Predictive Vision Model - a different way of doing deep learning

Filip Piekniewski
Koh Young Research America (KYRA), San Diego CA
(Previously Brain Corporation, San Diego CA)
Feb 2017

Summary

- **Where does deep learning succeed and where does it fail**
  The quest for Artificial Intelligence and elephants in the room
- Predictive vision model - alternative approach
  The predictive vision model as a way to bring machine learning to the dynamical, physical reality. Perception as prediction.
- Why it works
  Machine learning meets the issues of scalability, parallelism, real time execution, independence of particular implementation
- Next steps
  A path towards a truly “physical” machine learning implemented on a cortical processor

Artificial Vision Systems: many classes

- Custom Computer Vision algorithms
  Limited to narrow applications and not scalable
- Biologically detailed simulations of visual system
  Parameter hell, most of the time simply don't work
- Machine Learning algorithms
  Vaguely related to biology but typically work. High hopes, particularly with:
- Deep Learning (i.e., convolutional neural networks) recently made practical by GPUs
  Excellent performance on classifying photos taken by humans. Facebook and Google recently hired pioneers LeCun and Hinton (See LeCun, Bengio, & Hinton (2015) for a review in Nature).
  Somewhat narrow class of algorithms resembling neocognitron. Not suited for robotics and other online real world applications.

Shuffled images ->

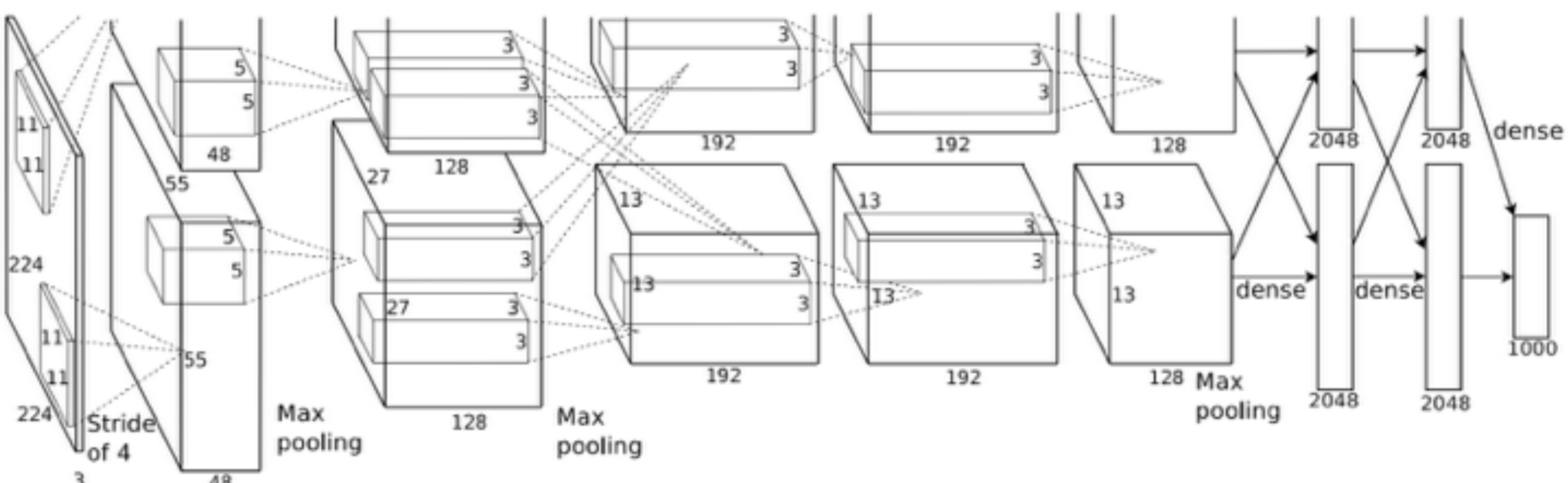
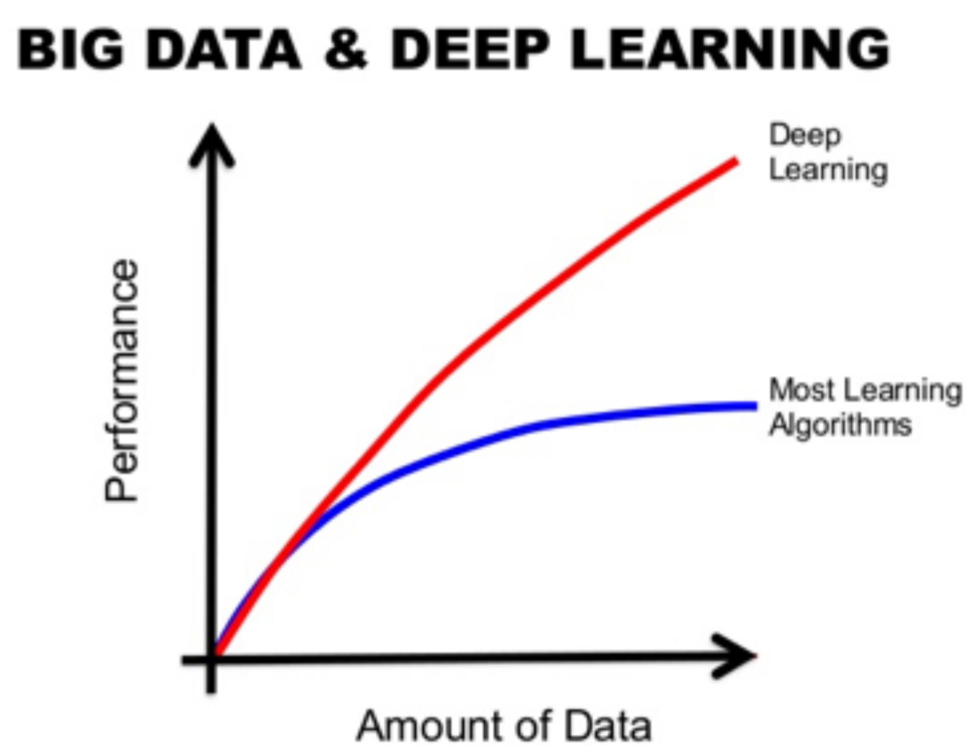


Image class <- labels

"Alexnet" by Krizhevsky, Sutskever & Hinton (2012)

Success of deep learning

- Deep learning has vastly improved object recognition in visual domain (e.g. ImageNet)
- Deep learning has improved speech recognition and machine translation



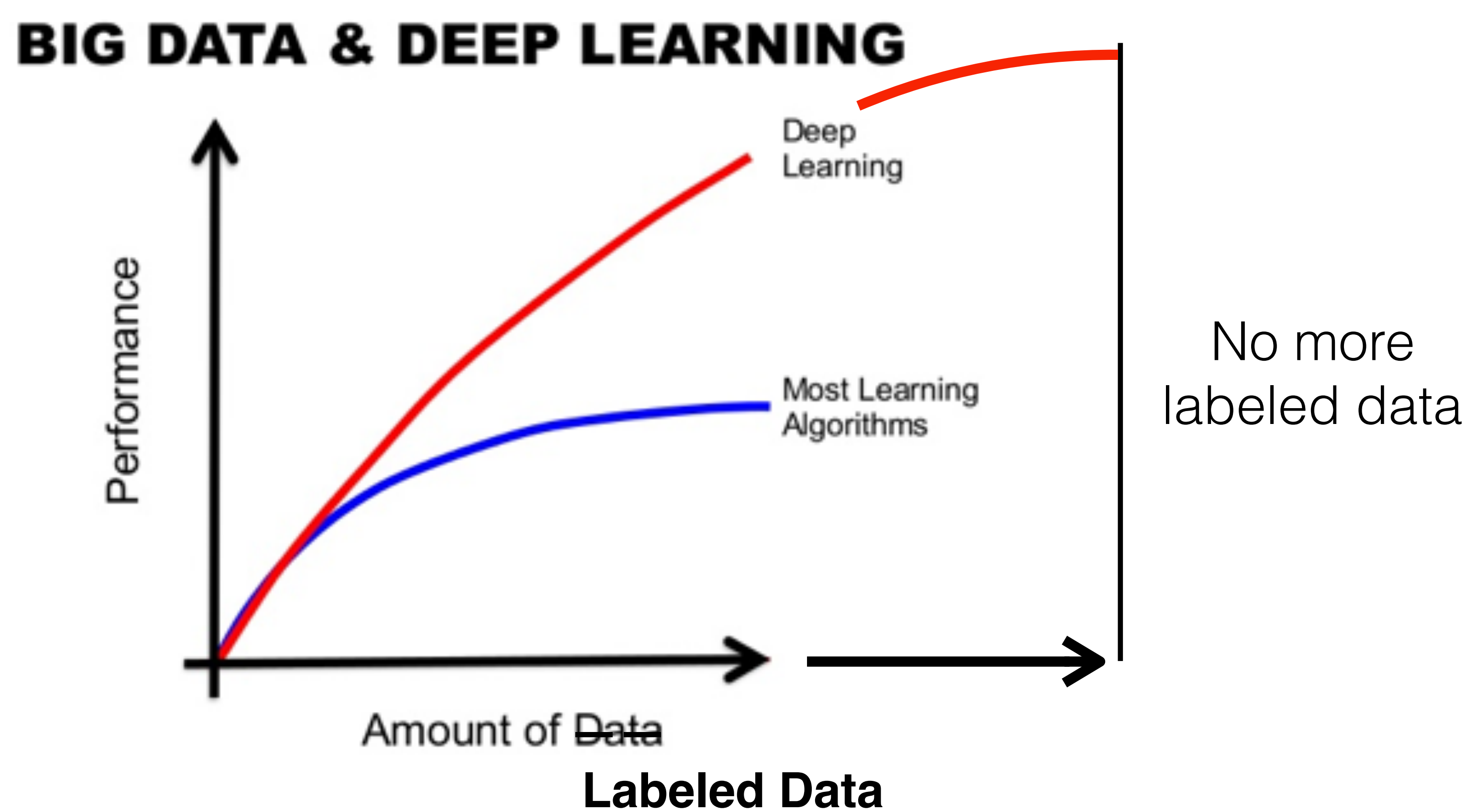
But…

- Deep learning has not yet enabled robotics
- In fact aside from clear benefits that the big data companies gained (such as Google or Facebook), deep learning is yet to find a big market application:

- Although impressive in some tasks, many real world applications remain out of reach

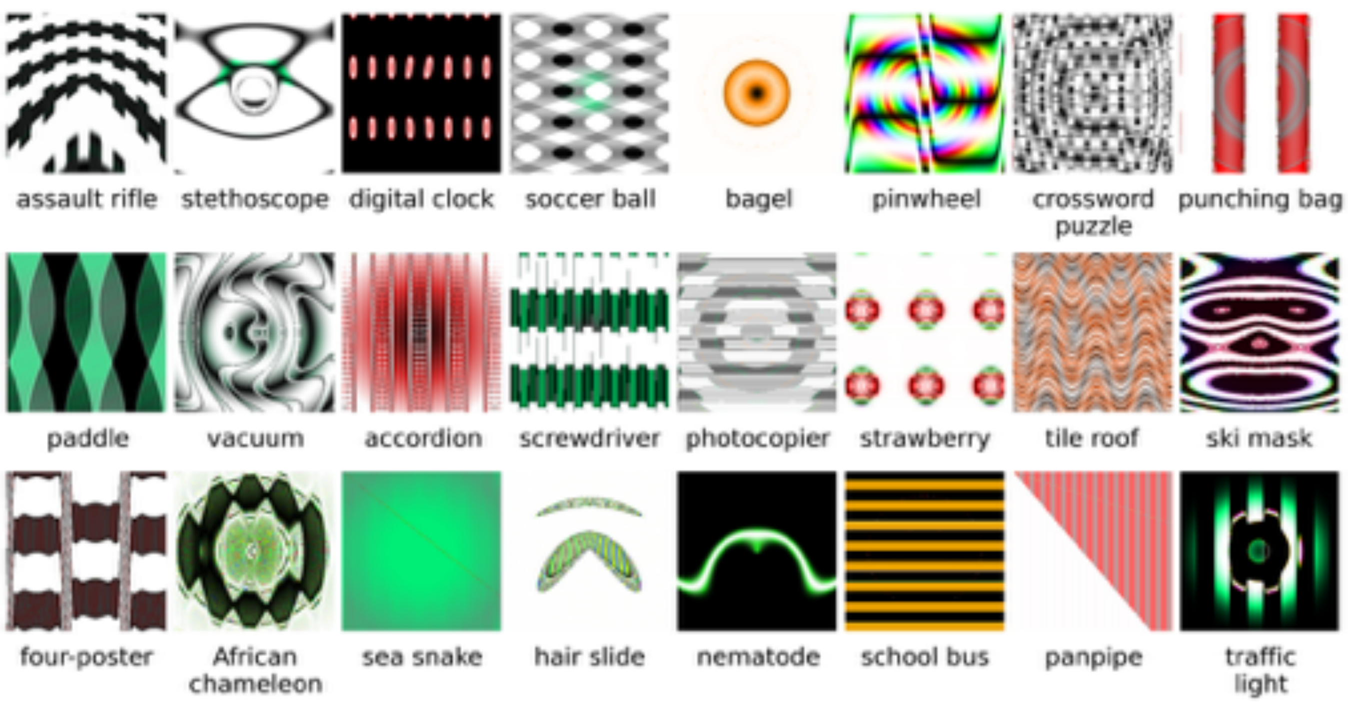
But

- Deep learning is currently primarily focused on supervised or reinforcement learning
- Both paradigms need a lot of labeled data. Reinforcement learning needs so much data that it is only effective in games/simulations.

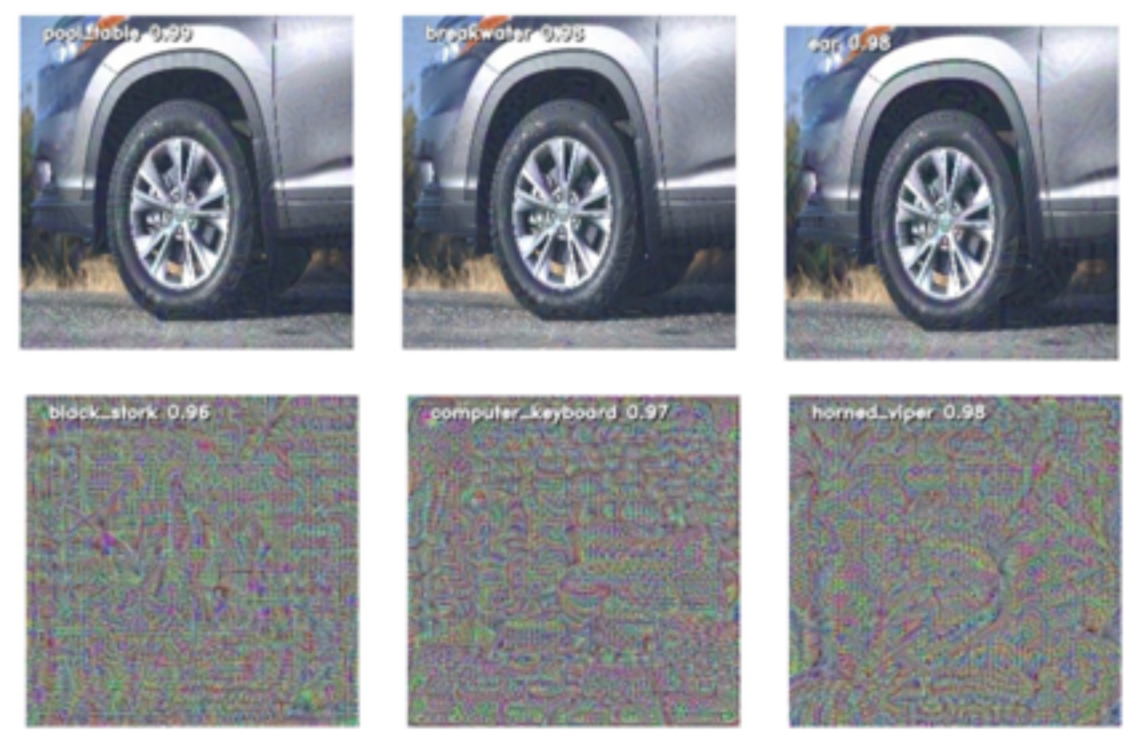


Artificial Vision Systems

- Deep Learning (i.e., convolutional neural networks)
  - Can be easily "fooled" (Nguyen, Yosinski, & Clune 2014)
  - Can have very rough category boundaries (Szegedy et al. 2013)
    - Theoretically "more data" could solve these problems... but not practically



