Preference-Guided Diffusion for Embodied Robot Morphology Evolution

1st Liming Xin
Shanghai University
Shanghai, China
xinliming@shu.edu.cn

2nd Yaowen Zhang
Shanghai University
Shanghai, China
iwannnn@shu.edu.cn

3rd Bin Sheng
Shanghai University
Shanghai, China
shengbin@shu.edu.cn

4th Hao Kong*
Shanghai University
Shanghai, China
haokong@shu.edu.cn

***Abstract*—Morphology optimization is critical for embodied robots but remains challenging due to the vast design space and the high cost of evaluating candidates with policy training. Recent diffusion-guided evolutionary methods improve sample efficiency by learning morphology priors. However, they typically rely on thresholded binary preference supervision and a fixed mutation-diffusion ratio. These choices yield coarse guidance and discard relative performance information, thereby failing to adapt to changing search dynamics. To address these limitations, we propose Preference-Guided Diffusion Evolution (PGDE), a framework that improves guidance reliability and search efficiency within a unified evolutionary loop. PGDE introduces rank-normalized soft preference learning to provide smoother targets and finer-grained guidance signals. This is complemented by a confidence-aware guidance mechanism that scales guidance strength based on the uncertainty of preference predictions to prevent misdirected sampling. To better match exploration and refinement across search stages, PGDE adopts an adaptive mutation-diffusion ratio that dynamically balances mutation-based local search and diffusion-based global search through online fitness feedback. Experiments on diverse EvoGym tasks show that the proposed framework consistently achieves higher reward compared to baseline approaches.**

***Index Terms*—Embodied intelligence, morphological intelligence, evolutionary algorithm, diffusion model**

I. INTRODUCTION

Biological organisms exhibit sophisticated embodied intelligence, with long-term morphological evolution progressively aligning their physical structures to environmental conditions and functional requirements [1]–[3]. Inspired by this principle, embodied robot design increasingly treats morphology optimization as a key mechanism for improving performance across diverse tasks. The design of embodied robots typically necessitates the joint optimization of morphology and control. Since morphology dictates system dynamics and action feasibility, the controller typically needs to be re-optimized through computationally expensive reinforcement learning. This tight coupling expands the combinatorial design space and makes candidate evaluation computationally prohibitive.

Traditional approaches [4], [5] commonly use population-based evolutionary algorithms to optimize morphologies across generations. In these methods, each candidate morphology is evaluated by training a task-specific controller and assigning a fitness value based on the resulting performance. Evolutionary algorithms naturally handle discrete and constrained design spaces, but their reliance on random mutation often leads to sample inefficiency because minor morphological changes can invalidate previously learned policies and necessitate costly policy retraining. While recent RL-based methods attempt to optimize morphology and control jointly [6]–[8], they often suffer from instability due to the non-stationary dynamics induced by continuous morphological updates. Diffusion models have emerged as a powerful tool for learning morphology priors and generating feasible structures, and Liu et al. [9] further introduced classifier guidance to bias diffusion sampling toward higher predicted fitness. Despite these advances, diffusion-guided evolution still relies on thresholded binary fitness labels and fixed operator mixing, discarding relative performance information and failing to adapt the balance between local search and global search as evolution progresses.

To address these limitations, we propose Preference-Guided Diffusion Evolution (PGDE), an evolutionary framework that improves the reliability of diffusion guidance and dynamically coordinates mutation and diffusion. PGDE introduces rank-normalized soft preference learning to train a preference regressor for scoring candidate morphologies. Instead of hard binary labels, PGDE uses continuous targets derived from relative fitness rankings, providing smoother supervision and reducing sensitivity to noisy evaluations. The regressor then guides diffusion sampling toward high-fitness morphologies. PGDE also incorporates a confidence-aware guidance mechanism that scales the guidance strength according to predictive uncertainty, reducing the influence of unreliable predictions. In addition, PGDE adopts an adaptive mutation-diffusion ratio driven by online fitness feedback, which reallocates evaluations between mutation-based local search and diffusion-based global search as evolution progresses. We evaluate PGDE on diverse EvoGym morphology optimization tasks and observe consistent reward improvements over strong evolutionary and diffusion-guided baselines.

The main contributions of this work are summarized as follows:

- We propose Preference-Guided Diffusion Evolution (PGDE), a diffusion-guided evolutionary framework for morphology optimization that improves guidance relia-

*Corresponding author: Hao Kong (haokong@shu.edu.cn).

bility and coordinates mutation and diffusion within a unified search loop.
- We introduce rank-normalized soft preference learning to train a preference regressor with relative-ranking targets and augment it with confidence-aware guidance to down-weight uncertain signals during diffusion sampling.
- We develop an adaptive mutation–diffusion ratio driven by online fitness feedback to balance mutation-based refinement and diffusion-based exploration, achieving consistent gains on diverse EvoGym tasks over evolutionary and diffusion-guided baselines.

II. RELATED WORK

The joint optimization of robot morphology and control has long been recognized as a fundamental problem in embodied intelligence. Pfeifer et al. highlight the impact of morphology on motion, learning, and performance, suggesting that body design is integral to intelligent behavior [10].

