Supplementary Material for
MOVE: Motion-Guided Few-Shot Video Object Segmentation

Kaining Ying*   Hengrui Hu*   Henghui Ding✉

Fudan University, China

https://henghuiding.com/MOVE/
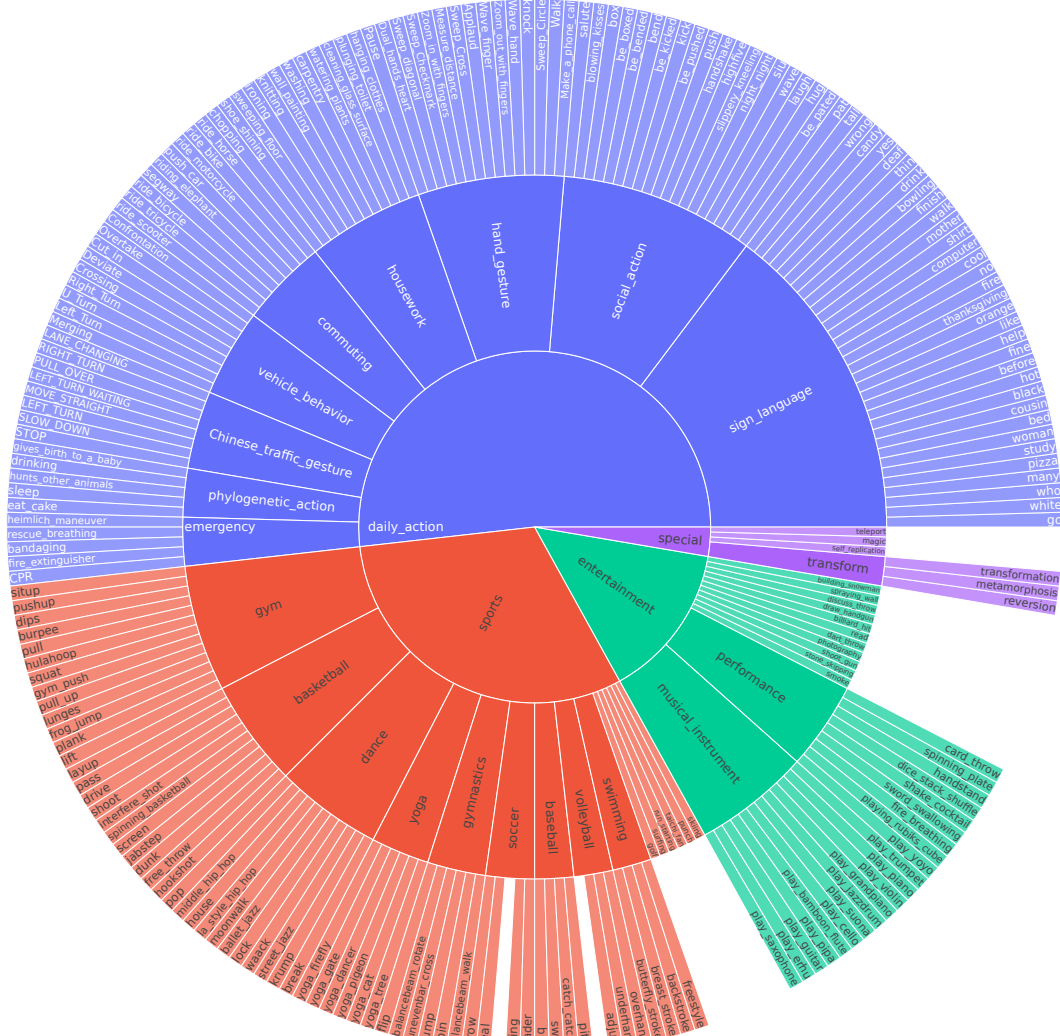
A. Dataset

A.1. Vocabulary Distribution

Figure I. The motion category distribution of our proposed MOVE.

As shown in Figure I, our motion categories are hierarchically distributed, with a total of 229 subcategories divided into 4 main categories. This ensures the diversity of our dataset.

---

*Equal contribution.   ✉Corresponding author (henghui.ding@gmail.com).

A.2. Dataset Source

We show the source of the dataset for each motion category in Tables I to V.

Table I. Dataset source of `MOVE` (part 1).

Node 1	Node 2	Node 3	Node 4	Source
sports	soccer	steal		MultiSports [8]
		throw	football throw	MultiSports [8]
			soccer throw	HAA500 [2]
		shoot	football shoot	MultiSports [8]
			soccer shoot	HAA500 [2]
		diving		MultiSports [8]
		header		HAA500 [2]
	basketball	pass		MultiSports [8]
		screen		MultiSports [8]
		3-point shot		MultiSports [8]
		free throw		MultiSports [8]
		block		MultiSports [8]
		dunk		HAA500 [2]
		hookshot		HAA500 [2]
		layup		HAA500 [2]
		jabstep		HAA500 [2]
		spinning		HAA500 [2]
	baseball	swing		HAA500 [2]
		catch		HAA500 [2]
		bunt		HAA500 [2]
		pitch		HAA500 [2]
	volleyball	underhand		HAA500 [2]
		set		HAA500 [2]
		overhand		HAA500 [2]
		pass		HAA500 [2]
	dance	break		AIST Dance [7]
		pop		AIST Dance [7]
		lock		AIST Dance [7]
		middle hip-hop		AIST Dance [7]
		LA style hip-hop		AIST Dance [7]
		house		AIST Dance [7]
		waack		AIST Dance [7]
		krump		AIST Dance [7]
		street jazz		AIST Dance [7]
		ballet jazz		AIST Dance [7]
		moonwalk		HAA500 [2]
	gym	push up	pushup	HAA500 [2]
			one_arm_push_up	HAA500 [2]
		situp	situp	HAA500 [2]
			situp	HMDB [6]
		burpee		HAA500 [2]
		dips		HAA500 [2]
		pull up		HAA500 [2]
		cartwheel		HAA500 [2]
		lift		HAA500 [2]

Table II. Dataset source of `MOVE` (part 2).

