big data and cognitive computing

MDPI

Article

Visual Safe Human-to-Humanoid Motion Imitation

Wenqi Cai [1,*], John Abanes [2], Nikolaos Evangeliou [1] and Anthony Tzes [1,3,*]

[1] Robotics and Intelligent Systems Control Lab, New York University Abu Dhabi (NYUAD), Abu Dhabi P.O. Box 129188, United Arab Emirates; nikolaos.evangeliou@nyu.edu
[2] Department of Electrical and Computer Engineering, New York University, New York, NY 10012, USA; john.abanes@nyu.edu
[3] Center for Artificial Intelligence and Robotics, New York University Abu Dhabi (NYUAD), Abu Dhabi P.O. Box 129188, United Arab Emirates
* Correspondence: wenqi.cai@nyu.edu (W.C.); anthony.tzes@nyu.edu (A.T.); Tel.: +971-55-6587107 (W.C.); +971-50-1644728 (A.T.)

Abstract

Safe human-to-humanoid motion imitation is crucial for shared environments, where direct motion retargeting may induce self-collision or human–humanoid collision due to embodiment mismatch, kinematic limits, perception uncertainty, and human proximity. This paper presents an online vision-aided safe human-to-humanoid motion imitation framework that integrates skeleton-based upper-body pose estimation, joint-space retargeting, and capsule-based Control Barrier Function Quadratic Program (CBF-QP) safety filtering. Human skeletal observations are mapped to a reduced eight-degree-of-freedom (8-DoF) humanoid upper-body command, while the CBF-QP layer computes a safety-corrected target that minimally modifies the nominal imitation command subject to robot self-collision and human–humanoid collision constraints. The framework is evaluated via simulation and hardware experiments under representative self-collision and human–humanoid interaction episodes, complemented by a comparative benchmark against velocity damping and potential field baselines. Furthermore, this work introduces an evaluation protocol combining geometric safety, command deviation, and local-link similarity metrics to systematically characterize the safety–imitation trade-off governed by CBF parameters. The results demonstrate that, within the tested moderate-speed regime, the proposed framework substantially reduces geometric collision violations while balancing imitation fidelity with online computational feasibility, thereby providing a viable foundation for safe human-guided humanoid motion deployment.

Keywords: humanoid robots; human–robot interaction; motion imitation; control barrier functions; collision avoidance; safety-critical control

Check for updates

Academic Editor: Jiaming Li

Received: 25 May 2026
Revised: 3 July 2026
Accepted: 22 July 2026
Published: 24 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Humanoid robots are increasingly expected to operate in environments originally designed for humans, including domestic assistance, healthcare support, industrial collaboration, disaster response, and service robotics. Their anthropomorphic morphology allows them to use human-oriented tools, move through human workspaces, and interact with people in a physically intuitive manner [1]. However, programming humanoid robots for diverse real-world tasks remains challenging because of their high-dimensional kinematics, complex dynamics, and strong dependence on task context [2,3]. Human-to-humanoid imitation and teleoperation provide a practical route for transferring human motion intent to humanoid platforms, enabling intuitive command generation without

requiring task-specific low-level programming [4,5]. This direction is also closely related to broader developments in human–robot interaction and safe deployment in human-centered environments [6,7].

A substantial body of work has investigated how human motion can be mapped to humanoid robots. Geometric and kinematic retargeting methods typically construct correspondences between human body segments and robot links, and then solve inverse kinematics, quadratic programming, or optimization problems to generate feasible robot references [8–10]. More recent vision-based approaches have explored monocular, RGB-D, and skeleton-based perception for human motion retargeting, where estimated human motion is converted into humanoid joint commands for real-time teleoperation [11,12]. In parallel, learning-based imitation and teleoperation methods have improved the expressiveness and robustness of humanoid motion generation by combining large-scale motion data, retargeting, reinforcement learning (RL), behavior cloning, and sim-to-real transfer [13–16]. These studies have significantly advanced human-to-humanoid motion transfer, but they primarily focus on motion feasibility, tracking quality, naturalness, and teleoperation performance.

Despite these advances, accurate imitation alone is not sufficient for deploying humanoid robots in shared environments. Human motions may be natural for the demonstrator but unsafe for the robot because of embodiment mismatch, different limb proportions, joint limits, actuator constraints, perception latency, and the presence of nearby humans [17]. For example, crossing arms, reaching across the torso, or moving close to another person can induce robot self-collision or human–humanoid collision after retargeting. These risks are particularly relevant in physical human–robot interaction and collaborative workspaces, where safety is not only a control requirement but also a prerequisite for practical deployment [18,19]. Existing safety studies in human–robot collaboration emphasize collision modeling, separation monitoring, speed limitation, and verification of safe interaction conditions. However, most human-to-humanoid imitation studies do not explicitly address the safety risks induced by the imitation process itself.

Control barrier functions (CBFs) provide a principled framework for enforcing safety constraints in real time. The CBF-QP formulation introduced by Ames et al. unifies safety conditions, expressed as forward-invariant sets, with performance objectives through quadratic programming [20]. A key advantage of CBF-QP methods is that they can minimally modify a nominal control input while enforcing state-dependent safety constraints [21,22]. This property is well suited to imitation control: the nominal retargeted command represents the desired human-like motion, while the CBF constraints provide a systematic mechanism for preventing unsafe configurations. Thus, the safety layer can be interpreted as an online filter that preserves imitation as much as possible, subject to formally specified safety inequalities.

Recent advances in safety-control methods have extended CBF-based and other related approaches to collision avoidance and interaction safety constraints in robotic systems [23–27]. To position the present work within the broader literature, Table 1 summarizes representative works in terms of their task focus, platform, imitation capability, self-collision safety, human–robot safety, and safety mechanism. These studies provide important foundations, but they primarily focus on their respective task settings, such as humanoid teleoperation, robot-internal safety control, human–robot separation, general collision avoidance, or contact-safe robot behavior, rather than on safety filtering within a real-time human-to-humanoid imitation pipeline. Moreover, the trade-off between safety enforcement and imitation fidelity introduced by such filtering has received limited systematic attention.

Table 1. A brief comparison of representative related works (not comprehensive).

Category	Works	Platform	Motion Imitation/Retargeting	Self-Collision Safety	Human–Robot Safety	Safety Mechanism	Key Focus/Distinction
Humanoid teleoperation and retargeting	[13–15]	Humanoid	Yes	Partial	No	Retargeting, constrained IK, joint-limit handling, or feasibility checking	Focuses on motion transfer quality and teleoperation feasibility, without explicit safety filtering.
Humanoid self-collision safety control	[28–30]	Humanoid/ legged robots	No	Yes	No	CBF constraints, Riemannian Motion Policy (RMP), or MPC-based collision handling	Focuses on robot self-collision or whole-body collision handling, without addressing vision-based imitation in shared space.
CBF-based safety control for HRC/HRI	[31–33]	Manipulators/ cobots	No	No/varies	Yes	CBF-QP constraints, speed regulation, and ISO-inspired separation monitoring	Focuses on enforcing human–robot safety in collaborative tasks, rather than preserving retargeted humanoid imitation behavior.
Reactive collision avoidance (Non-CBF)	[34–37]	Manipulators/ mobile robots	No	Varies	Varies	Safe sets, velocity damping, potential field, or predictive collision avoidance	Provides general collision-avoidance tools for navigation or manipulation, rather than a safety layer for humanoid imitation.
Soft-robot contact safety control	[38,39]	Soft robots	No	Partial	Contact safety	CBF-based contact, force, or material safety constraints	Addresses contact safety for compliant or deformable robots, differing from collision avoidance in rigid humanoid imitation.
This work	–	Humanoid	Yes	Yes	Yes	Capsule-based CBF-QP filtering	Integrates vision-based imitation, self-collision avoidance, and human–humanoid collision avoidance in one online humanoid framework.

This paper addresses the above gap by developing a safety-aware online human-to-humanoid upper-body imitation framework, as shown in Figure 1. The framework maps camera-observed human skeletal motion to an eight-degree-of-freedom (8-DoF) humanoid upper-body representation, generates nominal robot joint commands through retargeting, and applies a capsule-based CBF-QP safety filter before execution. By placing the safety layer between nominal retargeting and robot execution, the framework explicitly separates the nominal imitation command from the CBF-corrected command, making it possible to analyze how geometric safety enforcement affects imitation behavior. The evaluation therefore combines geometric safety metrics, including minimum clearance, violation time, and collision count, with imitation metrics based on joint-space deviation and local-link direction similarity after temporal alignment. These metrics are used to assess how effectively the framework reduces self-collision and human–humanoid proximity violations while preserving the structure of the demonstrated motion. They also provide a common basis for comparing the proposed CBF-QP with alternative safety filters (velocity-damping and potential-field baselines), and for analyzing how CBF margins and gains shape different safety–imitation operating regimes.

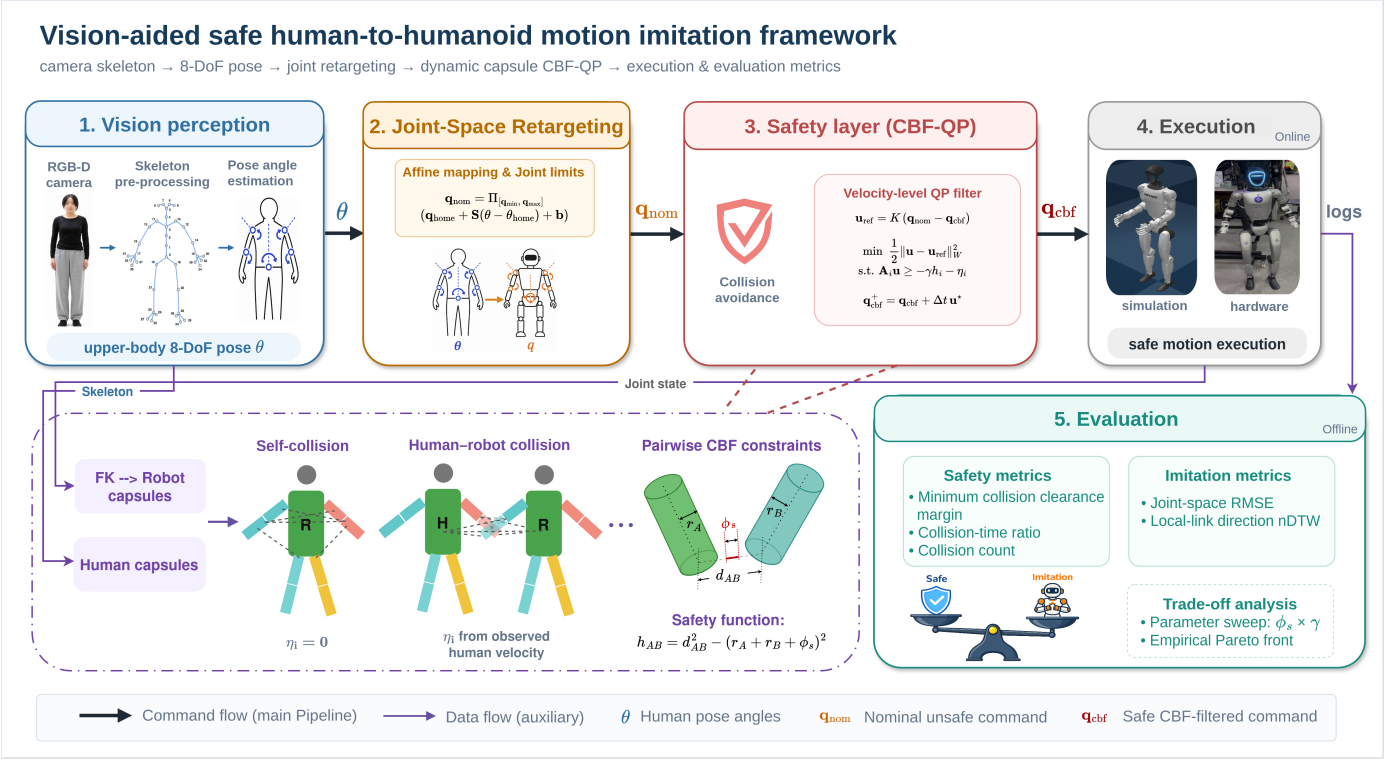


Figure 1. Overview of the proposed safe human-to-humanoid motion imitation framework.

The main **contributions** of this paper are summarized as follows:

- A complete online safe-imitation problem formulation is presented for vision-based human-to-humanoid motion imitation. Unlike studies that treat retargeting, self-collision avoidance, and human–humanoid separation separately, this work considers them within a single modular pipeline, from camera-observed human motion to safe humanoid execution. This modular structure also provides a baseline framework that can be extended to richer perception modules or higher-DoF humanoid platforms.
- A task-specific safety layer is developed for human-to-humanoid imitation by adapting capsule-based CBF-QP filtering to nominal retargeted joint commands. The safety layer enables the humanoid robot to avoid both self-collision and human–humanoid collision while preserving the qualitative structure of the imitated human motion.

- The proposed pipeline is evaluated through matched simulation and online hardware experiments in representative self-collision and human–humanoid interaction episodes. The results show that the CBF-QP safety layer substantially reduces unsafe behavior and enables safety-aware imitation in both simulation and real robot execution. A supplementary simulation comparison against alternative safety-filter baselines is also included to contextualize the CBF-QP behavior.
- A systematic safety–imitation evaluation protocol is introduced for CBF parameter analysis. By comparing nominal and CBF-filtered commands through complementary safety, imitation, and feasibility metrics, this protocol quantifies the safety–imitation trade-off and provides empirical guidance for selecting CBF margins and gains in safe imitation tasks.

The remainder of this paper is organized as follows. Section 2 presents the proposed safety-aware human-to-humanoid imitation framework. Specifically, Section 2.1 describes the vision-based upper-body motion representation and retargeting pipeline, Section 2.2 formulates the CBF-QP safety filter, and Section 2.3 defines the safety and imitation metrics used for quantitative evaluation. The simulation and hardware results are reported in Section 3, together with a safety–imitation trade-off analysis under different CBF parameters. Section 4 discusses the implications and limitations of the proposed framework, and finally, we conclude this work in Section 5.

2. Materials and Methods

2.1. Human-to-Humanoid Motion Retargeting

2.1.1. Vision-Based Pose Angle Estimation

This work estimates a reduced set of human upper-body joint variables from vision-based skeleton observations. Rather than solving a full inverse kinematics problem from noisy 3D keypoints, we directly extract an 8-DoF upper-body representation from camera-estimated 3D body-skeleton keypoints. The reduced representation captures the dominant torso and arm motions required for humanoid upper-body retargeting while remaining compact, interpretable, and structurally less sensitive to coordinate fluctuations.

Let $\boldsymbol{\theta} = \begin{bmatrix} \theta^{\text{to}}_{\text{roll}} & \theta^{\text{to}}_{\text{pitch}} & \theta^{\text{sh,L}}_{\text{pitch}} & \theta^{\text{sh,L}}_{\text{roll}} & \theta^{\text{el,L}}_{\text{pitch}} & \theta^{\text{sh,R}}_{\text{pitch}} & \theta^{\text{sh,R}}_{\text{roll}} & \theta^{\text{el,R}}_{\text{pitch}} \end{bmatrix}^T \in \mathbb{R}^8$ denote the reduced human upper-body pose. The first two components, $(\theta^{\text{to}}_{\text{roll}}, \theta^{\text{to}}_{\text{pitch}})$, describe torso inclination, while the remaining six components describe shoulder and elbow motion for the left and right arms, $(\theta^{\text{sh},i}_{\text{pitch}}, \theta^{\text{sh},i}_{\text{roll}}, \theta^{\text{el},i}_{\text{pitch}})$, with $i \in \{\text{L}, \text{R}\}$. This representation separates global upper-body orientation from local arm articulation and provides the input variables used by the subsequent robot-side retargeting map.

