UNIVERSITÀ DEGLI STUDI
DI MILANO

Corso di

Visione Artificiale

Laurea Magistrale in Informatica (F94)

Docente:

Raffaella Lanzarotti

Dipartimento di Informatica
Università degli Studi di Milano

Informazioni generali

- **Sito web del corso** https://github.com/lanzarotti/Visione-Artificiale-2025-26

- **Orario delle Lezioni**:
  - *martedì e giovedì 10:00 ÷ 12:30, Lab. LM 5 piano*

- **Materiale didattico**
  - Testi (disponibili online)
    - D.A. Forsyth, J. Ponce, Computer Vision – A Modern Approach – 2e, Pearson
    - C.M. Bishop, H. Bishop - Deep Learning - Foundations and Concepts - Springer
  - *Slide / materiale: sul sito del corso*

Informazioni generali

- **Modalità delle lezioni**
  - *Frontale*
  - *Demo in python*

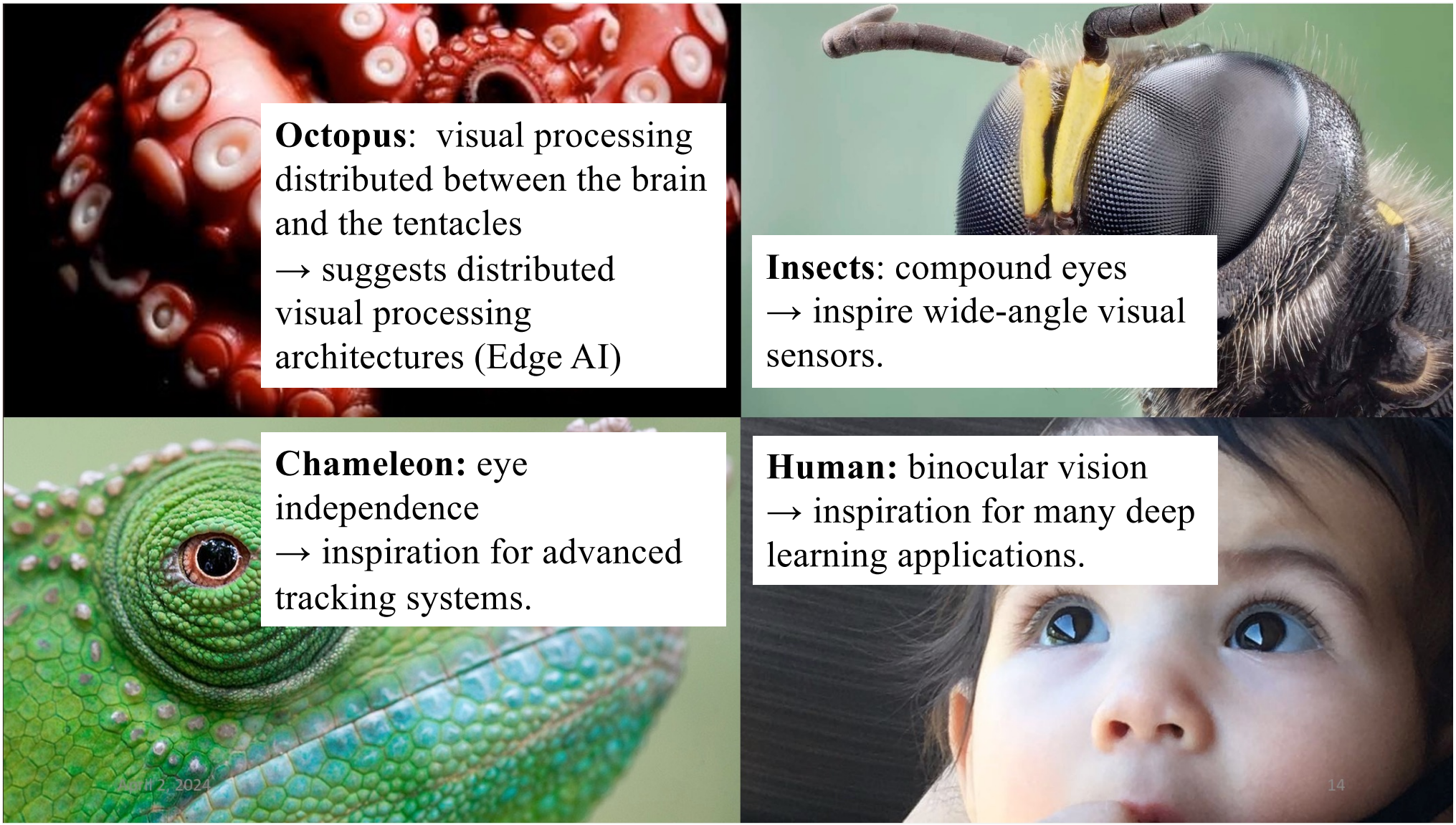
- **Modalità d'esame**
  - *Prova teorica scritta (consistente di domande aperte riguardanti tutti gli argomenti trattati a lezione)*

  OPPURE
  - *Progetto (consistente nell'implementazione in Python, la stesura di una relazione e la presentazione della soluzione proposta e dei risultati ottenuti. A fine corso verranno date le possibili tracce da seguire per realizzare il progetto)*

INTRODUCTION TO COMPUTER VISION

Credits: Luigi Cinque

What is Vision?



