Industrial Research Institute
2018-06-06

# The Power and Limits Of Deep Learning

Yann LeCun
Facebook AI Research
New York University
http://yann.lecun.com

facebook
Artificial Intelligence Research

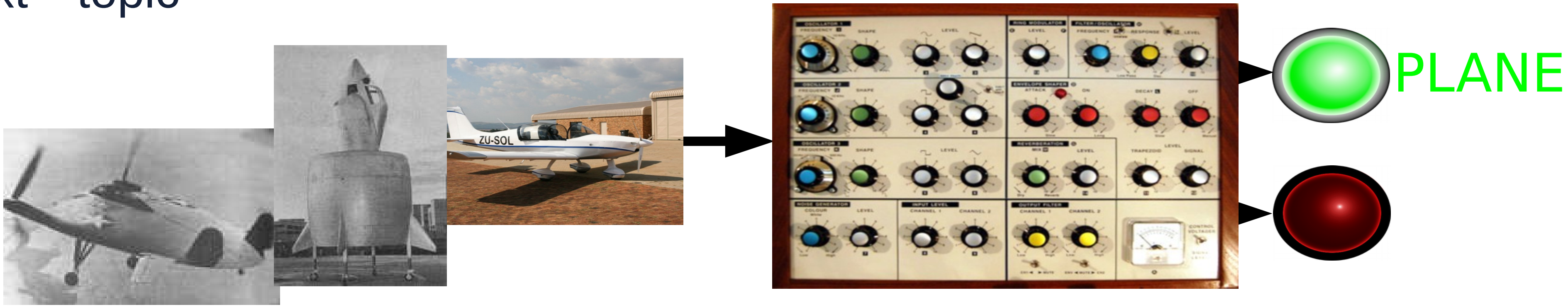
# Deep Learning Today

## History and State of the Art

facebook
Artificial Intelligence Research

# Supervised learning

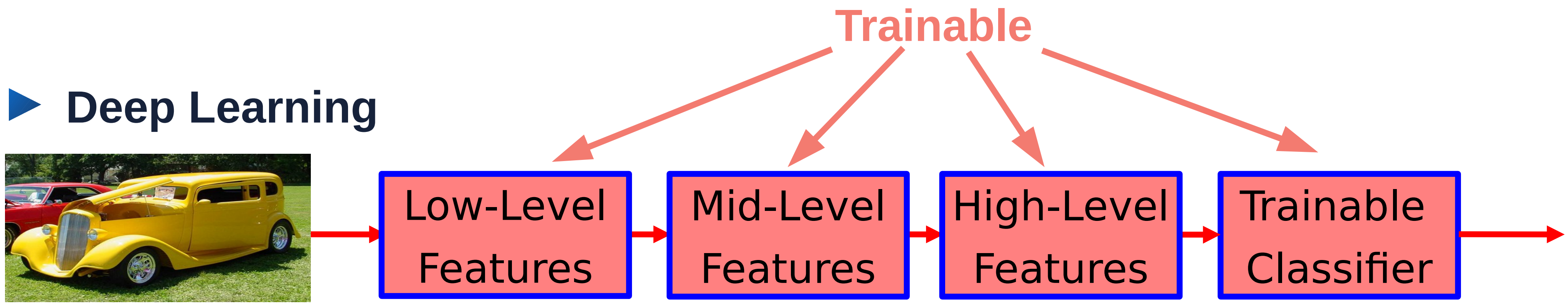
- **Training a machine by showing examples instead of programming it**
- **When the output is wrong, tweak the parameters of the machine**
- **Works well for:**
  - Speech → words
  - Image → categories
  - Portrait → name
  - Photo → caption
  - Text → topic
  - ....

# Deep Learning

- **Traditional Machine Learning**

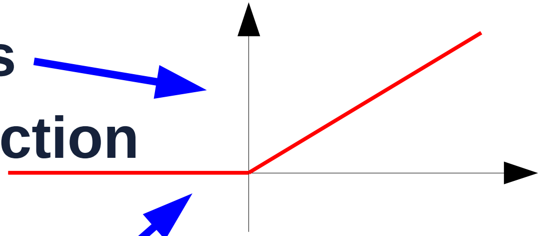

- **Deep Learning**



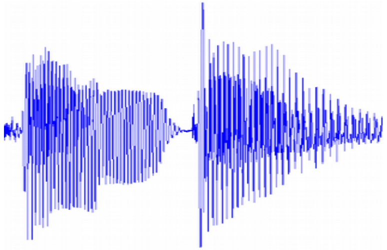
# Multi-Layer Neural Nets

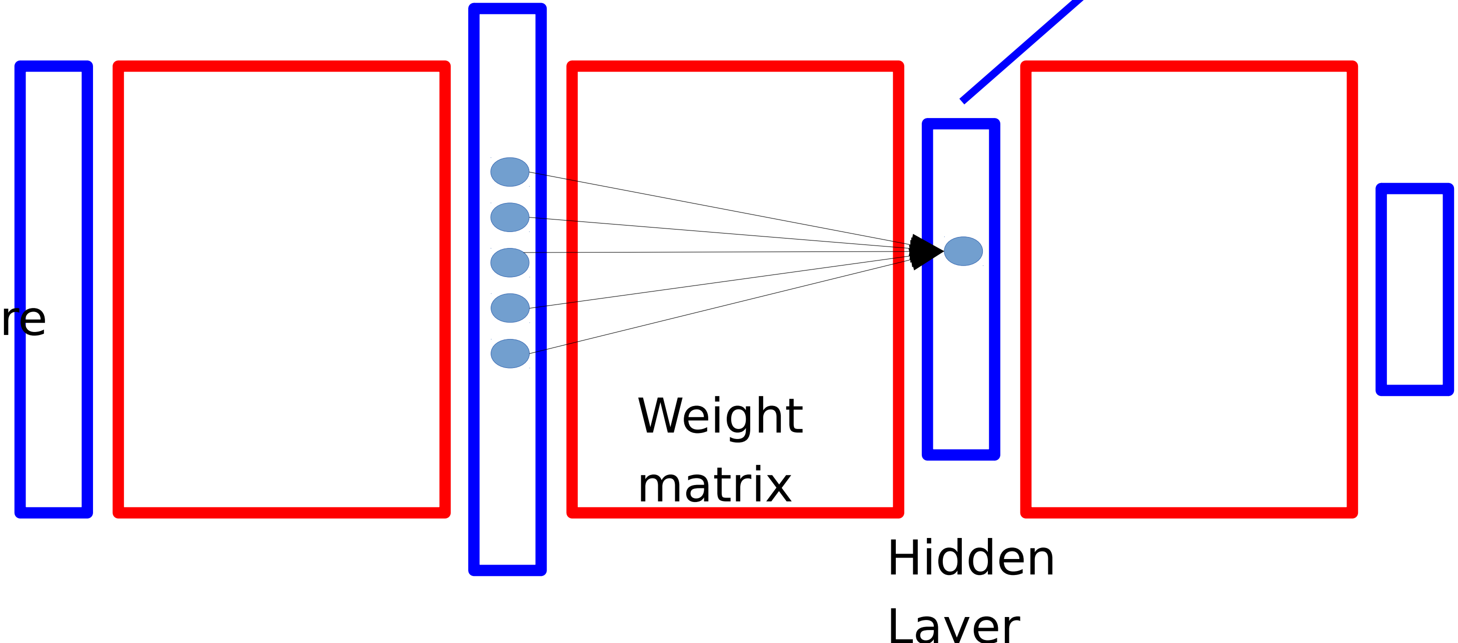
- **Multiple Layers of simple units**
- **Each units computes a weighted sum of its inputs**
- **Weighted sum is passed through a non-linear function**
- **The learning algorithm changes the weights**

$$ReLU(x) = max(x, 0)$$



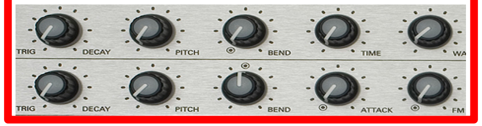
Ceci est une voiture





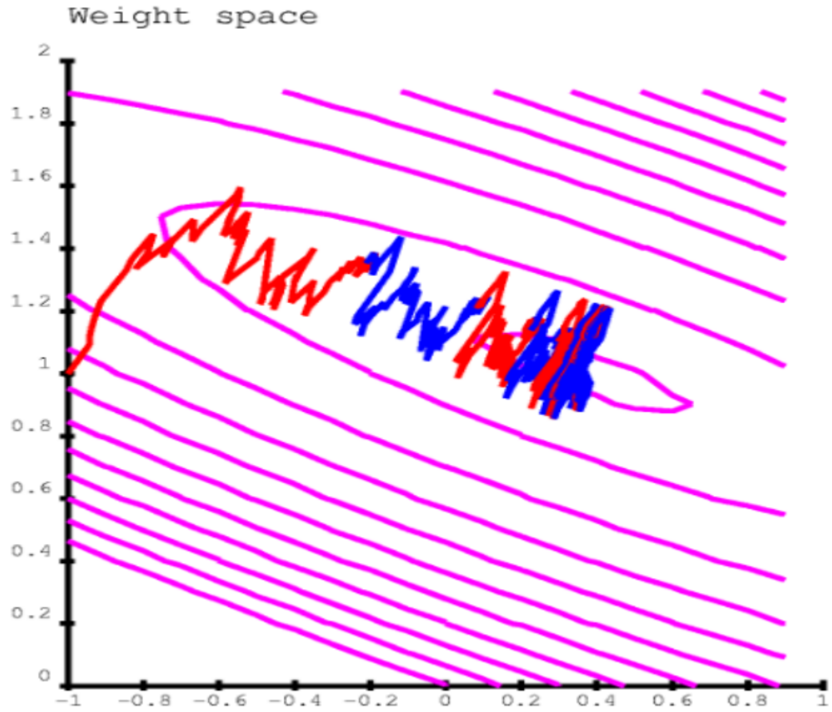
# Supervised Machine Learning = Function Optimization

Function with adjustable parameters




Objective Function

traffic light: -1

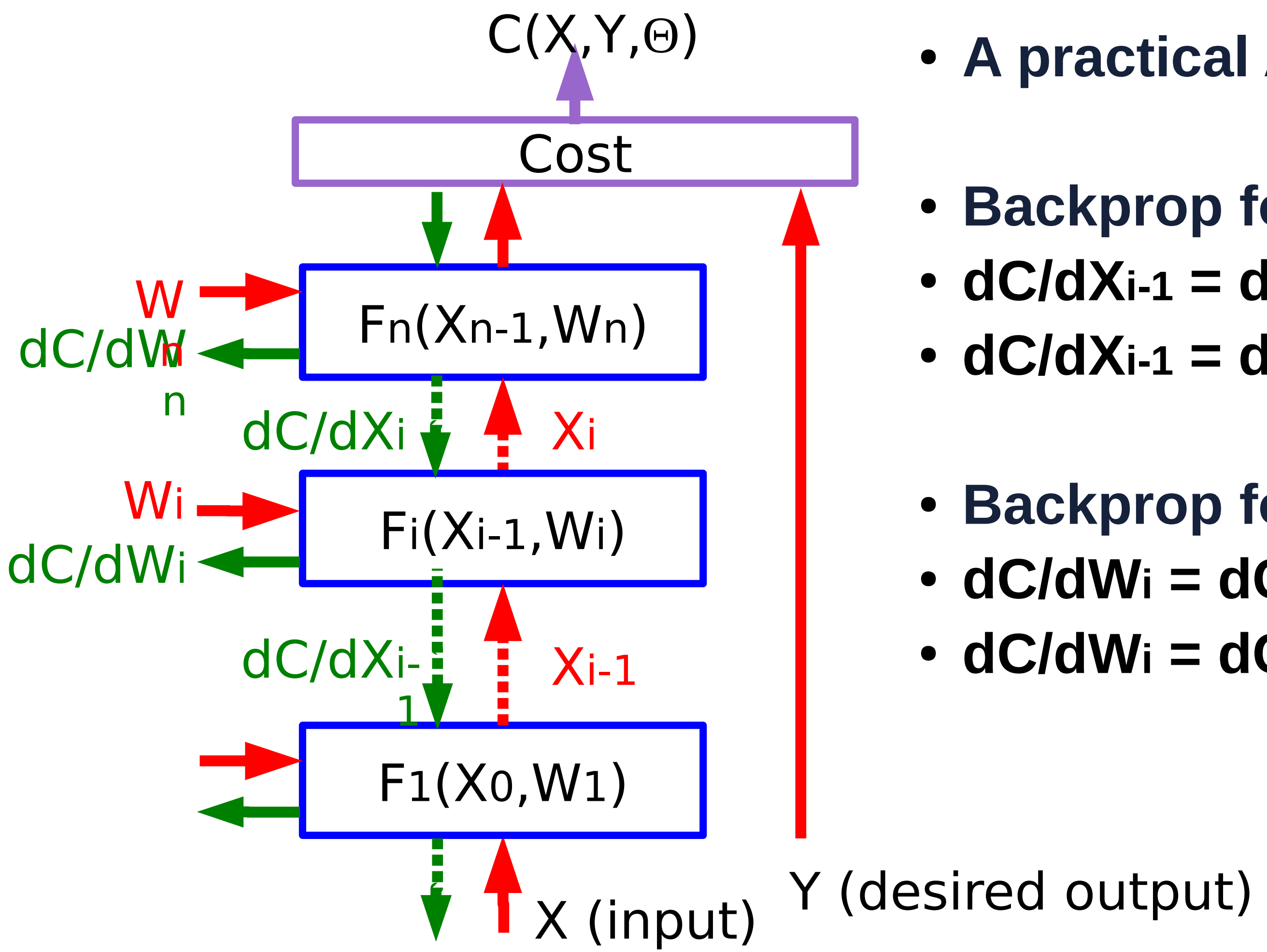



- **It's like walking in the mountains in a fog and following the direction of steepest descent to reach the village in the valley**
- **But each sample gives us a noisy estimate of the direction. So our path is a bit random.**

$$W_i \leftarrow W_i - \eta \frac{\partial L(W, X)}{\partial W_i}$$

- **Stochastic Gradient Descent (SGD)**

# Computing Gradients by Back-Propagation

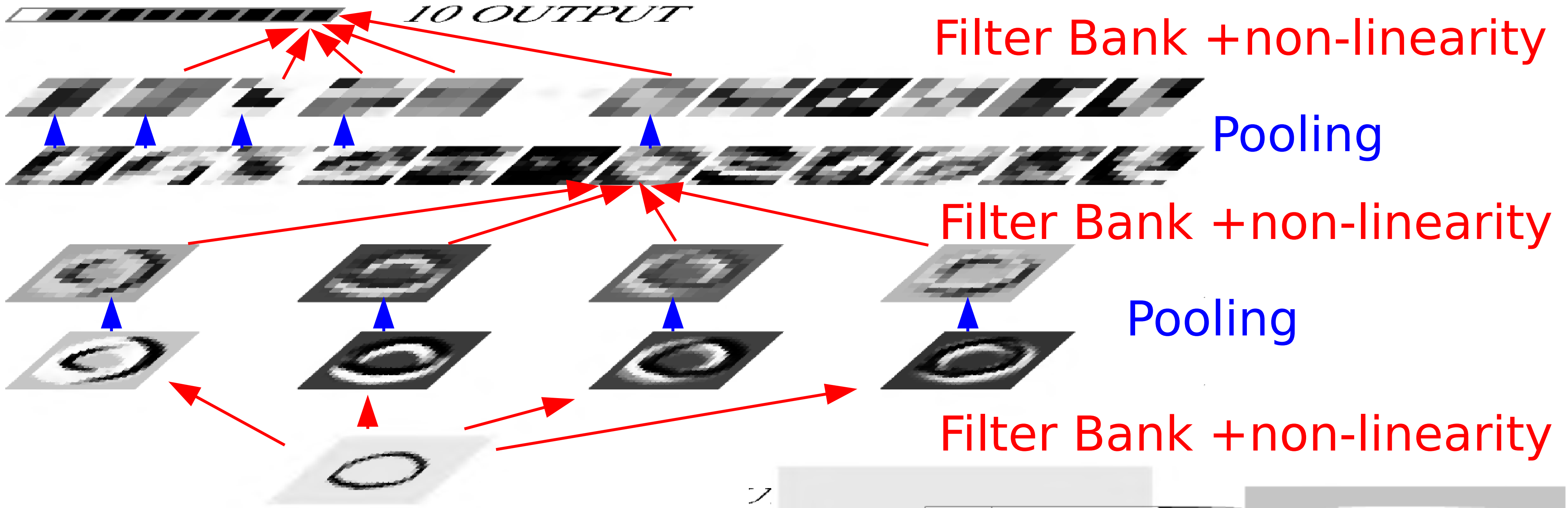



- **A practical Application of Chain Rule**

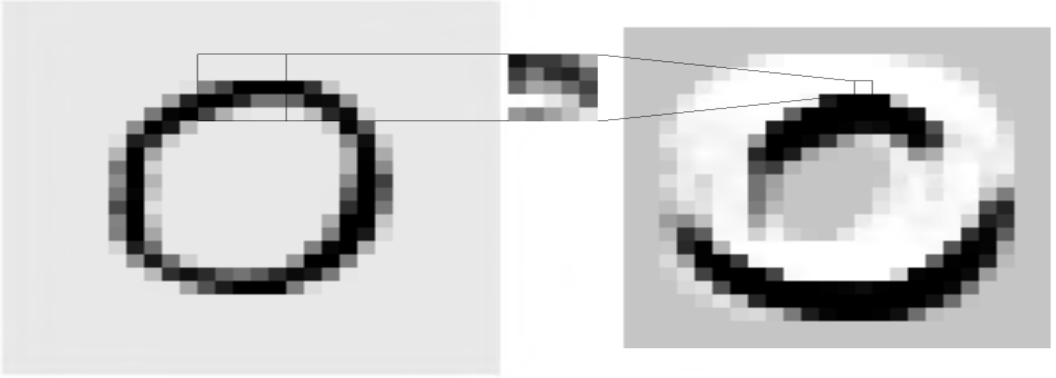
- **Backprop for the state gradients:**
- **$dC/dX_{i-1} = dC/dX_i \,.\, dX_i/dX_{i-1}$**
- **$dC/dX_{i-1} = dC/dX_i \,.\, dF_i(X_{i-1},W_i)/dX_{i-1}$**

- **Backprop for the weight gradients:**
- **$dC/dW_i = dC/dX_i \,.\, dX_i/dW_i$**
- **$dC/dW_i = dC/dX_i \,.\, dF_i(X_{i-1},W_i)/dW_i$**

# Convolutional Network Architecture [LeCun et al. NIPS 1989]



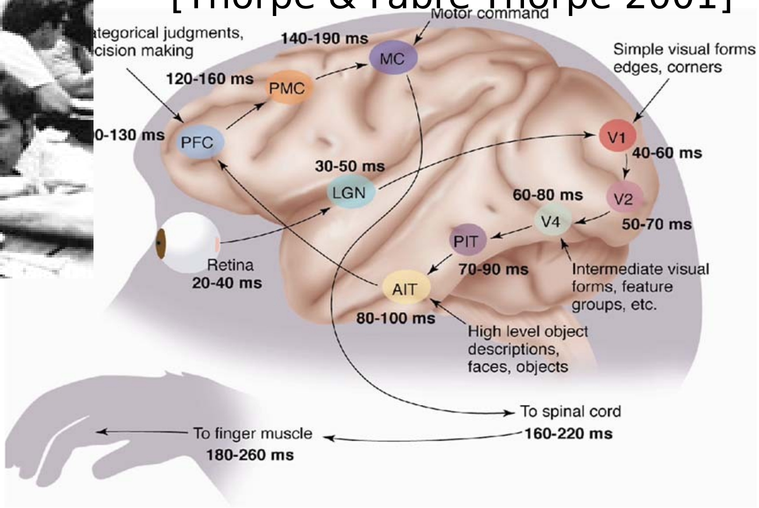

- **Inspired by [Hubel & Wiesel 1962] & [Fukushima 1982] (Neocognitron):**
  - simple cells detect local features
  - complex cells “pool” the outputs of simple cells within a retinotopic neighborhood.