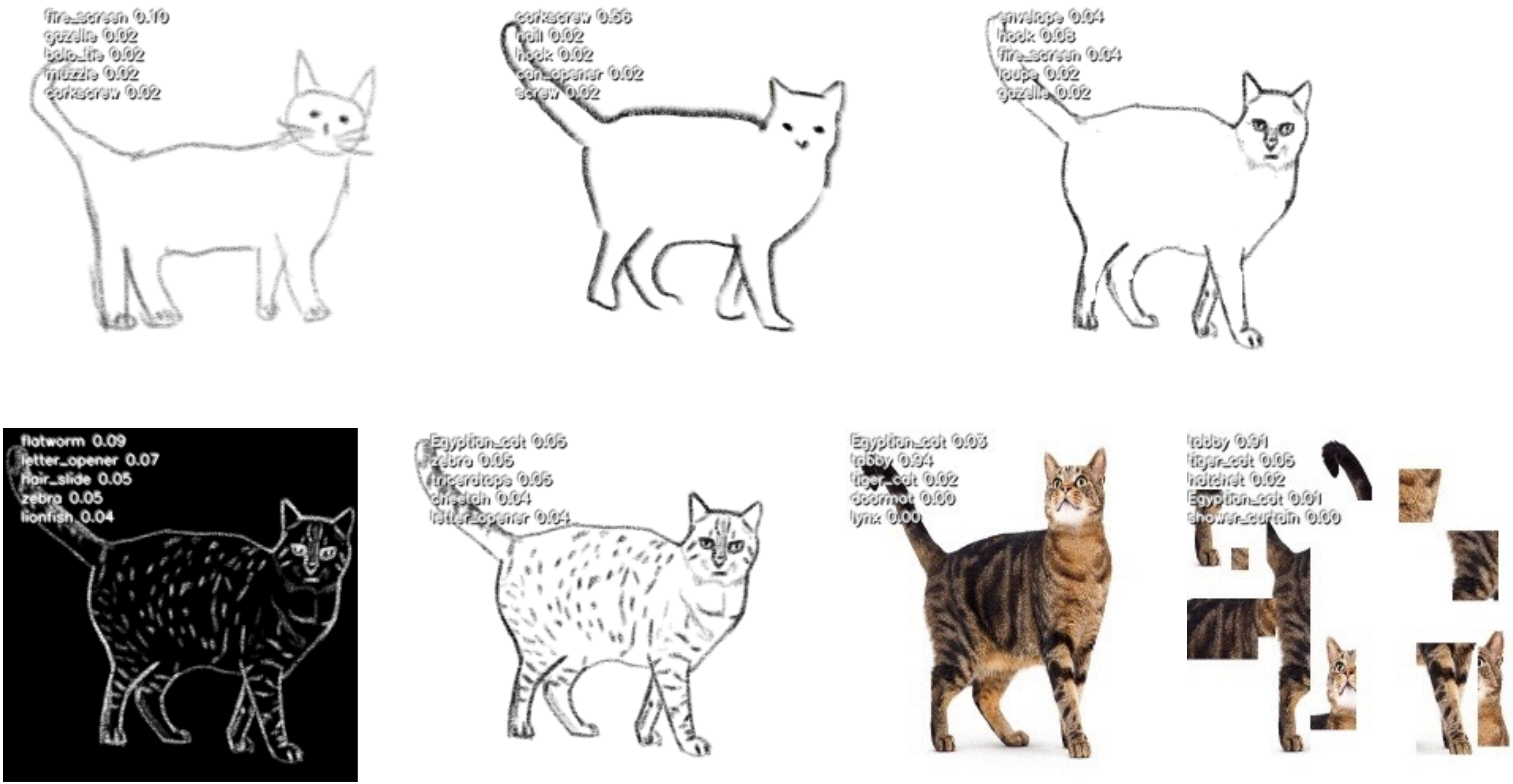
99.12% confidence



Subtle differences, large errors

Artificial Vision Systems

- Deep Learning image recognition relies on textures not outlines



Category: "Airplane"

- Labelling these images in a robust, generalizable way requires:
  - Contextual knowledge
  - Functional knowledge
  - Cultural knowledge
- In other words “**common sense knowledge**” !
- Rote memorization is an option (especially for a system with many parameters) but is under defined and doesn't generalize.




Source: ImageNet

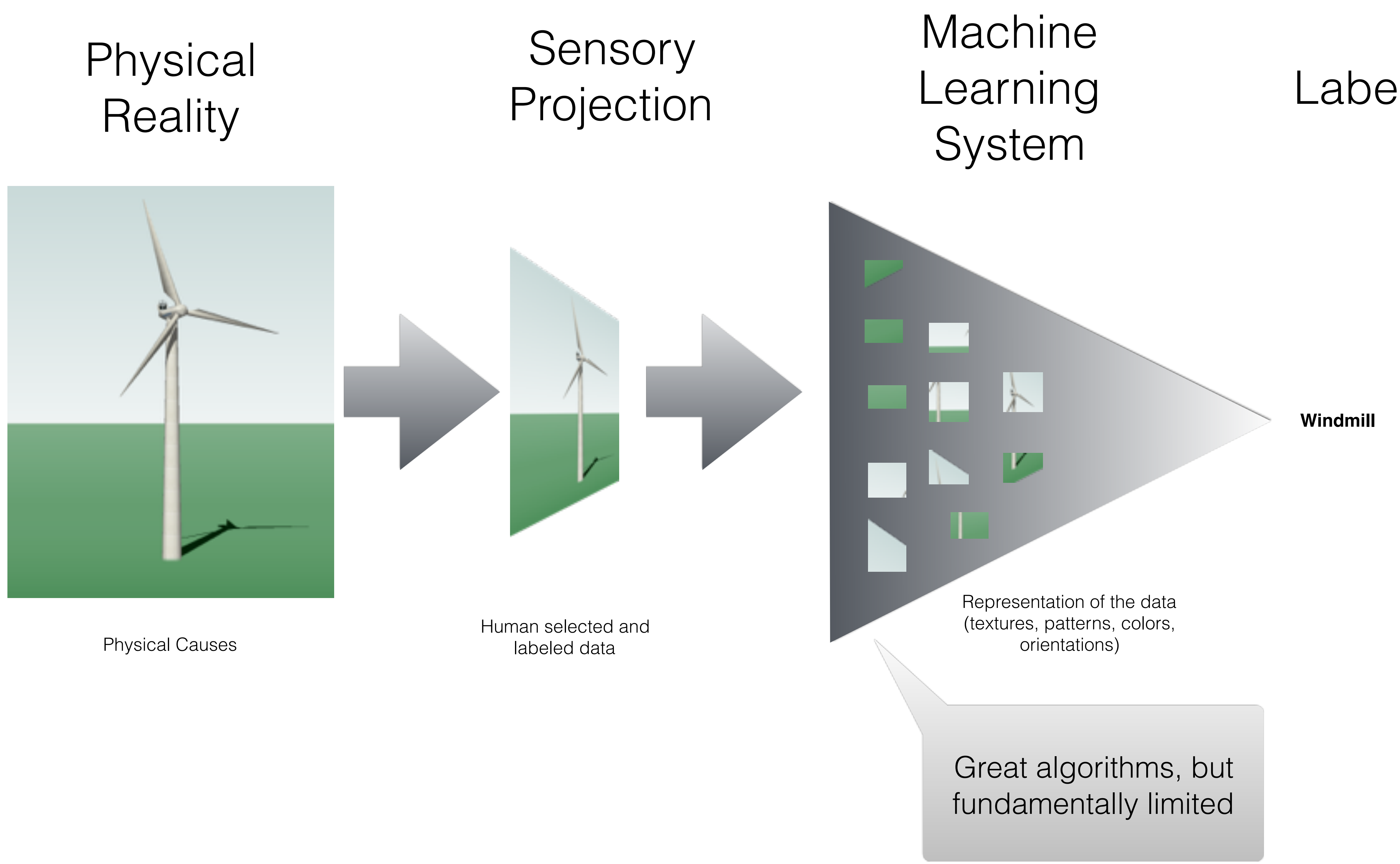


Summary

- Where does deep learning succeed and where does it fail
  The quest for Artificial Intelligence and elephants in the room
- **Predictive vision model - alternative approach**
  The predictive vision model as a way to bring machine learning to the dynamical, physical reality. Perception as prediction.
- Why it works
  Machine learning meets the issues of scalability, parallelism, real time execution, independence of particular implementation
- Next steps
  A path towards a truly “physical” machine learning implemented on a cortical processor

