Computer Vision

Pietro Perona
California Institute of Technology

DS@HEP 2017
Fermilab - 8 May 2017

What?
Where?

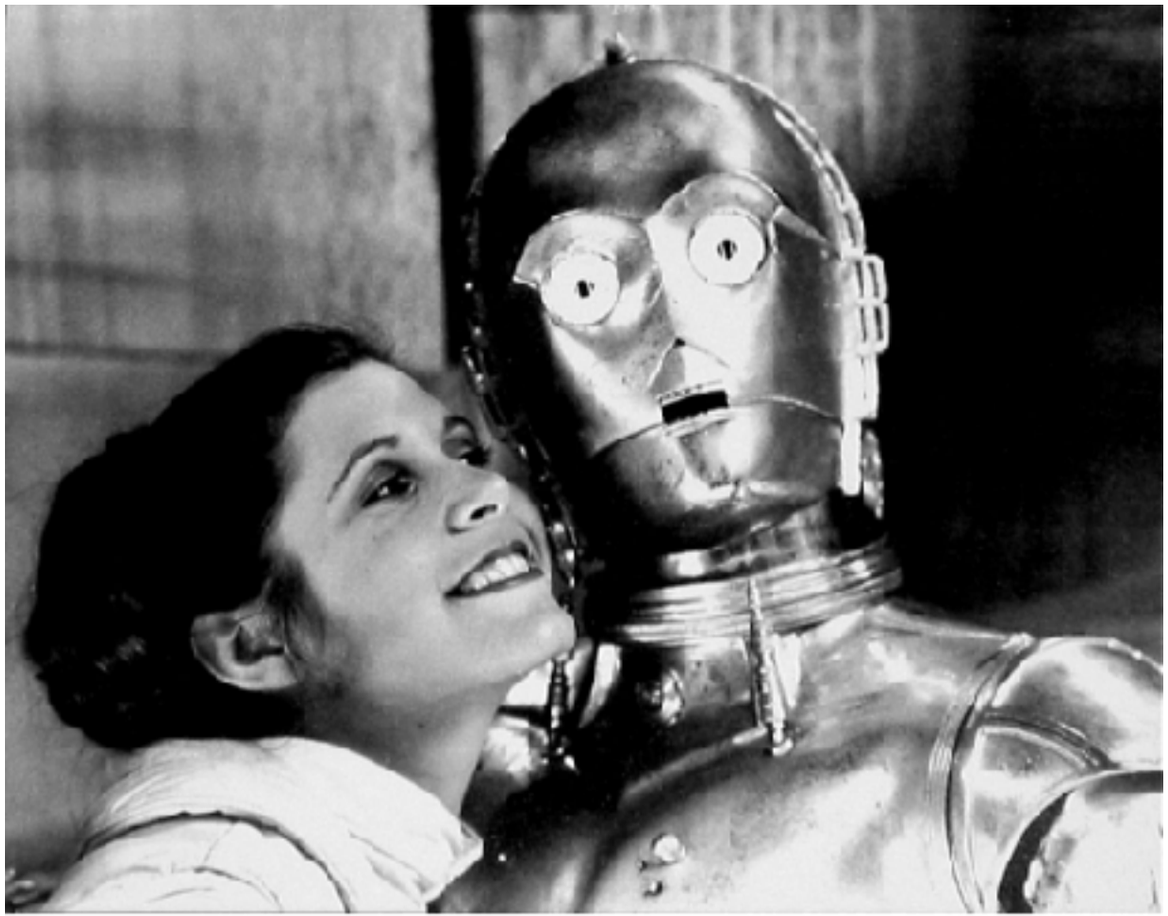
Scene understanding

Scene understanding

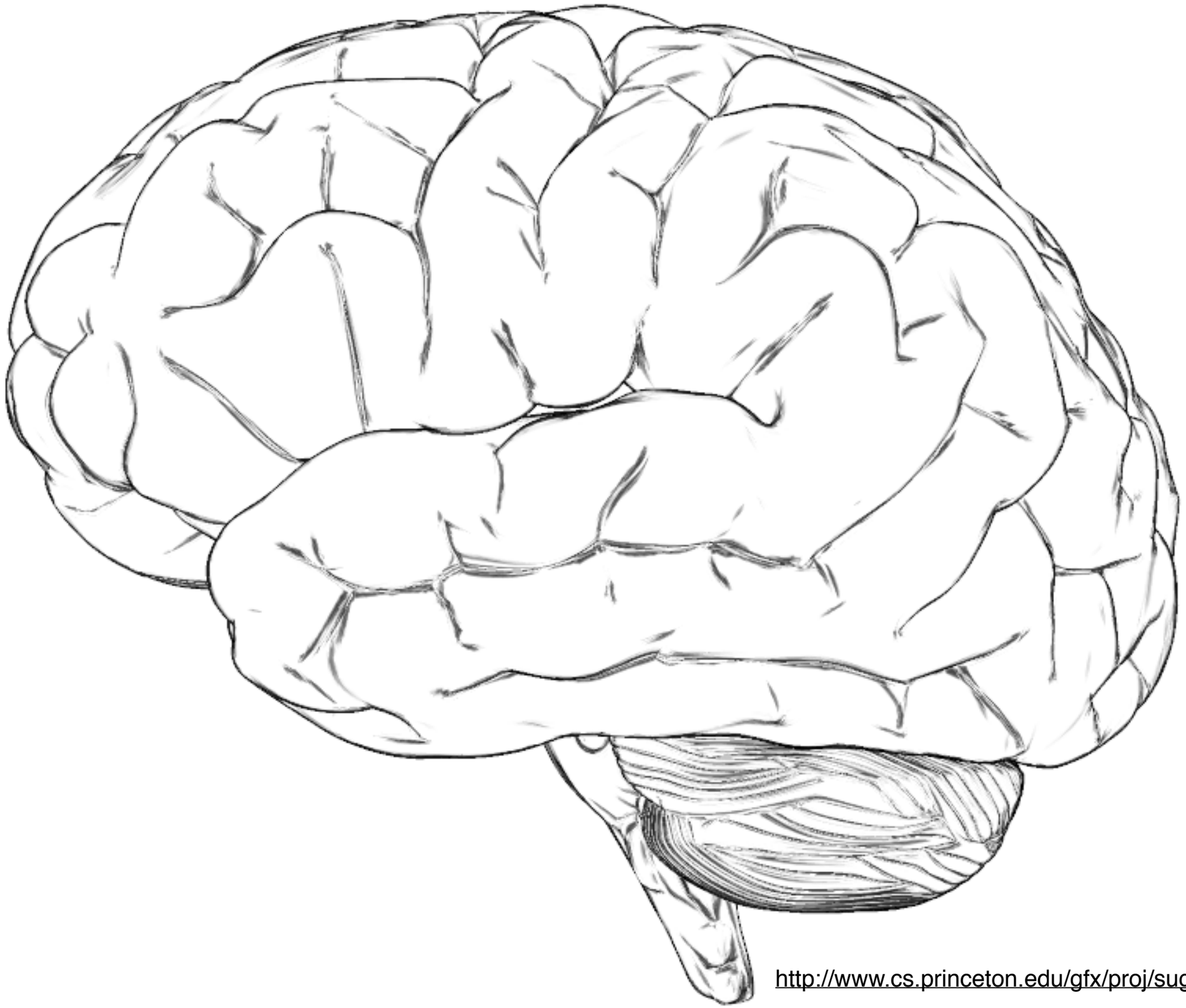
On the Italian Alps. Foggy but not too cold. Light breeze. Boy in danger of falling. Parent must be nearby.

Why study vision?

Many applications



Understanding the brain



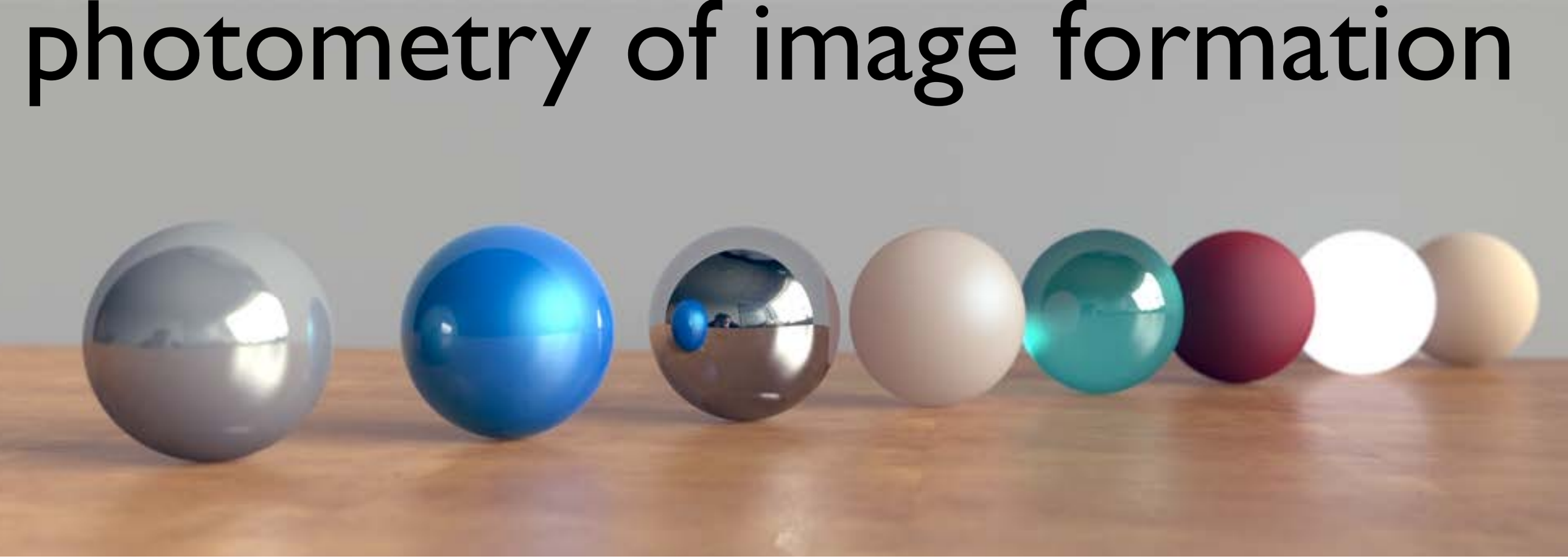
http://www.cs.princeton.edu/gfx/proj/sugcon/models/brain.png

Vision as an inverse problem

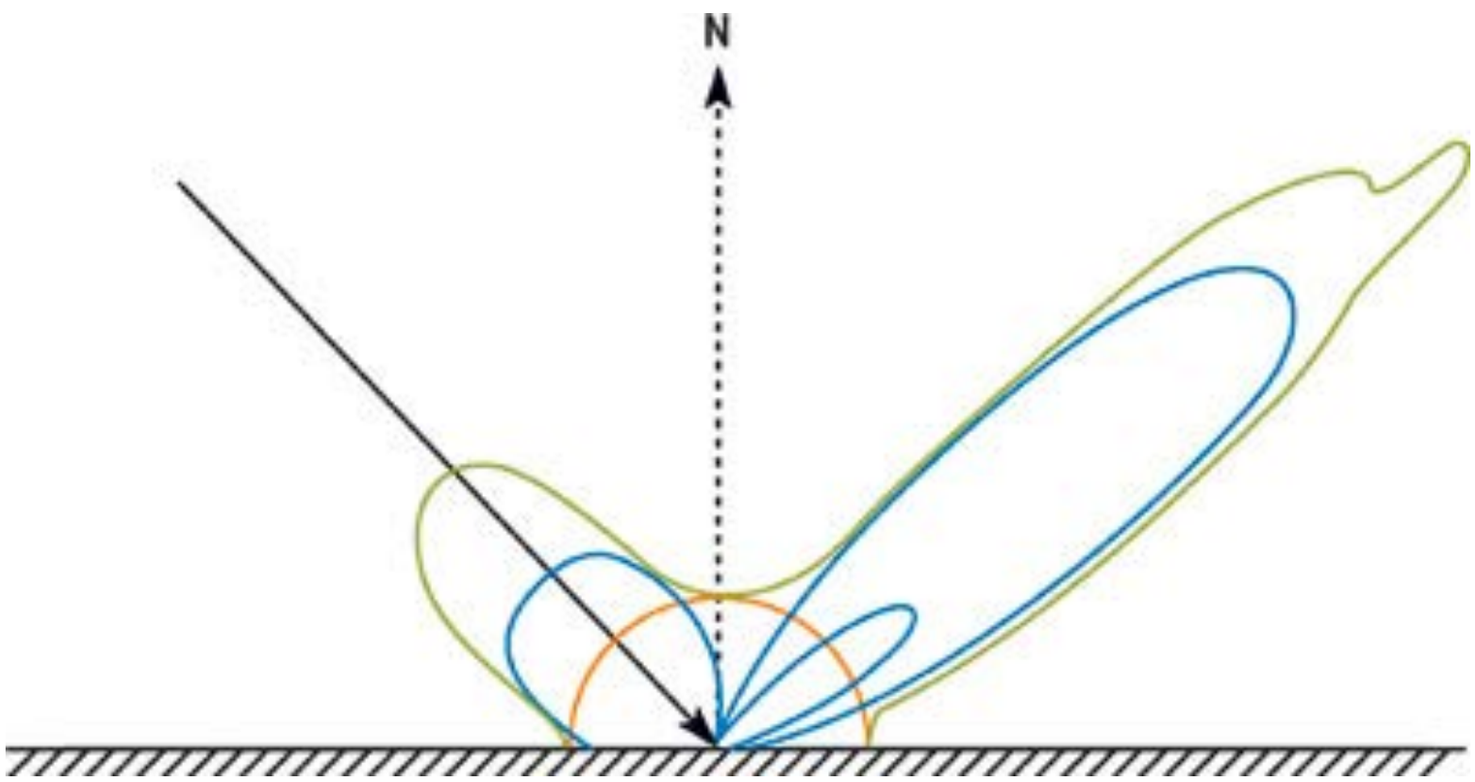
geometry of image formation

[A. Dürer 1525]

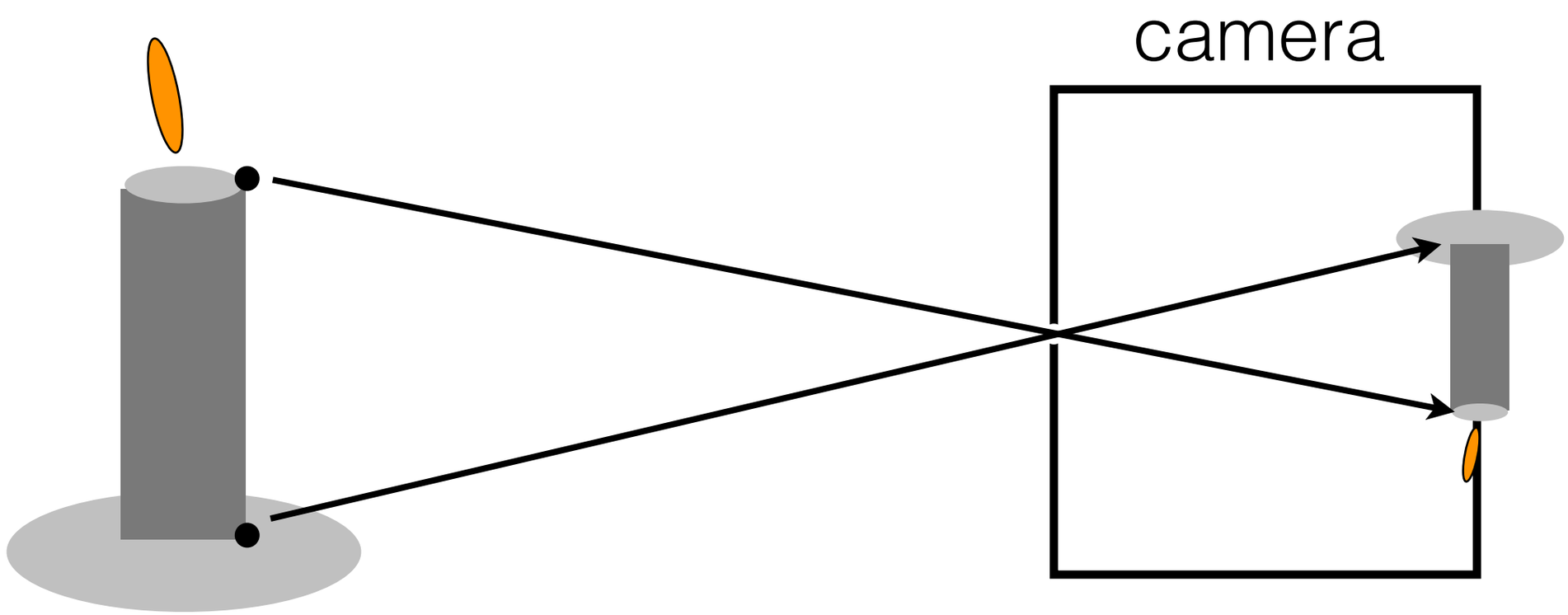
photometry of image formation



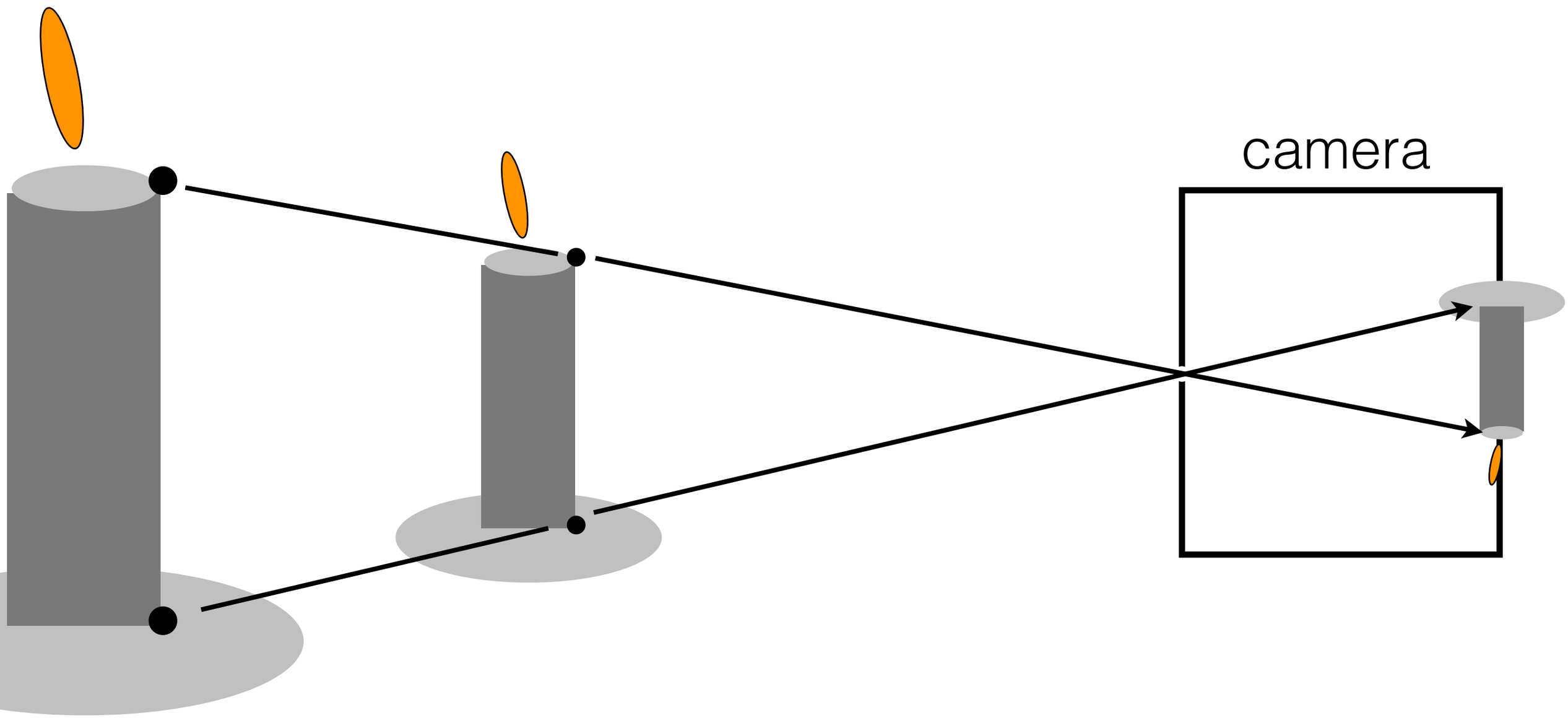
Bidirectional reflectance distribution function (BRDF)



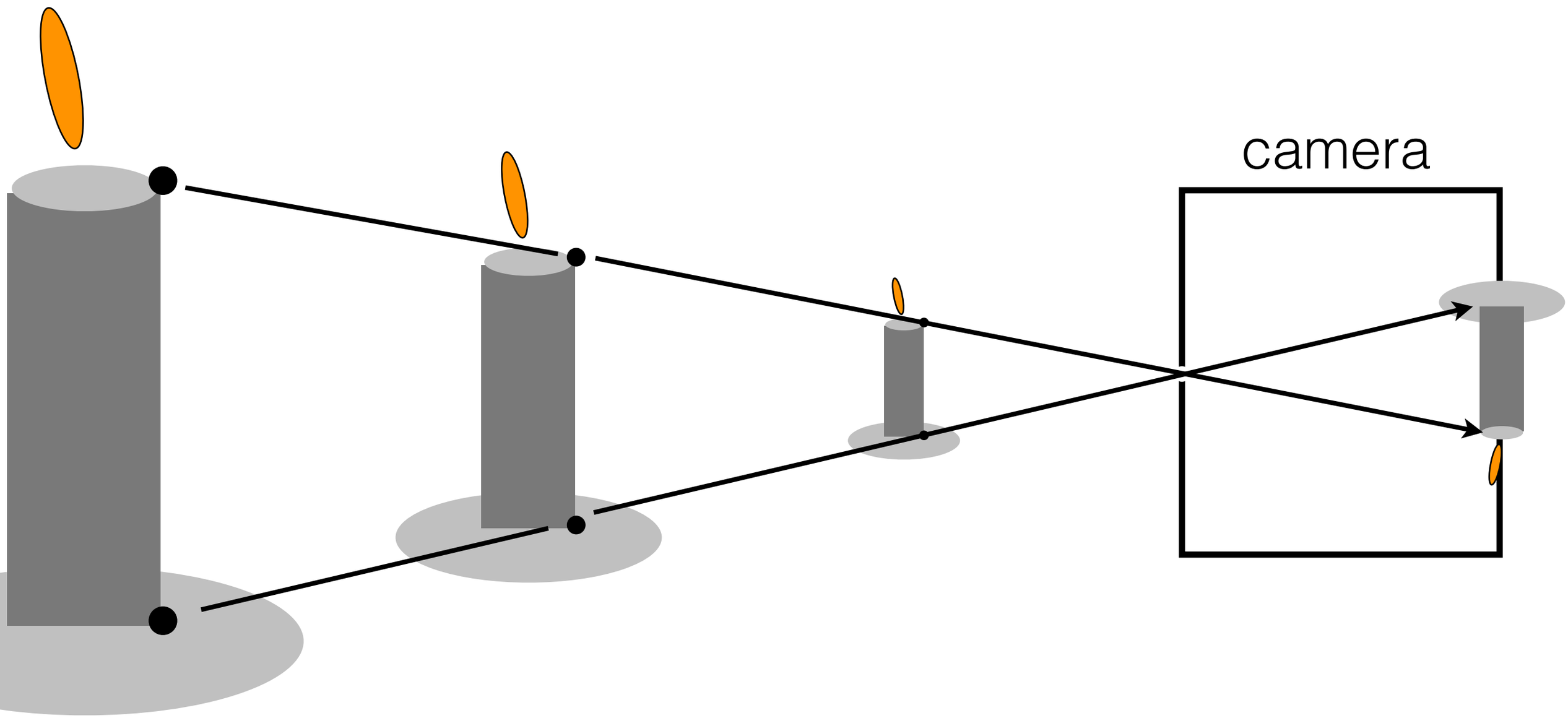
Many worlds, one image



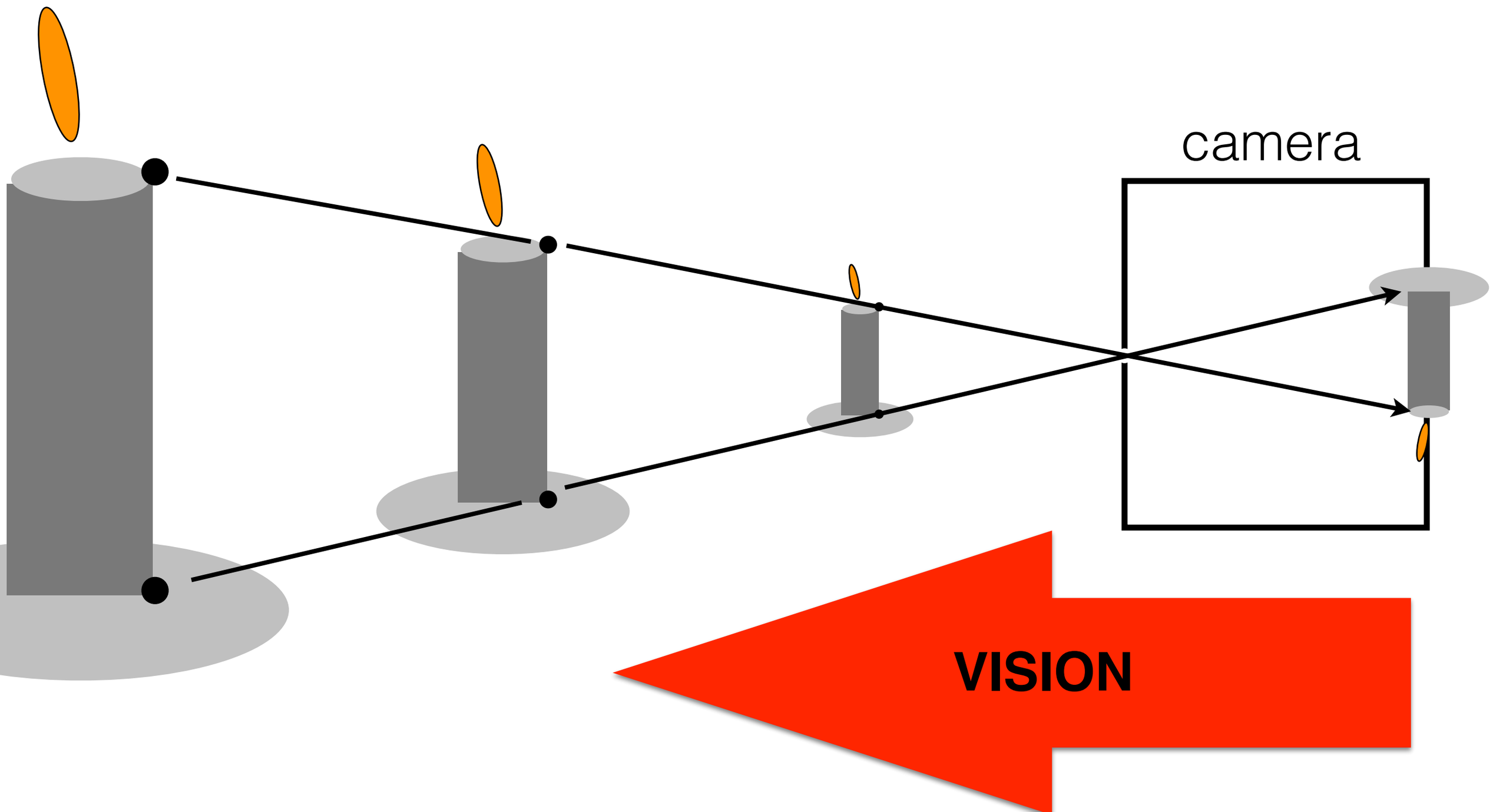
Many worlds, one image



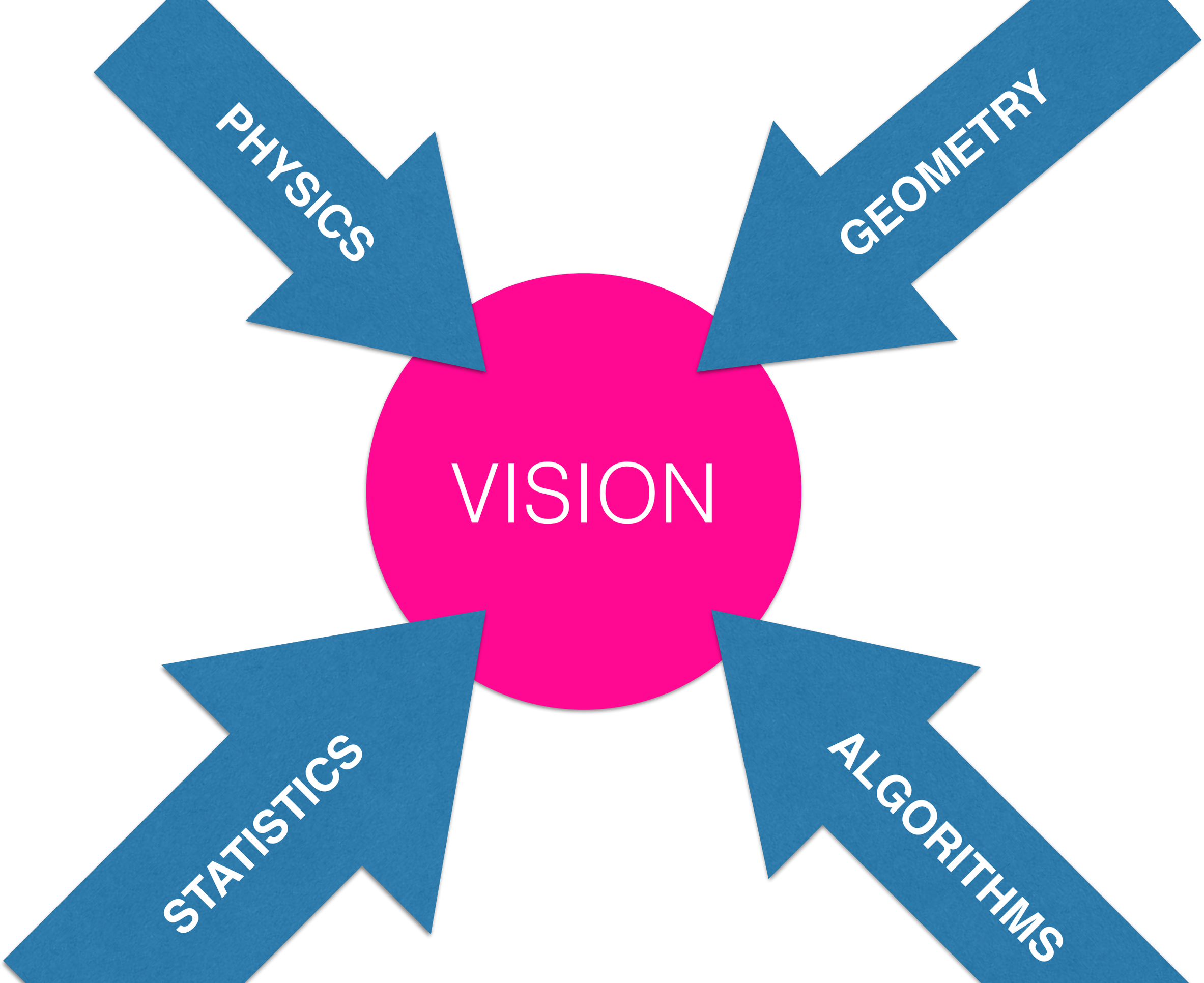
Many worlds, one image

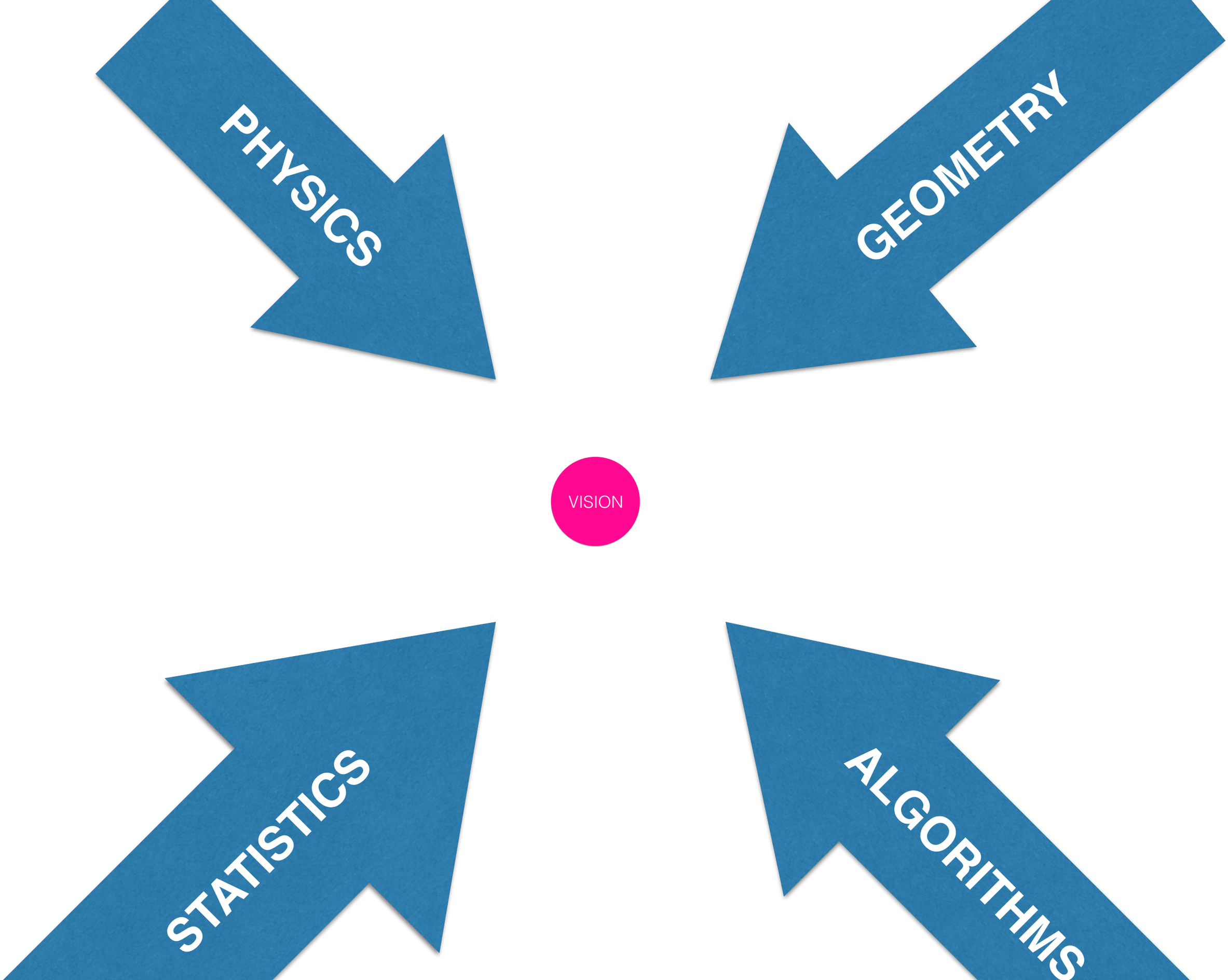


Many worlds, one image



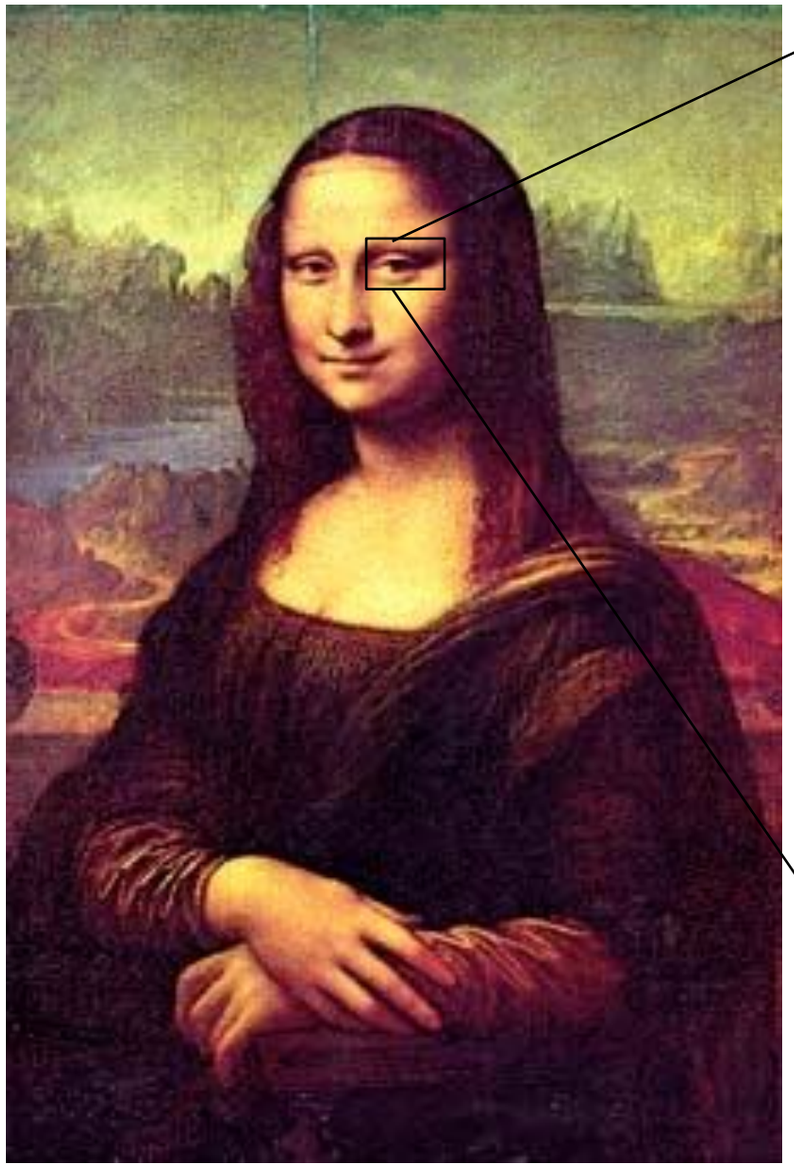
Shigeo Fukuda - Master of Deception
Underground Piano


PHYSICS
GEOMETRY
VISION
STATISTICS
ALGORITHMS


PHYSICS
GEOMETRY
VISION
STATISTICS
ALGORITHMS

A more modest proposal…

What is in an image?



