A vision-based humanoid compliant skill transfer framework: Application to robotic cutting tasks

MAI, Zhaohong <http://orcid.org/0009-0005-1391-8093>, ZENG, Chao <http://orcid.org/0000-0003-0799-6712>, WANG, Ning and YANG, Chenguang <http://orcid.org/0000-0001-5255-5559>

Biomimetic Intelligence and Robotics 6 (2026) 100280

Contents lists available at ScienceDirect

Biomimetic Intelligence and Robotics

journal homepage: www.keaipublishing.com/en/journals/biomimetic-intelligence-and-robotics/

Research Article

A vision-based humanoid compliant skill transfer framework: Application to robotic cutting tasks

Zhaohong Mai [a], Chao Zeng [b,*], Ning Wang [c], Chenguang Yang [b]

[a] Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China
[b] Department of Computer Science, University of Liverpool, Liverpool L69 3BX, United Kingdom
[c] School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield S1 2NU, United Kingdom

ARTICLE INFO

Article history:
Received 23 June 2025
Revised 3 August 2025
Accepted 1 September 2025
Available online 17 January 2026

Keywords:
Human–robot skill transfer
Compliant control
Skill generalization
Robotic cutting

ABSTRACT

Autonomously completing a contact-rich task for multiple manipulation objects remains a challenging problem for robots. To achieve this goal, learning from demonstration has emerged as an efficient method for transferring human-like skills to robots. Existing works primarily focus on trajectory or impedance learning to design force-impedance controllers for specific tasks, which require precise force sensing. However, visual perception plays a critical role in enabling humans to perform dexterous manipulation. To bridge the gap between vision and learning in the control loop, this work proposes a vision-based humanoid compliant skill transfer (VHCST) framework. Considering the lack of vision-impedance mapping, a hybrid tree is introduced as a planning bridge to encode skill parameters across multiple objects. To simplify skill transfer, an observation-wearable demonstration method is employed to capture the position and stiffness of human's arm. The decoupled learning model incorporates the geometric properties of stiffness ellipsoids, which reside on Riemannian manifolds. Finally, the proposed approach is validated through robotic cutting experiments involving multiple objects. Comparative experimental results demonstrate the effectiveness of the proposed framework.

1. Introduction

Compliance has emerged as an indispensable attribute for modern robotic systems, enabling them to achieve the precision, adaptability, and safety required for complex interactive tasks [1]. Significant advancements, such as hybrid position/force control and mechanical impedance have been achieved to enable reliable force and motion control applicable to both contact and non-contact scenarios [2]. Numerous prominent studies have enhanced impedance controller using force sensing [3], visual error feedback [4], or humanoid-inspired skills [5]. Robots necessitate robust policy frameworks and advanced task planning or learning methodologies to effectively execute contact-rich manipulation tasks that leverage environmental interactions [6], yet challenges persist in intelligent planning and seamless integration of multimodal perception and action.

One feasible way to efficiently deploy torque-controlled robots in contact-rich scenarios is to equip them with humanoid dexterous manipulation abilities through skill transfer (e.g., kinesthetic teaching, teleoperation, passive observation), also referred to as learning from demonstration (LfD) [7]. Multiple LfD approaches have been developed for manipulating contact-rich tasks, such as statistics-based methods (e.g., Gaussian mixture model–Gaussian mixture regression (GMM–GMR) [8]), primitive-based methods (e.g., dynamical movement primitives [9] and its extension [10]). To address robotic manipulation scenarios involving multiple reference frames or dynamically-evolving input–output relationships, task-parameterized motion generation methods have been proposed (e.g., Task-Parameterized GMM (TP-GMM)) [11,12].

Nevertheless, the aforementioned methods predominantly focus on data within Euclidean spaces, neglecting certain variables inherent to manifold spaces, such as orientation, stiffness and manipulability. Consequently, manifold-based learning models have been introduced in LfD to encode the manifold of symmetric positive definite (SPD) matrices [13,14] and other manifolds [15]. As previously established, the modulation of joint impedance is crucial for robots to effectively perform contact-rich tasks. However, prior works, such as [12,16], often compute stiffness matrices (measured on SPD) in Cartesian coordinates or decompose them into column-wise normalized eigenvector matrices and corresponding eigenvalue matrices [9], thereby underutilizing their computational potential within manifold spaces. The impedance regulation framework proposed in [17] decomposes the stiffness into configuration dependent stiffness (CDS) and common mode stiffness (CMS) and reproduces the distribution

* Corresponding author.
E-mail address: chaozeng@ieee.org (C. Zeng).

https://doi.org/10.1016/j.birob.2026.100280

characteristics of stiffness in various directions within the CDS, without transferring the magnitude and variation of human stiffness through the CMS to the robot. While, it may not be able to replicate the flexible adjustment of stiffness exhibited by humans during multi-stage variable stiffness movements. Inspired by [18], the proposed system in this paper represents stiffness as SPD matrix elements within manifold spaces for both learning and reproduction, leveraging GMM/GMR-based policy frameworks.

Meanwhile, current model-based learning approaches are highly dependent on demonstration data, significantly constrains their generalization capabilities in complex scenarios. Furthermore, task-parameterized learning methods lack effective integration with visual perception systems, hindering their ability to semantically interpret dynamic environments. To overcome these limitations, a vision-based learning model is proposed in this work to enhance the robot's generalization performance in compliant manipulation tasks.

In the domain of implicit force control, contemporary research prioritizes the tracking of desired interaction forces while dynamically adjusting impedance parameters through quadratic programming framework [8,19]. Alternatively, by estimating the geometric parameters of the contact surface in real time, the contact parameters and force constraints can be integrated into a gradient-based model predictive control framework [20] to achieve optimal force–position coordinated control. The above-mentioned model-based studies continue to rely heavily on force sensing, lacking an online generalized mapping model that maps visual features to impedance parameters. However, in task scenarios involving operations such as cutting or sawing, where handheld tools are utilized, it remains challenging to accurately measure the interaction forces during demonstrations. To address it, we propose a closed-loop framework which integrating visual feature understanding (e.g., material properties), dynamic characteristics (e.g., stiffness), and decision-making processes. A strategy selector and condition monitor are indispensable innovations in this work.

Behavior Trees (BT), which organize state and task-switching logic into a hierarchical tree, have been extensively utilized in robotic control. They can be generated via learning, planning, or manual design, making them suitable for managing complex task execution and multi-task scenarios [21]. Adaptive BT, by integrating strategy selection and condition monitoring, enable robots to dynamically adjust their behavior during task execution [22]. To construct BT efficiently and flexibly, LfD serves as an effective method for extracting knowledge. In [23], the authors propose a novel BT learning approach with logic factorization. Building on [23], this paper proposes a hybrid tree that integrates decision trees and behavior trees to enable strategy selection and condition monitoring through visual recognition capabilities.

The inherent complexity of tool-mediated contact-rich manipulation motivates robotic cutting as a validation benchmark. Imagine slicing food items (e.g., cucumber, potato), the robot arm starts at its initial pose and moves the knife until contact. Then the blade should be moved down until it cuts through the object completely and makes contact with the cutting board [24]. In this case, considering the food items differ greatly in both visual properties (e.g., shape, color, and texture) and mechanical properties (e.g., hardness, density, and friction), the robot will need to generalize its cutting skills across these varying attributes [25]. This simple yet intricate task illustrates the challenges of cutting multiple objects, which is introduces distinct technical difficulties compared to some contact-rich tasks (e.g., wipe, scoop, grind, pull a door and push a button). Previous studies have explored state estimation using multi-sensor systems [26] and motion planning for cutting actions based on elastic fracture mechanics [25]. LfD has also been utilized to capture cutting trajectory motions [24] and stiffness characteristics [27]. Given the destructive and irreversible nature of cutting tasks, differentiable simulators like Roboninja [28] have been employed to optimize cutting strategies under task-specific physical constraints.

