

OpenRoboCare: A Multimodal Multi-Task Expert Demonstration Dataset for Robot Caregiving

¹Xiaoyu Liang, ¹Ziang Liu, ³Kelvin Lin*, ¹Edward Gu*, ¹Ruolin Ye, ⁴Tam Nguyen
²Cynthia Hsu, ¹Zhanxin Wu, ¹Xiaoman Yang, ¹Christy Sum Yu Cheung
³Harold Soh, ²Katherine Dimitropoulou, ¹Tapomayukh Bhattacharjee

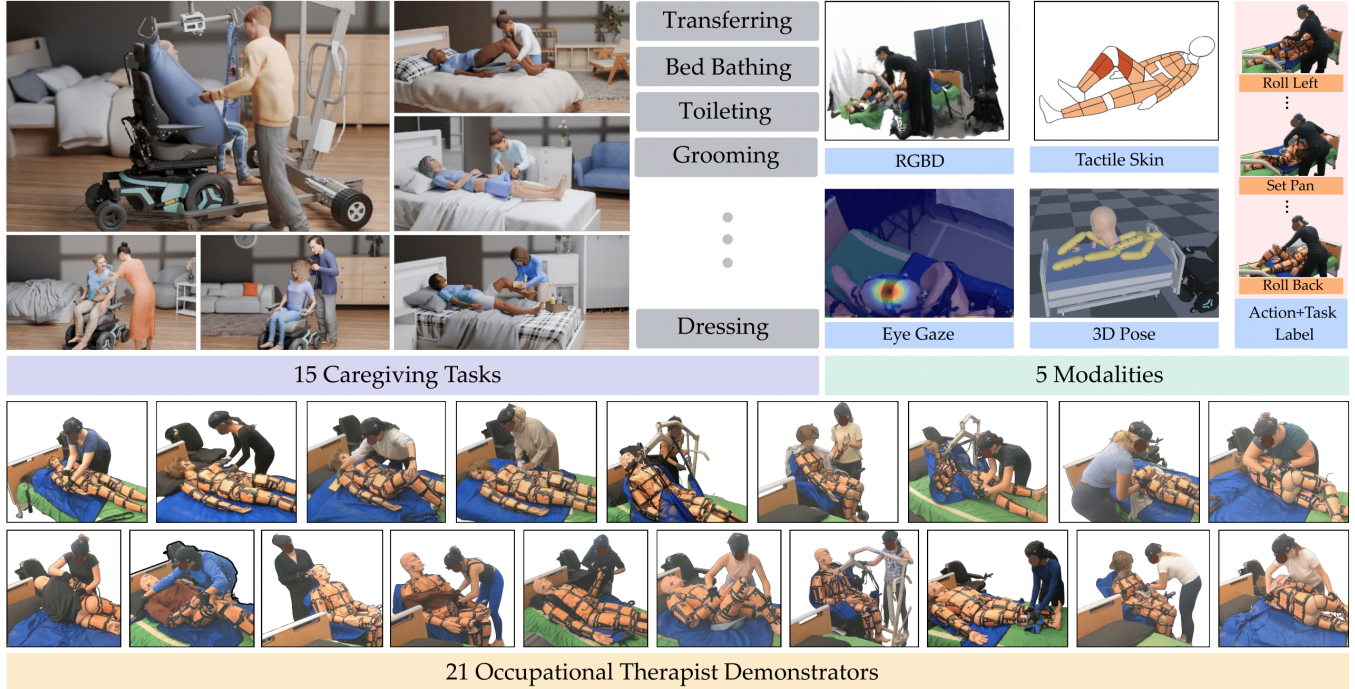


Fig. 1: Overview of OpenRoboCare dataset for robot caregiving, featuring 21 occupational therapists demonstrating 15 common caregiving tasks, captured across 5 data modalities. It consists of 315 sessions, totaling 19.8 hours, with a collection of 31,185 samples.

Abstract—We present OpenRoboCare, a multimodal dataset for robot caregiving, capturing expert occupational therapist demonstrations of Activities of Daily Living (ADLs). Caregiving tasks involve complex physical human-robot interactions, requiring precise perception under occlusions, safe physical contact, and long-horizon planning. While recent advances in robot learning from demonstrations have shown promise, there is a lack of a large-scale, diverse, and expert-driven dataset that captures real-world caregiving routines. To address this gap, we collect data from 21 occupational therapists performing 15 ADL tasks on two manikins. The dataset spans five modalities—RGB-D video, pose tracking, eye-gaze tracking, task and action annotations, and tactile sensing, providing rich multimodal insights into caregiver movement, attention, force application, and task execution strategies. We further analyze expert caregiving principles and strategies, offering insights to improve robot efficiency and task feasibility. Additionally, our evaluations demonstrate that OpenRoboCare presents challenges for state-of-the-art robot perception and human activity recognition methods, both critical for developing

safe and adaptive assistive robots, highlighting the value of our contribution. See our website for additional visualizations: <https://emprise.cs.cornell.edu/robo-care/>.

I. INTRODUCTION

According to the World Health Organization [7], approximately 1.3 billion people live with significant physical limitations, many of whom require assistance with Activities of Daily Living (ADLs) [8]. However, the demand for qualified caregivers and therapists far exceeds the number of available trained professionals [9]. Assistive robotics has potential to support the caregiving process and address these issues to some extent. Recent work in *robot caregiving* in feeding [10–13], dressing [14, 15], bathing [16, 17], and transferring [18] among other ADLs presents tremendous promise.

Despite these advancements, robot caregiving still faces significant technical challenges. For example, consider the requirements for assisted bed bathing [16, 17]: accurate perception of human state, often under occlusions; bimanual and mobile manipulation of human limbs with critical safety constraints; long-horizon planning under uncertainty; personalization to users’ physical function [19] and preferences [20]; and adaptation in response to human feedback.

To address these challenges, recent work has considered learning-based approaches for specific caregiving tasks [21–

¹Cornell University, Ithaca, NY, USA. ²Columbia University, New York City, NY, USA. ³National University of Singapore, Singapore. ⁴University of Massachusetts Lowell, Lowell, MA, USA

[†]This work was partly funded by NSF IIS #2132846, CAREER #2238792, and a Cornell–NUS Global Strategic Collaboration Award. We thank Yapeng Teng for data labeling tool setup. We thank Tasbolat Taunyazov, Crystal Liu, Moustafa Kassem, Xia Yan Zhao, Shannon Liu, Luke Kulm, Tasmin Sangha, Mandy Chen, Gavin Chen, Nancy Davila Belendez, Max Tweedale, Allyson Daos, Emily Avisado for their efforts in data processing and labeling.

