SWENETIX

COMPUTER VISION

Office Course

JUN 25, 2026

WHAT IS COMPUTER VISION

Computer Vision is a branch of Artificial Intelligence that enables computers to “see,” understand, and interpret images and videos in a way similar to humans.

GOAL: Replicate human vision capabilities

Human Vision

- Eyes capture
- Brains process
- Recognizes
- Understands
- Learns from experience

Computer vision

- Camera captures
- AI processes
- Algorithms
- Makes decisions
- Learns from data

SWENETIX

WHY COMPUTER VISION MATTERS

In today's digital world, visual data is generated every second from countless sources, including:

- Smartphones
- CCTV and security cameras
- Medical imaging systems
- Self-driving vehicles
- Social media platforms
- Industrial and surveillance systems

The amount of image and video data produced every day is far beyond what humans can manually review or analyze.
Computer Vision enables computers to automatically understand, process, and analyze visual information quickly, accurately, and at scale.

SWENETIX

REAL WORLD APPLICATIONS

Face Recognition

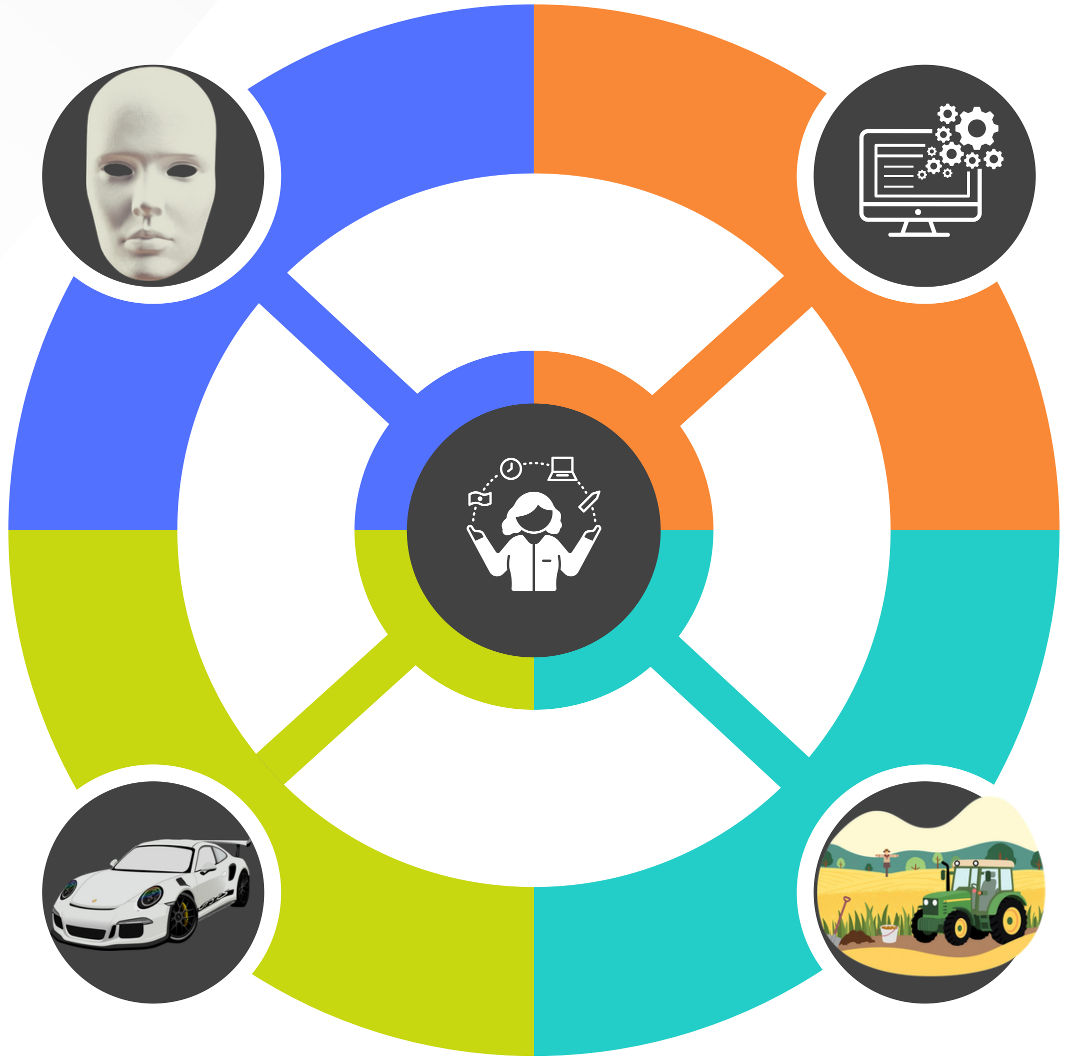
Used in:

- Phone unlocking
- Airport security
- Attendance systems

Medical Imaging

Hospitals use Computer Vision to analyze:

- X-rays
- MRI scans
- CT scans



Self-Driving Cars

Autonomous vehicles use Computer Vision to:

- detect roads
- identify pedestrians
- recognize traffic signs
- avoid obstacles

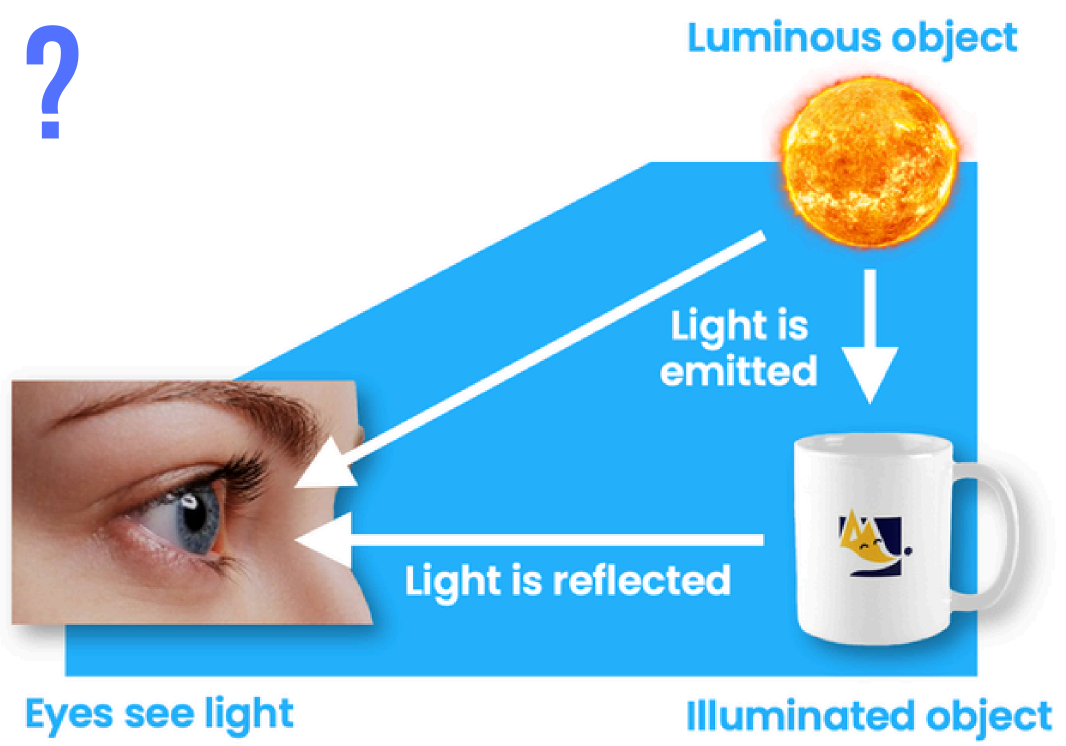
Agriculture

Farmers use drones and cameras to:

