MOVE, DON'T THINK: SAMPLE-EFFICIENT COLLISION AVOIDANCE WITH LOSSY PROXIMITY SENSORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Robots must learn to avoid collisions with as few demonstrations as possible to overcome the challenge of collecting large datasets of potentially destructive collision examples. By using egocentric proximity sensor observations as inductive biases, we train a humanoid robot to learn collision avoidance in fewer steps using lossy, low-dimensional sensors with wide overlapping coverage rather than with high-resolution, spatially accurate sensing of the robot's surrounding space. We also find that, perhaps unintuitively, anticipating when and where contact happens on the body of a robot harms its learning sample efficiency. We investigate eight candidate representations for sensor observations as an inductive bias for learning collision avoidance using reinforcement learning, which include latent compression, future contact anticipation, lossy corruption, and privileged state estimation. The candidate representations were tested by training over 200 humanoid robots to play a simplified version of dodgeball end-to-end under an intentionally constrained sample and compute budget.

1 INTRODUCTION

Robots deployed in unstructured environments must avoid collisions that could damage themselves or others. Data-driven models have enabled manipulation and locomotion tasks to be performed in unstructured environments (Kim et al., 2024b; Physical Intelligence et al., 2025), but learning collision avoidance through data is challenging because collision data can be damaging and costly to collect. This expensive data constraint motivates the need for sample-efficient ways to learn safe, reflexive behavior in as few demonstrations as possible.

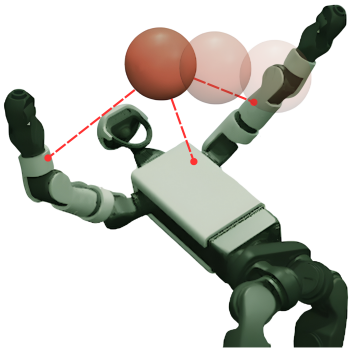

Figure 1: A humanoid robot dodges an incoming projectile using wide coverage, low resolution proximity sensors.

Forecasting future camera observations using generative world models emerged as a strategy for safely experiencing potentially dangerous scenarios (Agarwal et al., 2025; Assran et al., 2025), and forecasting has also been shown to dramatically improve task performance in dynamic tasks such as sports (Syed et al., 2026). However, occlusions and partial observability are a notorious problem for camera-based contact detection (Navarro et al., 2021), leading to contact anticipation with world models that rely on heuristic estimates and a growing demand for occlusion-free contact prediction (Liu et al., 2025).

Egocentric proximity sensors have concurrently existed as solutions to occlusion and partial observability due to their ability to sense directly from the surface of a robot (Scholz et al., 2024) and have been extensively used for safe collision avoidance (Navarro et al., 2021). A large variety of these whole-body proximity sensor constellations have emerged as a result of the many methods for estimating local distance through light ranging, capacitive sensing, acoustic sensing, and more (Scholz et al., 2024). Still, there is not a consensus on what properties of proximity sensors are ideal for sample-efficient learning.

The goal of this paper is two-fold: 1) to investigate how whole-body proximity sensor observations can be best represented as inductive biases for sample-efficient collision avoidance, and 2) to inves-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tigate if predicting future contact observations can lead to improved sample efficiency for learning collision avoidance. The contributions of this work are as follows:

1. We propose a design heuristic for using proximity sensors as sample-efficient inductive biases, which favors wide, overlapping coverage of low-resolution sensors and low signal dimensionality. We further show how a high-resolution proximity sensor can mimic a low-dimensional sensor using a simple lossy signal minimization to improve sample efficiency.
2. We propose that anticipating when and where contact will happen can create sample-efficiency deficits for learning collision avoidance, which we validate using a proximity world model (PWM) that forecasts future observations from whole-body proximity sensors to anticipate contact.
3. We rigorously ablate sensor observation representation, including signal processing, forecasting, and hardware properties, in a simulated collision avoidance task with a Unitree humanoid H1-2 robot equipped with 64 proximity sensors.

2 Proximity Signal Representation

We investigate how sensor observations can be represented to provide task-relevant information that a robot learning reflexive collision-avoidance behavior can readily exploit.

2.1 Related Works

Collision Avoidance: A reactive collision-avoidance policy $\pi_\theta(a_t \mid o_t)$ avoids contact that could damage the robot, where o_t is an observation and $x_t \subseteq o_t$ is an observation made by onboard sensors that can detect an incoming collision.

Zhang et al. (2025); Armleder et al. (2025); Escobedo et al. (2021) used quadratic programming (QP) with predefined constraints on o_t to achieve reactive control. Still, such algorithms rely on careful tuning of their dynamics and object estimations to maintain safety. Instead of manually tuning a controller, Yang et al. (2025); Kim et al. (2024a); Wang et al. (2024) trained a policy with learned parameters $\pi_\theta(a_t \mid o_t)$ trained on expert demonstrations. However, collecting real collision demonstrations can damage the robot or its environment, so policies trained on expert demonstrations typically also rely on privileged knowledge of object locations in the local environment, which limits the data domain. To avoid collecting expert demonstrations, Yang et al. (2026); Xue et al. (2026) use RL to train a reactive policy directly on observations guided by rewards, but their approach uses a simulation environment that needs to accurately emulate the real world for deployment and can require an intractable number of trial-and-error training steps to be successful end-to-end on real data. Therefore, we believe that the goal is to reduce the number of training steps needed for effective collision avoidance to support a future where robots adapt to collisions end-to-end, in real time, and without relying on manual tuning, expert demonstrations, or privileged simulators.

Egocentric Proximity Sensors: Egocentric proximity sensors, also sometimes referred to as artificial skins, have been used for collision avoidance including time-of-flight sensors that pulse light or sound waves, then measure distance by how long it takes the waves to return (Borelli et al., 2026; Kim et al., 2024a; Escobedo et al., 2021); capacitive sensors that measure distance by how long it takes to charge a conductive material based on nearby objects (Kohlbrenner et al., 2026; Liu et al., 2023; Moon et al., 2021); or stereo cameras that measure depth from pixel disparity (Hu et al., 2025). However, despite the wide variety of applicable proximity sensors, no metric exists to quantify how much of a proximity sensor's signal is task-relevant for collision avoidance.

2.2 Generic Proximity Sensor

Consider a proximity sensor that activates when any point on a nearby object intersects with the proximity sensor's coverage area. A proximity sensor's coverage normally depends on its underlying mechanisms, but we assume our modality-agnostic proximity sensor is limited only by range and can have either full spherical coverage around the origin or a rectangular window within that sphere, defined in this work by a horizontal and vertical field-of-view (FoV). Additionally, the proximity sensor can have varying resolution, defined by the density of smaller sensing regions within its

Rep. u_t	**Description**	**Dist.**	**Dir.**	**Dim.**	**Prec.**
x_t	Raw sensor observation.	–	–	$W \times H \times S$	$\mathbb{R}^{\text{dim}}$
$f_{\text{tp}}(x_t)$	Localizes the nearest object within sensing coverage relative to the base of the robot.	$\uparrow$	$\uparrow$	$\downarrow$ (3)	–
$f_{\text{min}}(x_t)$	Returns only the closest range measurement per sensor.	–	$\downarrow$	$\downarrow$ (S)	–
$f_{\text{bin}}(x_t)$	Returns only binary activations per signal.	$\downarrow$	–	–	$\downarrow$ $(\{0,1\}^d)$
$\mu_\psi(x_t)$	Returns a latent observation.	$\approx$	$\approx$	$\downarrow$	–
$\text{PWM}(x_t)$	Returns $[z_t, \hat{z}_t, c_{T_c}]$.	–	–	$\downarrow$	–
$\text{PWM}_{\text{min}}(x_t)$	Returns $[f_{\text{min}}(x_t), f_{\text{min}}(\hat{x}_{T_c}), c_{T_c}]$.	–	$\downarrow$	$\downarrow (2S+1)$	–
$\text{PWM}_0(x_t)$	Returns $[f_{\text{tp}}(x_t), f_{\text{tp}}(\hat{x}_{T_c}), c_{T_c}]$.	$\uparrow$	$\uparrow$	$\downarrow$ (6)	–

Table 1: Investigated representations of sensor observation x that modify the robot's access to location-revealing information. We assume the encoder representation f_ψ performs compression and is well trained, such that $H' \times W' \times S' \times e < W \times H \times S$, while retaining relevant information about nearby object distance and direction.

overall coverage, each measuring an individual signal when an object crosses its boundaries. We will denote W as the number of sensing regions along the horizontal FoV and H as the number of sensing regions along the vertical FoV.

