# Computer Vision

Lecture 6: Computer vision

Yuriy Kochura
iuriy.kochura@gmail.com
@y_kochura

# Today

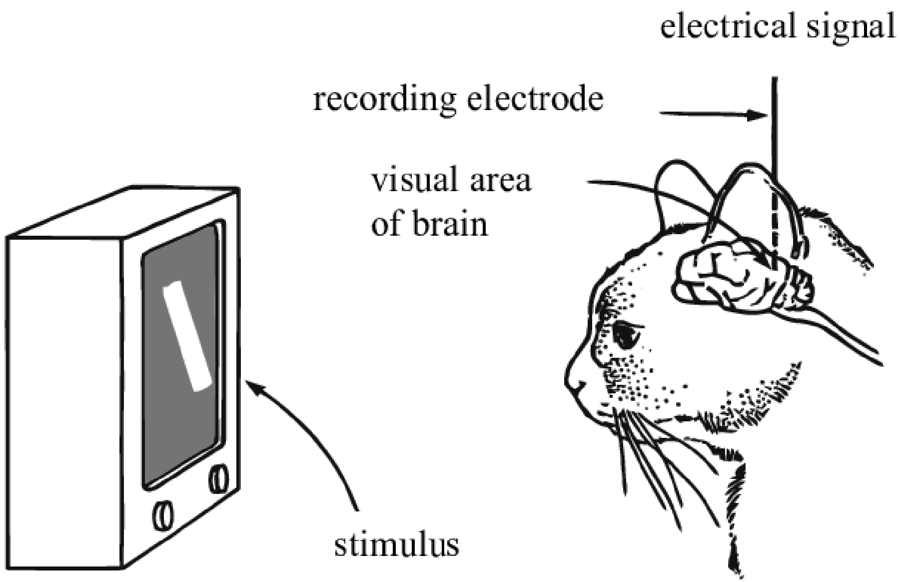
How to make neural networks see?

- Visual perception
- Human vision system
- Computer vision system
- Applications of computer vision

# Visual perception

# Visual perception

In 1959-1962, David Hubel and Torsten Wiesel discover the neural basis of information processing in the visual system. They are awarded the Nobel Prize of Medicine in 1981 for their discovery.

