HW5 Transfer Learning Report:

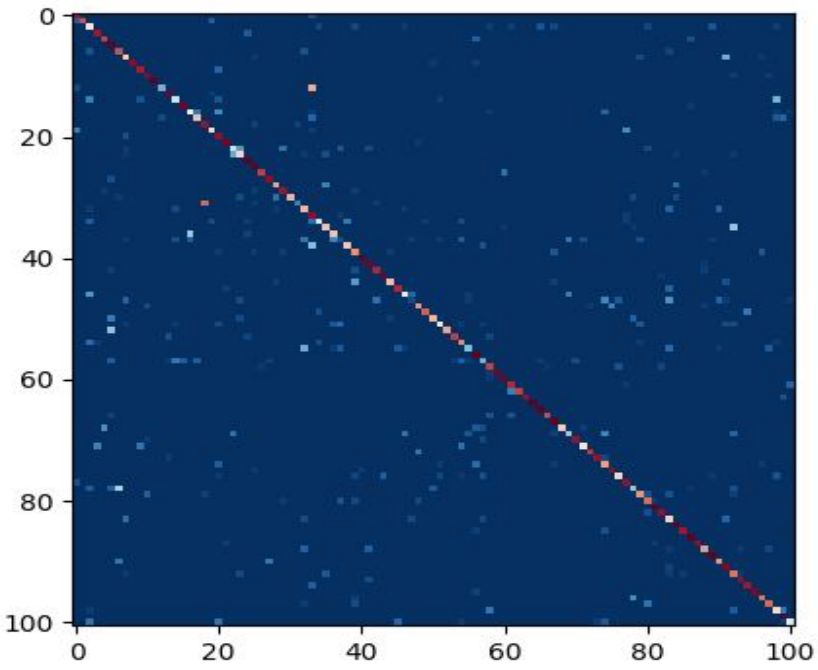
Model 1: (0.775311,0.936823,0.970394)
Model 2: (0.799636,0.947177,0.967213)
Combined: (0.851175,0.966057,0.973890)

	Top1 Accuracy	Top5 Accuracy	Top10 Accuracy
Model 1	77.53%	93.68%	97.04%
Model 2	79.96%	94.71%	96.72%
Combined	85.11%	96.61%	97.39%

The combined accuracy values are obtained by taking the average of the probabilities of each class for both models. The confusion matrices are plotted below and are generated using the confusion.py script.

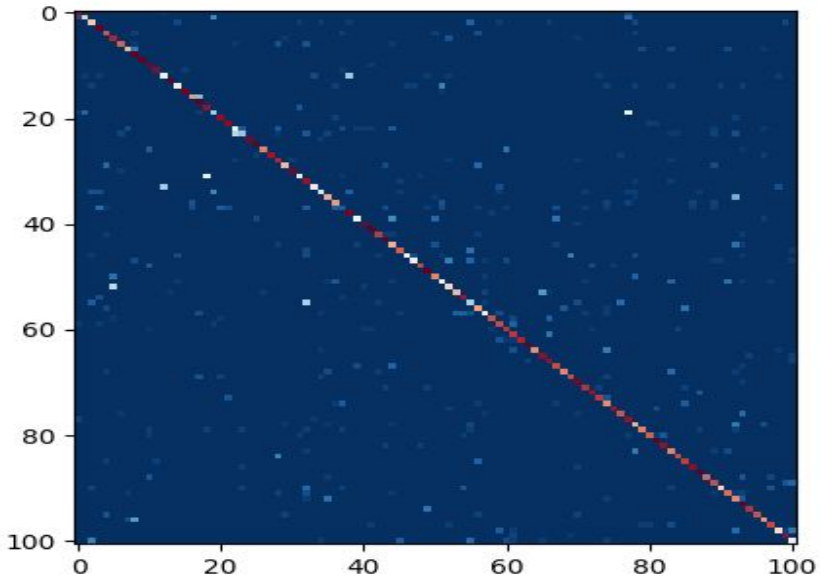
Model 1:

This is the plot of confusion matrix from the first model. The color scheme of cmap='RdBu_r' was used which indicates that the more red there is in the image the better the classification is for that class. Similarly blue indicates a worse classification.



