Introduction to Computer Vision

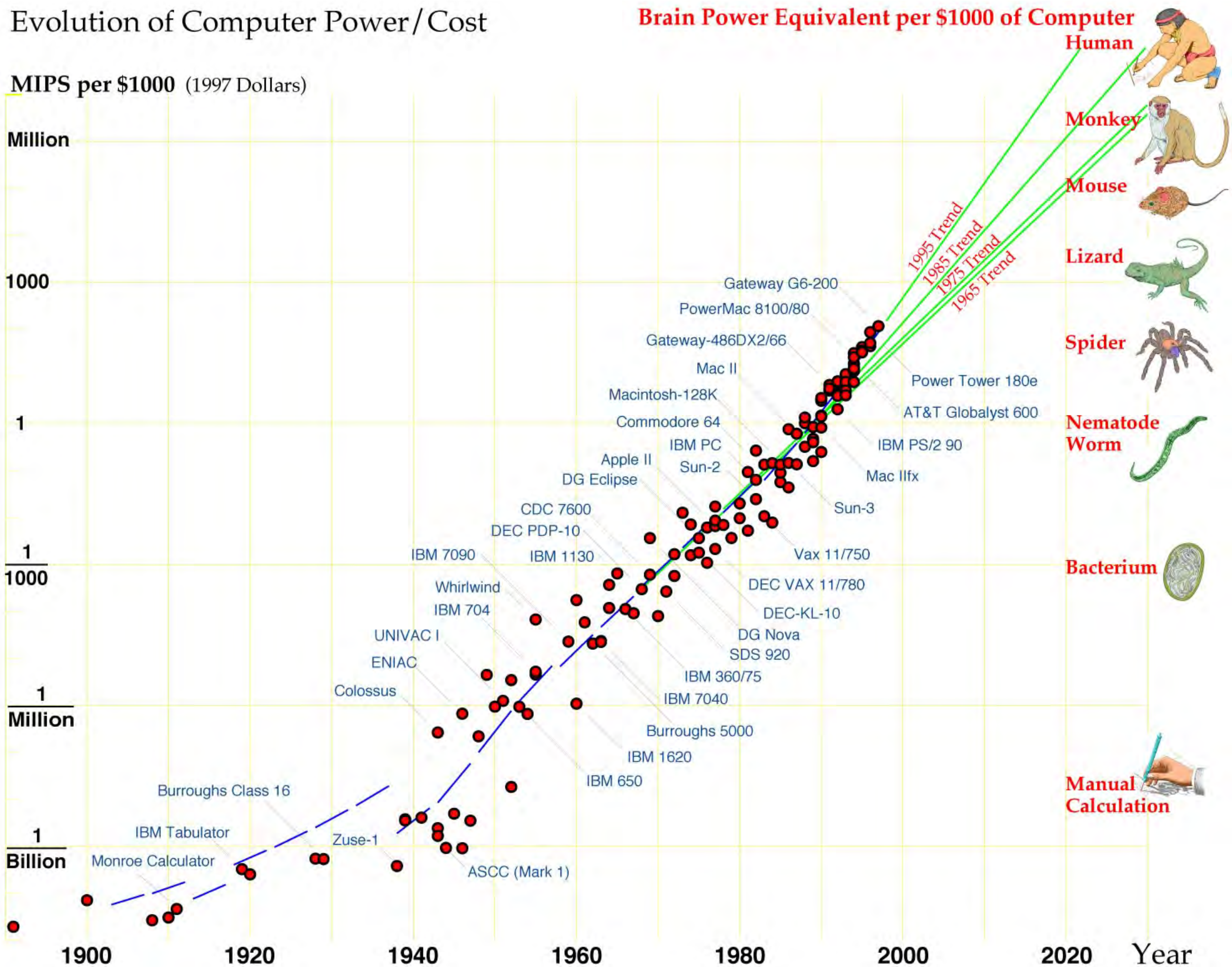
Lecture 1
Stella Yu

Slide Credit: Jitendra Malik

Moravec's argument(1998)

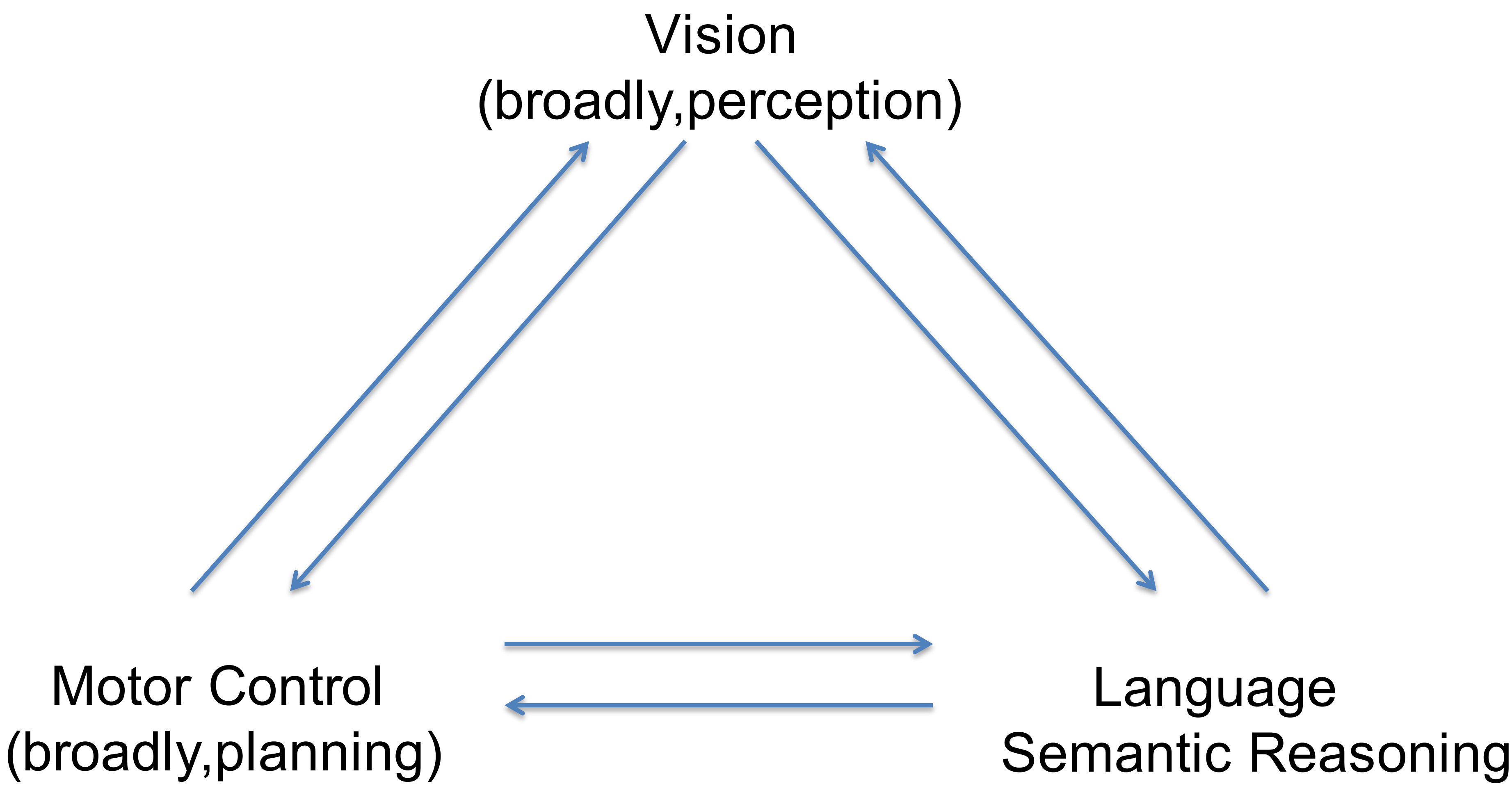
ROBOT: Mere Machine To Transcendent Mind

- 1 neuron = 1000 instructions/sec
- 1 synapse = 1 byte of information
- Human brain = 100 billion neurons
- Human brain then processes 10^14 IPS and has 10^14 bytes of storage
- In 2000, we have 10^9 IPS and 10^9 bytes on a desktop machine
- Assuming Moore's law we obtain human level computing power in 2025, or with a cluster of 100 nodes in 2015.



Evolution of Computer Power/Cost
Brain Power Equivalent per $1000 of Computer
MIPS per $1000 (1997 Dollars)
Million
1000
1
1/1000
1/Million
1/Billion
1900
1920
1940
1960
1980
2000
2020
Year
Human
Monkey
Mouse
Lizard
Spider
Nematode Worm
Bacterium
Manual Calculation
1995 Trend
1985 Trend
1975 Trend
1965 Trend
Gateway G6-200
PowerMac 8100/80
Gateway-486DX2/66
Mac II
Macintosh-128K
Commodore 64
IBM PC
Apple II
Sun-2
DG Eclipse
CDC 7600
DEC PDP-10
IBM 7090
IBM 1130
Whirlwind
IBM 704
UNIVAC I
ENIAC
Colossus
Burroughs Class 16
IBM Tabulator
Zuse-1
Monroe Calculator
ASCC (Mark 1)
Power Tower 180e
AT&T Globalyst 600
IBM PS/2 90
Mac IIfx
Sun-3
Vax 11/750
DEC VAX 11/780
DEC-KL-10
DG Nova
SDS 920
IBM 360/75
IBM 7040
Burroughs 5000
IBM 1620
IBM 650

Embodied Cognition



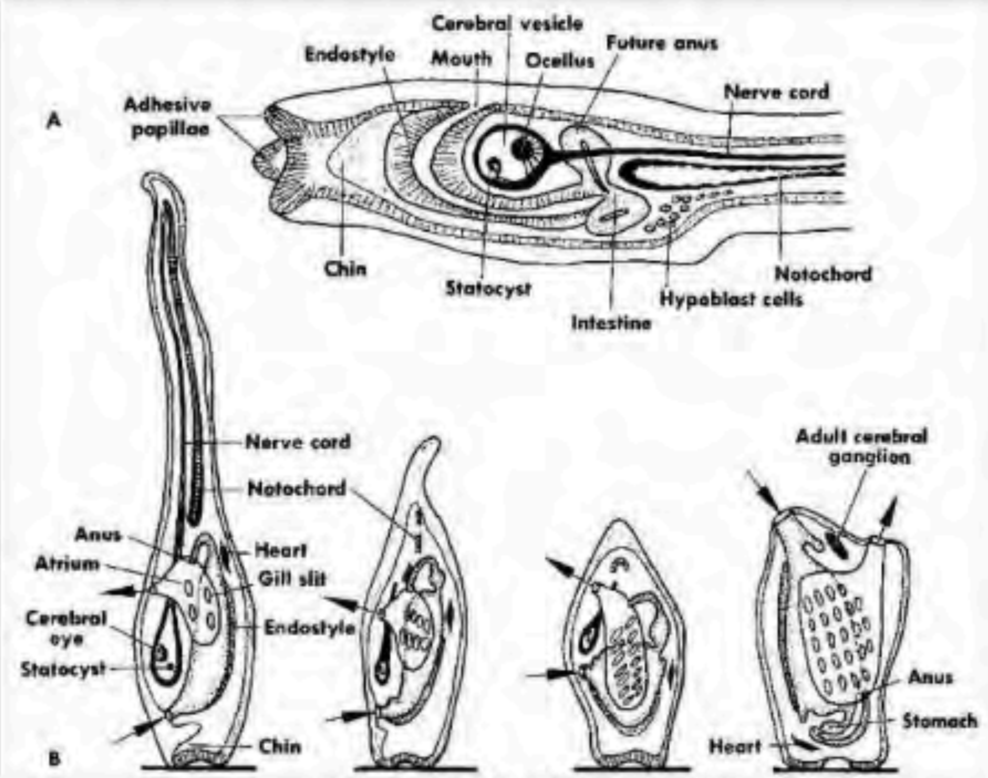
Phylogeny of Intelligence

- The Cambrian period (543-490 million yrs ago) led to the emergence of wide variety of animal life. These animals had vision and locomotion capabilities.
- Sensory systems provide great benefits only when accompanied by the ability to move - to find food, avoid predators etc.

If you don’t need to move you don’t need an eye or a brain!

https://goodheartextremescience.wordpress.com/2010/01/27/meet-the-creature-that-eats-its-own-brain/

The sea squirt larvae begin absorbing all the tadpole-like parts that made them *chordates*. Where the sea squirt larva once had gills, it develops the intake and exist siphons that will help it bring water and food into its body. It absorbs its twitching tail. It absorbs its primitive eye and its spine-like notocord. Finally, it even absorbs the rudimentary little “brain” (cerebral ganglion) that it used to swim about and find its attachment place.



Sea Squirt Life Cycle - UNM

So, yes, in common parlance, the sea squirt “eats its own brain,” such as it is. But since the sea squirt no longer needs its brain to help it swim around or to see, this isn’t a great loss to the creature. It needs this use this now superfluous body material to help develop its digestive, reproductive, and circulatory organs.

Hominid evolution in last 5 million years

- Bipedalism freed the hand for tool making. Dexterous hands coevolved with larger brains.

- Anaxagoras: It is because of his being armed with hands that man is the most intelligent animal

διὰ τὸ χεῖρας ἔχειν φρονιμώτατον εἶναι τῶν ζῴων ἄνθρωπον.

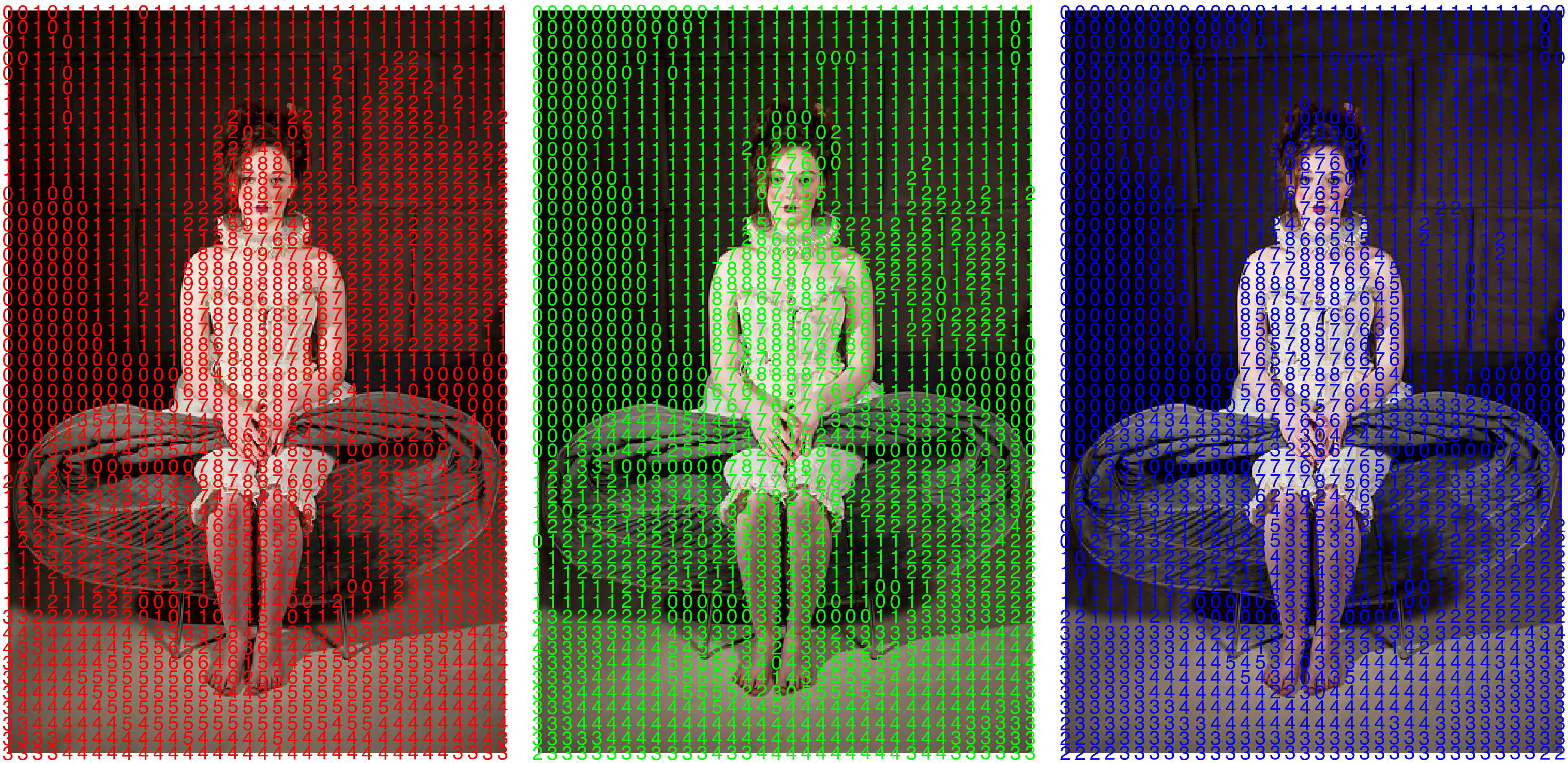
Arist. de partt. anim. I 10. 687a7; A102.

Origins of Language (from Trask)

The evolutionary progression

- Vision and Locomotion
- Manipulation
- Language

Successes in AI seem to follow the same order!