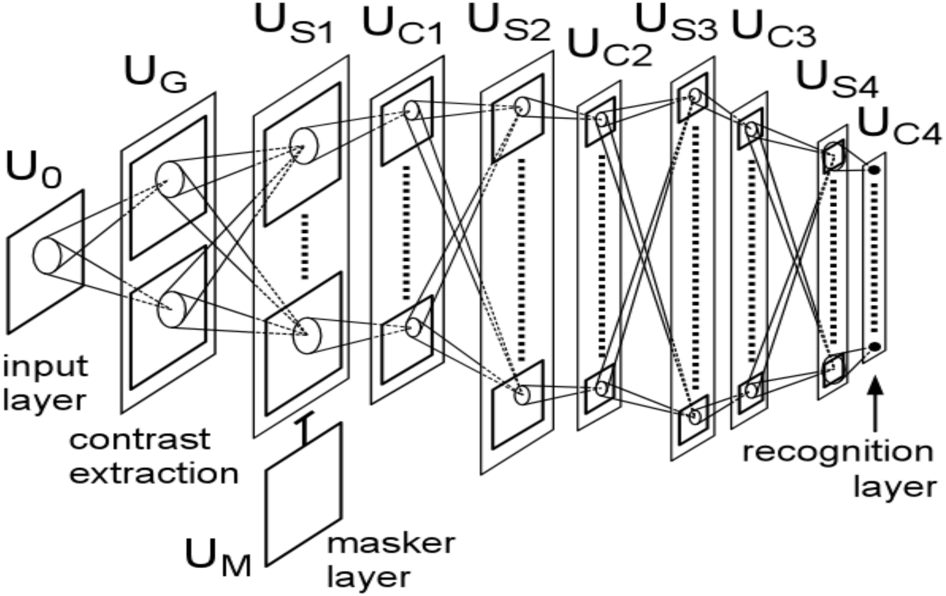


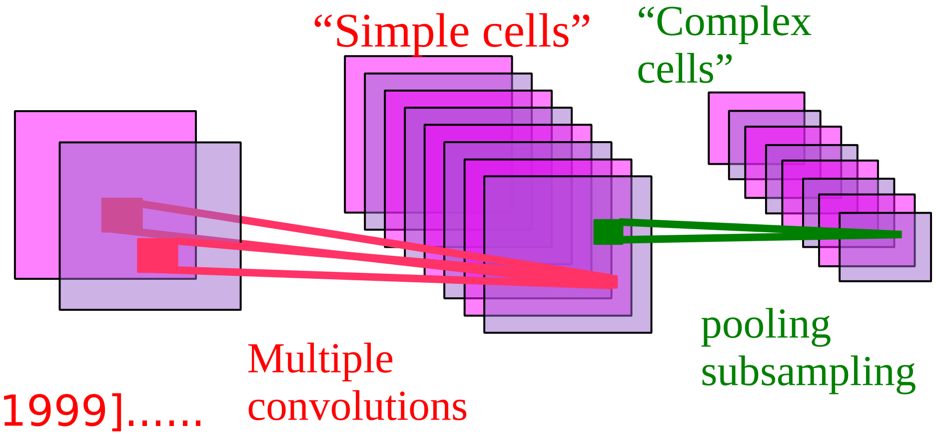
# Hubel & Wiesel's Model of the Architecture of the Visual Cortex

- **[Hubel & Wiesel 1962]:**
  - simple cells detect local features
  - complex cells "pool" the outputs of simple cells within a

[Thorpe & Fabre-Thorpe 2001]



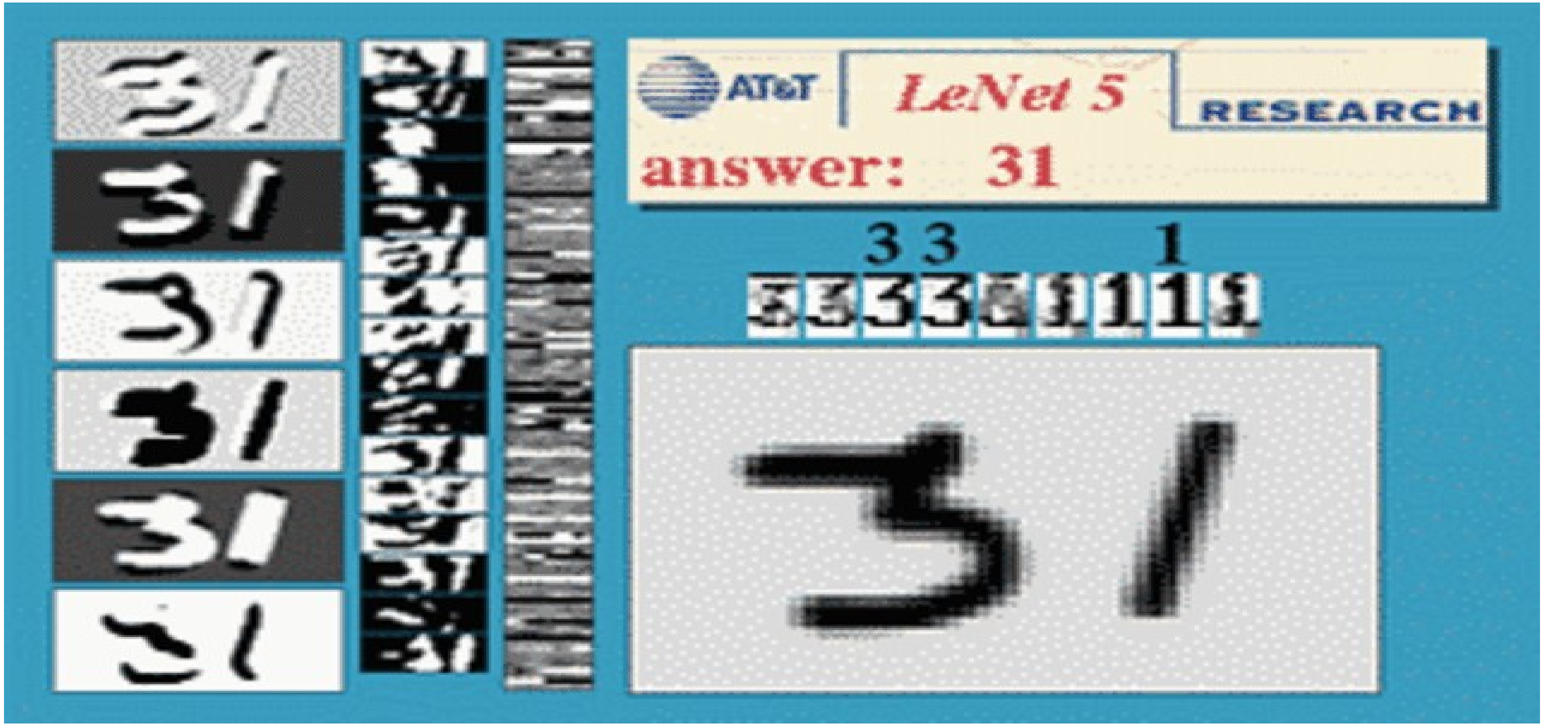

[Fukushima 1982][LeCun 1989, 1998],[Riesenhuber 1999]......

# Convolutional Network (LeNet5, vintage 1990)

- **Filters-tanh → pooling → filters-tanh → pooling → filters-tanh**

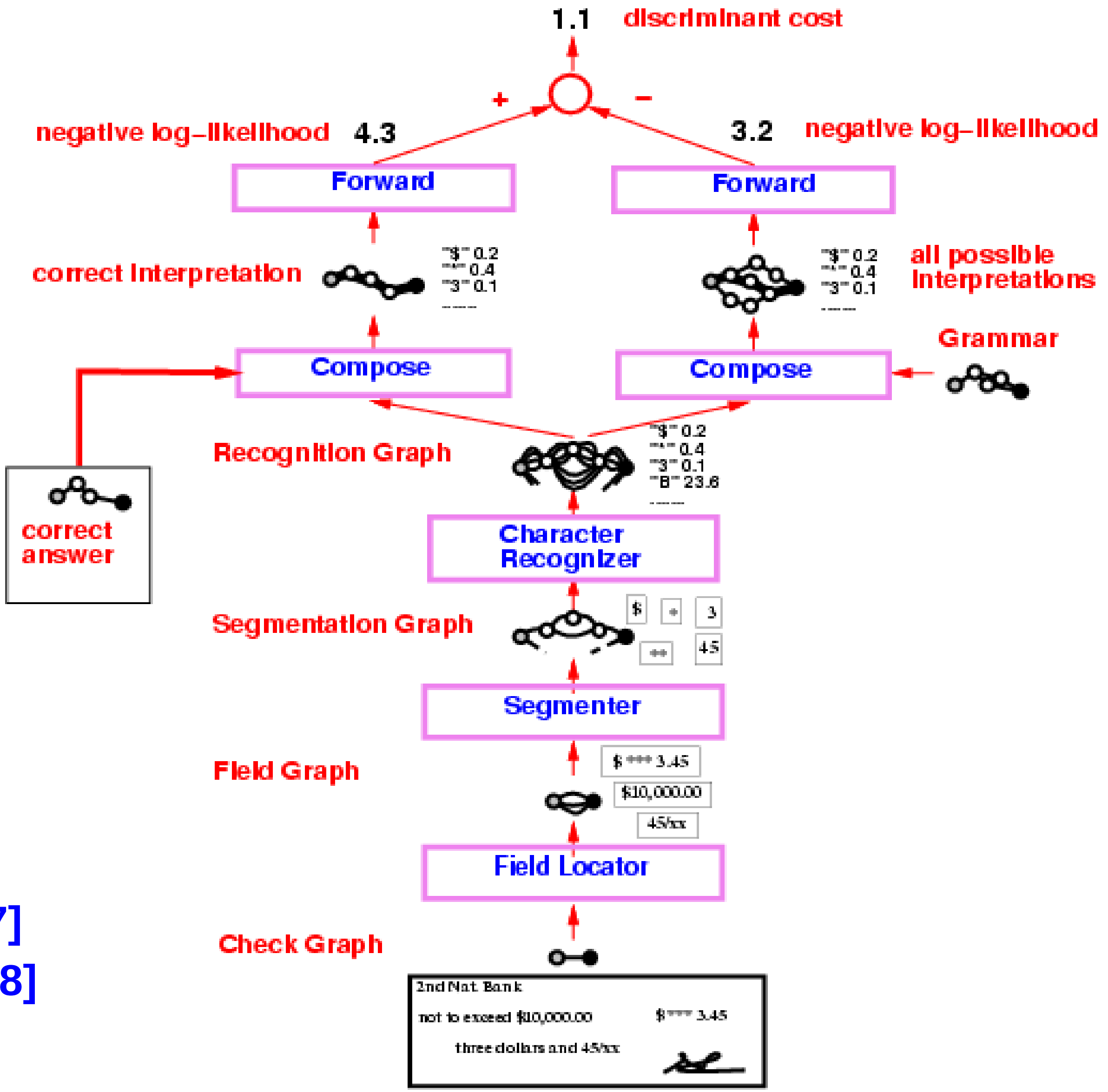
# ConvNets can recognize multiple objects

- **All layers are convolutional**
- **Networks performs simultaneous segmentation and recognition**



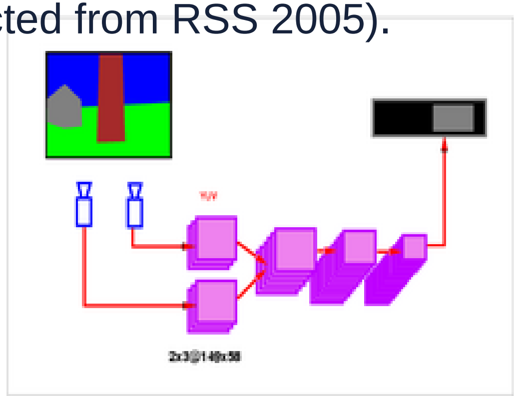
# Check Reader (AT&T 1995)

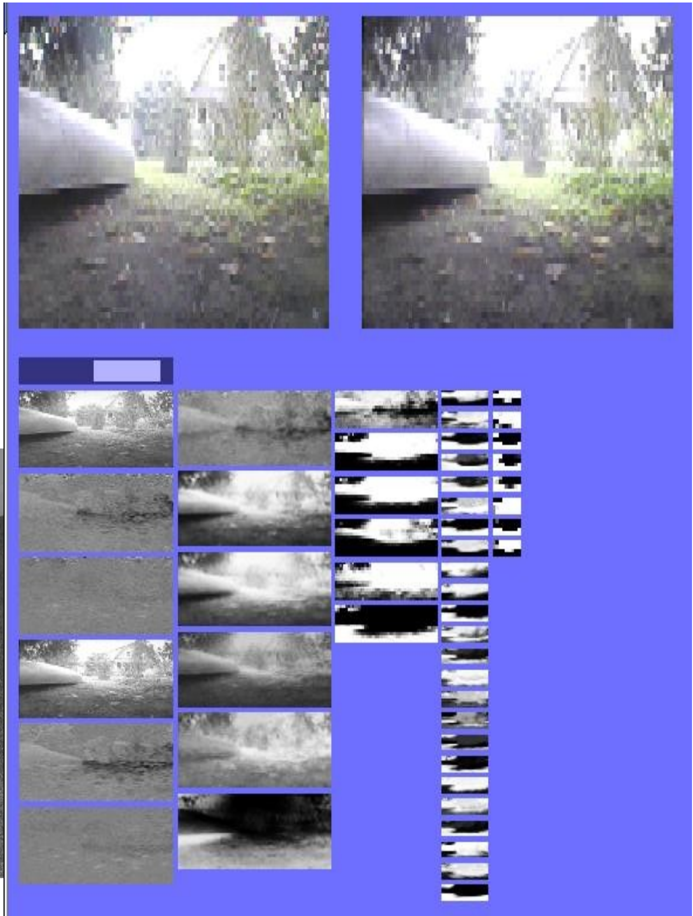
- Graph transformer network trained to read **check amounts.**
- Trained globally with Negative-Log-Likelihood loss (MMI).
- 50% percent correct, 49% reject, 1% error (detectable later in the process).
- **Fielded in 1996**, used in many banks in the US and Europe.
- Processed an estimated **10% to 20% of all the checks written in the US in the early 2000s.**
- [LeCun, Bottou, Bengio ICASSP1997]
  [LeCun, Bottou, Bengio, Haffner 1998]



# DAVE: obstacle avoidance through imitation learning

- **Fall 2003 project at Net-Scale Technologies (Urs Muller)**
  - [LeCun et al. NIPS 2005] (rejected from RSS 2005).
- **Human driver data**
- **Image →[convnet]→steering**
- **20 minutes of training data**
- **Motivated the DARPA LAGR project**

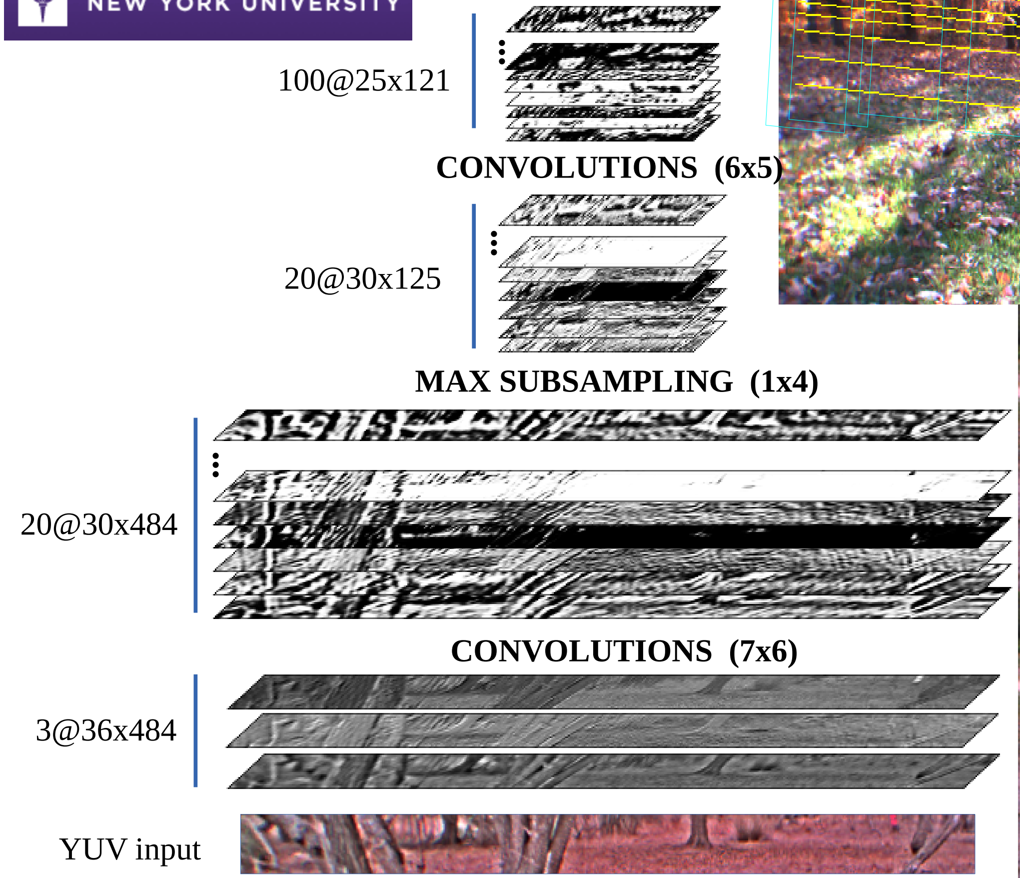


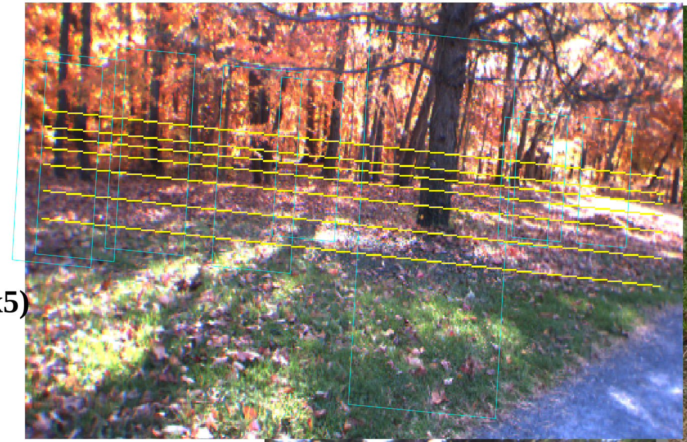



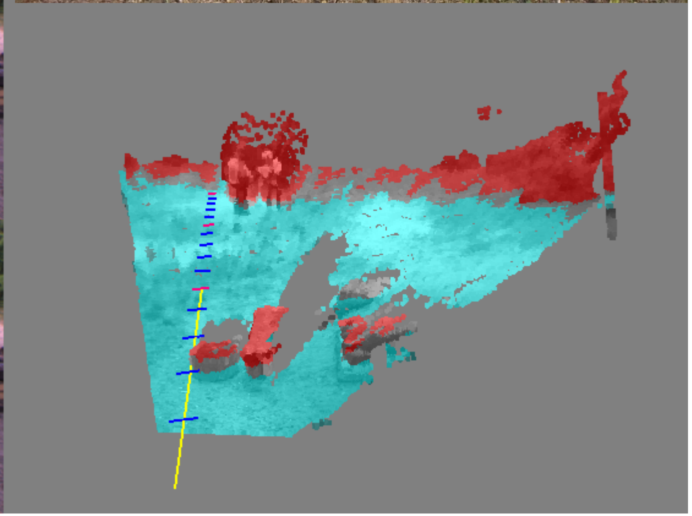
# DARPA LAGR: Learning Applied to Ground Robots

net SCALE Technologies, Inc.

NEW YORK UNIVERSITY

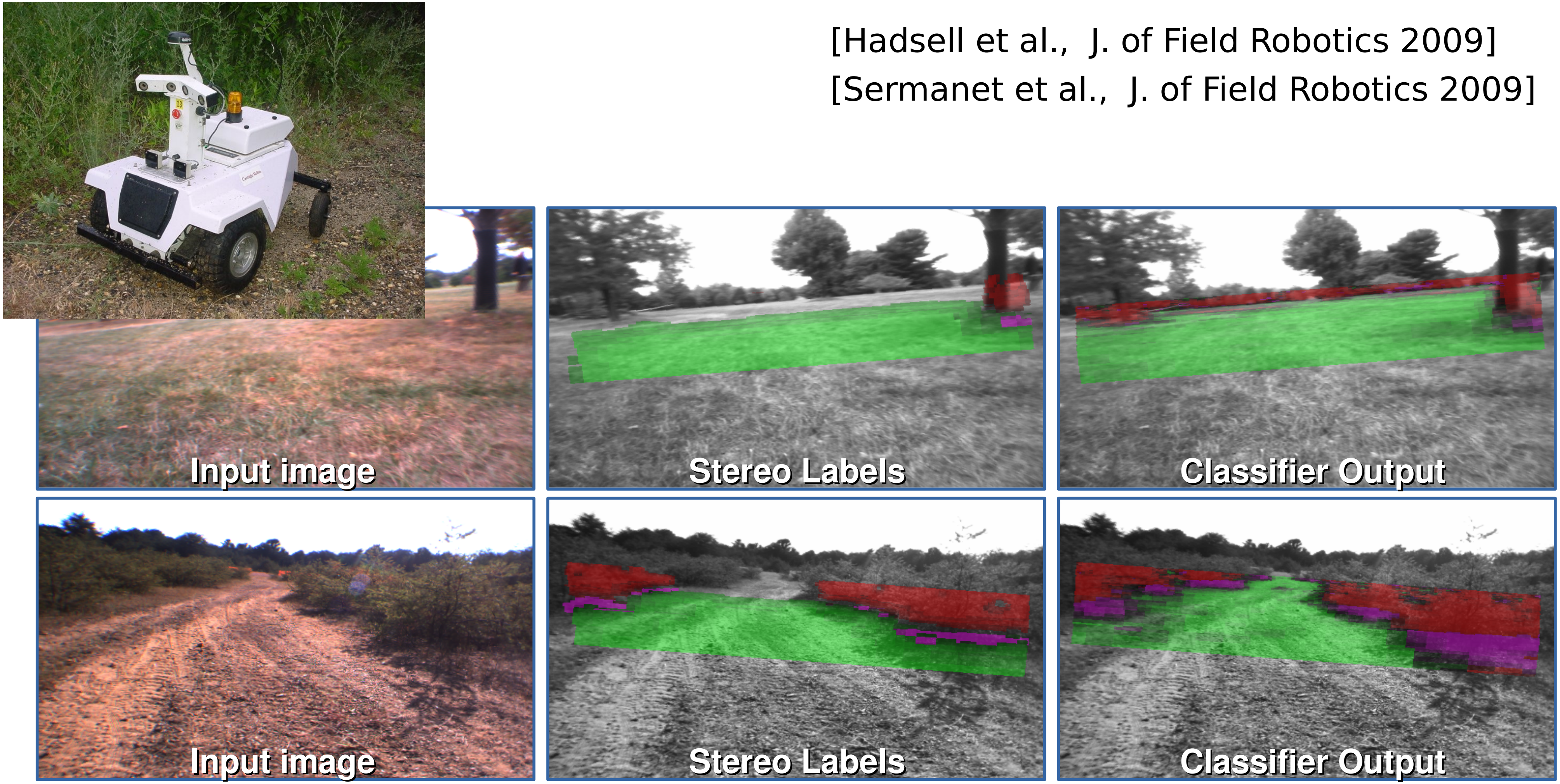


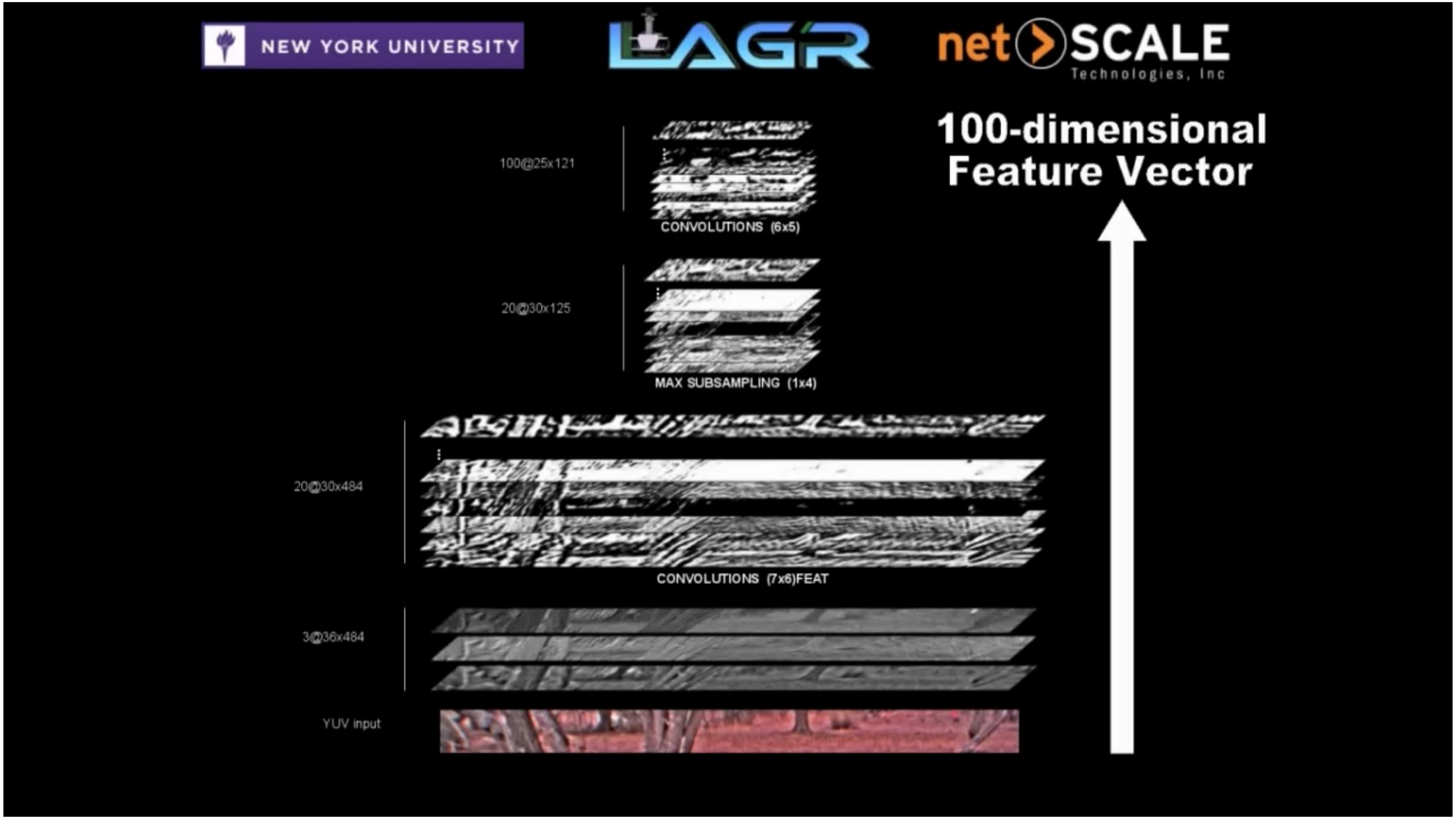


# Semantic Segmentation with ConvNet for off-Road Driving

[Hadsell et al., J. of Field Robotics 2009]
[Sermanet et al., J. of Field Robotics 2009]



