



# CSCI 1108



## Object Recognition



# What is this

- Interpreting the sensed environment is another crucial part in robotics
- The general idea is to match sensor patterns with expected patterns for specific objects
- Sensor data could come from many different type of modalities like vision, touch, themperture, etc

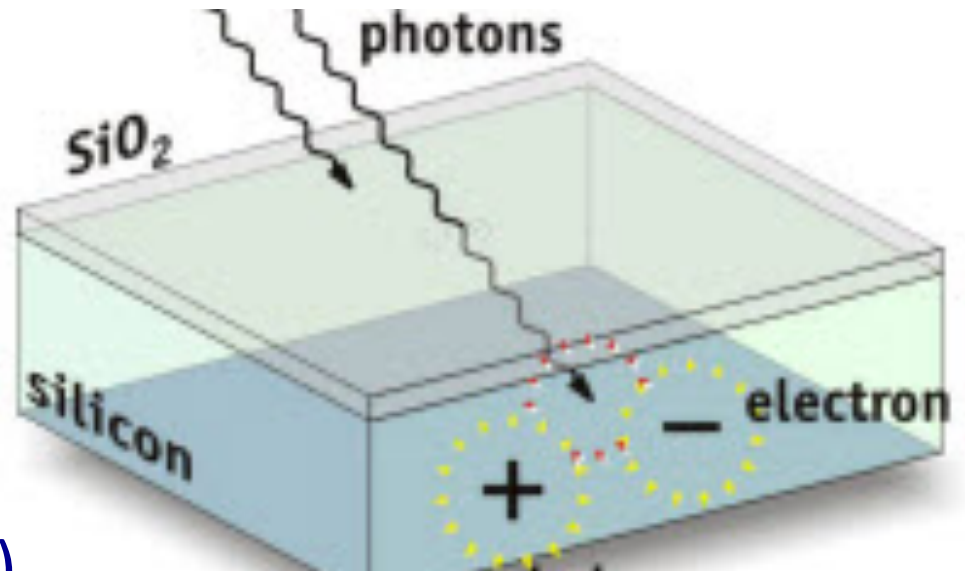
# Computer Vision

Computer vision considers the processing of data from image sensors. We discuss this in two steps:

1. How do we sense the physical world
- 2, How do we interpret the sensed world

# Photoelectric effect

Photons cause a small discharge, which is accumulated and then amplified to give a reading for the intensity of a pixel (picture element)

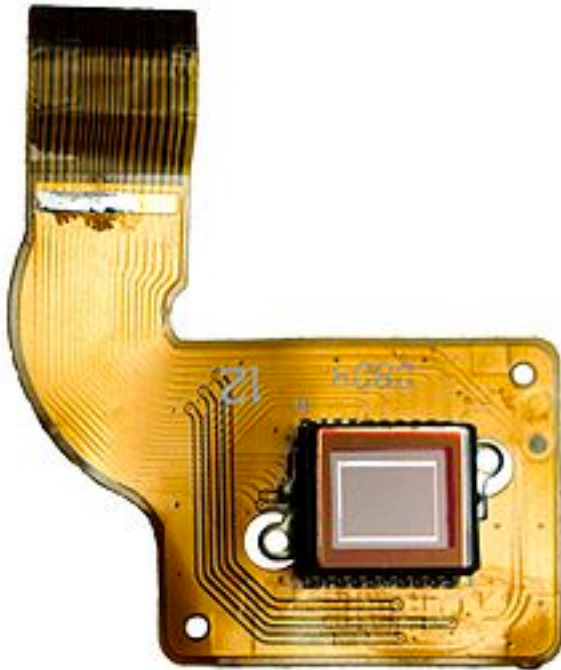
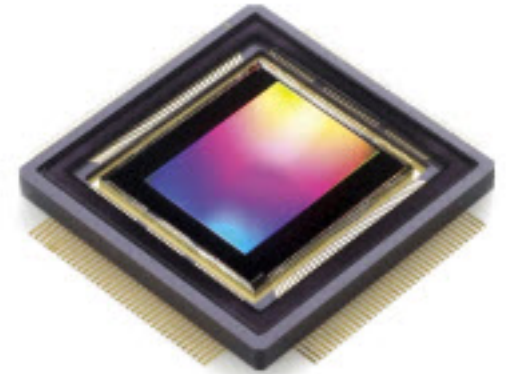