- monitor crops
- detect plant diseases
- estimate harvest quality

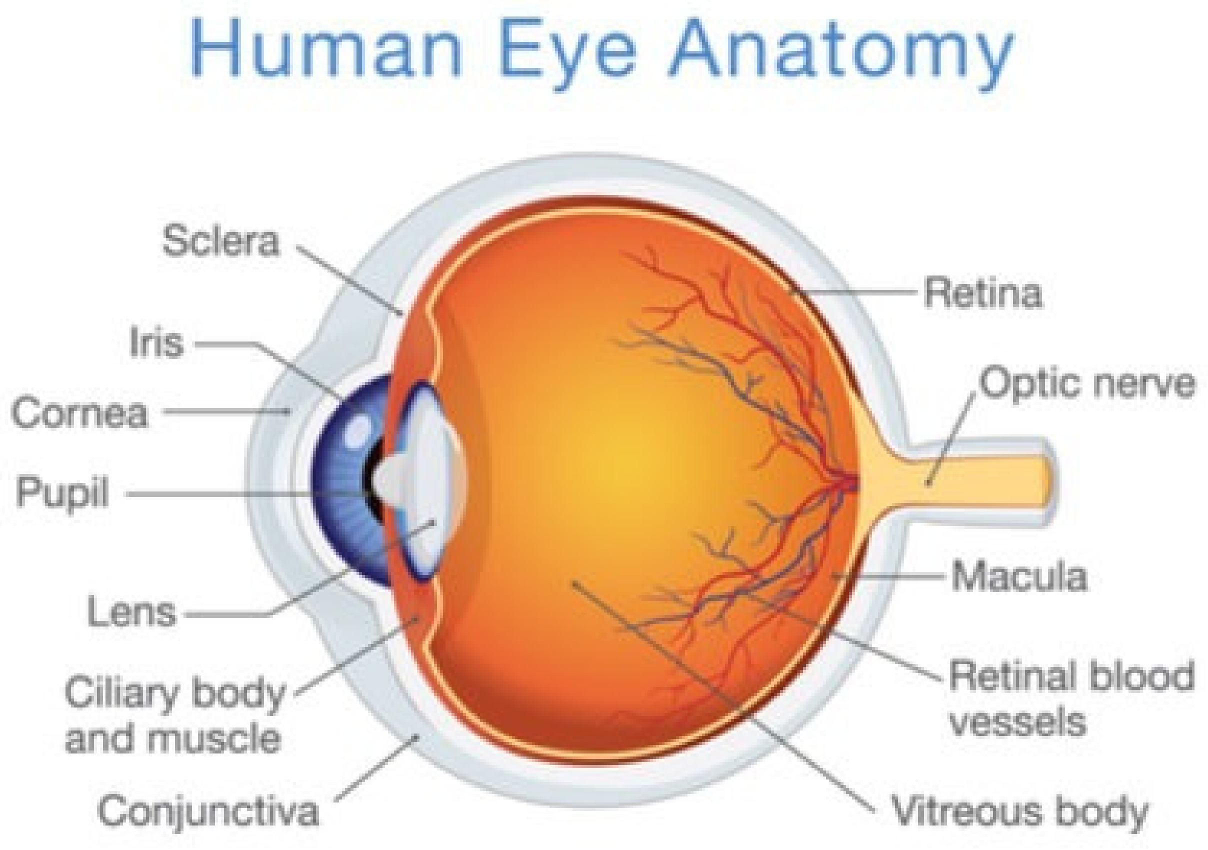
THE MIRACLE OF HUMAN VISION

HOW DID WE SEE ?



STEP 1: LIGHT ENTERS AND FOCUSES.

Light bounces off an object – say, this coffee cup – and travels to your eye. It passes through the cornea, the clear dome that does most of the bending, and then through the lens, which fine-tunes the focus. The goal is to project a crisp, upside-down image onto the back wall of your eye – the retina.



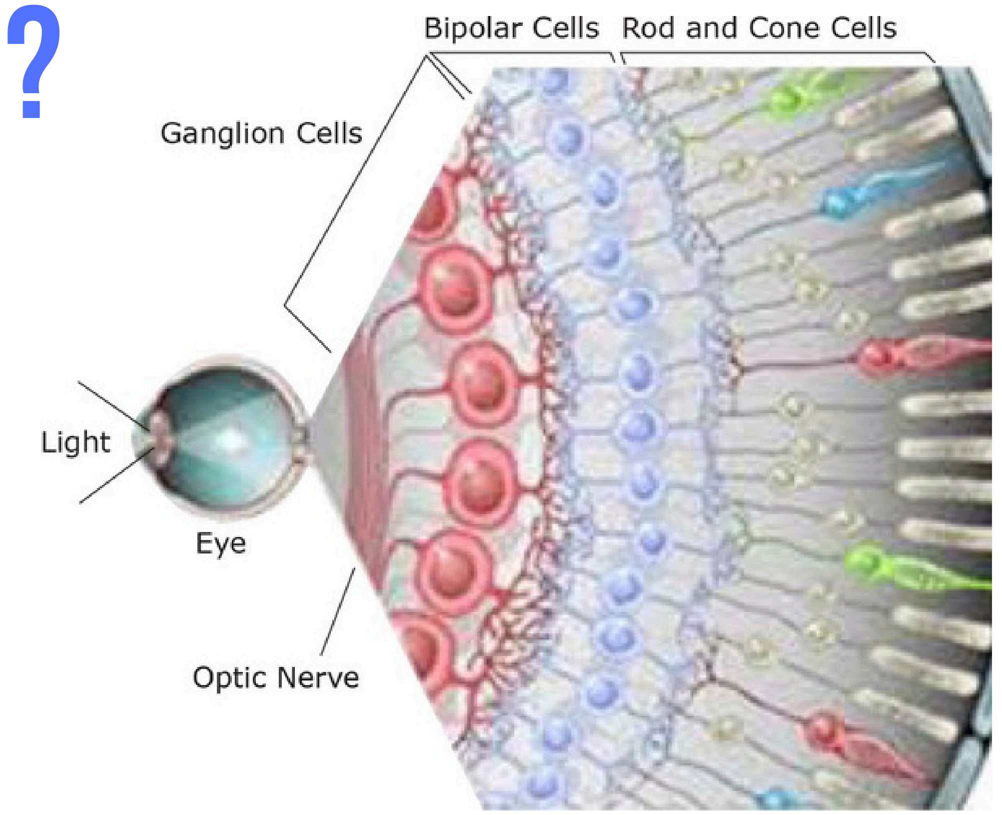
STEP 2: THE RETINA CONVERTS LIGHT INTO ELECTRICAL SIGNALS

The retina isn't just a passive screen. It's a piece of your brain that grew out to capture light. It's carpeted with about 120 million rod cells and 6 million cone cells. Rods let you see in dim light but only in black and white. Cones give you sharp, color vision in bright light. These cells turn light energy into electrical impulses – the universal language of the nervous system.

SWENETIX

THE MIRACLE OF HUMAN VISION

HOW DID WE SEE ?



STEP 3: EARLY PROCESSING STARTS IN THE EYE ITSELF.