Let a proximity sensor suite comprise S sensors. The observation x made by the sensors at time t is thus $x_t \in \mathbb{R}^{H \times W \times S}$. We assume all sensors are ideal, meaning they have zero phase offset, latency, and noise, because these are factors an engineer would choose to eliminate and are not design choices for sample efficiency. We will use this generic proximity sensor notation for the remainder of this work.

2.3 Sample-Constrained Usable Information

$\mathcal{V}$-information gives insight into the collision avoidance problem (Xu et al., 2020). Our motivating question is: *under a capacity-constrained learner and a limited training budget, which representation of proximity sensor data makes collision-relevant information easiest to exploit?*

Let X denote the random variable corresponding to a raw proximity observation in the sample space $\mathcal{X}$, let Y denote a collision-relevant target variable, and let $\mathcal{V}$ be a restricted family of predictors of Y used by the robot. $\mathcal{V}$ is a restricted family because the agent must predict Y in finite time (i.e., the robot cannot take its time to decide how to dodge an impending collision). The predictive $\mathcal{V}$-information $I_\mathcal{V}$ from X to Y is

$$I_\mathcal{V}(X \to Y) = H_\mathcal{V}(Y \mid \varnothing) - H_\mathcal{V}(Y \mid X), \tag{1}$$

where $H_\mathcal{V}$ is the $\mathcal{V}$-entropy,

$$H_\mathcal{V}(Y \mid X) = \inf_{q \in \mathcal{V}} \mathbb{E}(X,Y) \left[-\log q(Y \mid X) \right], \tag{2}$$

and $H_\mathcal{V}(Y \mid \varnothing)$ is the corresponding predictive entropy when the predictor is given no observation of X. Applying a deterministic transformation $f : \mathcal{X} \to \mathcal{X}$ on X would not increase the Shannon mutual information with Y, but under a restricted compute budget, a representation $f(X)$ that makes it easier for the robot to exploit Y has a higher $\mathcal{V}$-information with Y (Xu et al., 2020). We therefore seek representations that satisfy

$$I_\mathcal{V}(f(X) \to Y) > I_\mathcal{V}(X \to Y) \tag{3}$$

for the learner class and collision avoidance task. We treat Eq. (3) as an interpretive lens for the hypothesis rather than an assumed property, and evaluate its implications empirically through downstream collision-avoidance learning.

This predictive $\mathcal{V}$-information perspective motivates the representations laid out in Table 1. The sensor-level transformations $f_{\min}$ and f_{bin} deliberately discard aspects of x_t, f_ψ transforms x_t into a lower-dimensional latent representation intended to preserve collision-relevant structure, and the PWM forecasts future contact and future observations from x_t. We also consider an oracle that reveals privileged information about the incoming object's state as an upper-bound representation. Together, these representations provide insight into how best to transform raw sensory data for sample efficiency and what information is most relevant to preserve or decipher during collision avoidance.

For a given representation U and corresponding task success curve $S_U(n)$, we use steps to a given success rate, $N_\alpha(U) = \inf n \in [0, N] : S_U(n) \geq \alpha$, and normalized area under the curve (NAUC),

$$\text{NAUC}(U) = \frac{1}{N} \int_0^N S_U(n) dn, \tag{4}$$

as metrics to evaluate which representations are more readily exploitable by $\mathcal{V}$.

3 Contact Anticipation

Let $c_t \in \{0, 1\}$ denote the contact state at time t. The discrete objective of contact anticipation is to predict the conditional probability distribution of future contact at time k based on previous observations $o_{\leq t}$.

$$p(c_{t+1:t+k} \mid o_{\leq t}). \tag{5}$$

Early approaches estimated the velocity of nearby objects using camera and tactile sensors to predict time-to-contact (Roncone et al., 2016), but this approach assumes dynamics of the nearby object. He et al. (2025); Liu et al. (2025) proposed instead using a learned contact prediction model p_θ conditioned on camera images as prior observations to predict future contact states. Zhou & Negrut (2025); Team et al. (2026); Qin et al. (2026) further conditioned p_θ on tactile sensor observations and predicted future actions $\hat{a}_{t+1:t+k}$

$$\hat{c}_{t+1:t+k} \sim p_\theta(c_{t+1:t+k} \mid o_{\leq t}, \hat{a}_{t+1:t+k}), \tag{6}$$

However, their approach was validated in a manipulation setting, and anticipating future contact states at a whole-body scale for collision avoidance using proximity sensors has not been explored.

3.1 Proximity World Model

We propose a two-step proximity world model (PWM) that first generates future latent sensor observations, then uses those observations to predict contact. Following previous generative world model work (Agarwal et al., 2025), our world model comprises an encoder, dynamics model, and decoder to generate future observations, as shown in Fig. 7. We introduce an additional contact prediction module to anticipate which predicted future observations will lead to contact.

Our predictive world model starts by mapping x_t into a smaller dimensional latent vector $z_t \in \mathbb{R}^{H' \times W' \times S' \times e}$ through the encoder

$$\tilde{z}_t \sim q_\psi(z_t \mid x_t) \tag{7}$$

to improve sample efficiency during training, where q_ψ is a learned posterior distribution and e is an embedding dimension. At inference time, the encoder deterministically returns the latent observation as $z_t := \mu_\psi(x_t)$. The decoder model conversely samples a likely sensor observation conditioned on a given latent observation.

Next, the dynamics model r_ϕ samples a possible sequence of future latent sensor observations induced by the robot's latest action

$$\hat{z}_{t+1:t+k} \sim r_\phi(z_{t+1:t+k} \mid z_{\leq t}, a). \tag{8}$$

Accurate prediction of these future observations requires the world model to capture the relative dynamics between the robot and its environment. We then modify Eq. (6) to get the conditional probability distribution of contact conditioned on the predicted future observations and sample from a learned contact anticipation model p_η to get a candidate set of future contact states:

$$\hat{c}_{t+1:t+k} \sim p_\eta(c_{t+1:t+k} \mid \hat{z}_{t+1:t+k}, z_{\leq t}, a). \tag{9}$$

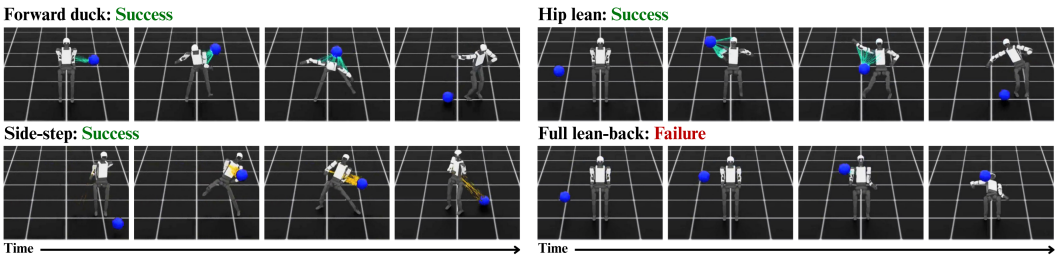

Figure 2: The robot is rewarded for learning collision-avoidant behavior and maintaining posture.

Thus, sampling a future trajectory from the world model and subsequently predicting its associated contact states provides a sample-based approximation to contact anticipation under a candidate future action sequence. The PWM used in this work was pretrained on a sample dataset of 6000 trajectories collected via expert demonstration; see Section A.3.1 for implementation details. For the same reason that sensors are represented without noise or latency, the large training data is intended to make the PWM well adjusted as an *idealized* contact predictor, that, while not perfect, serves as a higher-bound indicator of the influence of contact anticipation for sample-efficiency than a highly noisy PWM. It is assumed that contact prediction inaccuracy is ideally minimized in deployment.

