본 수업자료는 2025년도 과학기술 정보통신부 및 정보통신기획평가원의 ‘SW중심대학사업’ 지원을 받아 제작 되었습니다.

Computer Vision

Week1

2025-2

Mobile Systems Engineering

Dankook University

About Me

- **Education**
  - 2023, Ph. D., School of Integrated Technology, Yonsei University
    - Embedded Intelligent System (EIS) laboratory
    - Advisor: Dr. JeongGil Ko

- **Awards**
  - **Dissertation Fellowship,** Yonsei University, 2023
  - **Distinguished Paper Award**, ACM Ubicomp 2021
  - **Best Poster Award**, Korean Otological Society Conference 2021

- **Research Experience**
  - Commonwealth Scientific and Industrial Research Organization (CSIRO)
    - Research Intern, Brisbane, Australia, 2019
  - Advanced Multimedia Lab at SAMSUNG Mobile R&D Center
    - Senior Researcher, 2023 - 2024

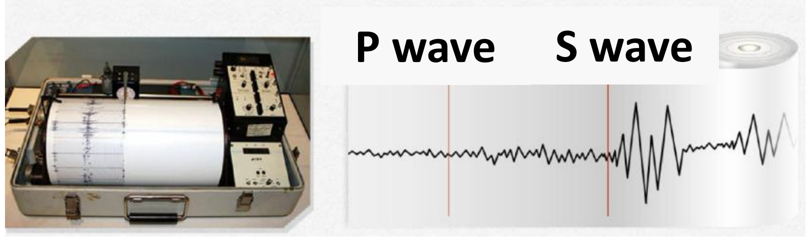
- **Research Community Services**
  - Korean Institute of Communication and Information Sciences (한국통신학회) 2025, Director
  - ACM The International Conference on Embedded Wireless Systems and Networks (EWSN) 2025, Technical Program Committee Member
  - IEEE/ACM International Conference on COMmunication Systems & NETworkS (COMSNETS) 2025, Technical Program Committee Member
  - International Conference on Information Networking (ICOIN) 2025, Technical Program Committee Member.
  - Technical Program Committee (TPC) Member,
    IEEE International Conference on Sensing, Communication, and Networking (SECON) 2024
  - Technical Program Committee (TPC) Member,
    IEEE Global Communications Conference (GLOBECOM) 2024
  - Web Co-Chair,
    ACM ACM International Conference on Mobile Systems, Applications, and Services (MobiSys) 2024
  - Technical Program Committee (TPC) Member,
    IEEE Global Communications Conference (GLOBECOM) 2023

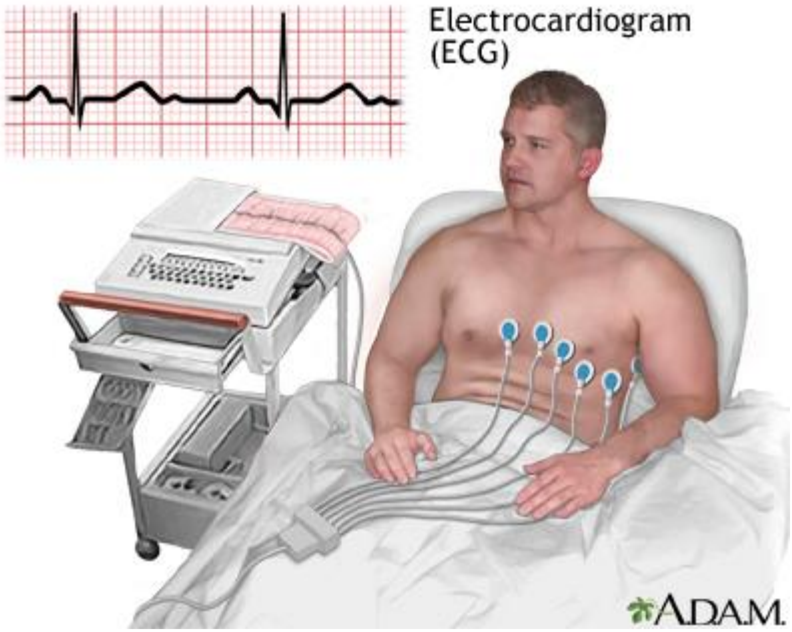
About me

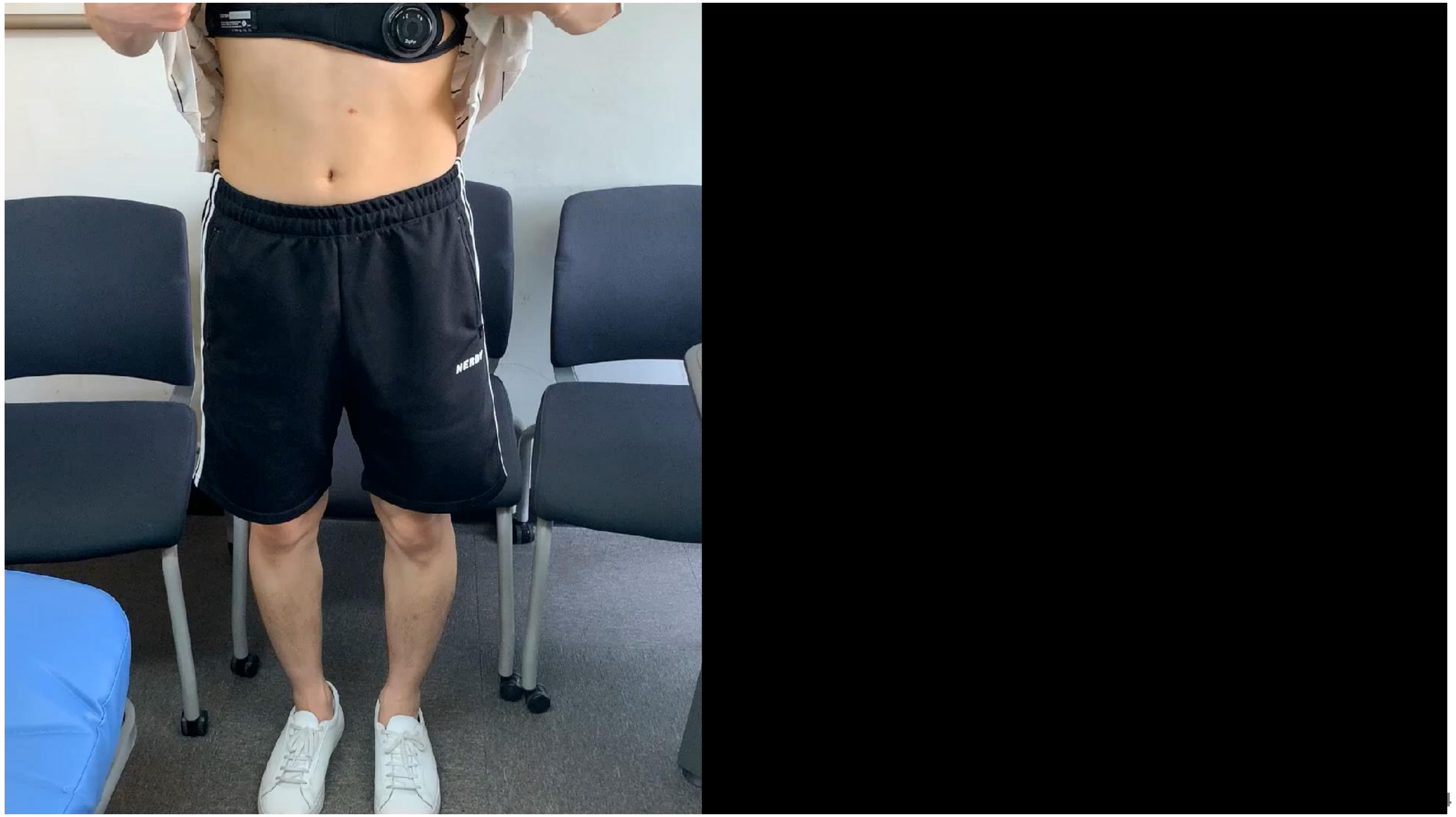
- **Working on Intelligent Mobile and Embedded Computing**
  - Research
    - **Mobile and Embedded Sensing System** for application-driven and large-scale data collection system design
    - **Machine Learning** for analyzing data collected from mobile and embedded devices and designing deep learning model architecture and mobile computing system-integrable frameworks
    - **Human-computer Interaction** for user-centric system design and evaluation based on human-affected factors in mobile computing