TABLE I: Comparison of OpenRoboCare with existing datasets.

Dataset	Context			Statistics				Data Modality					
	Care. ¹	Pop. ²	Demo. ³	#Subj. ⁴	#Tasks (ADLs)	#Hrs	Samples ⁵	IMU	Video	Pose CG CR ⁶	Gaze	Tactile	Act. ⁷
Bagewadi et al. [1]	✗	TD	User	33	1(0)	2.94	485	✓	RGB	M (- ✓)	✗	Partial	A,T
SBU Kinect [2]	✗	TD	User	7	8(0)	-	-	✗	RGB-D	V (- ✓)	✗	✗	T
TacAct [3]	✗	TD	-	50	12(0)	-	-	✗	✗	✗	✗	Partial	A
SONAR [4]	✓	E	Expert	14	23(5)	37.3*	36006	✓	✗	M (✓ ✗)	✗	✗	T
Kaczmarek et al. [5]	✓	ML	Expert	7	9(1)	0.7	132	✓	RGB	✗ ✗	✗	✗	T
HARMONIC [6]	✓	TD	User	24	1(1)	5	600	✓	RGB	V (- ✓)	✓	✗	T
OpenRoboCare (ours)	✓	SML	Expert	21	15(5)	19.8	31185	✗	RGB-D	M,V (✓ ✓)	✓	Full	A,T

¹Caregiving. ²Target population. (TD: typical developing. E: elderly. ML: mobility limitation. SML: severe mobility limitation.) ³Demonstration type. ⁴Number of subjects. ⁵Total number of samples (modalities $\times$ tasks $\times$ subjects $\times$ time in hours) ⁶Pose data. (M: motion capture, includes both optical and inertial. V: pose calculated from videos. CG: caregiver pose. CR: care recipient or user pose.) ⁷Action annotation. (T: task-level. A: action-level.) *Re-calculated after removing non-documented activities.

23]. These prior works collected their own datasets, typically in simulation, that are tailored to the task of interest. This *task-specific* data collection is in contrast to recent advances in general robot learning where large diverse datasets are used to train foundation models [24, 25]. Existing caregiving datasets are also limited in *data modalities*, often featuring vision [23] or haptics [5] but not both. Furthermore, with a few notable exceptions [4, 5], existing caregiving datasets do not feature data collected from *expert* human caregivers or occupational therapists. These datasets are therefore lacking the extensive practical knowledge accumulated by experts through years of experience and training.

To bridge this gap, we present **OpenRoboCare** (Fig. 1), the first *multi-task, multimodal, and expert-collected* dataset for robot caregiving. OpenRoboCare features expert demonstrations from 21 occupational therapists (OTs) in 15 distinct ADLs with data captured from five modalities: RGB-D video, tactile sensing, pose tracking, eye-gaze tracking, and action annotations. We provide tactile sensing with a custom whole-body piezo-resistive skin, pose tracking with motion capture, eye-tracking with Pupil Labs glasses, and action labeling via natural language during data collection. The OTs demonstrated their assistance on two manikins with different genders and body weights. Our data collection protocol was designed in collaboration with an expert occupational therapist and co-author, ensuring that the task procedures and setup closely resemble real-world caregiving scenarios.

In addition to releasing OpenRoboCare as an *open-source dataset*, we conduct a comprehensive quantitative analysis to inform task execution and physical interactions with care recipients in robot caregiving. In collaboration with an expert OT, we distill guiding principles in OT practice and identify specific techniques that exemplify these principles across tasks. We present OpenRoboCare as a benchmark for robot perception and human activity recognition in caregiving scenarios. While state-of-the-art methods perform poorly out of the box, fine-tuning on a small subset of OpenRoboCare leads to significant performance gains. Overall, our findings highlight the richness and complexity of expert caregiving strategies, positioning OpenRoboCare as a critical resource for advancing multimodal learning in robot caregiving.

II. RELATED WORK

Caregiving Datasets Previous works in rehabilitation and public health collected survey and interview-based data on

caregiving, e.g., for older adults [26–28] and those with spinal cord injuries [29]. These efforts focus on the health, social, and financial aspects of caregiving, rather than the physical caregiving process. Works close to ours either did not collect data from expert caregivers [6], or lacked multimodality [4, 5]. OpenRoboCare is the first dataset for multi-task, multimodal, expert caregiving (Table I).

Physical HHI & HRI Datasets Our work is also related to previous efforts in human-human and human-robot interaction that collected multimodal and multi-task data, e.g., for activity recognition [30, 31]. For example, Bagewadi et al. [1] collected data of human-robot hugging interactions using wearable sensors. The SBU Kinect Interaction [2] human-human dataset similarly considered hugging among other physical activities such as kicking and punching. TacAct [3] collected high-fidelity tactile data of a human touching a robot arm. We similarly collect high-fidelity tactile data, but placed sensors on the human body rather than the robot arm.

Human Activities Datasets Our work is also more broadly related to the literature on human activity recognition [32]. Most related are datasets for activity recognition that are situated in homes and hospitals [33–36]. In experiments, we evaluate a state-of-the-art method [37] for activity recognition on OpenRoboCare and show that our tasks present significant challenges and can drive further progress in the field.

III. TASK SELECTION AND CAREGIVING PROTOCOL

Our goal is to collect a multimodal, multi-task, expert-driven dataset that facilitates robot caregiving research. In this section, we describe the included tasks and the protocol for expert caregivers completing the tasks. This data collection protocol was approved by the Cornell IRB.

A. OT-in-the-loop Protocol Design

We selected tasks and designed our data collection in close consultation with an expert OT collaborator who also has extensive experience in OT education. Our setup follows standard clinical guidelines for training OTs [38, 39]. In designing the setup, we especially consider individuals with quadriplegia, a significant sensorimotor impairment that results in a lack of control and movement of the upper limbs, trunk, lower limbs, and pelvic organs and requires complete assistance with basic ADLs [40].

B. Task Selection

We consider five of the six basic Activities of Daily Living (ADLs): bathing, toileting, dressing, transferring, and grooming. Feeding is excluded as it requires substantially different caregiving skills and has been extensively studied [41, 42]. Within the five basic ADLs, we consider 15 tasks (Fig. 2): one each from bathing, toileting, and grooming; two from transferring; and 10 from dressing capturing diverse garment types and scenarios. We develop caregiving protocols that mirror real-life routines of care recipients with quadriplegia. We next describe the ADLs and tasks in detail, highlighting key aspects relevant to robot caregiving.