4 Experiments

We investigate signal representations in Table 1 empirically in a simulated collision avoidance task using a Unitree H1-2 humanoid robot equipped with 64 proximity sensors that were randomly distributed using GenTact (Kohlbrenner et al., 2025) to supply ample coverage of the surrounding space of the robot, as shown in Fig. 3. The humanoid robot is trained using proximal policy optimization (PPO) (Schulman et al., 2017) and is tasked with avoiding collisions with independently thrown balls launched at random times 1-2 seconds apart and set on a collision course with the robot. Fig. 2 shows a few successful and unsuccessful collision-avoidance policies for this task. More implementation details on the simulated environment are found in Section A.1

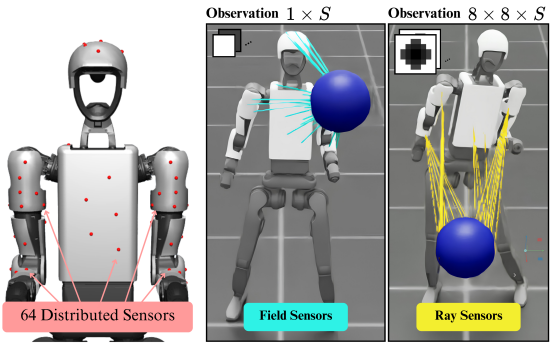

Figure 3: Example of distributed field-sensor activations (left) and ray-sensor activations (right). Lines connect to detected objects.

Termination: An episode can end under four conditions: 1) The ball collides with the robot, 2) the robot's root point (at the pelvis) falls below a 40 cm threshold from the ground, 3) the robot's root point pose exceeds an $80°$ angle from upright, and 4) the robot successfully survives for three seconds.

Rewards: The robot is rewarded for survival time to learn to avoid collisions and root point height to learn to stand, but is penalized for drifting from the starting point and expending energy in its movements. While the task reward drives the robot to discover agile dodging maneuvers to survive, the regularization terms are critical for suppressing high-frequency jitter and ensuring the learned behaviors are fluid, energy-efficient, and natural. We explicitly do not provide any reward on sensor observations to prevent a specific proximity sensor configuration from being unintentionally favored.

Observations: In addition to the sensor observation, the observation space is comprised of base angular velocity ω_b, relative joint positions q_r, joint velocities $\dot{q}$, gravity vector projected in the base

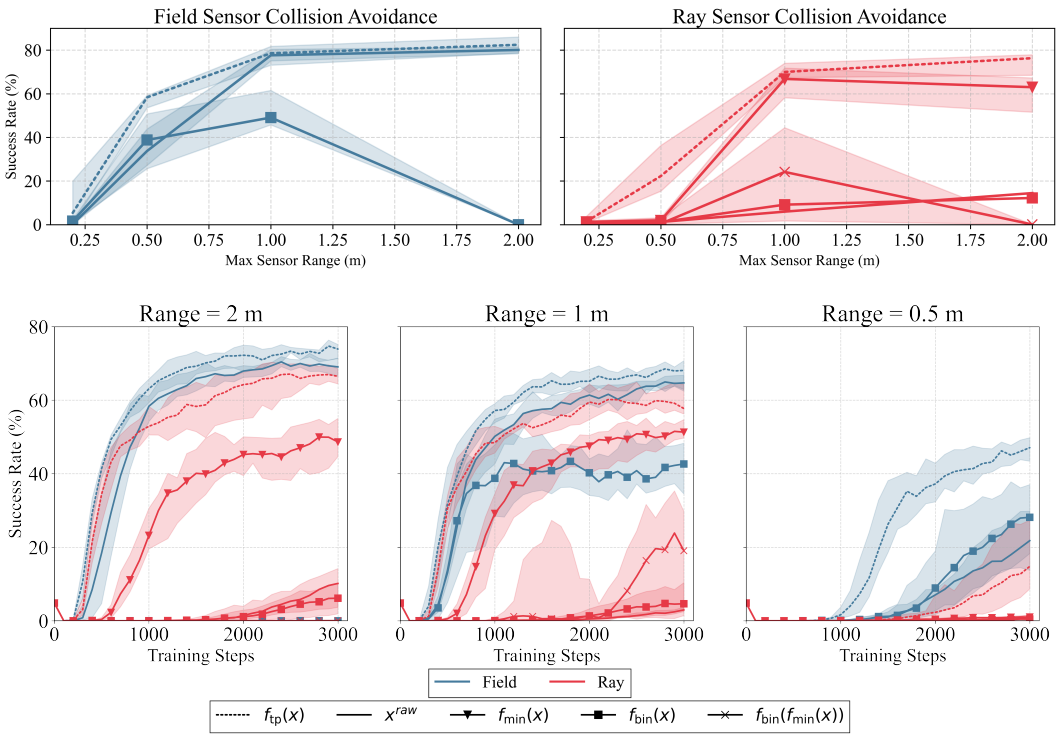

Figure 4: Sensor representation ablation results. Top: total task success for field and ray sensors across ranges of 0.2, 0.5, 1, and 2 meters. Bottom: task success over training steps for the field and ray sensor representations, where achieving higher success in fewer training steps indicates greater sample efficiency.

frame g_b, and the previous action for one-step action memory a_{t-1}, as

$$o_t = [x, \omega_b, g_b, q_r, \dot{q}, a_{t-1}]. \tag{10}$$

Additionally, we adopt an asymmetric actor-critic structure (Pinto et al., 2017) to improve training stability, where the critic receives a privileged linear base velocity observation v_b. The actor and critic are each two-layer feedforward networks with 32 units per layer. We chose this fixed, low-capacity architecture so differences in task performance can be attributed to the proximity sensor observation rather than model expressivity (Bengio et al., 2013).

Steps: We report sample efficiency in PPO learning iterations. Each iteration collects 16 steps from 4,096 parallel environments (65,536 environment interactions, reused for 5 epochs in mini-batches of 16,384) at a control rate of 60 Hz. All training runs share this budget, so an $m\times$ reduction in iterations is an $m\times$ reduction in environment interactions. Training runs for 3,000 iterations (1.966×10^8 interactions).

4.1 Sensor Representation Ablation

Inspired by the current dichotomy of the proximity sensing field, we define two extremes of the generic proximity sensor to use as test subjects: **field sensors** (low-resolution), which we define as a single sensing region ($H, W = 1$) with wide coverage and no explicit indication of where an object lies within the field of view, and **ray sensors** (high resolution), which use an 8×8 sensing grid over a narrow $45°$ field of view, where each signal in the grid is associated with a unique heading direction. We perform 10 seeded end-to-end training runs of the dodgeball task with a maximum compute budget of 3000 steps per run. Fig. 4a reports the median and interquartile range (IQR) of task success across seeds for these sensors and the signal representations in Table 1, and Fig. 4b reports the sample efficiency of each representation in task success over training steps.

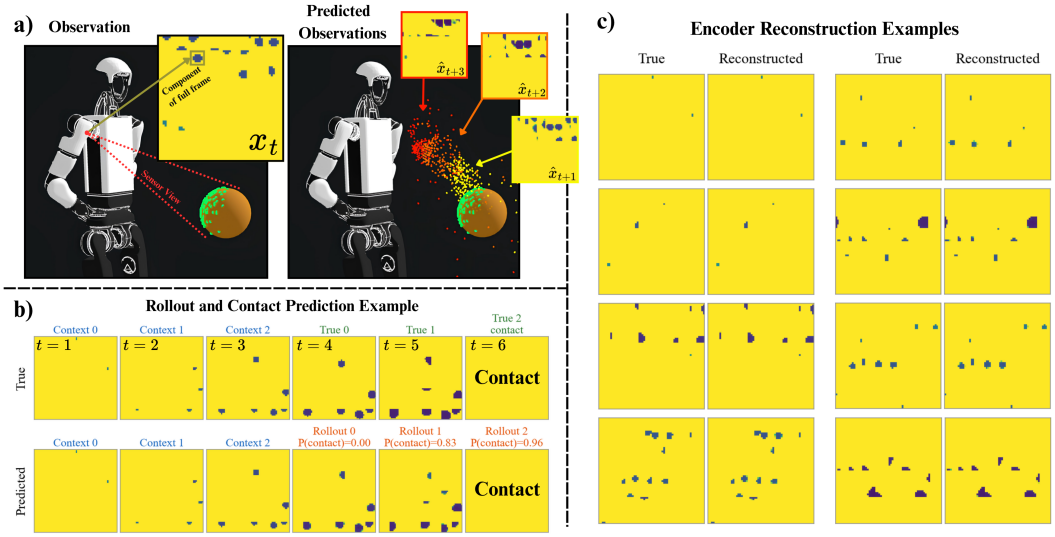

Figure 5: a) The full sensor observations are shown here as side-by-side frames of individual sensor depth images, where each sensor occupies an 8x8 grid of pixels in the observation frame. The green point cloud hits are points visible on the ball and correspond to the observation image through forward kinematics. When we roll out future observations, we can also visualize them in world space as a point cloud. Here we see the next three predicted observations $\hat{x}_{t+1:t+3}$. b) shows how a roll-out will end when a contact is predicted, and c) demonstrates the accuracy of the encoder and decoder from random samples of a validation set.

