# COMPUTER VISION

Course 1 · Foundations, Tasks and Modern Perspectives

Anca Ignat
Faculty of Computer Science
Alexandru Ioan Cuza University of Iași

# Course communication

- Microsoft Teams — course and laboratory
- Discord — laboratory and offline messages
- Course webpage: edu.info.uaic.ro/computer-vision/
- E-mail: use the addresses communicated in class

# Evaluation

- Course test: 3 points — on-site, week 15 or 16
- Laboratory activity: 6 points — six projects
- Article summary: 1 point
- Final requirement: overall grade ≥ 4.5
- Laboratory evaluation includes on-site and online components

# Recommended bibliography

- R. Szeliski — Computer Vision: Algorithms and Applications, 2nd ed. (2022)
- E. R. Davies — Computer Vision: Principles, Algorithms, Applications, Learning
- R. C. Gonzalez & R. E. Woods — Digital Image Processing, 4th ed.
- D. A. Forsyth & J. Ponce — Computer Vision: A Modern Approach
- Additional papers and online resources will be indicated during the semester

# Learning goals for Course 1

- Define Computer Vision and distinguish it from image processing and image analysis
- Recognize the major tasks of modern Computer Vision
- Understand how a digital image is represented
- Place classical, deep-learning and foundation-model approaches in context
- Identify common imaging modalities, applications, datasets and evaluation measures
- Recognize ethical issues associated with visual data and automated decisions

# What is Computer Vision?

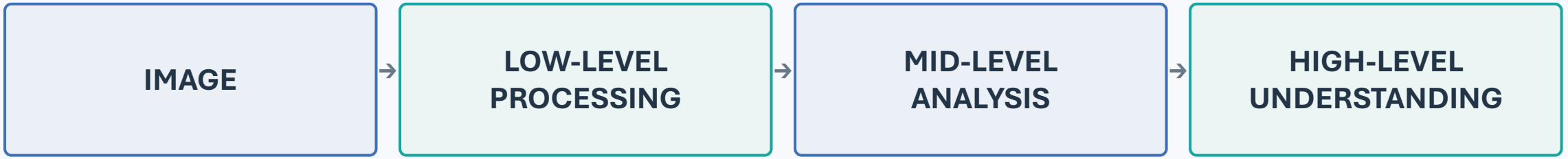
## Core idea

- Automatically extract information from images, video and other visual data
- Interpret objects, scenes, motion and spatial relationships
- Use visual information to support predictions, decisions or actions

## Human vision is an inspiration — not the definition

- Machines may solve visual tasks differently from humans
- Some sensors capture information humans cannot directly see
- Performance depends on data, task definition and evaluation

# Image Processing, Image Analysis and Computer Vision




**The boundaries are useful pedagogically, but modern systems often learn several stages jointly.**

# Three generations of Computer Vision

## Classical CV

Handcrafted features
+ explicit algorithms

**SIFT · SURF · HOG**

## Deep Learning CV

Learned features
+ task-specific models

**CNN · ResNet · YOLO**

## Foundation / multimodal CV

Large-scale pretraining
+ adaptation / prompting

**ViT · vision-language models**

These approaches coexist: modern systems often combine learned models with classical geometry, filtering or domain knowledge.

# Key tasks in Computer Vision

From image-level decisions to pixel-level and multimodal understanding

### Image classification

What is in the image?

### Object detection

What objects are present, and where?

### Semantic segmentation

Which class does each pixel belong to?

### Instance segmentation

Which pixels belong to each object instance?

### Tracking / action recognition

How do objects and actions evolve over time?

### Pose / landmarks

Where are key parts or joints?

### Depth / 3D vision

What is the geometry of the scene?

### Image retrieval

Which images are visually similar?

### Vision-language tasks

Describe, query or reason about visual content.

