Convolutional Neural Networks

Computer Vision: allows machines to interpret images or videos like humans, often involving matrices as a digital image is represented as a grid of pixels each with intensity values.

Edge Detection: basically what it says it references the boundaries or outer layer of objects in an image. There are different examples such as Canny Edge Detection which uses Gaussian Blur to remove noise, Sobel Operator to find gradients, Non-Maximum Suppression to keep only the strongest edges, and Hysteresis Thresholding to decide which edges to keep.

Padding: Adding extra pixels around an image to preserve the actual size of the image after convolution.

Strided Convolutions: moving the filter by more than one pixel at a time to reduce computation is faster but less detailed. Convolution over volume: is done over multi-channel images (RGB) convolution is applied over every single part separately. One Layer of a Convolutional Network: A single convolutional layer applies filters and activation functions to detect patterns think of it like a first glance to get a rough idea of the object(s) in the image.

Pooling Layers: reducing the size of an image while keeping important features.

Case Studies

Classic Networks: these are early CNN architectures that laid the foundation for modern deep learning models.

ResNets: Residual Networks (ResNets) introduce skip connections to solve the vanishing gradient problem in deep neural networks they work by allowing information to bypass certain layers, improving gradient flow and making deep networks trainable.

Networks in Networks and 1x1 Convolutions: These techniques enhance feature extraction and reduce computational cost by applying 1x1 convolutions within CNNs.

Inception Network: A CNN architecture that uses parallel convolutional layers with different kernel sizes to capture diverse spatial features. Inception Networks were designed to improve computational efficiency by using multiple filter sizes at each layer.

MobileNet: A lightweight CNN optimized for mobile and embedded vision applications, using depthwise separable convolutions. The architecture of MobileNet relies on depthwise separable convolutions to reduce computational complexity while maintaining accuracy.

EfficientNet: A family of CNNs that scale depth, width, and resolution systematically for optimal performance with minimal computational resources.

Practical Advice for Using ConvNets

Using Open-Source Implementation: Leveraging pre-trained CNN models and frameworks to speed up development and improve accuracy.

Transfer Learning: A technique where a pre-trained model is fine-tuned on a new dataset, reducing the need for extensive training data.

Data Augmentation: Enhancing training datasets by applying transformations such as rotation, flipping, and color changes to improve generalization.

Detection Algorithms

Object Localization: Identifying and pinpointing the location of an object in an image or video.

Landmark Detection: Detecting key points or landmarks in an image, often used in facial recognition and medical imaging.

Object Detection: Identifying objects in an image and classifying them while determining their locations.

Convolutional Implementation of Sliding Windows: Applying a sliding window approach using convolutional layers to efficiently detect objects.

Bounding Box Predictions: Predicting the coordinates of a rectangle that encloses the detected object.

Intersection Over Union: A metric used to evaluate object detection accuracy by comparing predicted and actual bounding boxes.

Non-max Suppression: A technique to remove duplicate bounding boxes by keeping only the most relevant ones.

Anchor Boxes: Predefined bounding boxes used to detect objects of different sizes and aspect ratios in object detection models.

YOLO Algorithm: "You Only Look Once" is a real-time object detection algorithm that divides an image into grids and predicts bounding boxes and class probabilities simultaneously.

Semantic Segmentation with U-Net: Assigning a class label to each pixel in an image, often used in medical and satellite imaging.

Transpose Convolutions: A type of convolution used to upsample feature maps, often in segmentation tasks.

U-Net Architecture Intuition: Understanding the U-Net model, which consists of an encoder-decoder structure for image segmentation.

U-Net Architecture: A CNN architecture designed for biomedical image segmentation, featuring a U-shaped structure with skip connections to preserve spatial information.

Face Recognition

What is Face Recognition?: Identifying or verifying individuals based on their facial features using deep learning techniques.

One Shot Learning: A machine learning technique where a model learns to recognize new classes with only one or a few training examples.

Siamese Network: A neural network architecture designed for comparing and learning similarity between two inputs, often used in face verification.

Triplet Loss: A loss function used to train models for facial recognition by minimizing the distance between an anchor and a positive example while maximizing the distance from a negative example.

Face Verification and Binary Classification: A task where a model determines whether two given face images belong to the same person, often framed as a binary classification problem.

Neural Style Transfer

What is Neural Style Transfer?: A deep learning technique that applies the artistic style of one image to the content of another using convolutional neural networks.

What are Deep ConvNets Learning?: Understanding the hierarchical features learned by deep convolutional networks in tasks like image classification and style transfer.

Cost Function: A function that quantifies the difference between the predicted output and the actual target, guiding the optimization process in neural networks.

Content Cost Function: A component of neural style transfer that ensures the content of the generated image closely resembles the input content image.

Style Cost Function: A component of neural style transfer that ensures the texture and patterns of the generated image match the artistic style of the reference image.

1D and 3D Generalizations: Extending convolutional neural networks (CNNs) to one-dimensional (e.g., speech processing) and three-dimensional (e.g., volumetric medical imaging) data.