NAME : JAITHRA VARMA
SEC : A
REG.NO : 200968254

REPORT -2

I will be using spatio temporal models for capturing both spatial and temporal information in the videos.

Models I will be using:

1) The first model I will be using is I3D. I3D proposes to take advantage of successful architectures such as Inception architecture and ResNet architectures to create Spatio-temporal models by transforming them into 3D CNNs. 2D filters and pooling kernels are converted into 3D filters and pooling kernels. Although the model works fine. One of the disadvantages I observed is that we have to choose temporal stride carefully. So we are basically doing the tradeoff between how well we want to capture the temporal information and how well we are capturing spatial information.

2) The second model I will be using is SlowFast Network. This overcomes the disadvantages of I3D and is expected to work better. In SlowFast Network we have two pathways. One is the Slow pathway which has a larger temporal stride The primary focus of the slow pathway is to capture finer spatial semantics such as the texture, colours, and edges. The other pathway is the fast pathway which have a lower temporal stride and expected to capture the temporal information better.

3) The third model I will be using is P3D. The P3D architecture proposes residual bottleneck blocks that decompose 3×3×3 filter size convolutions to combinations of 1×1×3 temporal and 3×3×1 spatial convolutions. This reduces the number of parameters compared to 3D CNNs, making them easier to train on small scale datasets. The P3D paper also presents a family of computationally feasible building blocks to perform 3D convolutions efficiently.