Rather than modeling the intricate dynamics of cutting interactions in detail, this work directly transfers impedance skills within Riemannian manifolds, enabling humans to impart prior knowledge to the robot. By employing task-oriented visual detection, we decompose the cutting task into multiple low-level skills (e.g., pose adjustment and stiffness modulation) and integrate them using hybrid tree. Interaction forces, robot torques, and energy consumption are jointly minimized with the proposed framework. The primary contributions are summarized as:

(1) We propose a vision-based humanoid compliant skill transfer framework, which integrates visual perception and learning-based control within a unified architecture. By incorporating task-parameterized decoupled learning model and robust visual feature detection, this approach enhances the generalization capability of robotic manipulation tasks across diverse scenarios.
(2) To address the interaction tasks for multiple manipulation objects, we propose a task-oriented hybrid tree strategy to dynamically determine critical task parameters, enabling the generation of compliant skills. The proposed framework establishes a principled mapping from visual perception to robot impedance and enhances the accuracy of compliant skill reproduction through manifold-based computations.
(3) To evaluate the effectiveness of our proposed skill transfer framework, we apply it to robotic cutting tasks. The results demonstrate that our framework allows the robot to successfully cut multiple objects, exhibiting good performance in terms of generalizability and flexibility.

2. Preliminary work

2.1. Stiffness ellipsoids on SPD manifold

In this paper, we consider a torque-controlled robot that physically interacts with an unknown environment. Accordingly, the robot dynamics is modeled in the joint space as

$$M(q)\ddot{q} + C(q,\dot{q}) + G(q) = \tau - J^T F_e(t) \tag{1}$$

where $q, \dot{q}, \ddot{q} \in \mathbb{R}^{n\times 1}$ denote the angle, angular velocity and angular acceleration of each joint. Similarly, $M \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^n$ and $G \in \mathbb{R}^n$ are the inertia matrix, centripetal vector and Coriolis and gravity vectors of the robot respectively. $J \in \mathbb{R}^{n\times 3}$ represents the Jacobian matrix, τ represents the control inputs, and F_e is the robot's interaction force with the environment, which can be defined as

$$M_d(\ddot{x}_d - \ddot{x}_0) + D_d(\dot{x}_d - \dot{x}_0) + K_d(x_d - x_0) = F_e \tag{2}$$

where $K_d \in \mathbb{R}^{3\times 3}$ and $D_d \in \mathbb{R}^{3\times 3}$ are the stiffness and damping factors and $M_d \in \mathbb{R}^{3\times 3}$ represents the inertia factors. $x_o \in \mathbb{R}^3$ is the actual trajectory in the Cartesian space and $x_d \in \mathbb{R}^3$ represents the desired trajectory. Considering the non-diagonally dominant inertia characteristic of robot, to simplify it, we choose the critically-damped condition ${D_d}^2 = 4\Lambda K_d$ and a fixed natural index Λ to guarantee the fastest convergence of the error without oscillations [29]. Thus, task variable K_d and x_d could be focused and learned from demonstration.

As the stiffness K of human's arm or robot is an SPD matrix that cannot be computed in Euclidean space due to the inadequacy of using Euclidean distance metric to measure the distance between two SPD matrices [30]. The SPD manifold S^l_{++},

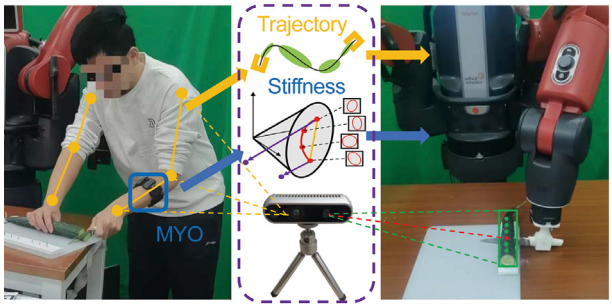

Fig. 1. The concept of transferring human compliant manipulation skills to a robot by considering both trajectories and stiffness profiles, using vision and EMG data, respectively.

which is a differentiable space for which each point $K \in \mathcal{M}$ locally resembles a Euclidean space and each tangent space $\mathcal{T}_K\mathcal{M}$ equips with a Riemannian metric, is used to denote the stiffness by $\mathcal{M}(\lambda) = \left\{K \in \mathbb{R}^{l\times l} \mid a^T K a > 0, a \in \mathbb{R}^l, a \neq 0\right\}$. It forms the interior of a convex cone in $\mathcal{T}_K\mathcal{M}$ with the inner product:

$$\langle W_1, W_2\rangle_{\mathcal{T}_K\mathcal{M}} = \left\langle K^{-\frac{1}{2}} W_1 K^{-\frac{1}{2}}, K^{-\frac{1}{2}} W_2 K^{-\frac{1}{2}}\right\rangle \tag{3}$$

where $\mathcal{T}_K\mathcal{M} = \left\{W \in \mathbb{R}^{l\times l} \mid W = W^T\right\}$ is the tangent space corresponding to the point $K \in \mathcal{M}$ and $S_1, S_2 \in \mathcal{T}_K\mathcal{M}$. To define the distance in the tensor space, an affine-invariant metric is introduced:

$$\begin{aligned}\operatorname{dist}(\Sigma_1, \Sigma_2) &= \left\|\log\left(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}\right)\right\| \\ &= \operatorname{dist}\left(A\Sigma_1 A^T, A\Sigma_2 A^T\right)\end{aligned} \tag{4}$$

where $\|\cdot\|$ represents the Frobenius norm of the matrix for any linear transformation A. Thus, there several operations on the S^l_{++} [31]. The logarithm map, which projects $K_2 \in \mathcal{M}$ to the tangent space $\mathcal{T}_{K_1}\mathcal{M}$ can be formulated as

$$W_2 = \log_{K_1}(K_2) = K_1^{\frac{1}{2}} \log\left(K_1^{-\frac{1}{2}} K_2 K_1^{-\frac{1}{2}}\right) K_1^{\frac{1}{2}} \tag{5}$$

The exponential map, which projects W_2 back to the $K_2 \in \mathcal{M}$, can be formulated as

$$K_2 = \exp_{K_1}(W_2) = K_1^{\frac{1}{2}} \exp\left(K_1^{-\frac{1}{2}} W_2 K_1^{-\frac{1}{2}}\right) K_1^{\frac{1}{2}} \tag{6}$$

2.2. The decoupled learning model with TP-GMM

Instead of kinesthetic teaching, we apply the proposed obser vation-wearable demonstration to record the wrist position and stiffness of the human's arm, as shown in Fig. 1. Building upon the impedance recognition framework proposed in [27], the endpoint stiffness K_g can be derived through a surface electromyography (sEMG)-based estimation methodology, which integrates intrinsic constant stiffness $\bar{K}_g$ and muscle activations extraction $\alpha(p)$. By performing parameter identification based on the results of the stochastic perturbation experiments, it is able to calculate the end-effector stiffness $K_g = \alpha(p) \cdot \bar{K}_g$.