# Classification vs. detection

## Classification

- One or more labels for an image
- Output: class probabilities / labels
- Example: normal vs. pneumonia chest X-ray

## Detection

- Class + location for each object
- Output: bounding boxes and scores
- Example: pedestrians and cars in a street scene

# Segmentation: semantic vs. instance

## Semantic segmentation

- Every pixel receives a semantic class
- Objects of the same class share one label
- Useful for road scenes, medical images, remote sensing

## Instance segmentation

- Separates individual object instances
- Combines localization and pixel-level masks
- Useful when distinct objects of the same class matter

# Beyond recognition

## Tracking & video

- Associate objects across frames
- Estimate motion and trajectories
- Recognize activities and temporal events

## 3D & spatial understanding

- Depth estimation
- Stereo / multi-view geometry
- 3D reconstruction and localization

# Vision meets language

## Typical tasks

- Image captioning
- Visual question answering
- Text-image retrieval
- Visual grounding

## Why it matters

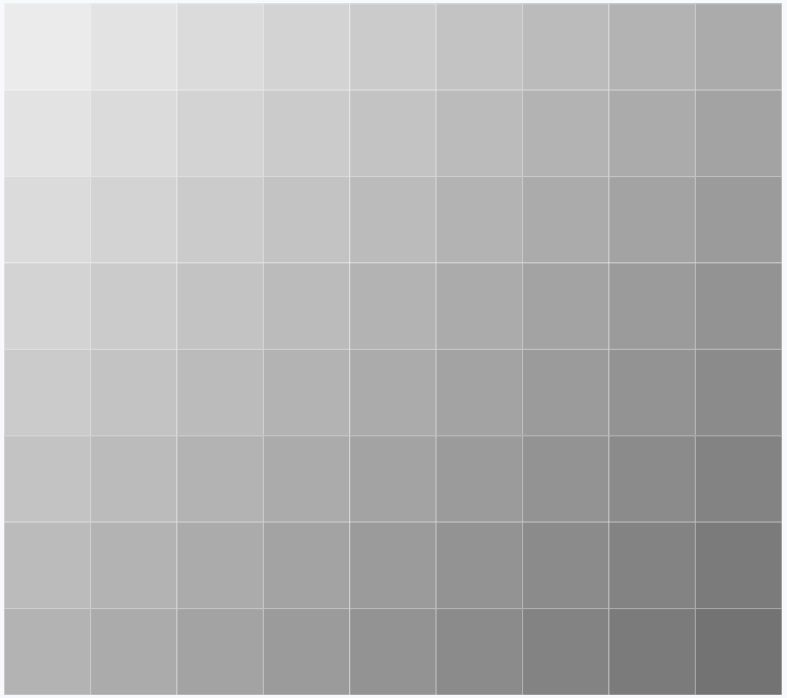
- Connects visual representations with natural language
- Supports zero-shot and prompt-based use cases
- Extends CV beyond fixed closed-set labels

# What is a digital image?

A digital image can be modeled as a discrete function

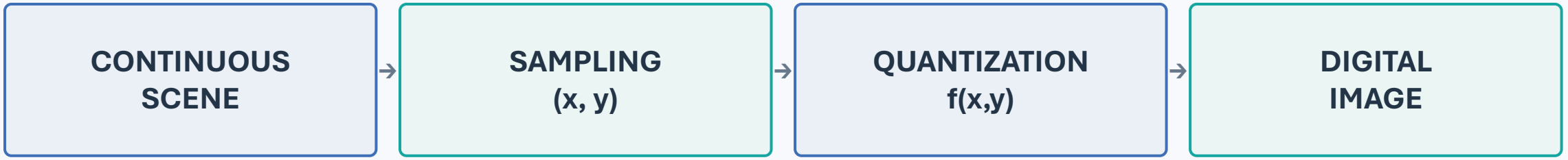
$$f(x, y)$$

x, y → spatial coordinates
f(x,y) → intensity or color value



**pixels = discrete samples**

# Sampling and quantization




**Sampling determines spatial resolution.**

**Quantization determines the number of representable intensity/color levels.**

Both processes discard information; the chosen resolution must be appropriate for the task.