Human keypoints are represented in the camera global frame $\mathcal{F}_Z$, which is used as the working world frame of the perception system. A local torso frame $\mathcal{F}_T$ is then constructed online from the pelvis and bilateral shoulder keypoints. This body-centered frame is used to express upper-arm motion relative to the human trunk. The torso frame $\mathcal{F}_T$ is constructed from three skeleton landmarks: pelvis, left shoulder, and right shoulder. Its origin is placed at the pelvis. The upward axis is defined from the pelvis to the shoulder center, the lateral axis is obtained from the shoulder line after orthogonalization with respect to the upward axis, and the remaining axis is computed by a right-handed cross product. This construction yields an orthonormal local frame that adapts to the current trunk orientation. The torso roll and pitch angles are computed from the orientation of the torso frame relative to the global vertical direction. The shoulder angles are computed from the upper-arm vectors expressed in $\mathcal{F}_T$, so that arm motion is measured relative to the trunk rather than the external world. The elbow-pitch angles are computed from the angle between the corresponding upper-arm and forearm segments. Accordingly, the full 8-DoF

representation is estimated directly from the seven keypoints $\{P, LS, RS, LE, RE, LW, RW\}$. Note that before retargeting, several preprocessing and filtering steps are applied to reduce the effect of noisy or invalid keypoint observations, including confidence-based rejection, validity checks, temporal smoothing, and rate limiting. A compact sensitivity analysis is provided in Appendix A.4.

The final quantities used in this work are the following:

$$\begin{aligned}
\theta^{\text{to}}_{\text{roll}} &= \text{atan2}\Big(-({}^{Z}\mathbf{e}_{T,y})^{T}({}^{Z}\mathbf{k}\times{}^{Z}\mathbf{e}_{T,z}),({}^{Z}\mathbf{k})^{T}{}^{Z}\mathbf{e}_{T,z}\Big),\\
\theta^{\text{to}}_{\text{pitch}} &= \text{atan2}\Big(-({}^{Z}\mathbf{e}_{T,x})^{T}({}^{Z}\mathbf{k}\times{}^{Z}\mathbf{e}_{T,z}),({}^{Z}\mathbf{k})^{T}{}^{Z}\mathbf{e}_{T,z}\Big),\\
\theta^{\text{sh,L}}_{\text{pitch}} &= \text{atan2}\Big(u^{T}_{L,y},-u^{T}_{L,z}\Big), \qquad \theta^{\text{sh,R}}_{\text{pitch}} = \text{atan2}\Big(u^{T}_{R,y},-u^{T}_{R,z}\Big),\\
\theta^{\text{sh,L}}_{\text{roll}} &= \text{atan2}\Big(-u^{T}_{L,x},-u^{T}_{L,z}\Big), \qquad \theta^{\text{sh,R}}_{\text{roll}} = \text{atan2}\Big(u^{T}_{R,x},-u^{T}_{R,z}\Big),\\
\theta^{\text{el,L}}_{\text{pitch}} &= \arccos\Big(\frac{({}^{Z}\mathbf{u}_{L})^{T}{}^{Z}\mathbf{f}_{L}}{\|{}^{Z}\mathbf{u}_{L}\|\,\|{}^{Z}\mathbf{f}_{L}\|}\Big), \qquad \theta^{\text{el,R}}_{\text{pitch}} = \arccos\Big(\frac{({}^{Z}\mathbf{u}_{R})^{T}{}^{Z}\mathbf{f}_{R}}{\|{}^{Z}\mathbf{u}_{R}\|\,\|{}^{Z}\mathbf{f}_{R}\|}\Big).
\end{aligned} \tag{1}$$

where ${}^{Z}\mathbf{k} = [0\ 0\ 1]^{T}$ is the global upward unit vector; ${}^{Z}\mathbf{e}_{T,x}$, ${}^{Z}\mathbf{e}_{T,y}$, and ${}^{Z}\mathbf{e}_{T,z}$ are the torso-frame axes expressed in $\mathcal{F}_Z$; and ${}^{Z}\mathbf{u}_{L}$, ${}^{Z}\mathbf{u}_{R}$, ${}^{Z}\mathbf{f}_{L}$, and ${}^{Z}\mathbf{f}_{R}$ denote the left/right upper-arm and forearm vectors expressed in $\mathcal{F}_Z$. The terms $u^{T}_{L,x}$, $u^{T}_{L,y}$, $u^{T}_{L,z}$, $u^{T}_{R,x}$, $u^{T}_{R,y}$, and $u^{T}_{R,z}$ are the components of the left/right upper-arm vectors expressed in $\mathcal{F}_T$. This reduced pose vector is used as the input to the human-to-humanoid retargeting module. Detailed coordinate-frame definitions and implementation details are provided in Appendix A.

2.1.2. Joint-Space Retargeting

The humanoid upper-body imitation space is defined as

$$\mathbf{q} = \begin{bmatrix} q_{\text{wr}} & q_{\text{wp}} & q^{L}_{\text{sp}} & q^{L}_{\text{sr}} & q^{L}_{\text{ep}} & q^{R}_{\text{sp}} & q^{R}_{\text{sr}} & q^{R}_{\text{ep}} \end{bmatrix}^{T} \in \mathbb{R}^{8}, \tag{2}$$

where q_{wr} and q_{wp} denote waist roll and pitch, respectively, and $q^{L/R}_{\text{sp}}$, $q^{L/R}_{\text{sr}}$, and $q^{L/R}_{\text{ep}}$ denote the left/right shoulder-pitch, shoulder-roll, and elbow-pitch joints. Let $\boldsymbol{\theta}_{\text{home}}$ denote the calibrated neutral human pose and $\mathbf{q}_{\text{home}}$ denote the robot home configuration. The nominal retargeting map is defined as

$$\mathbf{q}_{\text{nom}} = \text{clip}\Big(\mathbf{q}_{\text{home}} + \mathbf{S}\odot(\boldsymbol{\theta}-\boldsymbol{\theta}_{\text{home}}) + \mathbf{b},\ \mathbf{q}_{\min},\ \mathbf{q}_{\max}\Big) \tag{3}$$

where $\mathbf{S}\in\mathbb{R}^{8}$ is an element-wise retargeting vector accounting for both joint-axis alignment and range scaling. Its entries have unit magnitude, with signs selected to match the robot and human coordinate conventions. The offset vector $\mathbf{b}\in\mathbb{R}^{8}$ is set to zero, and the home configuration $\mathbf{q}_{\text{home}}$ is chosen as the neutral upper-body posture used by the controller to provide a repeatable robot-side baseline. At system startup, the user-specific neutral pose $\boldsymbol{\theta}_{\text{home}}$ is automatically established by averaging the measured human joint angles over a brief initialization window while the operator holds a relaxed posture. Consequently, the online retargeting mapping relies on the relative displacement $\boldsymbol{\theta}-\boldsymbol{\theta}_{\text{home}}$, reducing sensitivity to variations in the absolute standing posture across operators. The resulting $\mathbf{q}_{\text{nom}}$ is the nominal imitation command before safety filtering and is forwarded to the CBF module as the unfiltered upper-body joint command.

2.2. CBF-QP Safety Control

2.2.1. Capsule-Based Collision Model

To obtain a compact differentiable collision representation, the active CBF collision model represents both the robot and the human using seven main capsules: $\mathcal{B}_{\{\text{robot,human}\}} =$

{`torso, L/R upper arm, L/R forearm/hand, L/R thigh`}. Each capsule is parameterized by two endpoints and a radius, as detailed in Table 2. All capsule endpoints are expressed in the robot reference frame before distance evaluation. Robot endpoints are updated via forward kinematics, whereas human capsule lengths vary with the observed user skeleton keypoints. The capsule radii are fixed, with the configurable parameters chosen slightly conservatively in this study to provide a robust safety buffer across various user statures. Additional human primitives, such as the shins and a head sphere, are maintained for geometric completeness but are excluded from the active CBF pair list because they are not involved in the targeted upper-body interaction episodes. The monitored human–humanoid pairs focus on the dominant safety-critical proximities in this study, including robot arm–human arm, robot arm–human torso, robot torso–human arm, robot arm–human thigh, and torso–torso proximity. This targeted pair selection keeps the online QP compact while covering the collision modes evaluated in the experiments.

Table 2. Geometric parameters of the collision capsules.

Body Segment	Endpoint Definition	Human Radius	Robot Radius
Torso	pelvis–neck	0.15 m	0.10 m
Upper arm	shoulder–elbow	0.06 m	0.065 m
Forearm/hand	elbow–extended wrist	0.05 m	0.065 m
Thigh	hip–knee	0.08 m	0.065 m

2.2.2. Collision-Avoidance CBF Constraints

Consider a capsule pair A and B, with radii r_A and r_B, endpoints $(\mathbf{a}_A, \mathbf{b}_A)$ and $(\mathbf{a}_B, \mathbf{b}_B)$, and closest points $\mathbf{p}_A$ and $\mathbf{p}_B$. The closest-point distance is defined as $d_{AB}(\mathbf{x}_{AB}) = \|\mathbf{p}_A - \mathbf{p}_B\|_2$, where $\mathbf{x}_{AB} := [\mathbf{a}_A^\top, \mathbf{b}_A^\top, \mathbf{a}_B^\top, \mathbf{b}_B^\top]^\top$ collects the capsule endpoint positions. The corresponding capsule pair barrier function is defined as

$$h_{AB}(\mathbf{x}_{AB}) = d_{AB}^2(\mathbf{x}_{AB}) - (r_A + r_B + \phi_s)^2, \tag{4}$$

where $\phi_s > 0$ is the prescribed safety-distance margin. Therefore, $h_{AB} > 0$ indicates that the capsule pair satisfies the prescribed safety boundary, $h_{AB} = 0$ corresponds to the boundary, and $h_{AB} < 0$ indicates violation of the safety-distance margin.

The first-order CBF condition is imposed as

$$\dot{h}_{AB} + \gamma h_{AB} \geq 0, \tag{5}$$

where $\gamma > 0$ is the CBF gain. In this work, the pair-type parameters $(\phi_s^{rr}, \gamma^{rr})$ and $(\phi_s^{hr}, \gamma^{hr})$ are used for robot–robot self-collision pairs and human–humanoid pairs, respectively. The CBF condition constrains the evolution of h_{AB}: it permits approach when the pair remains safely separated, restricts further decrease near the safety boundary, and induces recovery when the prescribed margin is violated. For velocity-level safety filtering, Equation (5) is written as a constraint on the robot joint velocity $\mathbf{u} := \dot{\mathbf{q}}$. By the chain rule, the barrier derivative is

$$\dot{h}_{AB} = \frac{\partial h_{AB}}{\partial \mathbf{x}_{AB}} \dot{\mathbf{x}}_{AB}, \tag{6}$$

where the endpoint velocity $\dot{\mathbf{x}}_{AB}$ is determined by the capsule pair type, as detailed in the following two cases.

(1) Robot self-collision. For a robot–robot pair, both capsules are attached to the robot; hence, all capsule endpoint velocities are induced by the robot joint velocity $\mathbf{u}$. The endpoint velocity vector is defined as

$$\dot{\mathbf{x}}_{AB} = \mathbf{J}_{AB}(\mathbf{q})\mathbf{u}, \tag{7}$$

where the stacked endpoint Jacobian is $\mathbf{J}_{AB}(\mathbf{q}) = \begin{bmatrix} \mathbf{J}_{\mathbf{a}_A}^\top & \mathbf{J}_{\mathbf{b}_A}^\top & \mathbf{J}_{\mathbf{a}_B}^\top & \mathbf{J}_{\mathbf{b}_B}^\top \end{bmatrix}^\top$. Substituting Equation (7) into Equation (6) yields

$$\dot{h}_{AB} = \frac{\partial h_{AB}}{\partial \mathbf{x}_{AB}} \mathbf{J}_{AB}(\mathbf{q})\mathbf{u} = \mathbf{A}_{AB}(\mathbf{q})\mathbf{u}, \tag{8}$$

where $\mathbf{A}_{AB}(\mathbf{q}) = \frac{\partial h_{AB}}{\partial \mathbf{a}_A}\mathbf{J}_{\mathbf{a}_A} + \frac{\partial h_{AB}}{\partial \mathbf{b}_A}\mathbf{J}_{\mathbf{b}_A} + \frac{\partial h_{AB}}{\partial \mathbf{a}_B}\mathbf{J}_{\mathbf{a}_B} + \frac{\partial h_{AB}}{\partial \mathbf{b}_B}\mathbf{J}_{\mathbf{b}_B}$. Thus, robot self-collision constraints take the velocity-level CBF form, with the barrier derivative linear in $\mathbf{u}$.

(2) Human–humanoid collision. For a human–humanoid pair, the same barrier function is used, but only the robot capsule is controlled by $\mathbf{u}$. Let R denote the robot capsule and H denote the human capsule, and define $\mathbf{x}_{RH} := [\mathbf{a}_R^\top, \mathbf{b}_R^\top, \mathbf{a}_H^\top, \mathbf{b}_H^\top]^\top$. The robot endpoint velocities are induced by $\mathbf{u}$, whereas the human endpoint velocities $\mathbf{v}_{\mathbf{a}_H}$ and $\mathbf{v}_{\mathbf{b}_H}$ are estimated from consecutive observed capsule endpoints. The endpoint velocity vector is therefore written as

$$\dot{\mathbf{x}}_{RH} = \begin{bmatrix} \mathbf{J}_{\mathbf{a}_R}^\top & \mathbf{J}_{\mathbf{b}_R}^\top & \mathbf{0}^\top & \mathbf{0}^\top \end{bmatrix}^\top \mathbf{u} + \begin{bmatrix} \mathbf{0}^\top & \mathbf{0}^\top & \mathbf{v}_{\mathbf{a}_H}^\top & \mathbf{v}_{\mathbf{b}_H}^\top \end{bmatrix}^\top. \tag{9}$$

Substituting Equation (9) into Equation (6) gives

$$\dot{h}_{RH} = \mathbf{A}_{RH}(\mathbf{q})\mathbf{u} + \eta_{RH}, \tag{10}$$

where the robot-controlled term is defined as $\mathbf{A}_{RH}(\mathbf{q}) = \frac{\partial h_{RH}}{\partial \mathbf{a}_R}\mathbf{J}_{\mathbf{a}_R} + \frac{\partial h_{RH}}{\partial \mathbf{b}_R}\mathbf{J}_{\mathbf{b}_R}$, and the observed-human-motion term is defined as

$$\eta_{RH} = \frac{\partial h_{RH}}{\partial \mathbf{a}_H}\mathbf{v}_{\mathbf{a}_H} + \frac{\partial h_{RH}}{\partial \mathbf{b}_H}\mathbf{v}_{\mathbf{b}_H}. \tag{11}$$

The term η_{RH} represents the instantaneous change in the barrier value caused by the observed motion of the human capsule. This term shifts the admissible robot velocity set according to the measured human motion, thereby accounting for moving human obstacles without introducing human joint variables or increasing the QP decision dimension.

