

Computer Vision: Making Machines See

Alex Kendall

Department of Engineering



What do you see here?

128 037 246 247 255 254 255 255 255 250 240 255 049 253 037 108 079 190 134 106 042 022 255 023 182 036 034 084 128 036 238 071 226 031 036 193 026 076 156 143 255 236 252 248 251 255 119 216

192 039 250 249 255 255 255 255 241 242 211 254 252 255 112 031 180 098 127 066 043 021 242 243 113 033 033 076 247 135 241 015 129 030 054 173 029 090 084 150 251 255 251 239 255 184 212 072

120 255 255 254 252 255 255 255 255 254 117 104 237 255 212 048 173 068 222 125 048 021 181 255 050 029 039 090 247 252 128 167 145 039 073 043 041 080 187 040 157 253 249 233 255 198 255 255

028 172 254 251 246 218 244 255 255 247 106 223 240 050 255 124 051 002 237 194 027 021 032 255 049 028 011 107 197 152 017 253 253 037 120 048 117 103 016 127 197 255 252 026 024 063 254 155

030 189 254 246 252 255 210 255 255 255 255 248 235 151 168 072 067 255 129 249 055 021 031 247 040 030 021 039 100 211 021 255 224 039 020 023 113 017 147 246 163 097 255 255 255 254 252 087

026 202 209 242 231 253 069 255 240 255 241 208 249 033 031 196 255 255 221 255 247 023 020 254 036 028 065 180 025 178 114 151 246 035 181 186 028 122 076 202 193 243 254 255 255 222 255 190

022 255 255 226 240 138 162 249 255 255 254 254 255 055 038 112 151 189 134 128 229 030 021 255 035 025 146 236 053 052 111 134 176 028 026 130 068 152 033 223 255 253 235 229 255 197 246 223

023 055 044 053 050 063 066 080 254 255 255 255 196 255 106 253 253 255 174 250 236 036 017 025 037 029 181 029 242 019 189 236 081 029 095 069 210 078 061 189 155 181 255 211 233 252 138 174

068 038 049 091 057 145 051 242 226 255 255 253 197 252 105 255 052 255 230 219 048 019 020 023 031 025 218 232 253 022 173 226 255 032 028 021 034 090 100 064 156 169 255 255 255 230 185 230

036 076 030 119 118 132 131 231 066 053 050 046 042 048 031 032 026 106 047 232 254 023 031 022 032 030 024 242 026 020 152 142 255 027 015 026 034 006 164 255 136 255 255 226 240 178 255 191

032 034 068 055 110 135 133 085 046 049 044 038 044 042 032 032 037 025 036 033 041 029 035 027 027 032 030 041 031 019 150 138 255 028 020 037 025 046 255 091 211 255 177 253 247 169 253 251

042 061 039 037 042 055 109 097 095 108 076 074 053 072 055 093 058 040 045 061 050 026 034 031 032 029 050 033 037 019 042 045 036 030 037 027 031 121 168 255 129 255 255 255 236 240 215

044 073 070 086 086 089 057 088 084 068 067 071 075 066 026 059 067 053 075 066 048 026 037 037 028 035 034 023 065 023 060 062 026 030 044 028 031 039 220 226 170 255 255 255 208 246 190 245

068 065 084 085 051 095 070 045 087 062 067 054 039 039 036 049 032 047 047 040 098 062 038 032 025 038 035 033 028 023 054 052 050 031 058 075 027 029 182 208 243 255 255 255 254 255 200 199

057 072 058 078 043 062 076 037 079 067 070 071 070 065 036 035 052 094 123 106 161 069 032 029 027 046 080 032 024 021 058 057 027 027 032 064 027 024 027 015 255 255 255 255 254 255 241 038

035 073 065 082 084 087 074 081 075 059 071 065 066 040 030 056 050 047 051 039 103 085 038 026 023 043 041 030 058 026 066 048 025 026 045 066 030 044 189 059 255 255 255 255 255 253 247

085 080 058 071 054 066 066 076 117 077 064 046 078 029 030 058 044 030 052 044 073 068 034 023 036 041 046 030 048 048 069 020 024 036 026 043 031 040 141 249 255 255 255 255 208 198 255 255

052 039 075 074 073 065 065 084 104 063 058 056 078 029 052 062 045 032 030 043 082 034 029 023 036 040 042 027 034 040 062 034 019 018 030 037 034 034 055 202 255 255 255 255 068 094 201 252

052 034 044 050 078 051 040 050 039 039 053 038 064 027 063 062 042 048 035 037 101 057 030 023 026 034 041 027 027 036 063 024 029 019 022 031 030 053 081 255 255 255 210 067 128 128 090 054

037 044 045 061 077 094 062 081 080 045 048 085 062 027 047 044 039 037 044 038 198 044 026 026 035 022 033 028 060 039 059 022 025 025 025 029 031 058 171 253 235 067 061 093 129 133 126 025

032 058 059 058 049 051 044 043 035 039 051 118 054 027 036 044 034 048 060 032 034 026 022 027 044 035 022 060 031 059 023 032 022 021 017 022 158 186 217 055 134 104 072 117 110 021