Perfect Localization Oracle f_{tp}: We found that robots that could observe the true relative position of the incoming ball f_{tp} served as an upper-bound representation for both the field sensor and the ray sensor, achieving a NAUC of 0.584 at 2 m range. f_{tp} is believed to be the strongest representation because it compactly encodes both distance and direction in a three-dimensional signal.

Raw Distance Observation x_t: Robots with raw *field* sensor observations achieved a NAUC of 0.522 at 2 m range, which is 89.3% as strong as those with $f_{tp}(x_t)$ across all sensing ranges, requiring no modification to their representation for their performance. Although each field sensor provides no explicit directional information, we believe its distributed activations allowed the policy to infer object direction from the sensor constellation itself. In contrast, policies trained on raw ray sensor observations only reached a maximum NAUC of 0.02.

Minimum Distance Observation f_{min}: Applying the minimum-distance transformation $f_{\min}$ substantially improved both final performance and sample efficiency of the robots with ray sensors, despite discarding the within-sensor direction of the detected object. At $1\,\mathrm{m}$, $f_{\min}$ outperformed raw ray sensor observation NAUC by 5400%. By retaining only the nearest range measurement from each sensor, $f_{\min}$ effectively converts each ray sensor into a field-like observation within its sensing coverage. Additionally, $f_{\min}$ provides insight into how coverage itself influenced the robot's ability to learn. Robots with raw field sensors achieved an average NAUC 0.138 points higher than those with $f_{\min}(x^{ray})$ observations, despite $f_{\min}(x^{ray})$ observations occupying the same range and approximate signal properties to raw field sensors. Similarly, robots whom observed $f_{\text{tp}}(x_t^{\text{field}})$ performed an average NAUC of 0.12 higher than those with $f_{\text{tp}}(x_t^{\text{ray}})$. We attribute these marginal gains to the increased coverage of the field sensors. These two examples are the only representations that will be compared between sensor types. Field sensors were not evaluated with $f_{\min}$ due to only containing a single signal by default.

Binary Detection Observation f_{bin}: Binary detection performance was not linearly consistent across ranges. We observed that at the $2\,\mathrm{m}$ sensing range, the sensors activated well before contact and remained active as the ball approached, causing the representation to saturate. However, the field binary observation outperformed direct field proximity measurements at the $0.5\,\mathrm{m}$ sensing range, suggesting that distance becomes less important in the moments before contact.

Sensor Observation	Final Success Rate (%)	10% Success		25% Success		NAUC
		Steps	Speedup	Steps	Speedup	
$f_{\text{tp}}(x_t)$	66.5	338	8.8× faster	427	> 7.0× faster	0.508
$\text{PWM}_0(x_t)$	57.5	625	4.7× faster	770	> 3.9× faster	0.39
$\text{PWM}_{\text{min}}(x_t)$	49.2	1738	1.7× faster	2024	> 1.5× faster	0.158
$f_{\text{min}}(x_t)$	48.6	770	3.9× faster	1026	> 2.9× faster	0.299
$f_\psi(x_t)$	37.9	1613	1.8× faster	2082	> 1.4× faster	0.140
$\text{PWM}(x_t)$	29.9	1725	1.7× faster	2663	> 1.1× faster	0.099
x_t	10.2	2970	baseline	> 3000	baseline	0.020

Table 2: Median performance across sensor observation representations. Speedup represents how many less steps were needed to achieve success thresholds of 10% and 25% relative to raw ray sensor observations as the baseline.

4.2 Contact Prediction Ablation

We investigate using the PWM to improve the ray sensor's sample efficiency within the 3000-step allotted compute budget at the two-meter range, due to the large gap between $f_{\text{tp}}(x^{\text{ray}})$ and x^{ray}. Fig. 5 shows examples of how the current sensor observation is constructed using the ray sensors, as well as how future observations are propagated through the world model during a roll out.

We simulated the same dodge-ball task, this time with five seeded training runs, where the robot either received only the latent encoding of the current observation, $\mu_\psi(x_t)$, the latent encoding of the current and future observation k in addition to an anticipated contact label c_{t+k}, $\text{PWM}(x_t)$, the minimum distance of the current and future observation, $\text{PWM}_{\text{min}}(x_t)$, and finally an oracle observation that can see the true current and future ball position $\text{PWM}_0(x_t)$. The task success during training time of this study is shown in green in Fig. 6 and compared against the previous ray sensor representations from Section 4.1.

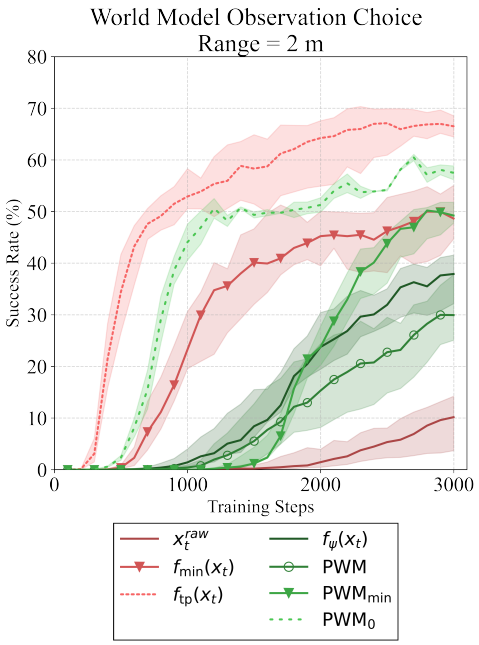

Figure 6: Robots that could anticipate contact (green) consistently underperformed variants that used only current observations (red).

Latent Observation $\mu_\psi(x)$: Learned compression with μ_ψ improved the sample efficiency relative to raw ray observations, but did not perform as well as f_{min}. This can be understood by the dimensionality of $f_\psi(x)$ being $d^x > d^{f_\psi(x)} > d^{f_{\text{min}}(x)}$, where the encoder was pretrained to output embeddings in the dimension of $H', W' = 2$, $S' = 16$, and $e = 8$, resulting in an embedding that is one-eighth the size of the raw observation.

Future Insight Oracle PWM_0: The oracle robot observed the true relative position of the ball in its current state and where it would be upon collision, but performed worse than the robot that only saw the current true position. Additionally, the oracle was unable to achieve any task success without additional information regarding where the ball came from. This result suggests that providing information about the future collision state may not carry strong, readily exploitable information for collision avoidance in this setting.

Current and Future Latent Observation PWM and PWM_{min}: Similarly to the oracle, providing predictive information about the future collision state harmed the agent's ability to learn collision avoidance quickly. Both the agent with PWM and PWM_{min} reached similar levels of task success as their non-predictive counterparts, but the additional predictive information reduced their ability to exploit Y in as many steps.

5 DISCUSSION

Our ablations point to two consistent phenomena: **signal dimensionality** was the biggest predictor of sample efficiency, and **anticipating contact distracted** the robot from exploiting task-relevant information in the demonstrated close-by, single impending collision task. Despite the the PWM being trained on thousands of expert demonstrations and having privileged pretraining data, it still did not beat a simple per-sensor minimum distance transformation.