8	193	199	228	238	226	242	244	244	247	235	229	243	218	195	188	162
	196	203	220	221	204	235	235	225	238	244	230	222	218	198	195	175
6	192	195	199	186	202	211	208	199	207	224	228	220	214	199	202	185
	164	163	160	146	122	98	102	114	111	124	160	183	194	189	202	195
4	132	134	143	133	131	115	110	104	71	43	53	79	127	139	172	178
	134	147	173	167	165	178	164	141	121	82	51	56	79	104	151	165
2	155	178	212	200	199	172	144	134	135	123	103	93	68	99	151	165
	158	184	217	197	175	69	19	33	37	51	68	58	47	78	131	144
0	52	82	145	197	157	44	1	5	16	9	2	8	3	7	23	46
	58	66	47	17	46	0	0	0	0	0	0	5	8	1	0	0
-2	67	115	112	75	26	0	0	0	0	8	29	30	16	9	0	0
	3	51	117	156	182	174	177	153	120	79	44	5	0	0	2	3
-4	0	0	24	47	103	99	97	74	44	11	4	0	0	0	8	12
	35	28	15	6	27	19	8	3	10	9	32	33	23	31	32	25
-6	27	26	45	63	73	74	60	58	74	73	85	67	71	72	63	48
	68	73	83	80	91	119	122	115	123	118	133	113	104	106	99	85
-8																
	-8	-6	-4	-2	0	2	4	6	8							

The challenge

8	193	199	228	238	226	242	244	244	247	235	229	243	218	195	188	162
	196	203	220	221	204	235	235	225	238	244	230	222	218	198	195	175
6	192	195	199	186	202	211	208	199	207	224	228	220	214	199	202	185
	164	163	160	146	122	98	102	114	111	124	160	183	194	189	202	195
4	132	134	143	133	131	115	110	104	71	43	53	79	127	139	172	178
	134	147	173	167	165	178	164	141	121	82	51	56	79	104	151	165
2	155	178	212	200	199	172	144	134	135	123	103	93	68	99	151	165
	158	184	217	197	175	69	19	33	37	51	68	58	47	78	131	144
0	52	82	145	197	157	44	1	5	16	9	2	8	3	7	23	46
	58	66	47	17	46	0	0	0	0	0	0	5	8	1	0	0
-2	67	115	112	75	26	0	0	0	0	8	29	30	16	9	0	0
	3	51	117	156	182	174	177	153	120	79	44	5	0	0	2	3
-4	0	0	24	47	103	99	97	74	44	11	4	0	0	0	8	12
	35	28	15	6	27	19	8	3	10	9	32	33	23	31	32	25
-6	27	26	45	63	73	74	60	58	74	73	85	67	71	72	63	48
	68	73	83	80	91	119	122	115	123	118	133	113	104	106	99	85
-8																
	-8		-6		-4		-2		0		2		4		6		8

The challenge

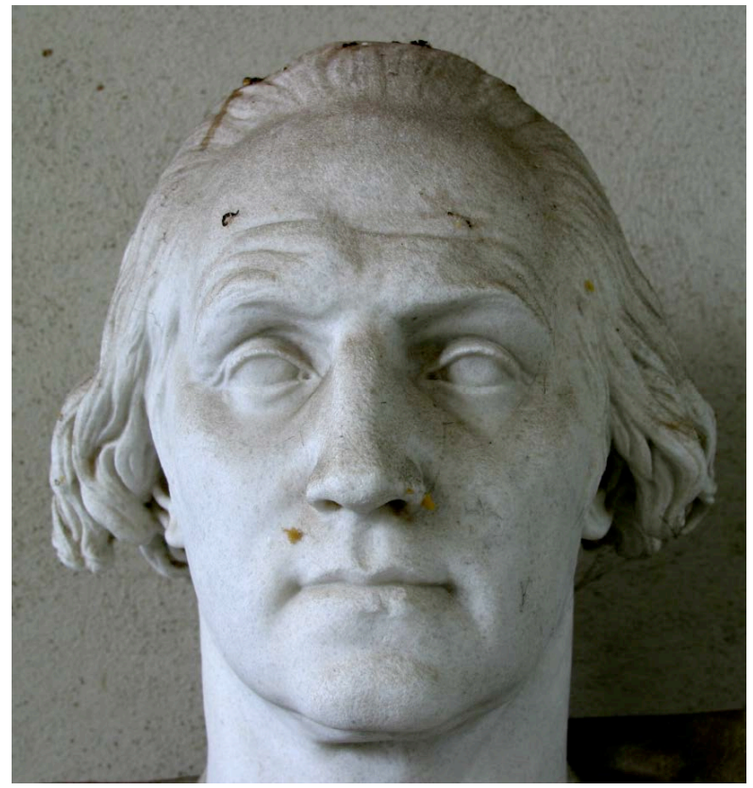
193	199	228	238	226	242	244	244	247	235	229	243	218	195	188	162
196	203	220	221	204	235	235	225	238	244	230	222	218	198	195	175
192	195	199	186	202	211	208	199	207	224	228	220	214	199	202	185
164	163	160	146	122	98	102	114	111	124	160	183	194	189	202	195
132	134	143	133	131	115	110	104	71	43	53	79	127	139	172	178
134	147	173	167	165	178	164	141	121	82	51	56	79	104	151	165
155	178	212	200	199	172	144	134	135	123	103	93	68	99	151	165
158	184	217	197	175	69	19	33	37	51	68	58	47	78	131	144
52	82	145	197	157	44	1	5	16	9	2	8	3	7	23	46
58	66	47	17	46	0	0	0	0	0	0	5	8	1	0	0
67	115	112	75	26	0	0	0	0	8	29	30	16	9	0	0
3	51	117	156	182	174	177	153	120	79	44	5	0	0	2	3
0	0	24	47	103	99	97	74	44	11	4	0	0	0	8	12
35	28	15	6	27	19	8	3	10	9	32	33	23	31	32	25
27	26	45	63	73	74	60	58	74	73	85	67	71	72	63	48
68	73	83	80	91	119	122	115	123	118	133	113	104	106	99	85

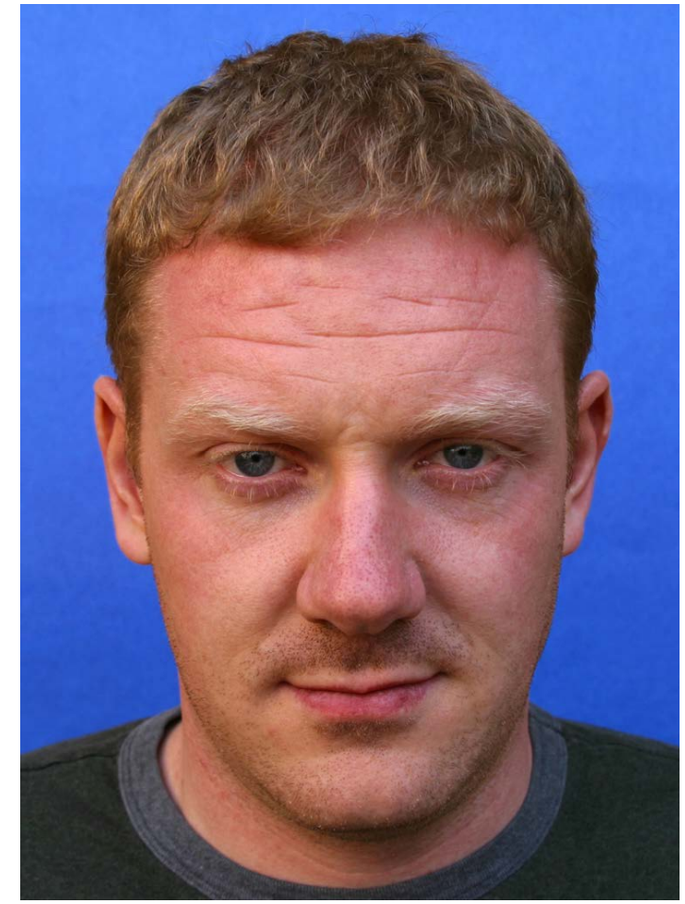
8 6 4 2 0 -2 -4 -6 -8

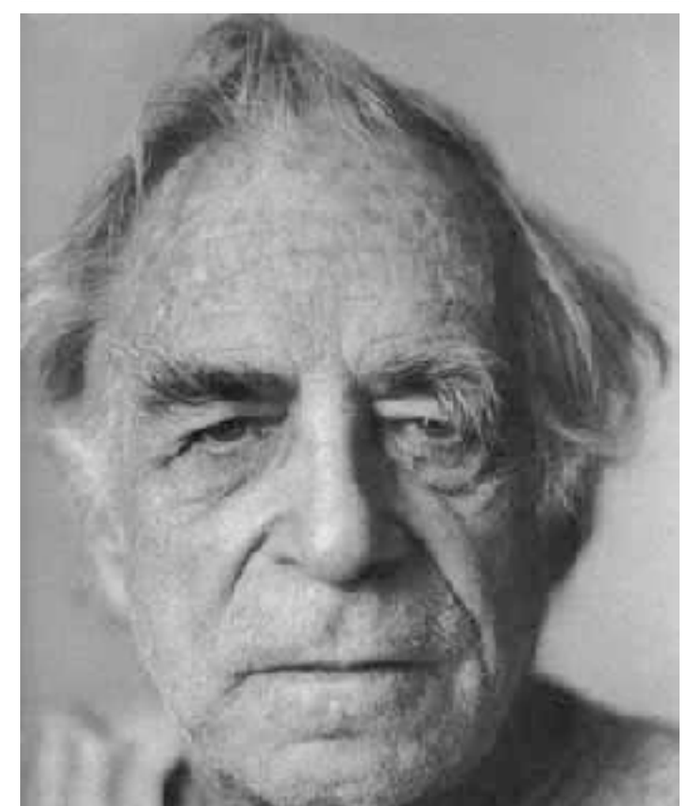
-8 -6 -4 -2 0 2 4 6 8

“Eye”

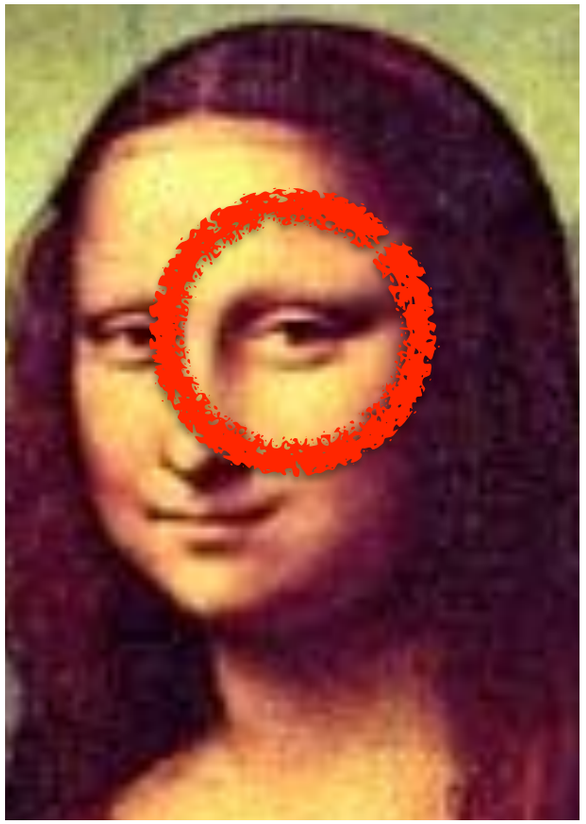
``Eye''

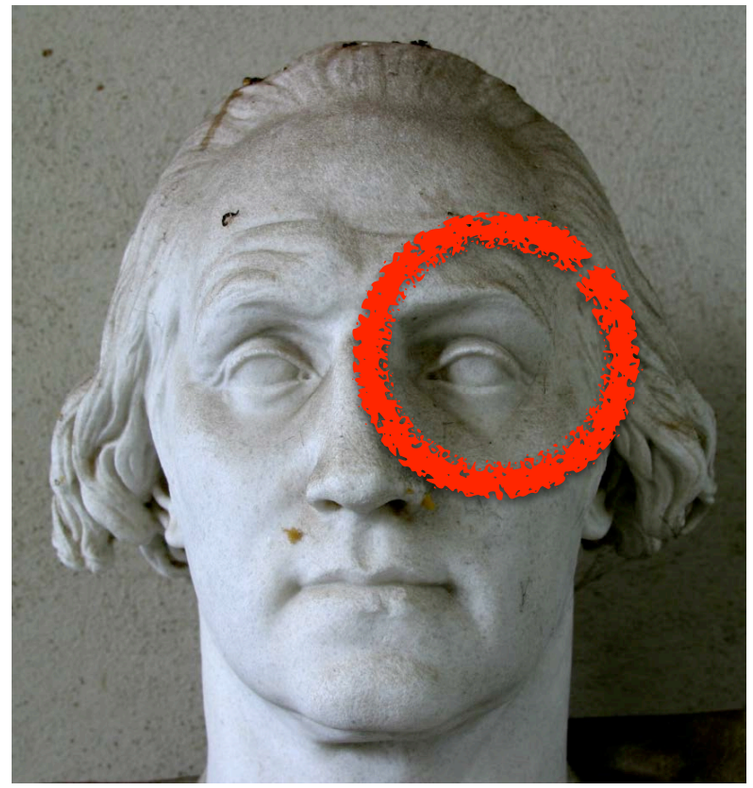


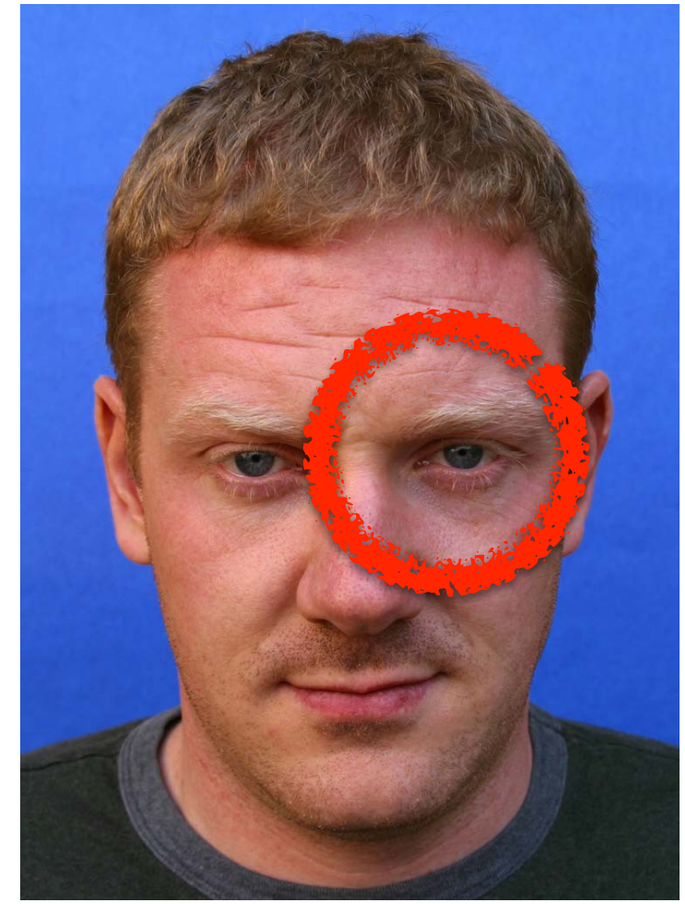




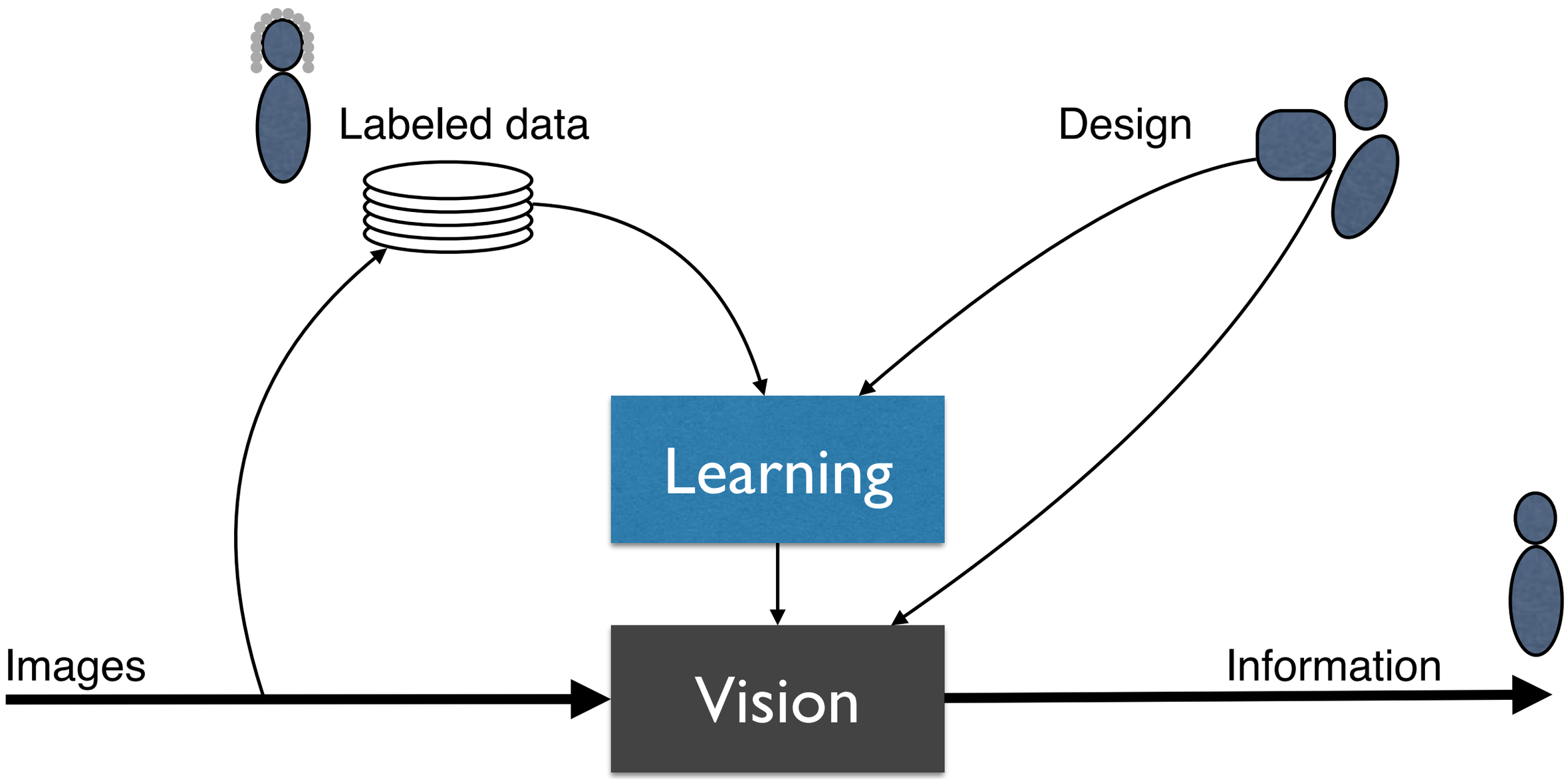
``Eye''

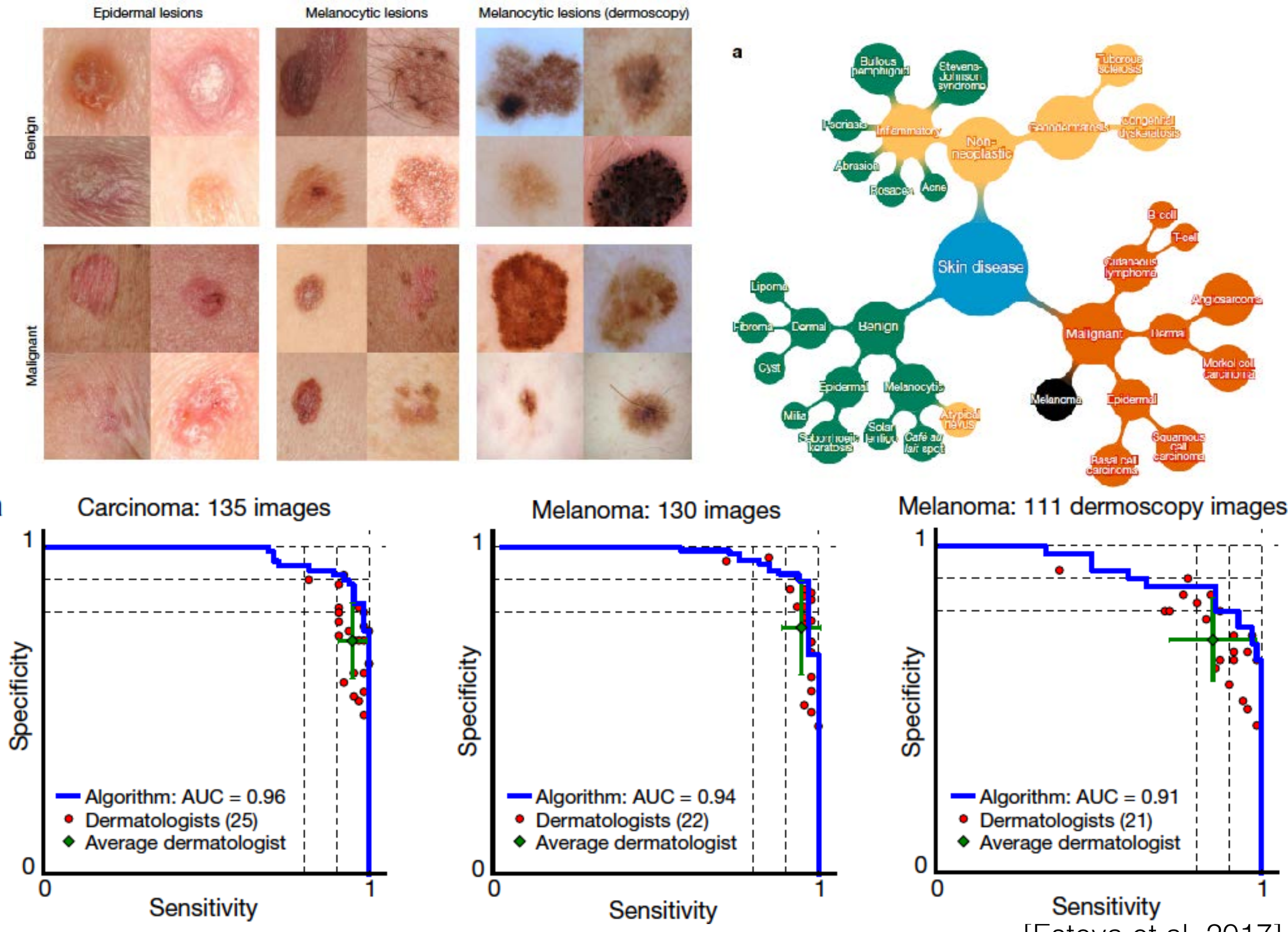




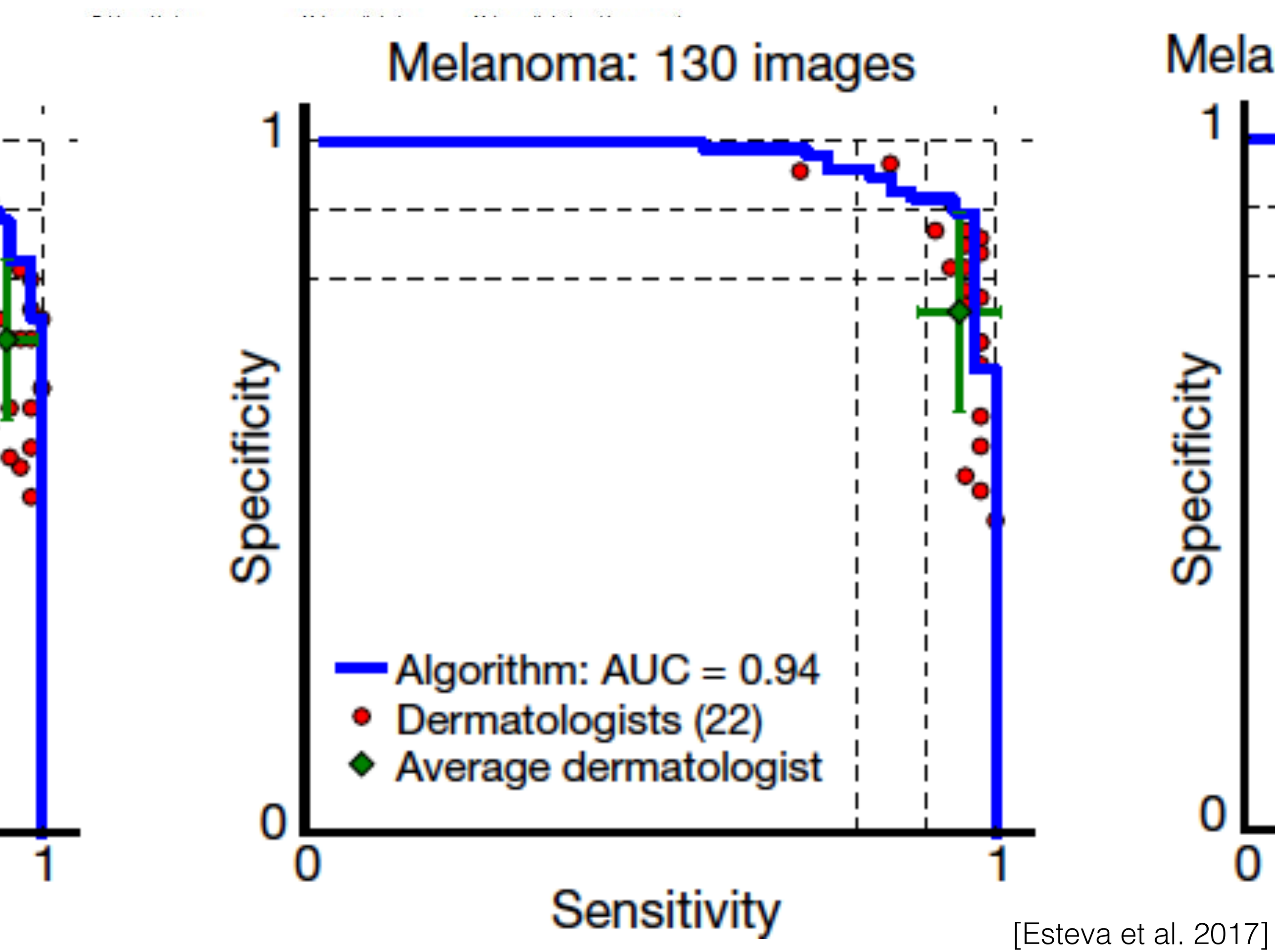


Learning-based approach

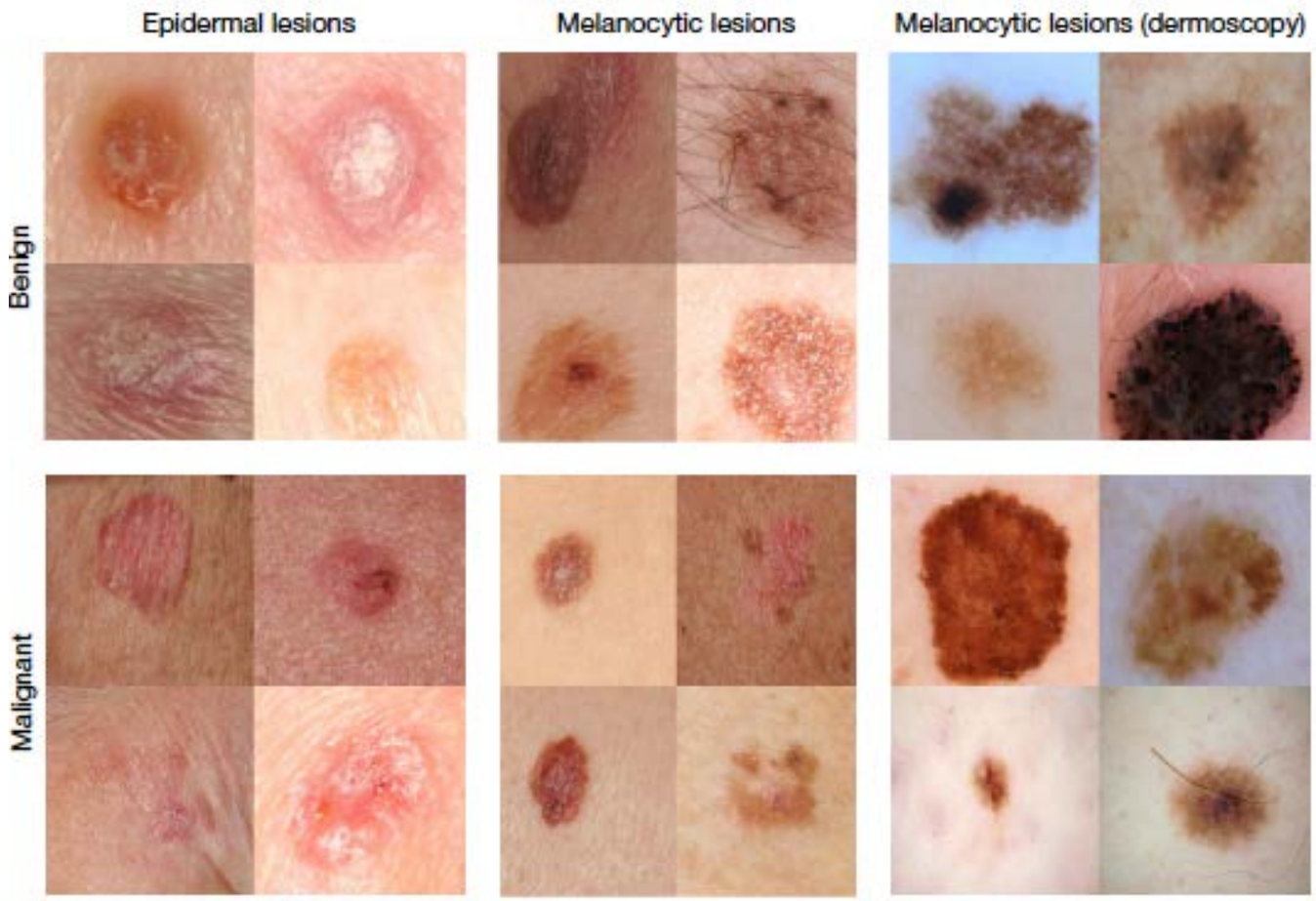


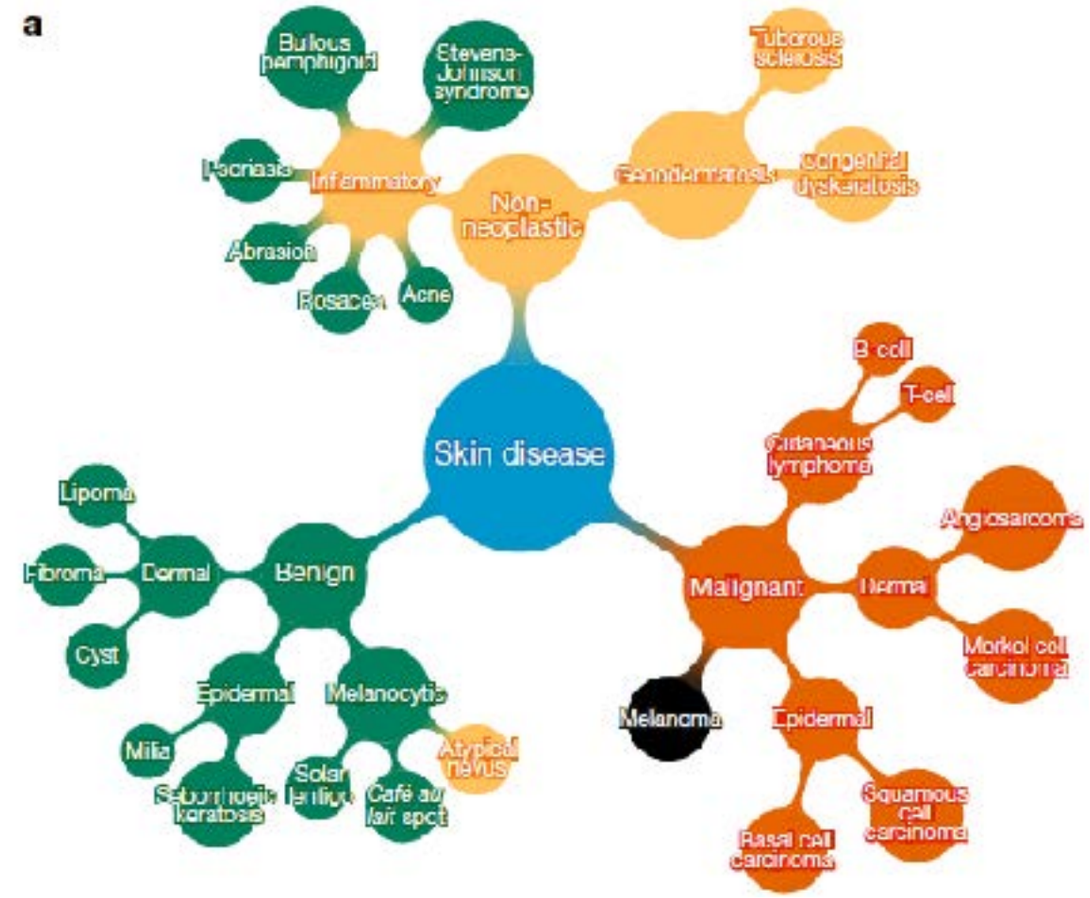


[Esteva et al. 2017]

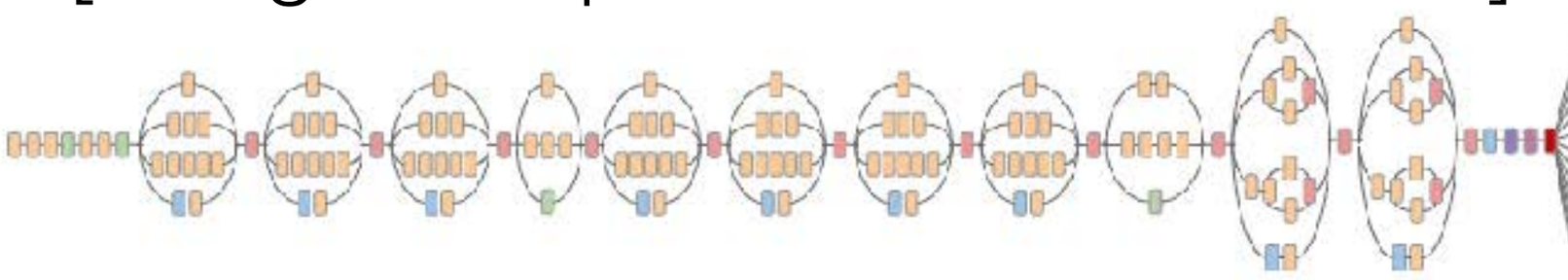


[Esteva et al. 2017]

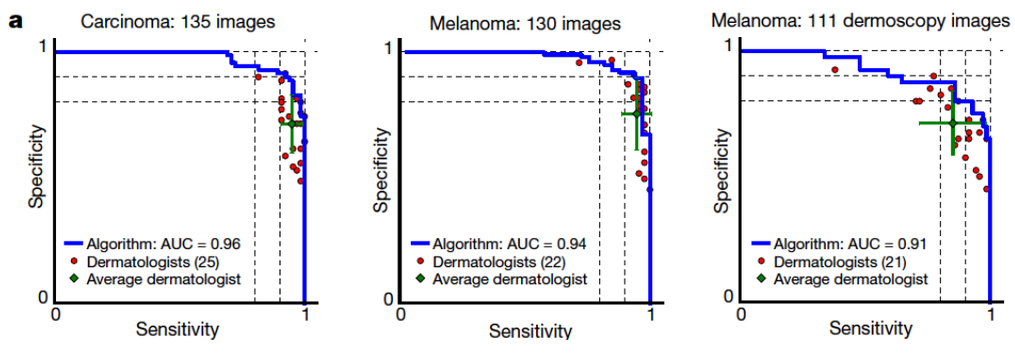




[Google inception architecture 2015]

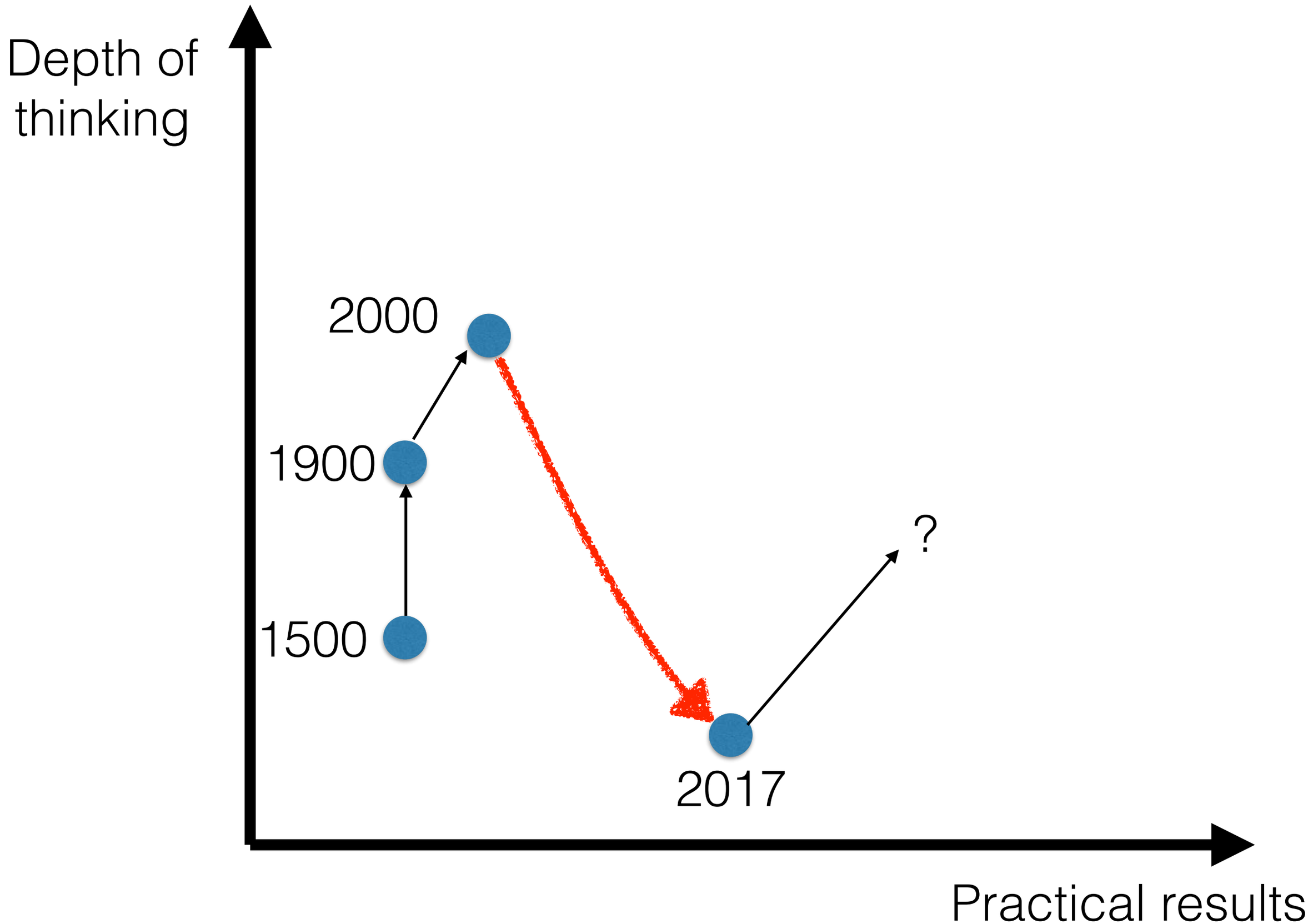


- Convolution
- AvgPool
- MaxPool
- Concat
- Dropout
- Fully connected
- Softmax



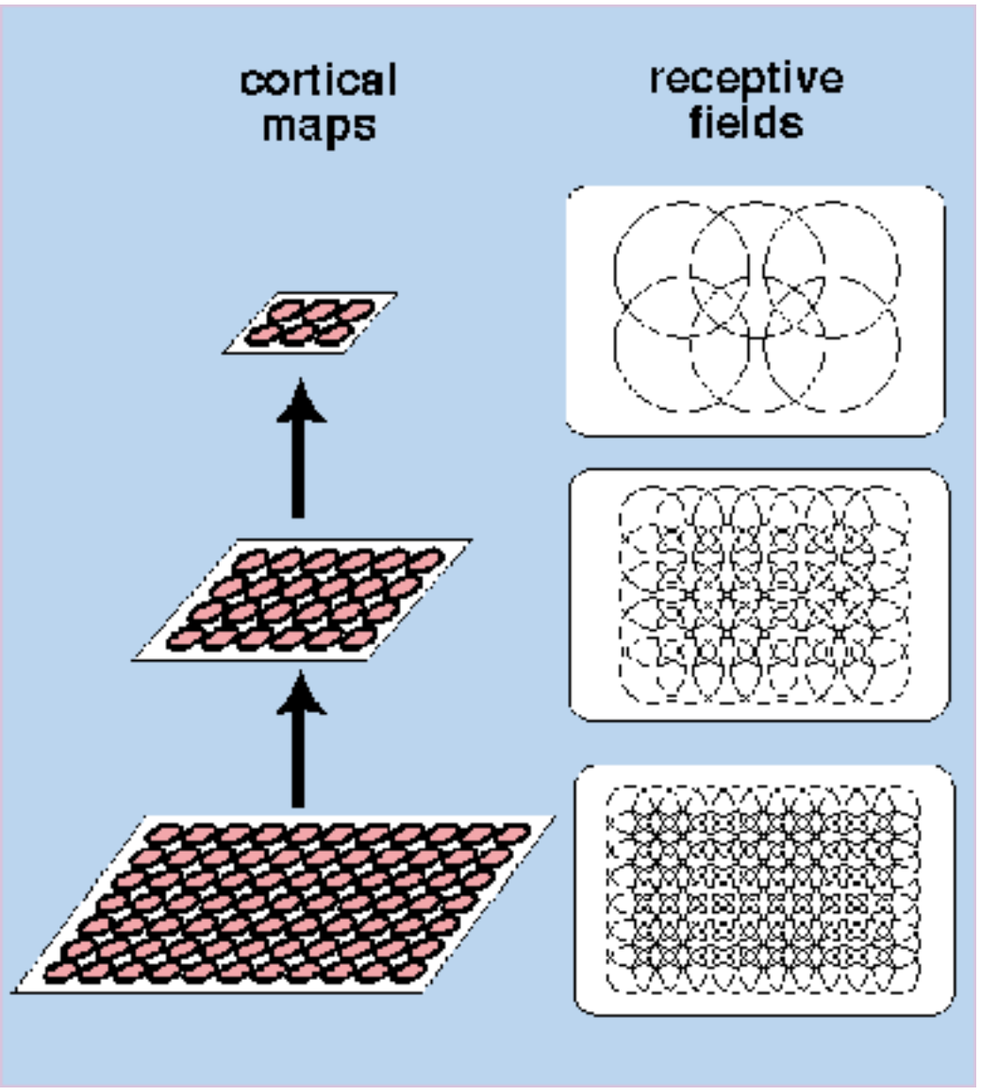
- Acral-lentiginous melanoma
- Amelanotic melanoma
- Lentigo melanoma
- ...
- Blue nevus
- Halo nevus
- Mongolian spot
- ...