Octopus: visual processing distributed between the brain and the tentacles
→ suggests distributed visual processing architectures (Edge AI)

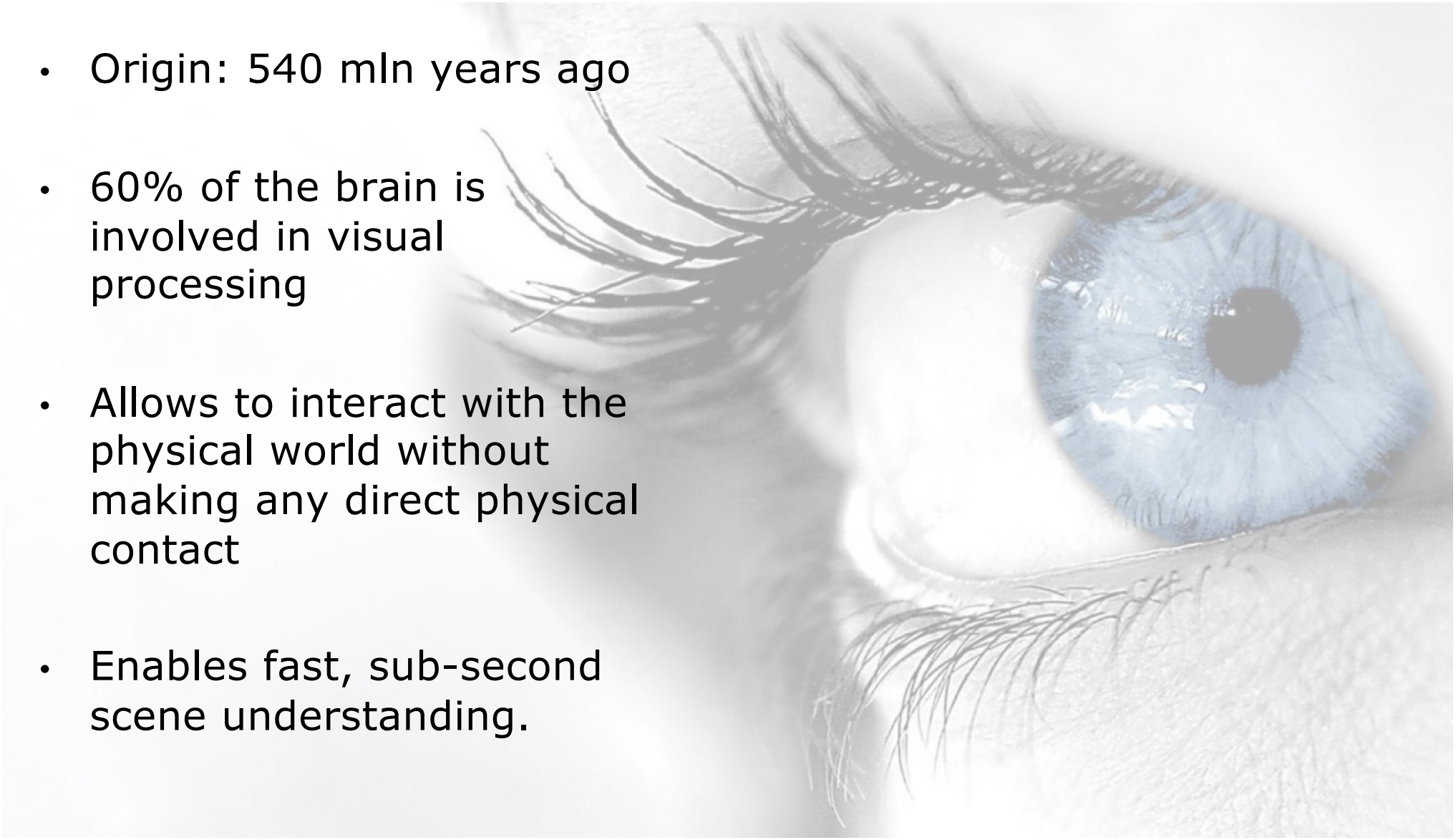
Insects: compound eyes
→ inspire wide-angle visual sensors.

Chameleon: eye independence
→ inspiration for advanced tracking systems.

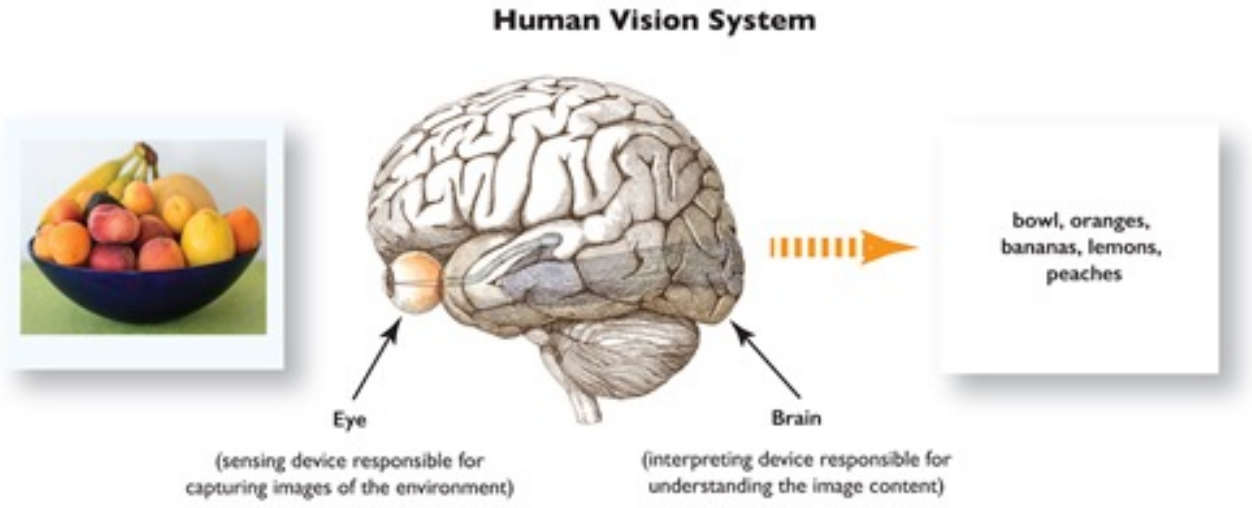
Human: binocular vision
→ inspiration for many deep learning applications.

