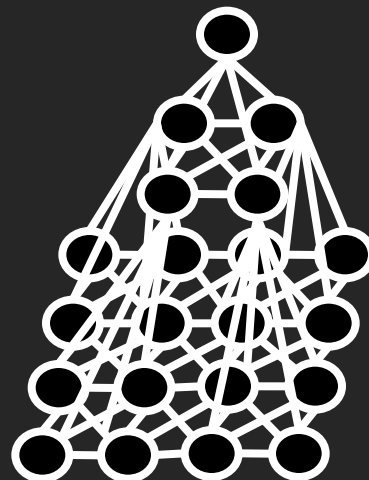




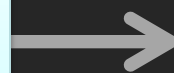
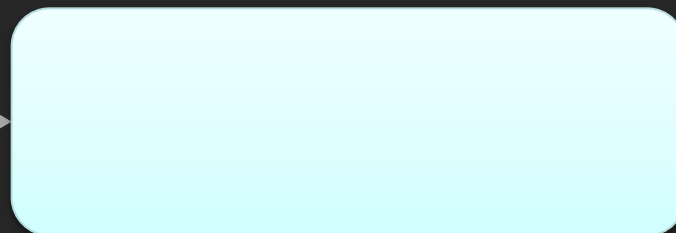
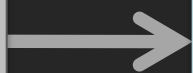
# Vision as understanding...

- Beyond perception for autonomous mobility: Generating descriptions that enable complex autonomous behaviors
- Naming objects and features
  - Identifying relations between them
  - Identifying activities and interactions
- Building block:
  - *Recognition and scene understanding*

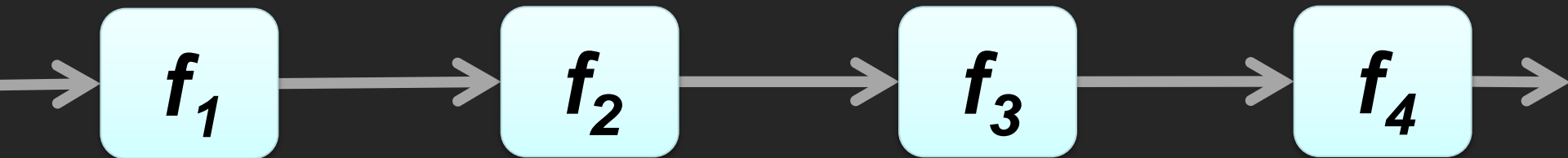
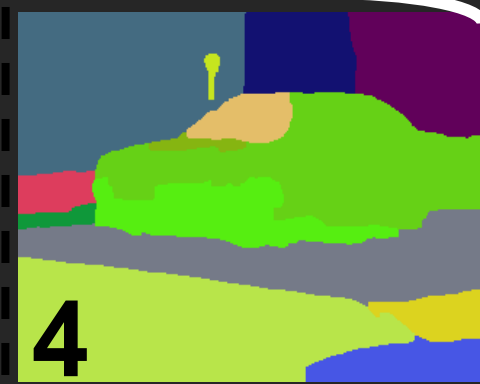
Efficient inference/learning  
tools *for perception*



**Input**

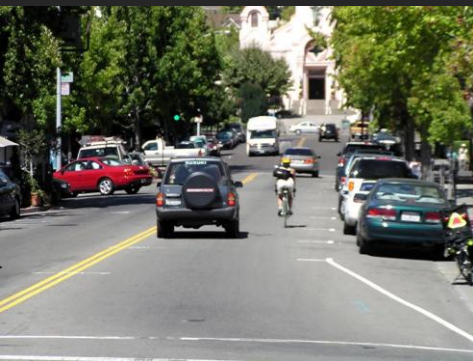


**Output**



Facing the uncertainty  
challenge?  
Dealing with time-bounded  
decisions?

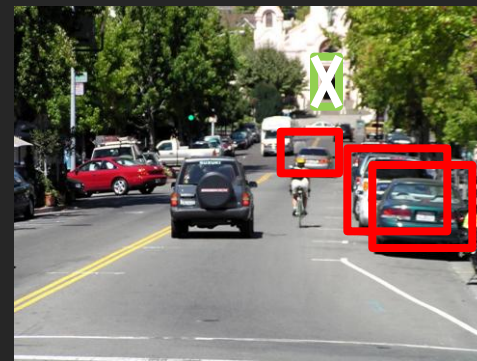
# Straight Interpretation Pipeline



Input image



Machine  
perception  
box



Labels



Planning  
Reasoning  
.....

