ProMix: Learning Optimal Data Mixtures for Robotic Imitation via Proxy-Reference Distillation

Qie Sima*, Wei Xue, Yike Guo†
Division of Emerging Interdisciplinary Areas (EMIA)
Hong Kong University of Science and Technology

***Abstract*—This paper introduces ProMix, an efficient data-mixture optimization framework designed to handle the heterogeneity of large-scale robotic datasets. While Vision-Language-Action (VLA) models benefit from diverse data, naïve uniform mixing often leads to performance saturation. ProMix adopts a proxy-reference distillation architecture that learns optimal per-domain sampling weights by minimizing the worst-case excess loss against a frozen reference model. This mechanism effectively mitigates the "over-pessimism" common in traditional distributionally robust optimization (DRO) when applied to complex robotic distributions. Experimental results on 2.7M real-world actions from the Open-X Embodiment dataset demonstrate that ProMix improves average success rates from 68.4% to 76.2%, with substantial gains (12–18%) in low-resource domains. Crucially, ProMix achieves a 17× reduction in computational overhead (11 vs. 192 GPU-hours) compared to existing adaptive mixing methods, providing a scalable pipeline for pre-training large-scale robotic foundation models.**

***Index Terms*—Vision-Language-Action model, data mixture optimization, proxy-reference model, imitation learning**

I. Introduction

Vision-Language-Action (VLA) models, such as RT-2 [1] and the π model family [2], have achieved impressive generalization by pre-training on massive, heterogeneous datasets like Open-X Embodiment. However, effectively blending data from disparate robot platforms and tasks remains a critical bottleneck. Naïve uniform mixing often fails to capture the varying importance of different domains, while current adaptive mixing strategies like Re-Mix [3] incur prohibitive computational costs and often suffer from suboptimal weight allocation.

A key limitation of existing robotics-focused data optimization is the omission of the "proxy-reference" architecture. Unlike the discrete token spaces of LLMs, robotic imitation learning involves continuous, high-dimensional action spaces with significant heteroscedastic noise. Without a dedicated reference model to provide a baseline for "achievable loss," a standalone optimizer tends to become **overly pessimistic**, disproportionately weighting inherently noisy trajectories rather than those most conducive to cross-domain generalization.

We propose **ProMix**, a framework that systematically adapts the complete proxy-reference methodology to robotics. By utilizing a lightweight proxy model guided by a frozen reference model, ProMix minimizes **excess loss** across domains to learn robust mixture weights. This approach aligns with the trend toward efficient, high-performance models like $\pi_{0.5}$ [4] by significantly reducing the computational barrier to data-mixture tuning.

Our key contributions are as follows:

- We provide a systematic adaptation of the "proxy-reference" architecture for robotics, demonstrating that the excess loss objective is essential for mitigating over-pessimism in heterogeneous datasets.
- We propose ProMix, a computationally efficient framework that consistently improves VLA success rates while reducing training wall-clock time by over 17× compared to prior adaptive mixing methods [3].
- We conduct experiments on 2.7M real-world actions, showing that ProMix significantly outperforms uniform and Re-Mix baselines, particularly in low-resource minority domains.

II. Related Work

VLA Models and Data Efficiency. Foundation models like RT-2 [1] and the π family [2], [4] demonstrate that sophisticated data curation enables high-performance policies. While recent works explore data efficiency through synthetic generation [5] or distillation [6], they often treat heterogeneous data equally. **ProMix** fills this gap by automating the discovery of optimal data blends across disparate sources.

Intelligent Data Selection. Curriculum and active learning [7], [8] identify informative samples but often require costly online interaction. Alternatively, data pruning and importance sampling [5] improve efficiency by discarding data, which may harm robustness in minority domains. In contrast, ProMix retains the full dataset but dynamically re-weights domains to ensure balanced generalization.

Distributionally Robust Learning. ProMix is grounded in Distributionally Robust Optimization (DRO). Recent studies like TaSIL [9] highlight DRO's role in handling distribution shifts. Unlike standard DRO which can be overly pessimistic in noisy environments, our "excess loss" objective—inspired by latent reasoning trends [10]—provides a more stable optimization target.