# LAGR Video


NEW YORK UNIVERSITY
LAGR
net SCALE
Technologies, Inc
100-dimensional
Feature Vector
100@25x121
CONVOLUTIONS (6x5)
20@30x125
MAX SUBSAMPLING (1x4)
20@30x484
CONVOLUTIONS (7x6)FEAT
3@36x484
YUV input

# Semantic Segmentation with ConvNets

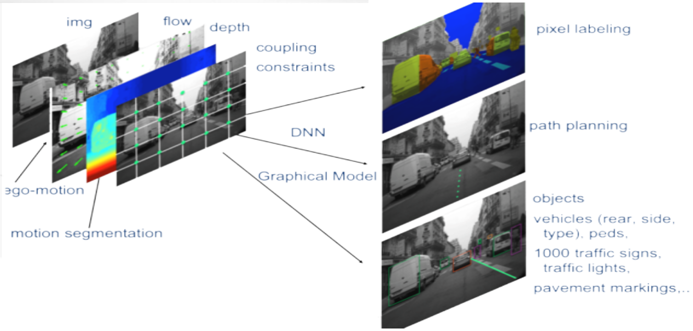
[Farabet et al. ICML 2011]
[Farabet et al. PAMI 2013]

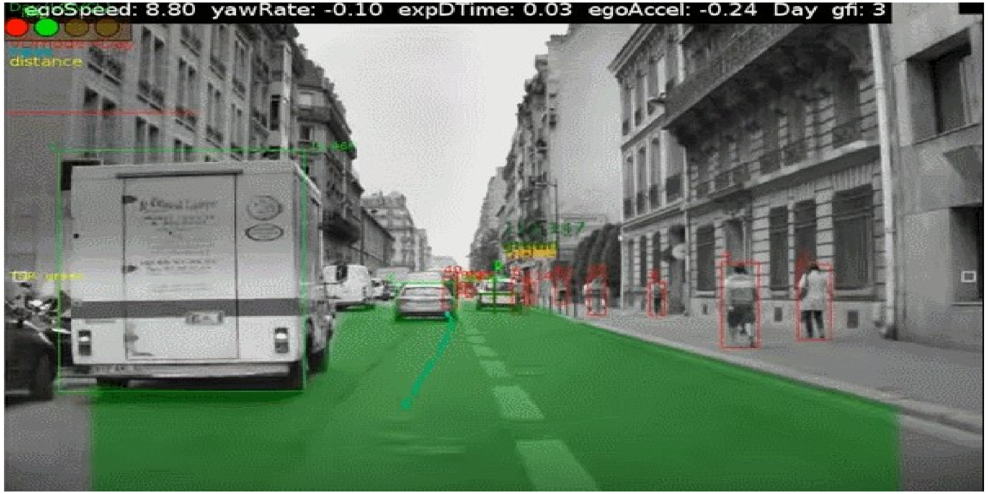
VIDEO

# Semantic Segmentation with ConvNets (33 categories)

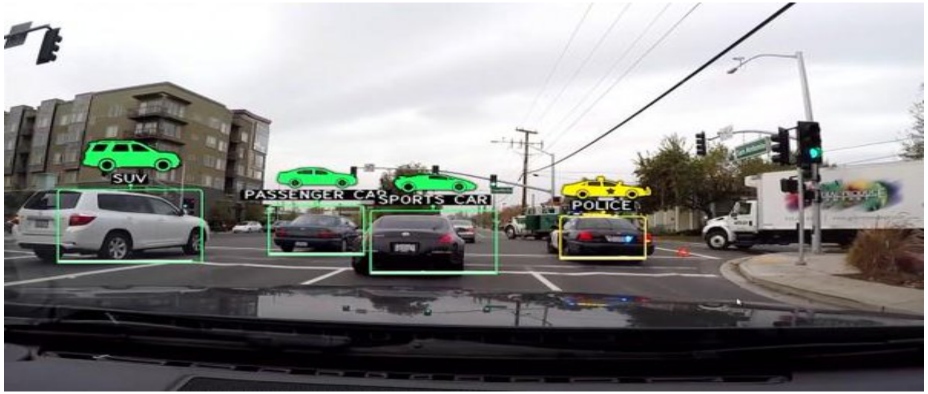
# Driving Cars with Convolutional Nets

- **MobilEye**



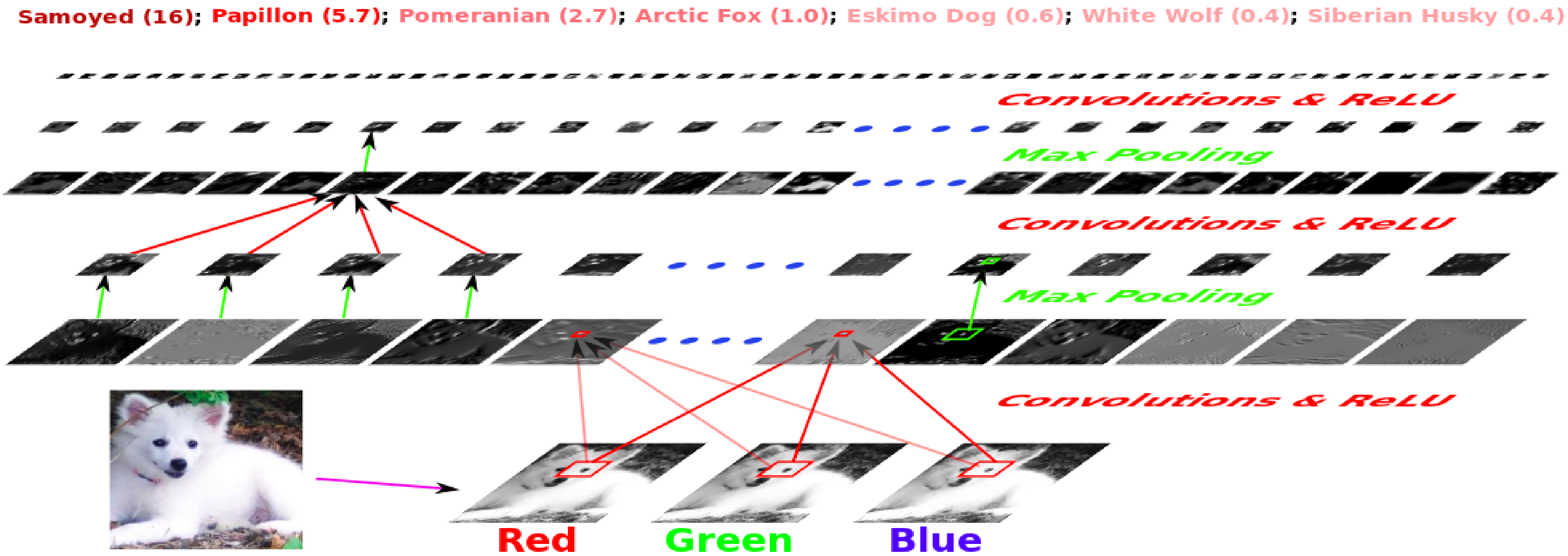




- **NVIDIA**




# Deep Convolutional Nets for Object Recognition

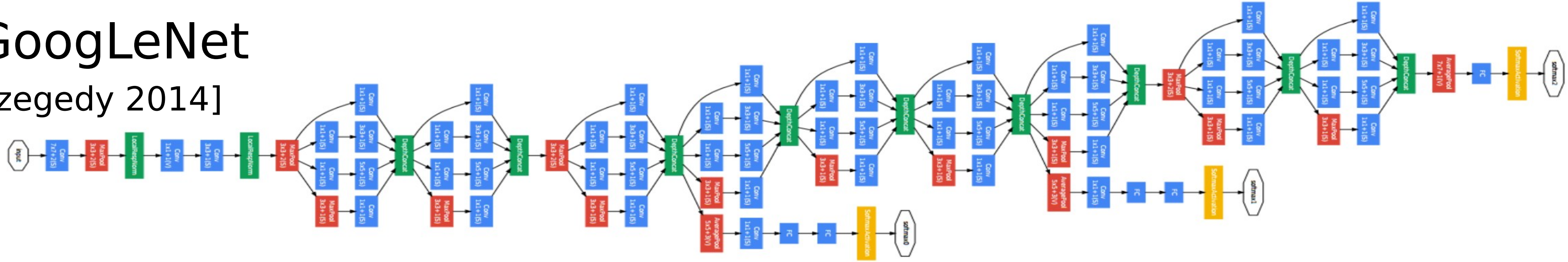
- AlexNet [Krizhevsky et al. NIPS 2012], OverFeat [Sermanet et al. 2013]
- 1 to 10 billion connections, 10 million to 1 billion parameters, 8 to 20 layers.



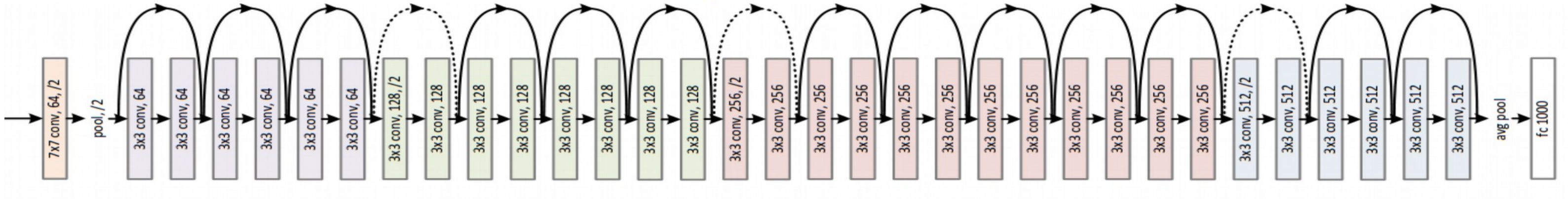
# Deep ConvNets (depth inflation)

VGG
[Simonyan 2013]

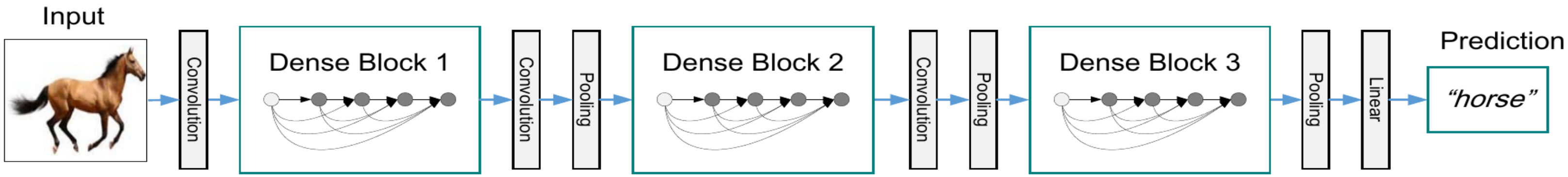

GoogLeNet
Szegedy 2014]



ResNet
[He et al. 2015]

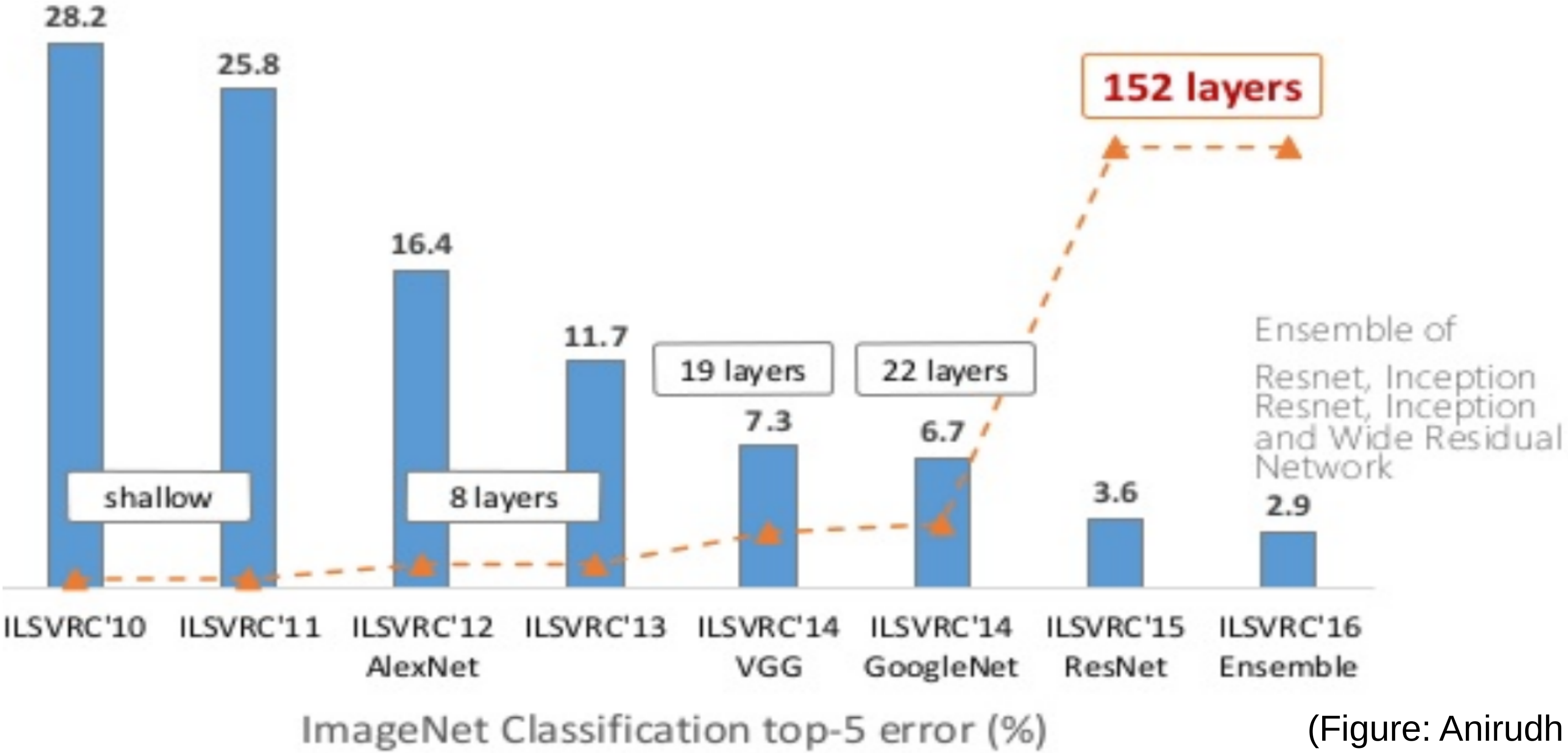



DenseNet
[Huang et al 2017]



# Error Rate on ImageNet

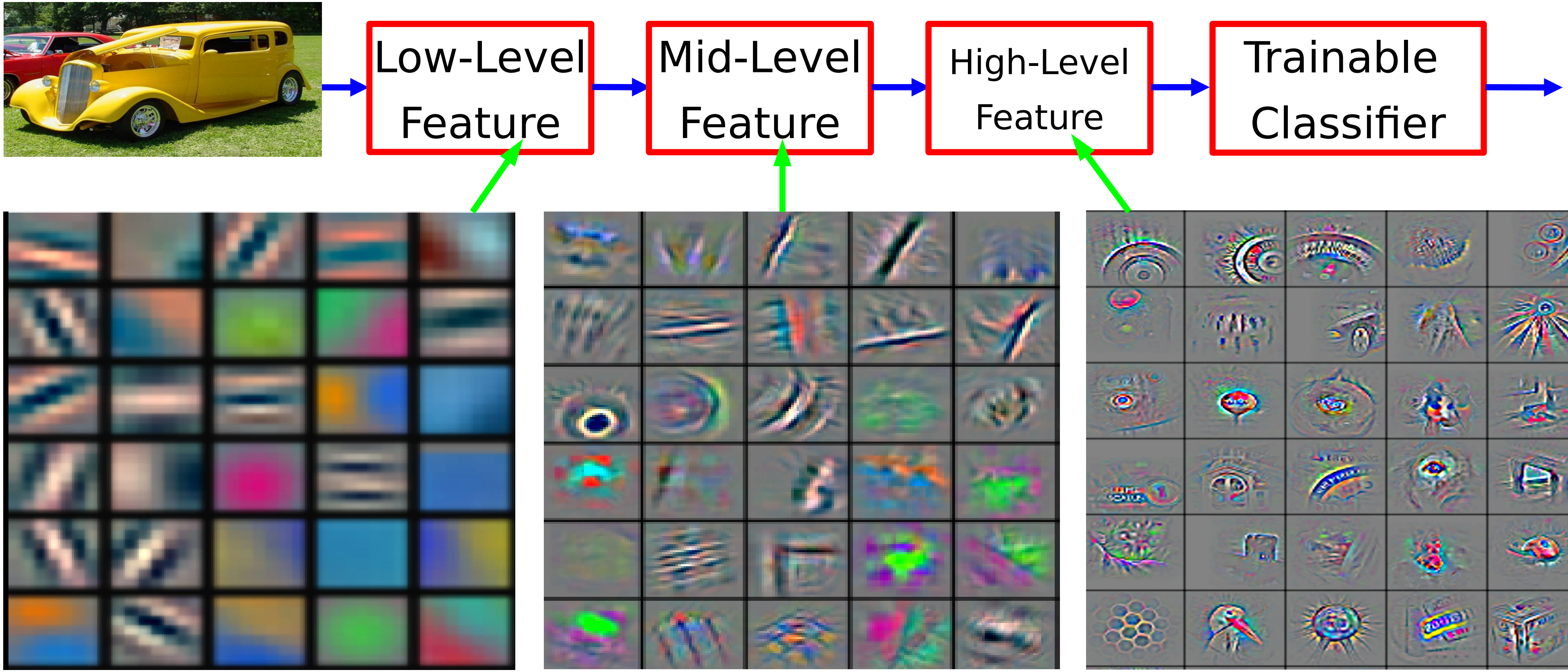
- **Depth inflation**




(Figure: Anirudh Koul)

# Multilayer Architectures == Compositional Structure of Data

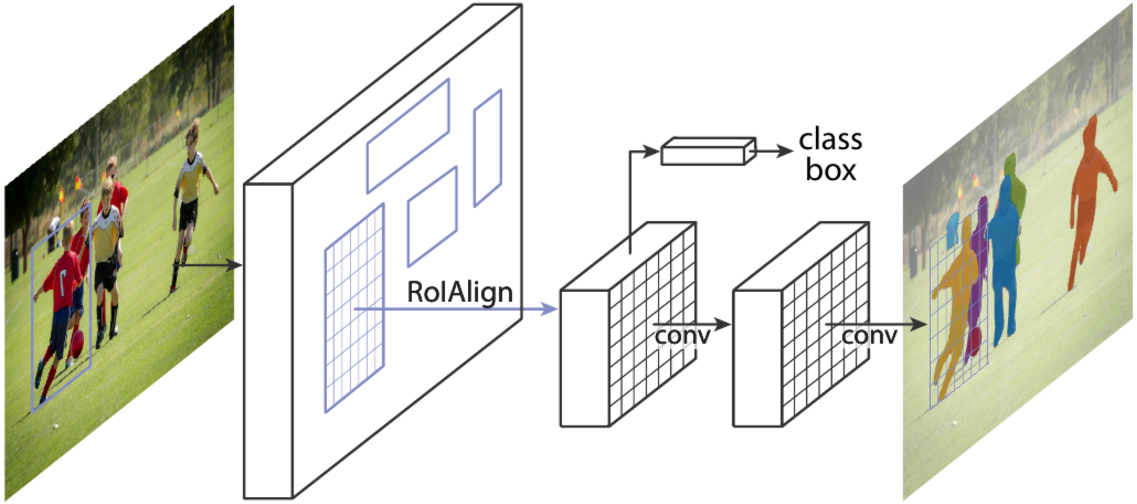
- **Natural is data is compositional => it is efficiently representable hierarchically**




Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

# Mask R-CNN: instance segmentation

- [He, Gkioxari, Dollar, Girshick arXiv:1703.06870]
- ConvNet produces an object mask for each region of interest
- Combined ventral and dorsal pathways




| | backbone | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| MNC [7] | ResNet-101-C4 | 24.6 | 44.3 | 24.8 | 4.7 | 25.9 | 43.6 |
| FCIS [20] +OHEM | ResNet-101-C5-dilated | 29.2 | 49.5 | - | 7.1 | 31.3 | 50.0 |
| FCIS+++ [20] +OHEM | ResNet-101-C5-dilated | 33.6 | 54.5 | - | - | - | - |
| **Mask R-CNN** | ResNet-101-C4 | 33.1 | 54.9 | 34.8 | 12.1 | 35.6 | 51.1 |
| **Mask R-CNN** | ResNet-101-FPN | 35.7 | 58.0 | 37.8 | 15.5 | 38.1 | 52.4 |
| **Mask R-CNN** | ResNeXt-101-FPN | **37.1** | **60.0** | **39.4** | **16.9** | **39.9** | **53.5** |

# Mask-RCNN Results on COCO dataset

- **Individual objects are segmented.**

# Mask-RCNN Results on COCO dataset

- Individual objects are segmented.

