Statistical Overview of the MicroG-4M Dataset

Lei Qi

The following tables summarize the statistics of the MicroG-4M dataset for the human action recognition task.

Category	Count
All videos	4759
Movie videos	1332
Real videos	3427
All actions	13261
Movie actions	3651
Real actions	9610
All action types	50
Movie action types	42
Real action types	46

Table 1: Summary of Video Categories, Actions, and Action Types

Group Nr.	Difference Amount	Corrected Amount
1	3	3
2	0	0
3	0	0
4	10	8
5	11	6
6	0	0
Label noise in the dataset: 17/600 ≈ 2.83%		

Table 2: Cross-check Data Summary

Label Type	Number of Labels	Ratio (%)
Object Manipulation	4986	37.60
Person Interaction	4288	32.34
Person Movement	3987	30.07

Table 3: Summary of Label Types and Their Distribution Across All Videos

Label Type	Number of Labels	Ratio (%)
Object Manipulation	1416	38.78
Person Interaction	1198	32.81
Person Movement	1037	28.40

Table 4: Summary of Label Types and Their Distribution Across Movies

Label Type	Number of Labels	Ratio (%)
Object Manipulation	3788	39.42
Person Interaction	2950	30.70
Person Movement	2872	29.89

Table 5: Summary of Label Types and Their Distribution Across Real Videos

Number of Persons	Number of Videos	Ratio (%)
1	3983	83.69
2	623	13.09
3	118	2.48
4	21	0.44
5	11	0.23
6	2	0.04
7	1	0.02

Table 6: Distribution of the Number of Persons per Video (All Videos)

Number of Persons	Number of Videos	Ratio (%)
1	816	61.26
2	395	29.65
3	94	7.06
4	17	1.28
5	8	0.60
6	1	0.08
7	1	0.08

Table 7: Distribution of the Number of Persons per Video (Movies)

Number of Persons	Number of Videos	Ratio (%)
1	3167	92.41
2	228	6.65
3	24	0.70
4	4	0.12
5	3	0.09
6	1	0.03

Table 8: Distribution of the Number of Persons per Video (Real Videos)

Table 9: Summary of Action Type Distribution in All Videos

ID	Action Name	Number of Actions	Ratio (%)
12	stand	3218	24.27
79	talk to (e.g., self, a person, a group)	3131	23.61
17	carry/hold (an object)	3126	23.57
80	watch (a person)	713	5.38
14	walk	369	2.78
59	touch (an object)	353	2.66
61	watch (e.g., TV)/any unspecified action	346	2.61
43	point to (an object)	323	2.44
11	sit	252	1.90
36	lift/pick up	188	1.42
74	listen to (a person)	148	1.12
69	hand wave	140	1.06
47	put down	138	1.04
62	work on a computer	110	0.83
30	enter	68	0.51
65	give/serve (an object) to (a person)	46	0.35
29	eat	45	0.34
27	drink	43	0.32
66	grab (a person)	41	0.31
34	hit (an object)	33	0.25
45	pull (an object)	32	0.24
38	open (e.g., a window, a car door)	32	0.24
26	dress/undress clothing	31	0.23
64	fight/hit (a person)	27	0.20
1	bend/bow (at the waist)	26	0.20
6	get up	25	0.19
46	push (an object)	24	0.18
28	operate spaceship	20	0.15
3	crouch/kneel	20	0.15
7	jump/leap	20	0.15
9	martial art	18	0.14
8	lie/sleep	17	0.13
60	turn (e.g., a screwdriver)	17	0.13
70	hug (a person)	16	0.12
78	take (an object) from (a person)	15	0.11
48	read	15	0.11
22	close (e.g., a door, a box)	13	0.10
10	run/jog	12	0.09
5	fall down	10	0.08
24	cut	9	0.07
57	text on/look at a cellphone	7	0.05
58	throw	4	0.03
68	hand shake	4	0.03
63	write	3	0.02
67	hand clap	3	0.02
76	push (another person)	3	0.02
41	play musical instrument	3	0.02
56	take a photo	2	0.02
20	climb (e.g., a mountain)	1	0.01
72	kiss (a person)	1	0.01

