Can We Tune Humanoid Behavior Foundation Models for Dynamic and Contact-Rich Tasks?

Author Names Omitted for Anonymous Review

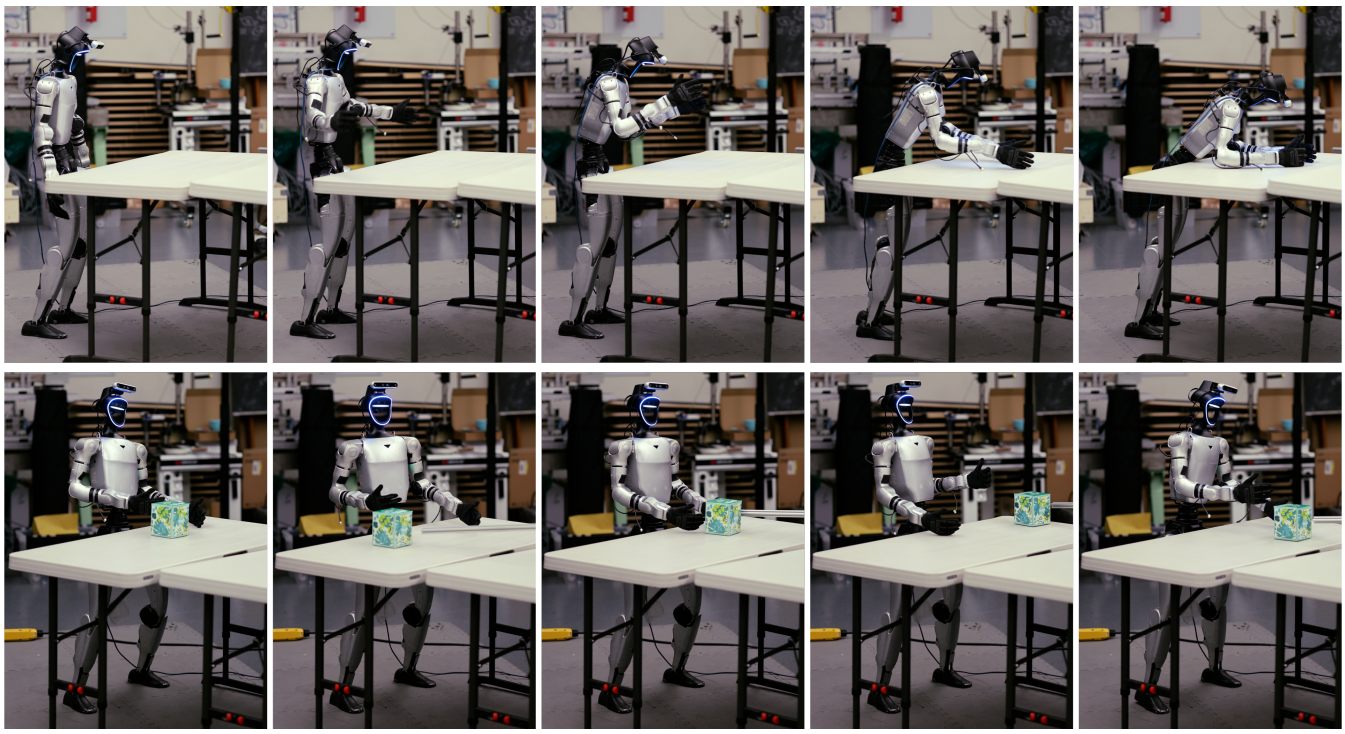

Fig. 1: **Overview of SONAR.** We adapt a pre-trained humanoid behavior foundation model to contact-rich and dynamic downstream tasks using a lightweight residual module. Top: the robot leans on a table while maintaining whole-body balance and environmental contact. Bottom: the robot reaches for a moving object using semantic-spatial observations from depth and dynamic object masks given a constant reference motion.

***Abstract*—While behavior foundation models (BFMs) offer a promising paradigm for humanoid whole-body control (WBC), their task-agnostic optimization for general motion tracking and lack of perception limits performance on specialized downstream tasks that require precise contacts, constraint satisfaction, and rapid response to moving objects. To bridge this gap, we present SONAR, a residual adaptation framework that customizes a pre-trained humanoid BFM, equipping it with perception via a lightweight plug-in module. SONAR augments the base policy with depth observations and a dynamic object-of-interest mask, allowing spatial geometry and explicit task semantics to enter the low-level control loop. Conditioned on these semantic-spatial observations, SONAR predicts a residual action superimposed on the original BFM action. By injecting these signals into low-level execution, SONAR enables rapid task-specific corrections without slower high-level feedback. We develop a structured training pipeline combining task-oriented motion augmentation, reinforcement learning, and teacher-student training trick for dynamic scenarios. We evaluate SONAR on two real-world G1 tasks spanning environmental contact and dynamic reaching. Our framework improves success, contact maintenance, and end-effector tracking precision while preserving the stability and generality of the underlying BFM.**

I. Introduction

Humanoid whole-body control (WBC) has progressed from carefully engineered controllers toward large, reusable policies capable of tracking diverse full-body motions. Recent systems trained on large-scale motion data have established a new paradigm known as Behavior Foundation Models (BFMs) [44, 17, 5, 24, 42, 46, 47]. These models provide a unified low-level interface for executing whole-body behaviors, allowing high-level planners, vision-language models, or manually designed reference motions to command complex humanoid actions without directly handling balance, contacts, and actuation constraints. In this sense, BFMs serve as reusable execution

priors that make whole-body humanoid behavior easier to specify, transfer, and compose.