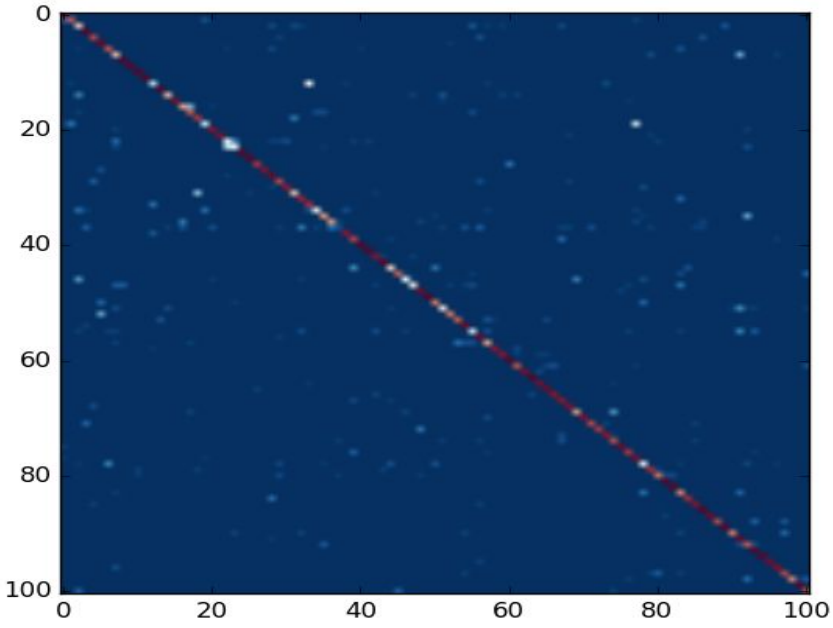
Model 2:

This is the plot of confusion matrix from the second model. The color scheme of cmap='RdBu_r' was used again.



Model 3 (Combined Model):

This is the plot of confusion matrix from the first model. The same color scheme of cmap='RdBu_r' from before was used.



Calculated difference in performance for each class between the part 1 model and the combined output model:

This is saved as performance_difference.csv file and was generated using the performance_difference.py script.

Report the 10 classes with the largest improvement and the 10 classes with worse performance (or smallest improvement if they all improved). Generated using report_top10.py file

The top 10 classes with largest improvement and 10 classes with worse performance are shown below respectively:

Classes that perform Better in (Model 1) / worse in Combined Model

Classes	Accuracy
SoccerPenalty	0.12195122
Rafting	0.10714286
CricketShot	0.10204082
ApplyLipstick	0.09375
PlayingDaf	0.07317072
BrushingTeeth	0.05555555
Lunges	0.05405402
JugglingBalls	0.05000001
BalanceBeam	0.0322581
Rowing	0.02777779

Classes that perform worse in (Model1)/ better in Combined Model:

Classes	Accuracy
BenchPress	-0.45833331
PullUps	-0.35714286
PushUps	-0.29999995
HeadMassage	-0.26829267
JumpRope	-0.26315788
WallPushups	-0.25714287
PommelHorse	-0.25714285
SoccerJuggling	-0.25641027
YoYo	-0.25
BoxingPunchingBag	-0.22448981

It seems that actions that involve motion through time like ‘BenchPress’ and ‘PullUps’ show improvements in the combined model whereas actions like ‘CricketShot’ can possibly be classified well by justing using spatial features and thus tend to do worse in the combined model (better in model1).

However, there are exceptions because ‘JugglingBalls’ does worse in the combined model even though it's an action involving motion through time. So it is not very clear as to the reason/intuition behind the better/worse performances of the models. Overall we can atleast say that the combined model has higher top_1 accuracy as shown above in the first table.