Combining the two cases, each active CBF constraint can be written as

$$\mathbf{A}_i(\mathbf{q})\mathbf{u} \geq -\gamma_i h_i(\mathbf{q}) - \eta_i, \tag{12}$$

where i indexes the active constraints constructed from the monitored capsule pair set, and

$$\eta_i = \begin{cases} 0, & \text{robot self-collision,} \\ \dfrac{\partial h_i}{\partial \mathbf{a}_H}\mathbf{v}_{\mathbf{a}_H} + \dfrac{\partial h_i}{\partial \mathbf{b}_H}\mathbf{v}_{\mathbf{b}_H}, & \text{human–humanoid collision.} \end{cases} \tag{13}$$

Thus, robot self-collision constraints correspond to $\eta_i = 0$, whereas human–humanoid constraints include the measured human endpoint-velocity contribution. Both cases remain linear in the robot joint velocity $\mathbf{u}$ and can therefore be incorporated directly into the CBF-QP safety filter.

2.2.3. CBF-QP Safety Filter

The retargeting module provides a nominal joint-position command $\mathbf{q}_{\text{nom}} \in \mathbb{R}^8$. The safety filter maintains an internal safe target $\mathbf{q}_{\text{cbf}} \in \mathbb{R}^8$, which is updated at the velocity level. At each update, a nominal reference velocity is generated as

$$\mathbf{u}_{\text{ref}} = K(\mathbf{q}_{\text{nom}} - \mathbf{q}_{\text{cbf}}), \tag{14}$$

where $K > 0$ is a proportional gain. This velocity drives the safe target toward the nominal retargeted command in the absence of active safety constraints.

To improve feasibility under constraint conflicts—such as during sudden human movements or when perception yields noisy skeleton estimates—the safe joint velocity is obtained by solving a soft-constrained CBF-QP as follows:

$$\begin{aligned} \mathbf{u}^\star, \boldsymbol{\xi}^\star = \arg\min_{\mathbf{u},\boldsymbol{\xi}} \quad & \frac{1}{2}\|\mathbf{u} - \mathbf{u}_{\text{ref}}\|_W^2 + \frac{\rho}{2}\|\boldsymbol{\xi}\|_2^2 \\ \text{s.t.} \quad & \mathbf{A}_i(\mathbf{q})\mathbf{u} + \xi_i \geq -\gamma_i h_i(\mathbf{q}) - \eta_i, \quad i = 1, \ldots, N_c, \\ & \xi_i \geq 0, \quad i = 1, \ldots, N_c, \\ & \mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max}, \end{aligned} \tag{15}$$

where $W \succ 0$ weights deviation from the nominal reference velocity, N_c is the number of active CBF constraints, and $\mathbf{u}_{\min}$, $\mathbf{u}_{\max}$ are joint-velocity limits. The non-negative slack variables $\boldsymbol{\xi} \in \mathbb{R}^{N_c}_{\geq 0}$ are heavily penalized by $\rho > 0$ and are used only when the hard CBF inequalities conflict with the joint-velocity bounds, allowing the QP to return a penalized least-violating command. As a system-level safeguard against severe anomalies (e.g., total perception loss or solver timeout), the controller rejects the new command and holds the previous safe target ($\mathbf{u}^\star = \mathbf{0}$). Additionally, the skeleton endpoint velocities in η_i are low-pass filtered to suppress measurement jitter in practical deployment. Nevertheless, it is worth noting that the proposed CBF-QP layer should be interpreted as a geometric separation filter rather than a complete physical safety certification, as it does not directly model impact dynamics, torque saturation, actuator power limits, or whole-body balance.

The internal safe target is then integrated as

$$\mathbf{q}^+_{\text{cbf}} = \mathbf{q}_{\text{cbf}} + \Delta t\, \mathbf{u}^\star, \tag{16}$$

and $\mathbf{q}^+_{\text{cbf}}$ is sent to the downstream position controller. Notably, for all experiments reported in this paper, the hard CBF constraints were feasible across all evaluated windows; thus, $\boldsymbol{\xi}^\star = \mathbf{0}$ holds throughout, meaning that no constraint relaxation was required for the reported trajectories.

2.3. Evaluation Metrics for Safety and Imitation

To quantitatively evaluate the proposed safety filter, we define two groups of metrics. The safety metrics quantify the collision-avoidance performance of the generated motion with respect to all monitored capsule pairs, whereas the imitation metrics characterize the similarity between the generated robot motion and the intended imitation behavior. Let T denote the number of sampled time steps and let $\mathcal{P}$ denote the set of monitored capsule pairs. We distinguish the safety-distance margins used in the CBF controller from the geometric quantity used for offline evaluation. The CBF-QP uses ϕ_s^{rr} and ϕ_s^{hr} as control-side safety-distance margins to enforce conservative separation. In contrast, the offline safety

metrics are computed with respect to the physical collision boundary. For each monitored capsule pair $i \in \mathcal{P}$, we define the collision clearance margin as

$$m_i(t) = d_i(t) - \left(r_{A_i} + r_{B_i}\right), \tag{17}$$

where $d_i(t)$ is the closest-point distance between the two capsule centerlines, and r_{A_i} and r_{B_i} are the corresponding capsule radii. Thus, $m_i(t) > 0$ indicates positive surface clearance, $m_i(t) = 0$ indicates contact, and $m_i(t) < 0$ indicates geometric collision.

2.3.1. Safety Metrics

Three complementary safety metrics are used: minimum collision clearance margin, collision-time ratio, and collision count.

(1) Minimum collision clearance margin. The minimum collision clearance margin is defined as the global minimum of the collision clearance margin over all monitored pairs and all time steps:

$$M_{\text{clear}} = \min_{t=1,\ldots,T} \min_{i \in \mathcal{P}} m_i(t). \tag{18}$$

This metric captures the worst-case geometric separation along the trajectory. A larger M_{clear} indicates greater clearance from physical contact, whereas $M_{\text{clear}} < 0$ indicates that at least one monitored capsule pair entered geometric collision.

(2) Collision-time ratio. The collision indicator at time t is defined as $I_{\text{unsafe}}(t) = \mathbb{I}[\exists i \in \mathcal{P},\ m_i(t) < 0]$. The collision-time ratio is then calculated as

$$M_{\text{ctr}} = \frac{1}{T} \sum_{t=1}^{T} I_{\text{unsafe}}(t). \tag{19}$$

This metric measures the fraction of the trajectory during which at least one monitored capsule pair is in geometric collision. Therefore, smaller values indicate safer behavior.

(3) Collision count. A collision event is counted when the trajectory enters a collision episode from a previously collision-free state. The collision count is defined as

$$M_{\text{cc}} = \sum_{t=2}^{T} \mathbb{I}[I_{\text{unsafe}}(t) = 1 \wedge I_{\text{unsafe}}(t-1) = 0]. \tag{20}$$

This metric measures how many separate collision episodes occur over the trajectory. It complements the collision-time ratio by distinguishing repeated short collisions from one continuous collision interval. Smaller values indicate fewer collision occurrences.

To support quantitative comparison across evaluated runs, the raw safety metrics are first converted into direction-consistent normalized scores. Let $\mathcal{R}$ denote the set of evaluated runs in a given comparison, and let $\epsilon > 0$ be a small constant used to avoid division by zero. For a metric value x_ρ from run ρ, the min–max normalization over $\mathcal{R}$ is defined as

$$\mathcal{N}(x_\rho) = \frac{x_\rho - \min_{\rho' \in \mathcal{R}} x_{\rho'}}{\max_{\rho' \in \mathcal{R}} x_{\rho'} - \min_{\rho' \in \mathcal{R}} x_{\rho'} + \epsilon}. \tag{21}$$

Since larger values of M_{clear} indicate safer behavior, its normalized safety contribution is $\widehat{M}_{\text{clear}} = \mathcal{N}(M_{\text{clear}})$. For the collision-time ratio M_{ctr} and collision count M_{cc}, smaller values indicate safer behavior. Their normalized safety contributions are therefore obtained by reversing the min–max normalization: $\widehat{M}_{\text{ctr}} = 1 - \mathcal{N}(M_{\text{ctr}})$ and $\widehat{M}_{\text{cc}} = 1 - \mathcal{N}(M_{\text{cc}})$. The composite safety score $S_{\text{safe}} \in [0, 1]$ is then a normalized, higher-is-better metric defined as

$$S_{\text{safe}} = w_1 \widehat{M}_{\text{clear}} + w_2 \widehat{M}_{\text{ctr}} + w_3 \widehat{M}_{\text{cc}}, \qquad w_1 + w_2 + w_3 = 1, \quad w_{1,2,3} \geq 0. \tag{22}$$

2.3.2. Imitation Metrics

Two complementary imitation metrics are used to evaluate different aspects of motion similarity. The first metric is a zero-lag joint-space measure that quantifies how much the CBF-QP modifies the nominal retargeted command at the command level. This metric directly reflects the intervention magnitude of the safety layer. The second metric is a local-link direction metric with temporal alignment. It evaluates whether the resulting upper-body motion remains geometrically similar to the human motion after allowing for time-warping effects caused by velocity-level safety filtering. These two metrics therefore answer different questions: RMSE_q^0 measures how much the safety layer changes the nominal robot command, whereas $\mathrm{nDTW}_{\mathrm{link}}$ measures how similar the local upper-limb geometry remains to the human motion after temporal misalignment is accounted for.

(1) Joint-space RMSE. Let $\mathbf{q}_{\mathrm{nom}}(t)$ denote the nominal retargeted joint command before safety filtering, and let $\mathbf{q}_{\mathrm{target}}(t)$ denote the safe joint target generated by the CBF-QP. The joint-space RMSE is defined as

$$\mathrm{RMSE}_q^0 = \sqrt{\frac{1}{TN_q}\sum_{t=1}^{T}\left\|\mathbf{q}_{\mathrm{target}}(t) - \mathbf{q}_{\mathrm{nom}}(t)\right\|_2^2}, \tag{23}$$

where $N_q = 8$ is the dimension of the upper-body joint space. This metric quantifies the direct command-level correction introduced by the safety layer at the same time instant. It intentionally does not compensate for temporal misalignment. A larger value therefore indicates that the CBF-QP had to deviate more strongly from the nominal imitation command to satisfy the safety constraints.

(2) Local-link direction nDTW. This metric follows the idea that imitation quality should not be evaluated only by global pose error or joint-space deviation. In human-to-robot motion imitation, local limb relationships are often more informative because the human and robot have different morphologies and limb lengths. Local-link-focused similarity measures have been used to compare human and humanoid motions at the body segment level [40]. In addition, dynamic time warping has been widely used to compare motion trajectories with nonuniform temporal alignment, including human–humanoid imitation evaluation [41,42]. We therefore combine these two ideas by applying DTW to trajectories of local upper-limb direction vectors. This produces a morphology-aware and timing-tolerant measure of geometric imitation quality.

For the human motion, four unit link-direction vectors are constructed at each frame: $\left[\hat{\mathbf{u}}_L^h(i),\quad \hat{\mathbf{f}}_L^h(i),\quad \hat{\mathbf{u}}_R^h(i),\quad \hat{\mathbf{f}}_R^h(i)\right]$, corresponding to the left upper arm, left forearm, right upper arm, and right forearm. For the robot motion, the corresponding unit link directions are obtained from forward kinematics: $\left[\hat{\mathbf{u}}_L^r(j),\quad \hat{\mathbf{f}}_L^r(j),\quad \hat{\mathbf{u}}_R^r(j),\quad \hat{\mathbf{f}}_R^r(j)\right]$. Using unit direction vectors makes the comparison focus on local limb orientation rather than absolute link length, which is important because the human and humanoid have different body proportions. The frame-to-frame local geometric cost between human frame i and robot frame j is computed as

$$c(i,j) = \sum_{\ell\in\{Lu,Lf,Ru,Rf\}} \arccos\Big(\mathrm{clip}\Big(\hat{\mathbf{v}}_\ell^h(i)^T\hat{\mathbf{v}}_\ell^r(j), -1, 1\Big)\Big), \tag{24}$$

where the clipping operation is used for numerical robustness. The cost $c(i,j)$ is the sum of angular discrepancies between corresponding local upper-limb directions. Thus, it is small when the human and robot upper arms and forearms point in similar directions in their respective local body frames. The DTW-link distance is $D_{\text{DTW-link}} = \min_\pi \sum_{(i,j)\in\pi} c(i,j)$,

where π is a valid monotonic DTW alignment path. The normalized local-link DTW metric is then defined as

$$\mathrm{nDTW}_{\mathrm{link}} = \frac{1}{|\pi^\star|} \sum_{(i,j)\in\pi^\star} c(i,j), \tag{25}$$

where $\pi^\star$ is the optimal DTW path. A smaller value indicates better time-aligned local geometric similarity.

Since both RMSE_q^0 and $\mathrm{nDTW}_{\mathrm{link}}$ are smaller-is-better imitation-error metrics, their normalized error scores are $\widetilde{\mathrm{RMSE}}_q^0 = \mathcal{N}\left(\mathrm{RMSE}_q^0\right)$ and $\widetilde{\mathrm{nDTW}}_{\mathrm{link}} = \mathcal{N}(\mathrm{nDTW}_{\mathrm{link}})$. The composite imitation score $S_{\mathrm{imit}} \in [0,1]$ is then a normalized, higher-is-better metric defined as

$$S_{\mathrm{imit}} = 1 - \left(\alpha_1 \widetilde{\mathrm{RMSE}}_q^0 + \alpha_2 \widetilde{\mathrm{nDTW}}_{\mathrm{link}}\right), \qquad \alpha_1 + \alpha_2 = 1, \quad \alpha_{1,2} \geq 0. \tag{26}$$

3. Results

The proposed framework is evaluated from two complementary perspectives: first, verifying the ability of the CBF-QP layer to prevent unsafe motions in representative self-collision and human–humanoid interaction scenarios; and second, quantifying how CBF parameters govern the safety–imitation trade-off via the metrics defined in Section 2.3. Throughout this evaluation, $\mathbf{q}_{\mathrm{nom}}$ denotes the nominal retargeted joint command before safety filtering, and $\mathbf{q}_{\mathrm{cbf}}$ represents the safe target produced by the CBF-QP layer. Both evaluation scenarios were initially recorded during live hardware executions of the complete pipeline, where the camera streamed human skeleton observations in real time to feed the online retargeting, CBF-QP filtering, and robot-command layers.

The resulting camera-captured skeleton sequences and their original timestamps were subsequently utilized as replay inputs for simulation experiments (including baseline comparisons) and extensive CBF parameter sweeps. This protocol preserves the exact human-motion profile observed during live experiments, thereby enabling strictly matched simulation analysis and repeatable parameter investigations under identical input conditions; as such, the replay mechanism is employed solely for controlled evaluation, while the hardware episodes themselves stem from genuine live online executions (see Video S1 and Video S2 in the Supplementary Materials). For technical reproducibility, the source code and experimental data are publicly available as specified in the Data Availability Statement, while supplementary computational benchmarks validating the capsule-based modeling choice are provided in Appendix B.