044 049 041 045 048 049 044 041 040 038 039 038 035 028 035 047 041 032 046 034 140 041 026 024 034 034 024 026 016 018 058 026 029 027 032 037 020 099 029 093 087 055 116 030 018 061 105 094

086 068 068 067 091 095 092 095 096 096 089 089 077 066 060 060 050 029 032 037 130 074 024 017 035 027 030 023 022 017 052 027 027 019 022 023 010 022 043 116 047 040 103 118 129 060 046 037

086 101 095 095 064 077 081 052 082 080 079 078 086 073 082 088 031 036 033 033 112 067 017 019 030 057 050 049 048 040 045 017 127 056 018 084 130 131 121 125 112 049 098 060 087 104 062 034

071 073 058 076 076 078 080 038 082 081 079 075 085 081 084 047 062 041 039 035 090 068 023 023 025 086 078 080 072 068 053 029 113 033 018 118 119 131 117 116 091 044 074 042 114 104 107 052

055 062 041 069 088 084 087 051 081 182 193 156 174 066 087 090 025 040 034 033 068 021 021 020 028 089 087 098 072 025 068 024 070 059 102 128 138 130 106 080 061 018 022 021 102 104 117 114

072 066 063 081 068 069 075 048 092 159 098 076 027 069 095 105 026 030 036 037 058 017 015 017 048 088 120 113 115 106 067 070 087 068 149 138 093 083 058 061 055 017 013 031 072 080 076 108

038 043 042 038 036 047 042 035 066 080 087 077 078 072 038 076 024 030 031 035 045 018 016 026 051 042 122 054 057 070 099 079 091 102 099 105 098 108 083 087 088 090 085 076 074 070 069 086

018 022 024 023 024 025 049 030 056 052 056 046 036 070 035 047 025 030 030 034 057 042 032 023 028 022 058 089 107 114 121 123 121 123 113 104 051 068 069 074 084 093 090 083 085 096 091 094

032 030 029 037 030 025 049 023 037 077 057 066 037 059 030 020 025 021 022 025 064 048 047 036 096 069 078 025 096 109 129 119 117 110 111 090 085 149 064 085 082 089 086 091 095 092 094 094 096

031 030 021 039 025 028 062 165 037 053 056 040 026 053 056 071 025 067 065 065 069 071 097 098 114 114 122 132 118 123 122 119 117 101 113 055 083 081 088 088 095 091 089 089 095 093 097 096

046 074 040 042 019 028 062 032 064 034 063 032 029 054 062 046 020 077 080 088 098 095 113 111 109 112 105 112 105 109 098 095 080 116 080 131 063 085 092 089 094 095 089 090 094 094 094 093

028 041 019 030 013 029 055 034 020 042 071 047 034 053 091 098 114 108 096 111 099 100 099 099 095 092 085 107 101 089 087 114 106 054 127 089 088 087 091 093 091 097 094 092 094 092 095 084

045 048 020 042 052 039 062 070 077 080 087 102 096 061 106 095 098 098 096 084 095 104 102 101 098 106 096 093 094 040 064 061 094 109 087 097 093 087 091 089 088 084 081 084 077 077 078 086

065 072 072 073 077 093 083 103 108 104 086 096 070 058 103 096 099 086 093 095 063 094 090 084 096 083 129 096 091 074 102 135 080 098 090 093 090 081 077 075 077 078 079 089 100 093 095 093

What do you see here?

128 246 255 255 255 240 049 037 079 134 042 255 182 034 128 239 226 036 026 156 255 252 251 119
120 255 252 255 255 117 237 212 173 222 048 181 050 039 247 128 145 073 041 187 157 249 255 255
030 254 252 210 255 255 235 168 067 129 055 031 040 021 100 021 224 020 113 147 163 255 255 252
022 255 240 162 255 254 255 038 151 134 229 021 035 146 053 111 176 026 068 033 255 235 255 246
066 049 057 051 226 255 197 105 052 230 048 020 031 218 253 173 255 028 034 100 156 255 255 165
032 068 110 133 046 044 044 032 037 036 041 035 027 030 031 150 255 020 025 255 211 177 247 253
044 070 086 057 084 067 075 026 067 075 048 037 028 034 065 060 026 044 031 220 170 255 208 190
057 058 043 076 079 070 070 036 052 123 161 032 027 080 024 058 027 032 027 027 255 255 254 241
065 058 054 066 117 064 078 030 044 052 073 034 036 046 048 069 024 026 031 141 255 255 208 255
052 044 078 040 039 053 064 063 042 035 101 030 026 041 027 063 029 022 030 081 255 210 128 090
032 059 049 044 035 051 054 036 034 060 150 026 027 035 060 059 032 021 022 186 055 104 117 100
086 068 091 092 096 089 077 060 050 032 130 024 035 030 022 052 027 022 010 043 047 103 129 046
071 058 076 080 082 079 085 084 062 039 090 023 025 078 072 053 113 018 119 117 091 074 114 107
072 063 068 075 092 098 027 095 026 036 058 015 048 120 115 087 087 149 093 058 055 013 072 076
018 024 024 049 056 056 036 035 025 030 057 032 028 058 107 121 121 113 051 069 084 090 085 091
031 021 025 062 037 056 026 056 025 065 069 097 114 122 118 122 117 113 083 088 095 089 095 097
028 019 013 055 020 071 034 091 114 096 099 099 095 085 101 087 106 127 088 091 091 094 094 095
065 072 077 083 108 086 070 103 099 093 093 090 096 129 091 102 080 090 090 077 077 079 100 095

What do you see here?