What is Human Vision?

- Origin: 540 mln years ago
- 60% of the brain is involved in visual processing
- Allows to interact with the physical world without making any direct physical contact
- Enables fast, sub-second scene understanding.



What is Computer Vision?



COMPUTER VISION:

- Emulating the (human) visual system

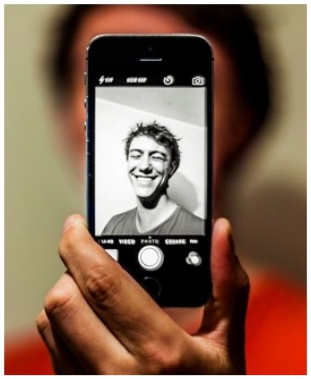
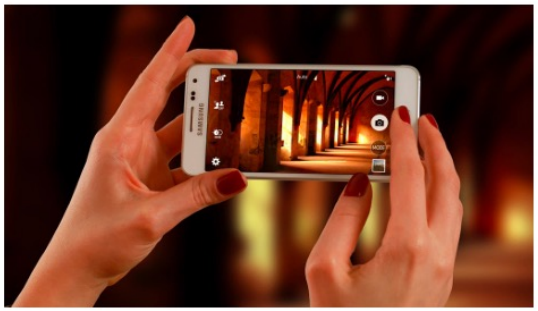
- Inputs extend beyond RGB (e.g., infrared, depth)

https://freecontent.manning.com/mental-model-graphic-grokking-deep-learning-for-computer-vision/

Why study computer vision?

- Computer vision is everywhere!