Our findings are further supported by the neurosciences, where fast multisensory neurons encode a peripersonal space (PPS): a compact representation of body-centric receptive fields that activate based on the proximity of nearby objects (Bufacchi & Iannetti, 2018). Our field-sensor representation has a similar computational structure: individual sensors provide little spatial detail, yet their overlapping, body-distributed activations encode sufficient direction and distance for avoidance. One interpretation of our results is therefore that sample-efficient collision avoidance benefits less from reconstructing the external world than from constructing a compact, action-oriented representation of peripersonal threat. In our experiment, range was a fixed value throughout training, but it is well known that the PPS adapts in size based on the speed of nearby objects (Noel et al., 2018). Future work should therefore explore a dynamically adapting *range* for improving sample efficiency.

6 CONCLUSION

We investigated eight candidate representations of proximity sensor signals for sample-efficient collision avoidance, including lossy transformations, latent compression, and world-model forecasting. We find in our experiments that wide coverage of low-resolution proximity sensors consistently resulted in higher sample efficiency against high-resolution alternatives, which are consistent with $\mathcal{V}$-information theory as having higher $\mathcal{V}$-predictive relevant information for reflexive contact avoidance under sample and compute constraints. Additionally, providing robots with additional insight into when and where a collision would happen in the future resulted in consistently lower sample efficiency and did not overcome the cost of extra signal dimensionality. Therefore, our findings that for learning sample-efficient collision avoidance with egocentric proximity sensors: 1) wide-overlap, low-resolution sensing that preserves relative distance information, and 2) conditioning learned policies only on current observations may be desirable for computationally constrained reflexive motion.

AI USE STATEMENT

In this work, we used generative AI tools for assistance and feedback setting up our synthetic dataset collection, pipeline implementation between the PWM and training environment, and mathematical verification on the synthesis of our results. We have not used generative AI tools for data analysis, hypothesis creation, or interpretations of results, and assistance with mathematical proofs and translation are not applicable to this work. Additionally, we used generative AI tools as a method of discovering relevant sources, figure editing, grammar and spelling refinement, and software editing. We have reviewed all AI-assisted work by inspecting our datasets for accuracy via custom tests. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ACKNOWLEDGMENTS

Non-archival preliminary results of this work were presented at [redacted until camera-ready].

REFERENCES

Niket Agarwal, Arslan Ali, Maciej Bala, Yogesh Balaji, Erik Barker, Tiffany Cai, Prithvijit Chattopadhyay, Yongxin Chen, Yin Cui, Yifan Ding, et al. Cosmos world foundation model platform for physical ai. *arXiv preprint arXiv:2501.03575*, 2025.

Simon Armleder, Florian Bergner, Julio Rogelio Guadarrama-Olvera, Jun Nakanishi, and Gordon Cheng. Real-time control of a humanoid robot for whole-body tactile interaction. *Advanced Intelligent Systems*, 7(12):e202500149, 2025.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

Simone Borelli, Francesco Giovinazzo, Francesco Grella, Giammarco Caroleo, Alessandro Albini, Perla Maiolino, and Giorgio Cannata. Sestosenso: Proxy-tactile perception for intelligent and safe human–robot interaction. In *2026 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1–8. IEEE, 2026.

Rory J Bufacchi and Gian Domenico Iannetti. An action field theory of peripersonal space. *Trends in cognitive sciences*, 22(12):1076–1090, 2018.

Caleb Escobedo, Matthew Strong, Mary West, Ander Aramburu, and Alessandro Roncone. Contact anticipation for physical human–robot interaction with robotic manipulators using onboard proximity sensors. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7255–7262. IEEE, 2021.

Partha Ghosh, Mehdi SM Sajjadi, Antonio Vergari, Michael Black, and Bernhard Schölkopf. From variational to deterministic autoencoders. *arXiv preprint arXiv:1903.12436*, 2019.

Zihao He, Hongjie Fang, Jingjing Chen, Hao-Shu Fang, and Cewu Lu. Foar: Force-aware reactive policy for contact-rich robotic manipulation. *IEEE Robotics and Automation Letters*, 10(6):5625–5632, 2025.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2017.

Songqiao Hu, Zeyi Liu, Shuang Liu, Jun Cen, Zihan Meng, Shihefeng Wang, Xiang Li, and Xiao He. Vlsa: Vision-language-action models with plug-and-play safety constraint layer. *arXiv preprint arXiv:2512.11891*, 2025.

Daehwa Kim, Mario Srouji, Chen Chen, and Jian Zhang. Armor: Egocentric perception for humanoid robot collision avoidance and motion planning. *arXiv preprint arXiv:2412.00396*, 2024a.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024b.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Carson Kohlbrenner, Caleb Escobedo, S Sandra Bae, Alexander Dickhans, and Alessandro Roncone. GenTact Toolbox: A computational design pipeline to procedurally generate context-driven 3d printed whole-body artificial skins. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4716–4722. IEEE, 2025.

Carson Kohlbrenner, Anna Soukhovei, Caleb Escobedo, Nataliya Nechyporenko, and Alessandro Roncone. Design, mapping, and contact anticipation with 3d-printed whole-body tactile and proximity sensors. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*, 2026.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Hang Liu, Yuman Gao, Sangli Teng, Yufeng Chi, Yakun Sophia Shao, Zhongyu Li, Maani Ghaffari, and Koushil Sreenath. Ego-vision world model for humanoid contact planning. *arXiv preprint arXiv:2510.11682*, 2025.

Zhe Liu, Diansheng Chen, Junlin Ma, Daiyu Jia, Tingyu Wang, and Qile Su. A self-capacitance proximity e-skin with long-range sensibility for robotic arm environment perception. *IEEE Sensors Journal*, 23(13):14854–14863, 2023.

Seung Jae Moon, Jinsol Kim, Hongsik Yim, Yeeun Kim, and Hyouk Ryeol Choi. Real-time obstacle avoidance using dual-type proximity sensor for safe human-robot interaction. *IEEE Robotics and Automation Letters*, 6(4):8021–8028, 2021.

Stefan Escaida Navarro, Stephan Mühlbacher-Karrer, Hosam Alagi, Hubert Zangl, Keisuke Koyama, Björn Hein, Christian Duriez, and Joshua R Smith. Proximity perception in human-centered robotics: A survey on sensing systems and applications. *IEEE Transactions on Robotics*, 2021.

Jean-Paul Noel, Olaf Blanke, Elisa Magosso, and Andrea Serino. Neural adaptation accounts for the dynamic resizing of peripersonal space: evidence from a psychophysical-computational approach. *Journal of neurophysiology*, 119(6):2307–2333, 2018.

NVIDIA, Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Muñoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, Lukasz Wawrzyniak, Milad Rakhsha, Alain Denzler, Eric Heiden, Ales Borovicka, Ossama Ahmed, Iretiayo Akinola, Abrar Anwar, Mark T. Carlson, Ji Yuan Feng, Animesh Garg, Renato Gasoto, Lionel Gulich, Yijie Guo, M. Gussert, Alex Hansen, Mihir Kulkarni, Chenran Li, Wei Liu, Viktor Makoviychuk, Grzegorz Malczyk, Hammad Mazhar, Masoud Moghani, Adithyavairavan Murali, Michael Noseworthy, Alexander Poddubny, Nathan Ratliff, Welf Rehberg, Clemens Schwarke, Ritvik Singh, James Latham Smith, Bingjie Tang, Ruchik Thaker, Matthew Trepte, Karl Van Wyk, Fangzhou Yu, Alex Millane, Vikram Ramasamy, Remo Steiner, Sangeeta Subramanian, Clemens Volk, CY Chen, Neel Jawale, Ashwin Varghese Kuruttukulam, Michael A. Lin, Ajay Mandlekar, Karsten Patzwaldt, John Welsh, Huihua Zhao, Fatima Anes, Jean-Francois Lafleche, Nicolas Moënne-Loccoz, Soowan Park, Rob Stepinski, Dirk Van Gelder, Chris Amevor, Jan Carius, Jumyung Chang, Anka He Chen, Pablo de Heras Ciechomski, Gilles Daviet, Mohammad Mohajerani, Julia von Muralt, Viktor Reutskyy, Michael Sauter, Simon Schirm, Eric L. Shi, Pierre Terdiman, Kenny Vilella, Tobias Widmer, Gordon Yeoman, Tiffany Chen, Sergey Grizan, Cathy Li, Lotus Li, Connor Smith, Rafael Wiltz, Kostas Alexis, Yan Chang, David Chu, Linxi "Jim" Fan, Farbod Farshidian, Ankur Handa, Spencer Huang, Marco Hutter, Yashraj Narang, Soha Pouya, Shiwei Sheng, Yuke Zhu, Miles Macklin, Adam Moravanszky, Philipp Reist, Yunrong Guo, David Hoeller, and Gavriel State. Isaac Lab: A gpu-accelerated simulation framework for multi-modal robot learning, 2025. URL https://arxiv.org/abs/2511.04831.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025.