What a computer sees

128 246 255 255 255 240 049 037 079 134 042 255 182 034 128 239 226 036 026 156 255 252 251 119
120 255 252 255 255 117 237 212 173 222 048 181 050 039 247 128 145 073 041 187 157 249 255 255
030 254 252 210 255 255 235 168 067 129 055 031 040 021 100 021 224 020 113 147 163 255 255 252
022 255 240 162 255 254 255 038 151 134 229 021 035 146 053 111 176 026 068 033 255 235 255 246
066 049 057 051 226 255 197 105 052 230 048 020 031 218 253 173 255 028 034 100 156 255 255 165
032 068 110 133 046 044 044 032 037 036 041 035 027 030 031 150 255 020 025 255 211 177 247 253
044 070 086 057 084 067 075 026 067 075 048 037 028 034 065 060 026 044 031 220 170 255 208 190
057 058 043 076 079 070 070 036 052 123 161 032 027 080 024 058 027 032 027 027 255 255 254 241
065 058 054 066 117 064 078 030 044 052 073 034 036 046 048 069 024 026 031 141 255 255 208 255
052 044 078 040 039 053 064 063 042 035 101 030 026 041 027 063 029 022 030 081 255 210 128 090
032 059 049 044 035 051 054 036 034 060 150 026 027 035 060 059 032 021 022 186 055 104 117 100
086 068 091 092 096 089 077 060 050 032 130 024 035 030 022 052 027 022 010 043 047 103 129 046
071 058 076 080 082 079 085 084 062 039 090 023 025 078 072 053 113 018 119 117 091 074 114 107
072 063 068 075 092 098 027 095 026 036 058 015 048 120 115 087 087 149 093 058 055 013 072 076
018 024 024 049 056 056 036 035 025 030 057 032 028 058 107 121 121 113 051 069 084 090 085 091
031 021 025 062 037 056 026 056 025 065 069 097 114 122 118 122 117 113 083 088 095 089 095 097
028 019 013 055 020 071 034 091 114 096 099 099 095 085 101 087 106 127 088 091 091 094 094 095
065 072 077 083 108 086 070 103 099 093 093 090 096 129 091 102 080 090 090 077 077 079 100 095

What you see



What a computer sees

128 246 255 255 255 240 049 037 079 134 042 255 182 034 128 239 226 036 026 156 255 252 251 119
120 255 252 255 255 117 237 212 173 222 048 181 050 039 247 128 145 073 041 187 157 249 255 255
030 254 252 210 255 255 235 168 067 129 055 031 040 021 100 021 224 020 113 147 163 255 255 252
022 255 240 162 255 254 255 038 151 134 229 021 035 146 053 111 176 026 068 033 255 235 255 246
066 049 057 051 226 255 197 105 052 230 048 020 031 218 253 173 255 028 034 100 156 255 255 165
032 068 110 133 046 044 044 032 037 036 041 035 027 030 031 150 255 020 025 255 211 177 247 253
044 070 086 057 084 067 075 026 067 075 048 037 028 034 065 060 026 044 031 220 170 255 208 190
057 058 043 076 079 070 070 036 052 123 161 032 027 080 024 058 027 032 027 027 255 255 254 241
065 058 054 066 117 064 078 030 044 052 073 034 036 046 048 069 024 026 031 141 255 255 208 255
052 044 078 040 039 053 064 063 042 035 101 030 026 041 027 063 029 022 030 081 255 210 128 090
032 059 049 044 035 051 054 036 034 060 150 026 027 035 060 059 032 021 022 186 055 104 117 100
086 068 091 092 096 089 077 060 050 032 130 024 035 030 022 052 027 022 010 043 047 103 129 046
071 058 076 080 082 079 085 084 062 039 090 023 025 078 072 053 113 018 119 117 091 074 114 107
072 063 068 075 092 098 027 095 026 036 058 015 048 120 115 087 087 149 093 058 055 013 072 076
018 024 024 049 056 056 036 035 025 030 057 032 028 058 107 121 121 113 051 069 084 090 085 091
031 021 025 062 037 056 026 056 025 065 069 097 114 122 118 122 117 113 083 088 095 089 095 097
028 019 013 055 020 071 034 091 114 096 099 099 095 085 101 087 106 127 088 091 091 094 094 095
065 072 077 083 108 086 070 103 099 093 093 090 096 129 091 102 080 090 090 077 077 079 100 095

- Vision is our most powerful sense
- Unlike other senses, vision is developed after birth
- We are born without the capability to **see**



4 months: focusing, hand-eye coordination and interest in faces

6 months: depth perception and colour vision

9 months: precision grasping and interaction

12 months: object recognition

How do we learn something as complex as vision?