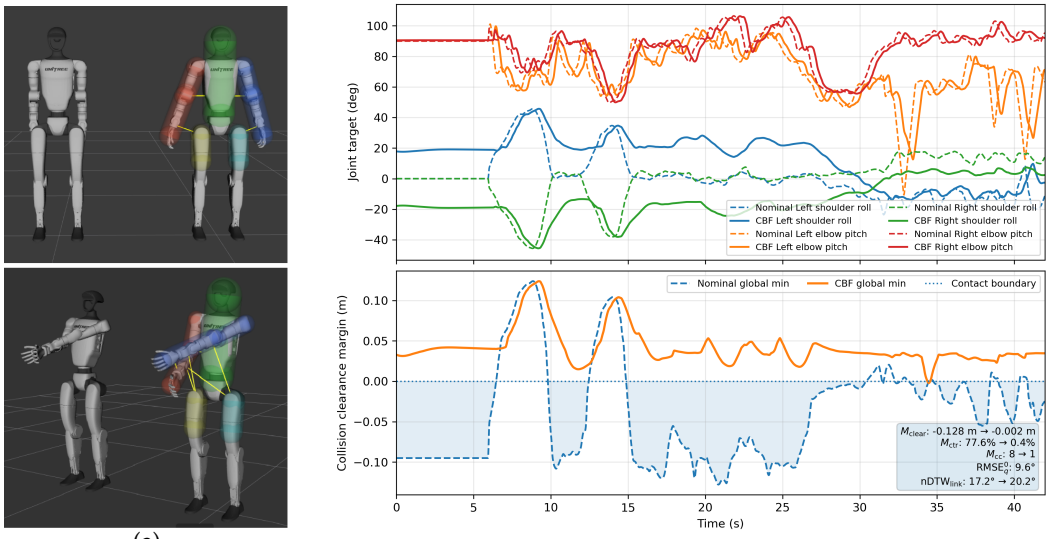
3.1. Safety in Motion Imitation

3.1.1. Simulation Studies

We first evaluated the proposed safety layer in simulation using the recorded skeleton streams from the two live hardware episodes described above. The snapshots in Figures 2 and 3 show representative moments within these episodes and the yellow segments indicate selected active capsule pairs through their closest-point connections.

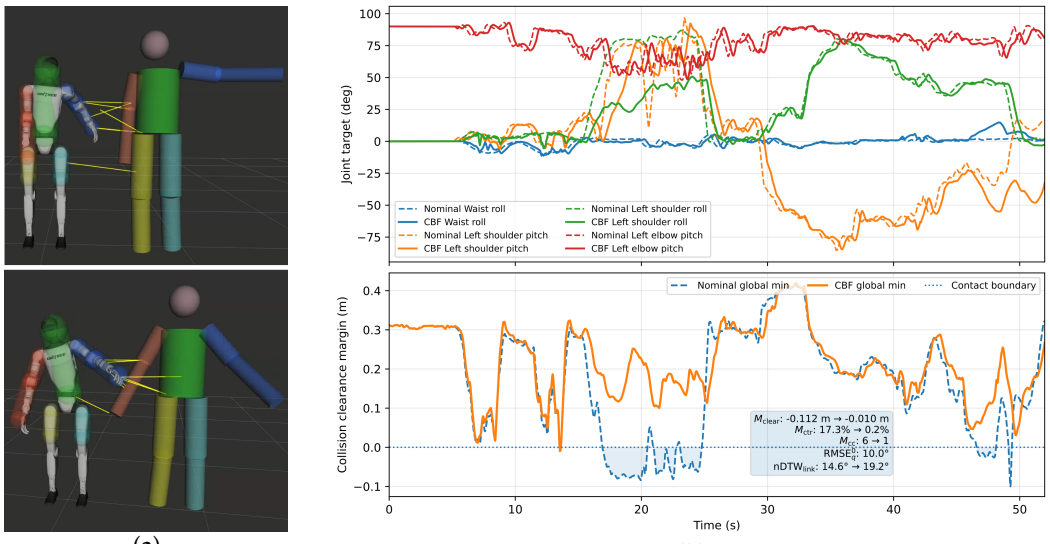
In the self-collision episode, the nominal retargeted command drives the humanoid arms across the torso and produces repeated violations of the physical collision boundary. As shown in Figure 2, the nominal global minimum clearance reaches $M_{\mathrm{clear}} = -0.128$ m, with a collision-time ratio of 77.6% and eight collision episodes. After CBF-QP filtering, the minimum clearance is improved to approximately -0.002 m, the collision-time ratio is reduced to 0.4%, and the collision count decreases to one. This indicates that the CBF-QP substantially suppresses persistent self-collision while preserving the qualitative structure of the retargeted arm motion. The corresponding imitation metrics show the expected cost

of safety filtering: the joint-space error is $\mathrm{RMSE}_q^0 = 9.6°$, and the time-aligned local-link error increases from 17.2° for the nominal motion to 20.2° for the CBF-filtered motion.



(**a**) (**b**)

Figure 2. Simulation self-collision avoidance with and without CBF-QP. (**a**) Representative moments from the self-collision episode, including a nominal cross-arm motion that produces unsafe robot self-contact. The ghost robot visualizes the nominal retargeted command, whereas the colored robot shows the CBF-filtered target. Yellow segments indicate selected monitored CBF constraint pairs. (**b**) Representative target joint trajectories and global minimum collision clearance margin. The CBF-QP strongly reduces the duration and number of self-collision episodes while preserving the main retargeted motion pattern.



(**a**) (**b**)

Figure 3. Simulation of human–humanoid collision avoidance with and without CBF-QP. (**a**) Representative moments from the human–humanoid interaction episode, including human-obstructed arm-raising and arm-crossing configurations. Yellow segments indicate selected monitored human–humanoid CBF constraint pairs. (**b**) Representative target joint trajectories and global minimum collision clearance margin. The CBF-QP reduces both the collision-time ratio and the number of collision episodes while retaining the main imitation behavior.

The human–humanoid interaction episode exhibits a different type of safety risk. Here, the nominal robot motion is kinematically plausible for imitation, but it intersects the nearby human capsule model during the arm-raising and -crossing portions of the episode.

As shown in Figure 3, the nominal command reaches $M_{\text{clear}} = -0.112$ m, with $M_{\text{ctr}} = 17.3\%$ and six collision episodes. With the CBF-QP enabled, the minimum clearance improves to -0.010 m, while the collision-time ratio decreases to 0.2% and the collision count decreases to one. Thus, the CBF-QP does not merely reduce the worst-case penetration; it also substantially shortens the duration over which any monitored pair remains in collision. The filtered trajectory introduces an RMSE_q^0 of 10.0°, and $\text{nDTW}_{\text{link}}$ changes from 14.6° to 19.2°, reflecting the safety-induced geometric deformation of the imitated arm motion.

These simulation results demonstrate two important properties of the proposed formulation. First, the same capsule-based CBF-QP structure can handle both robot self-collision and human–humanoid collision constraints. Second, the safety layer acts as a local deformation of the nominal imitation command rather than as a complete replacement of the retargeted motion. This motivates the hardware experiments and parameter analysis below, where the same command-level comparison between $\mathbf{q}_{\text{nom}}$ and $\mathbf{q}_{\text{cbf}}$ is used to quantify safety and imitation.

3.1.2. Hardware Experiments

Hardware experiments were conducted as live online executions of the complete perception–retargeting–CBF pipeline on the humanoid robot. During these runs, the camera streamed human skeleton observations in real time, and the proposed framework generated upper-body commands online. The robot lower body was controlled using the manufacturer-provided balance controller, while the proposed framework commanded the upper-body joints. The CBF parameters used in the hardware experiments were selected beforehand as a reasonable operating point based on prior tuning experience and preliminary observations, so that the experiments could provide a representative comparison between nominal imitation and CBF-filtered execution. The camera-captured skeleton sequences and command logs from these runs were recorded for the matched simulation and parameter-sweep analyses, as described at the beginning of this section.

Figure 4 shows the real-robot self-collision episode. The sequence includes standing, arm-lifting, arm-crossing, and recovery phases. Figure 5 reports the corresponding target joint trajectories, global minimum clearance, and runtime statistics. With $\phi_s^{rr} = 0.03$ m and $\gamma^{rr} = 2.0$, the CBF-QP successfully modifies the nominal self-collision-prone target so that the CBF-filtered target remains close to the physical contact boundary without producing the long collision intervals observed in the nominal command. Over the evaluation window, the CBF node runs at 12.9 Hz, the QP solve time is 1.11/4.56 ms for median/p95, and the QP success rate is 100%. These results indicate that the optimization layer itself remains reliable on hardware for the monitored self-collision constraints.

Figure 6 shows the hardware human–humanoid interaction episode. For qualitative illustration, the snapshots are taken from the complete safety configuration with both robot self-collision and human–humanoid collision constraints enabled. To isolate the effect of the human–humanoid CBF component, Figure 7 reports the corresponding hardware result with only the human–humanoid CBF constraints enabled. In this experiment, the human–humanoid CBF constraints are enabled to modify the upper-body target when the human body obstructs the nominal arm motion. The setting $\phi_s^{hr} = 0.15$ m and $\gamma^{hr} = 3.0$ yields stable operation with a CBF rate of 15.3 Hz, QP median/p95 solve time of 0.94/4.20 ms, and 100% QP success. Compared with the nominal command, the filtered target maintains greater separation from the human capsule model while preserving the main raising and crossing structure of the demonstrated motion. This confirms that the dynamic human CBF term in Equation (11) can be deployed in the real perception loop without increasing the QP decision dimension.

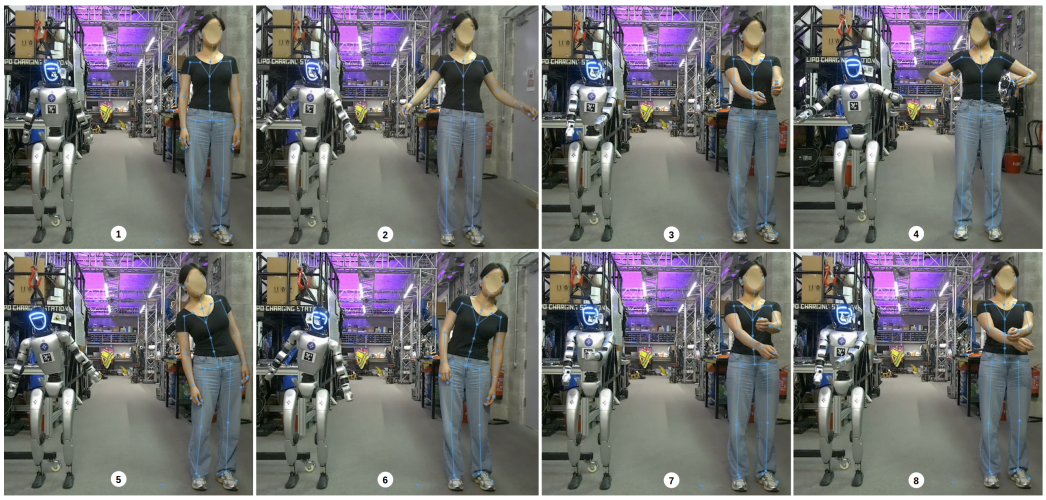


Figure 4. Hardware self-collision avoidance episode. The numbered snapshots show the temporal progression of the live online upper-body motion on the real robot: (1) neutral standing, with the robot maintaining a slightly abducted arm posture to avoid thigh contact; (2) bilateral arm abduction; (3, 4) arm motions; (5, 6) right and left torso-leaning motions, during which the robot tracks the demonstrated posture while maintaining arm–body separation; and (7, 8) arm-crossing motions, where the CBF-QP prevents robot self-collision between the upper limbs.

Settings	
ϕ_s^{rr}	0.03 m
γ^{rr}	2.0
d_{hr} (standstill)	1.2 m
Ctrl. rate	100 Hz
$k_{p-\text{arm}}$	50.0
$k_{d-\text{arm}}$	2.0
$k_{p-\text{waist}}$	100.0
$k_{d-\text{waist}}$	20.0
CBF runtime	
CBF rate	12.9 Hz
QP med/p95	1.11/4.56 ms
Tick med/p95	73.0/97.1 ms
Cons. med/max	5/5
QP success	100%

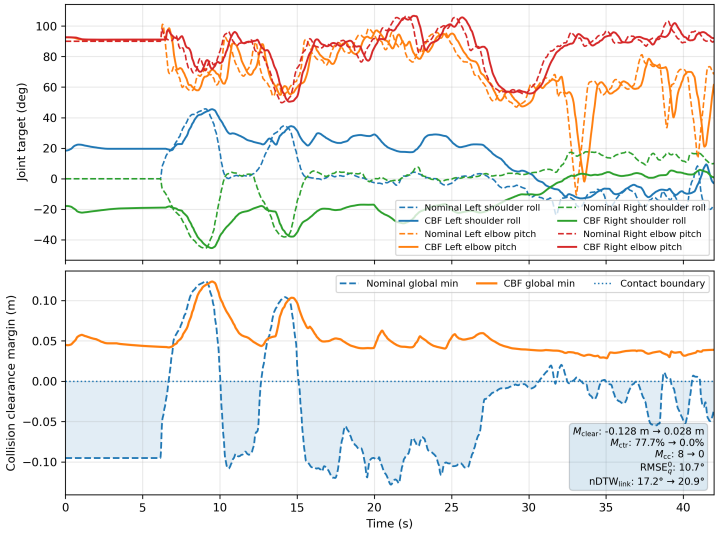


Figure 5. Hardware self-collision avoidance result. The left table summarizes controller settings and CBF runtime statistics over the evaluation window. The right plot shows target joint trajectories and the global minimum collision clearance margin for the nominal and CBF-filtered commands.

Finally, Figure 8 evaluates the complete safety configuration with both robot self-collision and human–humanoid collision constraints enabled. Because both constraint families are active, the number of monitored active constraints increases to a median/max of 10/13. While the full tick time increases to 117.13/160.50 ms for median/p95 (effective rate of $\approx$ 8.3 Hz), the core QP solve time remains low at 1.14/4.59 ms with a 100% success rate. This contrast indicates that the primary runtime cost stems from the perception, capsule generation, constraint construction, and callback pipeline rather than from the optimization itself. This integration latency introduces a delay-induced margin, where the p95 full tick time ($\tau = 0.1605$ s) would incur an additional buffer of 0.161 m at 1.0 m/s and 0.321 m at 2.0 m/s. Thus, while effective for the tested moderate-speed interactions, scaling to fast motions necessitates larger safety margins, higher-rate sensing, and predictive human-capsule construction.



Figure 6. Hardware human–humanoid interaction episode with both robot–robot and human–robot CBF constraints enabled. The numbered snapshots show the temporal progression of the live online shared-space interaction: (1) close standing configuration; (2–4) the human raises the right arm near the robot, and the robot laterally leans away while tracking the motion to maintain separation; (5) recovery to nominal following; (6, 7) the human raises the left arm while standing close to the robot, causing the robot to initially limit its arm elevation and then (8, 9) resume a larger arm raise as the human moves away; and (10–14) close human–robot arm interaction, where the robot follows the demonstrated motion while avoiding contact with the human.