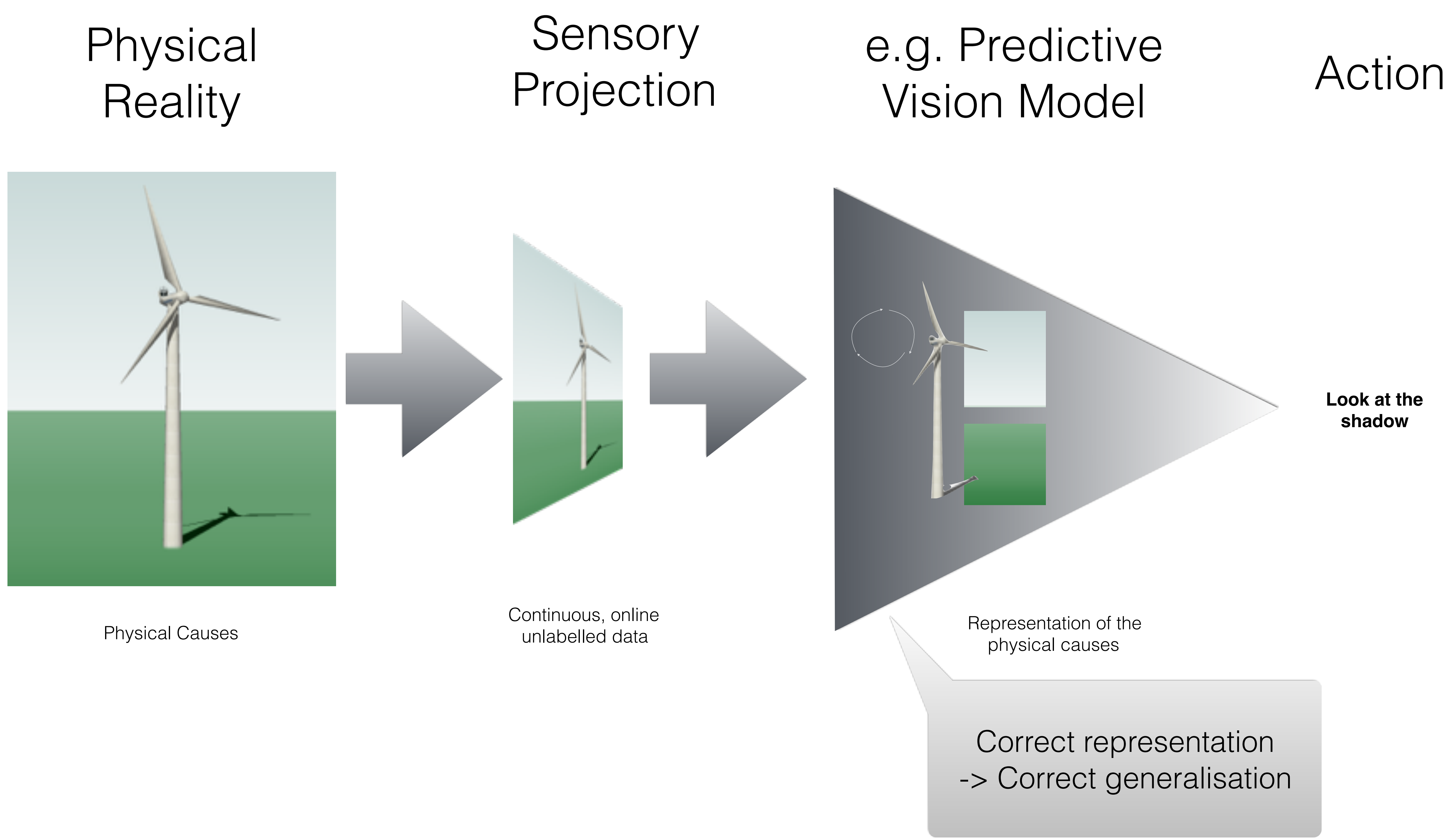
Perception should represent reality

State of the art in ML

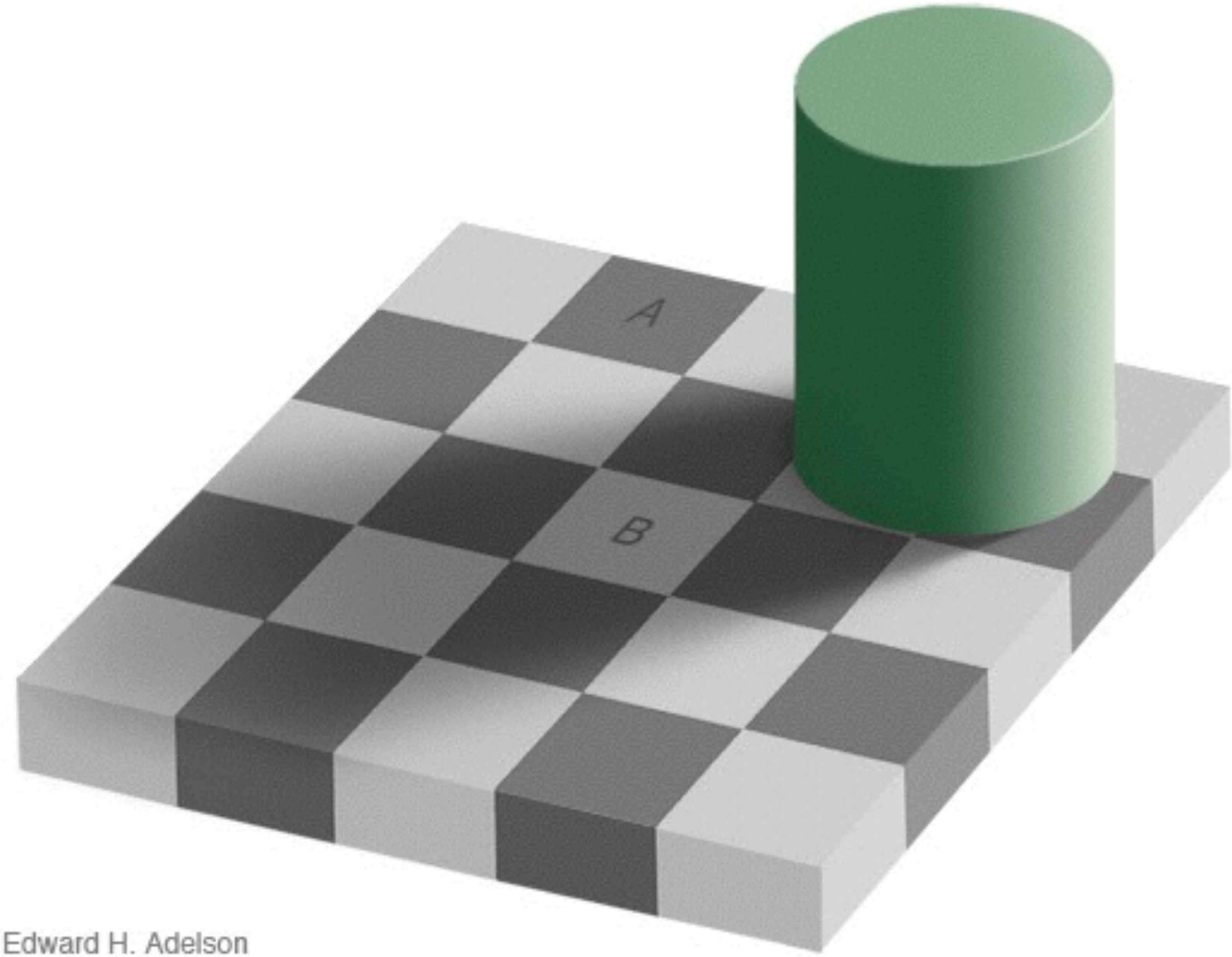


Perception should represent reality

Where ML should go

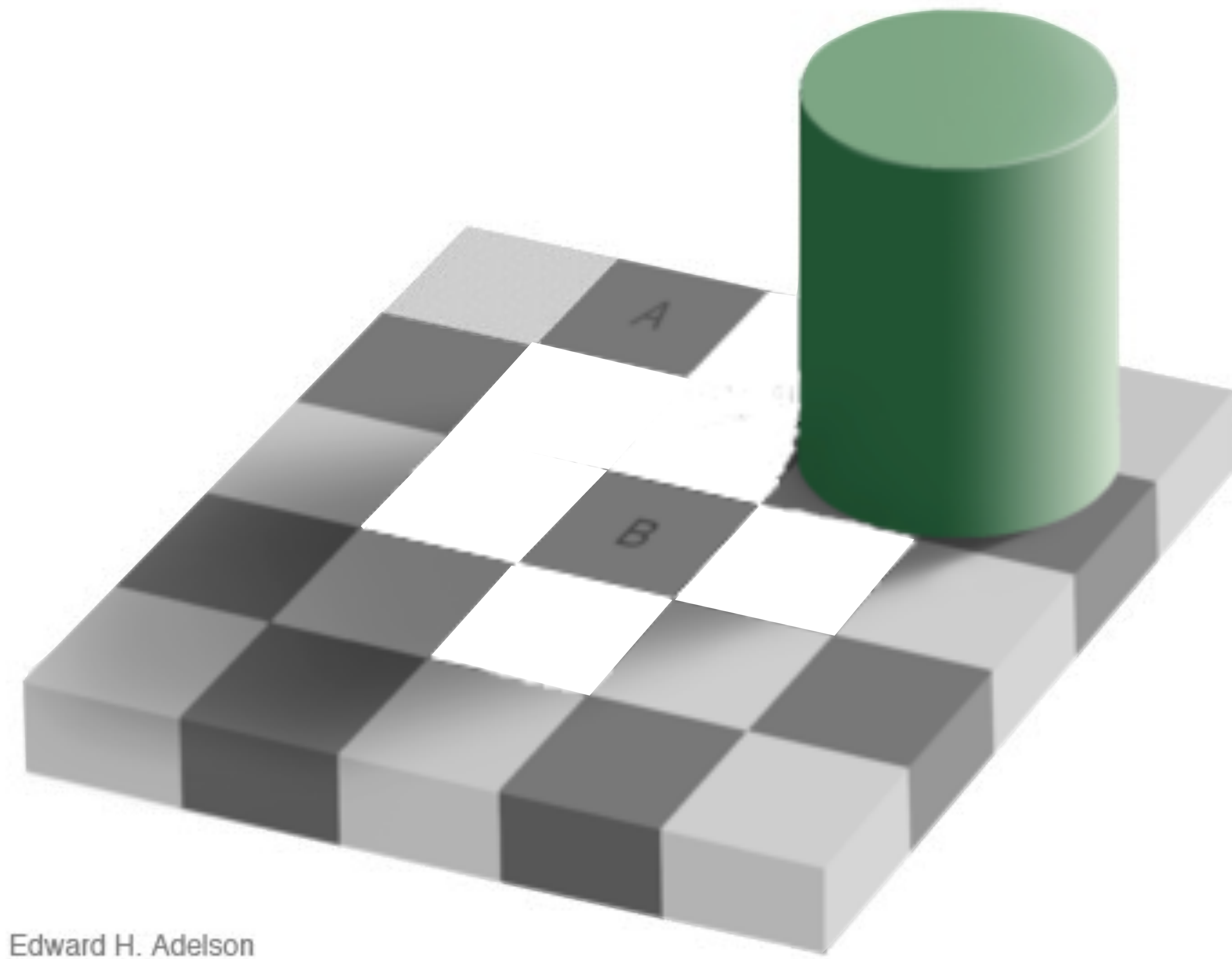


Perception depends on (physical) context



Edward H. Adelson

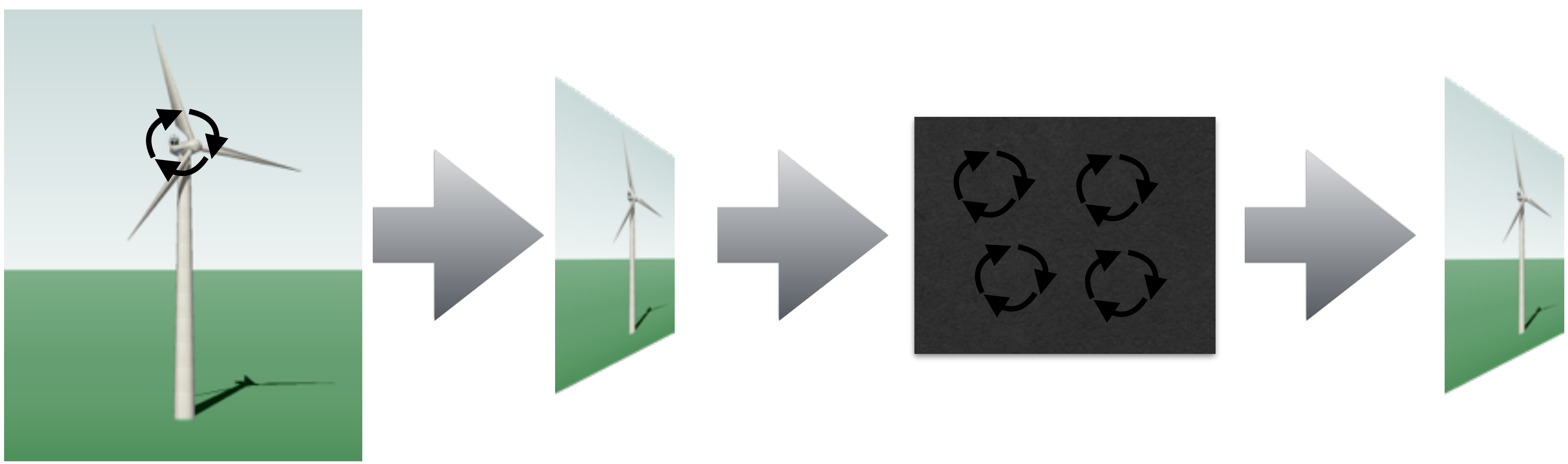
Perception depends on (physical) context



Perception should represent reality

How to build an AI for physical reality?

Physical Reality | Sensory Projection | Model | Prediction



Physical reality (dynamics)

Apparent view of the unfolding dynamics

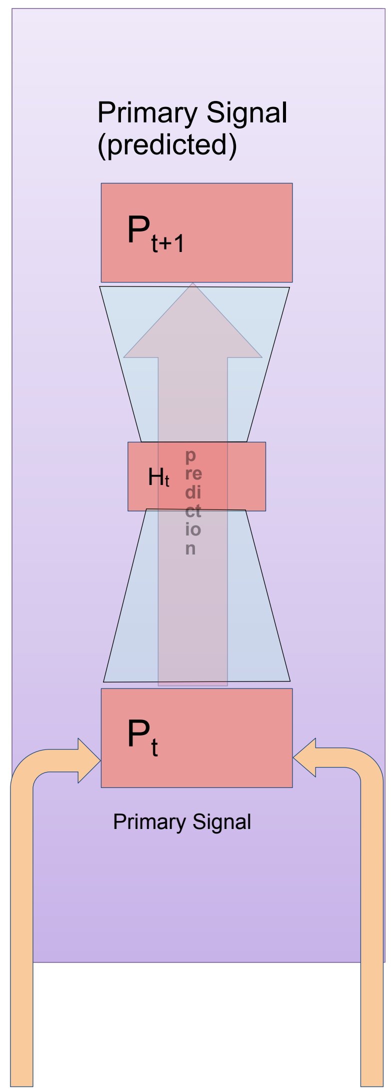
Prediction of immediate future evolution of the perceived system

If I can make good predictions, I have a good model of the phenomenon I observe. Otherwise my model needs to be modified. Prediction error is the supervising signal!

PVM - principles

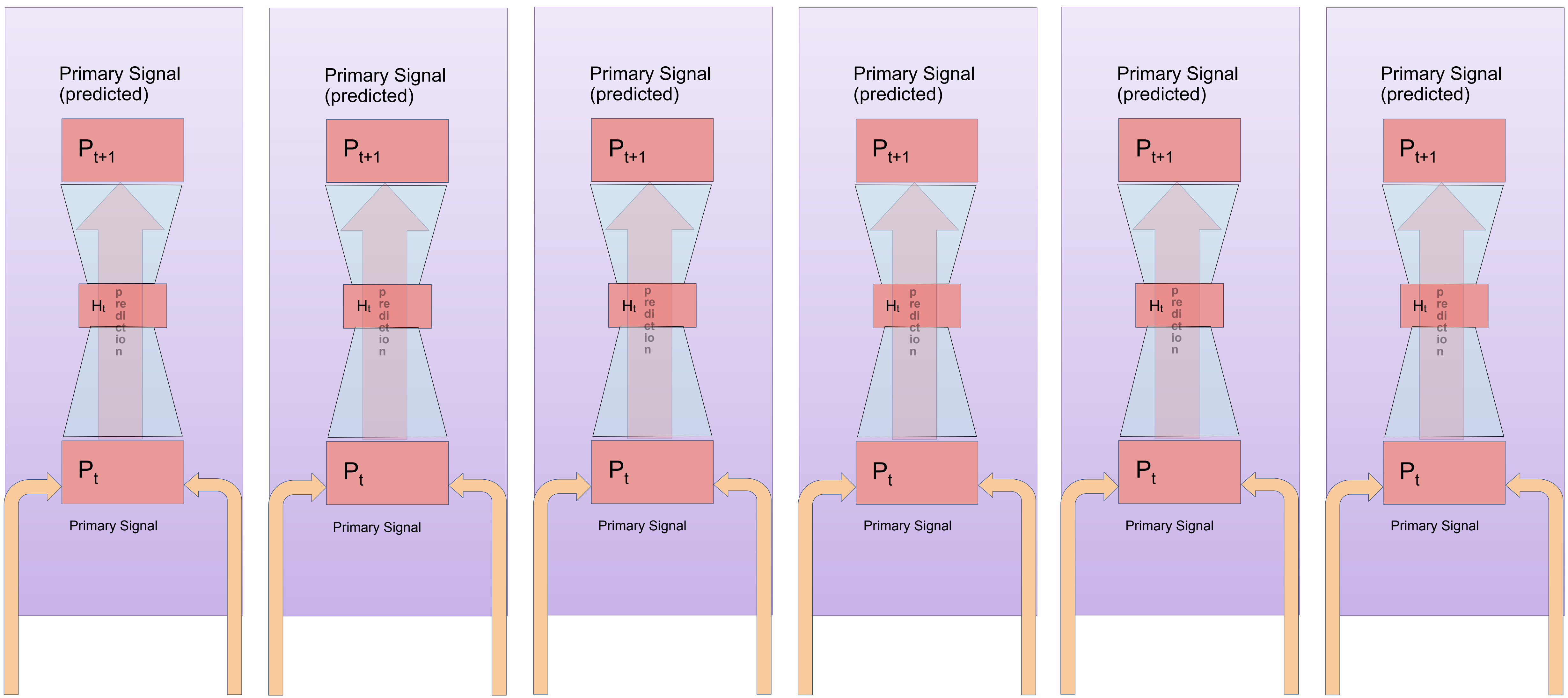
- Prediction error as the supervising signal
- Learning features at multiple levels of abstraction
- Recurrence since reality is complex
- Lateral interactions since in reality nearby phenomena affect each other
- Multi scale/scale free design, since in reality many scales are simultaneously intertwined
- Feedback across scales, since in reality phenomena at different scales affect each other
- Uniform parallel design for great scalability

Constructing PVM

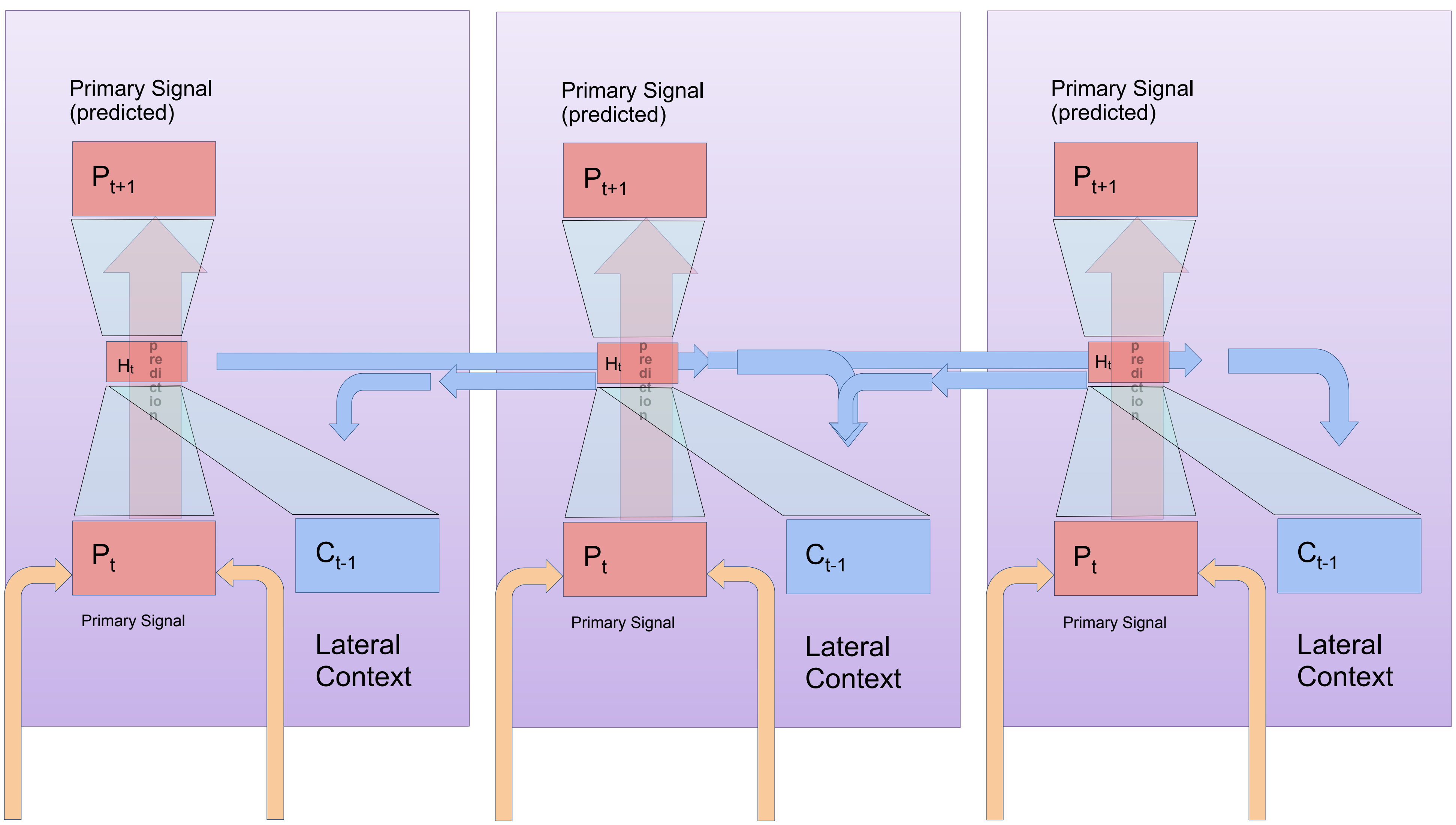


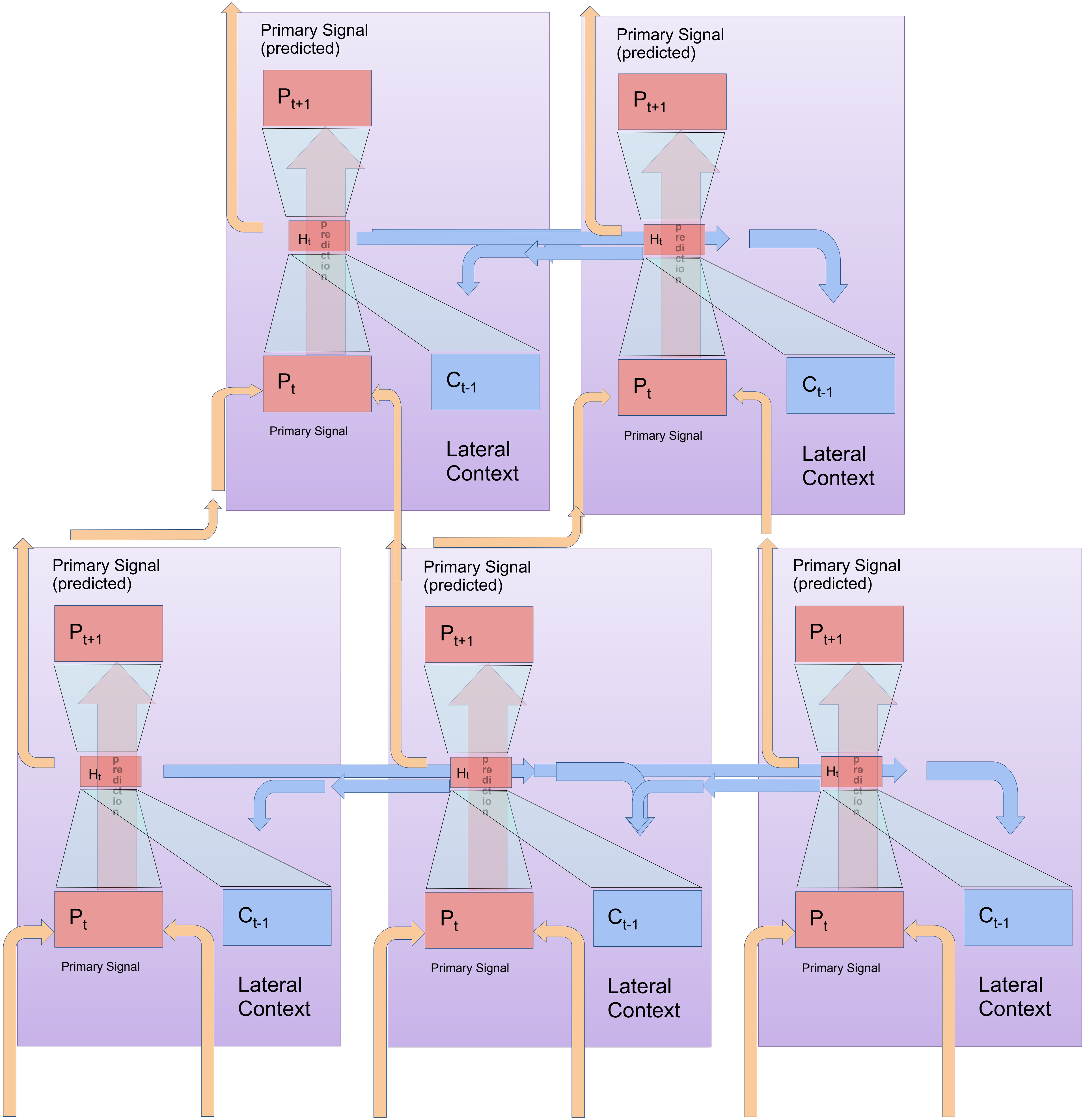
Associative memory module, binding signal now with signal in the near future. Maybe a “shallow perceptron”, “deep perceptron” or in fact anything else that is able to associate with a bottleneck (e.g. Boltzmann machine, spiking network)