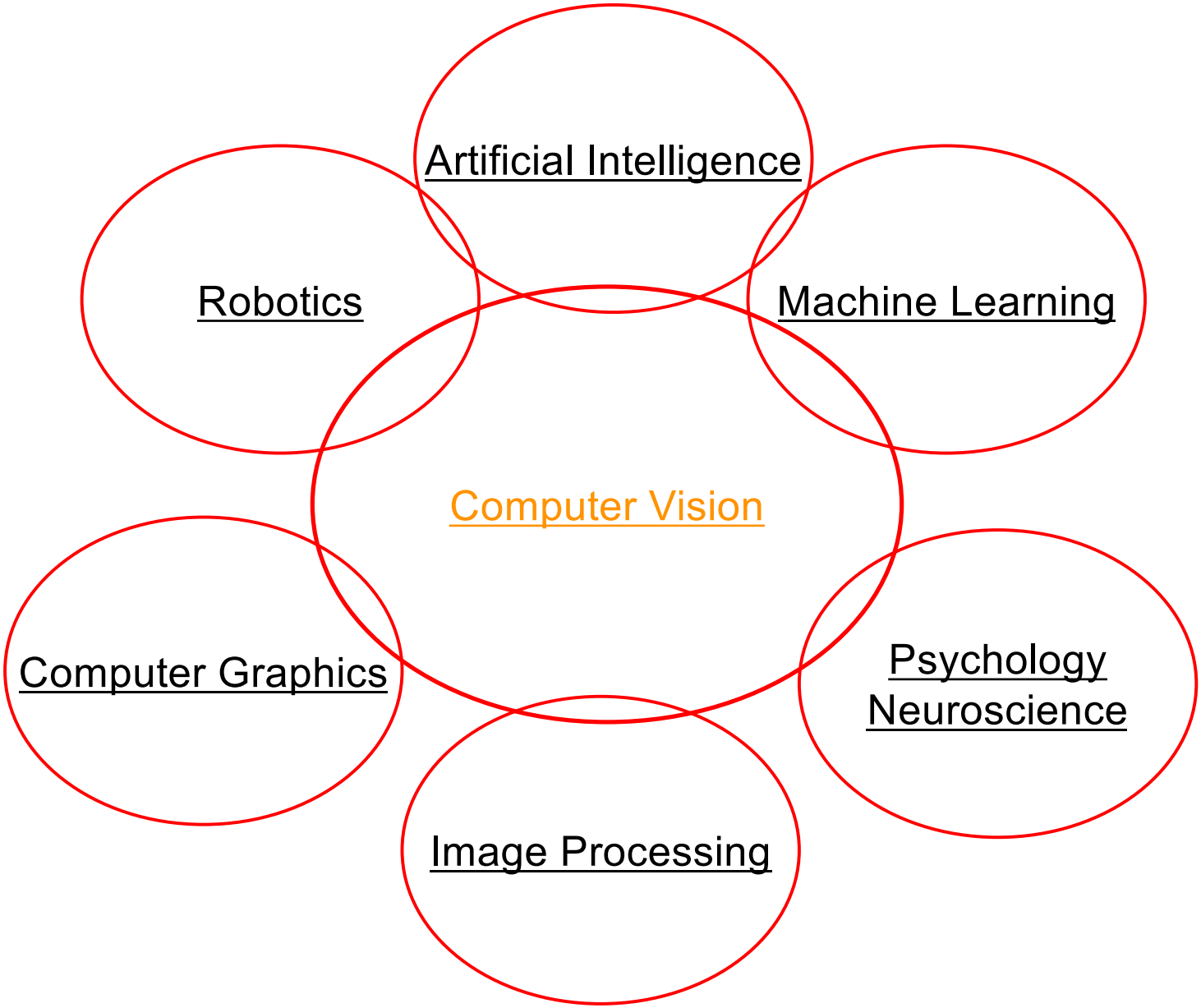


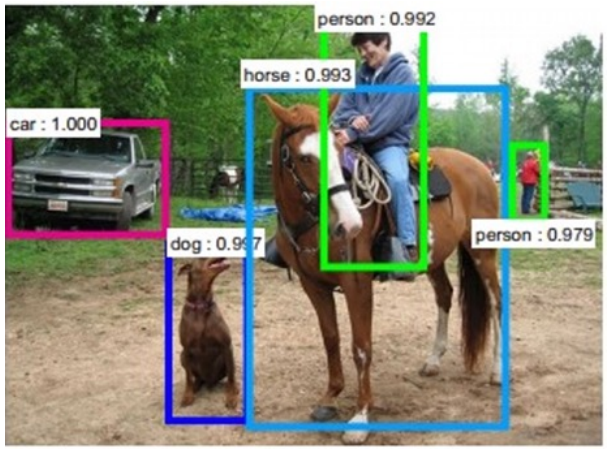





Connections to other disciplines



Applications of computer vision



Object detection

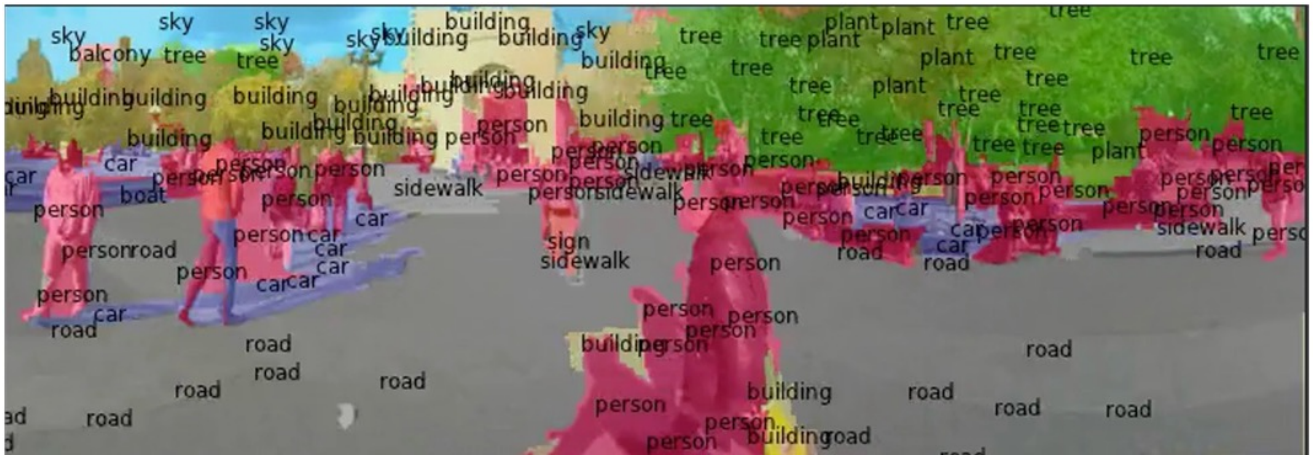


Image Segmentation

Human Skeleton

A white teddy bear sitting in the grass

Image Captioning

A sofa imitating a banana

Text-to-Image

The goal of computer vision

- To perceive the "world behind the picture"

153	156	148	152	149	147	139	146	142	150	146	144	137	125	120	119	136	146	151	164	172	175	183	188	196	200	205	208	214	214	219	217
159	151	150	148	140	138	139	129	119	104	86	82	89	97	107	115	118	130	128	132	128	144	160	168	179	188	200	208	213	220	212	214
149	146	153	147	147	146	132	99	73	78	87	96	105	120	138	151	145	157	163	171	165	161	146	126	157	184	190	201	215	212	214	214
145	150	154	148	148	126	93	67	72	78	96	107	117	127	131	134	127	154	166	167	183	194	200	195	143	140	175	190	197	203	206	207
151	153	151	147	120	85	67	75	84	83	94	92	81	78	78	91	83	117	126	144	178	200	201	203	208	175	127	159	185	196	195	206
146	144	139	123	79	66	74	83	79	69	64	62	58	50	46	54	54	66	60	80	86	108	141	191	184	200	187	123	144	175	198	199
135	130	115	87	64	77	90	79	78	85	81	63	55	57	56	53	70	62	61	68	59	58	84	105	168	194	196	183	131	151	185	197
128	116	92	71	82	94	103	101	83	101	88	66	70	90	80	42	39	53	88	73	76	82	116	87	97	144	188	195	190	166	171	203
135	120	84	83	108	127	135	115	100	92	79	49	85	74	59	0	0	0	50	69	52	79	157	141	100	84	136	187	206	204	189	200
144	103	91	115	139	147	127	91	87	80	72	44	61	84	25	0	0	0	50	181	45	69	142	164	167	113	93	130	193	199	208	203
139	102	123	143	137	131	109	85	93	84	68	47	77	86	31	0	3	0	51	156	53	75	141	169	199	151	171	108	143	181	199	208
141	135	153	142	114	104	97	97	83	98	77	42	77	96	79	21	0	23	58	46	56	77	155	199	212	161	194	193	164	187	202	205
160	172	164	141	128	112	98	95	100	96	91	73	68	86	75	73	64	65	54	69	77	115	190	212	193	181	174	188	210	194	202	207
179	189	160	140	139	116	97	97	108	103	110	99	75	80	72	83	50	55	54	95	98	174	205	185	179	188	185	190	193	217	217	224
189	183	152	130	121	105	105	117	114	108	107	115	110	81	85	85	87	81	81	124	183	202	175	180	178	171	173	204	225	215	219	225
178	161	149	135	120	115	122	129	137	145	131	121	125	115	109	91	92	111	132	159	173	170	184	176	184	190	191	217	210	226	228	223
187	159	139	127	125	115	118	121	121	131	133	134	140	137	134	139	140	152	141	154	170	163	195	194	176	198	216	209	219	224	223	226
185	164	140	122	116	110	109	108	113	118	115	116	123	127	135	148	154	162	165	170	171	160	183	198	201	210	223	216	221	222	221	226
188	175	150	130	118	117	113	110	108	115	117	123	130	132	138	150	157	158	174	182	189	186	198	221	224	221	227	221	223	218	218	222
187	179	158	141	124	127	125	127	126	129	130	135	139	141	150	165	175	172	185	195	207	210	212	226	229	222	224	224	223	218	219	221
188	184	172	159	138	135	135	143	143	143	144	146	145	147	160	174	184	191	199	207	211	213	217	224	227	223	223	221	221	218	224	223
192	191	187	174	153	139	140	147	146	149	157	162	160	159	165	174	181	198	201	210	212	216	223	224	225	225	220	215	217	215	224	224