Early studies mainly relied on population-based evolutionary algorithms to explore morphology spaces. Lipson et al. show that mutation and recombination can generate functional robot bodies without explicit human design [11]. This line of work has been extended to richer morphology spaces and more complex environments [2], [5], [12]. Several variants improve exploration efficiency by introducing structured representations. Grammar-based design constrains morphology generation with compositional rules and reduces invalid proposals and redundancy [4], [13]. Modular frameworks assemble morphologies from reusable components and enable scalable search over structured subspaces [14], [15]. Foundation model-based strategies further learn stronger priors from data to guide variation, improving offspring quality and search efficiency [5]. Despite these advances, evolutionary approaches remain evaluation intensive because each candidate still requires controller training, and even small morphology changes can necessitate costly re-optimization [16].

Another line of work introduces learning to assist morphology optimization. David Ha explores reinforcement learning to update morphology-related parameters together with control policies using task feedback [7], and Nygaard et al. further demonstrate real-world morphological adaptation driven by learning [17], [18]. Graph-based architectures also encode embodiment structure to improve policy generalization across morphologies [19], [20]. These approaches reduce reliance on purely black-box evolution, but scaling to large discrete morphology spaces remains difficult. Training can also be unstable because morphology updates change contact patterns and transition dynamics.

Building on learned priors, diffusion models have been explored to accelerate morphology search by generating feasible structures. Liu et al. propose CDMEO, which combines a task-independent diffusion model with task preference guidance during sampling [9]. CDMEO pretrains a diffusion model on feasible morphologies. It then trains a task-specific classifier online to guide sampling toward high-fitness designs, improving the efficiency of proposing feasible offspring over purely evolutionary search.

However, existing diffusion-guided evolution still faces two limitations. Guidance models are commonly trained with hard or binary labels, which can yield unreliable gradients when preference signals are noisy and weakly separated. Diffusion-generated offspring are also mixed with mutation using a fixed mutation–diffusion ratio, which limits stage-adaptive allocation between mutation and diffusion. Unlike existing approaches, we improve preference modeling with rank-normalized soft targets and confidence-scaled guidance. We also adjust the mutation–diffusion ratio based on online fitness feedback to yield a more efficient and robust search.

III. METHOD

A. Overview of PGDE

The morphology search space is large and evaluating each morphology is expensive. Diffusion-based generators mitigate this cost by sampling feasible candidates from a learned prior and can be guided toward high-fitness designs. However, existing diffusion-guided evolution often relies on thresholded binary preference supervision and a fixed mutation–diffusion allocation. This design discards relative performance information and limits adaptation of the exploration–refinement trade-off. To address these limitations, we propose Preference-Guided Diffusion Evolution (PGDE), a population-based evolutionary framework that improves guidance reliability and adaptively coordinates mutation and diffusion within a unified evolutionary loop, as summarized in Fig. 1.

PGDE updates an online preference model from evaluated morphologies and uses it to guide diffusion-based offspring generation. Specifically, PGDE constructs rank-normalized soft targets and trains a preference regressor to predict relative quality. The regressor also provides an uncertainty estimate. We use it to scale the guidance strength and stabilize diffusion sampling under immature preference learning or noisy fitness. Guided by this confidence-scaled signal, a pretrained diffusion model generates diverse offspring to support global exploration, while mutation perturbs selected parents to support local refinement. Offspring from the two operators are then merged into a single candidate set, corresponding to the generated offspring stage in Fig. 1. The merged candidates are subsequently evaluated and subjected to environmental selection to update the population for the next generation.

B. Soft Preference Learning for Diffusion Guidance

In diffusion-guided morphology evolution, a fitness-trained model is used to guide diffusion sampling toward high-performing morphologies. However, in early evolutionary stages the fitness values of candidate morphologies are often tightly clustered, so thresholding them into binary labels can amplify minor differences into contradictory 0/1 supervision and yield unstable gradients. This issue is further aggravated by task-dependent reward scales. Therefore, PGDE trains the guidance model with rank-normalized soft preference

