CSE471:DEEP LEARNING FOR COMPUTER VISION

L:2 T:0 P:2 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: remember the fundamental concepts of computer vision, image representation, preprocessing operations, and neural network basics used in deep learning.

CO2 :: understand the working principles of convolutional neural networks, feature extraction, pooling, regularization, and CNN training techniques.

CO3 :: apply transfer learning and fine-tuning methods using pre-trained models for domain-specific computer vision tasks.

CO4 :: apply object detection and localization techniques, including bounding box representation, IoU, NMS, and modern detection architectures.

CO5 :: apply image segmentation methods and transformer-based vision architectures for extracting meaningful visual information from images.

CO6 :: understand the principles of generative vision models, vision-language models, and interpretability techniques used in advanced computer vision applications.

Unit I

Foundations of Computer Vision and Deep Learning : Role of computer vision in modern AI applications, Traditional versus deep-learning-based vision approaches, image fundamentals, pixels, channels, color spaces, resizing, cropping and normalization, neural network basics, perceptron, feedforward networks, activation functions, loss functions, optimization methods

Unit II

Convolutional Neural Networks and Training Techniques : CNN Basics: convolutional neural networks, convolution operation, filters, feature maps, padding, stride, pooling, and receptive fields, Modern CNN architectures: LeNet, AlexNet, VGG, ResNet, and Inception networks, regularization techniques, batch normalization and dropout, data augmentation methods, learning rate scheduling, weight initialization and early stopping

Unit III

Transfer Learning and Fine-Tuning for Vision Tasks : Pre-trained architecture: transfer learning, ResNet, EfficientNet, MobileNet, feature extraction, fine-tuning, managing freezing and unfreezing of layers, applying layer-wise learning rates, domain adaptation and dataset

Unit IV

Object Detection and Localization : object detection: object detection principles ,algorithms, bounding box representation, intersection over union, non-maximum suppression, classical detection pipelines, RCNN, single-stage detectors, YOLO variants, SSD, RetinaNet with focal loss, Evaluation metrics: mean average precision, IoU thresholds

Unit V

Image Segmentation and Advanced Vision Architectures : Image Segmentation: semantic segmentation, instance segmentation, UNet, DeepLab architectures, instance segmentation using Mask R-CNN, Vision transformers: self-attention, multi-head attention, positional encoding, transformer-based models, ViT, Swin Transformer

Unit VI

Generative Vision Models : Generative deep learning: variational autoencoders, GAN architectures, DCGAN, CycleGAN, StyleGAN, image generation, image-to-image translation, super-resolution, vision-language models: CLIP for image-text alignment, interpretability techniques, Grad-CAM, saliency maps

List of Practicals / Experiments:

List of Practical's

- 1.Load an image using OpenCV, perform resizing, grayscale conversion, normalization, and visualize the results using Matplotlib in multiple subplots.
- 2.Implement a 2D convolution operation using NumPy, apply it to a grayscale image patch with different filters, and visualize the resulting feature maps.
- 3.Implement a 2D convolution operation using NumPy, apply it to a grayscale image patch with different filters, and visualize the resulting feature maps.

- 4.Train LeNet and ResNet-18 on MNIST and compare training time, accuracy, and model size to understand how architectural depth and skip connections influence performance.
- 5.Use a pretrained to extract image features and train a classifier on a new dataset such as CIFAR-10 or a small custom dataset (e.g., flower images). Evaluate the classifier's accuracy.
- 6.Fine-tune a pretrained MobileNet model on a small dataset of Cats vs Dogs. Experiment with freezing/unfreezing layers and observe the effect on training time and accuracy.
- 7.Use a pretrained YOLO model to detect objects in images or webcam/video streams, draw bounding boxes and class labels, and analyze detection confidence and performance
- 8.Fine-tune a pretrained Faster R-CNN model on a small custom dataset . Train for a few epochs and evaluate performance.
- 9.Use a pretrained UNet model to generate pixel-wise segmentation masks for sample images, overlay the predicted mask on the original image, and analyze the model’s ability to segment objects
- 10.Use a pretrained CNN model to classify images, then apply Grad-CAM to visualize which parts of the image contributed most to the prediction. Compare Grad-CAM outputs for correctly classified and misclassified samples.

Text Books:

1. FUNDAMENTALS OF DEEP LEARNING & COMPUTER VISION by NIKHIL SINGH, PARAS AHUJA, BPB PUBLICATIONS

References:

1. DEEP LEARNING by AMIT KUMAR DAS, SAPTRISHI GOSWAMI, PABITRA MITTRA, AMLAN CHAKRABARTI, Pearson Education India

2. ELEMENTS OF DEEP LEARNING FOR COMPUTER VISION by BHARAT SIKKA, BPB PUBLICATIONS

3. DEEP LEARNING FOR COMPUTER VISION: EXPERT TECHNIQUES TO TRAIN ADVANCED NEURAL NETWORKS USING TENSORFLOW AND KERAS by RAJALINGAPPAA SHANMUGAMANI, PACKT PUBLISHING