A Note on Main-Paper Table References

Due to a table float-ordering issue in the submitted paper, a few main-text table references are inconsistent with the printed table numbers. This does not affect any reported values, or experimental results, for clarity, the intended table mapping is as follows:

- In the "Zero-shot cross morphology transfer" paragraph, the PPO variant comparison refers to the table captioned "Zero-shot cross-morphology transfer".
- In the "Comparison to other Baselines" paragraph, the external-baseline comparison refers to the table captioned "External baselines vs. MorFiC".
- In the "Ablations" section, the conditioning-variant comparison refers to the table captioned "Ablation Table".

B Evaluation Metrics and Diagnostic Definitions

Normalized morphology distance. To analyze how transfer performance degrades with morphology shift, we have defined the normalized distance $d \in [0, 1]$ between target robot and training robot based on the 11-D morphology descriptor selected from Table I. We compute.

$$d(m_{\text{target}}, m_{\text{train}}) = \frac{1}{\sqrt{11}} \left\| \tilde{m}_{\text{target}} - \tilde{m}_{\text{train}} \right\|_2 . \tag{7}$$

This distance is used for transfer loss analysis and is not provided to the policy or critic during training (refer to Fig. 5).

Normalized Value Bias (NVB). NVB measures whether the critic systematically overestimates or underestimates returns, where close to 0 equals low systematic bias and higher value indicate higher bias and is given by:

$$\text{NVB}_\phi = \frac{\mathbb{E}_{(t,i)\in\mathcal{I}} \left[V_\phi(s_{t,i}, z_{m_i}) - \hat{G}_{t,i} \right]}{\sigma_G + \epsilon}. \tag{8}$$

Normalized Value Error (NVE). NVE measures the total critic prediction error, normalized by the return scale, lower value means better calibrated critic and is given by:

$$\text{NVE}_\phi = \frac{\sqrt{\mathbb{E}_{(t,i)\in\mathcal{I}} \left[\left(V_\phi(s_{t,i}, z_{m_i}) - \hat{G}_{t,i} \right)^2 \right]}}{\sigma_G + \epsilon}.. \tag{9}$$

Advantage Sign Agreement (ASA). ASA measures whether the critic-induced advantage has the same sign as the Monte-Carlo reference advantage, higher ASA is better and is given by:

$$\hat{A}^{\text{MC}}_{t,i} = \hat{G}_{t,i} - \frac{1}{|\mathcal{I}_\epsilon|} \sum_{(t',i')\in\mathcal{I}_\epsilon} \hat{G}_{t',i'}. \tag{10}$$

$$\hat{A}^{\phi}_{t,i} = \hat{G}_{t,i} - V_\phi(s_{t,i}, z_m). \tag{11}$$

$$\text{ASA}_\phi = \frac{1}{|\mathcal{I}_\epsilon|} \sum_{(t,i)\in\mathcal{I}_\epsilon} \mathbf{1} \left[\text{sign}\left(\hat{A}^{\phi}_{t,i} \right) = \text{sign}\left(\hat{A}^{\text{MC}}_{t,i} \right) \right]. \tag{12}$$

Table 5: **Morphology randomization parameters.** Each row reports one dimension of the 11-D morphology vector used by MorFiC. Values outside the training support are shown in bold with ↑ / ↓.

Parameter	Train min	Train max	Go2	Go1	A1	Cheetah	AlienGo	ANYmal	B1
Thigh mass	0.634	1.152	1.152	0.899	1.013	0.634	0.639	0.740	**4.300**↑
Calf mass	0.064	0.166	0.154	0.158	0.166	0.064	**0.207**↑	**0.337**↑	**1.050**↑
Hip mass	0.510	0.696	0.678	0.510	0.696	0.540	**1.000**↑	**0.740**↑	**2.150**↑
Foot mass	0.000	0.060	0.040	0.060	0.060	0.060	0.060	**0.250**↑	0.050
Base mass	3.300	6.921	6.921	4.800	6.000	3.300	**11.644**↑	**18.210**↑	**29.450**↑
Hip limit min	-1.100	-0.800	-1.047	-0.803	-0.803	**-1.600**↓	**-1.222**↓	**-0.720**↑	**-0.750**↑
Hip limit max	0.800	1.100	1.047	0.803	0.803	**1.600**↑	**1.222**↑	**0.720**↓	**0.750**↓
Thigh limit min	-1.650	-1.050	-1.571	**-1.047**↑	**-1.047**↑	**-2.600**↓	**-3.142**↓	-1.650	**-1.000**↑
Thigh limit max	3.490	4.050	3.491	**4.189**↑	**4.189**↑	**2.600**↓	**3.142**↓	4.050	3.500
Calf limit min	-2.720	-2.600	**-2.723**↓	-2.697	-2.697	-2.600	**-2.775**↓	-2.720	-2.600
Calf limit max	-0.950	-0.780	-0.838	-0.916	-0.916	-0.800	**-0.646**↑	-0.780	**-0.600**↑

