Artificial Intelligence CV

---

1. Introduction

Computer Vision (CV) is a field of AI that enables computers to **see, understand, and interpret visual information** from images or videos—just like humans.

CV allows machines to extract meaningful insights from:

- Photos
- Videos
- Live camera feeds
- Medical scans
- Satellite imagery

It’s used everywhere: face unlock, lane detection, barcode scanning, healthcare imaging, robotics, AR filters, and more.

---

2. What is Computer Vision?

Definition:
Computer Vision is a subfield of AI that trains machines to interpret and understand visual data (images/videos) using algorithms and deep learning models.

Simple Meaning:
CV = Making computers **see** + **understand** + **take action** based on visuals.

---

3. Need of Computer Vision

1 Automate visual tasks

- Detect defects in factories
- Monitor safety in workplaces

- Recognize products in retail

2 Improve accuracy

Humans get tired; AI doesn’t.
CV can process **millions** of images consistently.

3 Real-time decision-making

Used in:

- Self-driving cars
- Robotics
- Security surveillance

4 Unlock new experiences

- AR/VR
- Face filters
- Emotion detection

---

4. Real-World Applications of Computer Vision

A. Face Recognition

- iPhone Face ID
- Attendance systems
- Surveillance systems

B. Object Detection

- Detect pedestrians, traffic signs
- Retail shelf monitoring
- Sport analytics

C. Image Classification

- Cat vs dog
- Identifying cancers from MRI
- Quality checking in factories

D. Motion & Activity Recognition

- CCTV anomaly detection
- Fitness apps detecting posture
- Security intrusion detection

E. OCR (Optical Character Recognition)

- Scan documents
- Read number plates
- Bill/invoice digitization

F. Autonomous Vehicles

- Lane detection
- Signboard recognition
- Collision avoidance

G. AR/VR

- Snapchat filters
- Metaverse
- Virtual try-on (glasses, clothes)

H. Medical Imaging

- Tumor detection

- X-ray analysis
- Blood cell counting

---

5. Common Tasks in Computer Vision

1. Image Classification

Predict the category of an image.
Example: Dog, Cat, Car.

---

2. Object Detection

Locate and classify multiple objects.
Frameworks: YOLO, SSD, Faster R-CNN.

---

3. Semantic Segmentation

Color each pixel based on its class.
Example: Sky = blue, road = grey.

---

4. Instance Segmentation

Segmentation + object identity
Example: Separating two people standing close.

---

5. Keypoint Detection / Pose Estimation

Identify human body keypoints.
Example: Detecting workout posture.

---

6. OCR

Extract text from images.
Tools: Tesseract, Google Vision API.

---

7. Tracking

Follow an object across frames.
Example: Football player tracking.

---

8. Image Generation

- Stable Diffusion
- GANs
- DALL·E

---

6. Approaches Used in Computer Vision

A. Traditional CV (Pre-Deep Learning)

Before deep learning, CV used mathematical techniques:

- Edge detection (Canny)
- Handcrafted features (SIFT, HOG)
- Contours
- Color histograms

Pros: Fast, simple
Cons: Fails on complex images

B. Deep Learning-Based Computer Vision (Modern CV)

Uses Neural Networks to automatically learn features.

1. CNNs (Convolutional Neural Networks)

Models:

- LeNet

- AlexNet
- VGG
- ResNet
- EfficientNet

These revolutionized image recognition.

---

2. Vision Transformers (ViT)

Transformer architecture for images.
New SOTA in many tasks.

Pros:

- Better global understanding
- Works well with large datasets

---

3. GANs (Generative Adversarial Networks)

Used for image generation:

- Super-resolution
- Face generation
- Style transfer

---

4. Diffusion Models (Latest)

Used for:

- DALL·E
- Stable Diffusion
- Midjourney

High-quality image generation.

---

7. Computer Vision Pipeline (Simplified)

1. **Input Image/Video**
2. **Preprocessing**
   - Resize
   - Normalize
   - Augment
3. **Feature Extraction** (CNN/ViT)
4. **Model Prediction**
5. **Post-Processing**
   - NMS for bounding boxes
   - Thresholding
6. **Final Result**
   - Labels, boxes, masks

---

8. Challenges in Computer Vision

1. Lighting Variations

Image quality changes with brightness.

2. Occlusion

Objects getting partially blocked.

3. Real-time processing

Self-driving cars require millisecond decisions.

4. Bias in training data

Unbalanced datasets cause bad predictions.

5. Complex backgrounds

Objects blend with surroundings.

6. Data labeling cost

Annotating images is expensive/time-consuming.

7. High compute requirements

Training CV models needs GPUs.

---

9. Popular CV Frameworks & Tools

Libraries

- OpenCV
- PyTorch
- TensorFlow
- Keras

Pretrained Models

- YOLOv8, YOLO-NAS
- ResNet
- MobileNet
- ViT
- Detectron2
- Mask R-CNN

Cloud APIs

- Google Vision

- AWS Rekognition
- Azure Computer Vision

---

10. Real Interview-Level Concepts

1. CNN Architecture Concepts

- Convolutions
- Pooling
- Feature maps
- Stride
- Padding

2. Loss Functions

- Cross-entropy
- IOU loss
- Dice loss

3. Data Augmentation

- Flip
- Rotate
- Brightness shift

4. Evaluation Metrics

- Accuracy
- mAP (mean Average Precision)
- IOU

- F1 Score

---

11. Assignments (Hands-on)

Assignment 1 — Basic Classification

Use any dataset (MNIST/CIFAR-10):

- Train a CNN
- Show accuracy
- Predict on sample images

Assignment 2 — Object Detection

Using YOLOv8:

- Run inference on any image
- Draw bounding boxes

Assignment 3 — OCR

- Extract text from an image using Tesseract
- Clean and format output

Assignment 4 — Theory Questions

Explain in 4–5 lines each:

1. What is Computer Vision?
2. Difference between object detection & classification
3. What is a CNN and why is it used?

4. What is segmentation?

5. What is the difference between CNNs and Vision Transformers?