Bathing Caregivers perform a full-body sponge bath on a manikin lying on a hospital bed. They are instructed to pat the skin gently, as they would with a care recipient to minimize the risk of discomfort or injury. Key aspects of interest include the amount of force applied, techniques for cleaning hard-to-reach areas, and strategies for adjusting the manikin's position to access its back.

Toileting Caregivers assist a manikin with toileting using a bedpan while it is lying on a hospital bed, a common method for individuals with limited mobility or a high risk of injury that prevents the use of a regular toilet. The task requires caregivers to lift the manikin's hips to position the bedpan underneath and subsequently remove it for emptying. Key aspects of interest include the techniques used to lift and stabilize the manikin's hips when handling manikins of varying weights, as well as the hand placements and stabilization strategies employed.

Dressing Caregivers dress or undress a manikin. Dressing requires different strategy sequences and safety considerations for different body segments (upper/lower body), garment types, supporting surfaces (bed/wheelchair), and body position (lying down/sitting). To capture these variations, caregivers dress and undress the manikin in t-shirts, vests, and shorts while the manikin is lying on a bed or sitting in a wheelchair. In total, we define 10 tasks (see Fig. 2). Key aspects of interest include the ways caregivers prepare clothing, coordinate bimanual movements, and handle care recipients' body parts.

Transferring Caregivers transfer a manikin between a bed and a wheelchair using a Hoyer sling with a mechanical lift (see Fig. 2). We consider bed-to-wheelchair and wheelchair-to-bed transferring as two separate tasks. Key aspects of interest include the ways in which caregivers use the Hoyer sling, secure the sling properly, operate the lift, guide the manikin's position during transfer, and safely release the manikin after the transfer.

Grooming Caregivers brush the manikin's hair while it is seated in a wheelchair. Key aspects of interest include caregiver hand coordination, posture adjustments, and control strategies during the grooming procedure.

C. Caregiver Protocol

Data collection spanned two weeks and involved 21 OTs. Each participant performed 15 tasks on a male or female manikin, one trial per task, taking approximately one hour total. Before data collection, participants completed IRB consent forms and demographic surveys. Upon arrival, they

received a detailed briefing covering the study objectives, task instructions, equipment usage, and data collection procedures. Each caregiver was then fitted with motion capture gloves, a motion capture hat, and eye-tracking glasses. We calibrated the motion capture system and the eye-tracking glasses with an OT before each data collection session.

IV. DATA COLLECTION SETUP

In this section, we describe the environment and data recording setup. Data collection took place in an enclosed space measuring $3.68\text{m} \times 3.68\text{m}$, designed to simulate a realistic in-home caregiving setting. The hospital bed, wheelchair, and Hoyer sling were reset to the same initial position throughout the sessions. See Fig. 2 for an overview.

A. Hospital Manikins

In consultation with our OT expert, we select two hospital manikins. One male (Rescue Randy, 150 lbs, 6 ft 1 in) and one female (Simple Susie, 37.26 lbs, 5 ft 5 in), both with anthropomorphic dimensions and joints. Manikins are frequently used in OT clinical training to simulate real-life conditions [43]. They allow us to standardize caregiving tasks and collect reliable, repeatable data without inconveniencing real patients. While manikins lack partial agency or resistance, they effectively represent passive full-assistance scenarios that are common in OT practice. The strategies demonstrated by expert caregivers offer valuable insights for adapting to interactions with partially active patients.

B. Assistive Devices

The data collection setup involves various assistive devices to replicate realistic caregiving environments. The hospital bed (Invacare ETUDE HC Hi-Lo) and electric wheelchair (ROVI X3) are positioned next to each other in a fixed arrangement. A Hoyer sling (Invacare 9805P) is also included for transferring tasks. For specific caregiving tasks, assistive devices are provided: a bathing sponge for bathing, a bedpan for toileting, and a brush for grooming.

C. Sensing Modalities

RGB-D Videos We use three Intel RealSense D435i RGB-D cameras positioned around the scene to capture visual and geometric data of the caregiver's movements, interactions with the manikin and assistive devices, and the resulting manikin motion. Two cameras are placed at different angles facing the hospital bed, while a third camera is positioned behind the bed facing the wheelchair.

Tactile Skin We develop a custom tactile skin to fit the manikins and record physical interactions between the caregiver and the manikin. The sensor design is guided by three key considerations: (1) *customizability* to accommodate various manikin body shapes and sizes, (2) *flexibility* to ensure secure attachment to curved surfaces, and (3) *durability* to withstand pressure exerted by the manikin's weight.

Hardware Design: We design resistive tactile sensors using specialized fabrics that are lightweight, flexible, and durable (Fig. 3). Each sensor consists of a pressure-sensitive Velostat layer, sandwiched between two copper conductive fabric layers and secured with non-conductive electrical tape. The sensor's resistance decreases as force is applied, enabling force measurement. A voltage divider circuit converts resistance

