MARVA: Modular Architecture for Robust Visual Agents

Vanessa Schenkel[1] **, Gabriel de Oliveira Ramos**[1]

[1]Graduate Program in Applied Computing
Universidade do Vale do Rio dos Sinos, São Leopoldo - RS - Brazil

vanessaschenkel@edu.unisinos.br, gdoramos@unisinos.br

Abstract. *Generalization in visual RL is challenging: small visual shifts can degrade performance. We present MARVA, a dual-regularization extension of MaDi combining a GRL-based discriminator and a contrastive (InfoNCE) loss on masked views. On walker-walk, MARVA matches baseline performance in easier domains and improves robustness in video_hard and DistractingCS.*

1. Introduction

Reinforcement Learning (RL) has achieved remarkable progress in solving complex control tasks when trained and evaluated in identical environments. However, agents trained from raw pixels often fail to generalize: small changes in the background, color or visual perspective can drastically reduce performance even though the underlying task remains unchanged. This limitation is particularly pressing in Visual Reinforcement Learning (VRL) where robustness to distribution shifts is a prerequisite for deploying agents in real-world scenarios.

In this work we introduce MARVA (Modular Architecture for Robust Visual Agents), a dual-regularization extension built upon MaDi (Mask Distractions) [Grooten et al. 2024], combining adversarial and contrastive objectives for improved generalization. MARVA integrates a domain discriminator with a Gradient Reversal Layer and a contrastive (InfoNCE) objective on the masked views, ensuring that both latent representations and visual attention masks remain domain-invariant. Our experiments on the DeepMind Control Generalization Benchmark show that MARVA achieves performance comparable to MaDi in easier conditions while significantly improving robustness in the most challenging domains.

2. Related Work

Generalization in Visual Reinforcement Learning (VRL) remains one of the most pressing challenges in AI: small changes in background, color, or viewpoint can sharply degrade performance. A major line of work uses adversarial regularization to encourage domain-invariant features.

RARL [Pinto et al. 2017] frames training as a minimax game in the dynamics, with an adversary perturbing the environment while the protagonist maximizes reward. It improves robustness but is unstable and does not address visual shifts.

DARL [Li et al. 2021] brings adversarial learning to pixels by coupling an encoder to a domain discriminator via a Gradient Reversal Layer (GRL), improving zero-shot transfer over SAC [Haarnoja et al. 2018], RAD [Laskin et al. 2020], and DrQ [Yarats et al. 2021]. However, it requires multiple domains and careful loss balancing.

MaDi [Grooten et al. 2024] takes a lightweight path: a Masker before the encoder filters irrelevant pixels and is trained only via the critic's loss. MaDi matches or outperforms SVEA [Hansen et al. 2021], SGQN [Bertoin et al. 2022], and DrQ with $\sim$0.2% extra parameters, but sparse reward signals may still allow subtle domain cues.

In summary, adversarial methods enforce invariance but can be unstable and multi-domain dependent, whereas masking is efficient yet limited by sparse signals. Motivated by this gap, we propose a dual-regularization extension of MaDi that unifies both perspectives, suppressing domain-specific information while preserving task-relevant features.

3. Proposed Architecture: MARVA

Our proposed architecture, MARVA, builds upon a lightweight masking baseline [Grooten et al. 2024], extending it with adversarial regularization to explicitly enforce domain invariance. While the baseline demonstrated that lightweight masking can efficiently suppress task-irrelevant distractions, its reliance solely on sparse rewards signals limits its ability to prevent masks from encoding subtle domain-specific biases. MARVA addresses this limitation by combining masking with dual adversarial objectives, ensuring that both latent representations and attention masks are domain-agnostic.

The first component of this extension introduces a domain discriminator attached to the encoder via a Gradient Reversal Layer (GRL). During training, the discriminator attempts to classify the domain of origin for each observation, while the encoder is optimized to confuse it. This adversarial process forces the encoder to discard domain-specific information producing latent features that are invariant to background changes and other visual shifts.

The second component adds a contrastive (InfoNCE) objective on masked observations. Concretely, for each input we generate two masked views (one original and one augmented, e.g., with the overlay pipeline), pass them through the encoder and a projection head, and treat their projections as a positive pair while using the rest of the batch as negatives. The InfoNCE loss encourages alignment of the positive pair and separation from negatives, discouraging the Masker and encoder from retaining domain-specific cues in the masked features. This complements the adversarial discriminator by promoting domain-invariant representations at both the attention and latent levels.

An important aspect is that these extensions are integrated with minimal overhead. MARVA retains the lightweight nature of the baseline, requiring only the addition of a discriminator branch and a contrastive projection head with an InfoNCE loss, while preserving the same critic-driven training signal for the Masker. This modularity allows MARVA to remain computationally efficient while significantly improving robustness.

4. Experimental Evaluation

We updated several packages and modified parts of the code to ensure compatibility with both CPU and Apple's Metal Performance Shaders (MPS). To ensure comparability, we reran the official baseline [Grooten et al. 2024]. This step was necessary to confirm that our environment and code adjustments did not alter the expected behavior of the baseline agent. The reproduced baseline results were consistent with those in [Grooten et al. 2024], validating the fidelity of our setup.

