Imperial College London

Review Lecture

Dr Wenjia Bai

Department of Computing & Brain Sciences

We have learnt

Image classification,
object detection,
image segmentation,
action recognition

Image understanding

Reflection, optics,
visual system,
colour space,
image representation

Image formation

Image filtering,
edge detection,
interest point detection,
feature descriptor

Image processing

Optic flow,
object tracking,
image alignment

Motion

Mathematical foundation: calculus, linear algebra, statistics, geometry

You may want to learn more

Supervised learning,
unsupervised learning,
reinforcement learning,
geometric deep learning
......

Machine learning

Camera model,
Multiple view geometry,
3D reconstruction,
3D shape analysis
......

3D vision

Computer vision +
natural language
processing, robotics,
medical imaging
......

Applications

Mathematical foundation: calculus, linear algebra, statistics, geometry

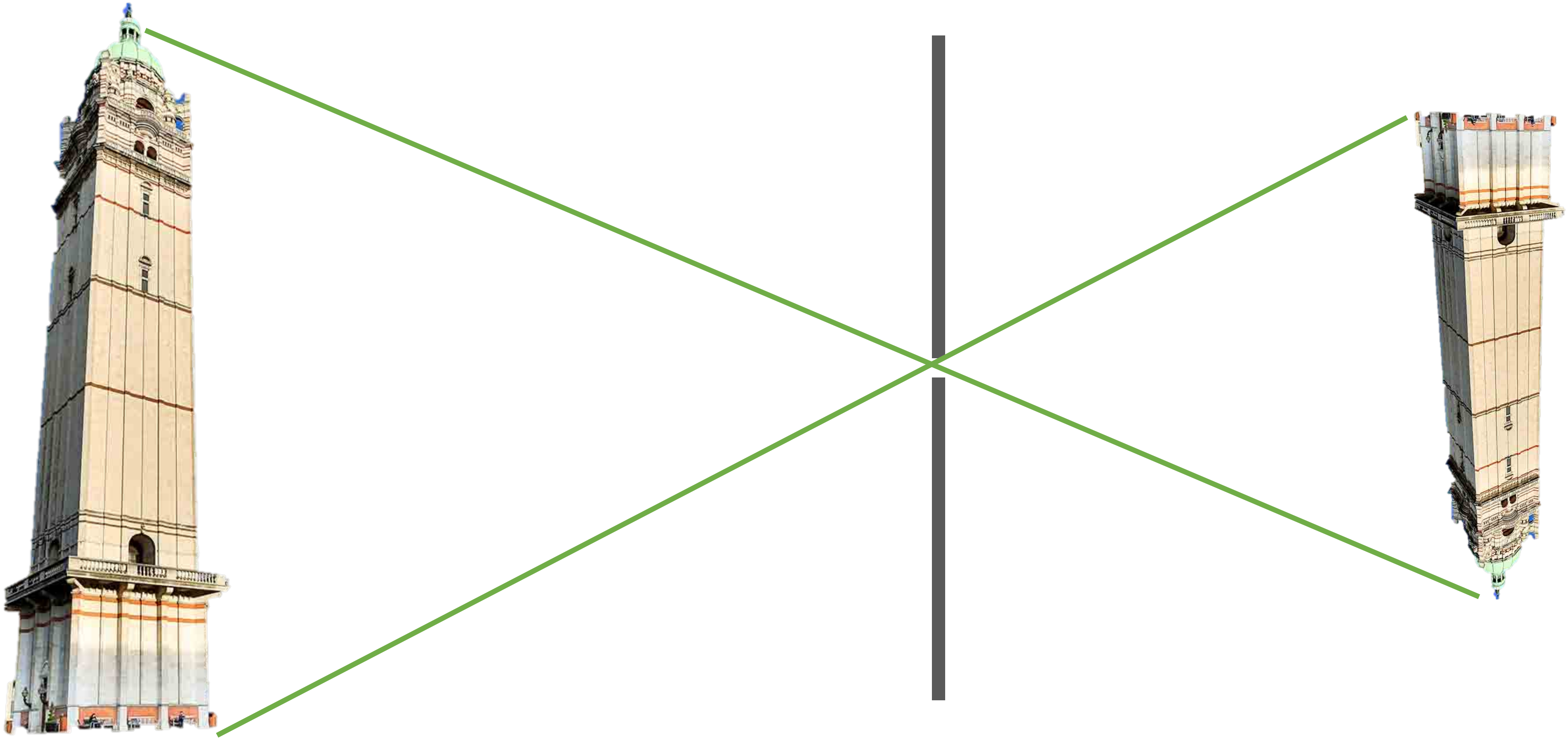
Object tracking in 3D world

https://www.youtube.com/watch?v=zRlzsi0OjDs

2D and 3D

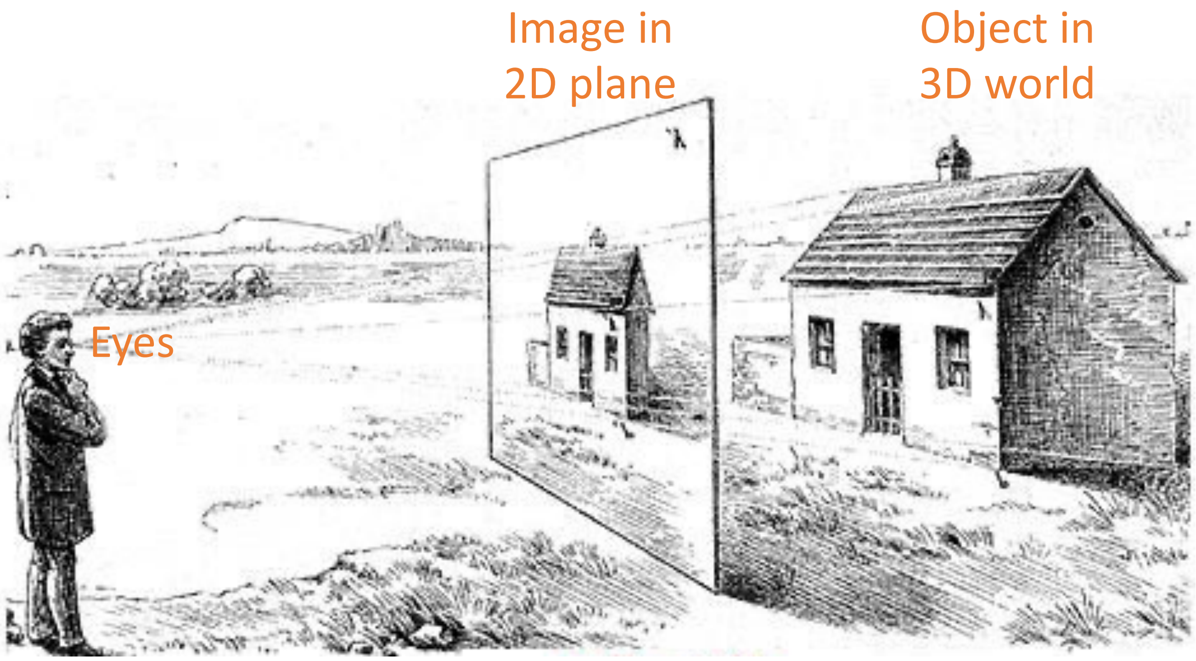
- The motion that we estimate from videos is in 2D.
- For 3D applications, if we need to estimate the 3D motion of an object, we need to understand
  - how camera works
  - how camera maps 3D coordinates to 2D coordinates.

Pinhole camera

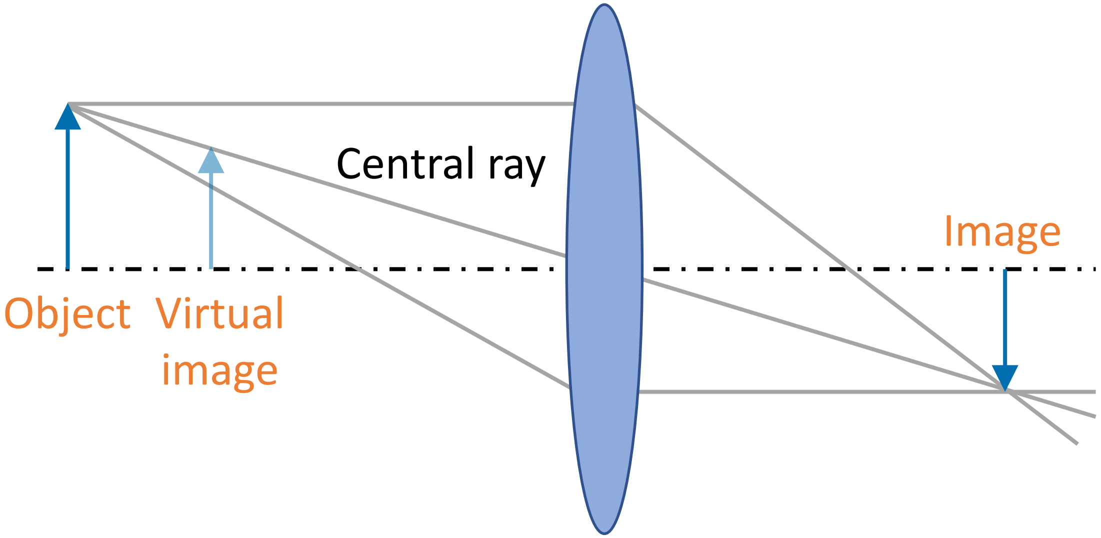


Perspective projection and lens camera

- We can find similar triangles when an object in 3D is mapped to an image in 2D.