Lerrel Pinto, Marcin Andrychowicz, Peter Welinder, Wojciech Zaremba, and Pieter Abbeel. Asymmetric actor critic for image-based robot learning, 2017. URL https://arxiv.org/abs/1710.06542.

Yan Qin, Yue Chen, Wenwei Lin, Shujia Liu, Chuqiao Lyu, Kailun Su, Chenze Yu, Ping Luo, Wenbo Ding, Tianxing Chen, et al. Dextouch-wm: Learning action-conditioned tactile world models from human touch for dexterous robot manipulation. *arXiv preprint arXiv:2609.20649*, 2026.

Alessandro Roncone, Matej Hoffmann, Ugo Pattacini, Luciano Fadiga, and Giorgio Metta. Peripersonal space and margin of safety around the body: learning visuo-tactile associations in a humanoid robot with artificial skin. *PloS one*, 11(10):e0163713, 2016.

Constantin Scholz, Hoang-Long Cao, Emil Imrith, Nima Roshandel, Hamed Firouzipouyaei, Aleksander Burkiewicz, Milan Amighi, Sebastien Menet, Dylan Warawout Sisavath, Antonio Paolillo, et al. Sensor-enabled safety systems for human–robot collaboration: A review. *IEEE Sensors Journal*, 25(1):65–88, 2024.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Shahram Najam Syed, Arthur Jakobsson, Haoran Hao, and Jeffrey Ichnowski. Intercepting the future: Latent-space predictive world model for dynamic vla manipulation. *arXiv preprint arXiv:2606.02486*, 2026.

NeoteAI Team, Fudan TEAI Team, et al. n_0-twam: Scaling tactile-native world-action model for contact-rich manipulation. *arXiv preprint arXiv:2607.23783*, 2026.

Naiyao Wang, Bo Zhang, Haixu Chi, Hua Wang, Seán McLoone, and Hongbo Liu. Duel: Depth visual ego-motion learning for autonomous robot obstacle avoidance. *The International Journal of Robotics Research*, 43(3):305–329, 2024.

Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, and Stefano Ermon. A theory of usable information under computational constraints. *arXiv preprint arXiv:2002.10689*, 2020.

Han Xue, Sikai Liang, Zhikai Zhang, Zicheng Zeng, Yun Liu, Yunrui Lian, Jilong Wang, Qingtao Liu, Xuesong Shi, and Li Yi. Collision-free humanoid traversal in cluttered indoor scenes. *arXiv preprint arXiv:2601.16035*, 2026.

Jiahui Yang, Jason Jingzhou Liu, Yulong Li, Youssef Khaky, Kenneth Shaw, and Deepak Pathak. Deep reactive policy: Learning reactive manipulator motion planning for dynamic environments. In *2nd Workshop on Safe and Robust Robot Learning for Operation in the Real World*, 2025. URL `https://openreview.net/forum?id=v2BNi6UQKr`.

Lizhi Yang, Junheng Li, and Aaron D Ames. Pac-man: Perception-aware cbf-rl for whole-body safety in humanoid dodgeball. *arXiv preprint arXiv:2607.28623*, 2026.

Zizhe Zhang, Yicong Wang, Zhiquan Zhang, Tianyu Li, and Nadia Figueroa. Viability-preserving passive torque control. *arXiv preprint arXiv:2510.03367*, 2025.

Zhenhao Zhou and Dan Negrut. Chronodreamer: Action-conditioned world model as an online simulator for robotic planning. *arXiv preprint arXiv:2512.18619*, 2025.

A Appendix

A.1 Dodgeball Task

The humanoid dodgeball task was simulated in Isaac Lab (NVIDIA et al., 2025) using the Unitree H1-2 humanoid robot platform. An artificial skin was generated for the upper-half of the robot using the GenTact Toolbox extension in Blender (Kohlbrenner et al., 2025). Proximity sensor locations were randomly placed using a Poisson-disk sampling algorithm to distribute them evenly over the body of the robot.

The ball propelled at the robot in the dodgeball task had a constant radius of $15\,\mathrm{cm}$. The ball was launched in intervals of 1-2 seconds apart and spawned between 2-3 m away from the robot, between 1-3 m above ground level, and had an initial speed between 4-6 m/s.

Task: The policy $\pi_\theta(a_t|o_t)$ outputs target joint positions $q_t^* \in \mathbb{R}^{21}$, which are converted into motor torques τ_t via a low-level PD controller:

$$\tau_t = K_p(q_t^* - q_t) + K_d(\dot{q}_t^* - \dot{q}_t), \tag{11}$$

where K_p and K_d are the proportional and derivative gains, and joint velocities $\dot{q}_t^*$ are approximated by zeros or finite differences. The policy commands to maximize the expected cumulative reward:

$$J(\theta) = \mathbb{E}_{r\sim p_{\pi_\theta}}\left[\sum_{t=0}^{T}\gamma^t r_t\right] \qquad \max_\theta J(\theta), \tag{12}$$

where γ is the discount factor and reward function r_t. Training is conducted using PPO and is tuned to favor compliant, stable motion as opposed to direct torque control.

Rewards: The policy is optimized at 60 Hz ($\Delta t = 1/120\,\mathrm{s}$ with decimation 2). Each step returns

$$r_t = w_h \exp\big(-\alpha(h_t - h^\star)^2\big) + w_a - w_d\|\Delta p_t\|_2^2 - w_e\|\tau_t\|_1. \tag{13}$$

Here h_t is the root height in meters, $h^\star$ is a target standing height, $\Delta p_t \in \mathbb{R}^2$ is the horizontal displacement from the reset pose, and τ_t is the vector of applied joint torques. The Gaussian rewards an upright posture, and w_a is received on every step until termination, so longer episodes are the objective that produces dodging. The drift and torque terms suppress leaving the spawn point and high-frequency motion. Eq. (13) depends only on the root state and the joint torques, so the reward does not favor a particular proximity-sensor representation. Contact, a root height below $0.4\,\mathrm{m}$, and a root tilt beyond 80° end the episode and stop the survival and height terms. In the reported experiments, $w_h = 10$, $\alpha = 5$, $h^\star = 1.04\,\mathrm{m}$, $w_a = 2$, $w_d = 1$, and $w_e = 0.01$.

A.2 Data Curation

The PWM is pretrained on a sample dataset of 6000 trajectories of expert demonstrations containing ray sensor observations and actions. To make sure the PWM can accurately predict contact for wide range of potential policies, one third of the experts were trained having seen only raw observations (x^{raw}), one third was trained having perfect localization of incoming balls (f_{tp}), and the remaining third was trained with only the minimum sensor distance (f_{min}). To ensure that the PWM could predict incoming projectiles during training and not just when the robot has full range of motion, 50% of the data contains intermediate training step demonstrations, comprised equally of the aforementioned experts at 500, 1000, 1500, 2000, and 2500 steps of training. The remaining 50% of the data was collected using the fully trained experts at 3000 steps. 15% of the dataset was used for validation, and an additional hold-out dataset of 1000 trajectories was collected using two experts that were not used for collecting the training dataset to verify that the PWM did not overfit.

The dynamics model learns the transition dynamics between time steps and can fail when the frame rate of the data is too high to capture local dynamics. To prevent this failure mode, the 60 Hz dataset was windowed in a 10 Hz stride, forcing the dynamics model to predict in $100\,\mathrm{ms}$ intervals.

Another failure mode of the PWM is too sparse of contact labels in the data to drive effective gradient passes, but by adding contact padding frames after an initial contact occurred, the windows could capture the sparse moments leading up to contact in the full $C + R$ sized window.