# Imaging chips

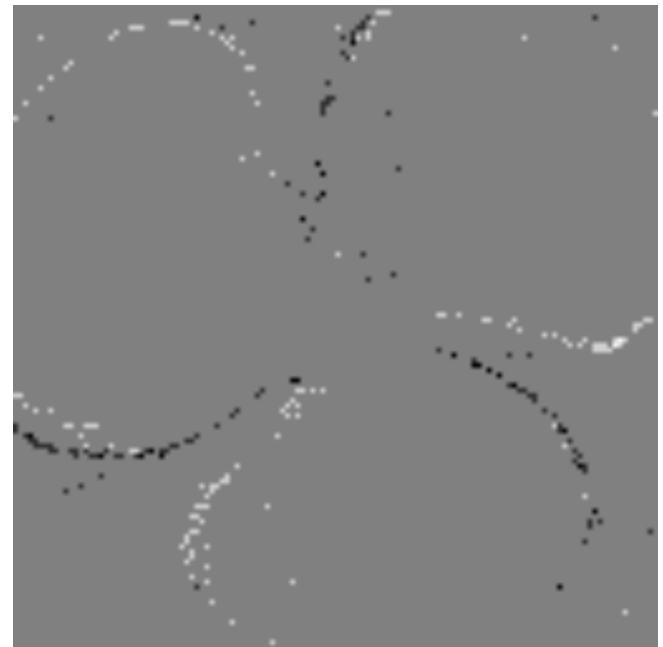
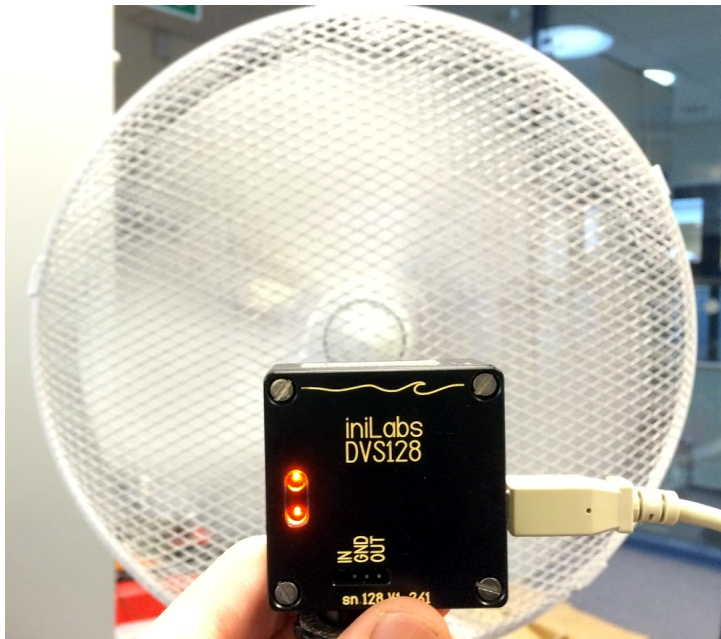
CCD  
(charge-coupled  
device)

and

CMOS  
(complementary  
metal-oxide  
semiconductor)