# Performance?

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbik	pers
a) base	.336	.371	.066	.099	.267	.229	.319	.143	.149	.124	.119	.064	.321	.353	.407
b) BB	.339	.381	.067	.099	.278	.229	.331	.146	.153	.119	.124	.066	.322	.366	.423
c) context	.351	.402	.117	.114	.284	.251	.334	.188	.166	.114	.087	.078	.347	.395	.431
d) rank	2	1	1	1	1	1	2	2	1	2	4	5	2	2	1
(UofCTTIUCI)	.326	.420	.113	.110	.282	.232	.320	.179	.146	.111	.066	.102	.327	.386	.420
CASIA Det	.252	.146	.098	.105	.063	.232	.176	.090	.096	.100	.130	.055	.140	.241	.112
Jena	.048	.014	.003	.002	.001	.010	.013		.001	.047	.004	.019	.003	.031	.020
LEAR PC	.365	.343	.107	.114	.221	.238	.366	.166	.111	.177	.151	.090	.361	.403	.197
MPI struct	.259	.080	.101	.056	.001	.113	.106	.213	.003	.045	.101	.149	.166	.200	.025
Oxford	.333	.246					.291			.125			.325	.349	
XRCE Det	.264	.105	.014	.045	.000	.108	.040	.076	.020	.018	.045	.105	.118	.136	.090

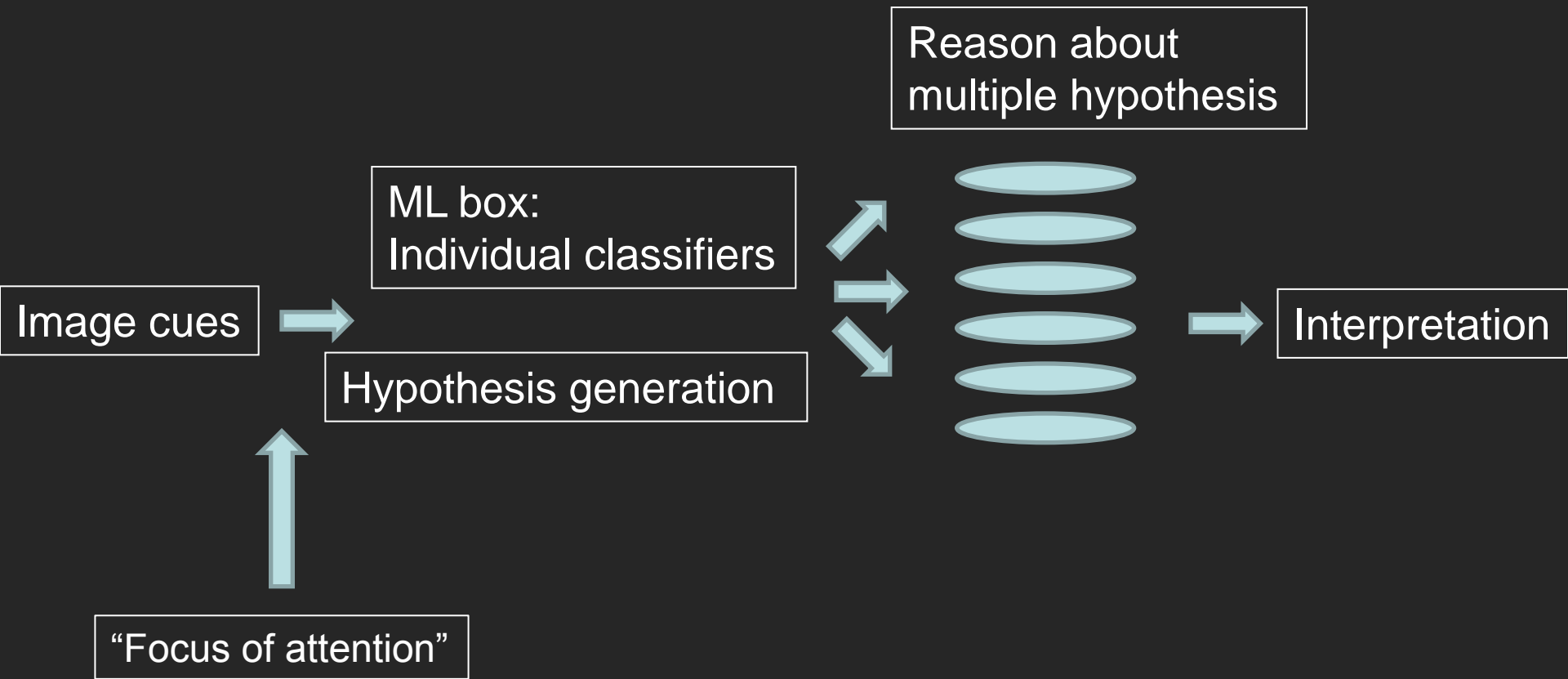
“We need a theory of performance guarantees”