However, the same design choices that make BFM-based controllers general also limit their performance on specialized downstream tasks. First, a broadly trained tracking policy is optimized for average tracking quality across many motions, not for the precision, contact timing, or forceful interactions required by a particular task. A reference motion that is semantically correct may still be insufficient when the robot must lean onto a table, step onto furniture, or reach a moving object with small end-effector error. Second, the common low-level interface of existing BFMs is often perception-light: the policy tracks a reference or action token while having little direct access to the task-relevant geometry and semantics of the surrounding scene. Such policies are closed-loop with respect to robot dynamics, but effectively open-loop with respect to task completion: they may stabilize the body and track the commanded motion, yet still miss the intended object, contact location, or scene change.

A natural solution is to customize a pre-trained BFM instead of training a new specialist controller from scratch. In language and vision-language-action models, pre-training followed by adaptation [9, 36, 31], steering [7, 3, 19], or fine-tuning [28, 18, 25, 4] is a standard recipe for specializing a general model. For humanoid BFMs, this recipe remains less developed. Directly fine-tuning a large low-level BFM can be computationally prohibitive and risks degrading the broad motion prior that made the model useful in the first place. And also it is unable to introduce extral modalities such as perception, due to the fixed input and output constraint. Another option is to place a high-level planner or vision-language model above the BFM to provide semantic feedback, but such systems typically operate at lower frequencies and can introduce delayed sensorimotor responses. A separate line of work injects depth images into specialist perceptive loco-manipulation policies [38, 15, 49, 51, 52], but they typically lack a reusable foundation-model interface due to highly task-specified velocity command, or encode geometry without explicit task semantics. This raises the central question of this paper: *Can we adapt a low-level WBC behavior foundation model with additional semantic and spatial information for improved downstream performance, especially in dynamic and contact-rich scenarios?*

We answer this question with **S**emantic **O**bject-aware **N**eural **A**daptation with **R**esiduals (`SONAR`), a residual adaptation framework built on top of SONIC [17], a representative humanoid behavior foundation model. Rather than replacing or fully fine-tuning the base policy, `SONAR` learns a lightweight plug-in residual module that operates in parallel with the frozen BFM. At each control step, the module receives the base policy's motion latents, robot proprioception, depth observations, and a dynamic object-of-interest mask, and predicts a residual action that is added to the original BFM action. Depth observations provide local spatial geometry for contact and obstacle interaction, while object masks provide explicit task semantics by indicating the relevant object or region. This design injects semantic-spatial feedback directly into the low-level execution loop, enabling fast task-specific corrections without sacrificing the stabilization and motion-tracking priors of the underlying controller.

Our contributions are threefold. First, we formulate residual adaptation as a practical mechanism for specializing humanoid behavior foundation models while preserving their original motion priors. Second, we introduce a low-level semantic-spatial perception interface that combines depth images with dynamic object masks, allowing the controller to correct task-relevant errors at high frequency rather than relying solely on slower high-level feedback. Third, we develop a structured training pipeline with task-oriented motion augmentation, reinforcement learning, and teacher-student distillation, and evaluate the framework on two real-world Unitree G1 tasks involving environmental contact, dynamic reaching, and contact-rich manipulation.

II. Related Work

A. Humanoid Behavior Foundation Models

Humanoid WBC aims to coordinate locomotion, balance, manipulation, and contact through a unified full-body controller, a problem complicated by high degrees of freedom, underactuation, and frequent contact transitions [26, 45]. Classical model-based controllers typically separate locomotion, posture, and upper-body objectives, which aids modularity but limits tightly coupled whole-body behaviors [26]. Recent reinforcement-learning-based methods instead learn to follow large-scale motion data while respecting balance, torque, and contact constraints [21, 22, 6], and when scaled across diverse behaviors they act as *behavior foundation models* (BFMs): pretrained controllers that encode broad behavioral priors and expose compact interfaces, such as reference motions, latent skills, or reward embeddings, to downstream tasks [23, 34]. Most such BFMs are trained by goal-conditioned tracking of reference motions [32, 8, 41, 17], while a separate line learns reward-free policies via forward-backward representations [35, 34, 12]. A parallel effort couples these controllers with high-level planners (LLMs, diffusion, or vision-language models) for long-horizon, language-conditioned tasks [20, 33, 27, 40].

However, most humanoid BFMs remain optimized for general tracking rather than task-specific interaction. They typically condition on reference motions, commands, or latent actions [32, 8], but have limited access to the spatial geometry and explicit object semantics that determine task success. Therefore, they may stabilize the robot and track plausible motions while still missing the intended contact, applying insufficient interaction force, or reacting slowly to scene changes. `SONAR` therefore adopts SONIC [17] as its base BFM, preserving it as a stabilizing prior while adding a task-conditioned residual module for semantic-spatial correction, closing the low-level control loop where precise contact timing and applied force are realized.

B. Perceptive Loco-Manipulation

Perceptive locomotion and loco-manipulation methods use onboard sensing to adapt motion to terrain, obstacles, and objects. Depth images, height maps, point clouds, and geometric encodings have enabled parkour [38, 50], climbing [49], stepping [15, 51], terrain traversal [29], and whole-body environmental interaction [43]. These works show that spatial perception is essential for reactive contact-rich behaviors.

Yet many perceptive loco-manipulation systems are trained as specialist controllers, making them difficult to reuse as general low-level interfaces. Moreover, geometry alone does not specify task relevance: a depth map can indicate where objects are, but not which object or region should guide the current behavior. Methods that do pair vision with whole-body control typically route it through a high-level policy that generates motion or pose targets for a separate low-level tracker [43], so object semantics reach the controller only indirectly. `SONAR` instead provides depth and dynamic object masks directly to low-level control, using depth for spatial geometry and masks as explicit semantic pointers, yielding a reusable low-level controller that is both spatially grounded and object-aware.