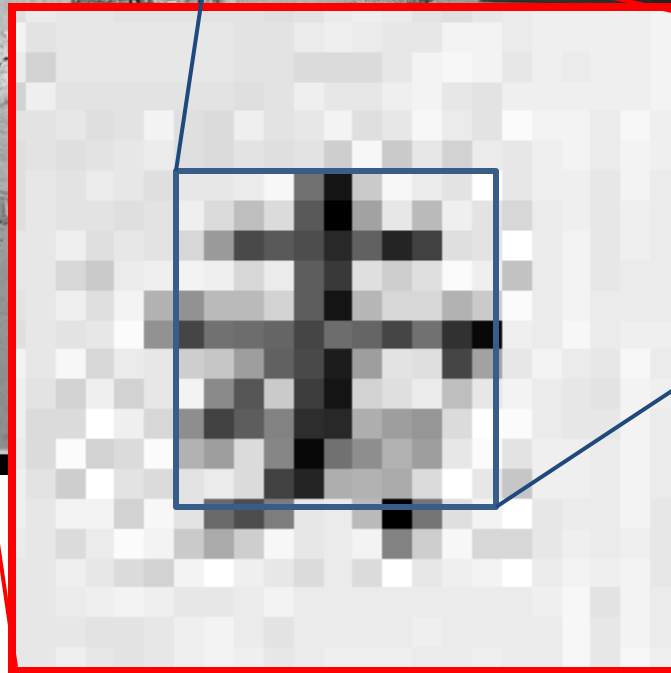


# (outlook) Event-based cameras





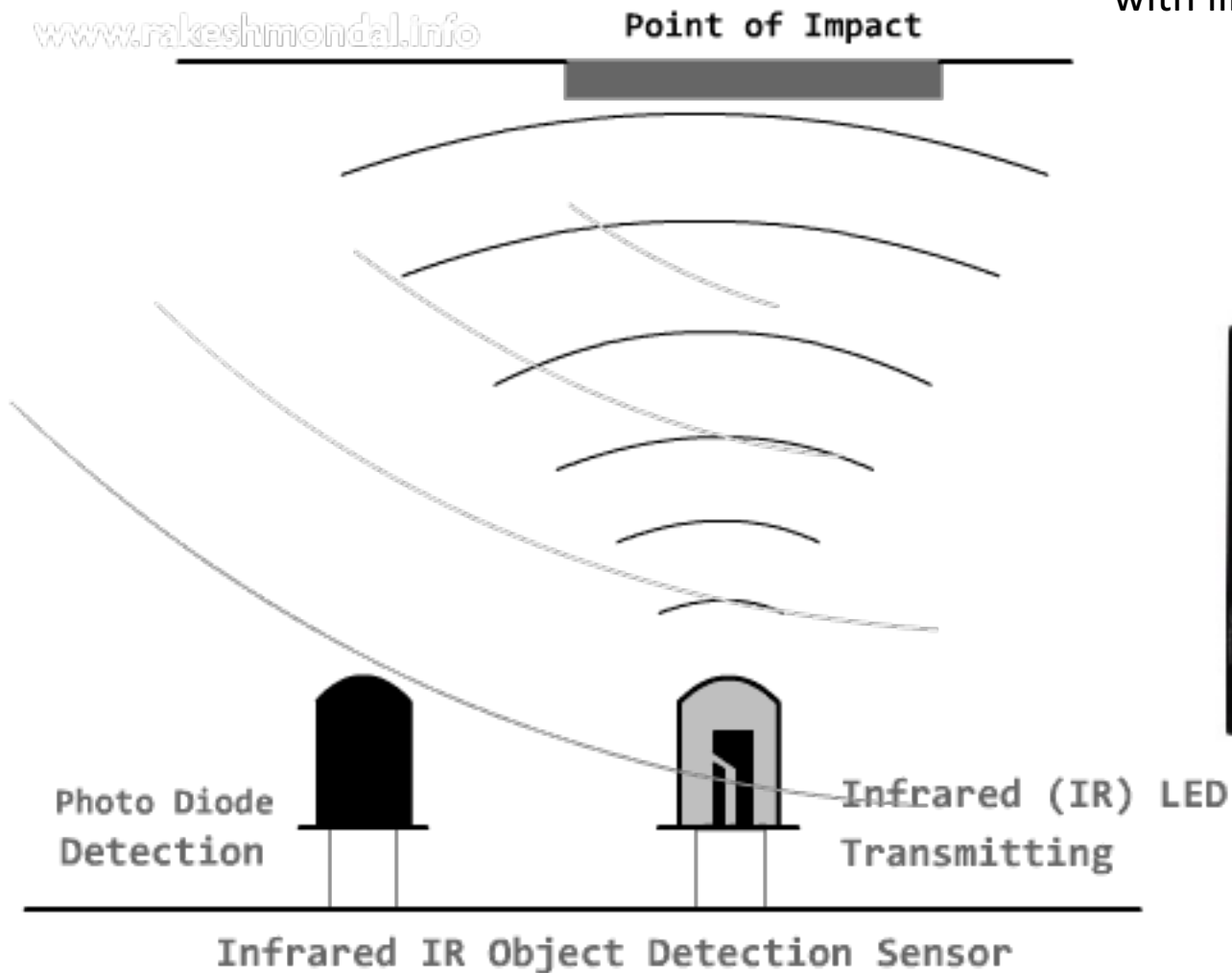
0.92	0.93	0.94	0.97	0.62	0.37	0.85	0.97	0.93	0.92	0.99
0.95	0.89	0.82	0.89	0.56	0.31	0.75	0.92	0.81	0.95	0.91
0.89	0.72	0.51	0.55	0.51	0.42	0.57	0.41	0.49	0.91	0.92
0.96	0.95	0.88	0.94	0.56	0.46	0.91	0.87	0.90	0.97	0.95
0.71	0.81	0.81	0.87	0.57	0.37	0.80	0.88	0.89	0.79	0.85
0.49	0.62	0.60	0.58	0.50	0.60	0.58	0.50	0.61	0.45	0.33
0.86	0.84	0.74	0.58	0.51	0.39	0.73	0.92	0.91	0.49	0.74
0.96	0.67	0.54	0.85	0.48	0.37	0.88	0.90	0.94	0.82	0.93
0.69	0.49	0.56	0.66	0.43	0.42	0.77	0.73	0.71	0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93