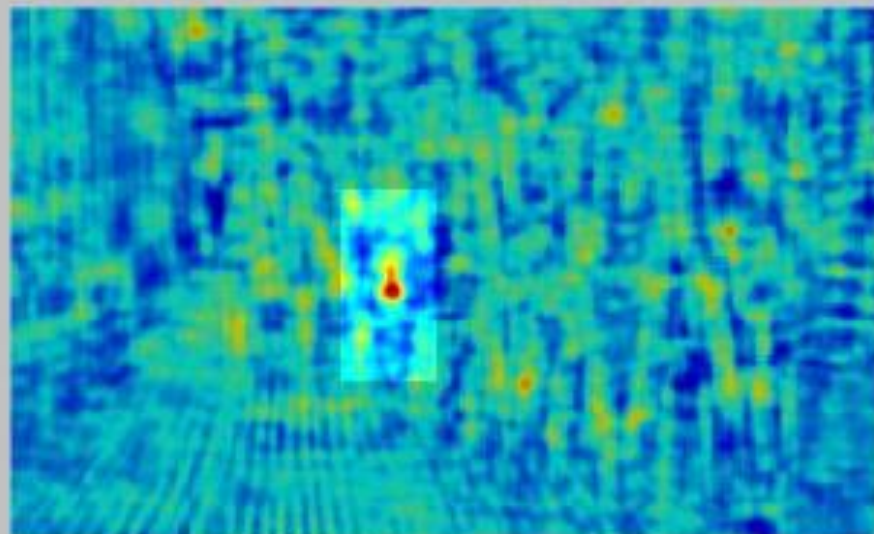
Stefano Soatto, August 22, 2011

“What is solved??” Don Geman, August 22, 2011

# Challenges

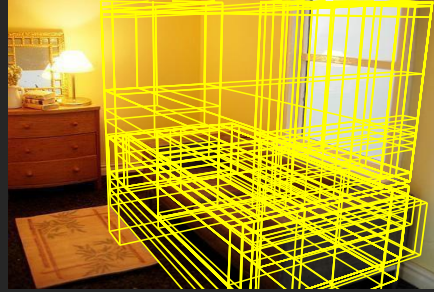
- Perception performance will always be limited:
  1. Provably impossible to drive down error rate to 0, in the absence of any other information
  2. The input maybe inherently ambiguous (given the training data and models)
  3. Intermediate (bounded computation) results are ambiguous
  4. Modeling everything that could be found in the data drives up the error rate
  5. It's wasteful and not necessarily useful to the task