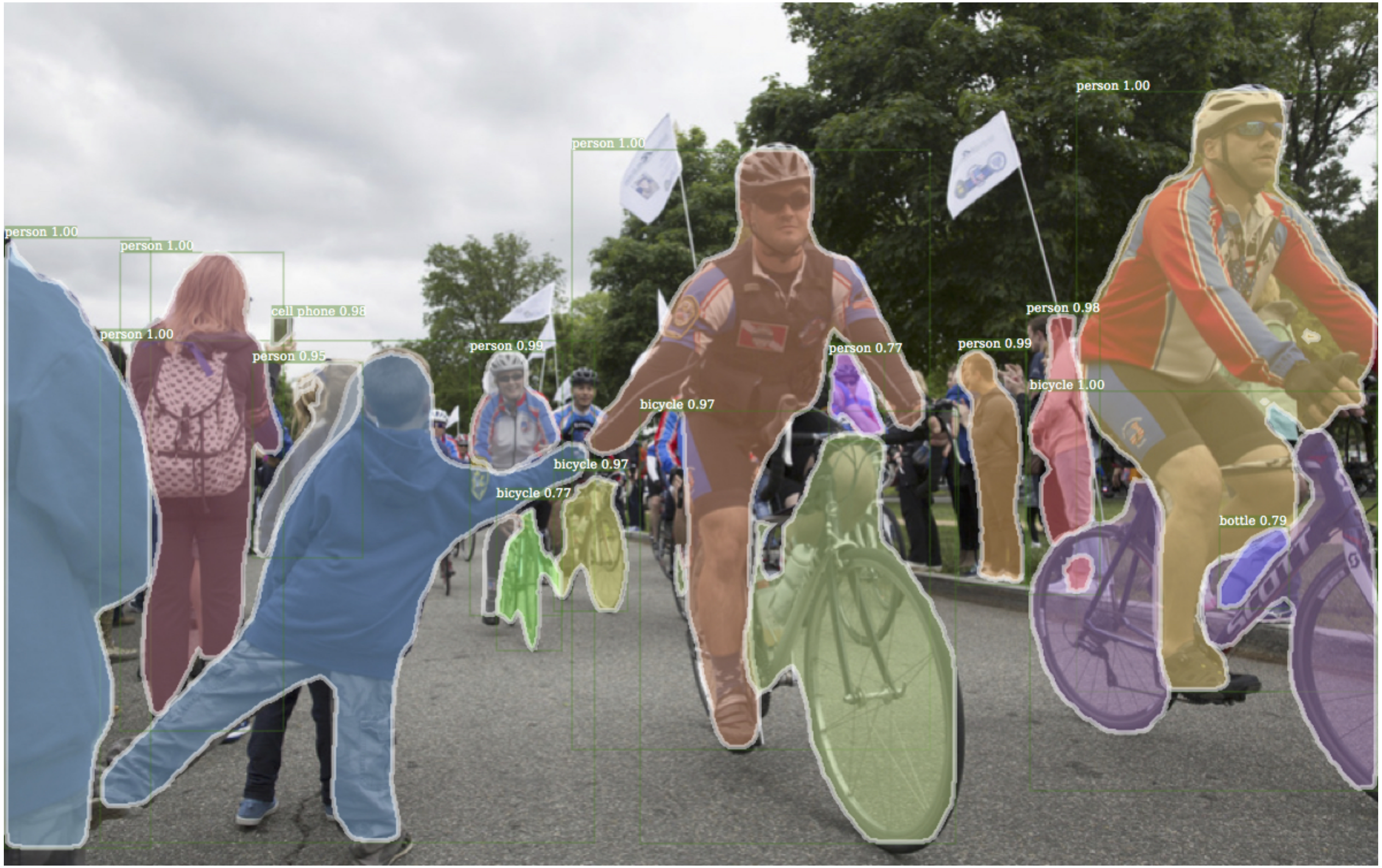
# Mask R-CNN Results on COCO test set

# Mask R-CNN Results on COCO test set

Figure 4. More results of **Mask R-CNN** on COCO test images, using ResNet-101-FPN and running at 5 fps, with 35.7 mask AP (Table 1).

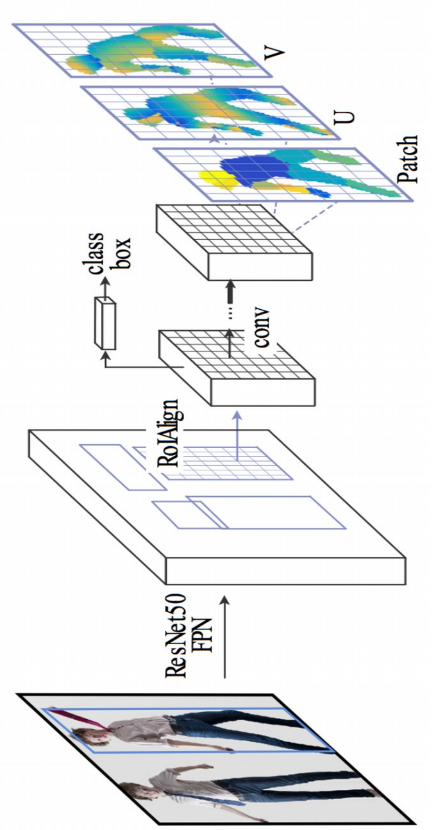
# Detectron: open source vision

https://github.com/facebookresearch/Detectron



# DensePose: real-time body pose estimation

- **[Guler, Neverova, Kokkinos CVPR 2018]** http://densepose.org
  - 20 fps on a single GPU



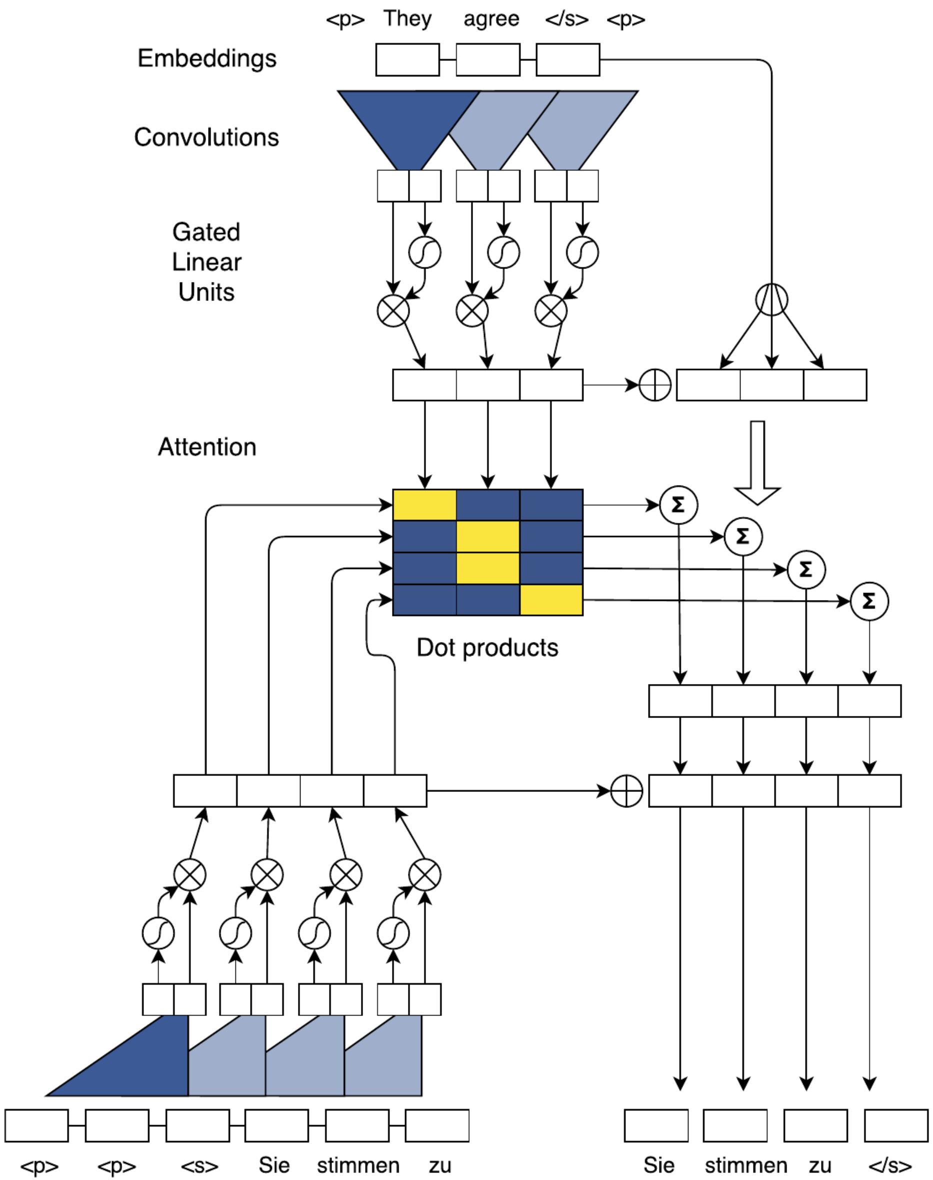

# FairSeq for Translation

- [Gehring et al. ArXiv:1705.03122]

| WMT'16 English-Romanian | BLEU |
|---|---|
| Sennrich et al. (2016b) GRU (BPE 90K) | 28.1 |
| ConvS2S (Word 80K) | 29.45 |
| ConvS2S (BPE 40K) | 29.88 |

| WMT'14 English-German | BLEU |
|---|---|
| Luong et al. (2015) LSTM (Word 50K) | 20.9 |
| Kalchbrenner et al. (2016) ByteNet (Char) | 23.75 |
| Wu et al. (2016) GNMT (Word 80K) | 23.12 |
| Wu et al. (2016) GNMT (Word pieces) | 24.61 |
| ConvS2S (BPE 40K) | 25.16 |

| WMT'14 English-French | BLEU |
|---|---|
| Wu et al. (2016) GNMT (Word 80K) | 37.90 |
| Wu et al. (2016) GNMT (Word pieces) | 38.95 |
| Wu et al. (2016) GNMT (Word pieces) + RL | 39.92 |
| ConvS2S (BPE 40K) | 40.46 |



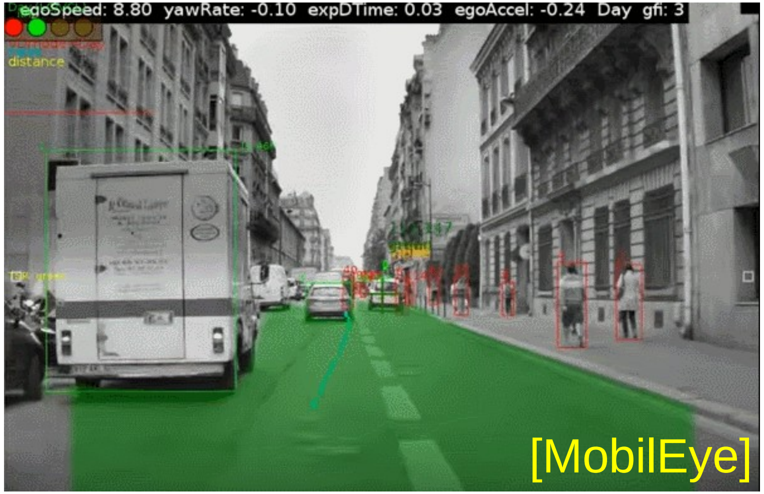
# Applications of ConvNets

- **Self-driving cars, visual perception**
- **Medical signal and image analysis**
  - Radiology, dermatology, EEG/seizure prediction….
- **Bioinformatics/genomics**
- **Speech recognition**
- **Language translation**
- **Image restoration/manipulation/style transfer**
- **Robotics, manipulation**
- **Physics**
  - High-energy physics, astrophysics
- **New applications appear every day**
  - E.g. environmental protection,….

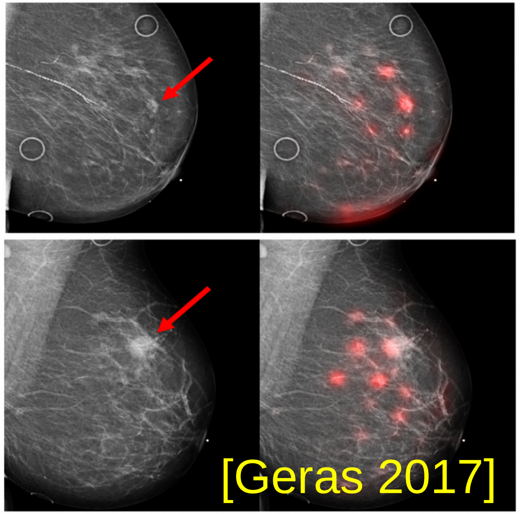
# Applications of Deep Learning

- **Medical image analysis**
- **Self-driving cars**
- **Accessibility**
- **Face recognition**
- **Language translation**
- **Virtual assistants***
- **Content Understanding for:**
  - Filtering
  - Selection/ranking
  - Search
- **Games**
- **Security, anomaly detection**
- **Diagnosis, prediction**
- **Science!**

[Mnih 2015]



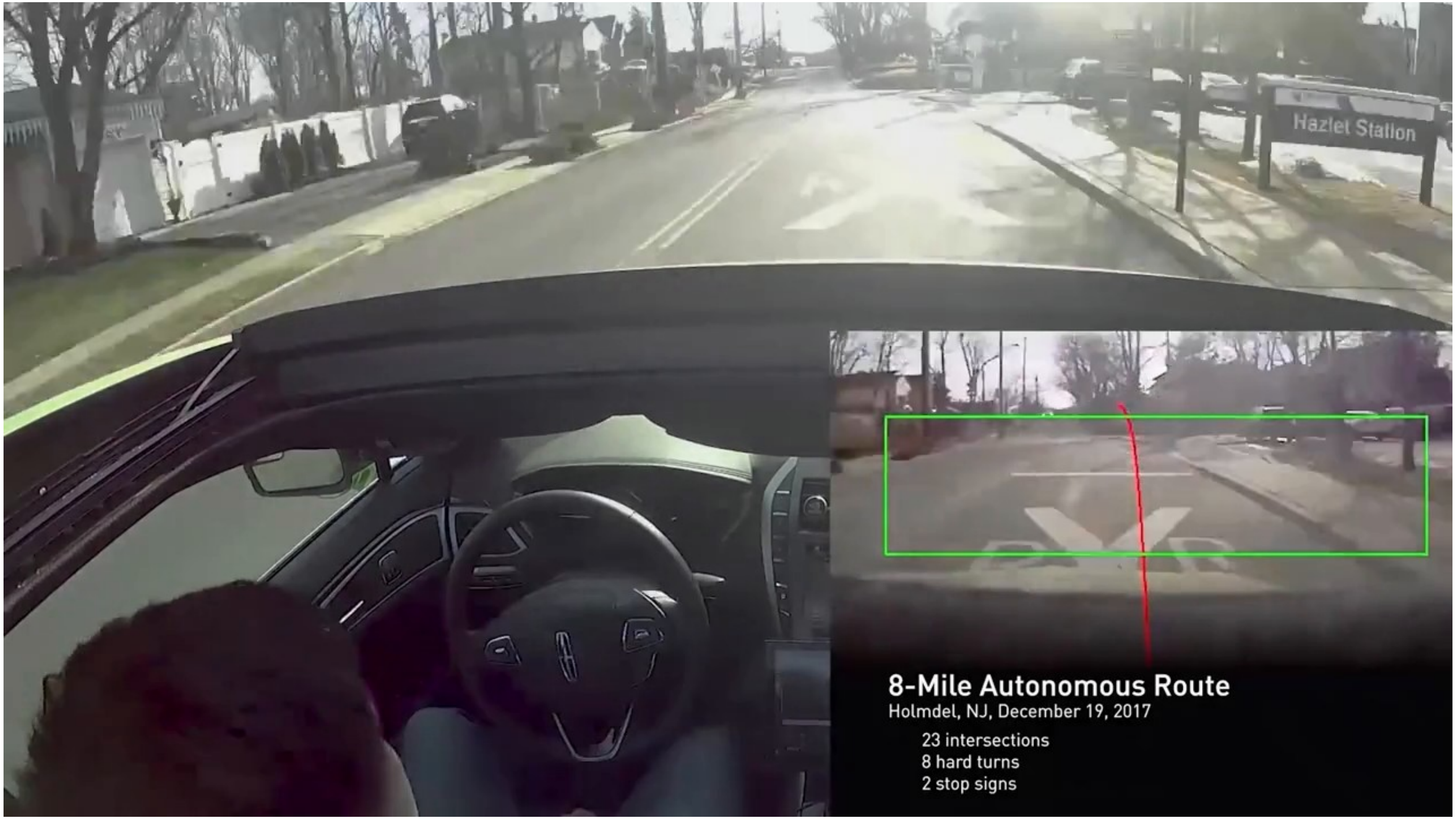
[MobilEye]


[Geras 2017]

[Esteva 2017]

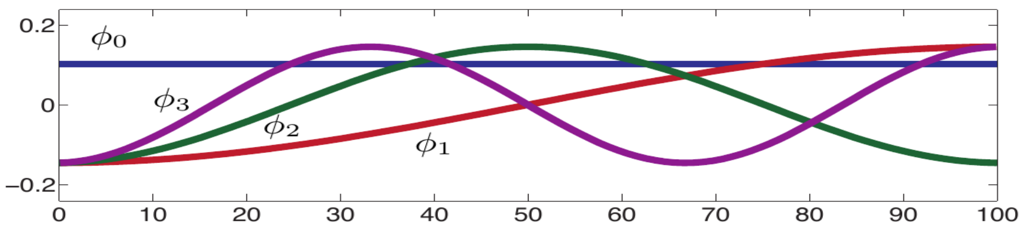
# NVIDIA Autonomous Driving Demo

- **In bucolic New Jersey**

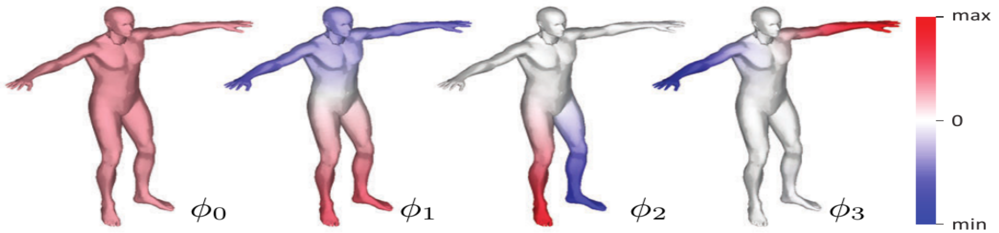


# Spectral Networks: Convolutional Nets on Irregular Graphs

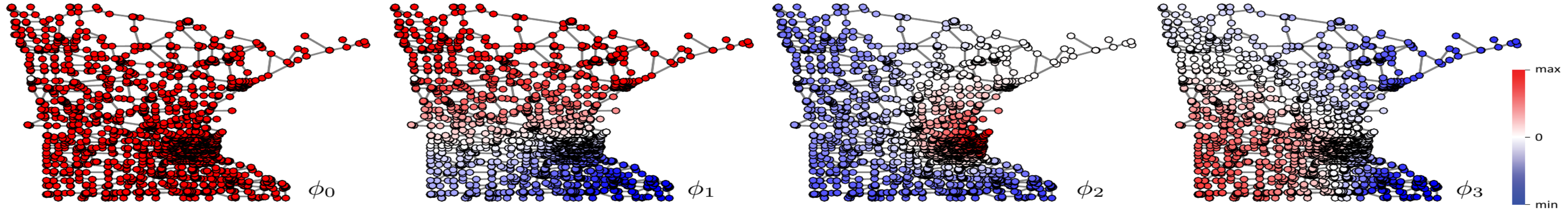
- **Convolutions are diagonal operators in Fourier space**
- **The Fourier space is the eigenspace of the Laplacian**
- **We can compute graph Laplacians**
- **Review paper: [Bronstein et al. 2016, ArXiv:1611.08097]**



Euclidean



Manifold



Graph

# What About (Deep) Reinforcement Learning?

It works great …

…for games and virtual environments

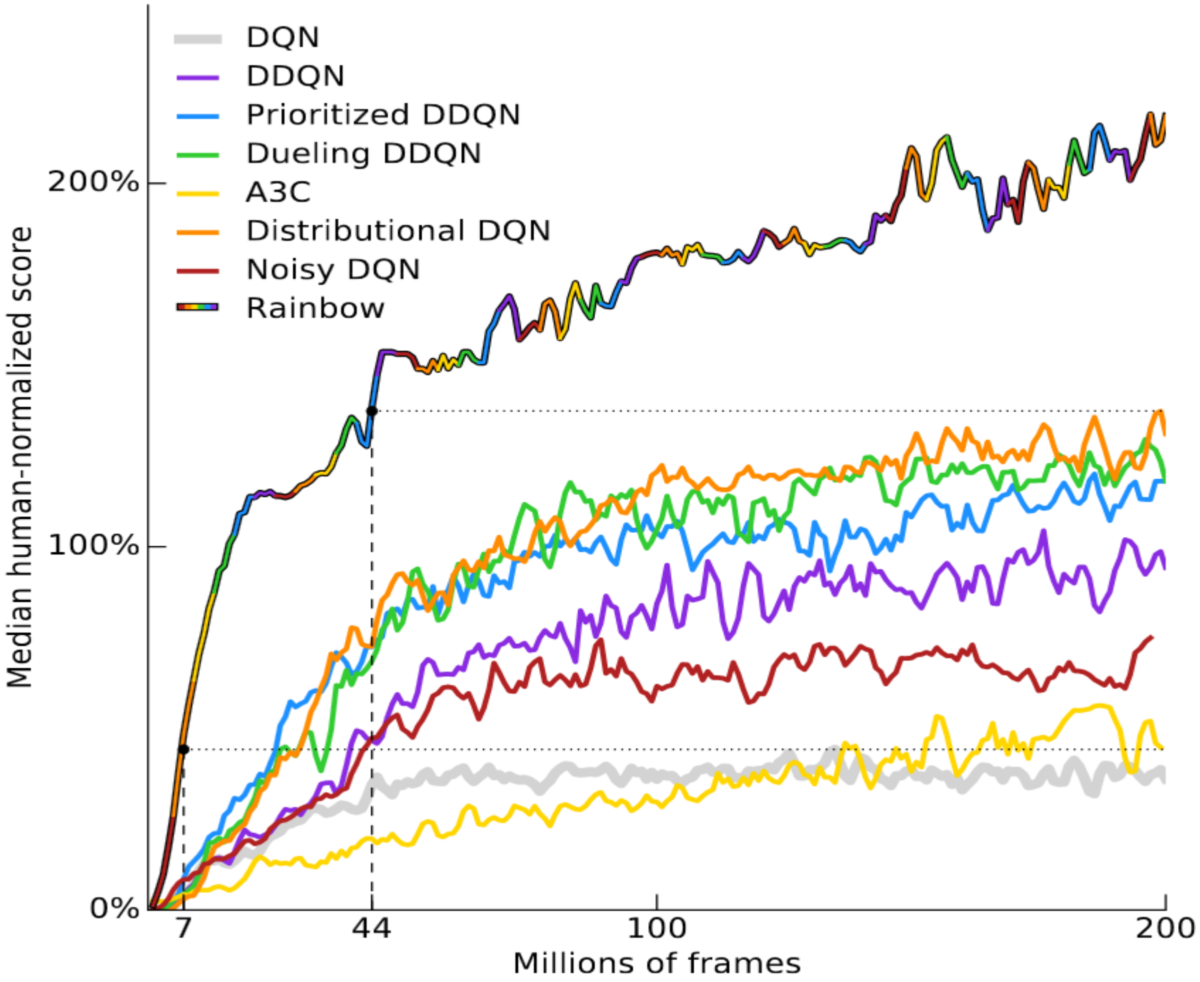
# Reinforcement Learning works fine for games



- **RL works well for games**
  - Playing Atari games [Mnih 2013], Go [Silver 2016, Tian 2018], Doom [Tian 2017], StarCraft (work in progress at FAIR, DeepMind….)
  - RL requires too many trials.
  - RL often doesn’t really work in the real world

# Pure RL requires many, many trials to learn a task

- **[Hessel ArXiv:1710.02298]**
- **Median performance on 57 Atari games relative to human performance (100%=human)**
- **Most methods require over 50 million frames to match human performance (230 hours of play)**
- **The best method (combination) takes 18 million frames (83 hours).**



# Pure RL is hard to use in the real world

- **Pure RL requires too many trials to learn anything**
  - it's OK in a game
  - it's not OK in the real world
- **RL works in simple virtual world that you can run faster than real-time on many machines in parallel.**

- **Anything you do in the real world can kill you**
- **You can't run the real world faster than real time**

# Facebook AI Research

- Open industrial research in the global Internet era
- A new relationship between industry and academia

# Facebook AI Research

- **Created in December 2013**
  - Machine learning, deep learning, AI would become critical to success
  - Mission: advance the science of intelligence and develop technology to apply it.
- **150 scientists, engineers, postdocs, resident PhD students**
  - 50% research scientists, 40% research engineers
- **4 main sites and 3 satellite sites**
  - Main sites: Paris, New York City, Menlo Park, Montréal
  - Satellite sites: Tel Aviv, Pittsburgh, Seattle
  - Good video conference system!