C. Adaptation and Residual Policies

Foundation models are commonly specialized through adaptation, steering, or lightweight test-time fine-tuning to mitigate deployment-time distribution shifts [39, 30, 2]. For humanoid BFMs, directly fine-tuning the full low-level controller can be costly and may degrade the broad motion prior, while relying only on high-level replanning or test-time guidance can introduce delayed sensory-motor feedback and may push generated trajectories away from the data manifold [10, 13, 50]. Residual learning offers a lightweight alternative: a residual policy learns an additive correction on top of an existing controller, allowing the base policy to maintain stability and reusable motion priors while the residual focuses on task-specific errors, perceptual biases, and local mismatches [48, 1, 37, 16]. `SONAR` applies this idea to humanoid BFM adaptation by learning a perception-conditioned residual over the base action, injecting semantic-spatial feedback directly into the low-level control loop for rapid adaptation to unseen complex terrains.

III. Method

`SONAR` adapts a pre-trained humanoid WBC foundation model to downstream contact-rich and perception-required tasks by learning an additive residual action. The base policy, denoted π_b, receives its original motion representation and produces an action a_t^b. Our residual module, denoted π_r, receives the base-policy representation together with proprioception, depth, and object-mask observations, and predicts Δa_t. The executed action is

$$a_t = a_t^b + \alpha \Delta a_t, \tag{1}$$

where α is a residual scale. This design intentionally makes the adaptation conservative: the base model remains responsible for stable whole-body motion, while the residual module learns task-specific corrections that require spatial or semantic information unavailable to the base policy.

A. Spatial and Semantic Perception

As shown in Figure 2, the residual module injects two forms of information that are missing from a standard tracking interface, to form a perceptive low-level control system: local scene geometry and task semantics. For geometry, we mount two depth cameras on the robot and vertically stack their depth images into a single observation. The stacked representation provides a compact egocentric encoding of near-field geometry such as tabletops or reachable objects.

For semantics, we introduce a dynamic object-of-interest mask. The mask indicates which pixels correspond to the object that the high-level task refers to, allowing the low-level residual policy to distinguish task-relevant objects from background geometry. In simulation, the object mask is obtained from privileged scene information. At deployment time, we use a perception pipeline operating on synchronized RGB-D streams. A text description of the target object is first used to obtain an RGB bounding box at low frequency with the open-vocabulary detector GroundingDINO [14]. After initialization, we use the visual tracker SiamRPN++ [11] to keep updating the bounding box at the camera rate. The tracked RGB box is then aligned to the depth image, and simple depth clustering within the box separates the foreground object from the background. The final binary mask is stacked with the depth observation and sent to the residual module.

This interface bridges high-level semantic intent and low-level control. The high-level system only needs to specify the target object category or description; the residual controller receives a dense, temporally updated mask that can be used for reactive correction. For tasks without a dynamic target, the mask can be empty or can encode the relevant environmental surface, depending on the task design.

B. Residual Adapter Architecture

The overall architecture is shown in Figure 3. We instantiate the base controller with SONIC. SONIC encodes the input motion into a latent representation, quantizes it through an action-token representation, and decodes the token into whole-body actions. `SONAR` treats this latent feature as an informative summary of the intended base motion. The residual adapter receives five inputs: (i) the SONIC motion latent, (ii) the SONIC action output, (iii) robot proprioception, including joint states, base information, and action history, (iv) stacked depth images, and (v) the object-of-interest mask. A visual CNN encoder processes the depth-mask observation, an MLP encodes proprioception and the base-policy representation, and the fused feature is decoded into a residual action with the same dimension as the base action.

The residual action can be interpreted as a task-aware correction to the foundation model. For environmental contact tasks, it adjusts body posture, limb placement, and contact timing according to nearby geometry. For dynamic reaching tasks, it shifts the end-effector trajectory toward the tracked object

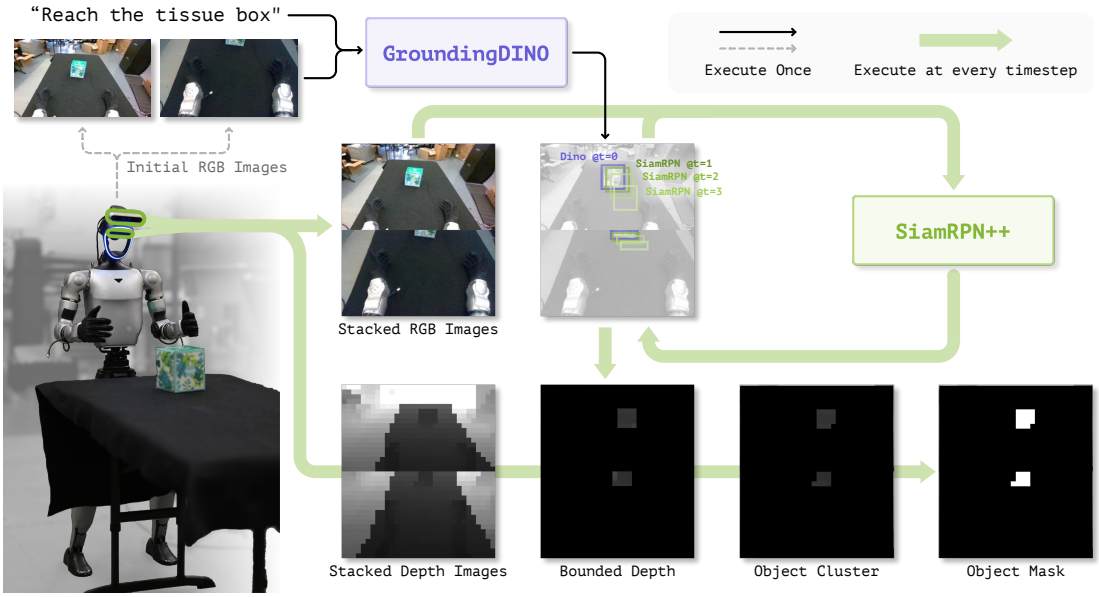

Fig. 2: **Spatial and semantic observations.** SONAR builds a low-level semantic-spatial observation from two RGB-D cameras. A text query initializes the target with GroundingDINO, and SiamRPN++ tracks it at every timestep. The tracked box is aligned with the stacked depth image, and depth clustering extracts a dynamic object-of-interest mask. The final depth-mask observation provides local geometry and explicit task semantics to the residual policy.

while preserving whole-body balance. For combined tasks, it allows the robot to maintain support on the environment while using the free limb to reach a target.