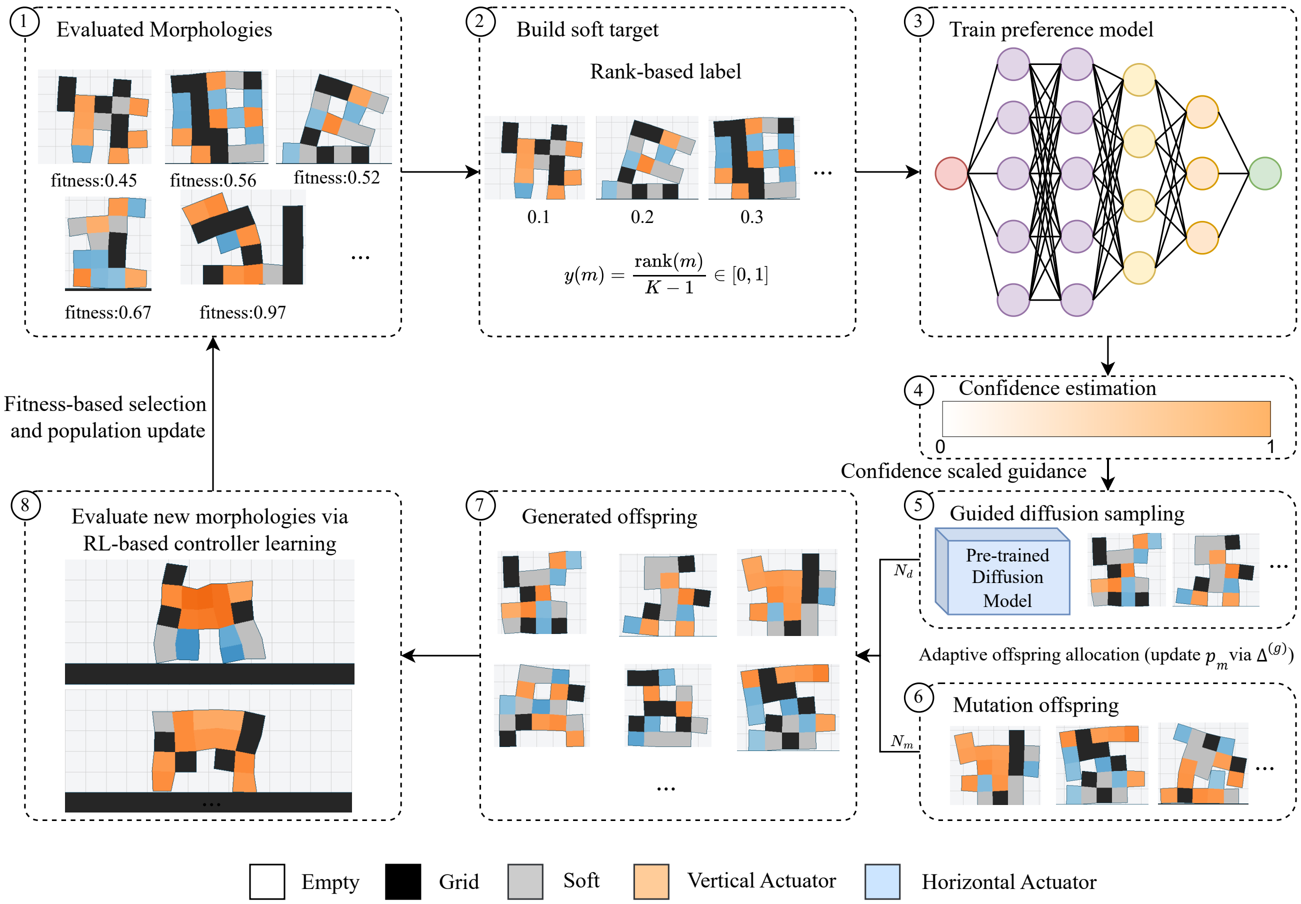

Fig. 1. Overview of PGDE. A task-agnostic diffusion model ϵ_θ provides a structural prior over feasible morphologies. During evolution, evaluated morphologies supervise an online soft preference regressor s_ϕ that guides diffusion sampling. An adaptive allocation mechanism updates p_m based on the best fitness improvement $\Delta^{(g)}$, and determines N_m and N_d for the next generation.

targets derived from within-generation comparisons to provide smoother supervision and preserve relative performance information. PGDE further introduces confidence-aware guidance to downweight uncertain predictions during sampling, which stabilizes diffusion guidance under noisy evaluations and early-stage preference learning.

In PGDE, diffusion guidance is trained and applied online within the evolutionary loop. At each generation, we collect evaluated morphologies and their fitness values and convert fitness feedback into soft preference targets via rank normalization:

$$y_i = \frac{\mathrm{rank}(f_i)}{K-1}, \quad y_i \in [0,1]. \tag{1}$$

Here, K is the batch size, f_i is the fitness of morphology m_i, and $\mathrm{rank}(f_i) \in \{0,\dots,K-1\}$ is the within-batch rank with larger values indicating better performance. This normalization yields smooth supervision that is less sensitive to noisy evaluations and reward-scale variations.

We train an online preference regressor conditioned on diffusion time to account for the distribution shift across diffusion steps, from heavily corrupted states at early steps to near-feasible samples at late steps. The regressor predicts a scalar preference score in $[0,1]$.

$$s_\phi(x_t,t) : \mathbb{R}^{1\times H\times W} \times \{0,\dots,T\} \to [0,1]. \tag{2}$$

Here, x_t is the noisy morphology state at diffusion step t, T is the total number of diffusion steps, and ϕ denotes the regressor parameters.