000 philg@mit.edu

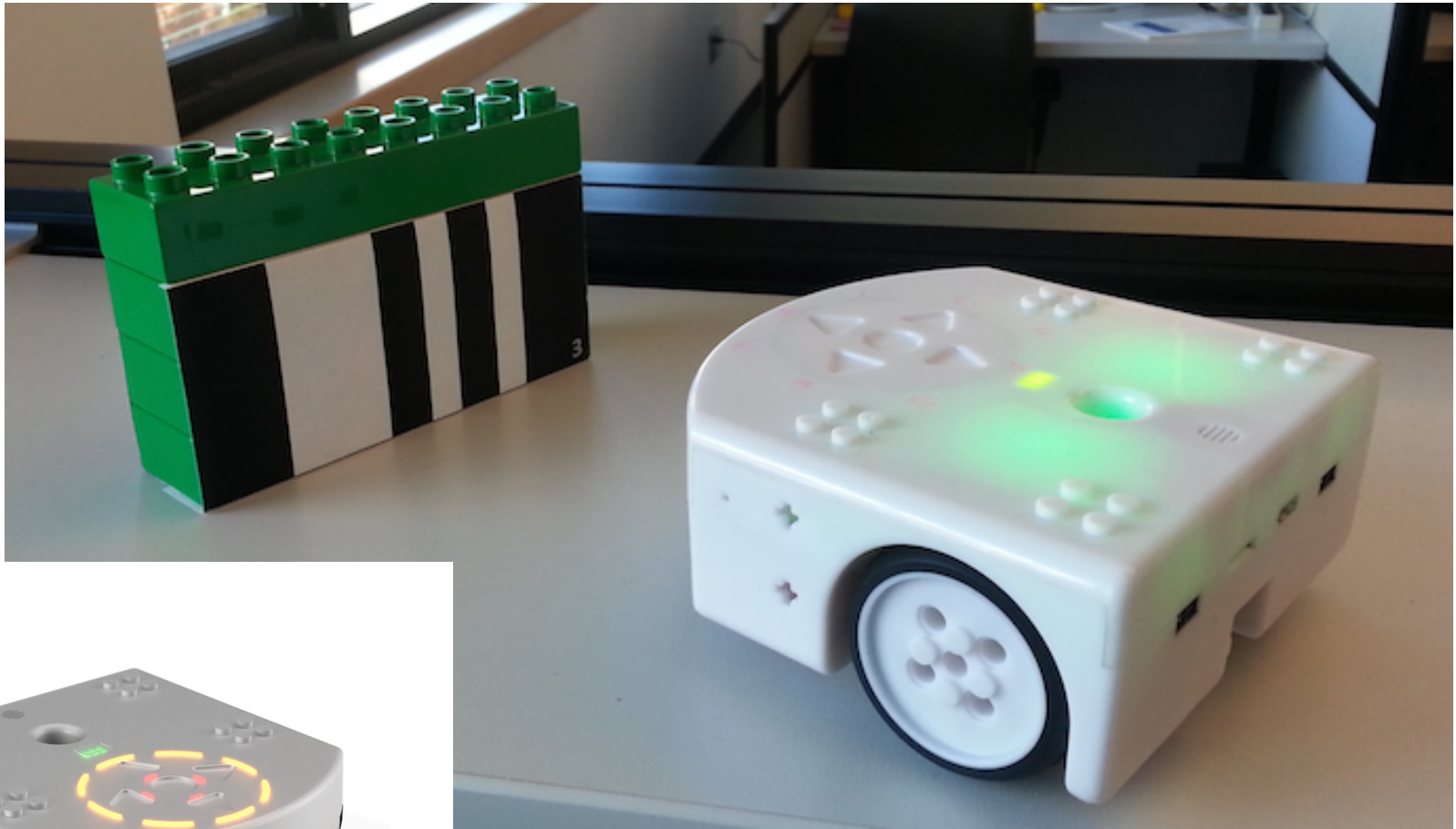
# Infrared Proximity Sensor

“one pixel infrared camera with illuminator”

