DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SUBJECT CODE: 23CSX502 **SEM :VI**

SUBJECT NAME: DEEP LEARNING AND NEURAL NETWORKS **YEAR: III**

UNIT II- CONVOLUTION NEURAL NETWORKS

FUNDAMENTALS

1. Foundations of Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep learning models designed specifically for processing visual data such as images and videos. CNNs are inspired by the human visual system, where neurons respond to local regions of the visual field and progressively build complex representations.

Digital images are represented as matrices of pixel values. A grayscale image is a 2D matrix, while a color image is a 3D volume consisting of Red, Green, and Blue (RGB) channels. Neighbouring pixels are spatially correlated, and CNNs exploit this property to efficiently learn meaningful visual patterns.

CNNs automatically learn features from raw images without manual feature engineering, making them powerful and scalable for real-world vision tasks.

2. Why CNNs are Needed & Limitations of Traditional ANNs

Limitations of Traditional Artificial Neural Networks (ANNs)

- **High Dimensionality**: Images contain thousands or millions of pixels, leading to a massive number of parameters in fully connected networks.
- **Loss of Spatial Information**: ANNs treat all pixels independently and ignore spatial relationships.
- **Overfitting**: Too many parameters cause the network to memorize training data.
- **Manual Feature Extraction**: Traditional systems rely on handcrafted features.
- **Poor Scalability**: Fully connected networks fail on large, high-resolution images.

Why CNNs are Effective

CNNs overcome these issues using three key ideas:

- **Local Receptive Fields**: Neurons connect to small local regions instead of the full image.
- **Weight Sharing**: The same filter is applied across the image to detect features anywhere.
- **Pooling**: Reduces feature map size and improves robustness.

3. Core Concepts of CNNs

3.1 Hierarchical Feature Learning

CNNs learn features in a hierarchical manner:

- **Early Layers**: Detect low-level features like edges, lines, and corners.

- **Middle Layers**: Learn shapes, textures, and object parts.
- **Deep Layers**: Recognize complete objects and semantic meaning.

This layered learning enables CNNs to understand complex visual patterns effectively.

3.2 Translation Equivariance and Invariance

- **Translation Equivariance**: Convolution layers ensure that when an object shifts, its feature map shifts accordingly.
- **Translation Invariance**: Pooling and deeper layers ensure that object recognition is independent of location.

Pooling layers and global average pooling help achieve approximate translation invariance.

4. CNN Architecture Overview

A CNN architecture is composed of multiple layers arranged sequentially to transform raw image pixels into class predictions.

Key Components of CNN Architecture

- **Convolution Layers**: Extract local features using filters.
- **Activation Layers**: Introduce non-linearity (ReLU commonly used).
- **Pooling Layers**: Reduce spatial dimensions and computation.
- **Fully Connected Layers**: Combine learned features.
- **Output Layer**: Produces final predictions using Softmax or Sigmoid.

CNNs progressively learn from simple patterns to complex representations.

5. Five Core Layers in CNN

5.1 Convolution Layer

- Applies filters over the input image
- Produces feature maps highlighting patterns
- Learns edges, textures, and shapes

5.2 Activation Layer

- Adds non-linearity to the model
- ReLU is most common: $f(x) = \max(0, x)$
- Enables learning of complex patterns

5.3 Pooling Layer

- Reduces feature map size
- Types: Max Pooling, Average Pooling
- Improves efficiency and generalization

5.4 Fully Connected Layer

- Flattens feature maps into vectors
- Learns high-level relationships

- Acts as decision-making layer

5.5 Output Layer

- Uses Softmax (multi-class) or Sigmoid (binary)
- Produces probability scores
- Final prediction is made here

6. Simple Convolutional Neural Network

A simple CNN consists of repeated blocks of:

Input → Convolution → ReLU → Pooling → Fully Connected → Output

The CNN reduces image complexity while preserving important features, enabling efficient classification.

7. Feature Extraction and Pooling

CNNs automatically extract features at different levels:

- Edges and gradients
- Shapes and textures
- Complete objects

Pooling layers reduce overfitting, computation, and sensitivity to small shifts in images.

8. Deep Convolutional Neural Networks (Deep CNNs)

Deep CNNs consist of many stacked convolutional layers. Each stage contains:

- Input volume
- Convolution + Bias
- Activation
- Feature maps
- Optional pooling

Examples include VGG-16, which uses repeated Conv + ReLU + Pooling blocks.

Volume convolution uses 3D kernels that operate across depth and spatial dimensions.

9. Popular CNN Architectures

9.1 AlexNet

- Introduced in 2012
- 5 convolution layers + 3 fully connected layers
- Uses ReLU, Dropout, GPU acceleration
- Won ImageNet 2012

Key Innovations:

- ReLU activation
- Dropout regularization
- Data augmentation

- Overlapping pooling

9.2 GoogLeNet (Inception V1)

- Introduced Inception modules
- Parallel 1×1, 3×3, 5×5 convolutions
- Uses global average pooling
- Very parameter efficient

Key Features:

- 1×1 convolutions for dimensionality reduction
- Auxiliary classifiers
- 22 layers deep

9.3 ResNet

- Introduced residual (skip) connections
- Solves vanishing gradient and degradation problems
- Enables very deep networks (50–152 layers)

Residual learning: $F(x) + x$

10. Advantages and Disadvantages of CNNs

Advantages

- Automatic feature learning
- Fewer parameters due to weight sharing
- High accuracy in vision tasks
- Robust to translation and noise

Disadvantages

- High computational cost
- Requires large labeled datasets
- Hard to interpret
- Risk of overfitting on small data

11. Applications of CNNs

- Image classification
- Object detection
- Medical imaging
- Face recognition
- Autonomous vehicles
- OCR and text extraction
- Quality inspection