Table 10: Summary of Action Type Distribution in Movies

ID	Action Name	Number of Actions	Ratio (%)
12	stand	698	19.12
80	watch (a person)	625	17.12
17	carry/hold (an object)	549	15.04
79	talk to (e.g., self, a person, a group)	504	13.80
61	watch (e.g., TV)/any unspecified action	316	8.66
11	sit	239	6.55
59	touch (an object)	136	3.73
74	listen to (a person)	135	3.70
14	walk	75	2.05
62	work on a computer	67	1.84
69	hand wave	44	1.21
66	grab (a person)	31	0.85
64	fight/hit (a person)	26	0.71
65	give/serve (an object) to (a person)	20	0.55
45	pull (an object)	19	0.52
28	operate spaceship	16	0.44
1	bend/bow (at the waist)	14	0.38
48	read	14	0.38
70	hug (a person)	13	0.36
38	open (e.g., a window, a car door)	13	0.36
78	take (an object) from (a person)	10	0.27
36	lift/pick up	10	0.27
60	turn (e.g., a screwdriver)	9	0.25
26	dress/undress clothing	9	0.25
46	push (an object)	8	0.22
47	put down	7	0.19
27	drink	5	0.14
30	enter	5	0.14
6	get up	4	0.11
22	close (e.g., a door, a box)	4	0.11
8	lie/sleep	4	0.11
63	write	3	0.08
34	hit (an object)	3	0.08
67	hand clap	3	0.08
68	hand shake	2	0.05
29	eat	2	0.05
3	crouch/kneel	2	0.05
76	push (another person)	2	0.05
43	point to (an object)	2	0.05
20	climb (e.g., a mountain)	1	0.03
5	fall down	1	0.03
72	kiss (a person)	1	0.03

Table 11: Summary of Action Type Distribution in Real Videos

ID	Action Name	Number of Actions	Ratio (%)
79	talk to (e.g., self, a person, a group)	2627	27.34
17	carry/hold (an object)	2577	26.82
12	stand	2520	26.22
43	point to (an object)	321	3.34
14	walk	294	3.06
59	touch (an object)	217	2.26
36	lift/pick up	178	1.85
47	put down	131	1.36
69	hand wave	96	1.00
80	watch (a person)	88	0.92
30	enter	63	0.66
62	work on a computer	43	0.45
29	eat	43	0.45
27	drink	38	0.40
61	watch (e.g., TV)/any unspecified action	30	0.31
34	hit (an object)	30	0.31
65	give/serve (an object) to (a person)	26	0.27
26	dress/undress clothing	22	0.23
6	get up	21	0.22
7	jump/leap	20	0.21
38	open (e.g., a window, a car door)	19	0.20
9	martial art	18	0.19
3	crouch/kneel	18	0.19
46	push (an object)	16	0.17
11	sit	13	0.14
45	pull (an object)	13	0.14
8	lie/sleep	13	0.14
74	listen to (a person)	13	0.14
10	run/jog	12	0.12
1	bend/bow (at the waist)	12	0.12
66	grab (a person)	10	0.10
5	fall down	9	0.09
22	close (e.g., a door, a box)	9	0.09
24	cut	9	0.09
60	turn (e.g., a screwdriver)	8	0.08
57	text on/look at a cellphone	7	0.07
78	take (an object) from (a person)	5	0.05
28	operate spaceship	4	0.04
58	throw	4	0.04
41	play musical instrument	3	0.03
70	hug (a person)	3	0.03
56	take a photo	2	0.02
68	hand shake	2	0.02
48	read	1	0.01
64	fight/hit (a person)	1	0.01
76	push (another person)	1	0.01