About me

- **Working on Intelligent Mobile and Embedded Sensing System**
  - Research
    - <u>JaeYeon Park</u>, Hyeon Cho, Rajesh K. Balan, JeongGil Ko. **"HeartQuake: Accurate Low-Cost Non-Invasive ECG Monitoring Using Bed-Mounted Geophones"**, **ACM UbiComp** 2020. Virtual Conference. **(*NRF-listed Top CS Conference - [Distinguished Paper Award]*)**



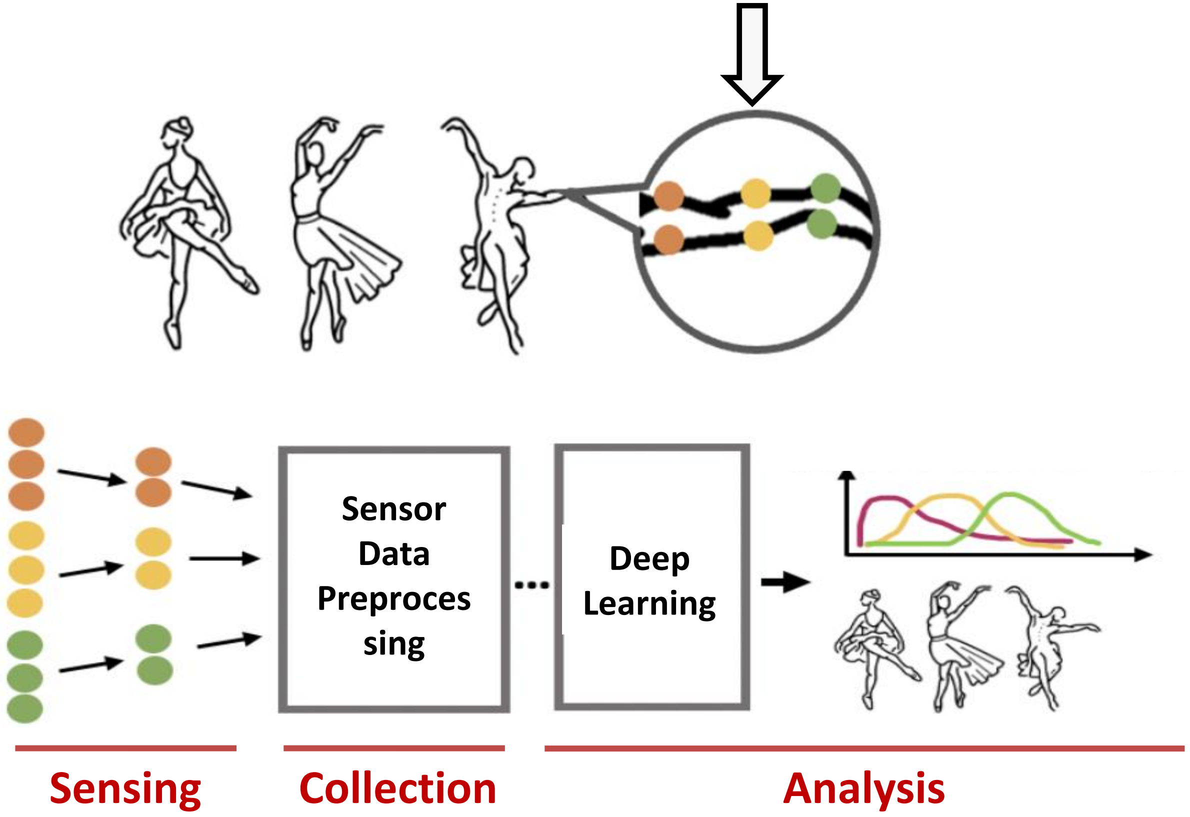




About me

- Working on Intelligent Mobile Computing
  - Research

Multiple sensors to multiple joints of a real B-Boy dancer



Contact Information

- **Office**
  - International Hall 611

- **Email**
  - jaeyeon.park@dankook.ac.kr

- **Office hours**
  - Wed, 09:00 am - 12:00 pm
  - Or send me an email for an appointment

- **Visiting office hours**
  - Mainly help on lecture material, concepts, etc.
  - Discussion on everything about mobile and embedded computing, sensing, and AI for them

Class info

- Lecture hours
  - Mon 10:30 – 12:00 / Tue 09:00 – 10:30
- Announcements, lecture notes
  - E-Campus
- Prerequisites
  - Linear Algebra, Basic and Applications of Machine Learning
- Methods of Grading
  - Midterm Exam: 30%
  - Final Exam: 35%
  - Assignment: 20%
  - Attendance: 15%

Class info

- **Goals**
  - Develop the ability to understand and apply both **classical image processing techniques** and **deep learning-based computer vision models** to solve **real-world problems**.
  - Gain skills to **select or design vision models** under various **constraints** (e.g., **mobile** or **real-time environments**) and propose **novel architectures** when necessary.
  - Enhance the capability to **analyze model architectures** and **performance**, interpret **experimental results**, and make **data-driven decisions** through **critical thinking**.

Class info

- **Weekly Schedule**

Week	Lecture Topic	Lecture Goals
1	Orientation & Introduction to Computer Vision	Course overview, history of computer vision, traditional techniques vs deep learning techniques
2	Fundamentals of Image Classification	Emergence of CNNs, automatic feature extraction from images
3	Advances in Deep Neural Network Architectures	Depth, parameter size, performance trade-offs, parameter efficiency
4	Extension of Deep Networks with Residual Learning	Vanishing gradient problem, concept of skip connection
5	Object Detection	Classification vs Detection, anchor boxes, ROI
6	Semantic / Instance Segmentation	Pixel-wise classification, difference between object and instance
7	Image Generation & Restoration	Autoencoder, denoising, inpainting techniques
8	**Midterm Exam**	

Class info

- **Weekly Schedule**

Week	Lecture Topic	Lecture Goals
9	Video Recognition	Temporal information processing, sequential frame handling
10	Attention Mechanisms and Self-Attention	Limitations of CNNs, introduction of attention mechanisms
11	Vision Transformer (ViT) Concepts	Patch embedding, position encoding
12	Hybrid Models: CNN + Transformer	Integration of CNNs and Transformers, efficiency comparison
13	Lightweight Models and Real-Time Vision	Model compression for mobile/embedded environments
14	Recent Trends	Introduction to SOTA models, paper reviews, industrial applications, Video-based Latest Technologies Study
15	**Final Exam**	