# Facebook AI Research: Modus Operandi

- **Scientist-driven open research**
  - Exploratory research: “bottom up” projects
    - involving a few scientists & students, and sometimes engineers.
  - Larger projects involve more engineering resources.
    - Some are in collaboration with engineering and product groups.
- **Open research**
  - All results are published, and systematically posted on ArXiv.org first
    - then submitted to a conference or journal.
  - Almost all code is open sourced
    - So others in academia and industry can build on it and contribute or collaborate.
  - Few patents.
    - Facebook has a policy of filing patents for defensive purpose only.

# Open Source Projects from FAIR

- **PyTorch: deep learning framework http://pytorch.org**
  - Many examples and tutorials. Used by many research groups.
- **FAISS: fast similarity search (C++/CUDA)**
- **ParlAI: training environment for dialog systems (Python)**
- **ELF: distributed reinforcement learning framework**

- **ELF OpenGo: super-human go-playing engine**
- **FastText: text classification, representation, embedding (C++)**
- **FairSeq: neural machine translation with ConvNets, RNN...**
- **Detectron / Mask-R-CNN: complete vision system**
- **DensePose: real-time body pose tracking system**
- **https://github.com/facebookresearch**

# FAIR: Why Open Research?

- **Why require scientists to publish and open source their code?**
  - it's good for their career and self-image. That's how we can attract the best and most scientifically ambitious people.
  - The results are more believable, reliable and reproducible internally.
  - It makes it easier to convince product groups to develop and deploy technology derived from our research.
  - It makes it easy for other labs to improve on our results
    - and it entices them to be open about it.
  - It's good for the reputation of the company
    - Good for recruiting in engineering divisions.

# FAIR: But wait! Aren't you giving out your best secrets?

- **Being first to invent has prestige value**
  - But that only works if you publish the invention and let people build on it.
- **Being first to deploy has market value and prestige value**
  - Requires an efficient process for tech transfer

    Research → Technology → Products (at scale)
- **Understanding intelligence and making progress in AI...**
  - ...is one of the greatest scientific & technological challenges of our times
    - along with understanding the universe and understanding life.
  - It will take the efforts of the entire research community
    - No single lab, as big as it is, has a monopoly on good ideas.
- **Every entity strives on advances from the whole community**
  - Not just on its own advances.

# Redefining the Academia ↔ Industry Relationship

- **More and more AI researchers in academia are joining industry.**
  - But many share their time between industry and academia
  - 80% / 20% ; 50% / 50% ; 20% / 80%
  - They maintain research and teaching activities at their school.
  - They have labs, PhD students, grants
  - In countries where academic salaries are abysmal, it helps financially
- **This is made possible by open research**
  - Possessive IP policies put barriers between industry and academia
  - Open research makes it easy.
  - Facebook has agreements with schools for resident PhD students, research funding, etc.

# What are we missing?

To get to “real” AI

# What current deep learning methods enables

- **What we can have**
  - Safer cars, autonomous cars
  - Better medical image analysis
  - Personalized medicine
  - Adequate language translation
  - Useful but stupid chatbots
  - Information search, retrieval, filtering
  - Numerous applications in energy, finance, manufacturing, environmental protection, commerce, law, artistic creation, games,…..
- **What we cannot have (yet)**
  - Machines with common sense
  - Intelligent personal assistants
  - "Smart" chatbots"
  - Household robots
  - Agile and dexterous robots
  - Artificial General Intelligence (AGI)

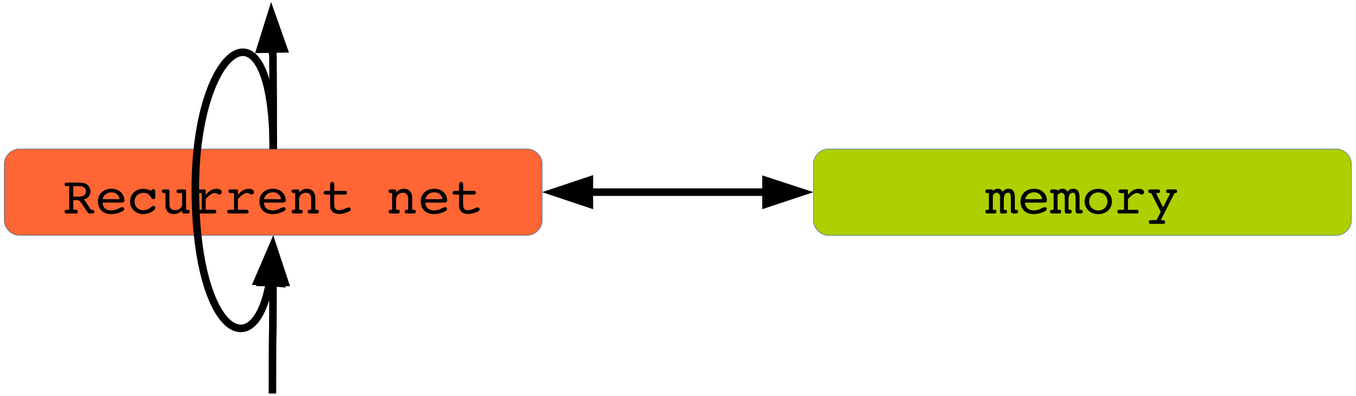
# Differentiable Programming: Marrying Deep Learning With Reasoning

Neural nets with dynamic, data-dependent structure,
A program whose gradient is generated automatically.

facebook
Artificial Intelligence Research

# Augmenting Neural Nets with a Memory Module

- **Recurrent networks cannot remember things for very long**
  - The cortex only remember things for 20 seconds
- **We need a “hippocampus” (a separate memory module)**
  - LSTM [Hochreiter 1997], registers
  - **Memory networks** [Weston et 2014] (FAIR), associative memory
  - **Stacked-Augmented Recurrent Neural Net** [Joulin & Mikolov 2014] (FAIR)
  - **Neural Turing Machine** [Graves 2014],
  - **Differentiable Neural Computer** [Graves 2016]

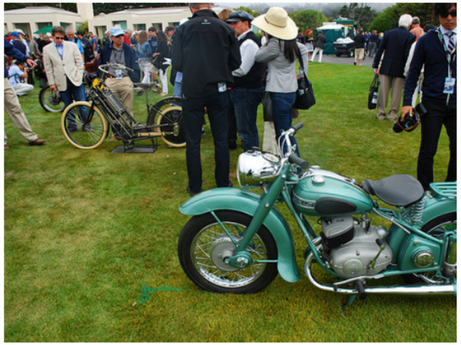


# Answering complex questions by running a program

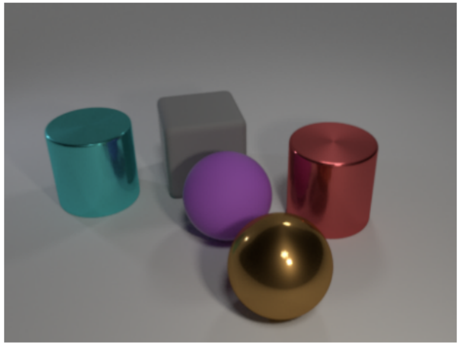
- [Johnson et al. ArXiv:1705.03633]

*How many chairs are at the table?*

*Is there a pedestrian in my lane?*



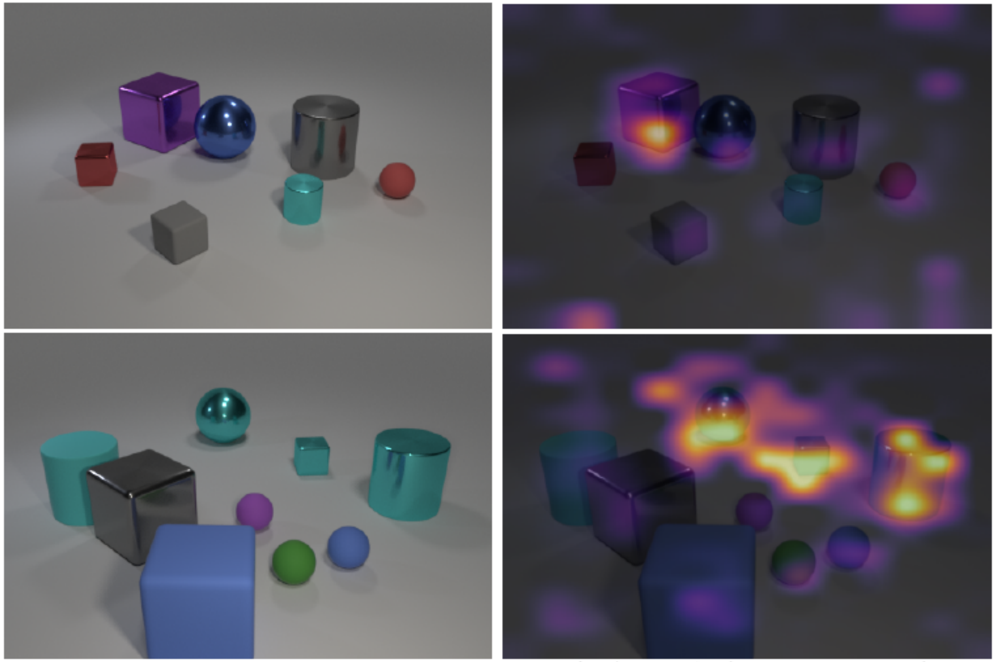
*Is the person with the blue hat touching the bike in the back?*



*Is there a matte cube that has the same size as the red metal object?*

**Q:** *What shape is the…* …*purple thing?*

**A:** *cube*

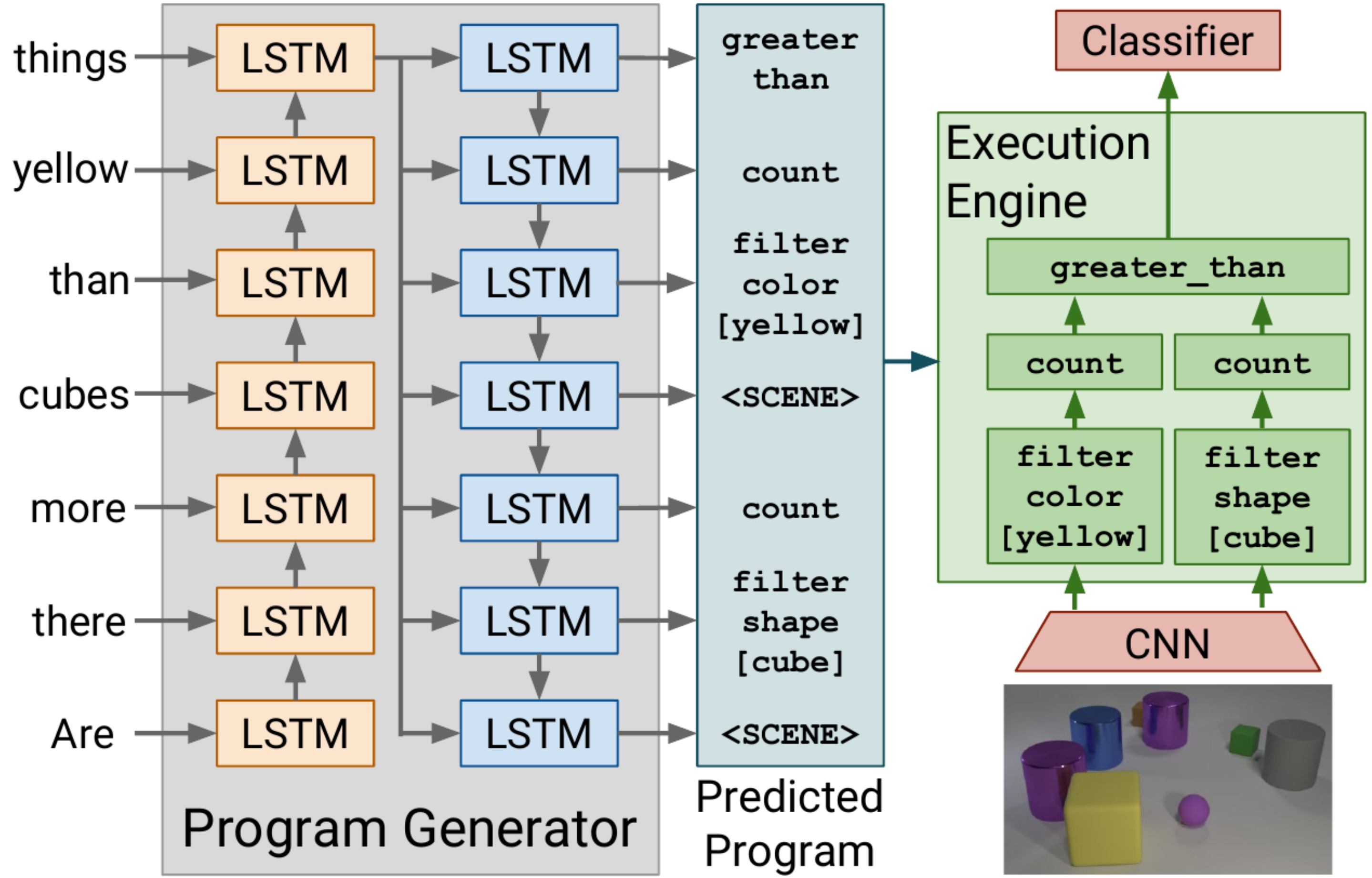


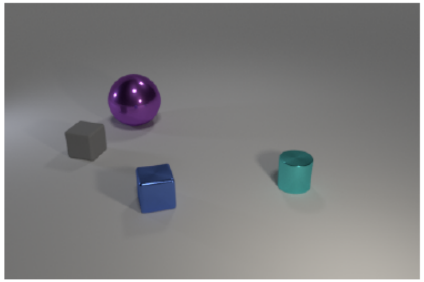
**Q:** *How many cyan things are…* …*right of the gray cube?*

**A:** *3*

# Inferring and executing programs for visual reasoning

**Question**: *Are there more cubes than yellow things?* **Answer**: *Yes*



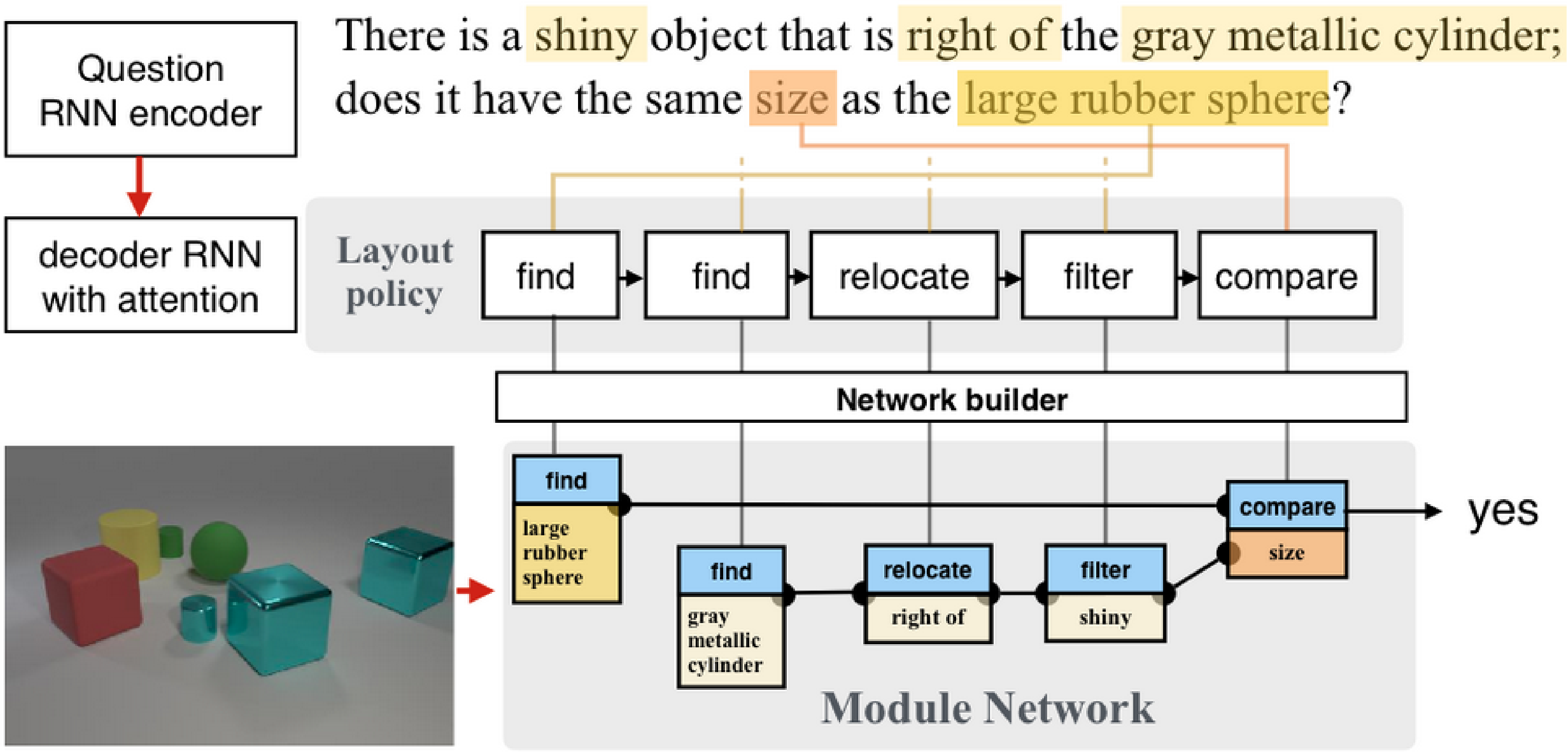
**Q:** *What shape object is* farthest *right?*
**A:** *cylinder*

**Predicted Program:**
query_shape
unique
relate[right]
unique
filter_shape[cylinder
filter_color[blue]
scene

**Predicted Answer:**
✓ *cylinder*

# Inferring and executing programs for visual reasoning

https://research.fb.com/visual-reasoning-and-dialog-towards-natural-language-conversations-about-visual-data/



# PyTorch: differentiable programming

- **Software 2.0:**
  - The operations in a program are only partially specified
  - They are trainable parameterized modules.
  - The precise operations are learned from data, only the general structure of the program is designed.

# How do Humans and Animal Learn?

So quickly

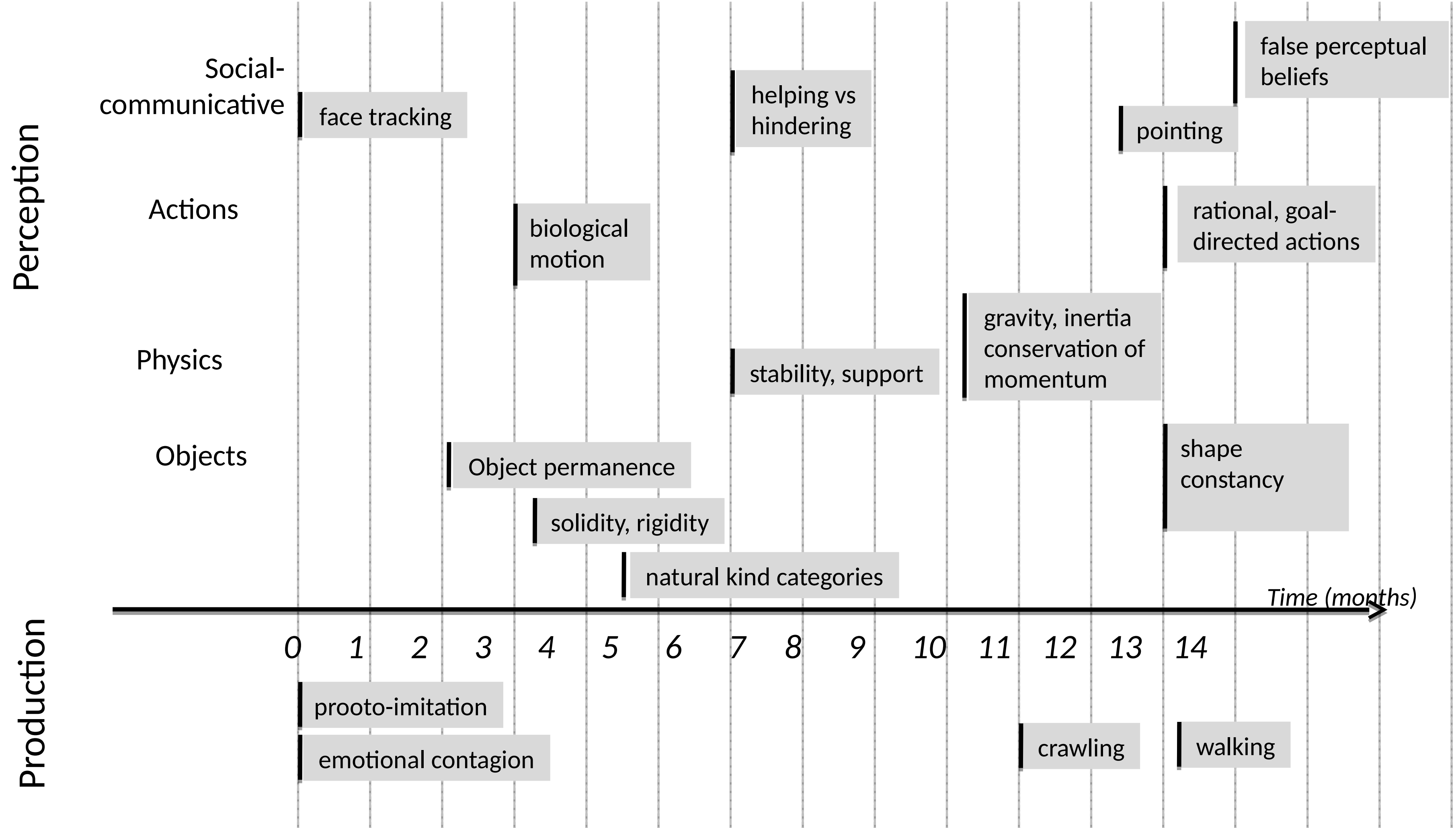
# Babies learn how the world works by observation