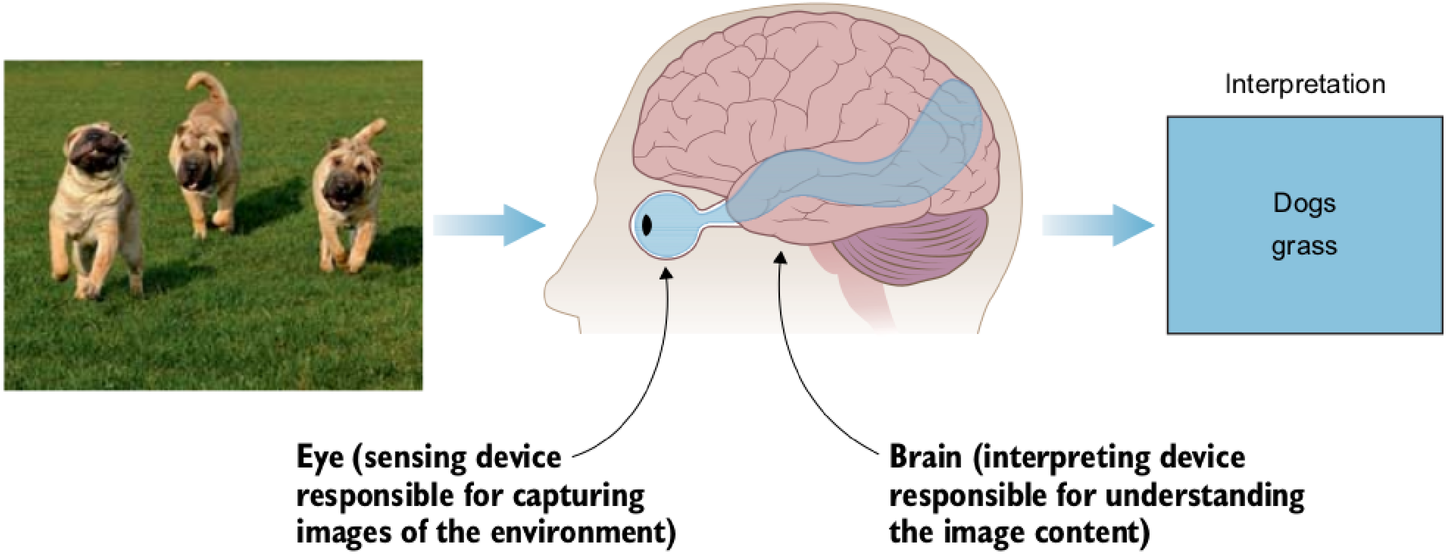



# Vision systems

In past decades, traditional image-processing techniques were considered CV systems, but that is not totally accurate. A machine processing an image is completely different from that machine understanding what's happening within the image, which is not a trivial task. Image processing is now just a piece of a bigger, more complex system that aims to interpret image content.

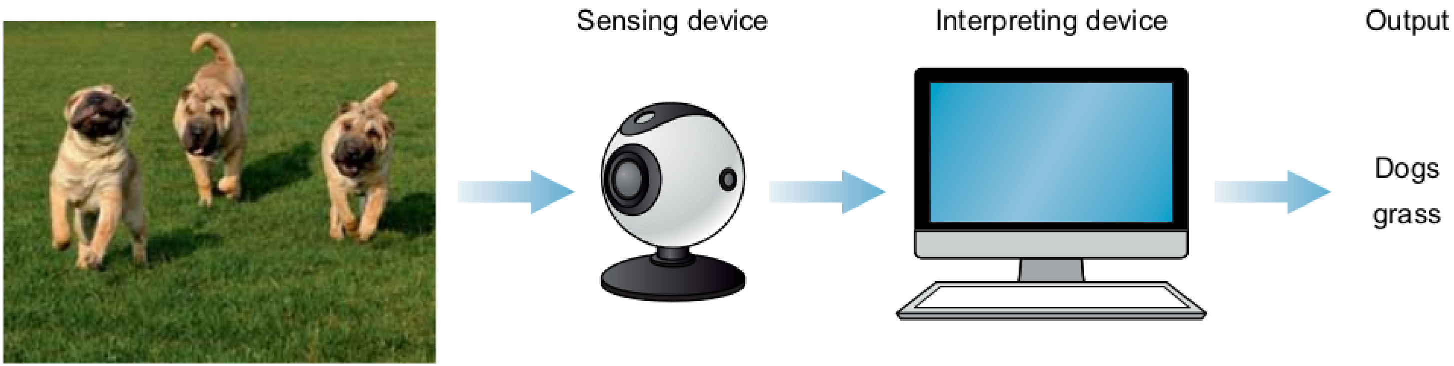
—
Credits: Mohamed Elgendy. Deep Learning for Vision Systems, 2020.

# Human vision system




Credits: Mohamed Elgendy. Deep Learning for Vision Systems, 2020.

# Computer vision system



# Sensing devices

Let's look at the autonomous vehicle (AV) example.

> The main goal of the AV vision system is to allow the car to understand the environment around it and move from point A to point B safely and in a timely manner. To fulfill this goal, vehicles are equipped with a combination of cameras and sensors that can detect 360 degrees of movement — pedestrians, cyclists, vehicles, roadwork, and other objects — from up to three football fields away.

Here are some of the sensing devices usually used in self-driving cars to perceive the surrounding area:

- Lidar, a radar-like technique, uses invisible pulses of light to create a high-resolution 3D map of the surrounding area.
- Cameras can see street signs and road markings but cannot measure distance.
- Radar can measure distance and velocity but cannot see in fine detail.

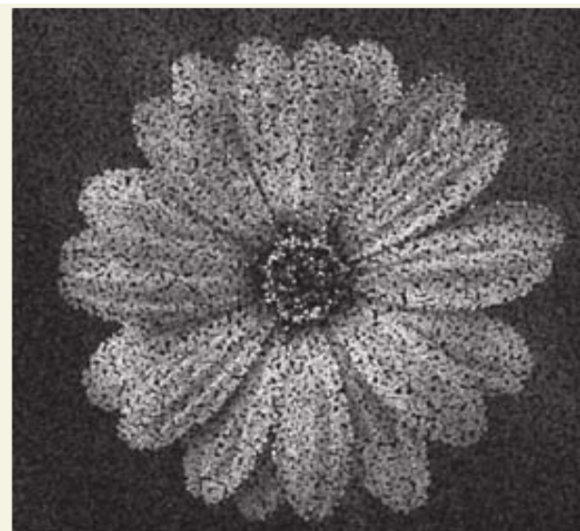
---

Credits: Mohamed Elgendy. Deep Learning for Vision Systems, 2020.