Node 1	Node 2	Node 3	Node 4	Source
sports	gym	push		HAA500 [2]
		lunges		HAA500 [2]
		plank		HAA500 [2]
		pull		HAA500 [2]
		squat		HAA500 [2]
		frog jump		HAA500 [2]
	yoga	cat		HAA500 [2]
		dancer		HAA500 [2]
		tree		HAA500 [2]
		firefly		HAA500 [2]
		gate		HAA500 [2]
		pigeon		HAA500 [2]
	gymnastics	unevenbar_cross		HAA500 [2]
		flip	unevenbar flip	HAA500 [2]
			balancebeam flip	HAA500 [2]
			horizontalbar flip	HAA500 [2]
		spin	unevenbar spin	HAA500 [2]
			balancebeam spin	HAA500 [2]
			horizontalbar spin	HAA500 [2]
		land	unevenbar land	HAA500 [2]
			horizontalbar land	HAA500 [2]
		jump	balancebeam jump	HAA500 [2]
			unevenbar jump	HAA500 [2]
			horizontalbar jump	HAA500 [2]
		balancebeam rotate		HAA500 [2]
		balancebeam walk		HAA500 [2]
	swim	backstroke		HAA500 [2]
		breast stroke		HAA500 [2]
		butterfly stroke		HAA500 [2]
		freestyle		HAA500 [2]
	golf	golf swing		HAA500 [2]
		golf		HMDB [6]
	surfing			HAA500 [2]
	skiing			HAA500 [2]
	punch	punching speedbag		HAA500 [2]
		punching sandbag		HAA500 [2]
	taichi fan			HAA500 [2]
	discuss throw			HAA500 [2]
	dart throw			HAA500 [2]

Table III. Dataset source of `MOVE` (part 3).

Node 1	Node 2	Node 3	Node 4	Source
daily action	phylogenetic actions	give birth		MammalNet [1]
		eat	apple	HAA500 [2]
			cake	HAA500 [2]
			burger	HAA500 [2]
			ice-cream	HAA500 [2]
			noodles	HAA500 [2]
		drinking		HAA500 [2]
		sleep		Internet
		hunt		MammalNet [1]
	hand gesture	Sweep Circle		EgoGesture [10]
		sweep cross		EgoGesture [10]
		sweep checkmark		EgoGesture [10]
		measure distance		EgoGesture [10]
		pause		EgoGesture [10]
		make a phone call		EgoGesture [10]
		walk		EgoGesture [10]
		applaud		EgoGesture [10]
		sweep diagonal		EgoGesture [10]
		dual hands heart		EgoGesture [10]
		wave hand		EgoGesture [10]
		wave finger		EgoGesture [10]
		knock		EgoGesture [10]
		zoom in with fingers		EgoGesture [10]
		zoom out with fingers		EgoGesture [10]
	sign language	drink		WLASL [3]
		computer		WLASL [3]
		before		WLASL [3]
		yes		WLASL [3]
		go		WLASL [3]
		who		WLASL [3]
		thin		WLASL [3]
		cool		WLASL [3]
		like		WLASL [3]
		mother		WLASL [3]
		candy		WLASL [3]
		no		WLASL [3]
		thanksgiving		WLASL [3]
		cousin		WLASL [3]
		deaf		WLASL [3]
		hot		WLASL [3]
		woman		WLASL [3]
		bed		WLASL [3]
		bowling		WLASL [3]
		kiss		WLASL [3]
		shirt		WLASL [3]
		study		WLASL [3]

Table IV. Dataset source of MOVE (part 4).

Node 1	Node 2	Node 3	Node 4	Source
daily action	traffic gesture	stop		chinese traffic [4]
		move straight		chinese traffic [4]
		left turn		chinese traffic [4]
		left turn waiting		chinese traffic [4]
		right turn		chinese traffic [4]
		lane changing		chinese traffic [4]
		slow down		chinese traffic [4]
		pull over		chinese traffic [4]
	social action	celebrate	siu	Internet
			night night	Internet
			slippery kneeling	Internet
		hug	hugging animal	HAA500 [2]
			hugging person	HAA500 [2]
			hug	BIT [5]
		handshake	handshake dog	HAA500 [2]
			handshake	BIT [5]
		highfive	dog highfive	HAA500 [2]
			high five	HAA500 [2]
			highfive	BIT [5]
		pat	pat	BIT [5]
			be pated	BIT [5]
		push	push person	BIT [5]
			be pushed	BIT [5]
		bend	bend	BIT [5]
			be bended	BIT [5]
		box	box	BIT [5]
			be boxed	BIT [5]
		kick	kick	BIT [5]
			be kicked	BIT [5]
		salute		HAA500 [2]
		laugh		HMDB [6]
		wave	wave	HMDB [6]
			arm wave	HAA500 [2]
	commuting	ride bike		HAA500 [2]
		ride motorcycle		HAA500 [2]
		ride scooter		HAA500 [2]
		segway		HAA500 [2]
		ride horse	ride horse	HAA500 [2]
			ride horse	HMDB [6]
			equestrian run	HAA500 [2]
		riding element		HAA500 [2]
		push car		HAA500 [2]
		drive car		Internet

Table V. Dataset source of MOVE (part 5).