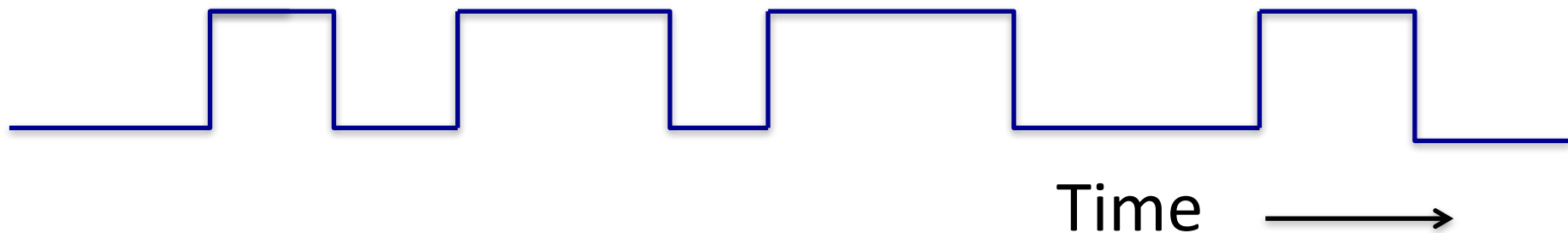
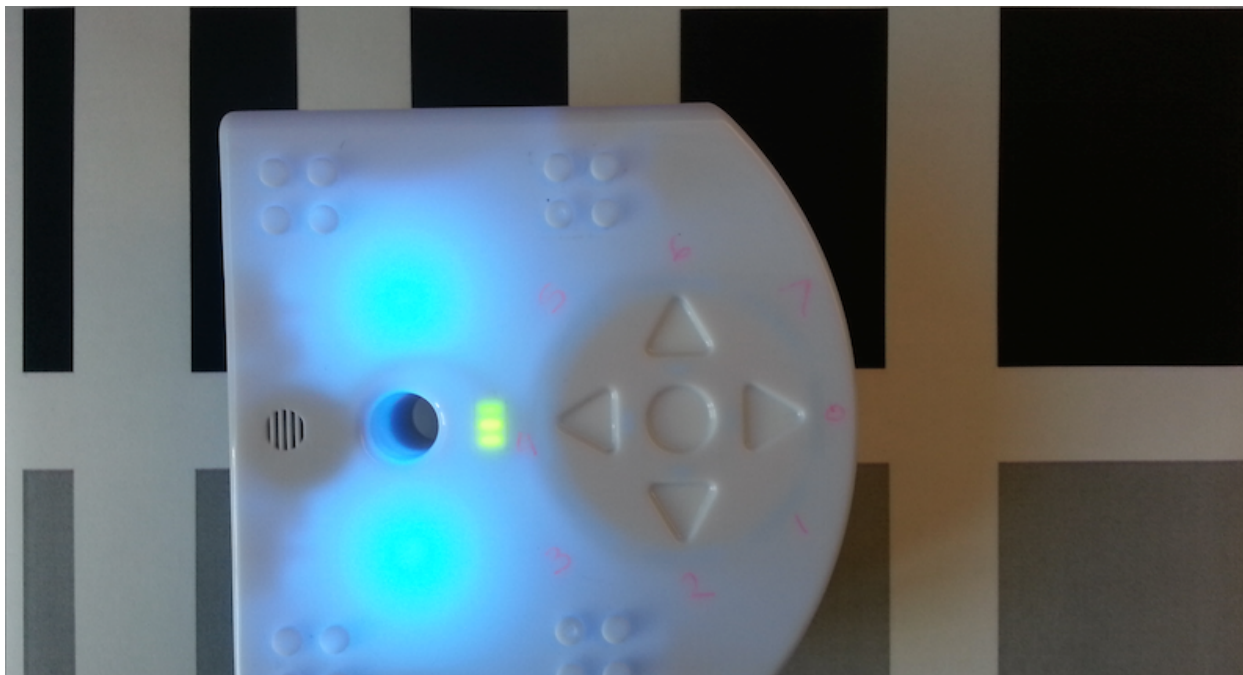




# A simple 5 pixel infrared camera



# Using pixels over time: A simple video recognition system



# How do we use sensor data: Image recognition

Simple pattern recognizer:

onevent prox

If `prox.horizontal[0] > THRESH` and `prox.horizontal[3] < THRESH` then

...

