

I want machines to see

Jonas Kubilius

Brain & Cognition / KU Leuven (Belgium)

DataLux / 2015-04-08



■ klab.lt

*except where otherwise noted, these slides are available under
the Creative Commons Attribution 4.0 International License*

The goal

How to get from an image to the knowledge about its contents as perceived by a human observer.

Compete against GoogLeNet

ILSVRC 2014 demo by Andrej Karpathy

Compete against GoogLeNet

ILSVRC 2014 demo by Andrej Karpathy

How did we get here?

Why is vision difficult?

1. The invariance problem
2. Discovering structure

The invariance problem



cc by 2.0 – Wendy Cope / Flickr

Discovering structure



cc by-nc 2.0 – Celeste RC / Flickr

Discovering structure

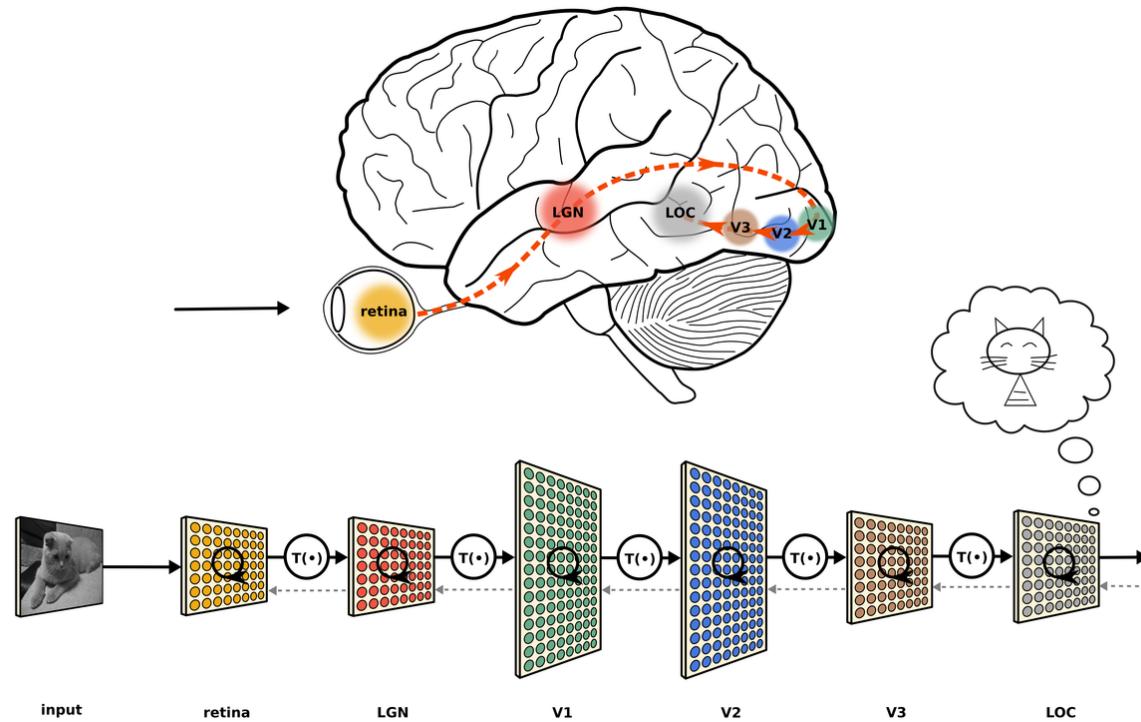
cc by-nc 2.0 – Celeste RC / Flickr

Discovering structure

198 198 198 199 199 195 202 201 203 203 202 202 204 201 202 197 196 197 193 193 189 178 168 154 137 119 108 087 095 098 105 095 086 082 077 060 046 039 039
198 198 198 200 197 199 204 204 203 203 202 203 204 204 201 200 201 198 197 196 189 179 165 145 125 111 098 086 079 088 091 091 086 072 056 053 033 025 023 029
198 198 198 199 203 204 199 197 199 187 188 193 203 204 201 202 201 196 193 188 181 172 153 126 099 086 075 068 069 079 079 064 059 048 032 030 022 019 015 019
201 201 202 201 199 193 176 169 171 166 182 178 169 184 192 197 196 190 185 177 160 148 132 103 069 061 059 060 069 071 062 049 037 031 024 027 035 026 020 020
202 203 203 201 193 168 174 145 137 130 133 135 147 165 170 177 187 186 170 153 132 118 100 077 058 054 055 061 069 065 057 053 047 038 040 051 065 061 051 047
202 203 204 201 191 179 144 117 117 130 136 127 117 121 145 166 161 163 153 129 111 099 083 079 063 057 071 078 080 079 077 076 074 065 062 073 083 086 075 071
201 202 204 202 191 165 128 126 120 114 117 133 137 131 125 133 147 146 154 136 098 114 097 100 091 094 093 096 096 089 085 073 067 063 066 074 077 070 067
206 203 198 195 192 198 189 192 114 116 124 136 139 136 134 128 126 136 139 190 155 125 108 117 120 116 111 099 087 085 071 058 049 053 058 060 059 061 063 059
202 203 198 186 172 120 121 118 123 133 130 138 147 155 150 143 126 124 132 149 151 157 136 123 125 124 110 092 077 065 039 033 037 053 067 066 073 082 080 071
205 203 196 189 144 099 094 101 103 130 153 153 150 154 162 159 153 137 123 122 134 154 159 148 128 113 091 073 055 048 050 047 054 078 094 091 098 107 106 093
203 203 197 180 135 098 089 106 101 113 120 128 136 139 137 140 138 133 146 150 127 129 143 163 140 110 081 061 061 062 074 076 080 095 106 108 111 117 113 100
200 203 200 190 122 104 111 110 118 124 132 150 151 149 157 146 142 142 136 144 138 139 146 150 149 147 107 080 074 091 092 094 093 113 119 116 113 117 105 092
203 201 196 185 129 107 095 113 133 152 157 159 165 176 170 156 157 154 148 131 130 132 134 139 149 150 147 108 099 097 103 113 114 120 121 113 107 096 079 067
201 198 196 172 118 099 105 105 110 110 121 137 144 163 182 188 185 183 184 178 163 154 140 144 151 149 150 154 131 107 113 126 130 124 113 093 084 071 051 035
198 197 195 180 133 115 085 090 076 103 125 150 165 180 199 198 196 178 180 171 187 184 175 158 149 119 126 145 165 142 116 120 129 112 093 070 060 058 048 034
197 196 196 180 105 091 072 082 093 110 153 178 176 166 156 151 133 148 152 170 158 148 147 149 168 149 118 107 125 150 160 120 110 112 081 071 069 072 081 072
197 197 195 180 125 099 078 086 101 127 129 137 156 169 183 189 167 167 157 150 157 161 159 156 142 152 150 121 103 133 148 178 124 113 104 104 108 102 103 091
204 203 199 190 121 083 075 087 090 106 136 149 143 156 164 162 168 165 163 172 172 187 182 167 159 140 146 151 125 107 114 135 164 129 119 125 124 115 110 104
203 200 193 194 139 086 068 089 078 095 136 143 146 146 168 191 190 189 187 185 181 197 173 149 145 148 135 123 135 143 120 105 125 170 142 119 108 105 104 100
204 200 198 187 151 101 075 077 086 123 144 141 165 187 191 194 208 190 172 170 163 169 155 160 153 156 152 142 143 142 127 123 100 123 166 152 103 085 088 090
202 198 190 174 149 112 074 073 103 100 120 169 163 188 178 168 149 157 156 168 175 184 178 166 138 148 146 140 128 111 141 118 099 138 167 175 121 079 061
201 194 178 160 163 101 049 081 089 105 112 129 146 119 124 138 160 155 174 171 167 174 184 192 179 172 146 151 160 144 141 123 129 110 117 136 152 177 129 132
199 187 172 151 153 102 080 076 099 082 106 113 100 112 138 157 171 177 200 187 173 160 151 164 169 173 171 156 132 115 121 091 116 117 113 100 118 159 150 179
196 181 161 137 150 134 084 086 100 093 072 082 102 134 150 176 182 188 160 155 169 166 163 157 146 142 148 137 119 098 093 104 124 118 101 085 103 143 139 147
188 167 145 127 140 144 111 089 098 095 070 063 073 096 123 109 121 124 132 137 158 166 182 155 164 168 160 138 123 101 091 075 085 134 094 100 153 110 140
176 152 131 124 119 147 092 096 112 106 083 061 061 064 076 096 128 148 173 158 144 171 189 189 191 186 173 159 138 125 128 107 063 063 118 126 113 145 138 157
158 135 120 113 117 148 111 069 101 109 090 073 080 089 092 110 108 128 184 216 199 208 202 193 174 166 166 147 136 118 128 117 087 101 112 121 148 127 161 177
141 126 114 111 114 137 134 070 099 093 099 096 090 109 119 161 202 209 201 189 170 177 171 165 170 173 189 159 147 141 148 142 120 124 154 146 176 166 167 162
126 119 118 117 114 124 151 095 094 094 093 098 121 143 171 166 161 167 179 195 190 186 198 199 171 147 115 144 137 152 140 181 151 179 168 174 163 199 188 214
125 125 132 128 117 127 124 150 133 095 088 103 109 110 128 177 204 202 212 216 212 217 201 200 192 185 175 161 137 156 165 176 189 176 155 156 168 175 201 174 187