Node 1	Node 2	Node 3	Node 4	Source
daily action	vehicle behavior	Confrontation		IDDX [9]
		Cut-in		IDDX [9]
		merging		IDDX [9]
		left turn		IDDX [9]
		deviate		IDDX [9]
		crossing		IDDX [9]
		overtake		IDDX [9]
		u-turn		IDDX [9]
		right turn		IDDX [9]
	emergency	fire extinguisher		HAA500 [2]
		bandaging		HAA500 [2]
		heimlich maneuver		HAA500 [2]
		cpr		HAA500 [2]
		rescue breathing		HAA500 [2]
entertainment	musical instrument	violin		HAA500 [2]
		piano		HAA500 [2]
		pipa		Internet
		cello		HAA500 [2]
		erhu		HAA500 [2]
		jazz drum		HAA500 [2]
		trumpet		HAA500 [2]
		suona		HAA500 [2]
		bamboo flute		Internet
		saxophone		HAA500 [2]
		guitar		HAA500 [2]
	performance	playing yoyo		HAA500 [2]
		sword swallowing		HAA500 [2]
		fire breathing		HAA500 [2]
		handstand		HAA500 [2]
		card throw		HAA500 [2]
		dice stack shuffle		HAA500 [2]
		shake cocktail		HAA500 [2]
		playing rubiks cube		HAA500 [2]
		spinning plate		HAA500 [2]
	smoke	smoking exhale		HAA500 [2]
		smoking inhale		HAA500 [2]
		smoke		HMDB [6]
	draw handgun			HAA500 [2]
	shooting handgun			HAA500 [2]
	shooting shotgun			HAA500 [2]
	stone skipping			HAA500 [2]
	building snowman			HAA500 [2]
	billiard hit			HAA500 [2]
special	transform	transformer	robot to car	Internet
			car to robot	Internet
		metamorphosis		Internet
	magic			Internet
	telesport			Internet
	self replication			Internet

A.3. Object Category Distribution

Table VI. Object categories in `MOVE` grouped into 5 main categories.

Main Category	Object Classes
Person	person, athlete, dancer, performer, woodworker, musician, soldier, chef, gunner, mage, horseman
Animal	cat, dog, panda, fox, cheetah, rabbit, lion, tiger, bear, horse, monkey, jaguar, pig, llama, mouse, elephant, squirrel, wildebeest, cow, deer, hamster, manatee, kapibala, giraffe, husky dog, sheep, caracal, alpaca, seal, malamute, frog, bulldog, camel, gibbon, snow leopard, chinchilla, lynx, red fox, tigger, terrier, red-panda, flying fox, donkey, buffalo, duck, raccoon, marmot, hyena, turtle, blue whale, killer whale, ermine, serval, lizard, caribou
Vehicle	motorcycle, car, truck, autorickshaw, bus, bicycle
Limbs	hand, mouth
Others	robot, extinguisher, toy, cartoon, sculpture, dinosaur, dragon, hamburger, wind, orcs, other
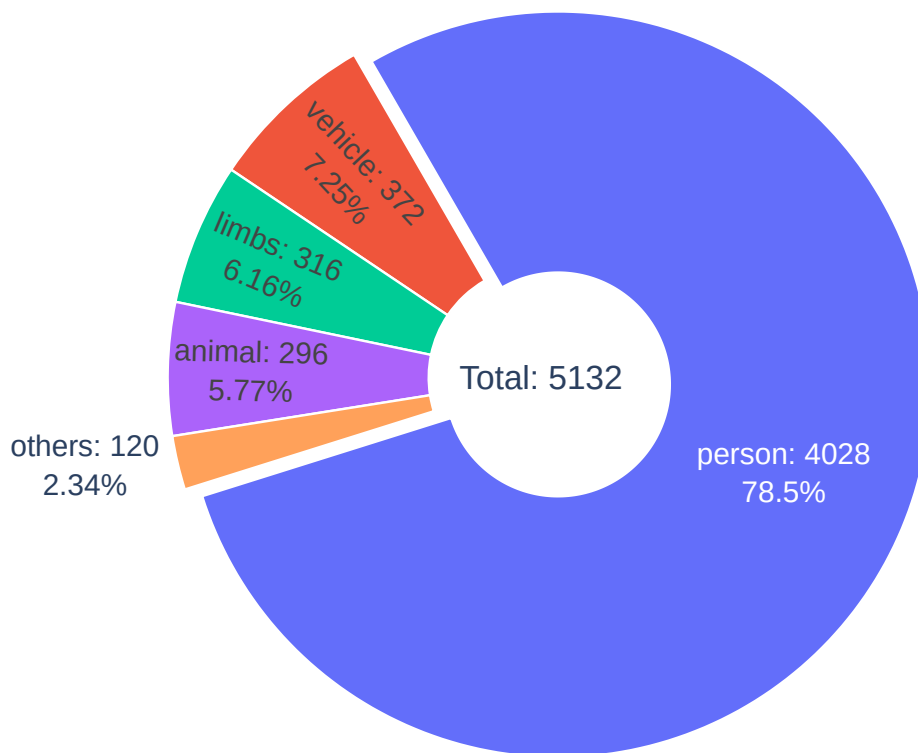
Figure II. The object category distribution of our proposed `MOVE`.

Our dataset contains 88 object categories that can be grouped into 5 main categories: person, animal, vehicle, limbs, and others. Table VI shows the detailed categorization of all object classes in our dataset. As shown in Figure II, the distribution of these object categories reflects the diversity and flexibility of the motion in our dataset.

A.4. Dataset Split

A.4.1. Overlapping Split

Table VII. Statistics of the overlapping split in `MOVE`.

