Deep Learning -Computer Vision (By Ratnesh Kumar Singh)

This work is designed to equip you with a deep understanding of computer vision, by mastering image processing, neural networks, and advanced models, you will gain the ability to analyze, interpret, and manipulate visual data.

Objectives

- Develop a strong foundation in image processing and deep learning for computer vision.
- Understand and implement Convolutional Neural Networks (CNN) for image classification.
- Apply advanced object detection techniques like YOLO and Faster RCNN.
- Explore vision transformers and how attention mechanisms enhance computer vision tasks.
- Master segmentation methods for precise image analysis using models like Unet and DeepLab.
- Implement and evaluate real-time object tracking systems with algorithms like SORT and DeepSORT.

Work Information

The Computer Vision course is an advanced program and requires previous competence in the following areas:

Programming Fundamentals:
Proficiency in Python, especially with libraries like NumPy, Pandas, and OpenCV.

Basic Machine Learning Concepts:
Familiarity with supervised and unsupervised learning, and experience in training foundational machine learning models.

Neural Network Basics:
A foundational understanding of neural networks, covering topics like neurons, layers, activation functions, and backpropagation, will help you follow advanced concepts taught in the course.

Module 1 – Computer Vision Introduction

This module introduces the foundational concepts of computer vision and image processing. You'll explore the basics of deep learning and how it applies to image data, including the structure of images, color models, and key image transformations like scaling, cropping, and rotating. By the end of this module, you'll understand how to manipulate images and apply initial classification techniques, setting the stage for more advanced methods in later modules.

Topics

Foundations of Image Processing

- Understanding Pixels
- Image Types

- EXIF

Color Models

- Color Models
- Color Thresholding
- Image Classification

Image Manipulation and Transformation

- Image Scaling
- Aspect Ratio
- Crop
- Image Flip
- Rotate

Image Features

- Contours
- Contours Processing

Module 2 – DL: Computer Vision I

This module dives into the fundamentals of deep learning applied to computer vision, focusing on neural networks. You'll learn about the essential components of neural networks, including neurons, layers, activation functions, and backpropagation. The module also covers the basic techniques for image classification, helping you understand how deep learning can be used to recognize simple patterns in visual data through digit recognition tasks with a vanilla neural network.

Topics

Deep Learning Concepts

- Types of Learning
- Understanding Image Data
- Data Variation: occlusion, scale variation, illumination, noise, background & other

Neural Network Fundamentals

- Components of Neural Network

Core Mechanisms of Neural Networks

- Activation Function
- Loss Function
- Optimizer

- Forward Propagation
- Backpropagation
- Learning Rate

Hands-on Practice

- Digit Recognition with Vanilla Neural Network

Module 2 – DL: Computer Vision I

This module dives into the fundamentals of deep learning applied to computer vision, focusing on neural networks. You'll learn about the essential components of neural networks, including neurons, layers, activation functions, and backpropagation. The module also covers the basic techniques for image classification, helping you understand how deep learning can be used to recognize simple patterns in visual data through digit recognition tasks with a vanilla neural network.

Topics

Deep Learning Concepts

- Types of Learning
- Understanding Image Data
- Data Variation: occlusion, scale variation, illumination, noise, background & other

Neural Network Fundamentals

- Components of Neural Network

Core Mechanisms of Neural Networks

- Activation Function
- Loss Function
- Optimizer
- Forward Propagation
- Backpropagation
- Learning Rate

Hands-on Practice

- Digit Recognition with Vanilla Neural Network

Module 3 Part 1 – DL: Computer Vision II

In this module, you'll delve into Convolutional Neural Networks (CNN), the cornerstone of modern computer vision. You'll explore why CNNs outperform traditional neural networks in image tasks and learn about key components like filters, pooling, and dense layers. By the end, you'll have a solid understanding of CNN architecture and its applications, and you'll be able to implement your own CNN model for basic image recognition tasks using architectures like LeNet.

Topics

Introduction to CNNs

- Convolution Neural Network
- Why CNN is Better than ANN
- Components of CNN
- Input Data

Core CNN Operations

- Convolution Layer
- Convolve Function
- Filters (Kernels)
- Kernel Size
- Stride
- Padding
- Feature Map
- Channels

Activation and Pooling

- Activation Function
- Why to Use Activation Function
- Pooling Layer
- Max + AVG Pooling
- 1x1 Convolution

Network Architecture

- Flattening
- Fully Connected Layer (Dense Layer)
- Dropout
- Batch Normalization
- Softmax

Module 3 Part 2 – DL: Computer Vision II

Topics

Training Mechanisms

- Loss Function
- Optimizer

- Forward Propagation
- Backpropagation

Output and Predictions

- Output

Module 4 – DL: Computer Vision III

This module focuses on advanced CNN architectures that have shaped the field of deep learning in computer vision. You'll explore key models like AlexNet, VGGNet, GoogLeNet, ResNet, and MobileNet, which revolutionized the way visual tasks are tackled. Understanding these architectures will give you the knowledge to choose and implement state-of-the-art models for various image classification challenges, optimizing for accuracy, efficiency, and scale.