Sign Flip Rate. Sign flip rate measures how often the critic-induced advantage has the opposite sign from the MC reference advantage and lower rate is better and is given by:

$$\mathrm{Flip}_\phi = \frac{1}{|\mathcal{I}_\epsilon|} \sum_{(t,i)\in\mathcal{I}_\epsilon} \mathbf{1}\left[\mathrm{sign}\left(\hat{A}^{\phi}_{t,i}\right) \neq \mathrm{sign}\left(\hat{A}^{\mathrm{MC}}_{t,i}\right)\right]. \tag{13}$$

$$\mathrm{Flip}_\phi = 1 - \mathrm{ASA}_\phi. \tag{14}$$

Gradient Cosine. Gradient cosine measures whether the PPO actor update direction produced by the critic-induced advantage matches the actor update direction produced by the MC reference advantage, where close to 1 means critic gives almost the same actor update as MC return, so a value higher than 0 and closer to 1 is better and is given by:

$$g(\hat{A}^{\phi}) = \nabla_\theta \mathcal{L}_{\mathrm{actor}}\left(\theta; \hat{A}^{\phi}\right), \qquad g(\hat{A}^{\mathrm{MC}}) = \nabla_\theta \mathcal{L}_{\mathrm{actor}}\left(\theta; \hat{A}^{\mathrm{MC}}\right). \tag{15}$$

$$\mathrm{CosGrad}_\phi = \frac{g(\hat{A}^{\phi})^\top g(\hat{A}^{\mathrm{MC}})}{\left\|g(\hat{A}^{\phi})\right\|_2 \left\|g(\hat{A}^{\mathrm{MC}})\right\|_2}. \tag{16}$$

Validity of survival-based metrics. Survival rate alone is not sufficient to indicate successful locomotion under cross-morphology transfer. In several in & out-of-distribution cases, a policy can avoid early termination while producing negligible forward motion, unstable stepping, collapsed postures, or physically invalid behaviors such as dragging or non-propulsive oscillations. And in some cases the policy do have valid motions while also producing behaviors like upside down leg movement, forward falling, disabled rear legs etc in some fraction of robot while the rate remains 1.00 or 100%. We therefore interpret survival jointly with achieved forward velocity, tracking ratio, and rollout validity. In the result tables, high survival with near-zero forward velocity is treated as a failure to transfer rather than successful locomotion.

This distinction is especially important for distant morphologies such as B1 and ANYmal-C, where policies may remain non-terminated but fail to generate a valid gait. High survival with near-zero forward velocity is not considered successful transfer; survival is interpreted jointly with achieved speed and rollout validity.

C Morphology Descriptor and Randomization

MorFiC uses an 11-D morphology vector containing link masses and joint-limit parameters. During training, each environment samples a morphology vector from the support in Table 5; this vector is applied in simulation and kept fixed for the episode. The same vector is encoded into the latent z_m used by the actor and critic. Table 5 also reports the descriptors of all evaluation robots, with

Table 6: **Effect of expanded morphology ranges and FiLM conditioning.** Each entry reports $v_x^{\max} \pm \sigma/S$, where S denotes survival rate. Bold indicates the higher value for each metric independently; ties are bolded for both methods.

Target	Expanded range $v_x^{\max} \pm \sigma/S$	FiLM $v_x^{\max} \pm \sigma/S$
Go2	$1.9013 \pm 0.3510/0.8531$	$\mathbf{2.0415 \pm 0.3670/0.9844}$
Go1	$1.9208 \pm 0.3087/\mathbf{0.7656}$	$\mathbf{2.0649 \pm 0.4552}/0.5469$
A1	$1.8241 \pm 0.2322/\mathbf{0.7596}$	$\mathbf{2.1404 \pm 0.3608}/0.5156$
AlienGo	$1.5499 \pm 0.5078/\mathbf{1.0000}$	$\mathbf{1.9812 \pm 0.5146/1.0000}$
B1	$0.2001 \pm 0.0215/\mathbf{1.0000}$	$\mathbf{0.2109 \pm 0.0768/1.0000}$
Cheetah	$1.2362 \pm 0.4805/0.3348$	$\mathbf{1.3922 \pm 0.6902/0.4219}$
ANYmal-C	$\mathbf{0.3079 \pm 0.1803/1.0000}$	$0.2505 \pm 0.2147/0.9688$

out-of-support values marked by $\uparrow$ or $\downarrow$. Go1 and A1 remain largely near the training support, while AlienGo, ANYmal-C, and especially B1 introduce larger shifts in base mass, limb masses, and joint limits.

C.1 Morphology Randomization

The randomization range was chosen to cover moderate morphology variation around the Go2 source template while preserving stable PPO training. This creates a controlled zero-shot setting: MorFiC is not trained on the target robot models, but observes morphology variations that partially cover nearby quadrupeds. B1 is the most distant target, with substantially larger link and base masses than the training support.