Here’s the first shocking fact: before the signal even leaves your eye, your retina does basic image processing. It detects edges, contrast, and motion. Ganglion cells in the retina fire when they spot a dark/light boundary moving in a particular direction. So your eye isn’t just a camera sensor; it’s a tiny pre-processor.

STEP 4: THE SIGNAL TRAVELS TO THE VISUAL CORTEX.

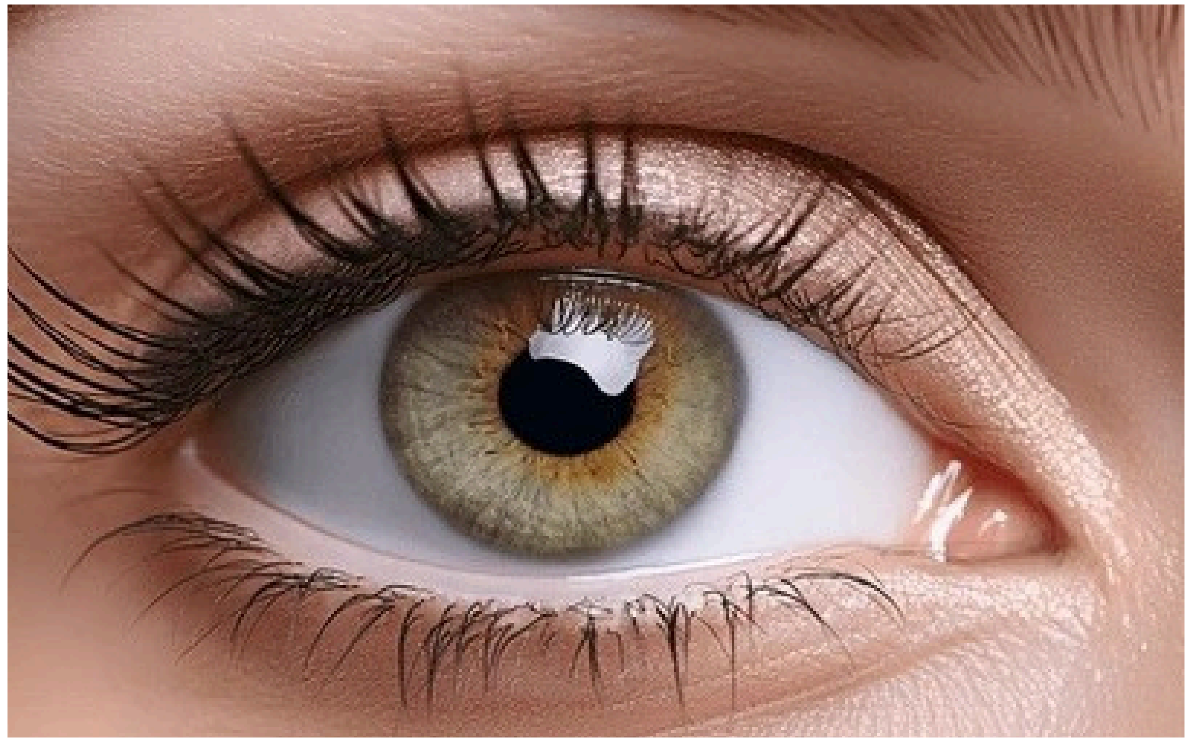
The bundled signals race through the optic nerve, cross at a relay station called the thalamus, and land in the visual cortex at the very back of your brain. This region is mapped like a pixel grid but with layers of increasing sophistication.

STEP 5: THE BRAIN BUILDS YOUR PERCEPTION IN STAGES.