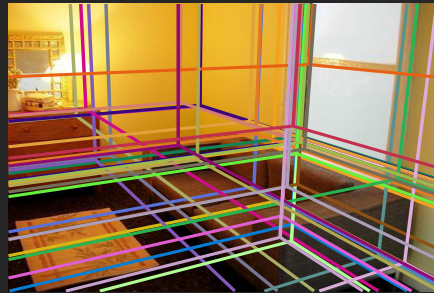




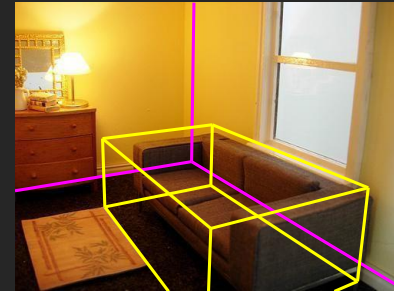
Input image



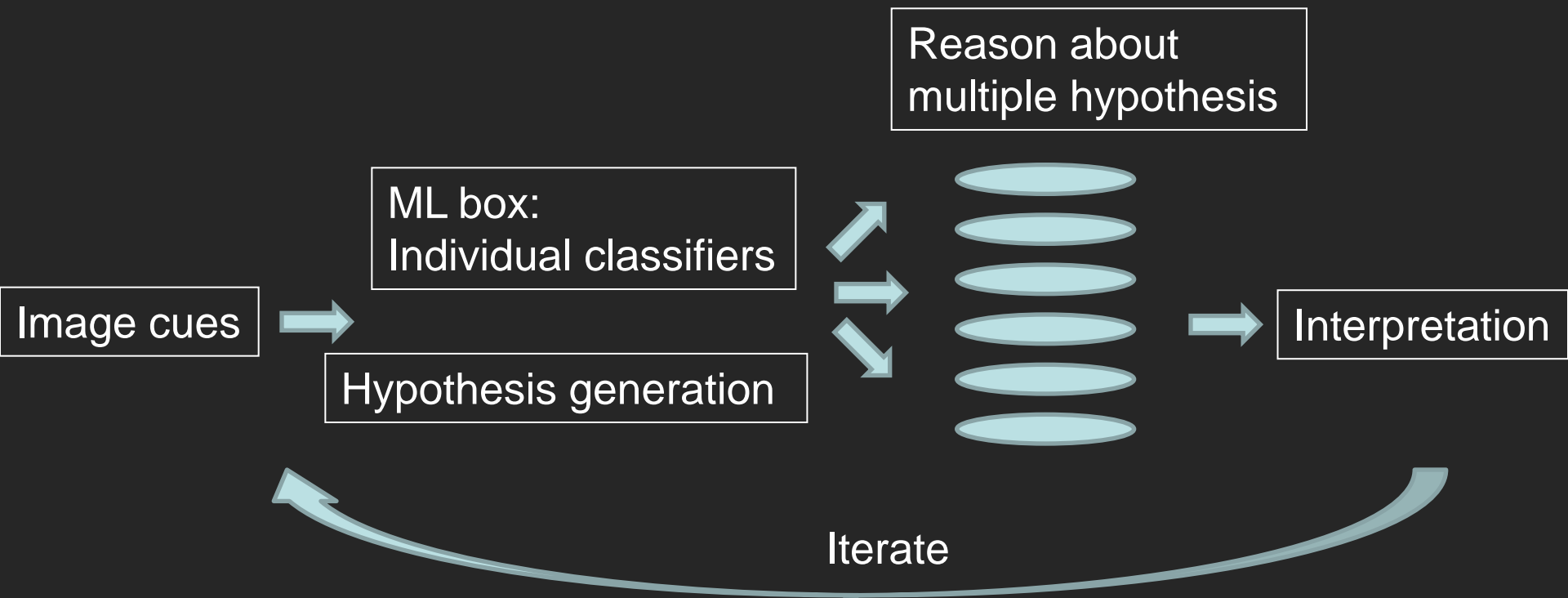
Object hypotheses



Spatial layout hypotheses



Final scene configuration



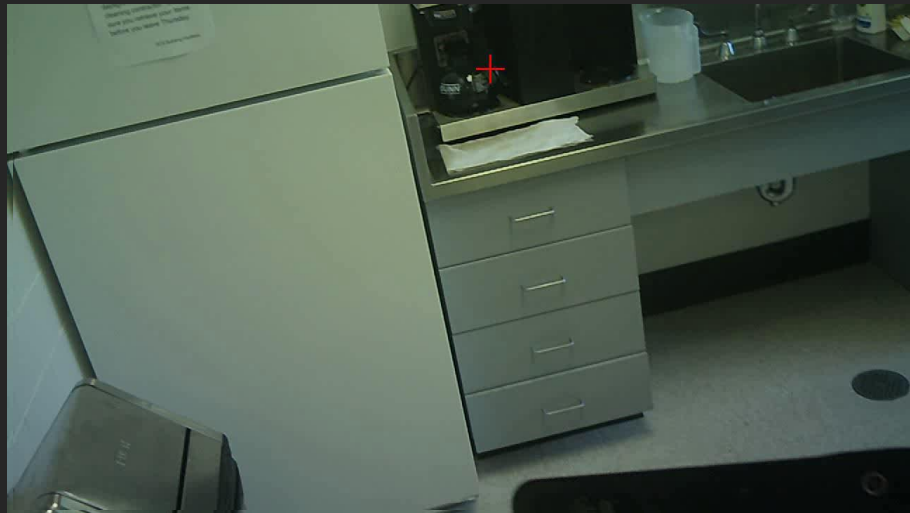
- Bounded computation and just in time output:
  - Refined interpretation available at any stage in the iteration