C.2 Effect of Expanding Morphology Randomization

We also tested whether widening the morphology range improves transfer to distant robots such as B1. Expanding the support to include much heavier morphologies made the policy more conservative, reducing forward tracking without resolving the B1 failure mode (see Table 6). This suggests that distant morphology transfer is not solved by naively widening randomization alone, and may require structured morphology curricula or limited multi-source training.

D Policy Architecture, Observation Streams, and PPO Hyperparameters

D.1 Observation Streams

Table 7 summarizes the observation streams used by MorFiC. The policy receives a history of proprioceptive observations, including base velocity, angular velocity, projected gravity, command, joint positions, joint velocities, and previous actions. The privileged stream contains simulator-side physical parameters used during training, while the morphology stream contains the 11-D descriptor encoded into z_m. All methods use the same observation history and privileged information; they differ only in how morphology information is injected into the actor–critic architecture.

D.2 Actor-Critic and Morphology Encoder Architecture

Table 8 & 9 reports the architecture used for MorFiC. We use separate actor and critic MLP trunks with ELU activations. The morphology vector is encoded using a lightweight MLP and used to condition the critic through residual scaled FiLM modulation at each hidden layer. The final FiLM layer is initialized to zero, so training starts close to an unconditioned critic and gradually learns morphology-dependent modulation.

Table 7: Observation streams used for policy and value learning.

Stream	Signals	Dim
o_t	$[\, v,\ \omega,\ g_{\text{proj}},\ c,\ q - q_0,\ \dot{q},\ a_{t-1}\,]$	48
h_t	$\{o_{t-k}\}_{k=1}^{15}$	15×48
p_t	$[\mu_{\text{fric}},\ e,\ m_{\text{base}},\ \Delta c,\ \mu_{\text{joint}}]$	18
z_m	morphology encoder output	$d = 11$

Table 8: Policy, critic, and morphology-conditioning architecture used in MorFiC. The PPO optimizer and rollout hyperparameters are shared across all methods and reported separately in Table 10.

Module	**Specification**
Actor MLP	[obs history, $\hat{z}_{\text{priv}}, z_m$] $\to [512, 256, 128] \to a$, ELU
Critic trunk	[obs history, z_{priv}] $\to [512, 256, 128] \to V$, ELU
Critic trunk (FiLM scaled)	[obs history, z_{priv}] $\to [1024, 512, 256] \to V$, ELU
Initial action std.	$\sigma_0 = 1.0$
Adaptation module	obs history $\to [256, 128] \to \hat{z}_{\text{priv}}$, ELU
Student latent dropout	$p = 0.2$ during training
Morphology encoder	$d_m \to 128 \to 64$, ELU
Main critic conditioning	Residual scaled FiLM at each critic hidden layer
FiLM generator	$64 \to 128 \to 2H$, ELU
FiLM update	$h \leftarrow h(1 + \gamma) + \beta$
FiLM scale	0.1
FiLM initialization	Last FiLM layer weights and biases initialized to zero
Decoder	Disabled

D.3 PPO Hyperparameters

Table 10 lists the PPO hyperparameters used across all baseline and MorFiC experiments. All variants share the same rollout setup, learning rate schedule, clipping parameters, entropy coefficient, value loss coefficient, and optimization settings. Thus, performance differences are attributable to the morphology-conditioning mechanism rather than changes in PPO training.

E Reward Terms and Weights

We shape our MorFiC rewards as a weighted sum of (i) command tracking terms for linear and angular velocities using exponential, (ii) stability rewards that encourage upright posture and dynamic balance via a center-of-pressure (CoP) proximity reward, (iii) gait/contact shaping terms that match desired contact schedules by penalizing unwanted swing forces and stance-phase foot velocities and a detailed definition and weights are given in Table V.

E.1 Reward structure

Our reward structure is divided into three main categories: Task/Tracking, Penalties and stability shaping rewards, and the total reward sum is given by (17).

$$r_t \;=\; \sum_{i \in T} w_i \, r_t^{(i)} \;-\; \sum_{j \in P} w_j \, p_t^{(j)} \;+\; \sum_{k \in S} w_k \, s_t^{(k)}. \tag{17}$$

Task/Tracking rewards encourage the robot to complete the commanded locomotion objective. Specifically, we tracked the commanded linear velocity and the angular velocity using exponential rewards. (Table 11)

Penalties are assigned as a way to discourage unsafe movement, which could lead to collisions, bad gait, or physically collapse(like jerks, jerky motions, base collision, joint limit violation, etc). For

Table 9: Morphology-conditioning variants evaluated in the ablation study. All variants use the same PPO settings and base actor–critic widths as Table 8.