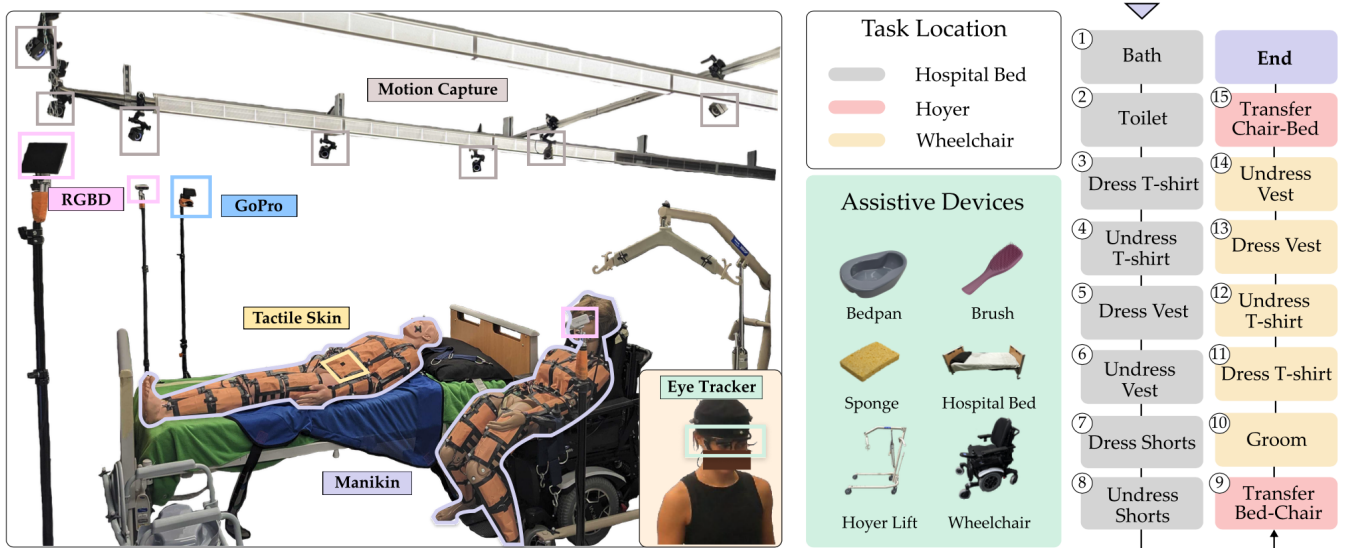


Fig. 2: Data collection setup and procedure. **Left:** setup of sensors and equipment. **Center:** assistive devices used by caregivers. **Right:** sequence of tasks performed by each caregiver.

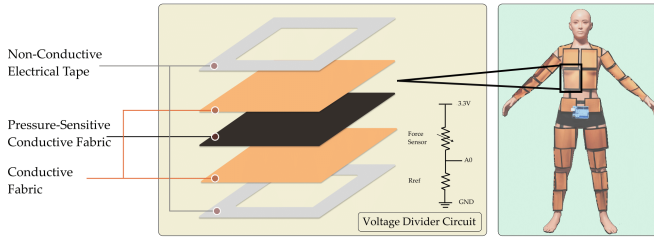


Fig. 3: Tactile skin design and layout of sensors on manikin.

changes into analog voltage signals, which are processed by an Arduino Uno. The sensor exhibits a nonlinear but stable voltage response across the 0.05 to 3 N/cm² pressure range. It has low hysteresis observed at forces below 5N and a maximum hysteresis error of 7.09% at higher loads. See website [44] for details.

Sensor Placement: A total of 88 resistive sensors are developed, with 44 sensors placed on each manikin (Fig. 3). The sensors are evenly distributed across the manikin’s body: 7 on each arm, 8 on each leg, and 14 across the front and back of the torso, aiming to maximize coverage. Each sensor covers an average area of 50 square inches, with gaps between adjacent sensors kept under 1 mm.

Calibration and Processing: Prior to data collection, each tactile sensor is calibrated using an ATI force/torque (F/T) sensor to accurately map voltage readings to force values. Before each task, the tactile skin is tared to eliminate any baseline offset, ensuring consistent force measurements.

Pose Tracking We use a motion capture system equipped with 12 OptiTrack PrimeX 13 cameras to track the movements of both the manikin and the caregiver. The caregiver wears a hat and gloves with motion capture markers to accurately track hand and head movements. For manikin pose tracking, we employ rigid body marker sets to define each body segment.

Occlusions pose a fundamental challenge in real-world caregiving and are critical for robotic systems to overcome. In our dataset, occlusions caused by clothing in dressing tasks and slings in transferring tasks often led to tracking

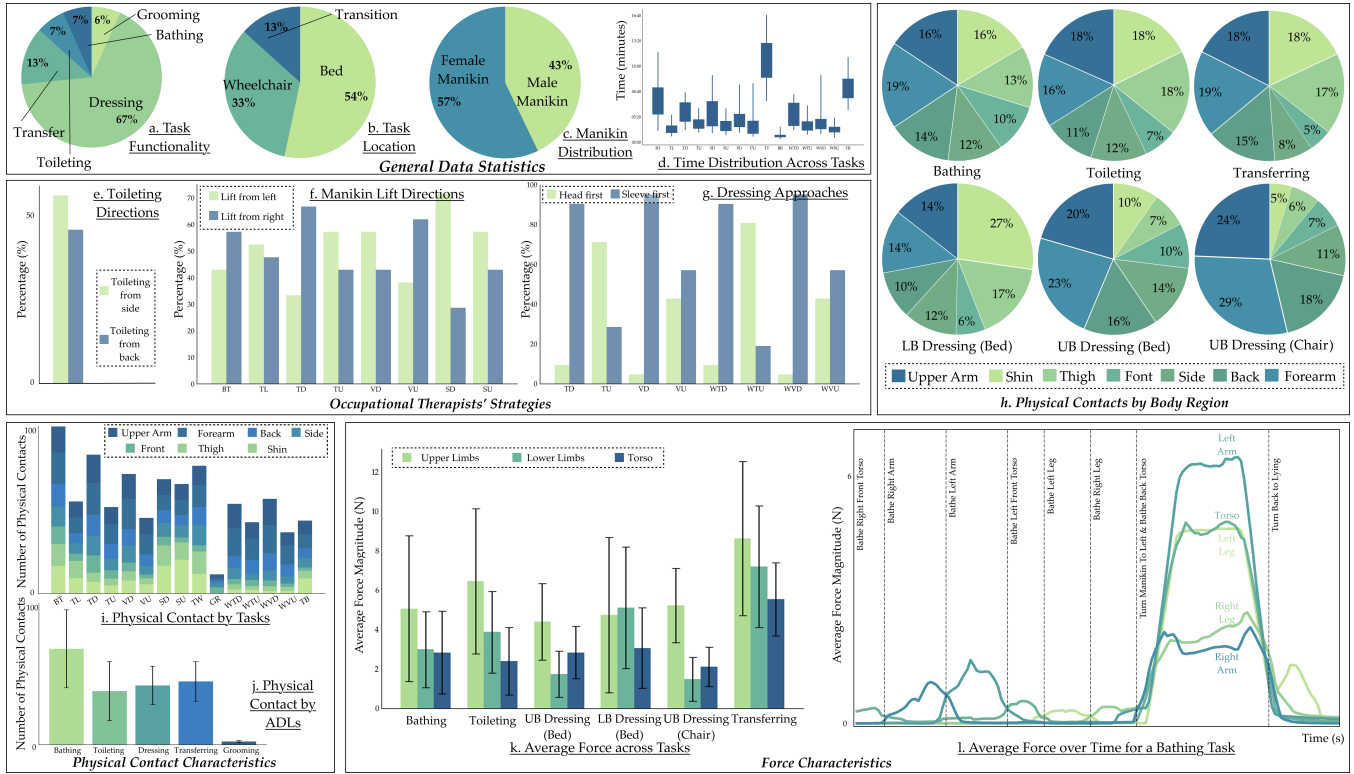
failures for both the manikin’s pose and the caregiver’s hand positions. To address this, we manually labeled body keypoints for a subset of the dataset using RGB images from three calibrated cameras, originally used for RGB-D video capture (details in Section IV-C). These annotations were then used to train a YOLOv11 [45] pose detector, which automatically labeled the remaining data with little human supervision. We estimated 3D positions by averaging triangulated results from all camera pairs. While limited to three views, this approach provides a practical and scalable solution for occlusion handling.

