5.2 Computer vision (overview)

Computer Vision is a field of Artificial Intelligence (AI) that focuses on enabling machines to interpret and make decisions based on visual information from the world, such as images and videos. It allows computers to "see" and analyze like humans but with speed, scale, and consistency.

Core Goals of CV

1.Perception – extracting useful information from raw images/videos.

2.Understanding – recognizing what objects are and how they relate.

3.Decision Making – using visual data for actions (e.g., autonomous driving).

Key Tasks in Computer Vision

Image Classification → Identifying what an image contains (e.g., "This is a cat").

Object Detection → Locating and labeling multiple objects in an image.

Image Segmentation → Dividing an image into regions (e.g., separating background from objects).

Facial Recognition → Identifying or verifying people based on facial features.

Pose Estimation → Detecting the position and orientation of people or objects.

Image Generation → Creating new, realistic images (GANs, diffusion models).

Video Analysis → Detecting actions, movements, and anomalies in video streams.

Techniques & Methods

Classical CV techniques: Edge detection, feature extraction (SIFT, HOG).

Deep Learning Methods

Convolutional Neural Networks (CNNs) → Backbone of modern CV.

R-CNN, YOLO, SSD → Object detection architectures.

Transformers (Vision Transformers – ViT) → Used in state-of-the-art models.

Preprocessing & Enhancement: Image filtering, denoising, resizing, augmentation.

Applications of Computer Vision

Healthcare: Medical imaging, tumor detection, X-rays, MRI scans.

Autonomous Vehicles: Lane detection, traffic sign recognition, pedestrian tracking.

Security & Surveillance: CCTV monitoring, facial recognition.

Retail & E-commerce: Visual search, automated checkout, product tagging.

Agriculture: Crop monitoring, pest detection, yield prediction.

Robotics: Navigation, object manipulation.

Entertainment & Media: AR/VR, motion capture, special effects.

Challenges in Computer Vision

Variability in lighting, angle, and background.

Occlusion (objects hiding behind others).

Real-time processing speed for videos.

Generalization across different environments.

Ethical issues (privacy in facial recognition, bias in datasets).

Future of Computer Vision

Integration with NLP → multimodal AI (e.g., captioning images, visual question answering).

Growth in edge computing → running CV models on mobile/IoT devices.

Stronger generative models for realistic image/video synthesis.

Wider adoption in healthcare, smart cities, and robotics.

In short: Computer Vision is the science of making computers “see” and understand the visual world, powering technologies from self-driving cars to medical diagnosis.