C. Training Pipeline

We train residual adapters for the task families implemented in our system. The first family is *dynamic object reaching*, where the nominal reference motion specifies a reaching intention but does not fully determine the object pose encountered at execution time. This setting is sensitive to small perception, calibration, or timing errors because the residual policy must correct the arm online while preserving whole-body balance. The second family is *environment interaction*, where the robot intentionally uses the scene as part of the behavior, such as leaning on a table or kneeling and climbing onto a bed. In both cases, the base SONIC policy provides a stable whole-body motion prior, and the residual policy learns task-specific corrections from augmented simulation experience.

a) Reference motion augmentation: To make the residual policy locally generalize within each base-motion modality, we augment reference motions by varying the reaching target while preserving the coarse whole-body behavior (Figure 4).

Starting from several clips of whole-body contact rich motion, such as leaning on table while reaching a target forward, we keep the support and contact structure of the original motion fixed, sample manipulation targets inside the intersection of the arm reachability region and the camera visibility region, and use inverse kinematics to get the arm motion. This gives feasible variations that remain in the same whole body task modality, while only varies on local motion.

For dynamic object reaching, we additionally create paired nominal and correction motions. The policy observes a nominal reference whose target stays fixed, while the simulated object follows a drifted trajectory. We compute a drift-corrected ground-truth motion that reaches the moving object, providing ground-truth reference for how the arm should compensate when the object deviates from the reference.

b) Residual policy optimization: We train the residual policy in simulation with reinforcement learning, combining motion tracking, task rewards, and regularization:

$$r_t = w_{\text{track}} r_t^{\text{track}} + w_{\text{task}} r_t^{\text{task}} + w_{\text{reg}} r_t^{\text{reg}}. \tag{2}$$

The tracking term keeps the residual close to the stable SONIC behavior, while regularization penalizes nonsmooth or high-energy corrections. The task reward depends on the behavior: contact rewards encourage intended support contacts for environment interaction, while dynamic reaching uses end-effector-to-object distance and arm tracking rewards against the drift-corrected ground truth.

c) Teacher-student auxiliary loss: To improve sample efficiency and assist reaching-intention learning, we add an auxiliary teacher-student loss for dynamic reaching. The teacher action is produced by running the base SONIC controller on the drift-corrected per-step reference, which represents the privileged action that would track the moving object if the true correction were known. The student is the deployable residual policy, which receives only the nominal reference and scene observations. During intervals where the object drifts, we add a mean-squared error loss between the teacher arm action and the student residual arm action. This auxiliary loss provides a dense training signal that guides the residual policy toward

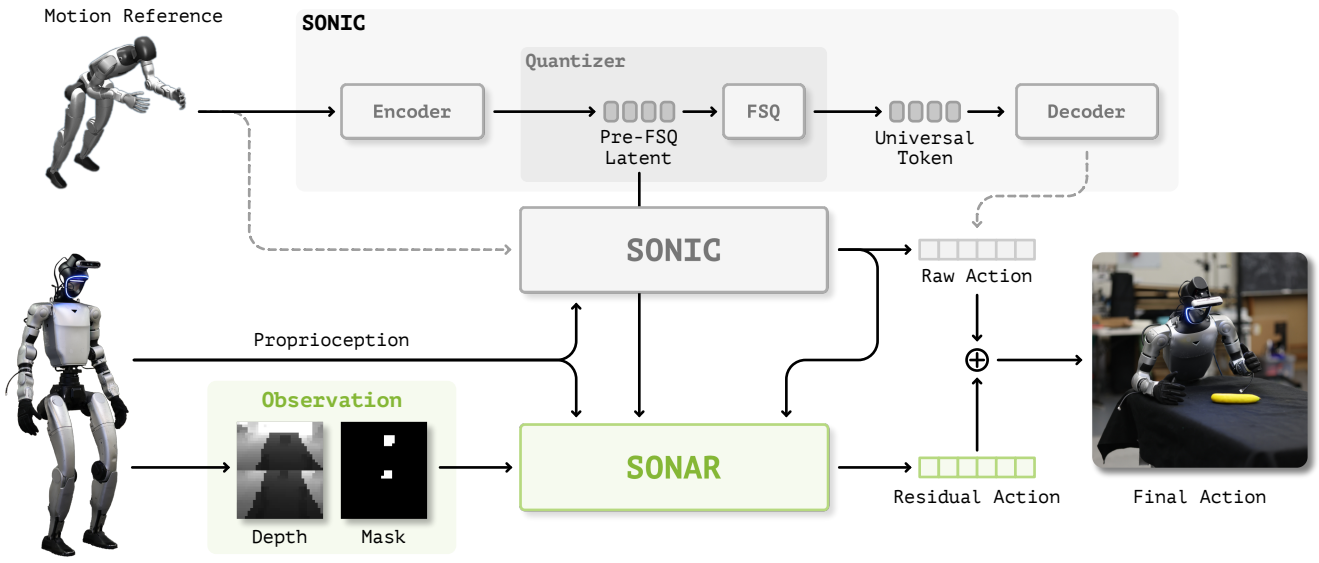

Fig. 3: **Residual adaptation pipeline.** The pre-trained SONIC controller remains frozen and produces a base action from the motion reference and proprioception. SONAR adds a lightweight residual pathway conditioned on the SONIC latent representation, proprioception, stacked depth images, and dynamic object masks. The final action is obtained by adding the residual action to the base action, enabling task-specific semantic-spatial corrections while preserving SONIC's motion prior.