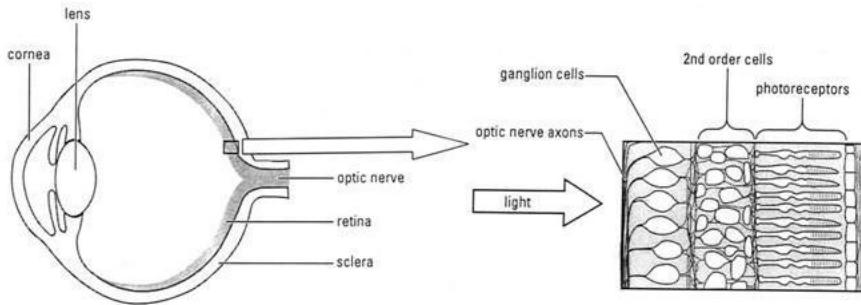
Human visual system

Human brain

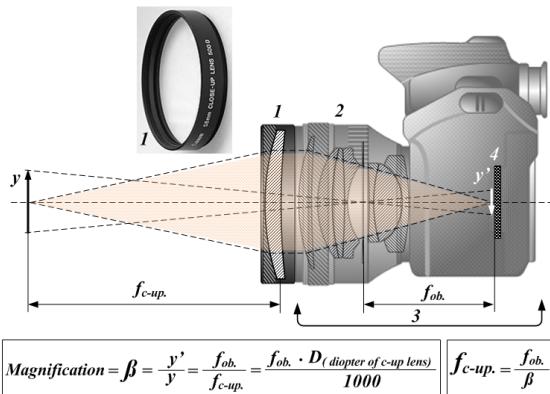


cc by 3.0 – Kubilius (figshare, 2013)

Retina

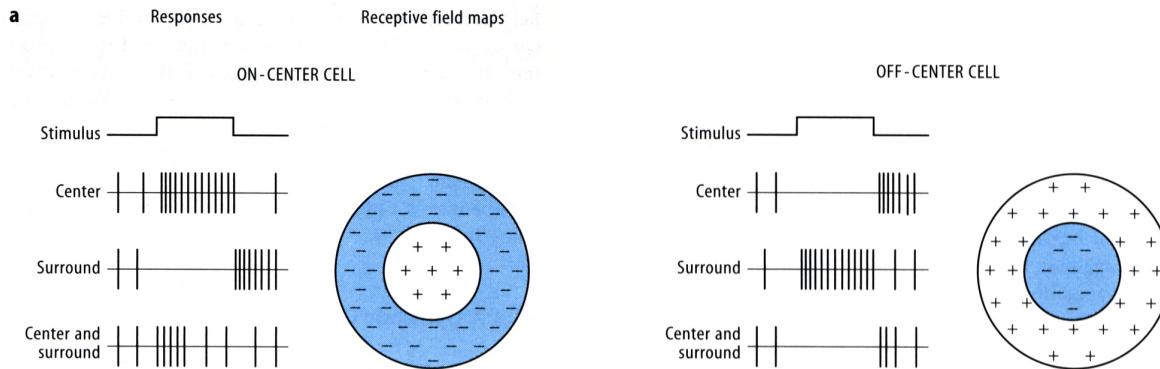


cc by-sa 3.0 – Dowling (Scholarpedia, 2007),



cc by-sa 3.0 – Tamasflex / Wikimedia Commons

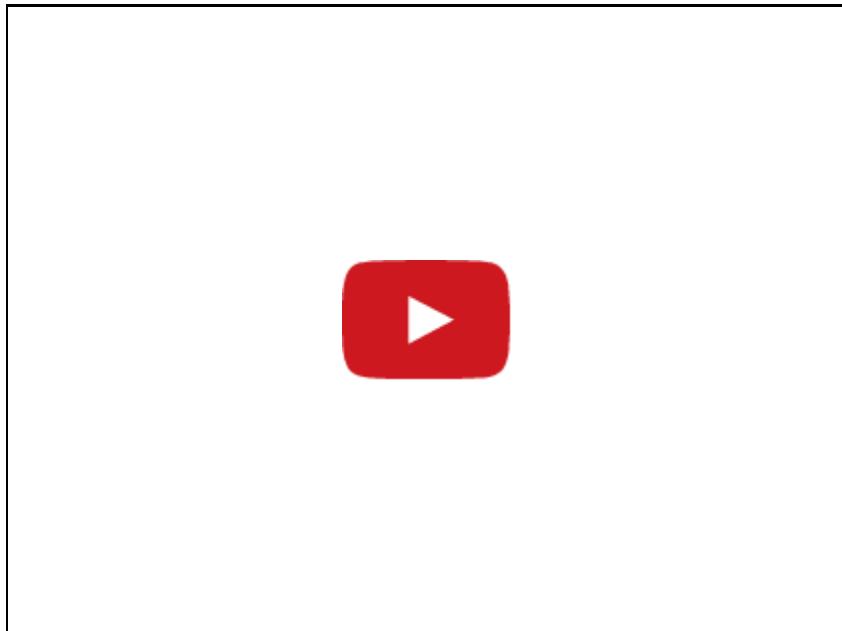
Retinal ganglion cells



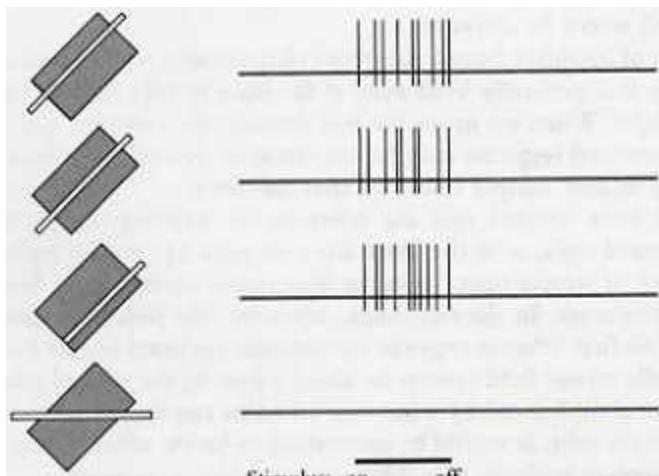
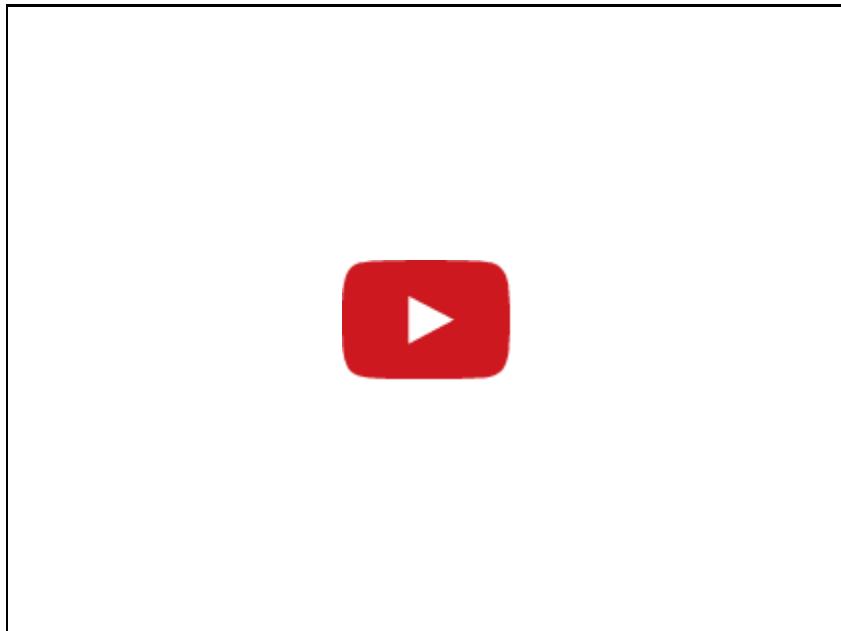
cc by-sa 3.0 -Dowling (Scholarpedia, 2007)