Folder	# Categories	# Objects	# Videos
Folder-1	56	1209	1110
Folder-2	56	1427	1234
Folder-3	56	1340	1267
Folder-4	56	1513	1287

As shown in Table VII, we divide our dataset into 4 folders for the overlapping split (OS), ensuring a balanced distribution of actions, objects, and videos across folders. Table IX and Table X show the detailed statistics of each folder. Each group contains 56 diverse actions spanning daily activities, sports, entertainment, and special effects. The distribution ensures that each group has a representative sample of the full dataset, allowing for effective cross-validation during model development and evaluation. In the overlapping split, actions from the same parent category can appear across different folders, which enables the model to learn generalizable features within similar action types while still maintaining diversity in each split.

A.4.2. Non-Overlapping Split

Table VIII. Statistics of the non-overlapping split in `MOVE`.

Folder	# Categories	# Objects	# Videos
Folder-1	56	1071	1070
Folder-2	60	1793	1232
Folder-3	53	1217	957
Folder-4	55	1051	1040

As shown in Table VIII, we divide our dataset into 4 folders for the non-overlapping split (NS), where we categorize actions based on their second-level parent nodes. Table XI and Table XII show the detailed statistics of each folder. Compared to the overlapping split, this approach presents a greater challenge as it creates larger domain gaps between different folders. This design intentionally tests models' ability to generalize across more distinct action categories.

Table IX. Motion categories in folder-1 and folder-2 of the overlapping split.

Folder	Motions
Folder-1	daily_action-phylogenetic_action-eat_cake, entertainment-performance-spinning_plate, sports-soccer-header, sports-volleyball-overhand, daily_action-hand_gesture-Zoom_in_with_fingers, sports-dance-house, daily_action-sign_language-cool, sports-dance-lock, sports-basketball-interfere_shot, daily_action-social_action-be_bended, daily_action-emergency-bandaging, daily_action-hand_gesture-Sweep_Checkmark, daily_action-social_action-celebrate-slippery_kneeling, entertainment-musical_instrument-play_suona, entertainment-performance-card_throw, daily_action-social_action-salute, daily_action-sign_language-black, entertainment-musical_instrument-play_violin, entertainment-musical_instrument-play_trumpet, daily_action-phylogenetic_action-sleep, daily_action-social_action-celebrate-night_night, daily_action-chinese traffic_gesture-RIGHT_TURN, sports-taichi_fan, daily_action-hand_gesture-Zoom_out_with_fingers, sports-yoga-yoga_cat, daily_action-chinese traffic_gesture-STOP, entertainment-draw_handgun, sports-gym-frog_jump, daily_action-sign_language-hot, daily_action-commuting-ride_scooter, sports-punch, daily_action-chinese traffic_gesture-LANE_CHANGING, entertainment-performance-dice_stack_shuffle, daily_action-social_action-bend, sports-gymnastics-balancebeam_walk, daily_action-hand_gesture-Sweep_Circle, daily_action-sign_language-fine, daily_action-sign_language-yes, daily_action-social_action-pat, sports-yoga-yoga_gate, special-transform-transformation, sports-baseball-swing, daily_action-vehicle_behavior-Deviate, daily_action-commuting-segway, sports-gym-lift, daily_action-sign_language-mother, sports-swimming-freestyle, entertainment-musical_instrument-play_cello, sports-basketball-screen, sports-dance-ballet_jazz, daily_action-commuting-ride_motorcycle, entertainment-building_snowman, entertainment-musical_instrument-play_grandpiano, entertainment-musical_instrument-play_pipa, daily_action-sign_language-wrong, daily_action-vehicle_behavior-U_Turn
Folder-2	daily_action-sign_language-walk, daily_action-hand_gesture-Make_a_phone_call, daily_action-sign_language-help, daily_action-social_action-box, daily_action-sign_language-before, daily_action-social_action-celebrate-siu, sports-dance-street_jazz, daily_action-sign_language-many, daily_action-sign_language-white, sports-basketball-free_throw, entertainment-billiard_hit, daily_action-phylogenetic_action-gives_birth_to_a_baby, sports-skiing, daily_action-sign_language-computer, daily_action-housework-plunging_toilet, daily_action-commuting-riding_elephant, daily_action-social_action-hug, sports-soccer-shoot, sports-volleyball-adjust, daily_action-sign_language-deaf, daily_action-chinese traffic_gesture-MOVE_STRAIGHT, daily_action-phylogenetic_action-hunts_other_animals, daily_action-vehicle_behavior-Crossing, daily_action-social_action-be_pated, daily_action-social_action-kick, sports-gymnastics-unevenbar_cross, sports-gym-gym_push, daily_action-sign_language-bed, sports-baseball-bunt, sports-gymnastics-jump, daily_action-housework-hanging_clothes, daily_action-commuting-ride_bicycle, entertainment-shoot_gun, sports-volleyball-pass, entertainment-read, daily_action-emergency-rescue_breathing, special-self_replication, sports-gym-squat, entertainment-performance-fire_breathing, entertainment-performance-playing_rubiks_cube, daily_action-sign_language-thanksgiving, daily_action-sign_language-shirt, entertainment-performance-handstand, daily_action-chinese traffic_gesture-PULL_OVER, daily_action-housework-shoe_shining, daily_action-vehicle_behavior-Confrontation, sports-dance-waack, daily_action-hand_gesture-Walk, daily_action-social_action-be_pushed, sports-baseball-pitch, daily_action-hand_gesture-Wave_finger, sports-baseball-catch_catcher, daily_action-vehicle_behavior-Merging, daily_action-sign_language-bowling, sports-swimming-backstroke, sports-gym-situp

Table X. Motion categories in folder-3 and folder-4 of the overlapping split.