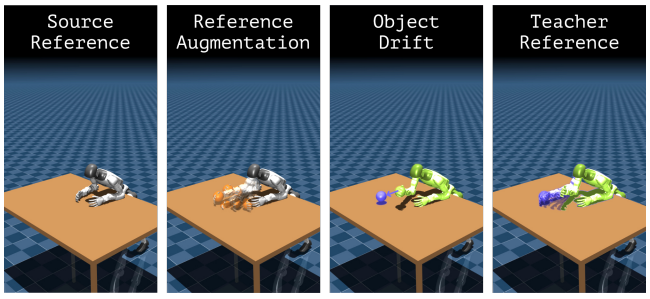

Fig. 4: **Data Augmentation.** Starting from a feasible source reference, we generate task-specific training data through reference augmentation and object drift. For contact-rich reaching tasks, we preserve the coarse whole-body contact structure while varying the manipulation target. For dynamic targets, we create paired nominal and drift-corrected references, where the corrected reference serves as a teacher signal. The residual policy is then trained with reinforcement learning and an optional teacher-student auxiliary loss for dynamic target response.

object-tracking corrections while the RL rewards enforce task success, balance, contact behavior, and smoothness.

D. Real-Robot Deployment

Our real-robot deployment is laptop-centric: all policy inference and image processing pipeline run on a laptop equipped with an NVIDIA GeForce RTX 4060 Laptop GPU. The robot only streams sensor data and receives low-level commands. A lightweight onboard relay sends synchronized RGB-D from two RealSense cameras, a head-mounted D455 and a face-mounted D435I, over TCP to the laptop. The relay is configured to stream both cameras at 60 Hz, using 424×240 RGB per camera and downsampled depth for lower latency. The laptop converts the dual-camera stream into the policy visual observation, a stacked normalized depth and object-mask tensor of size $36 \times 32 \times 2$ ($18 \times 32 \times 2$ for each camera), with the head view on top and the face view below.

On the laptop, the perception sidecar initializes the object with GroundingDINO and then tracks it using the PySOT SiamRPN MobileNetV2 L234 depthwise-correlation model. In our default setting, GroundingDINO is used for initialization and then stopped after the first successful detection; SiamRPN runs on each perception update, and the bounding box is converted to a mask by depth clustering. The perception sidecar runs asynchronously from the policy inference and publish the latest depth-mask tensor at up to 50 Hz. The control process subscribe to this and always uses the freshest frame available.

For teleoperation, we use SONIC's POSE-mode Pico teleoperation for user commands, feeding streamed SMPL human motion command to policy.

IV. Experiments

We design experiments to answer three questions. **Q1:** Can SONAR customize a WBC foundation model for contact-rich downstream tasks? **Q2:** Do depth images and dynamic masks effectively inject spatial and semantic information into low-level control? **Q3:** Can the system solve real-world tasks that have environmental contact and dynamic reaching?

TABLE I: **Foundation-model customization.** Base SONIC controller vs. SONAR on environmental-contact tasks. Success is the per-rollout completion rate (finished the motion, terminated only by timeout); Contact is the fraction of reference-required support contacts maintained; MPJPE is global mean per-joint position error (mm); Energy is mean actuator mechanical power (W). Metrics average over alive frames.

Method	Task	Success ↑	Contact ↑	MPJPE ↓	Energy ↓
SONIC	Bed kneel/climb	0.07	**0.85**	221.4	48.9
SONAR	Bed kneel/climb	**0.92**	0.76	**112.5**	**28.3**
SONIC	Table lean	**1.00**	0.15	98.3	23.1
SONAR	Table lean	**1.00**	**0.98**	**52.0**	**12.1**

A. Experimental Setup

Robot. We evaluate on a Unitree G1 humanoid robot. The base controller is SONIC, which provides the nominal whole-body action. SONAR adds the residual module described in Section III. The robot receives proprioception and synchronized RGB-D observations from two depth cameras.

Tasks. We evaluate two tasks that cover environmental geometry awareness and dynamic object interaction:

1) **Table lean:** the robot leans on the table, using the contact as an external support that keeps balance. Without contact support, the motion itself is physically implausible
2) **Bed kneel/climb:** the robot kneels one leg onto a bed-height surface and leans into a two-hand, one-knee multi-contact support, then lifts the remaining leg onto the bed, transferring support from the floor to the bed.
3) **Moving-object reach:** the robot reaches a moving object with its hand, under inaccurate reference motion.

Baselines and ablations. The primary baseline is the original SONIC policy without residual adaptation.

Metrics. We report task success rate, mean per-joint position error (MPJPE), control energy, and end-effector distance to the target object. For dynamic reaching, we additionally plot palm-to-object distance over time to measure responsiveness.

B. Foundation-Model Customization

The first experiment evaluates whether a residual adapter can specialize the base WBC model for tasks that require environmental contact. The table-lean and bed-kneel/climb tasks are difficult for a generic tracker. Because general trackers are typically not trained on scene interactions, they fail to leverage environmental support and tend to output overly conservative motions or trigger OOD terminations. Consequently, even when the reference motion is physically plausible, minute errors in body posture or contact timing can lead to catastrophic failure. We compare SONIC and SONAR on task success, tracking error, and energy.

The the result demonstrated that residual adaptation improves task success while keeping energy and tracking error controlled, as shown in Table I. For table leaning task, while SONIC tends not to lean on the table and always supports the bending pose via its waist effort, SONAR reduces energy consumption by leasing weight on the table. For bed task, SONIC always trigger OOD behavior once trying to kneel on the bed, so its ealuation result may be highly distorted and carry little statistical meaning.