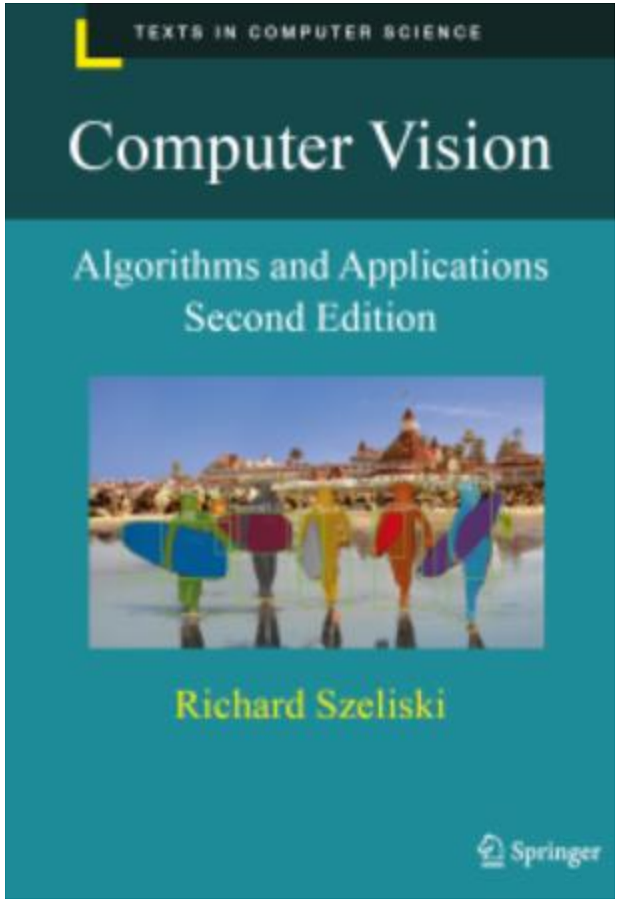
Class info

- **Grade Distribution**
  - Grade distribution guideline
    - A+/A0: 40%
    - B+/B0: 80%
    - Better grades are possible if you do really well!
      - ✓ Worse grades are also possible if you don't work hard!

- **‘F’ grade Policy: You may get a ‘F’ grade for….**
  - **Any kind of cheating or plagiarism**
  - **Absent for 10+ times** or more of the total classes
  - **Not taking an exam**, or **an exam score below 5** (out of 100) for any of the exams
  - **Inappropriate attitude during lectures**
  - Overall average score below **<u>20</u>** (out of 100)

Textbook and Assignment

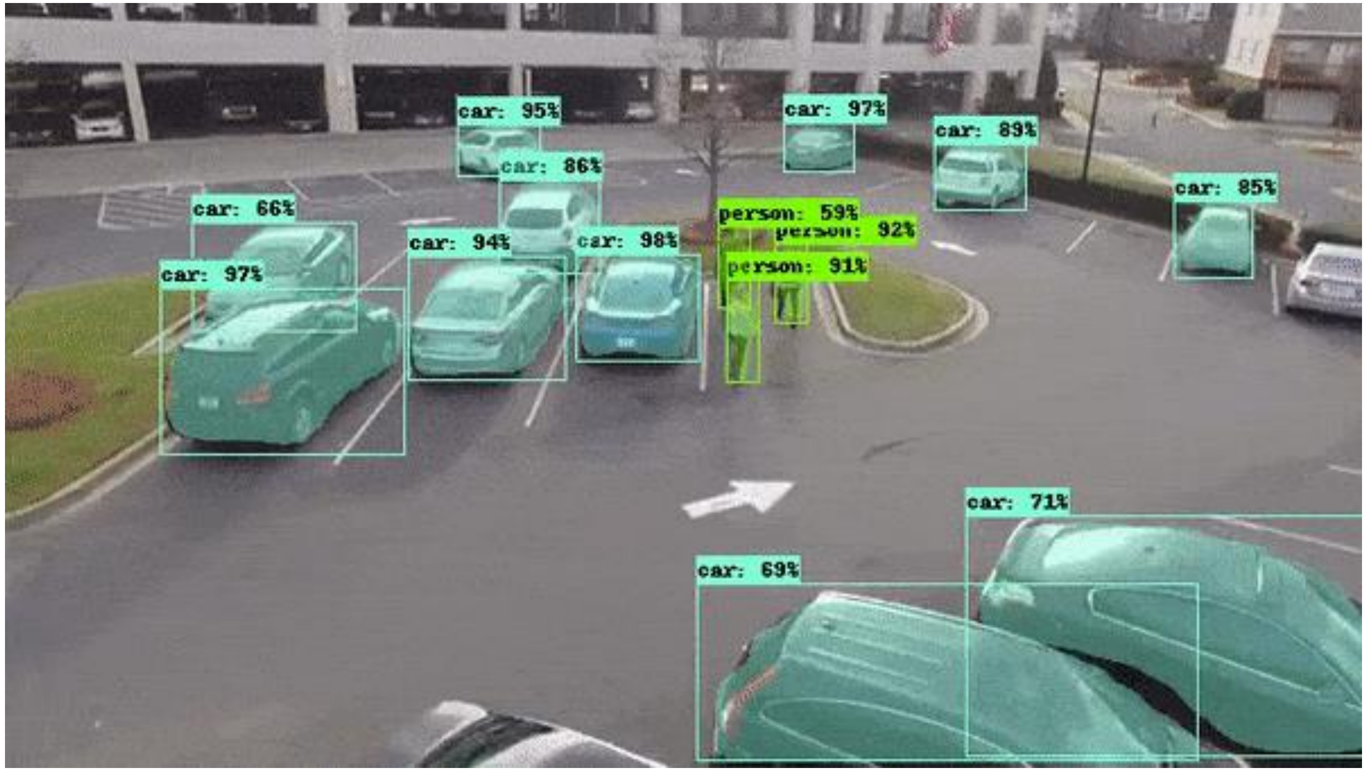
- **Textbook**
  - **Main**
    - PDF Teaching Material
  - **Reference**
    - Computer Vision: Algorithms and Applications, 2nd ed., Richard Szeliski



- **Assignment Policy**
  - About **6+ assignments** are planned
    - **1–2 assignments** will require **coding and writing reports**
    - **At least 5 assignments** will require **reading research papers** and writing a **critique**
  - Late submission
    - 30% penalty
    - No later than 1 weeks after the deadline or the end of grading period

