

Sample1. In day video (30 minutes) robot10

Event	Time	Range	Partial screenshot
E1_unleashed dog	44s	00:06-0:50	
E2_unleashed cat	56s	02:04-03:00	
E3_Jaywalking_looking_suspicious	14s	05:28-05:42	
E4_people falling	26s	07:03-07:29	

E5_people falling	12s	08:19-08-31	
E6_carrying suspicious bag_leave_bag	20s	10:45-11:05	
E7_people falling	30s	16:10-16:42	
E8_rob_stalkin g_weapon	44s	27:14-27:58	

Sample2. In night video(30 mintues) robot8

Event	Time	Range	Partial screenshot
E1_brawing	47s	03:44-04:23	
E2_arson	20s	06:55-07:10	
E3_arson	11s	9:49-10:-00	

E4_theft_public staff	18s	13:00 - 13:18	
E5_smoking	21s	15:03-15:24	
E6_assault_attack	17s	19:53-19:05	
E7_people falling	67s	23:53-25:00	

To mitigate location re-identification risk, we submit screenshots from two anonymized samples instead of raw videos.

In the future, we will replace these with links to short redacted event video clips and 20 full videos.