Considering that the coordinate systems of the target and the tool are not fixed, TP-GMM model (see [11] for details) is implemented and improved as decoupled learning model in this work. And the learning goal is to estimate the desired profiles from the demonstration dataset $\left\{X_i, K_{g,i}\right\}_{i=1}^N$. To achieve it, X is used to offline train the TP-GMM in Euclidean space and K_g is calculated on SPD manifold. Initially, a set of reference points is selected within the operational environment. Subsequently, the recorded demonstrations are transformed from the base coordinate system into the coordinate systems associated with these reference points via the following coordinate transformation process.

$$X^{(r)} = R^{(r)^{-1}}\left(X - d^{(r)}\right), K_g^{(r)} = R^{(r)^T} K_g R^{(r)} \tag{7}$$

where the position $X \in \mathbb{R}^{m\times 1}$ and the stiffness matrices $K_g \in \mathbb{R}^{l\times l}$ are concatenated form the dataset $Z_X = [t; \operatorname{vec}(X)] \in \mathbb{R}^{(m+1)\times 1}$ and $Z_K = \left[t; \operatorname{vec}\left(K_g\right)\right] \in \mathbb{R}^{(L+1)\times 1}$. The subscript r denotes the rth reference point, where $R^{(r)} \in \mathbb{R}^{m\times m}$ and $d^{(r)} \in \mathbb{R}^{m\times 1}$ represent the rotation matrix and the translation vector, respectively, associated with the r-th reference point relative to the base coordinates. Hence, the TP-GMM in the coordinate system is given as

$$p\left(Z^{(r)}\right) = \sum_{n=1}^{N} \pi_n^{(r)} \mathcal{N}\left(Z^{(r)} | \mu_n^{(r)}, \Sigma_n^{(r)}\right) \tag{8}$$

where $\pi \in [0, 1]$ is the prior weighting. It is worth noting that N represents the number of Gaussian models, which are expressed in the Riemannian manifolds space [32]:

$$\mathcal{N}(Z|\mu, \Sigma) = \left((2\pi)^L |\Sigma|\right)^{-\frac{1}{2}} \exp^{-\frac{1}{2}\log_\mu(Z)\Sigma^{-1}\log_\mu(Z)} \tag{9}$$

where $\mu \in \mathbb{R}^L$ represents the center and $\Sigma \in \mathbb{R}^{L\times L}$ is the covariance of the Gaussian model. Then the parameters in (9) for each reference point would be optimized by Expectation–Maximization algorithm to maximize the log-likelihood function in Eq. (8) and saved in skill tree.

Data is often represented in a vector form in the GMM model. To enable the conversion between stiffness matrices $K \in \mathbb{R}^{l\times l}$ and vectors $s \in \mathbb{R}^L$, the following is defined:

$$s_k = \begin{cases} M_{ii}, & \text{if} k \leq l \\ \sqrt{2} M_{ij}, & \text{if} k > l \end{cases} \tag{10}$$

where $L = l(l+1)/2$ is the dimension of stiffness vector. i and j represent the row and column of K.

3. Methodology

To enable the generalization of task trajectories, a cutting-point detection method is proposed to support task-parameterized learning. Concurrently, a hybrid tree structure is developed to efficiently map visual feature space to impedance space, facilitating effective planning. Subsequently, the generalized trajectories and impedance parameters are transformed into joint space to achieve precise control execution. The following section systematically introduces the four modules of the proposed framework (see Fig. 2).

3.1. Cutting-point detection

The 3D camera in this system is installed using a fixed-position mounting method, with the process beginning with preprocessing steps for eye-to-hand calibration. The ArUco marker detection method in OpenCV is employed to approximate the position of the robot's end-effector, followed by the calculation of the rotation matrix R_{cam}^{rob} and translation matrix T_{cam}^{rob} between the 3D camera and the robot coordinate system through the least squares method.

For recognition of the position and orientation of cutting objects, as shown in Fig. 4(a), the YOLOv8s-OBB model [33] is implemented to perform oriented bounding box object detection. The inference results for target type and individual cutting targets $B = (x_m, y_m, w, h, \theta)$ are obtained from the model, including central location, width, height and rotation angle. Then the pixel coordinates of the four vertices of the bounding box $P_i, i \in$

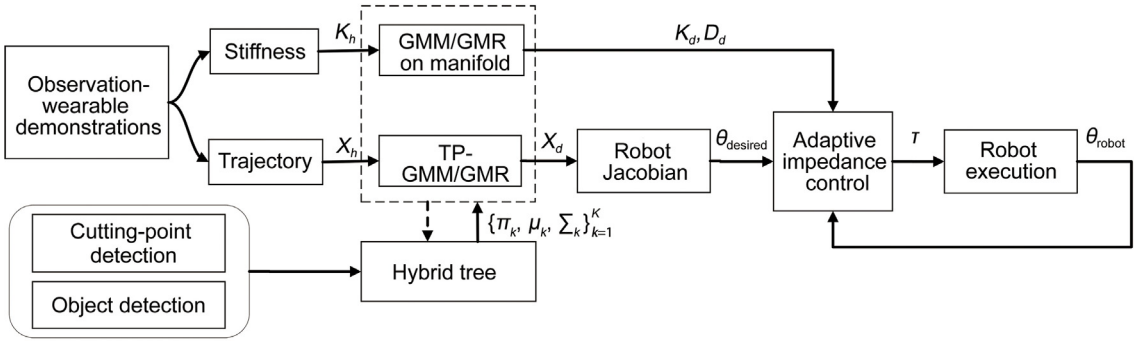

Fig. 2. The scheme of the proposed vision-based humanoid compliant skill transfer framework.

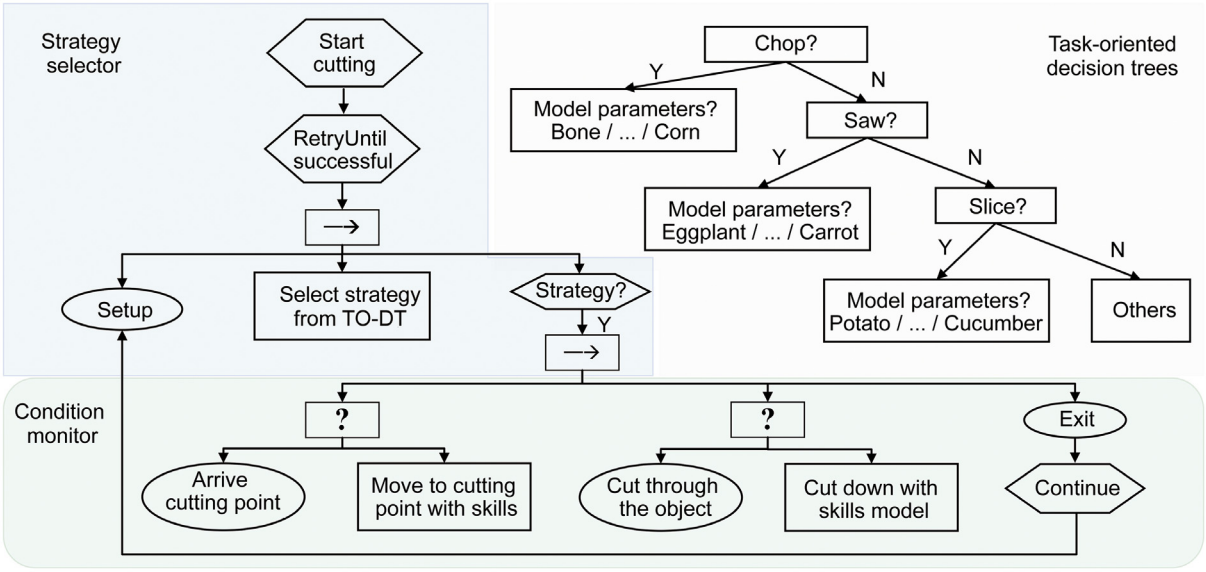

Fig. 3. Block diagram of the proposed hybrid tree structure. Strategy selector first decides which strategy to use for the next attempt according to the task-oriented decision tree. The condition monitor then instruct the robot to perform specific actions. The right-pointing arrow blocks are sequence nodes. The question mark arrow blocks are selector nodes.