Constructing PVM

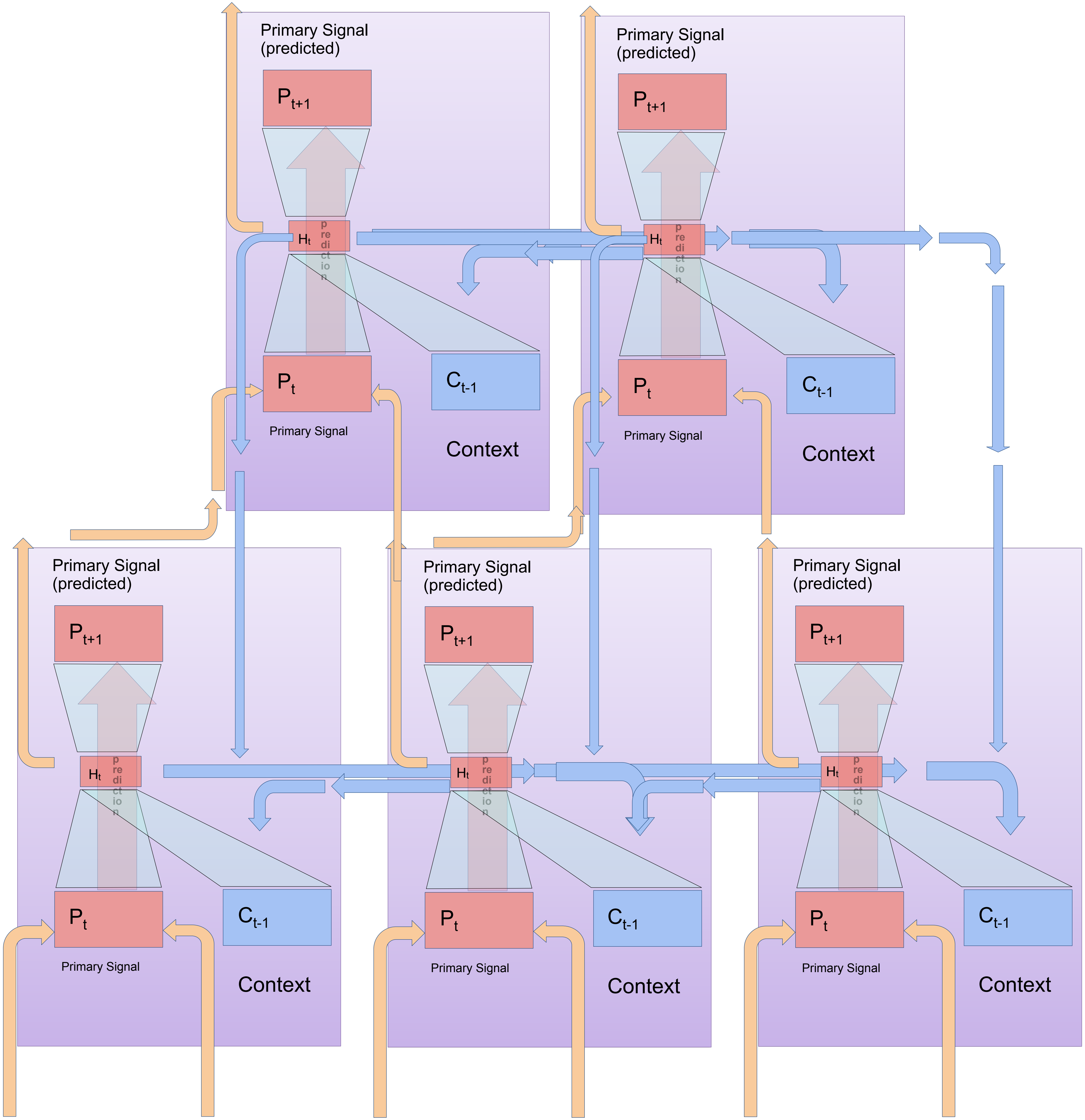


Constructing PVM



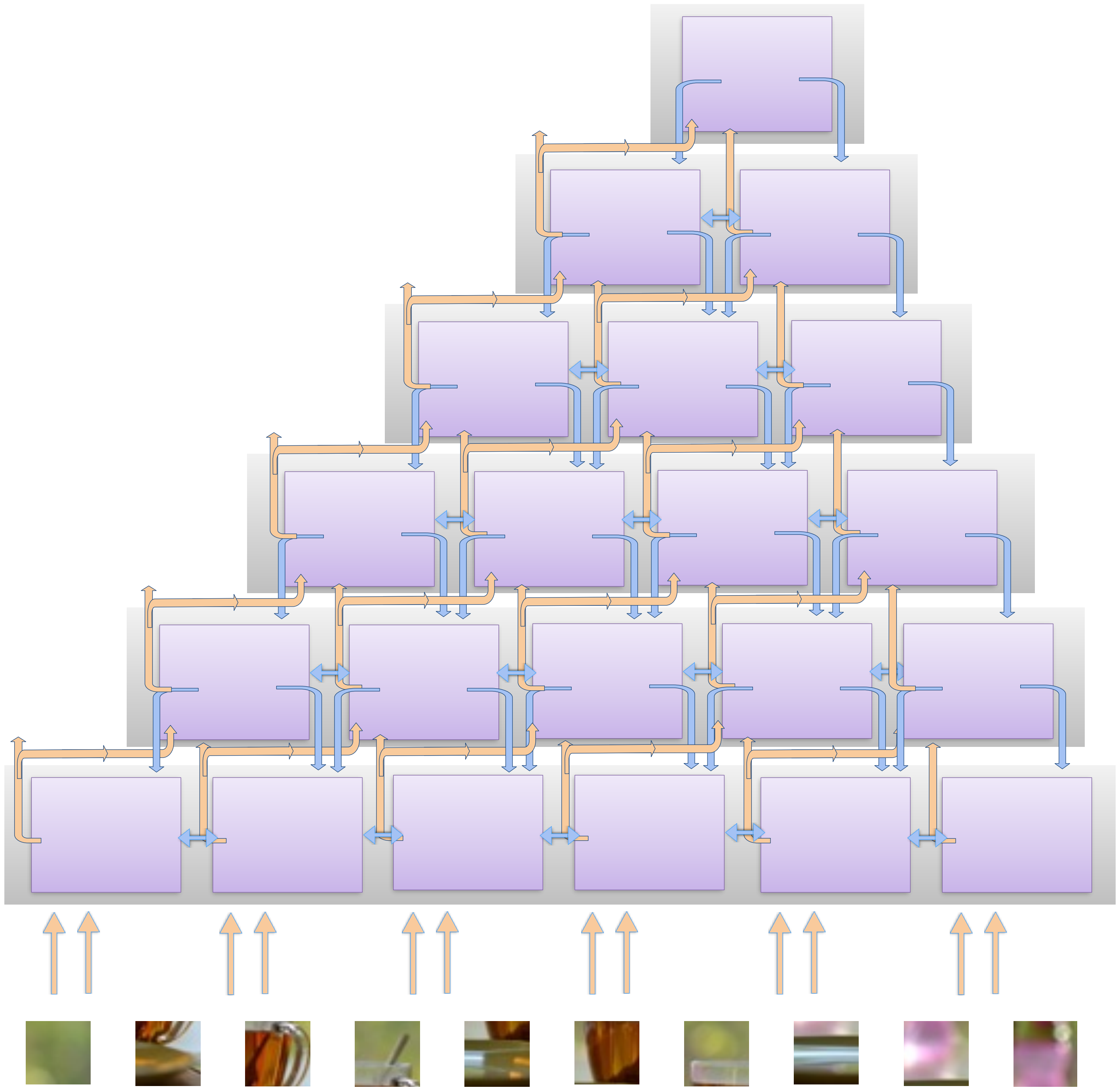


Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Lateral Context



Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Context
Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Context
Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Context
Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Context
Primary Signal (predicted)
P_{t+1}
H_t
prediction
P_t
Primary Signal
C_{t-1}
Context

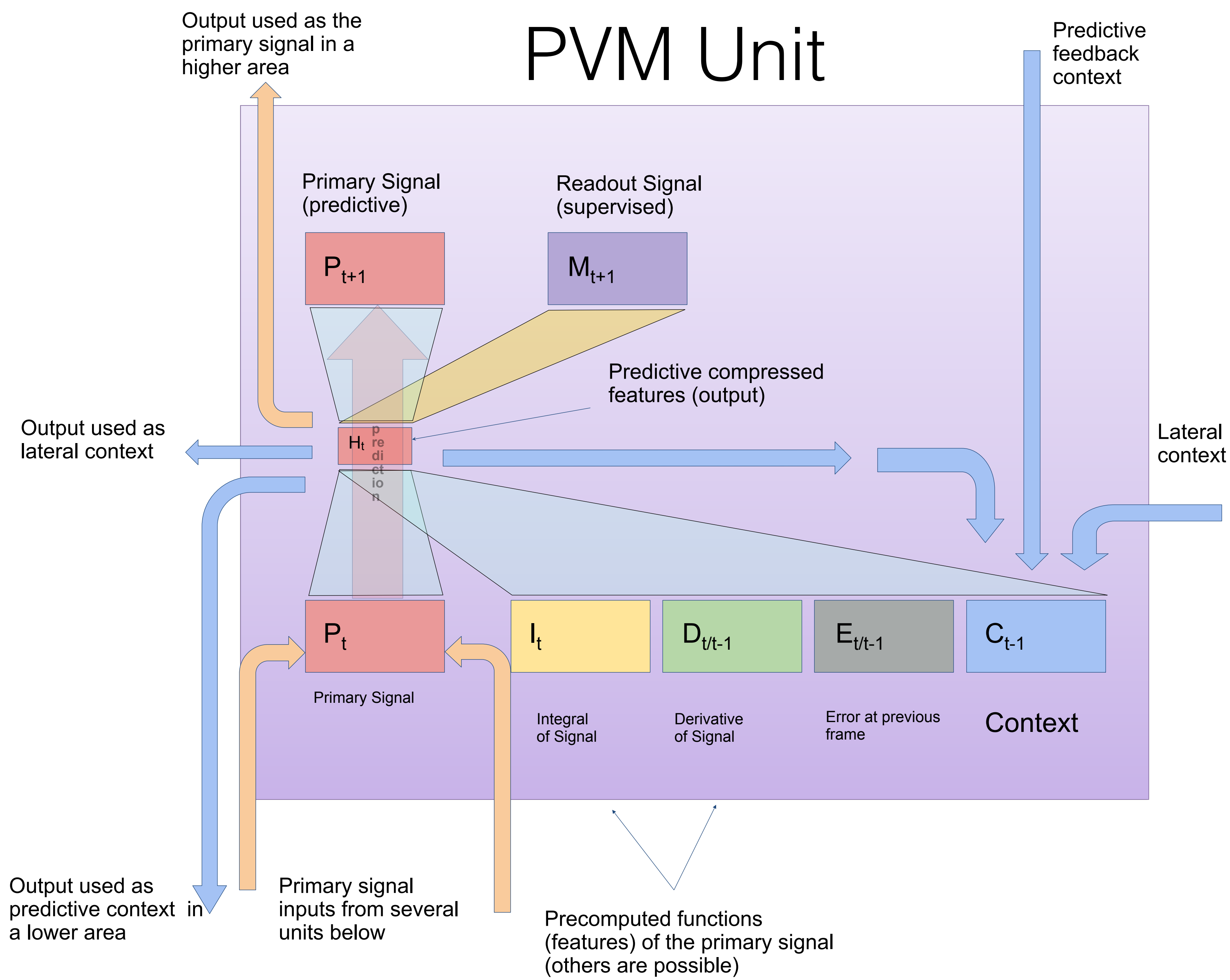
Architecture



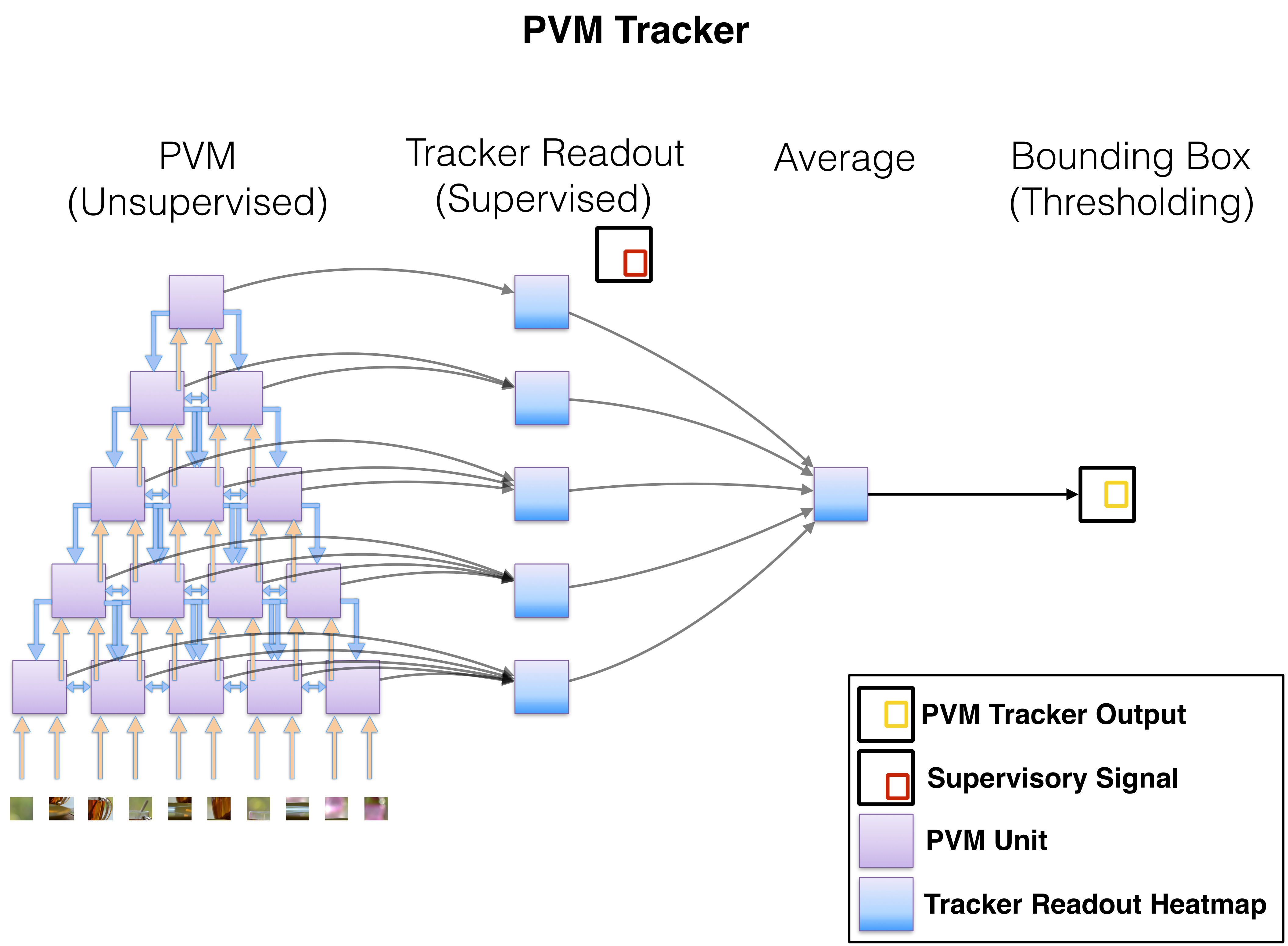
- Uniform structure
- Local learning
- Compression
- Not relying on any particular low level implementation
- Doesn't even have to be synchronous!
- Stable !