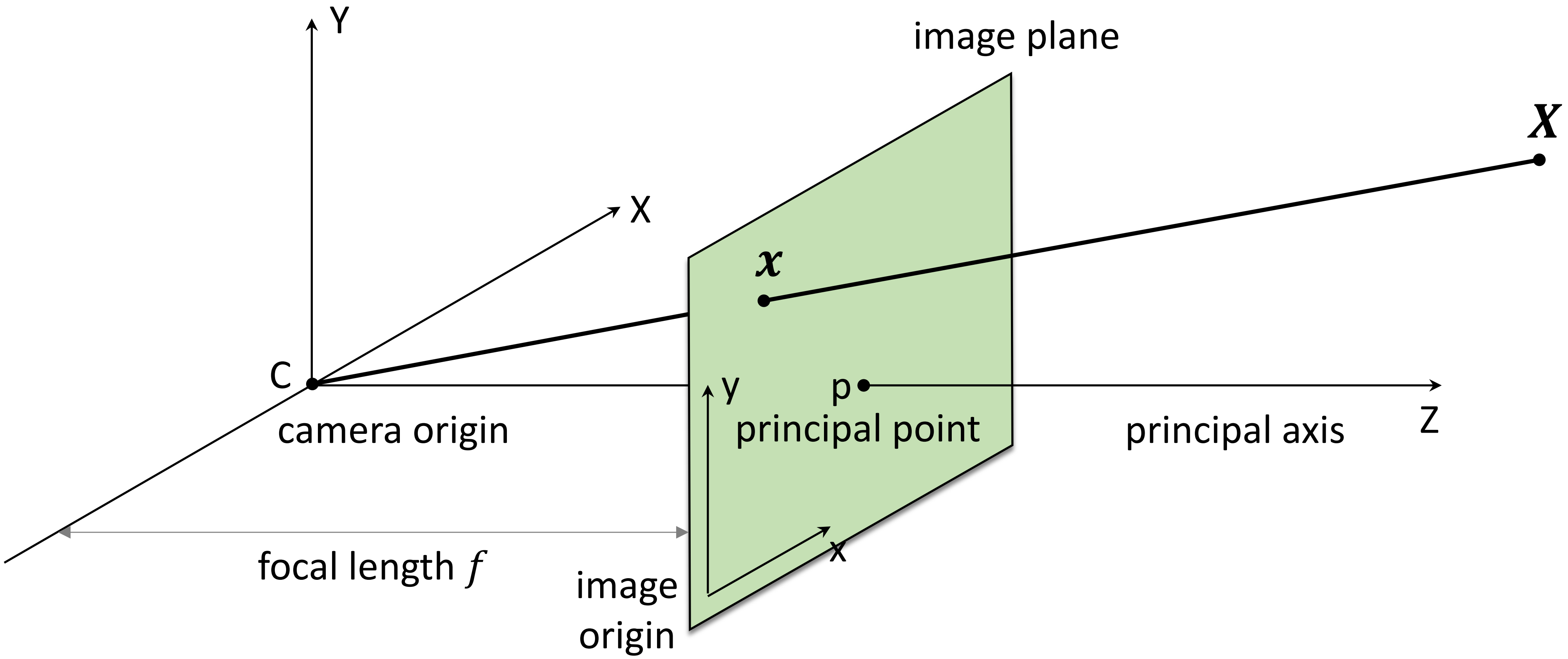


Perspective in painting



Lens camera

Camera model



The camera model describes the relationship between the 3D coordinate ***X*** and its 2D coordinate ***x*** on the imaging plane.

Camera model

- The camera model can be described by

$$\boldsymbol{x} = P\boldsymbol{X}$$

or

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- As long as we can determine the matrix P, the correspondence between 2D coordinate system and 3D coordinate system can be established.