# Image representation in practice

## Grayscale image

- One intensity value per pixel
- Common representation: H × W array
- Often 8-bit: values 0…255

## Color image

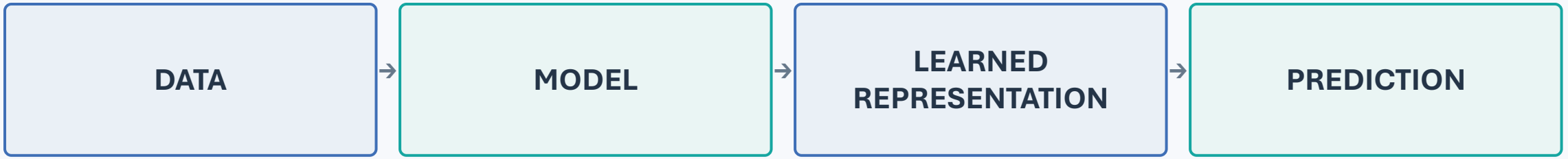
- Multiple channels per pixel
- RGB: H × W × 3
- Other spaces may separate luminance and chromatic information

# A classical Computer Vision pipeline

Example descriptors: edges, corners, texture, shape, SIFT/SURF/HOG

**Strength: interpretable stages and explicit control. Limitation: handcrafted choices may not generalize.**

# A learning-based Computer Vision pipeline




CNNs and Vision Transformers learn features from data rather than relying only on handcrafted descriptors.

**Training data, objective functions, architecture and evaluation protocol become central design choices.**

# Classical and learned methods are complementary

## Classical methods remain useful

- Filtering and enhancement
- Geometry and calibration
- Keypoints and local matching
- Small-data or constrained systems

## Learning-based methods excel when

- Large and representative datasets exist
- Features are difficult to engineer manually
- End-to-end optimization is beneficial
- Transfer learning is available