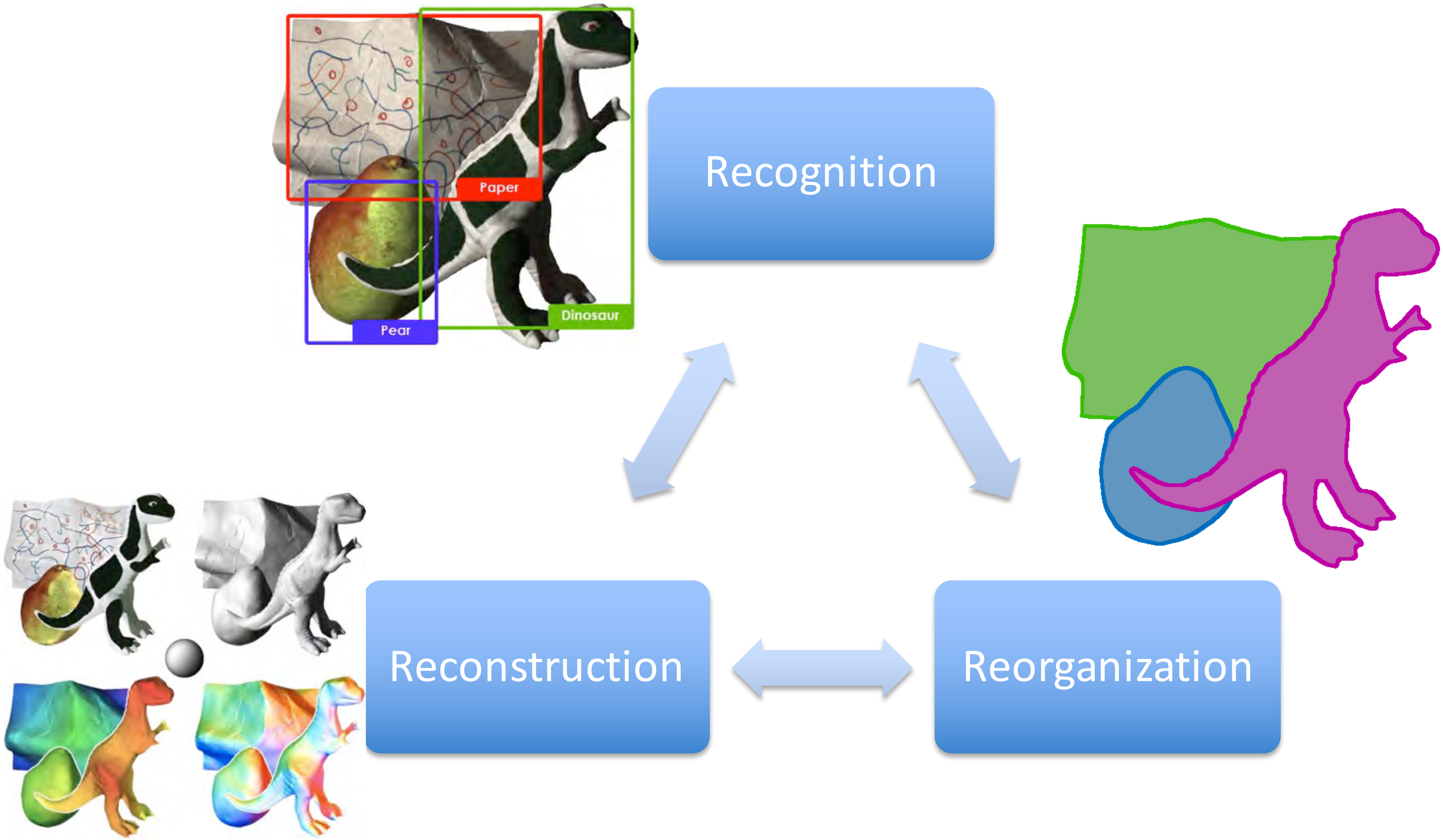


0 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 2 1 1 2 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 1 2 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 2 1 2 2 2 2 2 1 1 1 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 0 1 1 0 3 1 2 1 2 2 2 2 2 2 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 4 3 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 8 8 8 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 4 7 8 7 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2
0 1 1 0 0 1 1 1 1 1 1 1 1 1 2 8 8 8 7 7 2 2 2 2 2 2 2 2 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 2 2 2 2 8 7 7 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 2 2 2 5 9 8 7 4 6 2 2 2 2 2 2 2 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 8 7 7 6 6 6 2 2 2 2 2 2 2 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 1 8 8 7 9 9 7 7 7 6 2 2 2 2 2 1 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 8 9 8 8 9 9 8 8 8 7 2 2 2 2 2 1 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 9 9 9 9 8 8 8 8 8 7 7 2 2 2 2 1 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 2 1 1 9 7 8 6 8 6 8 8 7 6 7 2 2 2 2 0 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 1 8 7 8 8 8 8 8 8 7 6 7 2 2 2 2 1 2 2 2 2 2 1
0 0 0 0 0 0 0 1 1 1 1 1 8 7 8 7 8 5 8 7 7 5 7 2 2 2 2 1 2 2 2 2 1 1
0 0 0 0 0 0 0 1 1 1 1 1 8 8 6 8 8 8 8 7 8 9 7 2 2 2 2 2 2 2 2 2 1 1
0 0 0 0 0 0 0 0 0 0 1 1 8 8 7 8 8 8 7 7 8 8 8 1 1 1 1 1 1 1 1 1 0 0
0 0 0 0 0 0 0 0 0 0 0 8 8 4 8 8 8 8 8 8 8 6 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 6 8 8 8 8 8 8 8 7 6 5 4 1 1 1 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 7 7 8 8 7 6 8 7 6 5 3 4 3 3 3 3 3 2 1 0 0 0
0 0 0 0 0 3 5 4 4 4 5 4 4 4 6 8 7 8 8 2 3 3 3 3 3 3 3 3 1 1 2 4 0 0
0 0 0 3 4 4 4 4 4 4 3 3 3 2 4 8 6 3 7 2 4 4 4 2 1 3 3 2 2 3 1 3 4 0
0 0 2 3 3 0 4 4 4 3 5 5 4 4 3 3 6 9 7 8 3 1 1 0 0 0 0 0 0 0 3 1 3 0
1 2 1 3 3 1 0 0 0 0 0 0 0 7 8 7 7 8 8 7 7 6 1 3 2 2 2 2 3 4 1 2 3 2
2 2 3 2 1 2 1 0 3 3 3 3 2 8 8 7 8 8 7 6 6 6 2 3 2 2 3 3 4 3 2 3 3 2
1 2 2 2 1 2 3 3 3 3 4 3 3 7 4 7 8 4 6 8 7 2 2 2 3 3 3 2 3 3 3 4 3 2
1 2 0 2 2 2 4 4 5 5 4 3 4 4 6 5 6 6 6 5 6 1 2 2 2 3 2 2 3 4 3 3 3 2
1 2 2 3 3 3 1 2 2 3 2 1 3 4 6 4 5 6 5 5 3 1 1 2 2 2 2 3 2 3 3 3 4 2
1 2 2 2 2 3 4 2 2 1 2 1 3 3 6 5 5 6 5 5 1 1 1 1 1 2 3 2 3 3 3 5 3 3
1 1 3 3 2 2 3 2 3 3 2 3 2 3 5 4 5 6 5 4 1 1 1 1 1 2 2 3 3 3 2 3 3 3
1 1 1 2 2 2 2 3 2 2 1 2 1 1 5 4 4 5 4 4 1 1 1 1 1 2 2 2 4 3 3 3 3 3
1 1 1 1 1 1 2 3 2 2 3 2 2 1 5 4 4 4 4 4 1 1 1 0 0 1 2 3 3 3 3 3 3 3
1 1 1 2 1 1 2 2 2 0 0 0 1 0 4 4 4 4 4 0 0 1 2 1 1 1 2 3 3 3 3 3 3 3
3 3 2 2 2 2 2 2 2 0 1 0 1 1 0 4 4 5 4 0 1 1 1 1 2 3 3 3 3 3 3 3 3 3
4 4 3 4 4 4 4 4 4 4 3 3 3 3 3 5 4 5 4 3 3 3 3 3 3 3 3 3 3 3 5 4 4 5
4 3 3 4 4 4 4 4 5 5 5 6 5 2 4 6 3 6 4 4 4 4 2 5 3 5 5 5 5 5 4 4 4 4
3 3 4 4 4 4 4 5 5 5 5 6 6 6 4 6 1 5 4 4 5 5 5 5 5 5 5 5 5 4 4 4 4 4
3 4 4 4 4 5 5 5 5 5 5 5 6 6 5 6 2 5 5 6 5 5 5 5 5 5 5 5 5 5 4 4 4 4
3 4 4 4 4 4 5 5 5 5 5 5 5 5 4 4 4 3 5 5 5 5 5 5 5 5 5 5 4 4 4 4 4 4
3 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4
3 3 4 4 4 4 4 4 5 4 5 5 5 5 5 5 5 5 5 5 5 5 4 5 5 4 4 4 4 4 4 4 4 4
3 3 3 3 4 4 4 4 4 4 4 4 5 4 4 4 4 4 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3
3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3

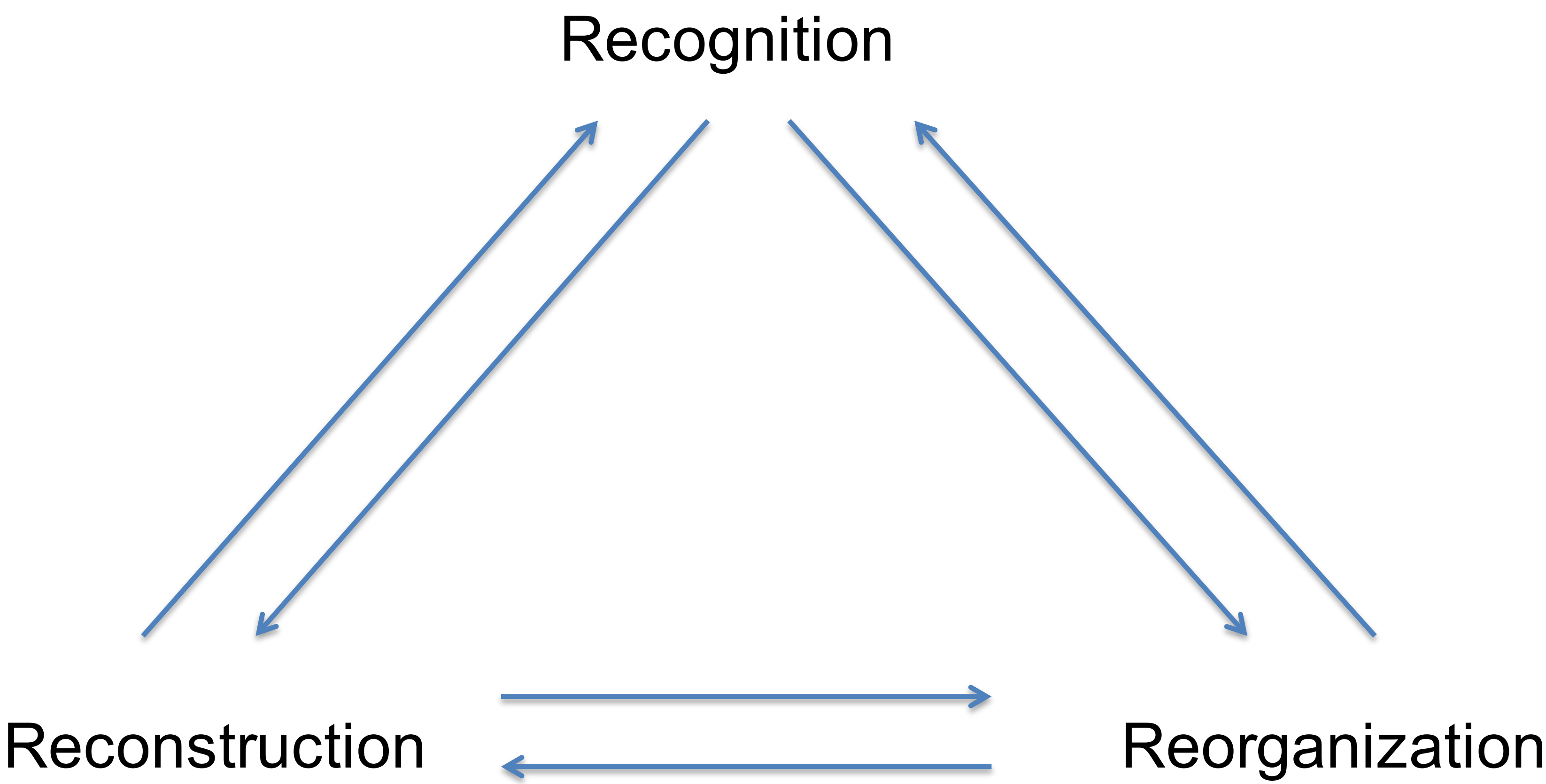
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0 1
0 0 0 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 2 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 7 7 6 0 0 1 1 1 1 1 2 1 1 1 1 1 1 1
0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 6 7 6 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 6 7 7 5 5 1 1 1 1 1 1 2 2 1 1 2 1 1 2
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 7 4 5 1 2 1 1 1 1 1 2 2 2 2 2 2 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 2 5 7 6 6 3 6 2 2 1 2 1 2 1 2 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 2 8 6 6 5 5 6 1 2 2 2 2 2 1 2 2 2 2 1 1
0 0 0 0 0 0 0 1 1 1 1 1 1 7 8 6 8 9 6 6 6 5 2 2 2 2 2 1 1 2 2 2 1 1
0 0 0 0 0 0 0 0 1 1 1 1 7 8 8 8 8 8 7 7 7 5 2 2 2 2 2 1 1 2 2 2 1 1
0 0 0 0 0 0 0 0 1 1 1 1 8 8 8 8 7 8 8 8 7 6 6 2 2 2 2 0 1 2 2 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 8 7 7 6 7 5 8 7 7 5 6 2 1 2 2 0 1 1 2 1 1 1
0 0 0 0 0 0 0 1 1 1 1 1 8 6 8 8 8 8 7 7 7 4 6 1 1 1 2 0 2 2 2 2 1 1
0 0 0 0 0 0 0 0 0 1 1 1 8 6 8 7 8 5 8 7 6 4 6 1 1 2 2 1 2 2 2 2 1 1
0 0 0 0 0 0 0 0 0 1 1 1 8 7 6 8 8 8 7 6 7 8 6 1 1 1 1 1 1 2 1 1 1 0
0 0 0 0 0 0 0 0 0 0 0 1 7 7 6 7 8 8 7 6 6 7 6 1 1 1 1 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 7 7 2 8 8 8 8 7 7 7 5 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 6 7 6 8 8 8 7 7 6 5 4 4 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 7 6 7 7 6 6 7 6 4 4 3 4 3 3 3 3 3 2 0 0 0 0
0 0 0 0 0 3 4 4 4 4 5 4 4 4 5 7 6 6 7 1 3 3 3 3 3 3 3 3 1 1 2 3 0 0
0 0 0 2 4 4 4 4 4 4 3 3 3 2 4 7 3 1 5 2 4 4 4 2 1 3 3 2 2 3 1 3 3 0
0 0 1 3 3 0 4 4 4 2 5 5 4 4 3 2 3 7 6 8 2 0 1 0 0 0 0 0 0 0 3 1 3 0
1 2 1 3 3 1 0 0 0 0 0 0 0 7 8 7 7 8 8 7 6 5 1 2 2 2 2 2 3 3 1 2 3 2
1 2 3 2 1 2 1 0 3 3 3 3 2 7 7 7 7 8 7 6 6 5 1 2 2 2 3 3 4 3 2 3 3 1
1 2 2 1 1 2 3 3 3 3 4 3 3 7 2 6 8 2 5 7 6 2 2 2 2 2 3 1 3 3 3 3 3 2
1 2 0 2 2 2 3 4 4 4 4 3 4 3 4 3 4 5 4 4 5 1 2 2 2 2 2 2 3 4 3 3 3 2
1 2 2 3 3 2 1 3 2 3 2 1 3 3 5 3 3 5 3 4 2 1 1 2 2 2 2 2 2 3 2 3 4 2
0 1 2 1 2 3 4 2 2 1 2 0 2 3 5 3 3 5 3 4 1 1 1 1 1 1 2 2 2 3 2 4 2 2
1 1 3 2 2 2 2 2 2 3 2 3 2 2 4 3 3 5 4 3 1 1 1 1 1 2 2 2 3 3 2 2 2 2
1 1 1 2 2 2 2 3 2 2 1 2 1 1 4 3 3 4 3 3 1 1 1 1 1 2 2 2 4 2 2 2 2 2
1 1 1 1 1 1 2 3 2 2 3 1 2 1 4 3 3 3 3 3 1 1 1 0 0 1 1 2 3 3 2 2 2 2
1 1 1 1 1 1 2 1 2 0 0 0 0 0 3 3 3 3 3 0 0 1 1 0 0 1 2 3 3 3 3 2 2 2
3 3 2 2 2 2 2 2 2 0 0 0 0 0 3 3 4 3 0 0 0 0 1 1 3 3 3 3 3 3 3 3 2 2
4 3 3 3 3 3 3 3 4 4 3 3 3 3 3 3 4 3 3 2 3 3 3 3 3 3 3 3 3 3 4 4 4 4
4 3 3 3 3 4 4 4 4 5 4 5 5 1 3 5 2 4 3 3 3 3 2 4 3 5 4 4 4 4 4 4 4 4
3 3 3 3 3 4 4 4 4 5 5 5 5 5 3 4 0 4 3 3 5 5 5 5 5 4 4 4 4 4 4 4 4 4
3 3 3 3 4 4 4 4 4 5 5 5 5 5 4 4 1 4 3 5 5 5 5 5 4 4 4 4 4 4 4 4 4 3
3 3 3 4 4 4 4 4 4 4 5 5 5 5 2 2 3 0 5 5 5 4 5 4 4 4 4 4 4 4 4 4 4 3
3 3 3 4 4 4 4 4 4 4 4 4 5 4 4 4 5 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3
3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3
3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3
2 3 3 3 3 3 3 3 3 3 3 3 4 3 3 4 4 4 4 4 4 4 4 4 4 3 4 4 3 3 3 3 3 3