2D and 3D

- With the camera matrix, we can relate the information from 3D world to 2D images.
  - Smartphone apps
  - Object tracking for drones

Apple ARKit

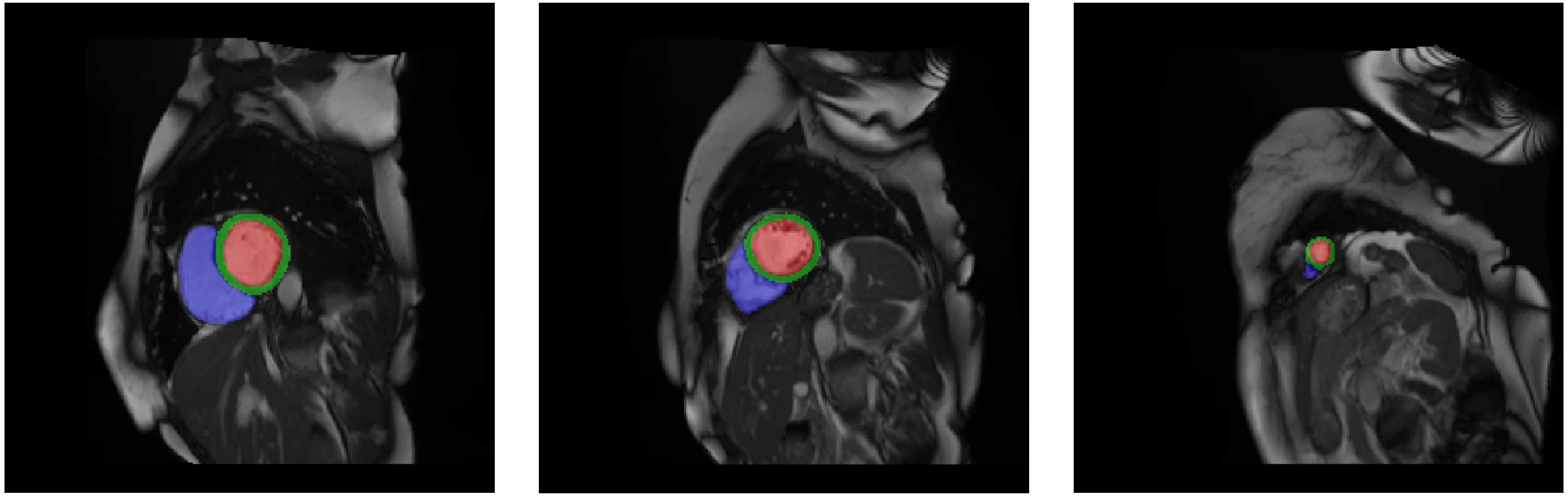
Pokemon Go

Applications

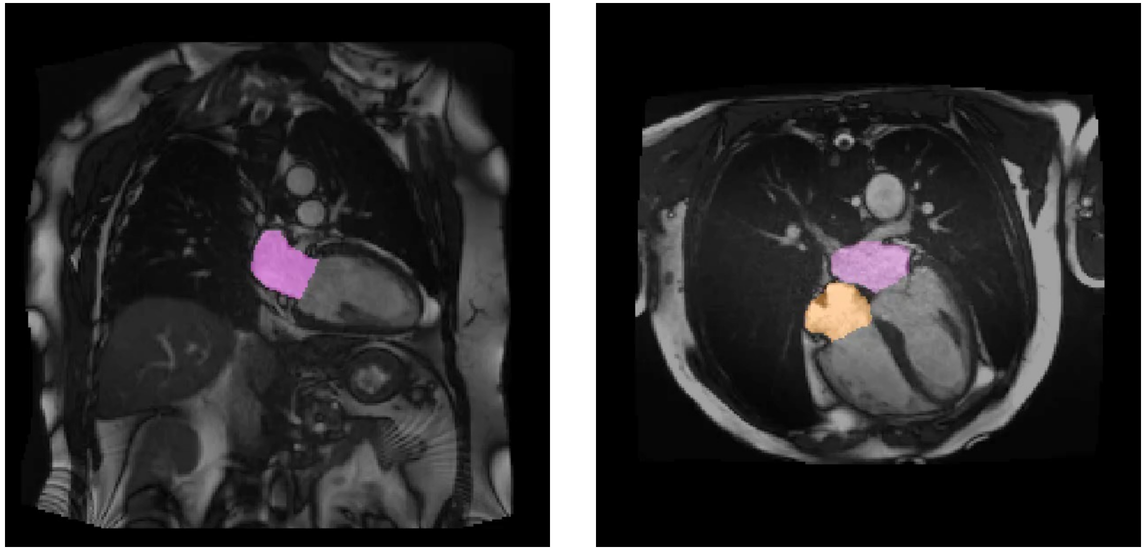
- Medical image computing
  - Computer vision + clinical domain knowledge + medical physics + …
- Autonomous driving
  - Computer vision + control + hardware + …
- Robotics
  - Computer vision + control + mechanics + materials + …