# A very short history

**1960s–70s** — Digital image processing, early vision, medical and space imaging

**1980s–90s** — Geometry, segmentation, handcrafted features

**2000s** — Local descriptors, bag-of-features, statistical learning

**2010s** — Deep CNNs transform recognition and detection

**2020s** — Vision Transformers, self-supervision, foundation and multimodal models

# Imaging beyond RGB

Computer Vision operates on many sensing modalities

**Visible / RGB**

consumer cameras · microscopy

**Infrared / thermal**

night vision · inspection

**X-ray / CT**

medicine · industry

**MRI / radio-frequency**

soft-tissue imaging

**Multispectral**

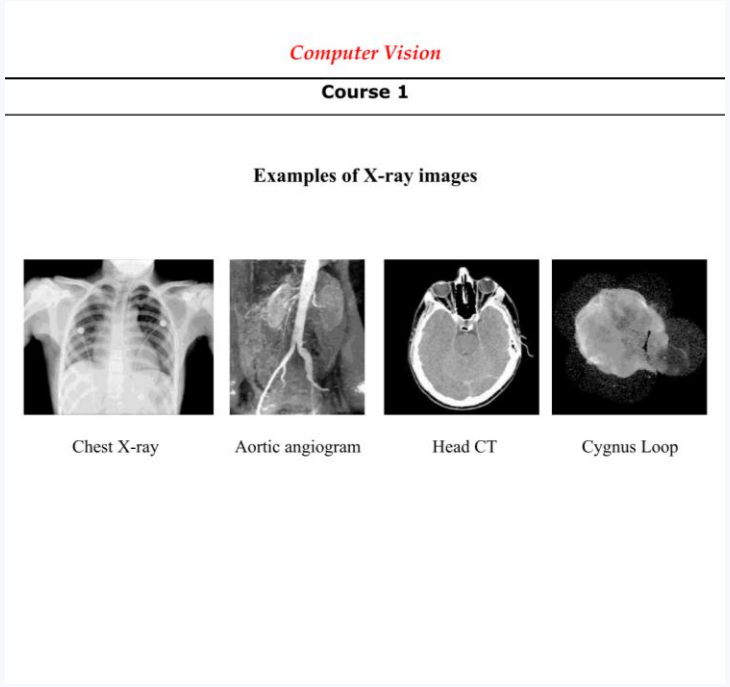
remote sensing · agriculture

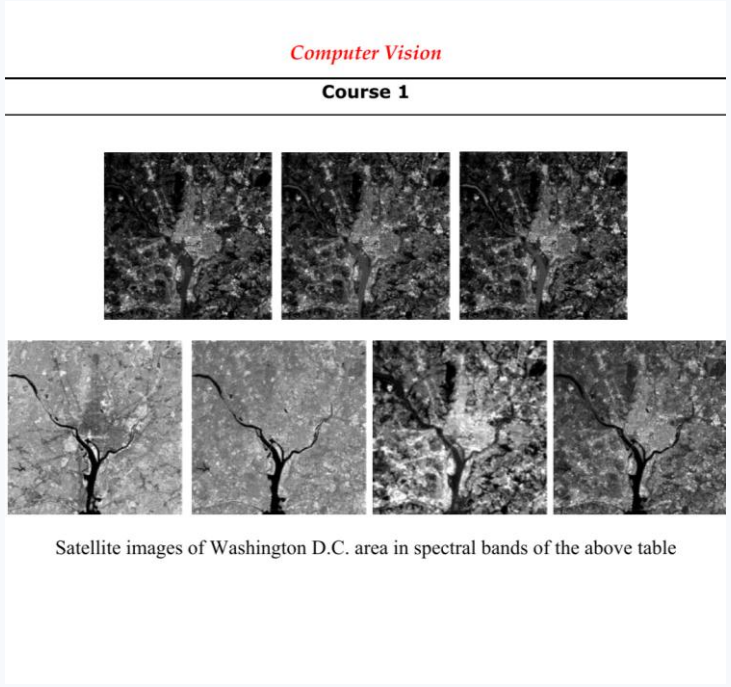
**Radar / LiDAR / depth**

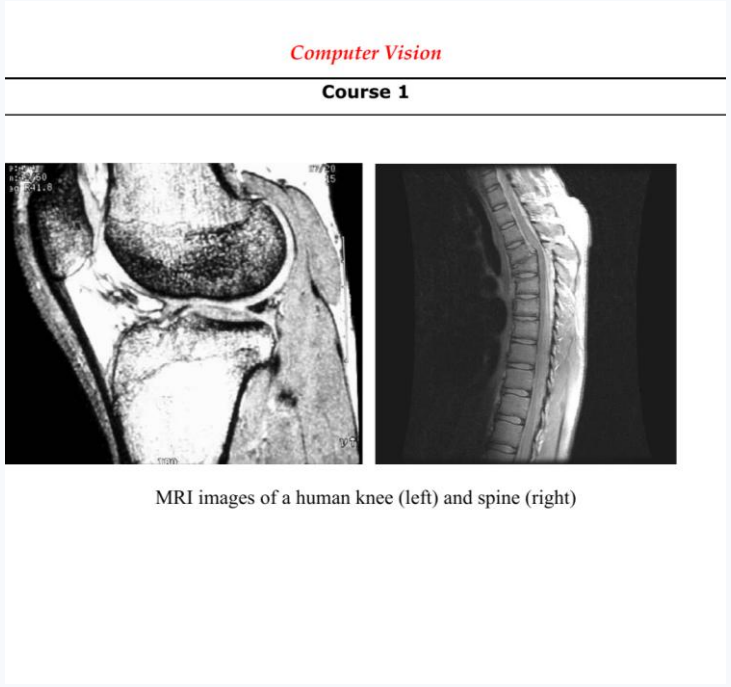