Data Mixture Optimization (DMO). Originally pioneered for LLMs by DoReMi [11], DMO uses a proxy-reference architecture to learn optimal mixtures. Re-Mix [3] adapted this for robotics but omitted the reference model, reducing

Qie Sima:qsima@connect.ust.hk †Corresponding author Yike Guo

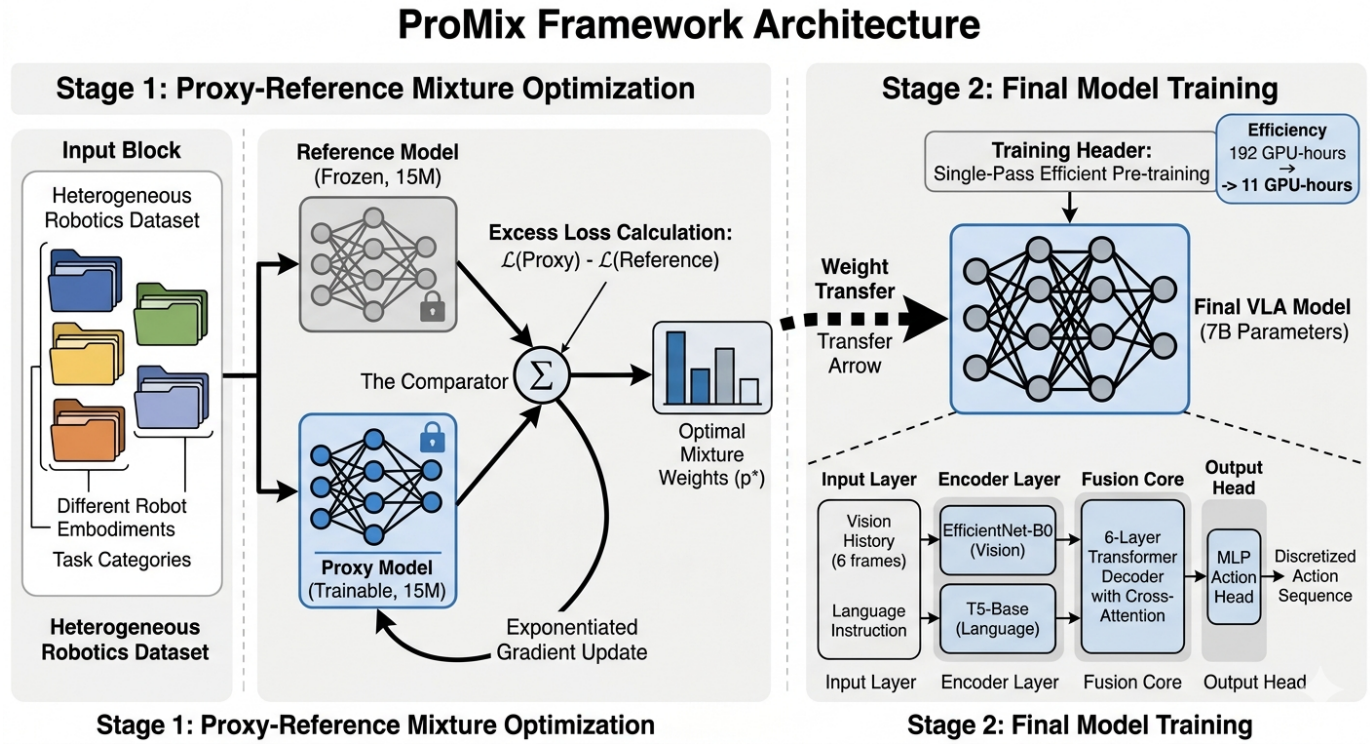

Fig. 1: **The ProMix Framework Architecture.** The pipeline consists of two stages: (Stage 1) A proxy-reference optimization loop that learns optimal domain weights p^* by minimizing the worst-case excess loss of a 15M trainable proxy against a frozen reference model; (Stage 2) Single-pass efficient training of the 7B VLA model using the transferred weights p^*, reducing total compute from 192 to 11 GPU-hours.

it to standard worst-case DRO. We argue that in robotics, the reference model is essential to prevent the optimizer from being distracted by inherently noisy or "unlearnable" trajectories. ProMix implements the complete methodology, achieving superior performance and 17× higher efficiency than Re-Mix.