V1: Simple cell



V1: Complex cell



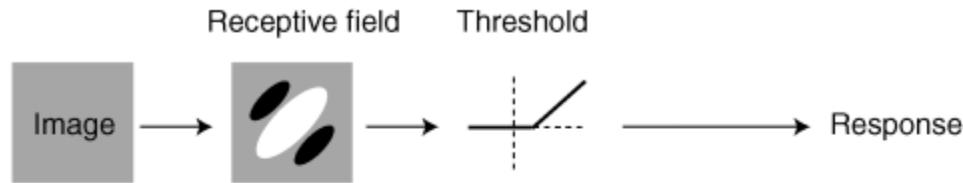
V1 function

Simple cells: **feature selectivity**

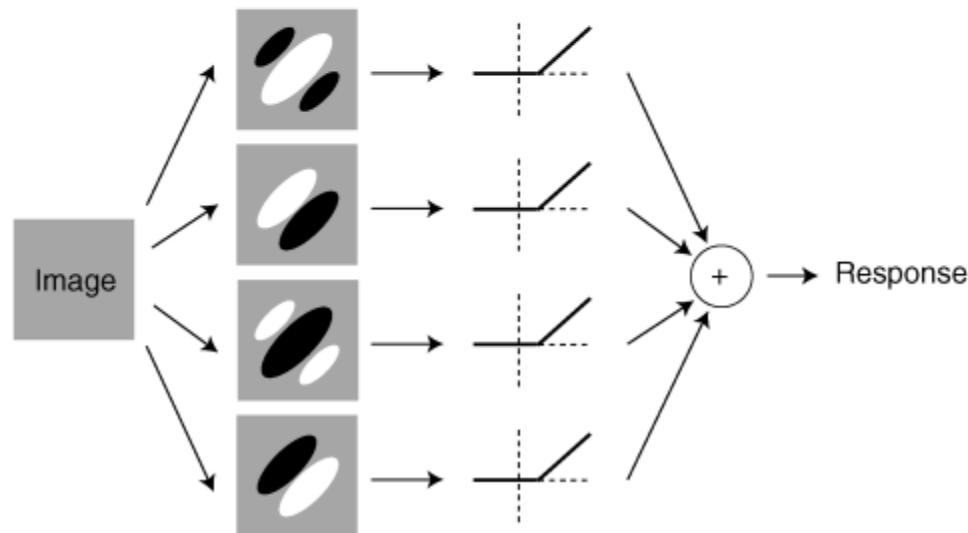
Complex cells: **feature invariance**

Complex cell connectivity

A Simple cell

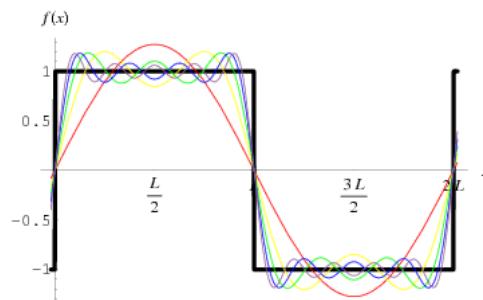


B Complex cell



fair use – Carandini (The Journal of Physiology, 2006)

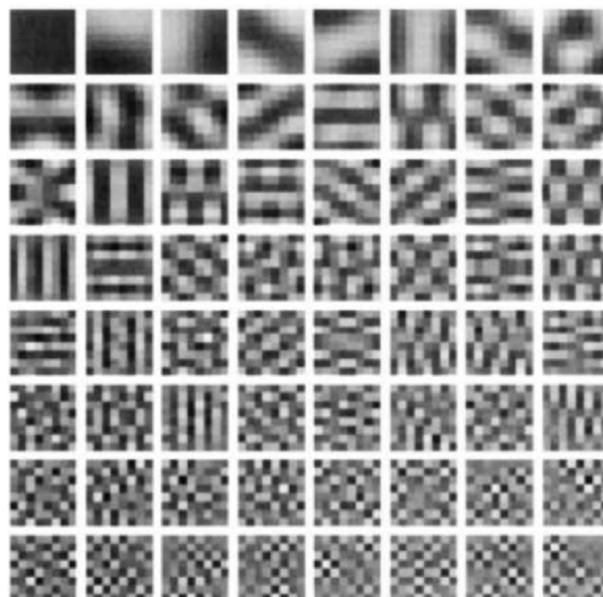
V1: Local frequency filtering



Weisstein / MathWorld



Principal components analysis



fair use - Olshausen & Field (Nature, 1996)

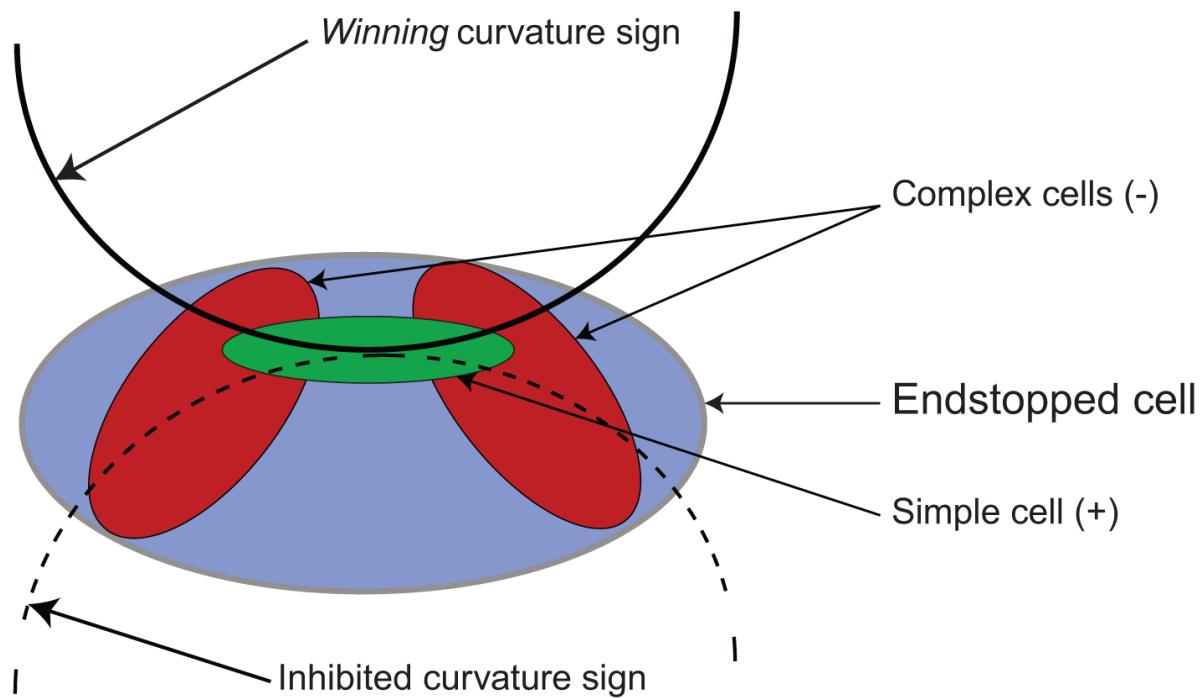
Sparse coding

Principal Components Analysis (PCA) + sparsity, or Independent Components Analysis (ICA)

V2: End-stopping

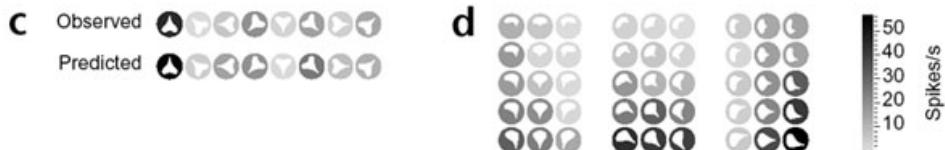
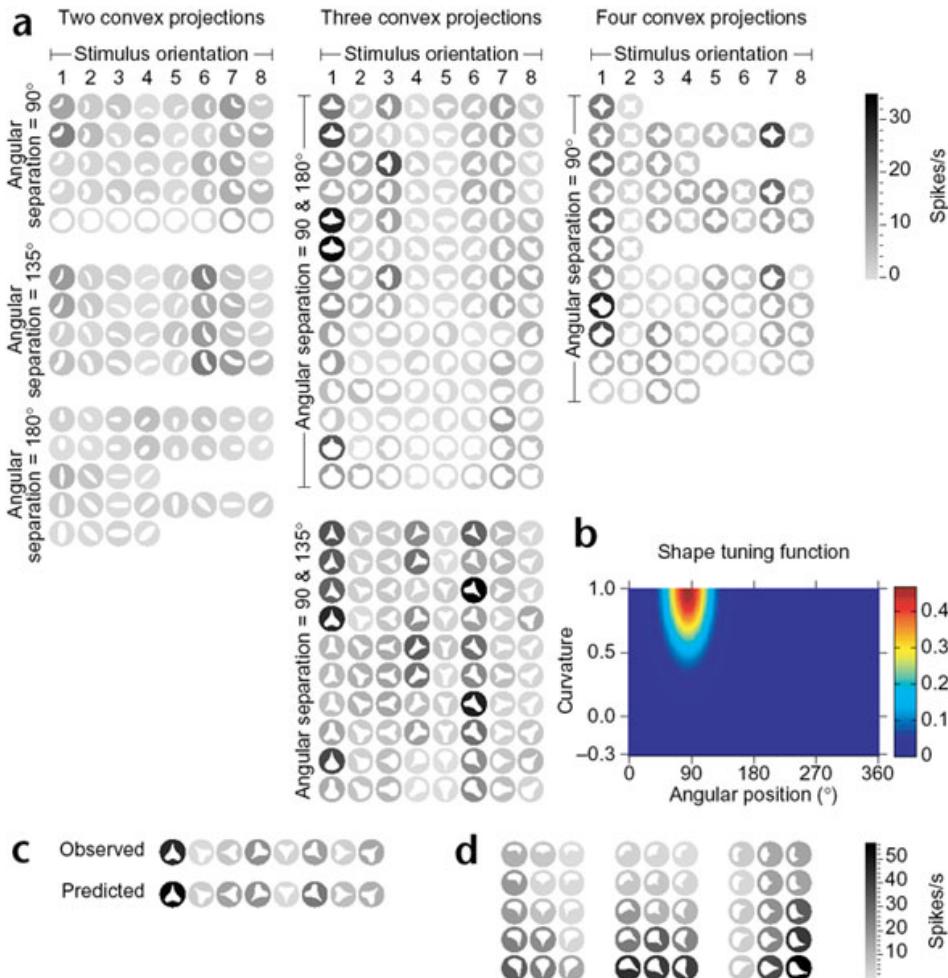