Are the features that make this architecture scalable

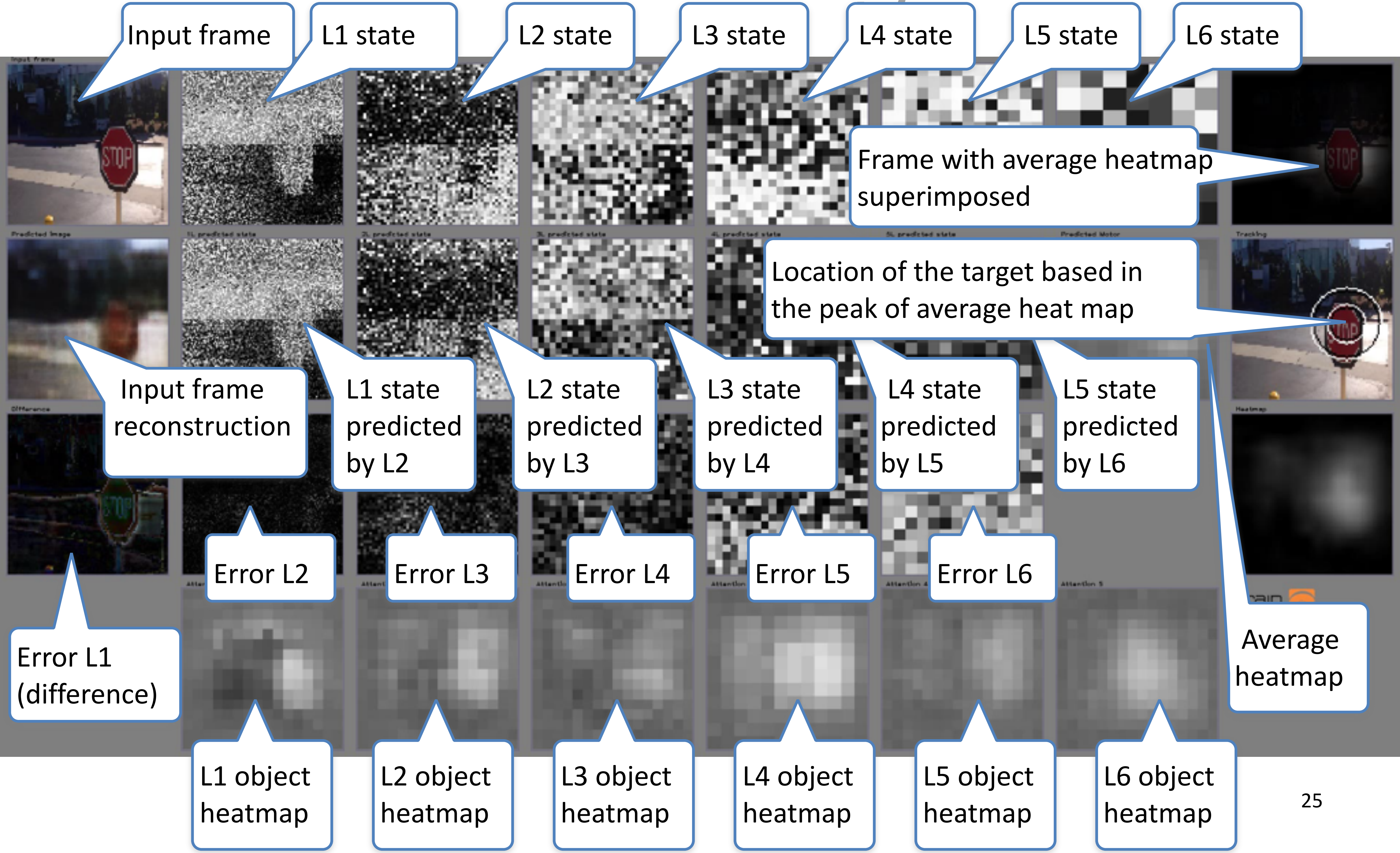
Mixing predictive and supervised



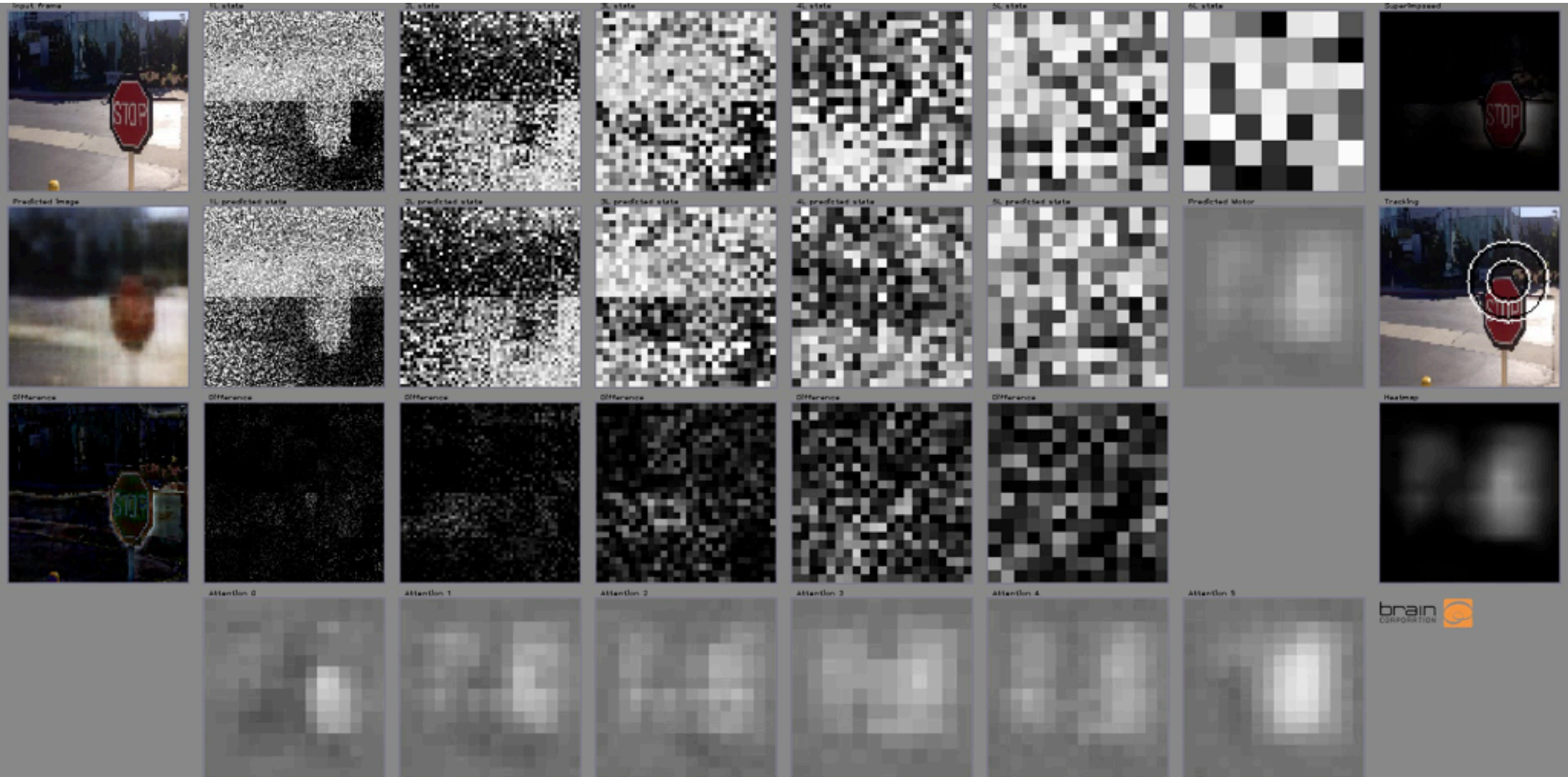
Predictive Vision Model Architecture



PVM Tracking task

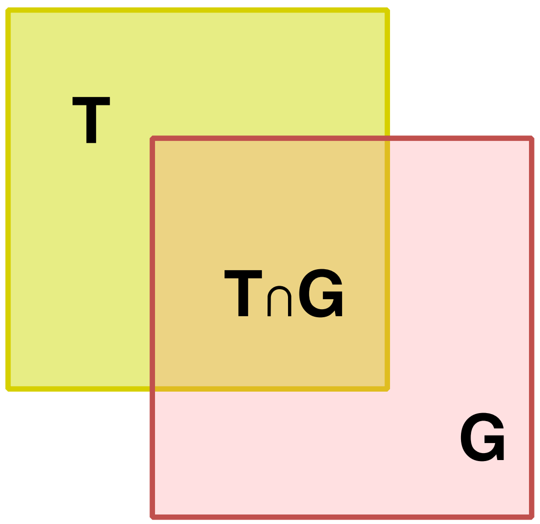


PVM Tracking task



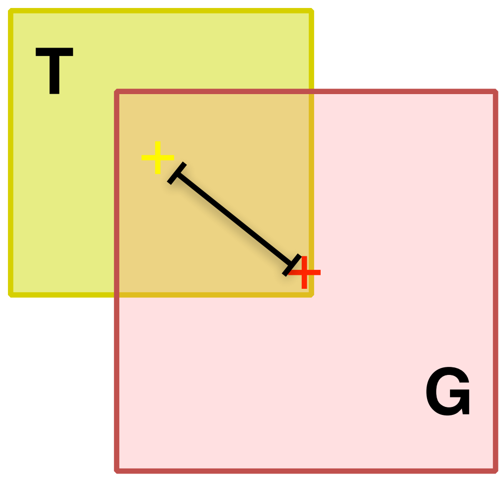
Quantifying Visual Object Tracking

Success (Overlap)



$$\%\left(\frac{T \cap G}{T \cup G} > \theta\right)$$

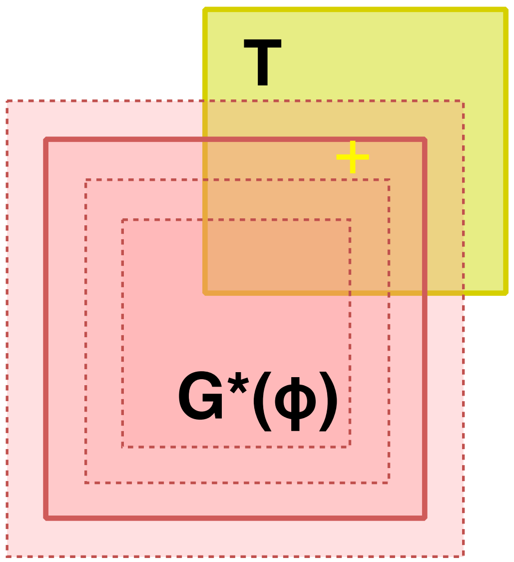
Precision



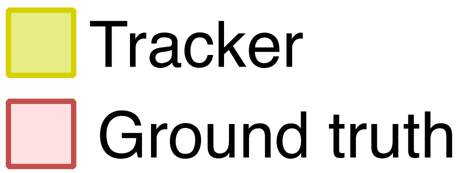
$$\%\left(T_{+} - G_{+} < \rho_{\text{pixels}}\right)$$

for true positives

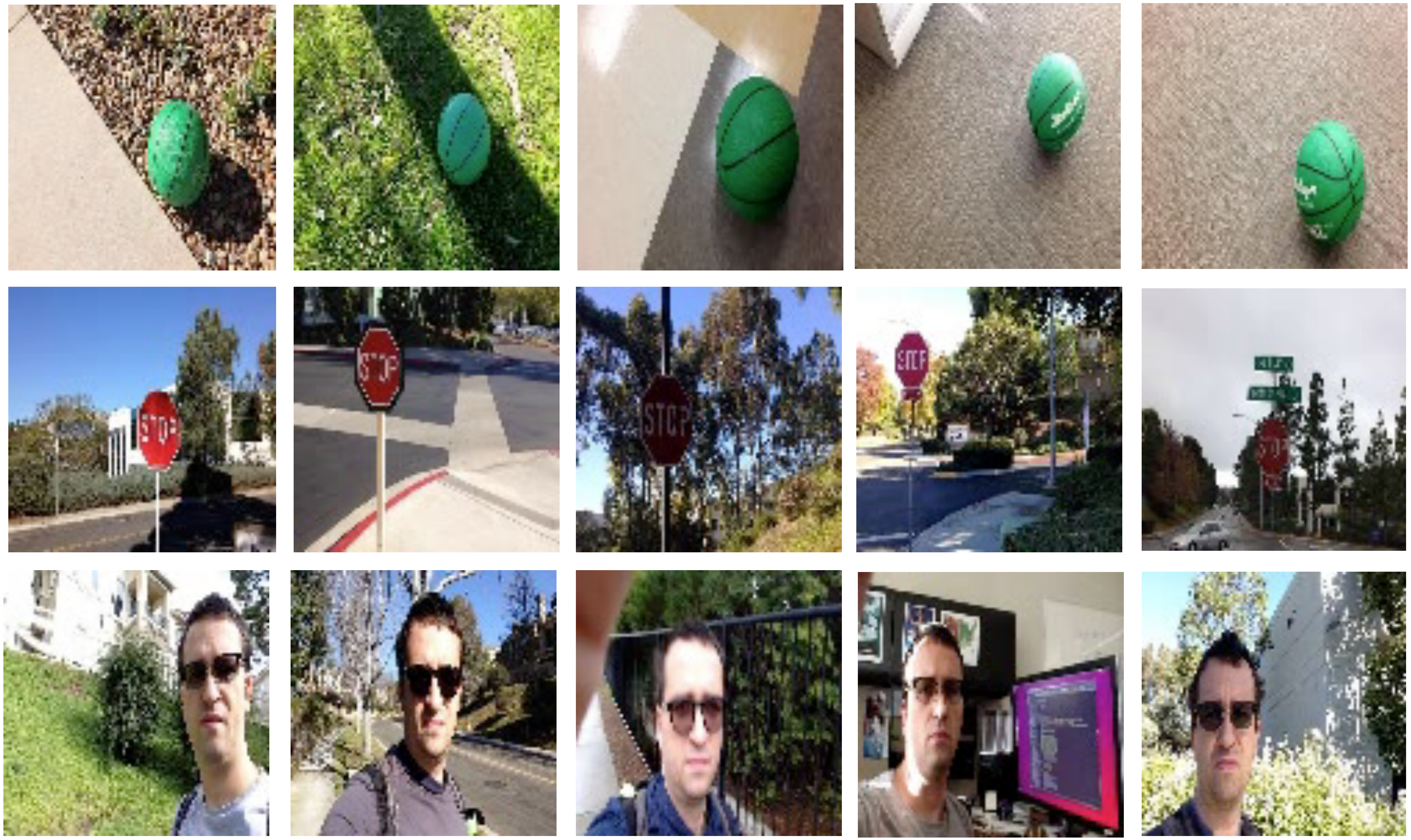
Accuracy



$$\%\left(\begin{array}{c} T_{+} \text{ in } G^{*}(\phi) \\ + \\ \text{true negatives} \end{array}\right)$$



Experiment 1: Datasets



Experiment 1: Datasets

Name	# sequences	Length (frames)	Approx. duration at 25fps
Green ball training	5	12,764	8.5 min
Green ball testing	29	34,791	23 min
Stop sign training	20	15,342	10 min
Stop sign testing	30	22,240	15 min
Face training	10	29,444	19.6 min
Face testing	25	34,199	23 min
Total training	35	57,550	38 min
Total testing	84	91,230	60 min

Pretty big compared to other visual tracking datasets.