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0
0 0 0 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 2 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 0 2 2 2 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 0 6 7 6 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 5 7 5 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 6 7 6 5 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 7 5 4 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 2 4 7 6 5 3 5 1 1 1 2 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 8 6 6 5 4 5 1 1 1 2 1 1 1 1 2 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 7 7 5 8 8 6 6 6 4 1 1 1 1 1 1 1 2 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 7 8 7 7 8 8 7 6 6 7 5 1 1 1 1 0 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 8 8 8 8 7 8 8 8 7 6 5 1 1 1 1 0 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 1 1 1 8 6 7 5 7 5 8 7 6 4 5 1 1 1 1 0 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 8 5 8 8 7 7 6 6 6 4 5 1 1 1 1 0 1 1 1 1 1 0
0 0 0 0 0 0 0 0 0 1 1 1 8 5 8 7 8 5 7 7 6 3 6 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 7 6 5 7 8 8 7 6 6 7 5 1 1 1 1 1 1 1 1 1 1 0
0 0 0 0 0 0 0 0 0 0 0 1 7 6 5 7 8 7 7 6 6 7 6 1 1 1 1 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 7 7 2 8 7 8 8 7 7 6 4 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 5 7 6 8 8 7 7 6 6 5 4 3 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 7 6 7 6 5 5 7 6 4 4 3 3 3 3 3 2 3 2 0 0 0 0
0 0 0 0 0 3 4 3 4 4 5 3 4 4 5 7 6 5 6 1 3 3 3 3 3 3 3 2 1 1 2 3 0 0
0 0 0 2 3 3 4 4 3 3 3 3 3 2 4 7 3 0 4 2 4 4 4 1 1 3 3 1 1 3 1 3 3 0
0 0 1 3 2 0 3 4 3 2 5 4 4 4 3 2 2 6 6 7 2 0 0 0 0 0 0 0 0 0 2 1 3 0
0 1 1 3 3 1 0 0 0 0 0 0 0 7 7 7 6 7 7 7 6 5 0 2 2 2 1 2 3 3 1 2 3 1
1 2 2 1 1 2 1 0 3 3 3 3 2 7 7 7 7 8 7 5 6 5 1 2 2 2 3 3 3 2 2 2 2 1
1 2 2 1 0 2 3 2 3 3 3 3 3 6 2 5 8 2 4 7 6 2 2 2 2 2 3 1 3 3 2 3 2 1
0 2 0 2 2 1 3 4 4 4 3 3 3 2 4 3 4 5 4 3 5 1 2 2 2 2 2 2 2 3 3 2 2 1
0 1 1 2 3 2 1 2 1 3 2 1 3 2 5 3 3 5 3 4 2 1 1 2 2 2 1 2 2 3 2 3 3 2
0 1 2 1 2 2 3 2 1 1 2 0 2 3 5 3 3 5 3 3 1 1 1 1 1 1 2 2 2 3 2 4 2 2
1 0 2 2 2 2 2 2 2 2 2 2 2 2 4 3 3 5 4 3 1 1 1 1 1 2 2 2 3 3 1 2 2 2
1 0 1 1 2 2 1 2 2 2 1 2 1 1 4 3 3 4 3 3 1 1 1 1 1 1 1 2 3 2 2 2 2 2
1 1 1 1 1 1 1 3 2 2 2 1 2 1 4 2 3 3 3 3 1 1 1 0 0 1 1 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 2 0 0 0 0 0 3 3 3 3 3 0 0 1 1 0 0 1 1 2 2 2 2 2 2 2
2 2 1 1 1 1 2 1 2 0 0 0 0 0 0 3 3 4 2 0 0 0 0 1 1 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 2 2 2 3 3 3 3 4 3 2 2 2 3 3 3 3 2 3 3 2 4 4 3 4
3 3 3 3 3 3 3 3 4 4 4 4 4 1 3 4 2 4 2 3 3 3 1 4 2 4 4 4 4 4 4 3 4 3
3 3 3 3 3 3 3 3 4 4 4 5 4 5 3 4 0 3 3 3 4 4 4 4 4 4 4 4 4 3 4 3 3 3
3 3 3 3 3 3 4 4 4 4 4 4 5 4 3 4 0 4 3 5 4 4 4 4 4 4 4 4 4 4 3 3 3 3
3 3 3 3 3 3 4 4 4 4 4 4 4 4 2 2 3 0 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3
3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3 3 3 3
2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 3 4 4 3 3 3 3 3 3 3 3 3 3
2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2

Recognition, Reconstruction & Reorganization



The Three R's of Vision



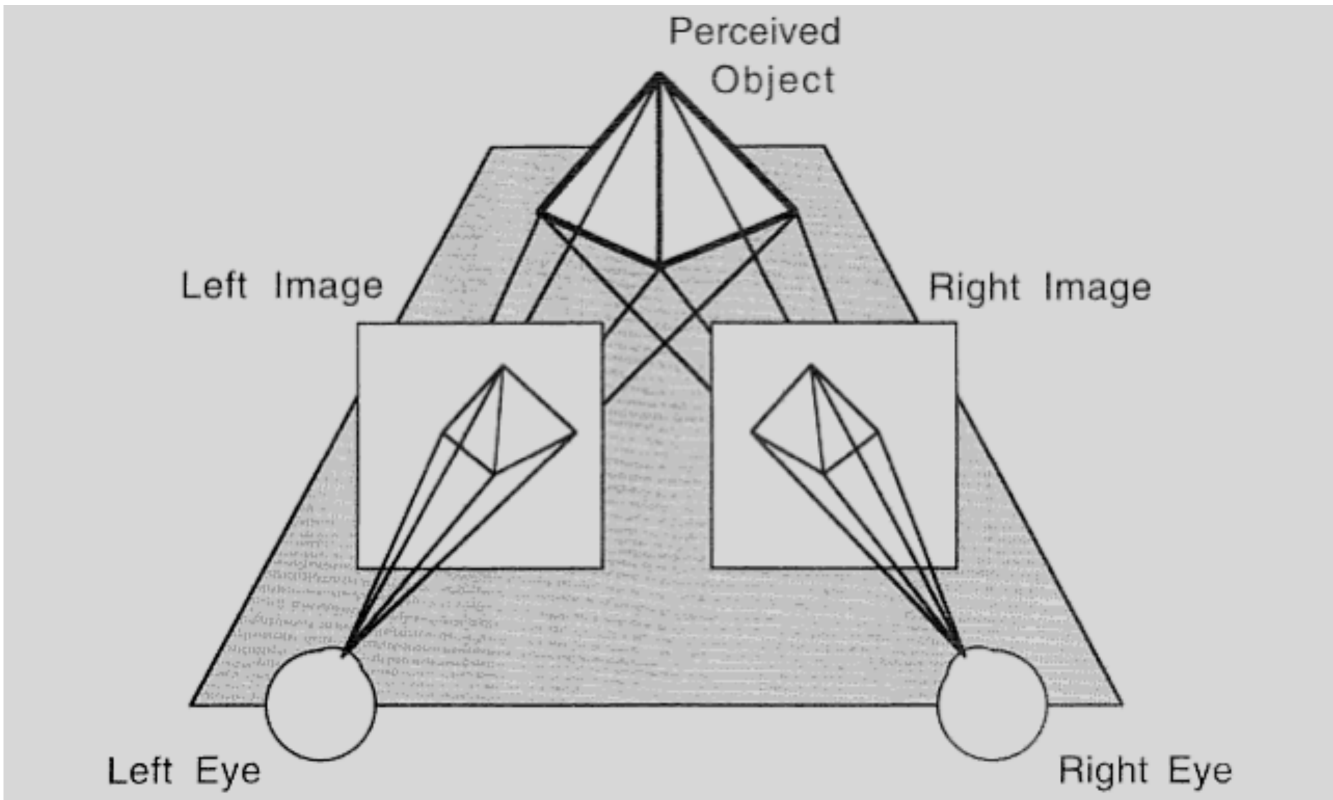
Each of the 6 directed arcs in this diagram is a useful direction of information flow

Some problems that we can solve…

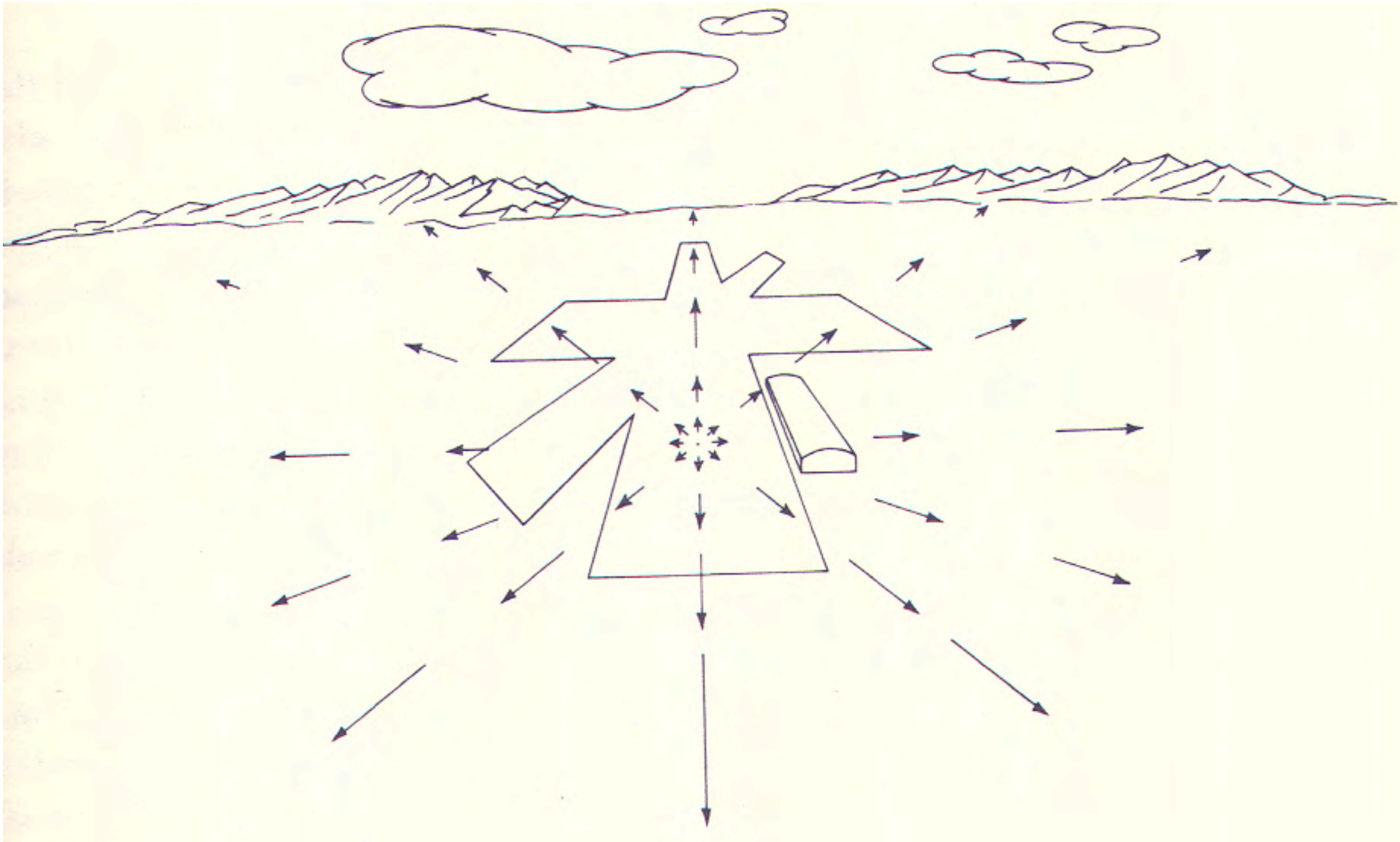
Six decades of computer vision

- 1960s: Beginnings in artificial intelligence, image processing and pattern recognition
- 1970s: Foundational work on image formation: Horn, Koenderink, Longuet-Higgins …
- 1980s: Vision as applied mathematics: geometry, multi-scale analysis, probabilistic modeling, control theory, optimization
- 1990s: Geometric analysis largely completed, vision meets graphics, statistical learning approaches resurface
- 2000s: Significant advances in visual recognition
- 2010s: Progress continues, aided by the availability of large amounts of visual data and massive computing power. Deep learning has become pre-eminent

