

	Accuracy	WUPS @0.9	WUPS @0.0
VSE			
GUESS	6.65	17.42	73.44
BOW	37.52	48.54	82.78
LSTM	36.76	47.58	82.34
IMG	43.02	58.64	85.85
IMG+BOW	55.92	66.78	88.99
VIS+LSTM	53.31	63.91	88.25
2-VIS+BLSTM	55.09	65.34	88.64
FULL	57.84	67.90	89.52
Proposed CNN without multimodal convolution layer	56.77	66.76	88.94
Proposed CNN without image representation	37.84	48.70	82.92
Proposed CNN	58.40	68.50	89.67