III. Problem Formulation

We consider training a large-scale VLA model $\boldsymbol{\theta}$ on a heterogeneous dataset $\mathcal{D}$ comprising N domains $\{\mathcal{D}_1, \ldots, \mathcal{D}_N\}$. Each domain $\mathcal{D}_i$ contains tuples of (s, a, l), where s, a, and l represent state, action, and language instruction, respectively. Standard imitation learning typically minimizes the uniform empirical risk:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \alpha_i L_{\mathcal{D}_i}(\boldsymbol{\theta}) \tag{1}$$

where $L_{\mathcal{D}_i}(\boldsymbol{\theta})$ is the average behavior cloning loss. In large-scale robotics, however, uniform mixing often causes performance saturation, as gradient updates can be dominated by massive but low-quality data, leading to suboptimal generalization on challenging "long-tail" domains.

To achieve cross-domain robustness, we frame data mixture optimization as a **Distributionally Robust Optimization (DRO)** problem. Traditional DRO ($\min \max \sum p_i L_i$) seeks to minimize the worst-case loss across domains. In the context of robotics, this formulation is often counter-productive: it inadvertently assigns the highest sampling weights to domains that are *inherently noisy* or "unlearnable" (e.g., chaotic human teleoperation or low-precision sensing), rather than those that are most beneficial for policy improvement.

We address this by defining a minimax objective based on **excess loss** relative to a frozen reference model $\boldsymbol{\theta}_{ref}$. The optimization seeks sampling weights $\boldsymbol{p} \in \Delta$ that minimize:

$$\min_{\boldsymbol{\theta}} \max_{\boldsymbol{p} \in \Delta} \sum_{i=1}^{N} p_i \cdot (L_{\mathcal{D}_i}(\boldsymbol{\theta}) - L_{\mathcal{D}_i}(\boldsymbol{\theta}_{ref})) \tag{2}$$

Here, $L_{\mathcal{D}_i}(\boldsymbol{\theta}_{ref})$ serves as an *achievability baseline* that captures the intrinsic difficulty of domain $\mathcal{D}_i$. By minimizing the "regret" against this baseline, the model is forced to prioritize domains where the current policy significantly lags behind its potential, effectively mitigating the **over-pessimism** that plagues standard DRO.

Solving Eq. (2) directly for a large-scale VLA is computationally prohibitive. **ProMix** decouples this optimization by employing a lightweight proxy model $\boldsymbol{\theta}_{proxy}$ to learn the optimal mixture $\boldsymbol{p}^*$, which is then utilized for efficient, single-pass final training.

IV. Methodology

ProMix introduces a two-stage pipeline designed to decouple data-mixture optimization from large-scale training.

This design leverages the theoretical observation that relative domain difficulties and optimal data mixtures identified by compact proxy models remain highly consistent when scaling to larger architectures [11]. As illustrated in Fig. 1, Stage 1 employs a dual-model distillation architecture where a proxy model tracks the "achievable" learning progress relative to a reference baseline. This feedback loop dynamically reshapes the sampling distribution p^*, prioritizing domains with high information gain while filtering out heteroscedastic noise.

A. The Proxy-Reference Framework

We employ a Reference Model ($\boldsymbol{\theta}_R$) and a Proxy Model ($\boldsymbol{\theta}_P$), both sharing a compact 15M-parameter VLA architecture to minimize search-phase overhead.