Settings	
ϕ_s^{hr}	0.15 m
γ^{hr}	3.0
d_{hr} (standstill)	0.8 m
Ctrl. rate	100 Hz
$k_{p-\text{arm}}$	50.0
$k_{d-\text{arm}}$	2.0
$k_{p-\text{waist}}$	100.0
$k_{d-\text{waist}}$	20.0
CBF runtime	
CBF rate	15.3 Hz
QP med/p95	0.94/4.20 ms
Tick med/p95	61.14/98.81 ms
Cons. med/max	5/9
QP success	100%

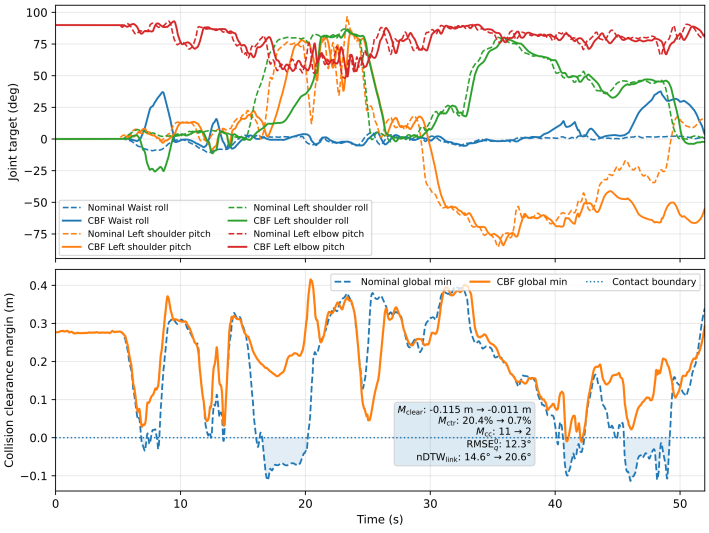


Figure 7. Hardware human–humanoid collision avoidance result with human–robot CBF constraints enabled. The left table summarizes the controller settings and CBF runtime statistics. The right plot shows the target joint trajectories and the global minimum human–humanoid clearance margin for the nominal and CBF-filtered commands.

Settings	
ϕ_s^{rr}	0.03 m
γ^{rr}	2.0
ϕ_s^{hr}	0.15 m
γ^{hr}	3.0
d_{hr} (standstill)	0.85 m
Ctrl. rate	100 Hz
$k_{p-\text{arm}}$	50.0
$k_{d-\text{arm}}$	2.0
$k_{p-\text{waist}}$	100.0
$k_{d-\text{waist}}$	20.0
CBF runtime	
CBF rate	8.3 Hz
QP med/p95	1.14/4.59 ms
Tick med/p95	117.13/160.50 ms
Cons. med/max	10/13
QP success	100%

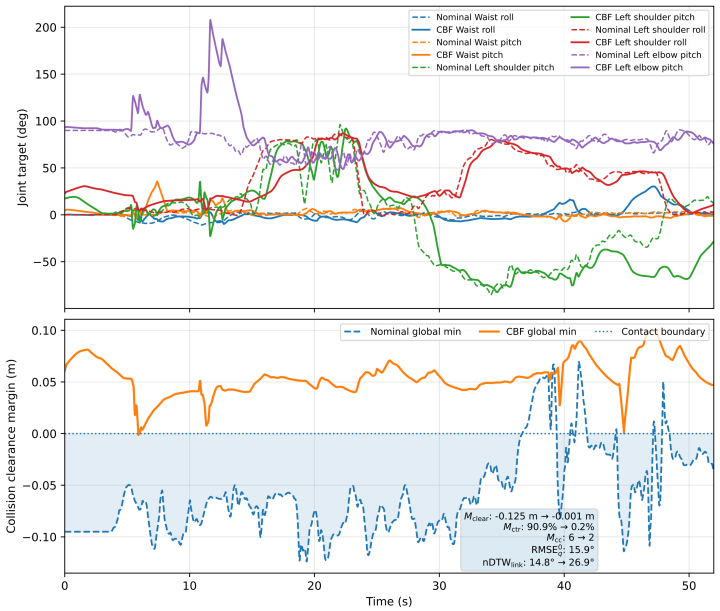


Figure 8. Hardware complete safety experiment with both robot–robot and human–humanoid CBF constraints enabled. The left table summarizes the controller settings and runtime statistics. The right plot shows target joint trajectories and the merged global minimum collision clearance margin over all monitored self-collision and human–humanoid pairs.

3.1.3. Quantitative Summary and Baseline Comparison

Table 3 summarizes the key quantitative metrics across all evaluated simulation and hardware scenarios, providing a centralized comparison between the nominal and CBF-filtered trajectories. As demonstrated, the CBF filter substantially reduces the collision-time ratio and collision count while increasing the minimum clearance toward the safe bounds. This safety enhancement incurs an expected decrease in imitation fidelity, reflected by the increased RMSE_q^0 and $\text{nDTW}_{\text{link}}$ values. Furthermore, the hardware logs show a median QP solve time below 1.15 ms and a 100% success rate. While the complete tick time currently limits deployment under high-speed interactions, these results validate the core efficacy and online feasibility of the safety filter within the tested moderate-speed regime.

To further contextualize the CBF-QP filter against representative non-CBF safety filters, we implement two classical reactive safety mechanisms for a benchmarking comparison: critical-pair *velocity damping* and *repulsive potential fields*. Crucially, both baselines are evaluated using the same recorded skeleton streams, capsule models, and evaluation metrics as the proposed CBF-QP method. The comparison is performed in replay-based simulation rather than on hardware, because these reactive baselines may generate non-smooth or insufficiently safe commands without extensive hardware tuning. The replay setting provides a controlled comparison under identical human motion, timing, and perception-noise conditions. Let $\mathbf{q}_b$ denote the baseline rollout state and $\boldsymbol{\nu}_{\text{ref}} = K(\mathbf{q}_{\text{nom}} - \mathbf{q}_b)$ the nominal tracking velocity. For capsule pair j, the surface clearance is $\chi_j = d_j - (r_{A_j} + r_{B_j})$, with gradient $\mathbf{g}_j = \partial\chi_j/\partial\mathbf{q}_b$. The velocity-damping baseline projects $\boldsymbol{\nu}_{\text{ref}}$ onto the half-space defined by the currently closest pair $j^\star = \arg\min_j \chi_j$, using a distance-dependent lower bound on the allowed approach velocity. The potential-field baseline instead adds a saturated sum of repulsive velocity corrections from all capsule pairs within the influence region. Both baselines use final joint-velocity clipping and the same safety threshold and influence distance for the corresponding capsule pairs.

Table 3. Summary results for nominal and CBF-filtered trajectories in simulation and hardware.

Scenario	Case	M_{clear} (m)	M_{ctr} (%)	M_{cc}	RMSE_q^0 (deg)	$\text{nDTW}_{\text{link}}$ (deg)	Runtime (ms) *	Active Const. †
Simulation Results (QP Success Rate: 100% for CBF-filtered runs)								
Self-collision	Nominal	−0.128	77.57	8	0.00	17.18	—	—
	CBF-QP	−0.002	0.43	1	9.61	20.24	0.74/38.22	5/5
Human–humanoid	Nominal	−0.112	17.35	6	0.00	14.59	—	—
	CBF-QP	−0.010	0.19	1	10.05	19.23	0.68/44.56	5/9
Hardware Results (QP Success Rate: 100% for CBF-filtered runs)								
Self-collision	Nominal	−0.128	77.67	8	0.00	17.22	—	—
	CBF-QP	0.028	0.00	0	10.72	20.94	1.11/73.02	5/5
Human–humanoid	Nominal	−0.115	20.42	11	0.00	14.63	—	—
	CBF-QP	−0.011	0.73	2	12.27	20.62	0.94/61.14	5/9
Both (merged)	Nominal	−0.125	90.92	6	0.00	14.79	—	—
	CBF-QP	−0.001	0.23	2	15.86	26.94	1.14/117.13	10/13

* Runtime reports the median QP solve time/median total CBF loop time in milliseconds, computed from the CBF diagnostics topic. Runtime is not applicable to nominal rows because no QP is solved. † Constraints report the median/maximum number of active CBF constraints during CBF-filtered execution.

As shown in Table 4, the reactive baselines improve some safety metrics but do not provide the same consistency as the CBF-QP filter under the tested multi-capsule constraints. The critical-pair velocity damper is effective in the human–humanoid episode, reducing M_{ctr} from 17.35% to 0.62%, but still leaves negative clearance and five collision episodes. In the self-collision episode, because it only reacts to the current closest pair, switching among competing capsule pairs leads to repeated short boundary crossings and increases the collision count to 59; this metric should therefore be interpreted together with collision-time ratio and minimum clearance. The potential-field baseline preserves imitation more closely than CBF-QP, but its safety improvement is limited, especially in the self-collision episode, where M_{ctr} remains 53.62%. In contrast, the CBF-QP achieves the strongest and most consistent geometric safety improvement among the tested filters, reducing M_{ctr} to 0.43% for self-collision and 0.19% for human–humanoid collision, with only one fleeting collision episode in each case. This improvement comes at the cost of larger imitation errors, so the comparison does not imply universal superiority over all possible safety filters or parameterizations. Nevertheless, under the same motion inputs, capsule model, and metrics, the results show that the CBF-QP formulation provides a clear advantage for safety-critical imitation: it coordinates multiple simultaneous geometric constraints within one optimization problem while yielding the feasible velocity closest to the nominal command.

Table 4. Simulation comparison with alternative non-CBF safety filters.

Scenario	Method	M_{clear} (m)	M_{ctr} (%)	M_{cc}	RMSE_q^0 (deg)	$\text{nDTW}_{\text{link}}$ (deg)
Self-collision	Nominal (joint-limit)	−0.128	77.57	8	0.00	17.18
	Velocity damping	−0.129	64.62	59	1.31	17.35
	Potential field	−0.117	53.62	7	2.11	17.49
	CBF-QP	−0.002	0.43	1	9.61	20.24
Human–humanoid	Nominal (joint-limit)	−0.112	17.35	6	0.00	14.59
	Velocity damping	−0.052	0.62	5	3.98	16.58
	Potential field	−0.067	11.23	6	1.57	15.20
	CBF-QP	−0.010	0.19	1	10.05	19.23

3.2. Safety–Imitation Trade-Off Analysis

The preceding experiments show that the CBF-QP layer can suppress unsafe nominal imitation commands, but this safety improvement is achieved by modifying $\mathbf{q}_{\text{nom}}$, so different CBF parameters may lead to different levels of imitation degradation. To characterize this safety–imitation trade-off, we performed hardware replay sweeps using the human–humanoid interaction episode with both robot self-collision and human–humanoid CBF constraints enabled. Each parameter setting was repeated three times, and each heatmap cell reports the mean over the repeated runs. Two sweeps are considered: the margin sweep varies ϕ_s^{rr} and ϕ_s^{hr} while fixing $\gamma^{rr} = 2.0$ and $\gamma^{hr} = 3.0$, whereas the gain sweep varies γ^{rr} and γ^{hr} while fixing $\phi_s^{rr} = 0.03$ m and $\phi_s^{hr} = 0.15$ m.

3.2.1. Effects of Safety-Distance Margins

The safety-distance margins directly shift the CBF safety boundary in Equation (5). Larger margins cause the safety filter to intervene earlier with respect to the physical collision boundary. Figure 9 reports the parameter-sweep results for different pairs of safety-distance margins $(\phi_s^{rr}, \phi_s^{hr})$. In each heatmap, the color gradient encodes the corresponding metric value, and the number in each cell gives the averaged value over repeated trials. The figure shows that the human–humanoid margin ϕ_s^{hr} is the dominant factor in this interaction episode. When $\phi_s^{hr} \in [0.09, 0.12]$ m, the minimum clearance remains negative across all tested ϕ_s^{rr} values, and the collision-time ratio remains approximately 0.8%–1.1%. Increasing ϕ_s^{hr} to 0.15 m produces a clear transition: the minimum clearance becomes positive for all tested ϕ_s^{rr}, and the collision-time ratio is reduced to zero in the repeated trials. Further increasing ϕ_s^{hr} to ≥ 0.17 m does not produce an equivalent improvement in all safety metrics because the merged metric is already near the collision-free regime. Instead, the main effect becomes an increase in imitation error. Both RMSE_q^0 and $\text{nDTW}_{\text{link}}$ increase with larger ϕ_s^{hr}, indicating stronger deformation of the nominal command. For example, the local-link nDTW error increases from approximately 19°–20° at the smallest human–humanoid margins to approximately 24°–27° at the largest margins. The joint-space RMSE follows the same trend, increasing from approximately 10°–12° to approximately 14°–16°. This pattern confirms that larger safety margins improve collision robustness primarily by sacrificing imitation fidelity.

The effect of ϕ_s^{rr} is weaker than that of ϕ_s^{hr} in this episode. This is expected because the replayed motion is a human–humanoid interaction episode in which the dominant unsafe behavior is induced by human proximity. Nevertheless, ϕ_s^{rr} remains relevant in the complete system: at very small robot–robot margins, short residual collision episodes can still occur even when the human–humanoid margin is large enough. This illustrates why the complete safety metric must consider both self-collision and human–humanoid collision pairs jointly. The hardware operating point $(\phi_s^{rr}, \phi_s^{hr}) = (0.03, 0.15)$ m, marked by the diamond, lies immediately after the observed safety transition, achieving a zero collision-time ratio without the larger imitation degradation seen at $\phi_s^{hr} \geq 0.17$ m.

3.2.2. Effects of CBF Gains

The CBF gains γ^{rr} and γ^{hr} regulate the barrier-value dynamics in Equation (5). Figure 10 reports the parameter-sweep results for different pairs of CBF gains $(\gamma^{rr}, \gamma^{hr})$ with the safety-distance margins fixed at the selected values. In each heatmap, the color gradient encodes the corresponding metric value, and the number in each cell gives the averaged value over repeated trials. This figure shows substantially weaker variation in safety than the margin sweep. Most parameter settings achieve a zero collision-time ratio, and the only remaining nonzero value is below 0.1%. This indicates that $\phi_s^{rr} = 0.03$ m and

$\phi_s^{hr} = 0.15$ m already define a sufficiently safe operating region for the evaluated episode, so the gains mainly affect how the safe target recovers toward the nominal command.

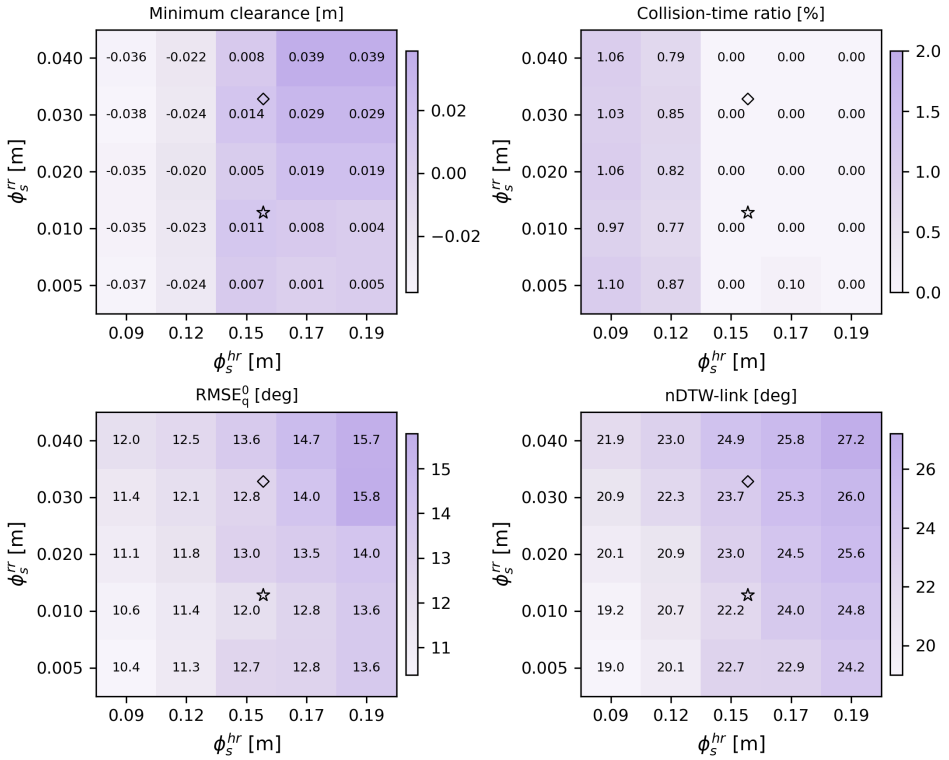


Figure 9. Effect of the safety-distance margins ϕ_s^{rr} and ϕ_s^{hr} on safety and imitation metrics in the complete hardware configuration. Both robot–robot and human–robot CBF constraints are enabled. Each cell reports the mean over three repeated replay trials with $\gamma^{rr} = 2.0$ and $\gamma^{hr} = 3.0$. Across all configurations, the trial variability remains bounded, with maximum standard deviations of 0.01 m for clearance, 0.18% for collision ratio, 1.71° for RMSE, and 0.94° for nDTW-link. The color gradient represents the value of the corresponding metric, and the numerical value is shown in each cell. The diamond marks the operating point used in the hardware experiments, whereas the star marks the preferred margin setting identified from the sweep results.

