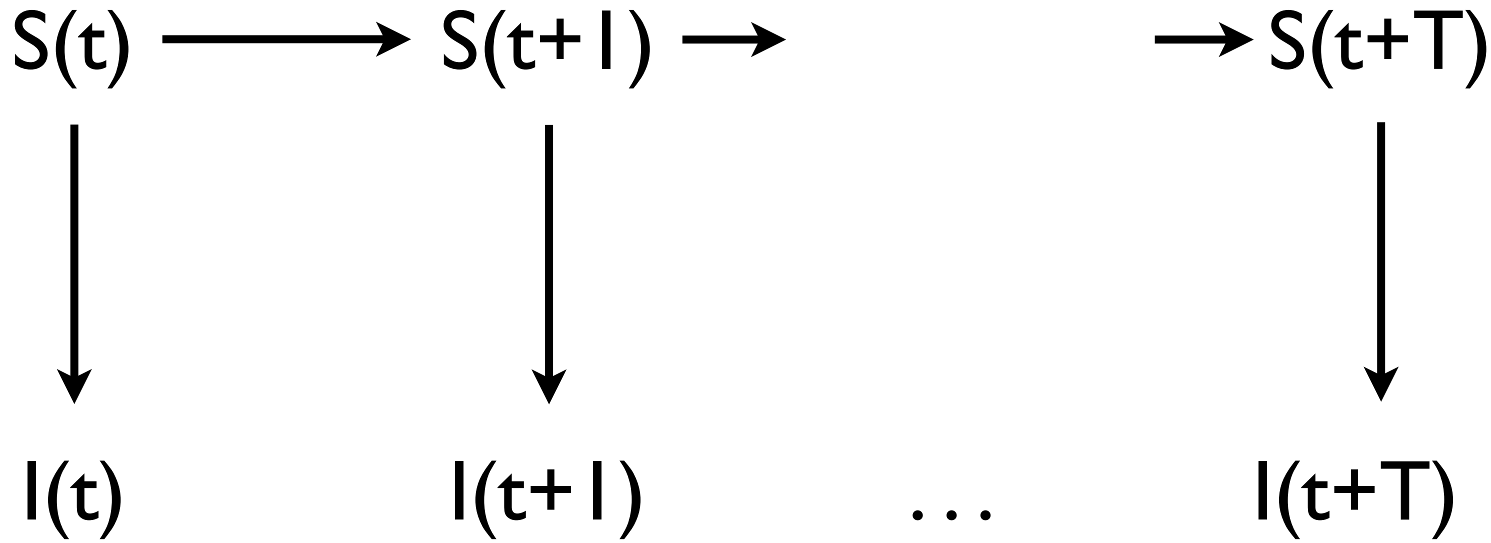
(2010)

Analysis-by-Synthesis in Perception



Kersten, NIPS 1998 Tutorial on Computational Vision

Analysis-by-Synthesis in Perception



Goal: $P(S|I) \propto P(I|S)P(S)$

Are probabilities necessary?

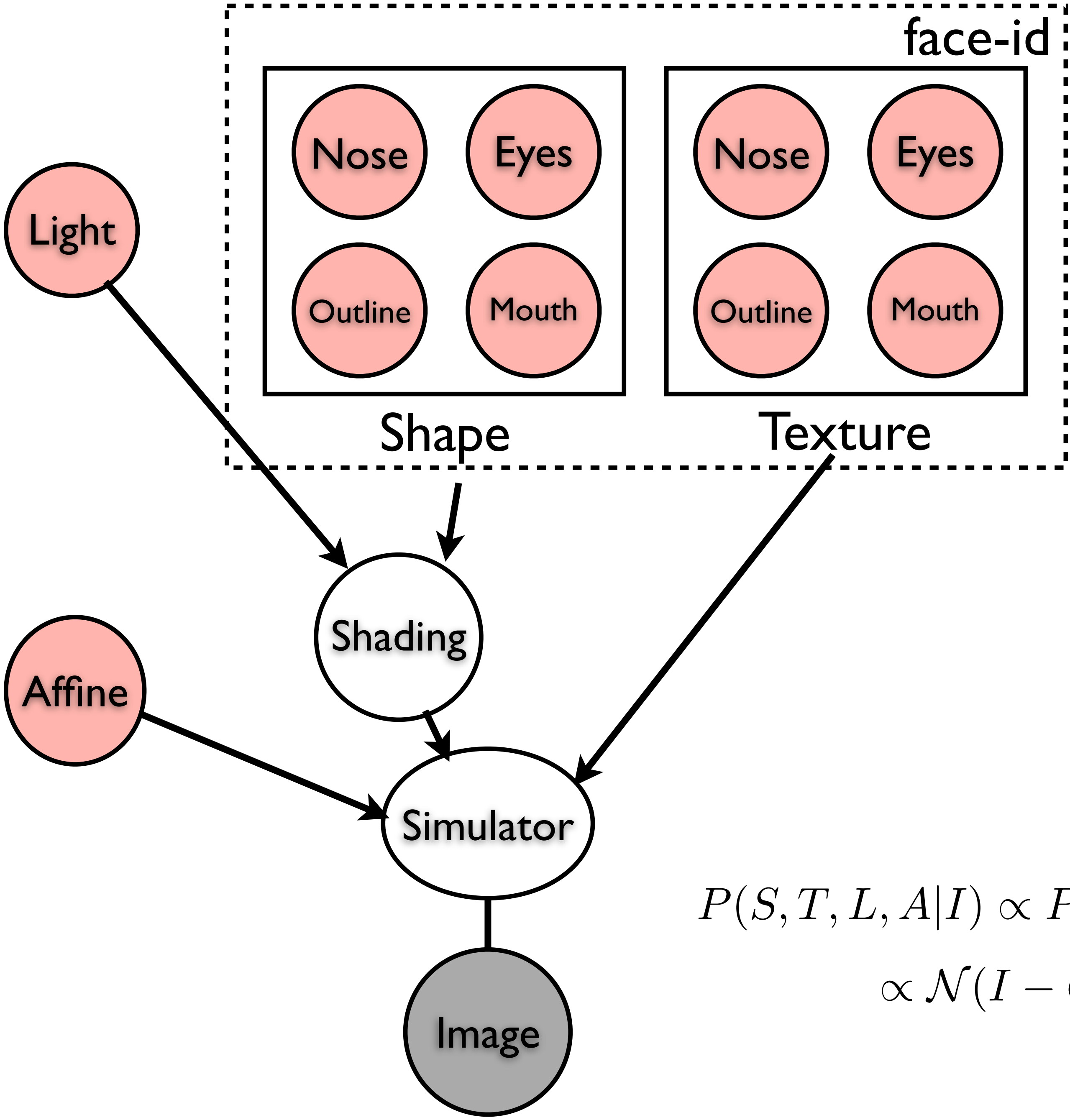
- Our internal models of reality are often incomplete. Therefore we need a mathematical language to handle uncertainty
- Probability theory is a framework to extend logic to include reasoning on uncertain information
- Probability need not have anything to do with randomness. Probabilities do not describe reality -- only our information about reality - *E.T. Jaynes*
- Bayesian statistics describes epistemological (*study of the nature and scope of knowledge*) uncertainty using the mathematical language of probability
- Start with prior beliefs and update these using data to give posterior beliefs

Are probabilities necessary?

Assumptions violated:
broad posterior

Assumptions satisfied:
narrower posterior

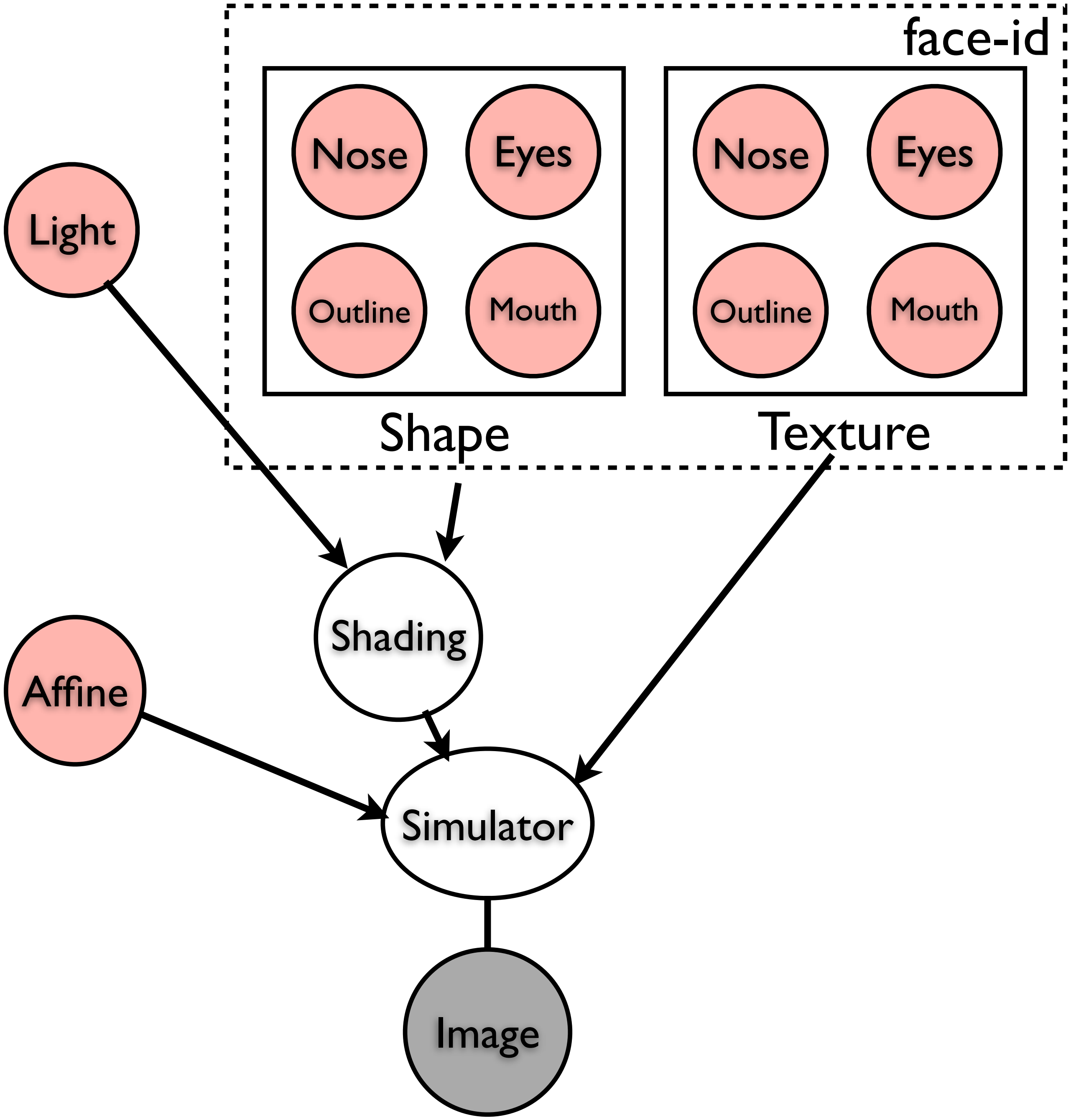
Probabilistic 3D Face Analysis



Inference Problem:

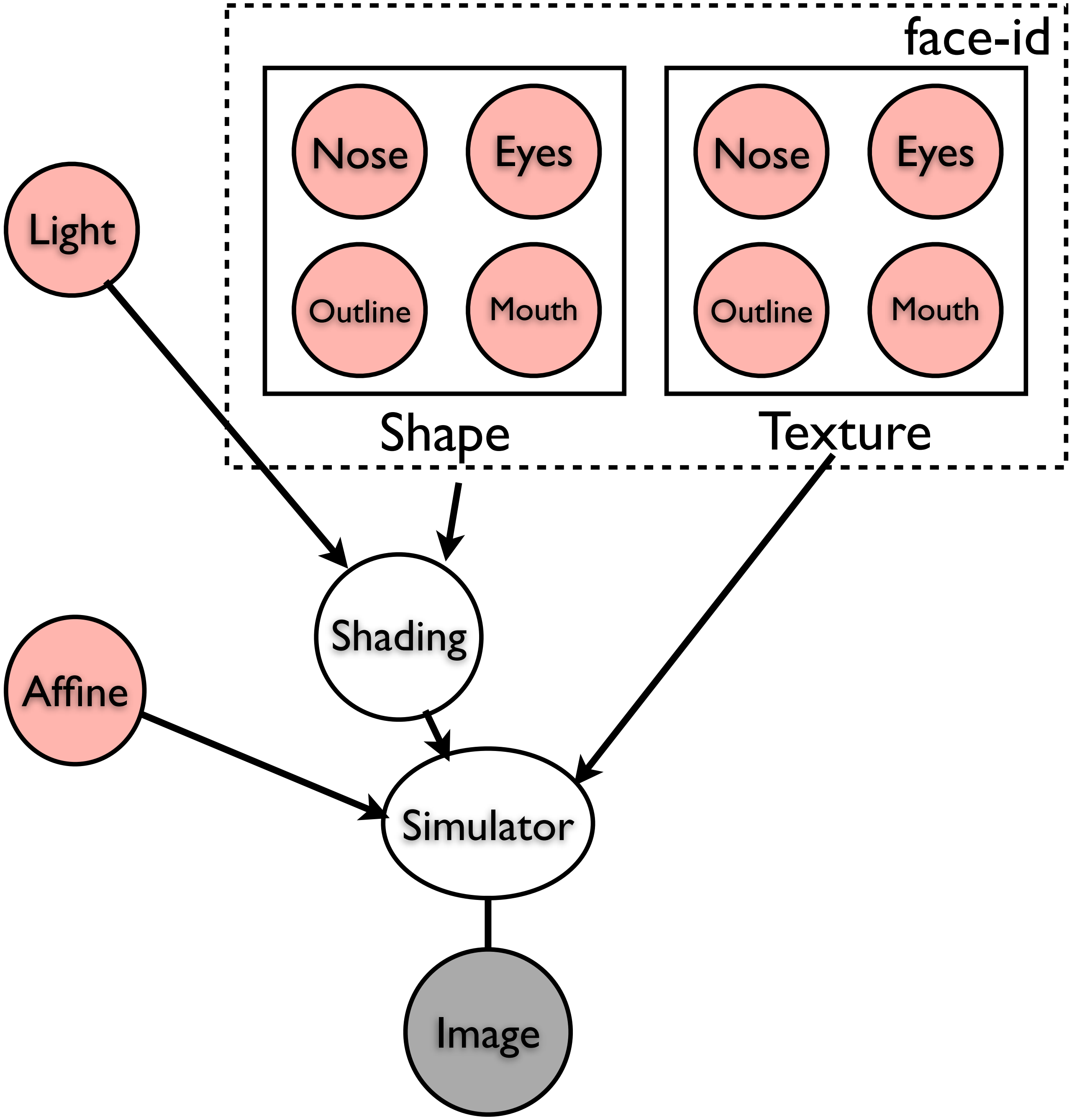
$$P(S,T,L,A|I) \propto P(I|S,T,L,A)P(L)P(S)P(T)P(A)$$
$$\propto \mathcal{N}(I-O;0,0.1)P(L)P(A)\prod_i P(S_i)P(T_i)$$