Definition of Computer Vision

- **What is Computer Vision?**

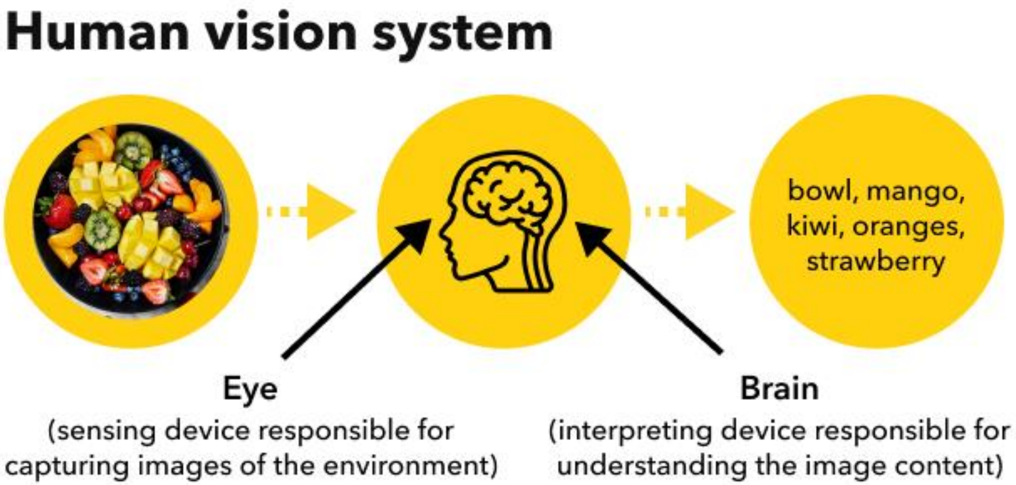


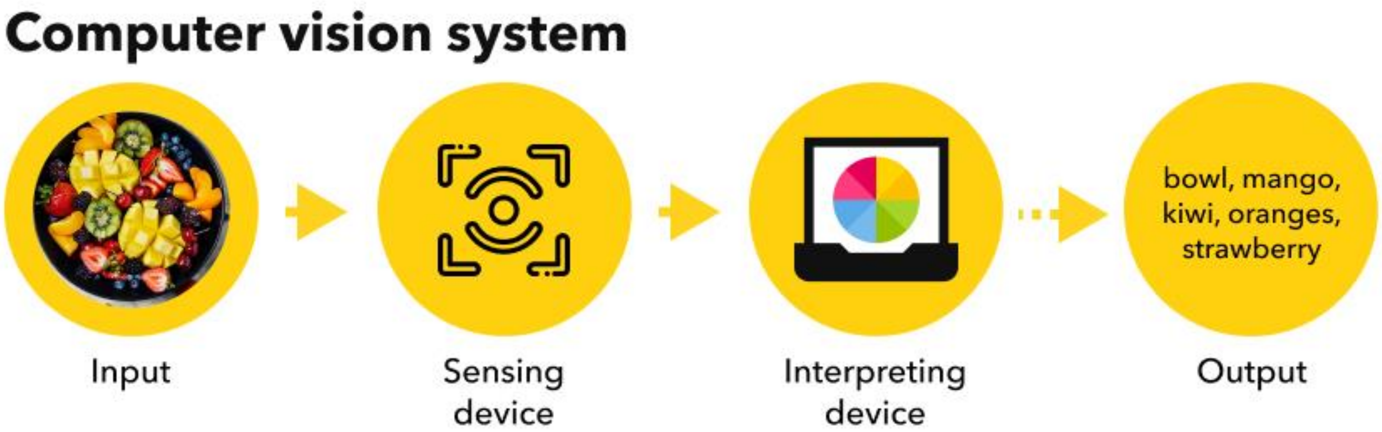
- **Computer Vision** is a field of Artificial Intelligence that enables machines to **see**, **analyze**, and **make sense** of images and videos, much like humans do with their eyes and brain.

- It combines computer science, mathematics, and cognitive science to allow understanding of visual inputs.

Human Vision vs. Computer Vision

- **How Do Humans and Computers See?**





Aspect	Human Vision	Computer Vision
Input	Light, depth, motion	Digital images (pixels)
Processing System	Eyes → Brain (visual cortex)	Pixels → Algorithms / Neural Networks
Learning Process	Experience, memory, intuition	Data-driven learning (e.g., deep learning)

Human Vision vs. Computer Vision

- **How Do Humans and Computers See?**

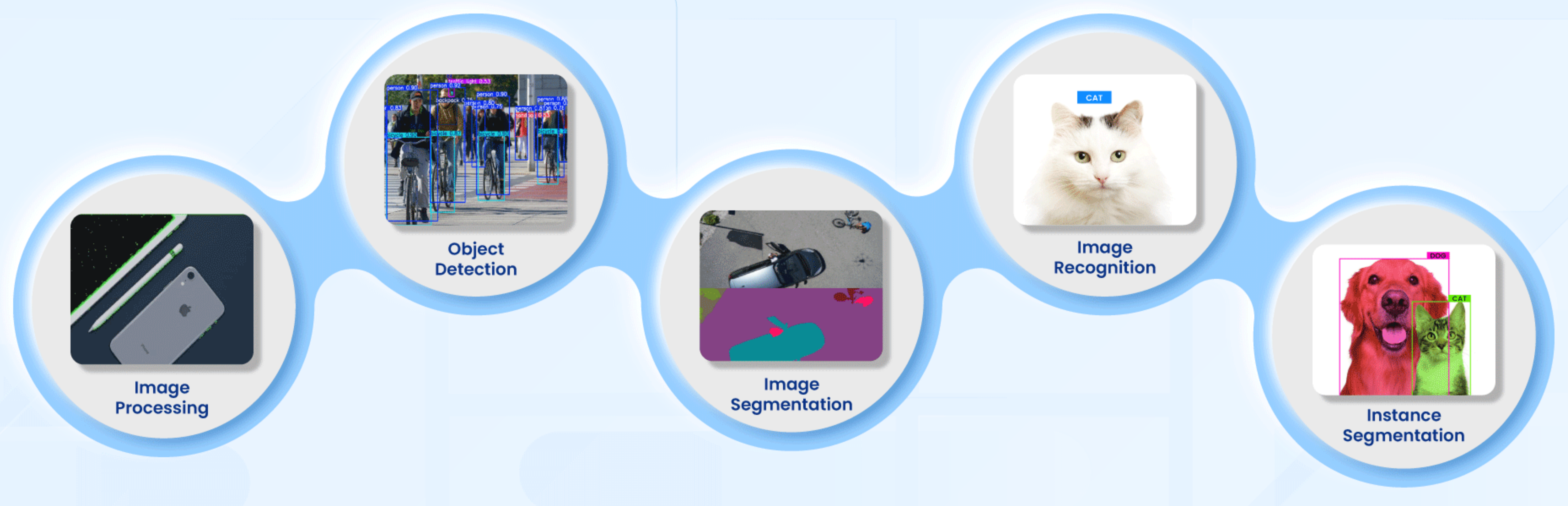
What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

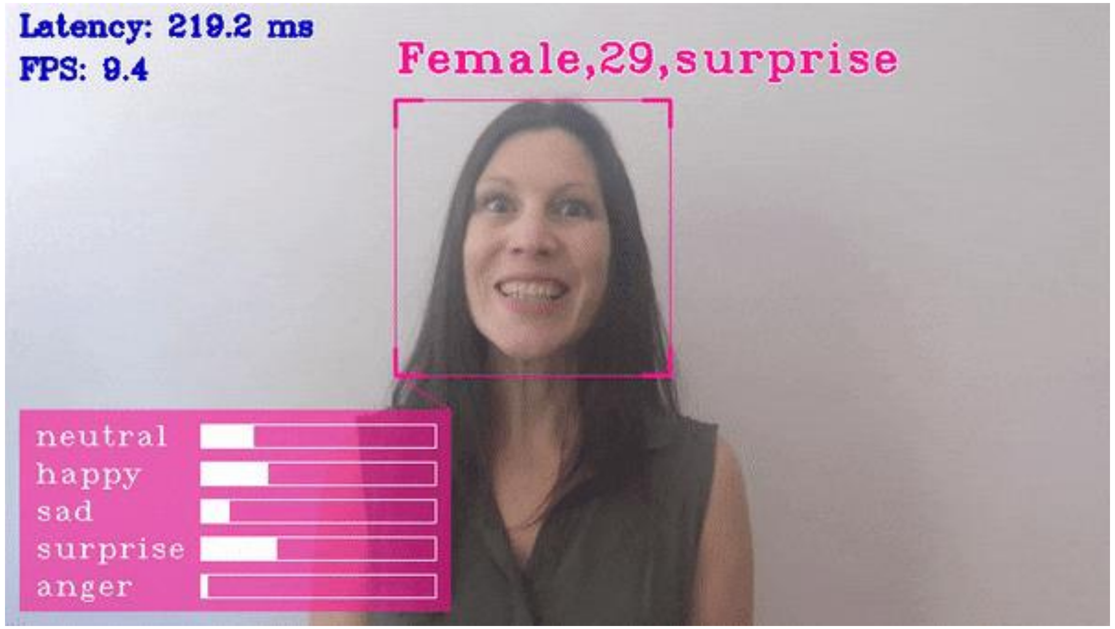
Typical Computer Vision Tasks

- **Main Tasks in Computer Vision**
  - **(1)** Image Processing, **(2)** Image Recognition (i.e., Image Classification), **(3)** Object Detection, **(4)** Image Segmentation, **(5)** Instance Segmentation, **(6)** Image Generation, etc.