The preference regressor is trained using diffusion-corrupted inputs and rank-normalized soft targets as follows:

$$x_t^{(i)} = \sqrt{\bar{\alpha}_t}\, x_0^{(i)} + \sqrt{1-\bar{\alpha}_t}\,\epsilon, \tag{3}$$

$$\mathcal{L}_{\text{soft}}(\phi) = \mathbb{E}_{(x_0^{(i)},y_i),\,t}\left[\left\| s_\phi(x_t^{(i)},t) - y_i \right\|_2^2\right]. \tag{4}$$

Here, $x_0^{(i)}$ denotes the normalized continuous representation of morphology m_i, $\epsilon \sim \mathcal{N}(0,I)$, and $\bar{\alpha}_t$ is the cumulative noise schedule coefficient. This objective yields stable gradients for online guidance learning.

To prevent unreliable regressor outputs in early evolution from producing overconfident guidance, we introduce a confidence-aware scaling mechanism. The key idea is that the regressor output should influence reverse diffusion strongly only when it is both directional and trustworthy. Since $s_\phi(x_t, t) \in [0, 1]$ can be interpreted as a soft preference score, its uncertainty can be characterized without training an additional variance head. Specifically, predictions near 0.5 indicate ambiguity between favorable and unfavorable morphologies, whereas predictions near 0 or 1 indicate a clearer preference signal. We therefore compute the guidance gradient from the regressor and modulate its strength using an entropy-based confidence score:

$$g_t = \nabla_{x_t} s_\phi(x_t, t), \tag{5}$$

$$\mathrm{conf}(x_t, t) = 1 - \frac{H\left(s_\phi(x_t, t)\right)}{\log 2}, \tag{6}$$

$$H(p) = -p \log p - (1-p)\log(1-p). \tag{7}$$

Here, $H(\cdot)$ is the binary entropy and $\mathrm{conf}(x_t, t) \in [0, 1]$ measures prediction confidence, with larger values indicating more reliable guidance. This choice is attractive because it is monotonic, bounded, and aligned with the semantics of a scalar preference output: the entropy is maximal at $p = 0.5$, where the regressor is most uncertain, and decreases smoothly as the prediction becomes more decisive. The final guidance strength is

$$\tilde{\lambda}_t = \lambda \cdot \max(\rho, \mathrm{conf}(x_t, t)), \tag{8}$$

where λ is the base guidance scale. The floor parameter ρ prevents guidance from vanishing completely when the regressor is still immature, so diffusion can retain a weak but nonzero task signal instead of degenerating to purely unguided sampling. In practice, we also clip $s_\phi(x_t, t)$ to $[\epsilon, 1-\epsilon]$ before evaluating entropy. This avoids numerical instability from $\log 0$ and prevents extremely saturated predictions from producing brittle confidence estimates. Finally, we apply confidence-aware guidance only when $t \le t_{\text{thresh}}$. At very early reverse-diffusion steps, samples are heavily corrupted and the preference regressor is less calibrated; delaying guidance to later steps makes the signal act on partially denoised structures, where task-relevant morphology cues are more reliable.

The proposed soft-target supervision and confidence-aware guidance form a unified pipeline for preference learning and diffusion sampling, improving guidance stability and robustness under noisy fitness evaluations.

C. Adaptive Mutation–Diffusion Ratio

Besides diffusion-based global search that proposes diverse candidates and is steered by preference guidance, mutation performs local search by perturbing selected parents. Since their utility varies across evolutionary stages, a fixed mixing ratio can waste evaluations by misbalancing exploration and refinement. PGDE therefore adapts the mutation–diffusion ratio using a lightweight fitness-improvement signal. It shifts allocation toward diffusion when progress stagnates to promote exploration, and toward mutation when rewards improve steadily to emphasize refinement. This stage-adaptive allocation improves sample efficiency and strengthens anytime performance under a fixed evaluation budget.

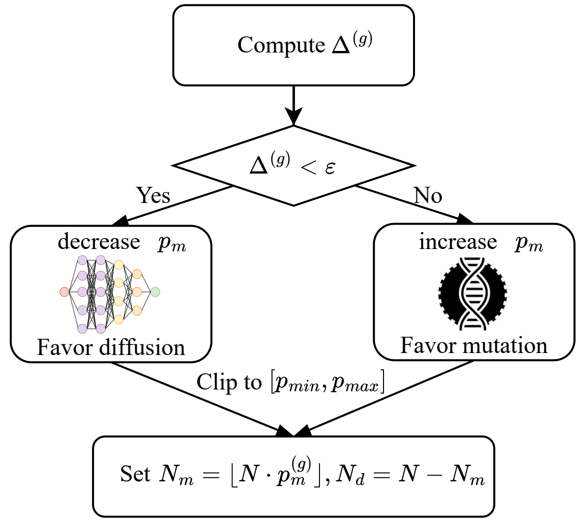

Fig. 2. Adaptive update of the mutation ratio. When improvement stagnates ($\Delta^{(g)} < \varepsilon$), we decrease p_m to favor diffusion; otherwise we increase p_m to favor mutation. The resulting p_m determines N_m and N_d under a fixed budget N.

