3. Deep Learning Categories

In this section below, we will discuss multiple deep learning architectures and explain their underlying algorithms. An up-to-date overview will be presented for each of the three main categories of neural networks, namely, Convolutional Neural Networks, Pretrained Unspervised Networks, and Recurrent/Recursive Neural Networks.

3.1 Convolutional Neural Networks (CNNs)

CNNs are inspired by biological processes and are designed to mimic the neural connectivity found in the brain's visual cortex. They requires considerably less data pre-processing compared to traditional image classification algorithms which require hand-engineered pre-processing filters [6]. CNNs have a large range of applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing (NLP).

3.1.1 CNN Structure

CNNs differ from conventional neural networks as they perform convolution instead of standard matrix multiplication in at least one of their layers (Fig. 1). They are famous for two distinct attributes: sparse interactions and parameter sharing. Sparse interactions or connectivity is achieved by making the model's kernel smaller than the size of the input. For example, in an image classification application, there may be millions of pixels representing a high resolution image. In this case, the kernel will be configured in a manner such that it only captures important features such as contrast and edges which are more indicative of the objects in the image. With fewer pixels of the image in consideration, there is a reduction in parameters to represent the image. This results in the reduction of memory utilization as well as computational overhead.