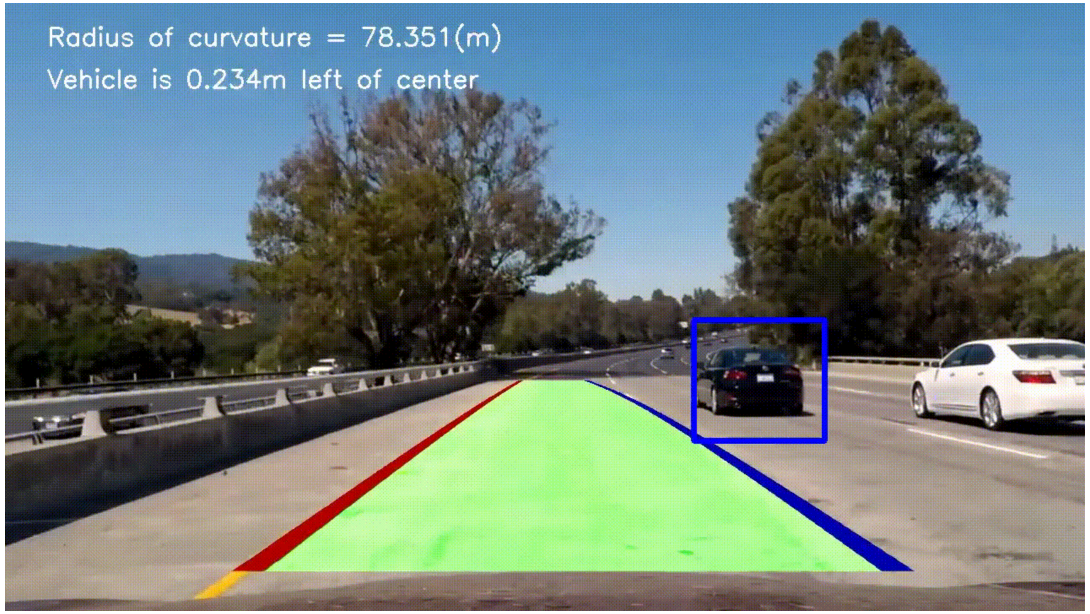
Examples of Applications

- **Where is Computer Vision Used?**
  - **Face Recognition**
    - Modern smartphones use your face as a **biometric key** to unlock the device.
    - Key Technologies: Face Detection / Face Recognition.
    - **Infrared cameras** and **depth sensors** help make it more secure



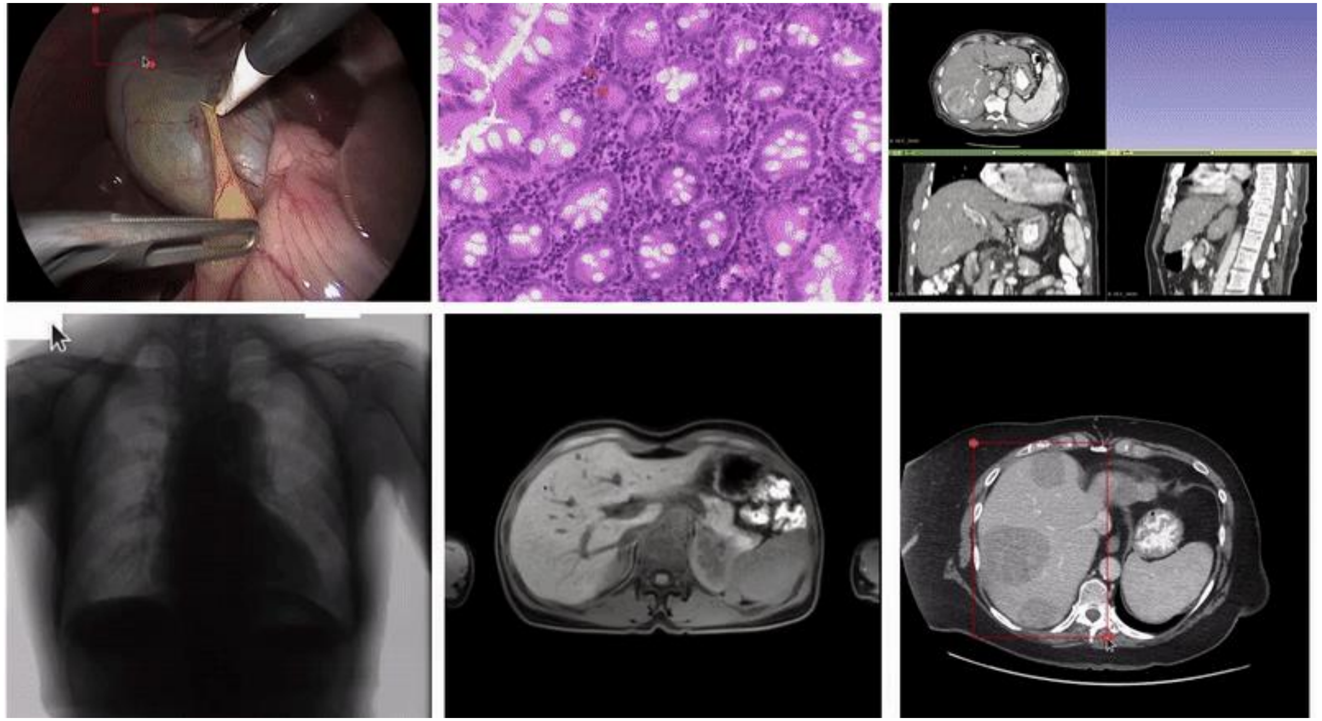
Examples of Applications

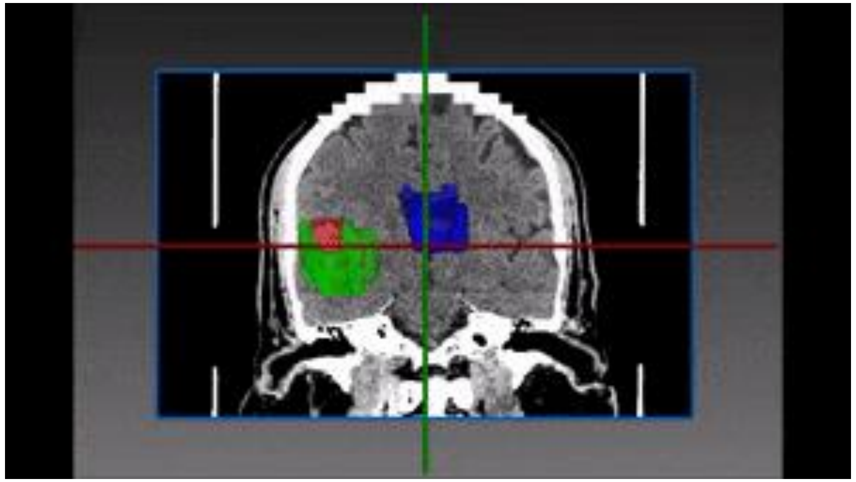
- **Where is Computer Vision Used?**
  - **Autonomous Cars – Detect Lanes, Pedestrians**
    - Self-driving cars "see" the road using cameras and sensors. They detect **lanes**, **pedestrians**, **vehicles**, and **traffic signs**.
    - Key Technologies: Object Detection / Semantic Segmentation / Sensor Fusion
    - Models like **Tesla's Occupancy Networks** combine 3D geometry and image input
    - **Bird's-eye-view** perception and **end-to-end driving policies** using Transformers