{0, 1, 2, 3} and the pixel coordinates of multiple (k) cutting points Q_t are calculated as follows:

$$P_i = \begin{bmatrix} x_m \\ y_m \end{bmatrix} + \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \pm w/2 \\ \pm h/2 \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} Q_t.x \\ Q_t.y \end{bmatrix} = (1-t)\left(\frac{P_0+P_1}{2}\right) + t\left(\frac{P_2+P_3}{2}\right) \tag{12}$$

with $t \in \{k/(n-1) \mid k = 0, 1, \ldots, n-1\}$. And the target cutting position of robot coordinate is calculated as follows:

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} = R_{cam}^{rob} \begin{bmatrix} (Q_t.x - c_x)\cdot d/f_x \\ (Q_t.y - c_y)\cdot d/f_y \\ d \end{bmatrix} + T_{cam}^{rob} \tag{13}$$

where d is the depth of the point $[Q_t.x, Q_t.y]$. And the camera's intrinsic parameters $[c_x, c_y, f_x, f_y]$ were obtained using Zhang's calibration method implemented in the MATLAB toolbox. Then the unit cutting direction vector in the horizontal plane is defined as follows:

$$V_{\text{unit}} = \left(\frac{-L_y}{\sqrt{L_x^2+L_y^2}}, \frac{L_x}{\sqrt{L_x^2+L_y^2}}, 0\right) \tag{14}$$

with $L_x = X_t - X_0, L_y = Y_t - Y_0$. Finally, the cutting pose quaternions of robot could be calculated.

3.2. Task-oriented hybrid tree

As illustrated in Fig. 3, the proposed hybrid tree is composed of three core stages. Firstly, leveraging the predefined database of skill transfer detailed in Section 2, the applied cutting methods (e.g., chop, saw, slice) and GMM model are selected based on the target detection results within the task-oriented decision tree (TO-DT). Secondly, the strategy selector determines the optimal strategy for the next attempt and initiates the process by integrating visual perception and TO-DT. Finally, the condition monitor encapsulates the subtree corresponding to the chosen strategy, checking a set of task and strategy-specific conditions concurrently to proper operation of the behavior tree. The categories of BT nodes utilized in this work are referenced in [22]. The influence of the perception module, oriented bounding boxes object recognition and cutting-point detection, on the hybrid tree is primarily evident in three key aspects:

- Object category identification and the determination of task-oriented decision tree.
- Cutting-point localization and behavior tree state awareness, thereby influencing manipulation skill transfer.
- Classification of object cutting (slicing or sawing) to constrain the workspace and limit the output torque of the robot.

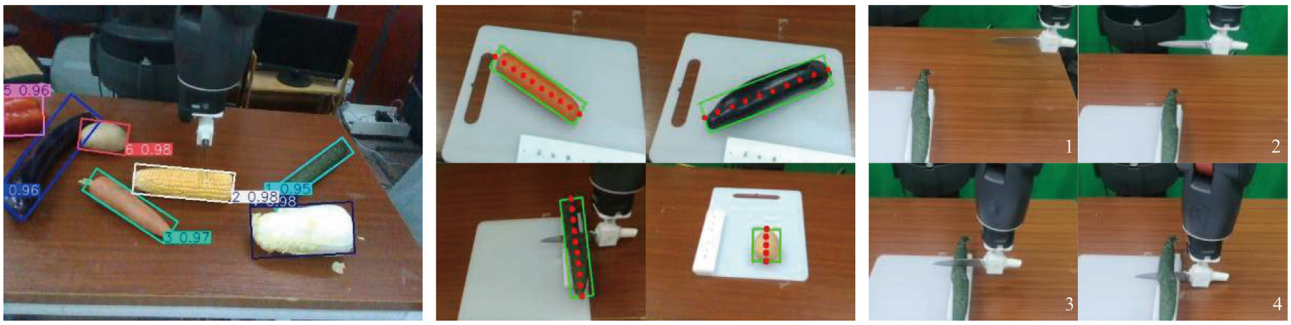

Fig. 4. (a) Examples of oriented bounding boxes object detection. (b) Examples of cutting-point detection. (c) The robot slices a cucumber.

3.3. Reproduction of position and stiffness

In the process of reproduction, the proposed decoupled TP-GMR method is utilized to replicate the sequence of robot end-effector position on Euclidean space and stiffness on manifold derived from TP-GMM [11]. With time serves as the input, the reproduced position and stiffness could be adjusted according to the temporal progression and cutting state of robot. Consequently, the reproduced data, alongside its center and covariance, can be expressed as follows:

$$Z = \begin{bmatrix} t_I \\ \mathrm{vec}\,(Y_O) \end{bmatrix}, \mu = \begin{bmatrix} \mu_I \\ \mu_O \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{II} & \Sigma_{IO} \\ \Sigma_{OI} & \Sigma_{OO} \end{bmatrix} \tag{15}$$

where Y_O can represent position X_O or stiffness K_O, $\mathrm{vec}\,(Y_O)$ is calculated by (10). The TP-GMR in the coordinate system of the r-th reference point is formulated as follows:

$$p(Y_O^{(r,i)}|t_I^{(r,i)}) = \sum_{n=1}^{N} h_n^{(r,i)} \mathcal{N}(\hat{\mu}_n^{(r)}, \hat{\Sigma}_n^{(r)}) \tag{16}$$

$$h_n^{(r,i)} = \frac{\pi_n^{(r)} \mathcal{N}(t_I^{(r,i)}|\mu_{I,n}^{(r)}, \Sigma_{II,n}^{(r)})}{\sum_{j=1}^{N} \pi_j^{(r)} \mathcal{N}(t_I^{(r,i)}|\mu_{I,j}^{(r)}, \Sigma_{II,j}^{(r)})} \tag{17}$$

where $p(Y_O^{(r,i)}|t_I^{(r,i)})$ represents the conditional probability distribution, $h_n^{(r,i)}$ is the weight of each Gaussian component, and $\mathcal{N}(\hat{\mu}_n^{(r)}, \hat{\Sigma}_n^{(r)})$ is a Gaussian distribution with mean $\hat{\mu}_n^{(r)}$ and covariance $\hat{\Sigma}_n^{(r)}$. For trajectory learned in Euclidean space, the final estimated outputs can be obtained as follow:

$$\hat{X}_O^i = \sum_i h_n^{(r,i)}(t) \left[\mu_{O,n}^{(r)} + \Sigma_{OI,n}^{(r)} \Sigma_{II,n}^{(r)^{-1}} \left(t - \mu_{I,n}^{(r)} \right) \right] \tag{18}$$