A.3 World Model Architecture

A.3.1 Overview

Encoder/Decoder Model: The encoder and decoder used in this work are three-dimensional variational autoencoders (3D-VAE), which stride over the H, W, and S (Kingma & Welling, 2013). A handful of variations to the standard evidence lower bound (ELBO) loss are employed in our approach to improve the fidelity of depth signals, such as a β regime to reduce pixel smoothing on the posterior (Higgins et al., 2017), a deterministic reconstruction term to improve decoder accuracy (Ghosh et al., 2019), and a sigmoid decoder trained using weighted soft-label binary cross-entropy (Section A.3.1).

Dynamics Model: The dynamics model uses conditional flow matching (Lipman et al., 2022) to generate a sequence of R likely future latent observations $z_{t+1:t+R}$ conditioned on C prior encoded observations, and candidate future actions $\hat{a}$ as

$$\hat{z}_{t+1:t+R} \sim r_\phi(z_{t+1:t+R} \mid z_{t-C+1:t}, a_t). \tag{14}$$

The dynamics model draws the full observation roll-out in a single step, making it not auto-regressive to reduce inference time during training. Since the dynamics model is not auto-regressive, we assume a zero-order hold on candidate actions based on the most recent action a_t.

Contact Anticipation Head : We attach a causal contact-anticipation head that maps the world model's imagined rollout to a first-contact stopping decision. It takes the fully-denoised context latents plus the jointly-generated rollout latents, pools each frame's sensor tokens to a single frame token, and processes the resulting context+rollout sequence with a causal transformer, yielding one

hazard probability $h_r \in [0, 1]$ per rollout step where h_r sees only history through r. At inference we threshold the logits under a sigmoid filter and select the first step above the threshold to define the first contact state in the rollout.

We train the contact anticipation head as a discrete-time first-contact hazard problem rather than per-frame classification: only pre-event steps contribute survival loss $-\log(1 - h_t)$, only the first contact step contributes event loss $-\log(h_k)$, and post-contact steps are masked, with an additional margin ranking term forcing the event logit above the hardest pre-event logit in the same window. Training is done on cached world-model generations rather than ground-truth latents, so the head learns the generator's artifacts directly.

The proximity world model is held fixed during policy learning. A frame is a stack of $S = 64$ depth images, $x_t \in [0, 1]^{1\times8\times8\times64}$, with millimeter ranges divided by 4000 so that 1 is the 4 m full scale. The model uses $C = 3$ context frames and predicts $R = 9$ future frames. At the 10 Hz training rate in Appendix A.2, the context covers the preceding 0.2 s and the rollout covers the next 0.9 s. The three modules are trained in the frozen order of Appendix A.4. Parameter counts are 2.77×10^7 for the encoder–decoder, 2.20×10^6 for the dynamics model, and 2.86×10^4 for the contact head.

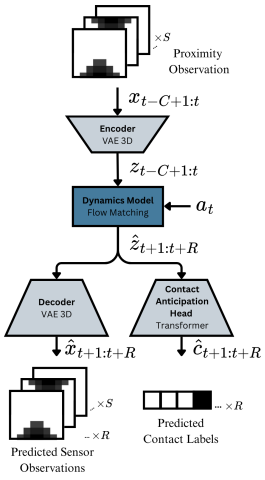

Figure 7: PWM architecture.

A.3.2 Implementation

Encoder: The encoder is a 3D convolutional VAE (Kingma & Welling, 2013). Two convolutions with kernel 4 and stride $(2, 2, 2)$ map $1 \to 256 \to 512$ channels and downsample every spatial axis by 4, followed by a stride-1 convolution that preserves 512 channels. A residual block with hidden width 128 (kernel-3 convolution, ReLU, pointwise convolution, residual add) is then applied eight times with tied weights. Separate $1 \times 1 \times 1$ convolutions read out the posterior mean and log-variance, with $\log \sigma^2$ clamped to $[-10, 10]$:

$$q_\psi(z_t \mid x_t) = \mathcal{N}\big(\mu_\psi(x_t),\, \mathrm{diag}\, \sigma^2_\psi(x_t)\big), \qquad \mu_\psi(x_t) \in \mathbb{R}^{8\times2\times2\times16}. \tag{15}$$

Flattening the spatial axes gives 64 tokens of dimension $e = 8$. The decoder mirrors the encoder with transposed convolutions and ends in a sigmoid, so each depth pixel is a Bernoulli probability. Downstream models consume $\mu_\psi(x_t)$ with no reparameterization noise. Sampling is used only for the stochastic reconstruction term during VAE training.

Depth images are bimodal: most pixels sit at full scale, and the ball occupies a few low values. Mean-squared error therefore blurs the near field. The reconstruction term is a weighted binary cross-entropy,

$$\mathcal{L}_{\mathrm{BCE}}(\hat{x}, x) = \mathbb{E}\big[w(x)\, \mathrm{BCE}(\hat{x}, x)\big], \qquad w(x) = \begin{cases} 10 & x < 0.4, \\ 1 & x \geq 0.4, \end{cases} \tag{16}$$

so returns closer than 1.6 m receive $10\times$ reconstruction weight. An ℓ_1 penalty on horizontal and vertical finite differences, applied inside each sensor and not across sensors, keeps range edges:

$$\mathcal{G}(\hat{x}, x) = \|\nabla_H \hat{x} - \nabla_H x\|_1 + \|\nabla_W \hat{x} - \nabla_W x\|_1. \tag{17}$$

Let $\hat{x}_z$ be the decode of a reparameterized sample and $\hat{x}_\mu$ the decode of $\mu_\psi(x)$. The training objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{VAE}} = &\big[\mathcal{L}_{\mathrm{BCE}}(\hat{x}_z, x) + \lambda_\mu \mathcal{L}_{\mathrm{BCE}}(\hat{x}_\mu, x)\big] \\ &+ \lambda_\nabla \big[\mathcal{G}(\hat{x}_z, x) + \lambda_\mu \mathcal{G}(\hat{x}_\mu, x)\big] + \beta\, D_{\mathrm{KL}}\big(q_\psi(z \mid x) \,\|\, \mathcal{N}(0, I)\big), \end{aligned} \tag{18}$$

with $\lambda_\mu = 1.5$, $\lambda_\nabla = 0.1$, and $\beta = 10^{-4}$. The KL is the mean per-dimension closed form, so β is a light regularizer: reconstruction, including the deterministic decode of μ (Ghosh et al., 2019), determines the latent, while a nonzero KL keeps dynamics samples in a region the decoder has seen (Higgins et al., 2017). The encoder is optimized with Adam at learning rate 10^{-4} for 100 epochs and batch size 512, on frames shuffled independently of time.

True class	Labels	Prediction accuracy
Contact	17,271	96.9% 16,737/17,271
Non-contact	121,131	83.8% 101,534/121,131
Total	138,402	85.5% 118,271/138,402

Figure 8: Contact prediction accuracy on the validation dataset.

Dynamics: Future latents are drawn jointly with conditional flow matching (Lipman et al., 2022). For a target window $z_1 = z_{t+1:t+R}$ and $z_0 \sim \mathcal{N}(0, I)$,

$$z_\tau = (1-\tau)z_0 + \tau z_1, \qquad \tau \sim \mathrm{Unif}(0,1), \qquad \mathcal{L}_{\mathrm{FM}} = \left\| v_\phi(z_\tau, \tau \mid z_{t-C+1:t}, a_t) - (z_1 - z_0) \right\|_2^2. \tag{19}$$

The conditional velocity v_ϕ is a two-layer diffusion transformer (DiT) of width 256, four heads, and feed-forward width 512, with dropout 0.001 (Peebles & Xie, 2023). Context tokens and noisy future tokens are concatenated. Each token is the sum of a linear map of its 8-dimensional latent, a learned frame index, a learned index among the 64 spatial tokens, and a learned role embedding that marks context versus future. The only action input is the 21-dimensional joint target a_t at the last context frame. A two-layer MLP ($21 \to 256 \to 256$, SiLU) maps a_t and adds it to every spatial token of every future frame, so the same action is held for all R steps in both training and sampling. Flow time τ is a sinusoidal embedding that produces adaptive layer-norm scale, shift, and residual gates inside each block, zero-initialized so each block starts as the identity (Peebles & Xie, 2023). Attention is bidirectional over the window. Only the future-token features are projected back to an 8-dimensional velocity.