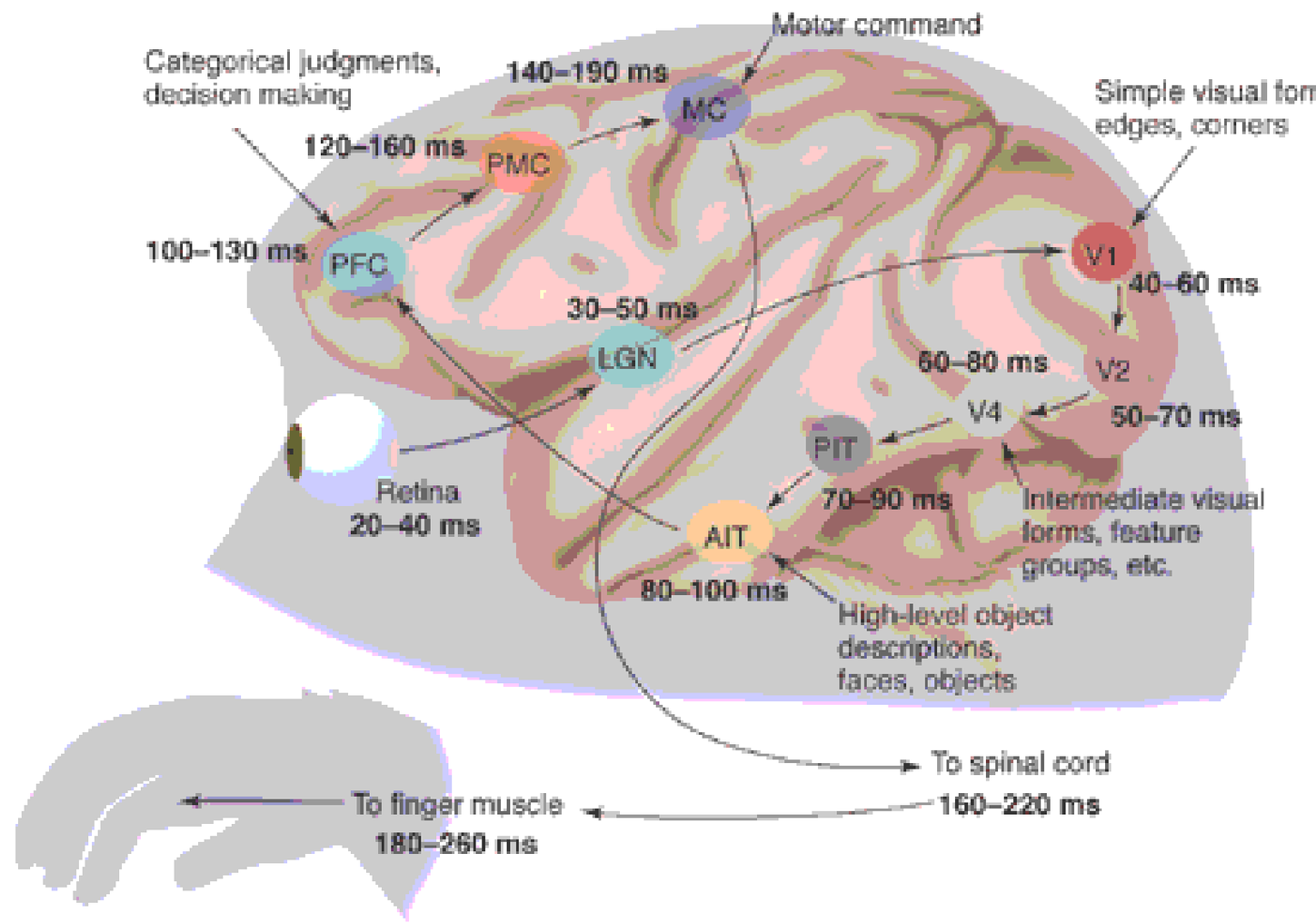
[Esteva et al. 2017]

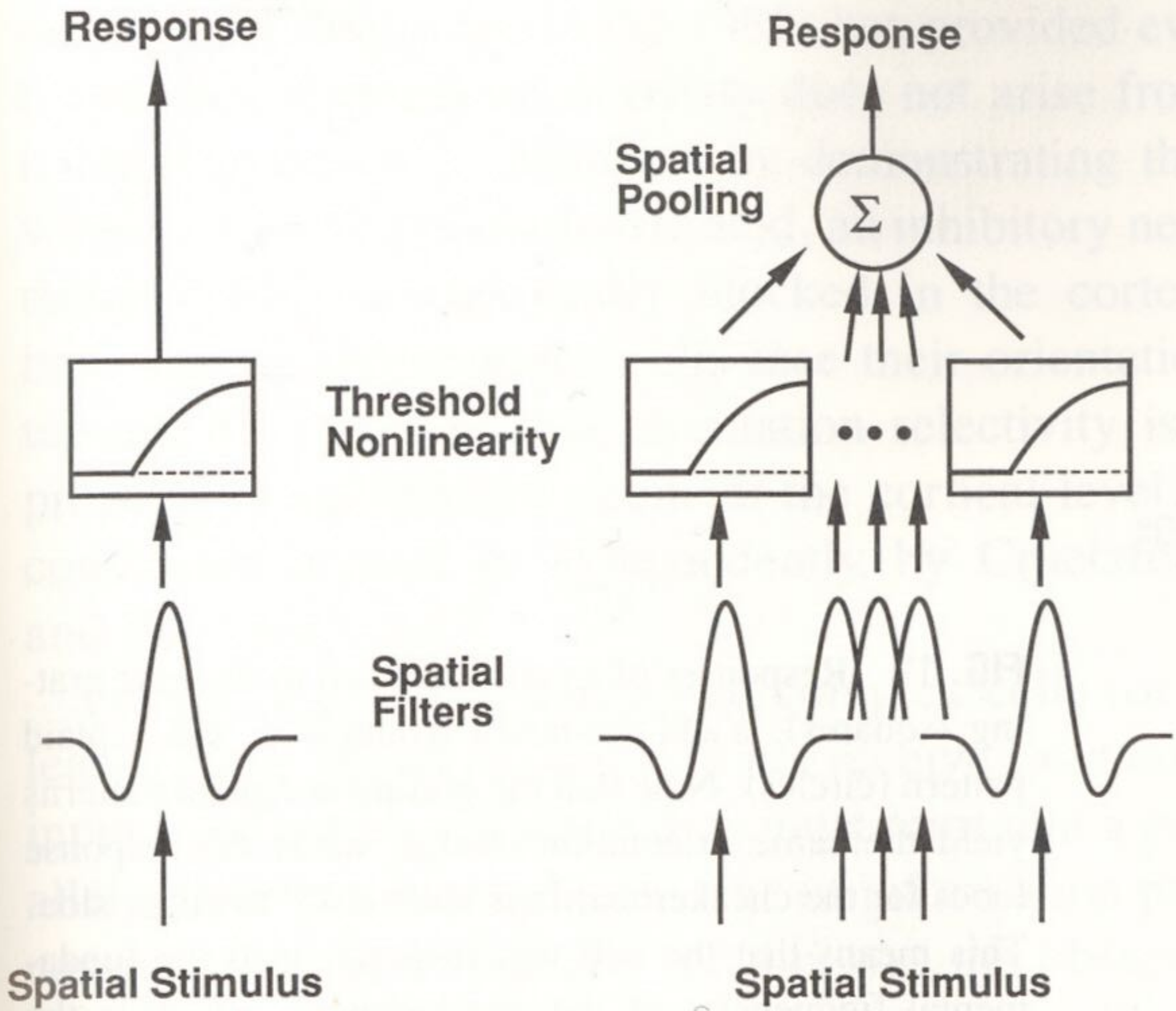

Depth of thinking
2000
1900
1500
?
2017
Practical results

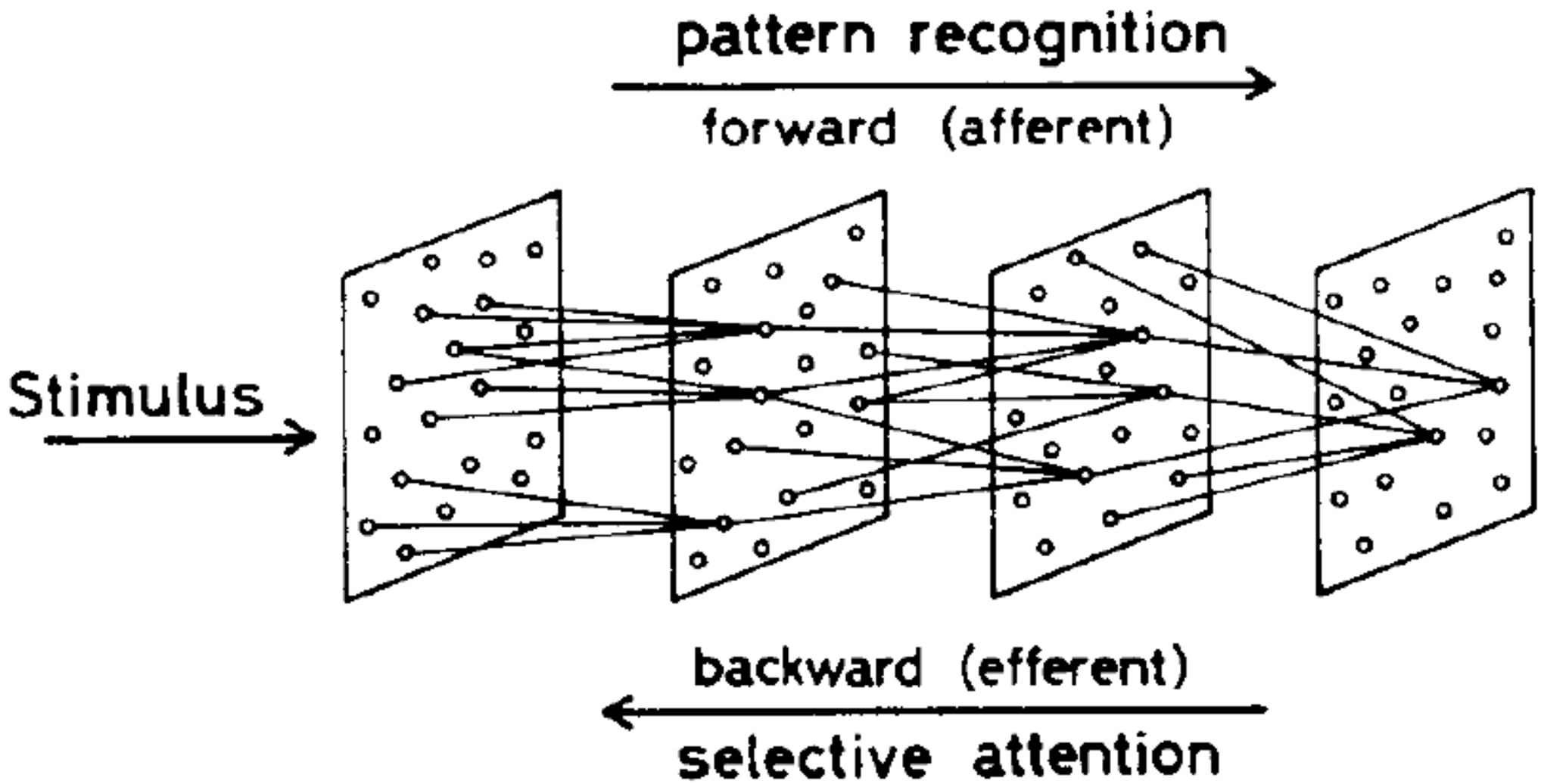
Deep networks

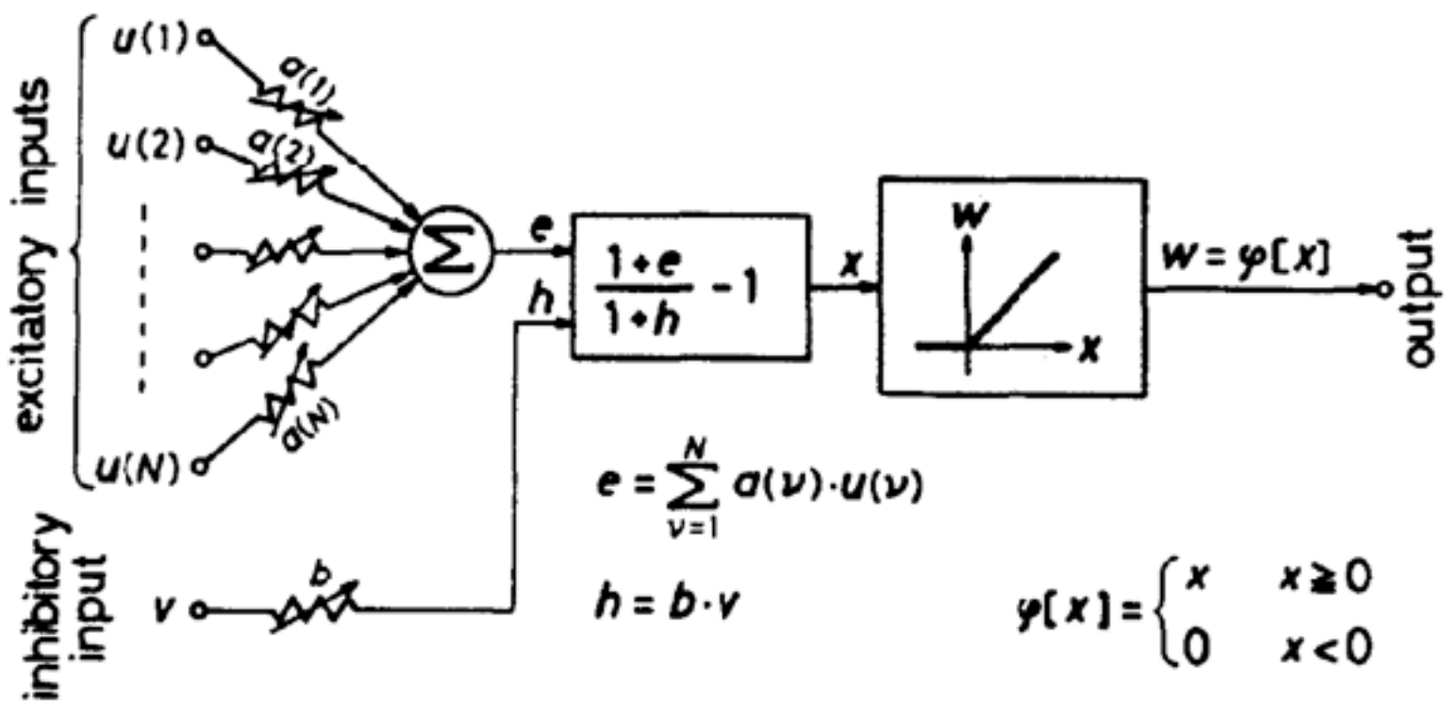
Visual system



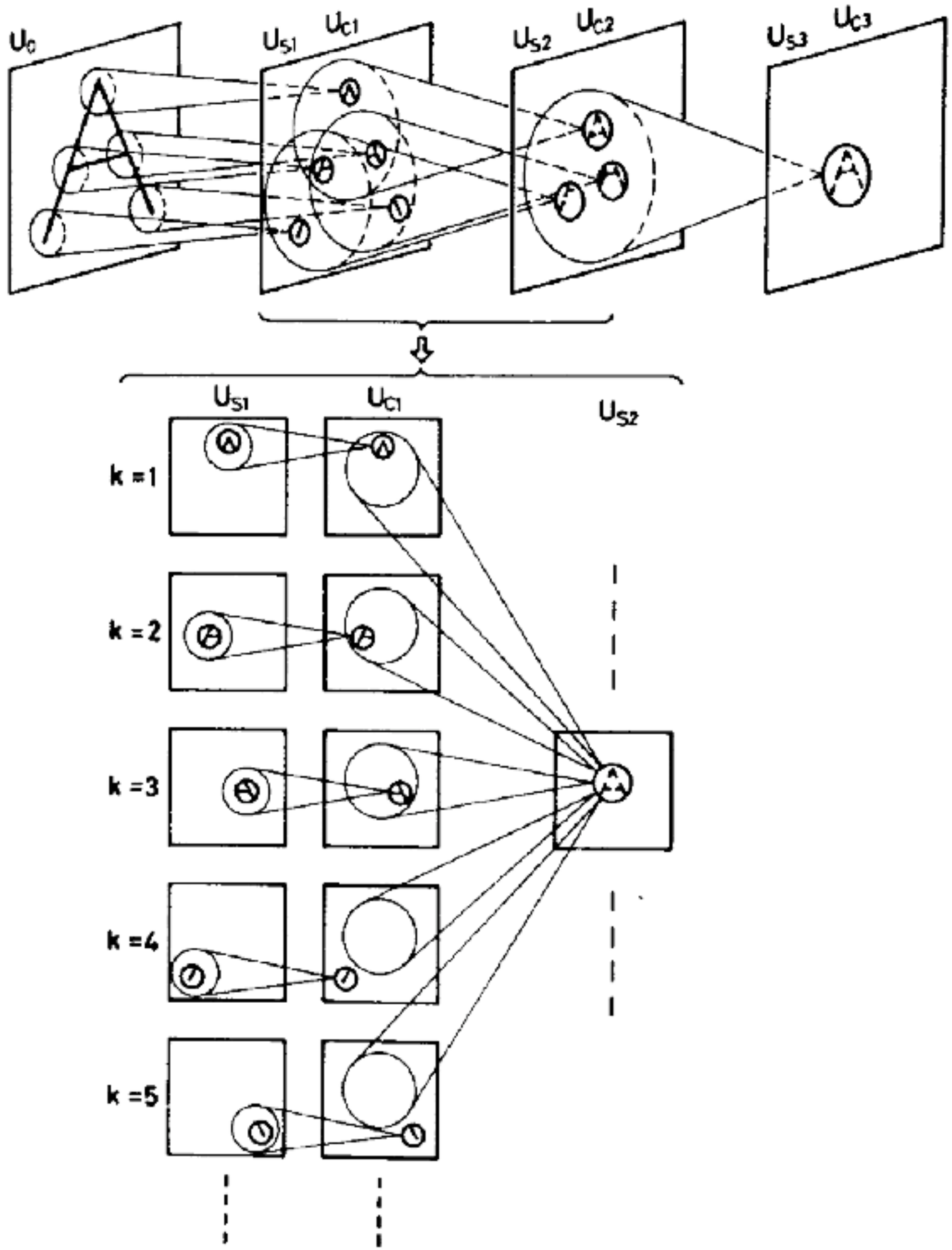



Response
Threshold
Nonlinearity
Spatial
Filters
Spatial Stimulus
Response
Spatial
Pooling
Σ
Spatial Stimulus





[Fukushima 1980]



[Fukushima 1980]

LeNet + backpropagation (1988)

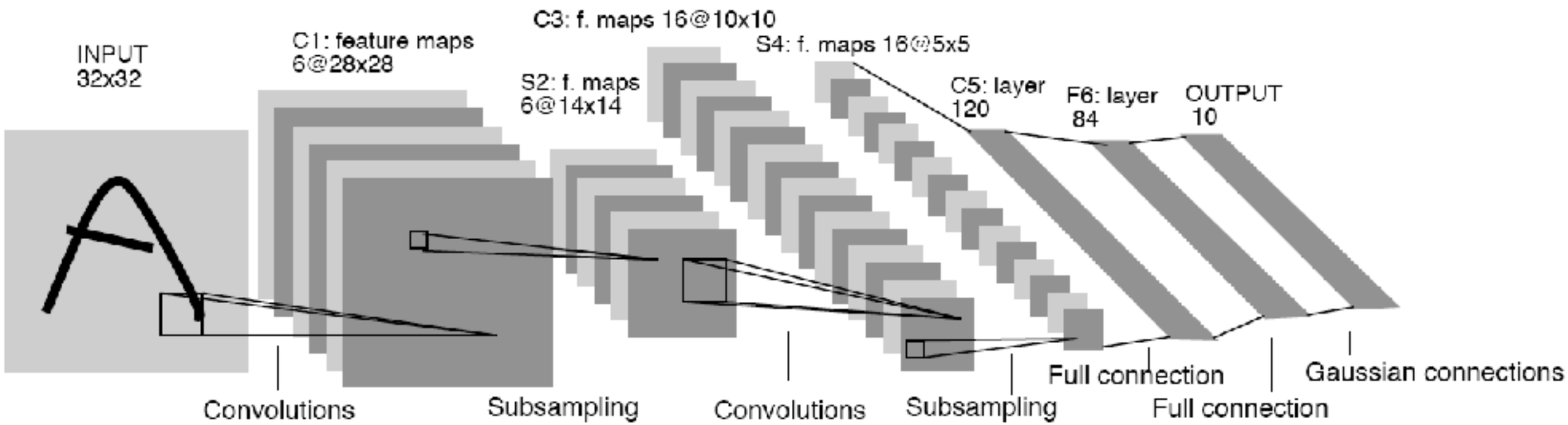


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

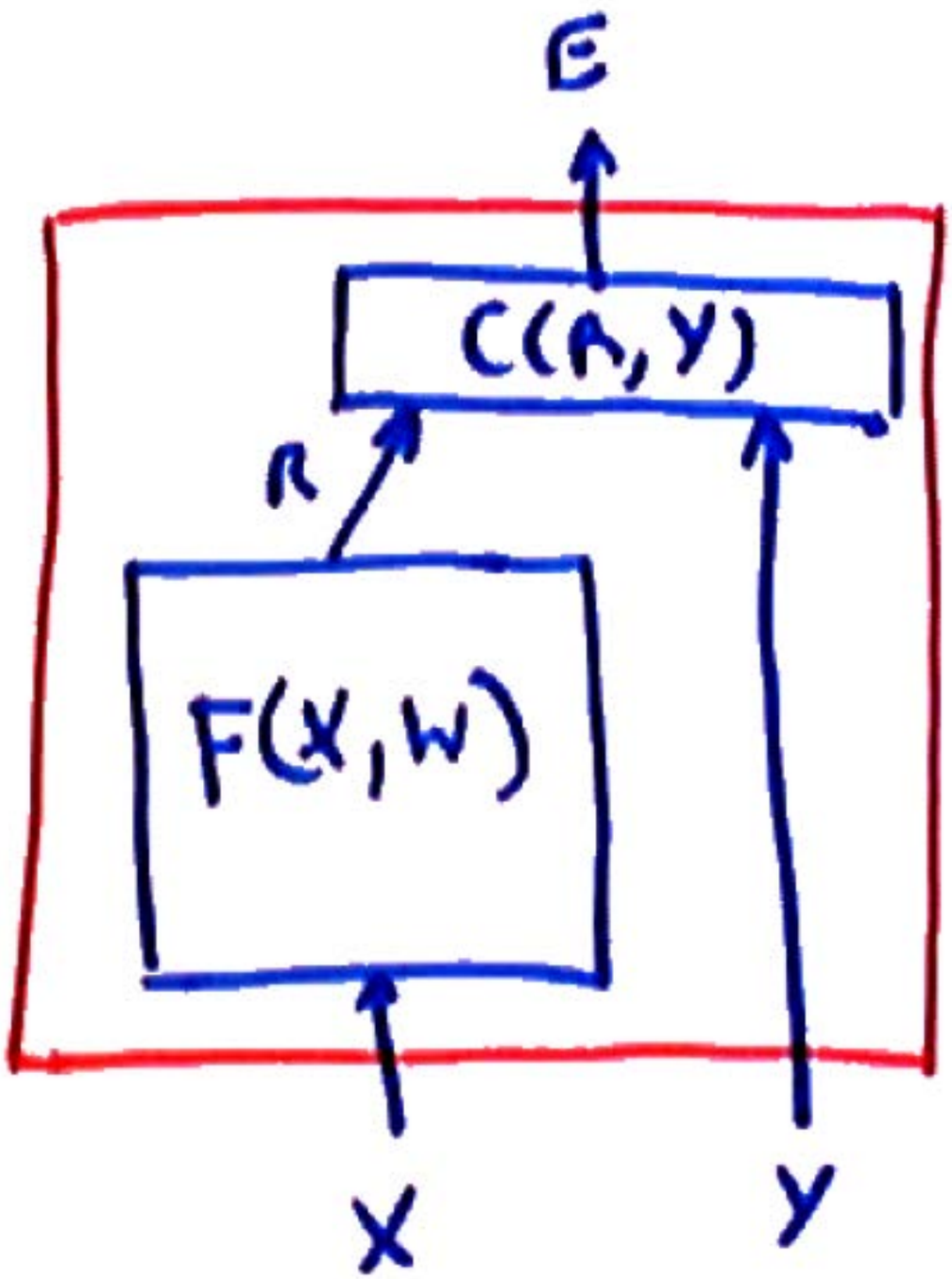
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

TABLE I

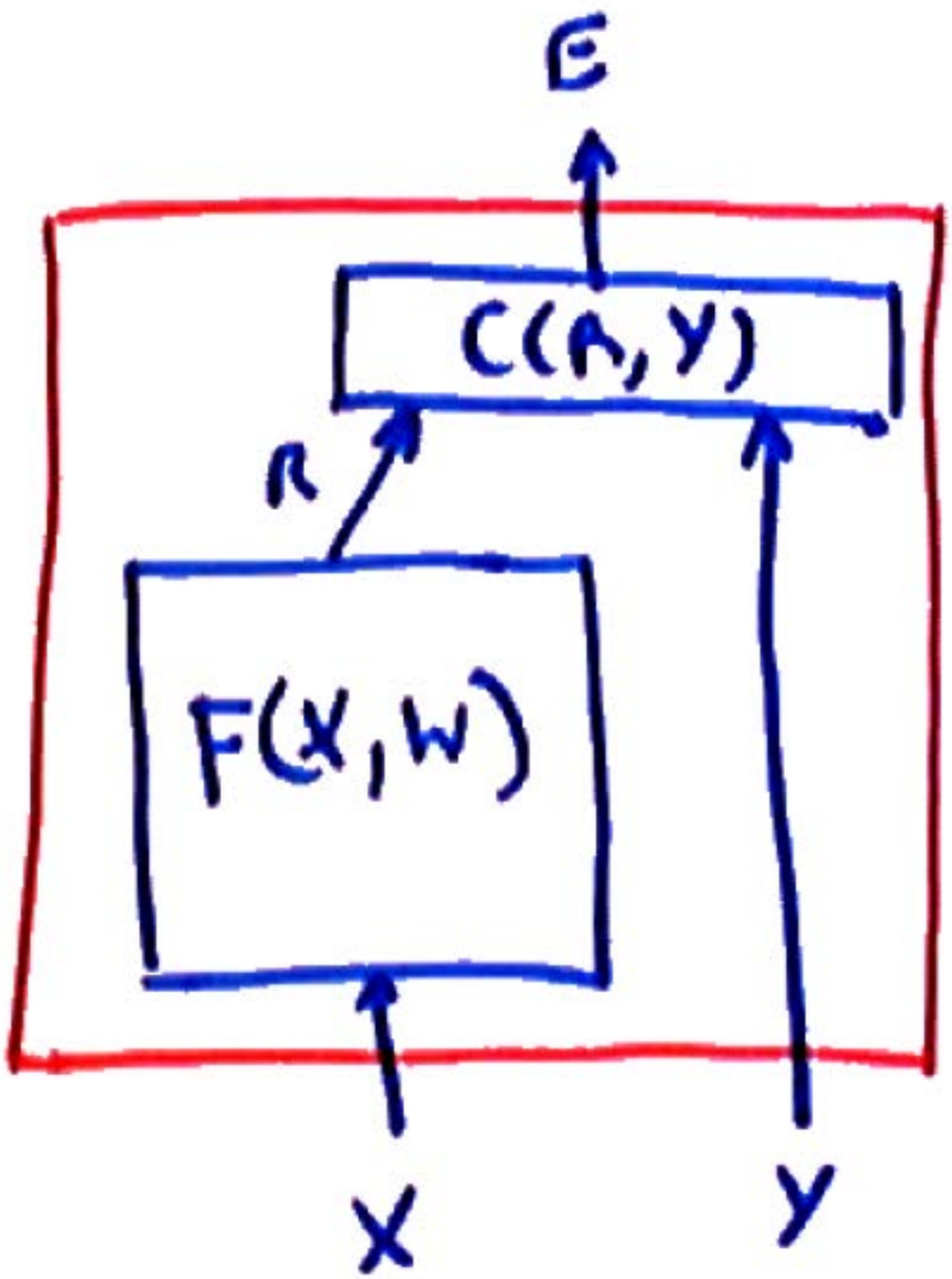
EACH COLUMN INDICATES WHICH FEATURE MAP IN S2 ARE COMBINED BY THE UNITS IN A PARTICULAR FEATURE MAP OF C3.

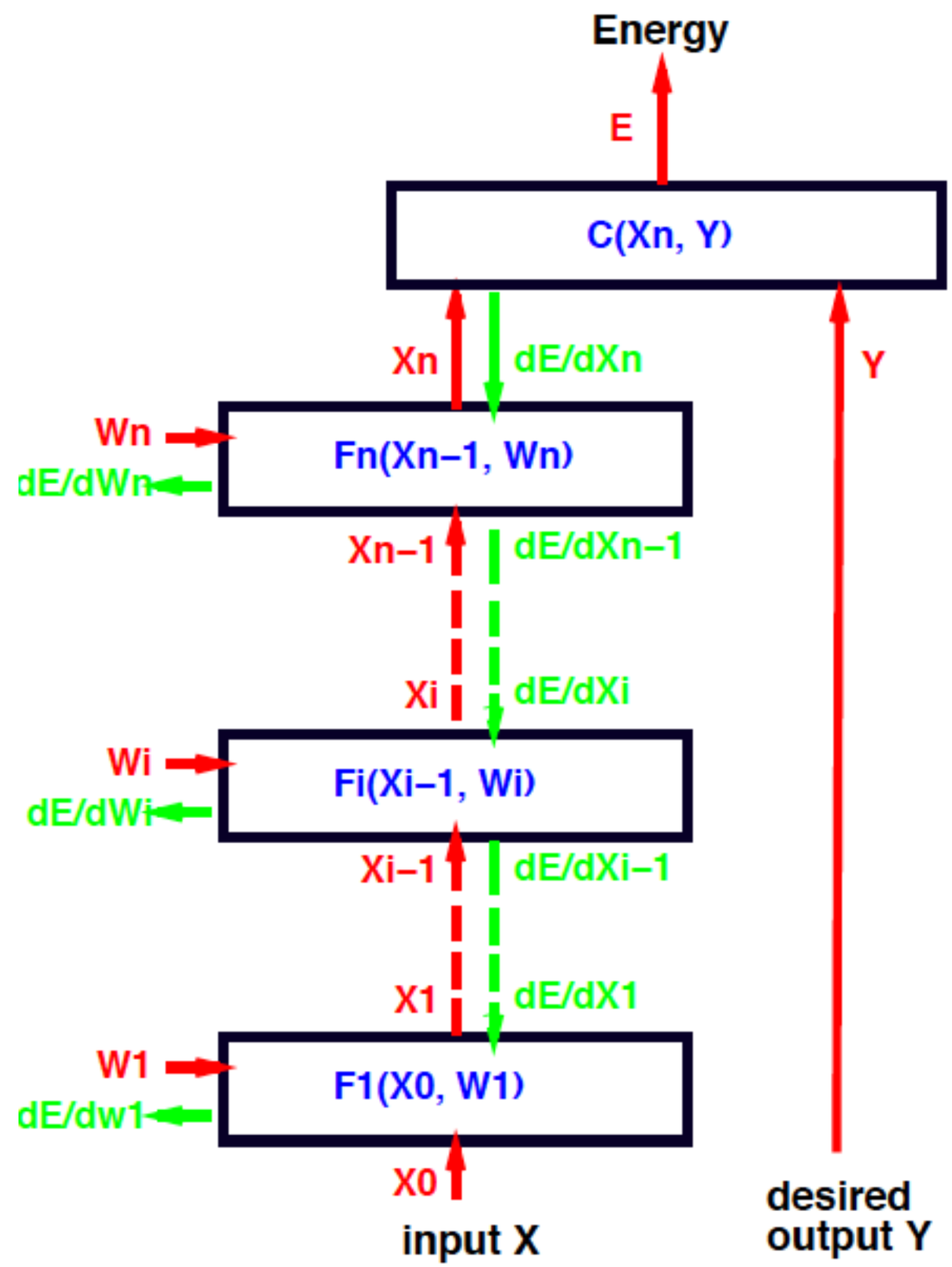
[LeCun et al. 1988]