Probabilistic 3D Face Analysis



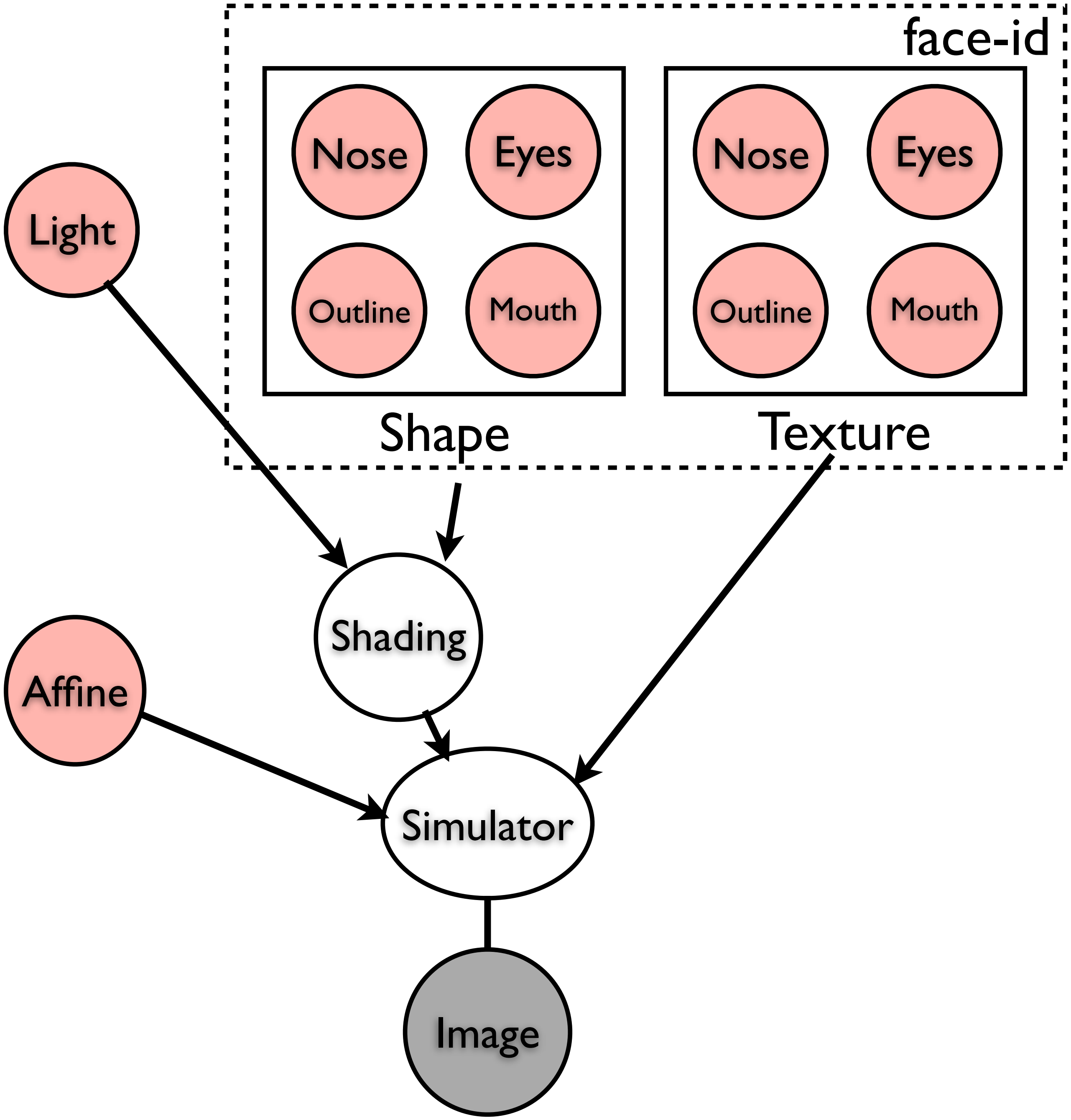
Random Draw

Probabilistic 3D Face Analysis



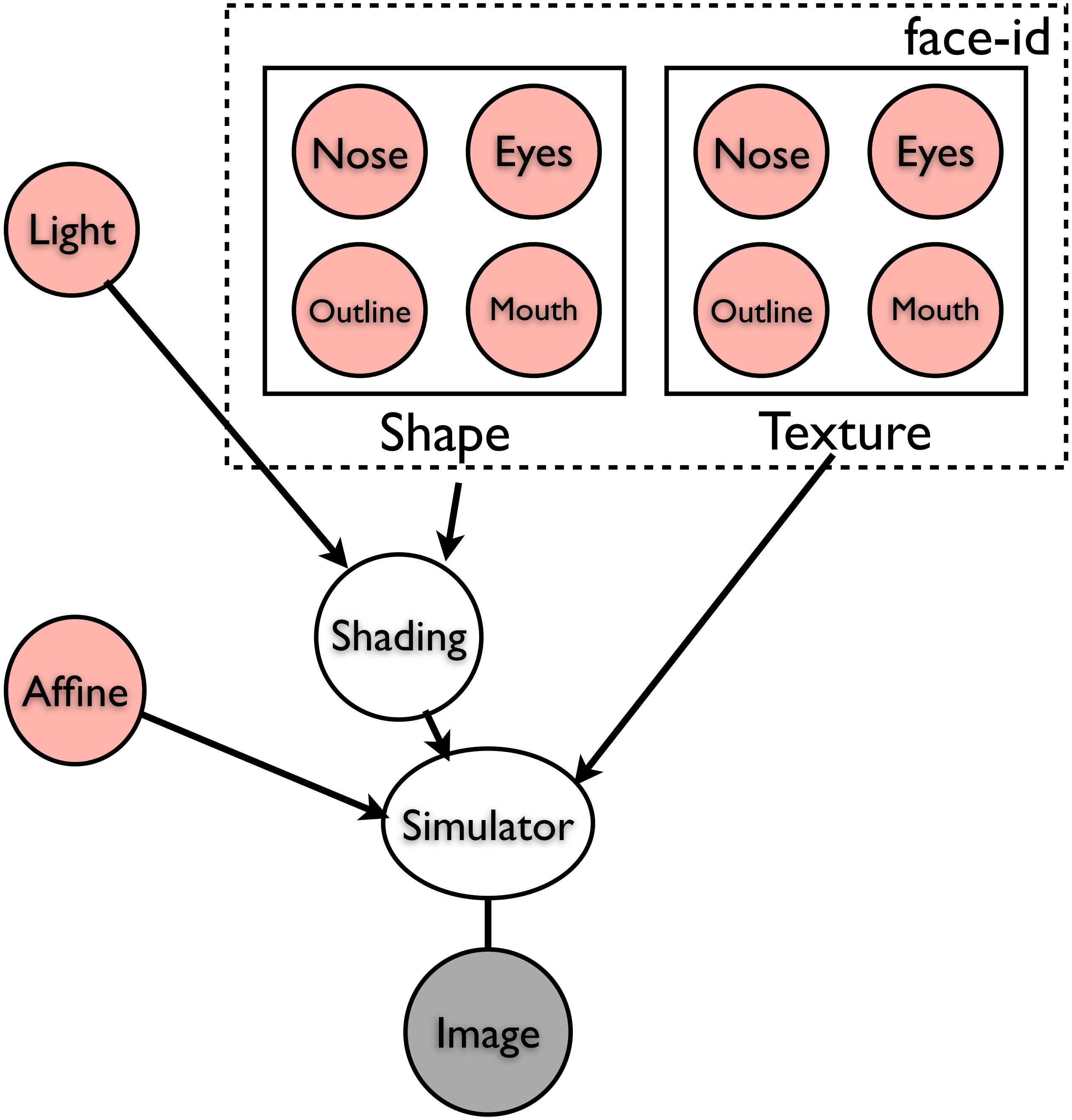
Random Draw

Probabilistic 3D Face Analysis



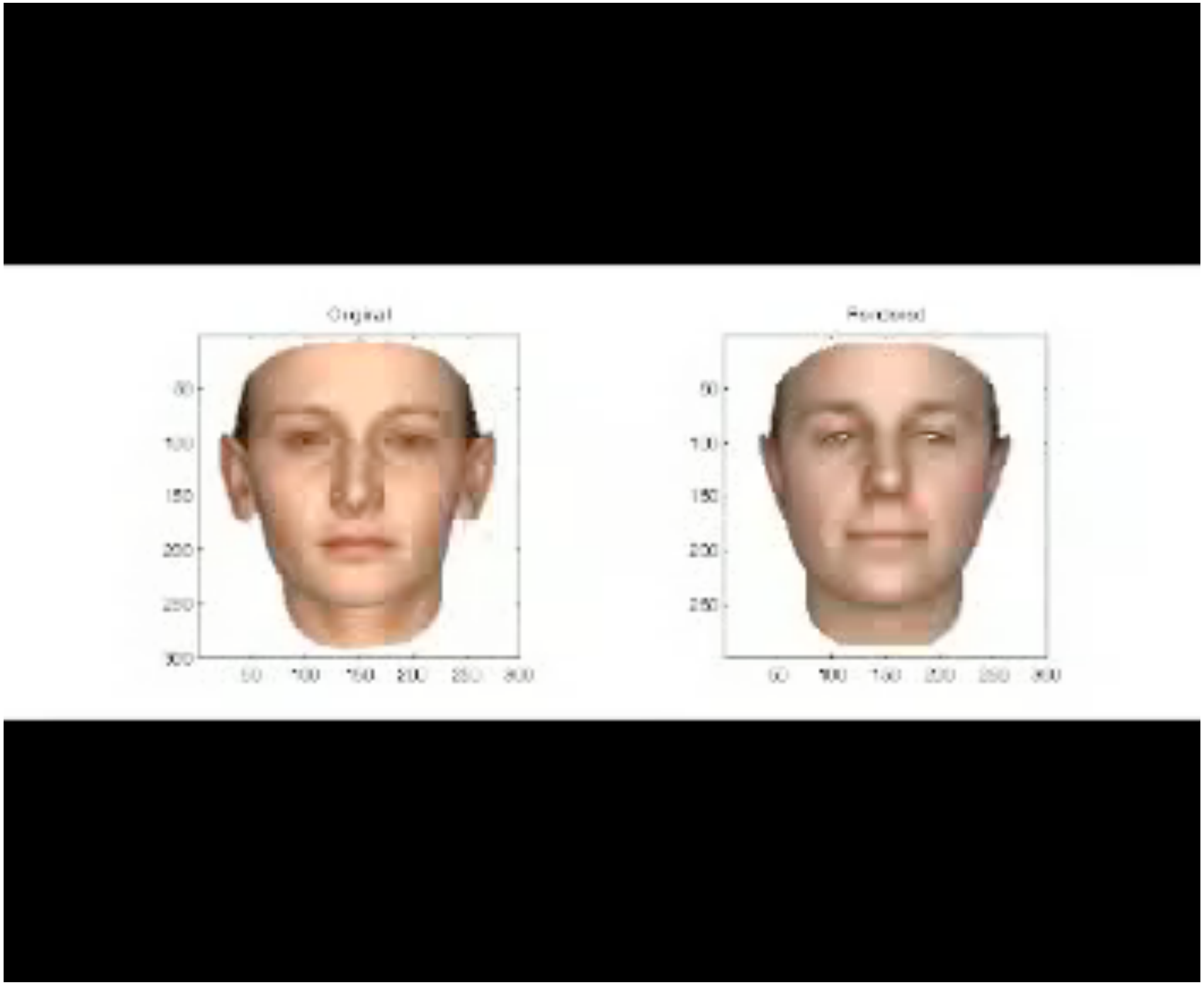
Random Draw

Probabilistic 3D Face Analysis

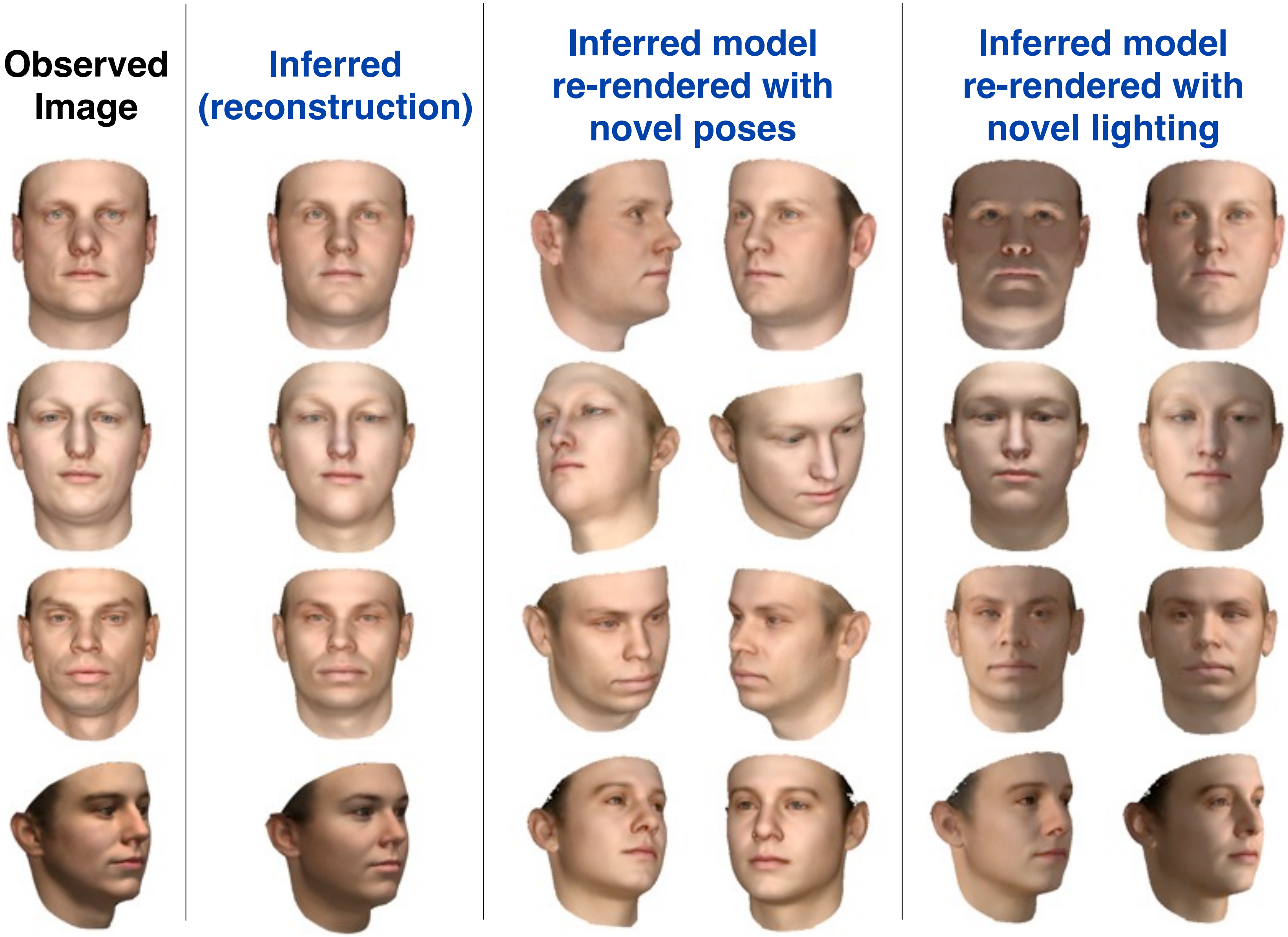


Random Draw

Probabilistic 3D Face Analysis



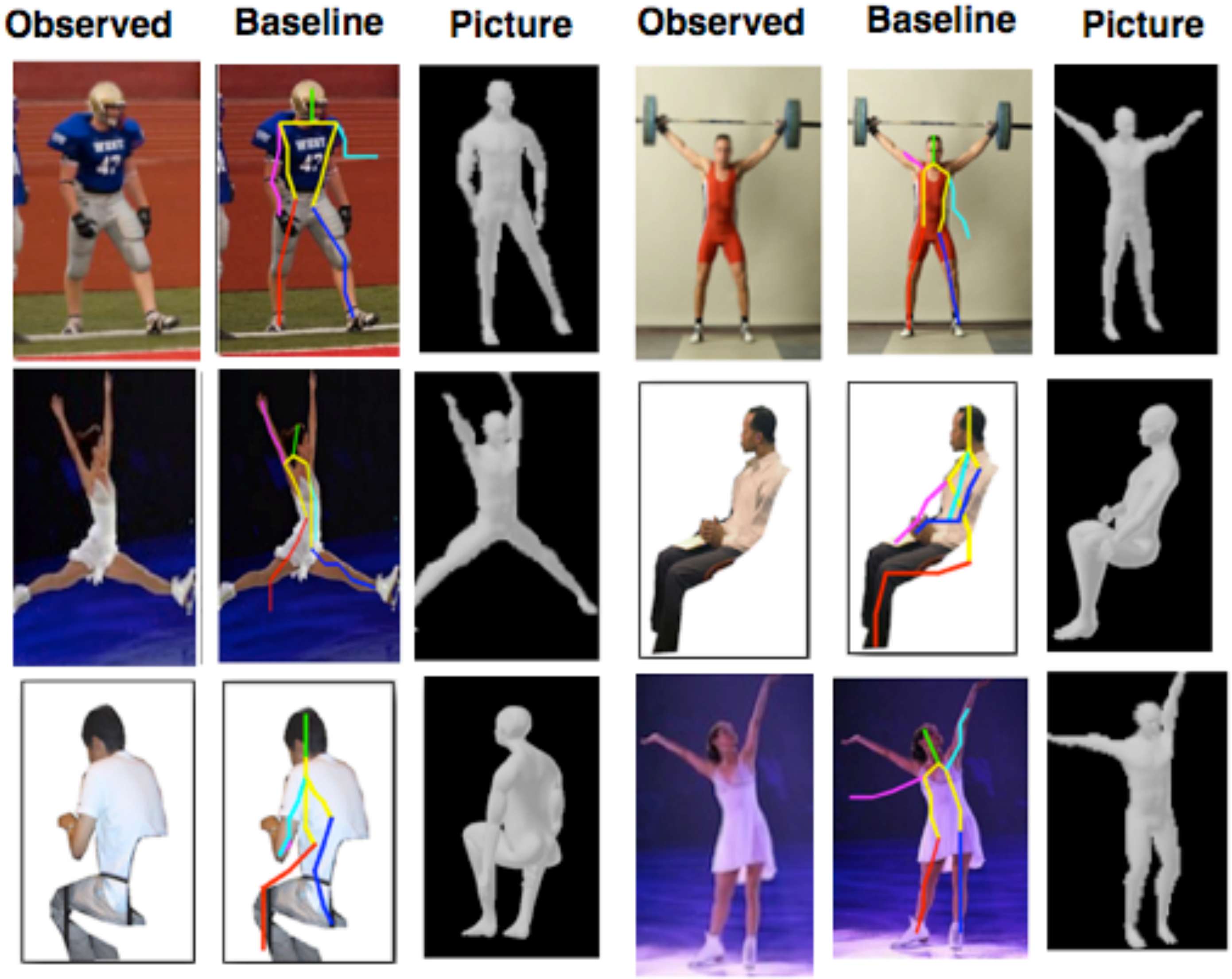
Probabilistic 3D Face Analysis



3D Human Pose

Test Image

Inference Trajectory



3D Shape Program

Test Image

Inference Trajectory

Quantitative Metrics		
METHOD	Z-MAE	N-MSE
Barron et al.	15.19	2.407×10^{-3}
Picture	11.40	1.704×10^{-3}