- **VLA Architecture:** Our base architecture features an EfficientNet-B0 vision encoder processing a history of 6 images, a frozen T5-Base language encoder, and a 6-layer Transformer decoder. The visual and language tokens are fused via cross-attention to predict discretized action sequences through an MLP action head.
- **Reference Model:** $\boldsymbol{\theta}_R$ is pre-trained once on the full dataset $\mathcal{D}$ using a uniform mixture and subsequently frozen. It serves as a static benchmark for the "inherent difficulty" of each domain $\mathcal{D}_i$. By using $\boldsymbol{\theta}_R$ as an anchor, we prevent the optimizer from over-prioritizing "unlearnable" noisy trajectories, a critical distinction from standard DRO.
- **Proxy Model:** $\boldsymbol{\theta}_P$ starts from a fresh initialization. It iteratively learns optimal domain weights $\boldsymbol{p}$ by minimizing the *excess loss* relative to $\boldsymbol{\theta}_R$. This ensures that the optimization effort is concentrated on domains where the policy has the most significant room for improvement.

B. Learning and Transferring Weights

ProMix learns sampling weights $\boldsymbol{p} = [p_1, \dots, p_N]$ by solving a distributionally robust minimax objective with entropy regularization to prevent weight collapse:

$$\min_{\boldsymbol{\theta}_P} \max_{\boldsymbol{p} \in \Delta} \sum_{i=1}^{N} p_i \cdot [L_{\mathcal{D}_i}(\boldsymbol{\theta}_P) - L_{\mathcal{D}_i}(\boldsymbol{\theta}_R)] + \lambda \sum_{i=1}^{N} p_i \log p_i \quad (3)$$

The term $L(\boldsymbol{\theta}_P) - L(\boldsymbol{\theta}_R)$ represents the per-domain regret. We solve this via a two-step iterative optimization:

1) **Model Update:** For fixed $\boldsymbol{p}$, we update the proxy parameters $\boldsymbol{\theta}_P$ via AdamW to minimize the weighted loss: $\boldsymbol{\theta}_P \leftarrow \boldsymbol{\theta}_P - \alpha \nabla_{\boldsymbol{\theta}_P} [\sum p_i L_{\mathcal{D}_i}(\boldsymbol{\theta}_P)]$, where $\alpha = 10^{-4}$.
2) **Weight Update:** For a fixed proxy $\boldsymbol{\theta}_P$, we update the weights p_i via exponentiated gradient ascent, yielding the following closed-form update that naturally maintains the simplex constraint:

$$p_i \leftarrow \frac{\exp\left(\eta \cdot [L_{\mathcal{D}_i}(\boldsymbol{\theta}_P) - L_{\mathcal{D}_i}(\boldsymbol{\theta}_R)]\right)}{\sum_{j=1}^{N} \exp\left(\eta \cdot \left[L_{\mathcal{D}_j}(\boldsymbol{\theta}_P) - L_{\mathcal{D}_j}(\boldsymbol{\theta}_R)\right]\right)} \quad (4)$$

C. Final Model Training

Once the optimal weights $\boldsymbol{p}^*$ converge, they are frozen and transferred to the final large-scale VLA model $\boldsymbol{\theta}_L$ (e.g., 7B parameters). The training objective follows:

$$\min_{\boldsymbol{\theta}_L} \sum_{i=1}^{N} p_i^* \cdot L_{\mathcal{D}_i}(\boldsymbol{\theta}_L) \quad (5)$$

Unlike iterative re-weighting methods like Re-Mix [3] that require multiple resource-intensive retraining passes of the target model, ProMix identifies the optimal blend using a 15M proxy and requires only a **single pass** for the 7B model. This architecture-agnostic transferability results in a reduction of total computational overhead from 192 to 11 GPU-hours, facilitating rapid iteration across the VLA pipeline.

V. Experiments

In this section, we present a comprehensive evaluation of **ProMix**. Our experiments aim to verify: (1) if ProMix outperforms existing data mixture strategies; (2) the necessity of the "proxy-reference" architecture; and (3) the transferability of learned weights across model scales.

A. Experimental Setup

We conduct experiments on a diverse subset of the **Open-X Embodiment Dataset [12]**, totaling approximately 2.7 million real-world actions. The data is partitioned into distinct domains based on robot embodiment (e.g., WidowX, UR5, Franka) and task categories (e.g., pick-and-place, door opening).

