Pre-trained Models in Computer Vision

Introduction

Pre-trained models are widely used in computer vision tasks such as **image classification, object detection, and segmentation**. These models are trained on large datasets (like **ImageNet** or **COCO**) and can be reused to solve new problems with minimal additional training.

They are categorized based on their **application type**, such as:

- Image Classification
- Object Detection
- Object Segmentation
- Image Segmentation

🔷 1. Pre-trained Models for Image Classification

These models assign a label (category) to an entire image.

- **VGG16 / VGG19** – simple CNN-based models trained on ImageNet.
- **ResNet (ResNet50, ResNet101, ResNet152)** – deeper CNNs with skip connections.
- **Inception (GoogLeNet, InceptionV3)** – efficient architecture using multi-scale filters.
- **DenseNet** – connects each layer to every other layer for feature reuse.
- **EfficientNet** – balances depth, width, and resolution for efficient training.
- **Vision Transformers (ViT, Swin Transformer)** – transformer-based models for classification tasks.

👉 **Use case:** Cat vs Dog classifier, medical image classification (X-ray normal vs abnormal).

🔷 2. Pre-trained Models for Object Detection

These models locate and classify **multiple objects** in an image with bounding boxes.

- **YOLO (You Only Look Once: v3, v4, v5, v7, v8)** – fast and real-time object detection.
- **SSD (Single Shot Detector)** – efficient one-stage detector.
- **Faster R-CNN** – region proposal network for accurate object detection.
- **RetinaNet** – balances speed and accuracy, uses focal loss for handling class imbalance.
- **DETR (DEtection TRansformer)** – transformer-based end-to-end detection model.

👉 **Use case:** Self-driving cars detecting pedestrians, traffic lights, and vehicles.

🔷 3. Pre-trained Models for Image Segmentation

These models classify **each pixel** of an image into a category.

- **FCN (Fully Convolutional Network)** – first CNN-based segmentation approach.
- **U-Net** – encoder-decoder model, widely used in medical imaging.
- **SegNet** – another encoder-decoder network for pixel-wise segmentation.
- **DeepLab (V2, V3, V3+)** – adds atrous (dilated) convolutions for better segmentation.
- **PSPNet (Pyramid Scene Parsing Network)** – captures context at multiple scales.

👉 **Use case:** Medical image segmentation (tumor detection), satellite image segmentation.

🔷 4. Pre-trained Models for Instance / Object Segmentation

These models detect **each object instance separately** and segment them at pixel level.

- **Mask R-CNN** – extension of Faster R-CNN for pixel-level object segmentation.
- **YOLACT** – real-time instance segmentation.
- **BlendMask** – combines object detection with segmentation.
- **SOLO / SOLOv2 (Segmenting Objects by Locations)** – instance segmentation without anchors.

👉 **Use case:** Identifying and separating multiple overlapping objects (e.g., detecting people in a crowd).

🔷 5. Specialized Vision Models

- **CLIP (OpenAI)** – connects vision and language, useful for zero-shot classification.
- **SAM (Segment Anything Model, Meta)** – general-purpose segmentation model.
- **DINO / DINOv2** – self-supervised vision transformers.

Sources of Pre-trained Models

- **TensorFlow Hub (tfhub.dev)** – image classification, detection, segmentation models.
- **PyTorch Hub (pytorch.org/hub)** – ResNet, YOLO, Faster R-CNN, etc.
- **Hugging Face (huggingface.co/models)** – hosts vision transformers, SAM, CLIP.
- **Model Zoos (Torchvision, Detectron2, MMDetection)** – large repositories of computer vision models.

Conclusion

Pre-trained models are crucial in computer vision, enabling rapid development for tasks like **image classification, object detection, image segmentation, and object/instance segmentation**.
They are widely used in fields such as healthcare, autonomous vehicles, security, and remote sensing.