4.1 Dataset

The Dataset used in this project is a hybrid dataset. Initially, it was a CBIR Dataset (Content Based Image Retrieval Dataset) from Kaggle. But later, to increase the size of the dataset and to diversify the dataset, we have used some multiple object datasets and merged all of them to build a huge workable dataset. After all the pre-processing, our hybrid image dataset is best fit for our project evaluation because of its large sample size, inferior data quality (when compared to other available datasets which suit our purpose since it helps us better generalize) and large variance; and also, because the output classes are balanced.









Fig 1: Sample Images of the Dataset

After discarding mistakes and scribbles, 4,141 images of different categories were included in the final dataset. The dataset consisted of 25+ categories of dataset, which consisted of different animals, fruits, cartoons and different regularly seen daily life objects were collected, digitized and pre-processed before running the feature extraction code. So that all the images are successfully pre-processed for smooth running of our particular algorithm. These images were pre-processed by applying colour-inversion, noise removal and edge-thickening filters. All the images of dataset were resized into 224 x 224px to fit them VGG16 pre-trained model.