Variant	Critic conditioning	Key detail
Vanilla PPO	None	Critic does not receive morphology information
Concat-z	Direct concatenation	z_m is appended to the critic input
FiLM	Feature-wise modulation	$h \leftarrow h(1+\gamma)+\beta$
Scaled FiLM	Residual FiLM with scale	FiLM outputs multiplied by 0.1
AdaLN	LayerNorm + FiLM	$h \leftarrow \mathrm{LN}(h)$ before FiLM modulation
HyperNet	Low-rank residual adapter	$h \leftarrow Wh + B_z A_z h$, rank 16, scale 0.1

Table 10: PPO hyperparameters used in all baseline and MorFiC experiments.

PPO setting	Value
Discount γ	0.99
GAE λ	0.95
Clip ϵ	0.2
Entropy coeff.	0.01
Value loss coeff.	1.0
Clipped value loss	True
Learning rate	1×10^{-3}
Schedule	adaptive (KL-targeted)
Target KL	0.01
Epochs / update	5
Mini-batches / update	4
Grad norm clip	1.0

example, we penalize body collisions

$$p_{\mathrm{col}} \;=\; \sum_{b\in B} 1.(\|\mathbf{f}_b\| > f_{\min})\,, \tag{18}$$

joint limit violations and is given by;

$$p_{\mathrm{lim}} \;=\; \sum_{i} \Big(\max(0,\; q_i - q_i^{\max}) + \max\big(0,\; q_i^{\min} - q_i\big) \Big). \tag{19}$$

These help the robot in not deviating from the desired locomotion gait.

Stability Shaping. Motivated by our repeated failure on unseen morphologies, we introduce 3 lightweight stability rewards namely upright stability, CoP stability, and feet-contact stability to reduce catastrophic tipping and improve recovery.

1. **Upright stability.** Encourages recovery from large pitch/tilt by penalizing deviation from projected gravity.

$$s_{\mathrm{upright}} = \exp\left(-\frac{\|\mathbf{g}_{xy}\|^2}{\sin^2(\theta_0)} \right). \tag{20}$$

2. **Center-of-pressure (CoP) stability.** Promotes balanced load distribution across the support polygon.

$$s_{\mathrm{CoP}} = \exp\left(-\frac{\|\mathrm{CoP} - \mathrm{center}\|}{\tau} \right) \cdot 1(n_{\mathrm{contact}} \geq 2)\,. \tag{21}$$

3. **Feet-contact / stance-width stability.** Penalize overly narrow instances during unstable locomotion (caused by large tilt or rapid movement when unstable). We gate this using eqns. 22, 23.

$$\mathrm{gate} = 1\Big(\|\mathbf{g}_{xy}\|^2 > \eta \;\vee\; \|\mathbf{v}_{xy}\| > v_{\mathrm{th}} \Big)\,, \tag{22}$$

Table 11: **Reward terms** used in our MorFiC locomotion objective. w_k are set in the reward configuration; fixed scalings appearing in code (e.g., the factor 4 in r_{lin}) are shown explicitly.

Term	Definition	Weight
r_{lin}	$4\exp\left(-\frac{\|\mathbf{v}_{xy}^{cmd}-\mathbf{v}_{xy}\|^2}{\sigma_{\text{lin}}}\right)$	w_{lin}
r_{yaw}	$\exp\left(-\frac{(\omega_z^{cmd}-\omega_z)^2}{\sigma_{\text{yaw}}}\right)$	w_{yaw}
p_{v_z}	$(v_z)^2$	w_{v_z}
$p_{\omega_{xy}}$	$\|\boldsymbol{\omega}_{xy}\|^2$	w_{ω}
p_{ori}	$\|\hat{\mathbf{g}}_{xy}\|^2$	w_{ori}
p_{τ}	$\|\boldsymbol{\tau}\|^2$	w_{τ}
$p_{\ddot{q}}$	$\left\|\frac{\dot{q}_{t-1}-\dot{q}_t}{\Delta t}\right\|^2$	$w_{\ddot{q}}$
$p_{\Delta a}$	$\|a_{t-1}-a_t\|^2$	$w_{\Delta a}$
p_{coll}	$\sum_{b\in\mathcal{B}}\mathbb{I}(\|\mathbf{f}_b\|>0.1)$	w_{coll}
p_{jl}	$\sum_i\left[(q_i-q_i^{\min})_- + (q_i-q_i^{\max})_+\right]$	w_{jl}
r_{upright}	$\exp\left(-\frac{\|\hat{\mathbf{g}}_{xy}\|^2}{\sin^2(\theta_0)}\right)$, θ_0=0.3	w_{upright}
r_{cop}	$\mathbb{I}(n_c\geq 2)\exp\left(-\frac{\|\text{CoP}-\text{center}\|}{\tau}\right)$, τ=0.10	w_{cop}
r_{width}	*(yaw-frame)* $\exp\left(-\frac{(w_{\min}-w)_+}{\sigma_w}\right)$ (gated)	w_{width}
r_{gaitF}	$-\frac{1}{4}\sum_i(1-d_i)\left(1-\exp(-f_i^2/\sigma_F)\right)$	w_{gaitF}
r_{gaitV}	$-\frac{1}{4}\sum_i d_i\left(1-\exp(-\|\dot{\mathbf{x}}_i\|^2/\sigma_V)\right)$	w_{gaitV}
p_{slip}	$\sum_i\mathbb{I}(c_i)\|\dot{\mathbf{x}}_{i,xy}\|^2$	w_{slip}
p_{impact}	$\sum_i\mathbb{I}(c_i)\,\text{clip}(\dot{z}_{i,t-1},-100,0)^2$	w_{impact}
p_f	$\sum_i\left(\|\mathbf{f}_i\|-f_{\max}\right)_+$	w_f
p_q	$\|q-q^0\|^2$	w_q
$p_{\dot{q}}$	$\|\dot{q}\|^2$	$w_{\dot{q}}$
r_{jump}	$-(z-(h^{cmd}+h_0))^2$	w_{jump}
r_{raibert}	$\|\mathbf{x}_{foot}^{des}-\mathbf{x}_{foot}\|^2$ (body yaw frame)	w_{raibert}