For stiffness matrices learned in SPD manifold, the reproduced K_O is calculated as

$$\Phi_n^{(r,i)} = \mathrm{Log}_{\mu_O^{(r)}}(\mu_{O,n}^{(r)}) + \Sigma_{OI,n}^{(r)} \Sigma_{II,n}^{(r)^{-1}} \mathrm{Log}_{t_I^{(r,i)}}(\mu_{I,n}^{(r)}) \tag{19}$$

$$K_O^{(r,i)} = \mathrm{Exp}_{\mu_O^{(r)}} \left(\sum_{n=1}^{N} h_n^{(r,i)} \Phi_n^{(r,i)} \right) \tag{20}$$

where $\Phi_n^{(r,i)}$ is the direction vector on the SPD manifold. Then, based on the inverse computation in Eq. (7), the primary output position X_O or stiffness K_O can be expressed as:

$$K_g^{(r,i)} = R^{(r)} K_O^{(r,i)} R^{(r)^T}, X_g^{(r,i)} = R^{(r)} X_O^{(r,i)} + d^{(r)} \tag{21}$$

3.4. Robot impedance control

Considering that excessive joint output torque may cause instability in the robot's joint movements, while insufficient torque may prevent the serial mechanism from operating properly, it is necessary to impose maximum and minimum limits on the joint output torque. Accordingly, the output torque generated during model reproduction is normalized and mapped to a reasonable operational range. With the cascaded structure of torque-controlled robot, combined with the desired end-effector position $X_{out}^i \in \mathbb{R}^{m\times 1}, i = 0, 1, \ldots, T$ and stiffness proposed in above-mentioned methods, variable stiffness $K_{jr} = J_e^{\mathrm{T}} K_g J_e$ in joint space would be optimized as follows:

$$K_j = K_{j,\min} + \frac{\left(K_{jr} - K_{j,\min}\right)^2}{K_{j,\max} - K_{j,\min}} \tag{22}$$

where $K_{j,\min}$ and $K_{j,\max}$ are the minimum and maximum stiffness of robot. Then, a torque controller is used in joint space as follows:

$$\tau_j = K_j (q_d - q) + D_j (\dot{q}_d - \dot{q}) + \tau_g \tag{23}$$

where $\tau_j \in \mathbb{R}^{n\times 1}$ is the control joint torque, $D_j = \zeta\sqrt{K_j}$ referred by [29]. $\tau_g \in \mathbb{R}^{n\times 1}$ represents the compensating torque. And the desired joint angles can be calculated as:

$$\begin{cases} \dot{q}_d = J^{\dagger}(X_d - X_I) \\ q_d(k+1) = q_d(k) + \dot{q}_d \Delta t \end{cases} \tag{24}$$

with $J^{\dagger} = J^T \left(JJ^T\right)^{-1}$ is the Moore–Penrose pseudoinverse of the Jacobian matrix.

4. Experiments and results

In this section, we demonstrate the performance of the vision-based humanoid compliant skill transfer framework. As shown in Fig. 5, robotic cutting tasks are performed to evaluate the effectiveness of the skill transfer method and the hybrid tree. In the experiment, the robot is enabled to cut different types of objects using two cutting styles: (1) slicing is used for cucumber and potato and (2) sawing is used for carrot and eggplant.

4.1. Experimental setup

The methodology is evaluated in a robotic cutting setup comprising a torque controlled manipulator of 7 DoF (the Rethink Robotic Baxter), a MYO armband from Thalmic Labs, a Microsoft Kinect V2 used for demonstration and the Intel Realsense d435i camera used for visual detection. The visual model training and real-time recognition are performed on a computer equipped with an NVIDIA GeForce RTX 4070 Laptop GPU. The results are transmitted to the robot's ROS-based control host via UDP communication.

To ensure consistency in the interaction during the cutting process, the experiment standardizes the use of the freshest vegetables and conducted multiple trials in the same time. As

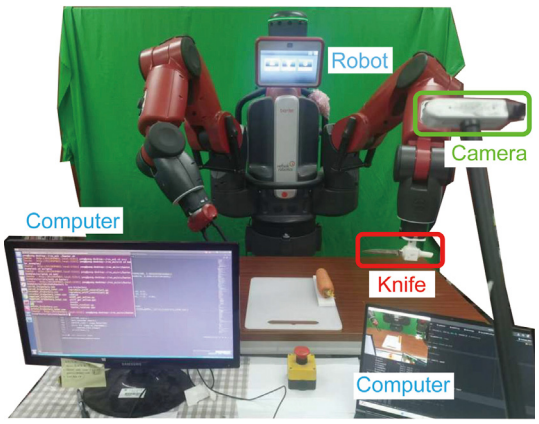

Fig. 5. Experimental scene diagram.

the study focuses on learning and controlling the cutting process through visual recognition and hybrid tree, the vegetables are securely fixed on the cutting board.

The proposed method utilizes a vision-based decoupled compliant learning model to generalize both the target trajectory and target stiffness for the cutting task. To validate its effectiveness, the robot performes the task using adaptive impedance control (AIC) derived from the generalized parameters. For comparison, the task is also executed under high impedance control (HIC) and low impedance control (LIC), which serve as baselines.

Given the limited force manipulability of the Baxter robot in a single direction and inspired by human cutting habits, the cutting behavior in this experiment is classified into two categories. The first category involves slicing soft or relatively brittle objects, such as potatoes and cucumbers, whereas the second category focuses on sawing harder objects or those with tough skins, such as carrots and eggplants. This classification facilitates the transfer of human-like compliant skills to the robot, allowing it to adapt its cutting strategy for various objects.

The four types of vegetables are selected as the experimental group for testing and focused analysis because their distinct physical properties are representative of a wide variety of vegetables. For example, cucumber is elongated and relatively brittle; potato has a uniform texture and a large cutting cross-section; carrot is a harder object with significant internal density variation; and eggplant has a tough outer skin, making it particularly difficult to cut through completely. Many other fruits and vegetables with similar physical properties exhibit comparable cutting requirements, thereby demonstrating the generalization capability of our proposed method to a broader range of objects.

4.2. Results of cutting tasks

The robot setup for collecting cutting data is shown in Fig. 1. With built-in sEMG sensor and Inertial Measurement Unit of MYO, as well as the open-source body tracking SDK of Kinect [34], humanoid cutting skills could be record precisely. Then the decoupled learning model allows the robot to automatically extract high-level features for generalizing across the different types of food items from demonstrations. The learned model parameters are subsequently integrated into the hybrid tree for sequencing skills. The target object is predicted from vision-based object detection. Then it is mapped to model parameters for adapting the slicing or sawing skills, which is initialized with the key information of cutting point detection. Some reproduced position and stiffness trajectories of slicing or sawing are shown in Fig. 8.