Computers see numbers.

The goal of computer vision

- To perceive the "world behind the picture"

Humans see objects and scenes; Computers see numbers.
Computer vision bridges this gap.

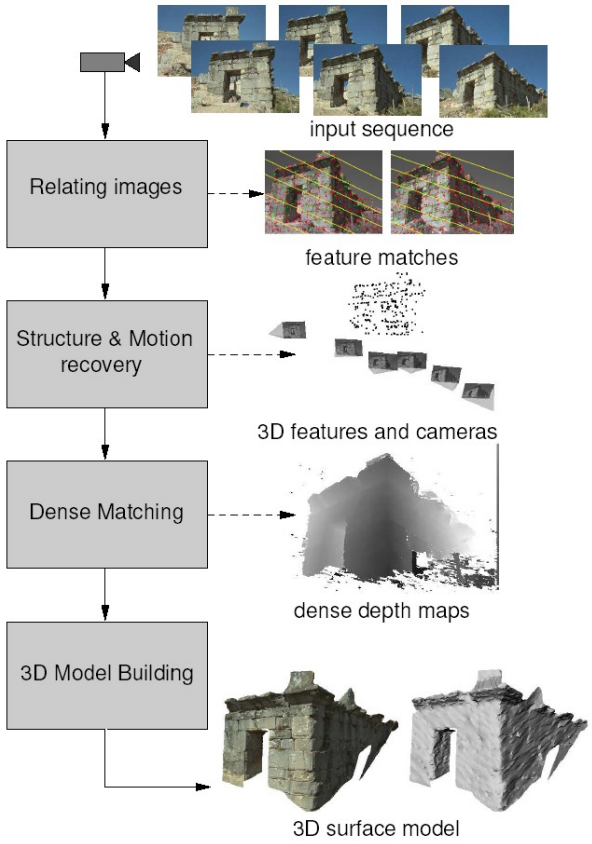
The goal of computer vision

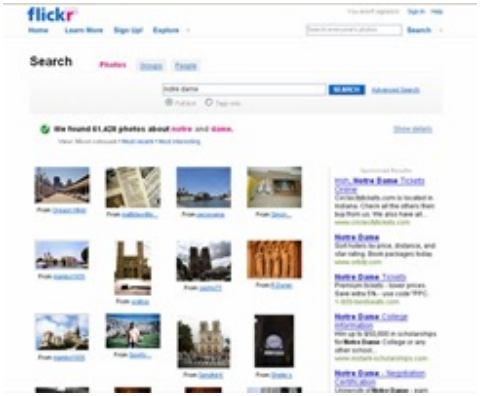
- To perceive the "world behind the picture"

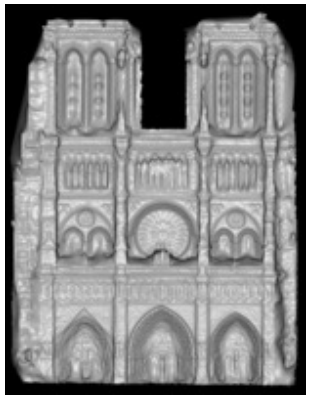
- What exactly does this mean?
  - *Vision as a source of metric 3D information*