Big data



“Infinite” amount of stored visual data

“Unlimited” recording capabilities  
Lifelong learning

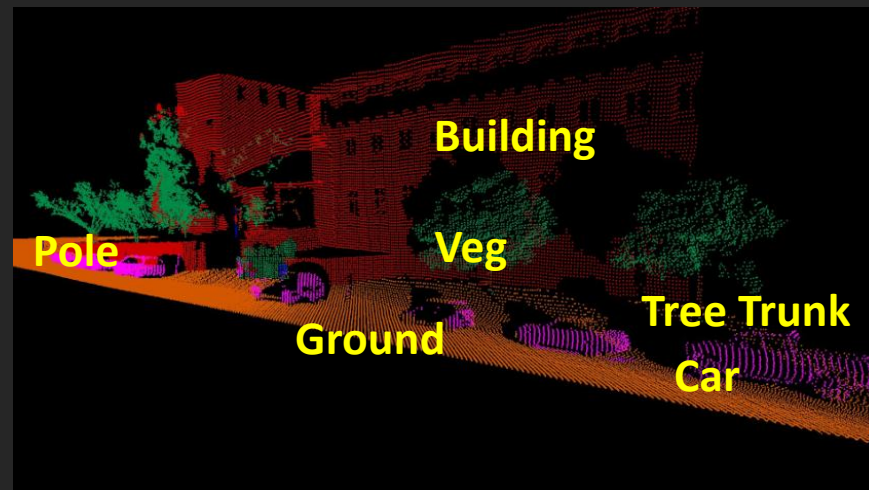
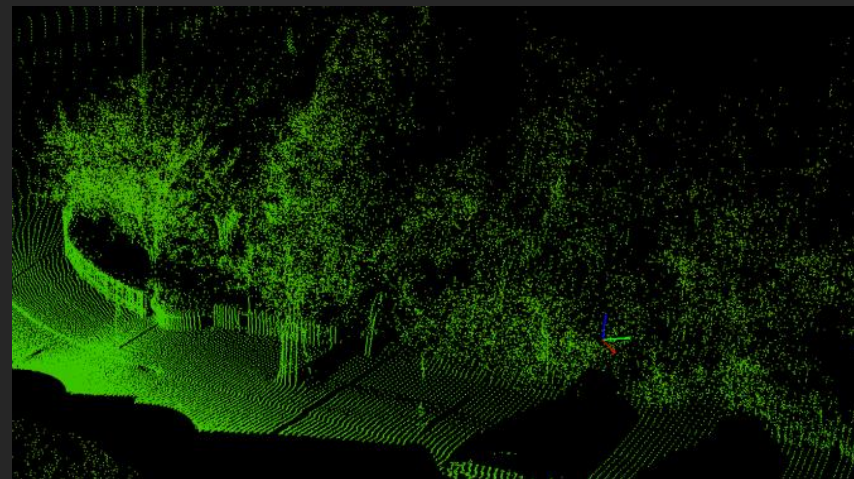
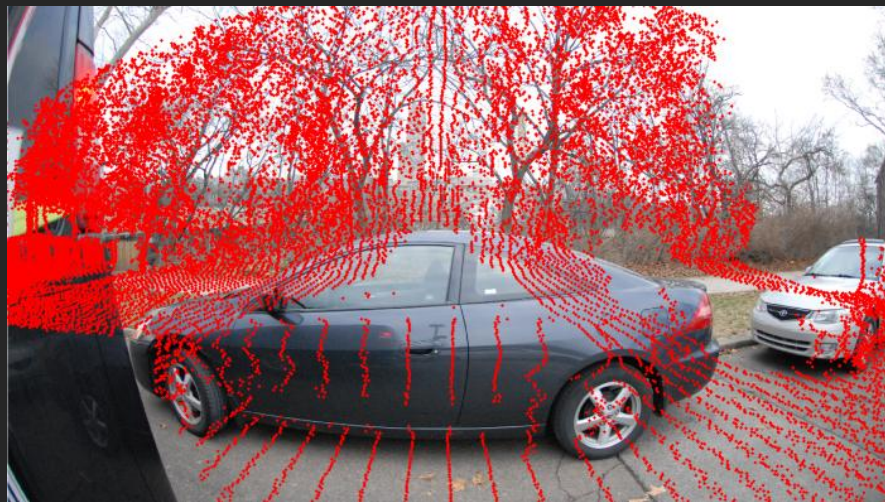