Inference

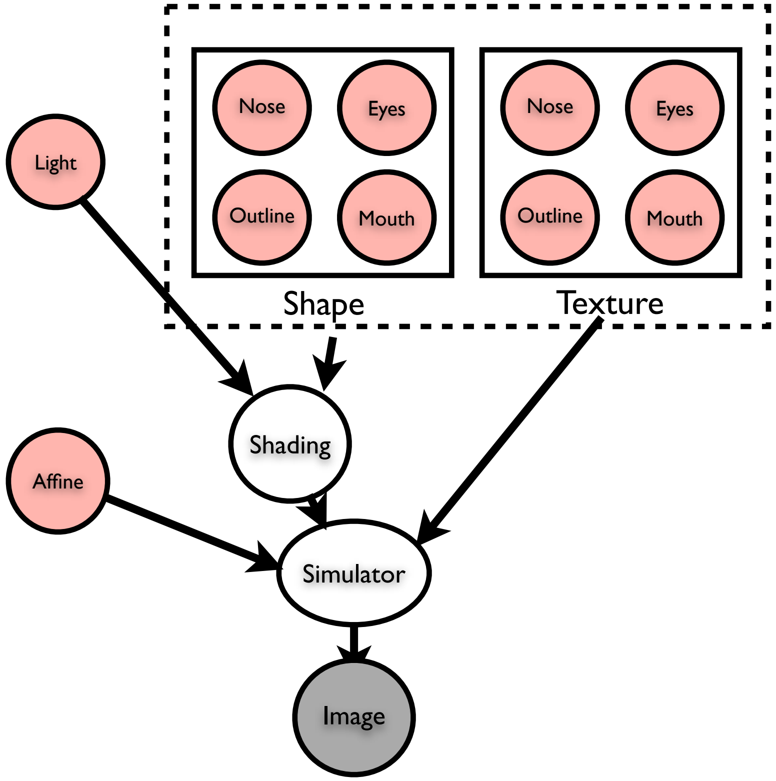
- Many inference strategies for generative models: MCMC, Variational, Message Passing, Particle filtering etc.
- Today we will discuss an algorithm that is simple and general (not necessarily efficient)

Inference

- Simplest MCMC Algorithm: Metropolis Hastings
- For simplicity, let us fix light and affine variables.

$$S_{\text{Nose}} \sim randn(50)$$
$$T_{\text{Nose}} \sim randn(50)$$
$$\vdots$$
$$S_{\text{Mouth}} \sim randn(50)$$
$$T_{\text{Mouth}} \sim randn(50)$$

$$P(I|S,T) \propto \text{Normal}(O - R; 0, \sigma_0)$$



MCMC

- Simplest MCMC Algorithm: Metropolis Hastings
- For simplicity, let us fix light and affine variables.

$S_{\text{Nose}} \sim randn(50)$

$T_{\text{Nose}} \sim randn(50)$

$\vdots$

$S_{\text{Mouth}} \sim randn(50)$

$T_{\text{Mouth}} \sim randn(50)$

$$P(I|S,T) \propto \text{Normal}(O - R; 0, \sigma_0)$$

Repeat until convergence:

(1) Let x be either S_i or T_i

We sample new $x' \sim randn(50)$

(2) $r = \dfrac{p(x')q(x|x')}{p(x)q(x'|x)}$

(3) Accept x' with probability: $\alpha = min\{1, r\}$

Otherwise, x'=x

How do we make inference faster?

Combine Probabilistic Inference with Neural Nets

Combine Probabilistic Inference with Neural Nets

- Inference often gets stuck in local minima and the only way out is if large set of variables are changed at once

Combine Probabilistic Inference with Neural Nets

- Inference often gets stuck in local minima and the only way out is if large set of variables are changed at once
- Helmholtz machine: Wake-Sleep Alg (Dayan, 1995)

Combine Probabilistic Inference with Neural Nets

- Inference often gets stuck in local minima and the only way out is if large set of variables are changed at once

- Helmholtz machine: Wake-Sleep Alg (Dayan, 1995)

- Informed Sampler (Jampani, 2015)

Combine Probabilistic Inference with Neural Nets

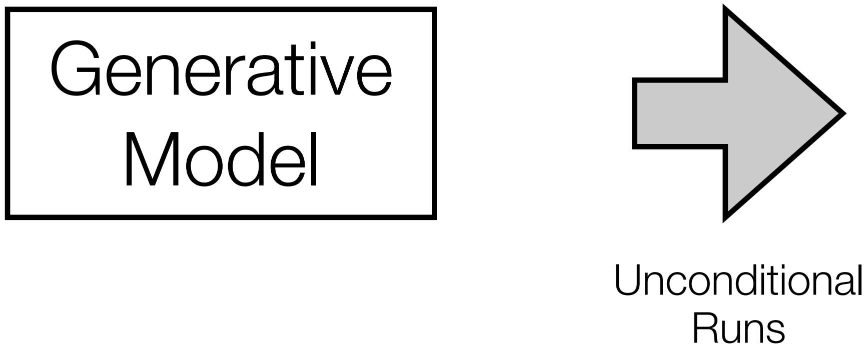
- Inference often gets stuck in local minima and the only way out is if large set of variables are changed at once
- Helmholtz machine: Wake-Sleep Alg (Dayan, 1995)
- Informed Sampler (Jampani, 2015)
- Use an external long-term memory to cache "hallucinations" synthesized from your generative models (*sleep*) and use them during perception (*wake)*

Combine Probabilistic Inference with Neural Nets

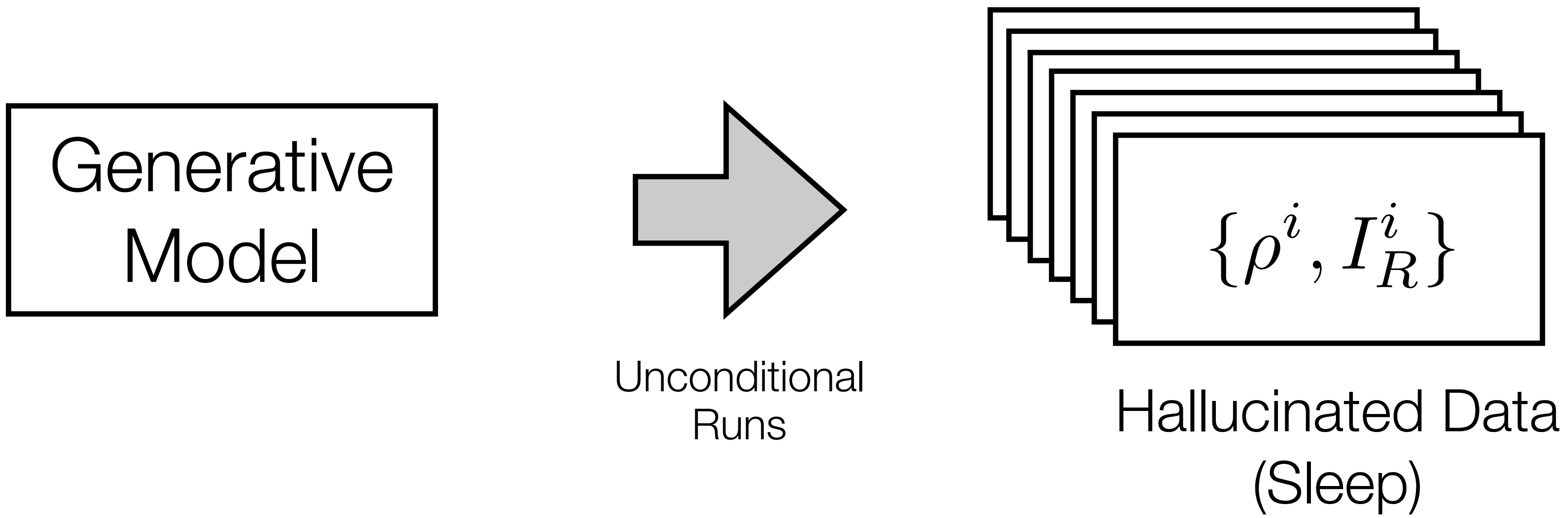
Combine Probabilistic Inference with Neural Nets

Generative Model

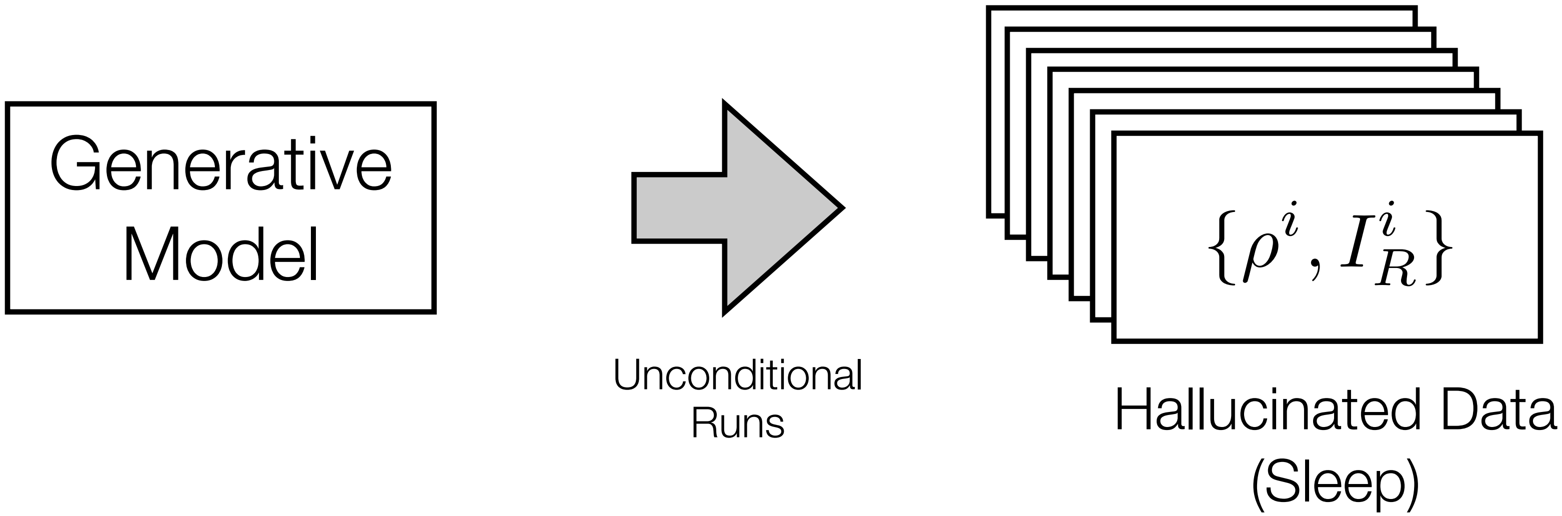
Combine Probabilistic Inference with Neural Nets