and define

$$s_{\text{width}} = \text{gate}\cdot\exp\left(-\frac{\max(0,\,w_{\min}-w)}{\sigma_w}\right)+(1-\text{gate}). \tag{23}$$

In addition to these components, we include a small set of rewards adapted from earlier work [28]. **All baselines were trained with the same reward**, so as to clearly isolate the performance difference caused by the poor approximation of the value function.

F Training Details

F.1 Random Seeds and Reporting

Unless stated otherwise, each internal PPO and MorFiC variant is trained with $n = 3$ independent random seeds. For simulation evaluation, each checkpoint is evaluated using the same target robots, command settings, rollout horizon, and number of parallel environments. Reported values are averaged across evaluation rollouts; standard deviations are reported over rollout-level measurements. For external baselines, we use the available released checkpoints or reproduced checkpoints as described in the corresponding baseline setup.

F.2 Simulation and Training Setup

All policies are trained in IsaacGym Simulator [29] with 4096 parallel environments with a time-step of dt = 0.005s; episode lasts roughly 20s. A command curriculum and domain randomization (joint friction, motor delays, sensor noise) are applied identically across all baselines. We used the same PPO hyperparameters, reward functions, command curriculum, morphology ranges, and network sizes across all baselines variants (See Appendix). And finally, each policy is trained a total of 400

Table 12: **External baselines vs. MorFiC at higher commanded speed.** Each entry reports $v_x^{\max} \pm \sigma / S$ (m/s; survival). All methods are evaluated at $v_x^{\mathrm{cmd}} = 2.5$ m/s, except GenLoco, which does not take a commanded forward velocity as input. Bold indicates the best value for each metric independently.

Target	URMA $v_x^{\max} \pm \sigma / S$	BodyTrans. $v_x^{\max} \pm \sigma / S$	GenLoco $v_x^{\max} \pm \sigma / S$	Embod. $v_x^{\max} \pm \sigma / S$	MorFiC-F $v_x^{\max} \pm \sigma / S$
Go2	$2.389 \pm 0.405 / \mathbf{1.000}$	$0.313 \pm 0.232 / 0.625$	$0.592 \pm 0.047 / 0.922$	$1.072 \pm 0.092 / 0.859$	$\mathbf{2.405 \pm 0.405} / 0.859$
Go1	$\mathbf{2.514 \pm 0.430} / \mathbf{1.000}$	$0.206 \pm 0.151 / \mathbf{1.000}$	$0.985 \pm 0.007 / \mathbf{1.000}$	$2.343 \pm 0.131 / 0.797$	$1.972 \pm 0.442 / 0.516$
A1	$\mathbf{2.230 \pm 0.336} / \mathbf{1.000}$	$0.374 \pm 0.183 / 0.515$	$0.919 \pm 0.011 / \mathbf{1.000}$	$2.143 \pm 0.116 / 0.719$	$1.595 \pm 0.467 / 0.156$
AlienGo	$- / 0.000$	$0.358 \pm 0.315 / 0.203$	$\mathbf{0.964 \pm 0.032} / \mathbf{1.000}$	$0.452 \pm 0.085 / 0.000$	$0.832 \pm 0.542 / \mathbf{1.000}$
B1	$- / 0.000$	$0.080 \pm 0.140 / 0.640$	$0.00 / 0.000$	$\mathbf{3.062 \pm 0.214} / \mathbf{0.797}$	$0.121 \pm 0.060 / 0.000$
Cheetah	$- / 0.000$	$- / \mathbf{1.000}$	$0.729 \pm 0.061 / 0.969$	$-0.118 \pm 0.091 / 0.000$	$\mathbf{1.128 \pm 0.986} / 0.234$
ANYmal-C	$\mathbf{2.159 \pm 0.354} / \mathbf{1.000}$	$0.014 \pm 0.081 / 0.843$	$0.708 \pm 0.005 / \mathbf{1.000}$	$0.359 \pm 0.150 / 0.453$	$0.303 \pm 0.422 / 0.984$

million time steps utilizing a Nvidia RTX 4090 laptop-based GPU, requiring roughly 2 hours of wall clock time. We have adapted our code mainly from open-source repositories [28, 3].

