Many recent methods self-supervised representation learning train feature extractors by maximizing an estimate of the mutual information (MI) between different views of the data. In this paper they discuss and provide evidence to analyze the extent to which maximizing Mutual Information (MI) between vectors has an effect on training feature extractors. They take 2 images, pass it through particular feature extractors and get 2 independent 2D-vectors for each of the images respectively. Following this MI is calculated using a predefined formula between these 2 vectors, using one among a variety of estimators (InfoNCE, InfoMAX, etc). The primary task is to maximize this MI. After noting down initial results they discuss if employing a better MI formula is the only way in which we could get better feature extractors. The authors go on to prove that the success of these methods cannot be attributed to the properties of MI alone, and that they strongly depend on the inductive bias in both the choice of feature extractor architectures, the parametrization of the employed MI estimators, the estimators and critics employed. This paper is to prove that maximization of mutual information, and focusing on improving solely that is not a necessary and/or sufficient condition for beating current SOTA results. It is further suggested that the use of triplet technique, in which we use representation of images, could help reduce the MI factor between similar images and increase it for distinct ones. It also suggests that while all of this is true we must continue to pursue better formulae to find MI as although it isn’t solely responsible for the performance of our experiment but it does play a vital role in it.