Backpropagation

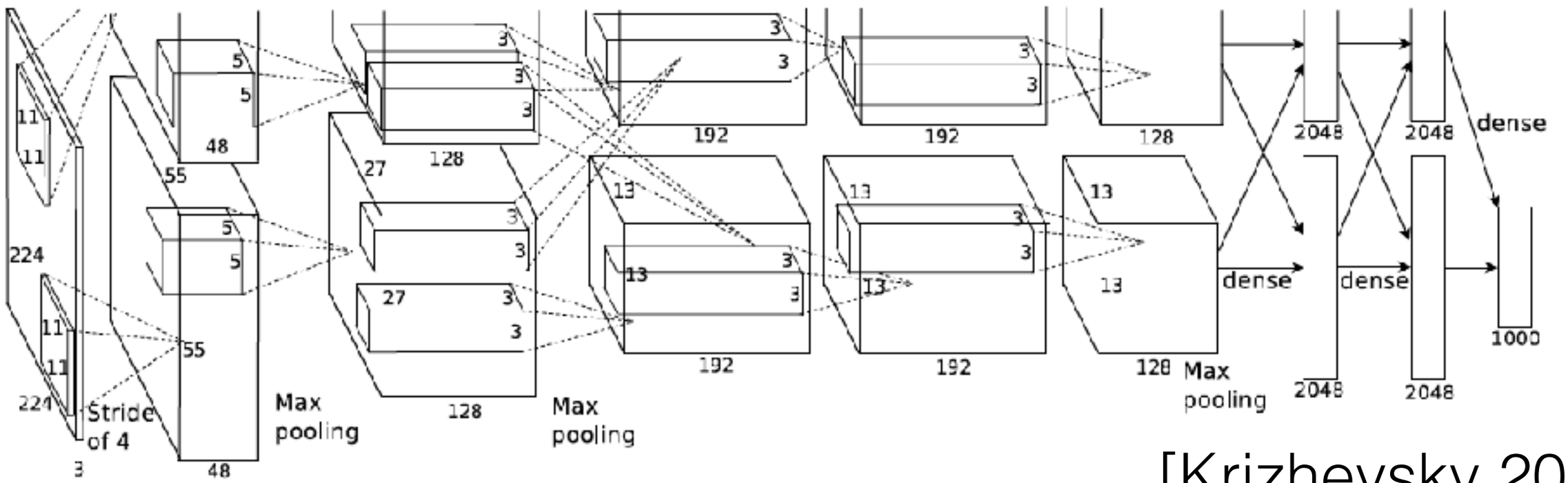


Backpropagation

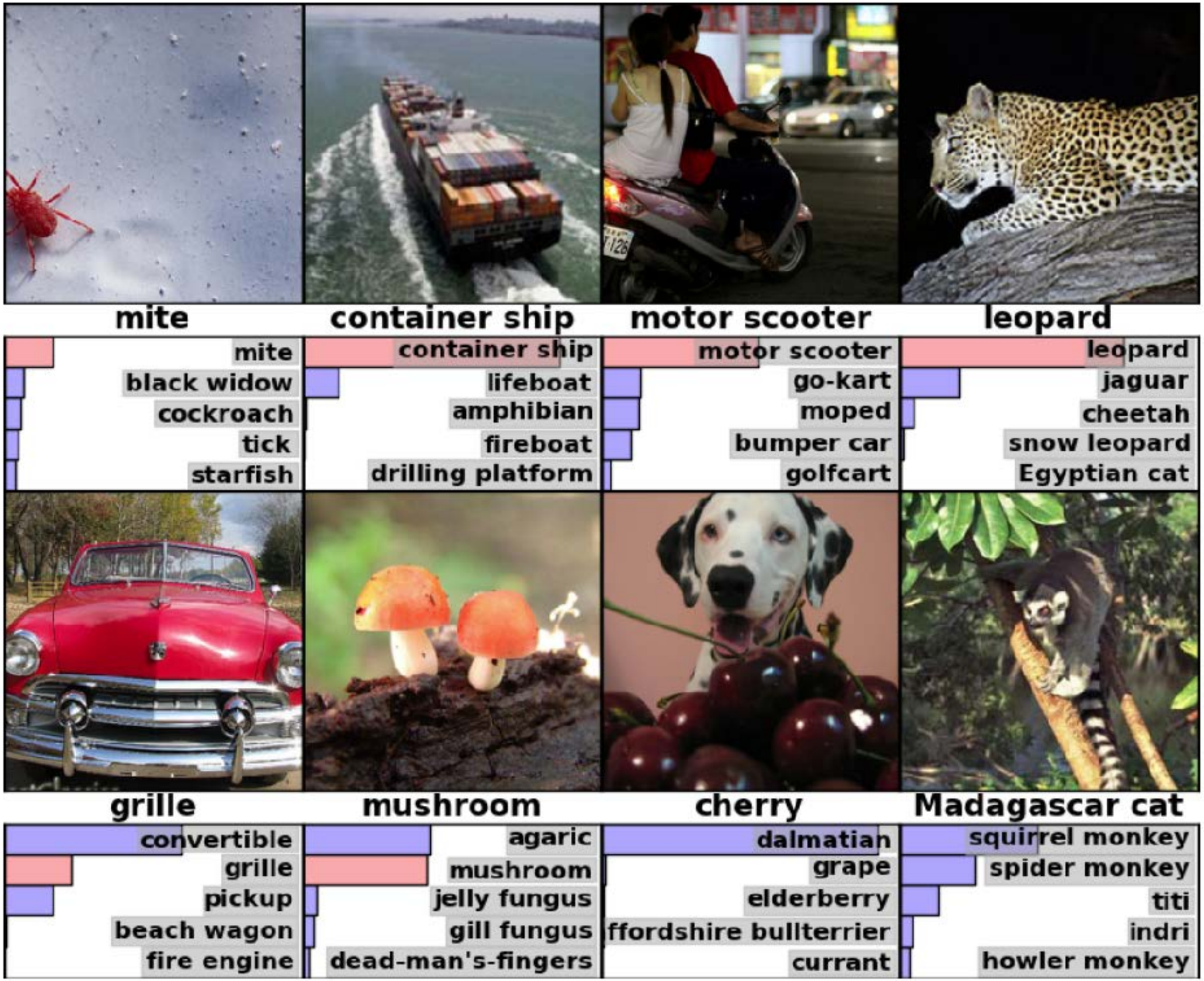


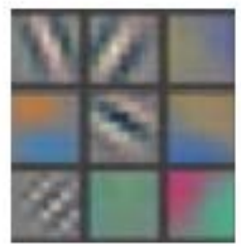


AlexNet 2012

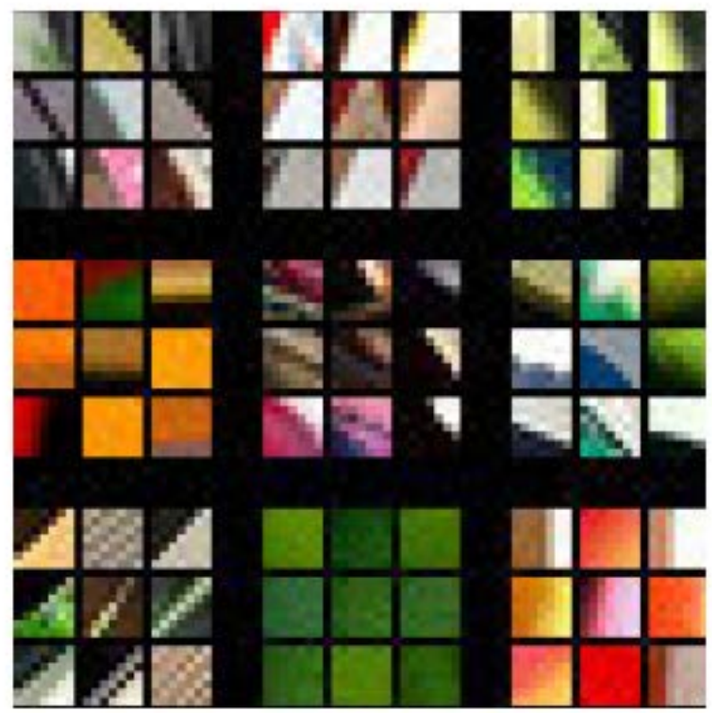


[Krizhevsky 2012]


mite
mite
black widow
cockroach
tick
starfish
container ship
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
motor scooter
go-kart
moped
bumper car
golfcart
leopard
leopard
jaguar
cheetah
snow leopard
Egyptian cat
grille
convertible
grille
pickup
beach wagon
fire engine
mushroom
agaric
mushroom
jelly fungus
gill fungus
dead-man's-fingers
cherry
dalmatian
grape
elderberry
ffordshire bullterrier
currant
Madagascar cat
squirrel monkey
spider monkey
titi
indri
howler monkey

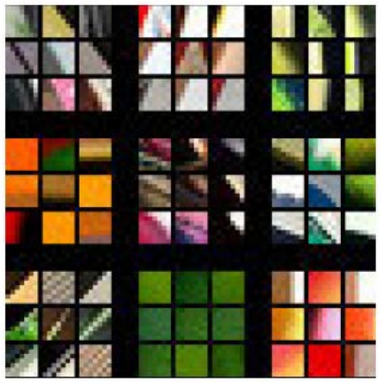


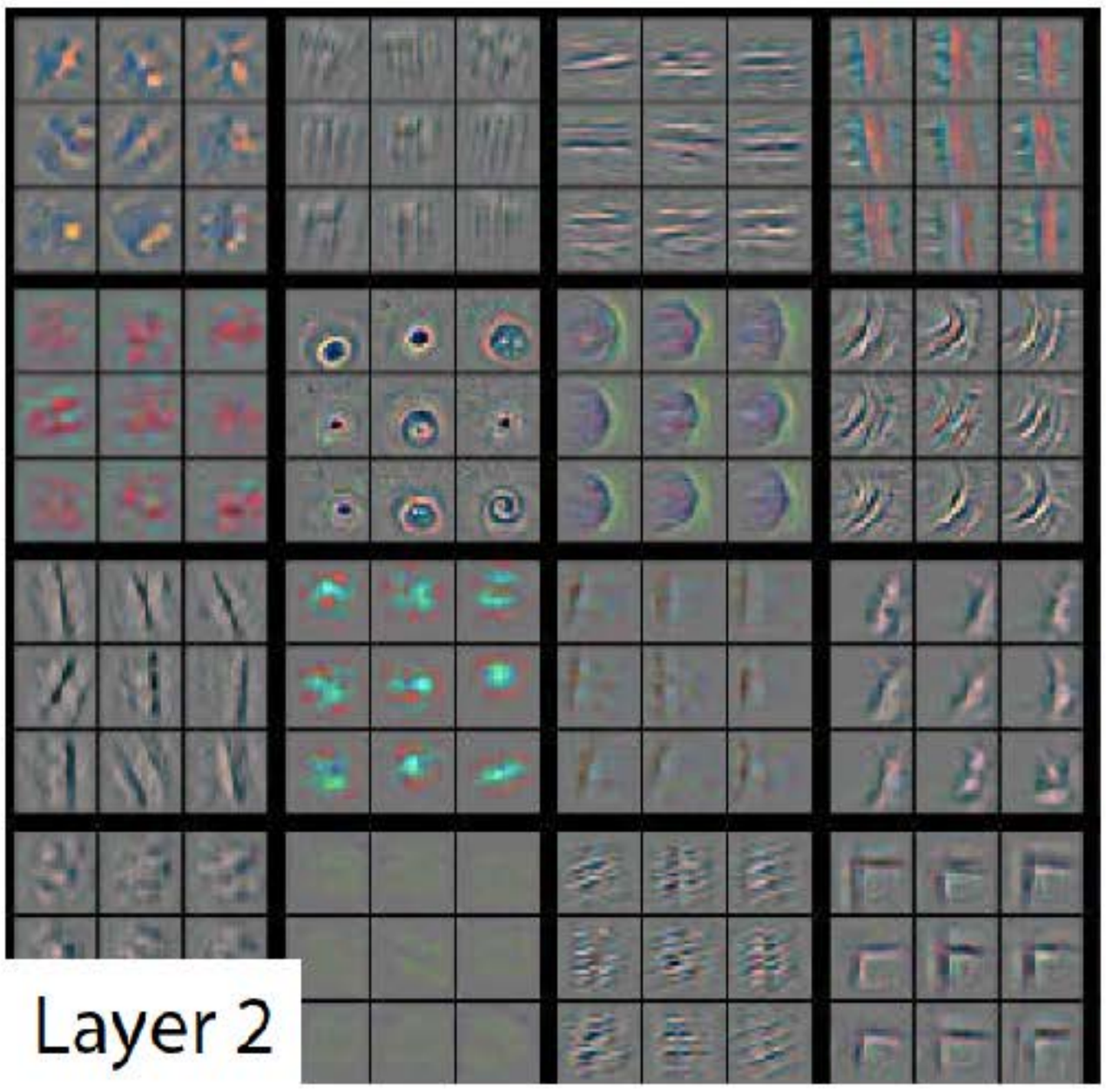
Layer 1

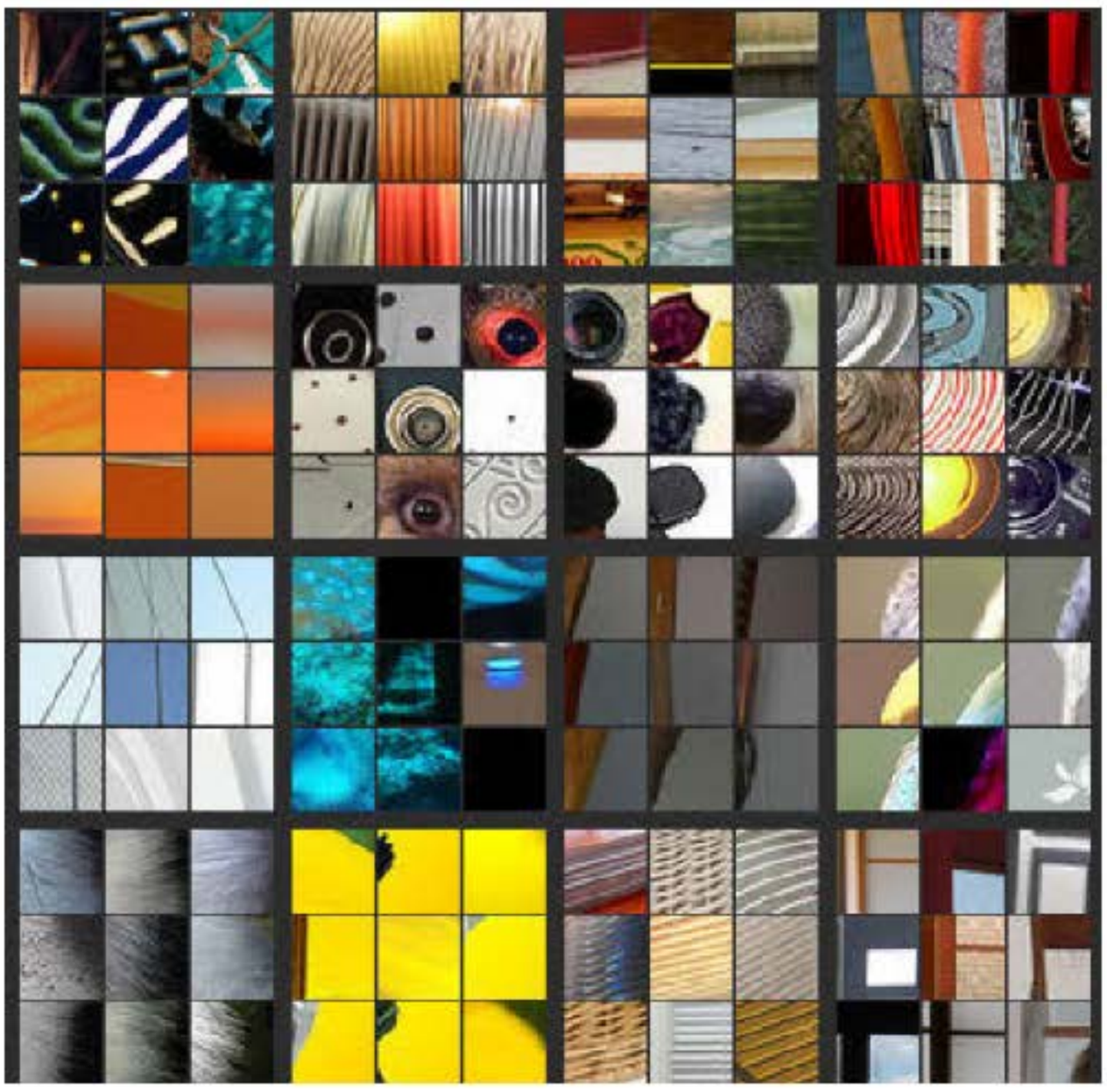


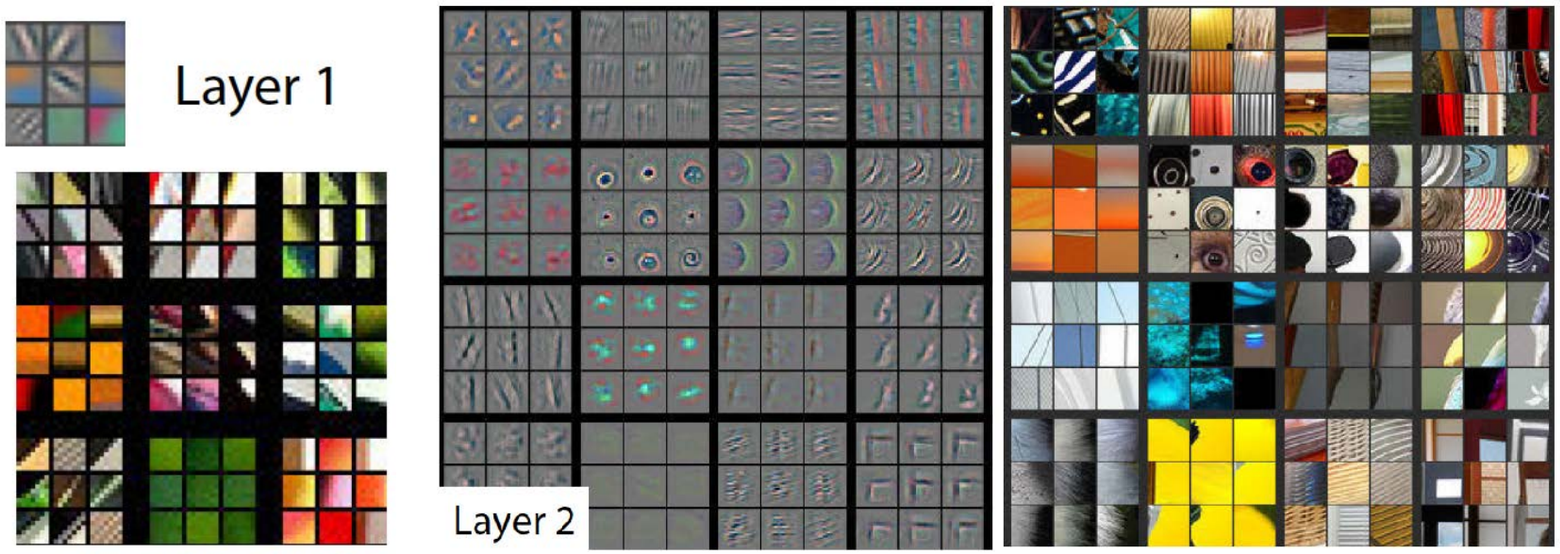
[Zeiler & Fergus 2014]

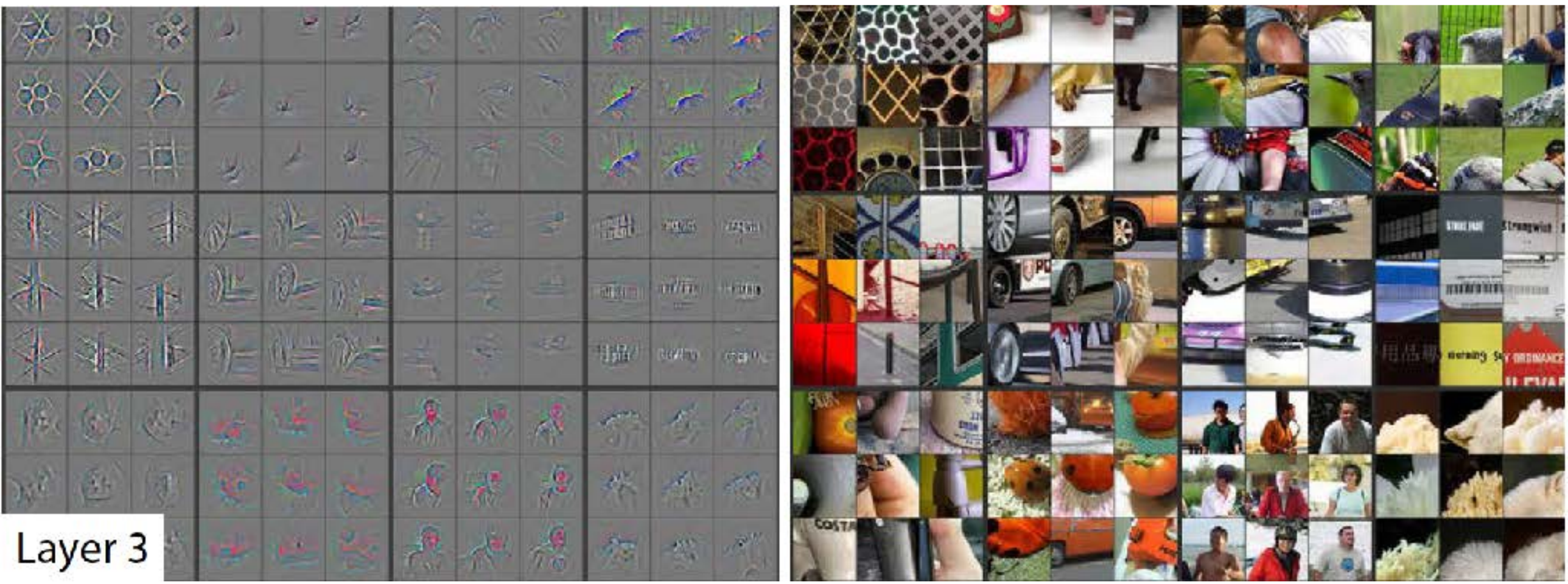
Layer 1

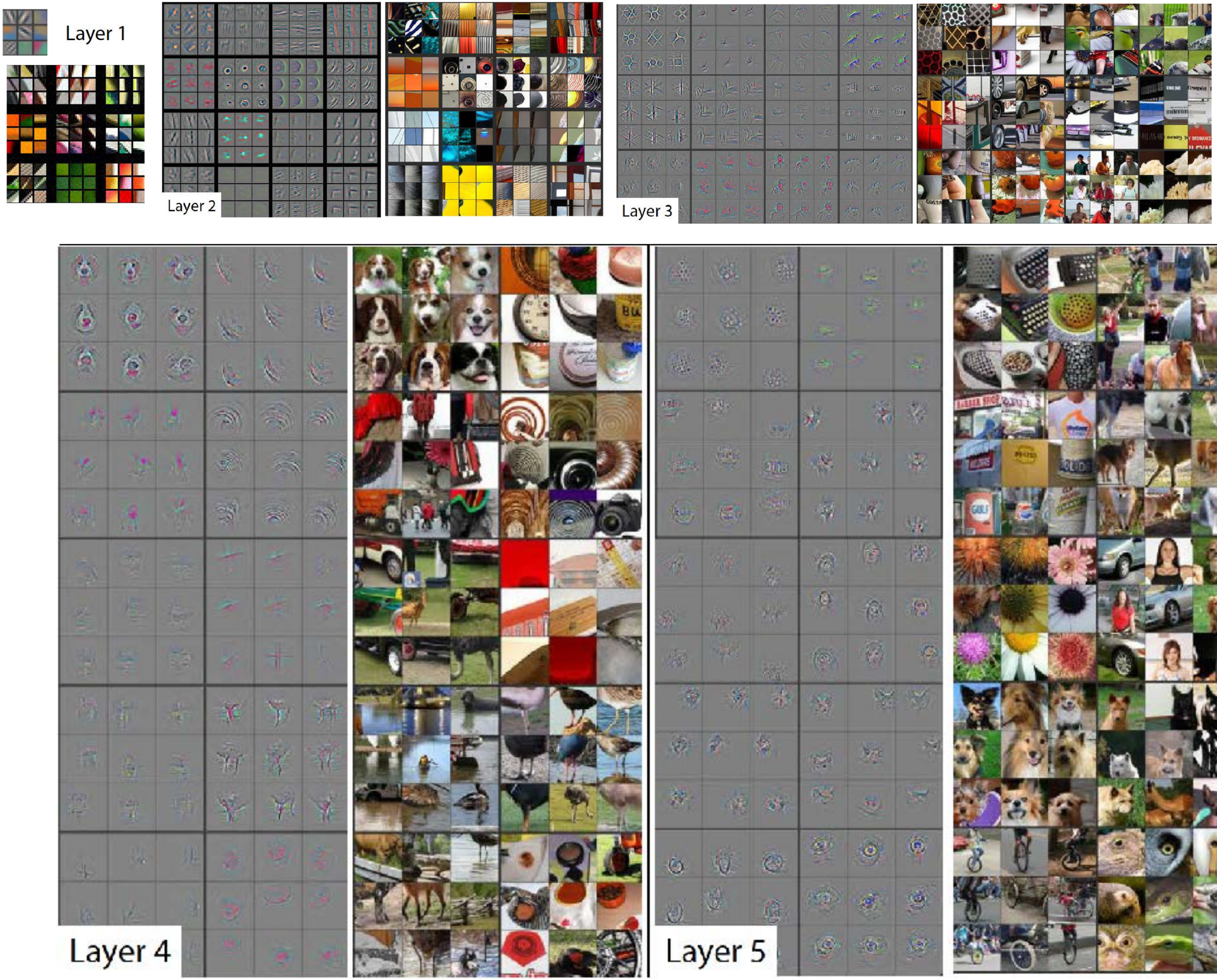




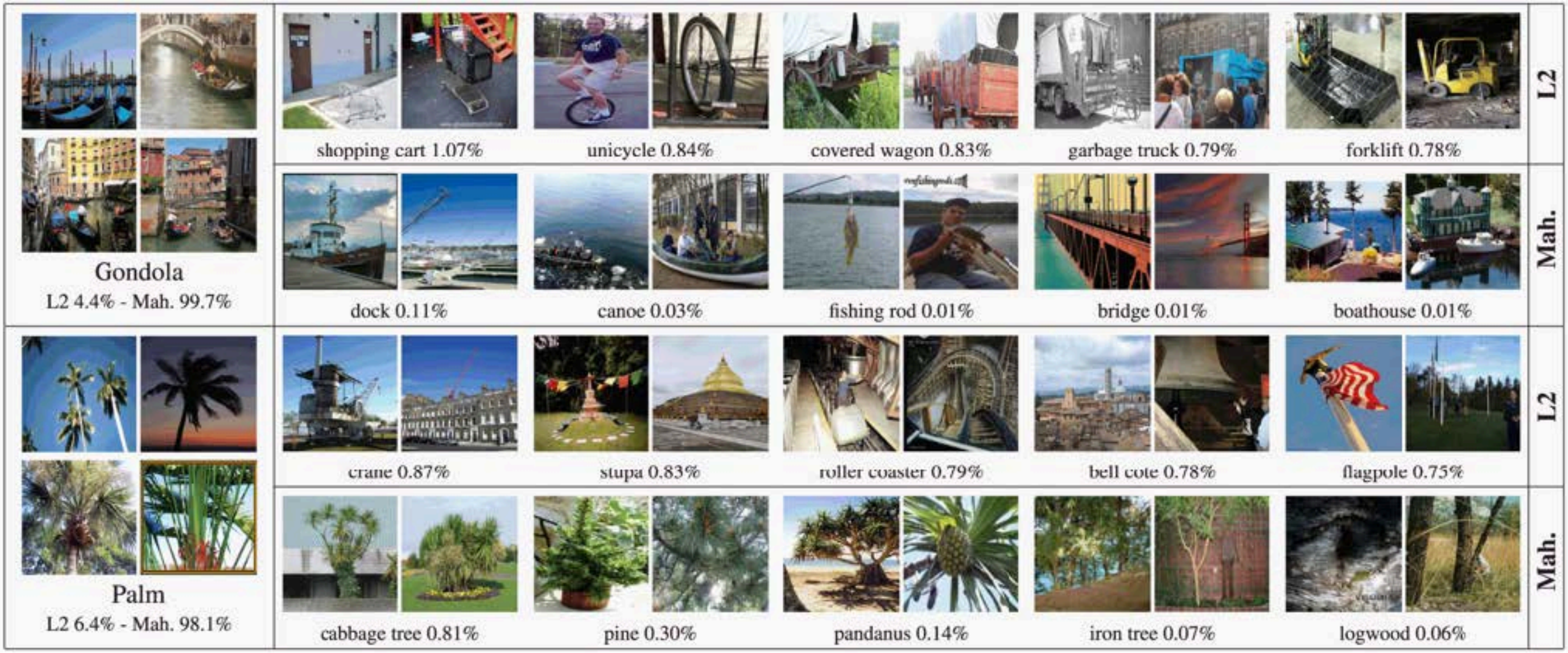



Layer 1
Layer 2


Layer 3


Layer 1
Layer 2
Layer 3
Layer 4
Layer 5

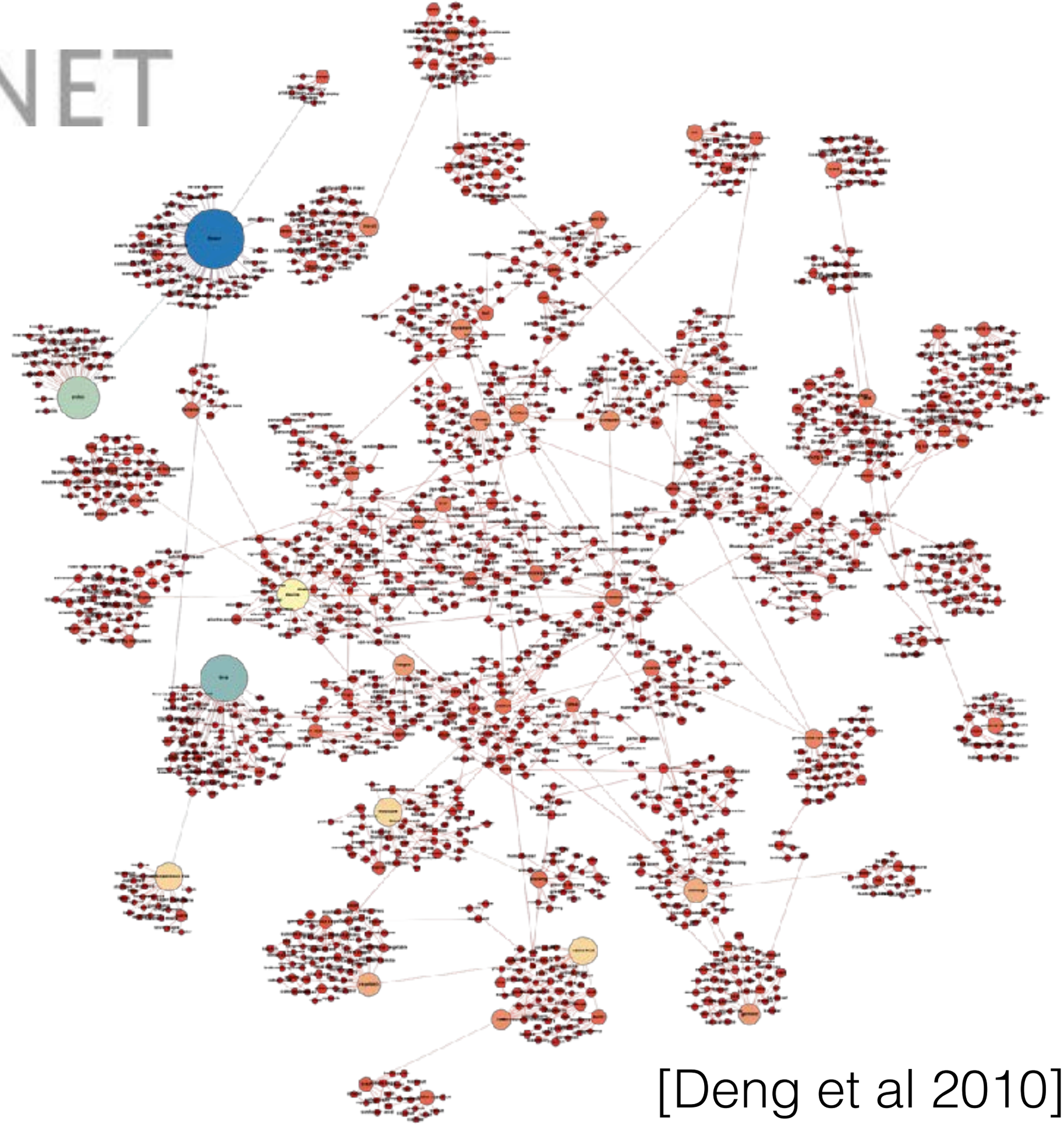
Catalyst: ImageNet 1000



~1K images for 1K categories = ~1M images

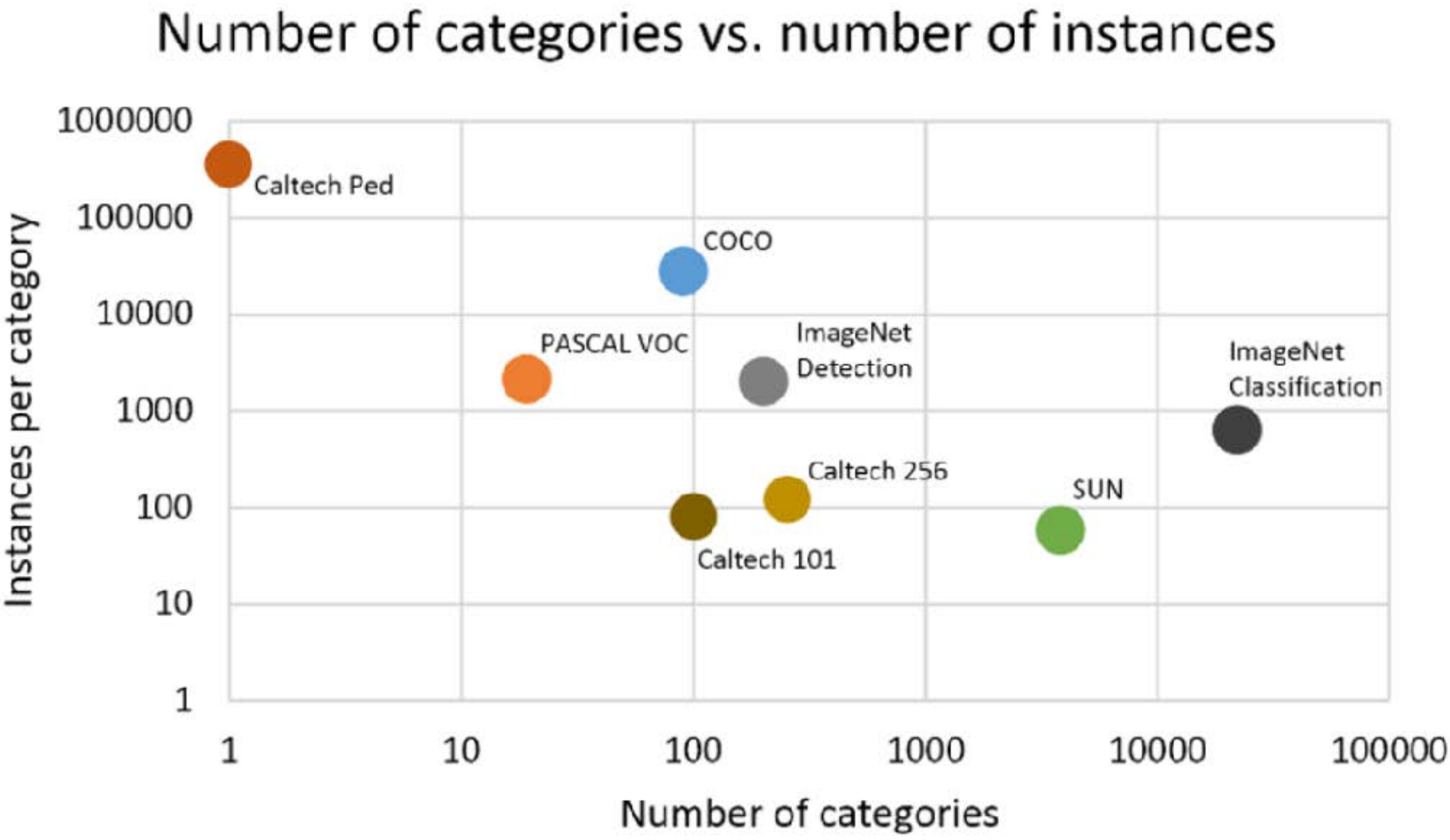
[Deng et al 2010]

IMAGENET

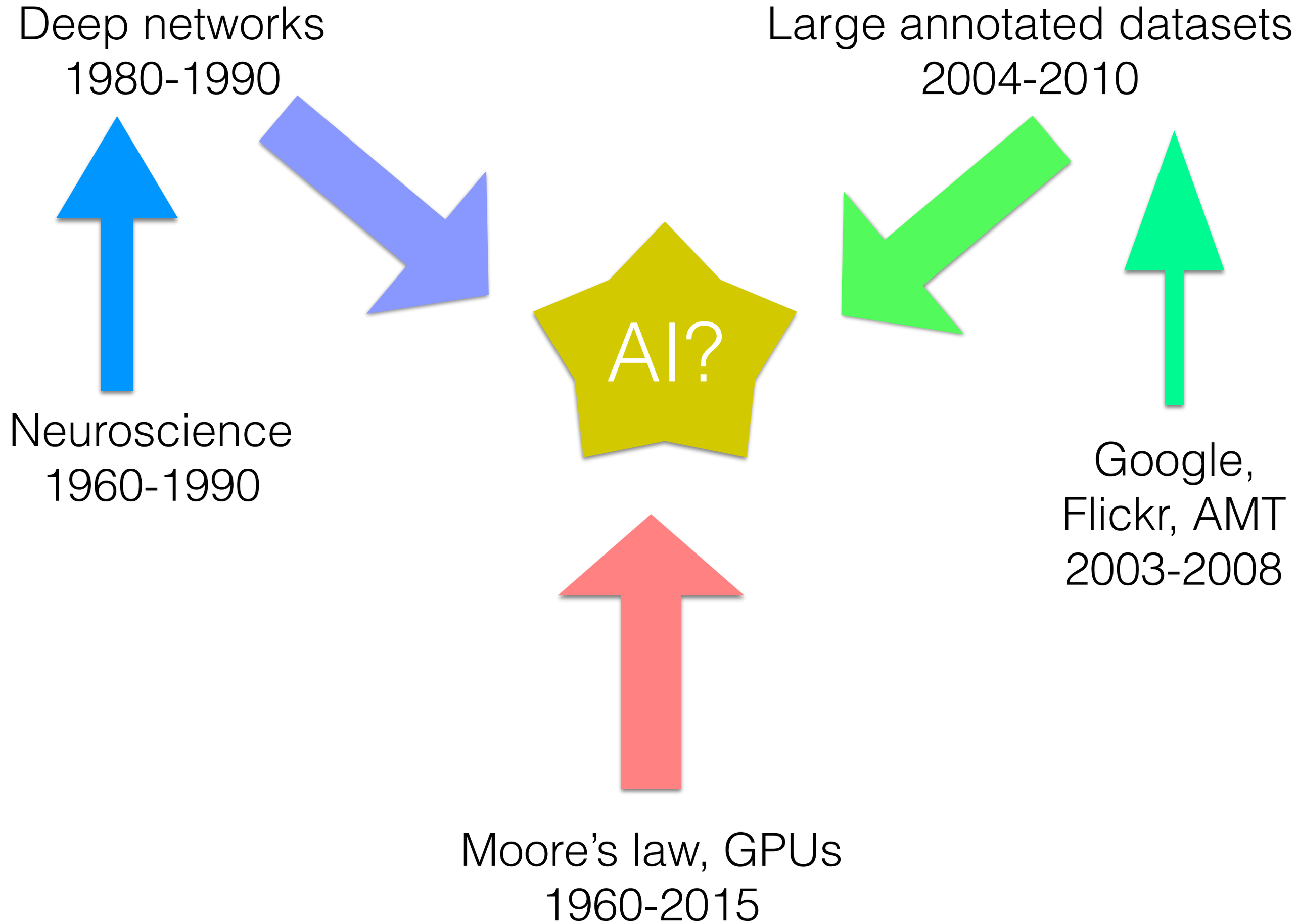


[Deng et al 2010]