F.3 Domain Randomization

We apply domain randomization during training to improve robustness and sim-to-real transfer. Randomization is applied periodically with the interval 10 s (rigid-body properties may be randomized after initialization). We randomize the properties of contact with the ground by sampling the friction coefficient uniformly from $[0.5, 1.25]$. We inject external disturbances by pushing the robot at fixed intervals of 15 s with a maximum planar push velocity of 1.0 m/s. To model actuation/communication delay, we randomize action latency using a history buffer of up to 6 timesteps (randomized lag timesteps enabled).

F.4 Evaluation Fairness

In order to maintain the fairness of the evaluation, we kept all the necessary parameters like morphological and domain randomization parameters for the training robot, including all the PPO hyperparameters, command curriculum ranges and reward functions, latent dimensions the same across all the variants.

G Additional Transfer Results

G.1 Additional Results MorFiC Vs. external baselines

Table 12 reports an additional evaluation at a higher commanded forward speed, $v_x^{cmd} = 2.5$ m/s. This setting tests whether the transfer trends from the main 2.0 m/s evaluation persist under a more aggressive command. MorFiC-F remains competitive on the source and nearby morphologies, while performance degrades on more distant targets such as B1 and ANYmal-C. As in the main results, survival is interpreted jointly with achieved forward velocity, since high survival with negligible motion does not indicate successful transfer.

H Extended Critic Calibration Diagnostics

The main paper reports explained variance as the primary critic-calibration metric. Here, we provide additional diagnostics that capture complementary aspects of value-function quality. Specifically, we report $|\mathrm{NVB}|$, NVE, and ASA for both internal PPO variants and external baselines. Lower $|\mathrm{NVB}|$ indicates smaller systematic value bias, lower NVE indicates smaller value prediction error, and higher ASA indicates better agreement between critic-induced and Monte-Carlo advantage signs.

Table 13 compares the internal PPO variants. The results show that additive morphology conditioning does not consistently resolve critic error across morphologies. FiLM improves the strongest OOD cases such as AlienGo and Cheetah, where it achieves lower NVE and higher ASA than the

Table 13: **Extended critic calibration diagnostics across PPO variants.** Each entry reports |NVB|/NVE/ASA, where NVB is the normalized value bias, NVE is the normalized value error, and ASA is the advantage-sign agreement. Lower |NVB| and NVE are better, while higher ASA is better. Best value for each metric within a target row is bolded.

Target	Vanilla PPO |NVB|/NVE/ASA	Concat-z Actor only	Concat-z Actor+Critic	FiLM |NVB|/NVE/ASA
Go2	1.944/1.981/0.610	**0.177**/1.168/0.689	0.535/**0.960**/0.670	1.768/1.943/**0.771**
Go1	1.591/1.817/0.612	**0.068**/**1.096**/0.662	1.623/1.834/0.674	1.375/1.653/**0.681**
A1	2.027/2.203/0.603	2.033/16.463/**0.754**	**0.199**/**0.936**/0.701	1.482/1.813/0.675
AlienGo	0.070/1.454/0.612	0.468/1.209/0.569	0.153/1.117/0.607	**0.048**/**0.835**/**0.706**
B1	0.182/1.231/0.618	0.358/1.187/0.570	**0.040**/1.134/**0.624**	0.645/**1.279**/0.618
Cheetah	0.362/1.869/0.428	**0.220**/1.086/0.638	1.085/1.344/0.730	0.464/**0.947**/**0.768**
ANYmal-C	**0.091**/1.229/0.514	0.411/**1.227**/0.528	0.385/1.307/**0.620**	0.120/1.229/0.578

Table 14: **Extended critic calibration diagnostics for external baselines.** Each entry reports |NVB|/NVE/ASA, where NVB is normalized value bias, NVE is normalized value error, and ASA is advantage-sign agreement. Lower |NVB| and NVE are better, while higher ASA is better. Best value for each metric within a target row is bolded.

Target	Embod. Scale |NVB|/NVE/ASA	URMA |NVB|/NVE/ASA	BodyTrans. |NVB|/NVE/ASA	GenLoco |NVB|/NVE/ASA	MorFiC-F |NVB|/NVE/ASA
Go2	3.437/3.646/0.657	3.005/3.178/**0.702**	3.319/3.492/0.652	5.458/5.674/0.561	**1.768**/**1.943**/0.671
Go1	2.569/2.772/0.607	3.170/3.333/0.692	4.771/4.983/**0.724**	3.402/3.594/0.623	**1.375**/**1.653**/0.681
A1	2.368/2.585/0.599	3.390/3.551/**0.686**	3.853/4.032/0.686	2.956/3.128/0.612	**1.482**/**1.813**/0.675
AlienGo	–	–	3.104/3.476/0.661	1.170/1.522/**0.709**	**0.098**/**0.835**/0.706
B1	3.330/3.590/0.657	–	3.571/3.781/**0.694**	7.428/7.729/0.667	**0.645**/**1.279**/0.618
Cheetah	–	–	–	5.294/5.511/0.594	**0.464**/**0.947**/**0.768**
ANYmal-C	3.589/3.812/0.667	2.538/2.740/0.728	3.368/3.529/0.674	1.107/2.324/**0.733**	**0.120**/**1.229**/0.578

vanilla and concatenation-based variants. This supports the claim that multiplicative critic conditioning improves the quality of the value baseline under morphology shift.