The imitation metrics reveal a clearer dependence on γ^{rr}. Very small robot–robot gains, especially $\gamma^{rr} = 0.5$, produce the largest joint-space and local-link errors. As γ^{rr} increases to 2.0, 3.0, and 3.5, both RMSE_q^0 and $\mathrm{nDTW}_{\mathrm{link}}$ decrease. This behavior is consistent with the velocity-level formulation in Equations (14)–(16): a small gain enforces slower recovery of the barrier dynamics and can keep $\mathbf{q}_{\mathrm{cbf}}$ farther from $\mathbf{q}_{\mathrm{nom}}$ for longer periods, whereas moderate gains allow the safe target to follow the nominal command more closely when the motion remains safe. The influence of γ^{hr} is comparatively weaker in this fixed-margin setting, suggesting that the human–humanoid safety boundary is already sufficiently conservative and that the gain mainly fine-tunes the recovery behavior.

Together, the margin and gain sweeps show that ϕ_s^{hr} determines the primary transition from unsafe to safe behavior in the evaluated human–humanoid interaction episode, whereas γ^{rr} and γ^{hr} mainly tune the deformation and recovery characteristics once the safety margins are adequate. This separation of roles is useful for parameter selection: margins should first be chosen to eliminate collision episodes, and gains can then be adjusted to improve imitation fidelity without reintroducing significant collision time.

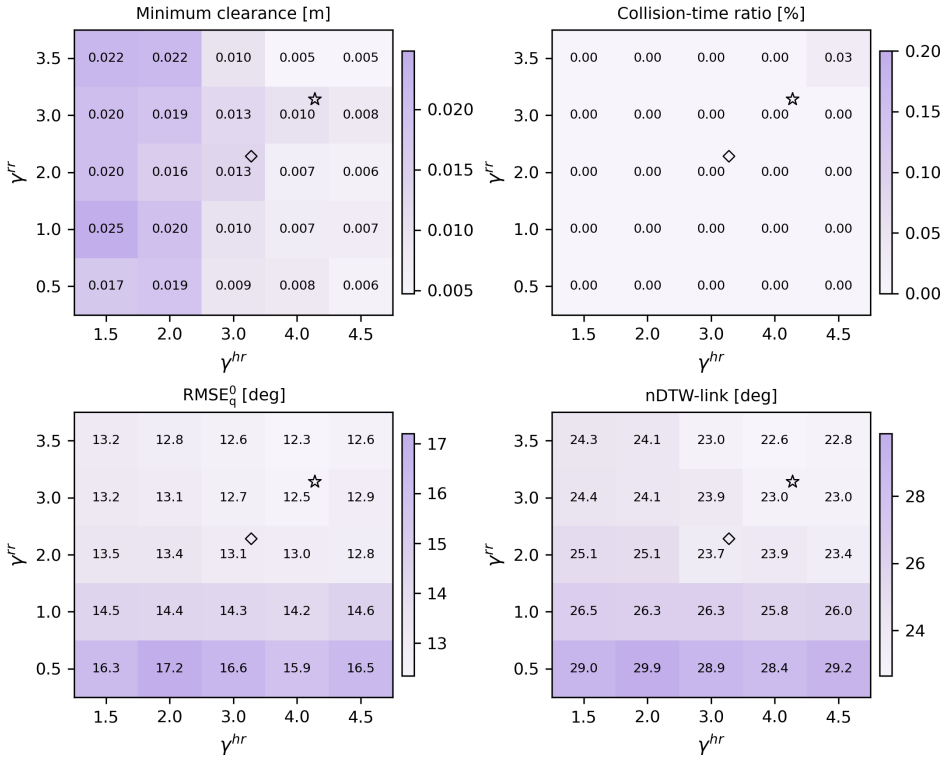


Figure 10. Effect of the CBF gains γ^{rr} and γ^{hr} on safety and imitation metrics in the complete hardware configuration. Both robot–robot and human–robot CBF constraints are enabled. Each cell reports the mean over three repeated replay trials with $\phi_s^{rr} = 0.03$ m and $\phi_s^{hr} = 0.15$ m. Across all configurations, the trial variability remains bounded, with maximum standard deviations of 0.01 m for clearance, 0.04% for collision ratio, 1.41° for RMSE, and 1.04° for nDTW-link. The color gradient represents the value of the corresponding metric, and the numerical value is shown in each cell. The diamond marks the operating point used in the hardware experiments, whereas the star marks the preferred gain setting identified from the sweep results.

3.2.3. Safety–Imitation Trade-Off

To summarize the overall effect of the four CBF parameters, Figure 11 plots the composite safety score S_{safe} against the composite imitation score S_{imit}. Each point represents one complete parameter setting $(\phi_s^{rr}, \phi_s^{hr}, \gamma^{rr}, \gamma^{hr})$, and the plotted location corresponds to the mean over repeated replay trials. The color encodes the human–humanoid safety-distance margin ϕ_s^{hr}, while the marker size encodes the robot–robot CBF gain γ^{rr}. The empirical Pareto front is obtained from the non-dominated settings in the $(S_{\text{imit}}, S_{\text{safe}})$ plane, where larger values are preferred for both scores. Specifically, these composite scores are computed using weights $(0.45, 0.45, 0.10)$ for S_{safe} and $(0.5, 0.5)$ for S_{imit}, reflecting the physical significance of each underlying metric. While these scores provide a relative ranking within this specific sweep campaign, a sensitivity check across six alternative weighting perturbations confirms that the top parameter selections remain unchanged, ensuring the robustness of the observed Pareto trends.

The resulting distribution reveals a clear trade-off structure. Settings with small human–humanoid margins tend to preserve imitation better but yield lower safety scores, consistent with the residual collision-time ratios observed in the margin sweep. Conversely, larger ϕ_s^{hr} values move the operating point toward higher safety scores, but excessive margins also reduce imitation fidelity because the CBF-QP must deform the nominal target more strongly. This trend is consistent with Figure 9, where increasing ϕ_s^{hr} beyond the safety transition improves or maintains clearance but increases both RMSE$_q^0$ and nDTW$_{\text{link}}$. The marker size

variation further shows that the gain parameters mainly refine the recovery behavior within the feasible region, rather than producing the dominant unsafe-to-safe transition.

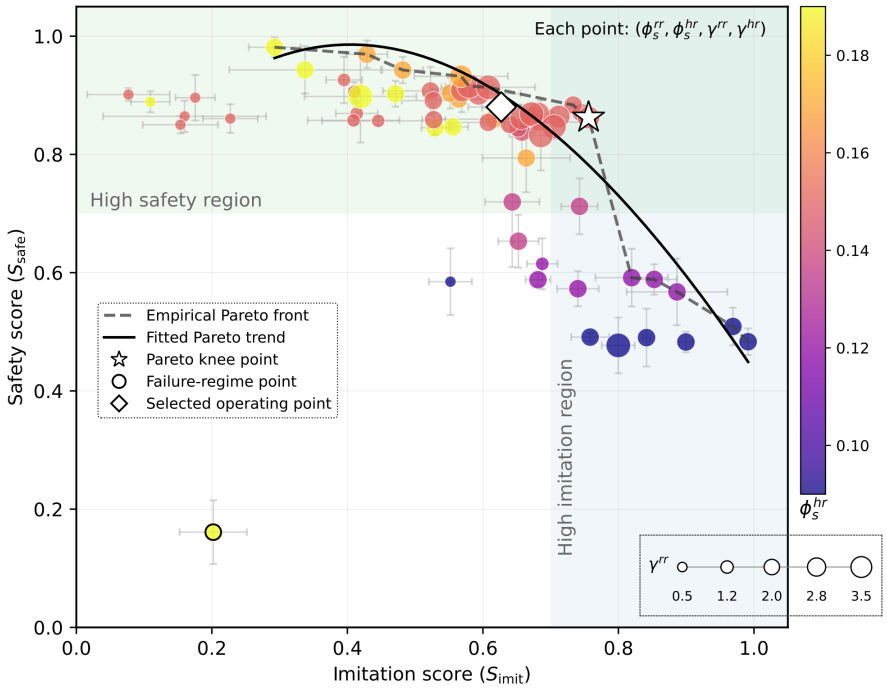


Figure 11. Composite safety–imitation trade-off under CBF parameter variation. Each point represents one complete parameter setting $(\phi_s^{rr}, \phi_s^{hr}, \gamma^{rr}, \gamma^{hr})$ and shows the mean over three repeated replay trials. Across all samples, the trial variability remains bounded, with maximum standard deviations of 0.102 for S_{imit} and 0.111 for S_{safe}. The color encodes the human–humanoid safety-distance margin ϕ_s^{hr}, and the marker size encodes the robot–robot CBF gain γ^{rr}. The dashed curve connects the empirical Pareto-efficient settings, and the solid curve provides a visual trend of the Pareto front. The diamond denotes the selected operating point $(0.03, 0.15, 2.0, 3.0)$ used in the hardware experiments, the star denotes the Pareto knee point $(0.01, 0.15, 3.0, 4.0)$ closest to the utopia point $(S_{\text{imit}}, S_{\text{safe}}) = (1, 1)$, and the circled marker denotes an intentionally selected failure-regime point $(0.03, 0.19, 2.0, 4.5)$. Composite scores are computed using $w = (0.45, 0.45, 0.10)$ for M_{clear}, M_{ctr}, and M_{cc}, and $\alpha = (0.5, 0.5)$ for RMSE_q^0 and $\text{nDTW}_{\text{link}}$.

The diamond marks the selected operating point $(0.03, 0.15, 2.0, 3.0)$ used in the main hardware experiments. It is not claimed to be globally optimal; rather, it lies in a high-safety and high-imitation region and provides a conservative operating choice for real-robot execution. The star marks the Pareto knee point $(0.01, 0.15, 3.0, 4.0)$, defined as the Pareto-efficient setting with the smallest Euclidean distance to the utopia point $(S_{\text{imit}}, S_{\text{safe}}) = (1, 1)$, and therefore represents the strongest empirical compromise within the sampled parameter set. The circled marker denotes an intentionally selected failure-regime point $(0.03, 0.19, 2.0, 4.5)$, illustrating that poorly balanced or extreme parameter combinations can degrade both objectives simultaneously, producing neither reliable safety nor faithful imitation. It helps identify unsuitable regions of the parameter space but does not affect the empirical Pareto front. Overall, the Pareto analysis provides a descriptive summary consistent with the heatmap trends: the safety margins should first be chosen to mitigate collision episodes, especially through ϕ_s^{hr}, and the CBF gains can then be adjusted to improve imitation fidelity without incurring substantial collision time.

To make the Pareto analysis directly actionable, Table 5 reports three representative operating regimes selected from the sampled parameter space: an imitation-prioritized setting, a balanced setting near the Pareto knee, and a safety-prioritized setting. The imitation-prioritized regime preserves the nominal motion most closely, as reflected by the lowest RMSE_q^0 and $\text{nDTW}_{\text{link}}$, but retains a negative minimum clearance and a nonzero

collision-time ratio. In contrast, the safety-prioritized regime achieves the largest minimum clearance and eliminates collision episodes, but at the cost of substantially increased imitation error. The balanced/Pareto-knee regime removes collision episodes while maintaining moderate imitation error, providing the strongest compromise among the three reported operating levels. Taken together, Table 5 shows that the composite scores are consistent with the raw geometric safety metrics: increasing safety improves clearance and removes collision episodes, while progressively increasing joint-space and local-link imitation errors. This supports the balanced/Pareto-knee setting as a practical operating choice when both collision avoidance and motion fidelity are required.

Table 5. Representative operating regimes selected from the CBF parameter sweep. Safety metrics are evaluated over combined robot–robot and human–robot pairs.

Regime	ϕ_s^{rr}	ϕ_s^{hr}	γ^{rr}	γ^{hr}	M_{clear} [m]	M_{ctr} [%]	M_{cc}	RMSE_q^0 [deg]	$\text{nDTW}_{\text{link}}$ [deg]	S_{safe}	S_{imit}
Imitation-prioritized	0.005	0.090	2.0	3.0	−0.0367	1.10	1.00	10.39	19.00	0.483	0.991
Balanced/Pareto knee	0.010	0.150	3.0	4.0	0.0110	0.00	0.00	11.92	22.10	0.862	0.756
Safety-prioritized	0.040	0.190	2.0	3.0	0.0389	0.00	0.00	15.66	27.21	0.981	0.293

4. Discussion

The results show that, in the evaluated episodes, safety-aware human-to-humanoid imitation can be achieved without replacing the nominal retargeting pipeline by a monolithic controller. Instead, the proposed framework separates perception, retargeting, and safety filtering into interpretable modules. This modularity is important because it allows the nominal command q_{nom} and the CBF-filtered target q_{cbf} to be compared directly, making the effect of safety enforcement measurable rather than only visually observable. In both simulation and hardware experiments, the CBF-QP layer substantially reduces persistent unsafe behaviors during self-collision and human–humanoid proximity episodes. In the simulation self-collision scenario, the collision-time ratio decreases from 77.6% to 0.4%, while in the human–humanoid interaction case, it decreases from 17.3% to 0.2%. Hardware experiments further demonstrate that the same formulation is computationally feasible to operate within an online perception–control loop, achieving a 100% QP success rate across all evaluated intervals. Finally, the simulation comparison with velocity damping and potential-field baselines reveals that the CBF-QP filter provides more consistent multi-constraint safety improvements. The two motion episodes used in the experiments were recorded from complete online hardware runs and reused as replay inputs for simulation reproduction and CBF parameter sweeps. This replay procedure serves as a standardization protocol for fair comparison, rather than a replacement for online execution. By using identical recorded input motions, the effects of different CBF parameters can be isolated without introducing trial-to-trial variations in human demonstration timing or posture. Within these representative scenarios, these findings suggest that CBF-QP filtering is a practical mechanism for adding explicit collision-avoidance behavior to vision-aided humanoid imitation while preserving the qualitative structure of the demonstrated motion. It is worth noting that although this study evaluates the method on an 8-DoF upper-body humanoid, the method is compatible with higher-DoF platforms. Extending it to such platforms would mainly require richer perception, retargeting of the additional joint motions, and a higher-dimensional CBF-QP with the corresponding safety constraints.