Folder	Motions
Folder-3	daily_action-sign_language-tall, sports-swimming-butterfly_stroke, sports-basketball-spinning_basketball, entertainment-performance-play_yoyo, daily_action-chinese traffic_gesture-LEFT_TURN_WAITING, daily_action-sign_language-orange, entertainment-performance-sword_swallowing, daily_action-social_action-be_kicked, daily_action-sign_language-go, daily_action-chinese traffic_gesture-SLOW_DOWN, sports-basketball-layup, entertainment-musical_instrument-play_jazzdrum, sports-basketball-drive, entertainment-performance-shake_cocktail, sports-basketball-jabstep, daily_action-sign_language-candy, daily_action-housework-watering_plants, daily_action-hand_gesture-Dual_hands_heart, entertainment-photography, sports-yoga-yoga_firefly, special-transform-reversion, daily_action-housework-washing, sports-gym-hulahoop, sports-dance-krump, daily_action-sign_language-fire, daily_action-commuting-push_car, entertainment-discuss_throw, daily_action-social_action-be_boxed, sports-gymnastics-balancebeam_rotate, daily_action-social_action-wave, sports-basketball-hookshot, sports-yoga-yoga_tree, daily_action-sign_language-who, daily_action-sign_language-cousin, daily_action-hand_gesture-Measure_distance, daily_action-housework-carpentry, daily_action-social_action-laugh, sports-run_starting, daily_action-social_action-handshake, sports-yoga-yoga_dancer, daily_action-hand_gesture-knock, sports-gym-dips, sports-gym-burpee, daily_action-sign_language-pizza, sports-dance-pop, daily_action-vehicle_behavior-Right_Turn, daily_action-sign_language-drink, sports-dance-middle_hip_hop, daily_action-vehicle_behavior-Overtake, sports-surfing, entertainment-musical_instrument-play_bamboon_flute, entertainment-smoke, special-magic, sports-golf, daily_action-sign_language-no, daily_action-vehicle_behavior-Left_Turn
Folder-4	sports-gymnastics-flip, entertainment-dart_throw, daily_action-hand_gesture-Wave_hand, sports-gym-pull, daily_action-social_action-highfive, sports-dance-la_style_hip_hop, daily_action-hand_gesture-Sweep_Cross, daily_action-sign_language-like, sports-dance-moonwalk, sports-gym-plank, sports-gym-pull_up, daily_action-housework-ironing, sports-soccer-diving, daily_action-hand_gesture-Applaud, daily_action-sign_language-study, sports-volleyball-underhand, daily_action-emergency-CPR, entertainment-stone_skipping, sports-yoga-yoga_pigeon, daily_action-sign_language-woman, daily_action-social_action-push, daily_action-emergency-heimlich_maneuver, sports-basketball-shoot, sports-basketball-dunk, sports-gym-pushup, daily_action-housework-chopping, sports-basketball-pass, daily_action-chinese traffic_gesture-LEFT_TURN, daily_action-social_action-blowing_kisses, daily_action-sign_language-finish, daily_action-vehicle_behavior-Cut_in, daily_action-commuting-ride_bike, entertainment-musical_instrument-play_guitar, daily_action-housework-knitting, sports-soccer-throw, sports-swimming-breast_stroke, sports-dance-break, daily_action-hand_gesture-Pause, daily_action-housework-sweeping_floor, daily_action-sign_language-thin, sports-gym-lunges, daily_action-emergency-fire_extinguisher, entertainment-musical_instrument-play_saxophone, daily_action-commuting-ride_horse, daily_action-hand_gesture-Sweep_diagonal, special-transform-metamorphosis, daily_action-phylogenetic_action-drinking, daily_action-housework-wall_painting, entertainment-spraying_wall, sports-gymnastics-spin, entertainment-musical_instrument-play_piano, daily_action-commuting-ride_tricycle, daily_action-housework-cleaning_glass_surface, special-teleport, sports-soccer-steal, entertainment-musical_instrument-play_erhu

Table XI. Motion categories in the folder-1 and folder-2 of the non-overlapping split.