SWENETIX

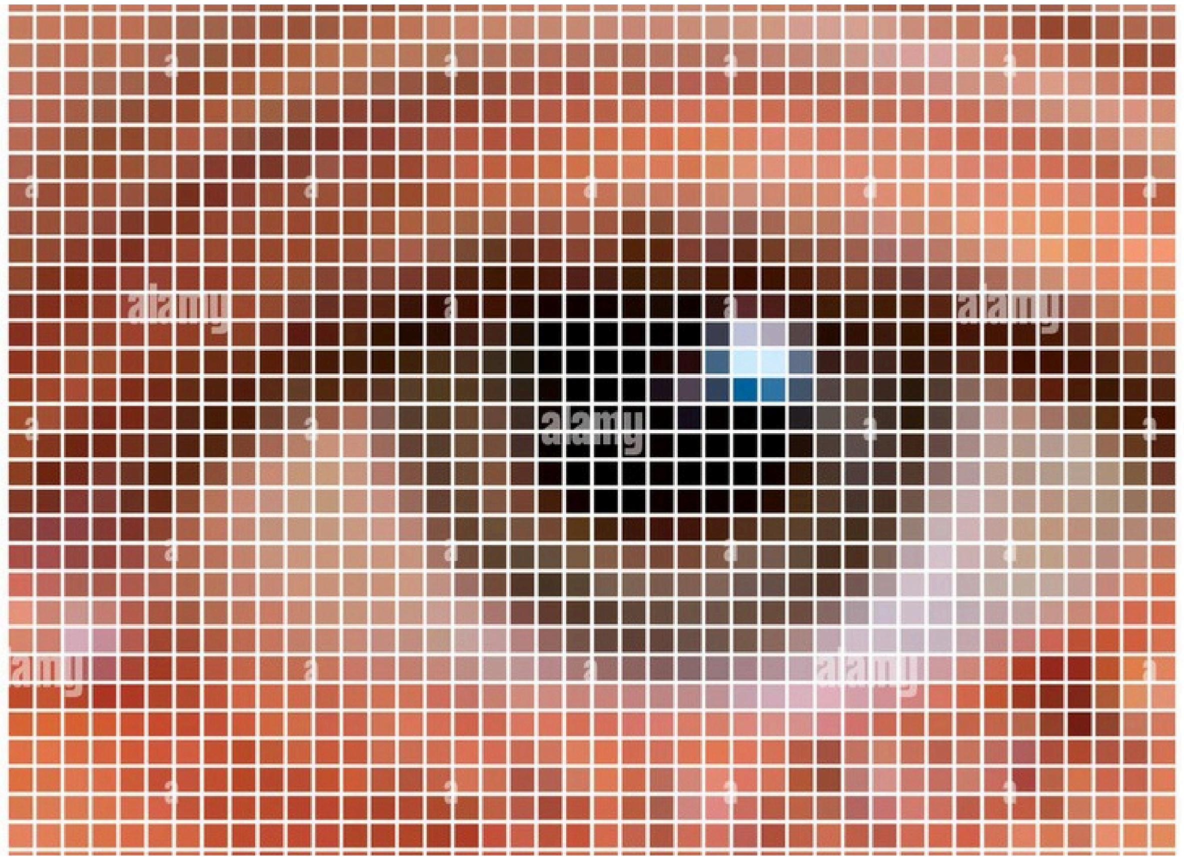
HOW DID COMPUTERS SEE ?

BASICS OF DIGITAL IMAGE



A digital image is a picture stored inside a computer.
Unlike humans, computers do not “see” images visually.
They only understand numbers.

PIXEL



A pixel is the smallest image component and thus shows the smallest detail
Each pixel contains values representing a particular color

An image is made of thousands or millions of tiny points called pixels.
Each pixel stores numerical information about:

- brightness
- color
- intensity

The computer analyzes these numerical values to understand the image.

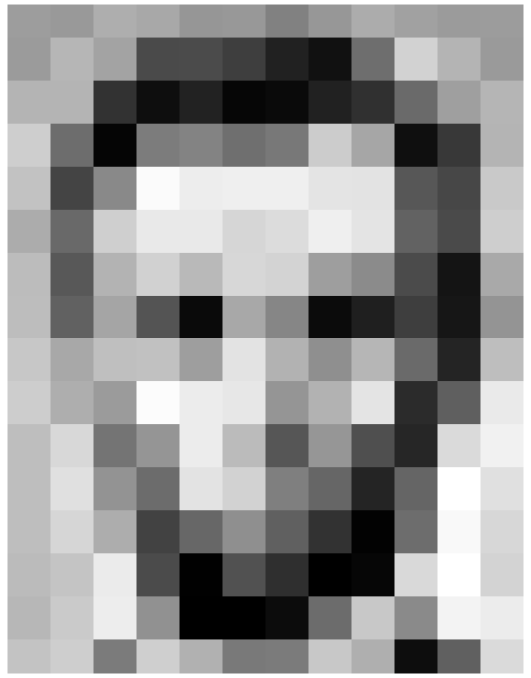
HOW COMPUTERS "SEE" THE PIPELINE

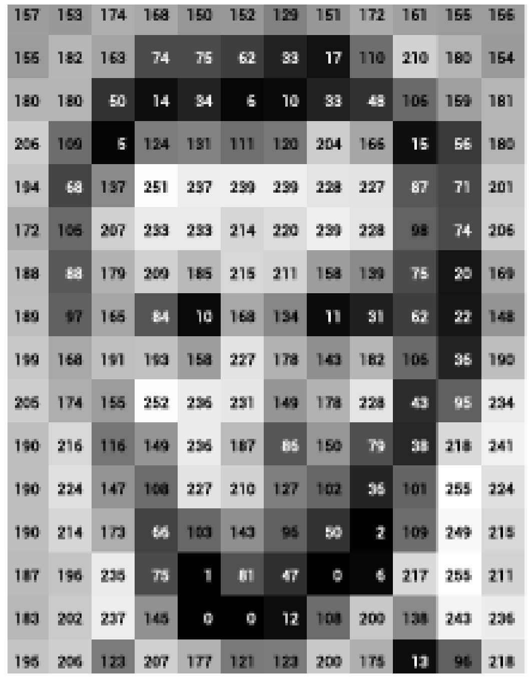
SWENETIX

STEP 1

Raw Image → Pixel

Every digital image is just a grid of tiny colored squares called pixels. Each pixel has three numbers - Red, Green, Blue values between 0 and 255.