The safety–imitation analysis also clarifies the role of the CBF parameters. The safety-distance margins ϕ_s^{rr} and ϕ_s^{hr} determine when the safety layer begins to intervene relative to the physical collision boundary, whereas the gains γ^{rr} and γ^{hr} mainly shape the recovery dynamics once the safety boundary is active. In the hardware parameter sweeps, the human–humanoid margin ϕ_s^{hr} was the dominant factor for the evaluated interaction

episode: increasing it to approximately 0.15 m produced the transition to a zero collision-time ratio across repeated trials, while larger values mainly increased imitation error. This behavior is consistent with the composite Pareto analysis and the representative operating regimes in Table 5. The imitation-prioritized setting preserved the nominal motion most closely but retained residual collision, whereas the safety-prioritized setting maximized clearance at the cost of a larger RMSE_q^0 and $\mathrm{nDTW}_{\mathrm{link}}$. The balanced/Pareto-knee setting removed collision episodes while maintaining moderate imitation error, indicating that the proposed metrics can support practical parameter selection rather than only post hoc performance reporting.

Several **limitations** remain. First, the current framework focuses on 8-DoF upper-body imitation. The robot lower body is handled by the manufacturer-provided balance controller, while the upper-body command generation uses only a simple waist-balance compensation strategy. Therefore, the system does not yet address full-body locomotion, coordinated whole-body balance, high-speed dynamic motions, or contact-rich task execution. These extensions would require tighter integration between upper-body imitation, lower-limb coordination, and balance recovery. Second, the current quantitative evaluation is limited to two representative skeleton-motion episodes recorded from a single human subject. Although these sequences were specifically selected as representative stress tests to cover core safety-critical scenarios, they do not constitute an exhaustive validation over different users, motion classes, approach distances, motion speeds, or environmental conditions. Extending the evaluation to these broader operational scenarios remains an important direction, primarily requiring richer multi-object perception, careful calibration, and scalable real-time optimization. Third, the present quantitative evaluation primarily isolates the behavior of the CBF-QP geometric safety-filtering layer by comparing q_{nom} and q_{cbf}. This choice is appropriate for analyzing the geometric safety–imitation trade-off introduced by the CBF parameters, but it does not fully characterize downstream physical safety factors such as actuator tracking error, low-level controller dynamics, communication delay, torque or force limits, contact dynamics, or hardware compliance. In particular, a target trajectory that is safe in the capsule-distance model may still require additional execution-level verification when the robot operates at higher speed, under larger payload variation, or in more dynamic human–humanoid interaction. A further limitation is that the present imitation space does not include detailed hand and finger motion. This simplification is appropriate for evaluating upper-body safety filtering, but hands are important in human–humanoid interaction, especially for manipulation, handover, and gestural communication. Extending the framework to hand motion would require more robust fine-scale perception and retargeting, as well as finer collision models beyond the current capsule-level upper-body representation.

Future work will extend the framework in three directions. The first direction is to move from upper-body imitation to learning-enhanced whole-body safe imitation in order to better support high-speed and dynamically coupled motions. Ongoing work comprises upgrading the system to a multi-camera perception setup processed on a high-performance computing node to reduce skeleton-estimation latency and improve tracking continuity. For faster interaction, the safety layer should also use predictive human capsules or Kalman/model-based predictors, so that CBF constraints are evaluated against anticipated human positions rather than only the current measured skeleton. On the control side, RL policies are being developed for lower-body balance control and learning-based upper-body tracking with improved gravity compensation and dynamic response. These components are expected to help coordinate the torso, waist, and lower body with arm imitation, thereby integrating safety filtering with whole-body stability. The second direction is to close the gap between target-level geometric safety filtering and execution-level physical behavior. This includes incorporating actuator tracking models, latency compensation,

measured joint feedback, and possibly force or contact sensing into the evaluation and control loop. These additions would allow the safety layer to account for physical hardware behavior more explicitly, rather than relying only on geometric clearance at the command level. The third direction is to extend the imitation representation to include hand motion and task-level interaction primitives. Learning-based perception and retargeting methods may be particularly useful for this extension, since hand motion involves fine-scale articulation, frequent occlusion, and strong task dependence. In such settings, the proposed safety–imitation metrics could be used not only for offline analysis but also for adaptive parameter tuning, where ϕ_s^{rr}, ϕ_s^{hr}, γ^{rr}, and γ^{hr} are adjusted online according to the current task, human proximity, and desired imitation fidelity.

5. Conclusions

This paper presented a vision-aided safe human-to-humanoid motion imitation framework that couples skeleton-based retargeting with a capsule-based CBF-QP safety filter. By inserting an interpretable optimization layer between nominal imitation and execution, the proposed formulation enforces robot self-collision and human–humanoid proximity constraints while preserving the qualitative structure of the demonstrated motion. Within the evaluated episodes, simulation and hardware results, along with the baseline comparison with alternative safety filters, demonstrate that the framework effectively reduces geometric collision violations and validates the feasibility of online operation within the perception–control loop. Furthermore, the parameter-sweep analysis formalizes the underlying safety–fidelity trade-off, making it quantitatively tunable via safety-distance margins and CBF gains to enable systematic selection of desired operating regimes.

While demonstrating the viability of modular and interpretable safety filtering, the current implementation operates within certain practical boundaries. The control framework is currently focused on upper-body kinematics and geometric proximity constraints rather than full physical safety certification, and its real-time performance remains bounded by end-to-end processing latency and the quality of vision-based skeleton tracking. Moreover, while the live hardware interactions support real-world feasibility, the evaluation remains a representative proof-of-concept restricted to a limited number of users and motion sequences, rather than an exhaustive validation across diverse statures, speeds, and environmental conditions. Rather than diminishing the core methodology, these limitations outline direct pathways for future research. This perspective provides a scalable foundation for extending the framework toward learning-enhanced whole-body dynamic balancing, multi-object perception under sensor noise, and high-rate, hand-aware human–humanoid interaction integrated with execution-level safety verification.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/bdcc10080246/s1, Video S1: online imitation without CBF; Video S2: online imitation with CBF (self-collision and human-humanoid collision).

Author Contributions: Conceptualization, W.C. and A.T.; methodology, W.C., J.A., N.E. and A.T.; software, W.C. and J.A.; validation, W.C. and J.A.; formal analysis, W.C., J.A., N.E. and A.T.; investigation, W.C. and J.A.; resources, A.T.; data curation, W.C. and J.A.; writing—original draft preparation, W.C.; writing—review and editing, W.C., J.A., N.E. and A.T.; visualization, W.C.; supervision, A.T.; project administration, A.T.; funding acquisition, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the NYUAD Center for AI and Robotics, funded by Tamkeen under the NYUAD Research Institute Award CG010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The person shown in Figures 4 and 6 is one of the authors of this paper. This individual is not involved as a human subject in the study. The figures are included for illustrative purposes to demonstrate the practical use of the proposed system and its interaction scenario. She was fully aware of the purpose of the images and has given consent for their publication in this academic paper.

Data Availability Statement: The source code developed for the proposed vision-aided safe human-to-humanoid motion imitation framework is publicly available at https://github.com/RISC-NYUAD/Safe_Human_Humanoid_Motion_Imitation (accessed on 21 July 2026). The experimental data supporting the simulation, hardware, and parameter-sweep results reported in this study are archived on Zenodo: https://zenodo.org/records/20322803 (accessed on 21 July 2026); DOI: 10.5281/zenodo.20322802.

Acknowledgments: The authors acknowledge the Kinesis Core Technology Platforms (Kinesis CTP) at New York University Abu Dhabi (NYUAD) for providing access to the experimental robotic platform used in this study, including the Unitree G1 humanoid robot. During the preparation of this manuscript, the authors partially used ChatGPT-5.5 solely for minor language refinement and grammar checking. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DoF	Degree of freedom
CBF	Control barrier function
QP	Quadratic program
DTW	Dynamic time warping
FK	Forward kinematics
HRC	Human–robot collaboration
HRI	Human–robot interaction
MPC	Model predictive control
ISO	International Organization for Standardization
RL	Reinforcement learning

Appendix A. Supplementary Details on Human Pose Angle Estimation

This appendix provides supplementary details on the coordinate frames, skeleton keypoints, and implementation steps used to compute the reduced human upper-body angles from the ZED-camera [43] skeleton observations. The notation ${}^{Z}\mathbf{p}_{A}$ denotes the position of keypoint A expressed in the ZED global frame $\mathcal{F}_Z$, and a leading superscript indicates the coordinate frame in which a vector is represented.

Appendix A.1. Coordinate Frames and Keypoints

The ZED global frame $\mathcal{F}_Z$ is configured as a right-handed frame with $+x_Z$ forward, $+y_Z$ left, and $+z_Z$ upward. A human torso frame $\mathcal{F}_T$ is constructed online from the pelvis and shoulder keypoints and is used to express arm motion relative to the trunk. The seven skeleton keypoints are $\mathcal{K} = \{P, LS, RS, LE, RE, LW, RW\}$, corresponding to the

pelvis, left/right shoulders, left/right elbows, and left/right wrists. The shoulder center and limb vectors are defined as

$$
\begin{aligned}
{}^{Z}\mathbf{p}_{SC} &= \frac{1}{2}\left({}^{Z}\mathbf{p}_{LS} + {}^{Z}\mathbf{p}_{RS}\right), \\
{}^{Z}\mathbf{u}_{L} &= {}^{Z}\mathbf{p}_{LE} - {}^{Z}\mathbf{p}_{LS}, \qquad {}^{Z}\mathbf{u}_{R} = {}^{Z}\mathbf{p}_{RE} - {}^{Z}\mathbf{p}_{RS}, \\
{}^{Z}\mathbf{f}_{L} &= {}^{Z}\mathbf{p}_{LW} - {}^{Z}\mathbf{p}_{LE}, \qquad {}^{Z}\mathbf{f}_{R} = {}^{Z}\mathbf{p}_{RW} - {}^{Z}\mathbf{p}_{RE}.
\end{aligned}
$$

Appendix A.2. Torso Frame Construction

The torso frame $\mathcal{F}_T$ is constructed from the pelvis and shoulder keypoints. Its origin is placed at the pelvis, and its vertical axis follows the pelvis–shoulder-center direction:

$$
{}^{Z}\mathbf{p}_{O_T} = {}^{Z}\mathbf{p}_{P}, \qquad {}^{Z}\mathbf{e}_{T,z} = \frac{{}^{Z}\mathbf{p}_{SC} - {}^{Z}\mathbf{p}_{P}}{\|{}^{Z}\mathbf{p}_{SC} - {}^{Z}\mathbf{p}_{P}\|}.
$$

The lateral axis is computed from the shoulder line $\mathbf{s} = {}^{Z}\mathbf{p}_{RS} - {}^{Z}\mathbf{p}_{LS}$ after removing its component along the vertical axis:

$$
{}^{Z}\tilde{\mathbf{e}}_{T,x} = \mathbf{s} - (\mathbf{s}^{T}{}^{Z}\mathbf{e}_{T,z}){}^{Z}\mathbf{e}_{T,z}, \qquad {}^{Z}\mathbf{e}_{T,x} = \frac{{}^{Z}\tilde{\mathbf{e}}_{T,x}}{\|{}^{Z}\tilde{\mathbf{e}}_{T,x}\|}.
$$

The third axis completes the right-handed frame, ${}^{Z}\mathbf{e}_{T,y} = {}^{Z}\mathbf{e}_{T,z} \times {}^{Z}\mathbf{e}_{T,x}$, yielding

$$
{}^{Z}\mathbf{R}_{T} = [{}^{Z}\mathbf{e}_{T,x}\ {}^{Z}\mathbf{e}_{T,y}\ {}^{Z}\mathbf{e}_{T,z}], \qquad {}^{T}\mathbf{R}_{Z} = ({}^{Z}\mathbf{R}_{T})^{T}.
$$

Thus, any vector ${}^{Z}\mathbf{v}$ is expressed in the torso frame by ${}^{T}\mathbf{v} = {}^{T}\mathbf{R}_{Z}{}^{Z}\mathbf{v}$.

Appendix A.3. Angle Computation Details

The reduced human upper-body angle vector $\boldsymbol{\theta} \in \mathbb{R}^{8}$ and its analytical definitions are provided in Section 2.1.1. This subsection records the additional vector transformations and numerical treatments used for the angle computation. The upper-arm vectors used for shoulder-angle estimation are first expressed in the torso frame as ${}^{T}\mathbf{u}_{L} = {}^{T}\mathbf{R}_{Z}{}^{Z}\mathbf{u}_{L}$ and ${}^{T}\mathbf{u}_{R} = {}^{T}\mathbf{R}_{Z}{}^{Z}\mathbf{u}_{R}$. The components of these vectors are denoted by

$$
{}^{T}\mathbf{u}_{L} = \begin{bmatrix} u_{L,x}^{T} & u_{L,y}^{T} & u_{L,z}^{T} \end{bmatrix}^{T}, \qquad {}^{T}\mathbf{u}_{R} = \begin{bmatrix} u_{R,x}^{T} & u_{R,y}^{T} & u_{R,z}^{T} \end{bmatrix}^{T}.
$$

These torso-frame components are used in the shoulder-pitch and shoulder-roll definitions in Section 2.1.1. The elbow-pitch angles are computed from the inter-segment angles between the corresponding upper-arm and forearm vectors. For numerical robustness, all direction vectors are normalized before angular computations, and the $\arccos(\cdot)$ arguments are clipped to $[-1, 1]$. Frames with missing keypoints or degenerate vectors with near-zero norms are excluded from offline metric computation or replaced by the latest valid estimate during online execution.

Appendix A.4. Perception-Side Sensitivity Analysis

A compact sensitivity analysis is conducted by perturbing the recorded skeleton streams before pose angle extraction. As summarized in Table A1, the implemented filtering, rate-limiting, and neutral-pose calibration reduce the effects of moderate keypoint noise, jitter, and calibration offsets, while invalid occlusion frames are conservatively rejected.

Table A1. Perception-side sensitivity analysis under representative perturbations.

Perturbation Type	Perturbation Level	Mean Angle Error	Peak Angle Error	Rejected Frames	Notes
Gaussian keypoint noise	$\sigma = 1.5\,\text{cm}$	2.2°	5.8°	0%	geometric vector estimation and EMA smoothing
Gaussian keypoint noise	$\sigma = 4.0\,\text{cm}$	4.9°	11.3°	0%	stable angle estimates, but with increased variance
Single joint keypoint occlusion	one elbow/wrist missing	6.1°	12.7°	20%	invalid frames are conservatively discarded
Skeleton jitter	5° high-frequency jitter	1.8°	8.2°	0%	rate limiting suppresses short-term spikes
Calibration offset	$\pm 3\,\text{cm}/\pm 3°$	3.1°	7.9°	0%	neutral-pose calibration reduces fixed biases

Appendix B. CBF Geometry Benchmarks

To motivate the use of capsule collision geometry, the performance is compared with other commonly used models, including spheres and bounding boxes, in the human–humanoid simulation environment. Table A2 shows the performance results of various collider geometries. For sphere decomposition, the center-to-center squared distance is used, implemented via [44]. For capsules, the signed distance between sphere-swept volumes method is used and implemented via [45]. For boxes, the convex polytope proximity method is used, implemented via [46]. These methods all provide differential proximity measurements that CBF can utilize. Benchmarks are performed on a computer with an Intel Core Ultra 9 275HX CPU and an NVIDIA RTX 5080 Laptop GPU. A visualization of these collision models is shown in Figure A1.