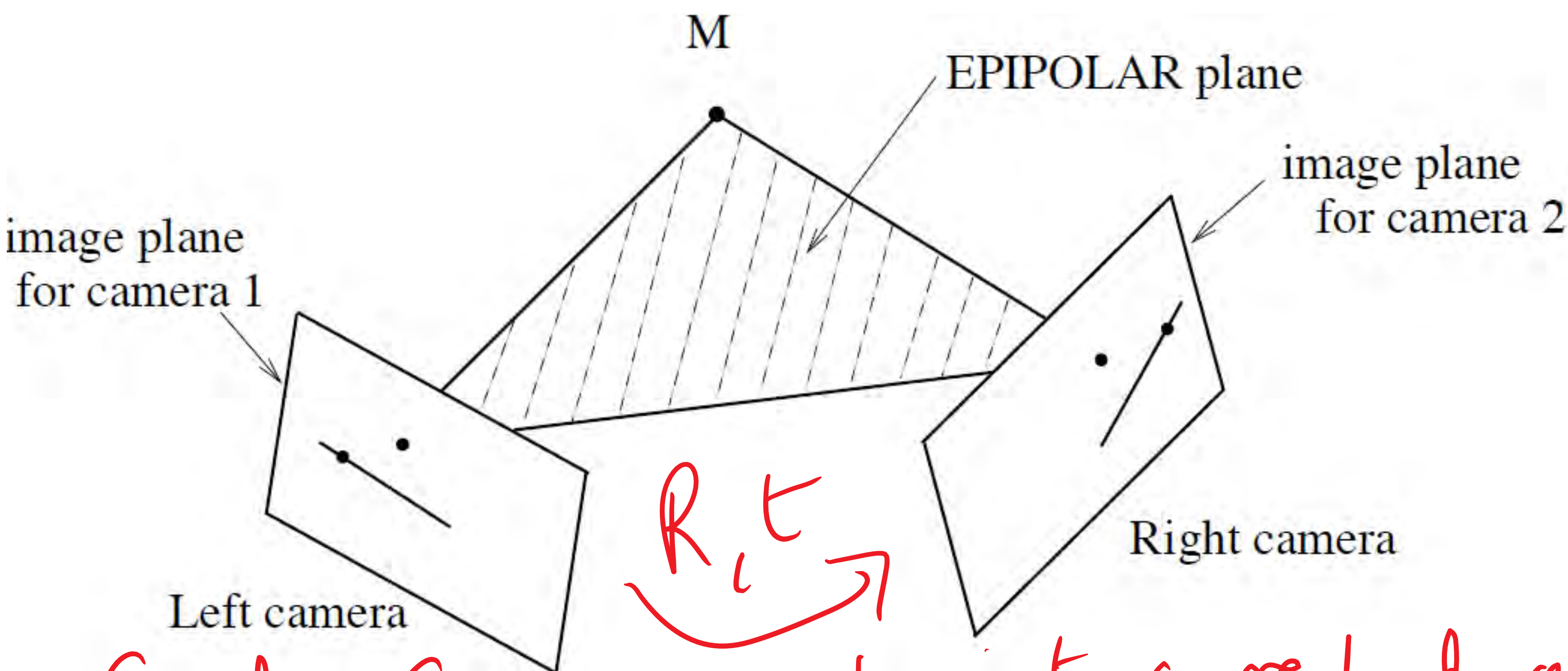
Binocular Stereopsis



Optical flow is a basic cue for all animals

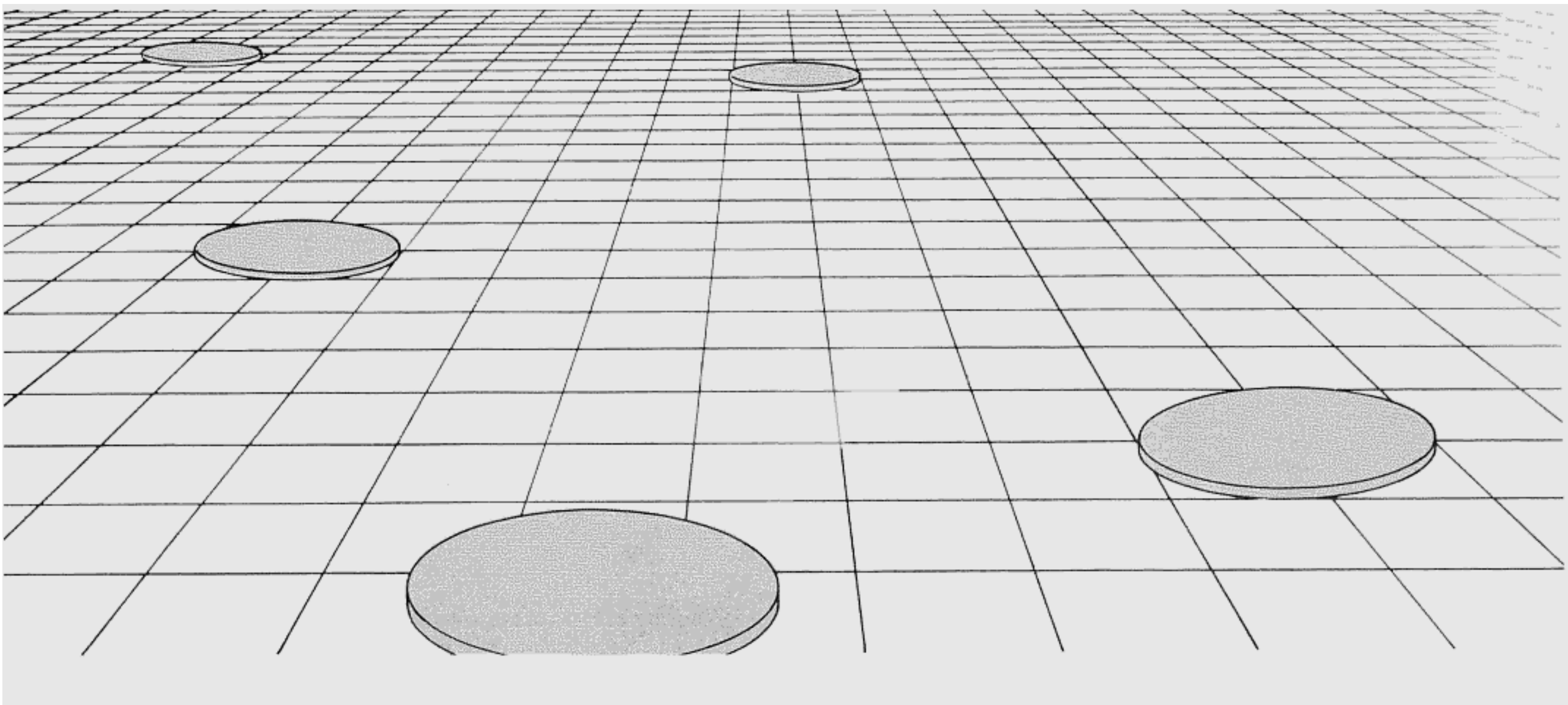


Epipolar geometry for cameras in general position

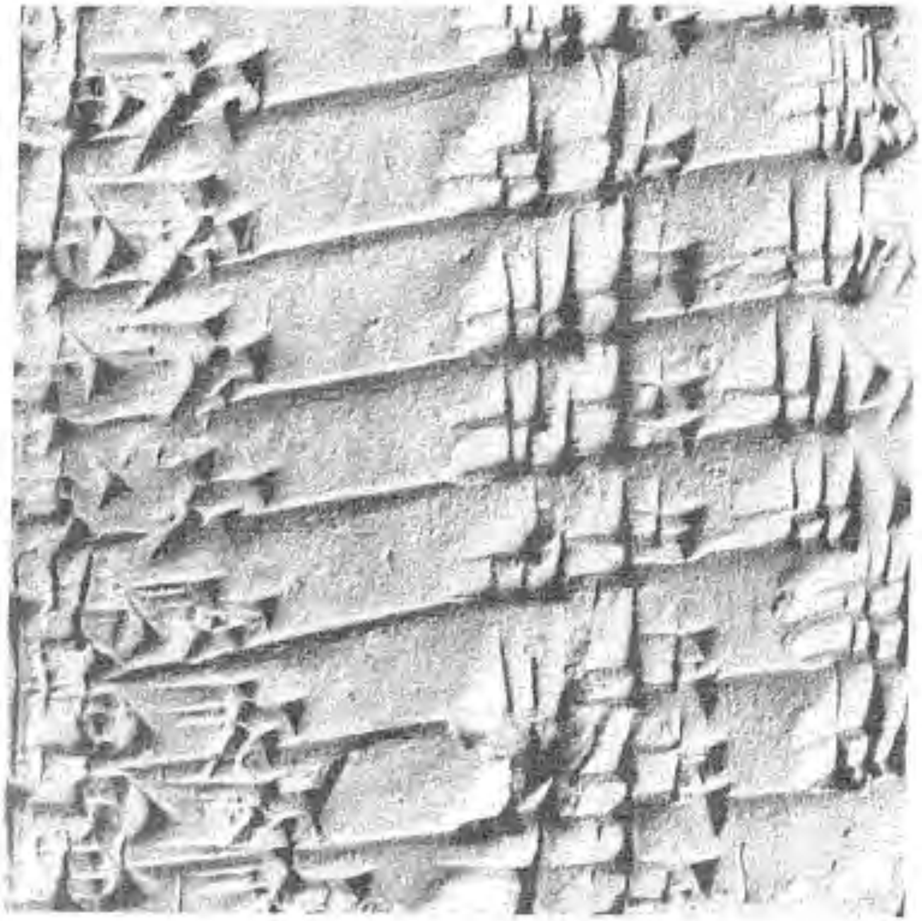


Goal: Given n point correspondences, estimate R, t and depths at the n points

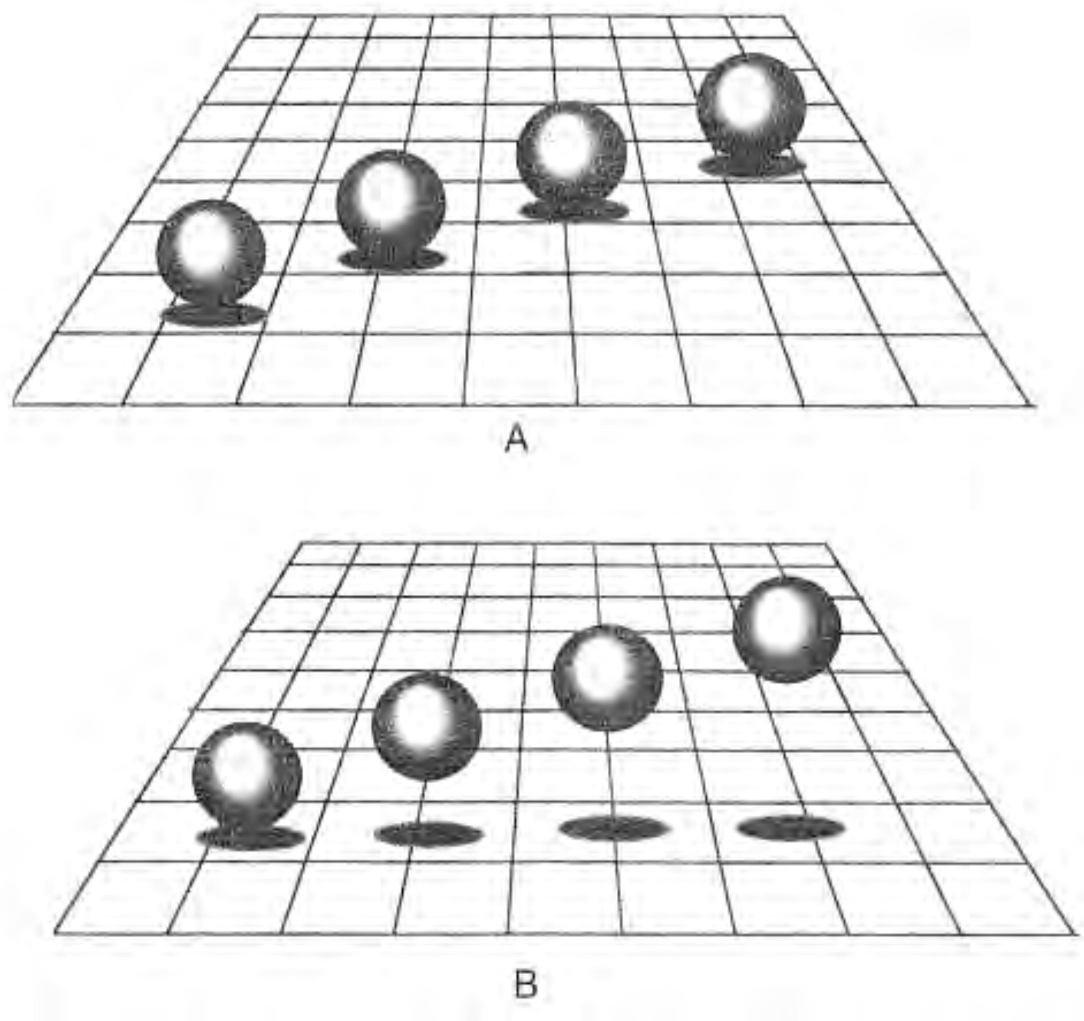
Some Pictorial Cues



Shading



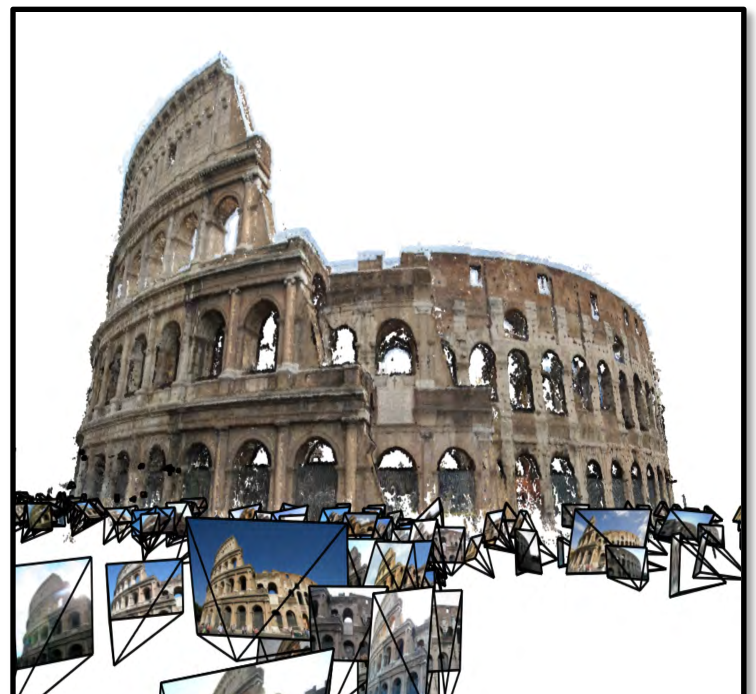
Cast Shadows



Reconstructing the world

Over the past 10 years, 3D modeling from images has made huge advances in scale, quality, and generality. We can reconstruct scenes…

… automatically from huge collections of photos downloaded from the Internet



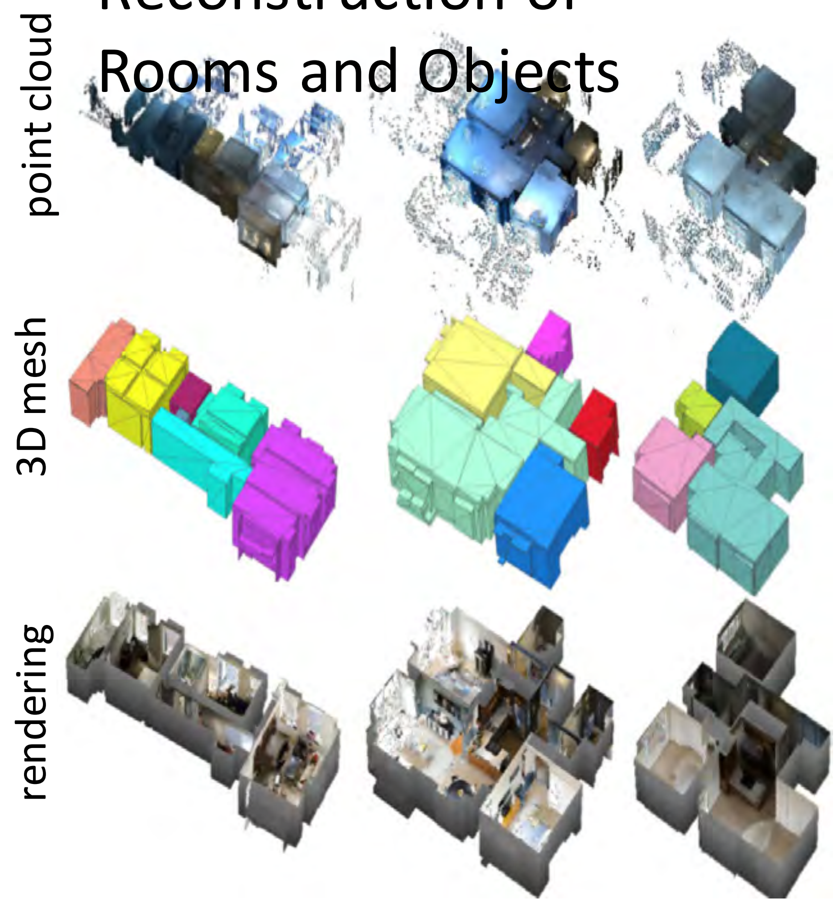
Snavely, Seitz, Szeliski.
Reconstructing the World from Internet Photo Collections.

Reconstructing the great indoors…

… using Depth Cameras

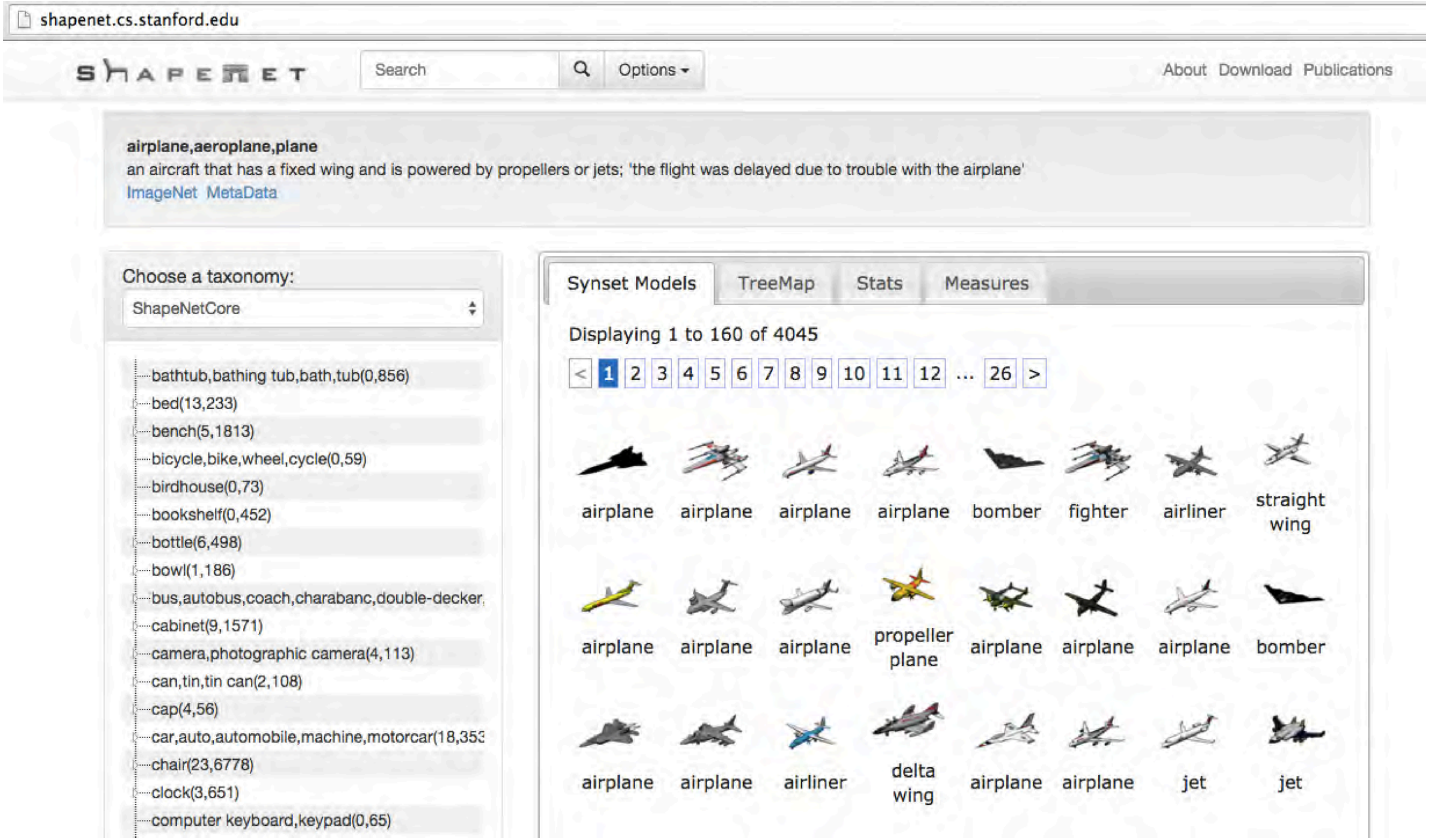
Choi, Zhou, Koltun.
Robust Reconstruction of Indoor Scenes.
CVPR 2015

… using Semantic Reconstruction of Rooms and Objects

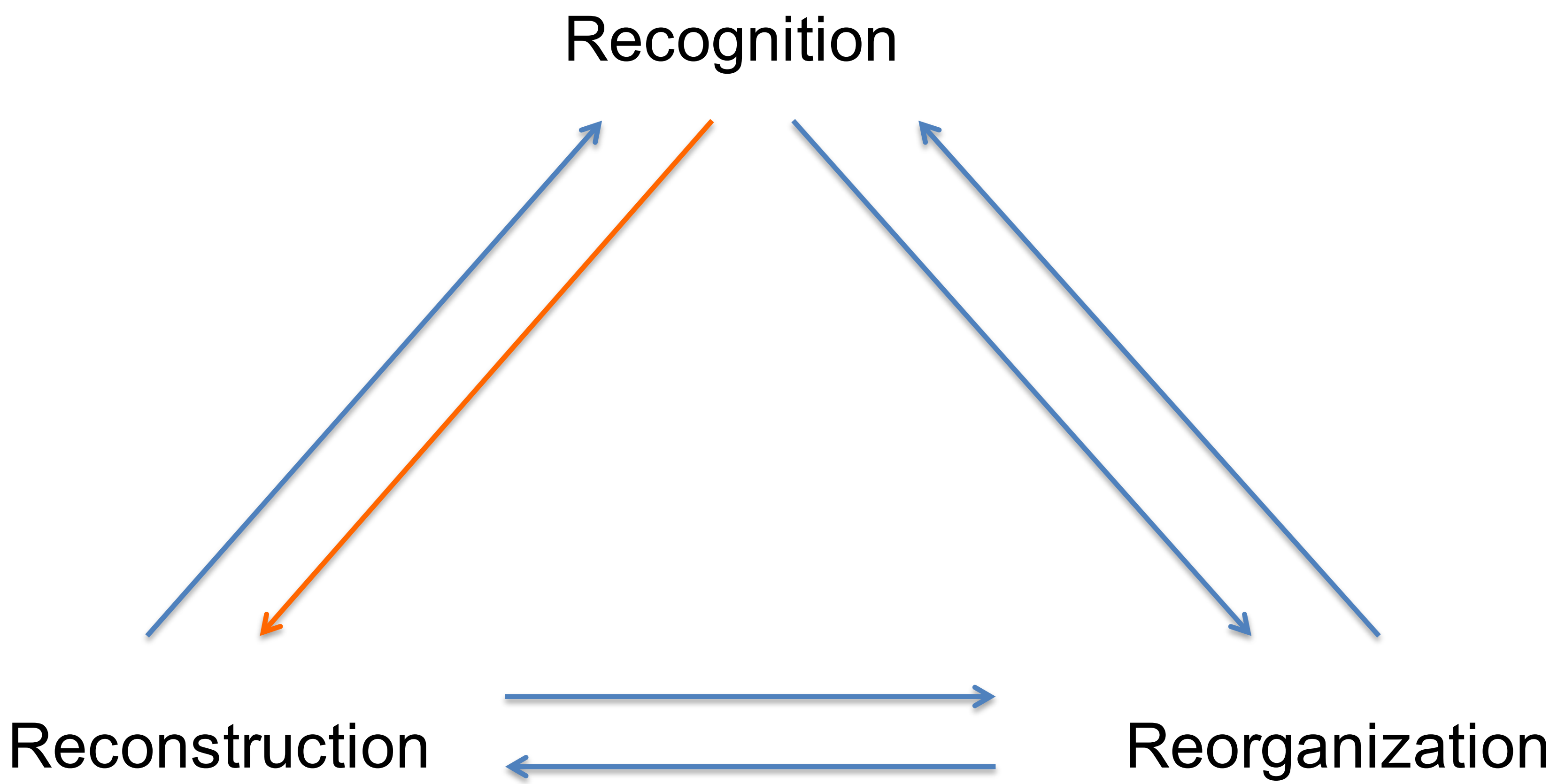


Ikehata, Yan, Furukawa.
Structured Indoor Modeling.
ICCV 2015

ShapeNet (Stanford & Princeton)