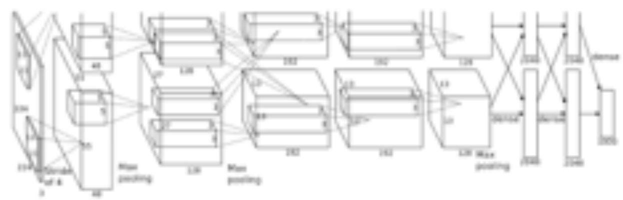


Combine Probabilistic Inference with Neural Nets



Combine Probabilistic Inference with Neural Nets

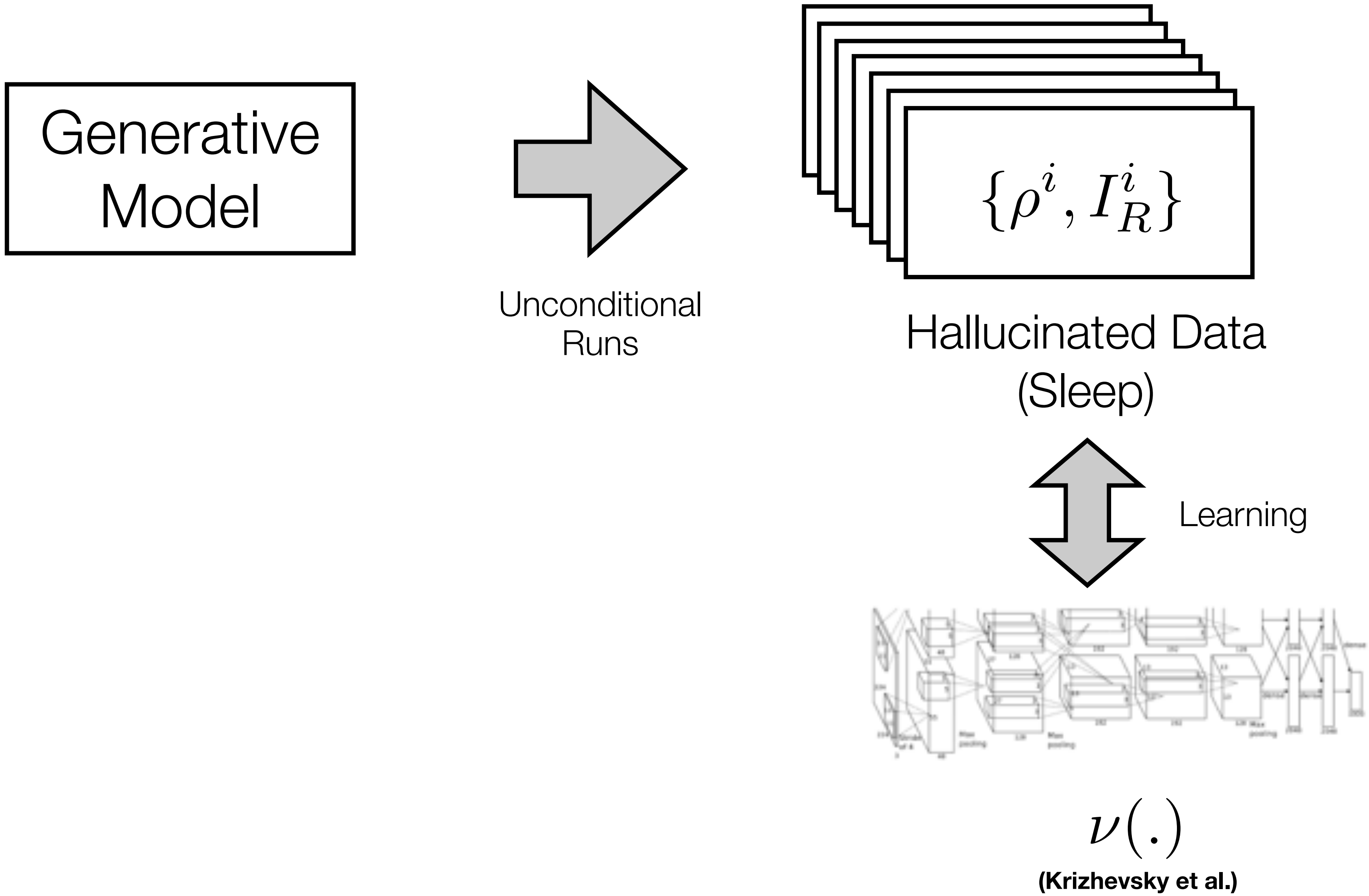




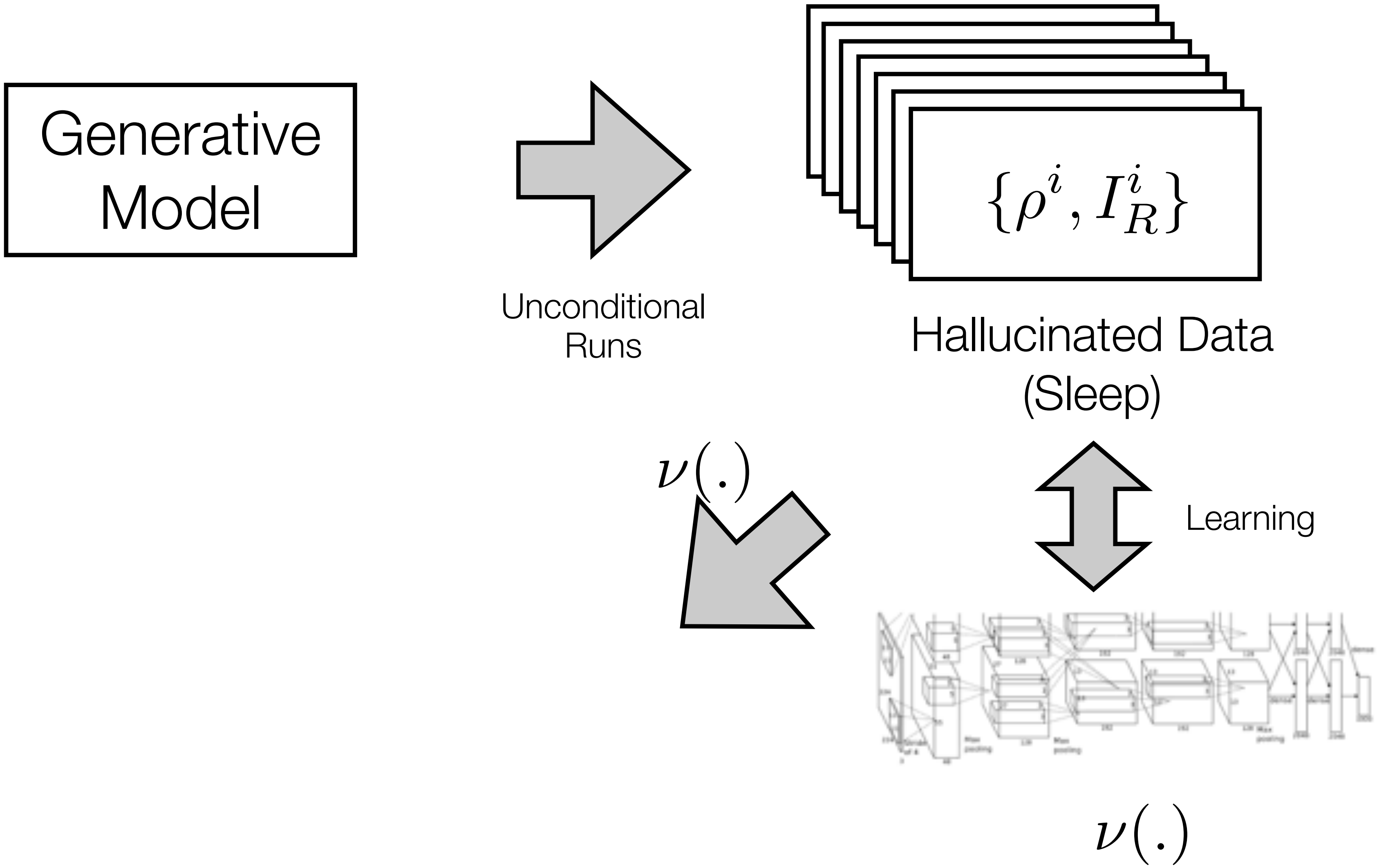
$\nu(.)$

(Krizhevsky et al.)

Combine Probabilistic Inference with Neural Nets

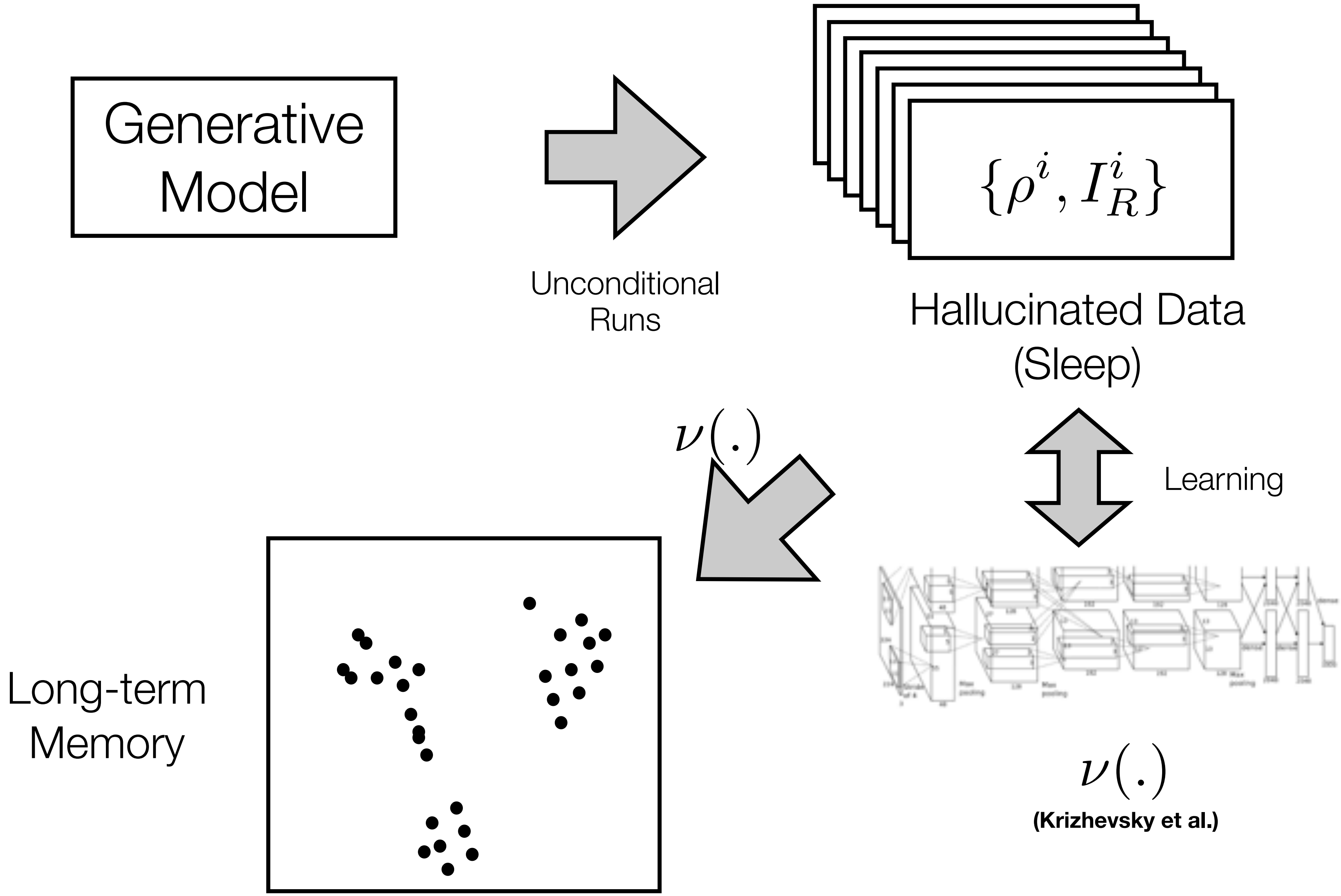


Combine Probabilistic Inference with Neural Nets

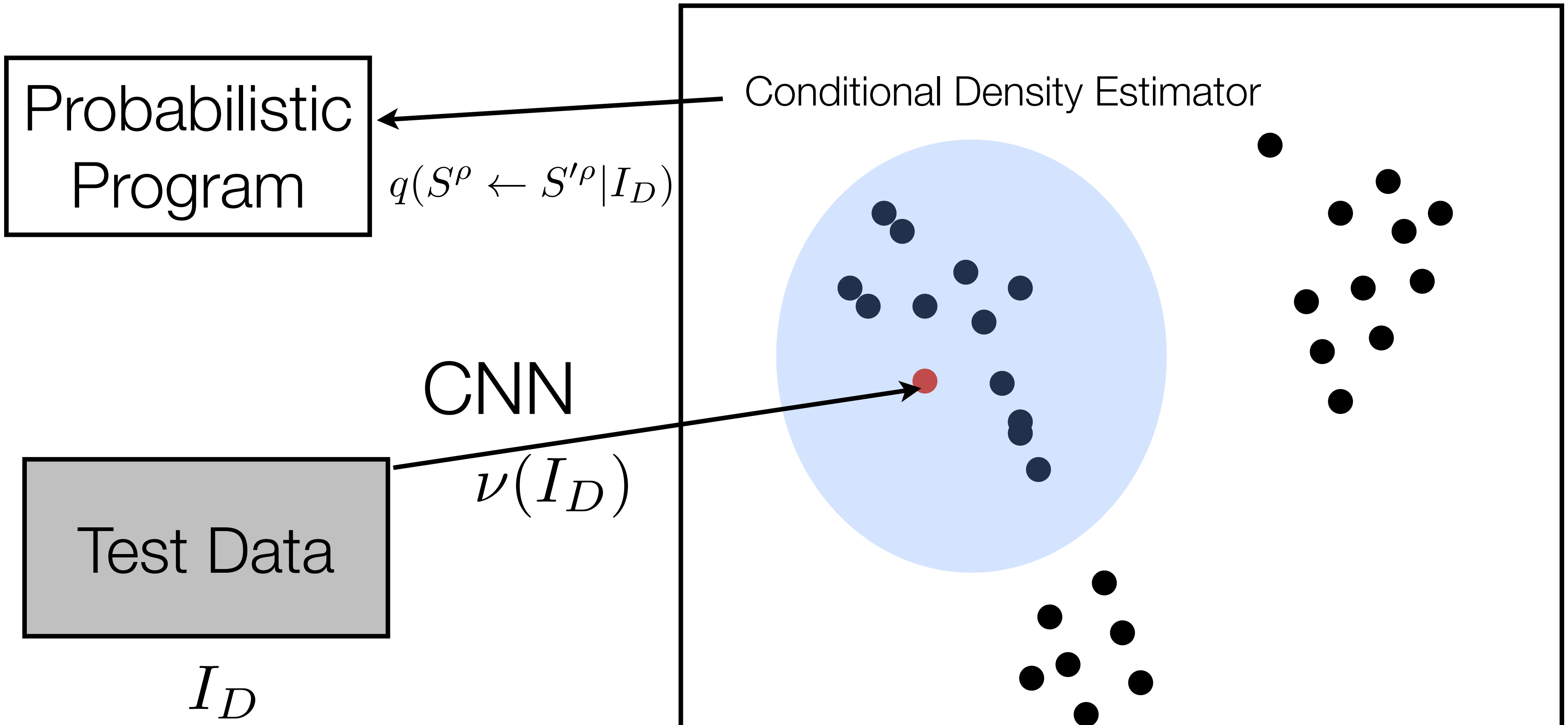


(Krizhevsky et al.)

Combine Probabilistic Inference with Neural Nets



Combine Probabilistic Inference with Neural Nets



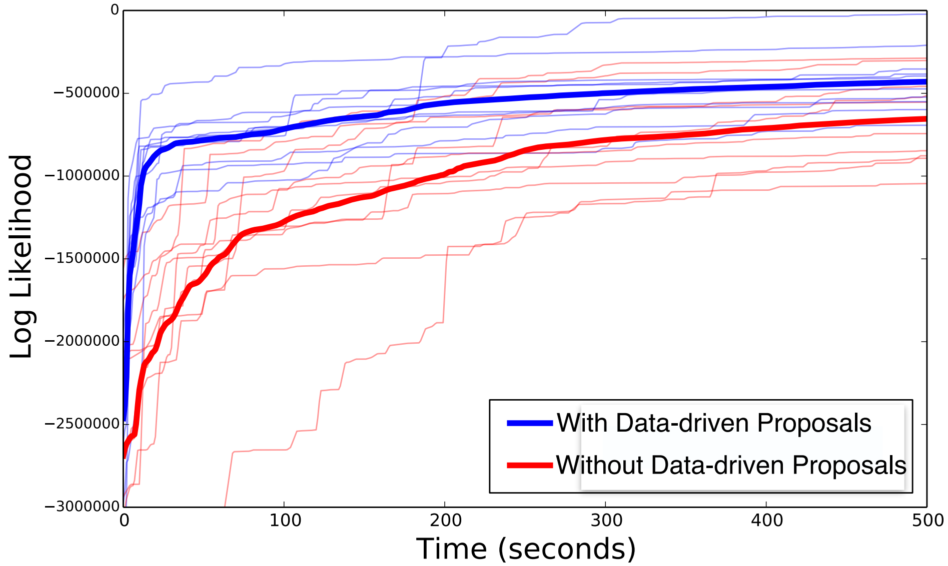
Long-term Memory
(Sleep)

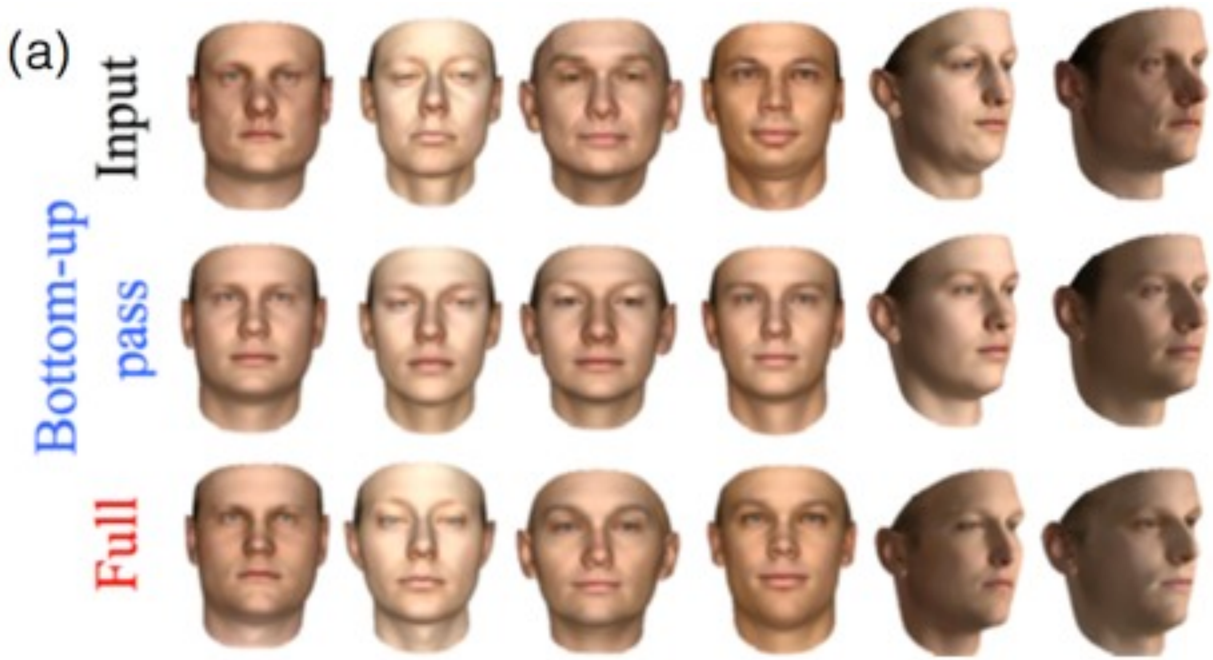
Now run inference

90%: Data-driven (Pattern Matching)

10%: Sampling/Search (Reasoning)