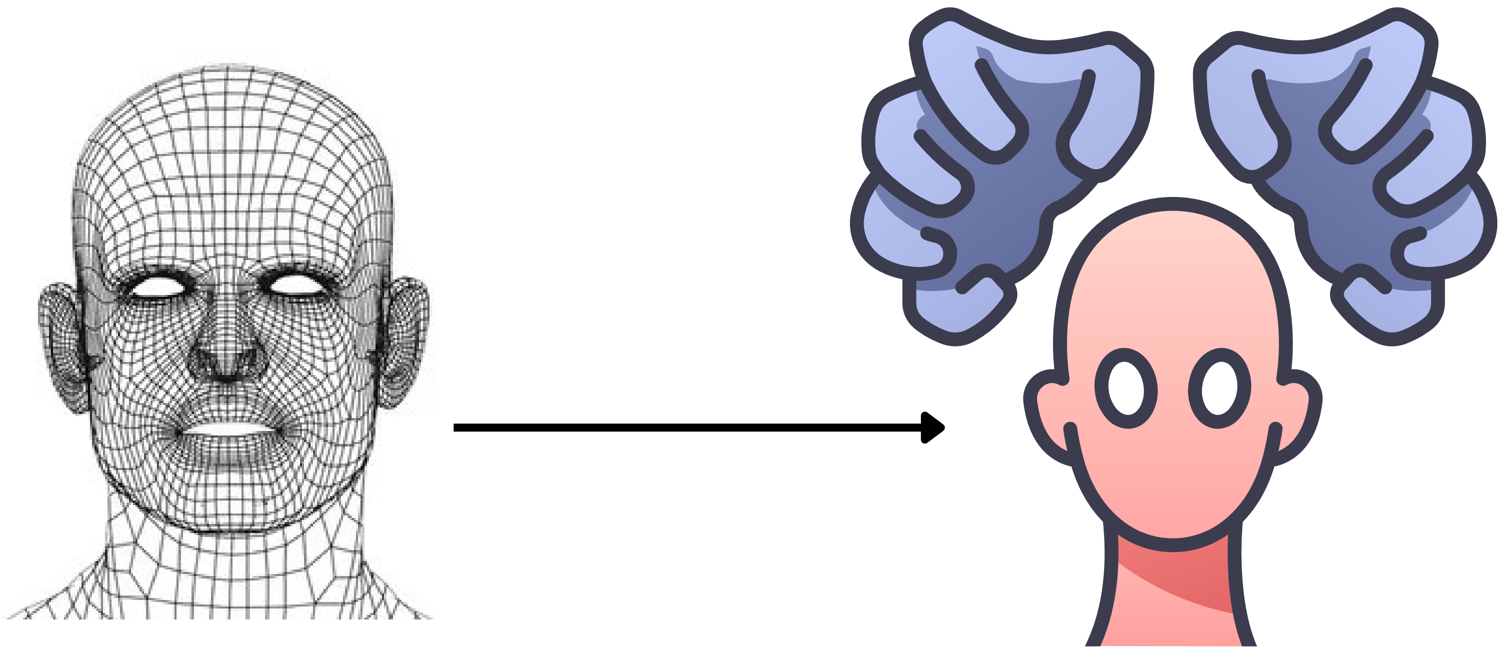
157	153	174	168	150	152	129	151	172	161	156
155	182	163	74	75	62	33	17	110	210	180
180	180	50	14	34	6	10	33	48	106	159
206	109	5	124	131	111	120	204	166	15	56
194	68	137	251	237	239	239	228	227	87	71
172	105	207	233	233	214	220	239	228	98	74
188	88	179	209	185	215	211	158	139	75	20
189	97	165	84	10	168	134	11	31	62	22
199	168	191	193	158	227	178	143	182	106	36
205	174	155	252	236	231	149	178	228	43	95
190	216	116	149	236	187	86	150	79	38	218
190	224	147	108	227	210	127	102	36	101	255
190	214	173	66	103	143	96	50	2	109	249
187	196	235	75	1	81	47	0	6	217	255
183	202	237	145	0	0	12	108	200	138	243
195	206	123	207	177	121	123	200	175	13	96

HOW COMPUTERS "SEE" THE PIPELINE

STEP 2

Image Processing Techniques

1. **Grayscale Conversion**
2. **Thresholding (Binarization)**
3. **Gaussian Blur (Noise Reduction)**
4. **Edge Detection**



HOW COMPUTERS “SEE” THE PIPELINE

GRAYSCALE CONVERSION

Why convert to grayscale? Color adds complexity (3 channels vs. 1) but often isn't needed for structure detection.

Real Example - Document Scanning:
When you scan a document, the scanner converts everything to grayscale or black-and-white because color contains no useful information for text recognition.

THRESHOLDING (BINARIZATION)

Thresholding converts grayscale images to pure black and white (binary images).