Cardiac image segmentation

Ventricles



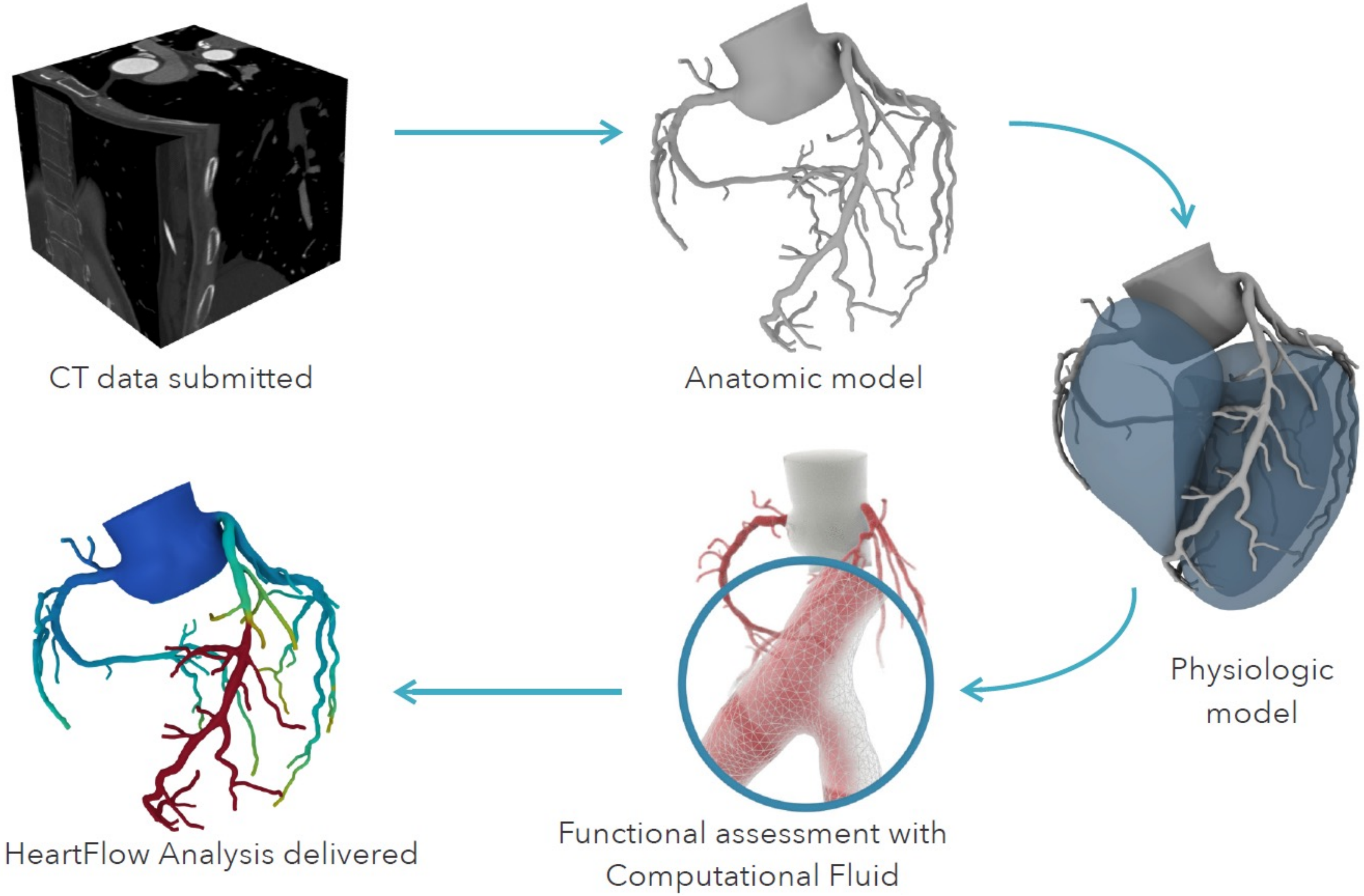
Atria



LV cavity RV cavity LA cavity RA cavity
LV myocardium

Bai et al. JCMR, 2018.