Table 14 reports the same diagnostics for external baselines. Although some external methods achieve reasonable locomotion performance on certain robots, their value predictions often remain poorly calibrated, with large |NVB| and NVE. MorFiC-F generally gives substantially lower value error, especially on OOD targets such as AlienGo, B1, Cheetah, and ANYmal-C. ASA is more mixed, so we interpret it jointly with |NVB| and NVE rather than as a standalone measure.

I Fixed-Rollout Critic Causality Analysis

To test whether morphology conditioning improves the actor through better critic estimates, we run a fixed-rollout diagnostic. For each trained policy, we freeze the sampled trajectories and compare the critic-induced advantage signs against an MC-return reference. We report the sign-flip rate, where lower values indicate fewer wrong-sign policy updates, and the gradient cosine against the MC-reference actor gradient, where higher values indicate better update alignment.

MorFiC generally yields lower flip rates and higher gradient alignment on near-support targets such as Go2, A1, AlienGo, and Cheetah, suggesting that FiLM conditioning improves the critic signal used by PPO. However, on distant morphologies such as B1 and ANYmal-C, the cosine can become negative, exposing a failure mode where the learned critic induces updates that are misaligned with the MC reference. This supports our diagnosis that cross-embodiment failure is not only an actor-transfer problem, but also a critic-miscalibration problem. (see Table 15)

J Training Compute Methodology

We report transition-normalized GPU-hours because the baselines use different training budgets. For each method, we profile the released training code on our local hardware for 100 iterations,

Table 15: **Fixed-rollout critic diagnostic across target morphologies.** Each entry reports Flip/Cos., where Flip is the sign-flip rate against the MC reference advantage (lower is better) and Cos. is the actor-gradient cosine with the MC-reference gradient (higher is better). Bold indicates the best learned method for each metric independently; the MC row is the reference and is not considered for bolding.

Src.	Go2		Go1		A1		AlienGo		B1		Cheetah		ANYmal-C	
	Flip	Cos.	Flip	Cos.	Flip	Cos.	Flip	Cos.	Flip	Cos.	Flip	Cos.	Flip	Cos.
MC	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
MorFiC	**0.227**	0.880	0.187	**0.972**	**0.239**	**0.785**	**0.255**	**0.366**	0.272	-0.182	**0.209**	**0.679**	**0.402**	-0.149
Concat	0.307	0.720	**0.151**	0.830	0.344	0.427	0.329	0.004	0.269	0.106	0.236	0.575	0.479	-0.419
Act-only	0.524	**0.890**	0.502	0.924	0.511	0.717	0.318	0.342	0.565	**0.647**	0.265	0.644	0.470	**-0.044**
Vanilla PPO	0.235	0.865	0.229	0.684	0.260	0.653	0.322	-0.113	**0.176**	0.601	0.253	0.264	0.461	-0.074

Table 16: **Training compute comparison.** We report total transitions, approximate PPO iterations, wall-clock seconds per iteration, total single-GPU-equivalent training hours, and GPU-hours normalized to a 400M-transition budget. Lower GPU-hours@400M indicates lower compute cost per transition. Expert-pool methods are separated because they train many specialized policies rather than one generalist policy.

Method	Transitions	Iters.	Sec/iter	GPU-h	GPU-h @400M	Notes
MorFiC	400M	4,000	**1.98–2.70**	**2.2–3.0**	**2.2–3.0**	Single generalist policy; local run
BodyTransformer	400M	8,000	17.25	38.3	39.0	Reproduced `model_8000`
URMA	1.600B	3,064	366–392	311–333	77.8–83.3	Original training budget
GenLoco	200M	1,526	125.8	53.3	106.6	200M setting
GenLoco	800M	6,104	125.8	213.5	106.8	800M setting
Expert-pool / selected-expert methods						
Embod. Scaling, Gendog8 expert	1.720B	17,499	0.94	4.5	1.05	One selected released expert
Embod. Scaling, full Gendog pool	570.6B	∼5.80M	0.94	1507	1.06	332 experts; ∼5 TB storage

measure the average wall-clock time per iteration, and compute the transitions per iteration from the method configuration. We then extrapolate total GPU-hours to the reported transition budget:

$$\text{GPU-hours} = \frac{T}{N_{\text{env}} H} \cdot \frac{\tau_{\text{iter}}}{3600},$$

where T is the total transition budget, N_{env} is the number of parallel environments, H is the rollout horizon, and τ_{iter} is the measured time per iteration. We also report GPU-hours normalized to 400M transitions for fair per-transition comparison. For BodyTransformer, we reproduce training to approximately the MorFiC budget; for larger-budget methods, we extrapolate using their reported transition counts.

