Sayantan Dutta

Question:

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?

Solution:

1) Single Frame Model

(a) top1_accuracy = 0.781025 (b) top5_accuracy = 0.947127 (c) top10_accuracy = 0.973287

2) 3D Model

(a) top1_accuracy = 0.837463   (b) top5_accuracy = 0.969845 (c) top10_accuracy = 0.985201

3) Combined Model

(a) top1_accuracy = 0.863492 (b) top5_accuracy = 0.976523 (c) top10_accuracy = 0.986154

From the results we can see that the combined model performed better than SingleFrame or 3D model.
All the accuracies increased in the combined model, the most notable increase is in top1_accuracy. It increased from around 78% in Single Frame and 84% in 3D model to 86% in the combined model. This makes sense as Combined model combines spatial and temporal information together to make predictions.

Question:

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?

Solution:

1) Single Frame Model

10 Classes with the Highest Performance:
Surfing, BasketballDunk, Rowing, Billiards, PlayingDaf, HorseRace, RockClimbingIndoor, PlayingTabla, FrisbeeCatch, Skijet

10 Classes with the Lowest Performance:
BodyWeightSquats, YoYo, JumpRope, JumpingJack, Shotput, HighJump, HandstandWalking, Nunchucks, CricketShot, BrushingTeeth

2) 3D Model

10 Classes with the Highest Performance:
JumpRope, Fencing, Drumming, HorseRiding, PoleVault, PlayingViolin, PlayingPiano, PlayingTabla, PlayingGuitar, BoxingSpeedBag

10 Classes with the Lowest Performance:
HandstandWalking, Lunges, CricketShot, HighJump, Nunchucks, FrontCrawl, YoYo, SoccerJuggling, ApplyLipstick, PommelHorse

3) Combined Model

10 Classes with the Highest Performance:
JumpingJack, PlayingDhol, BenchPress, Rafting, BasketballDunk, Bowling, BoxingSpeedBag, ParallelBars, RopeClimbing, PlayingViolin

10 Classes with the Lowest Performance:
HighJump, Lunges, YoYo, HandstandWalking, CricketShot, BrushingTeeth, FrontCrawl, Nunchucks, PizzaTossing, Haircut

From results we can see that combined model performed better than SingleFrame or 3D model. Classes like *CricketShot* has increased probability in combined model (0.56) than in 3D model (0.24)

and SingleFrame(0.47). This is because Combined model blends spatial and temporal information together to make predictions.

There are classes like *PlayingViolin* which has high correct classification probability among all three models. SingleFrame (0.964285), 3D model (1.0), Combined model (1.0)

There some classes like *JumpRope* which has been classified as top 10 lowest performing class by SingleFrame (0.052631), but its classified as top 10 best performing class in 3D model (1.0). So, it seems that *JumpRope* class needs more temporal information to be correctly identified than the spatial information.

There are also some classes like *Golfswing* which is not in the top 10 best/worst classes in SingleFrame or 3D model, but in SingleFrame its probability for correct classification is 0.736364, and in 3D model its probability is 0.533333. So, some classes like these require more spatial information than temporal information.

Question:

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?

Solution:

Top 10 confused classes for single frame model:

 (FrontCrawl, BreastStroke),
(BrushingTeeth, ShavingBeard),
(WallPushups, PullUps),
(HighJump, JavelinThrow),
(CricketShot, CricketBowling),
(PommelHorse, ParallelBars),
 (BoxingPunchingBag, BoxingSpeedBag),
 (Nunchucks, GolfSwing),
(ApplyEyeMakeup, ApplyLipstick),
(BalanceBeam, ParallelBars)

Top 10 confused classes for 3D model:

(Haircut, BlowDryHair),
(FrontCrawl, BreastStroke),
(CricketShot, CricketBowling),
(Nunchucks, TaiChi),
(HammerThrow, ThrowDiscus),
(ApplyLipstick, ApplyEyeMakeup),
(BrushingTeeth, ShavingBeard),
(BoxingPunchingBag, BoxingSpeedBag),
(ShavingBeard, ApplyEyeMakeup),
(YoYo, Archery)

Top 10 confused classes for combined model:

(Haircut, BlowDryHair),
(FrontCrawl, BreastStroke),
(CricketShot, CricketBowling),
(PommelHorse, ParallelBars),
(BrushingTeeth, ShavingBeard),
(BoxingPunchingBag, BoxingSpeedBag),
(Hammering, HeadMassage),
(HighJump, JavelinThrow),
(Nunchucks, TaiChi),
(HandstandPushups, HandstandWalking)

We can see that all the three models have similar confusing class pairs.
For all the three models for example the pairs (FrontCrawl, BreastStroke), (BrushingTeeth, ShavingBeard), (CricketShot, CricketBowling), (BoxingPunchingBag, BoxingSpeedBag), etc. have very confusing classes.