V2: Curvature detection

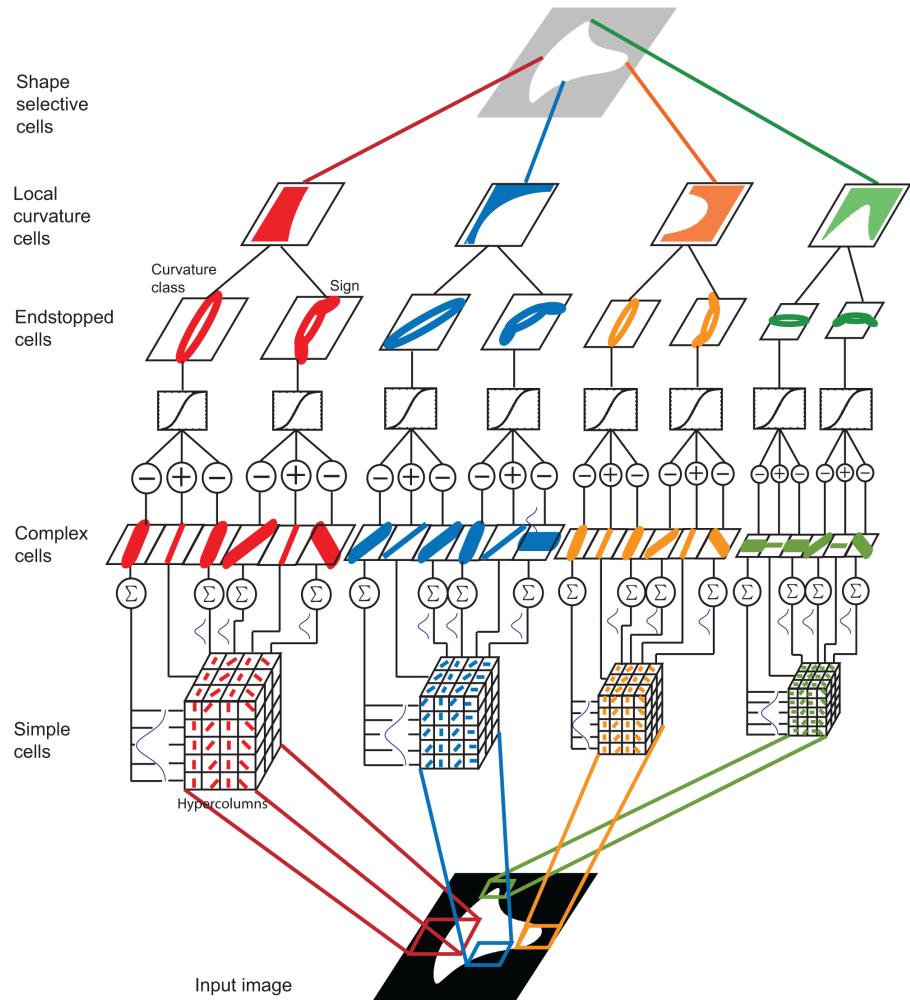


cc by - Rodríguez-Sánchez & Tsotsos (PLoS ONE, 2012)

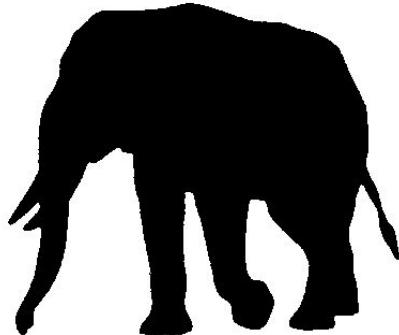
V4: Curvature detection



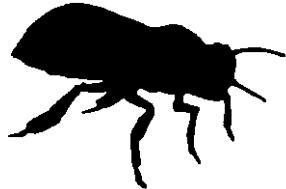
Putting it all together



Recognition from shape only?

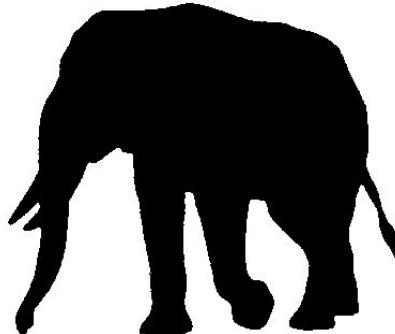


100%

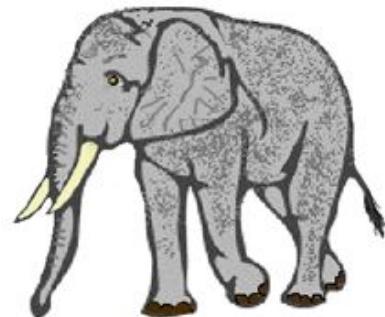


34%

Recognition from shape only?



100%

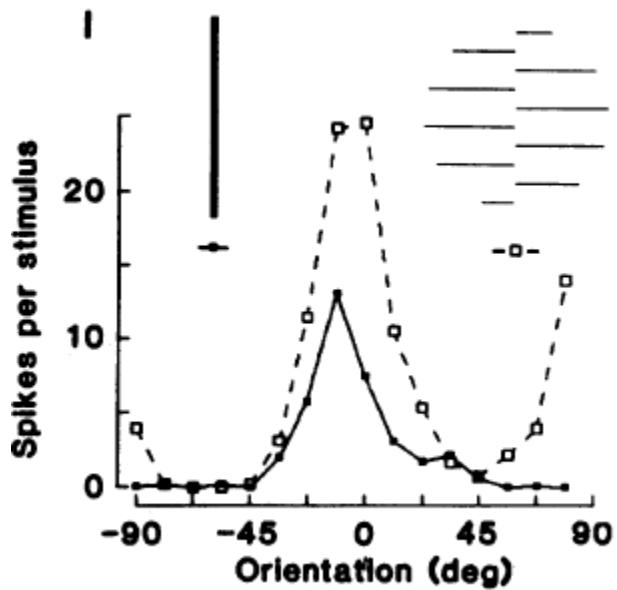


You won't get far with edges



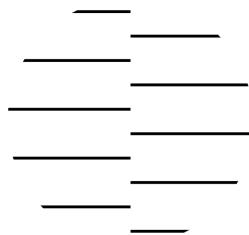
cc by 3.0 – Σ64 / Wikimedia Commons

What is V2 doing?

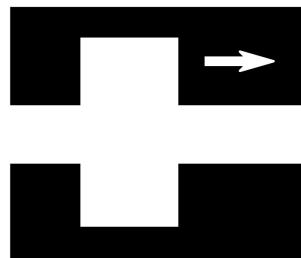


fair use – von der Heydt et al. (Science, 1984)

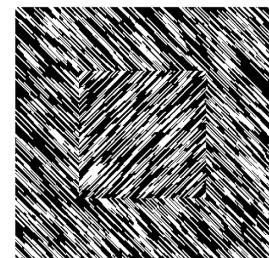
A



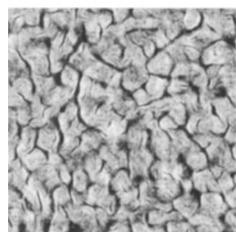
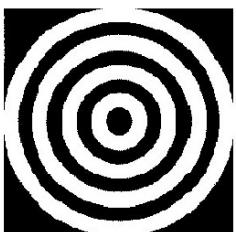
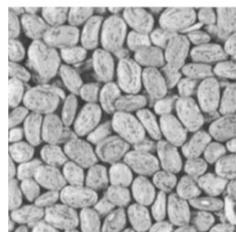
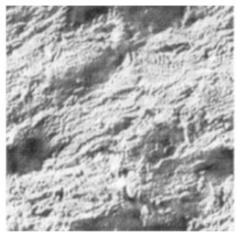
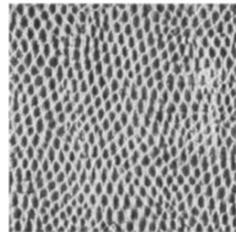
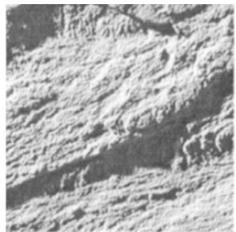
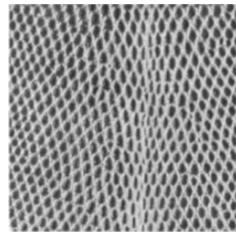
B