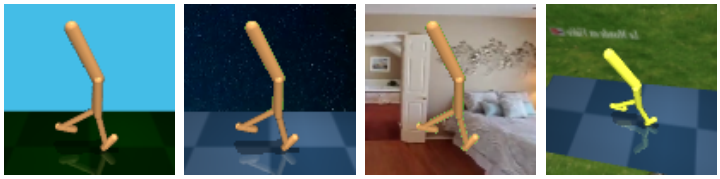

Figure 1. Frames from MARVA's evaluation rollouts in the four *walker-walk* domains: *color_hard*, *video_easy*, *video_hard*, and *DistractingCS*.

We ran a 20,000-step sanity check on *walker-walk* (adversarial, contrastive, and both regularizations; with and without DropBlock) to ensure stability. Training showed no divergence, consistent losses, and coherent masks, with DropBlock slightly accelerating early learning. DropBlock, a structured form of dropout that removes contiguous regions in feature maps, encourages spatial robustness and reduces overfitting. We then trained MARVA with the combined ("both") regularization (adversarial via GRL + contrastive via InfoNCE) for 500,000 steps, running three random seeds with and without DropBlock. Figure 1 shows frames from MARVA's evaluations across the four domains used by `eval_mode=four` (*color_hard*, *video_easy*, *video_hard*, *DistractingCS*).

Following the evaluation protocol of [Grooten et al. 2024], Table 1 reports the average undiscounted return over the last 10% of training for each evaluation environment. Although evaluation was performed in all four domains, the table reports only the three domains used in [Grooten et al. 2024] to enable direct comparison with the MaDi baseline. Results show that MARVA with dual regularization improves upon the MaDi baseline in the most challenging domains, *video_hard* and *DistractingCS*, while maintaining comparable performance in *video_easy* within the expected variance range. The variant with DropBlock achieved the highest returns in the hardest domains, suggesting that structured dropout helps stabilize learning and improve robustness under visual complexity and dynamic distractions.

Table 1. *walker-walk* at 500k steps; results averaged over three seeds (mean ± standard deviation).

Method	Video Easy	Video Hard	DistractingCS
MaDi (paper)	895 ± 24	504 ± 33	570 ± 49
MARVA w/o DropBlock	871.6 ± 35.5	550.3 ± 120.2	585.7 ± 80.3
MARVA w/ DropBlock	843.9 ± 76.6	599.3 ± 144.5	621.6 ± 49.1

Overall, these findings indicate that dual adversarial regularization strengthens robustness against visual shifts without harming easier domains and specifically mitigates the main weaknesses identified in the original MaDi baseline. Table 1 illustrate this effect clearly.

5. Conclusion and Future Work

In this work we introduced MARVA, a dual-regularization extension of MaDi designed to enforce domain invariance at both representation and masking levels by combining a Gradient Reversal Layer and a contrastive objective on masked views. Our evaluation on *walker-walk* shows that MARVA retains the efficiency of MaDi while improving robustness in the most challenging domains, *video_hard* and *DistractingCS*, and maintaining

comparable performance in *video_easy*. These results, obtained after 500,000 training steps and averaged over three seeds, validate the stability of the approach and open directions for future work, including ablations of the discriminator and contrastive branches, additional seeds for statistical confidence, and experiments on other benchmark tasks.

Even incremental progress in addressing the generalization gap in visual reinforcement learning brings us closer to real-world deployment. By showing that lightweight extensions such as MARVA can effectively reduce domain-specific biases, this work contributes to the long-term goal of making reinforcement learning both robust and practical for complex, dynamic environments.

Acknowledgments

This research was partially supported by CNPq (grants 443184/2023-2, 313845/2023-9, 445238/2024-0) and CAPES (Finance Code 001).

References

Bertoin, D., Zouitine, A., Zouitine, M., and Rachelson, E. (2022). Look where you look! saliency-guided q-networks for generalization in visual reinforcement learning. *Advances in neural information processing systems*, 35:30693–30706.

Grooten, B., Tomilin, T., Vasan, G., Taylor, M. E., Mahmood, R. A., Fang, M., Pechenizkiy, M., and Mocanu, D. C. (2024). Madi: Learning to mask distractions for generalization in visual deep reinforcement learning. In *AAMAS'24: 2024 International Conference on Autonomous Agents and Multiagent Systems*. IFAAMAS.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870.

Hansen, N., Su, H., and Wang, X. (2021). Stabilizing deep q-learning with convnets and vision transformers under data augmentation. *Advances in neural information processing systems*, 34:3680–3693.

Laskin, M., Lee, K., Stooke, A., Pinto, L., Abbeel, P., and Srinivas, A. (2020). Reinforcement learning with augmented data. *Advances in neural information processing systems*, 33:19884–19895.

Li, B., François-Lavet, V., Doan, T., and Pineau, J. (2021). Domain adversarial reinforcement learning. *arXiv preprint arXiv:2102.07097*.

Pinto, L., Davidson, J., Sukthankar, R., and Gupta, A. (2017). Robust adversarial reinforcement learning. In *Intl. Conf. on Machine Learning*, pages 2817–2826. PMLR.

Yarats, D., Kostrikov, I., and Fergus, R. (2021). Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. In *International conference on learning representations*.