Eye Tracking To analyze caregiver visual attention during tasks, we equip participants with Pupil Labs eye-tracking glasses to capture first-person video and 2D gaze data. We use 3D pose tracking of the caregiver’s head as a proxy for the eye-tracking glasses’ pose. During post-processing, we apply a low-pass filter to smooth the gaze data. We correct minor shifts in the glasses by re-aligning the gaze vector.

Task and Action Labeling ADL tasks like Hoyer sling transfers are long-horizon activities with multiple steps. Understanding caregiver actions at each stage enables segmentation into modular components. To support this, we record video and audio using a GoPro, with caregivers verbally describing their actions as they interact with the manikin. OTs then annotate the recordings, segmenting tasks into meaningful sub-tasks based on their expertise, providing insights into task decomposition and procedural flow.

D. Sensor Synchronization

Due to hardware limitations, each sensor operates at a different sampling rate: RGB-D cameras at 15 Hz, tactile skin at 60 Hz, motion capture at 150 Hz, and the eye tracker at 120 Hz. To achieve temporal synchronization, all computers are synchronized with the NIST Internet Time Servers using the chrony service on Ubuntu and w32tm on Windows. For data alignment, we use the RGB-D stream (15 Hz) as the reference timeline, and extract the closest timestamped samples from the other modalities for each RGB frame to produce synchronized frames at 15 Hz.



BT (Bathing), TL (Toileting), TD (Dressing T-shirt), TU (Undressing T-shirt), VD (Dressing Vest), VU (Undressing Vest), SD (Dressing Shorts), SU (Undressing Shorts), WTD (Wheelchair Dressing T-shirt), WTU (Wheelchair Undressing T-shirt), WVD (Wheelchair Dressing Vest), WVU (Wheelchair Undressing Vest), TW (Transfer to Wheelchair), TB (Transfer to Bed), and GR (Grooming)

Fig. 4: Analysis of Dataset Characteristics. We analyze the diversity of the dataset across different aspects: (a-c) general data collection statistics; (d-f) occupational therapists' strategies; (g) time duration across tasks; (h-j) physical contact characteristics; and (k-l) force magnitude information.

V. DATASET CHARACTERISTICS AND ANALYSIS

OpenRoboCare contains 315 sessions of caregiver expert demonstrations, totaling 19.8 hours of multimodal data across 5 modalities, collected by 21 occupational therapists. The task selection covers 5 out of 6 basic activities of daily living, with 15 task variations, for a total of 31,185 expert demonstration data samples. The dataset, usage documentation, and the fine-tuned pose estimation model is publicly available on our website [44]. In this section, we present the characteristics and unique insights within the dataset that can inform caregiving robot design.

A. Caregiver Demographics

We recruited 21 occupational therapists (OTs), including 19 final-year OT students and 2 licensed OT professionals. The most experienced OT has over 40 years of clinical experience. All participants are female, ages 22 to 75. Collectively, they have experience working with populations with neurological conditions, stroke, traumatic brain injury, spinal cord injury, muscular dystrophy, and cerebral palsy.

B. Guiding Principles for Caregiving

Throughout data collection, we observe various principles that OTs use to perform tasks efficiently and with minimal physical effort. We collaborate with an experienced OT to analyze our observations and distill underlying principles that can guide robot design for caregiving tasks.

Principle 1 - Pre-positioning (P1): OTs prioritize safety by carefully preparing the care recipient before initiating a task. They ensure that the care recipient's posture, stability,

joint angles, and supporting surfaces are appropriate for task execution. For example, before rolling the care recipient to the side, OTs align the trunk on the supporting surface to prevent unexpected limb trapping or unintended shifts in momentum due to weight redistribution.

Principle 2 - Anticipation (P2): OTs anticipate and position their body mechanics to support the entire task sequence, particularly for large-scale movements. They anticipate both the final position and the trajectory of the care recipient's body and limbs, which influences task execution decisions. For example, when rolling a care recipient on the bed, an OT may position their hands on the opposite side of the body before initiating the roll. Although this places the OT at a biomechanical disadvantage initially, it provides better control, allows for monitoring discomfort, and ensures the care recipient is positioned well for the next step.

Principle 3 - Efficiency (P3): OTs prioritize accuracy and timely completion of tasks to ensure efficiency. Care recipients with severe mobility limitations often have medical conditions, making efficient ADL execution crucial. Delays can lead to bradyarrhythmias, hypotension, or dizziness, while improperly placed garments may cause pressure ulcers. For example, during transfers with a Hoyer sling, OTs minimize the duration the care recipient is lifted to reduce discomfort and physiological stress.

C. Illustrative Caregiving Techniques

We connect the principles to concrete techniques observed in OpenRoboCare, illustrating how expert OTs ground these high-level strategies in real caregiving scenarios.

Technique 1 - Bridge Strategy: bending the care recipient's knees and applying pressure behind the knees at the top of the calf to momentarily elevate the pelvis, which involves anticipating the motion (P2). This technique, often used in bed toileting to position a bedpan, requires significant caregiver effort and is most efficient when the care recipient has a smaller body size than the caregiver (P3).

Technique 2 - Segmental Roll: gradually turning the care recipient's body. The OT bends the care recipient's opposite-side knee and applies pressure on the bent knee to initiate a progressive rolling motion toward the OT (P1). The pelvis moves first, followed by the upper body, shoulders, and finally the head. This technique allows for slow and controlled movement (P2), making it particularly useful for bed bathing and toileting, especially for care recipients prone to dizziness. Additionally, it benefits caregivers who are significantly smaller than the care recipient, as it reduces physical strain.

Technique 3 - Wheelchair Recline During Transfer: reclining the wheelchair to a 45-degree backward tilt before transferring the care recipient improves positioning of the care recipient on the chair, another example of pre-positioning (P1). This action also minimizes the need for post-transfer adjustments and reduces physiological stress, aligning with the efficiency principle (P3).

Technique 4 - Stabilizing Key Points of Control: The pelvic bone, shoulders, and head serve as the primary points of control for body movement and are essential in all ADL tasks. To facilitate movement, OTs place their hands on key control points—such as the left or right scapula and pelvis—to provide input to initiate, support, and control movement, aligning with pre-positioning and anticipation principle (P1, P2). This strategy maximizes the ability of the therapist to effectively and efficiently facilitate body movements, such as bed rolling, for the care recipient (P3).