3D Face Analysis

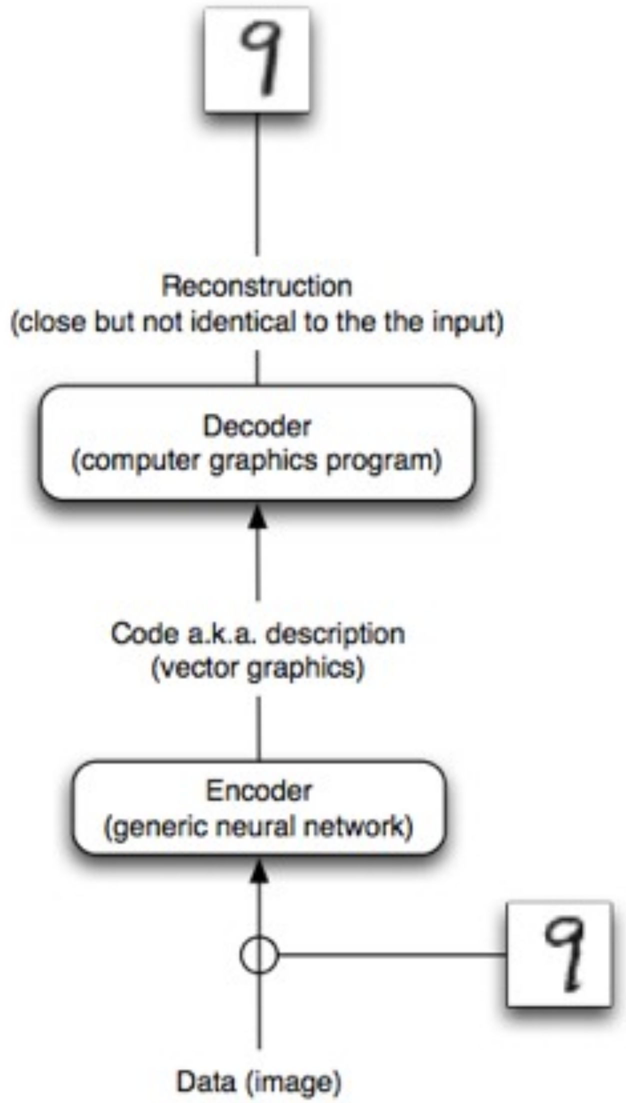




Learning parametrized generative models

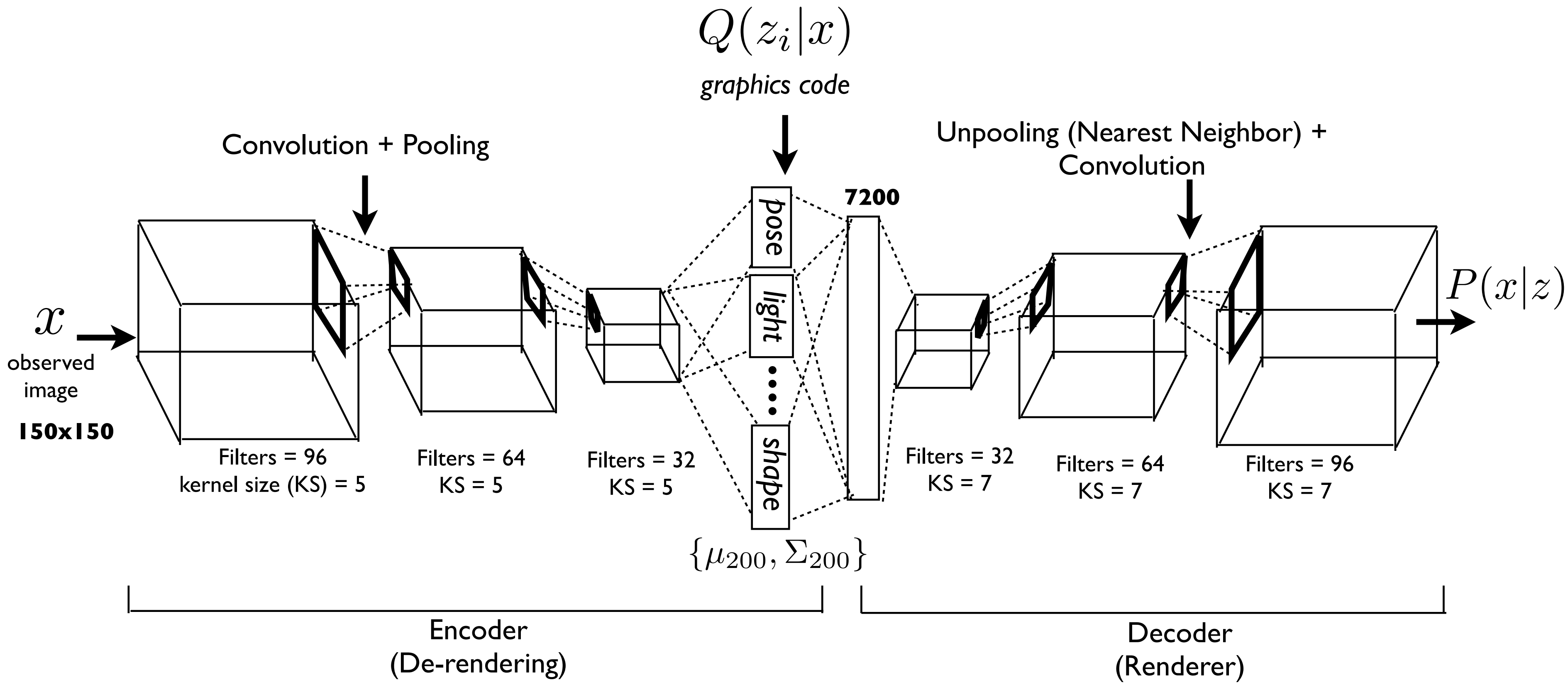
Tijmen Tielemen (Thesis, 2014)

Learning parametrized generative models



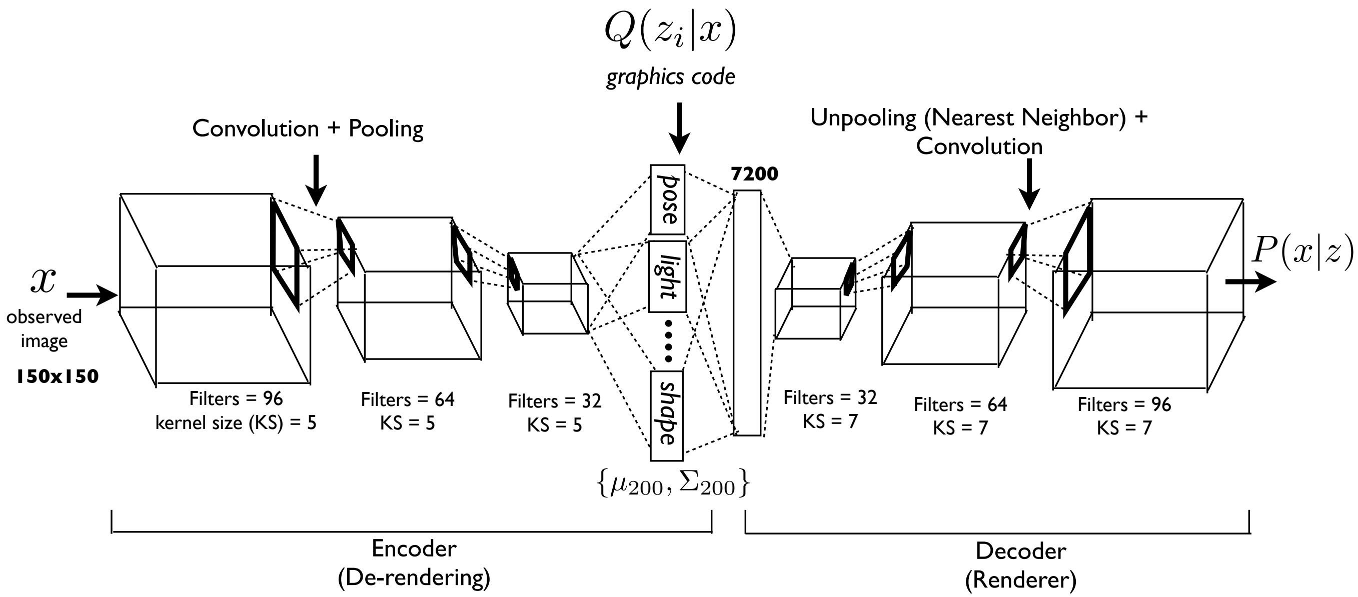
Tijmen Tielemen (Thesis, 2014)

Convolutional Inverse Graphics Network



'Inference' 'Generative Model'

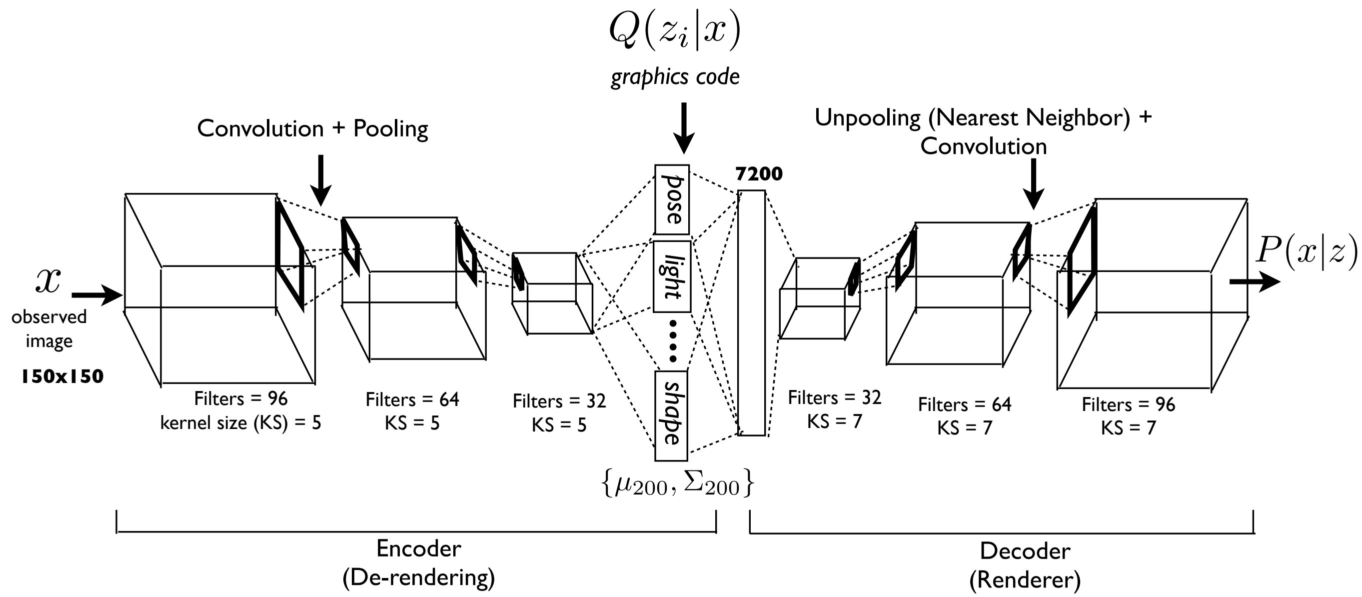
Convolutional Inverse Graphics Network



$$y_e = encoder(x)$$
$$\mu_{z_i} = W_e * y_e$$
$$\Sigma_{z_i} = \mathrm{diag}(\exp(W_e * y_e))$$
$$\forall i, z_i \sim \mathcal{N}(\mu_{z_i}, \Sigma_{z_i})$$

$\mathbf{z} =$	$\mathbf{z}_1$	$\mathbf{z}_2$	$\mathbf{z}_3$	$\mathbf{z}_{[4,n]}$
corresponds to	ϕ	α	ϕ_L	intrinsic properties (shape, texture, etc)

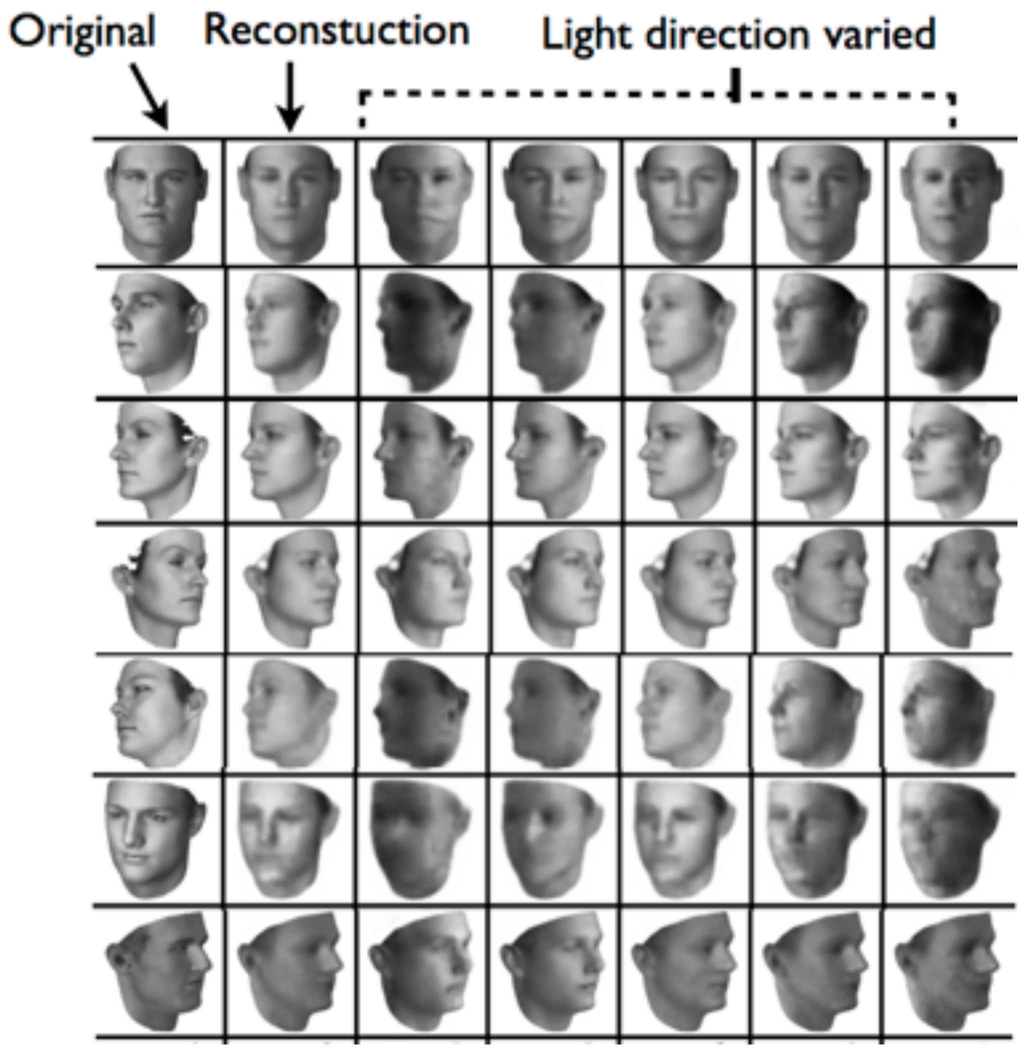
Variational Inference



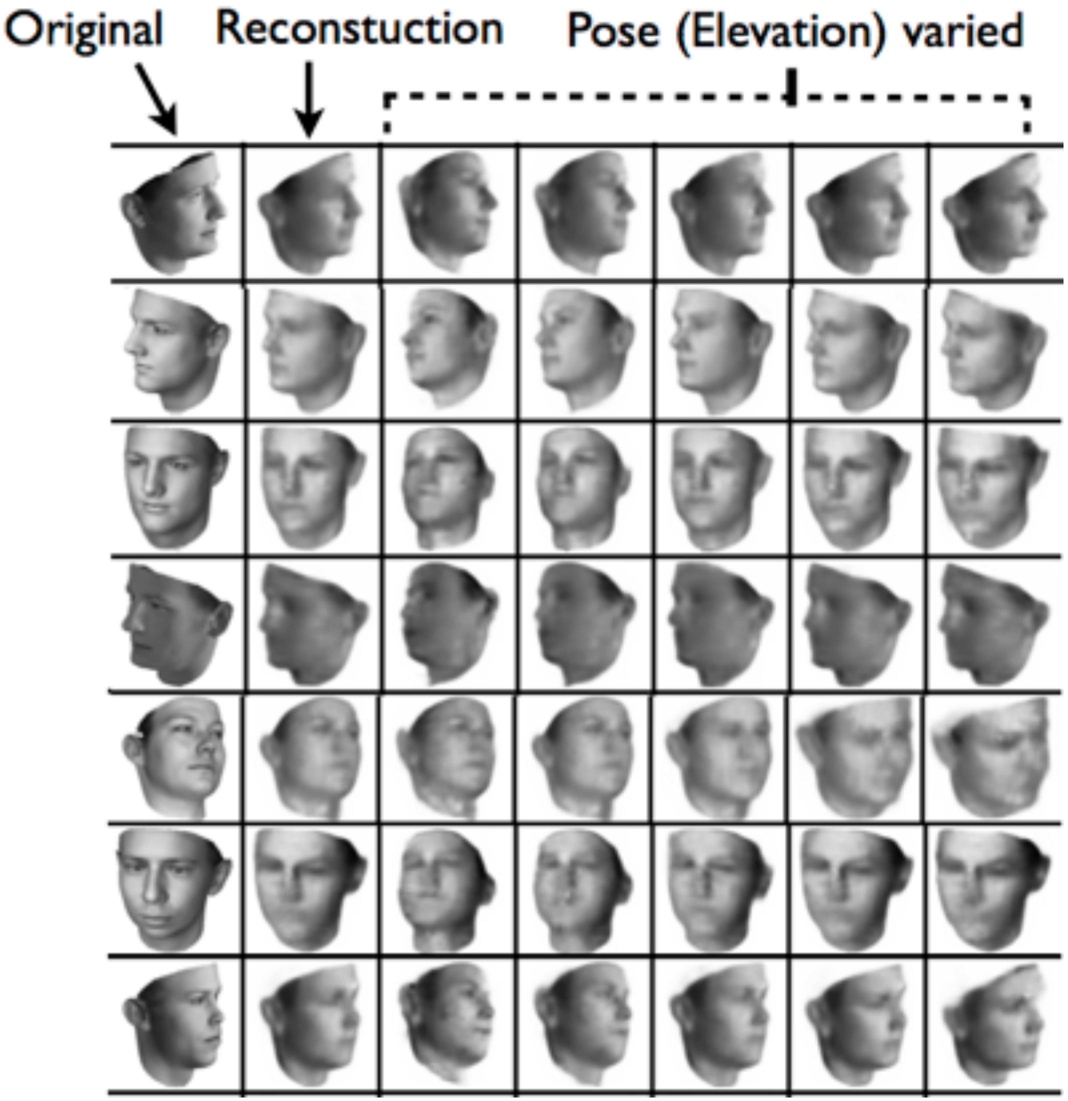
Objective Function: $-logP(x|Z) + KL(Q(Z|x)||P(Z))$