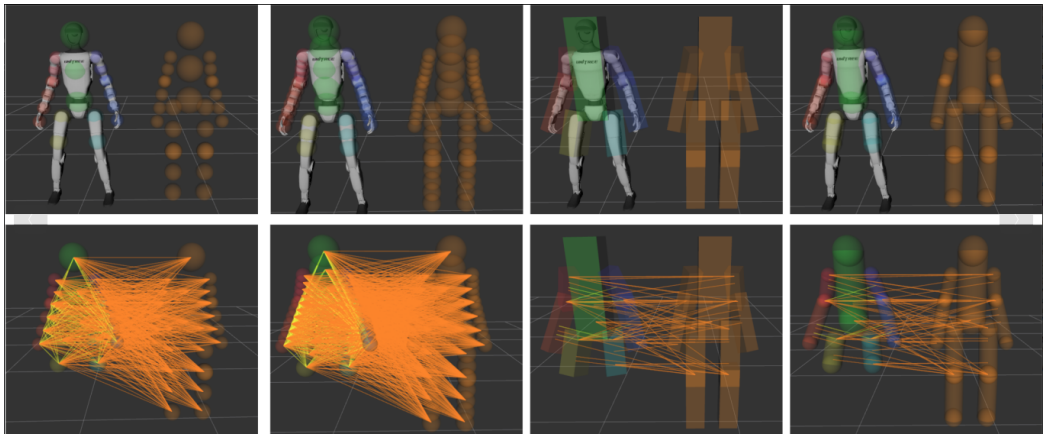


Figure A1. Benchmarked collision geometry. From left to right: Sphere decomposition with no subsampling, sphere decomposition with one subsampling, bounding box decomposition, capsule decomposition. Lines represent CBF constraints.

The computational performance of different implementation backends is also benchmarked. As shown in Table A2, utilizing the GPU results in significant performance uplift due to the number of parallel constraints. Differentiable kinematics are implemented in the CPU via Pinocchio [47], while the GPU methods utilize JAX [48]. The best overall performance was demonstrated by capsule geometry on the GPU, demonstrating both low solve time and minimal constraint blow-up.

Table A2. CBF benchmark statistics for various collision geometries.

	CPU			
Metric	Spheres ($k = 0$)	Spheres ($k = 1$)	Boxes	Capsule
Rate (Hz)	10.02	8.5	10.424	4.113
# Constraints	705	1988	72	72
Initial Compile Time (s)	0.02	0.05	0.4	0.7
Solve Time (s)	0.100	0.118	0.096	0.243
	GPU			
Metric	Spheres ($k = 0$)	Spheres ($k = 1$)	Boxes	Capsule
Rate (Hz)	30.5	11.549	3.937	33.084
# Constraints	705	1988	72	72
Initial Compile Time (s)	29.4	184.8	36.7	29.4
Solve Time (s)	0.033	0.087	0.254	0.030

Note: k denotes the degree of spherical subsampling used to approximate capsule geometry. All benchmarks use a QP solver with a maximum of 100 iterations and a solution tolerance of 10^{-3}. Results are reported for a worst-case setting in which CBF constraints are imposed between all external colliders and all other internal colliders, highlighting computational differences across geometry models.

References

1. Sheng, Q.; Zhou, Z.; Li, J.; Mi, X.; Xiang, P.; Chen, Z.; Xu, H.; Jia, S.; Wu, X.; Cui, Y.; et al. A comprehensive review of humanoid robots. *SmartBot* **2025**, *1*, e12008. [CrossRef]
2. Sugihara, T.; Morisawa, M. A survey: Dynamics of humanoid robots. *Adv. Robot.* **2020**, *34*, 1338–1352. [CrossRef]
3. Cai, W.; Vamvoudakis, K.G.; Gros, S.; Tzes, A. Cost-Matching Model Predictive Control for Efficient Reinforcement Learning in Humanoid Locomotion. *arXiv* **2026**, arXiv:2603.28243.
4. Darvish, K.; Penco, L.; Ramos, J.; Cisneros, R.; Pratt, J.; Yoshida, E.; Ivaldi, S.; Pucci, D. Teleoperation of humanoid robots: A survey. *IEEE Trans. Robot.* **2023**, *39*, 1706–1727. [CrossRef]
5. Vianello, L.; Penco, L.; Gomes, W.; You, Y.; Anzalone, S.M.; Maurice, P.; Thomas, V.; Ivaldi, S. Human-humanoid interaction and cooperation: A review. *Curr. Robot. Rep.* **2021**, *2*, 441–454. [CrossRef]
6. Ajoudani, A.; Zanchettin, A.M.; Ivaldi, S.; Albu-Schäffer, A.; Kosuge, K.; Khatib, O. Progress and prospects of the human–robot collaboration. *Auton. Robot.* **2018**, *42*, 957–975.
7. Cai, W.; Abanes, J.; Evangeliou, N.; Tzes, A. Safe Human-to-Humanoid Motion Imitation Using Control Barrier Functions. *arXiv* **2026**, arXiv:2604.11447.
8. Montecillo-Puente, F.J.; Narsipura Sreenivasa, M.; Laumond, J.P. On real-time whole-body human to humanoid motion transfer. In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*; SciTePress: Setúbal, Portugal, 2010; pp. 22–31.
9. Otani, K.; Bouyarmane, K. Adaptive whole-body manipulation in human-to-humanoid multi-contact motion retargeting. In *Proceedings of the 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*; IEEE: Piscataway, NJ, USA, 2017; pp. 446–453.
10. Darvish, K.; Tirupachuri, Y.; Romualdi, G.; Rapetti, L.; Ferigo, D.; Chavez, F.J.A.; Pucci, D. Whole-body geometric retargeting for humanoid robots. In *Proceedings of the 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*; IEEE: Piscataway, NJ, USA, 2019; pp. 679–686.
11. Baek, S.; Kim, A.; Choi, J.Y.; Ha, E.; Kim, J.W. Human motion retargeting to a full-scale humanoid robot using a monocular camera and human pose estimation. *Int. J. Control. Autom. Syst.* **2024**, *22*, 2860–2870. [CrossRef]
12. Jiang, X.; Wu, B.; Li, S.; Zhu, Y.; Liang, G.; Yuan, Y.; Li, Q.; Zhang, J. Multi-Humanoid Robot Arm Motion Imitation and Collaboration Based on Improved Retargeting. *Biomimetics* **2025**, *10*, 190. [CrossRef] [PubMed]
13. He, T.; Luo, Z.; Xiao, W.; Zhang, C.; Kitani, K.; Liu, C.; Shi, G. Learning human-to-humanoid real-time whole-body teleoperation. In *Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: Piscataway, NJ, USA, 2024; pp. 8944–8951.
14. He, T.; Luo, Z.; He, X.; Xiao, W.; Zhang, C.; Zhang, W.; Kitani, K.; Liu, C.; Shi, G. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. *arXiv* **2024**, arXiv:2406.08858.
15. Li, J.; Yang, L.; Ames, A.D. MIRROR: Visual Motion Imitation via Real-time Retargeting and Teleoperation with Parallel Differential Inverse Kinematics. *arXiv* **2026**, arXiv:2603.23995.

16. Ze, Y.; Chen, Z.; Araújo, J.P.; Cao, Z.a.; Peng, X.B.; Wu, J.; Liu, C.K. Twist: Teleoperated whole-body imitation system. *arXiv* **2025**, arXiv:2505.02833.
17. Lasota, P.A.; Fong, T.; Shah, J.A. A survey of methods for safe human-robot interaction. *Found. Trends® Robot.* **2017**, *5*, 261–349. [CrossRef]
18. Villani, V.; Pini, F.; Leali, F.; Secchi, C. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics* **2018**, *55*, 248–266. [CrossRef]
19. Zanella, R.; Califano, F.; Stramigioli, S. Physical Human-Robot Interaction: A Critical Review of Safety Constraints. *arXiv* **2026**, arXiv:2601.19462.
20. Ames, A.D.; Xu, X.; Grizzle, J.W.; Tabuada, P. Control barrier function based quadratic programs for safety critical systems. *IEEE Trans. Autom. Control* **2016**, *62*, 3861–3876. [CrossRef]
21. Xiao, W.; Cassandras, C.G.; Belta, C. *Safe Autonomy with Control Barrier Functions: Theory and Applications*; Springer: Cham, Switzerland, 2023.
22. Ames, A.D.; Coogan, S.; Egerstedt, M.; Notomista, G.; Sreenath, K.; Tabuada, P. Control barrier functions: Theory and applications. In *Proceedings of the 2019 18th European control conference (ECC)*; IEEE: Piscataway, NJ, USA, 2019; pp. 3420–3431.
23. Ferraguti, F.; Landi, C.T.; Singletary, A.; Lin, H.C.; Ames, A.; Secchi, C.; Bonfe, M. Safety and efficiency in robotics: The control barrier functions approach. *IEEE Robot. Autom. Mag.* **2022**, *29*, 139–151. [CrossRef]
24. Gonçalves, V.; Giakoumidis, N.; Moore, M.; Tzes, A. Leader-Follower Human-Cobot Improvised Dance Using Motion Capture Systems. In *Proceedings of the International Conference on ArtsIT, Interactivity and Game Creation*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 125–137.
25. Chaikalis, D.; Unlu, H.U.; Tzes, A.; Khorrami, F. Vision-aided Leader-Follower Collaborative Mobile Manipulation with Control Barrier Functions. *J. Intell. Robot. Syst.* **2026**, *112*, 24. [CrossRef]
26. Luo, R.; Jakobsen, J.M.; Roozing, W.; Califano, F.; Fang, C. Control Barrier Functions Solved with Hierarchical Quadratic Programming for Safe Physical Human-Robot Interaction. *arXiv* **2026**, arXiv:2604.23039.
27. Maithani, P.; Arab, A.; Khorrami, F.; Krishnamurthy, P. Proactive hierarchical control barrier function-based constraint prioritization to enhance safety in human-robot interaction. *Control Eng. Pract.* **2026**, *166*, 106624. [CrossRef]
28. Khazoom, C.; Gonzalez-Diaz, D.; Ding, Y.; Kim, S. Humanoid self-collision avoidance using whole-body control with control barrier functions. In *Proceedings of the 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*; IEEE: Piscataway, NJ, USA, 2022; pp. 558–565.
29. Marew, D.; Lvovsky, M.; Yu, S.; Sessions, S.; Kim, D. Integration of Riemannian motion policy with whole-body control for collision-free legged locomotion. In *Proceedings of the 2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–8.
30. Khazoom, C.; Hong, S.; Chignoli, M.; Stanger-Jones, E.; Kim, S. Tailoring solution accuracy for fast whole-body model predictive control of legged robots. *IEEE Robot. Autom. Lett.* **2024**, *9*, 11074–11081. [CrossRef]
31. Ferraguti, F.; Landi, C.T.; Costi, S.; Bonfè, M.; Farsoni, S.; Secchi, C.; Fantuzzi, C. Safety barrier functions and multi-camera tracking for human–robot shared environment. *Robot. Auton. Syst.* **2020**, *124*, 103388. [CrossRef]
32. Parma, F.; Tonola, C.; Pedrocchi, N.; Beschi, M. Embedding ISO 10218 Safety Compliance in Robots via Control Barrier Functions for Human-Robot Collaboration. *arXiv* **2026**, arXiv:2606.13203.
33. Shi, K.; Hu, G. Safety-guaranteed and task-consistent human-robot interaction using high-order time-varying control barrier functions and quadratic programs. *IEEE Robot. Autom. Lett.* **2023**, *9*, 547–554.
34. Liu, R.; Chen, R.; Liu, C. Safe interactive industrial robots using jerk-based safe set algorithm. In *Proceedings of the International Symposium on Flexible Automation 2022 International Symposium on Flexible Automation*; The Institute of Systems, Control and Information Engineers: Kyoto, Japan, 2022; pp. 196–203.
35. Xie, Z.; Dames, P. Drl-vo: Learning to navigate through crowded dynamic scenes using velocity obstacles. *IEEE Trans. Robot.* **2023**, *39*, 2700–2719. [CrossRef]
36. Marcucci, T.; Petersen, M.; Von Wrangel, D.; Tedrake, R. Motion planning around obstacles with convex optimization. *Sci. Robot.* **2023**, *8*, eadf7843. [CrossRef] [PubMed]
37. Haffemayer, A.; Jordana, A.; De Matteïs, L.; Wojciechowski, K.; Righetti, L.; Lamiraux, F.; Mansard, N. Collision avoidance in model predictive control using velocity damper. In *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: Piscataway, NJ, USA, 2025; pp. 16140–16146.
38. Dickson, A.K.; Garcia, J.C.P.; Anderson, M.L.; Jing, R.; Alizadeh-Shabdiz, S.; Wang, A.X.; DeLorey, C.; Patterson, Z.J.; Sabelhaus, A.P. Safe autonomous environmental contact for soft robots using control barrier functions. *IEEE Robot. Autom. Lett.* **2025**, *10*, 11283–11290. [CrossRef]
39. Wong, K.; Stölzle, M.; Xiao, W.; Della Santina, C.; Rus, D.; Zardini, G. Contact-aware safety in soft robots using high-order control barrier and lyapunov functions. *IEEE Robot. Autom. Lett.* **2025**, *10*, 12485–12492. [CrossRef]

40. Zhang, Z.; Niu, Y.; Yan, Z.; Lin, S. Real-time whole-body imitation by humanoid robots and task-oriented teleoperation using an analytical mapping method and quantitative evaluation. *Appl. Sci.* **2018**, *8*, 2005. [CrossRef]
41. Gong, L.; Chen, B.; Xu, W.; Liu, C.; Li, X.; Zhao, Z.; Zhao, L. Motion similarity evaluation between human and a tri-co robot during real-time imitation with a trajectory dynamic time warping model. *Sensors* **2022**, *22*, 1968. [CrossRef] [PubMed]
42. Deriso, D.; Boyd, S. A general optimization framework for dynamic time warping. *Optim. Eng.* **2023**, *24*, 1411–1432.
43. Stereolabs. ZED X Stereo Camera. 2026. Available online: https://www.stereolabs.com/en-si/store/products/zed-x-stereo-camera (accessed on 21 July 2026).
44. Morton, D.; Pavone, M. Safe, task-consistent manipulation with operational space control barrier functions. In *Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: Piscataway, NJ, USA, 2025; pp. 187–194.
45. Tracy, K.; Howell, T.A.; Manchester, Z. Diffpills: Differentiable collision detection for capsules and padded polygons. *arXiv* **2022**, arXiv:2207.00202.
46. Tracy, K.; Howell, T.A.; Manchester, Z. Differentiable collision detection for a set of convex primitives. In *Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: Piscataway, NJ, USA, 2023; pp. 3663–3670.
47. Carpentier, J.; Saurel, G.; Buondonno, G.; Mirabel, J.; Lamiraux, F.; Stasse, O.; Mansard, N. The Pinocchio C++ library – A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives. In *Proceedings of the IEEE International Symposium on System Integrations (SII)*; IEEE: Piscataway, NJ, USA, 2019.
48. Bradbury, J.; Frostig, R.; Hawkins, P.; Johnson, M.J.; Katariya, Y.; Leary, C.; Maclaurin, D.; Necula, G.; Paszke, A.; VanderPlas, J.; et al. JAX: Composable Transformations of Python+NumPy Programs. Version 0.3.13. 2018. Available online: http://github.com/jax-ml/jax (accessed on 21 July 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.