autonomy · mapping

**Different sensors reveal different physical properties — but many CV tasks remain the same.**

# Same scene, different information




**X-ray / CT**

**Multispectral remote sensing**

**MRI**

**Sensor choice is part of the Computer Vision problem definition.**

# Where is Computer Vision used?

## Healthcare

diagnosis · measurement · intervention

## Autonomous systems

perception · localization · navigation

## Industry

inspection · robotics · quality control

## Biometrics

identity verification · access control

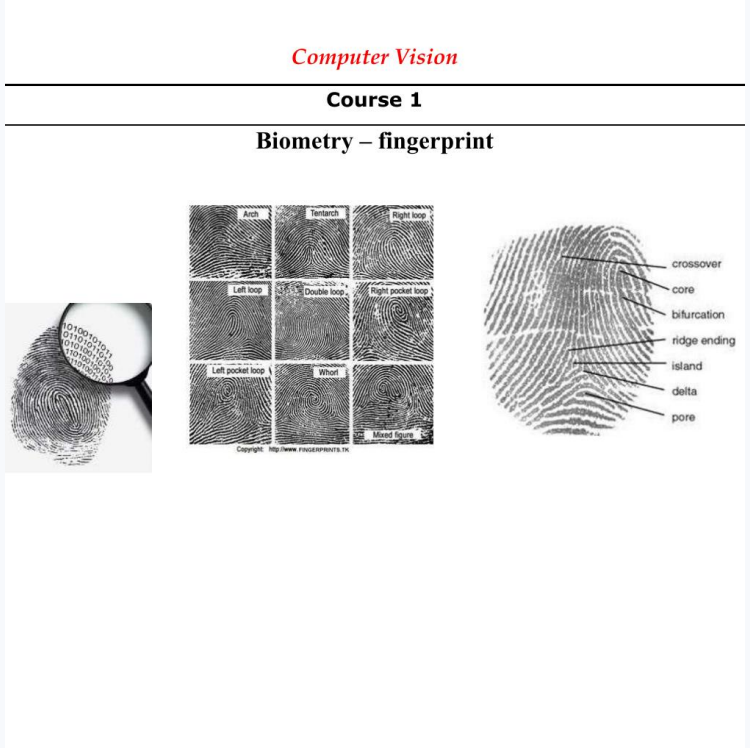
## Remote sensing

environment · agriculture · mapping

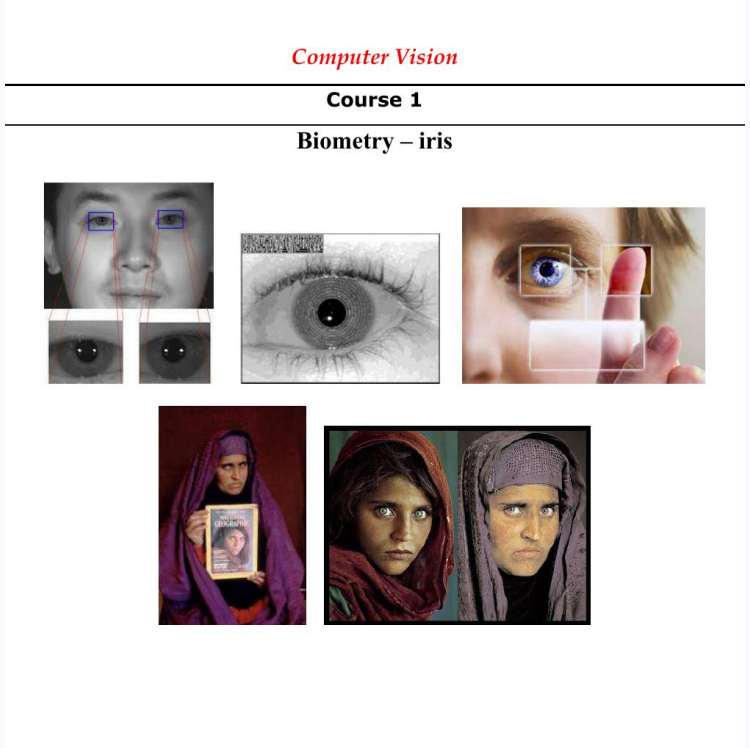
## Documents & media

OCR · retrieval · moderation · deepfake analysis

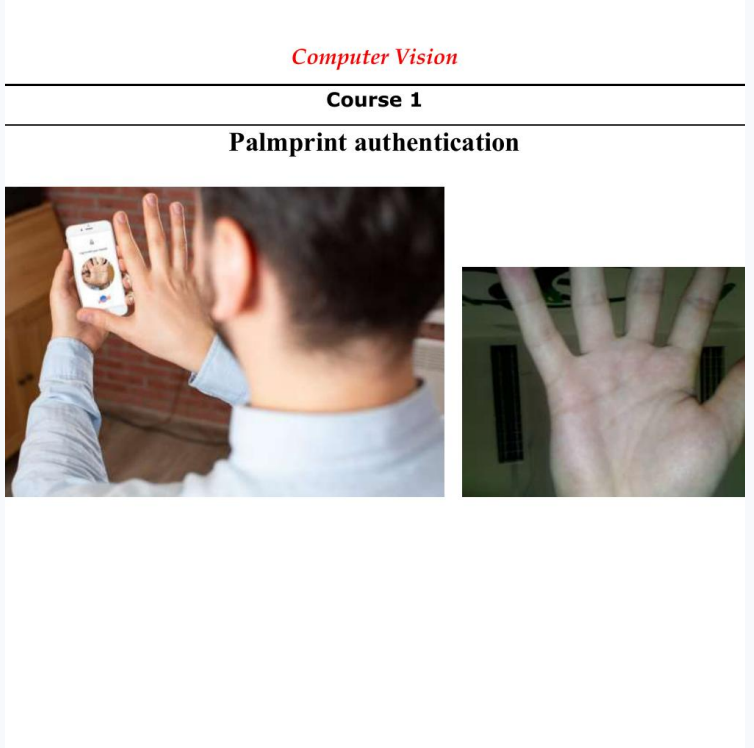
# Biometrics as a Computer Vision application