This would indicate that the residual module is not replacing the foundation model, but steering it toward contact configurations that the base tracker does not reliably reach.

TABLE II: **Effect of spatial and semantic inputs.** Base SONIC controller vs. SONAR on the stand-reaching drift task (held-out). Success is the drift-correction success rate (we define it as end-effector closing $\geq$ half the drift-induced gap); EE dist. is the mean palm-to-target (drifted-object) distance; MPJPE is global mean per-joint position error.

Method	Success ↑	EE dist. (m) ↓	MPJPE (mm) ↓
SONIC	0.03	0.089	**34.5**
SONAR	**0.99**	**0.008**	**34.5**

C. Spatial and Semantic Information

The second experiment tests whether depth and masks provide useful low-level task information. In the moving-object reaching task, the reference motion alone is insufficient because the object may move after the reference is generated. For example, in scenarios involving teleoperation, it is highly impractical for a human operator to provide instantaneous, micro-level trajectory adjustments to match a dynamically shifting target in real time.

We compare the full SONAR model with the base SONIC controller on a held-out stand-reaching drift task to evaluate online target correction. As detailed in Table II, SONIC fails to respond to target object drift, achieving a success rate of only 3% and leaving a massive mean palm-to-target distance of 0.089 m. It is expected since SONIC is a pure reference tracker. By contrast, SONAR achieves a 99% success rate, dynamically minimizing the end-effector distance to just 0.008 m. A visualization is in Figure 5, showing the distance between robot end effector and object. Crucially, this significant gain in reaching accuracy is achieved without degrading body posture tracking; the MPJPE for both methods remains identical at 34.5 mm, proving that our residual module isolates its corrections to the relevant task space without destabilizing the foundation model's global motion prior.

D. Real-World Experiments

We conduct real-world experiments on two representative settings: table-leaning reference-motion tracking and stand hand-reaching teleoperation with a dynamic object.

For table-leaning, we use robot motion as command input, and compare the behavior qualitatively against the base SONIC conrtoller, focusing on whether the robot establishes stable table contact, maintains balance through the lean, and completes the motion without falling or prematurely leaving the support surface. SONAR successfully lean on the table

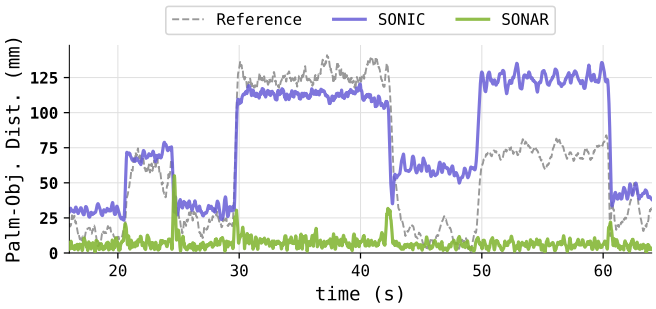

Fig. 5: **Dynamic reaching distance.** The curve shows the distance between object and reference motion, SONIC, and SONAR end effector, respectively. It can be seen that when reference motion have an approximately 10 cm gap from the actual object, SONAR could effectively close such gap via perception feedback.

firmly with its weights leased on table, and could support with one arm and spare another arm to do some reaching motion, while maintaining stable supporting.

For dynamic reaching, we use SONIC's POSE-mode Pico teleoperation interface. The operator provides smpl motion data, guiding the robot to reach an object placed on table. In the experiment, after the hand reaches the object, we move the object laterally while keeping the operator command unchanged. This simulates an under-specified teleoperation setting where the human provides only the coarse reaching intention, but not the fine-grained correction needed to follow a moving target. SONAR successfully uses the tracked depth-mask observation to generate an arm residual, moving the hand toward the new object position while preserving the standing whole-body posture. The image of these two experiments are shown in Figure 1.

These experiments show that the learned residual is effective as a foundation model customization module and a perception layer in the low-level control system.

V. Conclusion

We presented **SONAR**, a residual adaptation framework for customizing humanoid whole-body control foundation models. Instead of replacing the base controller, **SONAR** learns a perception-conditioned residual action that is added to the foundation-model action. This preserves the broad whole-body motion prior of the base policy while introducing task-specific spatial and semantic information through depth images and dynamic object masks. We described how to obtain these observations in simulation and on the real robot, how to train residual adapters through task-oriented motion augmentation and reinforcement learning, and how to deploy the combined policy on a Unitree G1 humanoid. Across environmental-contact and dynamic-reaching tasks, the framework is designed to improve task success and end-effector accuracy while retaining the stability of the underlying WBC model. More broadly, our results suggest that residual adaptation is a promising route for turning general humanoid WBC foundation models into practical task-aware controllers.

VI. Limitations

SONAR inherits both strengths and weaknesses from the base WBC foundation model. If the base model cannot produce a roughly feasible motion family for a task, a small residual correction may not be sufficient to recover the behavior. The method also requires task-specific reward design and task-relevant motion augmentation, which limits how automatically it can be applied to a new behavior. Finally, the quality of the residual policy depends on the perception stack: inaccurate object detection, tracking drift, poor depth alignment, or ambiguous masks can lead to incorrect corrections. Future work should explore more automatic task specification, stronger high-level planning integration, and uncertainty-aware residual control that can reason about perception failures.

References

[1] Minttu Alakuijala, Gabriel Dulac-Arnold, Julien Mairal, Jean Ponce, and Cordelia Schmid. Residual reinforcement learning from demonstrations. *arXiv preprint arXiv:2106.08050*, 2021.

[2] Ali Behrouz, Peilin Zhong, and Vahab Mirrokni. Titans: Learning to memorize at test time. *arXiv preprint arXiv:2501.00663*, 2024.