Examples of Applications

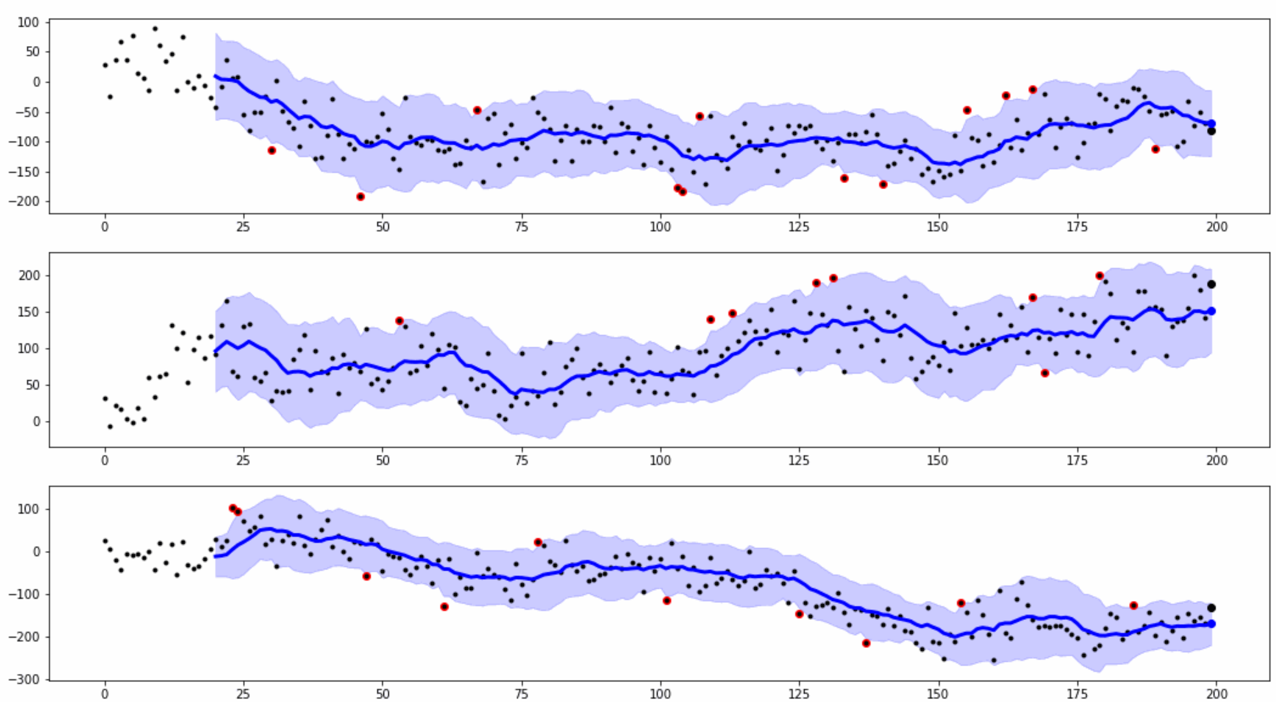
- **Where is Computer Vision Used?**
  - **Medical Imaging**
    - AI helps doctors by analyzing **X-rays**, **MRIs**, and **CT scans** to detect diseases like **tumors**, **fractures**, and **pneumonia**.
    - Key Technologies: Image Classification / Segmentation
    - Models can now outperform **junior radiologists** in some tasks.
    - Explainable AI (e.g., Grad-CAM) helps show **why** the AI thinks there's a tumor.





Examples of Applications

- **Where is Computer Vision Used?**
  - Security Systems – Detect Intruders
    - Cameras equipped with AI can detect **unusual activity**, **unauthorized access**, or **dangerous objects** in real time.
    - Key Technologies: Motion Detection / Pose Estimation / **Anomaly Detection** using spatio-temporal networks



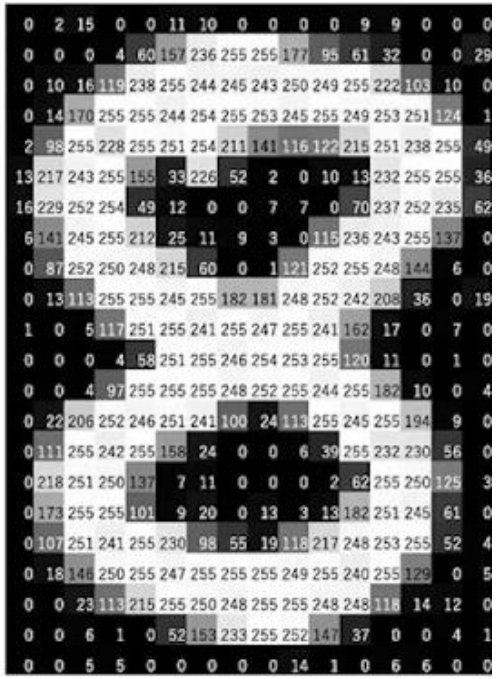
Why is Computer Vision Important?

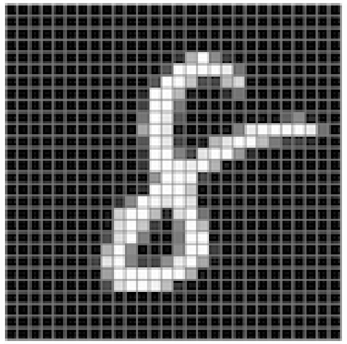
- **Key Message**
  - **"Computer Vision is a core technology that empowers machines to interpret visual data like humans."**

- **Why is it Important?**
  - It **enables machines to "see" and make decisions** based on visual input.
  - It powers real-world systems in **healthcare, self-driving cars, manufacturing, surveillance, and AR**.
  - It helps reduce human error, automate tasks, and generate insights from images and videos.

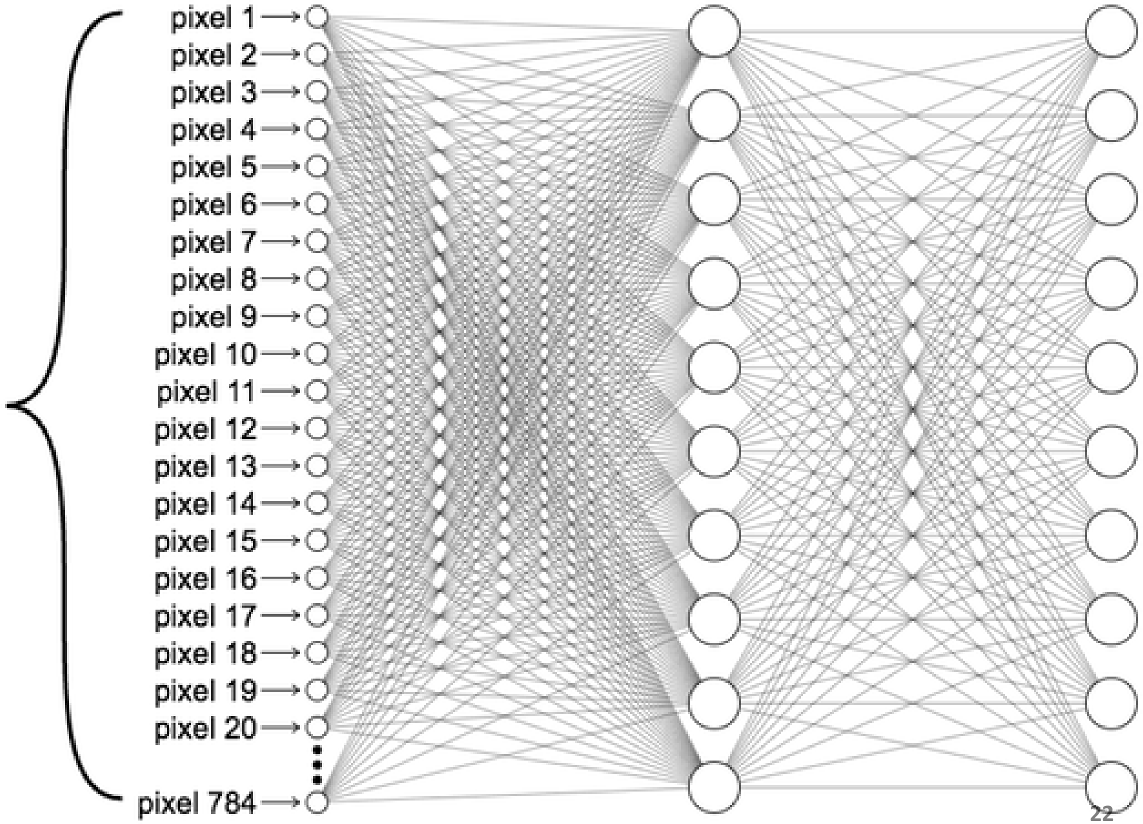
Image Data – The Foundation

- **What is an Image to a Computer?**
  - An image is a **grid of pixels** (e.g., 28×28 or 224×224)
  - Each pixel contains **intensity** (grayscale) or **RGB color** values
  - Computers view images as **numerical matrices**