Fingerprint

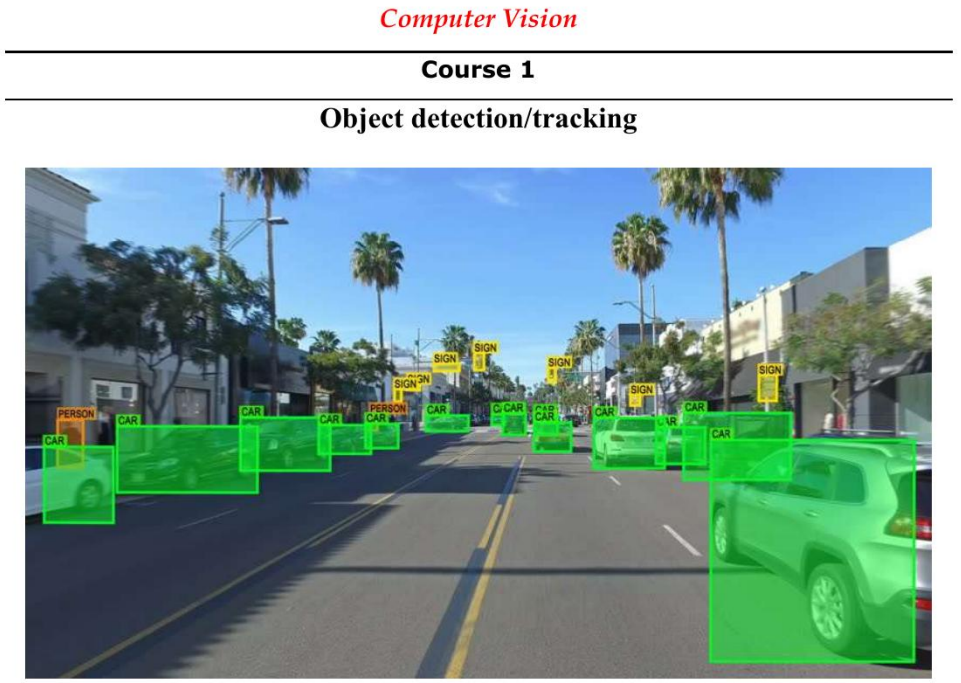



Iris




Palmprint

# Object detection and tracking




## Detection

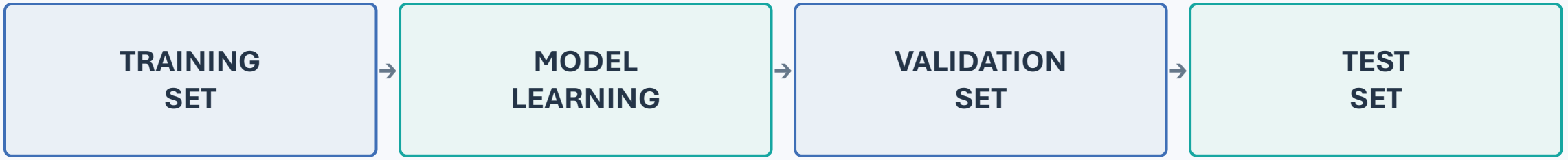
Find and classify objects in each frame.

## Tracking

Associate the same object across time and estimate its trajectory.