During this experiment, four groups are considered: potato, cucumber, carrot and eggplant. For each group, four demonstrations are conducted in total. The position trajectories are learned using TP-GMM/GMR with four Gaussian models and two candidate frames of reference (start point and end point), and the stiffness profiles are learned using TP-GMM/GMR on SPD with six Gaussian models and one candidate frames of reference (the camera coordinate system), the same as GMM/GMR on SPD. Significantly, trajectories between cutting point and end point on the board for slicing or sawing method are defined according to human demonstrations. The complete reproduced trajectory and stiffness profiles are then used to construct an adaptive impedance controller to perform cutting tasks on various objects. Additionally, two control groups with high impedance and low impedance settings are established to conduct the same tasks for comparison. Finally, both the HIC and the proposed AIC successfully complete the cutting tasks, whereas the LIC fails it. For both the HIC and AIC groups, during the period from the object is completely cut through to the end of the cutting action, the contact interface of the knife transitions from the object to the cutting board. Due to the change in environmental impedance, the contact force exhibits a rapid increase, as shown in Fig. 6. In contrast, for the LIC group, the output force is insufficient to complete the cut, resulting in the knife becoming embedded in the middle of the object. Subsequently, we will focus on the HIC and AIC groups.

It is indicated that the HIC and AIC achieve different performances in the contact force as is shown in Fig. 6. Compared to the HIC, the robot using the AIC demonstrates a significant reduction in both the average interaction force (potato: 51.18%, cucumber: 53.89%, carrot: 27.56%, eggplant: 13.33%) and the maximum interaction force (potato: 40.68%, cucumber: 53.66%, carrot: 26.26%, eggplant: 10.38%). Meanwhile, the robotic cutting in HIC mode occasionally fails to follow the reproduced trajectory or produce

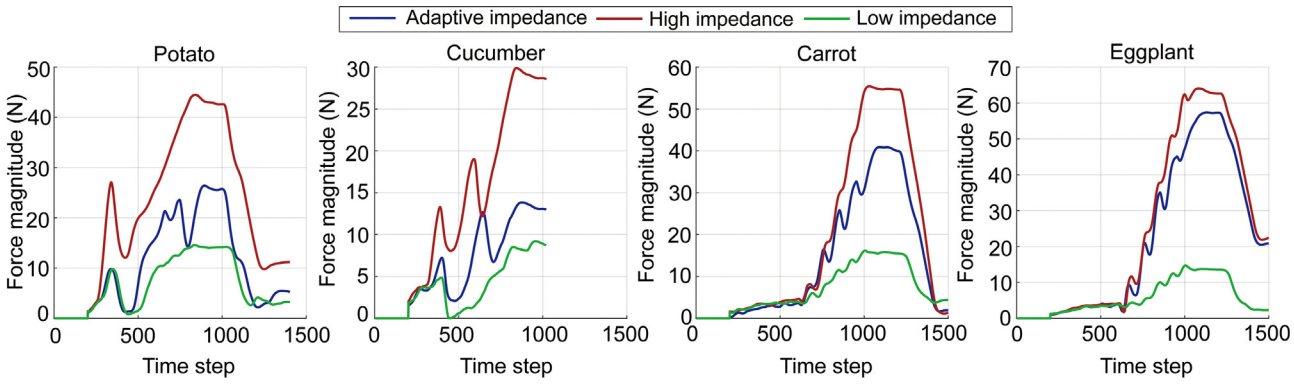

Fig. 6. The comparison results of interaction force using different control modes (HIC, AIC and LIC).

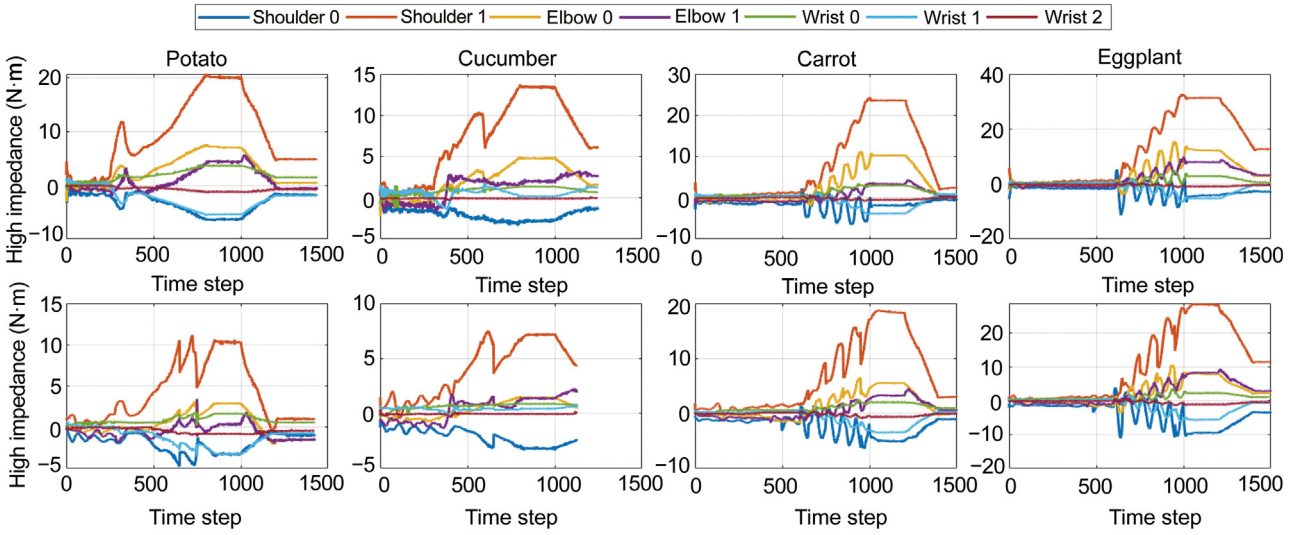

Fig. 7. The comparison results of robot joint torque using different control modes (HIC and AIC).

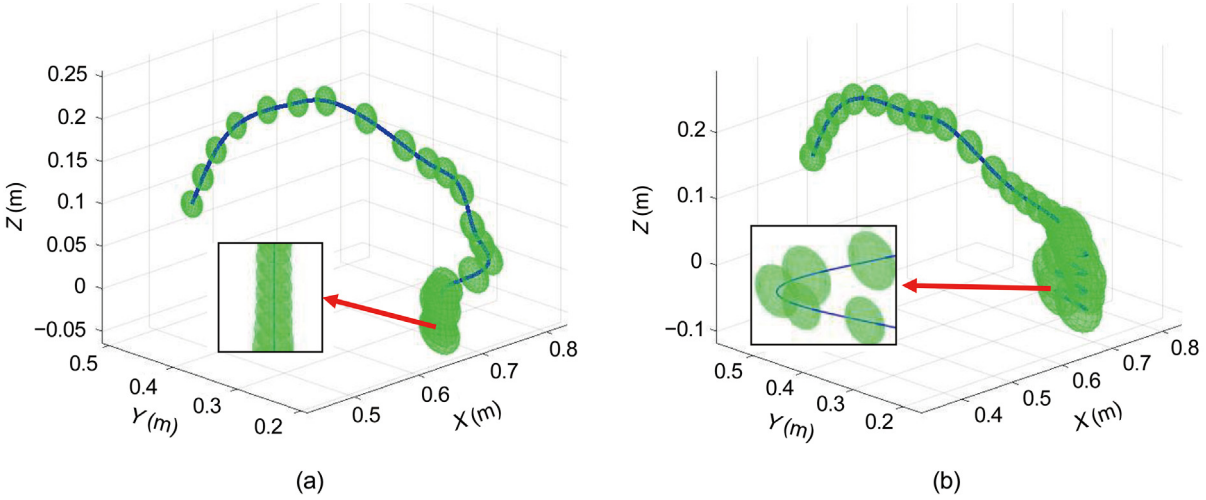

Fig. 8. The reproduced position trajectories and stiffness profiles using (a) slicing and (b) sawing cutting styles.