Real Example - Reading License Plates

HOW COMPUTERS “SEE” THE PIPELINE

GAUSSIAN BLUR (NOISE REDUCTION)

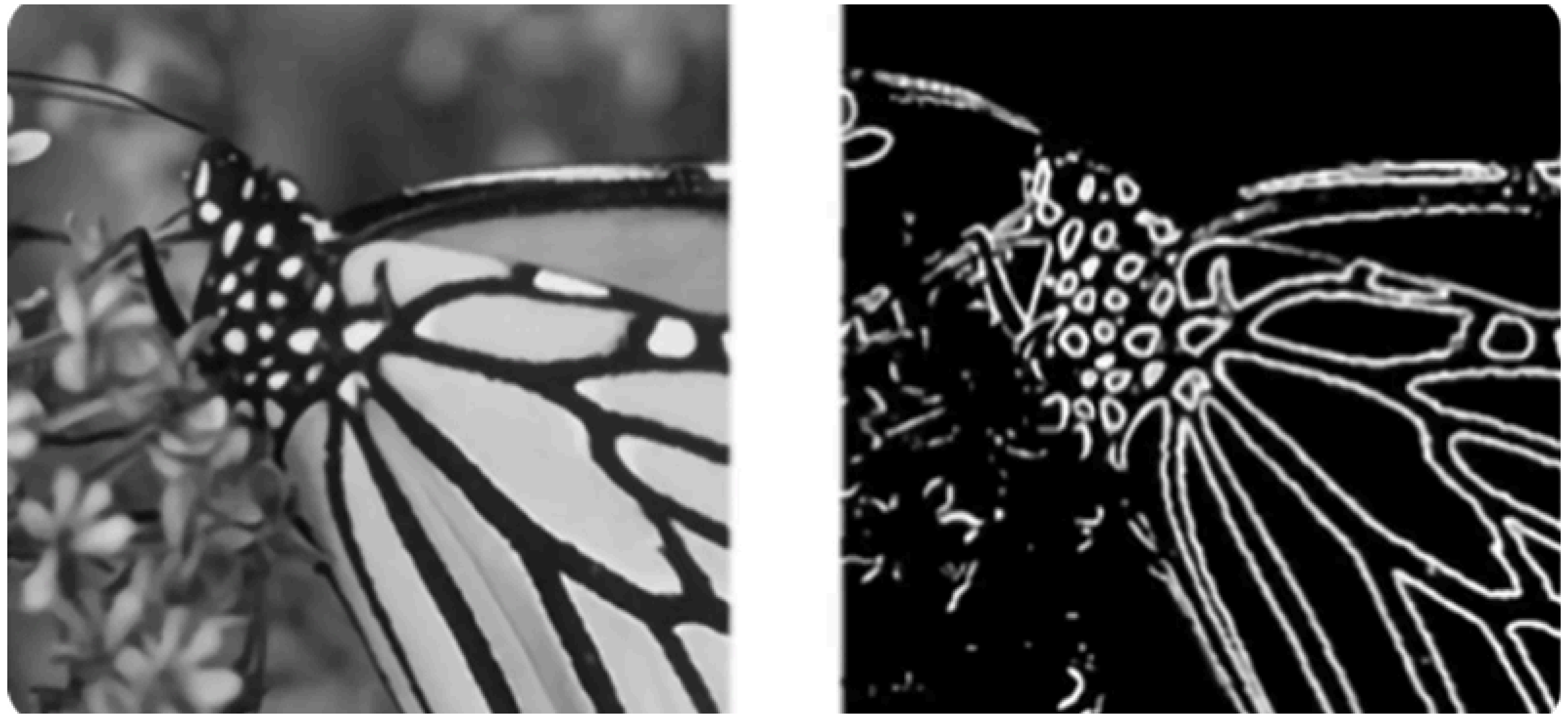
Named after Carl Friedrich Gauss (famous mathematician), this blur averages pixels with their neighbors using a weighted average (closer pixels matter more).

Real Example - Surveillance Video:
Security cameras often produce noisy images (grainy). Applying Gaussian blur before face detection improves accuracy by 15-20% because the noise is smoothed out.

EDGE DETECTION

Edges are where brightness changes sharply—these are usually object boundaries.