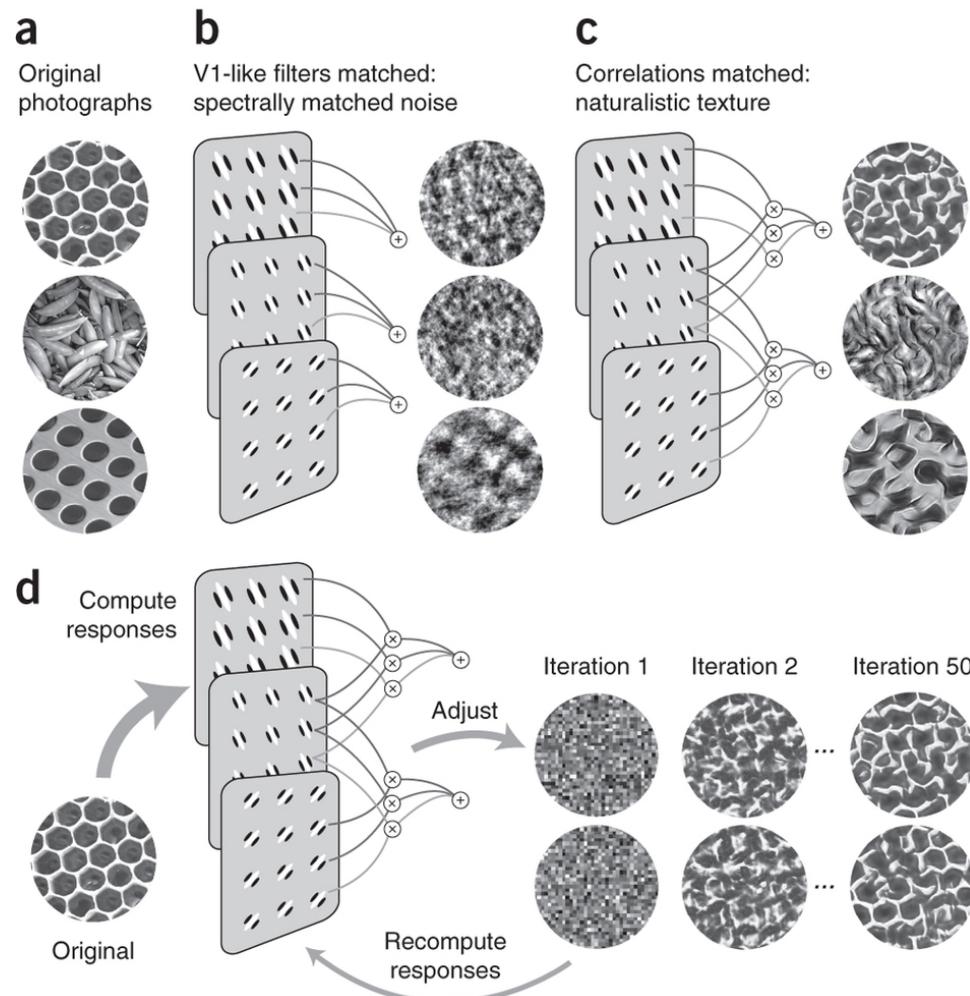
C



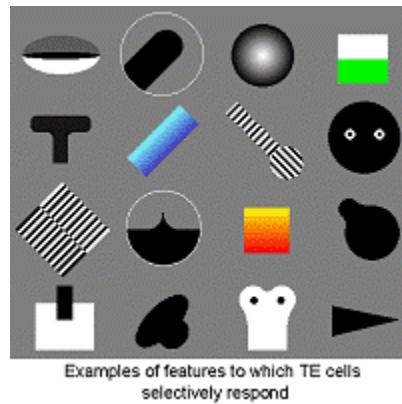
Texture processing



V2 processes textures

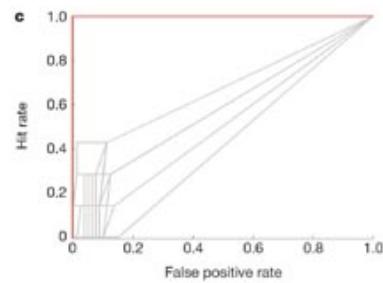
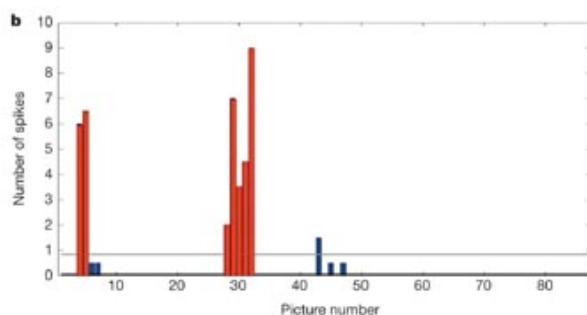
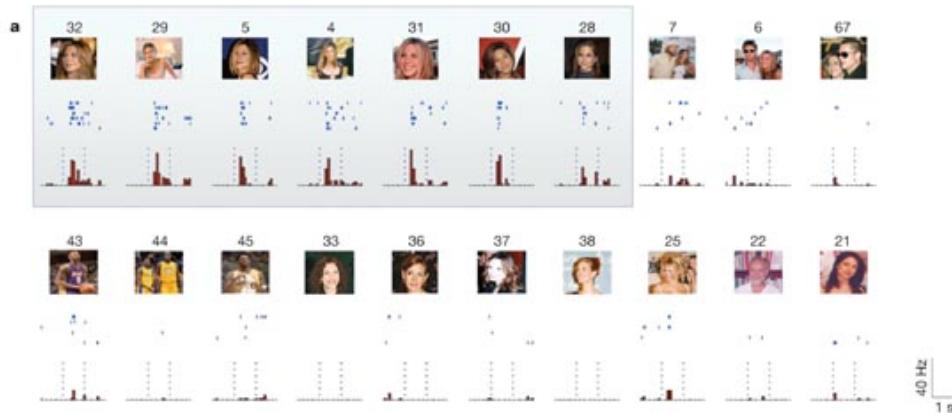


Later stages



fair use – Tanaka (Annual Review of Neuroscience, 1996) via RIKEN

Jeniffer Aniston cell

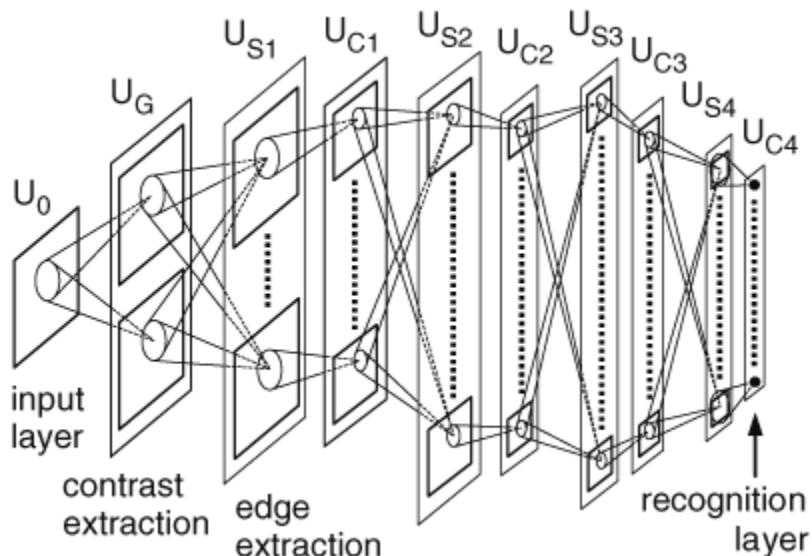


fair use - Quiroga et al. (Nature, 2005)

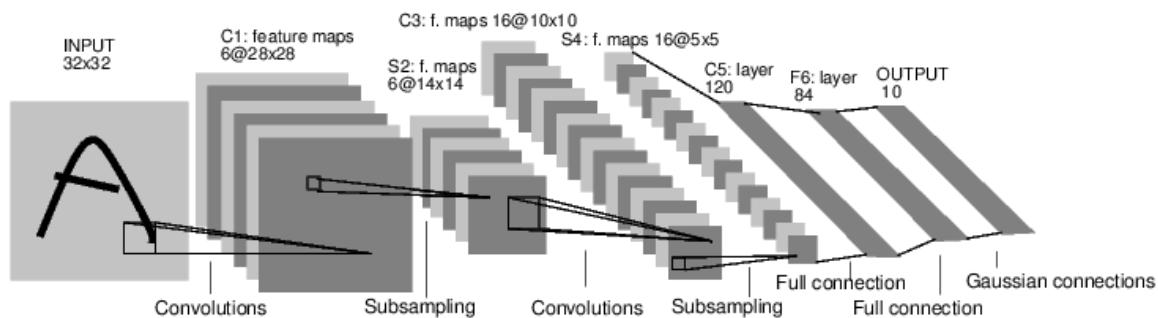
Deep Neural Networks

(actually, only convolutional)