Available at pvm.braincorporation.net

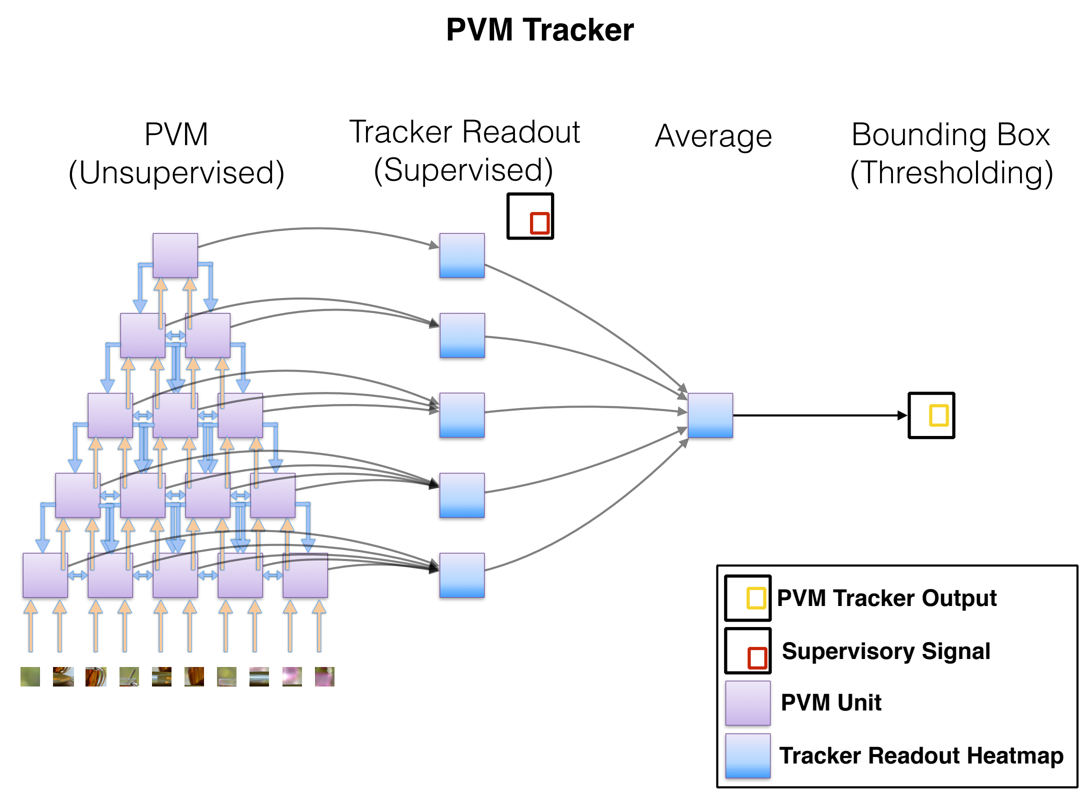
Experiment 1

- Simultaneous unsupervised and supervised training of PVM and tracker readout.

Experiment duration

Predictive training (unsupervised)

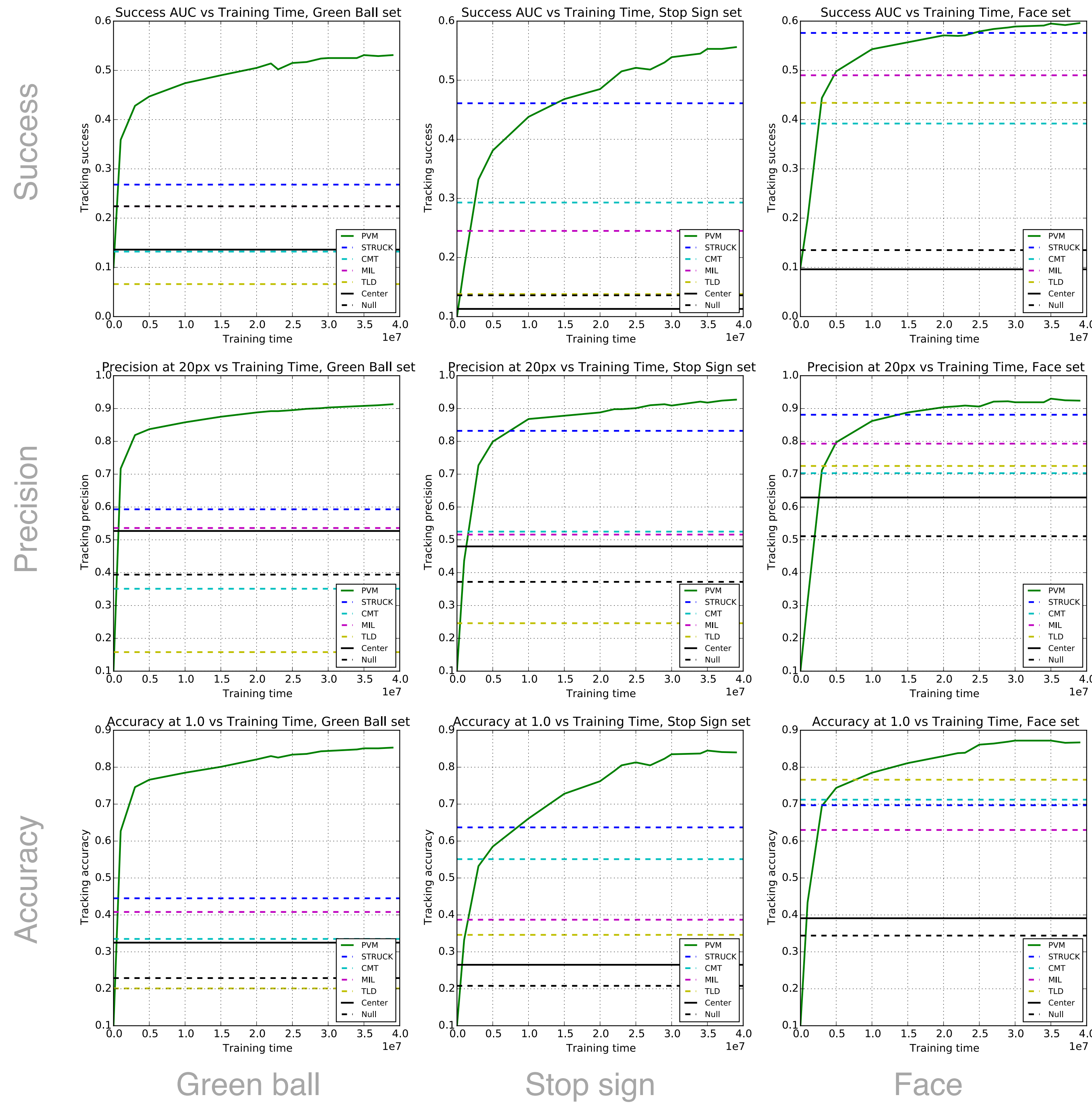
Readout training (supervised)



Experiment 1 Effect of Supervised + Unsupervised Training Time

- Performance on three datasets and three measures as a function of total training time

- In all 9 cases sufficient training allows to suppress state of the art



Experiment 1: Highlights Results Video

Predictive Vision Model
Online object tracking

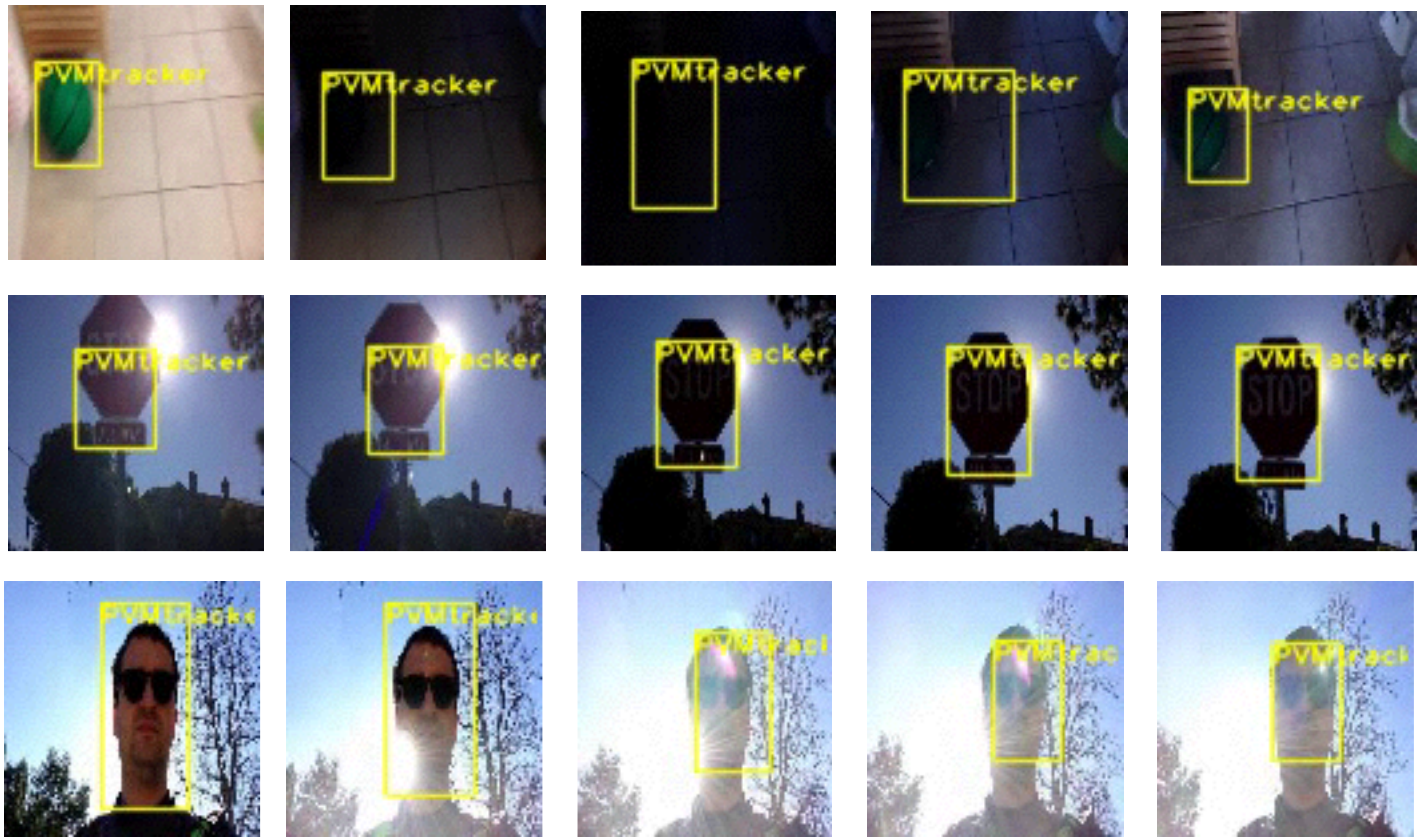
Examples from the test set - the model was not trained on any of the examples to follow.

Model processed 96x96 video frame. Red - human generated label, yellow - PVM

Brain Corporation 2016

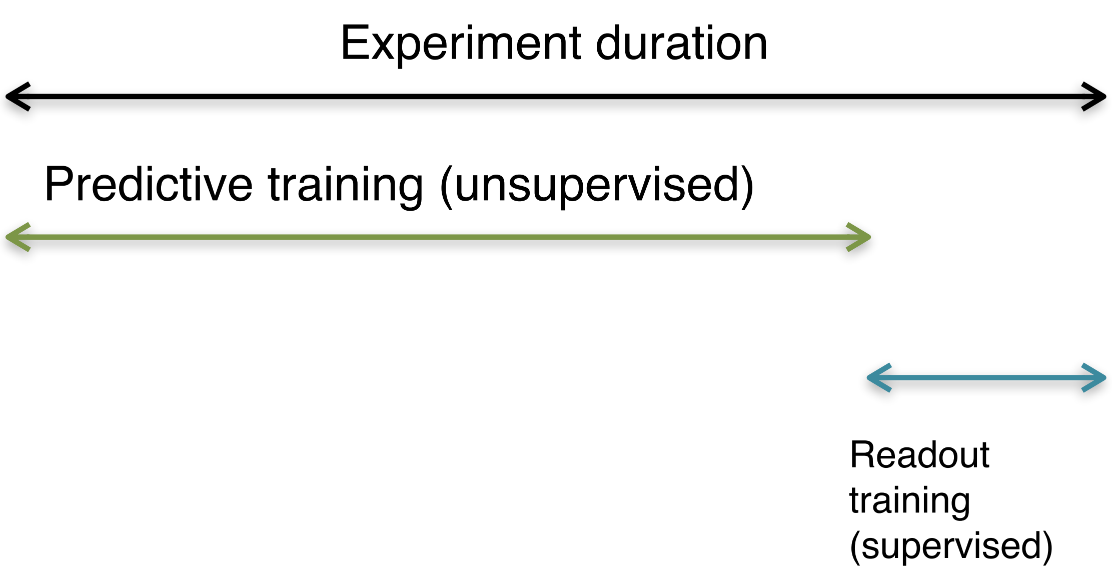
Challenging Visual Conditions

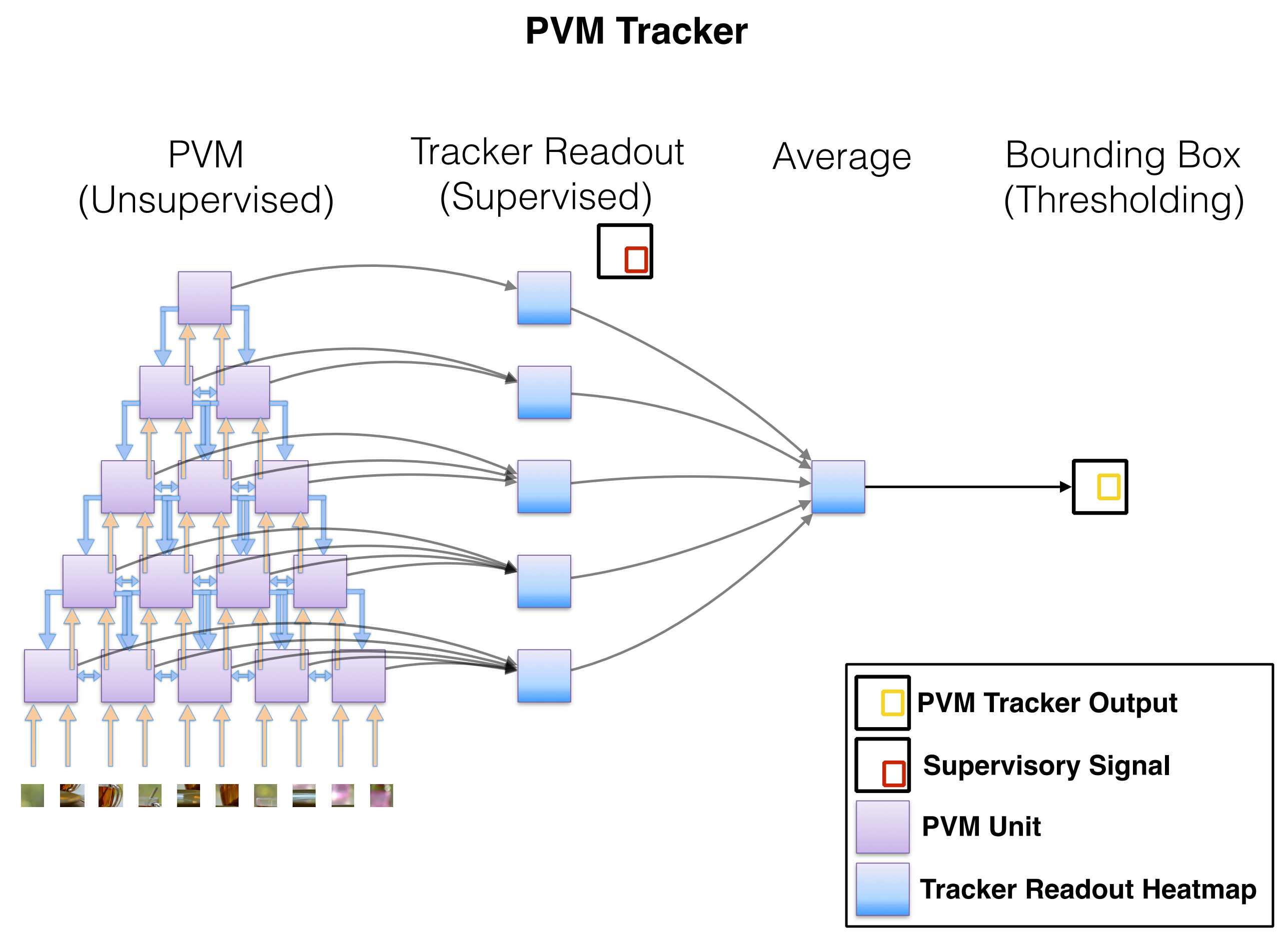
- A notable advantage of the PVM-based Tracker is that it can track objects robustly across real-world lighting changes, backlighting, lens flares, and shadows — **normally a challenge for robots operating in the real world.**



