Fig. 7: Examples of observations in three modes within the DMC-GB environment. Image courtesy of the SODA website.

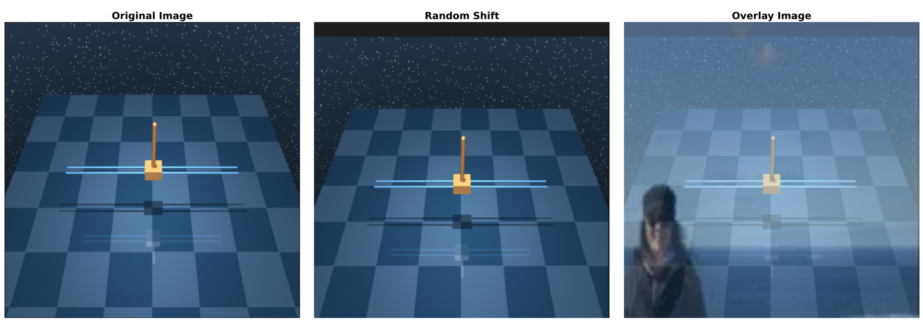

Fig. 8: Data augmentation pipeline. The image is randomly padded and shifted by a random value, and then mixed with a random image from the *Places 365* dataset.

A Environment Details

Fig. 7 shows three perturbed versions of the same observation in the DMC-GB environment [13], which we use to test the Generalization Ability. In the *color hard* mode, the colors of the subject and the background have been modified. In the *video easy* and *video hard* modes, the background is entirely replaced by unseen scenes.

B Implementation Details

For the Generalization Ability benchmark, we train the agent for 500,000 steps with 2 action repeats. All trainings are conducted on a single A100 GPU. The hyperparameters are listed in Tab. 3.

Table 3: Hyperparameters in the Generalization benchmark.

Hyperparameters	PromptAgent
Input size	84 x 84
Discount factor γ	0.99
Action repeat	2
Frame stack	3
Optimizer learning rate	$1e^{-4}$
Random shifting padding	4
Training step	500,000
Evaluation episodes	10
Optimizer	Adam
Replay buffer size	1,000,000
Mini-batch size	512 (Walker Walker, Walker Stand), 256 (others)

For the Sample Efficiency benchmark and ablation studies in Sec. 5.4, we maintain the aforementioned hyperparameters but reduce the training duration to 100,000 steps with 2 action repeats due to resource constraints.

Data Augmentation. Following the approach outlined in DrQ-v2 [41], we begin by applying the *RandomShift* algorithm as in [43], followed by integrating a randomly selected image $\mathcal{I}$ from the *Places 365* dataset linearly, $o = \alpha s + (1-\alpha)\mathcal{I}$ as in [13]. Fig. 8 shows an example from the data augmentation pipeline.

Critic Loss. The critic loss includes a regularization term $\mathcal{R}$ added to Eq. (1), as detailed in [12]. This term is constructed using an augmented version of the original image. The modified critic loss is now formulated as $\mathcal{L}(\phi) = \mathcal{F}(\phi) + \mathcal{R}(\phi)$ with $\mathcal{R}(\phi) = \mathbb{E}_{\tau \sim \mathcal{B}}\left[(Q_\phi(h_t^{'}, a_t) - y)^2\right]$ where $h_t^{'}$ is the latent representation of the augmented observation.