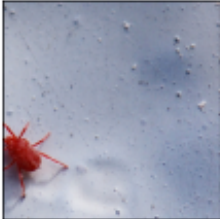



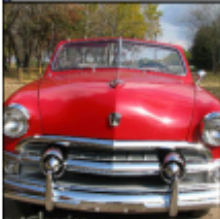

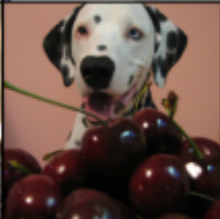
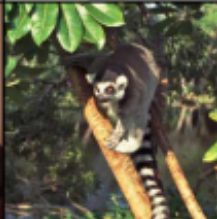
# Point & Find

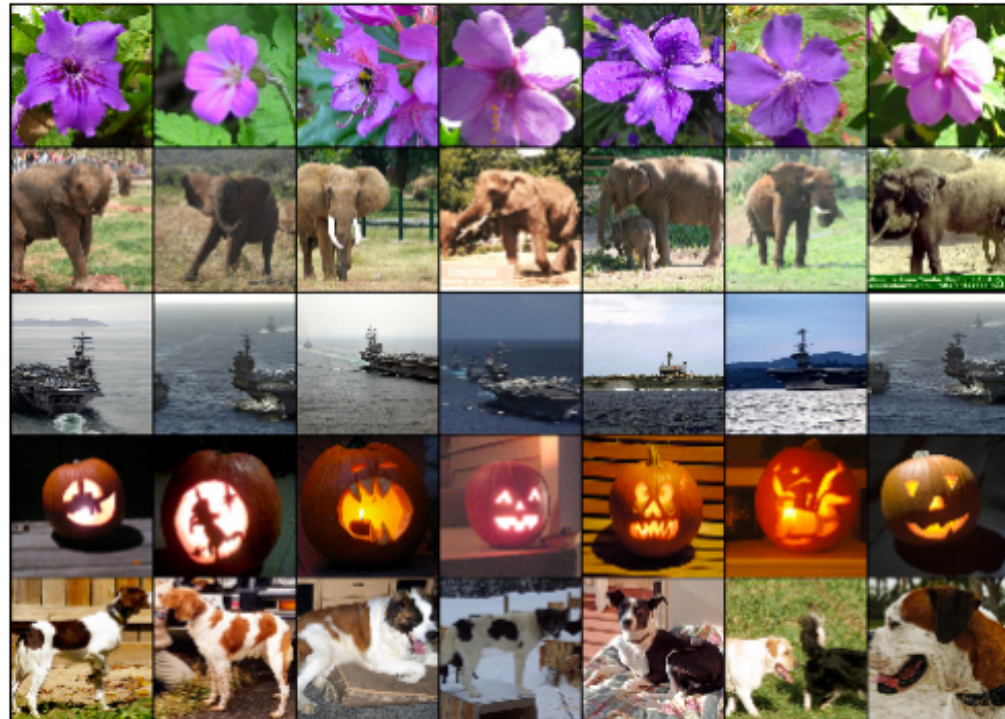


# German Traffic Sign Competition

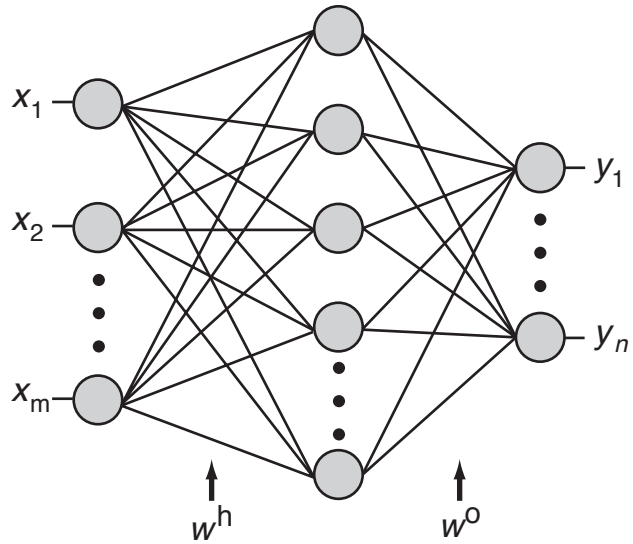


# ImageNet Competition

			
<b>mite</b>	<b>container ship</b>	<b>motor scooter</b>	<b>leopard</b>
<ul style="list-style-type: none"> <li>mite</li> <li>black widow</li> <li>cockroach</li> <li>tick</li> <li>starfish</li> </ul>	<ul style="list-style-type: none"> <li>container ship</li> <li>lifeboat</li> <li>amphibian</li> <li>fireboat</li> <li>drilling platform</li> </ul>	<ul style="list-style-type: none"> <li>motor scooter</li> <li>go-kart</li> <li>moped</li> <li>bumper car</li> <li>golfcart</li> </ul>	<ul style="list-style-type: none"> <li>leopard</li> <li>jaguar</li> <li>cheetah</li> <li>snow leopard</li> <li>Egyptian cat</li> </ul>
			