As illustrated in Fig. 2, PGDE updates the mutation–diffusion ratio once per generation based on the best-fitness improvement over a sliding window. The mutation ratio is bounded within a predefined range, and the diffusion ratio simply takes the remaining share of the budget. We track the best fitness of the current population and measure improvement over a sliding window:

$$b^{(g)} = \max_{m \in \mathcal{P}^{(g)}} f(m), \tag{9}$$

$$\Delta^{(g)} = b^{(g)} - b^{(g-k)}, \quad g \ge k. \tag{10}$$

For the first few generations, the initial ratio is kept unchanged. The update rule is:

$$p_m^{(g+1)} = \begin{cases} \max(p_{\min},\, p_m^{(g)} - \eta), & \Delta^{(g)} < \varepsilon, \\ \min(p_{\max},\, p_m^{(g)} + \eta), & \Delta^{(g)} \ge \varepsilon, \end{cases} \tag{11}$$

and the two operators generate offspring under a fixed per-generation budget:

$$N_m = \lfloor N \cdot p_m^{(g)} \rfloor, \quad N_d = N - N_m. \tag{12}$$

Here, k is the window size, η is the update step, and ε is the stagnation threshold, while $p_{\min}$ and $p_{\max}$ specify the ratio bounds. When improvement is small, the rule reduces $p_m^{(g)}$ and increases the diffusion share, which strengthens global exploration and helps escape local optima. When improvement is sustained, it increases $p_m^{(g)}$, which focuses evaluations on local refinement around strong morphologies.

Overall, the adaptive ratio provides a simple way to allocate expensive evaluations between exploration and refinement as the search progresses. When improvement stalls, it shifts more budget to diffusion sampling to explore new regions and escape local optima. When improvement is steady, it shifts more budget to mutation to refine strong morphologies. This makes offspring generation more efficient under a fixed evaluation budget. The resulting offspring are then evaluated by controller training and fed back into the evolutionary loop for selection in the next generation.

With the proposed soft preference learning, confidence-aware guidance, and adaptive operator allocation, PGDE improves the reliability and efficiency of diffusion-guided evolution under a fixed evaluation budget. Soft preferences provide smoother guidance than binary supervision, stabilizing preference learning under noisy fitness. Confidence-aware guidance and adaptive ratio control further reduce misdirected sampling and balance exploration with refinement as the search progresses. Together, these components produce higher-quality offspring and improve overall sample efficiency.

IV. EXPERIMENTS

A. Benchmark

We evaluate our methods on Evolution Gym (EvoGym) [21], a standard benchmark for voxel-based robot morphology and controller co-optimization. EvoGym represents each morphology as a discrete multi-material voxel grid. In our setting, each morphology is encoded as a 5×5 grid, where each cell is one of the available voxel types, including rigid, soft, horizontal actuator, vertical actuator, and empty. For each candidate morphology, we train a task controller using PPO [22], and use the resulting episodic return as the fitness value for morphology evaluation.

We evaluate PGDE on 9 EvoGym tasks covering diverse locomotion and manipulation objectives. We follow prior benchmarks and use evaluation budgets of 100, 150, and 200 fitness evaluations for easy, medium, and hard tasks, respectively. Our methods and baselines are run with 5 random seeds on each task, and we report the mean and standard deviation across seeds. For a fair comparison, all methods use the same PPO training budget and termination criteria for each morphology evaluation. We adopt the standard EvoGym PPO implementation and keep its settings fixed across all methods. Moreover, PGDE and CDMEO share the same pretrained diffusion prior, morphology encoding, population size, and per-generation evaluation budget; they differ only in preference supervision and offspring allocation, so the comparison isolates the effect of the proposed guidance and scheduling components.

B. Comparison with Baselines

We compare PGDE with three representative baselines, including a genetic algorithm (GA), an indirect encoding method (CPPN-NEAT) [23], and a diffusion-guided evolutionary method (CDMEO). GA searches voxel morphologies using selection and stochastic mutation, and evaluates each candidate by training a task controller and using the episodic return as fitness. CPPN-NEAT evolves an indirect encoding that maps voxel coordinates to element types, and we use the default EvoGym configuration. CDMEO is a recent diffusion-guided evolutionary method that injects diffusion-sampled feasible morphologies into the standard evolutionary loop and uses classifier guidance to bias sampling toward high-performing designs.

Table I reports results on nine EvoGym tasks grouped by difficulty. PGDE achieves the best performance on eight tasks and improves the average reward by 4.45% over CDMEO. Relative to the strongest baseline on each task, the average improvement is 3.76%, and it reaches 4.64% when averaged over the eight tasks where PGDE ranks first. The gains can be attributed to how PGDE addresses the main bottlenecks of diffusion-guided evolution. Preference-guided diffusion reduces invalid or low-quality proposals by sampling from a feasibility-aware prior. Soft preference learning and confidence-aware guidance stabilize the guidance signal, which improves task alignment during sampling. Adaptive ratio control further improves offspring generation efficiency by increasing diffusion when progress stagnates and increasing mutation when improvement is steady.