**Typical challenges: occlusion · scale change · illumination · crowded scenes**

# Computer Vision is data-driven




Training: fit model parameters

Validation: select settings / models

Test: final unbiased evaluation

**A model can only be evaluated meaningfully with a clearly defined dataset, split and protocol.**

# Examples of widely used datasets

## General-purpose

- ImageNet — image classification
- COCO — detection, segmentation, captions
- Cityscapes — urban scene understanding

## Domain-specific

- Medical imaging datasets
- Biometric datasets
- Remote-sensing datasets
- Task-specific research datasets

# How do we evaluate a CV system?

**Classification** Accuracy · Precision · Recall · F1

**Detection** IoU · Precision/Recall · mAP

**Segmentation** IoU · Dice coefficient

**Biometrics** FAR · FRR · EER

**Retrieval** Precision@k · Recall@k · mAP

**Efficiency** Latency · throughput · memory

**The metric must match the task — and should be interpreted together with the data distribution.**

# Human perception is not a pixel meter

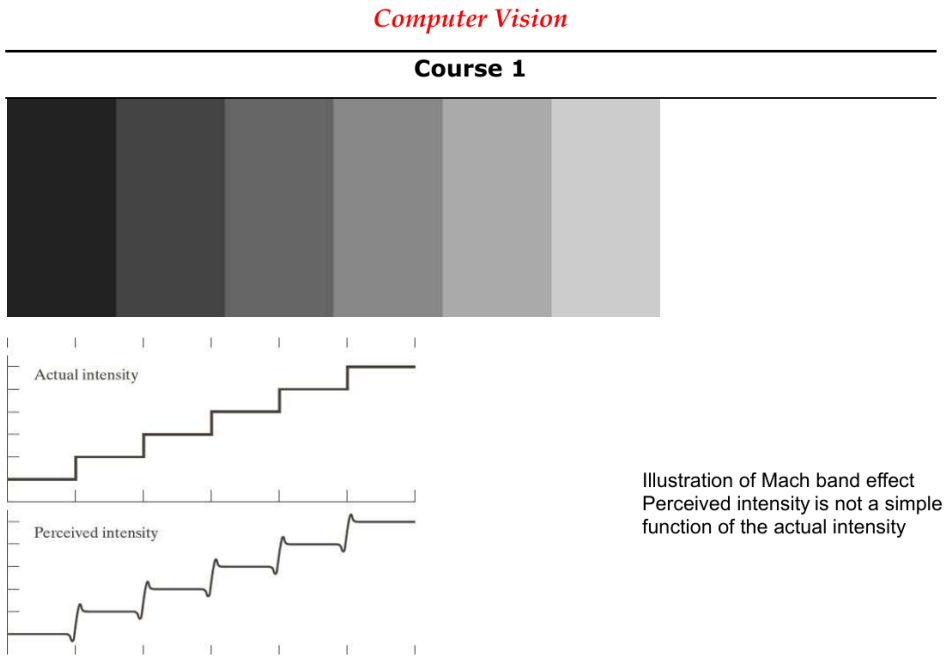


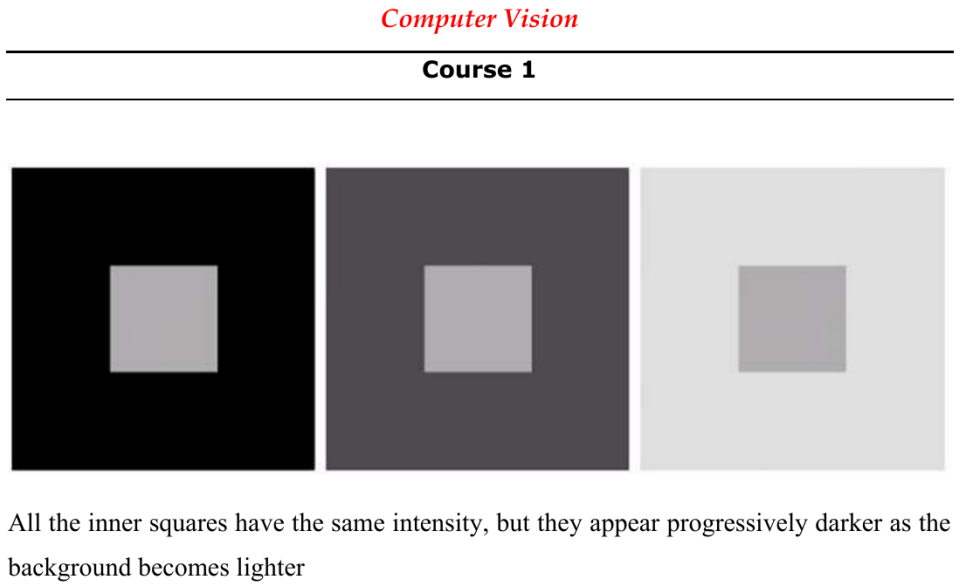

Illustration of Mach band effect
Perceived intensity is not a simple function of the actual intensity