<b>grille</b>	<b>mushroom</b>	<b>cherry</b>	<b>Madagascar cat</b>
<ul style="list-style-type: none"> <li>convertible</li> <li>grille</li> <li>pickup</li> <li>beach wagon</li> <li>fire engine</li> </ul>	<ul style="list-style-type: none"> <li>agaric</li> <li>mushroom</li> <li>jelly fungus</li> <li>gill fungus</li> <li>dead-man's-fingers</li> </ul>	<ul style="list-style-type: none"> <li>dalmatian</li> <li>grape</li> <li>elderberry</li> <li>ffordshire bullterrier</li> <li>currant</li> </ul>	<ul style="list-style-type: none"> <li>squirrel monkey</li> <li>spider monkey</li> <li>titi</li> <li>indri</li> <li>howler monkey</li> </ul>



# Neural Networks



Activity of “neurons”:

Sum of weighted inputs

Parameters of the model:

Weighted of connections

Learning:

Determine parameters from sample data

Goal: Generalization



Frank Rosenblatt  
(1928–1971)



David Rumelhart  
(1942-2011)

...



LeCun



Hinton



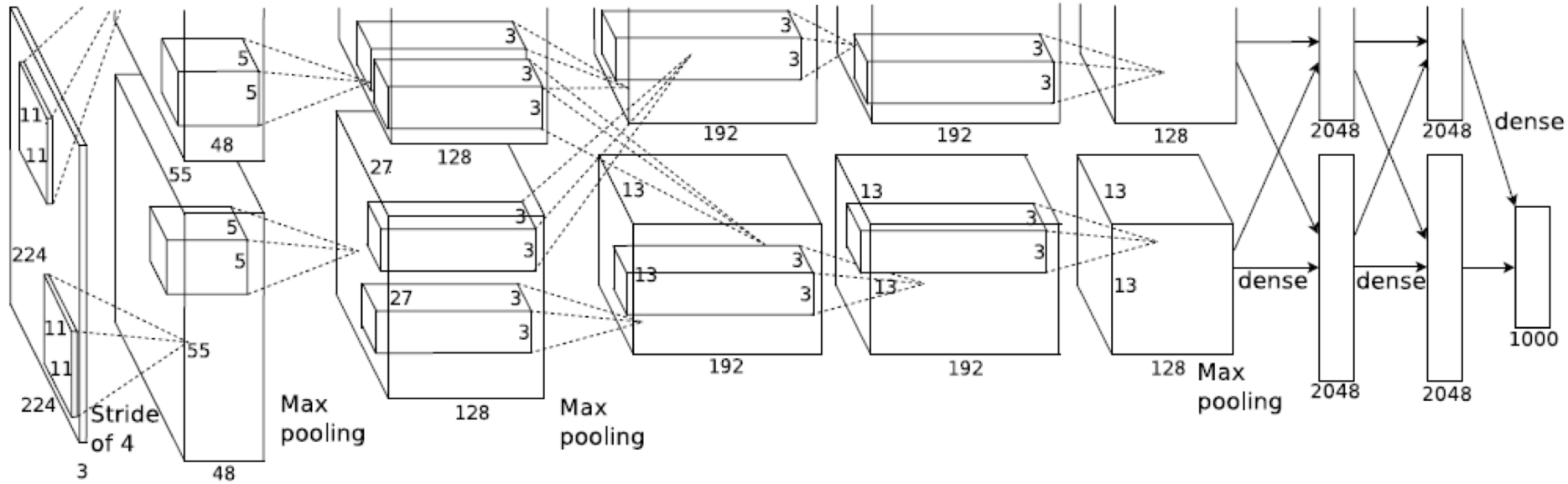
Bengio

...