Real Example - Self-Driving Cars



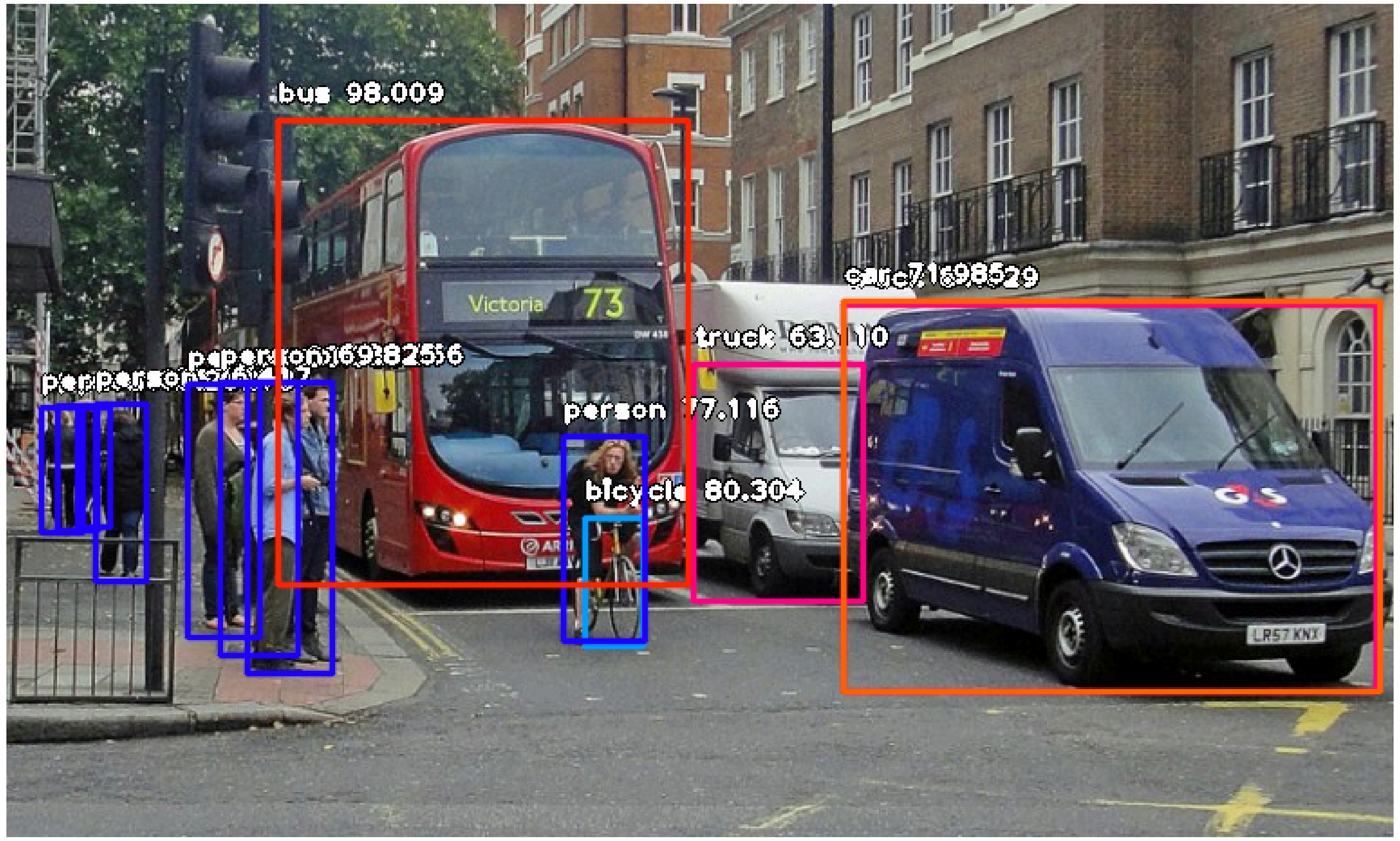
SWENETIX

HOW COMPUTERS “SEE” THE PIPELINE

STEP 3

Feature Detection & Descriptors

- **Corners**: Point where two edges meet (Harris Corner Detection)
- **Blobs**: Regions differing in brightness (SIFT, SURF)
- **Edges**: Boundaries between objects (Canny)
- **Ridges**: Elongated objects (vessels, roads)



TECHNOLOGIES

OpenCV

OpenCV (Open Source Computer Vision Library) is a BSD-licensed library containing 2500+ optimized algorithms for image and video analysis.

OpenCV helps computers:

- understand images
- process videos
- detect objects
- recognize faces
- track movement

- Lightweight – runs on CPU, no GPU required
- Real-time – processes 30+ FPS on a standard laptop
- Huge ecosystem – tutorials, bindings, pre-built binaries
- Portable – works identical on Windows / macOS / Linux

SWENETIX

TECHNOLOGIES

Mediapipe

MediaPipe is a google framework for building multimodal (video, audio, time series) applied ML pipelines. It provides pre-trained models for hand, face, pose, and object detection, optimised for real-time, on-device inference.

- No training required – models ready to use
- Fast – 30+ FPS even on CPU (using TFLite)
- Accurate – based on BlazePalm / FaceMesh research
- Easy integration – Python API as simple as OpenCV

SWENETIX

TECHNOLOGIES

deepFace

DeepFace is a lightweight Python deep learning framework for face recognition and attribute analysis (age, gender, emotion, race). It wraps several state-of-the-art models (VGGFace, FaceNet, ArcFace) into a simple API.

- No training required – model already knows what a smile looks like
- Multi-emotion – could later detect surprise, sadness, etc.
- Simple API – one line of code gives emotion scores
- Works in real-time – on CPU after initial model load (~200ms per face)

THANKS FOR YOUR ATTENTION

END OF SLIDE