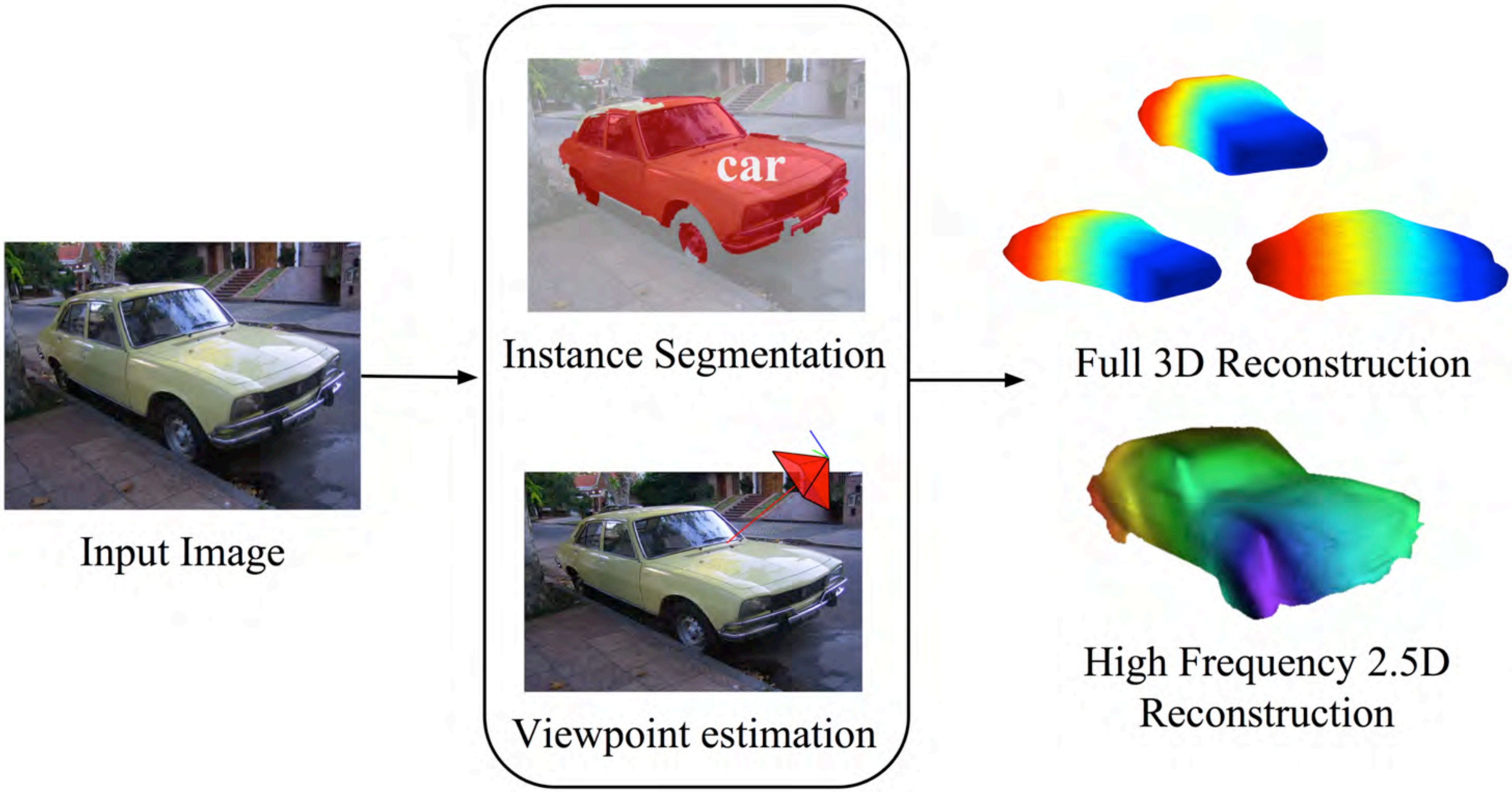


The Three R's of Vision

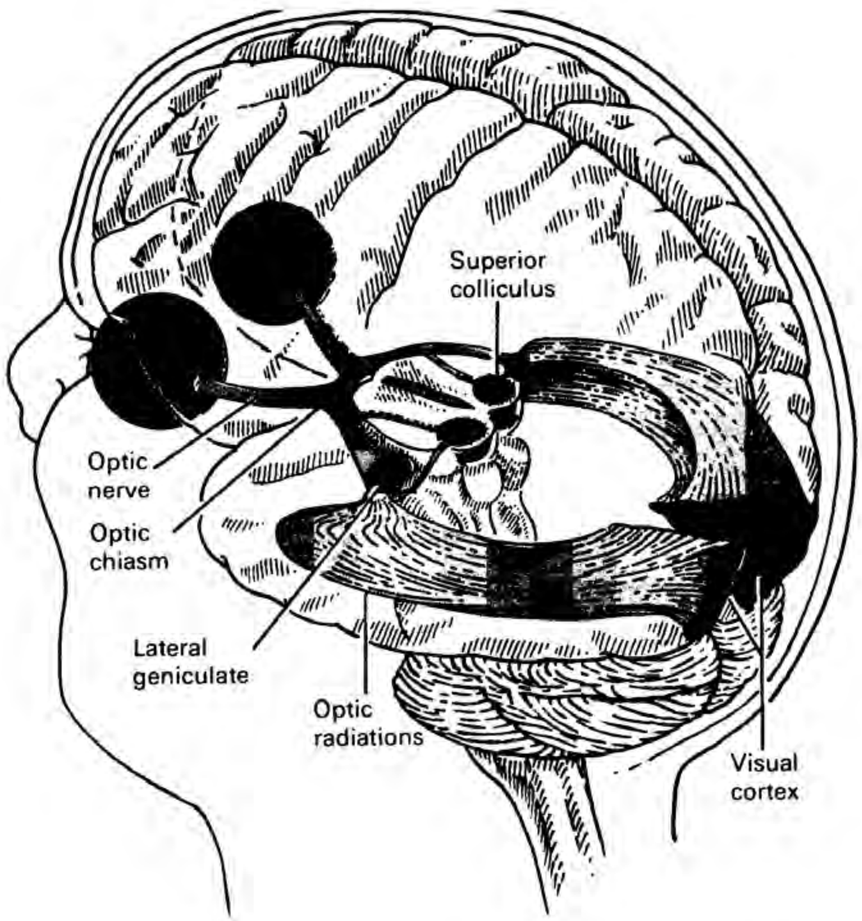


We have explored category-specific 3D reconstruction.

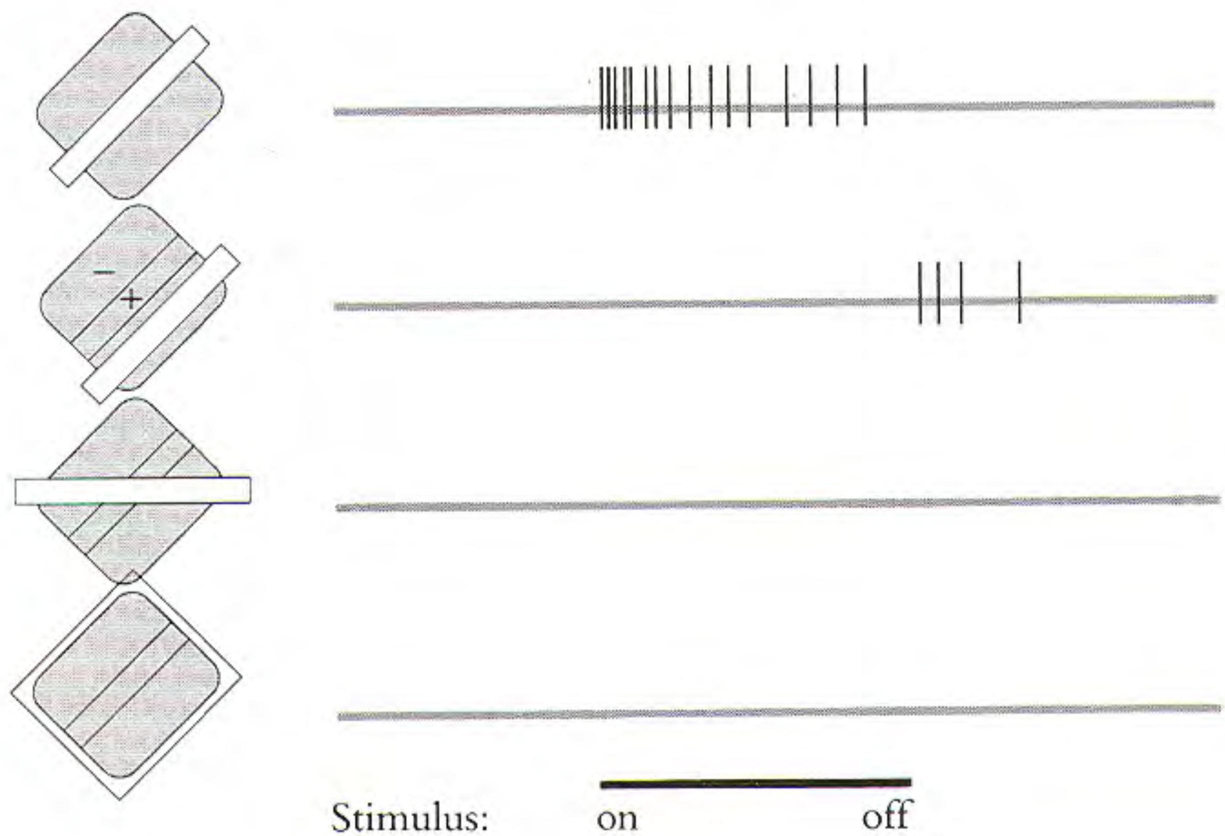
Category Specific Object Reconstruction
Kar, Tulisiani, Carreira & Malik



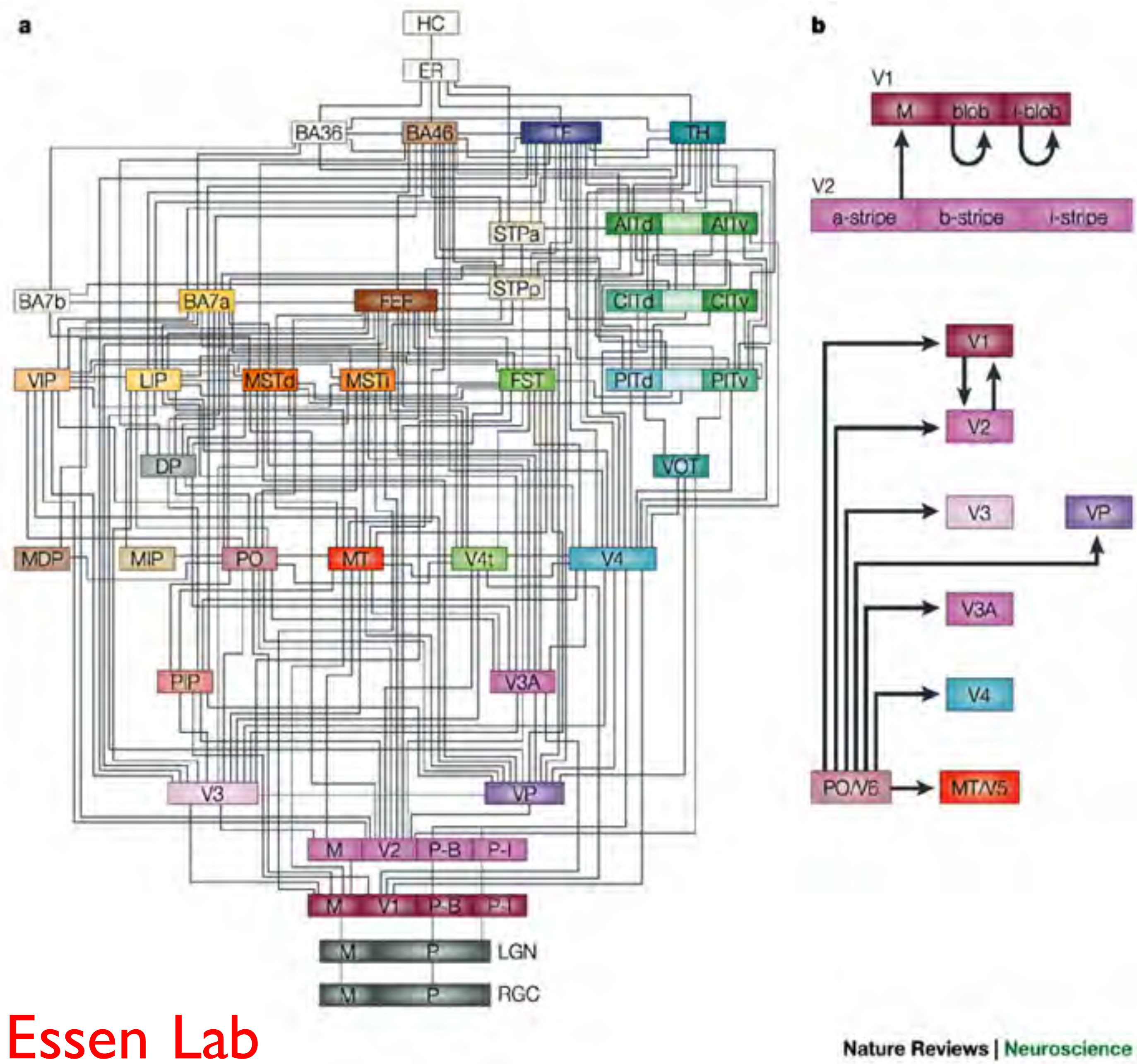
The Visual Pathway



Hubel and Wiesel (1962) discovered orientation sensitive neurons in V1

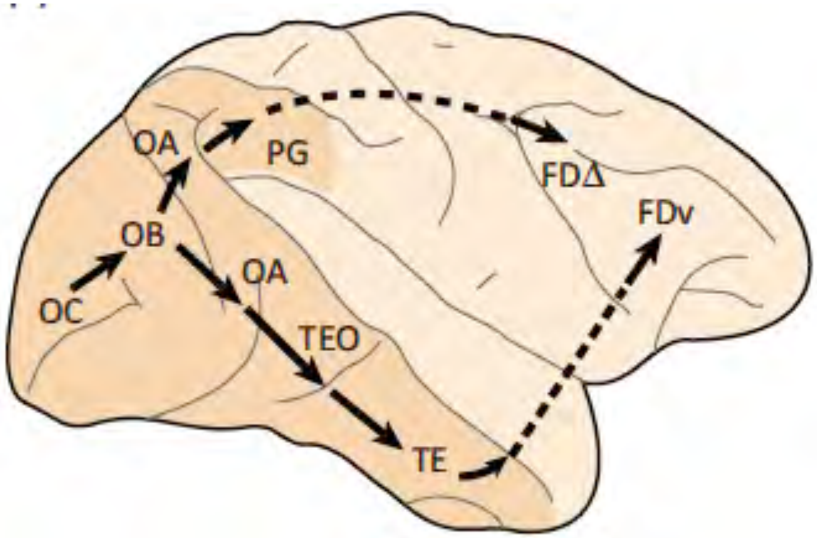


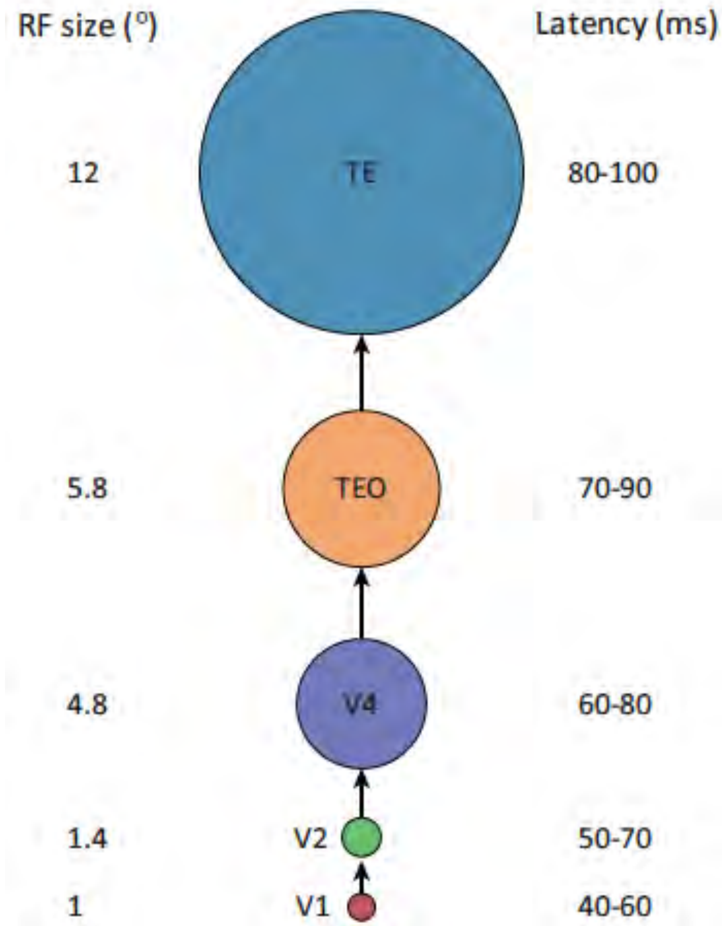
Block Diagram of the Primate Visual System



D. Van Essen Lab

Feed-forward model of the ventral stream





Kravitz et al, Trends in Cognitive Science 2013

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

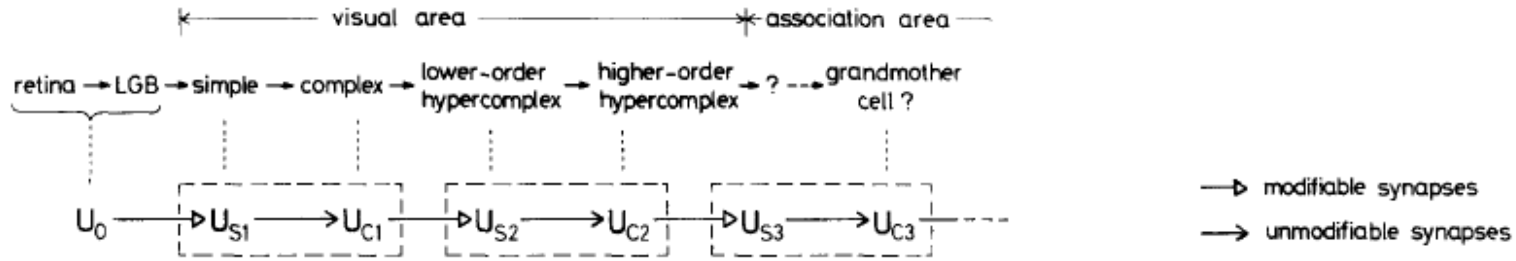


Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

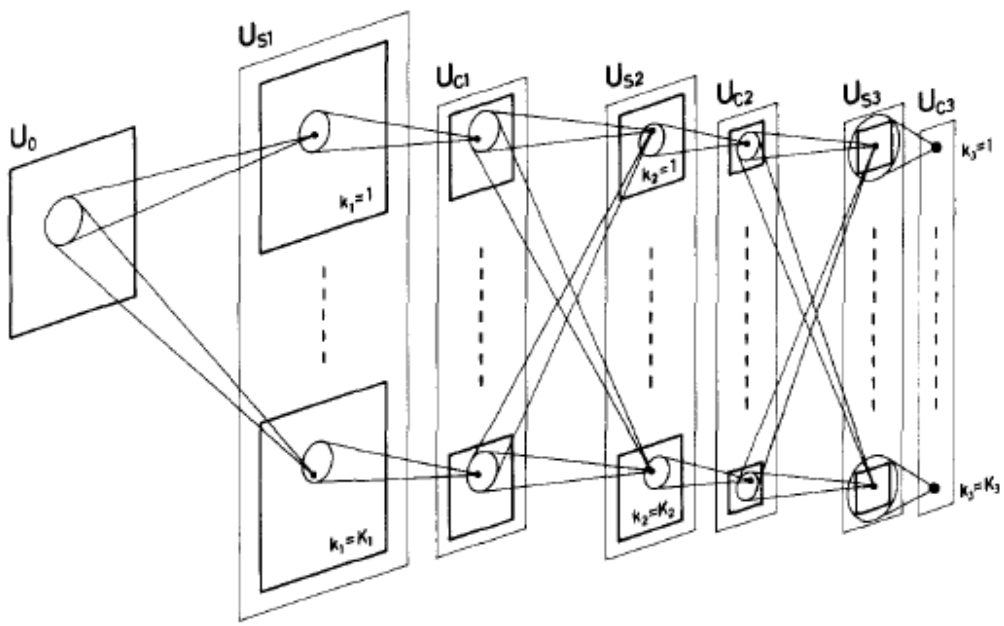


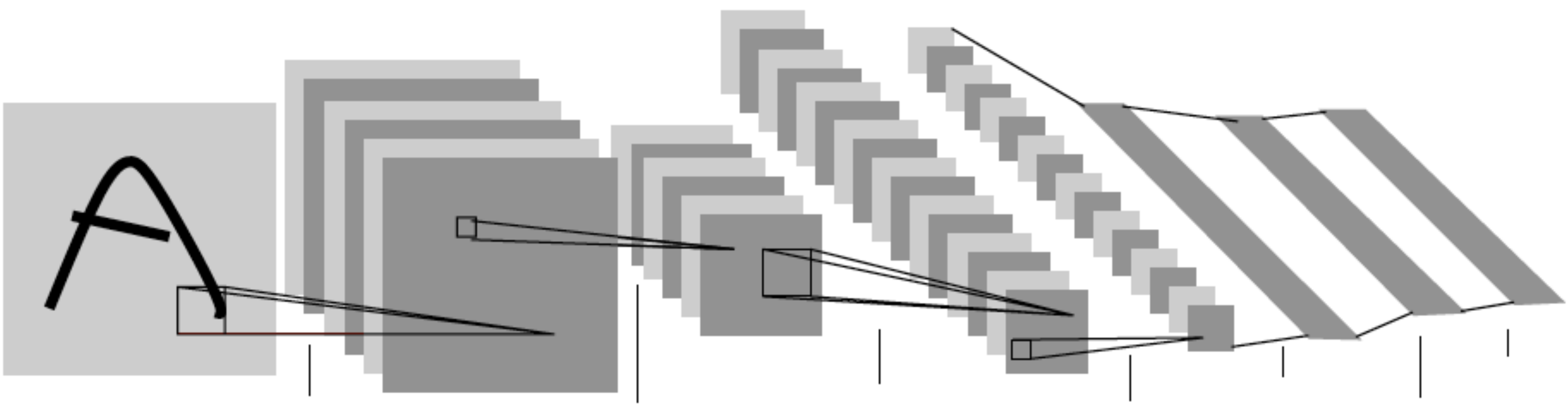
Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Biol. Cybernetics 36, 193–202 (1980)