  - *Vision as a source of semantic information*

Vision as measurement device







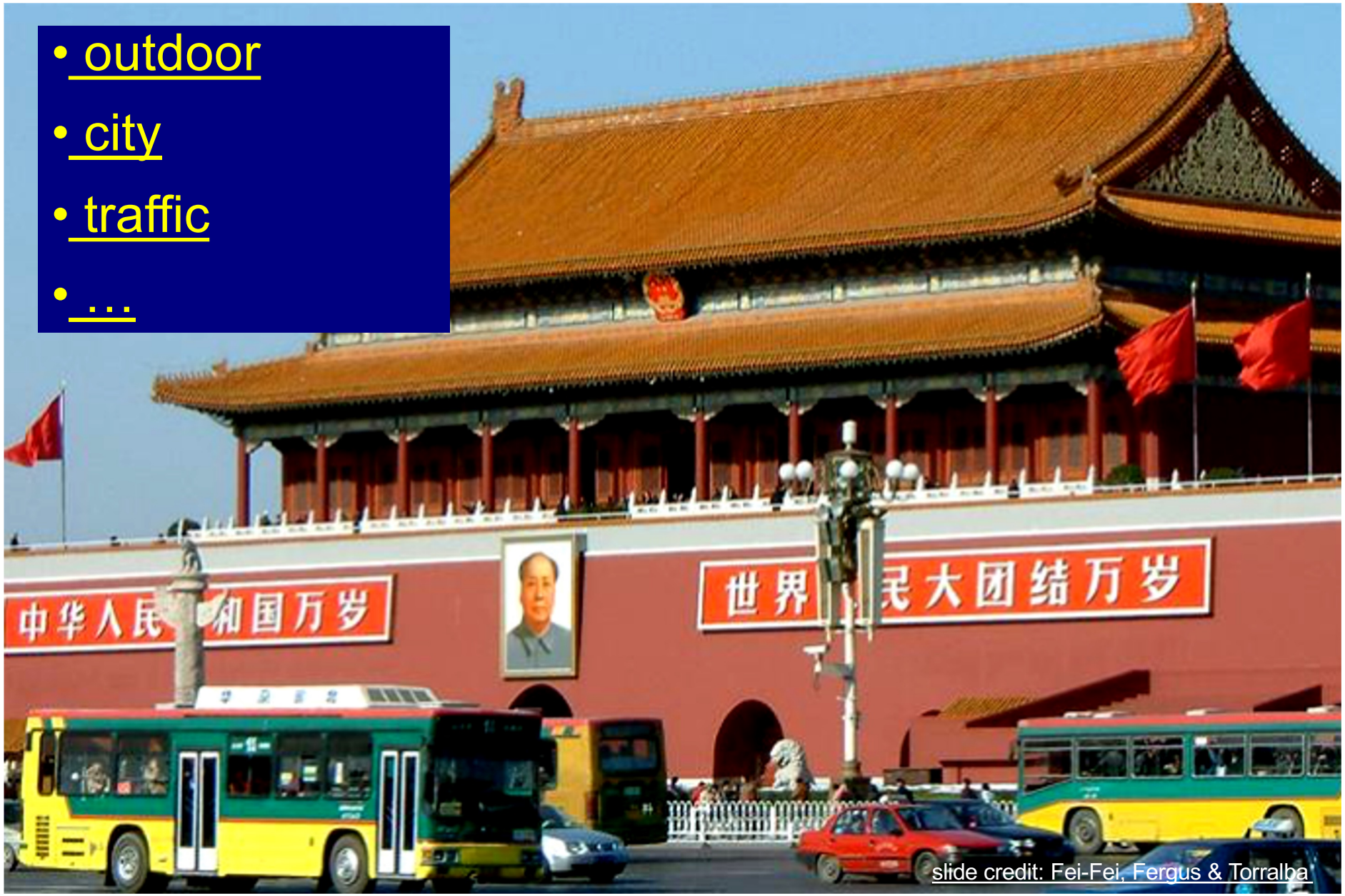
Amazing success story,
now even more incredible: SAM 3D:
https://www.aidemos.meta.com/segment-anything/editor/convert-image-to-3d

Vision as a source of semantic information

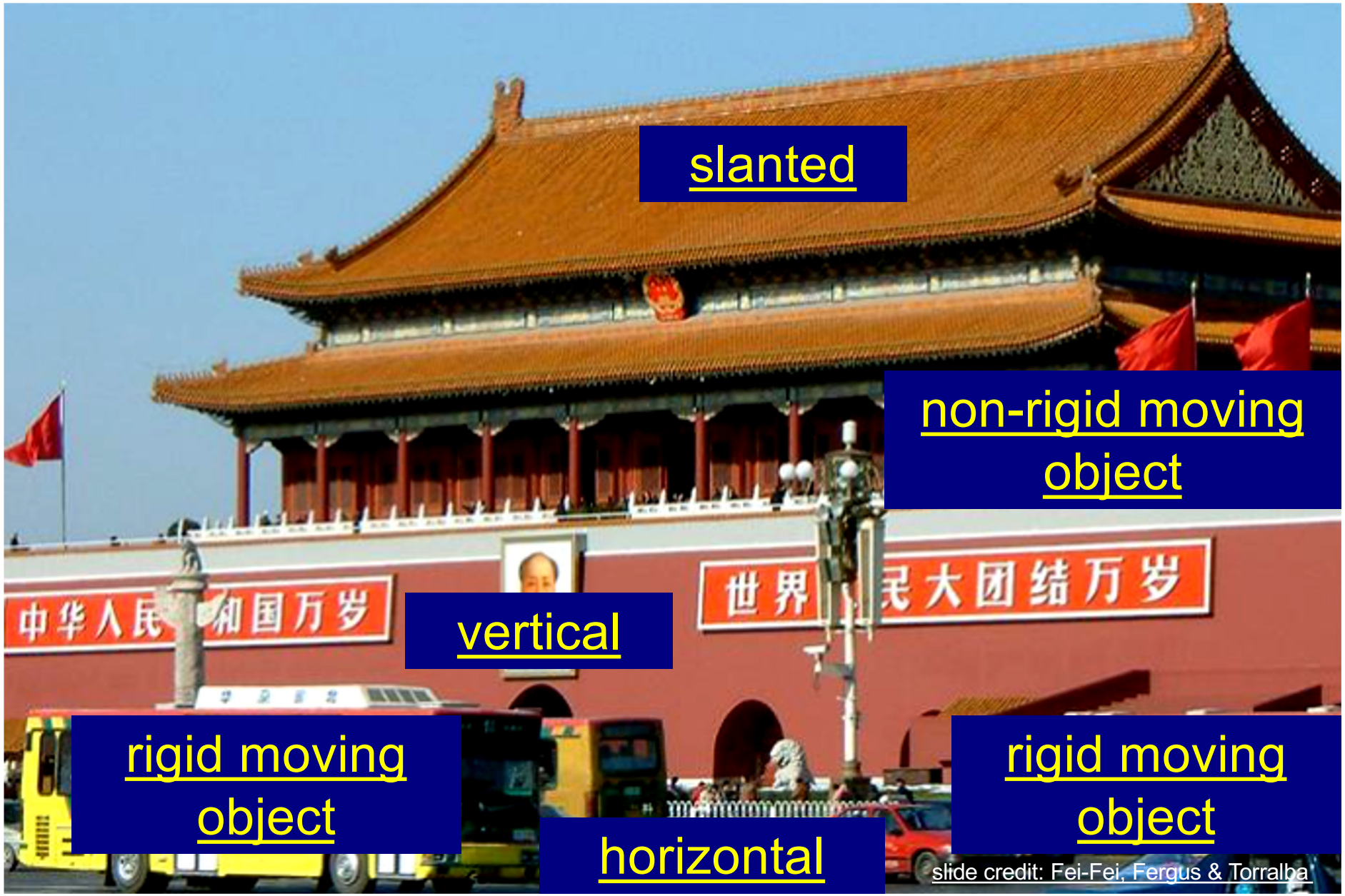
slide credit: Fei-Fei, Fergus & Torralba