Group	Motions
Folder-1	daily_action-hand_gesture-Make_a_phone_call, daily_action-hand_gesture-Walk, daily_action-hand_gesture-Sweep_Circle, daily_action-hand_gesture-knock, daily_action-hand_gesture-Wave_hand, daily_action-hand_gesture-Zoom_out_with_fingers, daily_action-hand_gesture-Wave_finger, daily_action-hand_gesture-Applaud, daily_action-hand_gesture-Sweep_Cross, daily_action-hand_gesture-Measure_distance, daily_action-hand_gesture-Zoom_in_with_fingers, daily_action-hand_gesture-Sweep_Checkmark, daily_action-hand_gesture-Sweep_diagonal, daily_action-hand_gesture-Dual_hands_heart, daily_action-hand_gesture-Pause, daily_action-sign_language-go, daily_action-sign_language-white, daily_action-sign_language-who, daily_action-sign_language-many, daily_action-sign_language-pizza, daily_action-sign_language-study, daily_action-sign_language-woman, daily_action-sign_language-bed, daily_action-sign_language-cousin, daily_action-sign_language-black, daily_action-sign_language-hot, daily_action-sign_language-before, daily_action-sign_language-fine, daily_action-sign_language-help, daily_action-sign_language-like, daily_action-sign_language-orange, daily_action-sign_language-thanksgiving, daily_action-sign_language-fire, daily_action-sign_language-no, daily_action-sign_language-cool, daily_action-sign_language-computer, daily_action-sign_language-shirt, daily_action-sign_language-mother, daily_action-sign_language-walk, daily_action-sign_language-finish, daily_action-sign_language-bowling, daily_action-sign_language-drink, daily_action-sign_language-thin, daily_action-sign_language-deaf, daily_action-sign_language-yes, daily_action-sign_language-candy, daily_action-sign_language-wrong, daily_action-sign_language-tall, daily_action-chinese traffic_gesture-LANE_CHANGING, daily_action-chinese traffic_gesture-RIGHT_TURN, daily_action-chinese traffic_gesture-PULL_OVER, daily_action-chinese traffic_gesture-LEFT_TURN_WAITING, daily_action-chinese traffic_gesture-MOVE_STRAIGHT, daily_action-chinese traffic_gesture-LEFT_TURN, daily_action-chinese traffic_gesture-SLOW_DOWN, daily_action-chinese traffic_gesture-STOP
Folder-2	daily_action-commuting-ride_horse, daily_action-commuting-ride_bike, daily_action-commuting-ride_motorcycle, daily_action-commuting-push_car, daily_action-commuting-riding_elephant, daily_action-commuting-segway, daily_action-commuting-ride_bicycle, daily_action-commuting-ride_tricycle, daily_action-commuting-ride_scooter, daily_action-vehicle_behavior-Confrontation, daily_action-vehicle_behavior-Overtake, daily_action-vehicle_behavior-Cut_in, daily_action-vehicle_behavior-Deviate, daily_action-vehicle_behavior-Crossing, daily_action-vehicle_behavior-Right_Turn, daily_action-vehicle_behavior-U_Turn, daily_action-vehicle_behavior-Left_Turn, daily_action-vehicle_behavior-Merging, daily_action-social_action-pat, daily_action-social_action-be_pated, daily_action-social_action-hug, daily_action-social_action-laugh, daily_action-social_action-wave, daily_action-social_action-celebrate-siu, daily_action-social_action-highfive, daily_action-social_action-handshake, daily_action-social_action-celebrate-night_night, daily_action-social_action-push, daily_action-social_action-be_pushed, daily_action-social_action-kick, daily_action-social_action-be_kicked, daily_action-social_action-bend, daily_action-social_action-be_bended, daily_action-social_action-be_boxed, daily_action-social_action-box, daily_action-social_action-celebrate-slippery_kneeling, daily_action-social_action-blowing_kisses, daily_action-social_action-salute, daily_action-housework-hanging_clothes, daily_action-housework-plunging_toilet, daily_action-housework-cleaning_glass_surface, daily_action-housework-watering_plants, daily_action-housework-carpentry, daily_action-housework-washing, daily_action-housework-wall_painting, daily_action-housework-knitting, daily_action-housework-ironing, daily_action-housework-sweeping_floor, daily_action-housework-shoe_shining, daily_action-housework-chopping, daily_action-phylogenetic_action-gives_birth_to_a_baby, daily_action-phylogenetic_action-drinking, daily_action-phylogenetic_action-hunts_other_animals, daily_action-phylogenetic_action-sleep, daily_action-phylogenetic_action-eat_cake, daily_action-emergency-heimlich_maneuver, daily_action-emergency-rescue_breathing, daily_action-emergency-bandaging, daily_action-emergency-fire_extinguisher, daily_action-emergency-CPR

Table XII. Motion categories in the folder-3 and folder-4 of the nonoverlapping split.

Group	Motions
Folder-3	sports-gym-situp, sports-gym-pushup, sports-gym-dips, sports-gym-burpee, sports-gym-pull, sports-gym-hulahoop, sports-gym-squat, sports-gym-gym_push, sports-gym-pull_up, sports-gym-lunges, sports-gym-frog_jump, sports-gym-plank, sports-gym-lift, sports-basketball-layup, sports-basketball-pass, sports-basketball-drive, sports-basketball-shoot, sports-basketball-interfere_shot, sports-basketball-spinning_basketball, sports-basketball-screen, sports-basketball-jabstep, sports-basketball-dunk, sports-basketball-free_throw, sports-basketball-hookshot, sports-yoga-yoga_firefly, sports-yoga-yoga_gate, sports-yoga-yoga_dancer, sports-yoga-yoga_pigeon, sports-yoga-yoga_cat, sports-yoga-yoga_tree, sports-baseball-bunt, sports-baseball-swing, sports-baseball-catch_catcher, sports-baseball-pitch, sports-gymnastics-flip, sports-gymnastics-balancebeam_rotate, sports-gymnastics-unevenbar_cross, sports-gymnastics-jump, sports-gymnastics-spin, sports-gymnastics-balancebeam_walk, sports-volleyball-pass, sports-volleyball-adjust, sports-volleyball-underhand, sports-volleyball-overhand, sports-soccer-throw, sports-soccer-steal, sports-soccer-shoot, sports-soccer-diving, sports-soccer-header, sports-swimming-butterfly_stroke, sports-swimming-breast_stroke, sports-swimming-backstroke, sports-swimming-freestyle
Folder-4	sports-dance-pop, sports-dance-middle_hip_hop, sports-dance-house, sports-dance-la_style_hip_hop, sports-dance-moonwalk, sports-dance-ballet_jazz, sports-dance-lock, sports-dance-waack, sports-dance-street_jazz, sports-dance-krump, sports-dance-break, sports-taichi_fan, sports-run_starting, sports-surfing, sports-golf, sports-skiing, sports-punch, special-transform-reversion, special-self_replication, special-transform-metamorphosis, special-magic, special-teleport, special-transform-transformation, entertainment-musical_instrument-play_saxophone, entertainment-musical_instrument-play_erhu, entertainment-performance-play_yoyo, entertainment-performance-playing_rubiks_cube, entertainment-musical_instrument-play_guitar, entertainment-smoke, entertainment-musical_instrument-play_pipa, entertainment-musical_instrument-play_bamboon_flute, entertainment-stone_skipping, entertainment-performance-fire_breathing, entertainment-performance-sword_swallowing, entertainment-musical_instrument-play_cello, entertainment-shoot_gun, entertainment-musical_instrument-play_suona, entertainment-photography, entertainment-dart_throw, entertainment-musical_instrument-play_jazzdrum, entertainment-read, entertainment-performance-shake_cocktail, entertainment-performance-dice_stack_shuffle, entertainment-musical_instrument-play_grandpiano, entertainment-musical_instrument-play_violin, entertainment-billiard_hit, entertainment-performance-handstand, entertainment-draw_handgun, entertainment-discuss_throw, entertainment-performance-spinning_plate, entertainment-performance-card_throw, entertainment-musical_instrument-play_piano, entertainment-spraying_wall, entertainment-building_snowman, entertainment-musical_instrument-play_trumpet