PGDE does not rank first on ObstacleTraverser-v0, where successful bodies often require abrupt topology changes to clear environmental barriers. In this setting, the online preference model is learned from a relatively small pool of evaluated morphologies, so a smoother guidance signal can occasionally favor conservative but feasible proposals over more exploratory candidates. This result indicates that PGDE is most beneficial when relative fitness ordering is informative enough to support stable soft guidance, rather than uniformly dominating every search landscape.

C. Ablation Experiment

We ablate three components of PGDE to quantify their contributions, including confidence modeling (Conf.), the adaptive mutation diffusion ratio (Adaptive Ratio), and soft preference learning (Soft Pref.). We evaluate four EvoGym tasks under the same evaluation budget and report mean and standard deviation over 5 random seeds in Tab. II. We further analyze the effect of ratio schedules in Fig. 3.

Soft preference learning is the most influential component. Removing it causes consistent drops on all tasks. The performance decreases by 0.70 on BridgeWalker-v0, 0.90 on Pusher-v0, 0.55 on DownStepper-v0, and 0.46 on BidirectionalWalker-v0. These results reflect our design choice in supervision. We replace hard thresholding with rank-normalized continuous targets. This preserves relative ordering inside each evaluated batch. It also reduces label noise caused by small return fluctuations. As a result, the preference regressor receives smoother targets and produces more reliable guidance for diffusion sampling.

Confidence modeling provides an additional and consistent gain. Compared with PGDE w/o Conf., PGDE improves by 0.52 on BridgeWalker-v0, 0.32 on Pusher-v0, 0.37 on DownStepper-v0, and 0.52 on BidirectionalWalker-v0. This improvement comes from how confidence is used in sampling. We scale the guidance strength by an uncertainty-aware confidence estimate. When the regressor output is uncertain, the guidance gradient is downweighted. This prevents noisy

TABLE I
PERFORMANCE COMPARISON ACROSS TASKS (MEAN(STD) OVER 5 SEEDS). BOLD INDICATES THE BEST RESULT ON EACH TASK.

Tasks	Levels	GA	CPPN-NEAT	CDMEO	PGDE (Ours)
BridgeWalker-v0	Easy	3.76(0.12)	3.98(0.53)	4.42(0.78)	**4.79(0.86)**
Pusher-v0		5.78(0.44)	6.87(0.46)	6.47(0.83)	**6.99(0.91)**
DownStepper-v0		4.68(0.48)	4.72(0.32)	5.34(1.04)	**5.64(0.72)**
BidirectionalWalker-v0	Medium	4.25(1.08)	4.31(0.54)	4.92(0.44)	**4.99(0.96)**
UpStepper-v0		3.06(0.42)	3.11(0.28)	3.19(0.58)	**3.45(0.76)**
ObstacleTraverser-v0		2.97(0.56)	2.67(0.31)	**3.28(0.41)**	3.17(0.56)
BeamSlider-v0	Hard	1.87(0.28)	1.95(0.15)	2.34(0.30)	**2.45(0.34)**
PlatformJumper-v0		2.38(0.67)	2.36(0.95)	2.72(0.69)	**2.86(0.21)**
GapJumper-v0		3.14(0.97)	2.91(0.05)	3.51(0.28)	**3.58(0.52)**

TABLE II
ABLATION STUDY OF PGDE ON FOUR EVOGYM TASKS (MEAN(STD) OVER 5 SEEDS). BOLD INDICATES THE BEST RESULT ON EACH TASK.

Tasks	PGDE w/o Conf.	PGDE w/o Adaptive Ratio	PGDE w/o Soft Pref.	PGDE (Ours)
BridgeWalker-v0	4.27(0.69)	4.52(0.95)	4.09(0.79)	**4.79(0.86)**
Pusher-v0	6.67(0.56)	6.85(0.56)	6.09(0.43)	**6.99(0.91)**
DownStepper-v0	5.27(0.55)	5.48(0.55)	5.09(0.87)	**5.64(0.72)**
BidirectionalWalker-v0	4.47(0.11)	4.40(0.98)	4.53(0.44)	**4.99(0.96)**

guidance from dominating the reverse diffusion trajectory. It improves the stability of diffusion-based generation and increases the chance of producing high-quality feasible offspring.

The adaptive mutation diffusion ratio improves both final performance and efficiency. In Tab. II, PGDE w/o Adaptive Ratio underperforms PGDE on all tasks, and the largest gap appears on BidirectionalWalker-v0 (4.40 vs 4.99). Fig. 3 explains why fixed ratios are suboptimal. A low mutation ratio such as $p_m = 0.3$ under-allocates mutation, which weakens refinement of promising morphologies and slows convergence. A high mutation ratio such as $p_m = 0.7$ under-allocates diffusion sampling, which reduces exploration and diversity and can trap the search in local optima. The adaptive strategy avoids committing to a single ratio: it decreases p_m when progress stagnates to increase exploration, and increases p_m when improvement is steady to strengthen refinement.