D. P. Kingma and M. Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

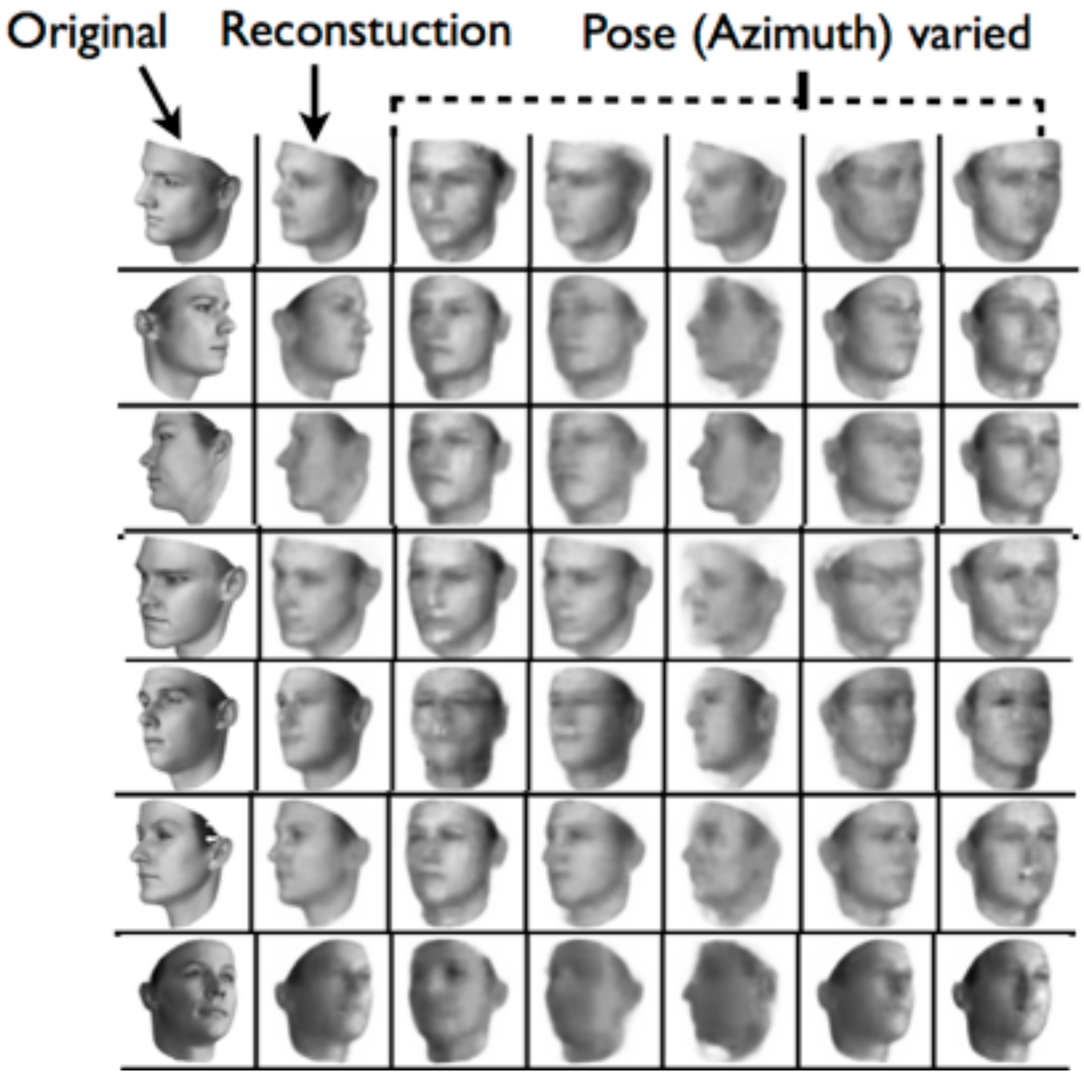
Results



Results

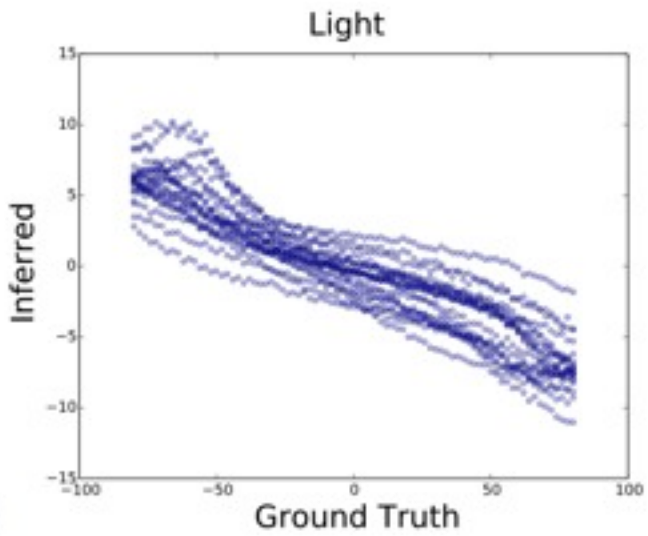


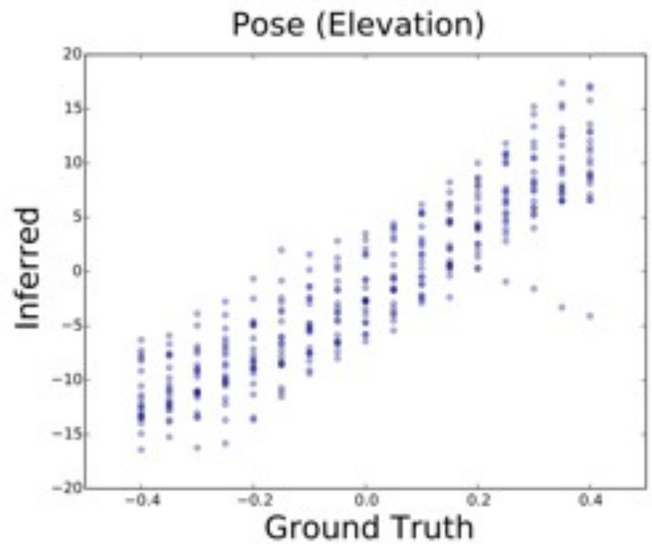
Results

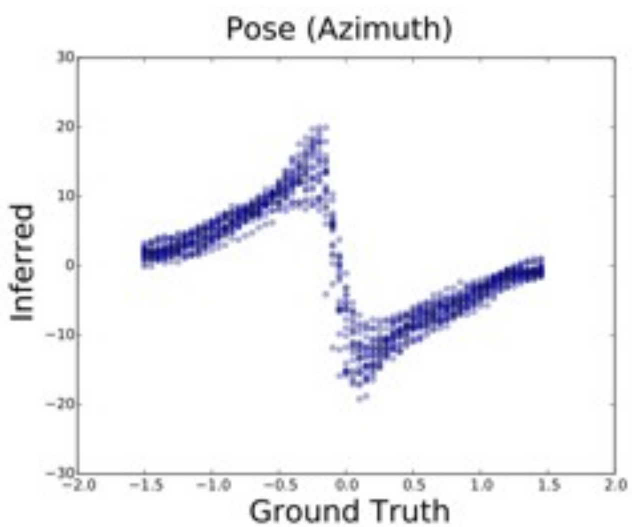


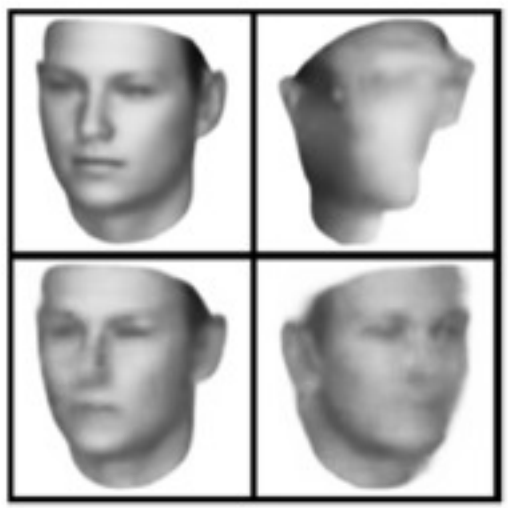
Results

Results

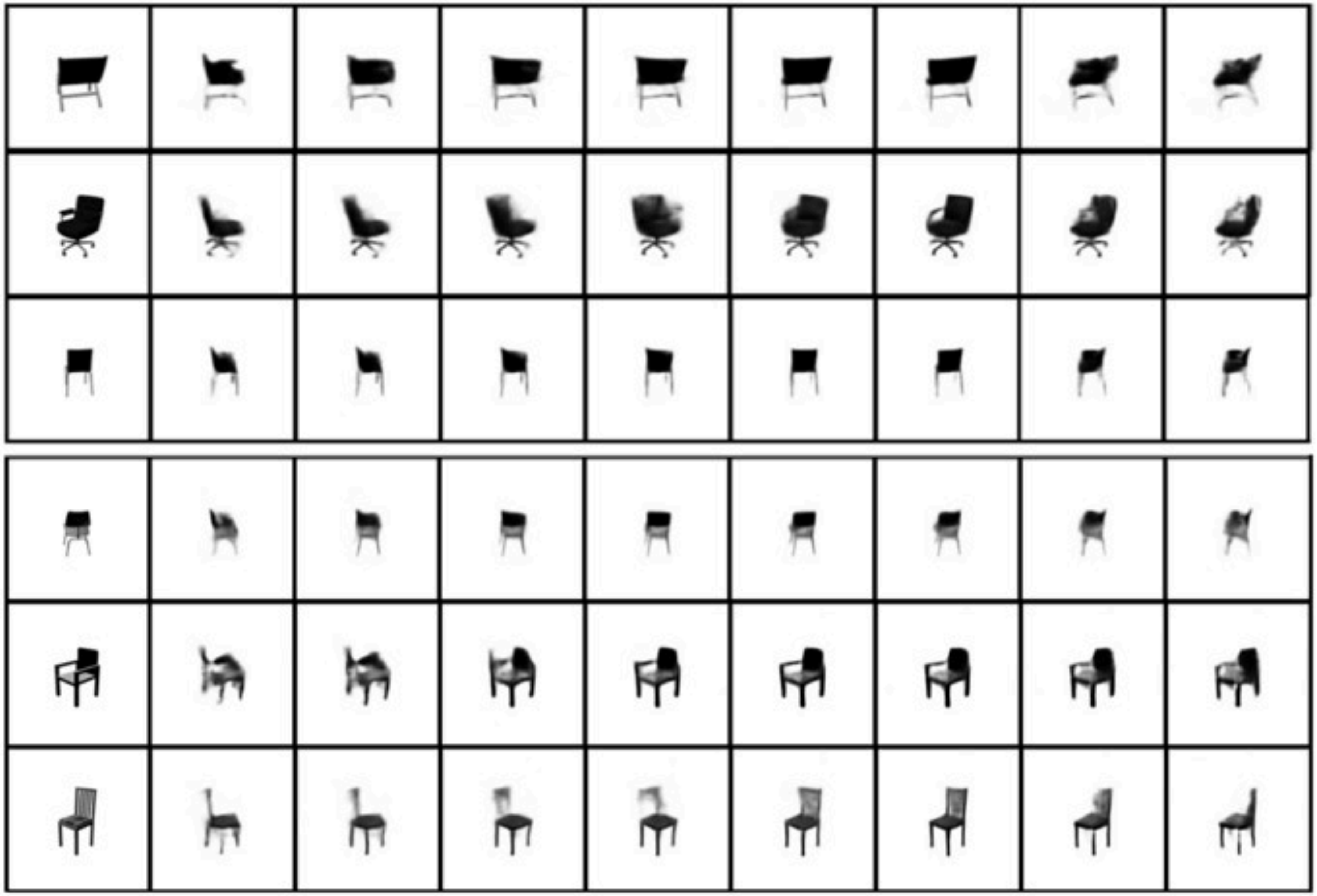

Light
Inferred
Ground Truth
15
10
5
0
−5
−10
−15
−100
−50
0
50
100


Pose (Elevation)
Inferred
Ground Truth
20
15
10
5
0
−5
−10
−15
−20
−0.4
−0.2
0.0
0.2
0.4

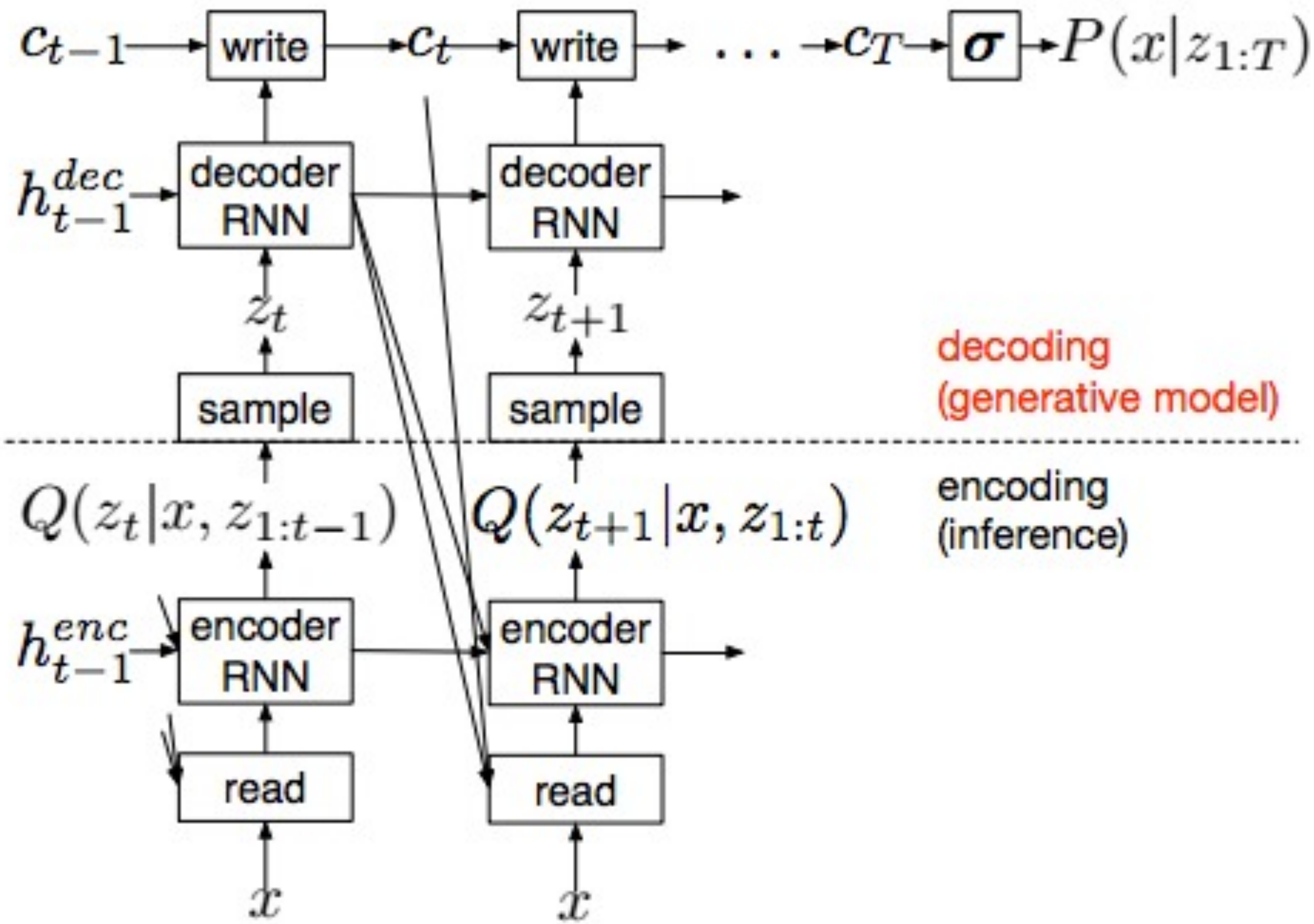

Pose (Azimuth)
Inferred
Ground Truth
30
20
10
0
−10
−20
−30
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0