- Suppose, a baby experiences 1 *saccade* per second, for 8 hours a day for 365 days
- $1 \times 60 \times 60 \times 8 \times 365 = 10,000,000$ training examples to learn to see

1. Why do machines need to learn to see?
2. How do we teach machines to see?
3. Two exciting applications

“What is where by looking” - Aristotle 384-322 BC

“Perception is our best guess
as to what is in the world,
given our current sensory
input and our *prior experience*”

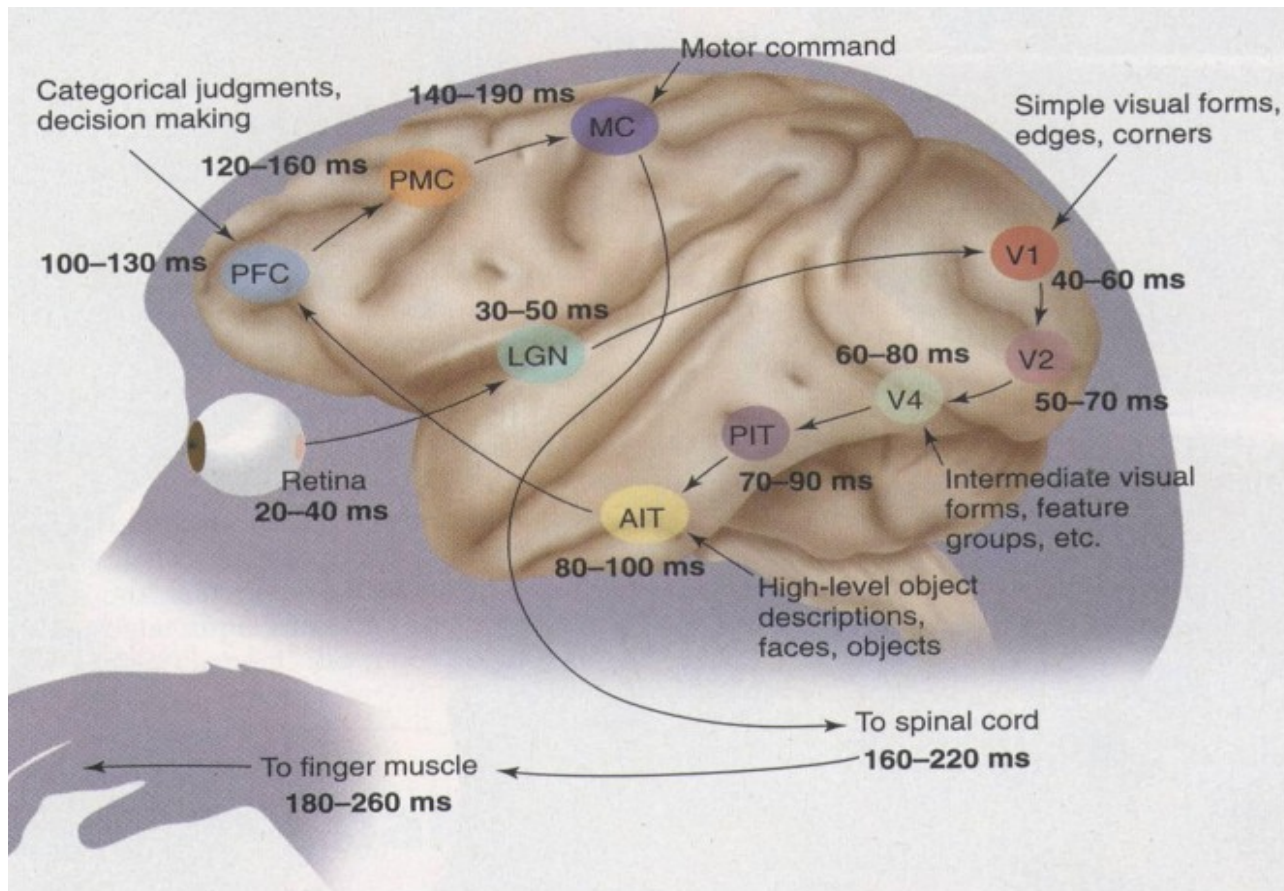
Helmholtz 1866

This prior experience is *learning*

Adelbert Ames' room in 1934



- Retinal image from eyes (100MP)
- Human visual system has over 60 billion neurons
- Visual cortex transforms what we see into high level understanding



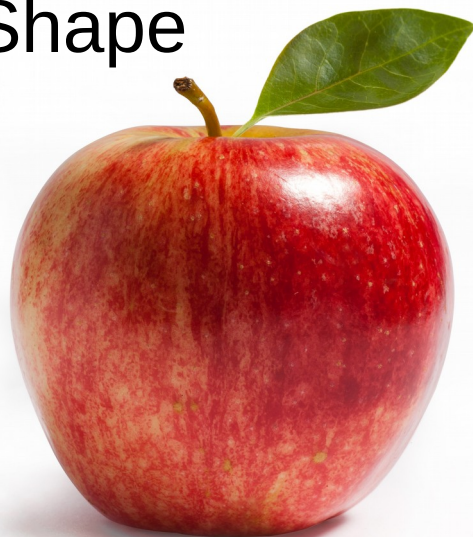
Let's design a basic visual recognition system!

How do we tell if something is an apple or an orange?