D. Insights from Task Execution

The dataset captures diverse task executions, offering valuable insights for training robots in caregiving-specific parameters, strategies, and workflow optimization.

Task Duration: We show the duration of each task in Fig. 4d. Care recipient transfer from bed to wheelchair is the most time-consuming task, followed by transfer from wheelchair to bed, both taking significantly longer than any other ADL. Transferring requires maneuvering multiple assistive devices and manipulating deformable objects and human limbs, making it a long-horizon task that can take up to 9 minutes even for expert OTs.

Toileting Approach: OTs take two distinct approaches for the toileting task (Fig. 4e): (1) rolling the manikin to the side and inserting the bedpan under the body; and (2) bending the knees and lifting the manikin's hips to place the bedpan underneath. OTs working with the heavier manikin chose the first technique, as it requires less physical effort.

Manikin Lift Side Preference: Fig. 4f shows the lifting-side preferences for on-bed tasks. We do not observe any significant trend in side preferences. For dressing tasks, we observe that OTs lift the manikin from both sides to prevent the cloth from getting stuck underneath. All but one OT

was right-handed. We observe no correlation between hand dominance and the preferred lifting side. In clinical practice, OTs also follow the care recipient's preferences for lifting in addition to their own movement preferences. To make this process efficient, OTs use specific biomechanical techniques that do not require their maximal strength.

Dressing Approach: Fig. 4g shows the distribution between two observed dressing approaches: (1) head-first, and (2) sleeve-first. Over 90% of OTs prefer to insert the sleeves first when dressing the T-shirt or vest for both hospital bed and wheelchair dressing tasks. Over 75% of OTs prefer to undress the T-shirt head-first. Dressing requires precise limb manipulation, making the sleeve-first approach preferable for better control. In contrast, undressing involves fewer constraints and does not demand precise limb guidance, making head-first approach quicker and more intuitive.

E. Insights from Physical Interactions

The dataset provides empirical insights into how human caregivers distribute force and interact with different body regions across tasks.

Physical Contact: Different tasks require physical interactions with different body regions (Fig. 4h). Lower-body dressing requires significantly more contact with the shin and thigh, whereas upper-body dressing requires more contact with the forearm and upper arm. Transferring results in near equal contact with body regions. Physical contact differs within ADL variations, with notable differences between transferring to a wheelchair versus a bed, dressing versus undressing, lying versus sitting, and different clothing types (Fig. 4i). Among all ADLs, bathing has the highest number of physical contacts, while grooming has the least (Fig. 4j).

Force Magnitude: Additionally, the magnitude of force exerted on the manikin varies significantly across tasks (Fig. 4k). Transferring requires the highest force, as the manikin must be completely lifted. In general, greater force is applied to the limbs than to the torso, as the limbs act as leverage points to turn the manikin and adjust its posture. Force magnitude also varies across body regions over time during a bathing task (Fig. 4l). It peaks when the caregiver bathes a specific area, while turning the manikin increases force across all regions.

F. Guidelines for Robot Caregiving

We distill insights from occupational therapist demonstrations to guide the development of caregiving robots. By analyzing gaze information, we can determine the caregiver's area of interest, allowing robots to identify where to act. Additionally, the caregiver's gaze shift speed provides an estimate of how fast the robot policy should be. Predictive gazes occur when the caregiver looks at a different body part before engaging in contact with it. Caregivers have a pre-emptive timing of around 2.02 seconds. In terms of robot policy planning, this delay indicates a possible look-ahead timing to predict the next trajectory, while the robot is still currently acting on its present task. Tactile sensing indicates the range and distribution of force the robot should apply. For example, gentle tasks such as bathing typically involve light contacts around 0.1–2 N, while physically demanding tasks such as repositioning or turning the body

can exceed 20–30 N. This wide range highlights the need for robots to be capable of both delicate touch and high-force interaction, requiring torque-sufficient, compliant, and backdrivable actuators, along with force sensors that offer high resolution and a broad dynamic range. Caregivers often use whole-arm contact, such as bracing the body during a roll, which differs significantly from typical robot manipulation techniques like pick-and-place. To support these interactions, robots should incorporate distributed sensing and compliance along the entire arm, not just at the end-effector, to enable safe and effective physical contact throughout the task. Observing caregivers’ workspaces helps define a reasonable range for designing caregiving robot hardware. Since caregiving strategies may vary based on a caregiver’s body shape, these strategies could also change depending on the robot’s embodiment. Robots can learn to coordinate multiple assistive devices by leveraging insights from human caregiving interactions. See our website for details [44].

VI. EVALUATION AND OPEN CHALLENGES

We evaluate the state-of-the-art perception (Sec. VI-A) and planning (Sec. VI-B) methods on OpenRoboCare, and discuss the open challenges to address the question: *What is the performance gap in existing approaches for the robot caregiving domain?* We also highlight the potential for this dataset to help advance robot vision and planning methods.

A. Perception: Human pose estimation

We evaluate SOTA pose detection methods on RGB images for tracking the manikin’s pose. For 2D pose estimation (Table II), we use mAP_{50–95}, and for 3D pose estimation (Table III), we use Mean Per Joint Position Error (MPJPE).

For 2D pose estimation, the off-the-shelf YOLOv11 [45] performs poorly, but fine-tuning on even a small subset of our labeled dataset leads to large gains, especially in occlusion-heavy tasks like dressing and transfer. These results demonstrate the potential of OpenRoboCare in advancing robust pose perception for real-world caregiving.

Occlusion: Physical interactions in close proximity between caregivers and care recipients, along with the presence of assistive devices, result in frequent heavy occlusion. This makes pose estimation particularly challenging, due to partial visibility of the body.

Distribution Shift for Real-world Caregiving: The caregiving domain exhibits a substantial distribution shift compared to general human pose datasets. It involves close physical interaction between two individuals, necessitating multi-body pose estimation. Prior datasets [2] typically assume simpler poses (e.g., standing) and fixed camera placements that minimize occlusion. Additionally, caregiving involves many unique postures that rarely occur in non-caregiving ADL scenarios, further complicating pose estimation and generalization. Finally, caregiving contains multimodal variability in task planning for multiple tasks, much of which originates from OT expertise. These factors pose challenges for existing models for robot-assisted caregiving. Our dataset has the potential to bridge this gap.

TABLE II: 2D Pose Detection Performance (mAP_{50–95}).