Model Architectures: We utilize a consistent VLA architecture: a frozen EfficientNet-B0 vision encoder (6-frame history), a frozen T5-Base language encoder, and a 6-layer Transformer decoder. The **Reference** (θ_R) and **Proxy** (θ_P) models are compact (15M parameters), while the **Final Model** (θ_L) scales up to 50M and 7B parameters.

Training Details: Models are trained using AdamW with weight decay 0.01. The learning rate for parameters is 1×10^{-4} with cosine decay, and for domain weights $\eta = 10^{-2}$. Proxy optimization runs for 100k steps on an NVIDIA H100 GPU. We report mean $\pm$ 95% CI over 5 seeds; paired t-tests confirm $p < 0.05$ for ProMix's gains.

B. Main Results: Comparison with Baselines

Table I summarizes the performance against baselines including Uniform Mixing, Proportional Sampling, Expert Weights, and Re-Mix [3]. **ProMix** achieves an average success rate of 76.2%, outperforming Re-Mix by $+1.4\%$ and uniform mixing by $+7.8\%$. Notably, the minimum domain success rate improves from 51.4% (uniform) to 68.5% (+17.1%), validating the minimax formulation's ability to lift challenging, low-resource domains.

C. Ablation and Cross-Scale Generalization

We conduct ablation studies to verify the role of the reference model. Removing θ_R causes the proxy to minimize raw loss, leading to a 3.8% drop in average SR and a 14.3% crash in min-domain performance (Table II). This confirms

TABLE I: Success Rate (%) Comparison on Open-X Embodiment

Method	Avg. SR	Min Domain	GPU-h
Uniform Mixing	68.4 ± 1.2	51.4	-
Proportional Sampling	64.1 ± 1.5	47.2	-
Expert Weights	70.3 ± 1.3	54.6	-
Re-Mix [3]	74.8 ± 0.9	65.4	192
ProMix (Ours)	**76.2 ± 0.7**	**68.5**	**11**

that the excess loss objective is essential for mitigating over-pessimism toward noisy data.

Furthermore, we transfer the weights p^* learned by the 15M proxy to larger models. As shown in Table II, ProMix yields consistent gains across all scales, reaching a $+11.4\%$ relative improvement for the 50M model. This justifies our "compute once, deploy anywhere" workflow.

TABLE II: Ablation Study (Top) and Scaling Results (Bottom)

Ablation Variant	Avg. SR (%)	Min SR (%)
w/o Reference	72.4 ± 1.1	54.2
w/o Entropy Regularization	73.8 ± 1.4	48.7
ProMix (Full)	**76.2 ± 0.7**	**68.5**
Model Scale	*Uniform SR*	*ProMix SR*
15M (Proxy)	64.2 ± 1.5	**68.9 ± 1.1**
50M (Medium)	68.4 ± 1.2	**76.2 ± 0.7**
7B (Large)	72.5 ± 0.9	**79.8 ± 0.6**

D. Qualitative Analysis

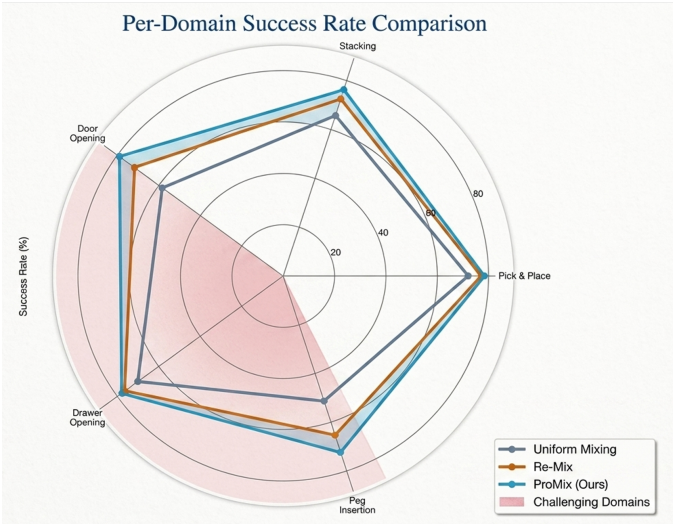

Fig. 2: Radar plot of per-domain success rates across strategies.