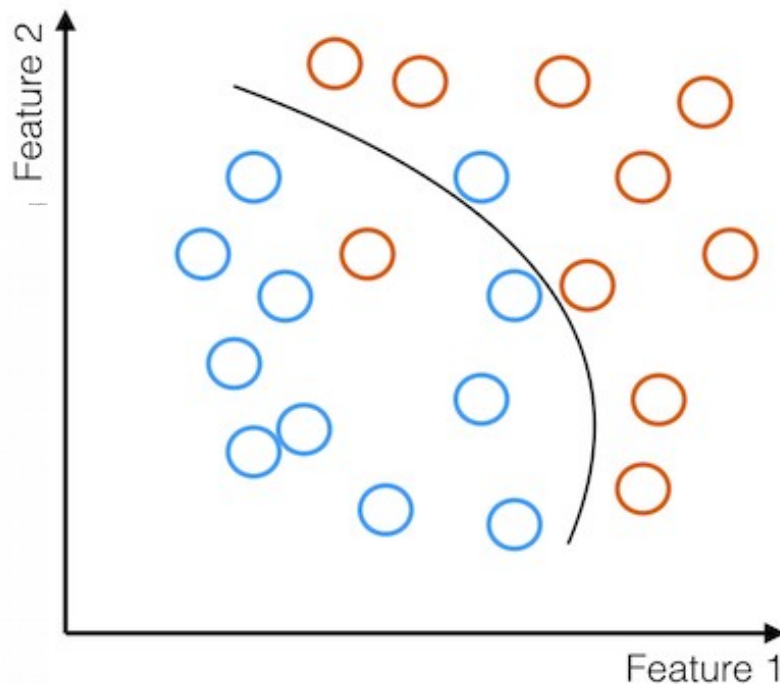


We need to extract features

- Colour
- Shape
- Size
- Texture



We can design classification rules for each feature



Now what?



- How do we design these parameters for 1,000+ different objects?
- How do we choose the features which describe these?
- How do we understand the confidence in our model?

We can't design this by hand – we need to learn!

- Machine learning allows us to learn features and classification rules
- Deep learning allows us to learn a powerful hierarchical representation
- Deep learning is used today in speech, image and language recognition; autonomous driving, big data analysis and artificial intelligence

Why is it important to learn deep, hierarchical representations?

It allows us to understand complex data:

Edges/Corners



Lines/Textures



Objects



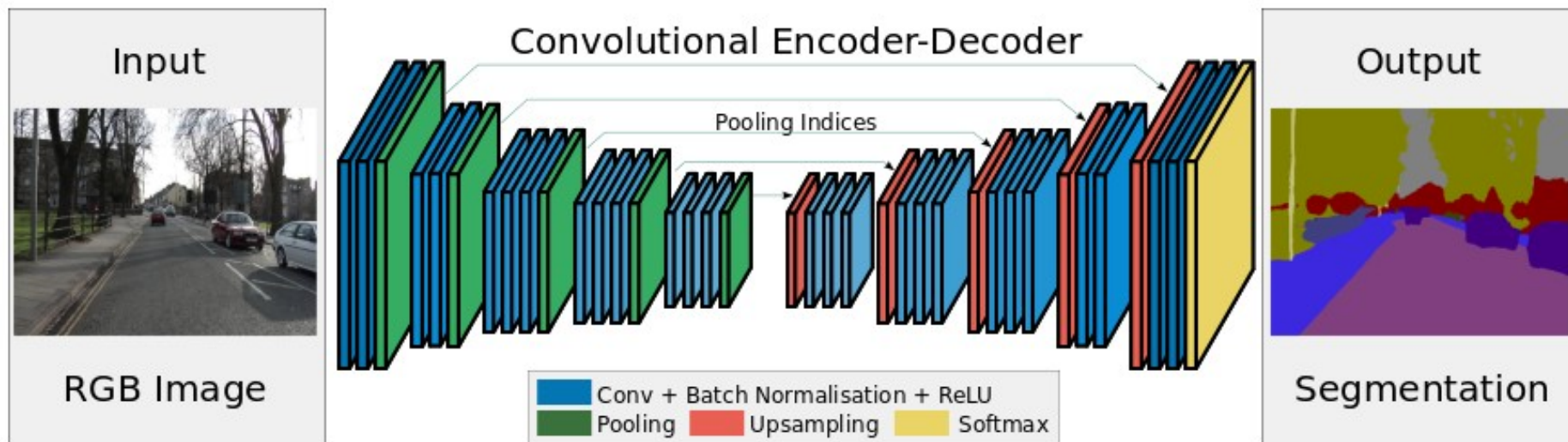
Scene



Deep convolutional neural network

Layers map input to higher dimensional representations

26 layers, 30 million parameters



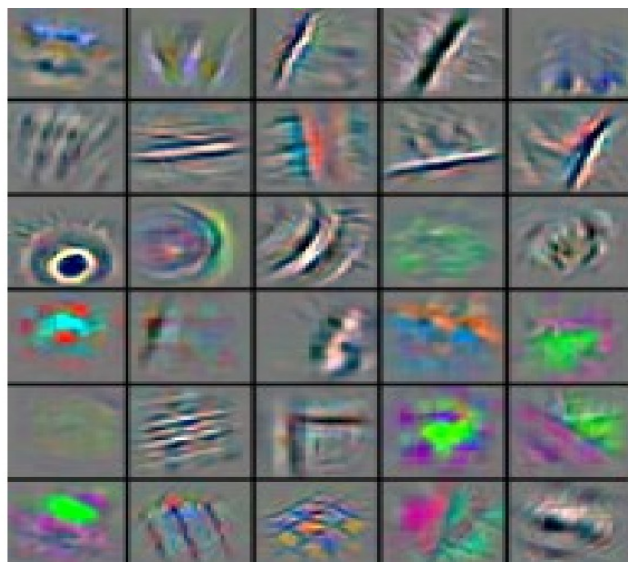
Model is trained with *Stochastic Gradient Descent*