Experiment 2

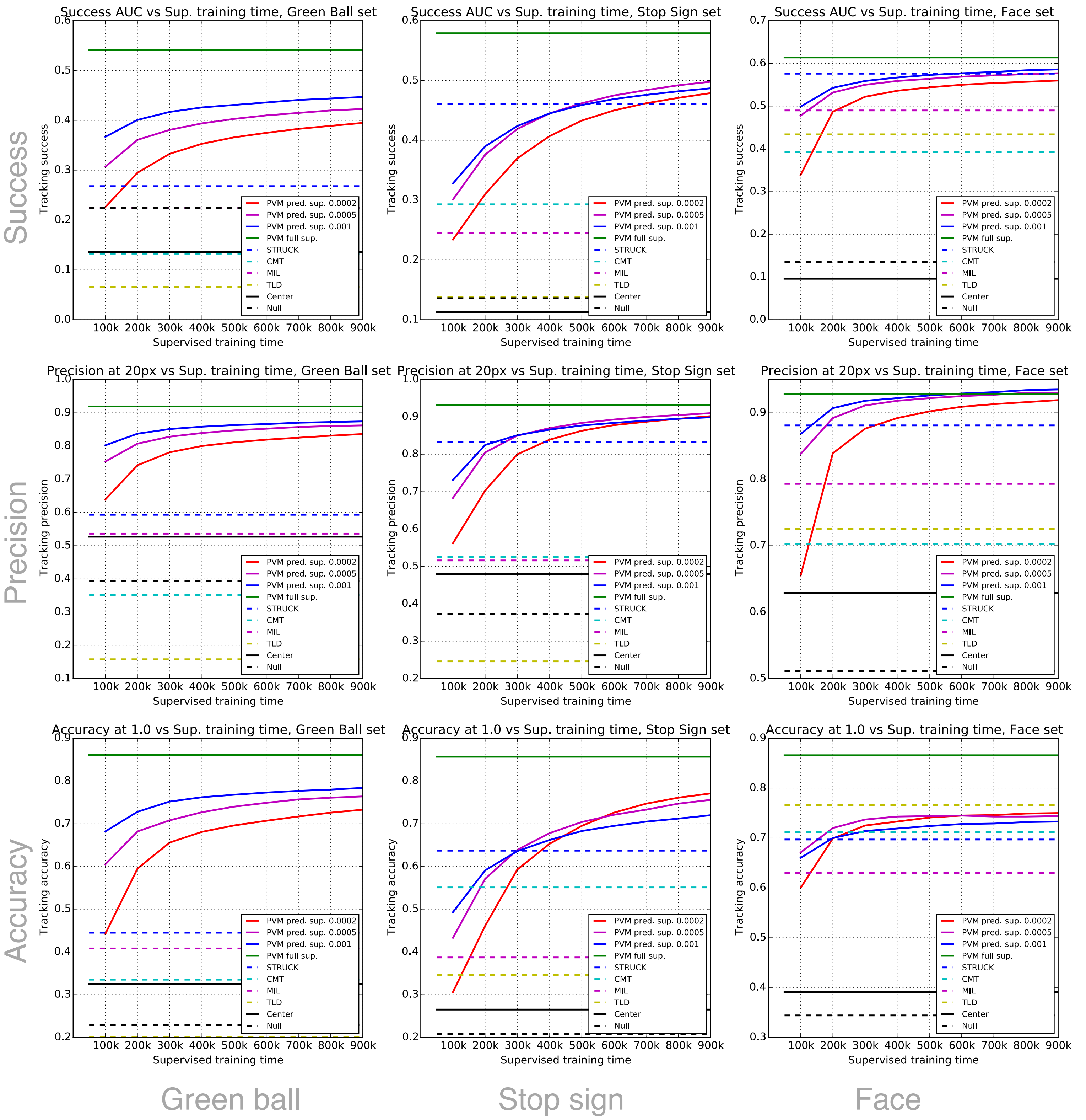
- Unsupervised-only Training of PVM, then Shorter Supervised "Priming" Segment of PVM Tracker



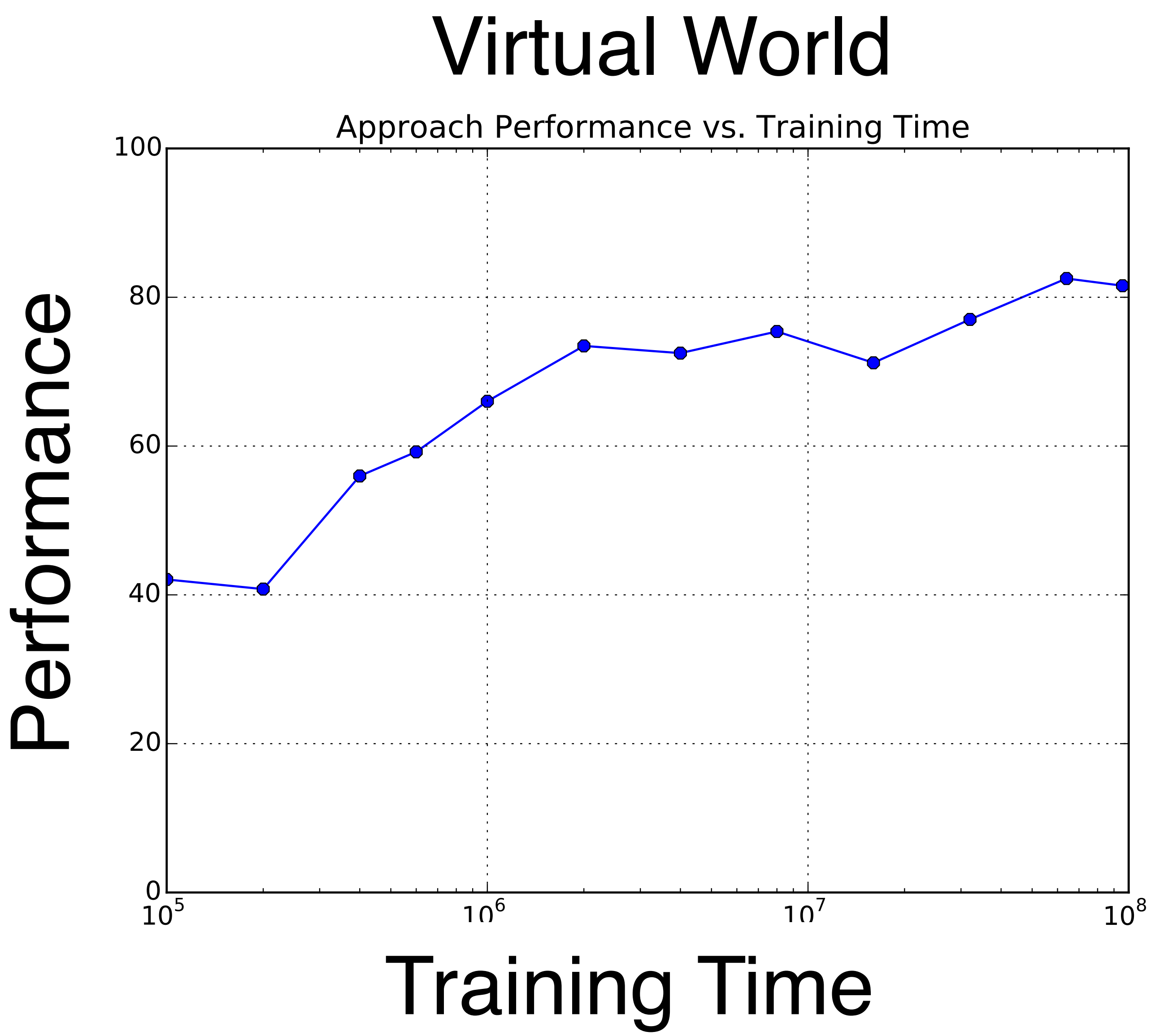


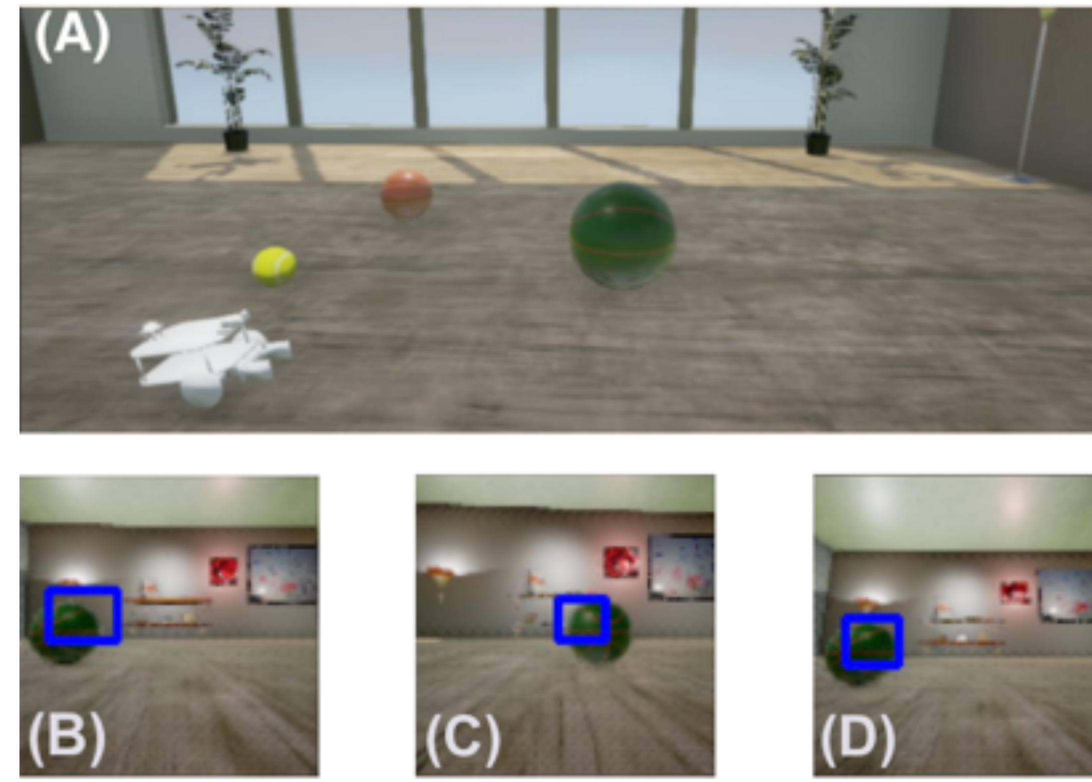
Experiment 2 Effect of Supervised Training Time

- Performance on three datasets and three measures as a function of short supervised training (after much longer unsupervised)
- In most cases the performance surpass state of the art



Experiment 3





The robot is placed in random positions in the virtual room and has to approach the ball within given amount of time.

Experiment 3 in the real world

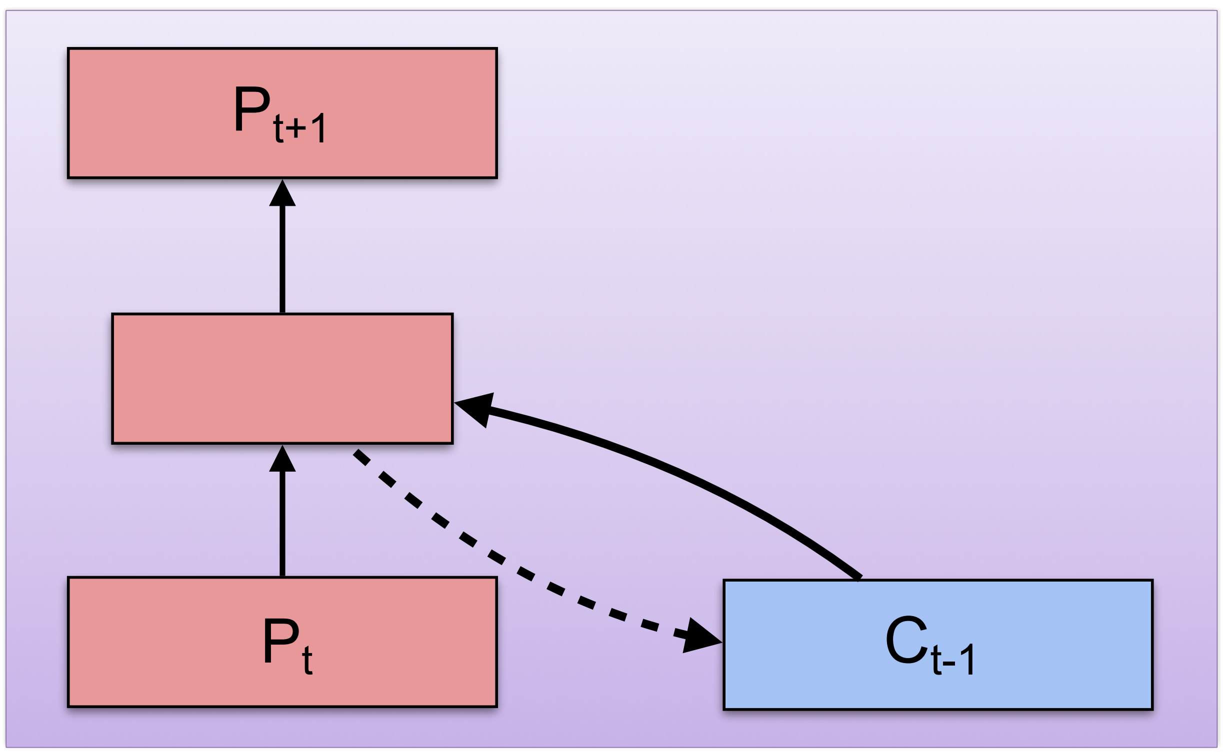
Summary

- Where does deep learning succeed and where does it fail
  The quest for Artificial Intelligence and elephants in the room
- Predictive vision model - alternative approach
  The predictive vision model as a way to bring machine learning to the dynamical, physical reality. Perception as prediction.
- **Why it works**
  Machine learning meets the issues of scalability, parallelism, real time execution, independence of particular implementation
- Next steps
  A path towards a truly "physical" machine learning implemented on a cortical processor

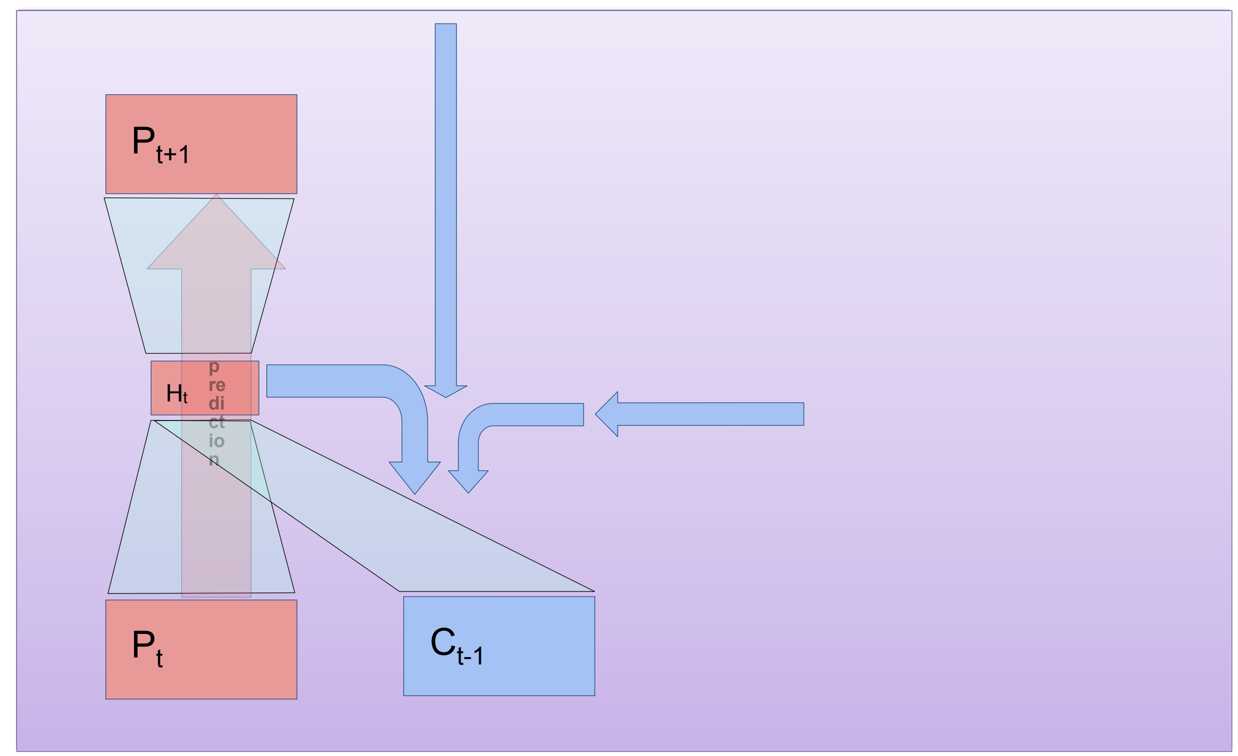
Analysis of PVM's recurrent structure

- PVM architecture is constructed as a hierarchy with feedback.
- Re-conceptualizing the network recursively, PVM units are organized as self-similar, nested simple recurrent networks.
  - PVM units **use each other** as deep SRN "context layers"