Object categorization

Scene and context categorization



Qualitative spatial information



Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

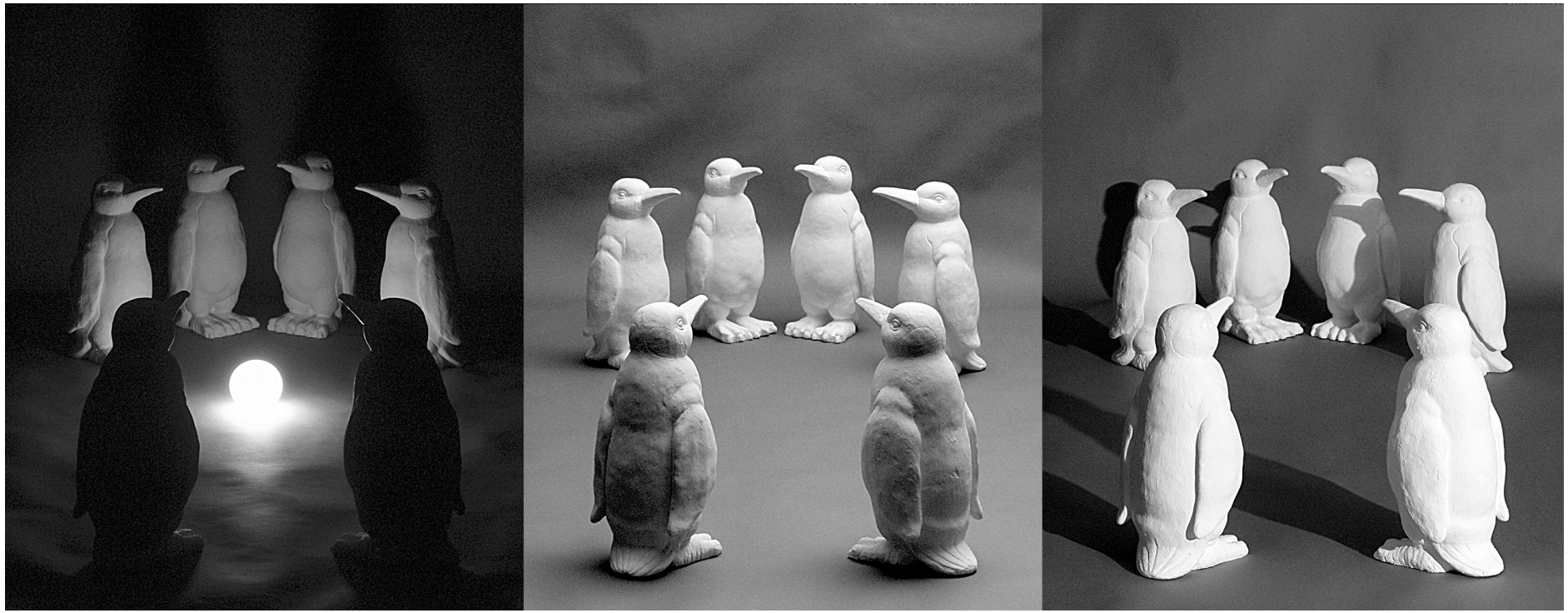
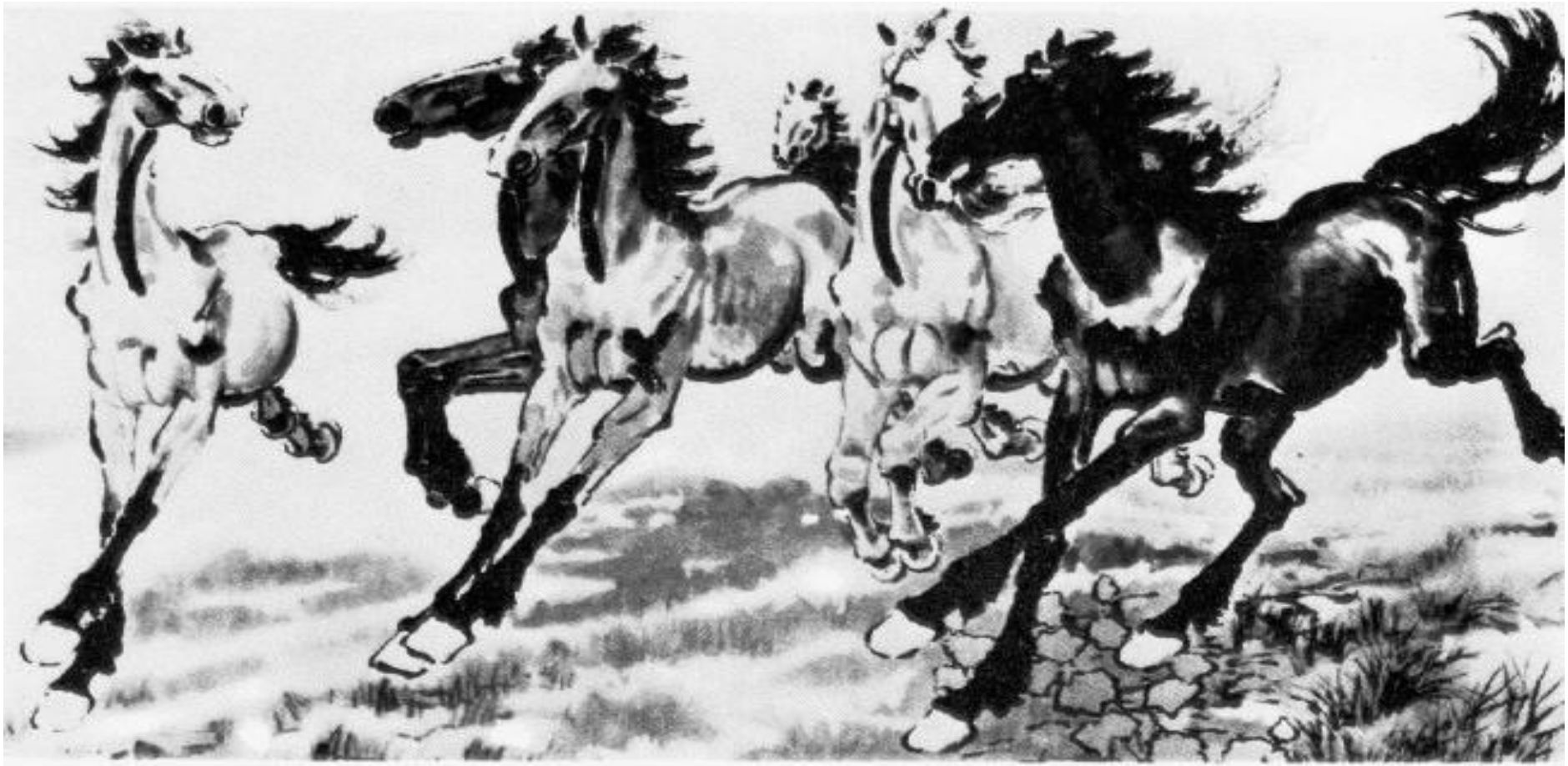


