

We introduce a self-supervised representation learning method based on the task of temporal alignment between videos. The method trains a network using ten