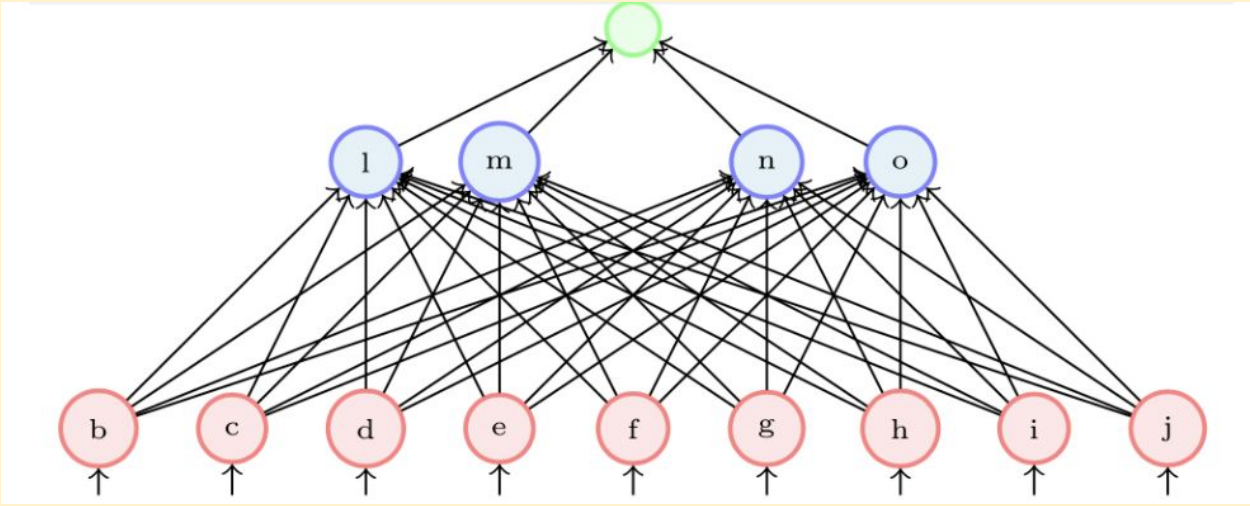
PadhAI: From Convolution Operation to Neural Network

One Fourth Labs

Training CNNs

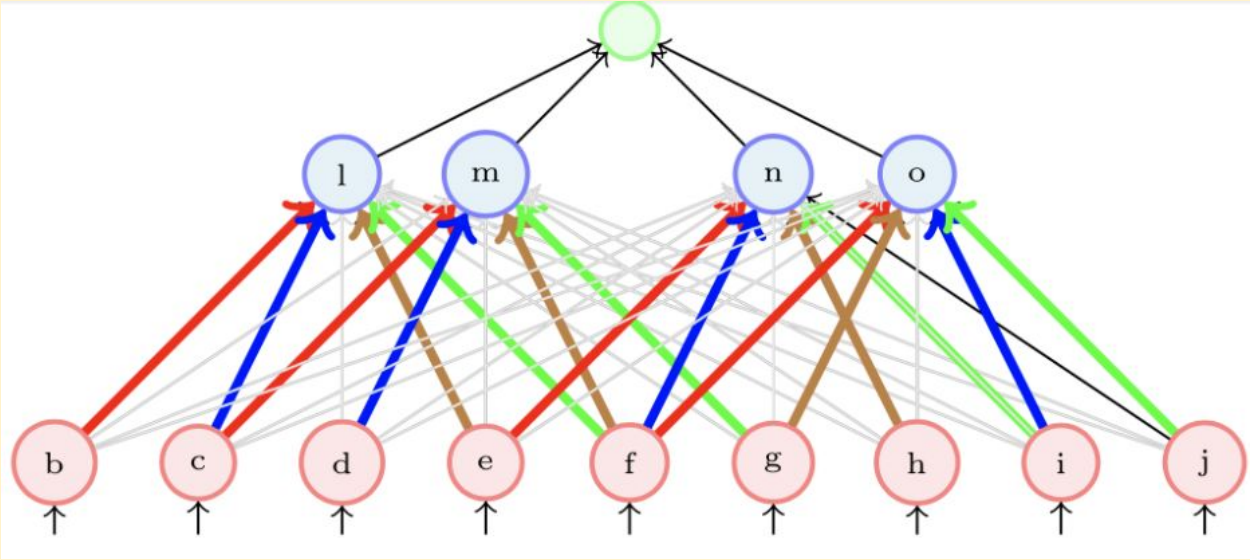
How do we train a Convolutional Neural Network?

1. Let us consider a regular Feedforward Network



   a. Here, we obtain ŷ as a prediction, and use it to calculate the loss.
   b. Using the loss value, we backpropagate to calculate the gradients w.r.t each of the parameters
   c. Using the gradient descent update rule, we update the weights such that they minimize the loss.

2. Now, we can carry over all of these processes to a CNN, with a few small modifications



   a. A CNN can be implemented as a feedforward network wherein only a few weights (coloured) are active
   b. The rest of the weights (grey) remain zero
   c. Thus, we can train a CNN using backpropagation by thinking of it as a FFN with sparse connections

3. However, in practice we don't do this, as most of the weights in the matrix end up being zero. Frameworks like PyTorch and Tensorflow don't end up creating such large matrices and only focus on dealing with the weights that are to be updated.