Teaching machines to see
A quest to visual intelligence

Lamberto Ballan
University of Padova, Italy
Artificial Intelligence

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Eyes and Vision
What is (computer) vision?

Image / video  

Sensing device  

“Interpreting” device  

Interpretations  

dog, flower, grass, green, spring, ...

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
The goal of computer vision

• Bridge the gap between pixels and “meaning”
Why study computer vision?

• Big visual data: almost 90% of web data is visual

Image credit: L. Fei-Fei
Applications

Digit recognition

Biometric

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Applications

Face detection

Face recognition

Face ID

Your face is your secure password.

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”

Applications

Visual search

Healthcare
Applications (not far from now…)

Retail/Supermarket

Self-driving cars

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Vision is challenging: *let’s play a game* (1)
Vision is challenging: *let’s play a game* (1)
Feb 2015: people across social media have been arguing about whether this picture depicts a dress as blue with black lace fringe or white with gold lace fringe.
Vision is challenging: *let’s play a game* (2)
Vision is challenging: *let's play a game* (2)

Lamberto Ballan: "Teaching machines to see: a quest to visual intelligence"
Vision is challenging: *let’s play a game* (3)
Vision is challenging: *let’s play a game* (3)
Vision is challenging: *let's play a game* (3)
Vision is challenging: *let’s play a game* (3)
Vision is challenging: let's play a game (3)
Vision begins with the eyes…

… but truly occurs in the brain!

**Functional specialization:**
- V1: primary visual cortex
- V4: color, V3/VP: recognition
- MT/V5: motion
Teaching machines to see

Input → Output

“dog”
Teaching machines to see

Input → Output

“dog”

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Teaching machines to see

Input

Machine Learning

Output

"car"
"dog"
"cat"
"person"
"tree"
"chair"
"Bike"
"sky"
"computer"
...
Teaching machines to see

Deep Learning
Convolutional Neural Network

Input

15M images, 22K classes

Output

“car”
“dog”
“cat”
“person”
“tree”
“chair”
“Bike”
“sky”
“computer”

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”
Teaching machines to see

Deep Learning
Convolutional Neural Network

Neuron-like unit

\[
\begin{aligned}
  x_1 & \\
  x_2 & \\
  x_3 & \\
\end{aligned}
\]

\[ f(\cdot) \rightarrow y \]
Teaching machines to see

Deep Learning
Convolutional Neural Network

Neuron-like unit
Teaching machines to see

Convolutional Neural Network

Lamberto Ballan: “Teaching machines to see: a quest to visual intelligence”

24 M nodes, 140 M parameters, 15 B connections
Teaching machines to see
Contact

✉️ lamberto.ballan@unipd.it
🏠 www.math.unipd.it/~ballan
@ twitter.com/lambertoballan

Thanks!