[3] Yuanpu Cao, Tianrong Zhang, Bochuan Cao, Ziyi Yin, Lu Lin, Fenglong Ma, and Jinghui Chen. Personalized steering of large language models: Versatile steering vectors through bi-directional preference optimization. *Advances in Neural Information Processing Systems*, 37: 49519–49551, 2024.

[4] Yuxin Chen, Devesh K Jha, Masayoshi Tomizuka, and Diego Romeres. Fdpp: Fine-tune diffusion policy with human preference. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12010–12016. IEEE, 2025.

[5] Zixuan Chen, Mazeyu Ji, Xuxin Cheng, Xuanbin Peng, Xue Bin Peng, and Xiaolong Wang. Gmt: General motion tracking for humanoid whole-body control. *arXiv preprint arXiv:2506.14770*, 2025.

[6] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. In *Robotics: Science and Systems (RSS)*, 2024.

[7] Yi Dong, Zhilin Wang, Makesh Sreedhar, Xianchao Wu, and Oleksii Kuchaiev. Steerlm: Attribute conditioned sft as an (user-steerable) alternative to rlhf. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11275–11288, 2023.

[8] Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, Linxi Fan, and Yuke Zhu. HOVER: Versatile neural whole-body controller for humanoid robots. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9989–9996. IEEE, 2025.

[9] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

[10] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. *arXiv preprint arXiv:2205.09991*, 2022.

[11] Bo Li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4282–4291, 2019.

[12] Yitang Li, Zhengyi Luo, Tonghe Zhang, Cunxi Dai, Anssi Kanervisto, Andrea Tirinzoni, Haoyang Weng, Kris Kitani, Mateusz Guzek, Ahmed Touati, Alessandro Lazaric, Matteo Pirotta, and Guanya Shi. BFM-Zero: A promptable behavioral foundation model for humanoid control using unsupervised reinforcement learning. *arXiv preprint arXiv:2511.04131*, 2025.

[13] Qiayuan Liao, Takara E Truong, Xiaoyu Huang, Yuman Gao, Guy Tevet, Koushil Sreenath, and C Karen Liu. Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. *arXiv preprint arXiv:2508.08241*, 2025.

[14] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European conference on computer vision*, pages 38–55. Springer, 2024.

[15] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9997–10003. IEEE, 2025.

[16] Jing Yuan Luo, Yunlong Song, Victor Klemm, Fan Shi, Davide Scaramuzza, and Marco Hutter. Residual policy learning for perceptive quadruped control using differentiable simulation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2025.

[17] Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Sirui Chen, Fernando Castaneda, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, et al. Sonic: Supersizing motion tracking for natural humanoid whole-body control. *arXiv preprint arXiv:2511.07820*, 2025.

[18] David McAllister, Songwei Ge, Brent Yi, Chung Min Kim, Ethan Weber, Hongsuk Choi, Haiwen Feng, and Angjoo Kanazawa. Flow matching policy gradients. *arXiv preprint arXiv:2507.21053*, 2025.

[19] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[20] Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, and Jingbo Wang. TokenHSI: Unified synthesis of physical human-scene interactions through task tokenization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5379–5391, 2025.

[21] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018.

[22] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. AMP: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (TOG)*, 40(4):1–20, 2021.

[23] Matteo Pirotta, Andrea Tirinzoni, Ahmed Touati, Alessandro Lazaric, and Yann Ollivier. Fast imitation via behavior foundation models. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

[24] Zekun Qi, Xuchuan Chen, Dairu Liu, Chenghuai Lin, Yunrui Lian, Sikai Liang, Zhikai Zhang, Yu Guan, Jilong Wang, Wenyao Zhang, et al. Humanoid-gpt: Scaling data and structure for zero-shot motion tracking. *arXiv preprint arXiv:2606.03985*, 2026.

[25] Allen Ren, Justin Lidard, Lars Ankile, Anthony Simeonov, Pulkit Agrawal, Anirudha Majumdar, Benjamin Burchfiel, Hongkai Dai, and Max Simchowitz. Diffusion policy policy optimization. In *International Conference on Learning Representations*, volume 2025, pages 77288–77329, 2025.

[26] Luis Sentis and Oussama Khatib. A whole-body control framework for humanoids operating in human environments. In *2006 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2641–2648. IEEE, 2006.

[27] Yiyang Shao, Xiaoyu Huang, Bike Zhang, Qiayuan Liao, Yuman Gao, Yufeng Chi, Zhongyu Li, Sophia Shao, and Koushil Sreenath. LangWBC: Language-directed humanoid whole-body control via end-to-end learning. In *Robotics: Science and Systems (RSS)*, 2025.

[28] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[29] Wandong Sun, Baoshi Cao, Long Chen, Yongbo Su, Yang Liu, Zongwu Xie, and Hong Liu. Learning perceptive humanoid locomotion over challenging terrain. *arXiv preprint arXiv:2503.00692*, 2025.

[30] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020.

[31] Octo Model Team, Dibya Ghosh, Homer Walke, Karl

Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[32] Chen Tessler, Yunrong Guo, Ofir Nabati, Gal Chechik, and Xue Bin Peng. MaskedMimic: Unified physics-based character control through masked motion inpainting. *ACM Transactions on Graphics (TOG)*, 43(6), 2024.

[33] Guy Tevet, Sigal Raab, Setareh Cohan, Daniele Reda, Zhengyi Luo, Xue Bin Peng, Amit H. Bermano, and Michiel van de Panne. CLoSD: Closing the loop between simulation and diffusion for multi-task character control. In *The Thirteenth International Conference on Learning Representations (ICLR)*, pages 69704–69718, 2025.