image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Challenges: occlusion

Magritte, 1957

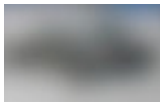
slide credit: Fei-Fei, Fergus & Torralba

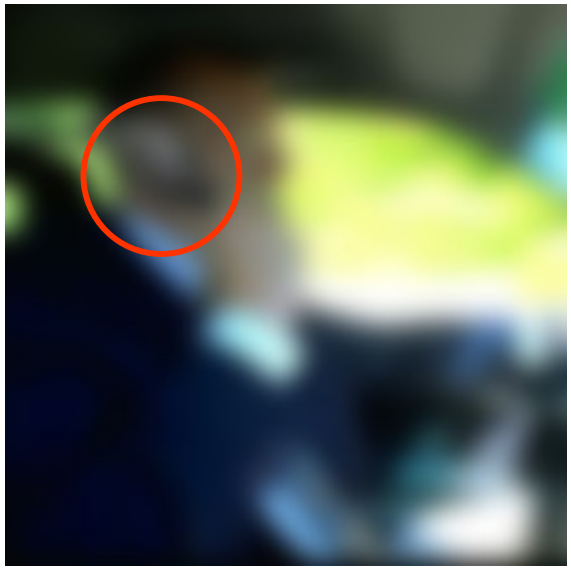
Challenges: background clutter

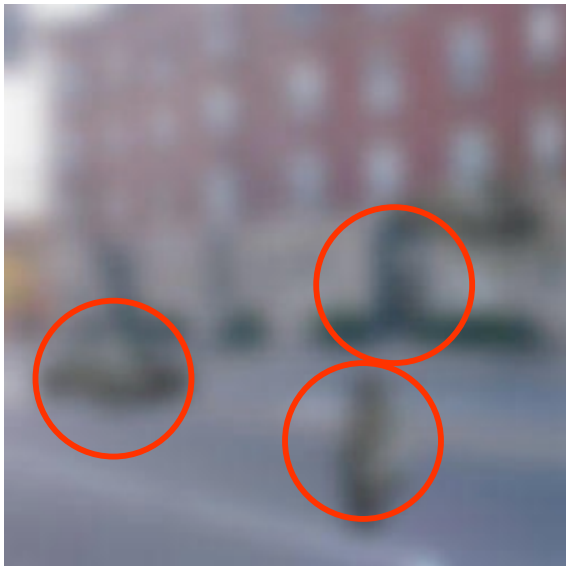
Challenges: object intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

Challenges: local ambiguity







slide credit: Fei-Fei, Fergus & Torralba

Course Program (tentative)

First part: Image formation and Early vision

- Image formation
  - Geometric Camera Models
  - Color spaces
- Image Processing
  - Punctual and spatial processing
  - Feature Extraction
- Reconstruction
  - Camera calibration
  - Stereo Vision
  - Structure from Motion and RGB-d Cameras
  - Optical flow and Tracking

Second part: Machine learning for CV

- Linear Neural Network
- Multi Layer Perceptron
-
- Convolutional Neural Networks
- Recurrent Neural Networks
- Transformers
- Variational Auto-Encoders
- Generative Adversarial Networks
- Graph Neural Networks
- Self-supervised Learning
- Vision Language Models