Results on Chair Dataset



Generative models with attention

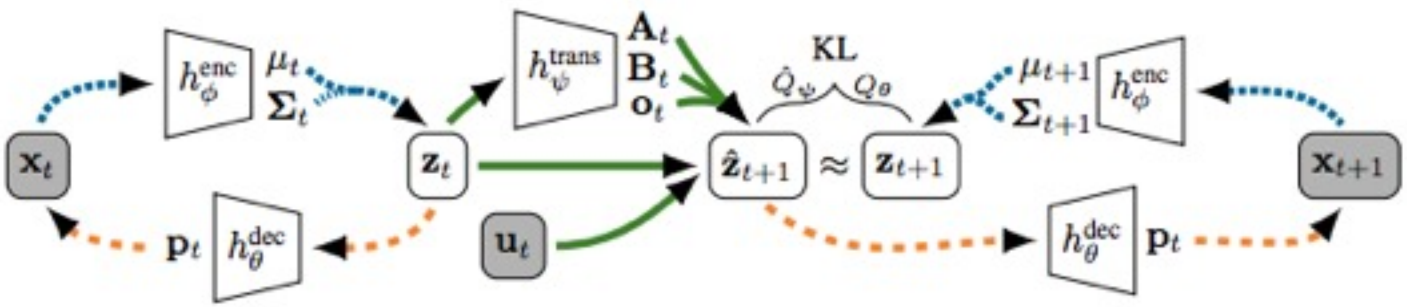


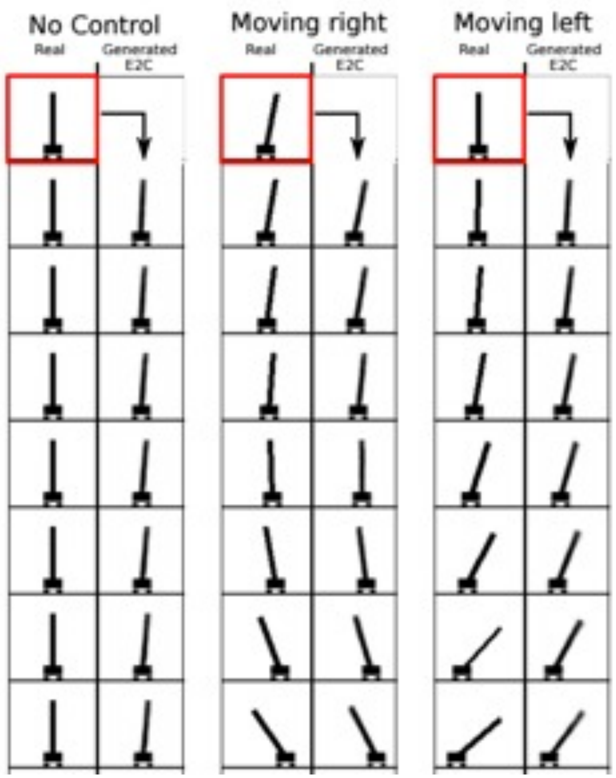
Gregor et al.

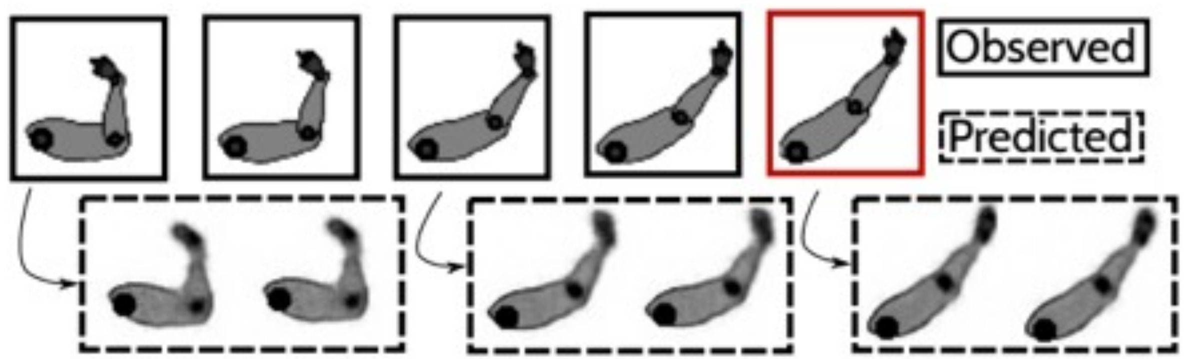
Generative models with attention

Gregor et al.