The learned weights p^* (Fig. 2) converge to a stable distribution that systematically depresses large, homogeneous domains while elevating smaller, high-variance tasks. This results in a 12–18% absolute gain in the most challenging scenarios, ensuring robust zero-shot generalization across the VLA pipeline.

VI. Conclusion

In this work, we presented **ProMix**, a computationally efficient framework for optimizing data mixtures in large-scale VLA pre-training. By systematically adapting the *proxy-reference* distillation architecture to the robotics domain, we demonstrated that optimal sampling weights can be learned on a lightweight 15M model and successfully transferred to 7B+ architectures. Our results on Open-X Embodiment show a 7.8% improvement in average success rate and a **17× reduction** in optimization overhead compared to prior art.

ProMix offers a practical and scalable solution for the VLA pipeline, making high-performance data curation accessible without industrial-scale compute. Future work will explore integrating ProMix with automated dataset pruning and active learning to further enhance the data efficiency of embodied AI systems.

References

[1] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," in *Conference on Robot Learning*. PMLR, 2023, pp. 2165–2183.

[2] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter *et al.*, "π_0: A vision-language-action flow model for general robot control," *arXiv preprint arXiv:2410.24164*, 2024.

[3] J. Hejna, C. Bhateja, Y. Jiang, K. Pertsch, and D. Sadigh, "Re-mix: Optimizing data mixtures for large scale imitation learning," *arXiv preprint arXiv:2408.14037*, 2024.

[4] P. Intelligence, K. Black, N. Brown, J. Darpinian, K. Dhabalia, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai *et al.*, "$\pi_{0.5}$: a vision-language-action model with open-world generalization," *arXiv preprint arXiv:2504.16054*, 2025.

[5] B. Wang, X. Meng, X. Wang, Z. Zhu, A. Ye, Y. Wang, Z. Yang, C. Ni, G. Huang, and X. Wang, "Embodiedreamer: Advancing real2sim2real transfer for policy training via embodied world modeling," *arXiv preprint arXiv:2507.05198*, 2025.

[6] J. Y. Zhu, C. G. Cano, D. V. Bermudez, and M. Drozdzal, "Incoro: In-context learning for robotics control with feedback loops," *arXiv preprint arXiv:2402.05188*, 2024.

[7] M. Hou, K. Hindriks, A. Eiben, and K. Baraka, "Active robot curriculum learning from online human demonstrations," in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2025, pp. 810–818.

[8] Y. Yin, Z. Wang, Y. Sharma, D. Niu, T. Darrell, and R. Herzig, "In-context learning enables robot action prediction in llms," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 8972–8979.

[9] A. Gahlawat, A. Aboudonia, S. Banik, N. Hovakimyan, N. Matni, A. D. Ames, G. Zardini, and A. Speranzon, "Distributionally robust imitation learning: Layered control architecture for certifiable autonomy," 2025. [Online]. Available: https://arxiv.org/abs/2512.17899

[10] S. Bai, J. Lyu, W. Zhou, Z. Li, D. Wang, L. Xing, X. Zhao, P. Wang, Z. Wang, C. Chi, B. Chen, and S. Zhang, "Latent reasoning vla: Latent thinking and prediction for vision-language-action models," *arXiv preprint arXiv:2602.01166*, 2026.

[11] S. M. Xie, H. Pham, X. Dong, N. Du, H. Liu, Y. Lu, P. S. Liang, Q. V. Le, T. Ma, and A. W. Yu, "Doremi: Optimizing data mixtures speeds up language model pretraining," *Advances in Neural Information Processing Systems*, vol. 36, pp. 69 798–69 818, 2023.

[12] A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain *et al.*, "Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6892–6903.