- **Largely by observation, with remarkably little interaction.**



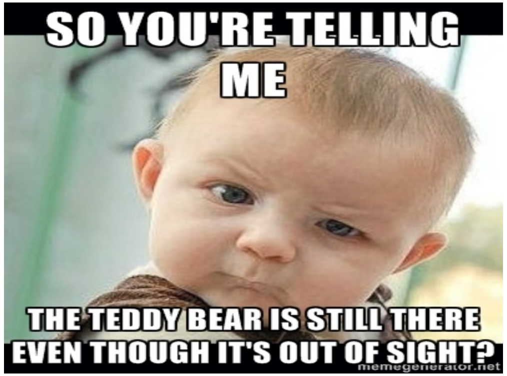
**Photos courtesy of Emmanuel Dupoux**

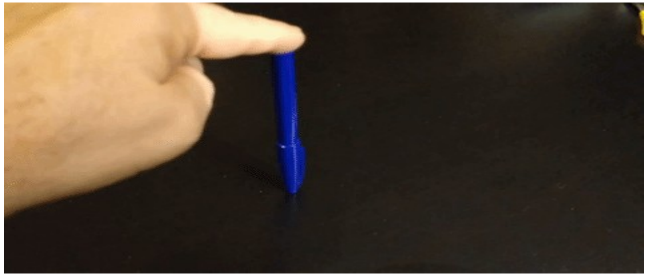
# Early Concept Acquisition [after Emmanuel Dupoux]

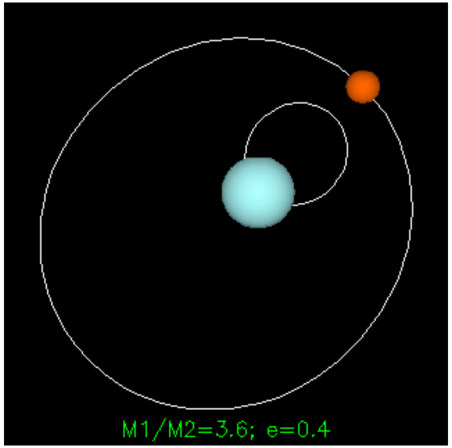


# Prediction is the essence of Intelligence

- **We learn models of the world by predicting**

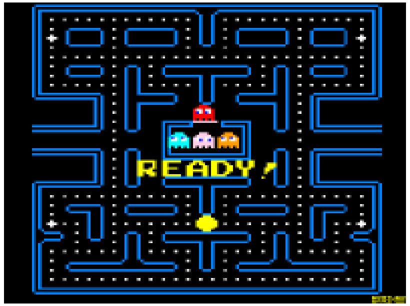


# Three Types of Learning

- **Reinforcement Learning**
  - The machine predicts a scalar reward given once in a while.
  - **weak feedback**
- **Supervised Learning**
  - The machine predicts a category or a few numbers for each input
  - **medium feedback**
- **Self-supervised Predictive Learning**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **A lot of feedback**





# How Much Information is the Machine Given during Learning?

- **“Pure” Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

# Two Big Questions on the way to “Real AI”

- **How can machines learn as efficiently as humans and animals?**
  - By observation
  - without supervision
  - with very little interactions with the world
- **How can we train machines to plan and act (not just perceive)?**
  - Where inference involves a complex iterative process

- **Learning predictive forward models of the world under uncertainty**
  - Learning hierarchical representations of the world unsupervised
  - Enabling long-term planning using the model
  - Enabling learning in the real world with few interactions

# The Next AI Revolution


With thanks
To
Alyosha Efros

# Common Sense is the ability to fill in the blanks

- Infer the state of the world from partial information
- Infer the future from the past and present
- Infer past events from the present state

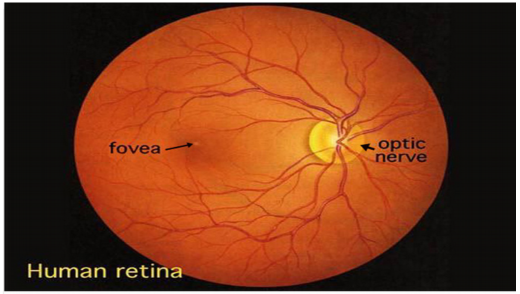


*Fig. 1. Human retina as seen through an opthalmoscope.*

- Filling in the visual field at the retinal blind spot
- Filling in occluded images, missing segments in speech
- Predicting the state of the world from partial (textual) descriptions
- Predicting the consequences of our actions
- Predicting the sequence of actions leading to a result

- **Predicting any part of the past, present or future percepts from whatever information is available.**

- That's what **self-supervised predictive learning** is
- But really, that's what many people mean by unsupervised learning

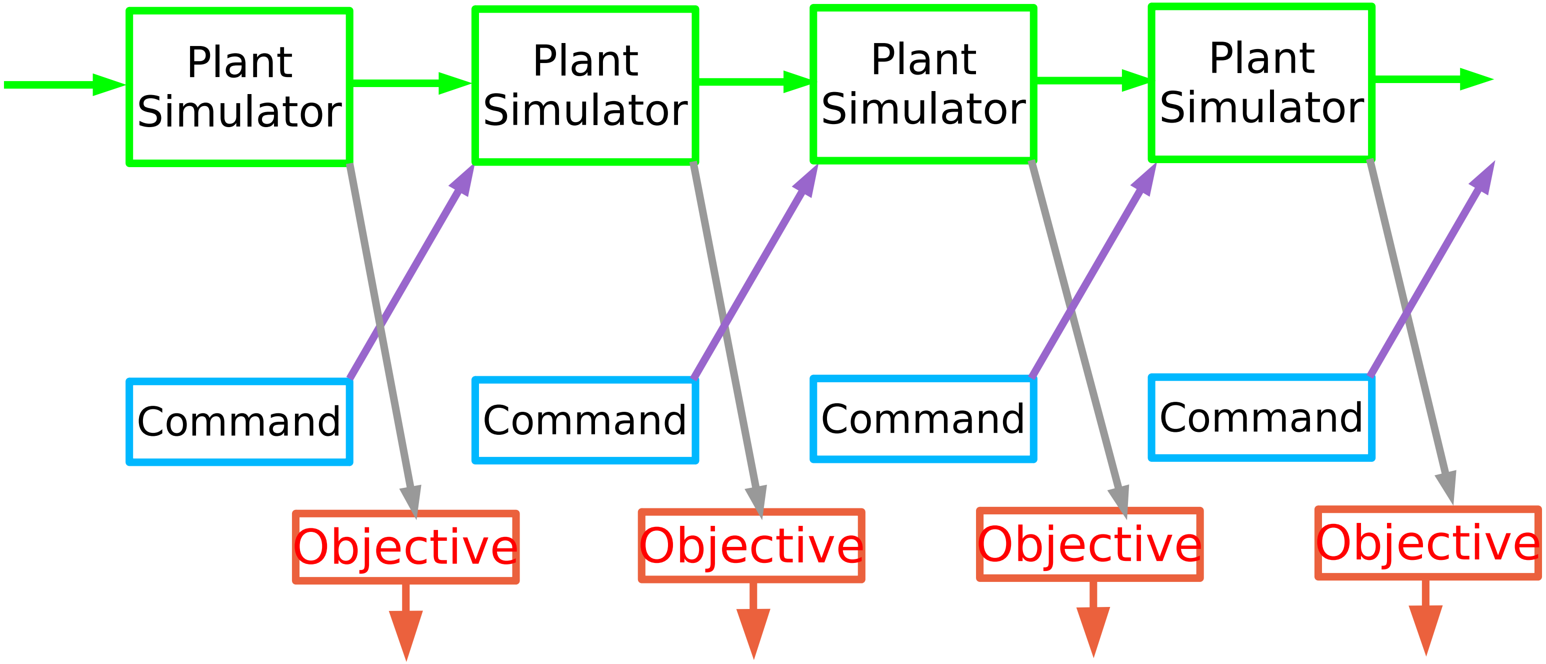
# Learning Predictive Models of the World

## Learning to predict, reason, and plan, Learning Common Sense.

facebook
Artificial Intelligence Research

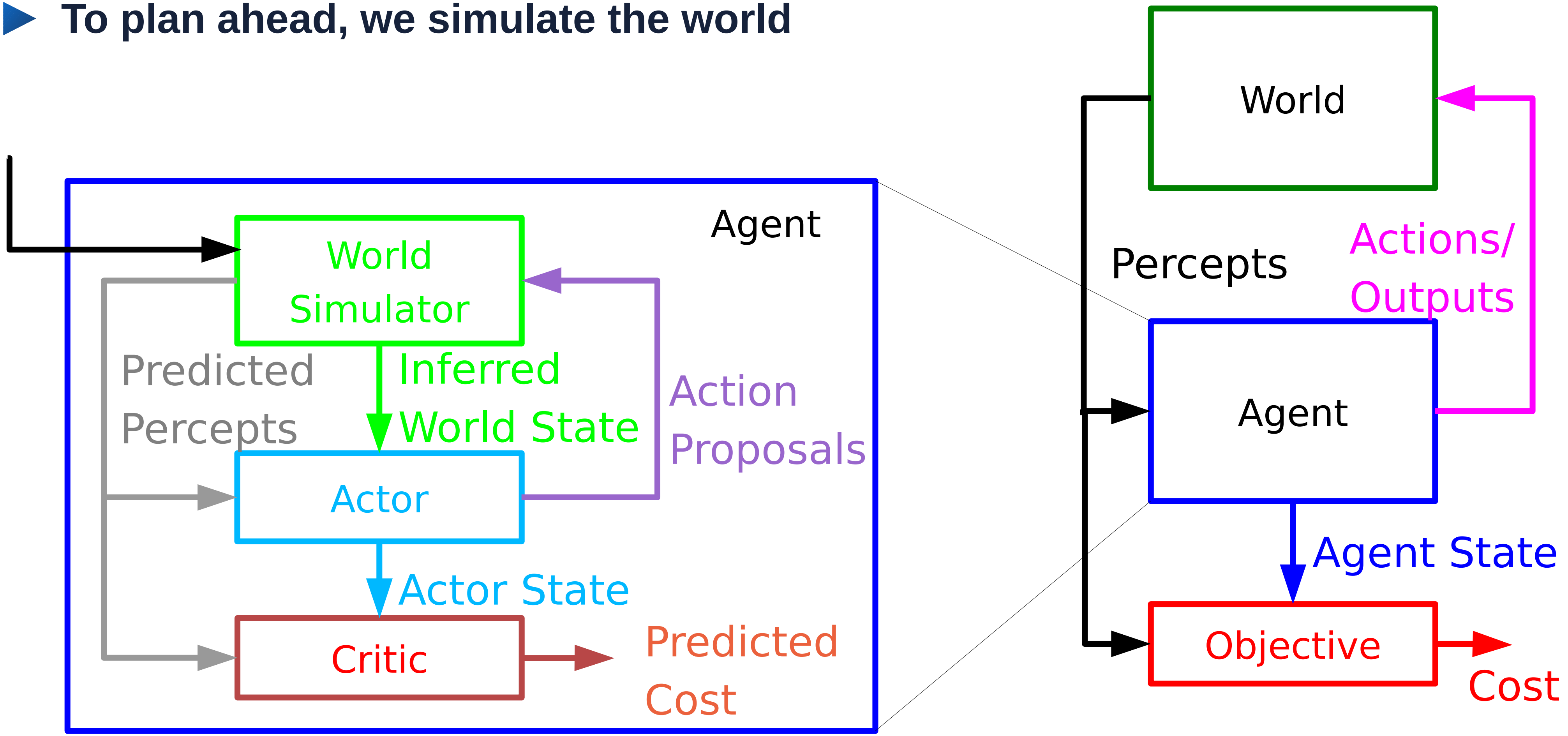
# Classical model-based optimal control

- **Simulate the world (the plant) with an initial control sequence**
- **Adjust the control sequence to optimize the objective through gradient descent**
- **Backprop through time was invented by control theorists in the late 1950s**
  - it's sometimes called the adjoint state method
  - [Athans & Falb 1966, Bryson & Ho 1969]



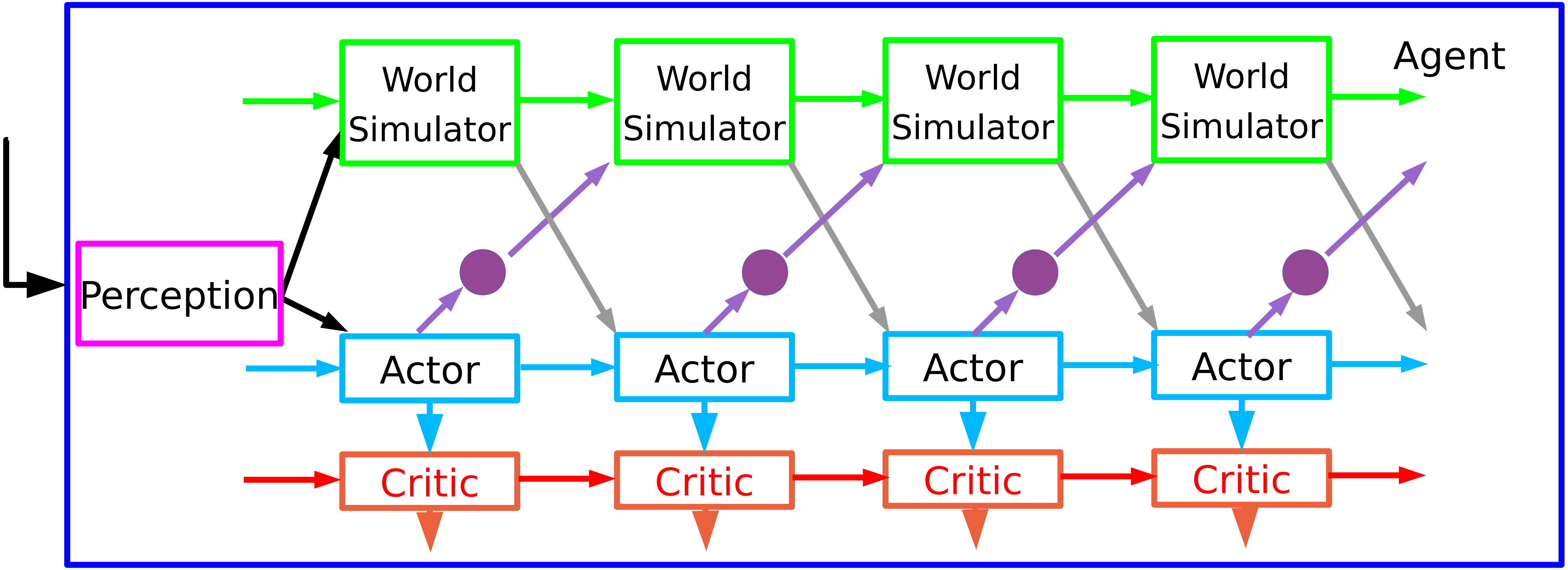
# Planning Requires Prediction

- **To plan ahead, we simulate the world**



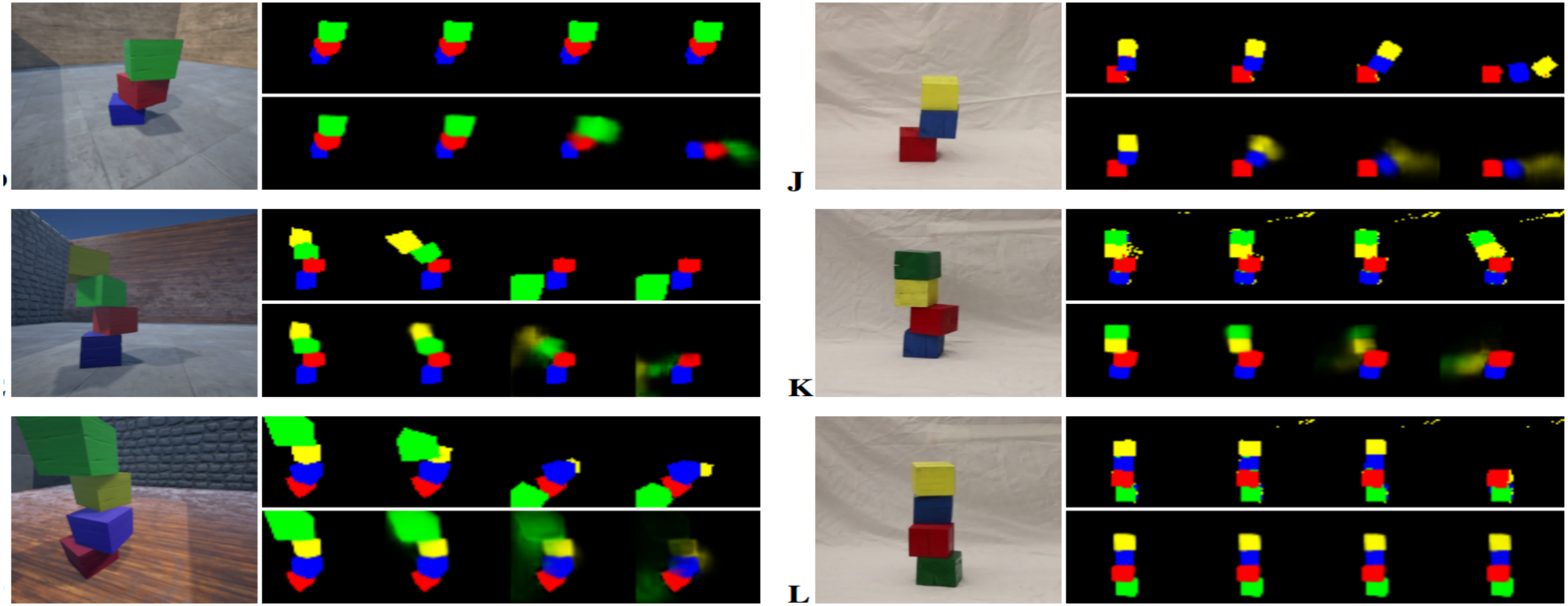
# Training the Actor with Optimized Action Sequences

- **1. Find action sequence through optimization**
- **2. Use sequence as target to train the actor**
  - Over time we get a compact policy that requires no run-time optimization



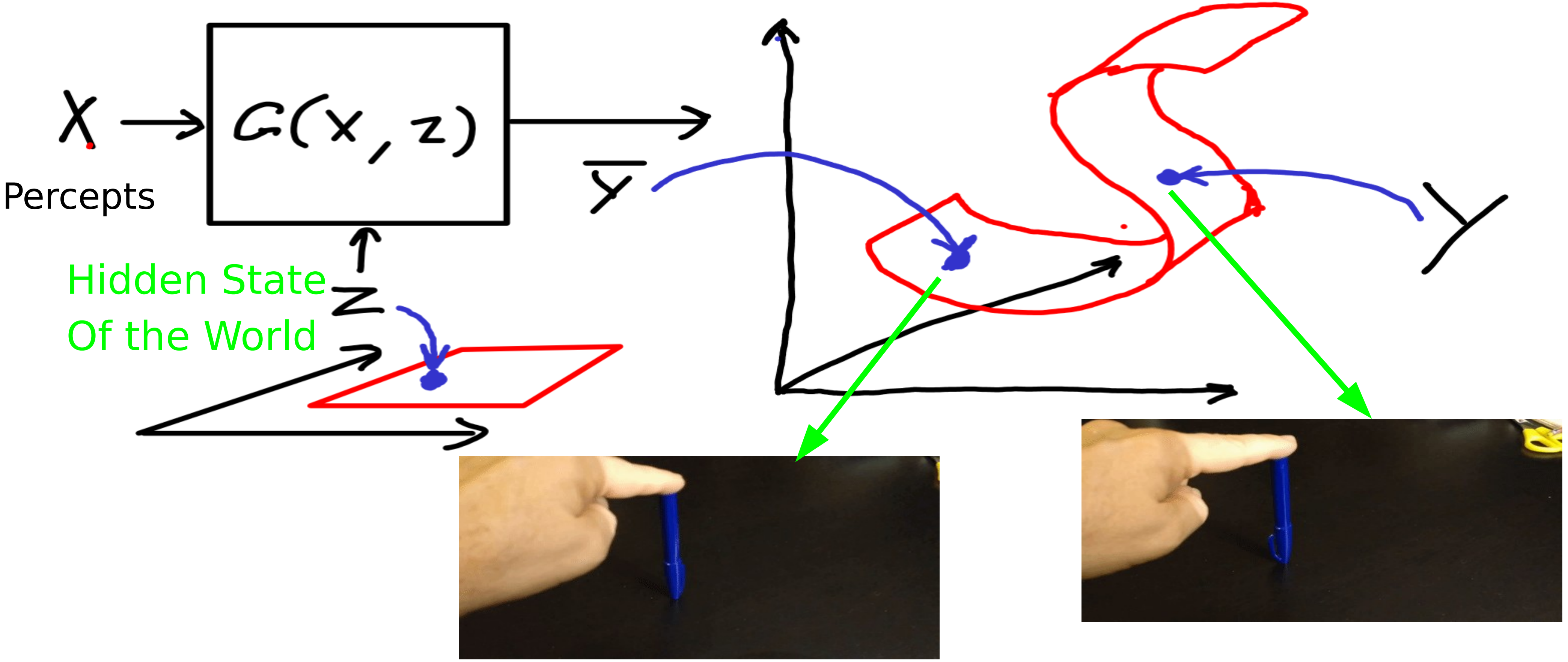
# Learning Physics (PhysNet)

- **[Lerer, Gross, Fergus ICML 2016, arxiv:1603.01312]**
  - ConvNet produces object masks that predict the trajectories of falling blocks. Blurry predictions when uncertain