Optimises these millions of parameters towards an objective

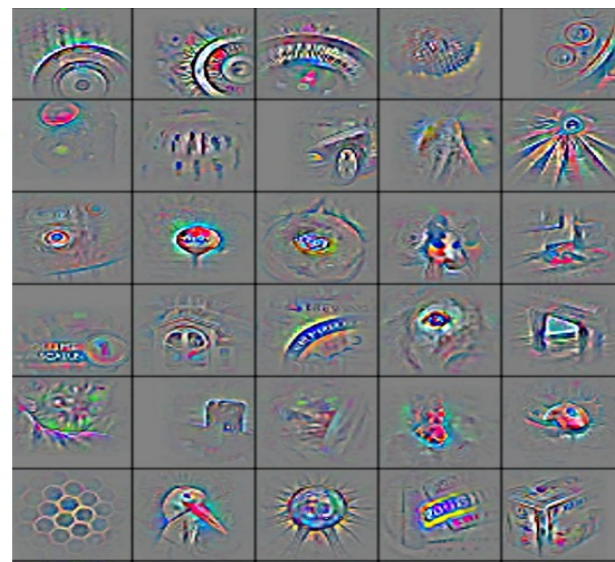
Low-level features



Mid-level features

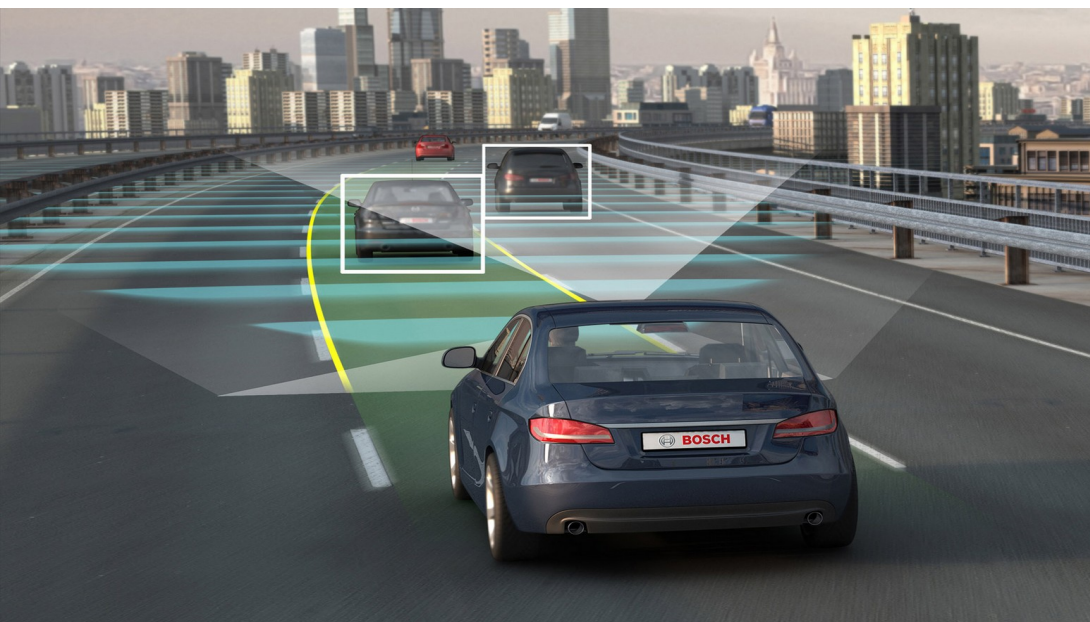


High-level features



Source: Zeiler and Fergus, 2014. “Visualizing and Understanding Convolutional Networks”

- Where am I?
- What is around me?