Method (YOLOv11 Variant)	Bathing	Dressing	Transfer
Pretrained	0.0244	0.0259	0.0218
Fine-tuned (1 OT)	0.5711	0.7052	0.5964
Fine-tuned (5 OTs)	0.7757	0.8228	0.6648

TABLE III: 3D Pose Estimation Performance (MPJPE in mm)

Method	RTMOPose3D [46]	MixSTE [47]	HoT [48]	MHFormer [49]
MPJPE (mm)	119.9	122.9	142.2	162.7

B. Planning: Long-horizon task recognition

We run VidChapters-7M [37] on a subset of 21 videos. Qualitative results are available on our website [44]. While existing methods can recognize some subtasks, a significant gap remains in the caregiving domain. The lack of training data contributes to errors—for example, the model misidentifies “positioning the Hoyer” as “positioning the foyer,” likely due to unfamiliarity with caregiving terminology. The long task horizons make it difficult for current models to recognize full procedures. Finally, the diversity in task plans introduces challenges in decomposing long-horizon task plans.

VII. DISCUSSION

In this work, we proposed OpenRoboCare, the first multi-task, multimodal, expert-collected dataset for robot caregiving. While the dataset is already large, future work could consider supplementing it with partial sensory data that is easier to obtain. Another limitation of OpenRoboCare is its focus on fully passive care recipients. Future work will consider partially mobile individuals who actively participate in ADLs. These efforts are crucial to advance learning-based approaches, ultimately enabling more adaptable and capable robot caregivers.

REFERENCES

- [1] K. Bagewadi, J. Campbell, and H. B. Amor, “Multimodal dataset of human-robot hugging interaction,” *CoRR*, vol. abs/1909.07471, 2019. [Online]. Available: <http://arxiv.org/abs/1909.07471>
- [2] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras, “Two-person interaction detection using body-pose features and multiple instance learning,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 28–35.
- [3] P. Wang, J. Liu, F. Hou, D. Chen, Z. Xia, and S. Guo, “Organization and understanding of a tactile information dataset tacact for physical human-robot interaction,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 7328–7333.
- [4] O. Konak, V. Döring, T. Fiedler, L. Liebe, L. Masopust, K. Postnov, F. Sauerwald, F. Treykorn, A. Wischmann, S. Kalabakov, H. Gjoreski, M. Lustrek, and B. Arnrich, “Sonar, a nursing activity dataset with inertial sensors,” *Scientific Data*, vol. 10, 10 2023.
- [5] S. Kaczmarek, M. Fiedler, A. Bongers, S. Wibbeling, and R. Grzeszick, “Dataset and methods for recognizing care activities,” in *Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence*, ser. iWOAR ’22. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3558884.3558891>
- [6] B. A. Newman, R. M. Aronson, S. S. Srinivasa, K. Kitani, and H. Admoni, “Harmonic: A multimodal dataset of assistive human-robot collaboration,” *The International Journal of Robotics Research*, vol. 41, no. 1, pp. 3–11, 2022. [Online]. Available: <https://doi.org/10.1177/02783649211050677>
- [7] W. H. Organization, *Global report on health equity for persons with disabilities*. World Health Organization, 2022.
- [8] P. F. Edemekong, D. L. Bomgaars, S. Sukumaran, *et al.*, *Activities of Daily Living*, StatPearls, Ed. Treasure Island, FL: StatPearls Publishing, 2023, [Updated 2023 Jun 26]. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK470404/>