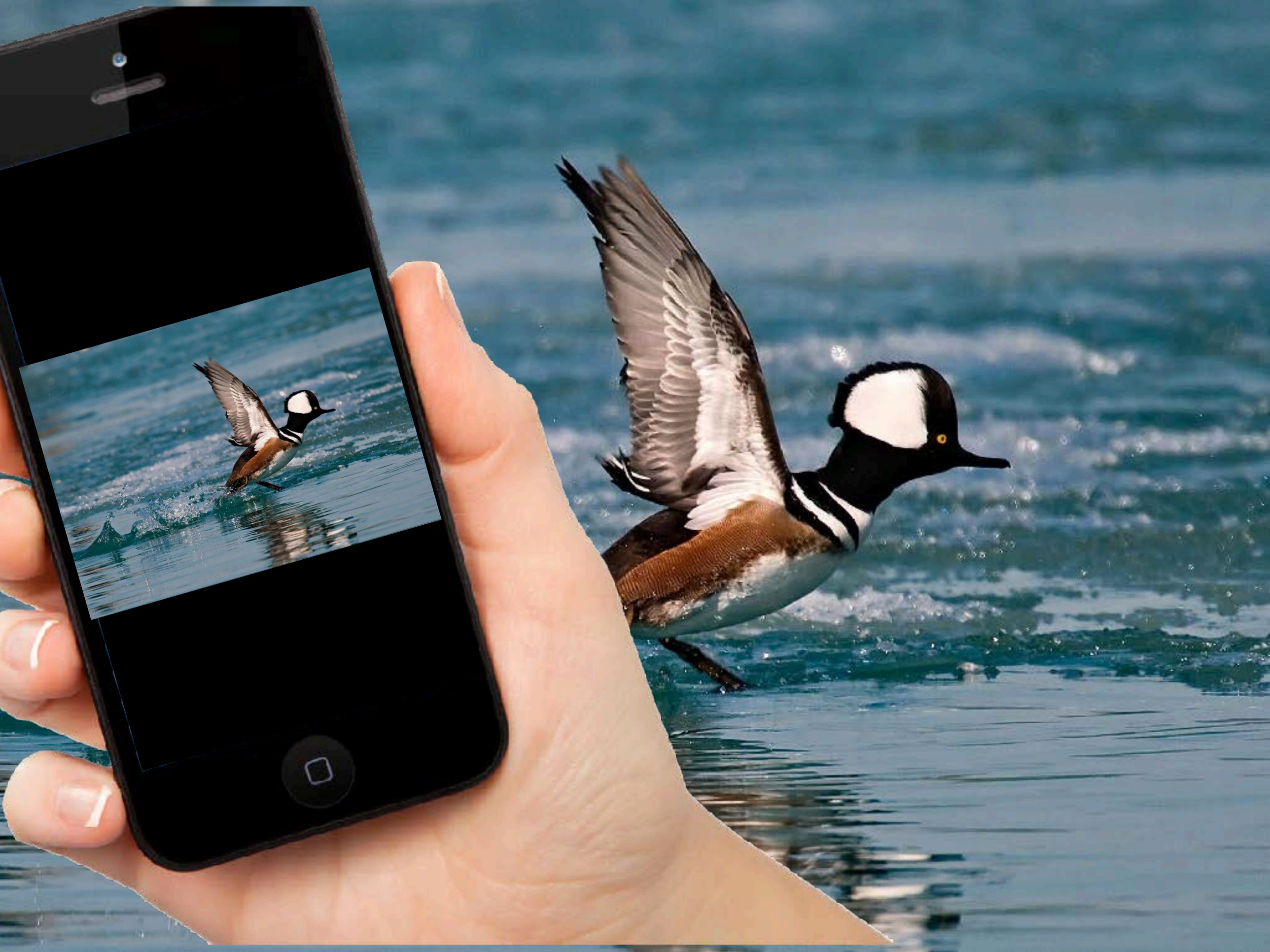
Large annotated datasets



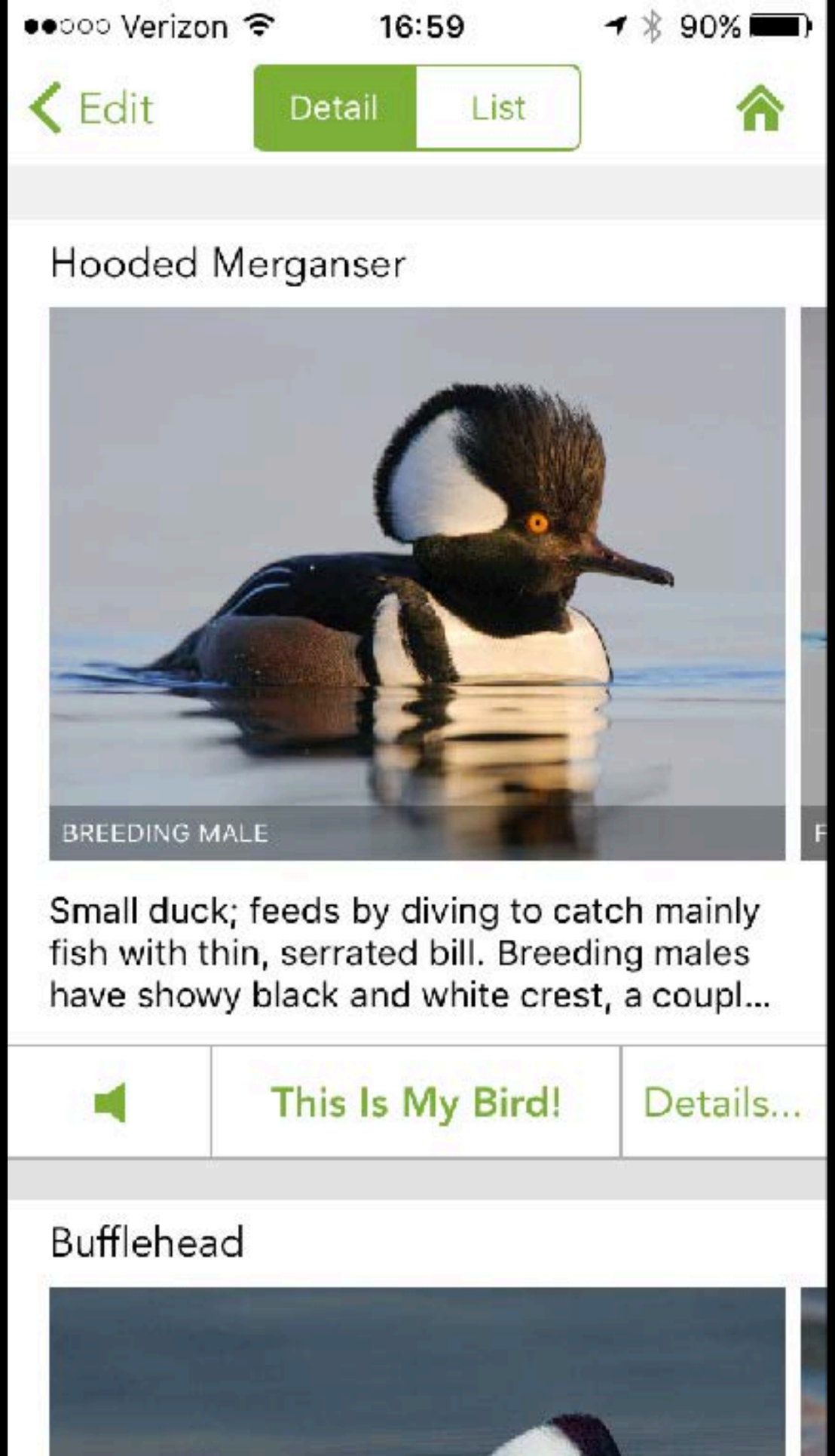
[Lin et al. 2015]


Deep networks
1980-1990
Large annotated datasets
2004-2010
AI?
Neuroscience
1960-1990
Google,
Flickr, AMT
2003-2008
Moore's law, GPUs
1960-2015

success stories

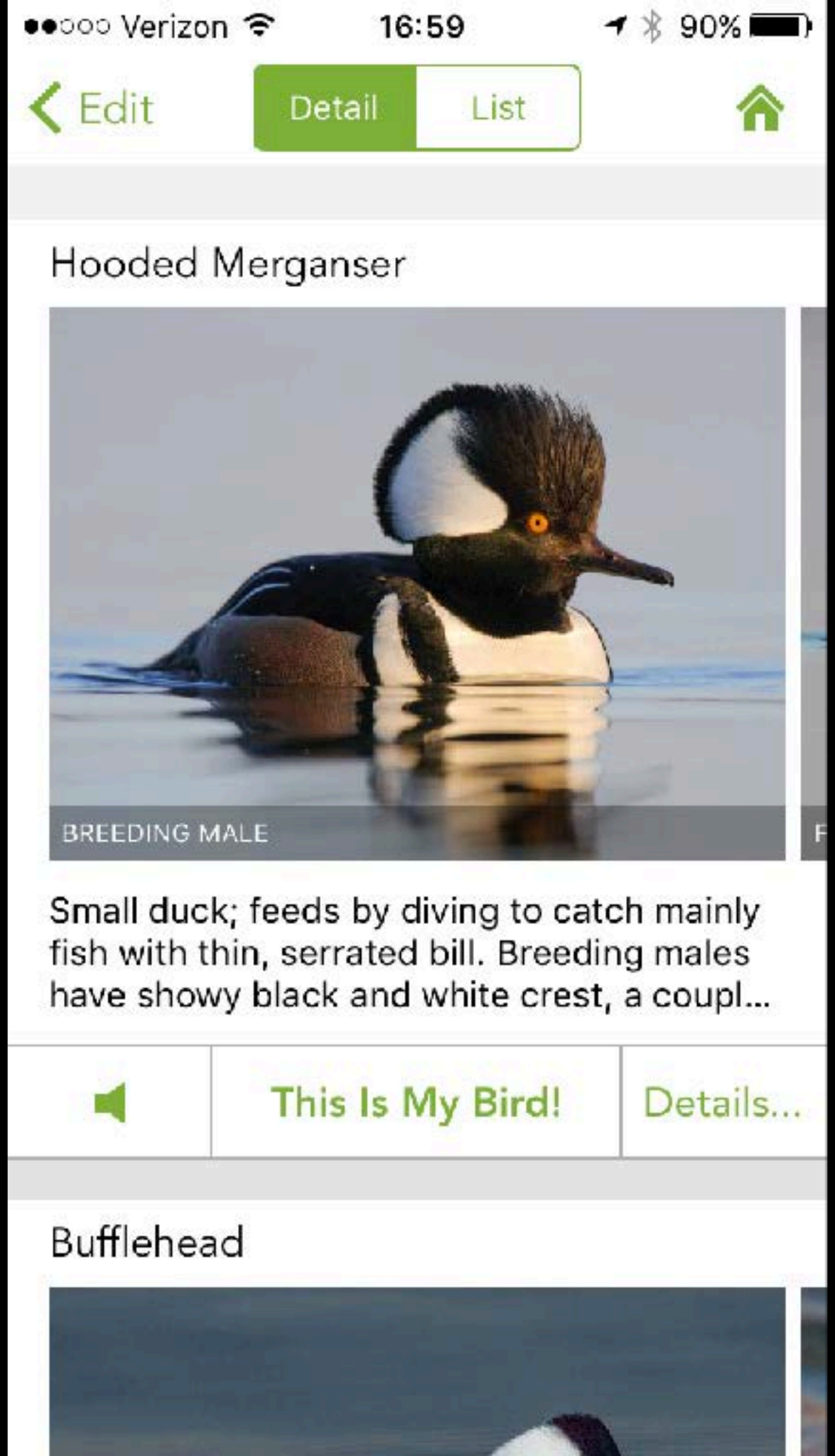


550 species of N. American birds



[van Horn et al 2014,2016]

550 species of N. American birds



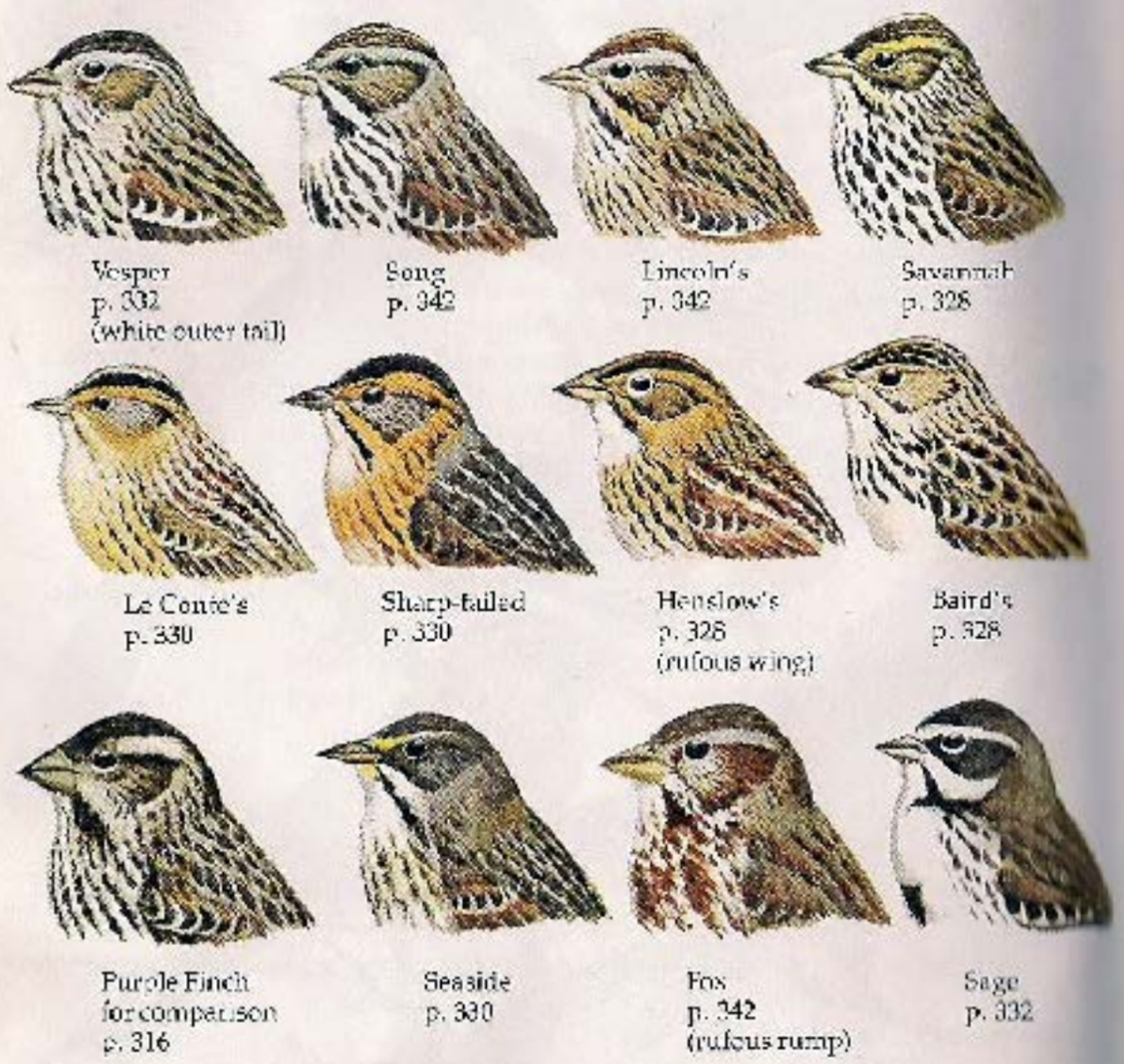
App store: ``Merlin Bird ID''

[van Horn et al 2014,2016]

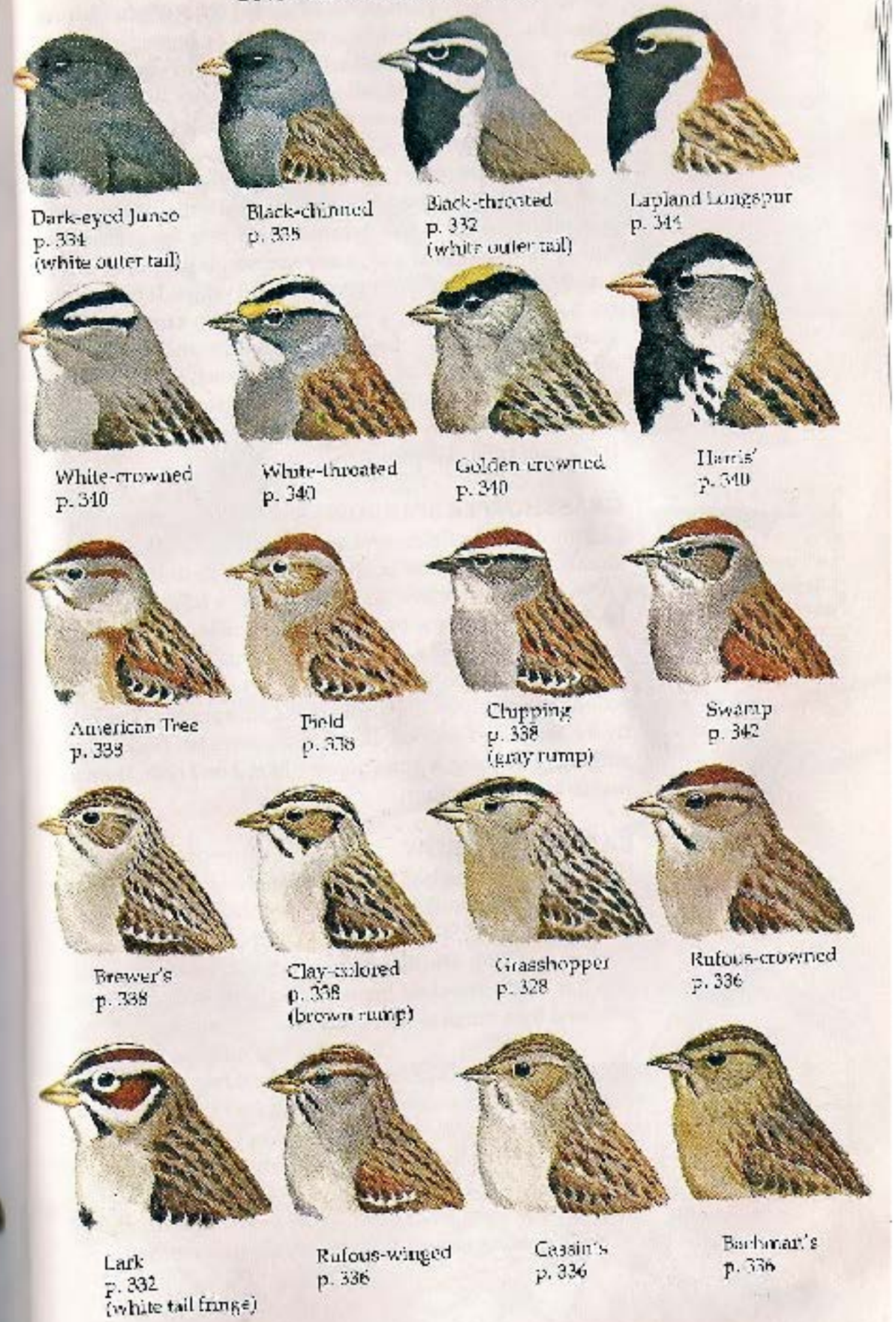
Sparrows

- **SPARROWS** are small brown-bodied birds with streaked backs and short conical beaks. Their food, mostly seeds except during the nesting season, is obtained on or near the ground. When not nesting, most are see[illegible] these may be [illegible] identifica- tion [illegible] es females are [illegible] pecies are sho[illegible] presented by one head each; the striking Lark Bunting, the Snow Bunting, the towhees, and the Olive Sparrow are omitted. Immatures of some species are much duller, especially those species with black or rufous on the head. Songs and chips of sparrows are often more easily distinguished than are their plumages. See pp. 328-345 for further details.

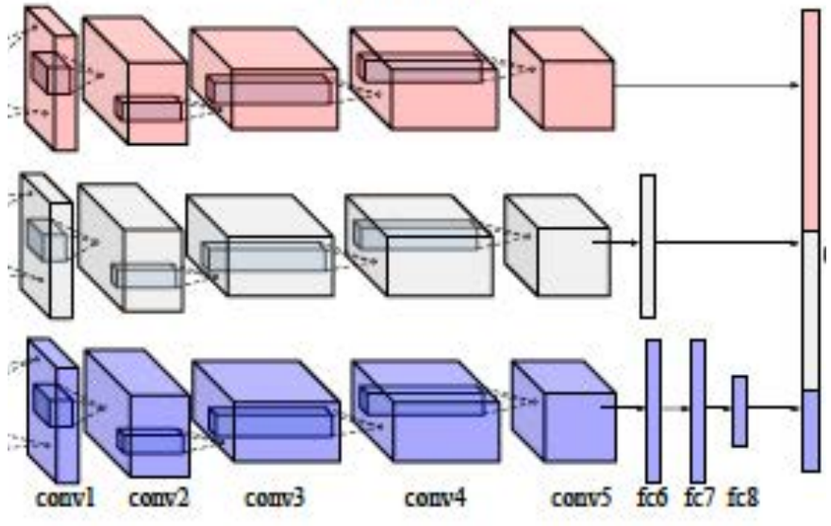
STREAKED BREASTS

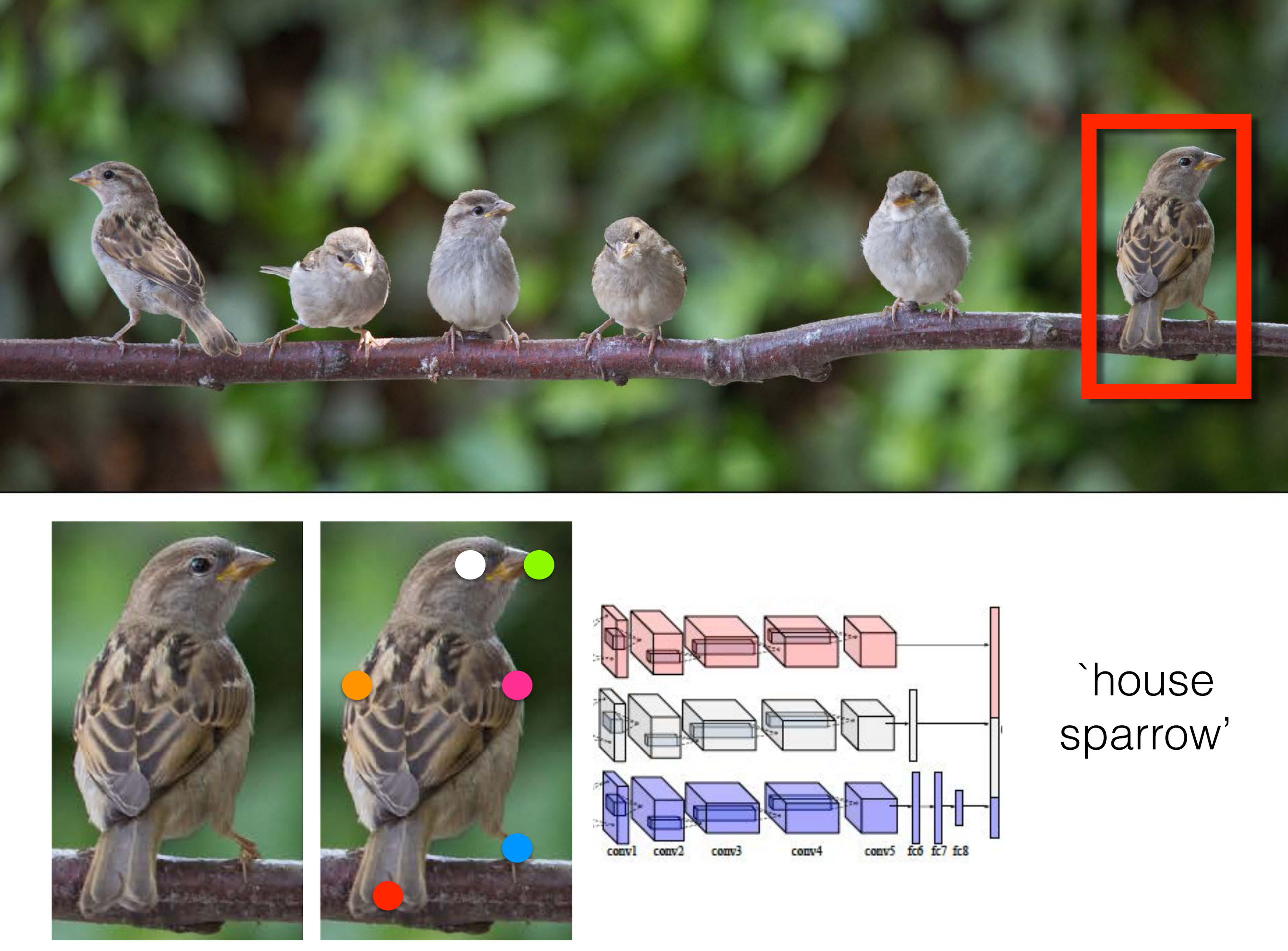

Vesper p. 332 (white outer tail)
Song p. 342
Lincoln's p. 342
Savannah p. 328
Le Conte's p. 330
Sharp-tailed p. 330
Henslow's p. 328 (rufous wing)
Baird's p. 328
Purple Finch for comparison p. 316
Seaside p. 330
Fox p. 342 (rufous rump)
Sage p. 332

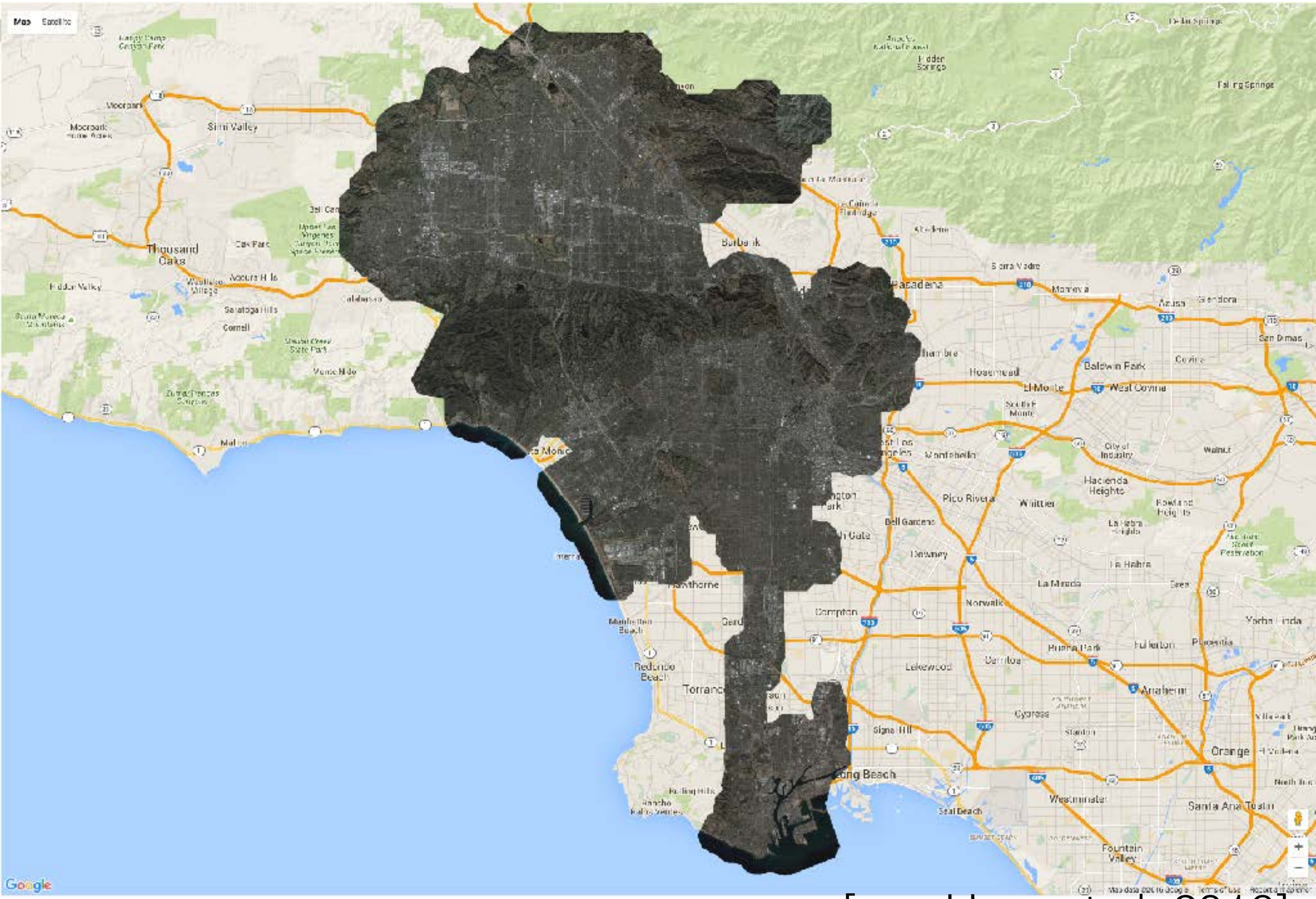
UNSTREAKED BREASTS


Dark-eyed Junco p. 334 (white outer tail)
Black-chinned p. 335
Black-throated p. 332 (white outer tail)
Lapland Longspur p. 344
White-crowned p. 340
White-throated p. 340
Golden crowned p. 340
Harris' p. 340
American Tree p. 338
Field p. 338
Chipping p. 338 (gray rump)
Swamp p. 342
Brewer's p. 338
Clay-colored p. 338 (brown rump)
Grasshopper p. 328
Rufous-crowned p. 336
Lark p. 332 (white tail fringe)
Rufous-winged p. 336
Cassin's p. 336
Bachman's p. 336

Invariance to pose and background

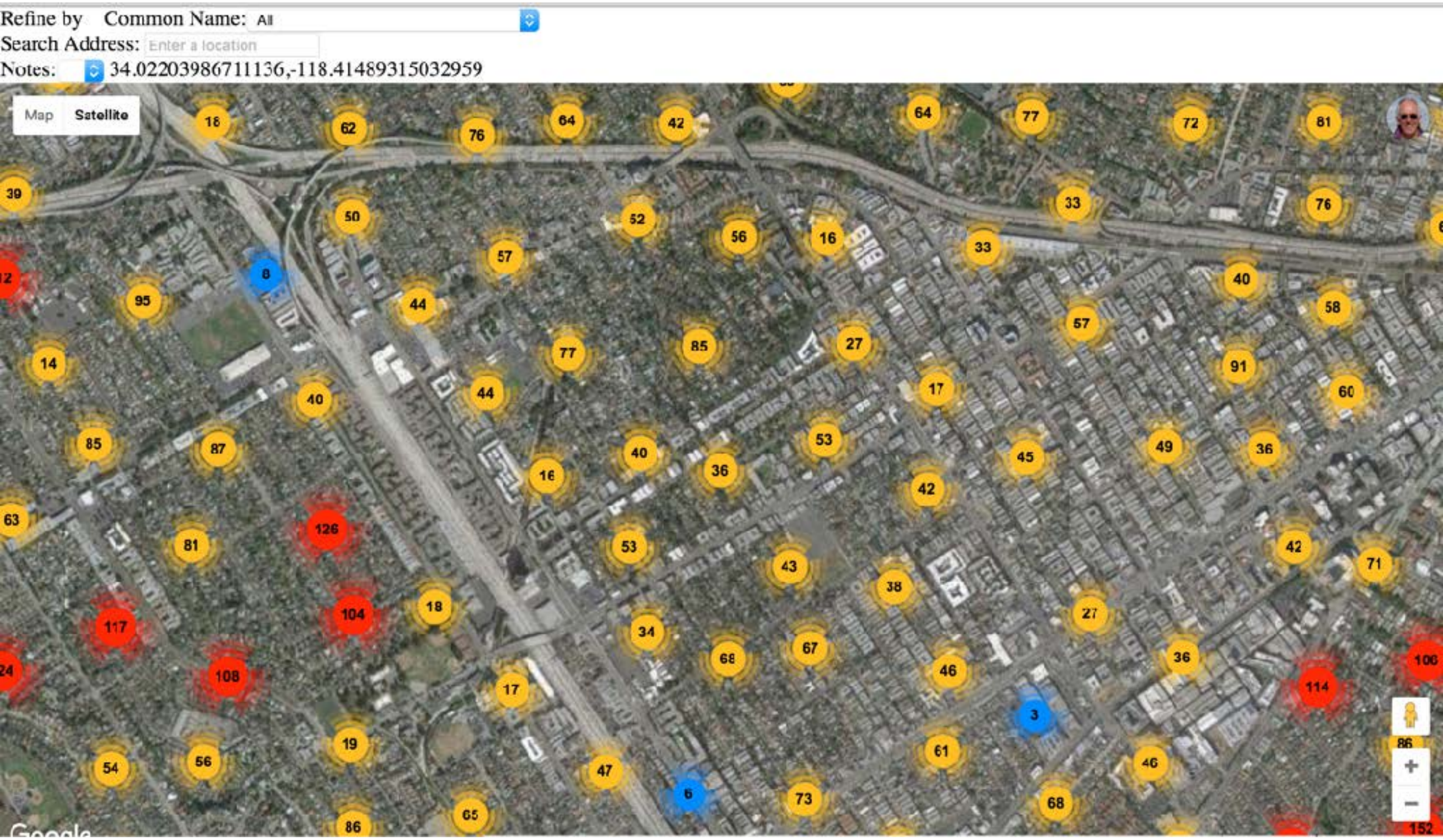

conv1
conv2
conv3
conv4
conv5
fc6
fc7
fc8


conv1
conv2
conv3
conv4
conv5
fc6
fc7
fc8
`house sparrow’



[van Horn et al. 2016]

Los Angeles = 1M trees



ID: 548
Location: 34.03278128145396E, 118.42148597780546
Species List: Jacaranda(2.01), Chinese sweetgum(-0.384), Dracaena palm(-1.01), Magnolia(-1.018), Rough Shell Macadamia(-1.095)
Trunk Diameter: 15
Detection Score: 0.222
Street View: N8coolmRTT1J8C4o4V3cxUQ
Localization Error (m): N/A
Update Note
Registree Caltech