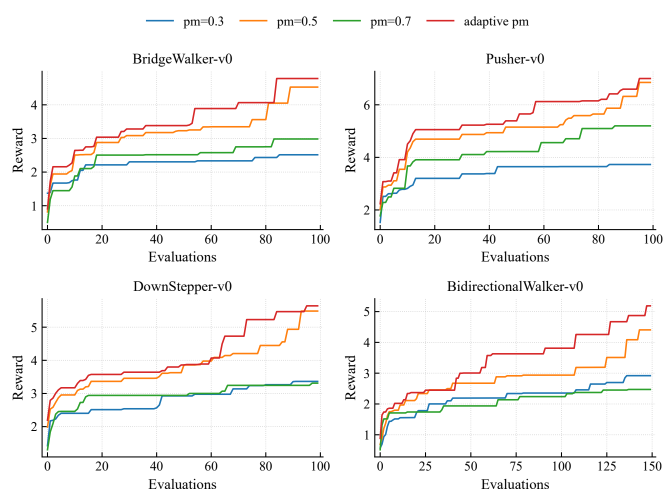

Fig. 3. Ablation on mutation ratio p_m across four tasks.

D. Sensitivity Analysis

We conduct a sensitivity analysis on the hyperparameters of the adaptive mutation–diffusion ratio controller. Specifically, we sweep the improvement tolerance ϵ and the update step size η. Together, these parameters control the stagnation criterion and the magnitude of ratio updates once stagnation is detected. We evaluate BridgeWalker-v0 under the same evaluation budget and PPO settings as in prior experiments. Fig. 4 reports the final reward as a heatmap over a 7×7 grid, covering all tested hyperparameter combinations.

Overall, the controller exhibits a clear and smooth optimum region centered around $\epsilon = 10^{-3}$ and $\eta = 0.10$, where the best performance reaches 4.79. Nearby configurations remain competitive, indicating that the adaptive allocation is not overly sensitive to moderate perturbations. For example, fixing $\epsilon = 10^{-3}$ yields 4.62 at $\eta = 0.08$ and 4.72 at $\eta = 0.12$, while slightly smaller or larger step sizes lead to gradual degradation.

The trends across both axes are consistent with the controller mechanics. When ϵ is too small (e.g., 10^{-5} to 10^{-4}), minor fluctuations are treated as progress, so the allocation is updated less effectively toward correcting stagnation, resulting in lower reward. Conversely, when ϵ is too large (e.g., 5×10^{-3} to 10^{-2}), stagnation is declared too often, biasing the allocation toward excessive diffusion and weakening mutation-based refinement. Similarly, a very small η under-reacts to stagnation, whereas a large η makes the allocation changes abrupt and can induce oscillation. Therefore, intermediate

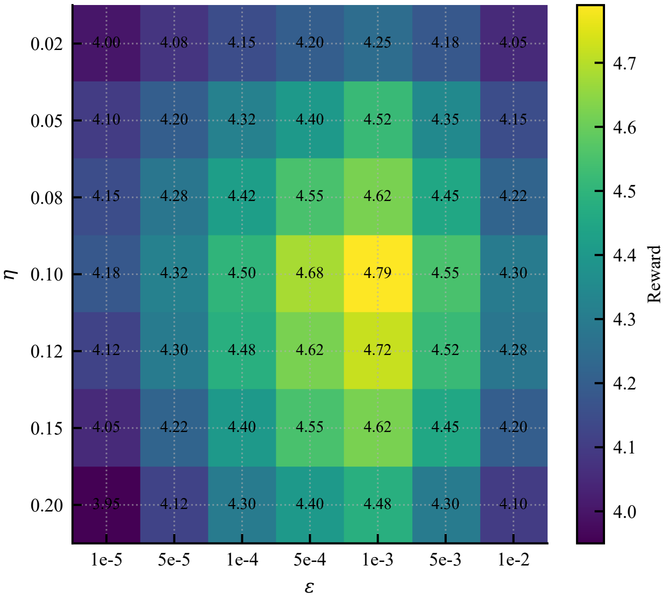

Fig. 4. Sensitivity analysis of the adaptive ratio controller with respect to the tolerance ϵ and update step size η on BridgeWalker-v0. Each cell reports mean final reward over 5 seeds under the same budget as our main experiments.

values $\epsilon \approx 10^{-3}$ and $\eta \approx 0.10$ provide the best balance between responsiveness and stability, and we adopt them as the default in our main experiments.

V. Conclusion

We presented Preference-Guided Diffusion Evolution (PGDE), a diffusion-guided evolutionary framework for robot morphology optimization under expensive controller training. PGDE replaces hard labels with rank-normalized soft targets, scales sampling updates with a confidence estimate, and adapts the mutation–diffusion ratio using online fitness improvement. Experiments on 9 EvoGym tasks show that PGDE improves performance and robustness over evolutionary and diffusion-guided baselines, while ablations confirm the contribution of each component. PGDE still inherits the cost of controller retraining and depends on the quality of online fitness feedback, so we view it as a practical enhancement to existing morphology-search pipelines rather than a complete replacement for transferable priors. The source code of PGDE is publicly available at https://github.com/Iwannnn/pgde.git.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62402302, in part by Shanghai Pujiang Program under Grant 24PJD032, in part by the Science and Technology Commission of Shanghai Municipality project under Grant 24ZR1422200. The authors would also like to thank the reviewers for their careful reading and constructive comments. Their insightful suggestions have helped improve the quality and clarity of this paper.