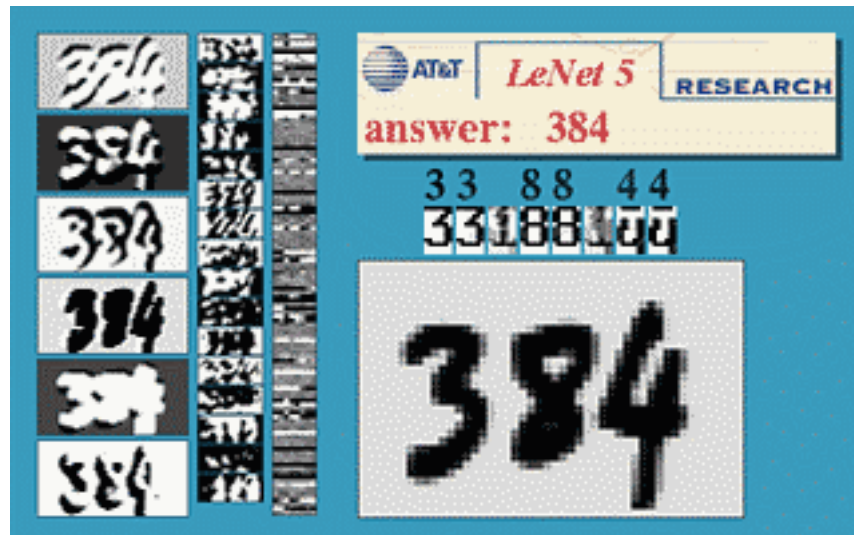


# Perceptron and deep learning

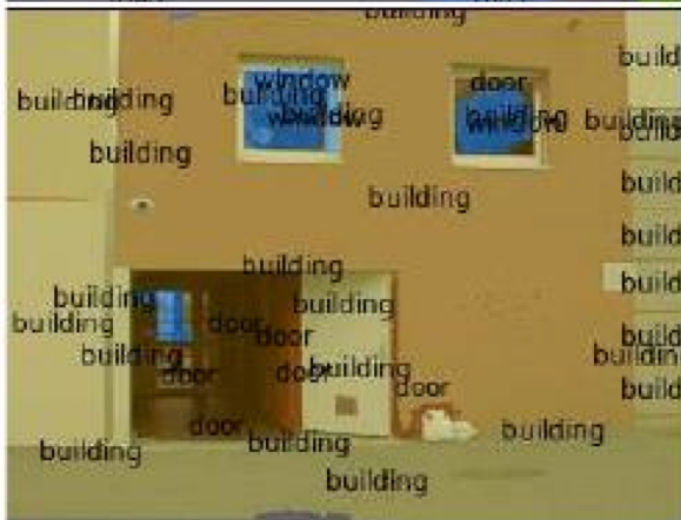
## Alex Net



## LeNet (LeCun):







Scene labeling  
(Farabet et al. 2013)



Object recognition and localization  
(Sermanet et al. 2014)

# Semantic segmentation

[https://www.youtube.com/watch?v=CxanE\\_W46ts](https://www.youtube.com/watch?v=CxanE_W46ts)



Andrej Karpathy, Li Fei-Fei, CVPR 2015

# Deep Visual-Semantic Alignments for Generating Image Descriptions



"man in black shirt is playing guitar."



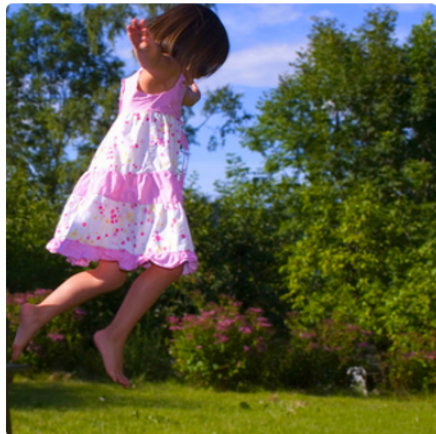
"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."



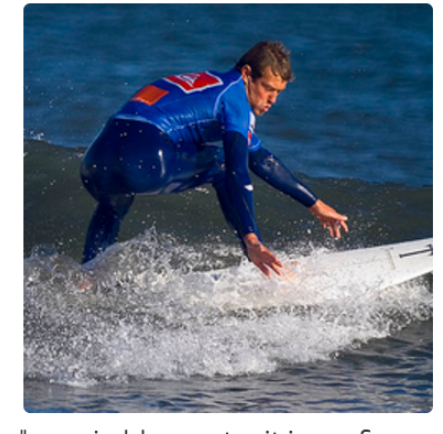
"girl in pink dress is jumping in air."



"black and white dog jumps over bar."



"young girl in pink shirt is swinging on swing."



"man in blue wetsuit is surfing on wave."