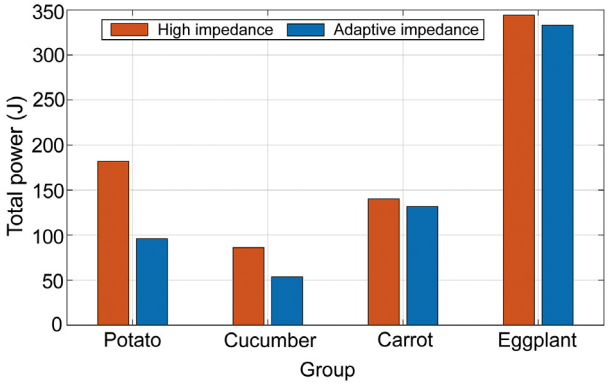

Fig. 9. The comparison of energy consumption between HIC and AIC.

a flat cutting surface due to intense vibrations. Considering that the high contact forces in HIC lead to unstable experimental performance, the proposed AIC not only reduces the interaction forces between the robot and the environment but also improves task execution, enhances system stability, and protects the robot itself. What is more, the AIC would provide a lower torque of each joint compared with HIC, as shown in Fig. 7. During the slicing of eggplant and carrot, both the joint torques of the robot and the interaction forces exhibit noticeable oscillations. This phenomenon arises because slicing is a short-range, back and forth motion, resulting in periodic changes in the end-effector torque and interaction force along the cutting plane. And these oscillations do not affect the overall stability of the system.

To comparatively evaluate the effects of HIC and AIC modalities on robotic power consumption, we formulate the robot joint operation power estimation model as $P_j = \sum_{i=1}^{n} \pi_i \cdot \tau_i \cdot \omega_i$, where $\pi = [0.1, 0.4, 0.2, 0.15, 0.05, 0.05, 0.05]$ is the associated weight of each joint torque defined by the robot structure. τ_i and ω_i are the joint torque and angular velocity. The power consumption results calculated based on the above equations are shown in Fig. 9. Given the inherent variability in the physical properties of the test objects, such as cutting surface length and area, it is ensured that both the control group and the experimental group conducted the experiments at locations on the same object that are in very close proximity and have nearly identical cross-sectional areas. It can be observed that the AIC is capable of reducing torque output and reducing energy consumption while completing the task. Specifically, it can be observed that the power consumption

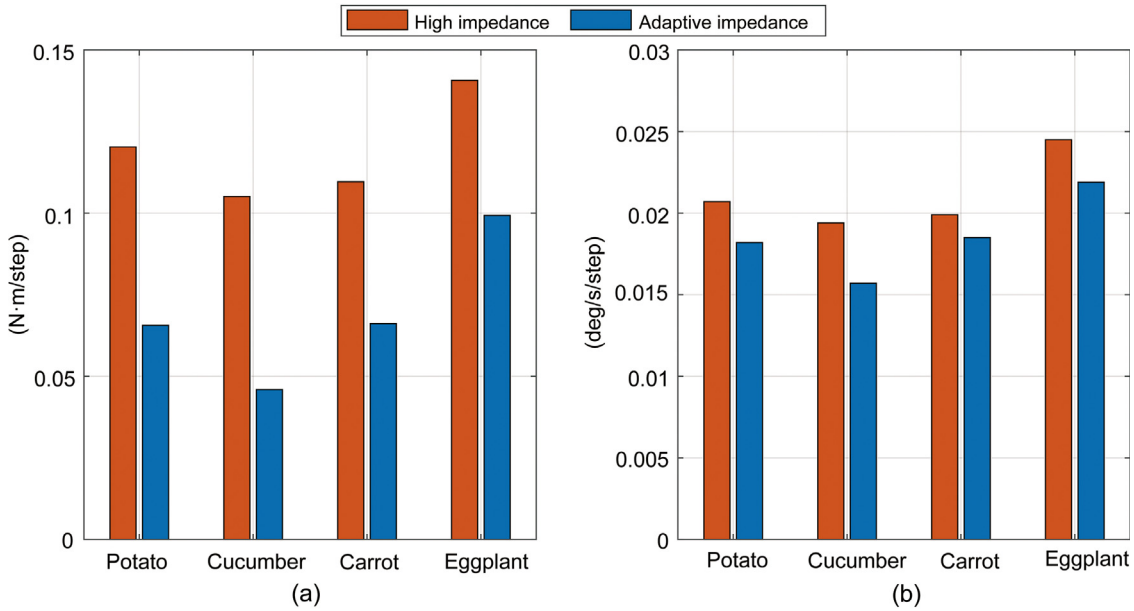

Fig. 10. The comparison of (a) joint torque variation rates (b) angular accelerations between HIC and AIC.

for slicing eggplant is significantly higher than that of the other samples. This can be attributed to the relatively tough skin of the eggplant, which requires a greater slicing force and higher tool stiffness to achieve a successful cut. Furthermore, the greater distance between the knife and the cutting point requires the robot to traverse a longer path during the slicing process.

In addition, the comparisons of joint torque variation rates and angular accelerations are presented in Fig. 10. These results further demonstrate that the proposed method produces a natural and smooth motion trajectory during the cutting process, effectively reducing wear on the robot's movements.

4.3. Discussion

The experimental results demonstrate the superior performance of our proposed method in achieving compliant robot-environment interaction. Notably, we observe that when processing objects with heterogeneous internal density or material composition (e.g., carrots), while greater cutting force was applied, the HIC mode occasionally fails to penetrate the dense core region, whereas the AIC mode successfully completes it. This comparative analysis substantiates the practical significance of implementing variable impedance control in robotic cutting applications, particularly when contrasted with position-controlled robotic systems. Moreover, our vision-based hybrid tree methodology exhibits distinct advantages over previous research in cutting scenarios, manifested through several principal aspects.

The proposed approach eliminates the requirement for precise modeling of cutting forces, frictional dynamics, or fracture mechanics predictions during the cutting process. Instead of depending on accurate measurements of material thickness and deformation, our methodology achieves safe multiple objects cutting through effective skill transfer of the demonstrator's stiffness to the robot.

Human intuitively employ distinct cutting strategies for different objects, as exemplified by the slicing method for potatoes and cucumbers versus the sawing method for carrots and eggplants in above-mentioned experiments. Training a monolithic network to encompass all cutting skills risks compromised accuracy and excessive complexity. To address this, our framework implements a decoupled learning model that separately encodes object-specific cutting strategies with minimal parameterization.

The proposed framework employs visual object detection to facilitate compliant manipulation tasks, circumventing the dependency on high-precision force sensors for perceptual input. Experimental results from cutting trials demonstrate the framework's potentiality in addressing tool-mediated manipulation tasks, as well as applications inherently compatible with implicit force control methodologies.

While the proposed method demonstrates promising performance in the targeted scenarios, it is subject to two practical limitations. First, given that the present study concentrates on implementing the proposed compliant skill transfer framework in contact-rich operations, the issue of workpiece immobilization has been deliberately simplified. However, practical cutting applications may necessitate the coordinated control of additional robotic arms to address this challenge effectively. Additionally, a current limitation of the framework lies in its absence of online relearning capabilities, potentially reduce the accuracy of positioning when encountering substantial changes between initial and target positional parameters. Future work plans to address the above issues and achieve the peeling task via dual-arm tool-in-hand manipulation.