# Recognizing images

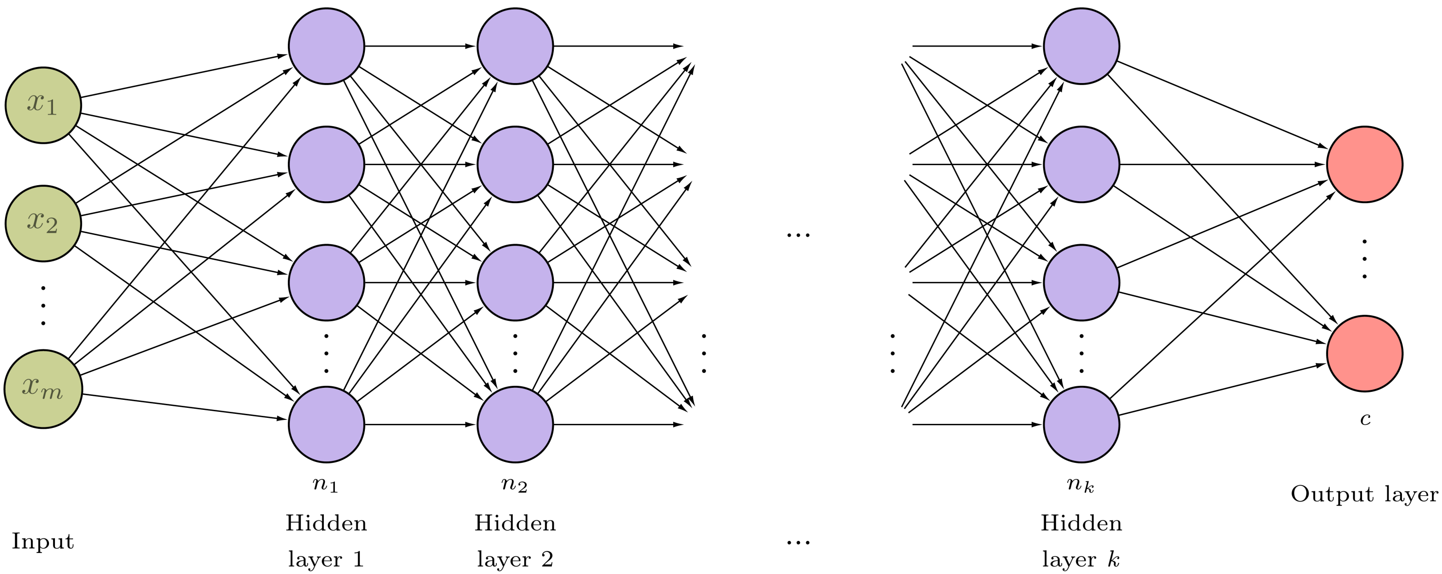
Animals, humans, and insects all have eyes as sensing devices. But not all eyes have the same structure, output image quality, and resolution. They are tailored to the specific needs of the creature. Bees, for instance, and many other insects, have compound eyes that consist of multiple lenses (as many as 30,000 lenses in a single compound eye). Compound eyes have low resolution, which makes them not so good at recognizing objects at a far distance. But they are very sensitive to motion, which is essential for survival while flying at high speed. Bees don't need high-resolution pictures. Their vision systems are built to allow them to pick up the smallest movements while flying fast.

Compound eyes



How bees see a flower

# Interpreting devices




```
import tensorflow as tf

model = tf.keras.Sequential([
    tf.keras.layers.Dense(n1),
    tf.keras.layers.Dense(n2),
    .
    .
    .

    tf.keras.layers.Dense(nk),
    tf.keras.layers.Dense(c)
])
```

## Can machine learning achieve better performance than the human brain?

Recent AI and DL advances have allowed machines to surpass human visual ability in many image classification and object detection applications, and capacity is rapidly expanding to many other applications.

—
Credits: Mohamed Elgendy. Deep Learning for Vision Systems, 2020.