Integrating actions with generative models

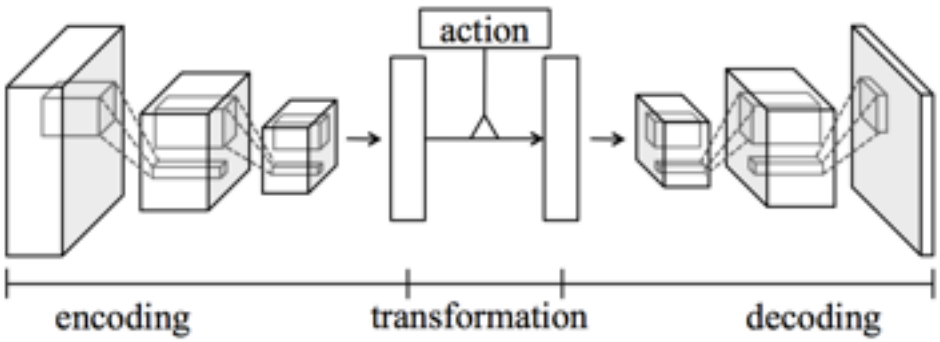






Watter, Springenberg et. al

Results on Atari

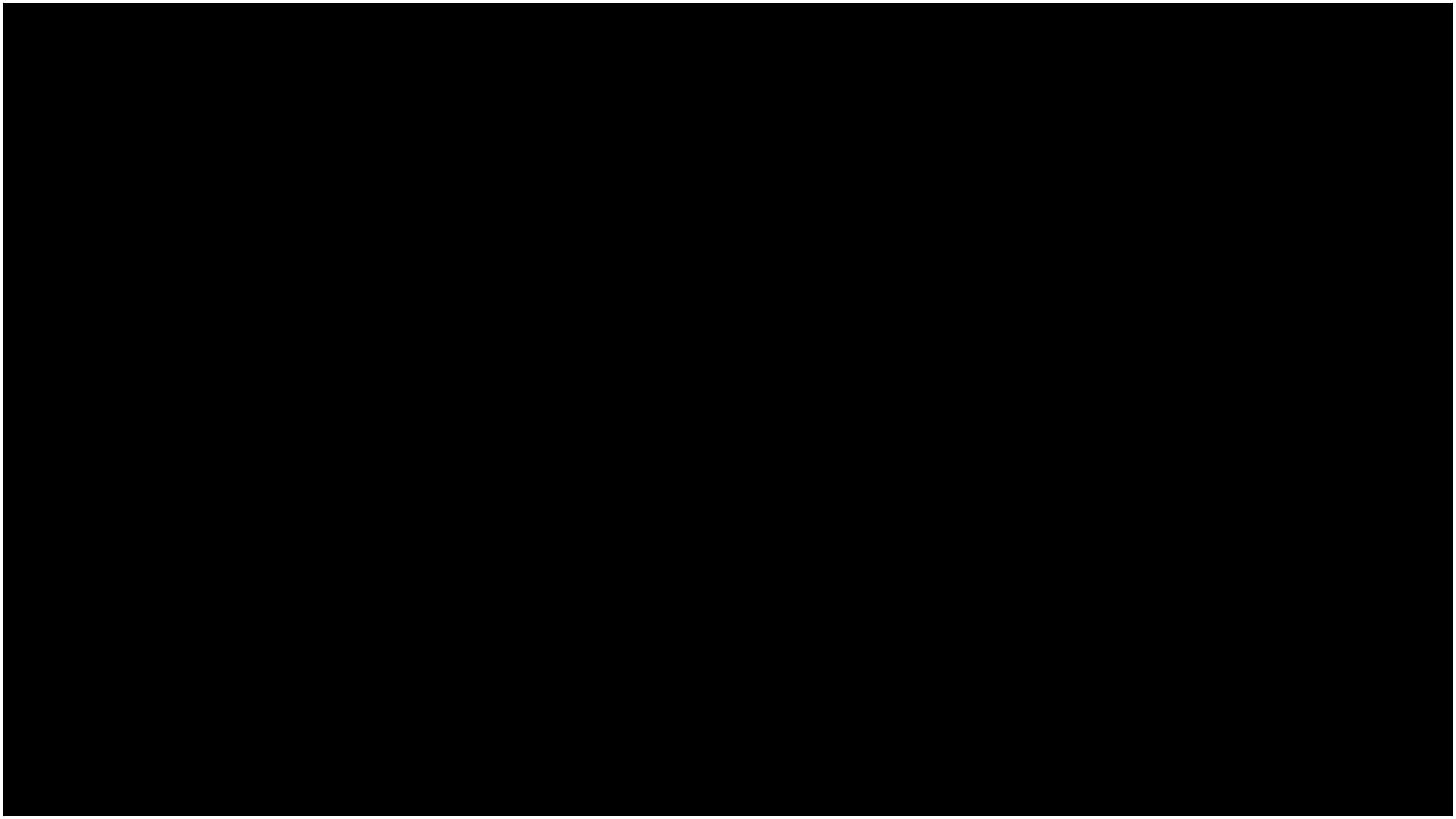


(a) Feedforward encoding

(b) Recurrent encoding

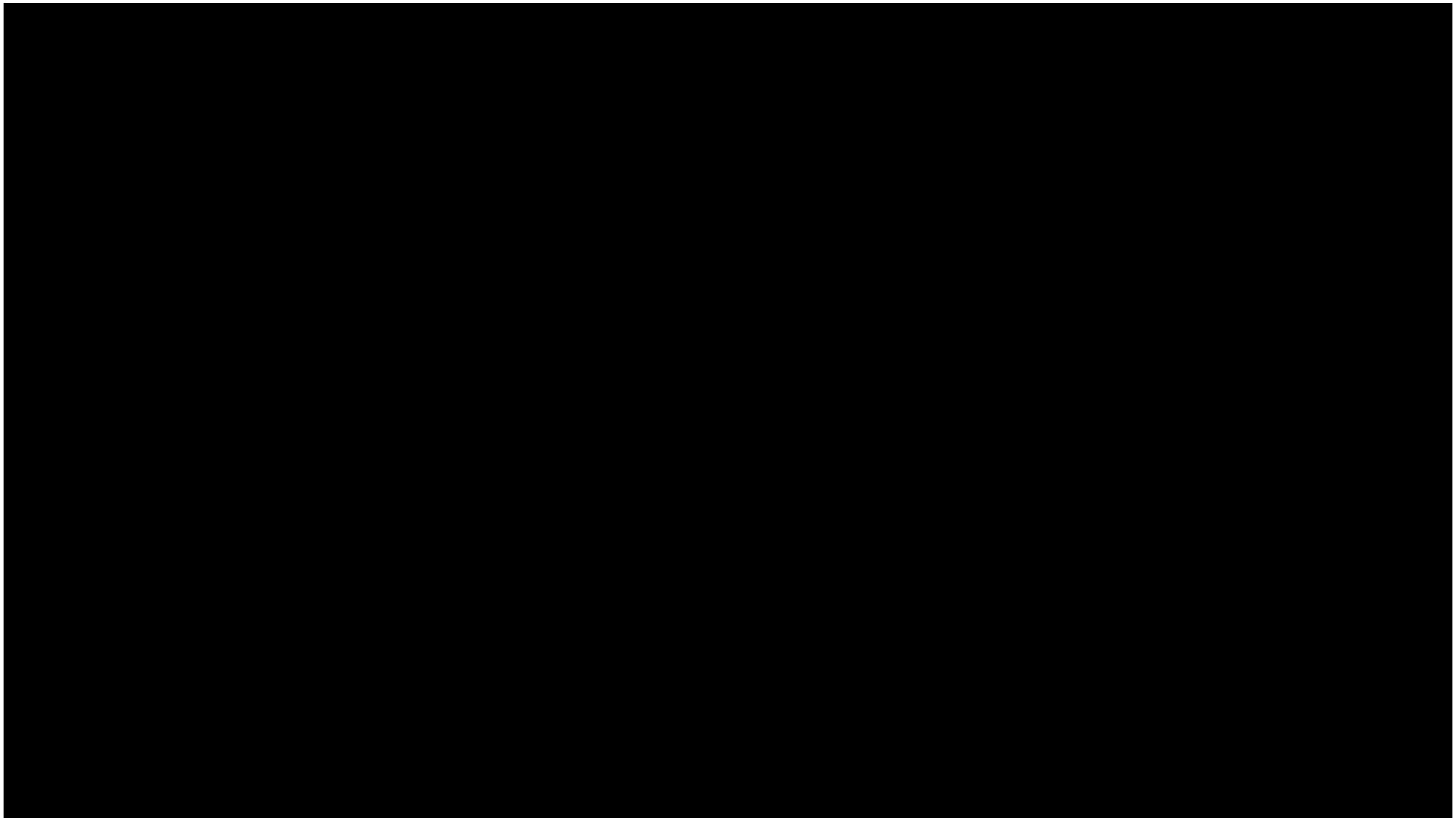
Junhyuk Oh et. al, NIPS 2015

Results on Atari



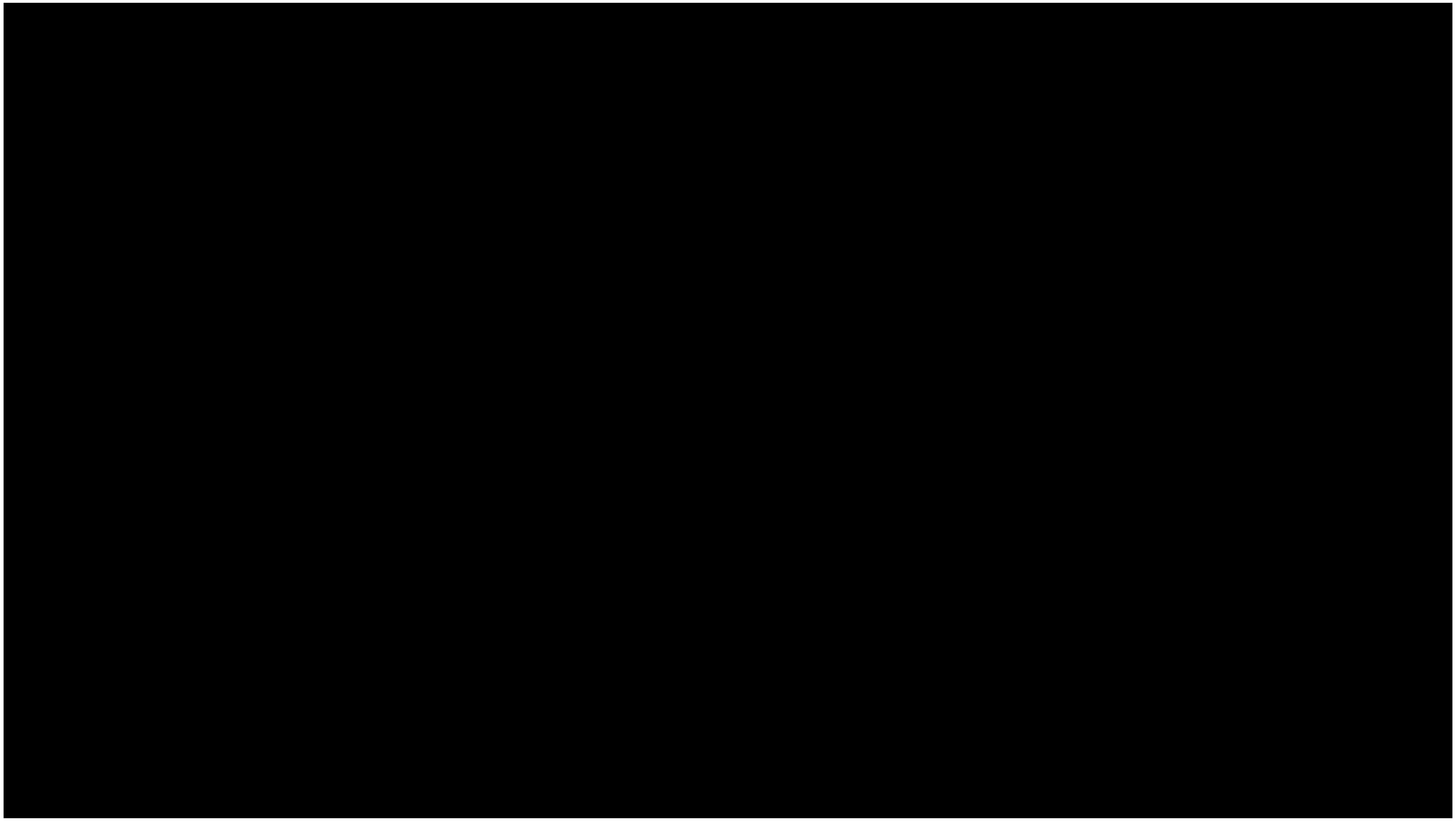
Junhyuk Oh et. al, NIPS 2015

Results on Atari



Junhyuk Oh et. al, NIPS 2015

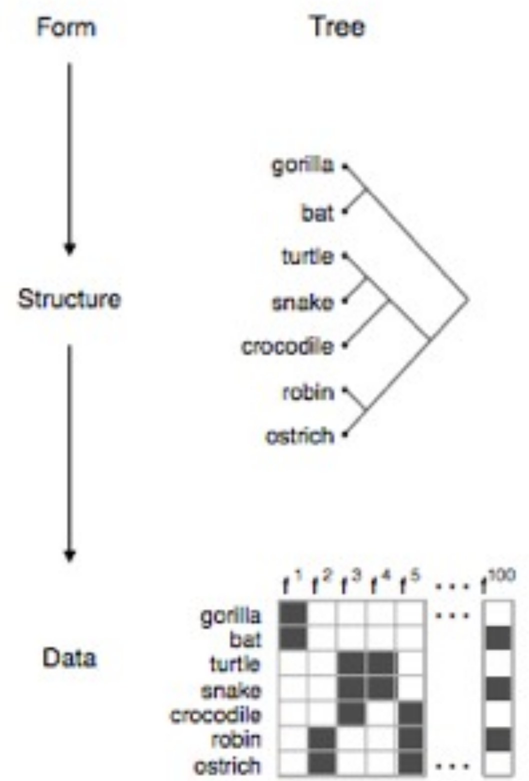
Results on Atari



Junhyuk Oh et. al, NIPS 2015

Generative models with growing structure

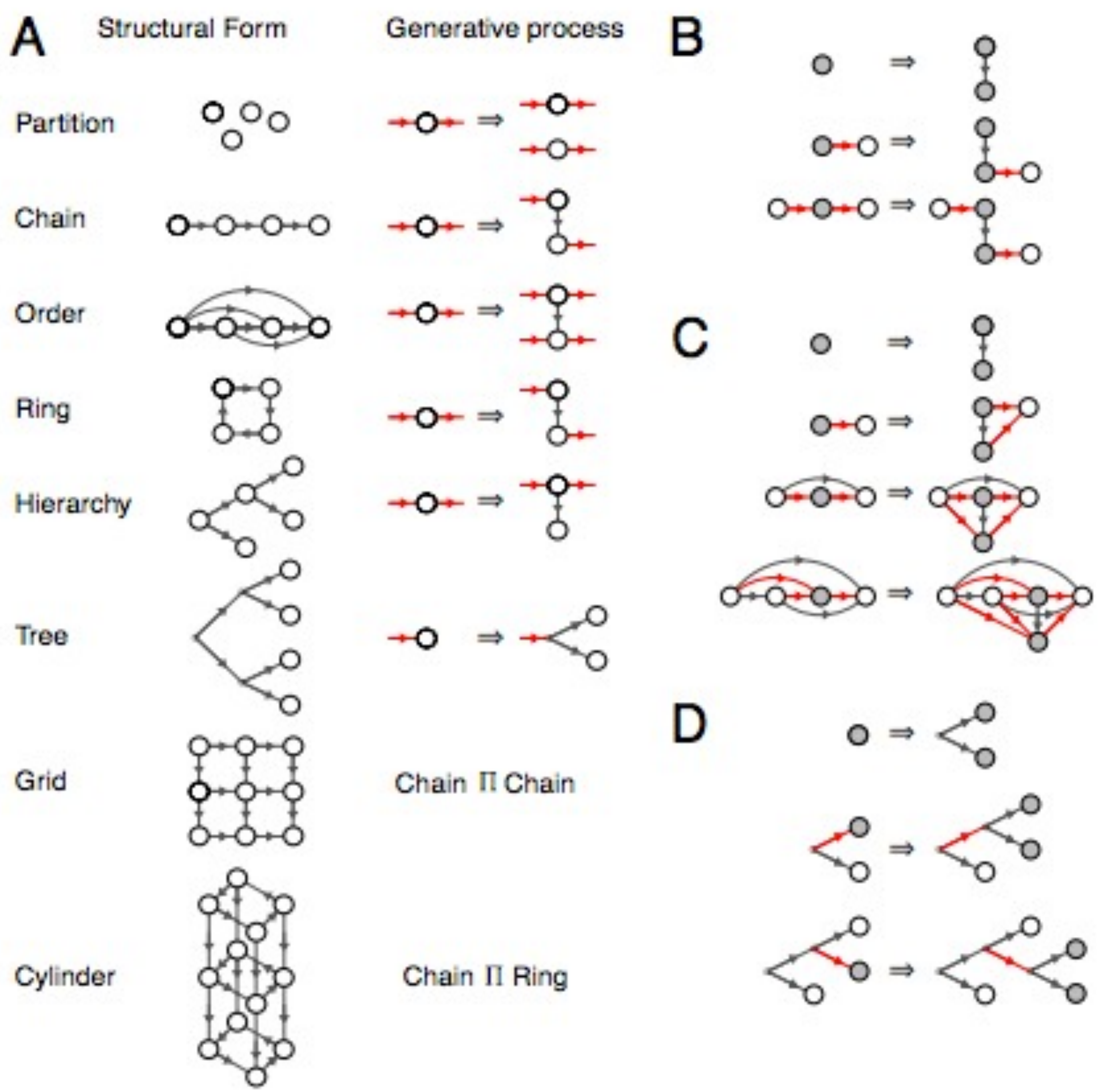
- Bayesian modeling allows us to move beyond parameter estimation to infer structure of the model itself
- Depending on the data, humans use different structural forms to form abstractions.
- Structure Learning can be cast as posterior inference problem to obtain the most likely model form/structure under observations



$$P(S, F|D) \propto P(D|S)P(S|F)P(F).$$

Kemp et. al., The discovery of structural form, PNAS 2008

Generative models with growing structure



Kemp et. al., The discovery of structural form, PNAS 2008

Generative models with growing structure

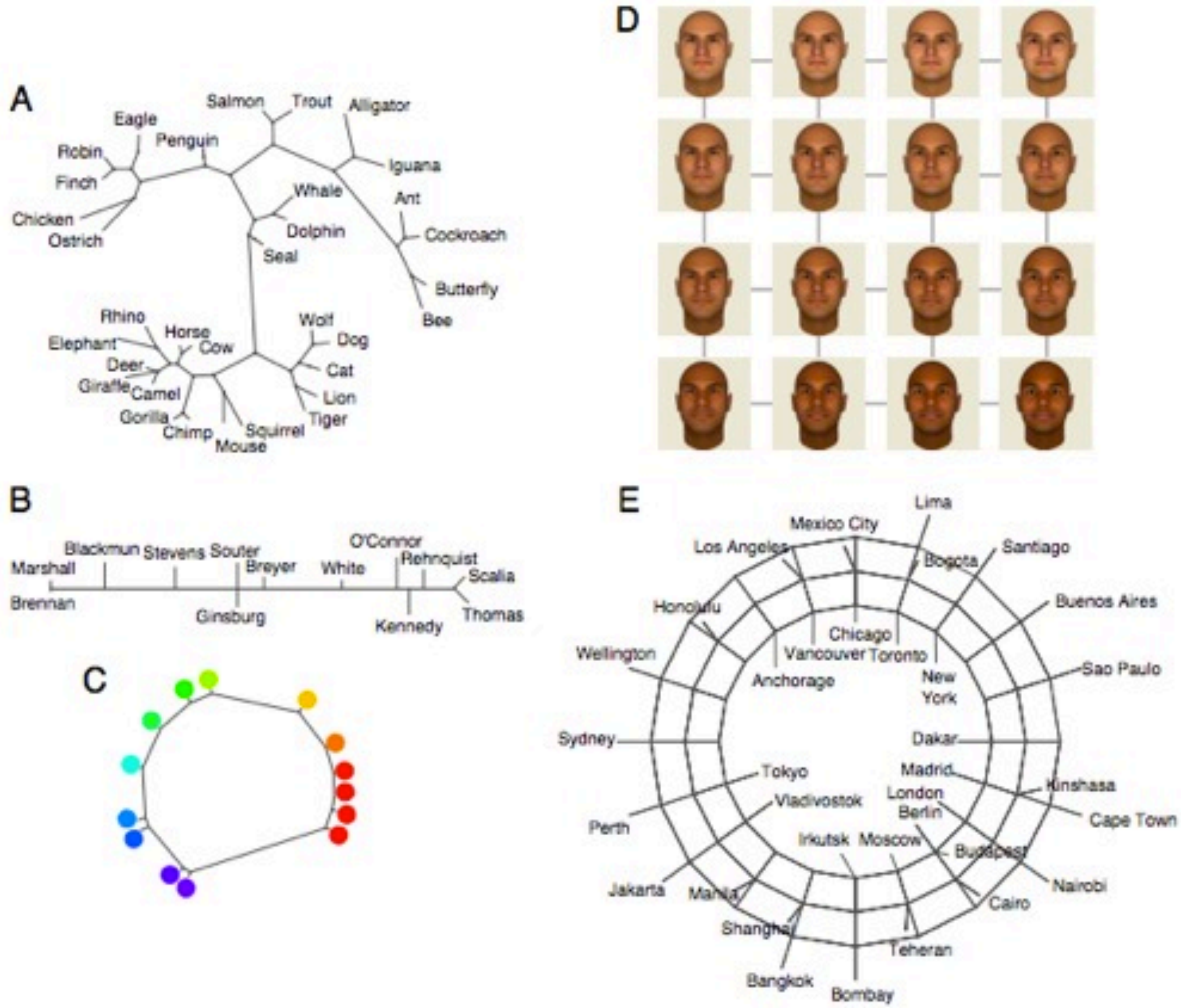
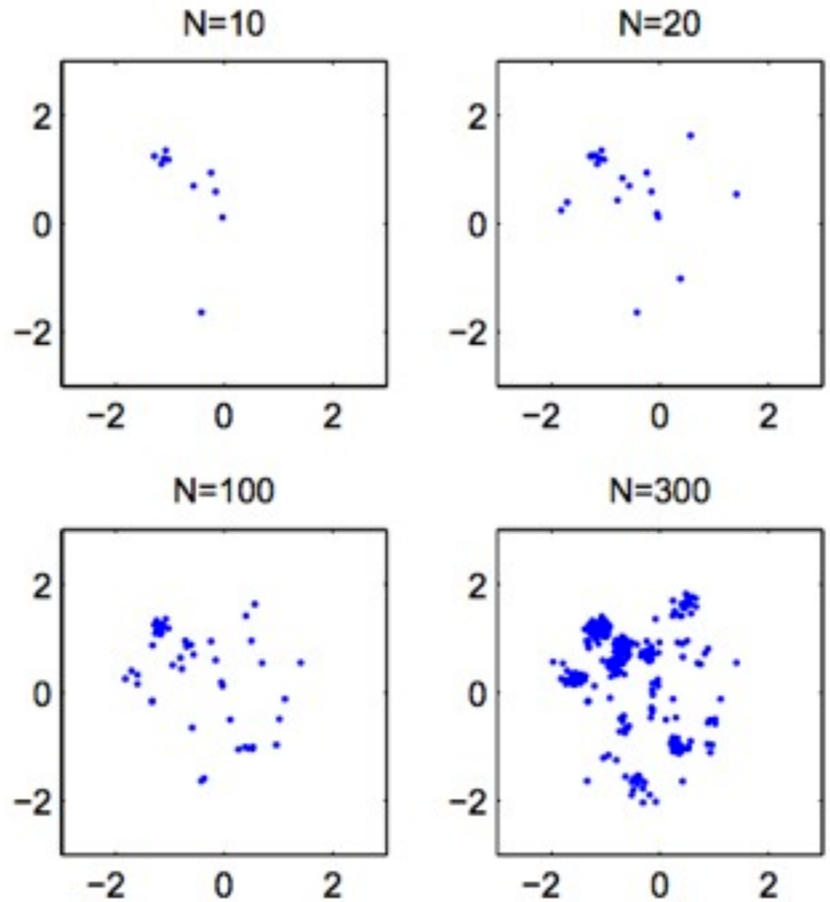


Fig. 3. Structures learned from biological features (*A*), Supreme Court votes (*B*), judgments of the similarity between pure color wavelengths (*C*), Euclidean distances between faces represented as pixel vectors (*D*), and distances between world cities (*E*). For *A*–*C*, the edge lengths represent maximum a *posteriori* edge lengths under our generative model.

Kemp et. al., The discovery of structural form, PNAS 2008

Generative models with growing structure

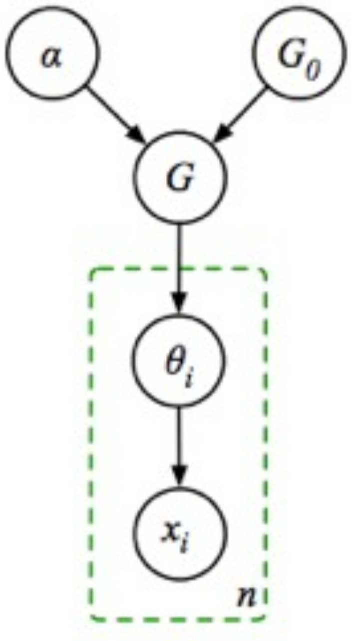
- Humans can grow abstractions in an arbitrary way to fit data
- Bayesian Non-parametric models naturally increase the number of parameters with data. Therefore no issue of over or under fitting.
- Many non-parametric processes: Chinese Restaurant Process, Gaussian Process, HDP, Indian Buffet Process etc.



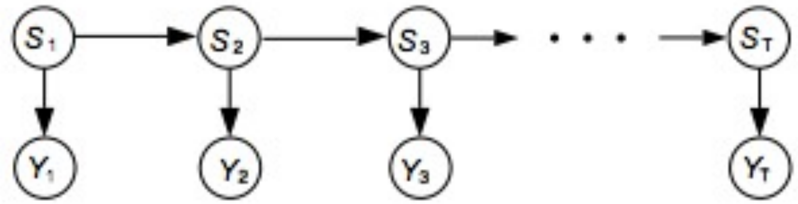
K = ?

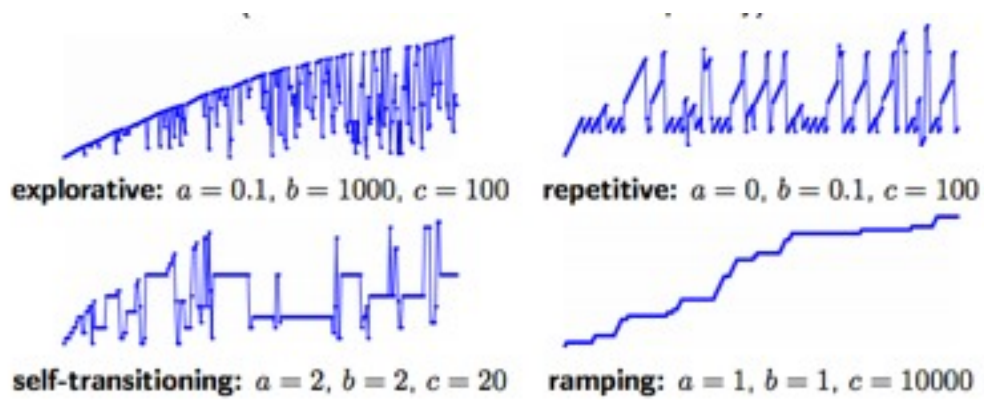
Ref: http://mlg.eng.cam.ac.uk/zoubin/talks/uai05tutorial-b.pdf

Generative models with growing structure



DP Mixture





Infinite
HMM

Ref: http://mlg.eng.cam.ac.uk/zoubin/talks/uai05tutorial-b.pdf

Questions?