b 8

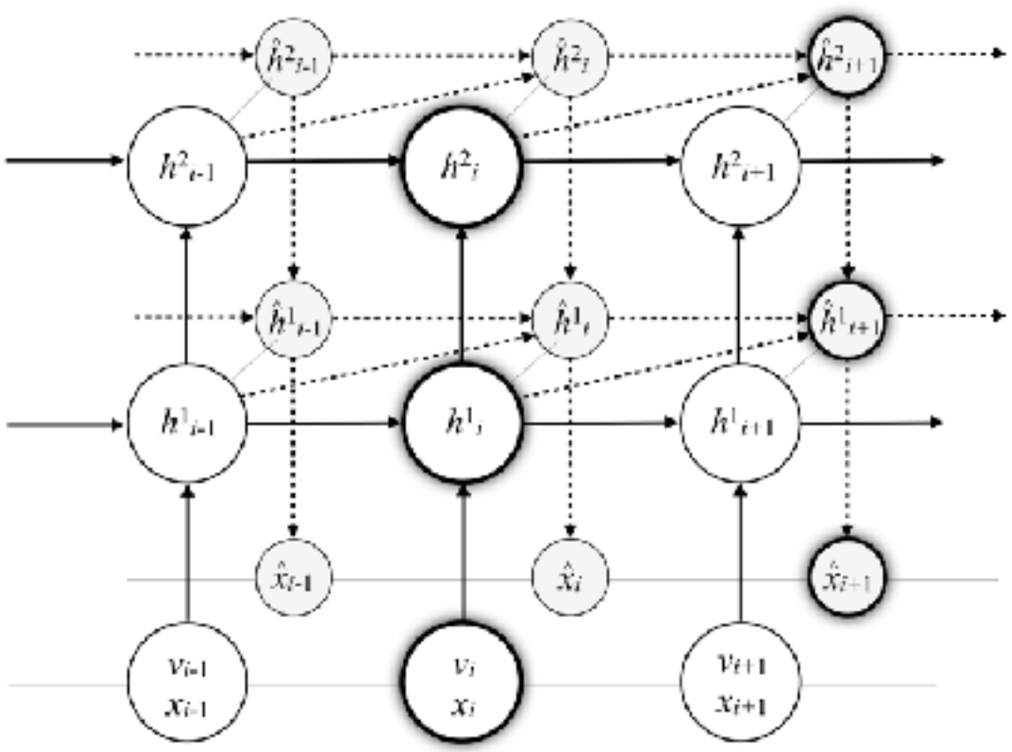
b 8

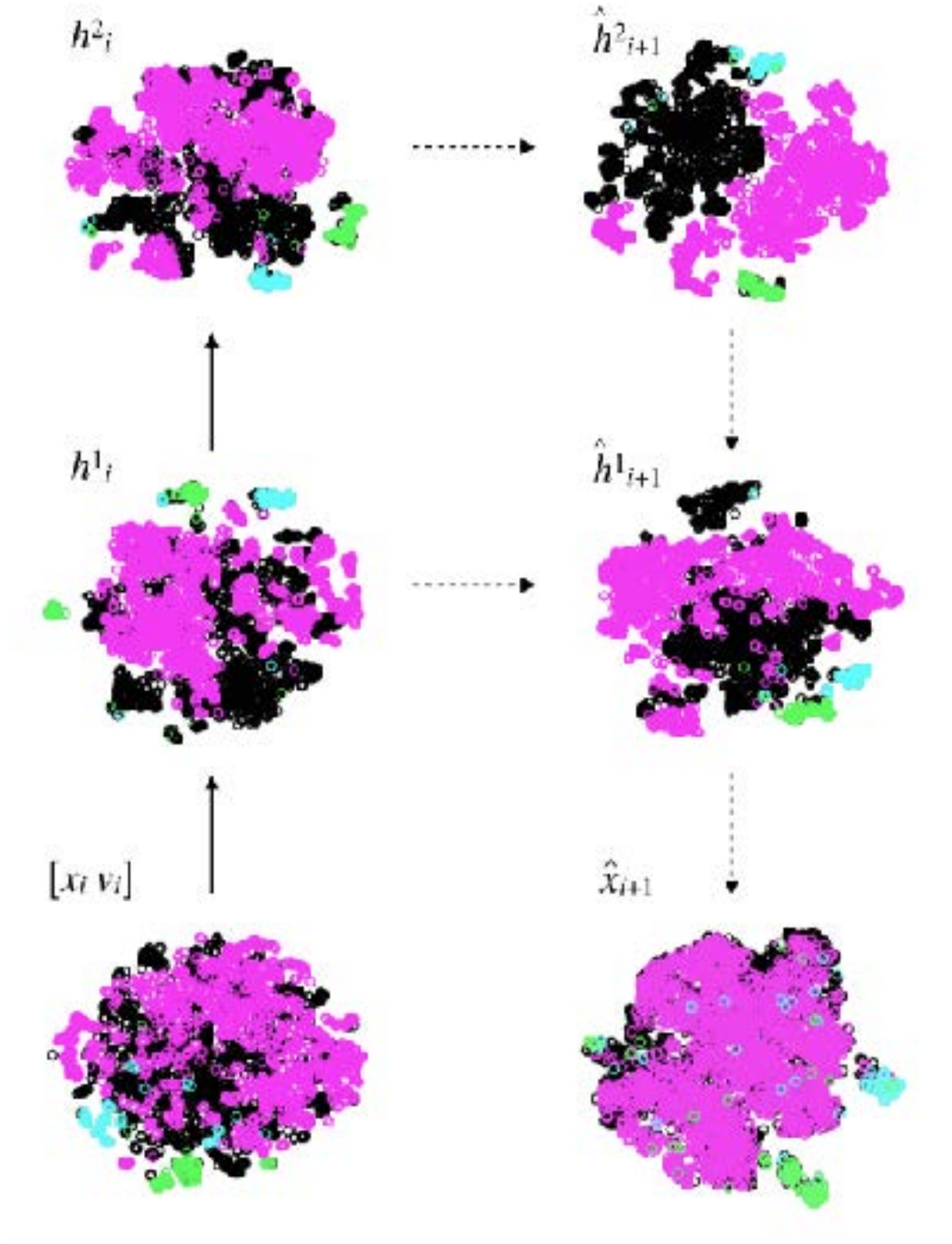
Discovery: FlyBowl

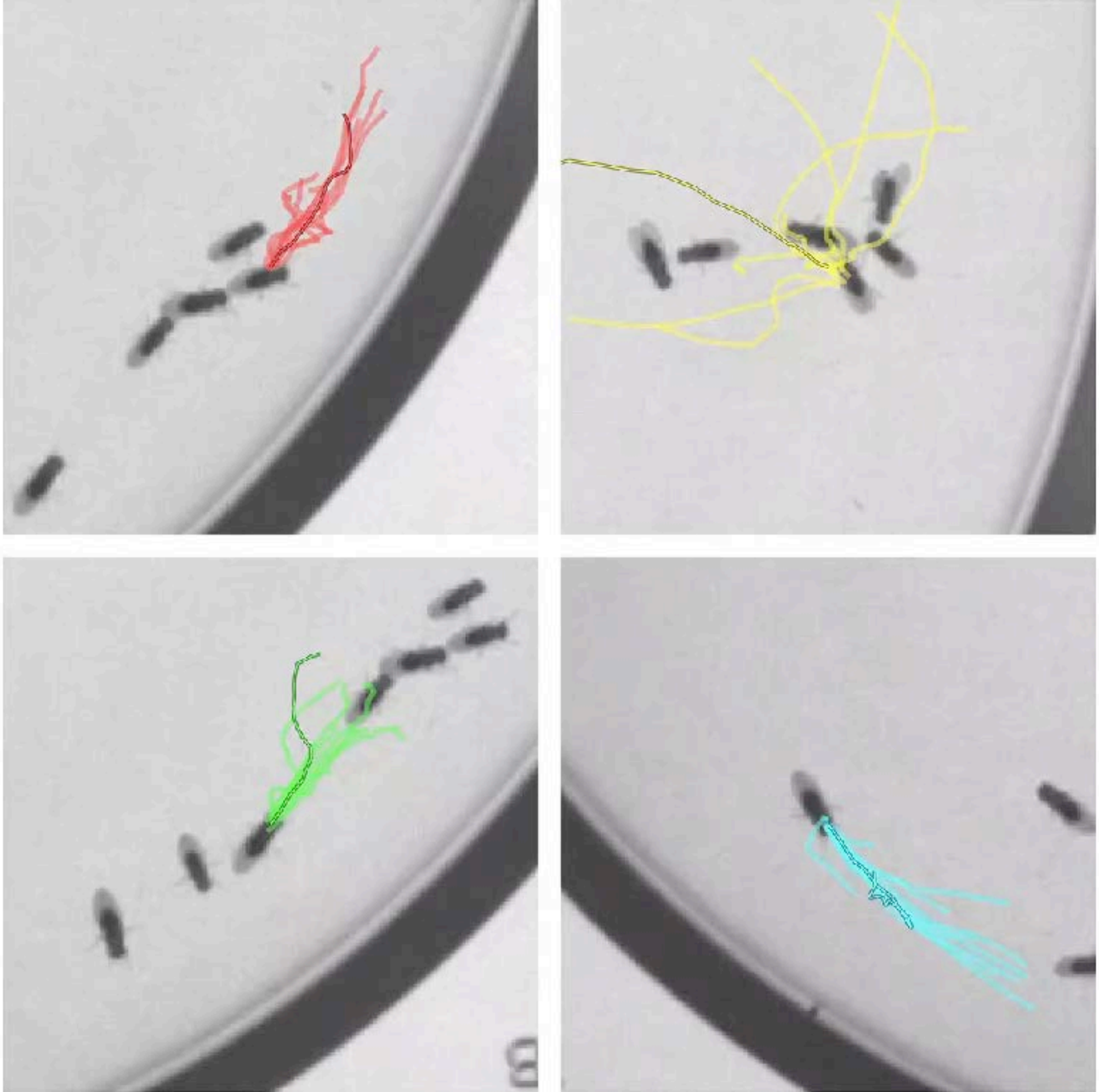
tSNE dimensionality reduction

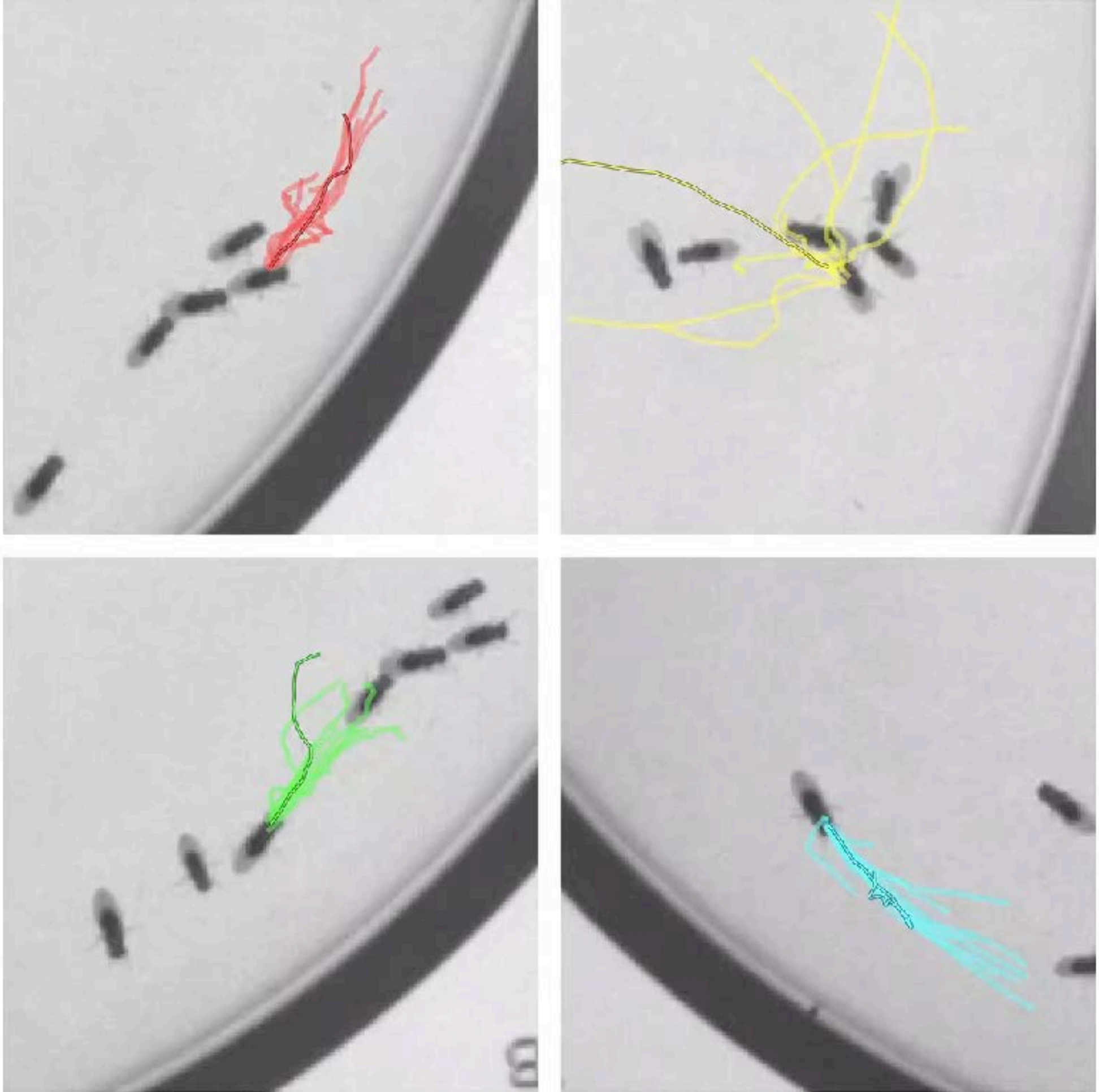
- female
- male
- right wing extension
- left wing extension