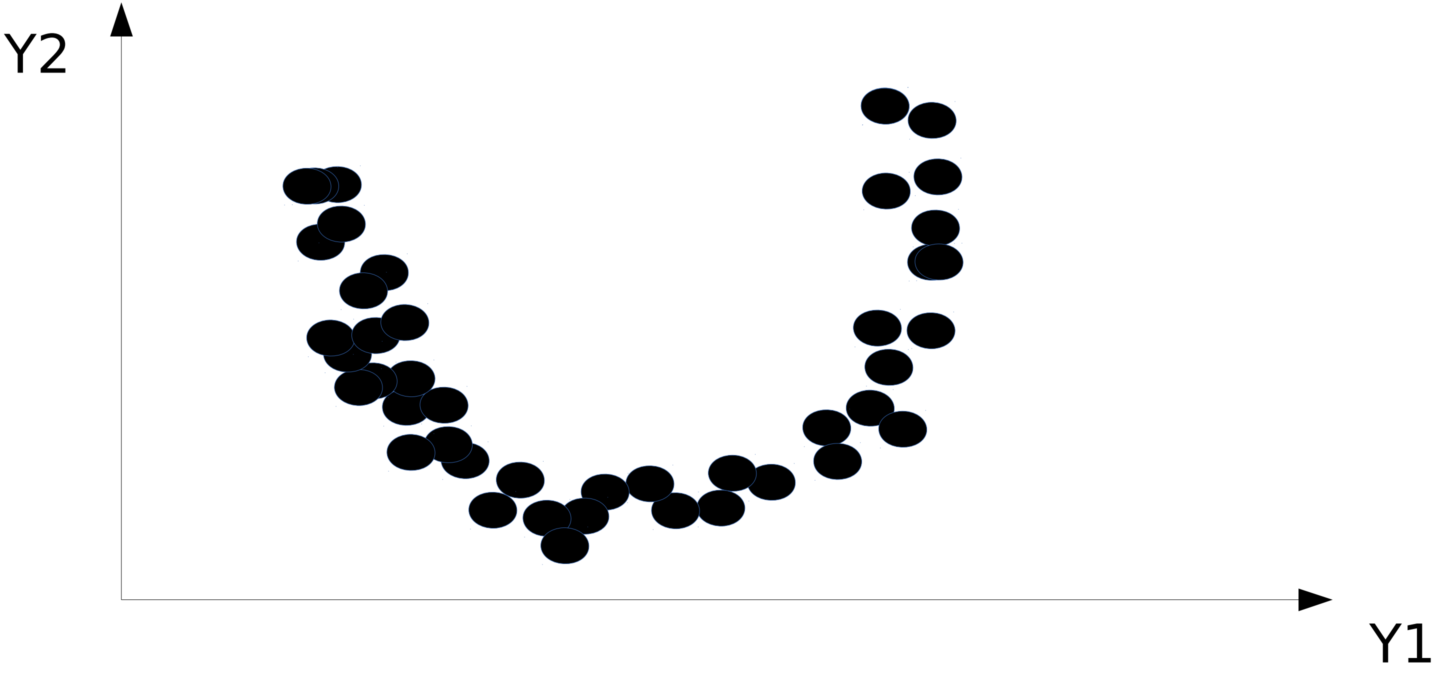
# The Hard Part: Prediction Under Uncertainty

- **Invariant prediction: The training samples are merely representatives of a whole set of possible outputs (e.g. a manifold of outputs).**



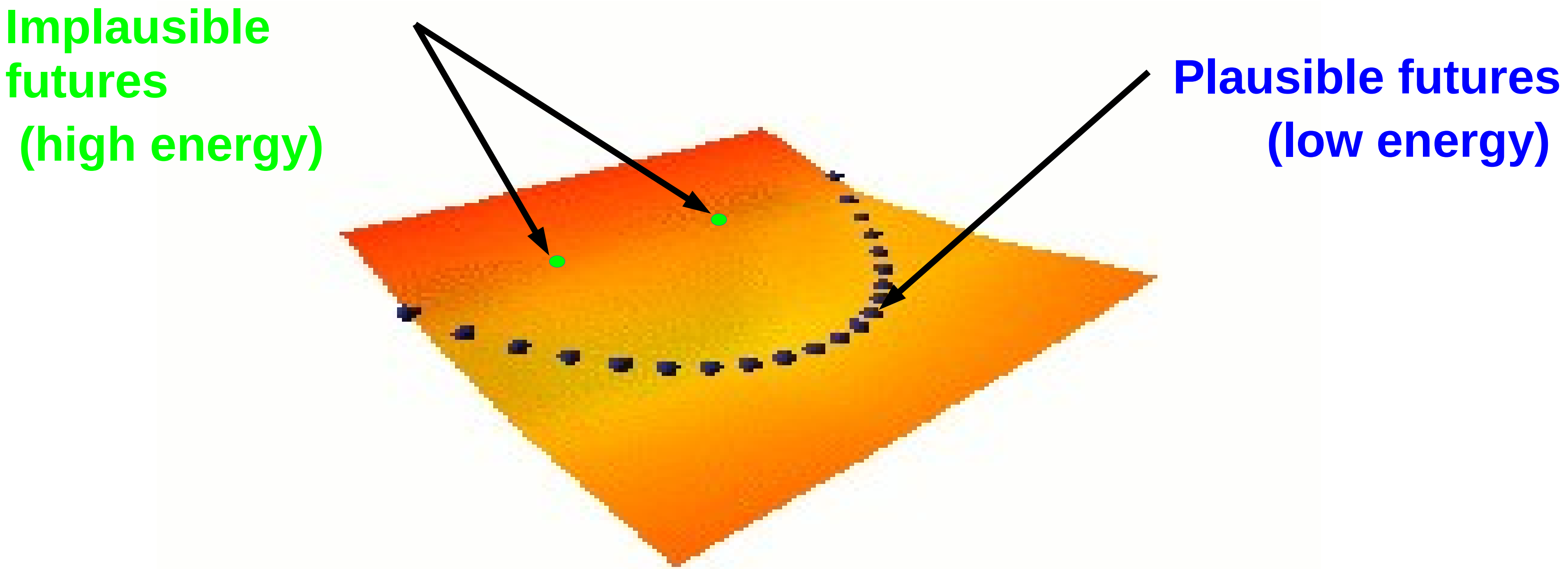
# Learning the "Data Manifold": Energy-Based Approach

- **Learning an energy function (or contrast function) that takes**
  - Low values on the data manifold
  - Higher values everywhere else



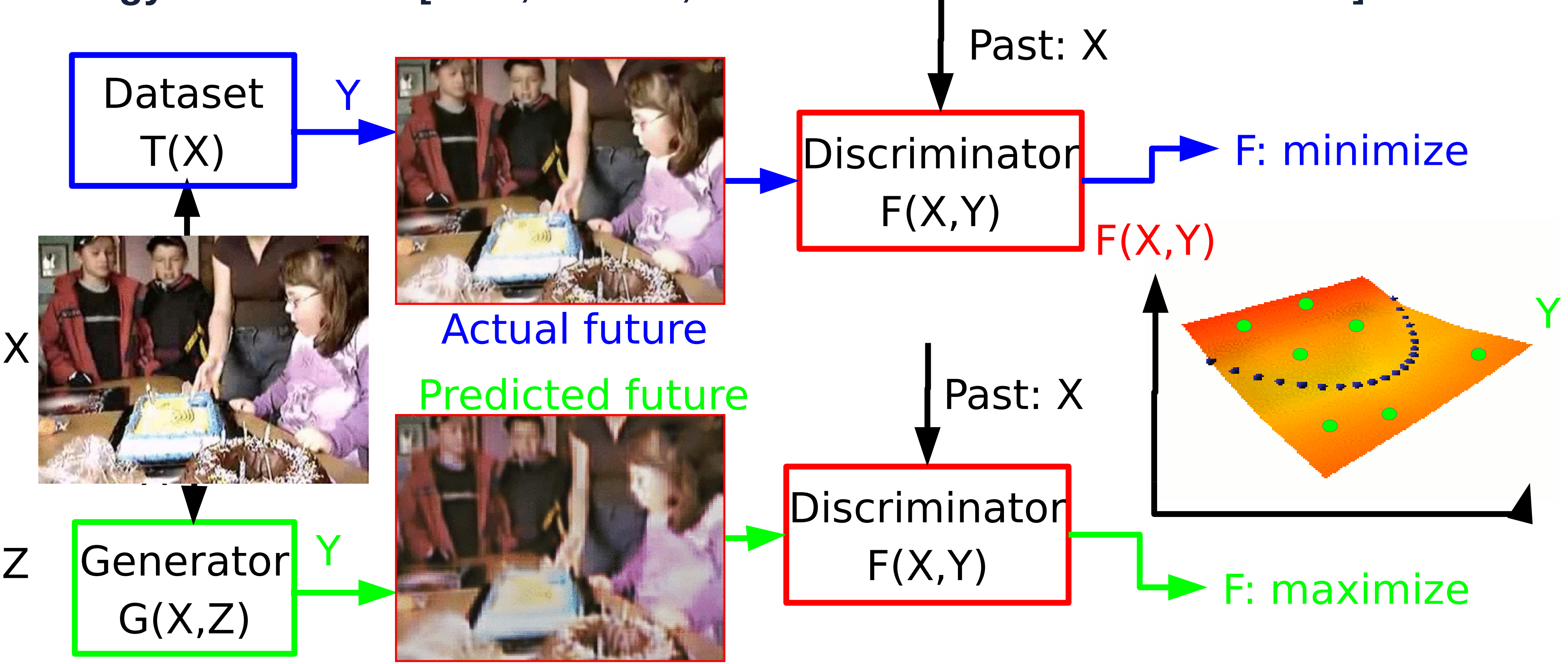
# Energy Function for Data Manifold

- **Energy Function:** **Takes low value on data manifold, higher values everywhere else**
- **Push down on the energy of desired outputs. Push up on everything else.**
- **But how do we choose where to push up?**



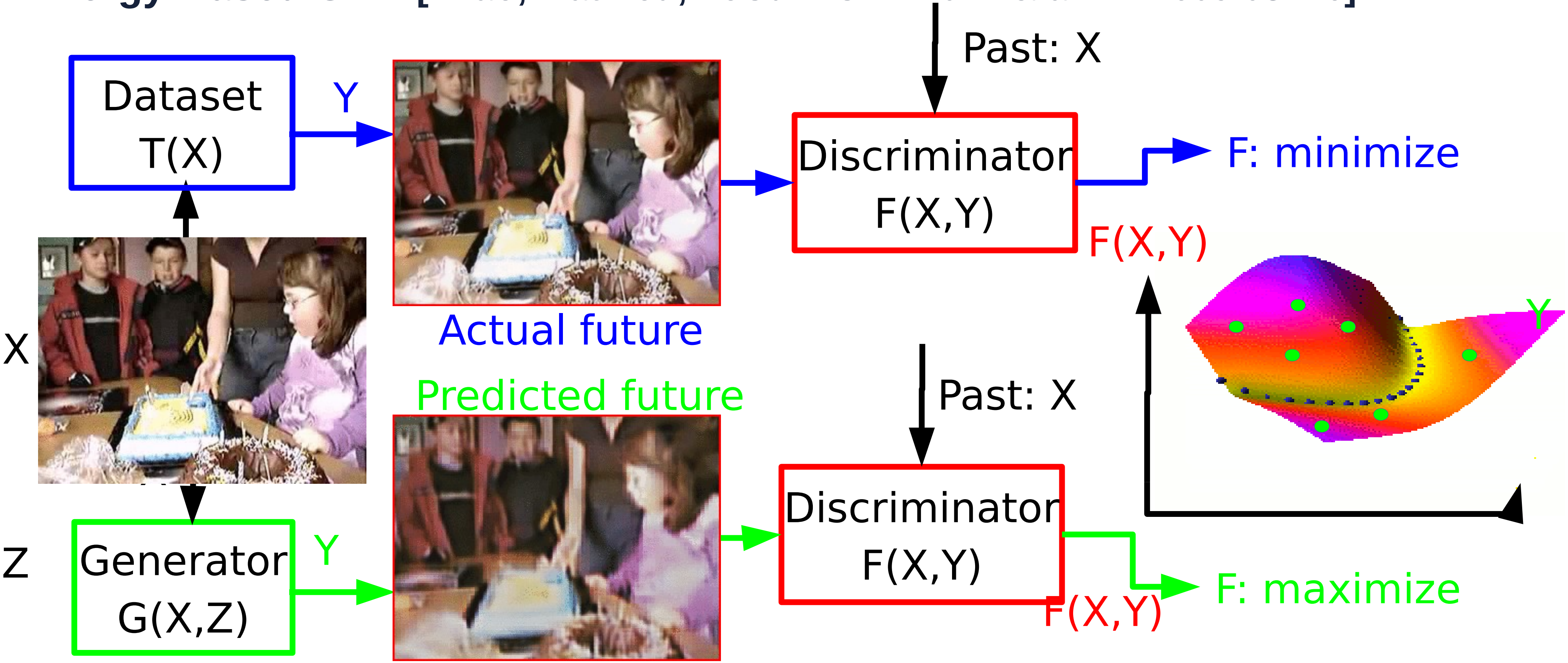
# Adversarial Training: the key to prediction under uncertainty?

- **Generative Adversarial Networks (GAN) [Goodfellow et al. NIPS 2014],**
- **Energy-Based GAN [Zhao, Mathieu, LeCun ICLR 2017 & arXiv:1609.03126]**



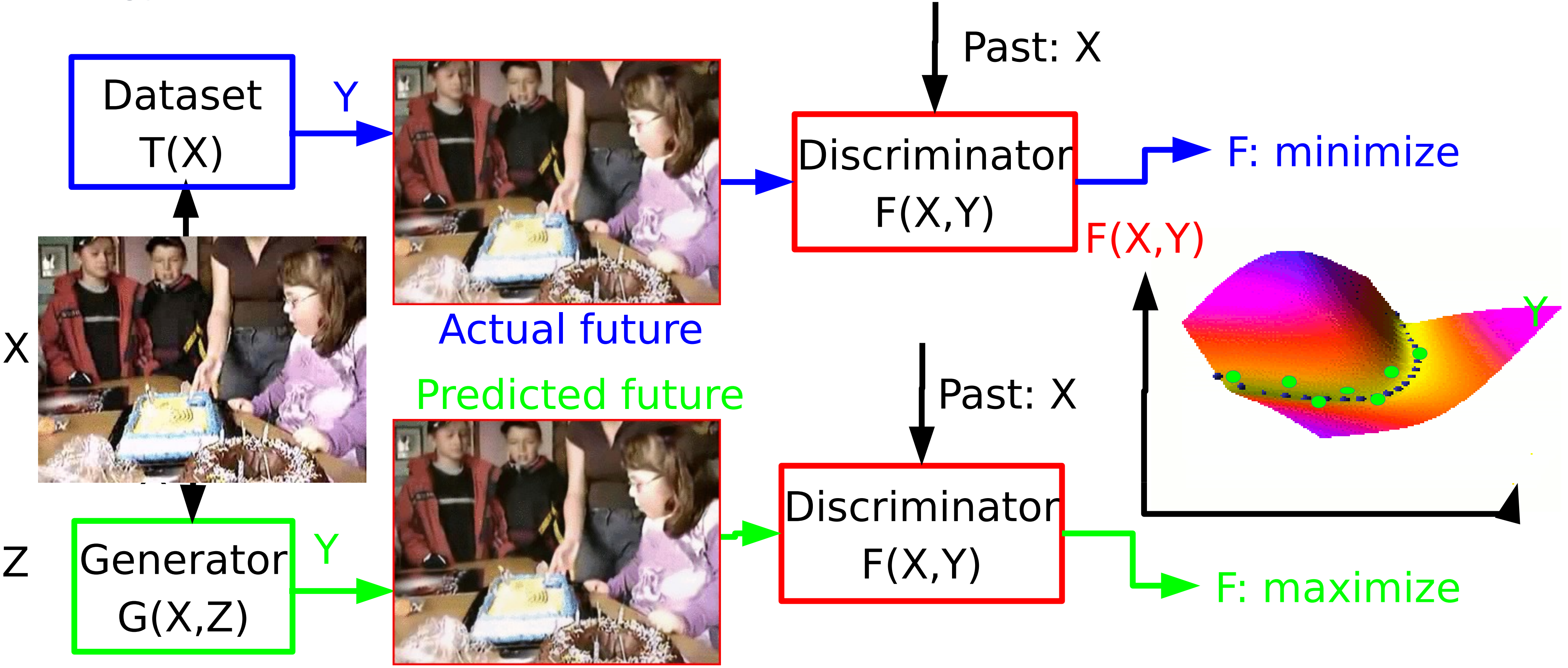
# Adversarial Training: the key to prediction under uncertainty?

- **Generative Adversarial Networks (GAN) [Goodfellow et al. NIPS 2014],**
- **Energy-Based GAN [Zhao, Mathieu, LeCun ICLR 2017 & arXiv:1609.03126]**



# Adversarial Training: the key to prediction under uncertainty?

- **Generative Adversarial Networks (GAN) [Goodfellow et al. NIPS 2014],**
- **Energy-Based GAN [Zhao, Mathieu, LeCun ICLR 2017 & arXiv:1609.03126]**



# DCGAN: “reverse” ConvNet maps random vectors to images

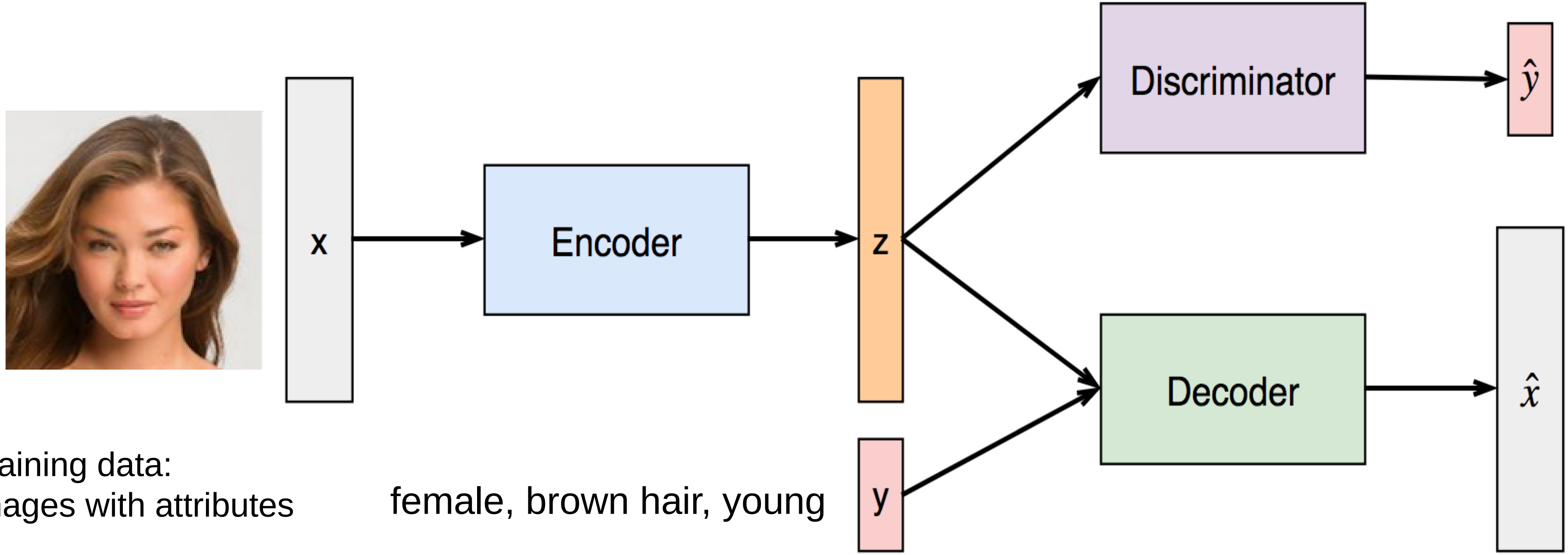
- **DCGAN: adversarial training to generate images.**
- **[Radford, Metz, Chintala 2015]**
  - Input: random numbers; output: bedrooms.

# Faces "invented" by a neural net (from NVIDIA)

- **From random numbers [Karras et al. ICLR 2018]**

# Fader Network: Auto-Encoder with two-part code

- **[Lample, Zeghidour, Usunier, Bordes, Denoyer,  Ranzato arXiv:1706.00409]**
- **Discriminator trains Encoder to remove attribute information Y from code Z**
  - Discriminator trained (supervised) to predict attributes.
  - Encoder trained to prevent discriminator from predicting attributes




Training data:
images with attributes

# Varying Attributes

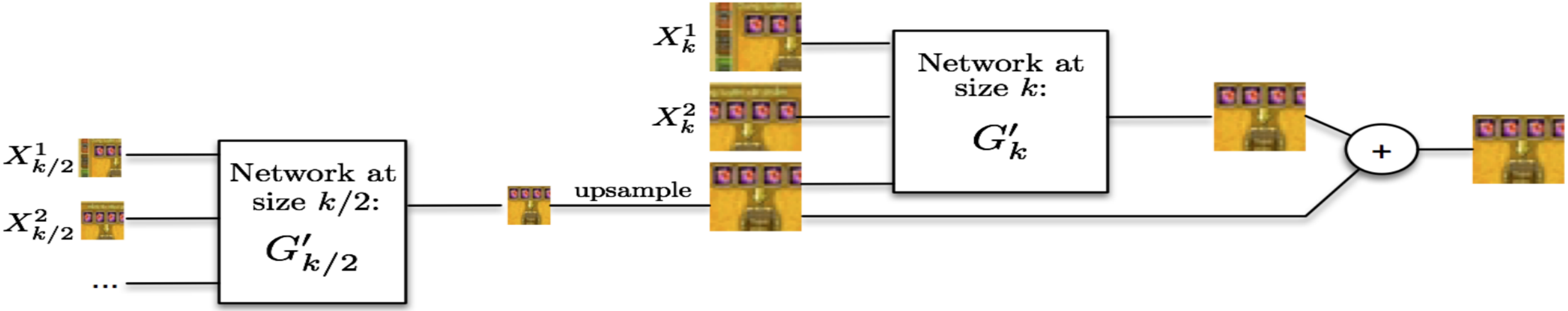
- **Young to old and back, male to female and back**

# Video Prediction with Adversarial Training

[Mathieu, Couprie, LeCun ICLR 2016]
arXiv:1511:05440

# Multi-Scale ConvNet for Video Prediction

- 4 to 8 frames input → ConvNet → 1 to 8 frames out
- Multi-scale ConvNet, without pooling
- If trained with least square: blurry output

# Predictive Unsupervised Learning

- **Our brains are “prediction machines”**
- **Can we train machines to predict the future?**
- **Some success with “adversarial training”**
  - [Mathieu, Couprie, LeCun arXiv:1511:05440]
- **But we are far from a complete solution.**

# Video Prediction: predicting 5 frames

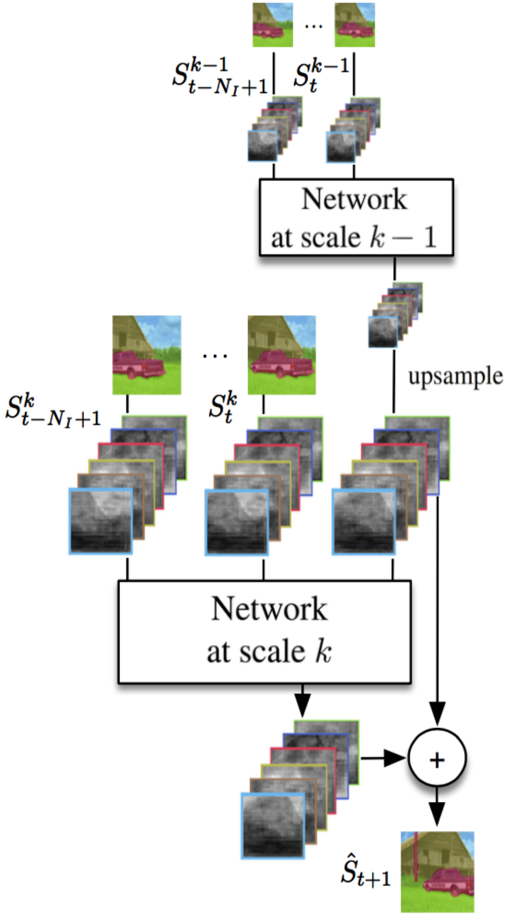
# Video Prediction in Semantic Segmentation Space

