Model	Appearance	Motion	Sampling method	Frame num.	
Previous Work					
TA	GoogleNet (0.5x)	C3D (2x)	uniform sampling 26 frames	26 (4x)	0.5×2
S2VT	VGG (0.5x)	OF (2x)	uniform sampling 80 frames	80 (13x)	$0.5 \times 2 \times$
LSTM-E	VGG (0.5x)	C3D (2x)	uniform sampling 30 frames	30 (5x)	0.5×2
p-RNN	VGG (0.5x)	C3D (2x)	uniform sampling 30 frames	30 (5x)	0.5×2
HRNE	GoogleNet (0.5x)	C3D (2x)	first 200 frames	200 (33x)	$0.5 \times 2 \times$
BA	ResNet (0.5x)	C3D (2x)	every 5 frames	72 (12x)	$0.5 \times 2 \times$
Our Models					
Baseline	ResNet (1x)	×	uniform sampling 30 frames	30 (5x)	1
Random	ResNet (1x)	×	randomly sampling	15 (2.5x)	1×2
k-means (k=6)	ResNet (1x)	×	k-means clustering	6 (1x)	1
PickNet (V)	ResNet (1x)	×	picking	6 (1x)	1
PickNet (L)	ResNet (1x)	×	picking	6 (1x)	1
PickNet (V+L)	ResNet (1x)	×	picking	6 (1x)	1