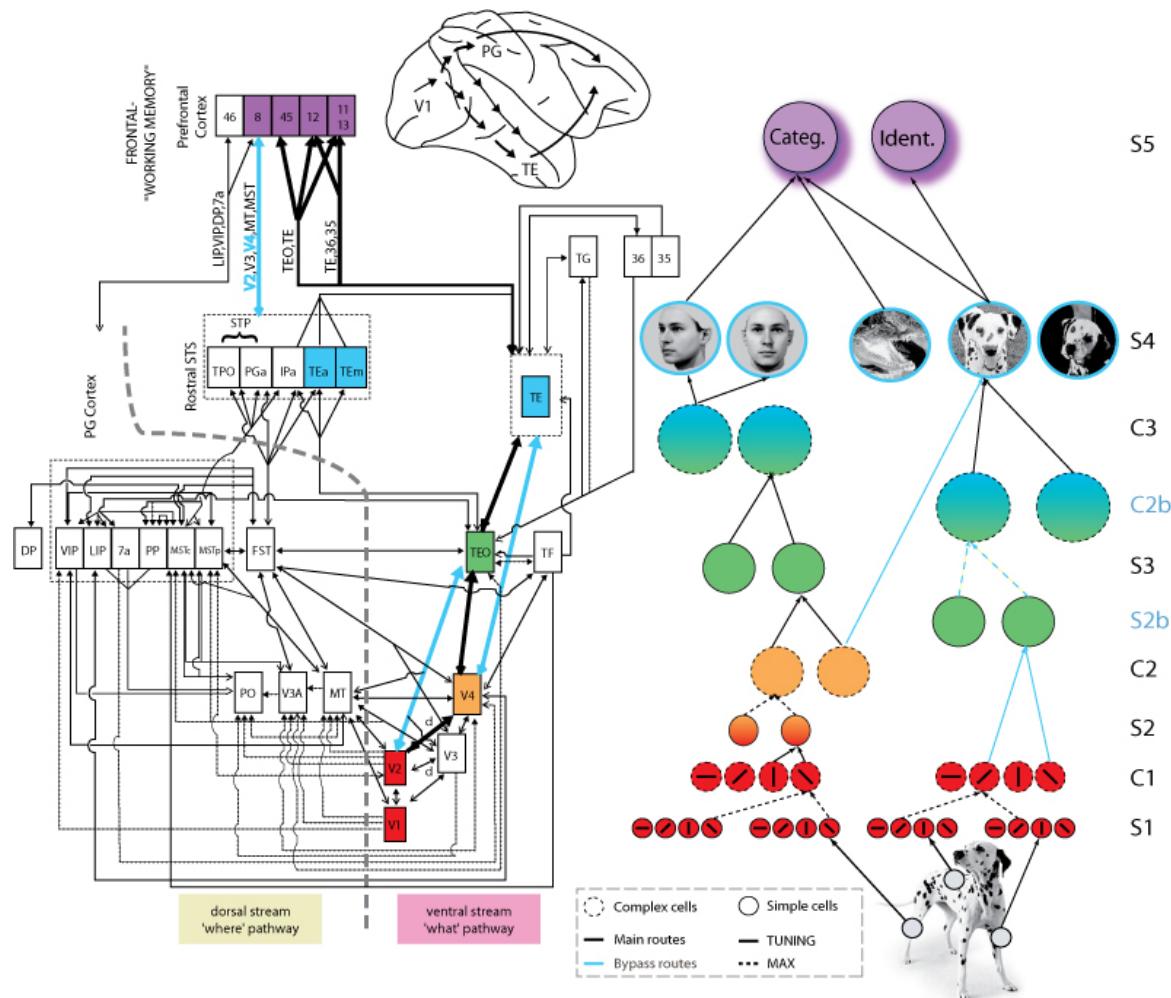
Early architectures



cc by-nc-sa 3.0 – Fukushima (Scholarpedia, 2007)

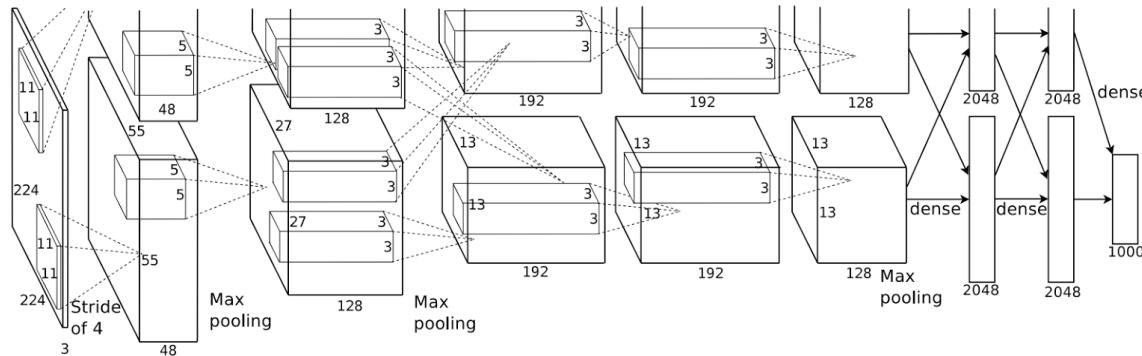


HMAX



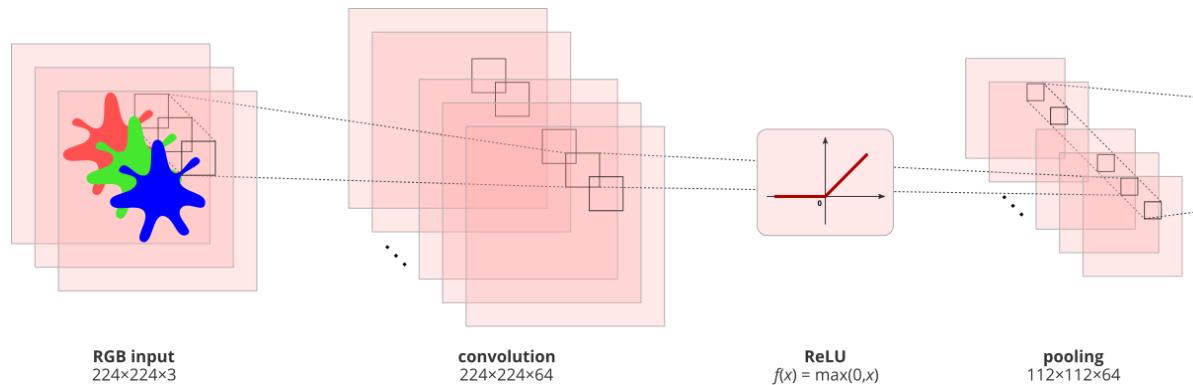
cc by 3.0 – Serre et al. (PNAS, 2007) via Kreiman (Scholarpedia, 2008)

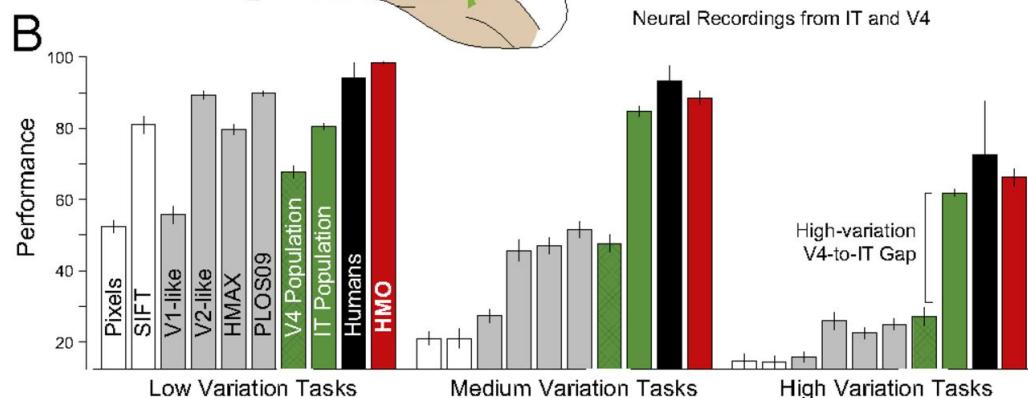
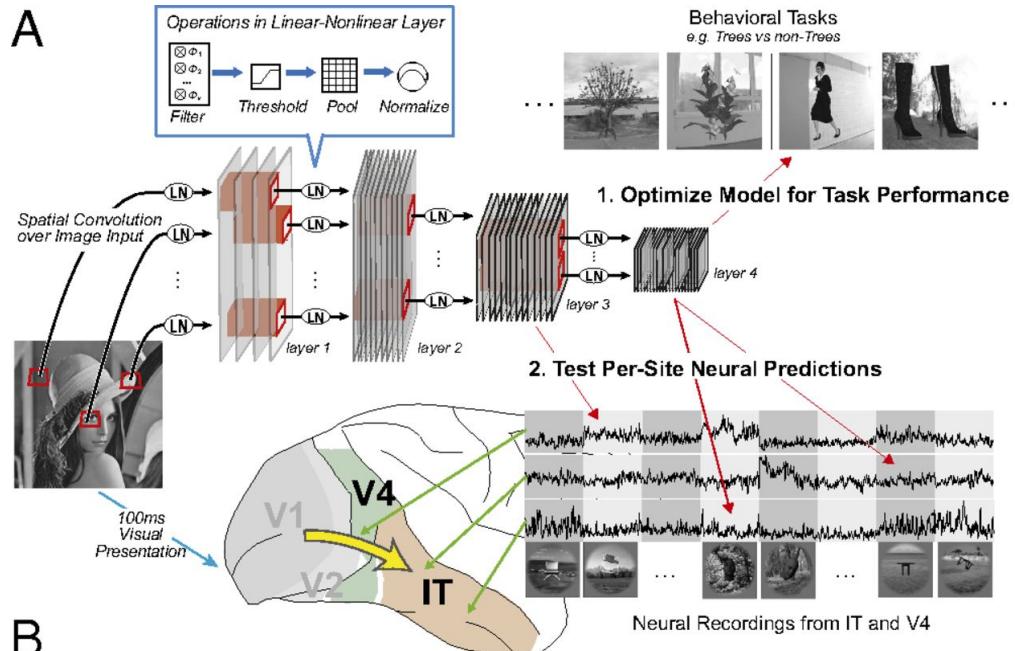
AlexNet



fair use – Krizhevsky et al. (NIPS, 2012)