Topics

Early CNN Architectures

- LeNet
- AlexNet
- VGGNet

Advanced and Deeper Networks

- ResNet
- Inception-v3

Efficient and Mobile-Friendly Architectures

- DenseNet
- MobileNet

Module 5 – DL: Computer Vision IV – Computer Vision with Attention

This module introduces the concept of attention mechanisms in computer vision, particularly through Vision Transformers (ViT). You'll learn how transformers, originally designed for NLP, are transforming vision tasks by capturing relationships across an entire image. The module also covers ConvNeXt, a hybrid approach combining CNNs and transformers, providing you with cutting-edge techniques to enhance model performance in visual tasks.

Topics

Introduction to Transformers in Vision

- Why Use Transformers to Solve Vision Tasks
- Vision Transformers (ViT)

Input Processing and Representation

- Input Representation
- Positional Encoding
- Class Token

Core Transformer Mechanisms

- Multi-Head Self-Attention
- Feed-Forward Network
- Layer Normalization
- Residual Connections

Model Architecture

- Encoders
- Output Head

Module 6 Part 1 – DL: Computer Vision V – Object Detection

This module focuses on object detection, a key technique for identifying objects within an image. You'll explore both two-stage (e.g., Faster RCNN) and single-stage (e.g., YOLO) detectors, along with techniques like Region Proposal Networks (RPN) and anchor boxes.

Topics

Introduction to Object Detection

- What is Object Detection
- Classification
- Regression

Core Components of Object Detection

- Selective Search
- Region Proposal Network
- CNN – Feature Extractor
- Pre-trained Backbones
- Feature Pyramid Network
- RoI Pooling
- Anchor Boxes
- Bounding Box Regression & Classification Head

Object Detection Algorithms

- Two-Stage Detectors: RCNN, Fast RCNN, Faster RCNN

- Single-Stage Detectors: YOLO
- Object Detection using YOLOv5 & YOLOv11

Advanced Detection Methods

- Non-Maximum Suppression
- Advanced Loss Functions

Module 6 Part 2 – DL: Computer Vision V – Object Detection

Topics

- Hands-On Creating Our Own Object Detection Algorithm

Module 7 – DL: Computer Vision VI – Segmentation

This module explores segmentation, which focuses on classifying individual pixels in an image to distinguish objects and regions.

Topics

Introduction to Segmentation

- What is Segmentation
- Semantic Segmentation
- Instance Segmentation

Core Concepts in Segmentation

- Downsampling
- Upsampling / Transposed Convolution
- Skip Connections
- Atrous Convolutions
- Conditional Random Fields
- Loss Functions (Softmax with Cross-Entropy, Dice Loss)
- Evaluation Metrics

Popular Architectures & Framework

- Unet
- DeepLab v3
- MaskRCNN
- MMDetection

Practical Implementation

- Implementing Unet from Scratch
- Popular Datasets to Get Started

Module 8 – DL: Computer Vision VII – Object Tracking

This module covers object tracking, a dynamic task in computer vision that focuses on following objects as they move through video frames.

Topics

Introduction to Object Tracking

- What is Object Tracking

Tracking Methods

- Filter-Based Tracking
- CNN-Based Tracking

Key Algorithms in Object Tracking

- Kalman Filter
- SORT
- DeepSORT
- Re-ID

Hands-On

- Using YOLO and ByteSort to Track Objects

Module 9 – PRO Module: Generative Models for Vision Applications

Topics

Introduction to Generative AI

- Overview of Generative AI in Vision

Applications

- Image Synthesis
- Style Transfer
- Segmentation

Key Models and Techniques

- CLIP

- SAM
- Stable Diffusion
- CycleGAN

Hands-On

- Text-to-Image Generation for Creative Media
- Image Segmentation for Data Labeling
- Style Transfer and Domain Adaptation for Synthetic Data

This module dives into generative AI models specialized for vision applications, including cutting-edge tools such as **CLIP, SAM2, Stable Diffusion, and CycleGAN**. You'll explore how these models are used for tasks like text-to-image generation, segmentation, style transfer, and more. By mastering these techniques, you'll be able to create and manipulate visual content with precision and apply these skills in fields like digital art, automated content creation, and synthetic data generation for training other models.

>>

Tech Stack

◇ Core Computer Vision

computer-vision

image-processing

opencv

vision-ai

image-classification

◇ Deep Learning

deep-learning

neural-networks

cnn

convolutional-neural-networks

pytorch

◇ Object Detection

object-detection

yolo

yolov5

yolov8

faster-rcnn

mask-rcnn

◇ Segmentation

image-segmentation

semantic-segmentation

instance-segmentation

unet

deeplab

◇ Object Tracking

object-tracking

sort

deepsort

bytetrack

kalman-filter

◇ Transformers & Attention

vision-transformer

transformers

attention-mechanism

vit

◇ Generative AI (Vision)

generative-ai

diffusion-models

stable-diffusion

clip

segment-anything

sam

◇ Engineering / MLOps Ready

machine-learning

ai-engineering

ml-pipelines

model-training

model-evaluation

real-time-systems

>> Short Version >>>>>>>>>>>>>>>>>>>

🧠 Computer Vision & Vision-AI – End-to-End

From pixels to production — a complete Computer Vision & Vision-AI engineering repository

---

📌 Overview

This work is designed to equip you with a **deep understanding of Computer Vision**, covering **classical image processing, deep learning, advanced vision models, and Generative AI**.
By mastering these techniques, you will gain the ability to **analyze, interpret, and manipulate visual data** for real-world applications.