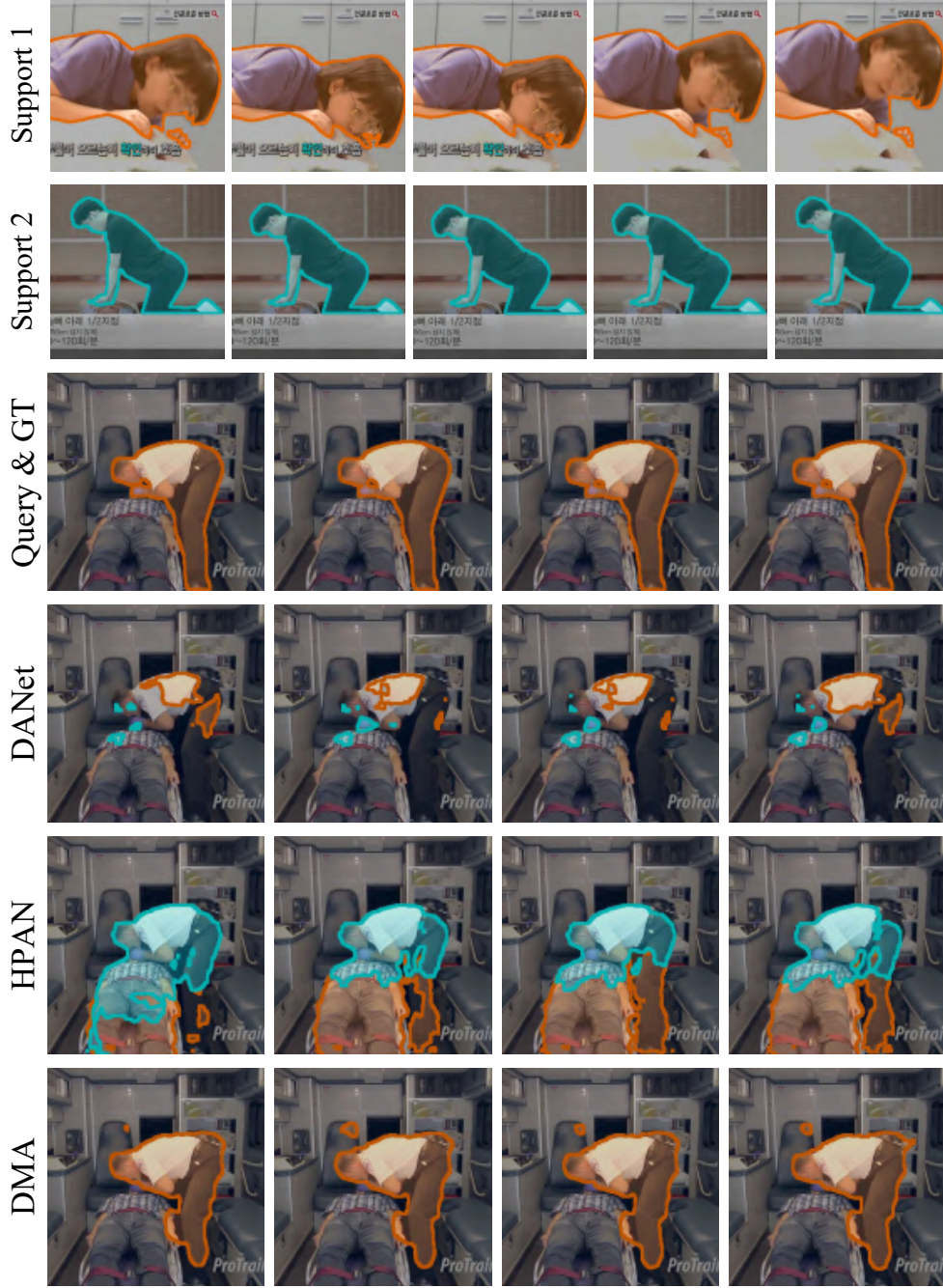## B. Qualitative Results and Failure Cases

Figure III. Qualitative comparison between existing models and DMA.

In this section, we provide some supplement details of qualitative experiment and failure analysis. As shown in Figure III, DMA correctly segments the person performing the motion **rescue breathing**. This comparison demonstrates the model's solid performance against significant differences in spatial position and perspective between support and query pairs. DMA performs particularly well when the support instances act as the active party in interaction scenes.