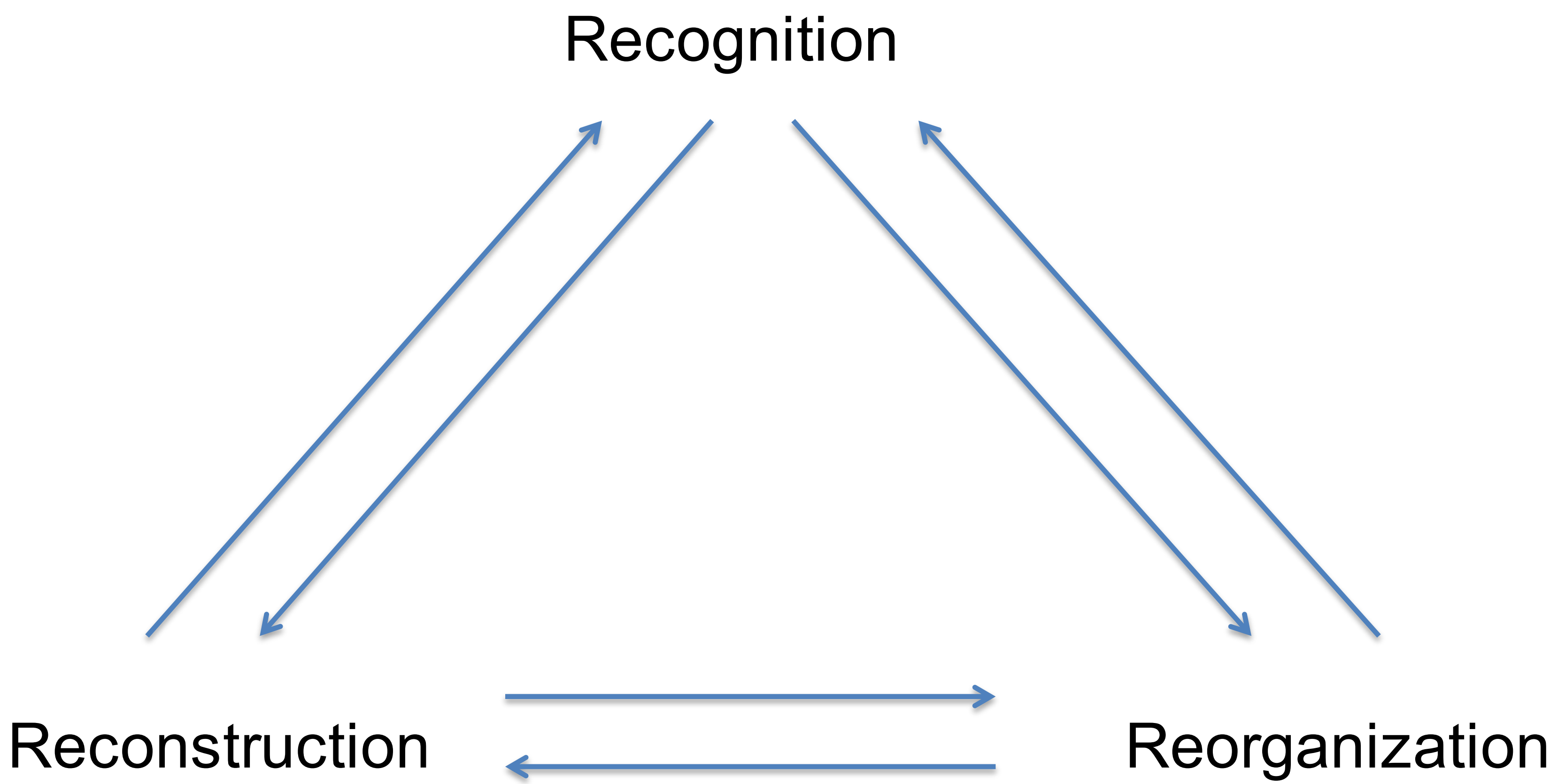
Convolutional Neural Networks (LeCun et al)

Used backpropagation to train the weights in this architecture

- First demonstrated by LeCun et al for handwritten digit recognition(1989)
- Applied in sliding window paradigm for tasks such as face detection in the 1990s.
- However was not competitive on standard computer vision object detection benchmarks in the 2000s.
- Krizhevsky, Sutskever & Hinton showed effectiveness for full image classification on ImageNet Challenge (2012)

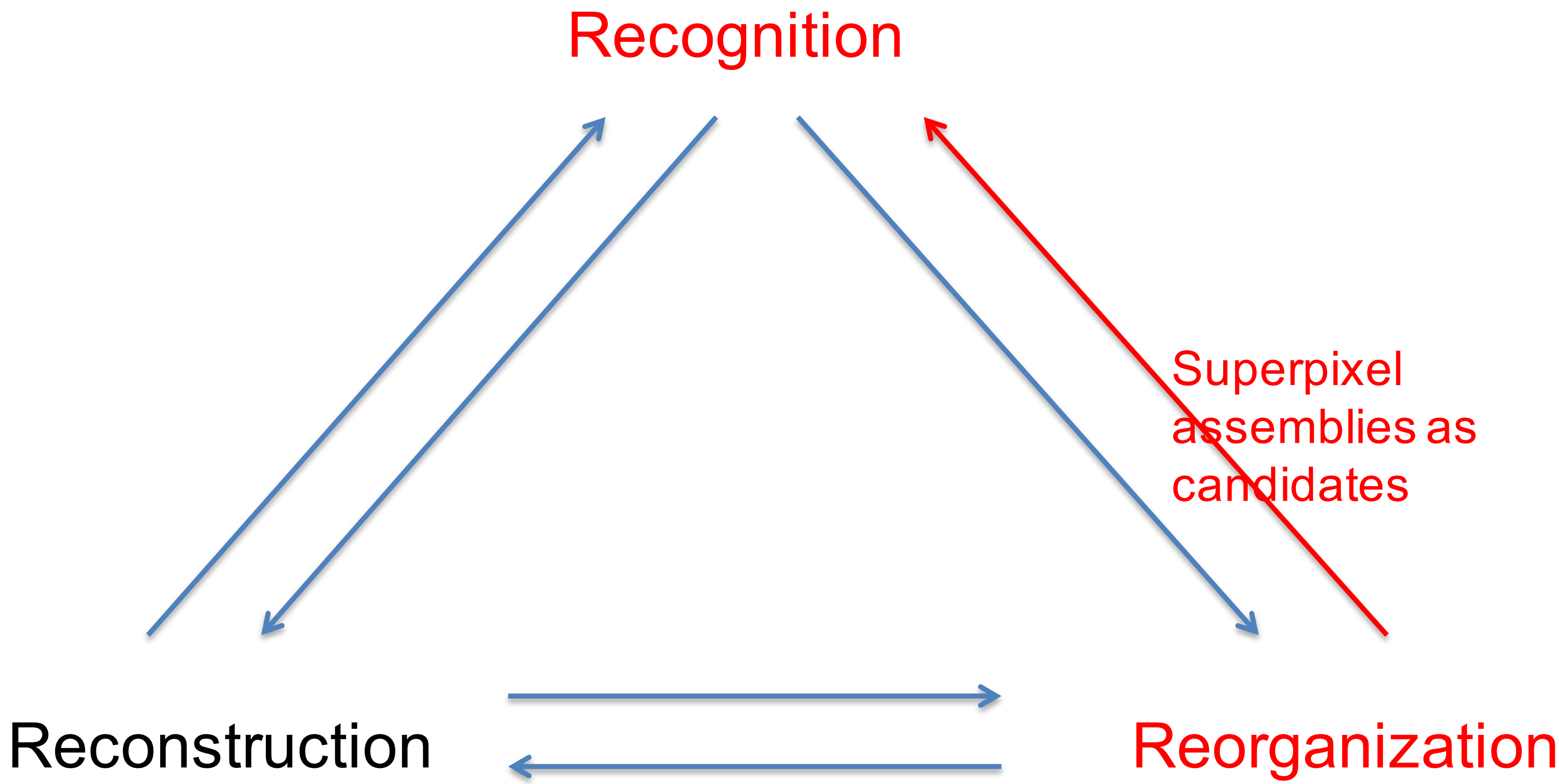


The Three R's of Vision

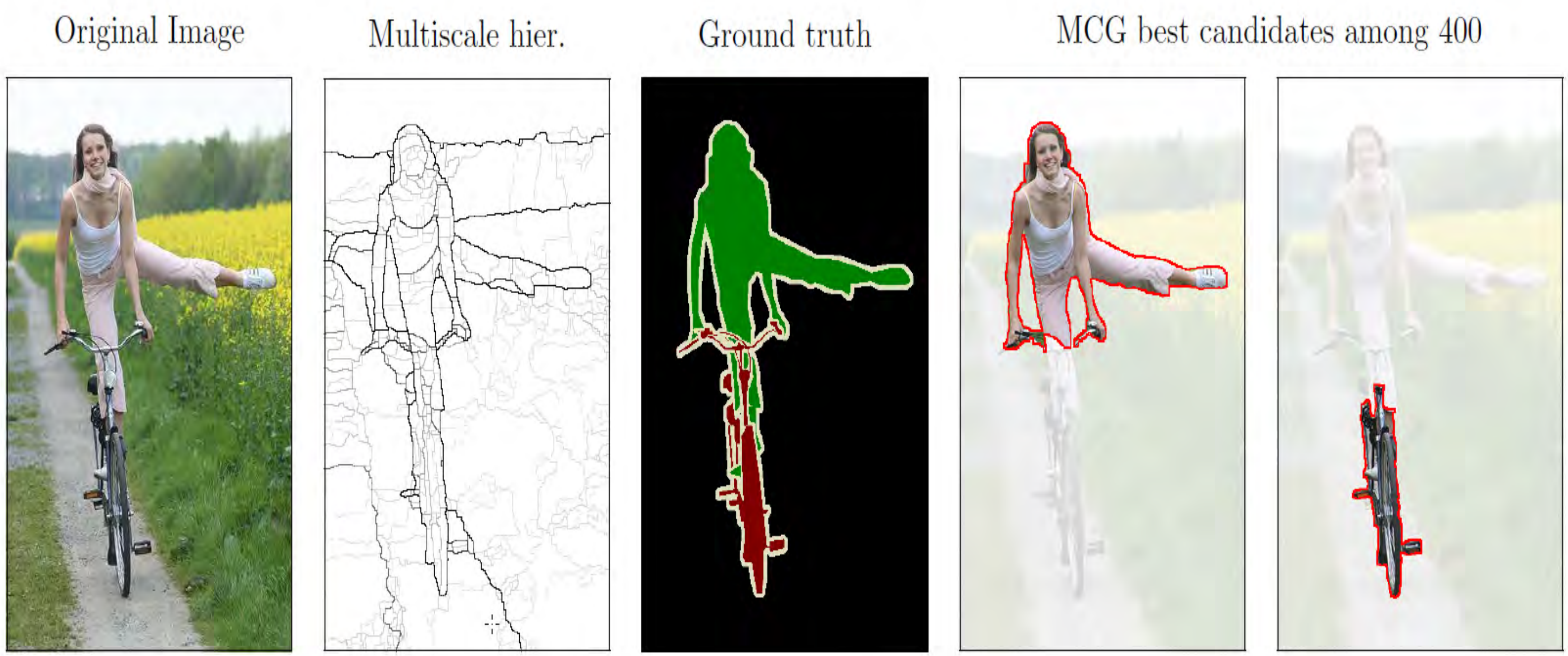


Each of the 6 directed arcs in this diagram is a useful direction of information flow

The Three R’s of Vision



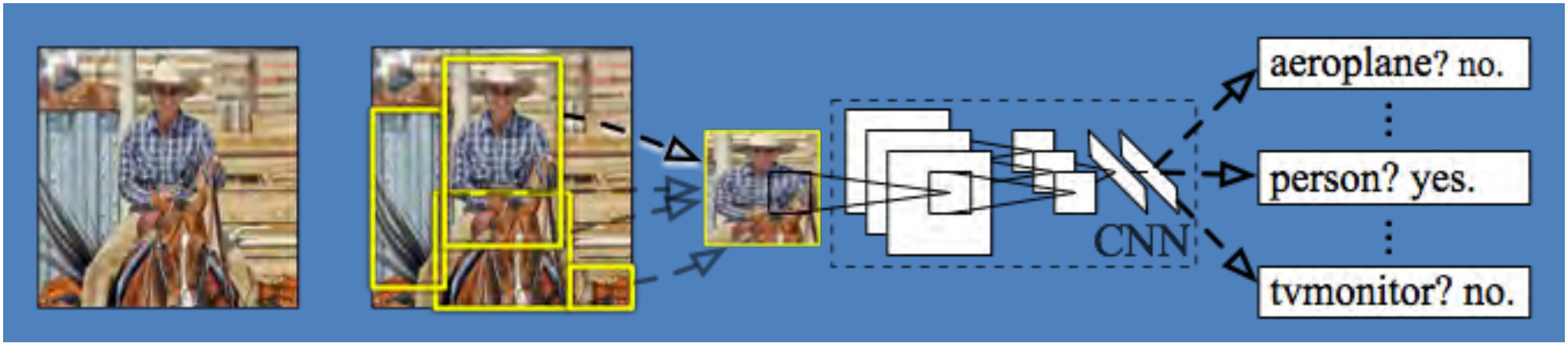
Bottom-up grouping as input to recognition



We produce superpixels of coherent color and texture first, then combine neighboring ones to generate object candidates

R-CNN: Regions with CNN features

Girshick, Donahue, Darrell & Malik (CVPR 2014)



Input image

Extract region proposals (~2k / image)

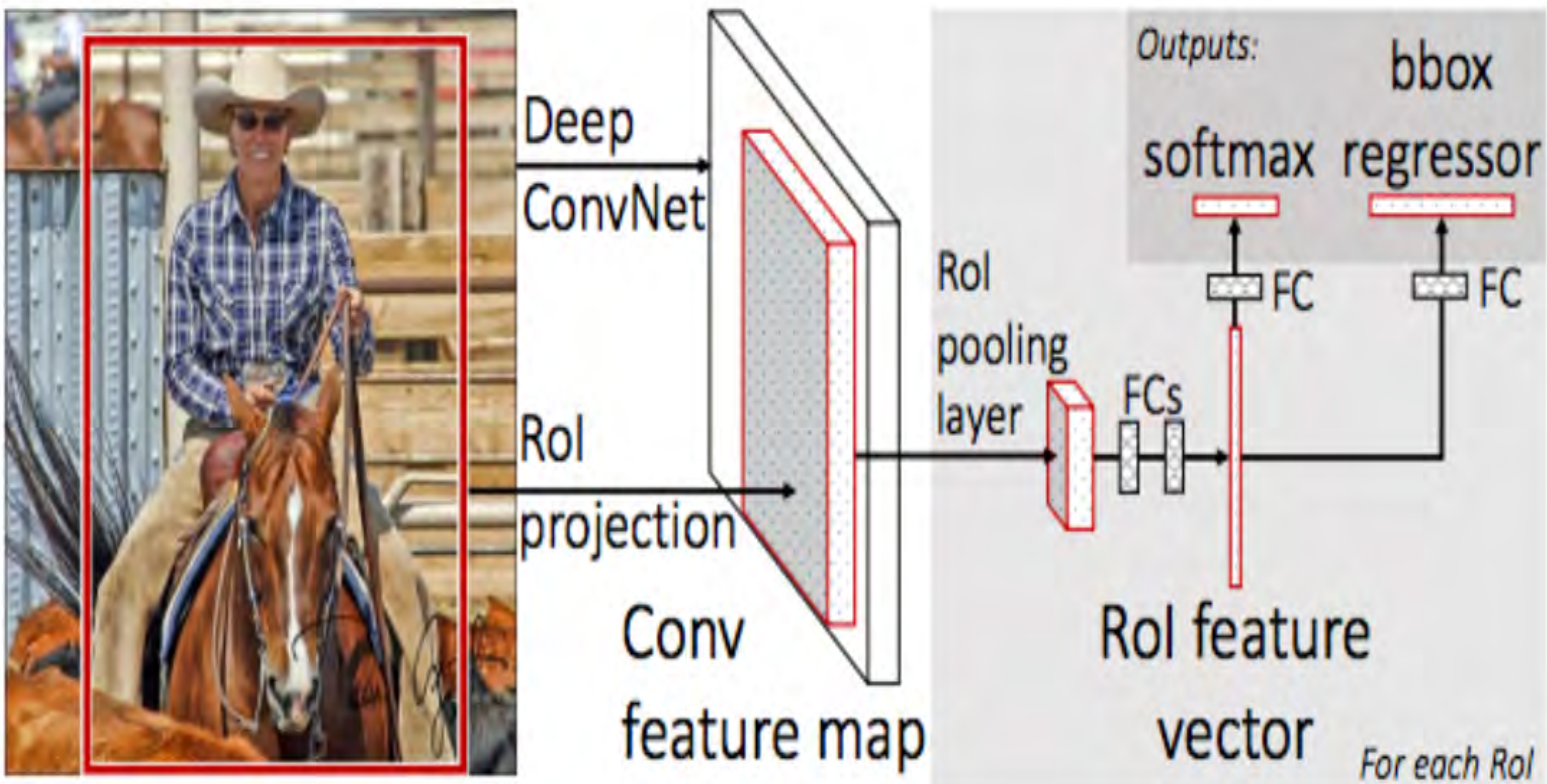
Compute CNN features

Classify regions (linear SVM)