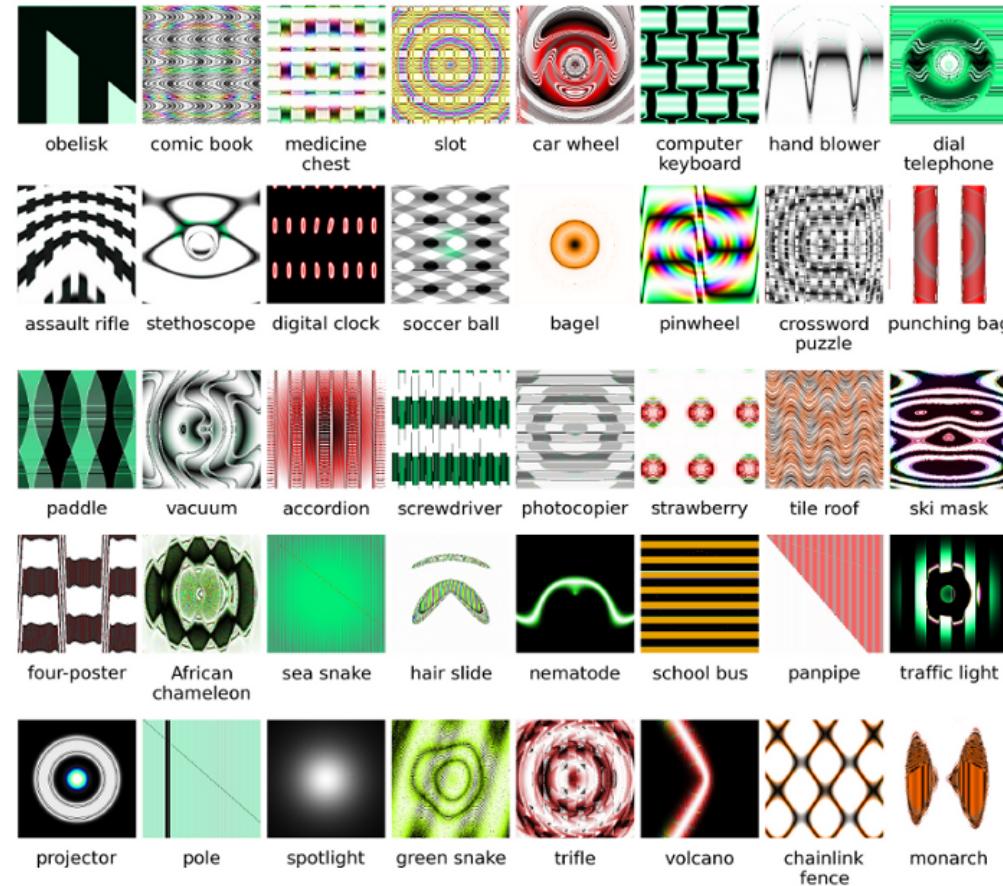
Basic building blocks





Fooling ConvNets

With 99.12% confidence, this is what I see...

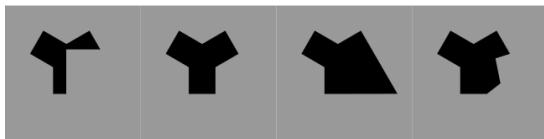


**Is this what we fought
for?**

Occlusion

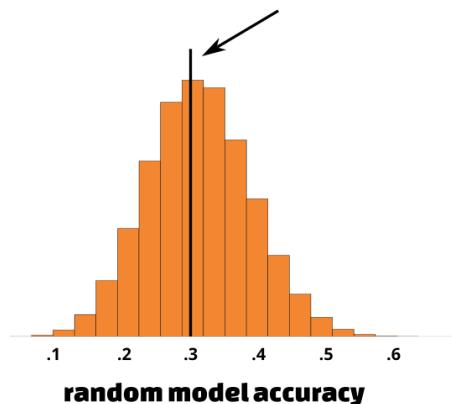


which shape is present?



a dataset of partially occluded
and unoccluded shapes

GaborJet, HMAX, & HMO performance



I want machines to ~~see~~

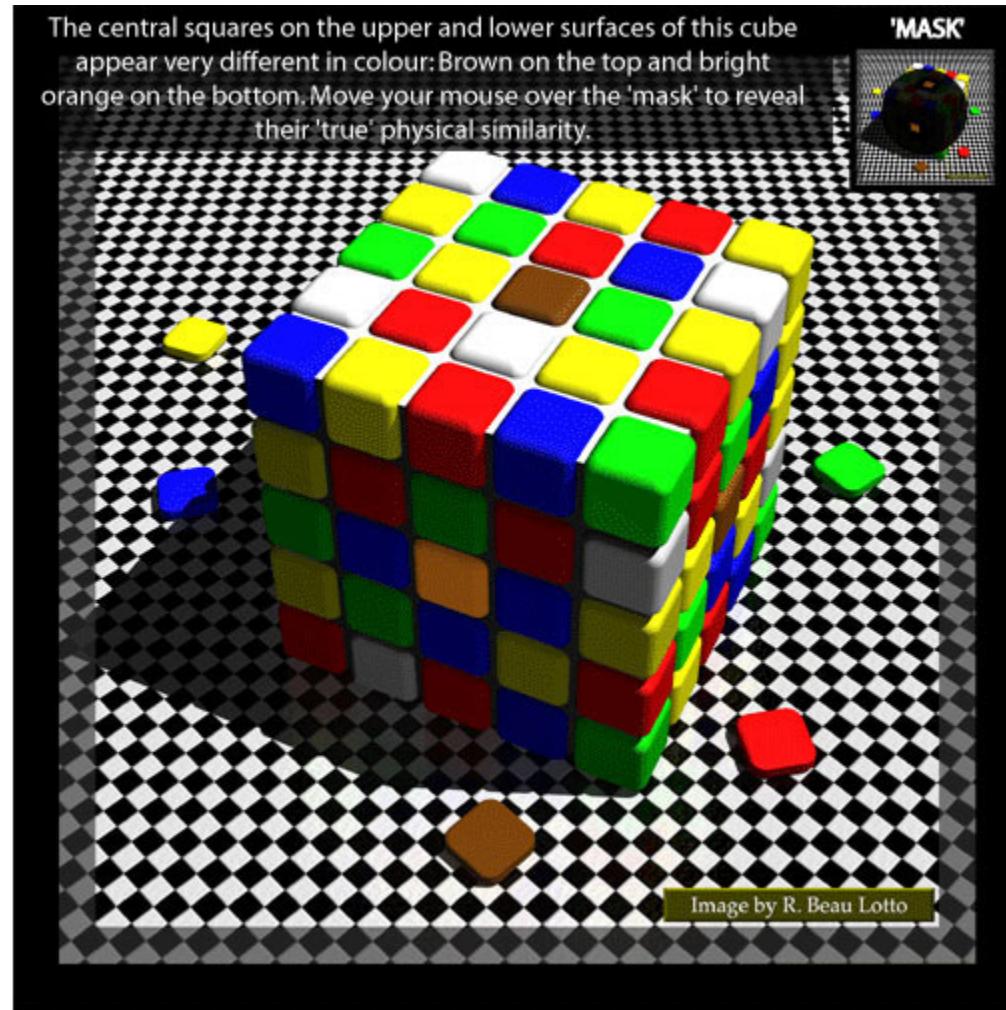
perceive

The dress

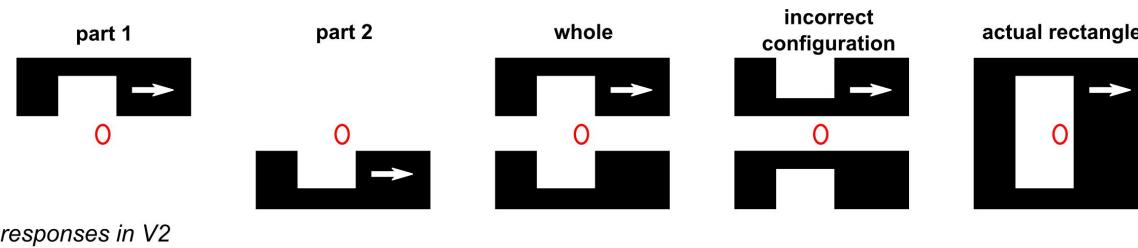


fair use - [Wikipedia.png](#)