References

[1] J. Bongard, “Why morphology matters,” *The horizons of evolutionary robotics*, vol. 6, pp. 125–152, 2014.

[2] A. Gupta, S. Savarese, S. Ganguli, and L. Fei-Fei, “Embodied intelligence via learning and evolution,” *Nature communications*, vol. 12, no. 1, p. 5721, 2021.

[3] L. Xin, Z. Huang, Y. Wang, J. Peng, H. Kong, Y. Liu, and Y. Sun, “Directed morphological evolution for fast robot structure optimization,” *IEEE Transactions on Evolutionary Computation*, 2026.

[4] A. Zhao, J. Xu, M. Konaković-Luković, J. Hughes, A. Spielberg, D. Rus, and W. Matusik, “Robogrammar: graph grammar for terrain-optimized robot design,” *ACM Transactions on Graphics (TOG)*, vol. 39, no. 6, pp. 1–16, 2020.

[5] K. Qiu, K. Ciebiera, P. Fijałkowski, M. Cygan, and Ł. Kuciński, “Robomorph: Evolving robot morphology using large language models,” *arXiv preprint arXiv:2407.08626*, 2024.

[6] C. Schaff, D. Yunis, A. Chakrabarti, and M. R. Walter, “Jointly learning to construct and control agents using deep reinforcement learning,” in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 9798–9805.

[7] D. Ha, “Reinforcement learning for improving agent design,” *Artificial life*, vol. 25, no. 4, pp. 352–365, 2019.

[8] Y. Wang, “Learn and evolve to optimize robot morphologies,” in *4th International Conference on Internet of Things and Smart City (IoTSC 2024)*, vol. 13224. SPIE, 2024, pp. 85–89.

[9] S. Liu, J. Yan, H. Wang, and Y. Jin, “Morphology evolution for embodied robot design with a classifier-guided diffusion model,” *IEEE Transactions on Evolutionary Computation*, 2025.

[10] R. Pfeifer and C. Scheier, *Understanding intelligence*. MIT press, 2001.

[11] H. Lipson and J. B. Pollack, “Automatic design and manufacture of robotic lifeforms,” *Nature*, vol. 406, no. 6799, pp. 974–978, 2000.

[12] T. F. Nygaard, C. P. Martin, D. Howard, J. Torresen, and K. Glette, “Environmental adaptation of robot morphology and control through real-world evolution,” *Evolutionary Computation*, vol. 29, no. 4, pp. 441–461, 2021.

[13] J. Hu, J. Whitman, and H. Choset, “Glso: Grammar-guided latent space optimization for sample-efficient robot design automation,” in *Conference on Robot Learning*. PMLR, 2023, pp. 1321–1331.

[14] J. Whitman, M. Travers, and H. Choset, “Modular mobile robot design selection with deep reinforcement learning,” in *NeurIPS Workshop on ML for engineering modeling, simulation and design*, vol. 2, 2020, pp. 7–2.

[15] J. Hu, J. Whitman, M. Travers, and H. Choset, “Modular robot design optimization with generative adversarial networks,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4282–4288.

[16] J. T. Carvalho and S. Nolfi, “The role of morphological variation in evolutionary robotics: Maximizing performance and robustness,” *Evolutionary Computation*, vol. 32, no. 2, pp. 125–142, 2024.

[17] T. F. Nygaard, C. P. Martin, J. Torresen, K. Glette, and D. Howard, “Real-world embodied ai through a morphologically adaptive quadruped robot,” *Nature Machine Intelligence*, vol. 3, no. 5, pp. 410–419, 2021.

[18] J. Hou, Y. Hong, S. Liu, Q. Pan, J. Zhang, Q. Xu, Q. Nie, Z. Wang, L. Xin, Y. Wang *et al.*, “A fully self-powered digital wearable system for the auxiliary treatment of plantar fasciitis,” *Advanced Science*, p. e21682, 2026.

[19] Y. Yuan, Y. Song, Z. Luo, W. Sun, and K. Kitani, “Transform2act: Learning a transform-and-control policy for efficient agent design,” *arXiv preprint arXiv:2110.03659*, 2021.

[20] H. Dong, J. Zhang, T. Wang, and C. Zhang, “Symmetry-aware robot design with structured subgroups,” in *International Conference on Machine Learning*. PMLR, 2023, pp. 8334–8355.

[21] J. Bhatia, H. Jackson, Y. Tian, J. Xu, and W. Matusik, “Evolution gym: A large-scale benchmark for evolving soft robots,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 2201–2214, 2021.

[22] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[23] K. O. Stanley, “Compositional pattern producing networks: A novel abstraction of development,” *Genetic programming and evolvable machines*, vol. 8, no. 2, pp. 131–162, 2007.