[34] Andrea Tirinzoni, Ahmed Touati, Jesse Farebrother, Mateusz Guzek, Anssi Kanervisto, Yingchen Xu, Alessandro Lazaric, and Matteo Pirotta. Zero-shot whole-body humanoid control via behavioral foundation models. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[35] Ahmed Touati and Yann Ollivier. Learning one representation to optimize all rewards. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:13–23, 2021.

[36] Yihao Wang, Pengxiang Ding, Lingxiao Li, Can Cui, Zirui Ge, Xinyang Tong, Wenxuan Song, Han Zhao, Wei Zhao, Pengxu Hou, et al. Vla-adapter: An effective paradigm for tiny-scale vision-language-action model. In *Proceedings of the AAAI conference on artificial intelligence*, volume 40, pages 18638–18646, 2026.

[37] Yunshen Wang, Shaohang Zhu, Peiyuan Zhi, Yuhan Li, Jiaxin Li, Yong-Lu Li, Yuchen Xiao, Xingxing Wang, Baoxiong Jia, and Siyuan Huang. Omnixtreme: Breaking the generality barrier in high-dynamic humanoid control. *arXiv preprint arXiv:2602.23843*, 2026. URL https://arxiv.org/abs/2602.23843.

[38] Zhen Wu, Xiaoyu Huang, Lujie Yang, Yuanhang Zhang, Xi Chen, Pieter Abbeel, Rocky Duan, Angjoo Kanazawa, Carmelo Sferrazza, Guanya Shi, et al. Perceptive humanoid parkour: Chaining dynamic human skills via motion matching. *arXiv preprint arXiv:2602.15827*, 2026.

[39] Zehao Xiao and Cees GM Snoek. Beyond model adaptation at test time: A survey. *arXiv preprint arXiv:2411.03687*, 2024.

[40] Haoru Xue, Xiaoyu Huang, Dantong Niu, Qiayuan Liao, Thomas Kragerud, Jan Tommy Gravdahl, Xue Bin Peng, Guanya Shi, Trevor Darrell, Koushil Sreenath, and Shankar Sastry. LeVERB: Humanoid whole-body control with latent vision-language instruction. *arXiv preprint arXiv:2506.13751*, 2025.

[41] Yufei Xue, Wentao Dong, Minghuan Liu, Weinan Zhang, and Jiangmiao Pang. A unified and general humanoid whole-body controller for fine-grained locomotion. In *Robotics: Science and Systems (RSS)*, 2025.

[42] Kangning Yin, Weishuai Zeng, Ke Fan, Minyue Dai, Zirui Wang, Qiang Zhang, Zheng Tian, Jingbo Wang, Jiangmiao Pang, and Weinan Zhang. Unitracker: Learning universal whole-body motion tracker for humanoid robots. *IEEE Robotics and Automation Letters*, 2026.

[43] Shaofeng Yin, Yanjie Ze, Hong-Xing Yu, C. Karen Liu, and Jiajun Wu. VisualMimic: Visual humanoid loco-manipulation via motion tracking and generation. *arXiv preprint arXiv:2509.20322*, 2025.

[44] Mingqi Yuan, Tao Yu, Wenqi Ge, Xiuyong Yao, Dapeng Li, Huijiang Wang, Jiayu Chen, Bo Li, Wei Zhang, Wenjun Zeng, et al. A survey of behavior foundation model: Next-generation whole-body control system of humanoid robots. *IEEE transactions on pattern analysis and machine intelligence*, 2025.

[45] Mingqi Yuan, Tao Yu, Wenqi Ge, Xiuyong Yao, Dapeng Li, Huijiang Wang, Jiayu Chen, Bo Li, Wei Zhang, Wenjun Zeng, Hua Chen, and Xin Jin. A survey of behavior foundation model: Next-generation whole-body control system of humanoid robots. *arXiv preprint arXiv:2506.20487*, 2026.

[46] Yanjie Ze, Siheng Zhao, Weizhuo Wang, Angjoo Kanazawa, Rocky Duan, Pieter Abbeel, Guanya Shi, Jiajun Wu, and C Karen Liu. Twist2: Scalable, portable, and holistic humanoid data collection system. *arXiv preprint arXiv:2511.02832*, 2025.

[47] Zhikai Zhang, Jun Guo, Chao Chen, Jilong Wang, Chenghuai Lin, Yunrui Lian, Han Xue, Zhenrong Wang, Maoqi Liu, Jiangran Lyu, et al. Track any motions under any disturbances. *arXiv preprint arXiv:2509.13833*, 2025.

[48] Siheng Zhao, Yanjie Ze, Yue Wang, C Karen Liu, Pieter Abbeel, Guanya Shi, and Rocky Duan. Resmimic: From general motion tracking to humanoid whole-body loco-manipulation via residual learning. *arXiv preprint arXiv:2510.05070*, 2025.

[49] Siheng Zhao, Yuanhang Zhang, Ziqi Lu, Pieter Abbeel, Rocky Duan, Koushil Sreenath, Yue Wang, C Karen Liu, and Guanya Shi. Ladderman: Learning humanoid perceptive ladder climbing. *arXiv preprint arXiv:2606.05873*, 2026.

[50] Shaoting Zhu, Baijun Ye, Jiaxuan Wang, Jiakang Chen, Ziwen Zhuang, Linzhan Mou, Runhan Huang, and Hang Zhao. Ttt-parkour: Rapid test-time training for perceptive robot parkour. *arXiv preprint arXiv:2602.02331*, 2026.

[51] Shaoting Zhu, Ziwen Zhuang, Mengjie Zhao, Kun-Ying Lee, and Hang Zhao. Hiking in the wild: A scalable perceptive parkour framework for humanoids. *arXiv preprint arXiv:2601.07718*, 2026.

[52] Ziwen Zhuang, Shaoting Zhu, Mengjie Zhao, and Hang Zhao. Deep whole-body parkour, 2026. URL https://arxiv.org/abs/2601.07701.