There are three sets of results for comparison: 1.) single-frame model, 2.) 3D model, 3.) combined output of the two models. For each of these three, report the following:

- **(top1_accuracy,top5_accuracy,top10_accuracy): Did the results improve after combining the outputs?**

Number of epochs: 10

	Top_1 Accuracy (avg)	Top_5 Accuracy (avg)	Top_10 Accuracy (avg)
Single Frame	0.7634	0.9468	0.9726
Sequence Model	0.8069	0.9543	0.9812
Combined Model	0.8160	0.9652	0.9899

The results improved after combining the two models and the combined model seems to give the best accuracy.

Single Frame Model:

Training Accuracy : 96.9042553191489

Test Accuracy: 72.89189189189189

Sequence Model:

Training Accuracy : 98.9618288590604

Test Accuracy: 80.61440677966101

- **Use the confusion matrices to get the 10 classes with the highest performance and the 10 classes with the lowest performance: Are there differences/similarities? Can anything be said about whether particular action classes are discriminated more by spatial information versus temporal information?**

Single Frame Model:

High Performing Classes:

SoccerPenalty 1.0 41.0
Skijet 1.0 28.0
BabyCrawling 1.0 35.0
Rowing 1.0 36.0
PlayingTabla 1.0 31.0
BreastStroke 1.0 28.0
BasketballDunk 1.0 37.0
Billiards 1.0 40.0
TableTennisShot 1.0 39.0
PlayingFlute 1.0 48.0

Low Performing Classes:

JumpRope 0.02631579 38.0
HandstandWalking 0.0882353 34.0
BodyWeightSquats 0.16666667 30.0
FrontCrawl 0.1891892 37.0
YoYo 0.19444445 36.0
PizzaTossing 0.27272728 33.0
CricketShot 0.30612245 49.0
Nunchucks 0.4 35.0
PullUps 0.42857143 28.0
PommelHorse 0.45714286 35.0
JavelinThrow 0.48387095 31.0
Hammering 0.4848485 33.0

- **Use the confusion matrices to get the 10 most confused classes. That is, which off-diagonal elements of the confusion matrix are the largest: Are there any notable examples?**

Most Confused Classes:

The first indices indicate the most confused classes here.

[[0.67567569, [31, 18]] : FrontCrawl, BreastStroke (FronCrawl being misclassified as BreastStroke)
[0.36363637, [33, 12]]: Haircut, BlowDryHair
[0.36363637, [0, 1]]: ApplyEyeMakeup, ApplyLipstick
[0.30612245, [23, 22]]: CricketShot, CricketBowling
[0.28947368, [47, 42]]: JumpRope, HulaHoop
[0.27906978, [77, 19]],: ShavingBeard,BrushingTeeth
[0.26530612, [16, 17]]: BoxingPunchingBag, BoxingSpeedBag
[0.25, [19, 77]]: BrushingTeeth, ShavingBeard
[0.22857143, [68, 56]]: PommelHorse, ParallelBars
[0.21621622, [39, 44]]]: HighJump, JavelinThrow

Sequence Model:

High Performing Classes:

BasketballDunk 1.0 37.0
WritingOnBoard 1.0 45.0
Biking 1.0 38.0
Billiards 1.0 40.0
PlayingPiano 1.0 28.0
CleanAndJerk 1.0 33.0
HeadMassage 1.0 41.0
Diving 1.0 45.0
Rowing 1.0 36.0

FrisbeeCatch 1.0 37.0
Bowling 1.0 43.0
Fencing 1.0 34.0
PoleVault 1.0 40.0

Low Performing Classes:

ApplyLipstick 0.09375 32.0
Lunges 0.21621622 37.0
Nunchucks 0.31428573 35.0
HandstandWalking 0.32352942 34.0
CricketShot 0.3469388 49.0
JavelinThrow 0.3548387 31.0
PullUps 0.35714287 28.0
FrontCrawl 0.4054054 37.0
Haircut 0.42424244 33.0
GolfSwing 0.46153846 39.0
Swing 0.47619048 42.0

Most Confused Classes:
Most misclassified classes:

[0.625, [1, 19]]: ApplyLipstick
[0.54054052, [31, 18]]: FrontCrawl
[0.45714286, [68, 56]]: PommelHorse
[0.39393941, [33, 12]]: Haircut
[0.3877551, [23, 22]]: CricketShot
[0.36363637, [34, 38]]: Hammering
[0.35714287, [69, 74]]: PullUps
[0.30555555, [48, 72]]: Kayaking
[0.26666668, [35, 92]]: HammerThrow
[0.25806451, [44, 50]]: JavelinThrow

As can be seen, the high performing and low performing classes are different for both models. The spatial model seems to perform well on videos that involve more objects. It underperforms on activities. Temporal model works well on activities like playing instrument, writing etc.

Th most confused classes seem to be similar in both cases like FrontCrawl, ApplyLipstick, Haircut because these involve complicated activities and thus the classifier cannot classify them reasonably.