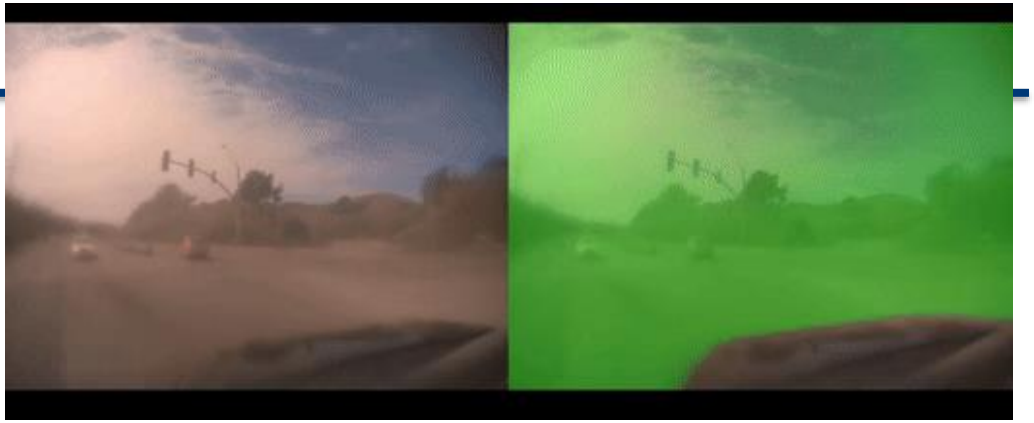




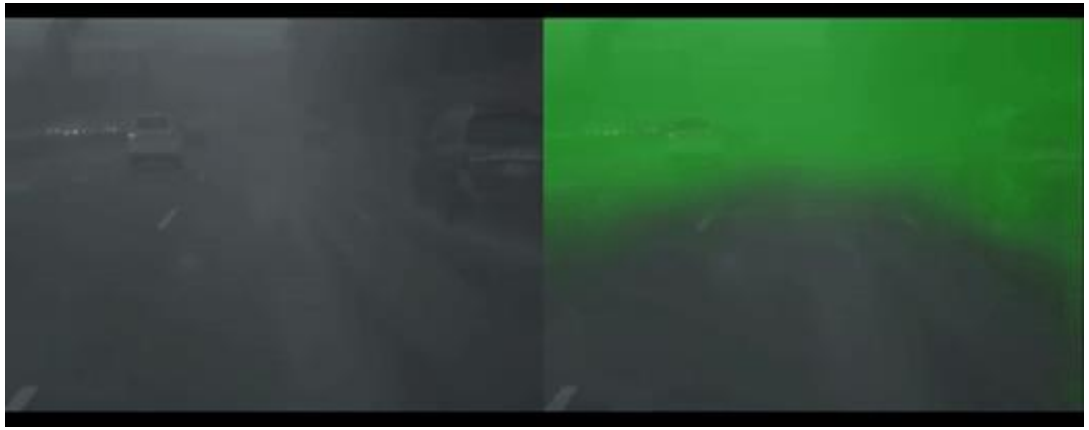
Grayscale image: 1 value per pixel

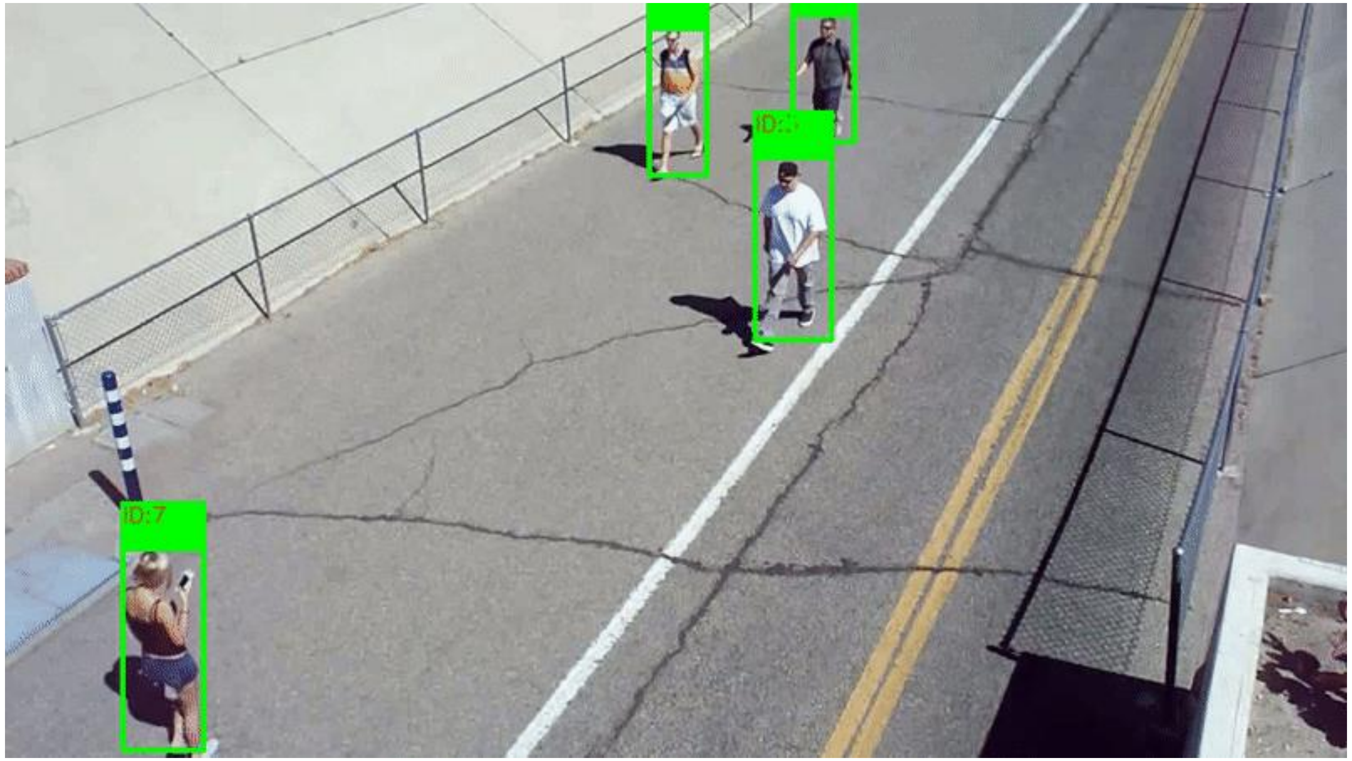


Challenges in Computer Vision



- **Why is Computer Vision Hard?**
  - **Lighting Changes** – Bright vs. dark environments
  - **Occlusion** – Object is partially hidden
  - **Deformation** – Object changes shape or pose
  - **Viewpoint Variation** – Different angles show different appearances
  - **Noise** – Blurry or poor-quality input