5. Conclusions

This paper proposes a novel vision-based humanoid compliant skill transfer framework. The intuitive demonstration interface combined with a decoupled modeling framework enables efficient and precise acquisition of human skills, while the integration of visual recognition with hybrid tree ensures reliable and rapid skill reproduction. Experimental results demonstrate that the adaptive impedance cutting method based on decoupled skill transfer position trajectories and stiffness ellipsoids can effectively reduce the interaction force, reduce the energy consumption and improve the stability of the system. Notably, the visual-impedance mapping within the learning for control structure significantly enhances human-in-the-loop robot learning efficacy, enabling adaptability across a contact-rich task for multiple objects.

CRediT authorship contribution statement

Zhaohong Mai: Writing – original draft, Software, Methodology. **Chao Zeng:** Writing – review & editing, Supervision, Funding acquisition. **Ning Wang:** Validation, Formal analysis. **Chenguang Yang:** Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the UKRI Guarantee funding for Horizon Europe MSCA Postdoctoral Fellowships (EP/Z00117X/1).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.birob.2026.100280.

References

[1] Botao Lin, Shuang Song, Jiaole Wang, Variable stiffness methods of flexible robots for minimally invasive surgery: A review, Biomim. Intell. Robot. (2024) 100168.
[2] Moses C. Nah, Johannes Lachner, Neville Hogan, Robot control based on motor primitives: A comparison of two approaches, Int. J. Robot. Res. 43 (12) (2024) 1959–1991.
[3] Sami Haddadin, Erfan Shahriari, Unified force-impedance control, Int. J. Robot. Res. 43 (13) (2024) 2112–2141.
[4] Vincenzo Lippiello, Giuseppe Andrea Fontanelli, Fabio Ruggiero, Image-based visual-impedance control of a dual-arm aerial manipulator, IEEE Robot. Autom. Lett. 3 (3) (2018) 1856–1863.
[5] Chenguang Yang, et al., Human-like adaptation of force and impedance in stable and unstable interactions, IEEE Trans. Robot. 27 (5) (2011) 918–930.
[6] Markku Suomalainen, Yiannis Karayiannidis, Ville Kyrki, A survey of robot manipulation in contact, Robot. Auton. Syst. 156 (2022) 104224.
[7] Harish Ravichandar, et al., Recent advances in robot learning from demonstration, Annu. Rev. Control. Robot. Auton. Syst. 3 (1) (2020) 297–330.
[8] Chao Zeng, et al., Hierarchical impedance, force, and manipulability control for robot learning of skills, IEEE/ASME Trans. Mechatronics (2024).
[9] Zhiwei Liao, et al., Simultaneously learning of motion, stiffness, and force from human demonstration based on riemannian dmp and qp optimization, IEEE Trans. Autom. Sci. Eng. (2024).
[10] Chao Zeng, et al., An approach for robotic leaning inspired by biomimetic adaptive control, IEEE Trans. Ind. Inform. 18 (3) (2021) 1479–1488.
[11] Sylvain Calinon, A tutorial on task-parameterized movement learning and retrieval, Intell. Serv. Robot. 9 (2016) 1–29.
[12] Chenzui Li, et al., Towards Robo-Coach: Robot interactive stiffness/position adaptation for human strength and conditioning training, in: 2024 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2024, pp. 860–866.
[13] Martijn J.A. Zeestraten, et al., An approach for imitation learning on Riemannian manifolds, IEEE Robot. Autom. Lett. 2 (3) (2017) 1240–1247.
[14] Sylvain Calinon, Gaussians on Riemannian manifolds: Applications for robot learning and adaptive control, IEEE Robot. Autom. Mag. 27 (2) (2020) 33–45.
[15] Fares J. Abu-Dakka, Matteo Saveriano, Ville Kyrki, A unified formulation of geometry-aware discrete dynamic movement primitives, Neurocomputing 598 (2024) 128056.
[16] Rui Wu, He Zhang, Jie Zhao, Robot variable impedance skill transfer and learning framework based on a simplified human arm impedance model, IEEE Access 8 (2020) 225627–225638.
[17] Yuqiang Wu, et al., A framework for autonomous impedance regulation of robots based on imitation learning and optimal control, IEEE Robot. Autom. Lett. 6 (1) (2020) 127–134.
[18] Noémie Jaquier, Sylvain Calinon, Gaussian mixture regression on symmetric positive definite matrices manifolds: Application to wrist motion estimation with sEMG, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2017, pp. 59–64.
[19] Luca Beber, et al., A passive variable impedance control strategy with viscoelastic parameters estimation of soft tissues for safe ultrasonography, in: 2024 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2024, pp. 1298–1304.
[20] Haochen Zheng, et al., Interaction model estimation-based robotic force-position coordinated optimization for rigid–soft heterogeneous contact tasks, Biomim. Intell. Robot. 5 (1) (2025) 100194.
[21] Matteo Iovino, et al., A survey of behavior trees in robotics and ai, Robot. Auton. Syst. 154 (2022) 104096.
[22] Jacques Cloete, Wolfgang Merkt, Ioannis Havoutis, Adaptive manipulation using behavior trees, 2024, arXiv preprint arXiv:2406.14634.
[23] Simona Gugliermo, et al., Learning behavior trees from planning experts using decision tree and logic factorization, IEEE Robot. Autom. Lett. 8 (6) (2023) 3534–3541.
[24] Kevin Zhang, et al., Leveraging multimodal haptic sensory data for robust cutting, in: 2019 IEEE-RAS 19th International Conference on Humanoid Robots, Humanoids, IEEE, 2019, pp. 409–416.
[25] Prajjwal Jamdagni, Yan-Bin Jia, Robotic cutting of fruits and vegetables: Modeling the effects of deformation, fracture toughness, knife edge geometry, and motion, IEEE Trans. Robot. (2024).
[26] Ryan Wright, et al., Safely and autonomously cutting meat with a collaborative robot arm, Sci. Rep. 14 (1) (2024) 299.
[27] Chenguang Yang, et al., Interface design of a physical human–robot interaction system for human impedance adaptive skill transfer, IEEE Trans. Autom. Sci. Eng. 15 (1) (2017) 329–340.
[28] Zhenjia Xu, et al., RoboNinja: Learning an adaptive cutting policy for multi-material objects, in: Proceedings of Robotics: Science and Systems, RSS, 2023.
[29] Mathew Jose Pollayil, et al., Choosing stiffness and damping for optimal impedance planning, IEEE Trans. Robot. 39 (2) (2022) 1281–1300.
[30] Rajendra Bhatia, Positive Definite Matrices, Princeton University Press, 2009.
[31] Xavier Pennec, Pierre Fillard, Nicholas Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis. 66 (2006) 41–66.
[32] Liwen Situ, et al., Human multi-dimensional stiffness skills transfer for robot teleoperation system, in: 2024 IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE, 2024, pp. 321–327.
[33] Glenn Jocher, Ayush Chaurasia, Jing Qiu, Ultralytics YOLOv8, 2023, URL: https://github.com/ultralytics/ultralytics.
[34] Juan R. Terven, Diana M. Córdova-Esparza, Kin2. a kinect 2 toolbox for MATLAB, Sci. Comput. Program. 130 (2016) 97–106.