Autonomous Vehicles



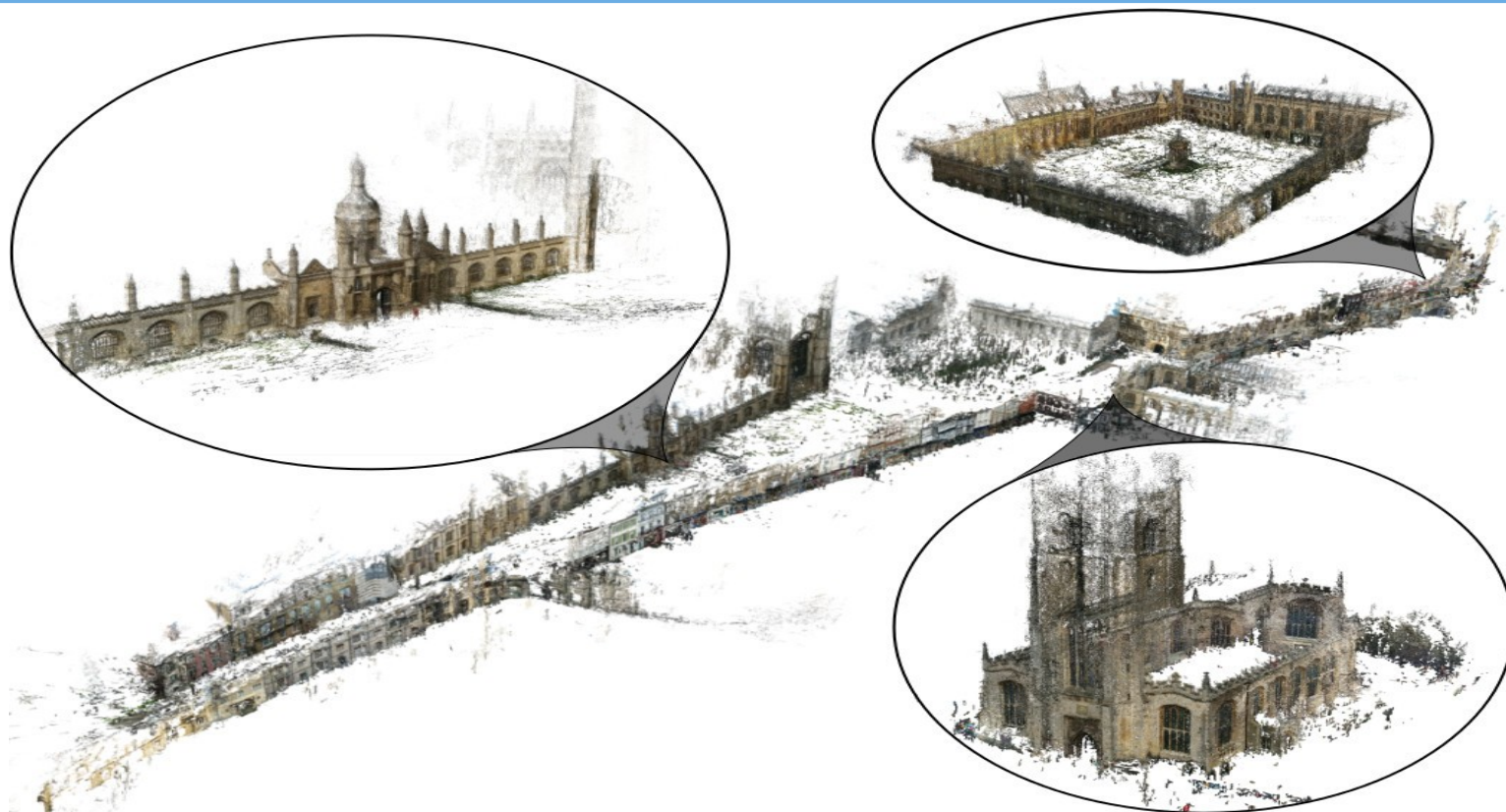
Augmented Reality

Application 1: Localisation



- Relocalisation is the problem of determining where you are in a previously explored area
- Also known as the *Robot Kidnap Problem*
- Training data: images and their location
- How do we annotate images with their location?
Surveying/GPS?

Let's collect some training data!



Code demonstrated from ElasticFusion:

<http://www.imperial.ac.uk/dyson-robotics-lab/downloads/elastic-fusion/>

LOCALISATION DEMO FOR CENTRAL CAMBRIDGE, UNITED KINGDOM

Paste an image url here

SUBMIT

 OR UPLOAD AN IMAGE FILE

Or use one of these example images that we obtained from the internet:



We have localised your input image on the map below! The **blue arrow** shows where we think you are. The image must have been taken within the blue highlighted region on the map.

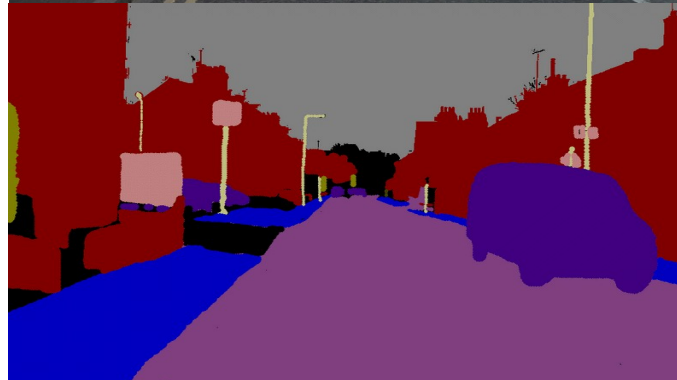
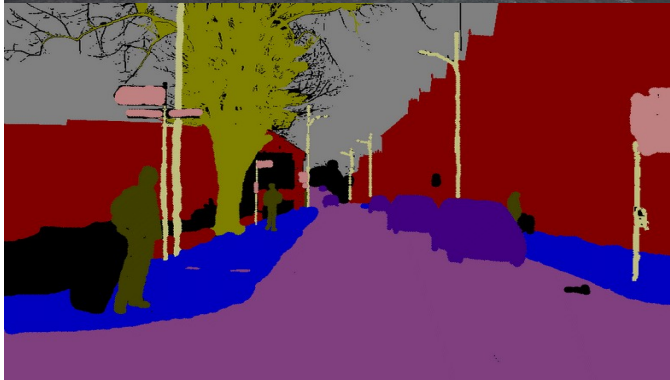
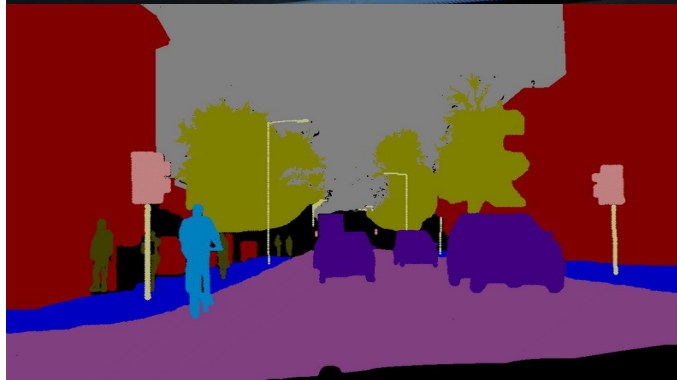