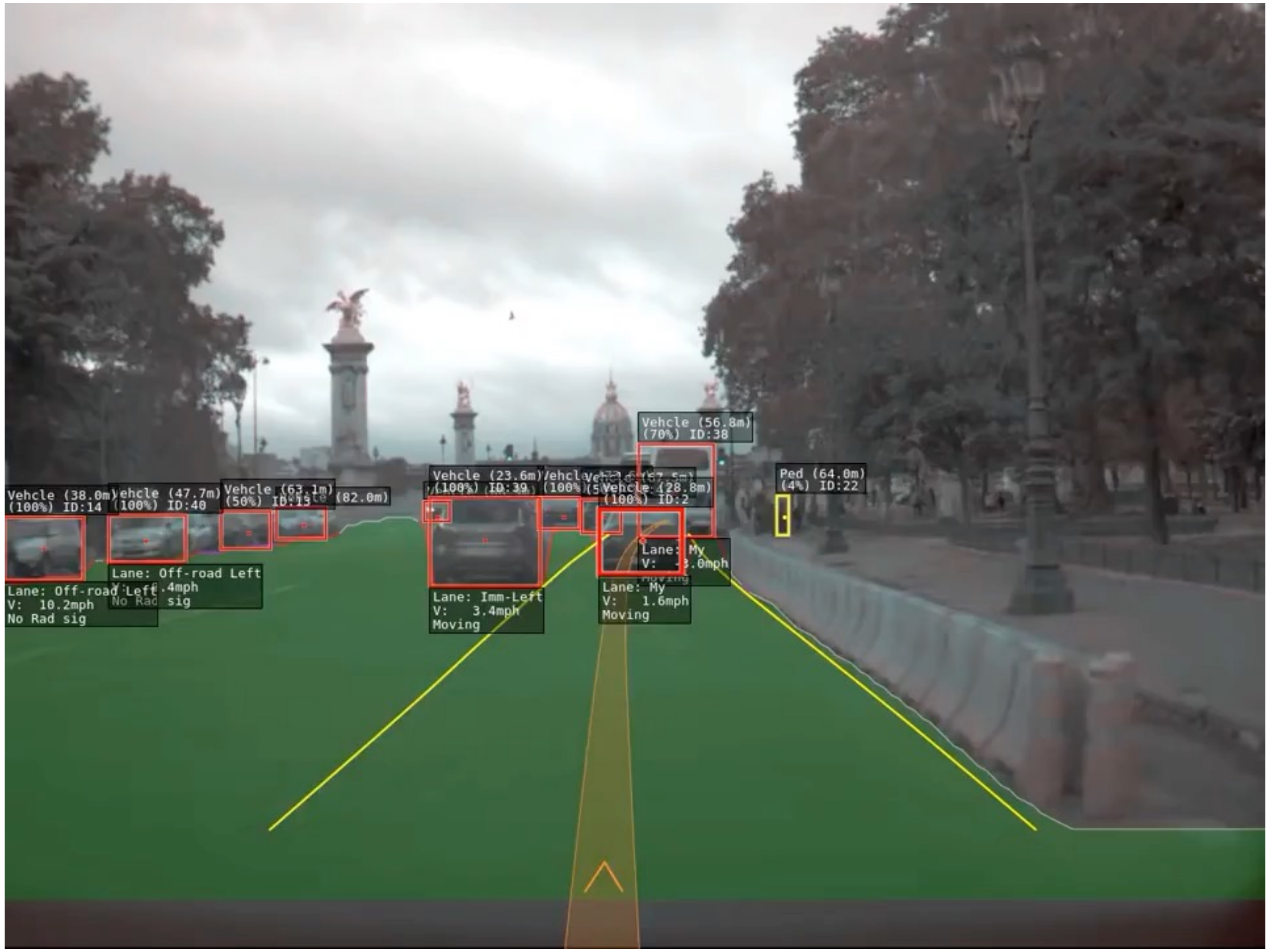
Coronary artery segmentation



The segmented arteries provide a physiologic model for disease treatment planning. ©HeartFlow

https://on-demand-gtc.gputechconf.com/gtcnew/sessionview.php?sessionName=s7520-deeplumen