Transfer data between domains

Interacting with the environment

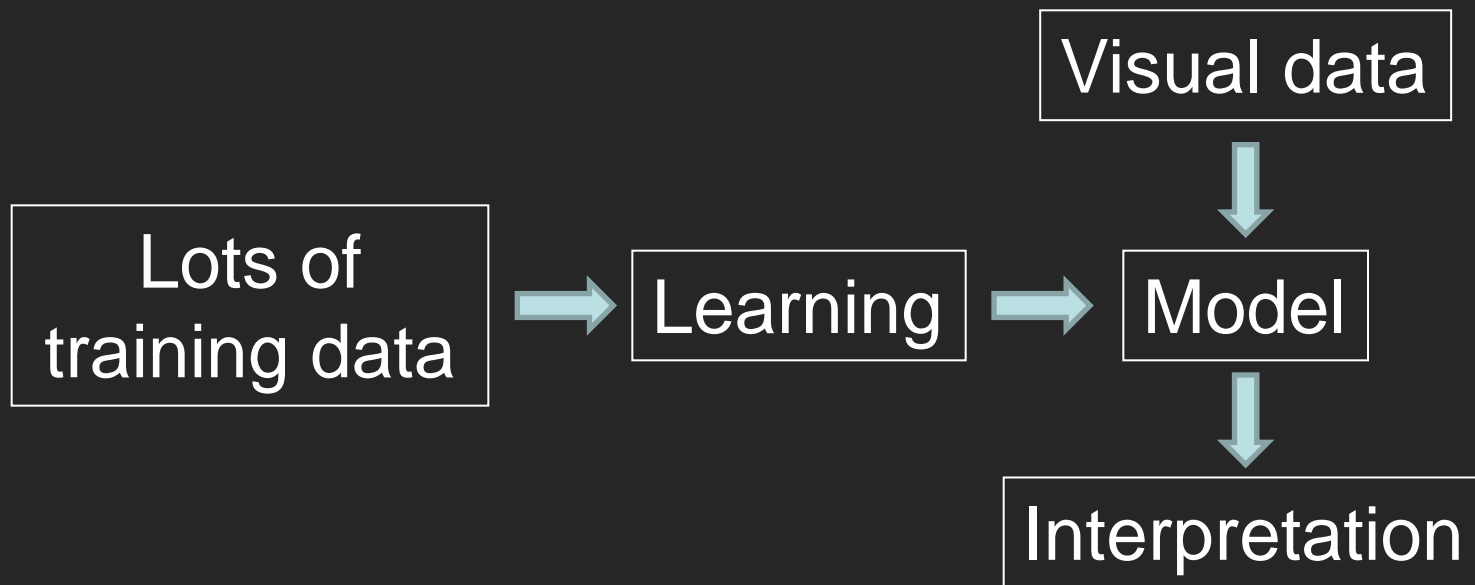


# Multi-modal



3D, text, ..., others?

Incorporating domain  
knowledge; Reasoning tools



- Many important pieces of knowledge do not fit easily in the statistical approach
- Physical and geometrical rules
- Cars drive on the right side of the road  
(or other prior knowledge about the local environment)
- Driving is forbidden in this area because of holiday  
(or other knowledge about current events in the local environment)
- Objects of type X may be threats and have high priority  
(or other knowledge about mission and intelligence pieces)

- Efficient inference/learning tools
- Facing the uncertainty challenge?  
Dealing with time-bounded decisions?
- Big data
- Multi-modal
- Incorporating domain knowledge;  
Reasoning tools