🎯 Objectives

- Build a strong foundation in **image processing & deep learning**
- Design and implement **Convolutional Neural Networks (CNNs)**
- Apply **advanced object detection** techniques (YOLO, Faster R-CNN)
- Understand **Vision Transformers (ViT)** and attention mechanisms
- Perform **semantic & instance segmentation** (UNet, DeepLab, Mask R-CNN)
- Develop **real-time object tracking systems** (SORT, DeepSORT, ByteTrack)
- Work with **Generative AI models for vision**

🧩 Work Information & Prerequisites

This is an **advanced-level Computer Vision program** and assumes prior knowledge of:

🔹 **Programming Fundamentals**

- Python
- NumPy, Pandas
- OpenCV

🔹 **Machine Learning Basics**

- Supervised & Unsupervised Learning
- Model training & evaluation

🔹 **Neural Network Foundations**

- Neurons & Layers
- Activation Functions
- Loss Functions
- Backpropagation

🧩 Breakdown

◇ Module 1 – Computer Vision Introduction

Foundations of image processing & visual data manipulation

Topics

- Pixels, Image Types, EXIF
- Color Models & Thresholding
- Image Scaling, Cropping, Rotation, Flipping
- Feature Extraction using Contours

◇ Module 2 – DL: Computer Vision I

Neural Networks for Vision

Topics

- Types of Learning
- Image Data Variations
- Neural Network Components
- Activation, Loss, Optimizers
- Forward & Backpropagation

Hands-On

- Digit Recognition using a Vanilla Neural Network

◇ Module 3 – DL: Computer Vision II (CNNs)

Part 1 – CNN Fundamentals

- CNN vs ANN
- Convolution, Filters, Stride, Padding
- Pooling (Max & Avg)
- Dropout, BatchNorm, Softmax
- LeNet-style Architectures

Part 2 – Training & Inference

- Loss Functions
- Optimizers
- Model Outputs & Predictions

◇ Module 4 – DL: Computer Vision III

Advanced CNN Architectures

- LeNet
- AlexNet
- VGGNet
- ResNet
- Inception-v3
- DenseNet
- MobileNet

◇ Module 5 – Computer Vision with Attention

Transformers for Vision

- Vision Transformers (ViT)
- Positional Encoding & Class Tokens
- Multi-Head Self-Attention
- Feed-Forward Networks
- Residual Connections
- Encoder Architectures

◇ Module 6 – DL: Computer Vision V – Object Detection

Part 1 – Detection Theory

- Object Detection Fundamentals
- Two-Stage Detectors: RCNN, Fast RCNN, Faster RCNN
- Single-Stage Detectors: YOLO
- Anchor Boxes & RPN
- Feature Pyramid Networks
- Non-Maximum Suppression

Part 2 – Hands-On

- Build a Custom Object Detection Algorithm from Scratch

◇ Module 7 – DL: Computer Vision VI – Segmentation

- Semantic vs Instance Segmentation
- UNet, DeepLab v3, Mask R-CNN
- Upsampling & Skip Connections
- Dice Loss & Cross-Entropy
- Popular Datasets & Evaluation Metrics

◇ Module 8 – DL: Computer Vision VII – Object Tracking

- Filter-Based vs CNN-Based Tracking
- Kalman Filter
- SORT & DeepSORT
- Re-Identification (Re-ID)

Hands-On

- Object Tracking using YOLO + ByteTrack

◇ Module 9 – PRO Module: Generative Models for Vision

Generative AI for Vision Applications

- Image Synthesis
- Style Transfer
- Segmentation

Key Models

- CLIP

- SAM / SAM2
- Stable Diffusion
- CycleGAN

Hands-On

- Text-to-Image Generation
- Image Segmentation for Data Labeling
- Style Transfer & Domain Adaptation

🛠 Tech Stack

◇ Core Computer Vision

computer-vision · image-processing · opencv · vision-ai · image-classification

◇ Deep Learning

deep-learning · neural-networks · cnn · pytorch

◇ Object Detection

object-detection · yolo · faster-rcnn · mask-rcnn

◇ Segmentation

image-segmentation · semantic-segmentation · instance-segmentation · unet · deeplab

◇ Object Tracking

object-tracking · sort · deepsort · bytetrack · kalman-filter

◇ Transformers & Attention

vision-transformer · transformers · attention-mechanism · vit

◇ Generative AI (Vision)

generative-ai · diffusion-models · stable-diffusion · clip · sam

◇ Engineering / MLOps

machine-learning · ai-engineering · ml-pipelines · real-time-systems

---