[Luc, Neverova, Couprie, Verbeek, & LeCun ICCV 2017]

# Temporal Predictions of Semantic Segmentations

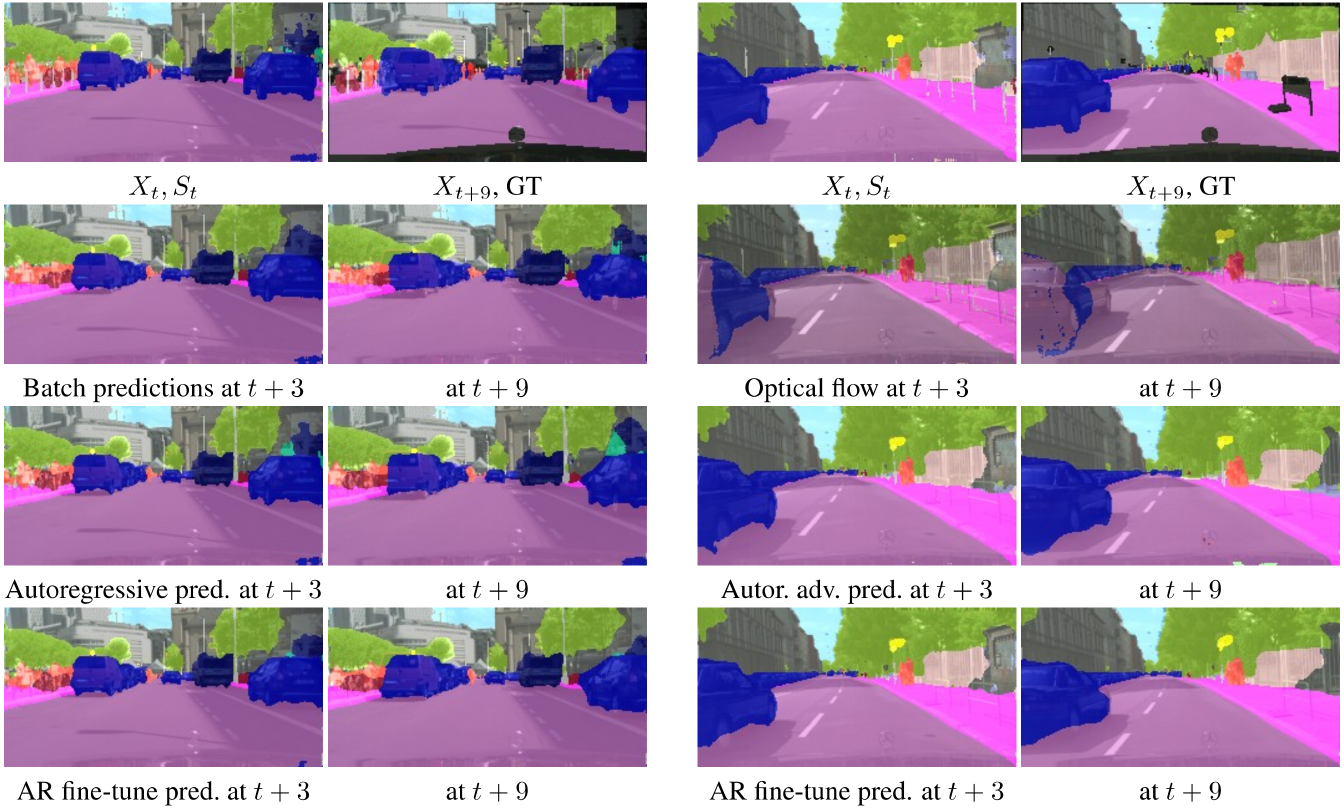
- **Predictions a single future frame**
  - CityScape dataset [Cordt et al. CVPR 2016]

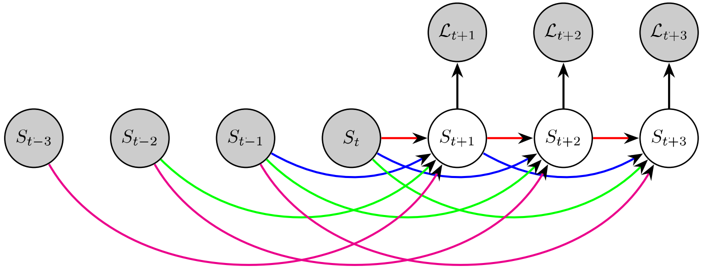
| Method | PSNR | SSIM | IoU GT | IoU SEG | IoU-MO GT | IoU-MO SEG |
|---|---|---|---|---|---|---|
| Copy last input | 20.6 | 0.65 | 49.4 | 54.6 | 43.4 | 48.2 |
| Warp last input | 20.9 | 0.67 | 50.4 | 55.5 | 44.9 | 49.8 |
| Model X2X | 24.0 | **0.77** | 23.0 | 22.3 | 12.8 | 11.4 |
| Model S2S | — | — | **58.3** | 64.9 | 53.8 | 59.8 |
| Model S2S-adv. | — | — | **58.3** | **65.0** | **53.9** | **60.2** |
| Model XS2X | **24.2** | **0.77** | 22.4 | 22.5 | 10.8 | 10.0 |
| Model XS2S | — | — | 58.2 | 64.6 | 53.7 | 59.9 |
| Model XS2XS | 24.0 | 0.76 | 55.5 | 61.1 | 50.7 | 55.8 |



# Temporal Predictions of Semantic Segmentations

- **Prediction 9 frames ahead (0.5 seconds)**
- **Auto-regressive model**

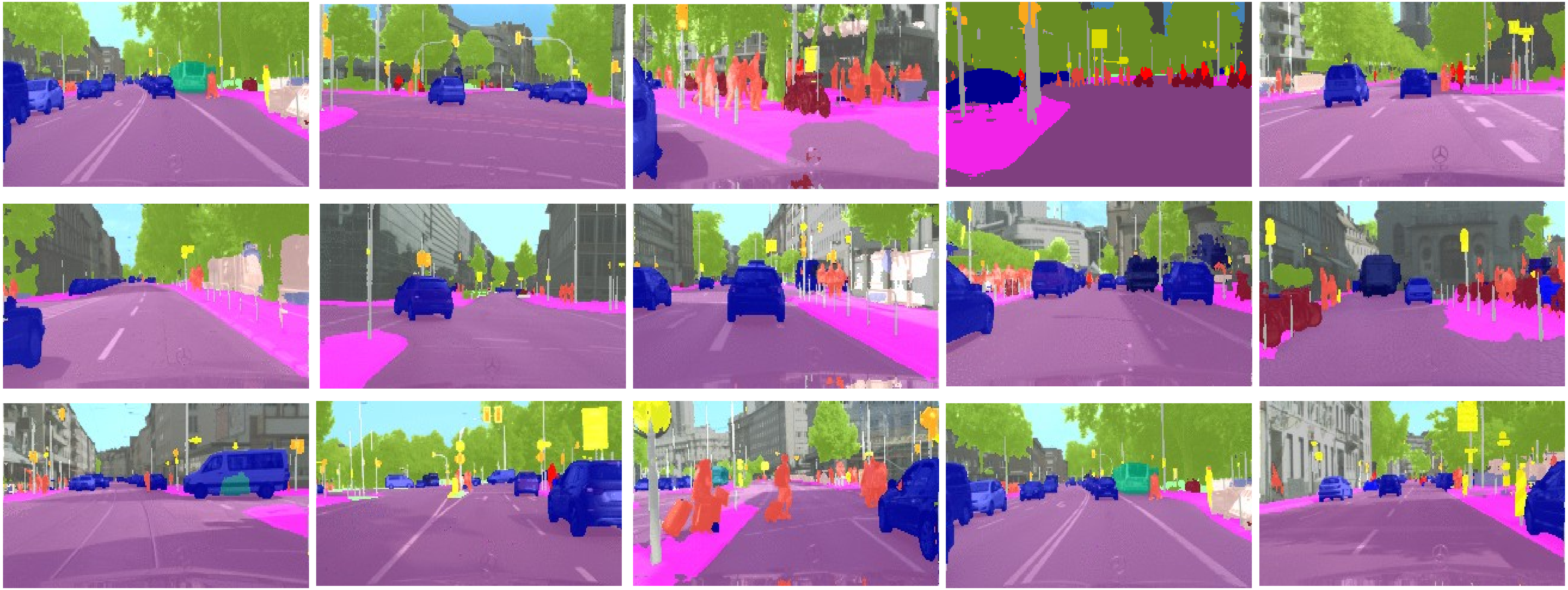



| Model | IoU GT | IoU SEG | IoU-MO GT |
|---|---|---|---|
| Copy last input | 36.9 | 39.2 | 26.8 |
| Warp last input | 37.5 | 39.5 | 27.9 |
| S2S, AR | 45.3 | 47.2 | 36.4 |
| S2S-adv, AR | 45.1 | 47.2 | 37.3 |
| S2S, AR, fine-tune | **46.7** | **49.7** | **39.3** |
| XS2XS, AR | 39.3 | 40.8 | 27.4 |
| S2S, batch | 42.1 | 44.2 | 32.8 |
| XS2S, batch | 42.3 | 44.6 | 33.1 |
| XS2XS, batch | 41.2 | 43.5 | 31.4 |

# Temporal Predictions of Semantic Segmentations

- Prediction 9 frames ahead (0.5 seconds)
- Auto-regressive model



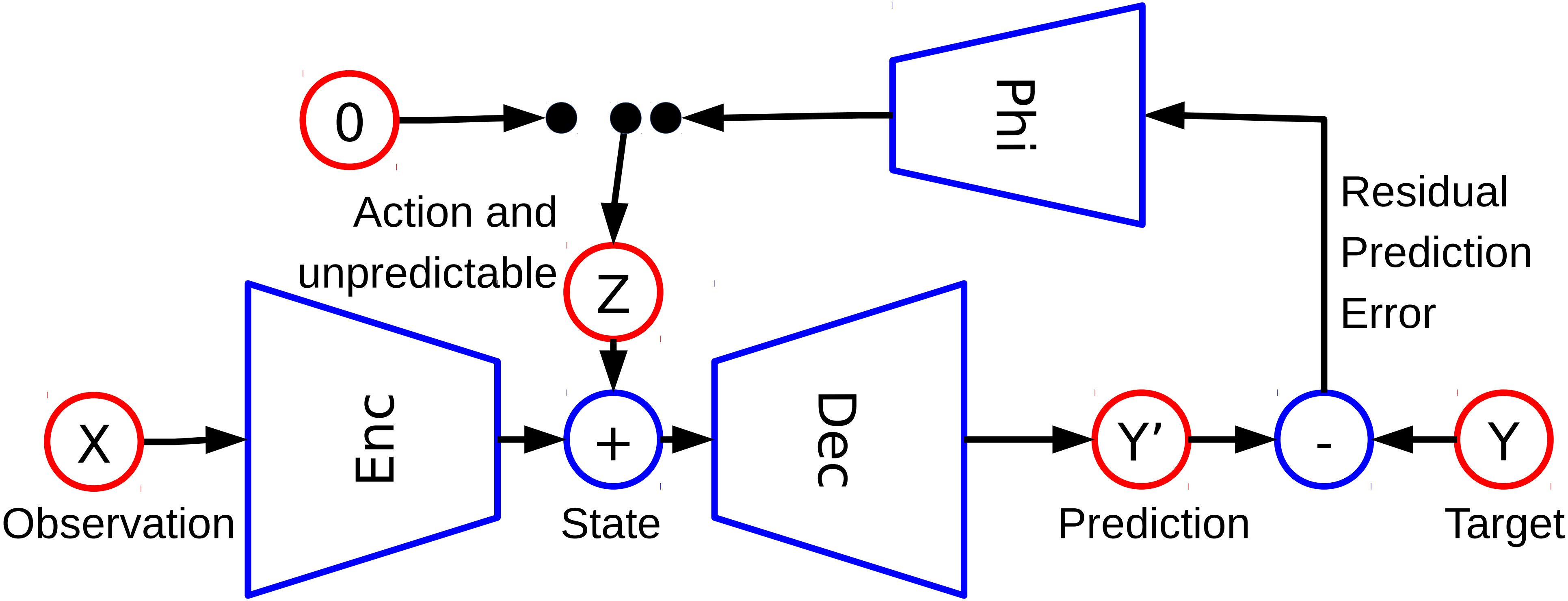
# Trained Forward Models
# for
# Planning and Learning Skills

[Henaff, Zhao, LeCun ArXiv:1711.04994]
[Henaff, Whitney, LeCun Arxiv:1705.07177]

facebook
Artificial Intelligence Research

# Error Encoding Network:
## Forward model that infers actions & unpredictable latent variables.

- **[Henaff, Zhao, LeCun ArXiv:1711.04994]**
- **Y' = Dec( Enc(X) + Z)     with Z=0 or   Z = Phi(Y-Y')**

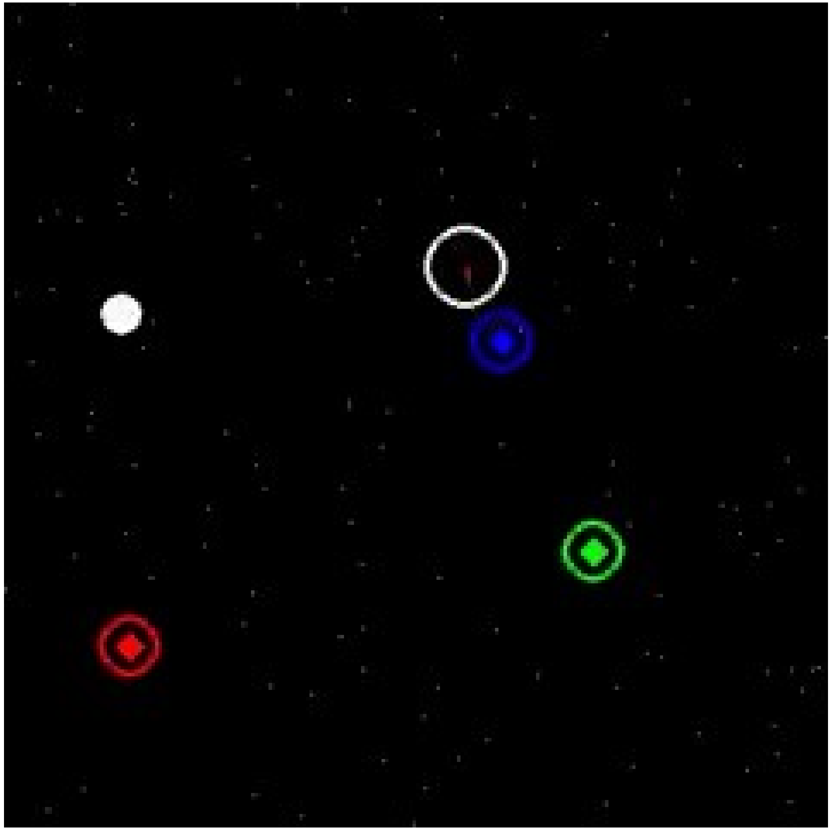


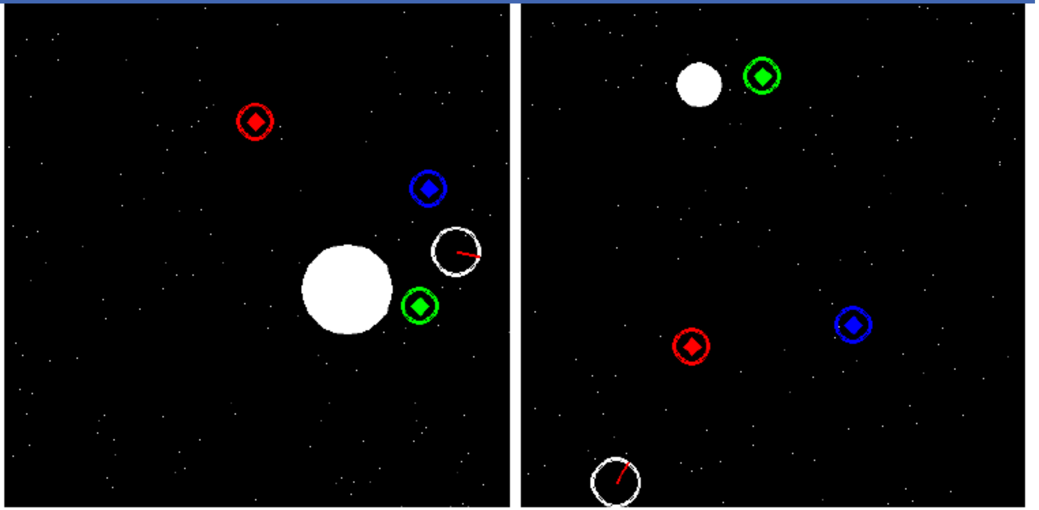
# Forward model that infers the action

- **Video: predictions as Z varies**

# Spaceship control

- **Planet with gravity, targets,**
- **Ship with orientable thruster**





| Method | Average Reward | Time (s) | Env. Steps |
|---|---|---|---|
| Random | -62.7 | - | 0 |
| A2C | -19.2 | 0.01 | 3.8M |
| GBP | 11.1 | 0.19 | 800K |
| DistGBP | 12.2 | 0.01 | 800K |

# The Future Impact of AI

facebook
Artificial Intelligence Research

# Promising Areas for Research

- **Marrying deep learning and (logical) reasoning**
  - Replacing symbls by vectors and logic by algebra
- **Self-supervised learning of world models**
  - Dealing with uncertainty, high dimensionality
- **Learning hierarchical representations of control space**
  - Instantiating complex/abstract action plans into simpler ones
- **Theory!**
- **Compilers for differentiable programming.**

# Technology drives & motivates Science (and vice versa)

- **Science drives technology, but technology also drives science**
- **Sciences are born from the study of technological artifacts**
  - Telescope → optics
  - Steam engine → thermodynamics
  - Airplane → aerodynamics
  - Calculators → computer science
  - Telecommunication → information theory
- **What is the equivalent of thermodynamics for intelligence?**
  - Are there underlying principles behind artificial and natural intelligence?
  - Are there simple principles behind learning?
  - Or is the brain a large collection of “hacks” produced by evolution?

# Authentic human experience > material goods

- **Material goods:**
  - BlueRay player: $47
  - Handmade ceramic bowl: $750
- **Mozart's opera Die Zauberflöte**
  - Downloadable recording: $7
  - Ticket at the NYC Met: up to $807
- **Bright future for jazz musicians and artisans?**

Samsung

Samsung Smart Curved Design Blu-Ray Disc 1080p Player With Wired Ethernet Content Streaming Manufacturer Refurbished

19 customer reviews | 8 answered questions

Price: $46.88 & FREE Shipping


Mozart: Die Zauberflöte
Wiener Philharmoniker
January 1, 2012

19 customer review

Start your 30-day free trial of Unlimited to Prime pricing.

See all 50 formats and editions

| Streaming | MP3 | Audio CD |
|---|---|---|
| Unlimited | $6.99 | $8.98 |

| Section | QTY | Price | |
|---|---|---|---|
| CENTER ORCHESTRA Row B-EE Email delivery by: 09/26/17 | 5 | $751.00 ea. | BUY |
| ORCH Row B-O Email delivery by: 09/26/17 | 4 | $772.00 ea. | BUY |
| CENTER ORCHESTRA Row A-DD Email delivery by: 09/26/17 | 5 | $786.00 ea. | BUY |
| ORCH Row A-N Email delivery by: 09/26/17 | 4 | $807.00 ea. | BUY |

Scallop Bowl

$750.00 USD

Dimensions: 14"w x 7"h

H3-40

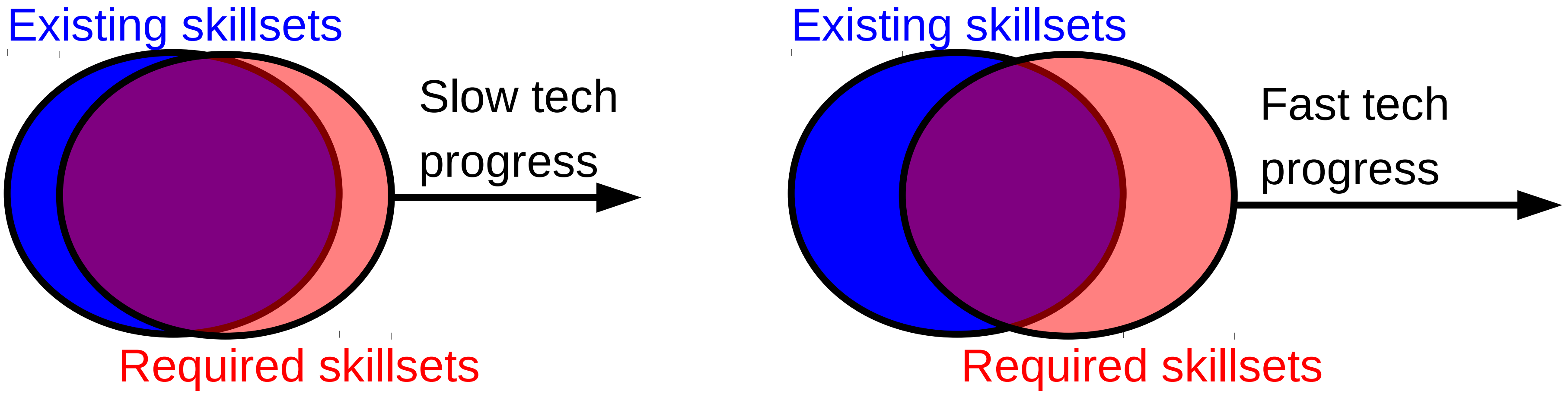
ADD TO CART

ADD TO REGISTRY

Tweet  Like 0

# AI is a "General Purpose Technology" (GPT)

- **GPT: steam engine, electricity, computer...**
  - [Bresnahan & Trajtenberg 1995] "GPTs 'Engines of growth'?". J. Econometrics.
- **AI will affect many sector of the economy**
- **But it will take 10 or 20 years before we see the effect on productivity**
- **AI/automation → job displacement → technological unemployment**
- **Technology deployment is limited by how fast workers can train for it**



# When will the “True AI” revolution occur?

- **We won’t have household robots and good digital friends (or assistants) until machines acquire common sense.**
- **This won’t happen until we get machines to learn predictive world models**
- **Discovering the principles of it may take 2, 5, 10 or 20 years.**
- **Developing practical technology from it may take another 10 years**
  - The emergence of “true AI” will not be a singular event as in Hollywood movies.
- **We work on the assumption that there is “simple” principle (and a few algorithms) for AI, as there is for flight (aerodynamics) or engines (thermodynamics).**

# What will super-intelligent AGI be like?

- **Will the “singularity” happen?**
  - No. Nothing is exponential forever
- **Future AI systems will have emotions and moral values**
  - How to align AI values with human values?
- **Will it take our jobs?**
  - No. But our jobs will change. Human experience will have high value.
  - it will empower humanity by amplifying our intelligence
- **Will it want to take over the world?**
  - No, the desire to dominate is not correlated with intelligence but with testosterone

# Thank you

facebook
Artificial Intelligence Research