Evolving Paradigm in Computer Vision – From CNN to Transformers

- **From Spatial to Spatio-Temporal to Global Attention**
  - [“As visual tasks became more complex, deep learning models have evolved beyond traditional CNNs.”]()

  - **Past: CNNs for Spatial Understanding**
    - **CNNs** (Convolutional Neural Networks) were the backbone of image classification, detection, and segmentation tasks.
    - Highly effective in learning **local patterns** (e.g., edges, textures, objects).
    - Examples: **LeNet**, **VGG**, **ResNet**

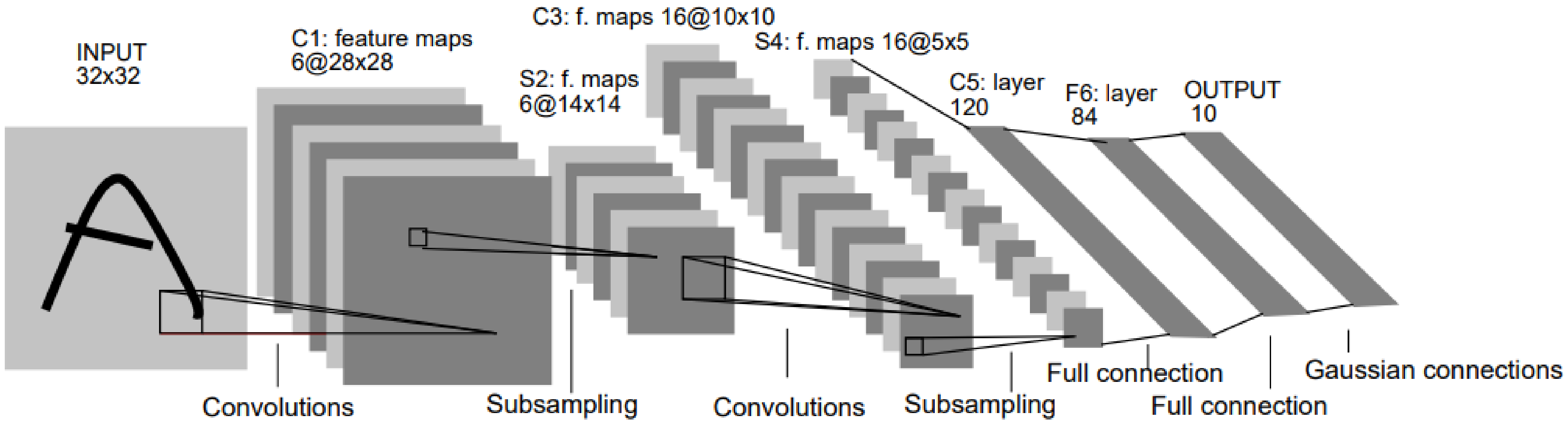


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Evolving Paradigm in Computer Vision – From CNN to Transformers

- **From Spatial to Spatio-Temporal to Global Attention**
  - **“As visual tasks became more complex, deep learning models have evolved beyond traditional CNNs.”**
  - **Transition: Spatio-Temporal Models for Video**
    - Real-world videos require understanding **how things change over time**.
    - Introduced models like:
      - ✓ **3D CNNs** – Extended convolution to time dimension
      - ✓ **ConvLSTM** – Integrated convolution with temporal recurrence

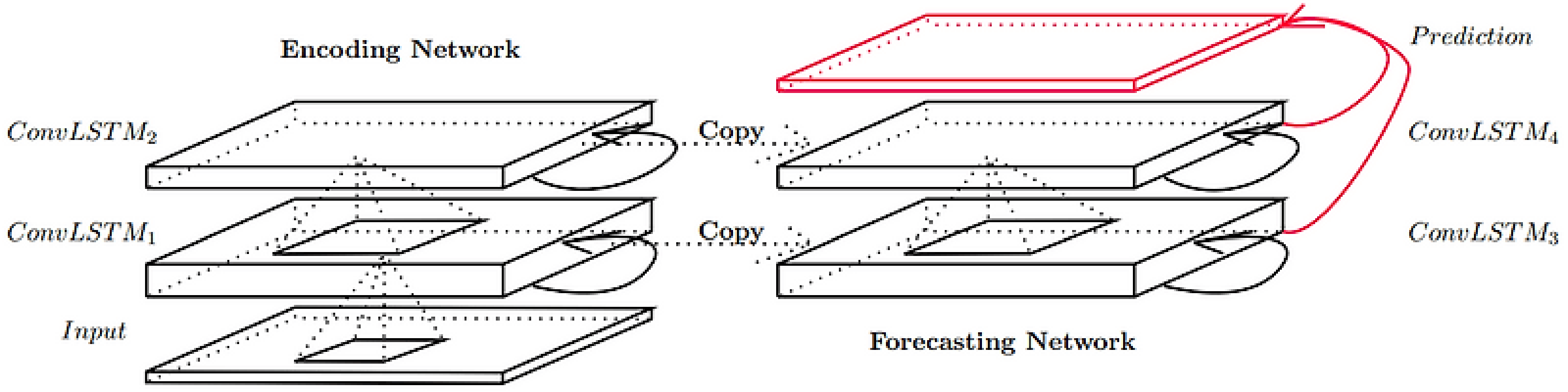
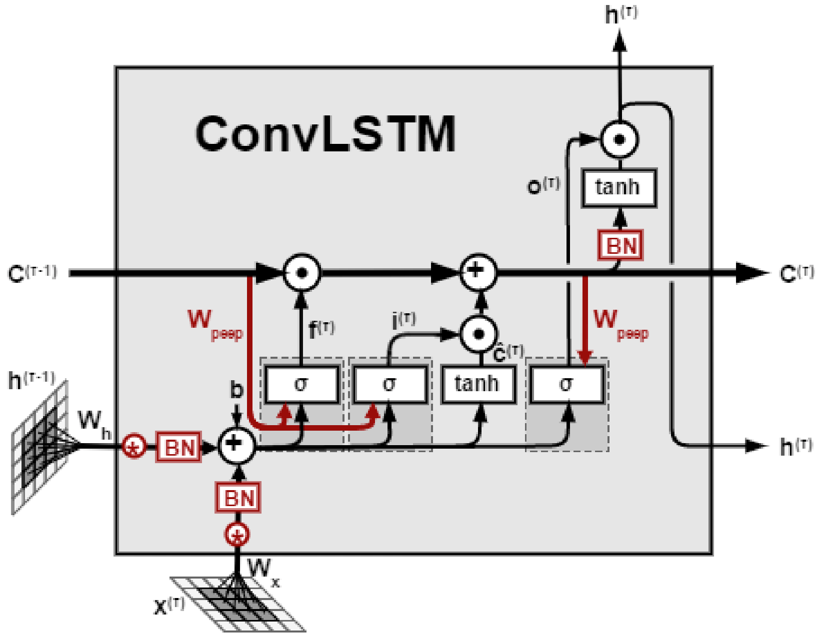


Figure 3: Encoding-forecasting ConvLSTM network for precipitation nowcasting



Evolving Paradigm in Computer Vision – From CNN to Transformers

- **From Spatial to Spatio-Temporal to Global Attention**
  - **“As visual tasks became more complex, deep learning models have evolved beyond traditional CNNs.”**

  - **Now: Transformers for Vision**
    - Originally from NLP, ***Transformers*** are now reshaping Computer Vision.
    - Key innovation: **Self-Attention**, which enables models to capture **global dependencies**.
    - Widely used in:
      - ✓ **Vision Transformers (ViT)** for classification
      - ✓ **Video Swin Transformer, TimeSformer** for video understanding
      - ✓ **Segment Anything**, **DINOv2**, **CLIP** for general vision tasks