This and the Multibox work from Google showed how to apply these architectures for object detection

Fast R-CNN (Girshick, 2015)

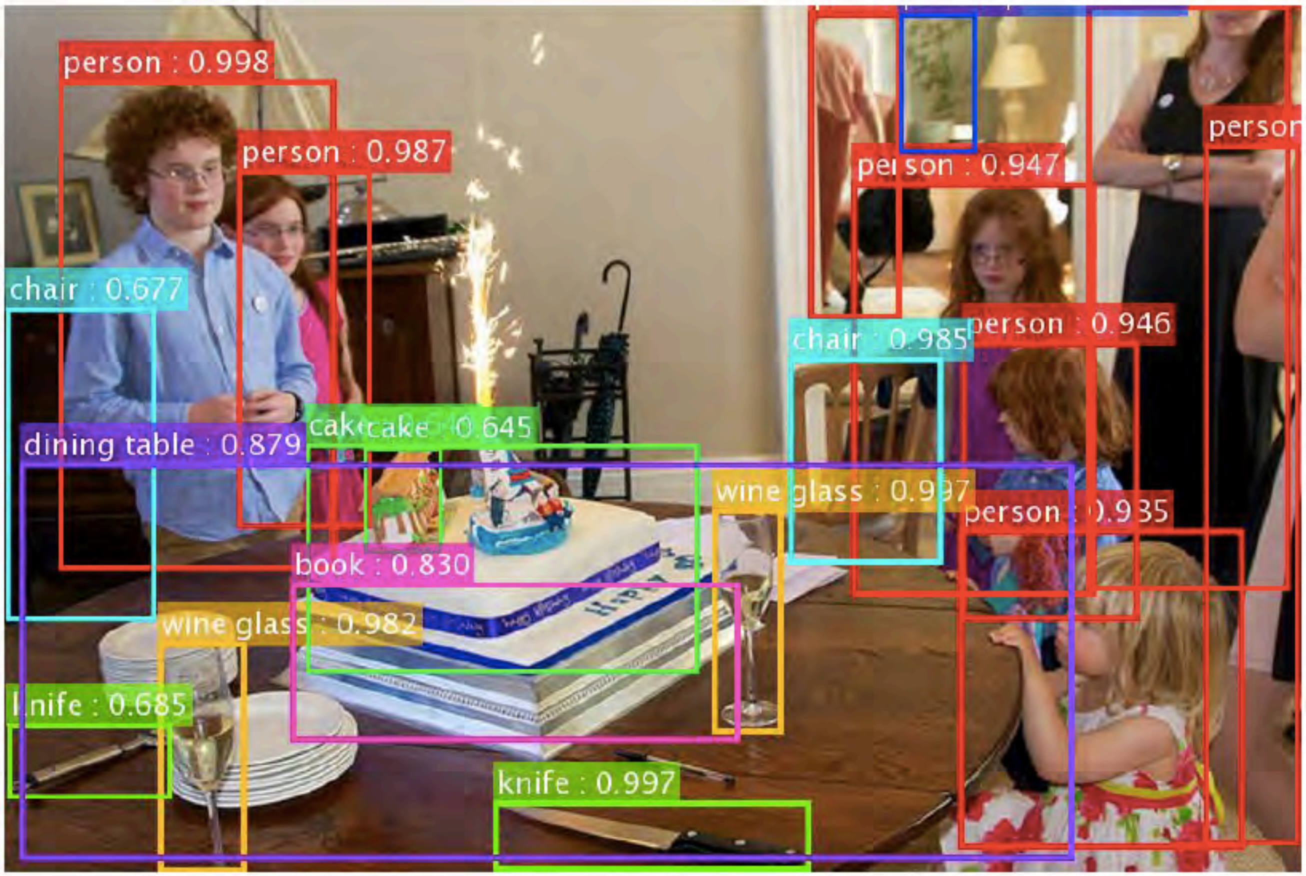
R-CNN with SPP features, no need to warp individual windows



There is also Faster R-CNN
which doesn't require external proposals

Current systems can do remarkably well in detecting objects

Microsoft Research

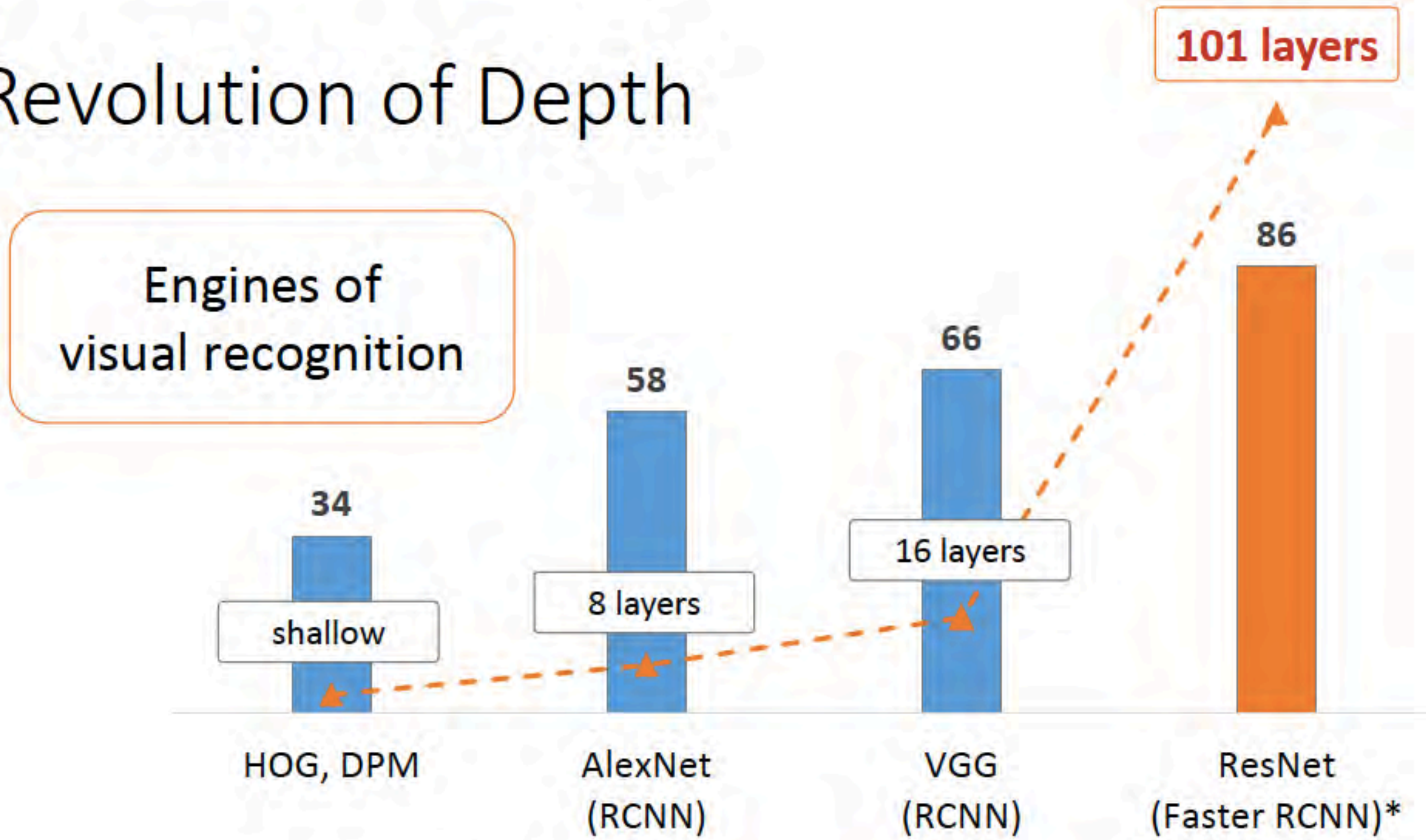


Our results on COCO – too many objects, let's check carefully!

*the original image is from the COCO dataset

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.
Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". NIPS 2015.

Revolution of Depth

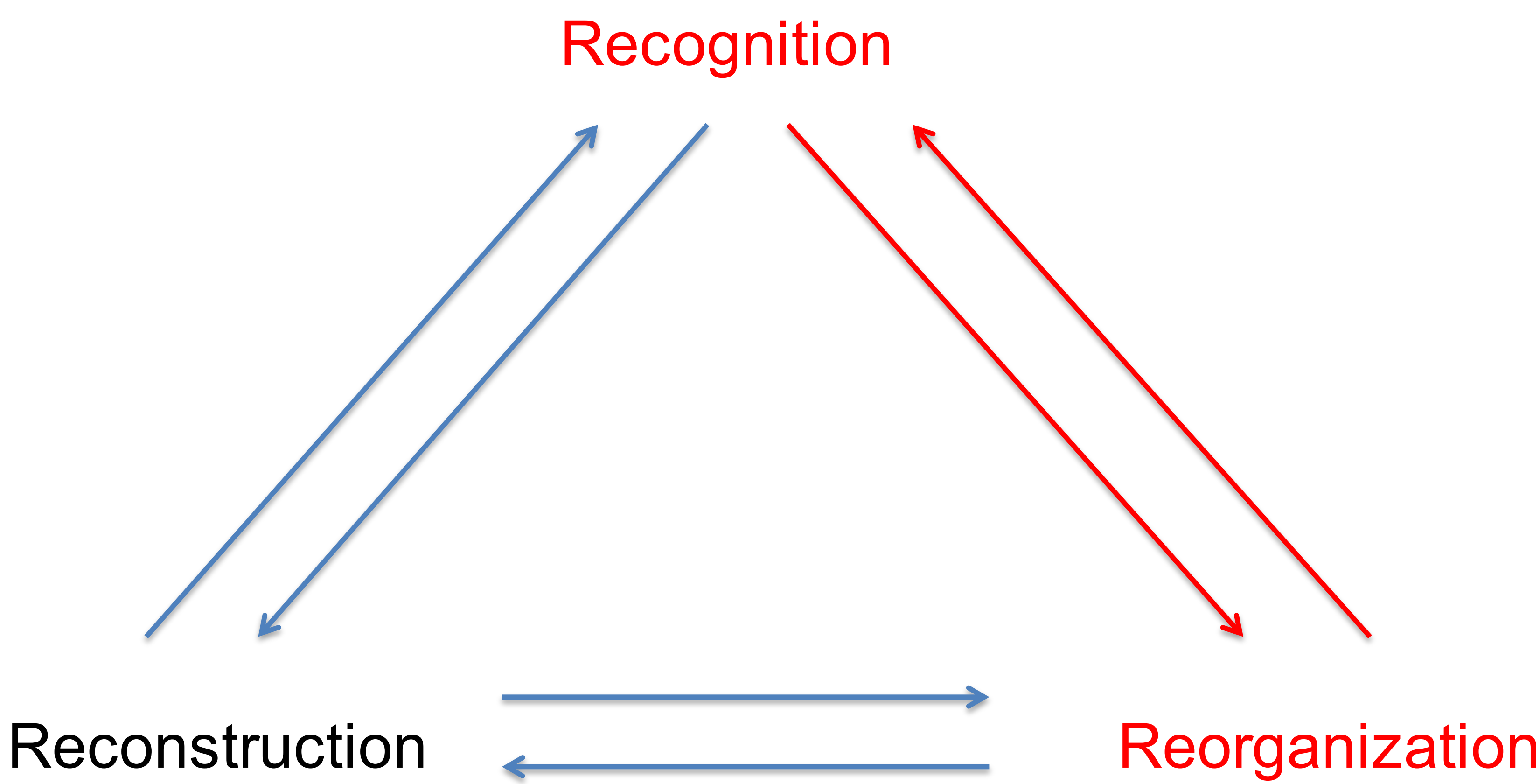


PASCAL VOC 2007 **Object Detection** mAP (%)

*w/ other improvements & more data

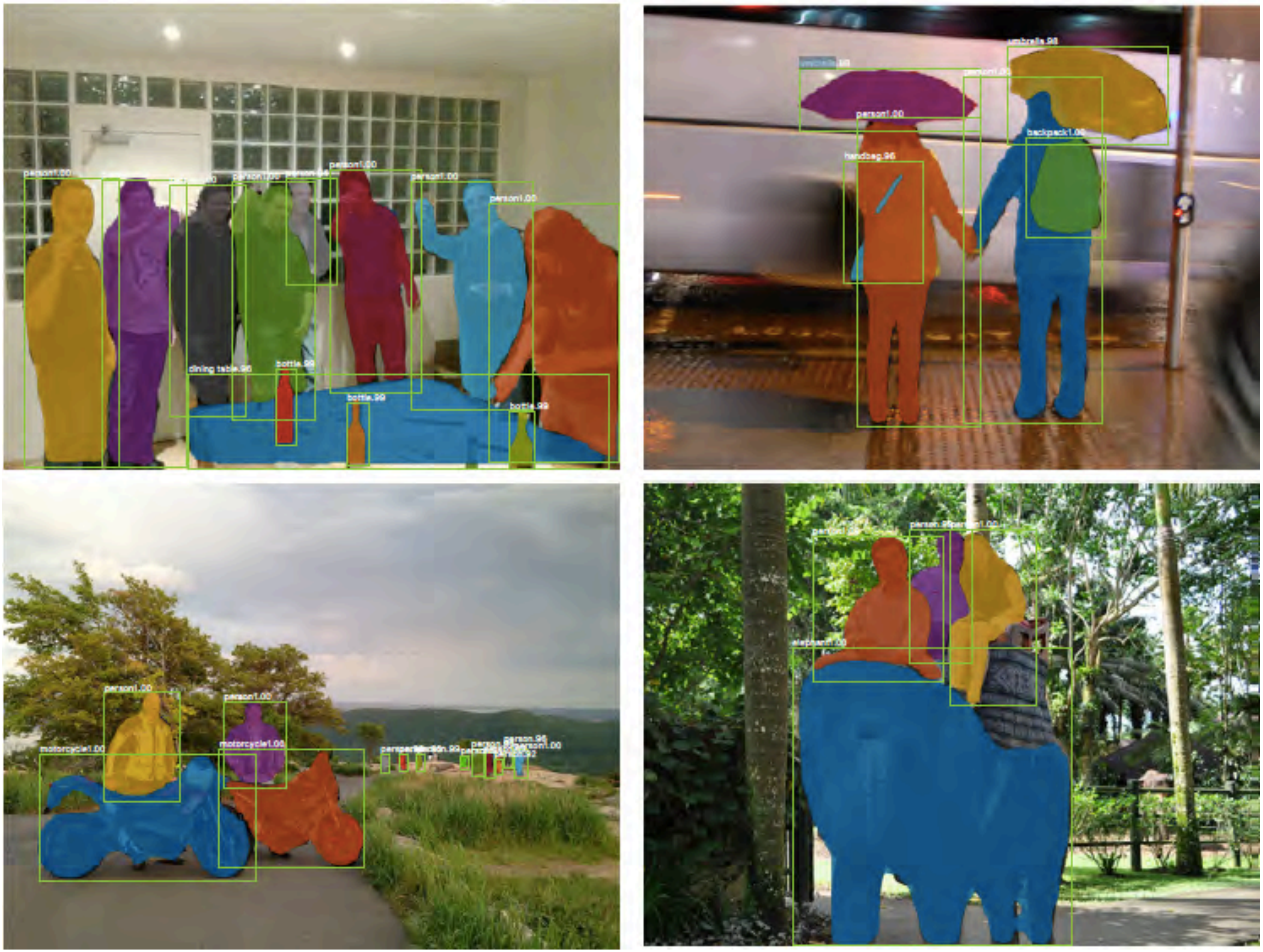
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

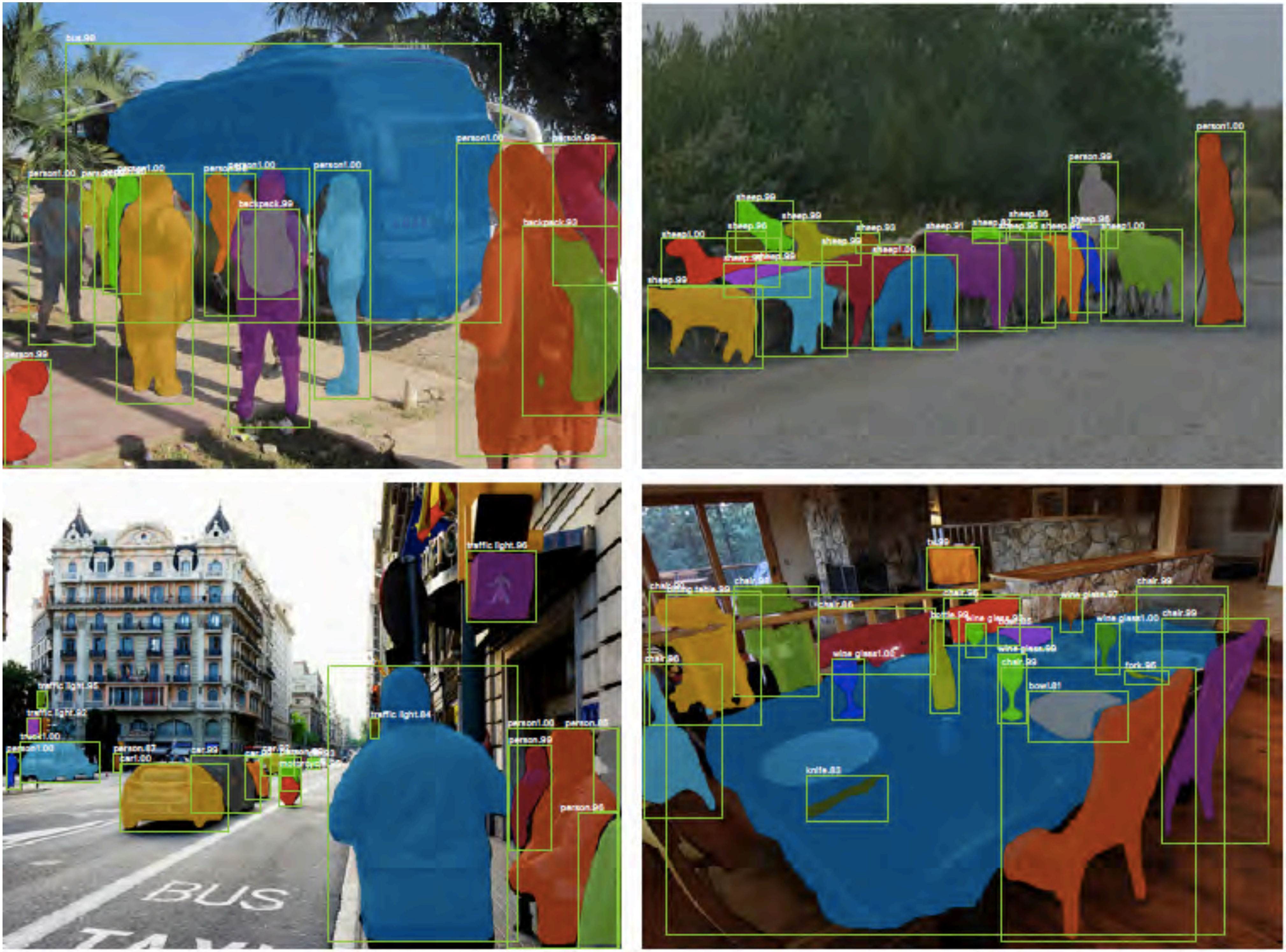
How about the other direction…



Recognition Helps Reorganization

Mask R-CNN : He, Gkioxari, Dollar & Girshick (2017)