- [9] K. Scales, "Understanding the direct care workforce," 2024, accessed: 2025-02-28. [Online]. Available: <https://www.phinational.org/policy-research/key-facts-faq/>
- [10] R. K. Jenamani, P. Sundaresan, M. Sakr, T. Bhattacharjee, and D. Sadigh, "Flair: Feeding via long-horizon acquisition of realistic dishes," *arXiv preprint arXiv:2407.07561*, 2024.
- [11] E. K. Gordon, R. K. Jenamani, A. Nanavati, Z. Liu, D. Stabile, X. Dai, T. Bhattacharjee, T. Schrenk, J. Ko, H. Bolotski, *et al.*, "An adaptable, safe, and portable robot-assisted feeding system," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 74–76.
- [12] R. K. Jenamani, D. Stabile, Z. Liu, A. Anwar, K. Dimitropoulou, and T. Bhattacharjee, "Feel the bite: Robot-assisted inside-mouth bite transfer using robust mouth perception and physical interaction-aware control," in *2024 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2024, pp. 313–322.
- [13] R. K. Jenamani, T. Silver, B. Dodson, S. Tong, A. Song, Y. Yang, Z. Liu, B. Howe, A. Whitneck, and T. Bhattacharjee, "Feast: A flexible mealtime-assistance system towards in-the-wild personalization," in *Robotics: Science and Systems (RSS)*, 2025.
- [14] A. Jevtić, A. F. Valle, G. Alenyà, G. Chance, P. Caleb-Solly, S. Dogramadzi, and C. Torras, "Personalized robot assistant for support in dressing," *IEEE transactions on cognitive and developmental systems*, vol. 11, no. 3, pp. 363–374, 2018.
- [15] J. Zhu, M. Gienger, G. Franzese, and J. Kober, "Do you need a hand?—a bimanual robotic dressing assistance scheme," *IEEE Transactions on Robotics*, vol. 40, pp. 1906–1919, 2024.
- [16] R. Madan, S. Valdez, D. Kim, S. Fang, L. Zhong, D. T. Virtue, and T. Bhattacharjee, "Rabbit: A robot-assisted bed bathing system with multimodal perception and integrated compliance," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 472–481.
- [17] C.-H. King, T. L. Chen, A. Jain, and C. C. Kemp, "Towards an assistive robot that autonomously performs bed baths for patient hygiene," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 319–324.
- [18] Z. Huang, A. Nagata, M. Kanai-Pak, J. Maeda, Y. Kitajima, M. Nakamura, K. Aida, N. Kuwahara, T. Ogata, and J. Ota, "Robot patient for nursing self-training in transferring patient from bed to wheel chair," in *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management: 5th International Conference, DHM 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings 5*. Springer, 2014, pp. 361–368.
- [19] Z. Liu, Y. Ju, Y. Da, T. Silver, P. N. Thakkar, J. Li, J. Guo, K. Dimitropoulou, and T. Bhattacharjee, "Grace: Generalizing robot-assisted caregiving with user functionality embeddings," in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2025, pp. 686–695.
- [20] T. Silver, R. K. Jenamani, Z. Liu, B. Dodson, and T. Bhattacharjee, "Coloring between the lines: Personalization in the null space of planning constraints," *Under Review*, 2025.
- [21] P. Sundaresan, S. Belkale, and D. Sadigh, "Learning visuo-haptic skewering strategies for robot-assisted feeding," in *6th Annual Conference on Robot Learning*, 2022.
- [22] N. Ha, R. Ye, Z. Liu, S. Sinha, and T. Bhattacharjee, "REPEAT: A real2sim2real approach for pre-acquisition of soft food items in robot-assisted feeding," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 7048–7055.
- [23] R. Ye, S. Chen, Y. Yan, J. Yang, C. Ge, J. Barreiros, K. Tsui, T. Silver, and T. Bhattacharjee, "CART-MPC: Coordinating assistive devices for robot-assisted transferring with multi-agent model predictive control," in *ACM/IEEE International Conference on Human Robot Interaction (HRI)*, 2025.
- [24] A. Khazatsky, K. Pertsch, S. Nair, A. Balakrishna, S. Dasari, S. Karamcheti, S. Nasiriany, M. K. Srirama, L. Y. Chen, K. Ellis, *et al.*, "Droid: A large-scale in-the-wild robot manipulation dataset," *arXiv preprint arXiv:2403.12945*, 2024.
- [25] A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, *et al.*, "Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6892–6903.
- [26] AARP and N. A. for Caregiving, "Caregiving in the united states 2020," 2020.
- [27] R. M. Goodwin, R. L. Utz, C. E. Elmore, K. A. Ornstein, D. L. Tay, L. Ellington, K. R. Smith, and C. E. Stephens, "Leveraging existing datasets to advance family caregiving research: Opportunities to measure what matters," *Journal of Aging & Social Policy*, vol. 36, no. 4, pp. 562–580, 2024.
- [28] L. Hyejin, O. Bumjo, K. Sunyoung, and L. Kiheon, "Adl/iadl dependencies and unmet healthcare needs in older persons: A nationwide survey," *Archives of Gerontology and Geriatrics*, vol. 96, p. 104458, 2021.
- [29] C. S. Wilson, S. DeDios-Stern, C. Bocage, A. A. Gray, B. M. Crudup, and H. F. Russell, "A systematic review of how spinal cord injury impacts families," *Rehabilitation psychology*, vol. 67, no. 3, p. 273, 2022.
- [30] A. Olivares-Alarcos, S. Foix, and G. Alenyà, "On inferring intentions in shared tasks for industrial collaborative robots," *Electronics*, vol. 8, no. 11, 2019. [Online]. Available: <https://www.mdpi.com/2079-9292/8/11/1306>
- [31] F. Pastor, D. Lin, J. Gomez-de Gabriel, and A. Garcia, "Dataset with tactile and kinesthetic information from a human forearm and its application to deep learning," *Sensors*, vol. 22, p. 8752, 11 2022.
- [32] G. Saleem, U. I. Bajwa, and R. H. Raza, "Toward human activity recognition: a survey," *Neural Computing and Applications*, vol. 35, no. 5, pp. 4145–4182, 2023.
- [33] T. Alshammari, N. Alshammari, M. Sedky, and C. Howard, "Simadl: Simulated activities of daily living dataset," *Data*, vol. 3, no. 2, 2018. [Online]. Available: <https://www.mdpi.com/2306-5729/3/2/11>
- [34] M. Patel and S. Chernova, "Proactive robot assistance via spatio-temporal object modeling," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 881–891. [Online]. Available: <https://proceedings.mlr.press/v205/patel23a.html>
- [35] E. L. Tonkin, M. R. Whitehouse, H. Song, N. Twomey, T. Diethe, M. Kull, M. P. Nieto, M. Camplani, S. Hannuna, X. Fafoutis, Z. Nian, P. Woznowski, G. J. L. Tourte, R. Santos-Rodríguez, P. A. Flach, and I. J. Craddock, "A multi-sensor dataset with annotated activities of daily living recorded in a residential setting," *Scientific Data*, 2023.
- [36] D. Sánchez, M. Tentori, and J. Favela, "Activity recognition for the smart hospital," *IEEE intelligent systems*, vol. 23, no. 2, pp. 50–57, 2008.
- [37] A. Yang, A. Nagrani, I. Laptev, J. Sivic, and C. Schmid, "Vidchapters-7m: Video chapters at scale," in *NeurIPS*, 2023.
- [38] F. Stein and K. Haertl, *Pocket Guide to Intervention in Occupational Therapy*. Routledge, 2024.
- [39] N. Bell and E. McGeggen, "Activities of daily living," *Early's physical dysfunction practice skills for the occupational therapy assistant E-book*, vol. 194, 2021.
- [40] M. F. Rybski, "Rehabilitation approach," in *Kinesiology for Occupational Therapy*. Routledge, 2024, pp. 367–438.
- [41] B. A. Newman, R. M. Aronson, S. S. Srinivasa, K. Kitani, and H. Admoni, "Harmonic: A multimodal dataset of assistive human-robot collaboration," *The International Journal of Robotics Research*, Dec 2021. [Online]. Available: <https://doi.org/10.1177/02783649211050677>
- [42] T. Bhattacharjee, H. Song, G. Lee, and S. S. Srinivasa, "A Dataset of Food Manipulation Strategies," 2018. [Online]. Available: <https://doi.org/10.7910/DVN/8TTXZ7>
- [43] C. E. Mortimer, "Comparison of manikin-based simulators and patient monitor simulators within paramedic education: the student perspective," *BMJ Simulation & Technology Enhanced Learning*, vol. 4, no. 2, p. 65, 2018.
- [44] "OpenRoboCare Website," <https://emprise.cs.cornell.edu/robo-care/>.
- [45] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics YOLO," Jan. 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [46] T. Jiang, X. Xie, and Y. Li, "Rtmw: Real-time multi-person 2d and 3d whole-body pose estimation," *arXiv preprint arXiv:2407.08634*, 2024.
- [47] J. Zhang, Z. Tu, J. Yang, Y. Chen, and J. Yuan, "Mixste: Seq2seq mixed spatio-temporal encoder for 3d human pose estimation in video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 13 232–13 242.
- [48] W. Li, M. Liu, H. Liu, P. Wang, J. Cai, and N. Sebe, "Hourglass tokenizer for efficient transformer-based 3d human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 604–613.
- [49] W. Li, H. Liu, H. Tang, P. Wang, and L. Van Gool, "Mhformer: Multi-hypothesis transformer for 3d human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 13 147–13 156.