## Mach band effect

Perceived intensity is influenced by neighboring intensities. The visual system emphasizes local contrast near boundaries.

**Lesson for CV: image measurements and human perception are related, but not identical.**

# Context changes what we perceive




All the inner squares have the same intensity, but they appear progressively darker as the background becomes lighter

**The inner squares have the same intensity.**

Yet they appear different because the surrounding background changes.

**Visual interpretation depends on context — a recurring theme in both human and machine vision.**

# Responsible Computer Vision: “Could I?” and “Should I?”

- Is the visual data collected with appropriate consent?
- Does the dataset represent the population where the system will be used?
- Can errors disproportionately affect particular groups?
- Could the system enable intrusive surveillance or misuse?
- Can results be audited and reproduced?
- Are image-processing and model-generation steps documented?

# Scientific image integrity

## Good practice

- Keep the original acquired data
- Document every processing step
- Report processing that affects interpretation
- Use reproducible workflows

## Risky practice

- Removing or adding content without disclosure
- Selective enhancement that changes interpretation
- Undocumented AI-generated or AI-edited content
- Presenting processed data as raw data

# Benchmark images also have a history

*Computer Vision*

**Course 1**

**Meet Lena!**

**The First Lady of the Internet**

https://www.g4media.ro/o-fotografie-playboy-din-1972-interzisa-de-o-comunitate-stiintifica-fotografia-cu-modelul-lena-forsen-a-devenit-standardul-jpeg-e-considerata-acum-o-imagine-sexista.html

**The “Lena” image became a long-standing image-processing benchmark.**

Its origin and continued use later prompted discussion about inclusiveness, consent and the culture of technical communities.

**Technical conventions are not socially neutral.**

# Deepfakes: Computer Vision as both generator and detector

## Generation

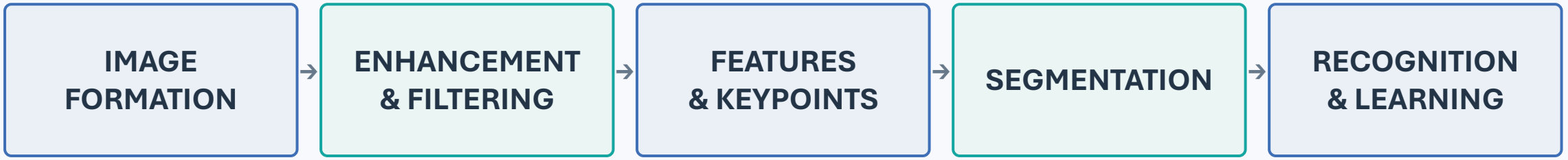
Models can synthesize or modify faces, voices and scenes with increasing realism.

## Detection

CV can analyze visual artifacts, temporal consistency and provenance signals.

**The same technological progress can create new applications and new risks.**

# Where this course is going




Classical foundations ↔ Machine learning / deep learning ↔ Applications

**Goal: understand not only how algorithms work, but also when and why to use them.**

# Key takeaways

- Computer Vision turns visual data into information, predictions and actions.
- Modern CV includes classical methods, deep learning and foundation/multimodal models.
- The task determines the output: label, box, mask, trajectory, depth, retrieval result or language.
- Image acquisition and representation matter: sensor, sampling and quantization affect what can be recovered.
- Data and evaluation protocols are as important as the model.
- Responsible use requires attention to privacy, bias, integrity, provenance and reproducibility.

# Discussion

## Choose one Computer Vision application.

What is the input?
What is the desired output?
What data would be required?
How would you evaluate success?
What could go wrong?

**These five questions will reappear throughout the course.**

# Computer Vision

From pixels to understanding