Comprehensive Overview of CNN Concepts

1. CNN Workflow

The CNN workflow typically involves the following steps:

- Input Image: The raw input data, such as an image.
- Convolution Layer: Extracts features from the input using filters.
- Pooling Layer: Reduces spatial dimensions to simplify the feature maps.
- Fully Connected Layer: Performs classification or regression tasks using the extracted features.

These layers work together to learn both low-level and high-level features for tasks like image classification.

2. Convolution Layer

- Filters perform convolution operations, sliding over the input to extract features.
- Key hyperparameters include:
  - Filter size (F): Determines the spatial dimensions of the filter.
  - Stride (S): Defines the step size of the filter as it slides over the input.
  - Padding (P): Adds zeros around the input to control output dimensions.
- The output feature map size is computed as:

  $O = (I - F + 2P) / S + 1$

3. Pooling Layer

- Reduces spatial dimensions of feature maps while retaining important information.
- Common types:
  - Max Pooling: Takes the maximum value within a region.

  - Average Pooling: Computes the average value within a region.

- Helps make the network invariant to minor translations in the input.

4. Fully Connected Layer

- Operates on flattened inputs where every input connects to every neuron.

- Typically found towards the end of the network, used for classification tasks.

- Number of parameters:

  (N_in + 1) * N_out

  - N_in: Number of input neurons.

  - N_out: Number of output neurons.

5. Filter Hyperparameters

- A filter of size F x F x C extracts features from inputs with C channels.

- Using K filters results in an output feature map size of O x O x K.

- Hyperparameters allow tuning of model complexity and feature extraction.

6. Zero-Padding

- Adds zeros around the input to control the size of the output feature map.

- Modes:

  - Valid: No padding, output size shrinks.

  - Same: Pads to keep output size the same as input.

  - Full: Maximum padding for complete input coverage.

7. Receptive Field

- Represents the region of the input that affects a particular neuron in the feature map.

- Formula:

  $R_k = 1 + \sum(F_j - 1) * \prod(S_i)$

  - F_j: Filter size at layer j.

  - S_i: Stride at layer i.

- The receptive field grows with depth in the network.

8. Activation Functions

Common activation functions:

- ReLU: Introduces non-linearity, outputs max(0, z).

- Leaky ReLU: Allows small gradients for negative values.

- ELU: Smooth activation in the negative region.

- Softmax: Converts raw scores into probabilities for classification.

Summary

This document provides an overview of CNNs, including layers, hyperparameters, activation functions, and concepts like receptive fields. These components work together to enable deep learning models to perform complex tasks such as image classification, object detection, and more.