1. CNNs are the most widely used models in Deep Neural Networks (DNNs) [1]. They have demonstrated excellent performance in several fields, such as Image Recognition [2], Semantic Segmentation [3], Speech Recognition [4], Object Detection [5] and Natural Language Processing [6]. Since CNN models require great amount of data and computational skills, their success was attributed to dramatic growth in the amount of data as well as an increase in the computational speed [7], [8], [9]. Based on these factors, researchers and practitioners were able to train deeper networks ([10], [11], [12]), which resulted in a higher increase in the number of the CNN models.

2. CNN comprises forward and backward passes. Both of them greatly depend on matrix multiplication. In the forward pass, each hidden layer executes multiplication between a specific part of the input and kernel weight matrix. It creates feature maps, which act as inputs for the next layer. Regarding back-propagation [16], the same matrix multiplication operations are executed in the opposite direction. A typical sort of CNN has a large number of layers, thus there are numerous matrix multiplication operations in it. The only available parameter that requires modification is the weight matrix.

3. Convolutional Neural Networks (CNNs) have made a great impact on attaining state-of-the-art results in image task classification. Weight initialization is one of the fundamental steps in formulating a CNN model. It determines the failure or success of the CNN model. In this paper, we conduct a research based on the mathematical background of different weight initialization strategies to determine the one with better performance. To have smooth training, we expect the activation of each layer of the CNN model follow the standard normal distribution with mean 0 and standard deviation 1. It prevents gradients from vanishing and leads to more smooth training.