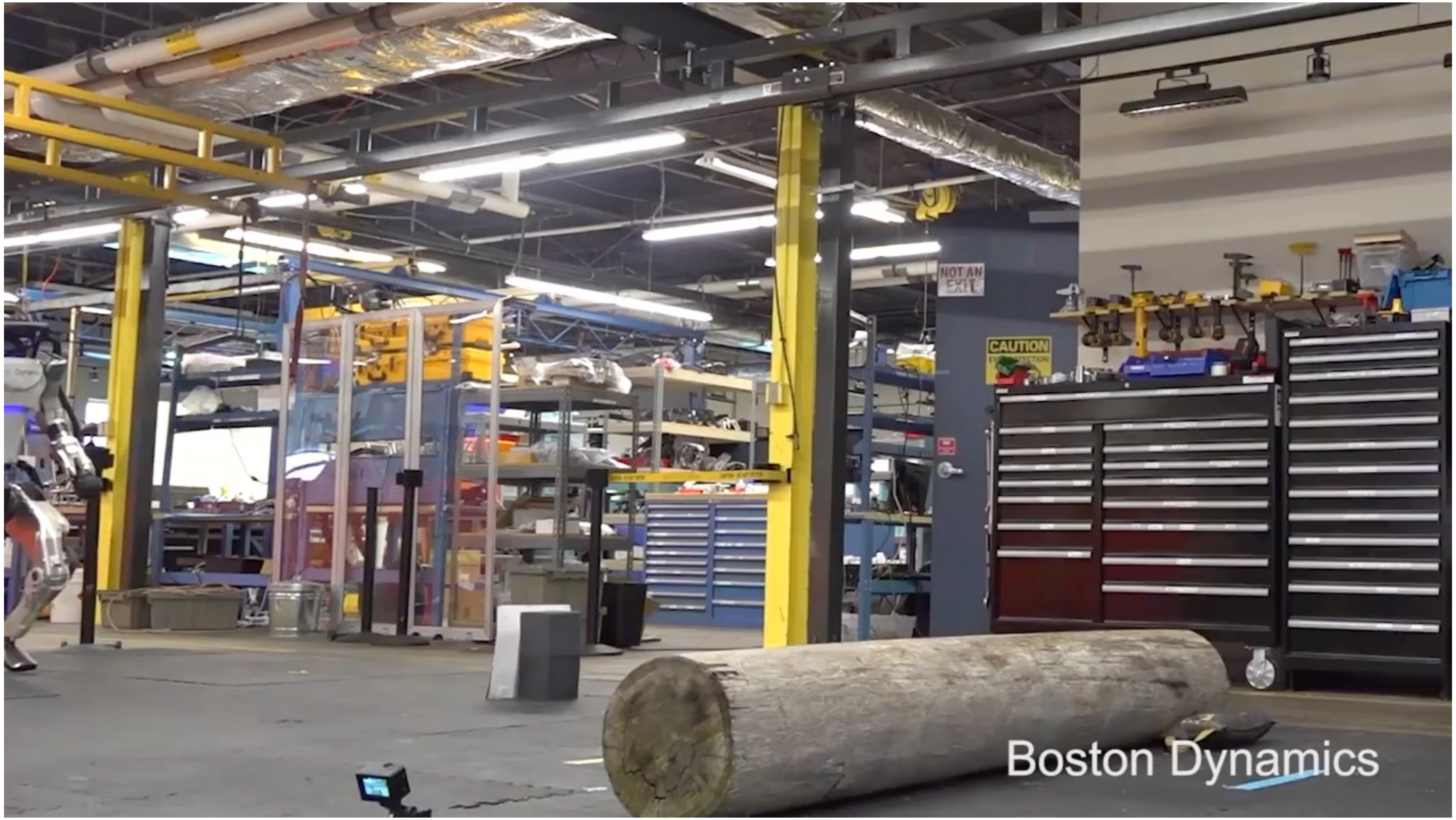
Autonomous driving



Paris streets in the eyes of Tesla Autopilot

https://www.youtube.com/watch?v=_1MHGUC_BzQ

Robotics



Atlas uses computer vision to locate itself with respect to visible markers on the approach. @BostonDynamics

https://www.youtube.com/watch?v=LikxFZZO2sk

Limitations

- Many modern computer vision algorithms are based on deep neural networks, which are very powerful in learning features.
- However, these approaches might have certain limitations:
  - Data hungry.
  - May be over-parameterised, not easy for mobile deployment.
  - May not generalise well, especially on data from an unseen domain.
  - May not be well integrated with prior knowledge.
  - Lack of interpretability.
  - ...
- Therefore, this is still a research active area, waiting for you to explore!

Exam preparation

- Timed open book exam
  - You have access to the course materials, textbooks and the internet.
  - The format will be consistent with other exams, which is to upload images/pdfs of your working.
  - You should have received emails from the department regarding the timed remote assessments (TRAs).
  - Timetable: https://exams.doc.ic.ac.uk/prog/generaltimetable.cgi

Exam preparation

- Most part of the exam paper will be similar to previous exam papers (https://exams.doc.ic.ac.uk) and tutorial questions.
- Types of questions
  - Do some math (e.g. calculus, linear algebra, convolution)
  - Explain concepts in computer vision
  - Solve practical problems with computer vision

What do I need to know?