$h^l{}_i$: hidden states of unsupervised model

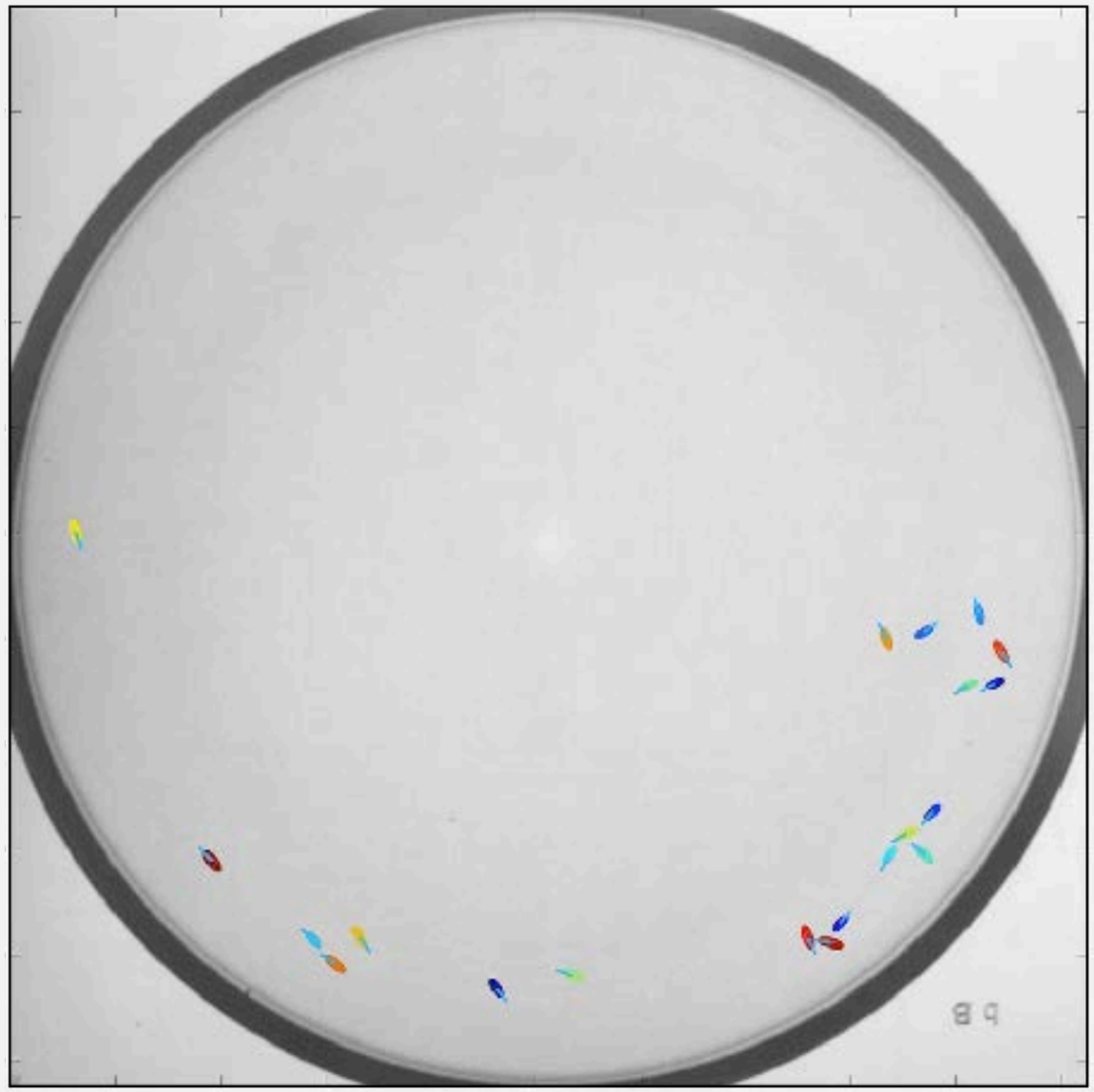


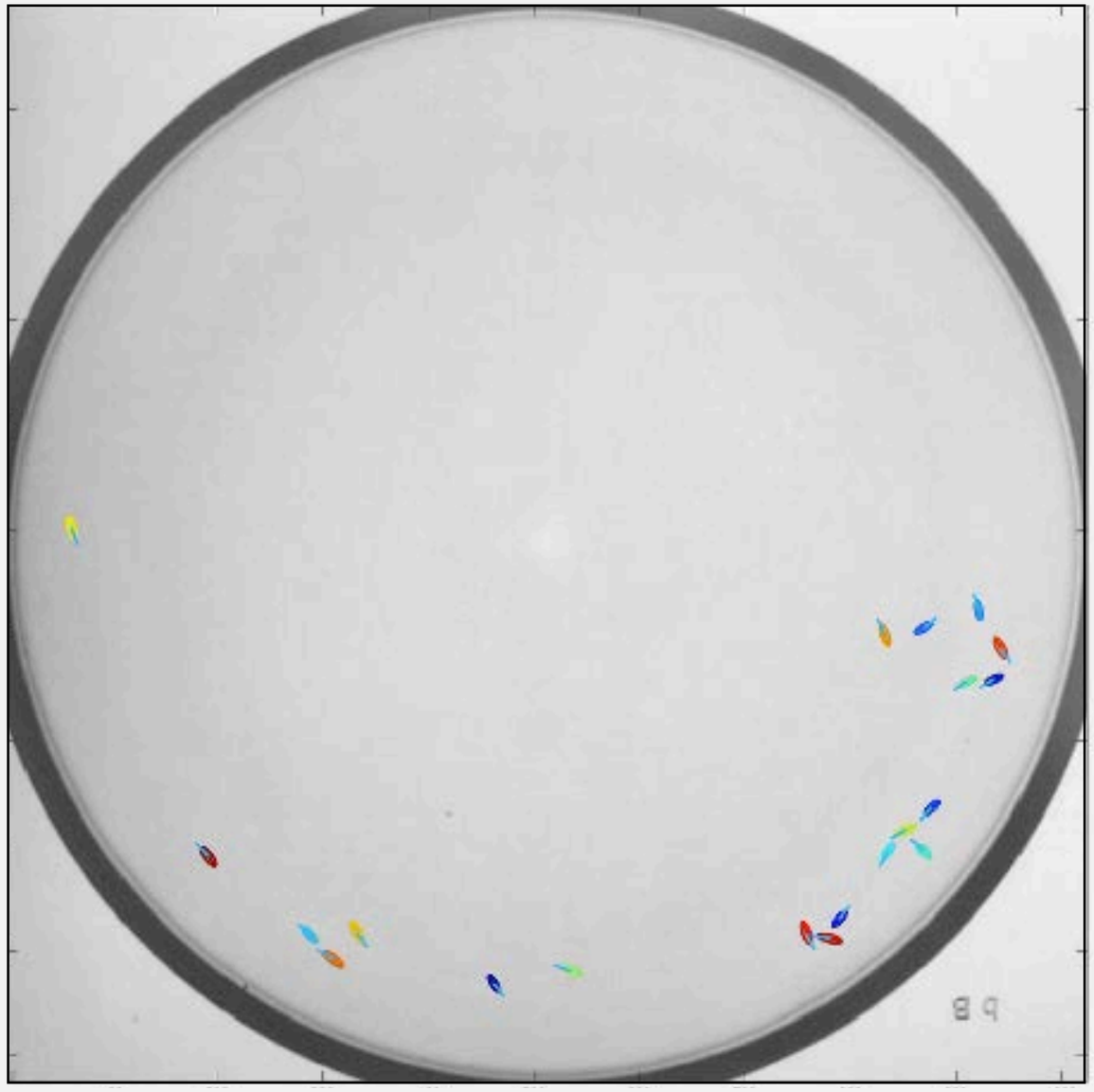




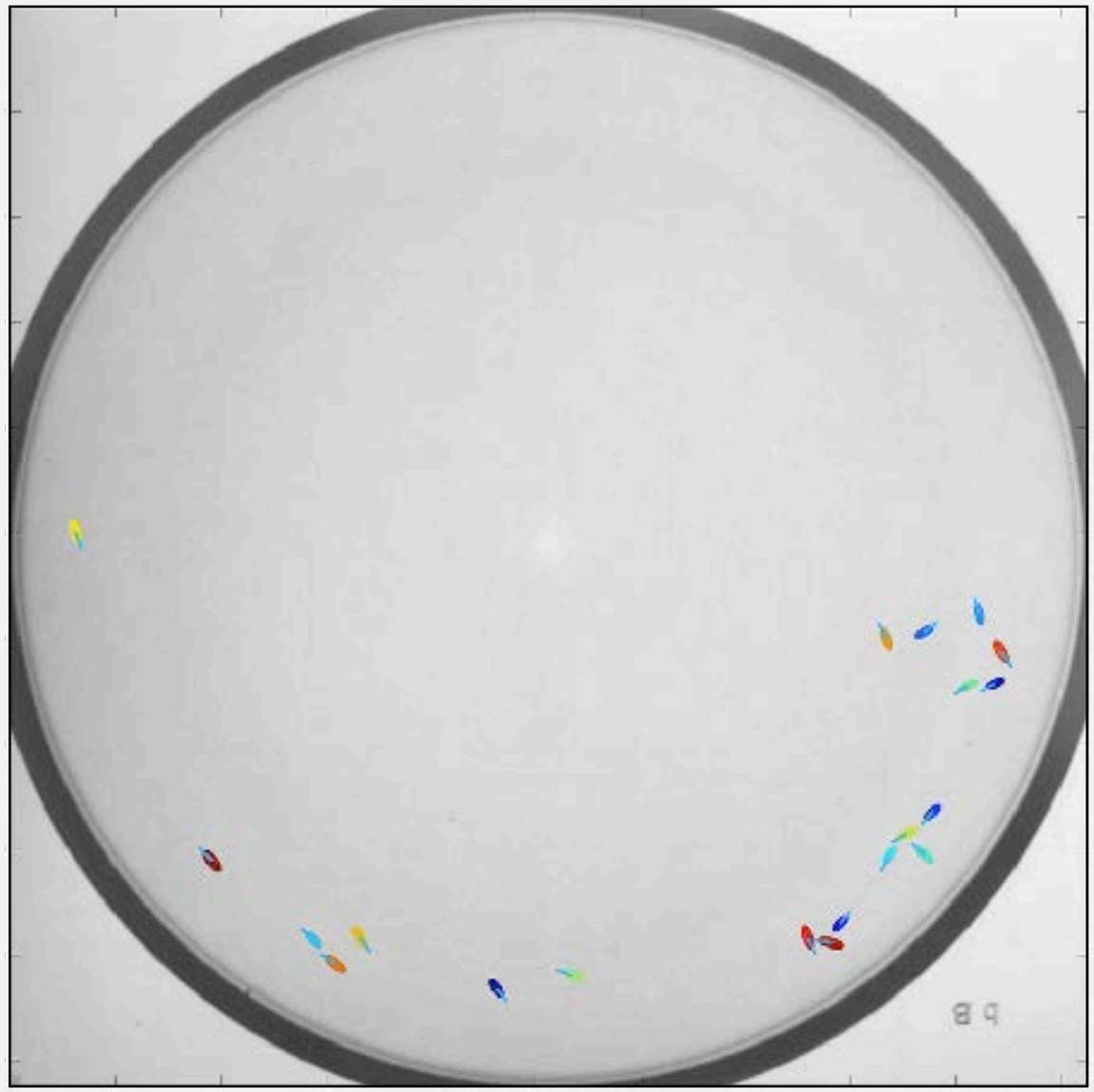


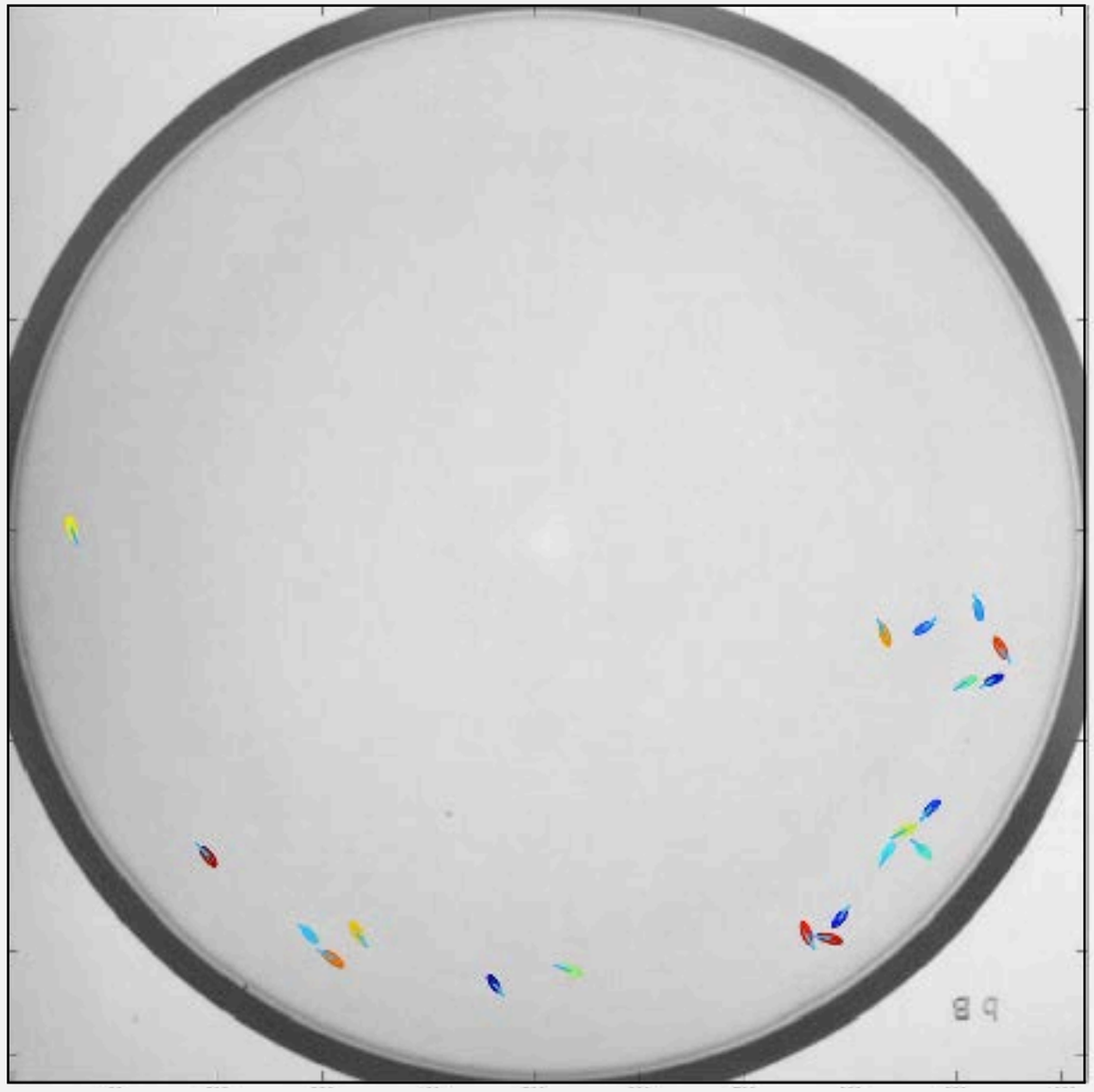
Simulation: FlyBowl



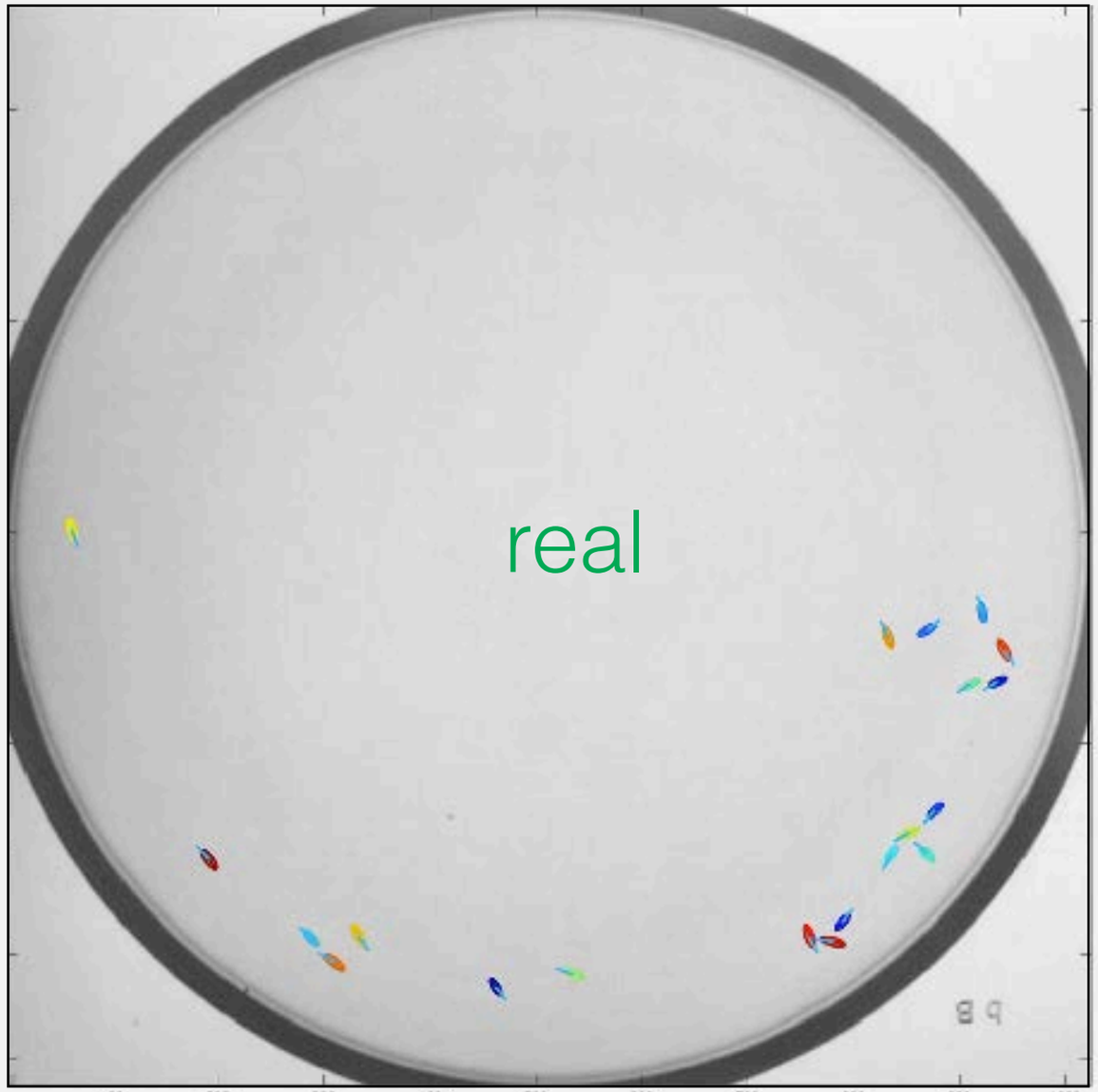


Simulation: FlyBowl



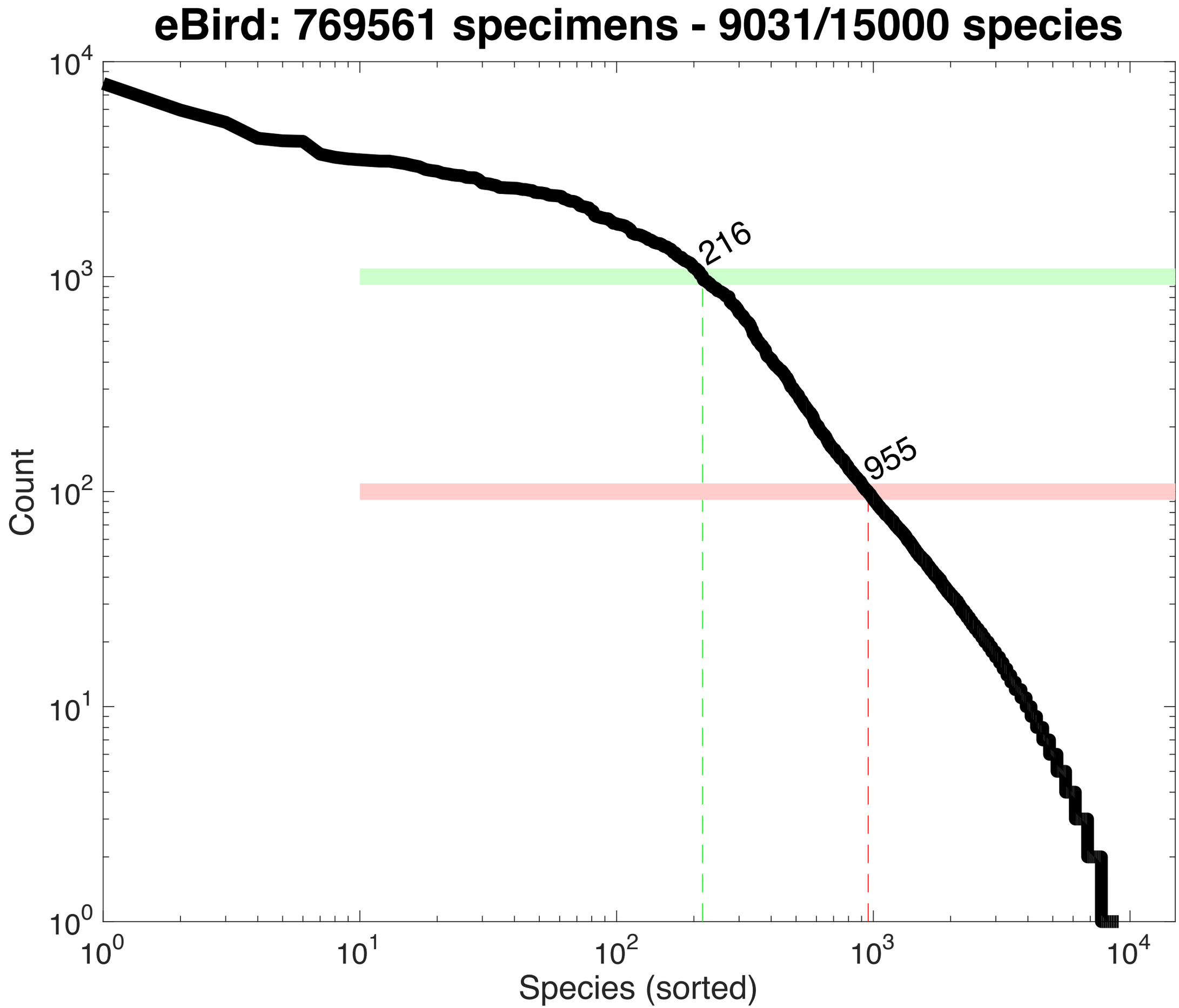


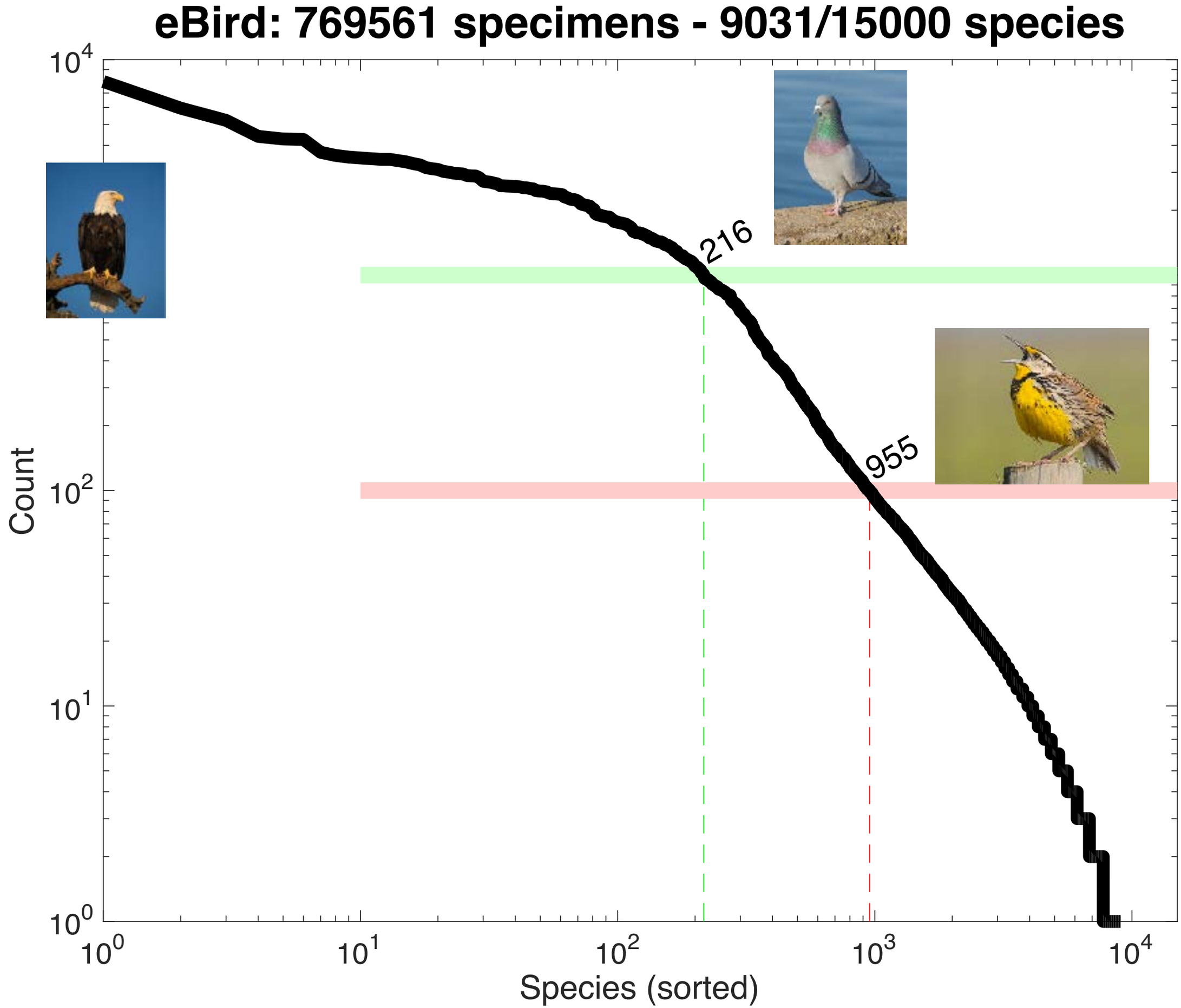
Simulation: FlyBowl

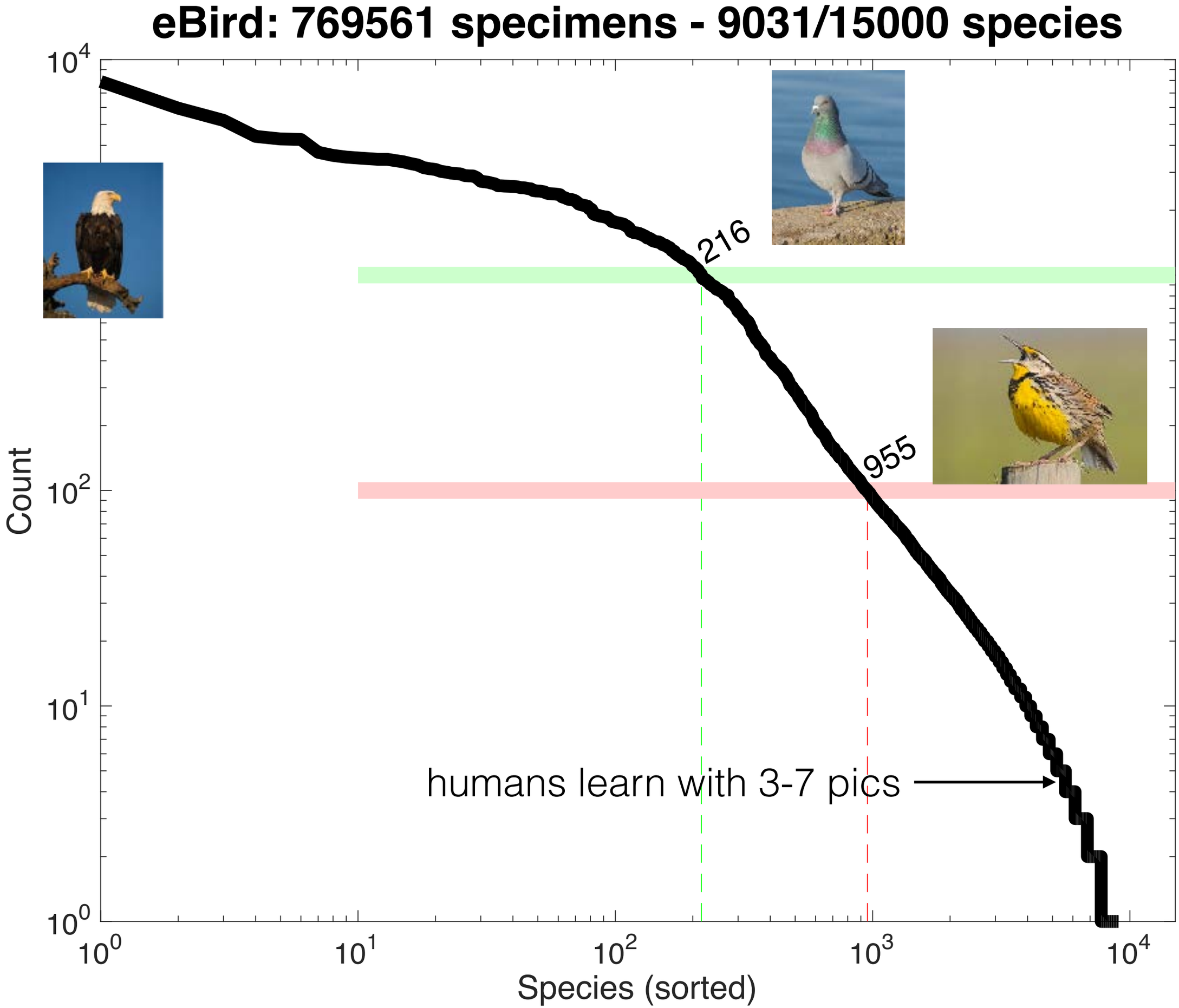


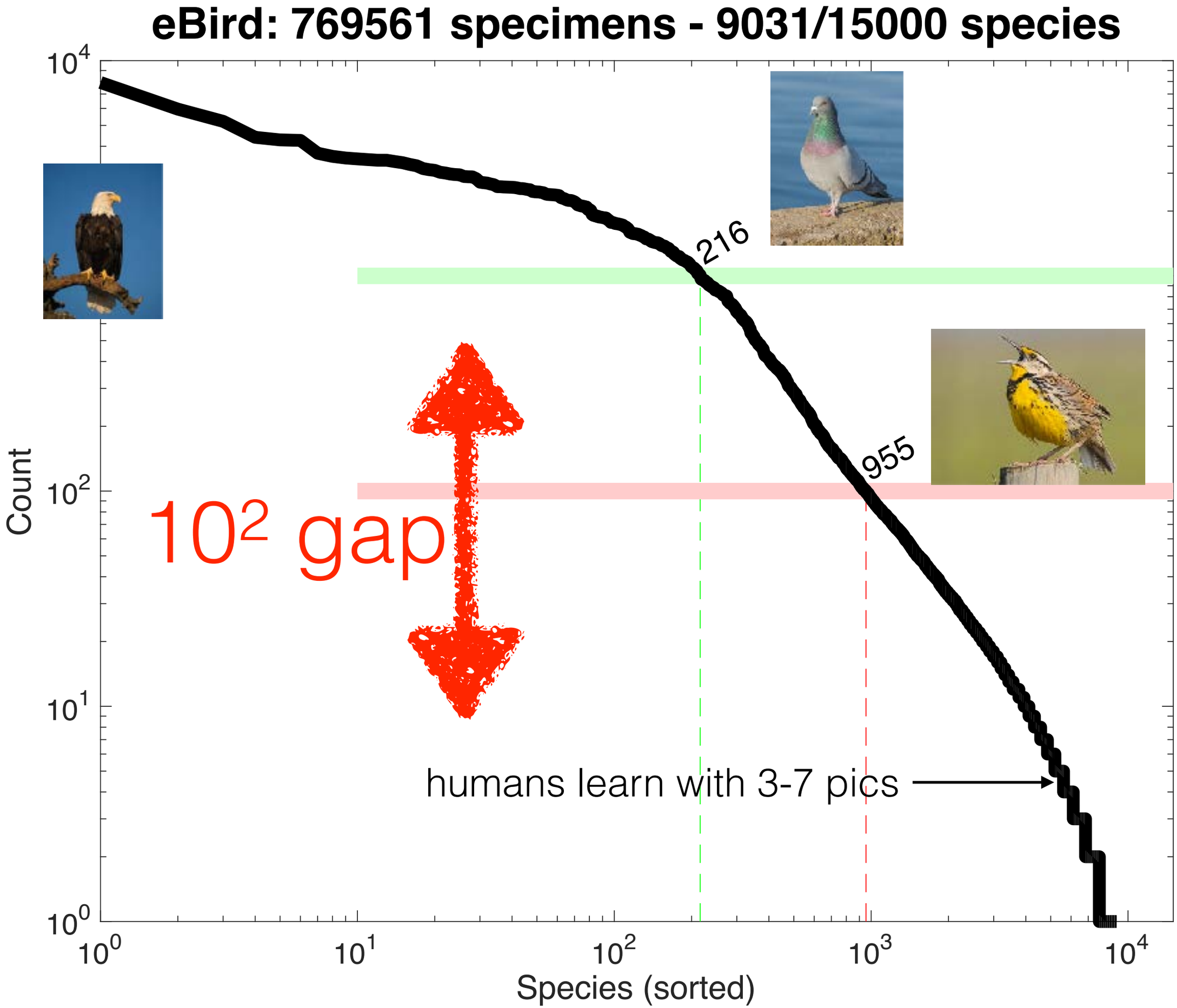
Challenges

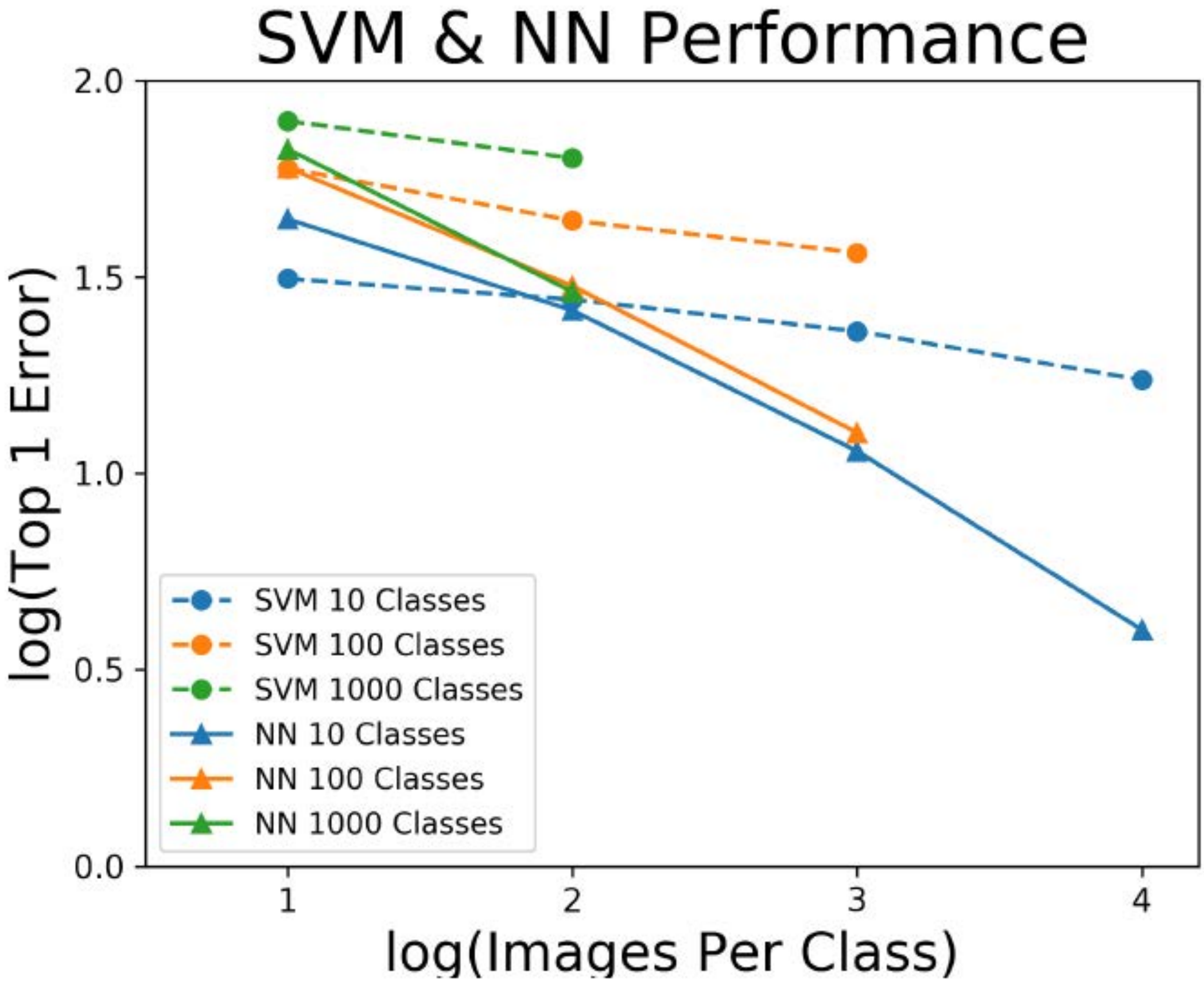
long tails (N->0)

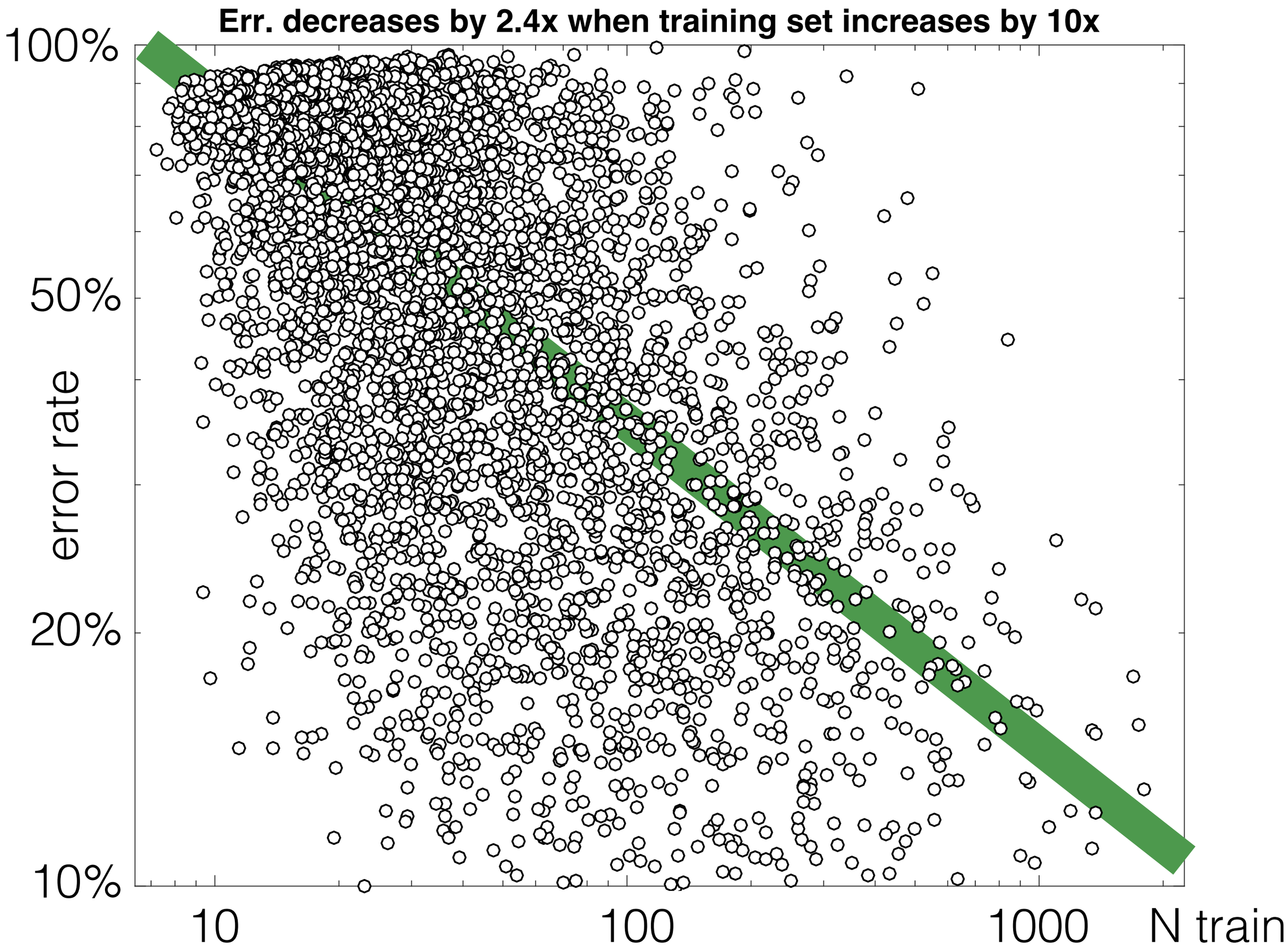

eBird: 769561 specimens - 9031/15000 species
Count
Species (sorted)
216
955
10^0
10^1
10^2
10^3
10^4


eBird: 769561 specimens - 9031/15000 species
Count
Species (sorted)
216
955
10^0
10^1
10^2
10^3
10^4


eBird: 769561 specimens - 9031/15000 species
Count
Species (sorted)
10^0
10^1
10^2
10^3
10^4
216
955
humans learn with 3-7 pics


eBird: 769561 specimens - 9031/15000 species
Count
Species (sorted)
10^0
10^1
10^2
10^3
10^4
216
955
10^2 gap
humans learn with 3-7 pics

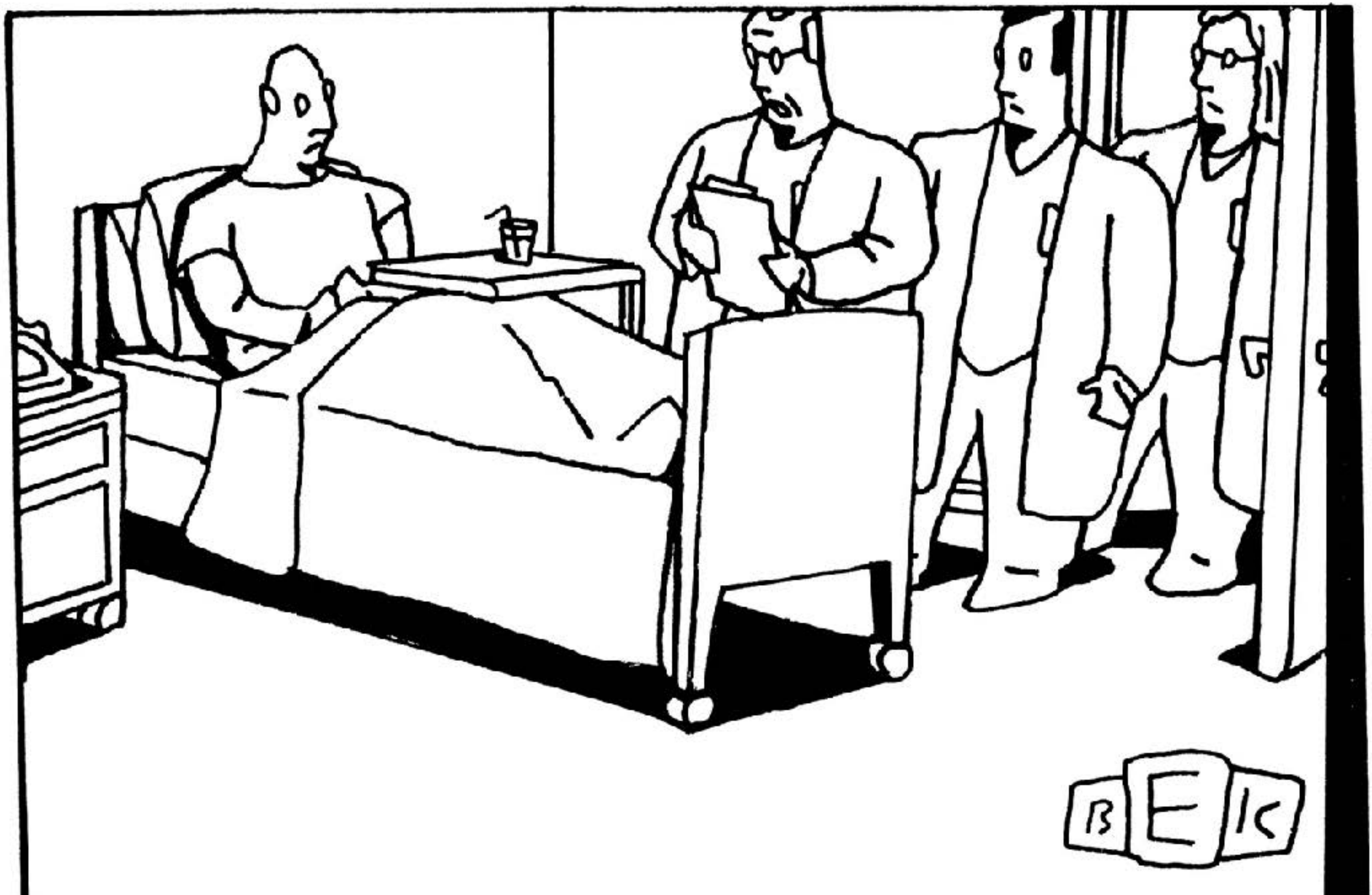

SVM & NN Performance
log(Top 1 Error)
log(Images Per Class)
2.0
1.5
1.0
0.5
0.0
1
2
3
4
SVM 10 Classes
SVM 100 Classes
SVM 1000 Classes
NN 10 Classes
NN 100 Classes
NN 1000 Classes


Err. decreases by 2.4x when training set increases by 10x
100%
50%
20%
10%
error rate
10
100
1000
N train

Levels of understanding

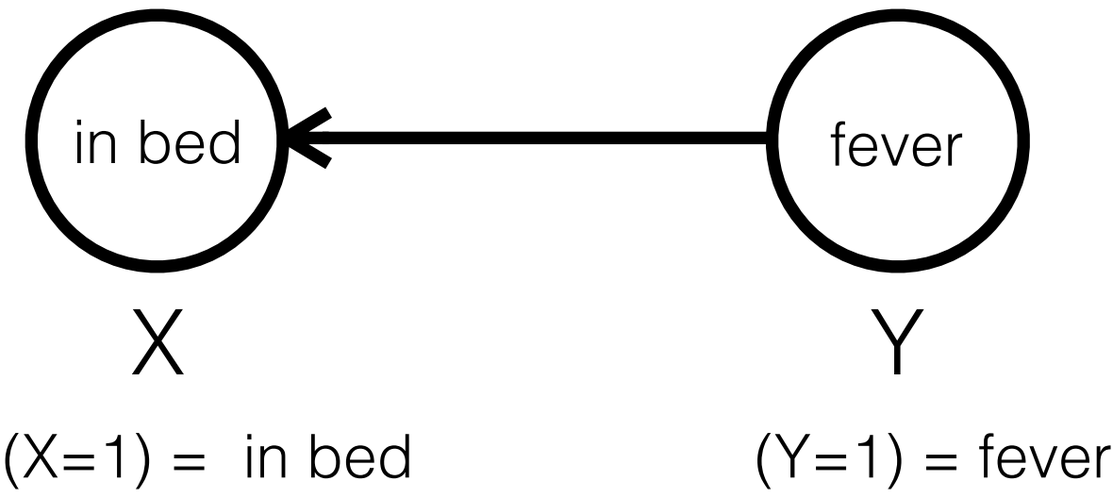
- Memorizaton / recall
- Generalization / prediction
- Mechanisms / intervention

Correlation vs causation



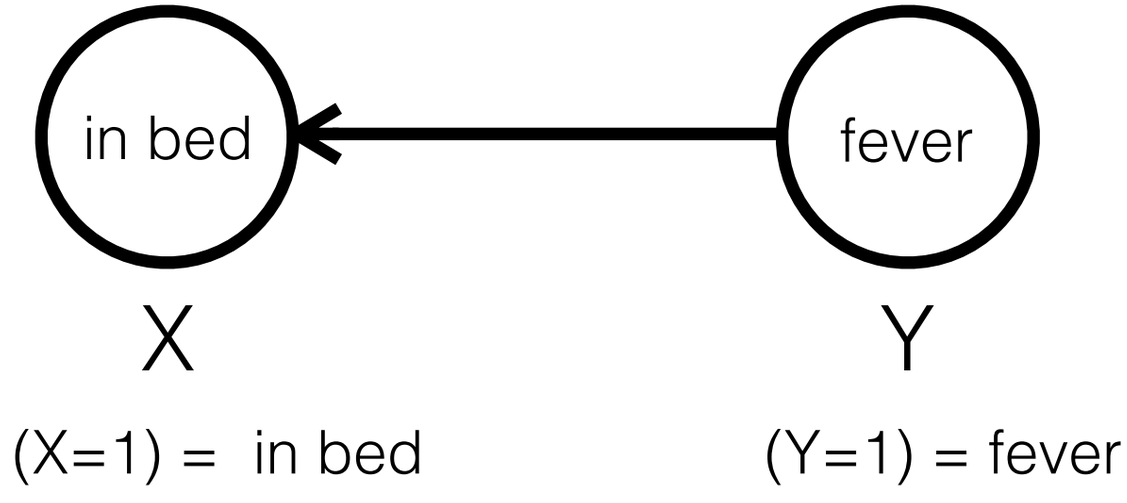
$$P(Y=1|X) = \begin{cases} 0.03 & \text{if} \quad X=0 \\ 0.95 & \text{if} \quad X=1 \end{cases}$$

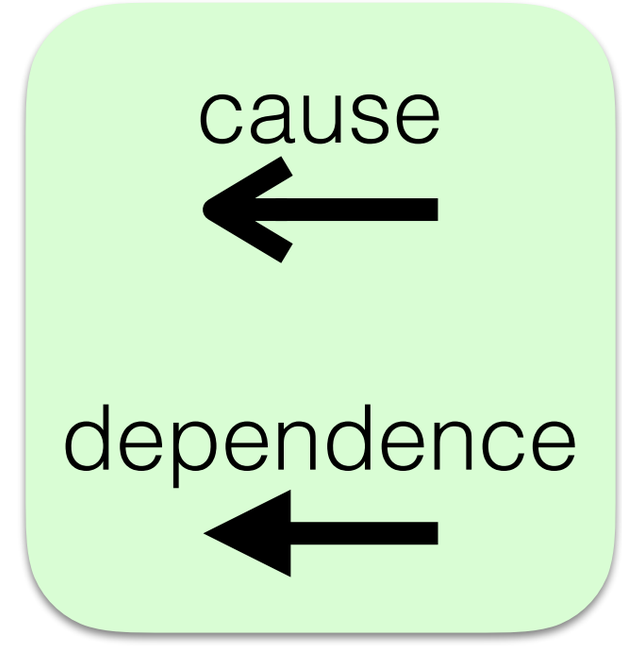
Causation



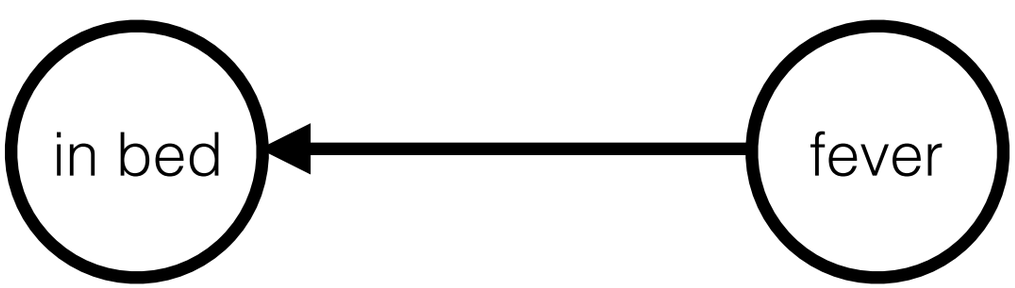
$$P(Y=1|X) = \begin{cases} 0.03 & \text{if} \quad X=0 \\ 0.95 & \text{if} \quad X=1 \end{cases}$$

Causation

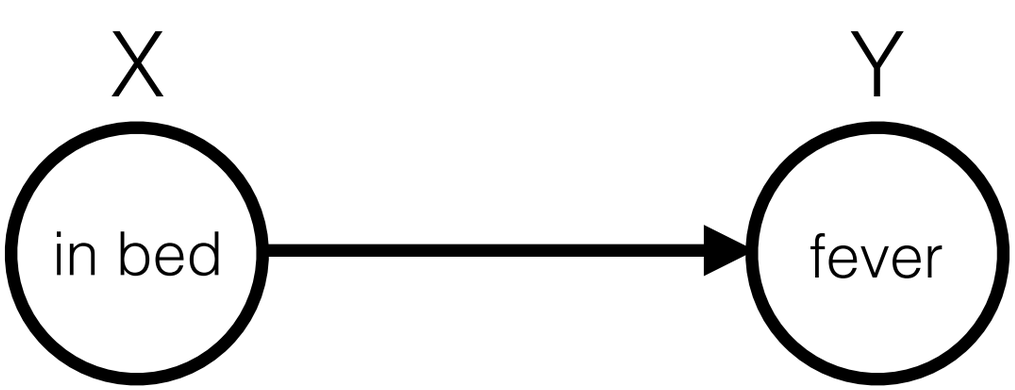




Correlation

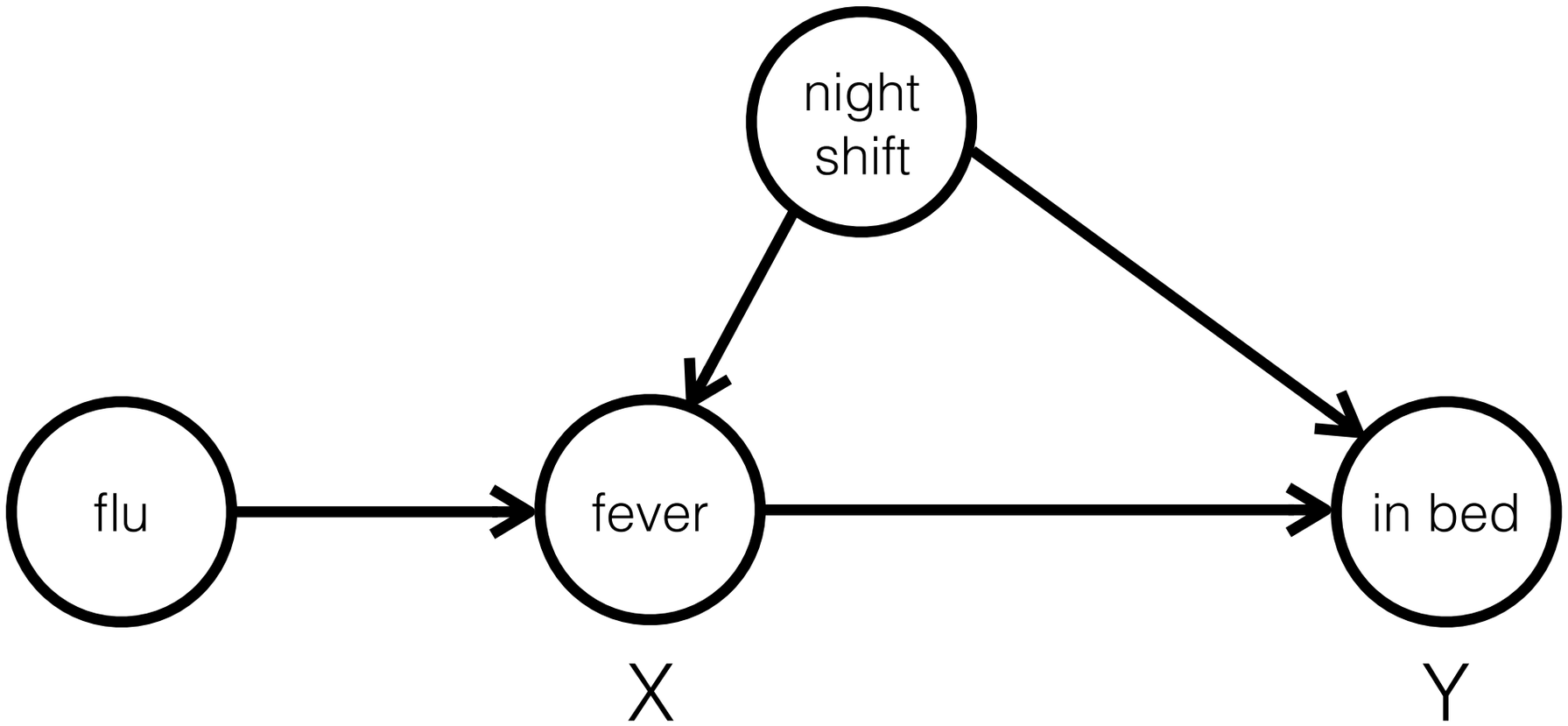


$$P(X=1|Y) = \begin{cases} 0.01 & \text{if} \quad Y=0 \\ 0.9 & \text{if} \quad Y=1 \end{cases}$$



$$P(Y=1|X) = \begin{cases} 0.03 & \text{if} \quad X=0 \\ 0.95 & \text{if} \quad X=1 \end{cases}$$

Definition of causation

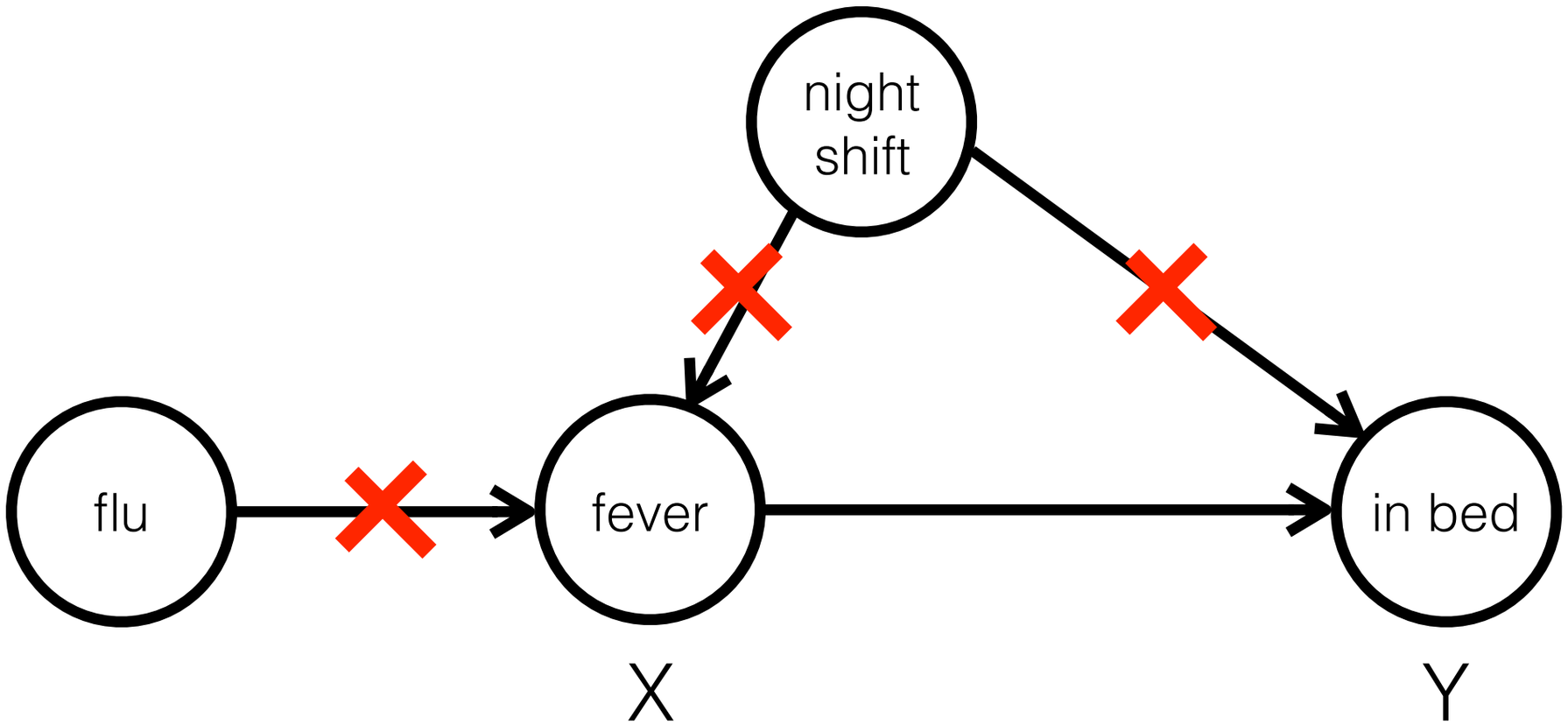


$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

[Pearl 2000]

Definition of causation

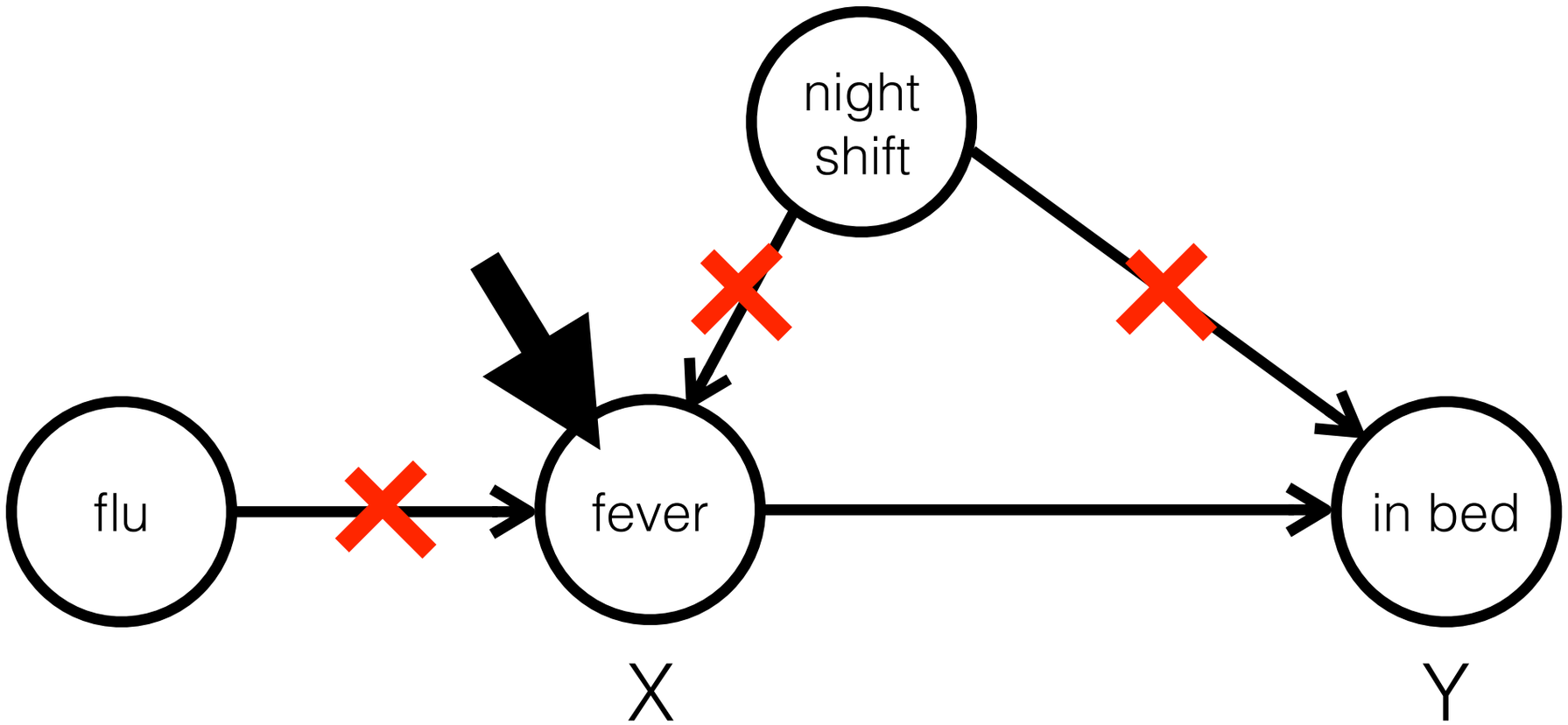


$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

[Pearl 2000]

Definition of causation

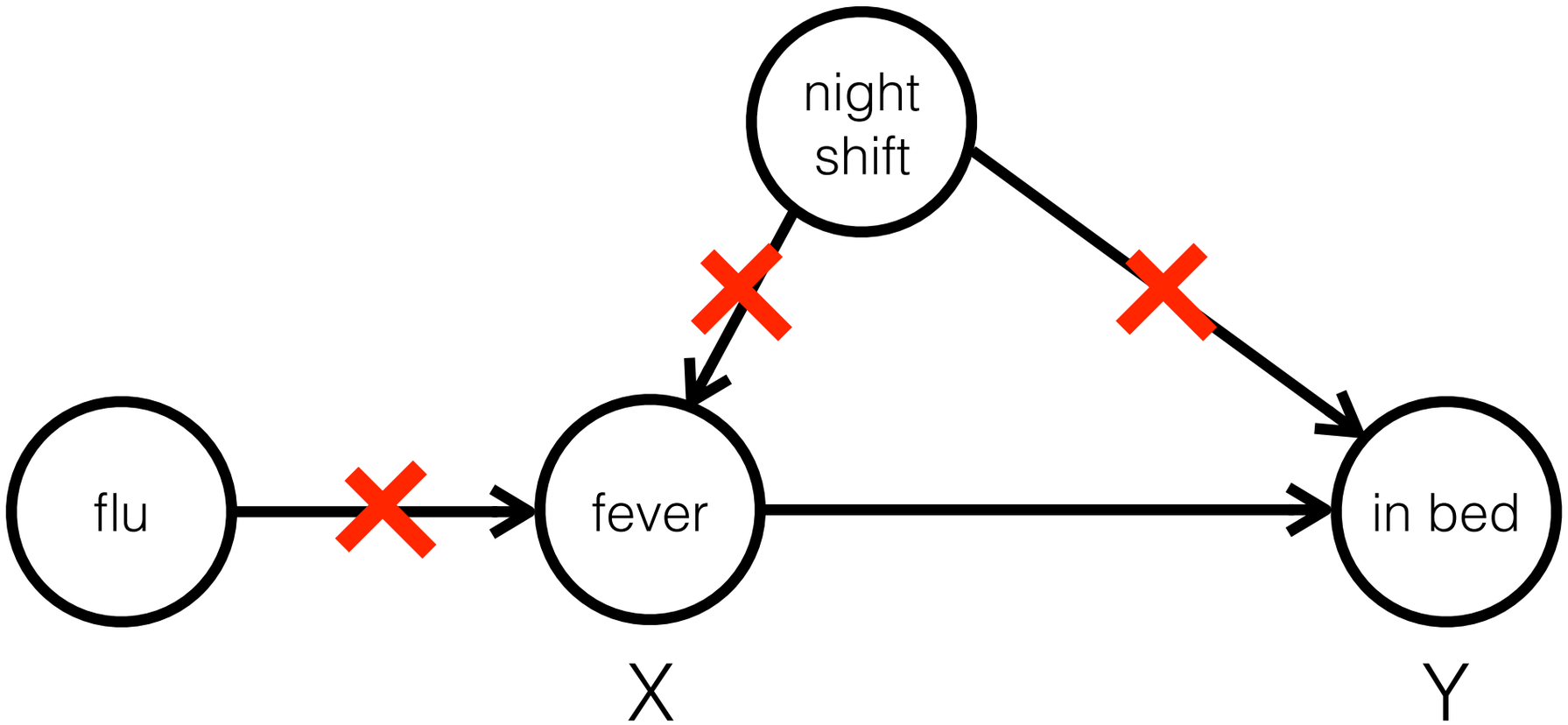


$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

[Pearl 2000]

Definition of causation

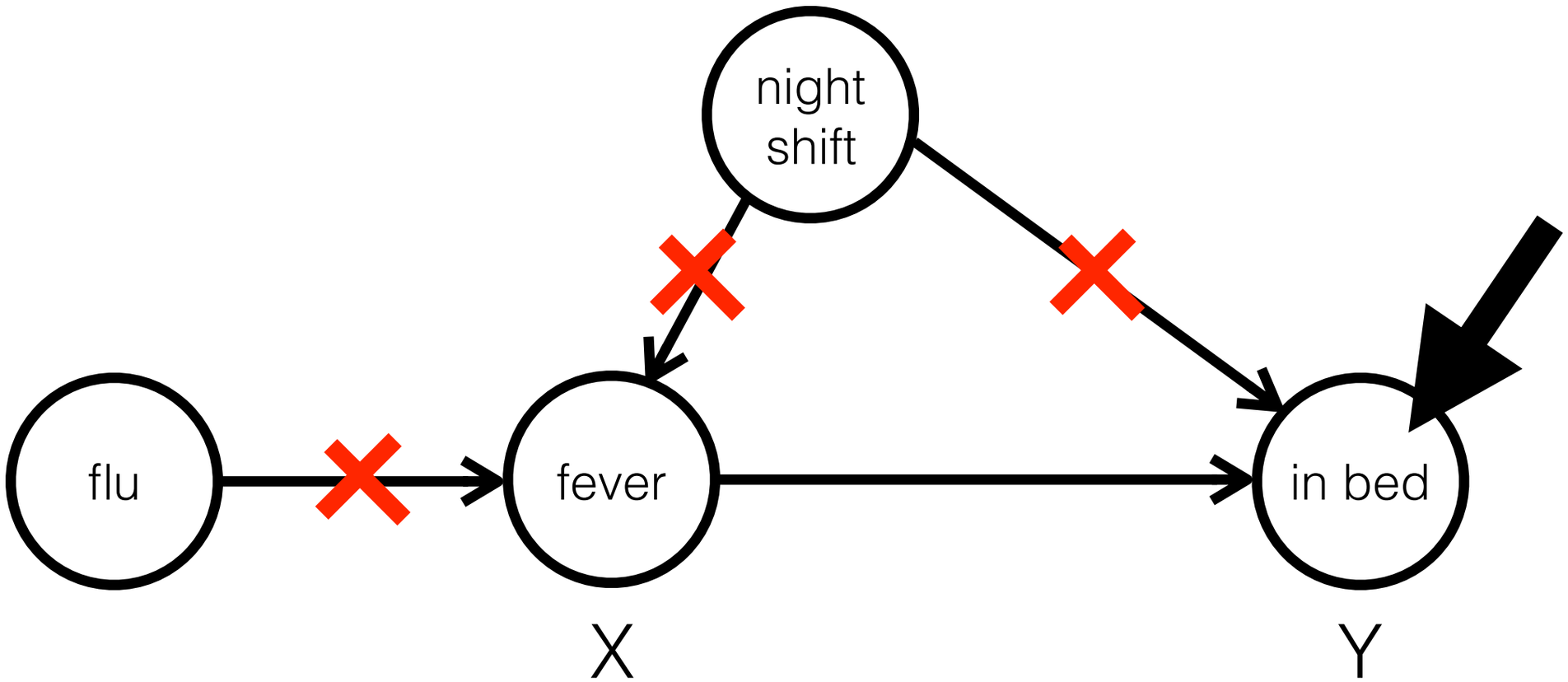


$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

[Pearl 2000]

