Model	Appearance	Motion	Sampling method	Frame num.	
Previous Work					
ruc-uva	GoogleNet (0.5x)	C3D (2x)	every 10 frames	36 (4.5x)	0.5
Aalto	GoogleNet (0.5x)	C3D+IDT (2x)	one frame every second	36 (4.5x)	0.5
DenseCap	ResNet (0.5x)	C3D (2x)	uniform sampling 30 frames	30 (3.5x)	0.5
MS-RNN	ResNet (1x)	C3D (2x)	uniform sampling 40 frames	40 (5x)	
Our Models					
Baseline	ResNet (1x)	×	uniform sampling 30 frames	30 (3.8x)	
Random	ResNet (1x)	×	randomly sampling	15 (1.9x)	
k-means (k=8)	ResNet (1x)	×	k-means clustering	8 (1x)	
PickNet (V)	ResNet (1x)	×	picking	8 (1x)	
PickNet (L)	ResNet (1x)	×	picking	8 (1x)	
PickNet (V+L)	ResNet (1x)	×	picking	8 (1x)	