- How to perform image filtering?
- What is convolution?

x

y

1	0	-1
1	0	-1
1	0	-1

Prewitt filter along x-axis

1	1	1
0	0	0
-1	-1	-1

Prewitt filter along y-axis

x

y

1	0	-1
2	0	-2
1	0	-1

Sobel filter along x-axis

1	2	1
0	0	0
-1	-2	-1

Sobel filter along y-axis

What do I need to know?

- Why do we need to perform image filtering?
- What are the commonly used filters?
- How to apply padding?

What do I need to know?

- How to describe features for points and for images?
- How to perform image matching using the interest points?
- What is a typical pipeline for image classification?
- Concepts in image classification, object detection and image segmentation

What do I need to know?

- Multi-layer perceptron (MLP)
- Convolutional neural network (CNN)
- Receptive fields
- Loss functions
- Evaluation

What do I need to know?

- Optic flow constraint equation
- Motion estimation
- Multi-scale framework
- Object tracking

What do I not need to know?

- You do not need to know
  - Derivation of the backpropagation algorithm
  - Functional optimisation in Horn-Schunck method
  - Slides that says “statement without proof”

Understanding

- Think about applications of computer vision in daily life.
- What algorithms are perhaps used for these applications?

Tutorial question 1

- Given image f and filter h, can you perform the convolution $f * h$?
- You can use the flip- and-shift method.

$h[n] =$

1	0	-1

$h[-n] =$

-1	0	1

step 1: flip the kernel

$*$ $*$ $*$

step 2: shift the kernel, multiply at each position

$f[n] =$

10	10	10	50	20	0	0

	0	40	10	-50	-20	

step 3: sum over the support of the kernel

Tutorial question 2

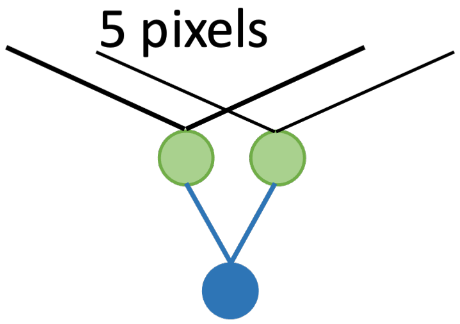
- Can you calculate the receptive field of each neuron for the following convolutional neural networks?

layer	0	1	2	3	4	5	6	7
type	input	conv	pool	conv	pool	conv	conv	loss
filter shape	-	5x5x1	2x2	5x5x20	2x2	4x4x50	1x1x500	-
#filters	-	20	-	50	-	500	10	-
stride	-	1	2	1	2	1	1	-
pad	-	0	0	0	0	0	0	-
data shape	28x28 x1							
receptive field	1x1	5x5						

Receptive field

- To calculate the receptive field of a neuron at Layer $i+1$, we account for the receptive field at Layer i and the kernel size at Layer $i+1$.

- For example, for a neuron at Layer 2,



Input layer

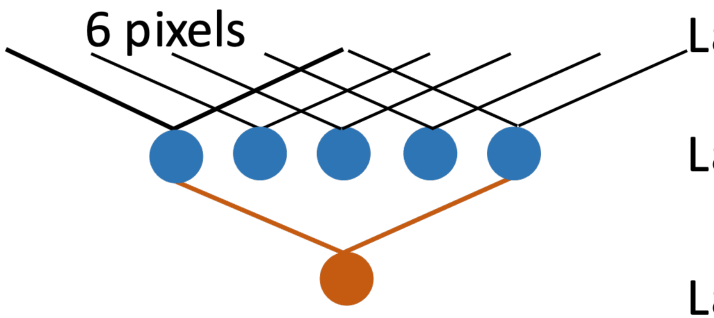
Layer 1: Each neuron sees 5 pixels.

Layer 2: Each neuron sees 2 neurons at Layer 1, which are 1 pixel apart.
Therefore, it see 5 + 1 = 6 pixels in the input image.

Receptive field

- To calculate the receptive field of a neuron at Layer $i + 1$, we account for the receptive field at Layer i and the kernel size at Layer $i + 1$.

- Then, for a neuron at Layer 3,



Input layer

Layer 1

Layer 2: Each neuron sees 6 pixels.

Layer 3: Each neuron sees 5 neurons at Layer 2. The distance between the first neuron and the fifth neuron is 8 pixels (due to the stride of 2). Therefore, it sees 6 + 8 = 14 pixels in the input image.

Any questions?

I hope you enjoy the course.
Good luck!