Definition of causation

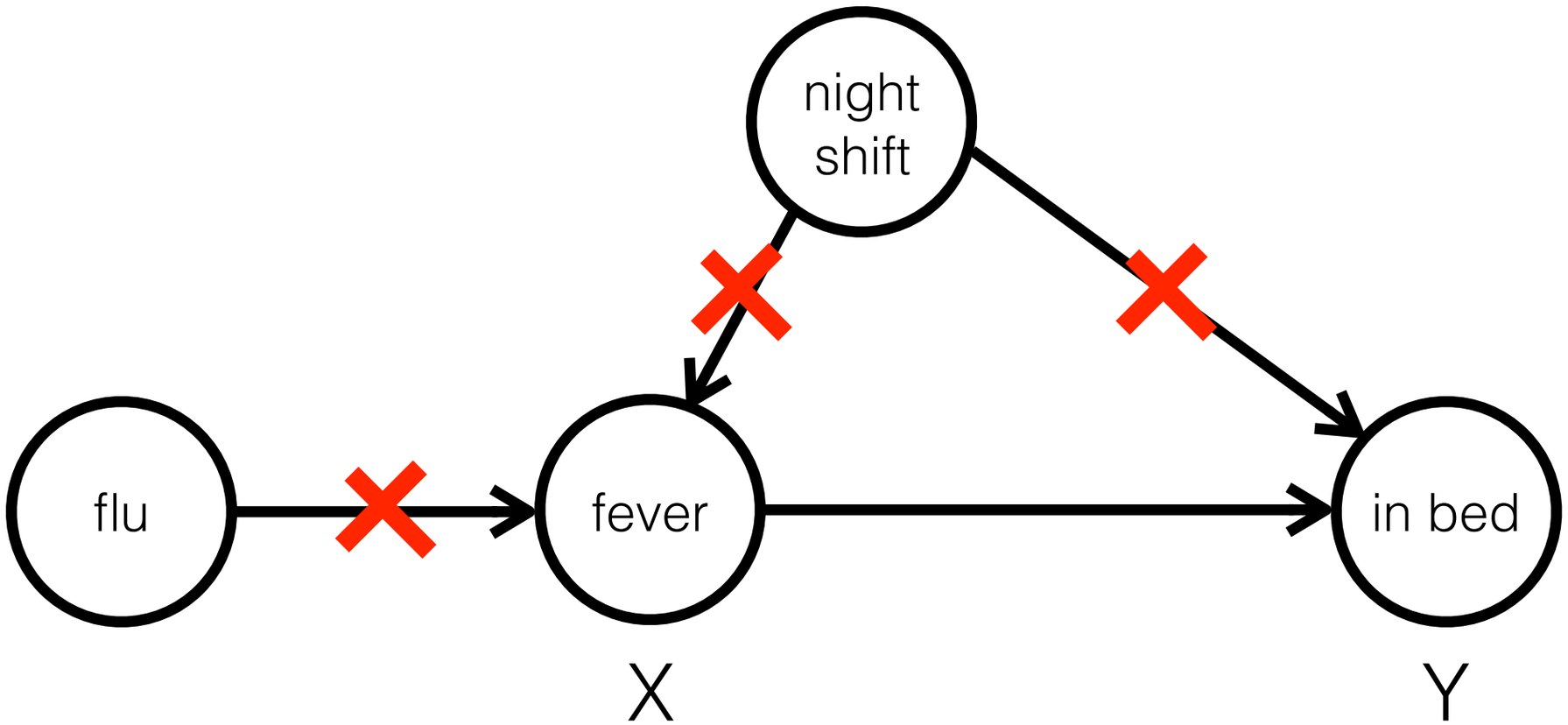


$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

[Pearl 2000]

Definition of causation

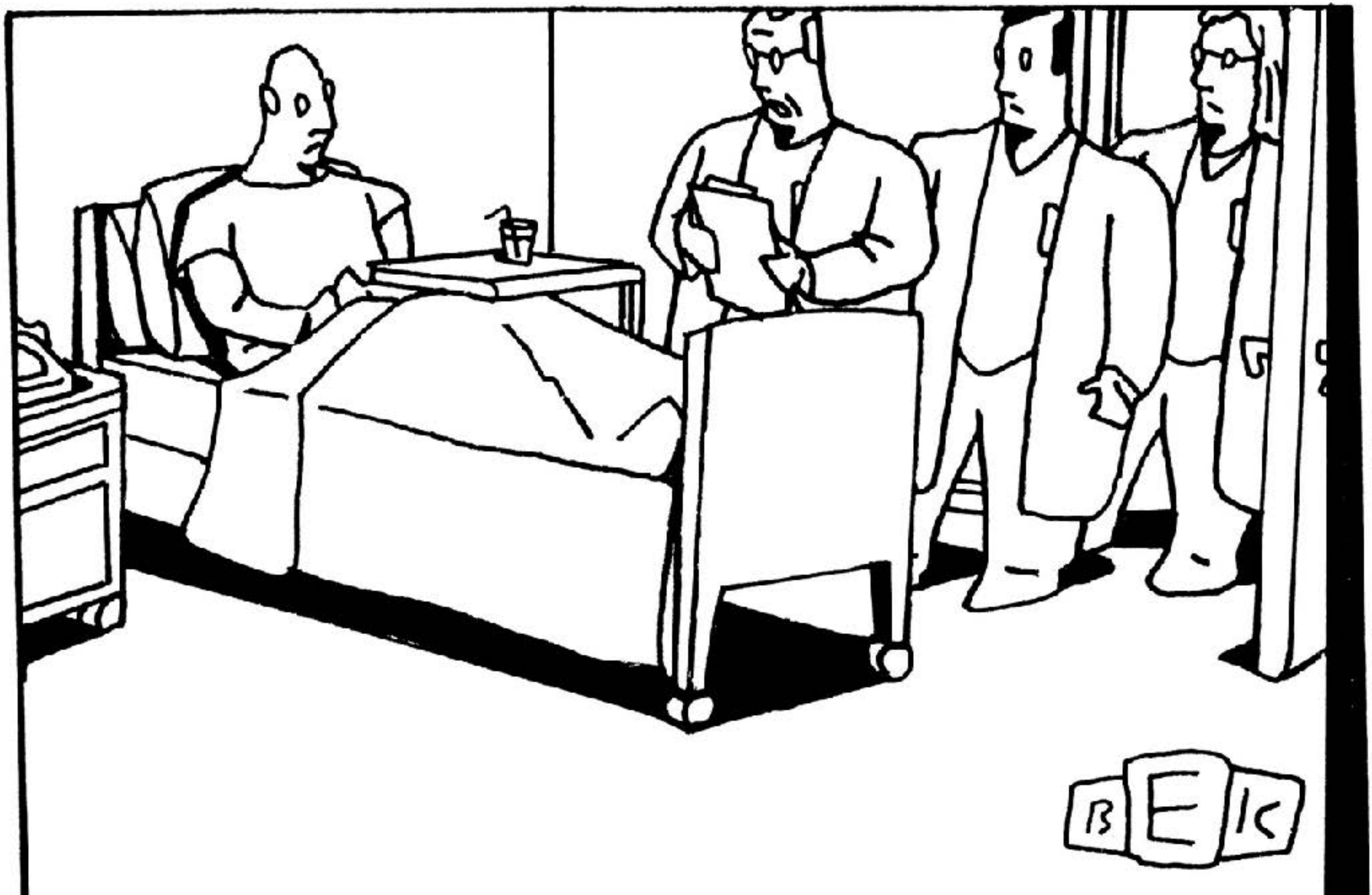


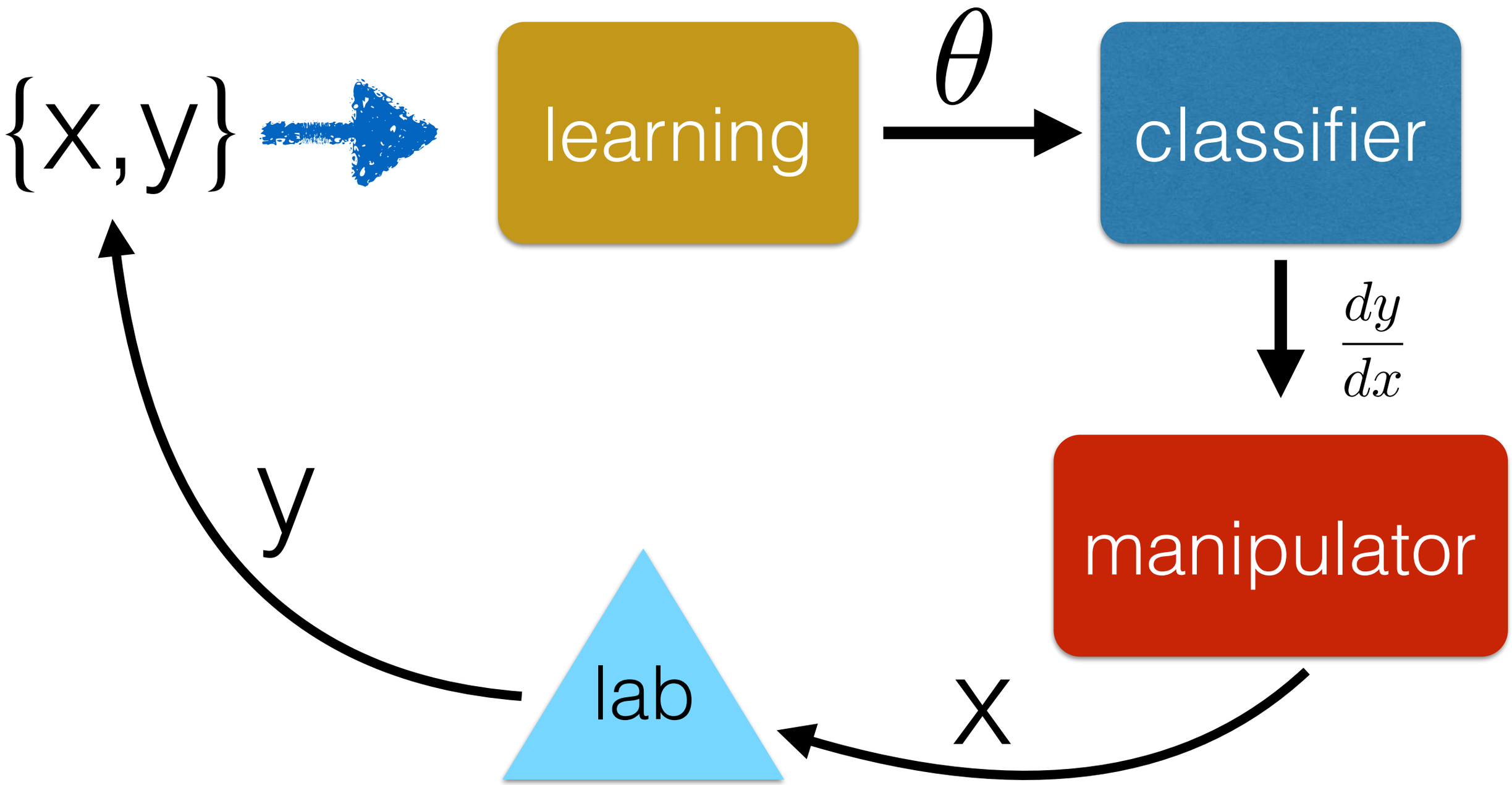
$$X \not\perp\!\!\!\perp Y \Leftrightarrow P(Y \mid X) \neq P(Y)$$

$$X \rightarrow Y \Leftrightarrow P(Y \mid \mathrm{do}(X)) \neq P(Y)$$

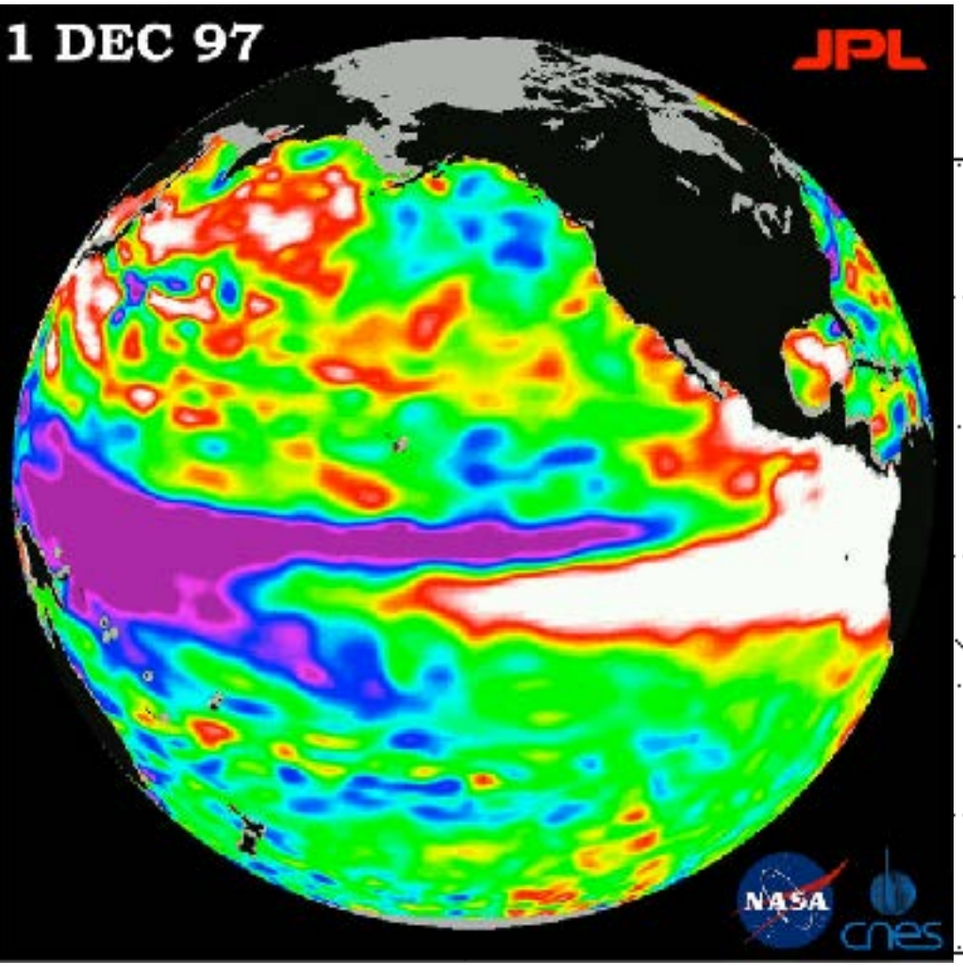
[Pearl 2000]

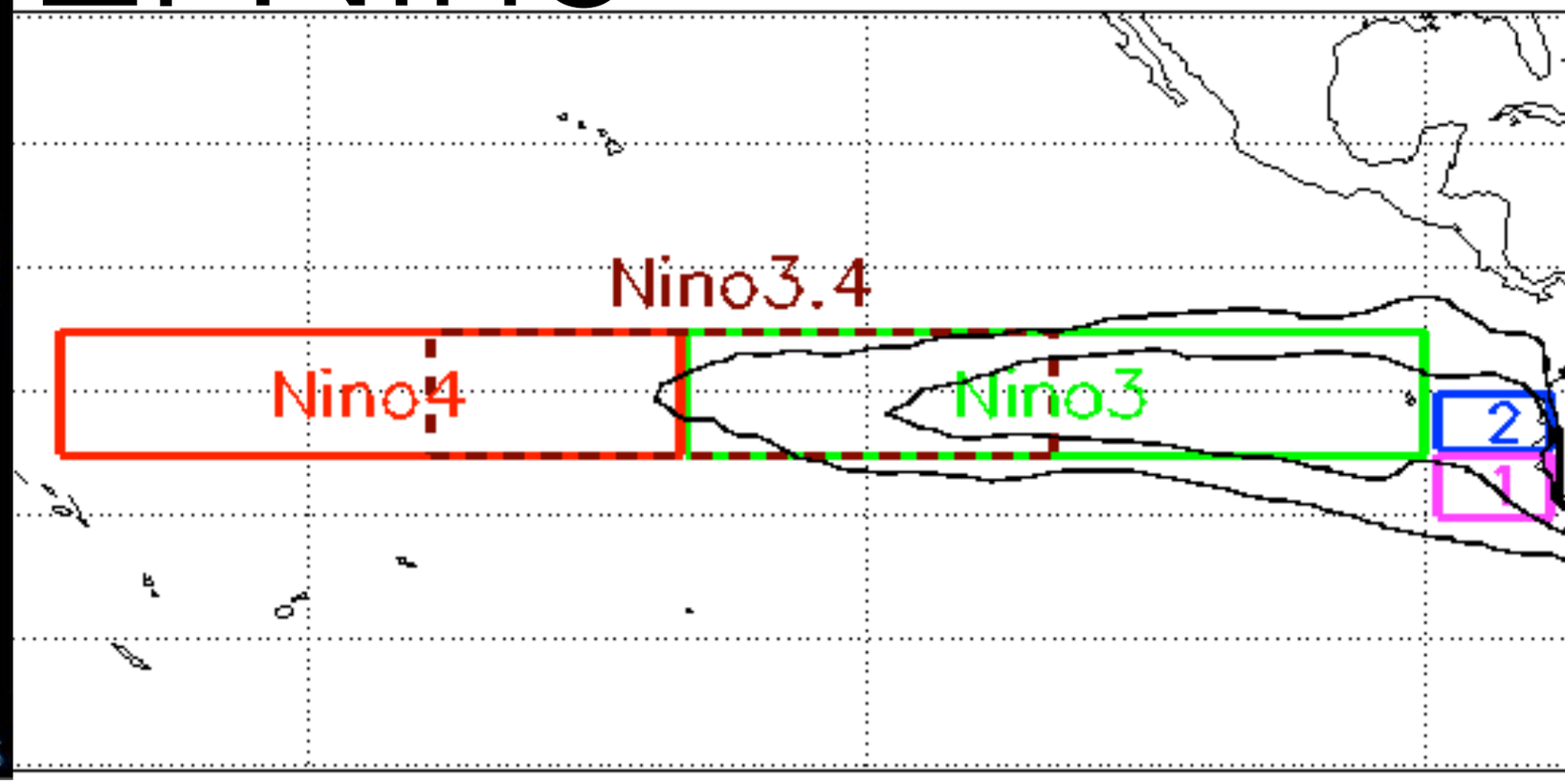
Prediction vs intervention

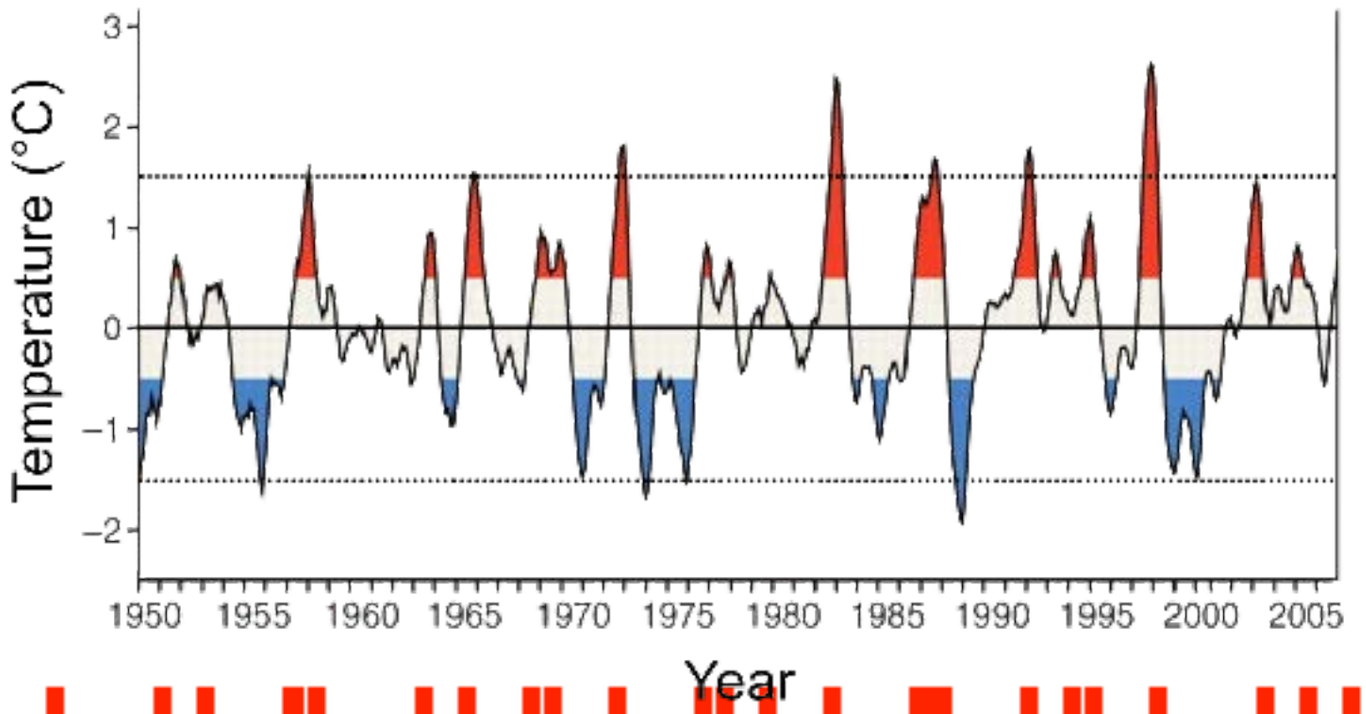


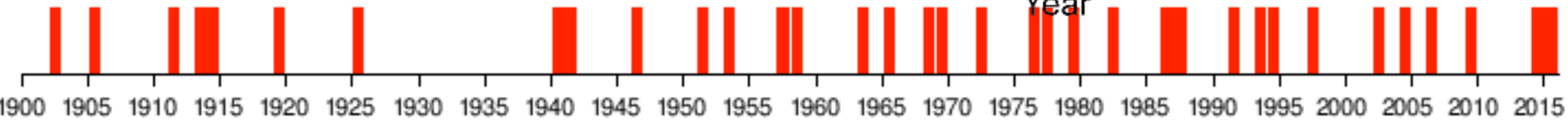

{x,y}
learning
θ
classifier
$\frac{dy}{dx}$
manipulator
x
lab
y

El Nino


1 DEC 97
JPL
NASA
cnes


Nino3.4
Nino4
Nino3
2
1


Temperature (°C)
3
2
1
0
−1
−2
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
Year


1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

clarifai

ABOUT PRICING DEVELO

cow milk agriculture farm cattle livestock dairy beef hayfield field grass mammal pasture calf farmland rural animal pastoral bull grassland

cow beef agriculture cattle milk pasture mammal livestock farmland grass farm hayfield rural herd dairy pastoral grassland field calf bull

cow mammal pasture grass animal no person nature agriculture livestock hayfield cattle farm rural field milk grassland beef pastoral countryside

clarifai

ABOUT PRICING DEVELO

cow milk agriculture farm cattle livestock dairy
beef hayfield field grass mammal pasture calf
farmland rural animal pastoral bull grassland

cow beef agriculture cattle milk pasture mammal
livestock farmland grass farm hayfield rural herd
dairy pastoral grassland field calf bull

cow mammal pasture grass animal no person nature
agriculture livestock hayfield cattle farm rural field
milk grassland beef pastoral countryside

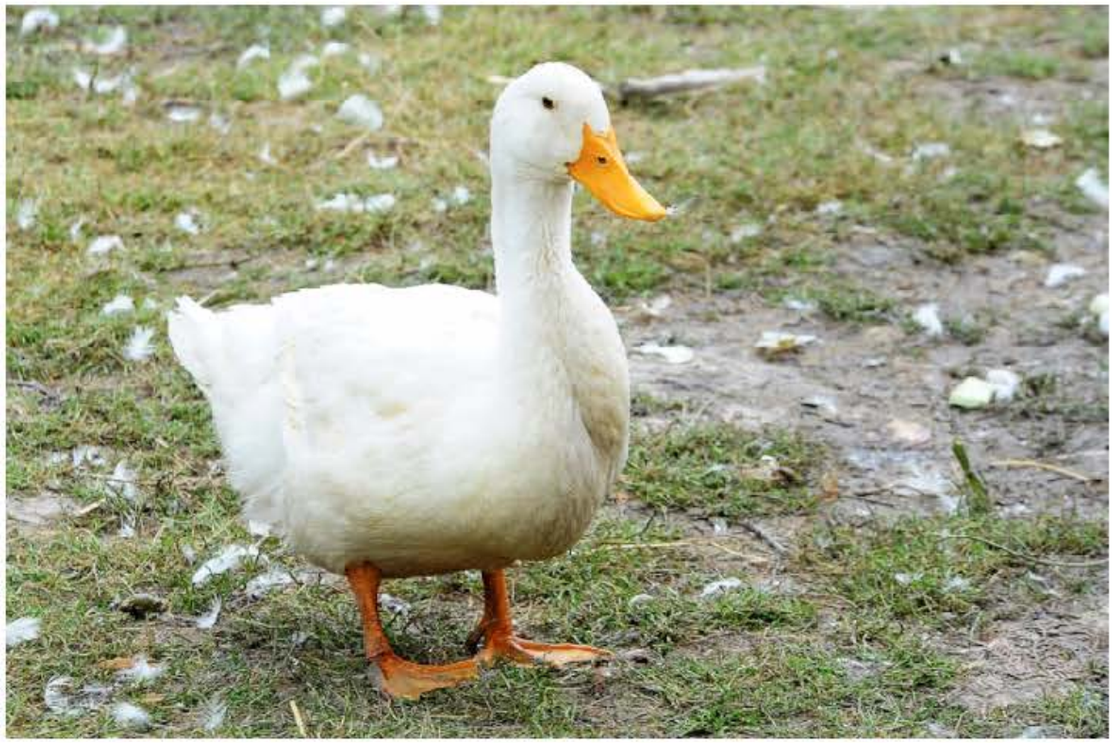
clarifai

ABOUT PRICING DEVELOPE

beach sand travel no person water sea seashore summer sky outdoors ocean nature

no person water mammal cattle outdoors cow landscape travel sky livestock

water no person beach seashore sea sand mammal outdoors travel ocean surf sky



bird duck goose poultry wildlife
waterfowl animal no person feather nature
beak outdoors water wild lake pool

bird no person one

Schoolbus

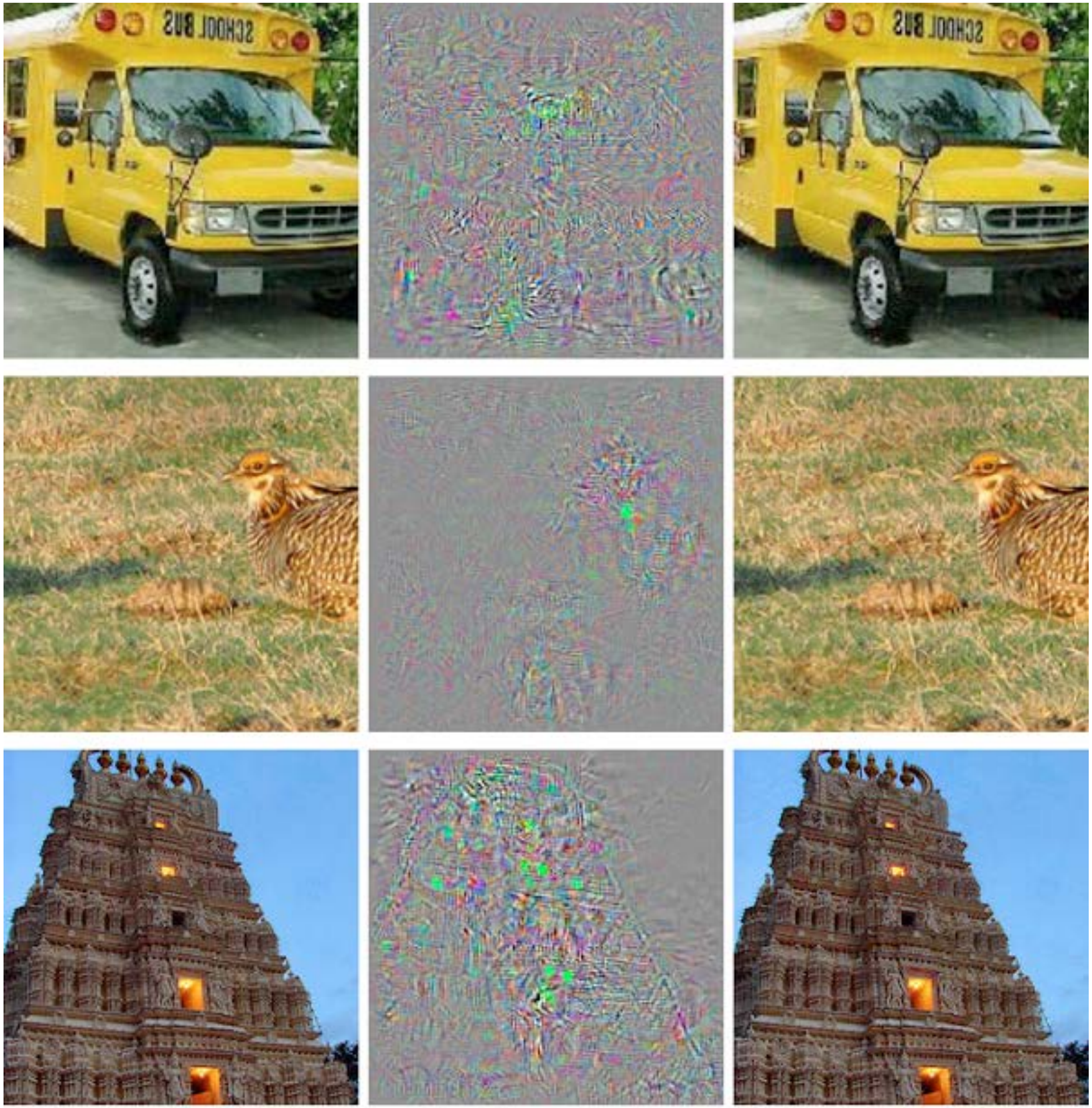
Ostrich

Grouse

Ostrich

Pyramid

Ostrich

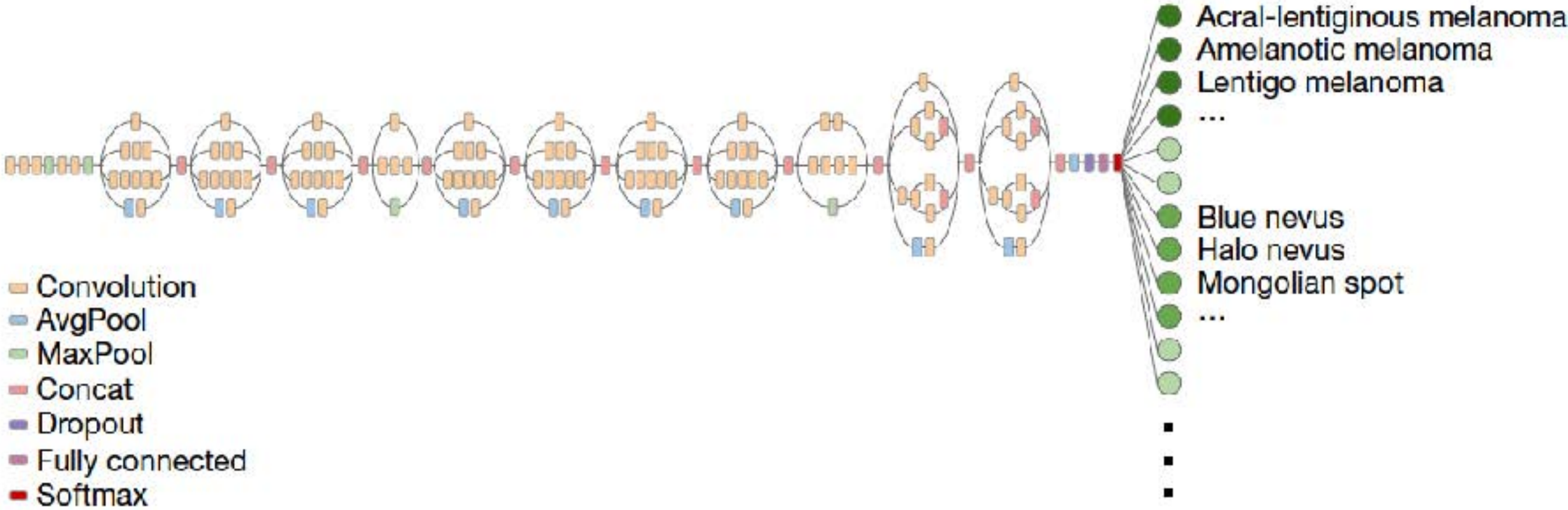


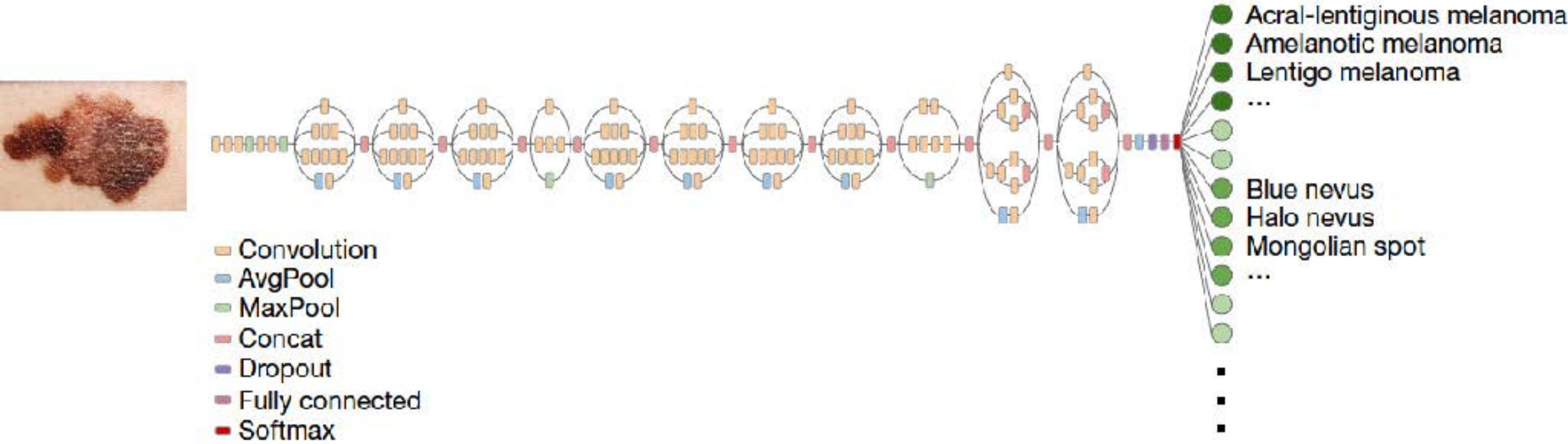
[Szegedy et al. 2014]

Levels of understanding

- Memorizaton / recall
- Generalization / prediction
- Mechanisms / intervention

theory-driven design


Acral-lentiginous melanoma
Amelanotic melanoma
Lentigo melanoma
...
Blue nevus
Halo nevus
Mongolian spot
...
Convolution
AvgPool
MaxPool
Concat
Dropout
Fully connected
Softmax



Questions:

Basic modules

Architecture

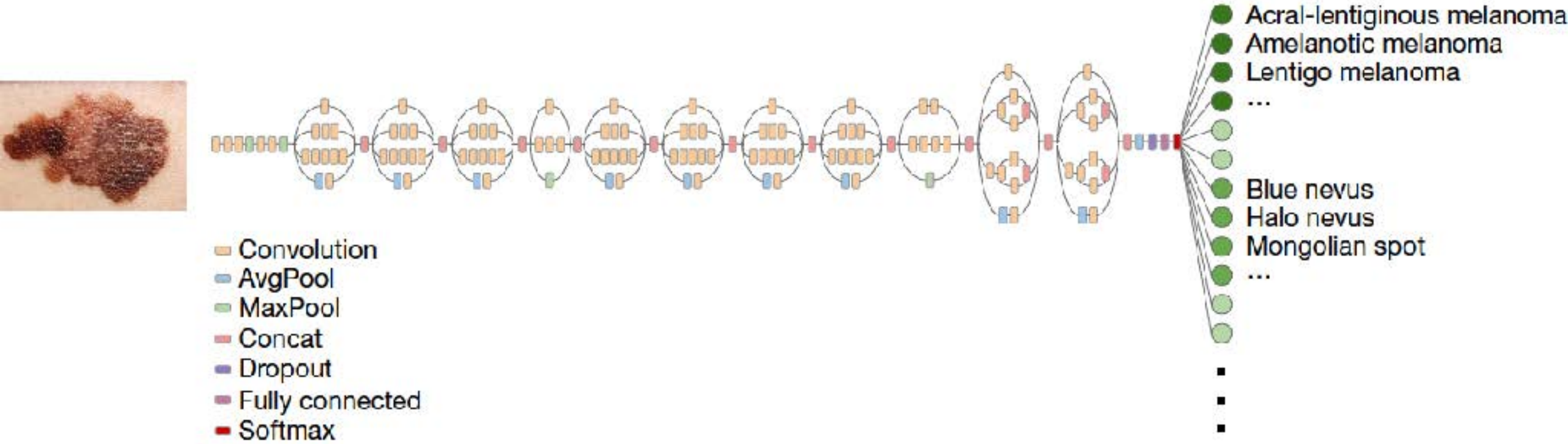
Optimization

N->0

Structure of data

Performance bounds

….



Questions:
Basic modules
Architecture
Optimization
N->0
Structure of data
Performance bounds
....

Better understanding needed

Conclusions

- Computer vision
- Learning-based approach
- Deep networks
- Practical results
- Open problems: N->0, causality, design
- Need theory