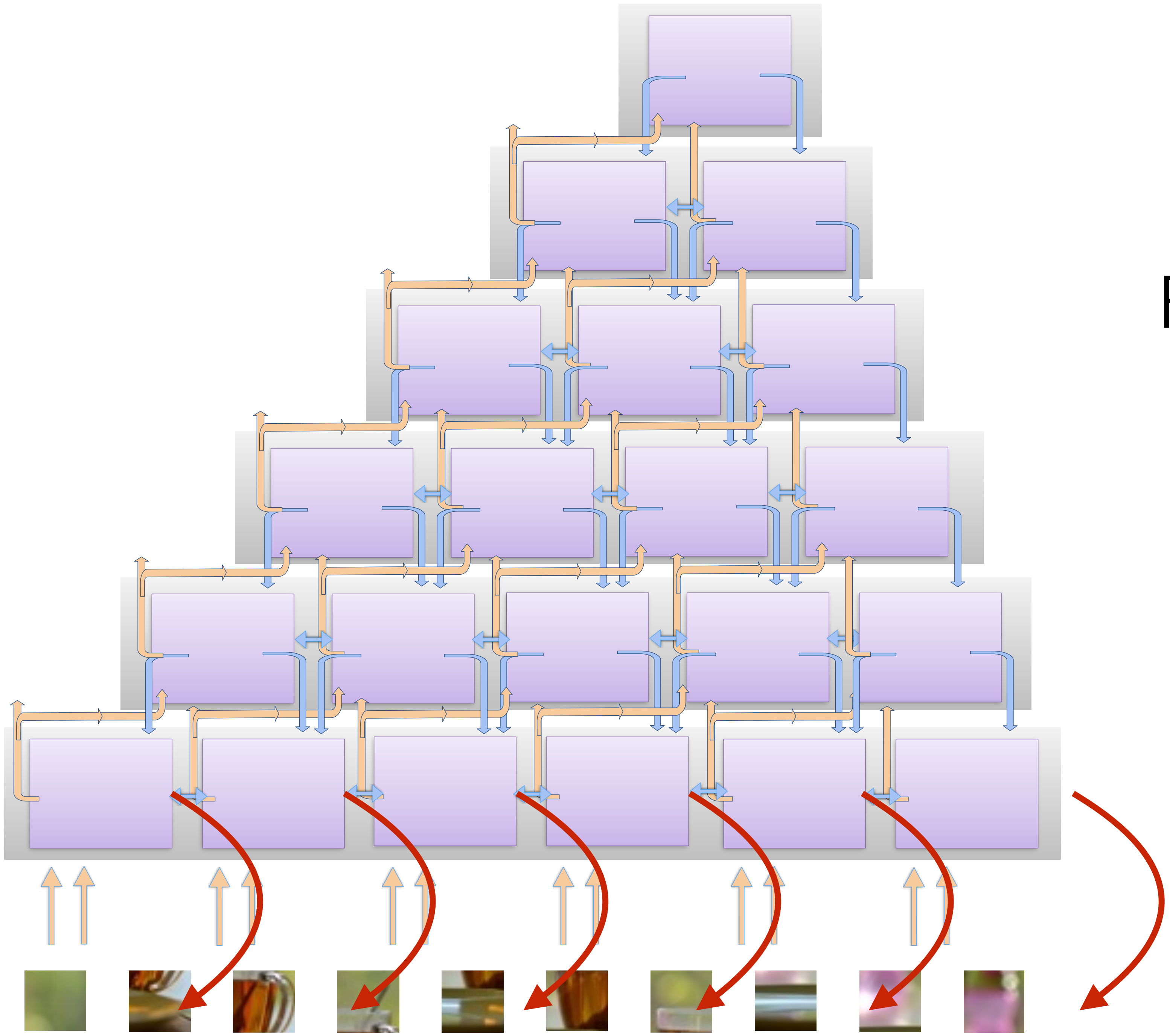
(A) Simple Recurrent Network



(B) Predictive Vision Model Unit*



Generative PVM



Feed prediction as new input

Explore the "predictive trajectory"

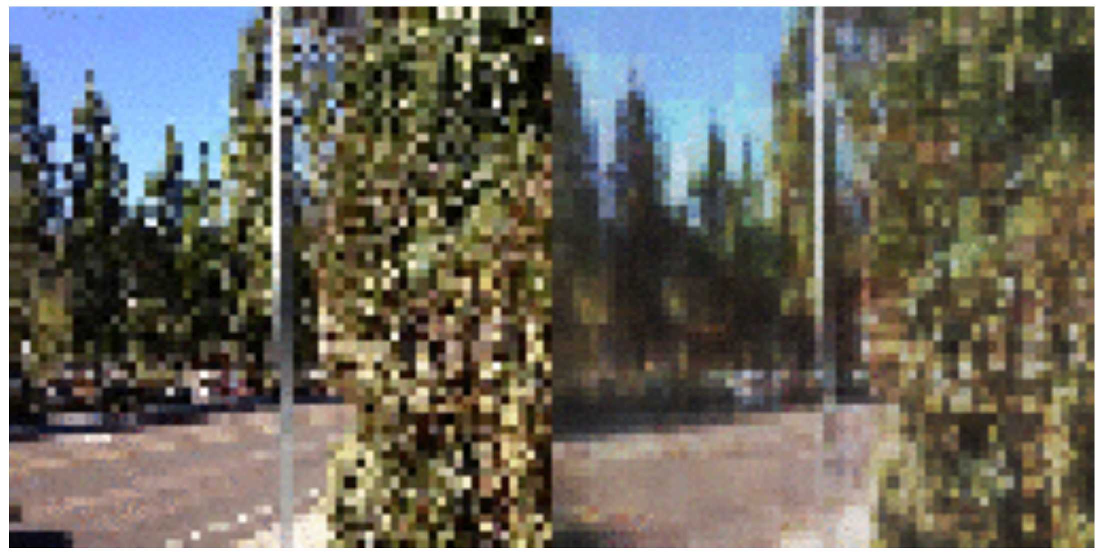
Recurrent dreams

“Dream” Actual Difference

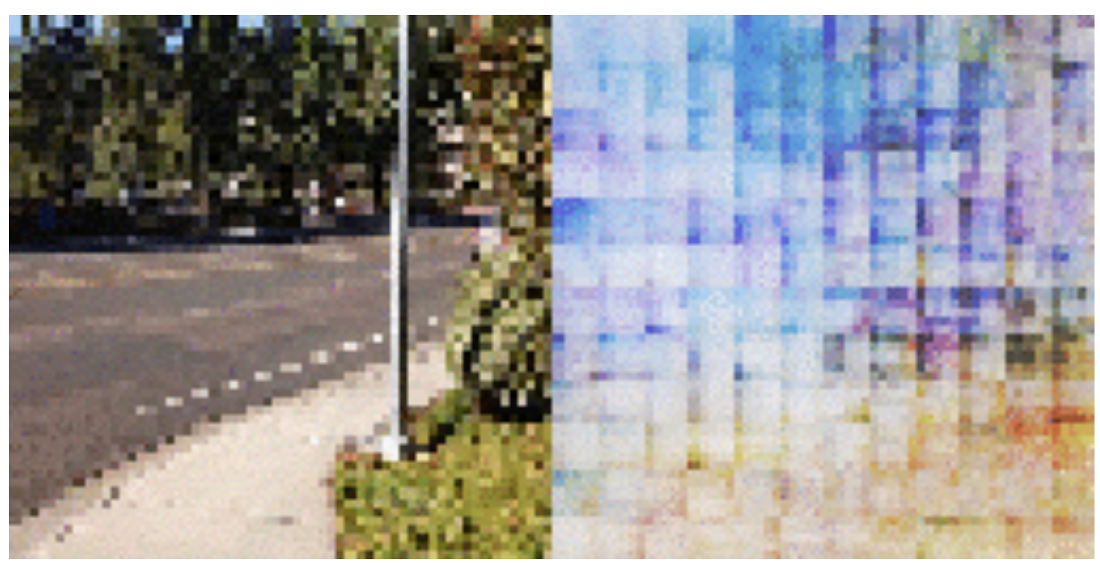
“Dream” Actual Difference

Contextual fill-in

Reconstruction behind an occluder.



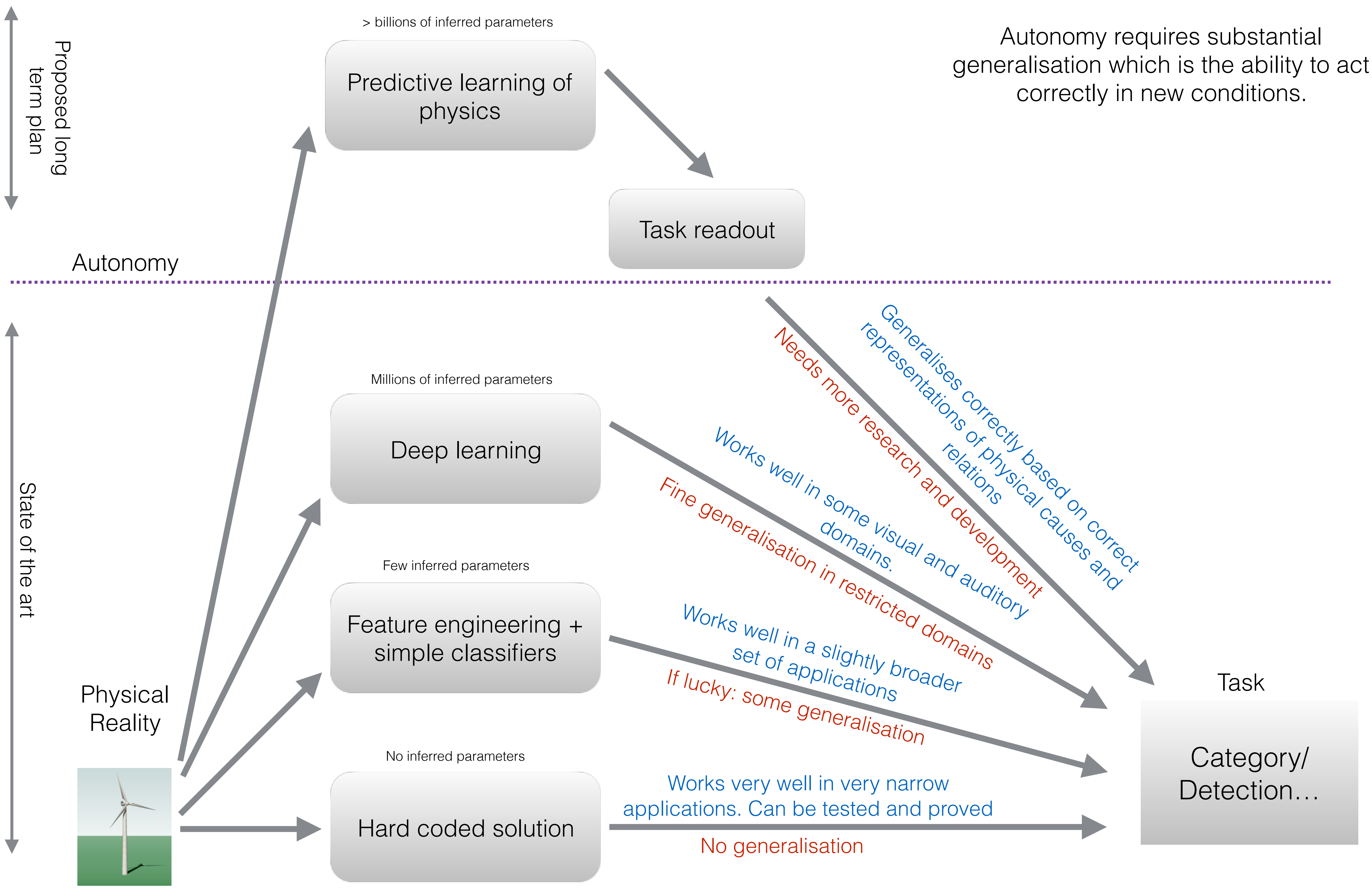
Reconstruction with feedback and lateral connections disabled.



Summary

- Where does deep learning succeed and where does it fail
  The quest for Artificial Intelligence and elephants in the room
- Predictive vision model - alternative approach
  The predictive vision model as a way to bring machine learning to the dynamical, physical reality. Perception as prediction.
- Why it works
  Machine learning meets the issues of scalability, parallelism, real time execution, independence of particular implementation
- **Next steps**
  A path towards a truly “physical” machine learning implemented on a cortical processor

Path to autonomy



Next steps

Rough plan:

0. Reimplement PVM for GPU
1. Explore applications, e.g. autonomous cars or multimodal perception
2. Scale up number of parameters, training time, modalities
3. Explore meta-parameters
4. Experiment with various neuromorphic implementations
5. Apply in closed loop setting (control problem)
6. Add action selection/reinforcement learning
7. Scale up
8. Scale up…

Moravec's Paradox

"It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers...

...and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."

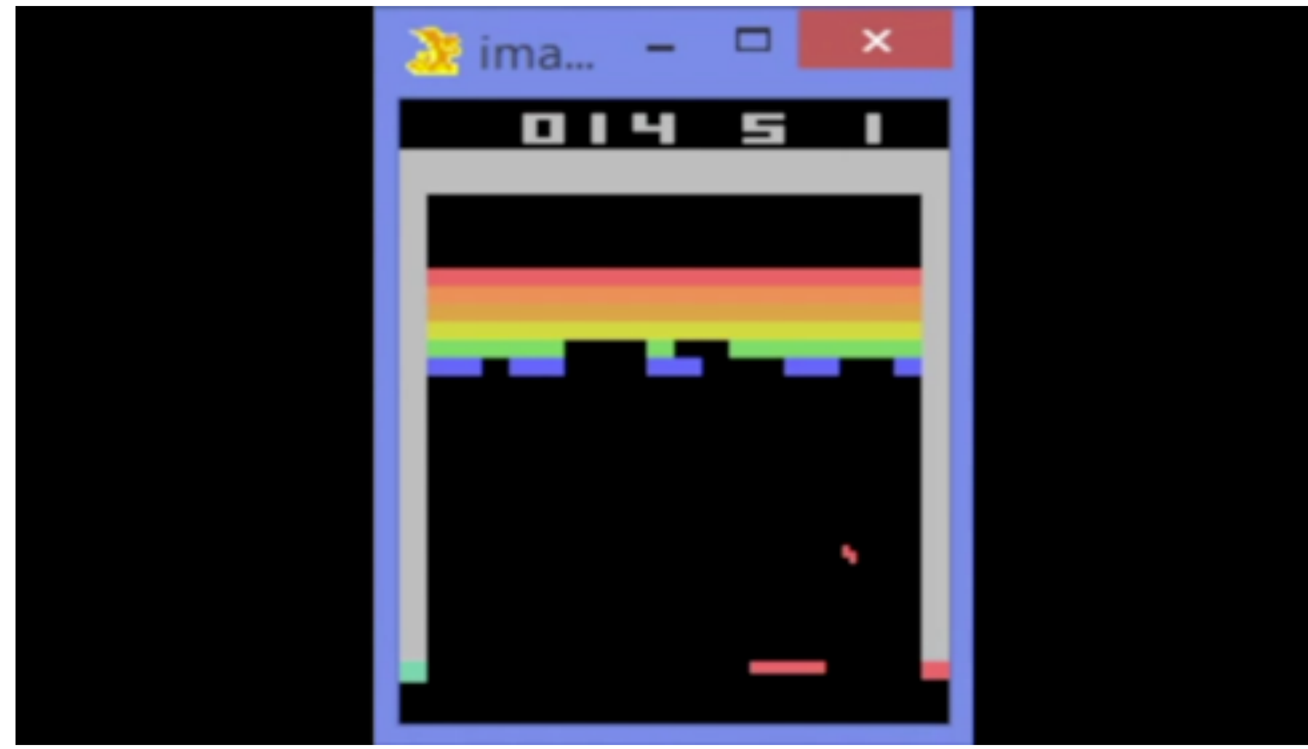
—Hans Moravec (1988)

Moravec's Paradox AD 2016

2015 DARPA Robotics Challenge

Source: youtube

Various recent AI achievements



How can we not see this elephant?

Moravec’s paradox is not a problem for roboticists.
It is the central problem for AI that has long been neglected.

Acknowledgements

- AFRL and DARPA Cortical Processor seedling contract FA8750-15-C-1078
- People involved:
  - **Todd Hylton** - Management support
  - **Patryk Laurent**, **Csaba Petre** - PVM and robotics experiments
  - **Micah Richert**, **Dimitry Fisher** - ILCM
  - **Filip Piekniewski** - PI, PVM, implementation
- **Dan Hammerstrom** who started the program
- Countless **robots** developed and sacrificed for the sake of science
- **Hans Moravec** for stating his paradox so clearly

Papers and results

More details and source code available in two papers recently released:

“Unsupervised Learning from Continuous Video in a Scalable Predictive Recurrent Network”
https://arxiv.org/abs/1607.06854

“Fundamental principles of cortical computation: unsupervised learning with prediction, compression and feedback”
https://arxiv.org/abs/1608.06277

Code (CPU)
http://github.com/braincorp/PVM

And a blog:
http://blog.piekniewski.info

Thank you!