Sampling integrates $\mathrm{d}z/\mathrm{d}\tau = v_\phi$ with Euler's method for 3 uniform steps from $\tau = 0$ to $\tau = 1$, starting at Gaussian noise. 3 steps was found to be a sweet spot between computation and prediction accuracy for our application (Fig. 9). All R frames are produced by that one integration. The dynamics model is optimized with AdamW at learning rate 5×10^{-4}, weight decay 10^{-3}, and gradient clip 1, for 100 epochs and batch size 512, on posterior means from the frozen encoder.

Contact prediction head: The head maps a fully denoised context and a generated rollout to a first-contact decision. Within each frame, tokens are projected to width 32, the same frame, spatial, and role embeddings are added, and a learned query cross-attends to the 64 tokens. A residual MLP of width 64 yields one frame vector. A two-layer causal transformer (width 32, two heads, feed-forward width 64, dropout 0.05) then runs over the $C + R$ frame vectors. The hazard logit ℓ_r for future step $r \in \{1, \ldots, R\}$ is read from the state that has attended to the context and to generated frames $1, \ldots, r-1$. The frame being scored is excluded, so a generated latent cannot explain its own contact label. The linear head is bias-initialized to -1.4, i.e. a prior hazard probability $\sigma(-1.4) \approx 0.20$.

Labels come from the simulator contact flag and may stay positive after impact. Training uses only the first positive step. If that step is T_c, steps $t < T_c$ are survival steps, step T_c is the event, and steps $t > T_c$ are masked. A window with no contact contributes only survival terms. With $h = \sigma(\ell)$,

$$\mathcal{L}_{\mathrm{event}} = \mathbb{E}_{t=k}\big[-\log h_t\big], \qquad \mathcal{L}_{\mathrm{surv}} = \mathbb{E}_{t<k}\big[-\log(1-h_t)\big], \tag{20}$$
$$\mathcal{L}_{\mathrm{rank}} = \mathbb{E}\Big[\mathrm{softplus}\big(\max_{t<k} \ell_t - \ell_k + m\big)\Big], \qquad \mathcal{L}_{\mathrm{hazard}} = \mathcal{L}_{\mathrm{event}} + \lambda_s \mathcal{L}_{\mathrm{surv}} + \lambda_r \mathcal{L}_{\mathrm{rank}}.$$

Averages are taken over the positions or windows that contribute to each term, so the single event step is not diluted by the longer survival prefix. Windows whose contact falls on the first rollout step have no pre-event logit and skip the ranking term. We use $\lambda_s = 0.8$, $\lambda_r = 0.3$, and $m = 0$. The head is optimized with Adam at learning rate 3×10^{-4} for 200 epochs and batch size 1024, on rollouts cached from the frozen dynamics model, so it learns that generator's errors.

At inference, $h_r = \sigma(\ell_r)$ and the selected future frame is

$$r^\star = \min\{r : h_r \geq 0.5\}, \tag{21}$$

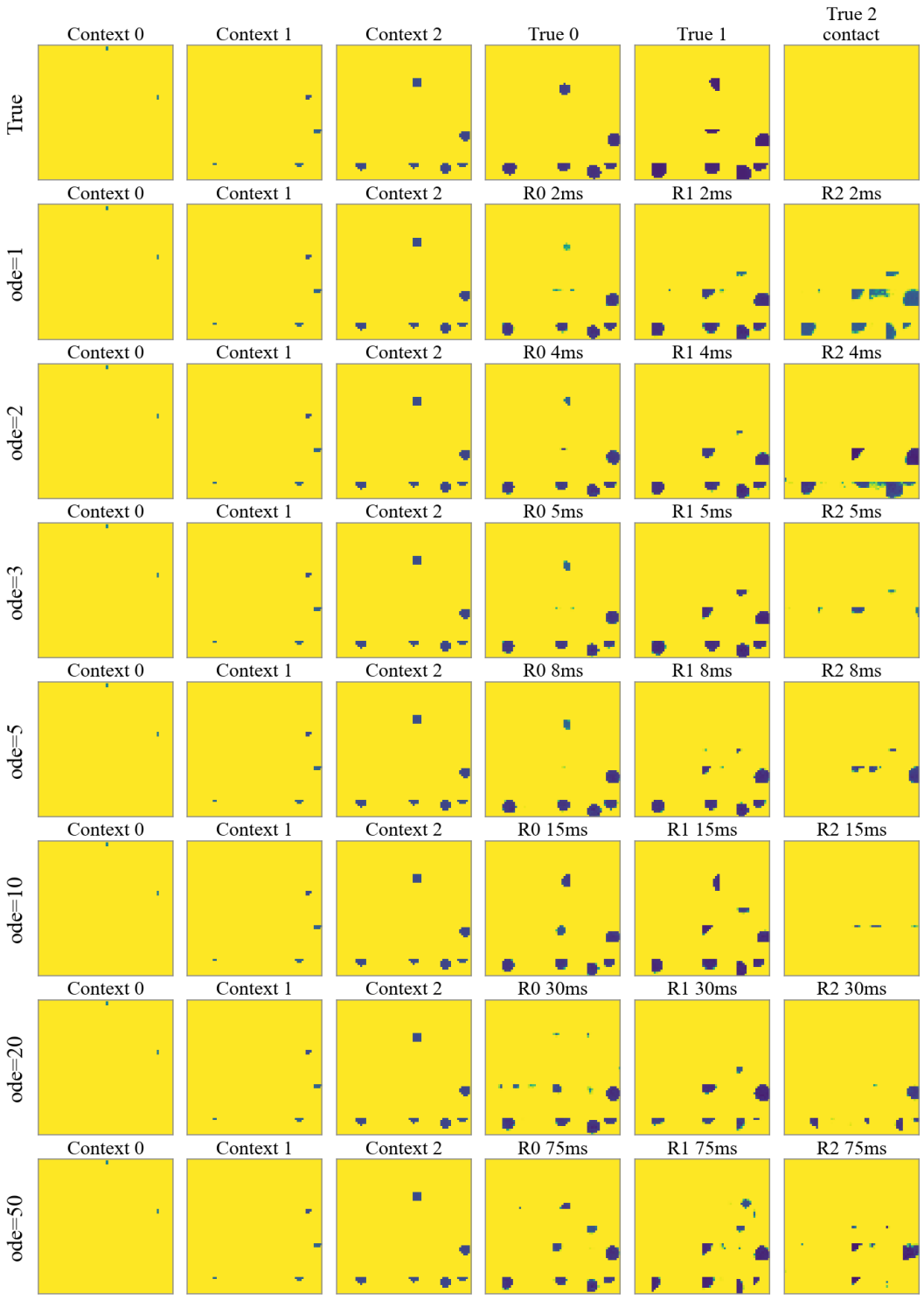

Figure 9: A comparison of Euler step sizes and computation for the dynamics model.

or $r^\star = R$ when no step crosses 0.5. The policy receives the latent of that frame.

The contact head achieved a 96.9% prediction accuracy on future contact, and an 83.8% accuracy on non-contact labels (Fig. 8).

A.4 TRAINING REGIME

The encoder and decoder were trained together without the dynamics and contact prediction models, then held frozen while the dynamics model and contact anticipation head were trained. The dynamics model was also held frozen during training of the contact prediction head.

The encoder and decoder were trained on a shuffled set of raw sensor observations. The dynamics model was trained on temporally consistent latent observations generated by the encoder model, and the contact anticipation head was trained on rollouts of latent vectors generated by the dynamics model.

Limitations: We note that a large dataset of trajectories was needed to train the PWM itself, rendering it not sample-efficient for adaptive end-to-end collision avoidance on real data. However, we maintain the stance that our investigation reveals *how* representing future sensor observations can influence downstream sample-efficiency during training. Additionally, the goal of this paper is to achieve sample-efficient behavior learning through sensor representation choice, measured via relative improvements in sample efficiency on a dynamic collision avoidance task that has a humanoid surrounded at all sides by impending collision. However, *we do not claim that the training example presented is in itself an absolute, sample-efficient approach*. We hope that this work advances egocentric proximity sensors do be designed for sample-efficiency downstream and end-to-end on real systems.