

IE 534 Homework 9 Report - Video Recognition

Hanwen Hu

November 29, 2018

This homework asks to finish the following tasks:

1. Classify videos from UCF101 datasets with a random frame from each, trained with ResNet-50;
2. Classify videos from UCF101 datasets with a bunch of continuous frames (16 in our case), trained with ResNet-101 (3-dimension form).

1 Hyperparameter Settings

For the first part, the batch size is set as 100; I adopt Adam as the optimizer with initial learning rate $1e-4$; I activate 8 CPU cores and run 30 epochs when the model begins to be overfitting.

For the second part, the batch size is set as 16; I adopt Adam as the optimizer with initial learning rate $1e-4$; I activate 8 CPU cores and run 11 epochs in total.

As for the combination part, I combine the outputs with a weighted average between two parts, which gives 0.1 to part 1 and 0.9 to part 2. This gives a better performance than each of the model in top 5 and top 10 accuracy, while sacrifices little accuracy in top 1.

2 Accuracy

For the first part, accuracy of top-1 is 0.785884; accuracy of top-5 is 0.936558; accuracy of top-10 is 0.966957.

For the second part, accuracy of top-1 is 0.830029; accuracy of top-5 is 0.959820; accuracy of top-10 is 0.966957.

For the combination part, accuracy of top-1 is 0.822099; accuracy of top-5 is 0.960349; accuracy of top-10 is 0.978588.

3 Result Evaluation

3.1 Best and worst classes

For the first part, the best classes that are all classified correctly are:

'BasketballDunk', 'Billiards', 'Bowling', 'BreastStroke', 'Diving', 'FloorGymnastics', 'HorseRace', 'HorseRiding', 'MilitaryParade', 'PlayingDaf', 'PlayingTabla', 'PoleVault', 'RockClimbingIndoor', 'Rowing', 'Skijet', 'SkyDiving', 'SoccerPenalty', 'SumoWrestling', 'Surfing', 'WritingOnBoard'.

The worst 10 classes are:

'JumpRope', 'HandstandWalking', 'YoYo', 'Nunchucks', 'BodyWeightSquats', 'HighJump', 'FrontCrawl', 'PullUps', 'WallPushups', 'Lunges'.

As for the second part, the best classes that are all classified correctly are:

'BabyCrawling', 'BalanceBeam', 'BasketballDunk', 'Billiards', 'Bowling', 'BoxingSpeedBag', 'Fencing', 'FloorGymnastics', 'HorseRace', 'HorseRiding', 'JumpingJack', 'Knitting', 'PlayingDhol', 'PlayingGuitar', 'PlayingPiano', 'PlayingTabla', 'PoleVault', 'RockClimbingIndoor', 'Skijet', 'SoccerPenalty', 'StillRings', 'VolleyballSpiking', 'WallPushups'.

The worst 10 classes are:

'Nunchucks', 'Lunges', 'JavelinThrow', 'YoYo', 'Shotput', 'HandstandWalking', 'MoppingFloor', 'SoccerJuggling', 'PizzaTossing', 'HighJump'.

As for the third part, the best classes that are all classified correctly are:

'BabyCrawling', 'Billiards', 'BoxingSpeedBag', 'Fencing', 'FloorGymnastics', 'HorseRiding', 'JumpingJack', 'Knitting', 'PlayingDhol', 'PlayingGuitar', 'PlayingPiano', 'PlayingTabla', 'PoleVault', 'RockClimbingIndoor', 'Skijet', 'SoccerPenalty', 'StillRings', 'VolleyballSpiking'.

The worst 10 classes are:

'Nunchucks', 'YoYo', 'Lunges', 'JavelinThrow', 'HandstandWalking', 'Shotput', 'MoppingFloor', 'CricketBowling', 'PizzaTossing', 'SoccerJuggling'.

As I observe, there are 20 classes best classified by the first model, 23 by the second model, 18 by the third model. This implies that the second model works a bit better than the first model in UCF101 dataset, while the combination doesn't work very well in top-1 prediction.

In addition, among the two lists of best classes, there are in total 11 classes are best classified by both parts. This implies that both temporal and spatial information about these classes are properly classified.

'BasketballDunk', 'Billiards', 'Bowling', 'FloorGymnastics', 'HorseRace', 'HorseRiding', 'PlayingTabla', 'PoleVault', 'RockClimbingIndoor', 'Skijet', 'SoccerPenalty'.

As for the worst 10 classes, the two models share almost all the same classes: *'Nunchucks', 'YoYo', 'Lunges', 'JavelinThrow', 'HandstandWalking', 'Shotput', 'MoppingFloor', 'PizzaTossing', 'SoccerJuggling'.* This implies they are discriminated by both spatial and temporal information.

The classes that are best classified by the first model, but not the second:

'BreastStroke', 'Diving', 'MilitaryParade', 'PlayingDaf', 'Rowing', 'SkyDiving', 'SumoWrestling', 'Surfing', 'WritingOnBoard',

which means these classes are discriminated by their temporal information.

The classes that are best classified by the second model, but not the first:
'BabyCrawling', 'BalanceBeam', 'BoxingSpeedBag', 'Fencing', 'JumpingJack', 'Knitting', 'PlayingDhol', 'PlayingGuitar', 'PlayingPiano', 'StillRings', 'VolleyballSpiking', 'WallPushups'.

which means these classes are discriminated by their spatial information.

3.2 Most confused pairs of classes

For the first part, the most confused 10 pairs of classes are:

TrueLabel	Prediction
FrontCrawl	BreastStroke
Haircut	BlowDryHair
PommelHorse	ParallelBars
HighJump	JavelinThrow
BrushingTeeth	ShavingBeard
CricketShot	CricketBowling
HandstandWalking	PushUps
JumpRope	HulaHoop
HammerThrow	ThrowDiscus
BodyWeightSquats	WallPushups

Table 1: 10 most confused pairs of classes in Part 1

And for the second part, the most confused 10 pairs of classes are:

TrueLabel	Prediction
Haircut	BlowDryHair
JavelinThrow	LongJump
Rowing	Skijet
YoYo	JugglingBalls
HighJump	LongJump
FrontCrawl	BreastStroke
MoppingFloor	HandstandWalking
BoxingPunchingBag	BoxingSpeedBag
Hammering	HeadMassage
JumpRope	JumpingJack

Table 2: 10 most confused pairs of classes in Part 2

For the combination part, the most confused 10 pairs of classes are:

TrueLabel	Prediction
Haircut	BlowDryHair
JavelinThrow	LongJump
Rowing	Skijet
YoYo	JugglingBalls
FrontCrawl	BreastStroke
MoppingFloor	HandstandWalking
HighJump	LongJump
BoxingPunchingBag	BoxingSpeedBag
CricketShot	CricketBowling
ApplyLipstick	BrushingTeeth

Table 3: 10 most confused pairs of classes in Part 3

As I observe, Haircut videos will always be confused as BlowDryHair in all outputs. Since both behaviors involve hair, it's reasonable to say that the extracted information from videos by our models mainly focuses on hair rather than the behavior on it. But I notice that part 1 and part 2 outputs do not have any other common most confused pairs of classes.

As for the combination part, since more weights are put on Part 2 outputs, it shares a lot of the same confused pairs with Part 2. However, the pair 'CricketShot'-'CricketBowling' is introduced from Part 1, and the pair 'HighJump'-'LongJump' becomes less confused than Part 2.