BUS

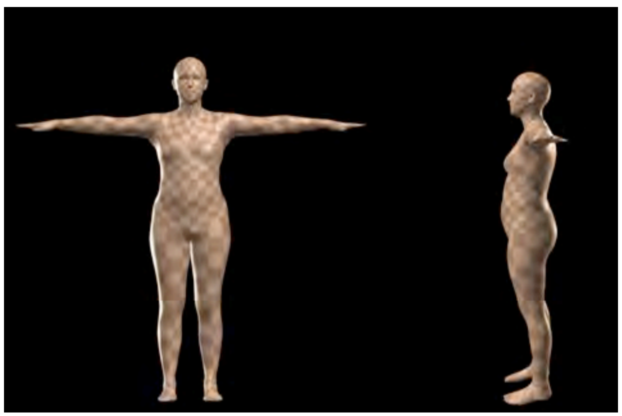
End-to-end Recovery of Human Shape and Pose

Angjoo Kanazawa, Michael Black, David Jacobs, Jitendra Malik

Our representation

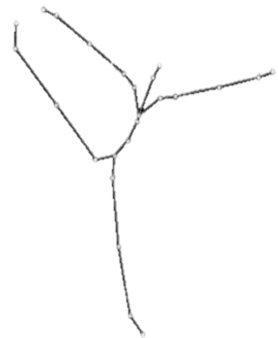
Shape: PCA coefficients



$\vec{\beta}$

$+$

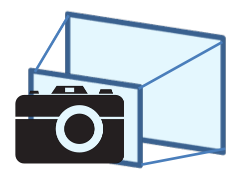
Pose: Rotation of joints



$\vec{\theta}$

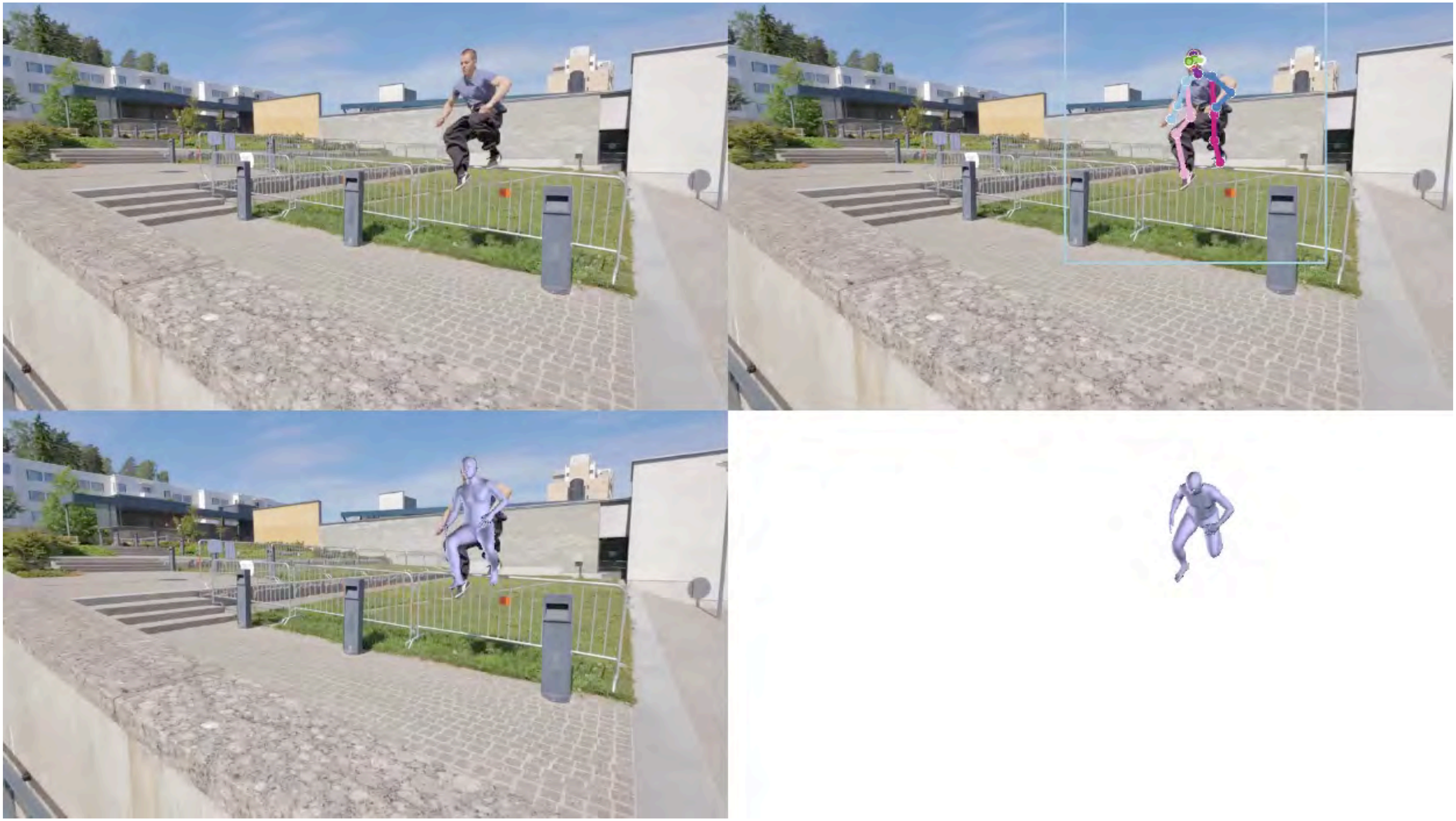
$=$

Mesh



Camera

SMPL [Loper et al. SIGGRAPH ASIA '16]



R*CNN (Gkioxari et al)

R*CNN on PASCAL VOC Action

Visual Semantic Role Labeling
Gupta & Malik (2015)

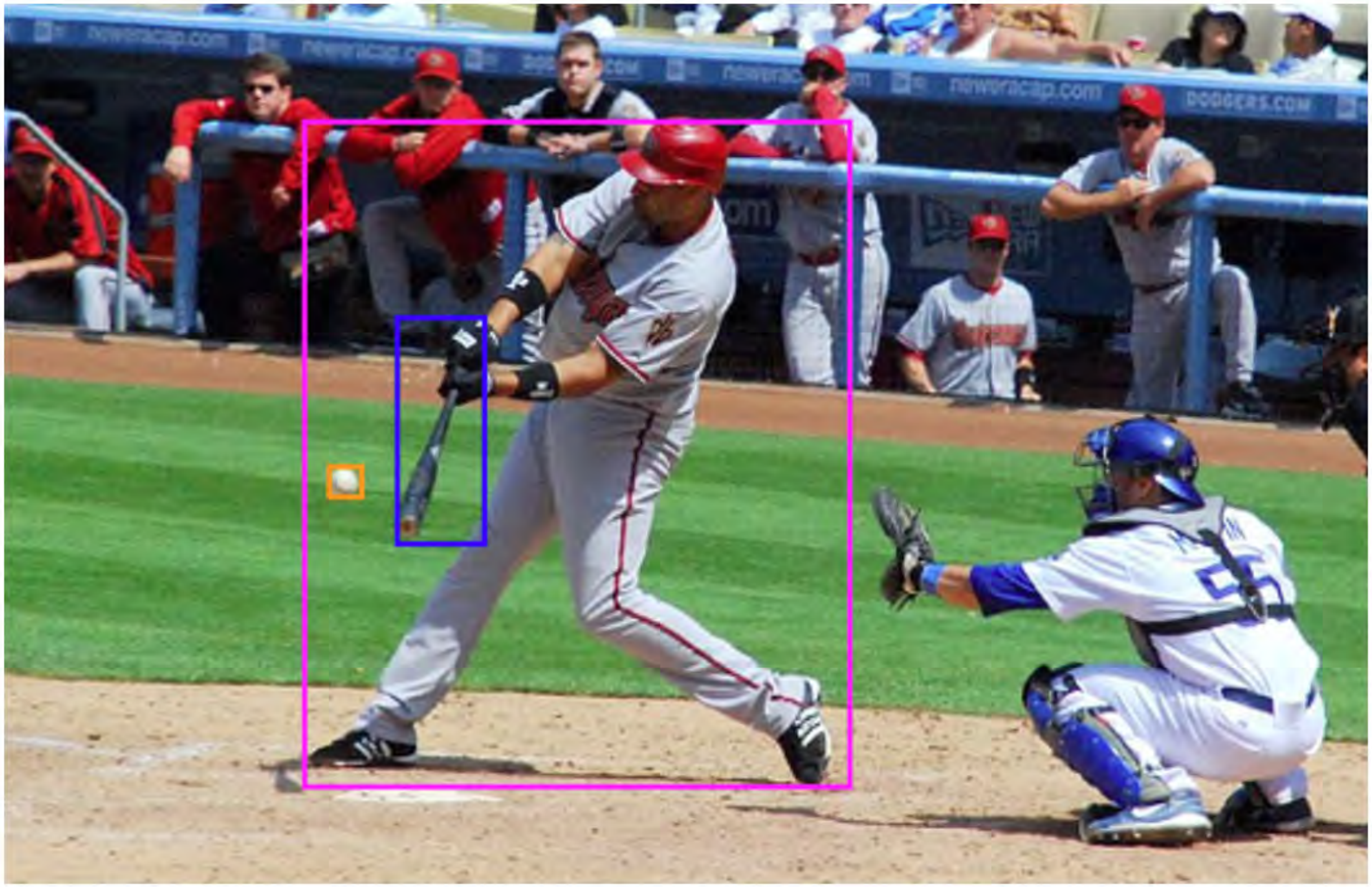


Figure 1. **Visual Semantic Role Labeling**: We want to go beyond classifying the action occurring in the image to being able to localize the agent, and the objects in various semantic roles associated with the action.

What we would like to infer…

Will person B put some money into Person C's tip bag?

Social Perception

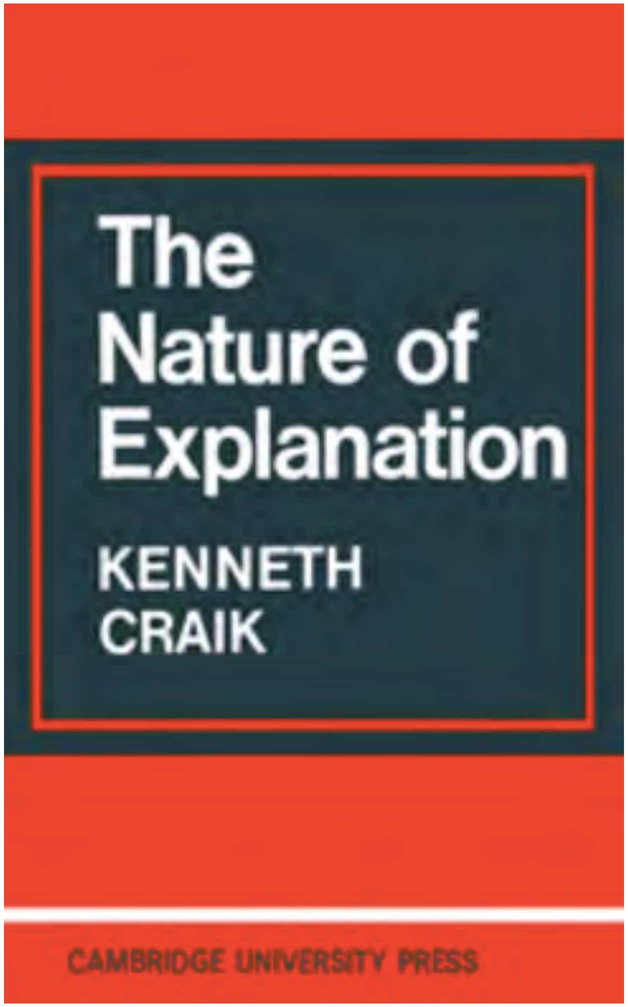
- Computers today have pitifully low “social intelligence”
- We need to understand the internal state of humans as they interact with each other and the external world
- Examples: emotional state, body language, current goals.

What we can’t do (yet)

- The hierarchical structure of human behavior- movement, goals, actions and events

ACTION = MOVEMENT + GOAL

On Mental Models



If the organism carries a `small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it (Craik, 1943,Ch. 5, p.61)

Modern Control theory (Kalman et al) uses a state space formalism to achieve this.

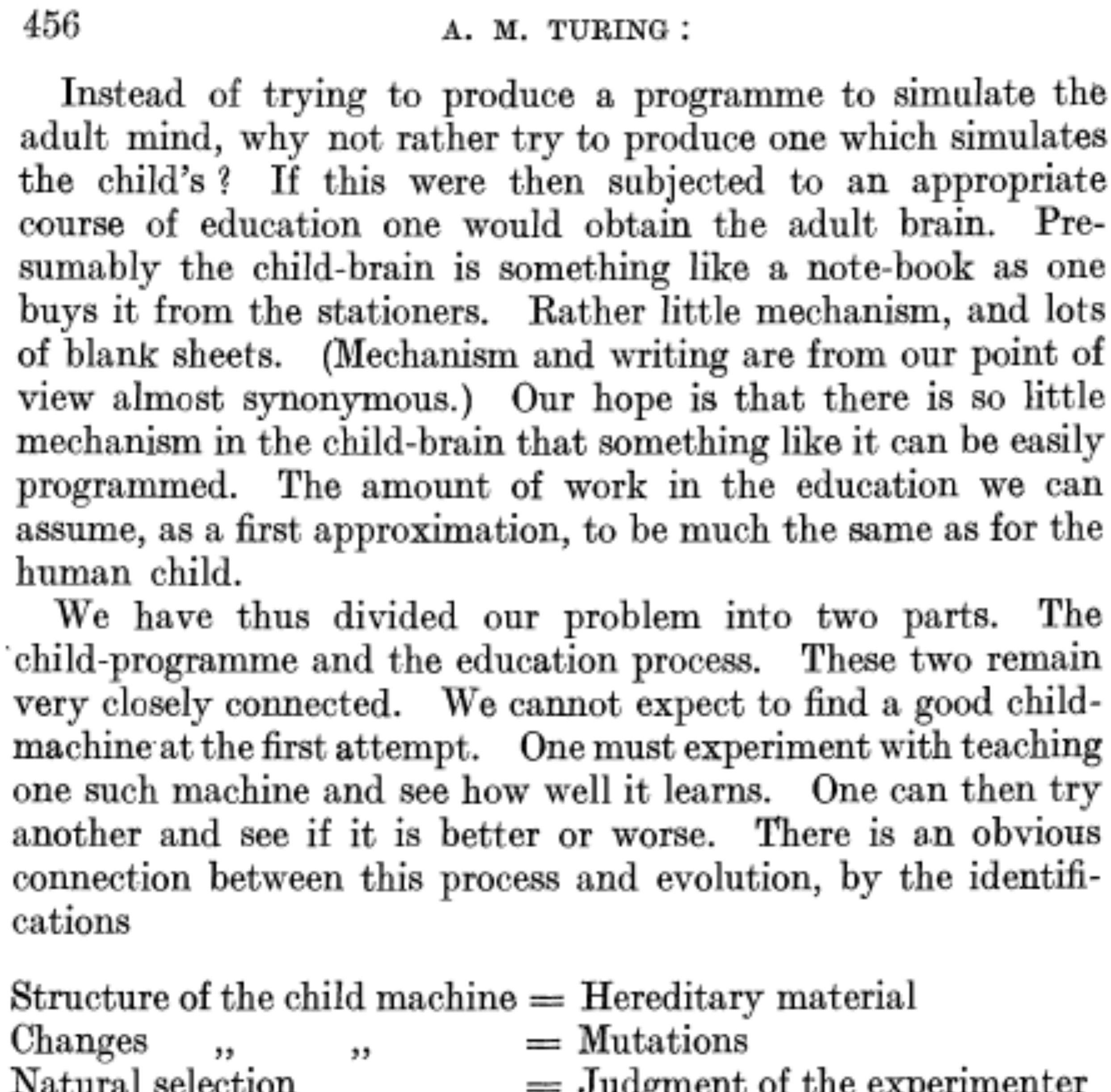
External
Teacher
Signal

External
Supervision

Internal
Teacher
Signal

Self
Supervision

Computing Machinery and Intelligence
Turing (1950)


456 A. M. TURING:

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child-brain is something like a note-book as one buys it from the stationers. Rather little mechanism, and lots of blank sheets. (Mechanism and writing are from our point of view almost synonymous.) Our hope is that there is so little mechanism in the child-brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child.

We have thus divided our problem into two parts. The child-programme and the education process. These two remain very closely connected. We cannot expect to find a good child-machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications

Structure of the child machine	=	Hereditary material
Changes ,, ,,	=	Mutations
Natural selection	=	Judgment of the experimenter

The Development of Embodied Cognition: Six Lessons from Babies

Linda Smith & Michael Gasser

Abstract. The embodiment hypothesis is the idea that intelligence emerges in the interaction of an agent with an environment and as a result of sensorimotor activity. In this paper we offer six lessons for *developing* embodied intelligent agents suggested by research in developmental psychology. We argue that starting as a baby grounded in a physical, social and linguistic world is crucial to the development of the flexible and inventive intelligence that characterizes humankind.

The Six Lessons

- Be multi-modal
- Be incremental
- Be physical
- Explore
- Be social
- Use language

We hope you enjoy the course!