However, our model also exhibits certain limitations. In the case shown in Figure IV, while our proposed DMA successfully discriminates the query instance's motion category in a fine-grained support pair, it fails to accurately mark
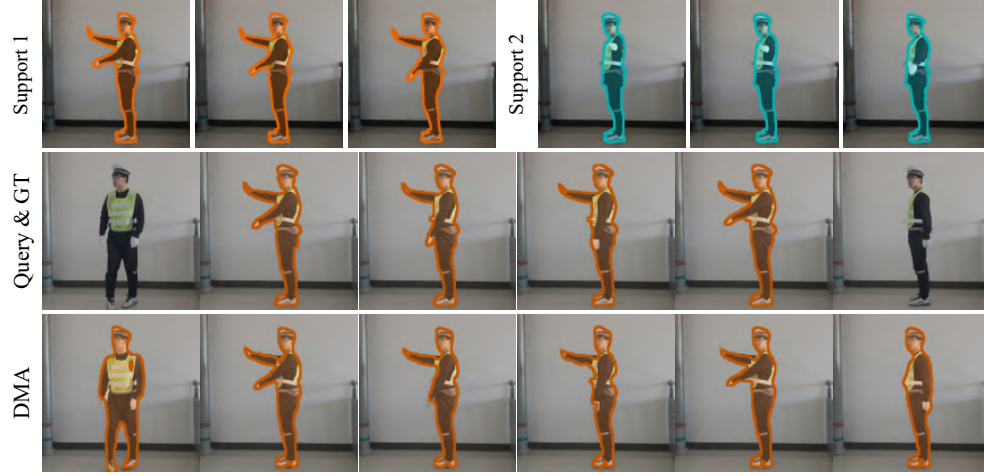
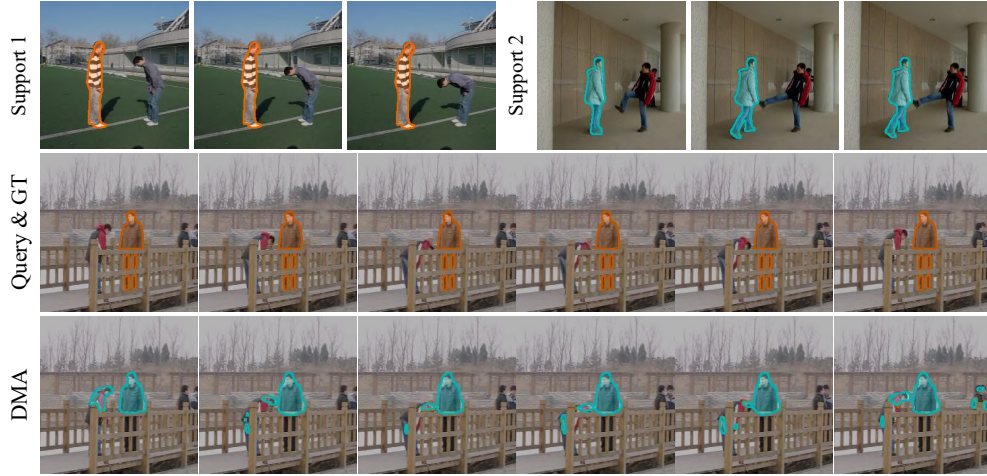Figure IV. Failure case for fine-grained motion with boundaries.

Figure V. Failure case for passive action.

the motion duration. Furthermore, Figure V reveals that DMA lacks sufficient capacity to deeply comprehend detailed interactions and complex relations between different instances. It is challenging for DMA to distinguish between passive actions such as **be bended** and **be kicked**. This limitation stems from DMA's high concentration on motion information solely within the masked region, causing it to ignore the motions of background objects. Future work could address these limitations by incorporating temporal boundary refinement mechanisms and developing more sophisticated attention mechanisms that better capture interactions between foreground and background objects.

References

[1] Jun Chen, Ming Hu, Darren J Coker, Michael L Berumen, Blair Costelloe, Sara Beery, Anna Rohrbach, and Mohamed Elhoseiny. Mammalnet: A large-scale video benchmark for mammal recognition and behavior understanding. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 4

[2] Jihoon Chung, Cheng-hsin Wuu, Hsuan-ru Yang, Yu-Wing Tai, and Chi-Keung Tang. Haa500: Human-centric atomic action dataset with curated videos. In *Int. Conf. Comput. Vis.*, 2021. 2, 3, 4, 5, 6

[3] Hongdong Li Dongxu Li. Wlasl (world level american sign language) video, 2022. 4

[4] HE Jian and WANG Weidong. Visual recognition of chinese traffic police gestures based on spatial context and temporal features. *Acta Electronica Sinica*, 2020. 5

[5] Yu Kong, Yunde Jia, and Yun Fu. Learning human interaction by interactive phrases. In *Eur. Conf. Comput. Vis.*, 2012. 5

[6] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. Hmdb: a large video database for human motion recognition. In *Int. Conf. Comput. Vis.*, 2011. 2, 3, 5, 6

[7] Ruilong Li, Shan Yang, David A Ross, and Angjoo Kanazawa. Ai choreographer: Music conditioned 3d dance generation with aist++. In *Int. Conf. Comput. Vis.*, 2021. 2

[8] Yixuan Li, Lei Chen, Runyu He, Zhenzhi Wang, Gangshan Wu, and Limin Wang. Multisports: A multi-person video dataset of spatio-temporally localized sports actions. In *Int. Conf. Comput. Vis.*, 2021. 2

[9] Chirag Parikh, Rohit Saluja, CV Jawahar, and Ravi Kiran Sarvadevabhatla. Idd-x: A multi-view dataset for ego-relative important object localization and explanation in dense and unstructured traffic. In *IEEE Int. Conf. Robot. Autom.*, 2024. 6

[10] Yifan Zhang, Congqi Cao, Jian Cheng, and Hanqing Lu. Egogesture: a new dataset and benchmark for egocentric hand gesture recognition. *IEEE Trans. Multimedia*, 2018. 4