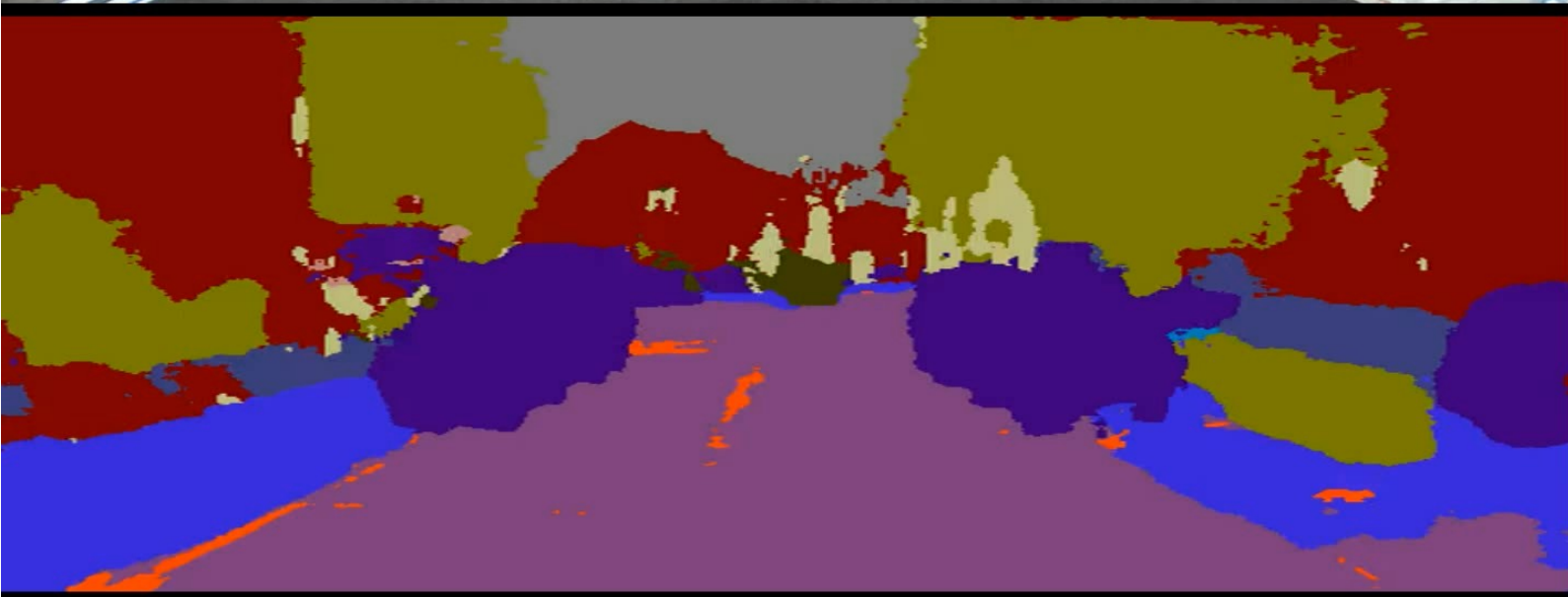
This is the closest view in Google Maps Street View to the blue arrow.










Application 2: Scene Understanding



Training data = images and labelled examples





- | | |
|---|--------------|
|  | Sky |
|  | Building |
|  | Pole |
|  | Road Marking |
|  | Road |
|  | Pavement |
|  | Tree |
|  | Sign Symbol |
|  | Fence |
|  | Vehicle |
|  | Pedestrian |
|  | Bike |



SegNet

[Alex Kendall](#), [Vijay Badrinarayanan](#), [Roberto Cipolla](#)

Try the demo

SegNet


A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling

Use a random image, upload your own, search for a place, or click on one of the example images in the gallery below. SegNet is trained to classify each pixel of an urban street image to be one of twelve classes.

Select a Country



Get Random Image

 Upload an Image File

OR Paste Image URL

OR Search for a Place, e.g. Trinity Street Cambridge

Process with SegNet

Thank you to Prof. Roberto Cipolla and my colleagues at the University of Cambridge

- Tomaso Poggio, MIT Center for Brains, Minds and Machines
cbmm.mit.edu/views-reviews/article/what-if
- Alex Kendall's Webpage
mi.eng.cam.ac.uk/~agk34/