

M. ASJAUN

1103210181

#### Member 4: Computer Vision Hugging Face Course (Chapters 1-4)

Link yt : <https://youtu.be/mOWp9bL4InY>

### 1. Vision dan Computer Vision

- **Pengertian Vision:**
    - Proses menangkap cahaya, mengubahnya menjadi sinyal listrik di mata, lalu merekonstruksi di otak untuk memahami lingkungan.
    - Penting untuk evolusi manusia dan perkembangan kognitif.
  - **Pendekatan Pemrograman Tradisional vs Machine Learning:**
    - Pemrograman berbasis aturan sulit menangkap konteks dan nuansa objek.
    - Machine Learning memungkinkan sistem belajar dari data untuk memahami pola.
- 

### 2. Gambar dalam Computer Vision

- **Definisi Gambar:**
    - Representasi visual dari objek, adegan, atau konsep.
    - Secara matematis digambarkan sebagai fungsi 2D:  $F(X,Y)$ .
  - **Jenis Gambar:**
    - 2D, 3D, dan gambar khusus seperti RGB, biner, atau 4D untuk biomedis.
  - **Representasi Gambar di Komputer:**
    - Matriks numerik 2D atau graf untuk pemrosesan data.
- 

### 3. Teknik dan Proses dalam Imaging

- **Akuisisi Gambar:**
    - Melibatkan sensor untuk menangkap energi (cahaya atau gelombang) dan mengubahnya menjadi sinyal digital.
  - **Teknologi Sensor:**
    - CCD, Strip Sensor, atau teknologi pemindaian 3D seperti MRI.
  - **Kompresi Gambar:**
    - Mengurangi ukuran data dengan metode seperti Huffman Coding dan Run-Length Encoding.
-

#### 4. Computer Vision: Definisi dan Evolusi

- **Definisi:**
    - Mengembangkan metode untuk mengakuisisi, memproses, dan memahami data visual.
  - **Evolusi:**
    - Dari metode berbasis aturan sederhana ke CNN dan deep learning modern.
  - **Aplikasi Utama:**
    - Scene recognition, object detection, image captioning, dan pelacakan objek.
- 

#### 5. Model dan Arsitektur dalam Computer Vision

- **GoogLeNet:**
    - Arsitektur dengan modul Inception untuk efisiensi parameter dan menangkap skala fitur.
  - **MobileNet:**
    - Menggunakan depthwise separable convolutions untuk efisiensi pada perangkat dengan sumber daya terbatas.
  - **Swin Transformer:**
    - Model berbasis hierarki patch untuk menangkap konteks global dan lokal.
- 

#### 6. Multimodal Models

- **Konsep Multimodalitas:**
  - Penggabungan berbagai modalitas (teks, gambar, suara) untuk memahami informasi lebih baik.
- **Model CLIP:**
  - Melatih teks dan gambar dalam ruang embedding yang sama melalui contrastive learning.
- **BLIP (Bootstrapping Language-Image Pre-training):**
  - Meningkatkan kinerja melalui penyaringan data gambar-teks (CapFilt).