# Applications of computer vision

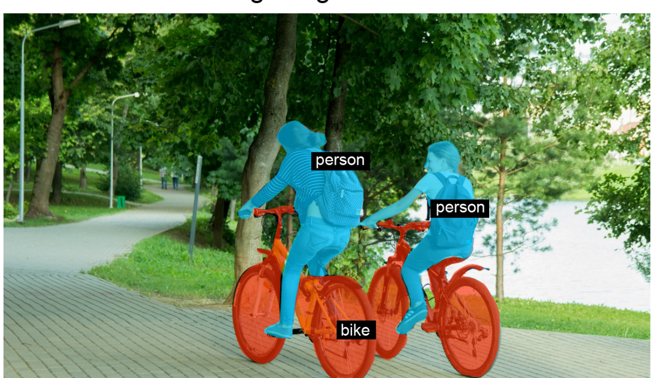
Single-label multi-class classification

- (•) Biking
- ( ) Running
- ( ) Swimming

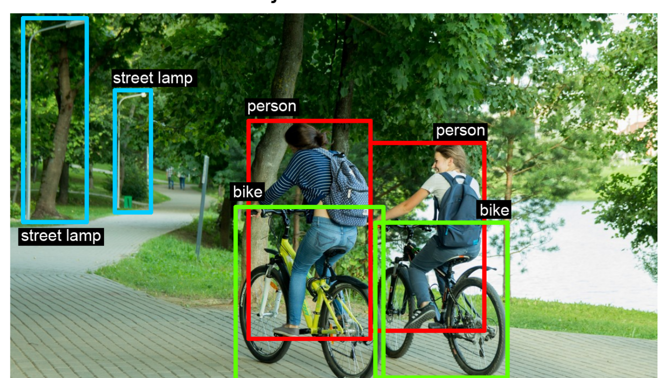
Multi-label classification

- [x] Bike
- [x] Person
- [ ] Boat
- [x] Tree
- [ ] Car
- [ ] House

Image segmentation




Object detection




Credits: Francois Chollet. Deep Learning with Python, 2021.

# Image classification

The goal is to assign one or more labels to an image. It may be either single-label classification (an image can only be in one category, excluding the others), or multi-label classification (tagging all categories that an image belongs to, as seen in the figure on the previous slide).

For example, when you search for a keyword on the Google Photos app, behind the scenes you're querying a very large multilabel classification model—one with over 20,000 different classes, trained on millions of images.

—
Credits: Francois Chollet. Deep Learning with Python, 2021.

# Object detection

The goal is to draw rectangles (called bounding boxes) around objects of interest in an image, and associate each rectangle with a class.

A self-driving car could use an object-detection model to monitor cars, pedestrians, and signs in view of its cameras, for instance.

# Image segmentation

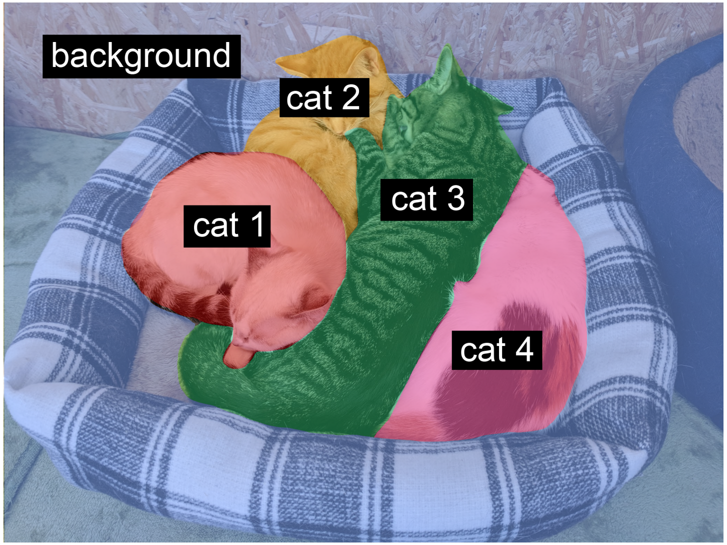
The the goal is to "segment" or "partition" an image into different areas, with each area usually representing a category (as as seen in the figure on the Applications of computer vision slide).

For instance, when Zoom or Google Meet diplays a custom background behind you in a video call, it's using an image segmentation model to tell your face apart from what's behind it, at pixel precision.

—

Credits: Francois Chollet. Deep Learning with Python, 2021.

# An image segmentation example


Semantic segmentation

Instance segmentation

—
Credits: Francois Chollet. Deep Learning with Python, 2021.

# Demo

Computer Vision Explorer

The end