K Real-Robot Rollout Diagnostics

To examine whether the learned critic remains meaningful during real deployment, we record reward and value trajectories from Go1 hardware rollouts. Figure 5 shows that MorFiC produces smoother value estimates that follow the trend of the discounted return despite noisy instantaneous rewards. In contrast, the non-MorFiC critic exhibits higher variance and less stable predictions. This suggests that morphology-conditioned value estimation can improve critic stability beyond simulation and remains useful under real-world rollout noise.

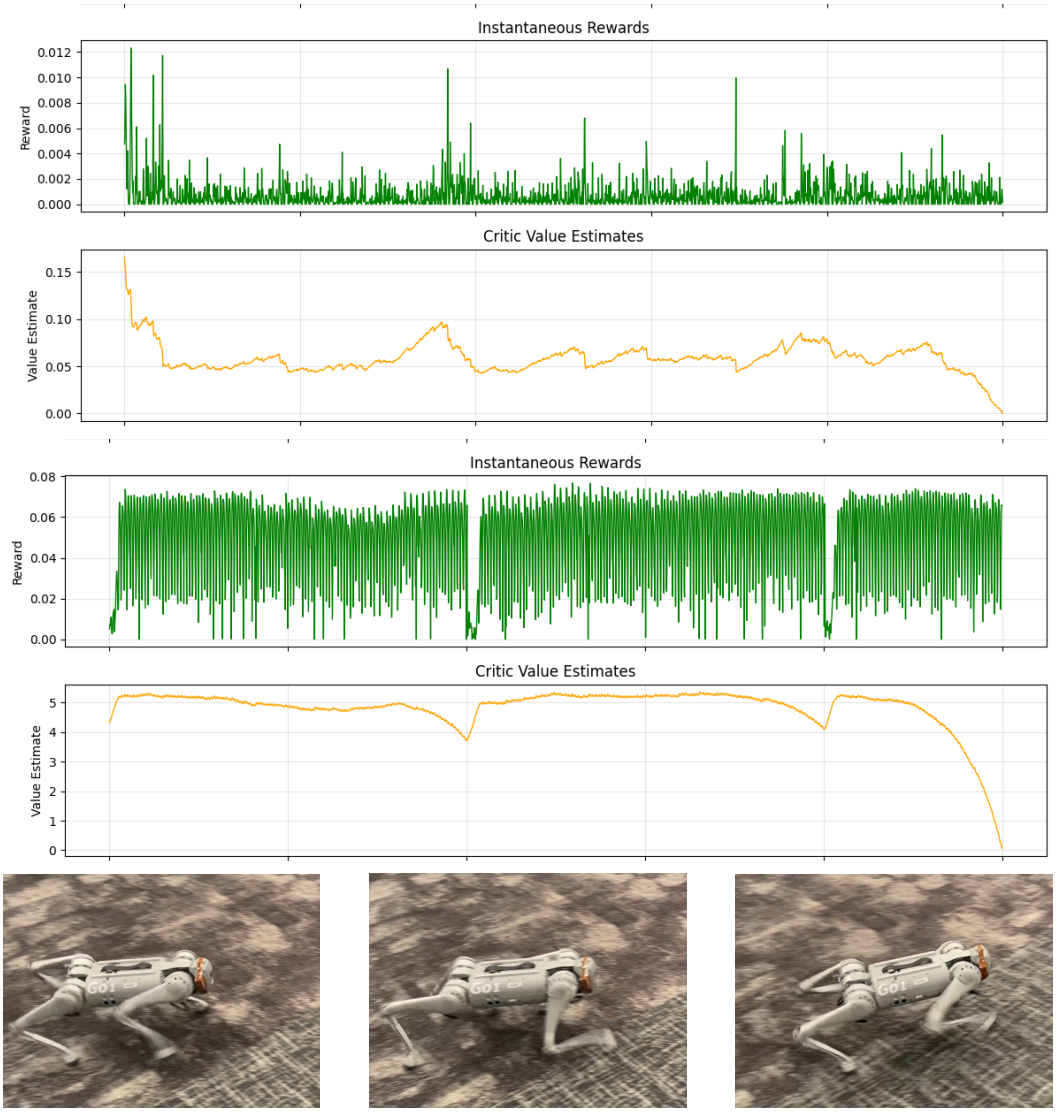


Figure 5: Above figure shows representative reward and value trajectories during deployment on the Go1 robot. The MorFiC produces smooth value estimates (orange) (second row) that track the expected discounted returns from noisy instantaneous rewards (green). While non-MorFic value estimates are noisy (first row), MorFiC's morphology-conditioned critic maintains stable value predictions even on unseen morphologies.