Color tile



Illusory contours

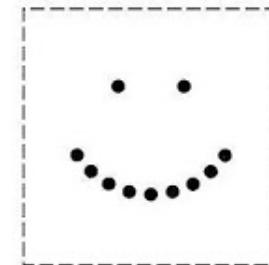
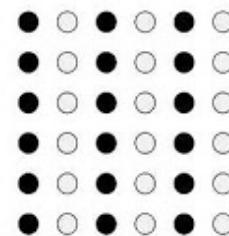
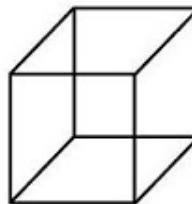
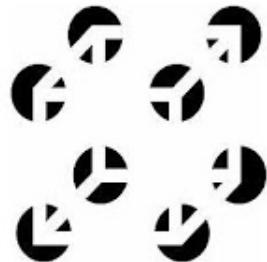
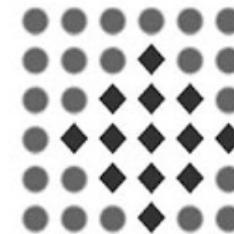
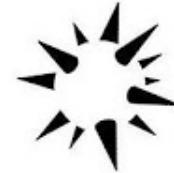
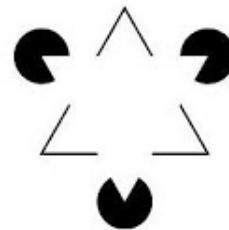
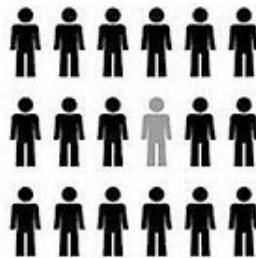
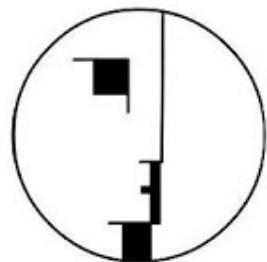


von der Heydt et al. (Science, 1984) / via Kubilius et al. (Cortex, in press)

Integration

- Need to put things together
- Feedforward architectures are not sufficient
- Must work on multiple tasks

Gestalts

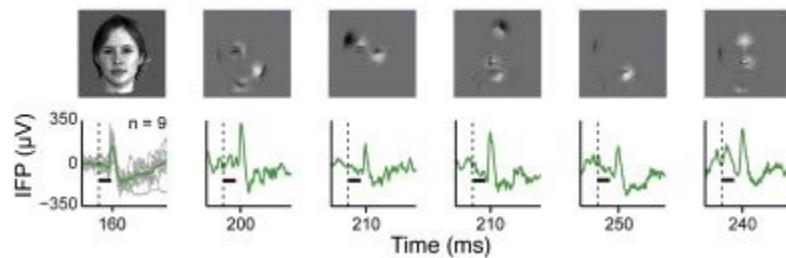


cc by-sa 3.0 - Impronta / Wikimedia Commons

Border-ownership

It takes time to perceive

A



fair use – Liu et al. (Neuron, 2014)

What does it look like?

fair use - LabelMe

What does it look like?

fair use - LabelMe

More tasks

The visual system does a lot of things:

- Categorization
- Navigation
- Acting on environment
- ...



A



B



C



D

cc by 4.0 – Kubilius et al. (Frontiers in Computational Neuroscience, 2014)

Visual Turing test



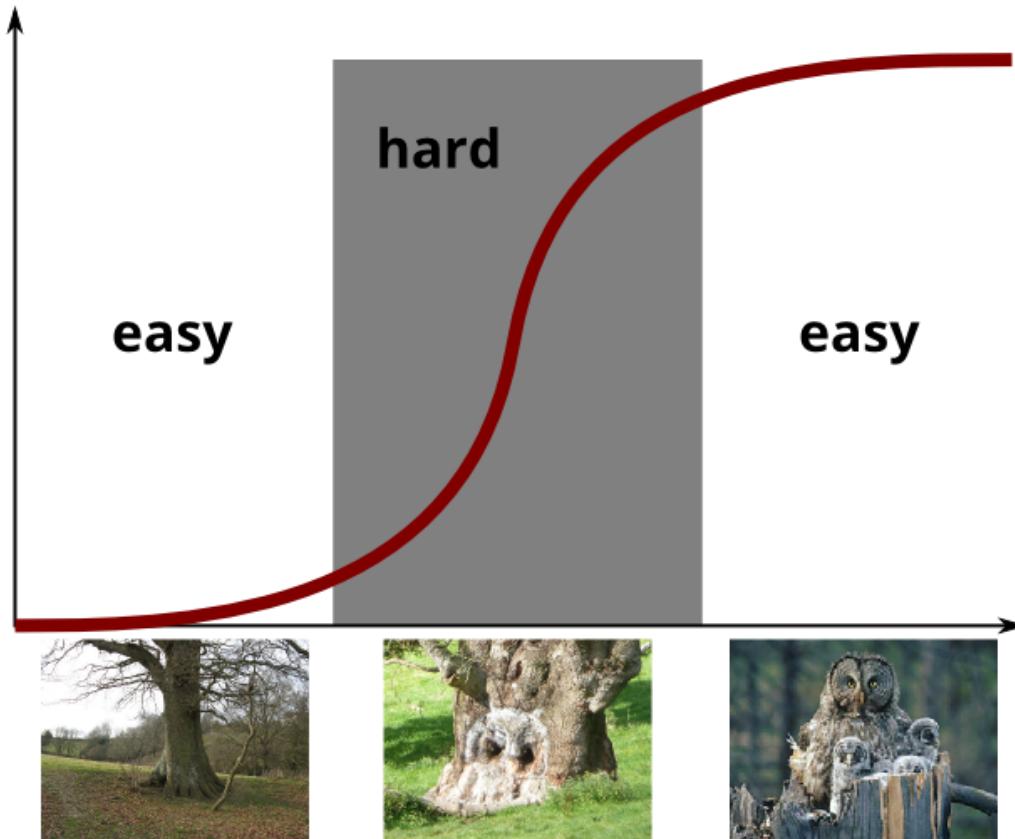
- | | |
|--|---------|
| 1. Q: Is there a person in the blue region? | A: yes |
| 2. Q: Is there a unique person in the blue region?
(Label this person 1) | A: yes |
| 3. Q: Is person 1 carrying something? | A: yes |
| 4. Q: Is person 1 female? | A: yes |
| 5. Q: Is person 1 walking on a sidewalk? | A: yes |
| 6. Q: Is person 1 interacting with any other object? | A: no |
| ⋮ | |
| 9. Q: Is there a unique vehicle in the yellow region?
(Label this vehicle 1) | A: yes |
| 10. Q: Is vehicle 1 light-colored? | A: yes |
| 11. Q: Is vehicle 1 moving? | A: no |
| 12. Q: Is vehicle 1 parked and a car? | A: yes |
| ⋮ | |
| 14. Q: Does vehicle 1 have exactly one visible tire? | A: no |
| 15. Q: Is vehicle 1 interacting with any other object? | A: no |
| 17. Q: Is there a unique person in the red region? | A: no |
| 18. Q: Is there a unique person that is female in the red region? | A: no |
| 19. Q: Is there a person that is standing still in the red region? | A: yes |
| 20. Q: Is there a unique person standing still in the red region?
(Label this person 2) | A: yes |
| ⋮ | |
| 23. Q: Is person 2 interacting with any other object? | A: yes |
| 24. Q: Is person 1 taller than person 2? | A: amb. |
| 25. Q: Is person 1 closer (to the camera) than person 2? | A: no |
| 26. Q: Is there a person in the red region? | A: yes |
| 27. Q: Is there a unique person in the red region?
(Label this person 3) | A: yes |
| ⋮ | |
| 36. Q: Is there an interaction between person 2 and person 3? | A: yes |
| 37. Q: Are person 2 and person 3 talking? | A: yes |

A machine who mistook a knife for a cat

**or why understanding human perception
matters**

Decision making

Decisions are difficult when only partial or ambiguous information is available.

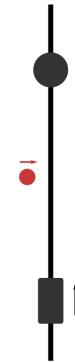


left: cc by-sa 2.0 – Dave Spicer / middle: cc by-sa 2.0 – Chris Downer / right: cc by 2.0 – Chief Trent

The trolley problem



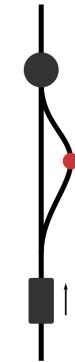
the switch
Foot, 1967



the fat man
Thomson, 1976



the fat villain



the loop
Costa, 1987



the man in the yard
Unger, 1992

CC0 – Jonas Kubilius / Wikimedia Commons

Tools must act clever

1. Tools are too limited (e.g., default location in Google Maps)
2. Programmers often are bad at accounting for human nature (e.g., awful GUIs)

Thank you!

slides available at klab.it