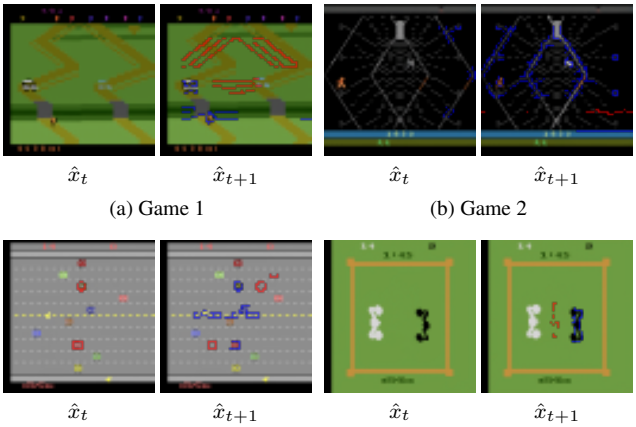
DIAMOND-LoL: Enforcing Lieb-Robinson Locality in Diffusion World Models for Long-Horizon Consistency

Anonymous ICCV submission

Paper ID 14

Abstract

*The world model enables agents to perform reasoning and planning through learning in the simulator, improving the sample efficiency in reinforcement learning. However, while diffusion-based models circumvent detail loss by operating in pixel space, their standard ℓ_2 loss introduces a critical physical inconsistency. Specifically, by averaging over plausible futures in partially observable scenarios, it generates blurry boundaries and acausal displacements, artifacts that violate the environment's fundamental principle of finite-speed propagation. To address this challenge, we propose **DIAMOND-LoL**, a diffusion training framework, which adds a **L**ieb-**Ro**binson **L**ocality loss (LoL loss) to enforce the finite speed propagation of pixel dynamics. Based on the Lieb-Robinson bound, LoL loss penalizes structural changes outside the data-driven light cone radius, keeping the predictions within the reachable range of the environment and avoiding mode averaging interpolation. Moreover, we prove that LoL loss is zero only when the prediction boundary is within the finite propagation set, and we show that it converts the long-term error growth from exponential form to linear form. Experiments demonstrate that DIAMOND-LoL provides a principled and physically consistent training objective for diffusion world models, especially having significant value in safety-critical scenarios.*



(a) Game 1 (b) Game 2 (c) Game 3 (d) Game 4

Figure 1. Visualization of acausal transitions between consecutive frames within a single rollout from DIAMOND. Each pair displays two sequential predicted frames from the same imagined trajectory, $\hat{x}_t$ (left) and $\hat{x}_{t+1}$ (right), to highlight the sudden, physically inconsistent mutations that occur in a single step. The overlays on the right frame ($\hat{x}_{t+1}$) diagnose these violations. Specifically, **red** areas denote new boundaries that materialize abruptly, outside the light-cone reachable from the preceding frame $\hat{x}_t$. Conversely, **blue** areas represent boundaries present in $\hat{x}_t$ that disappear in the very next step without a physically plausible successor. This teleportation is a direct consequence of the l_2 objective, which can cause the model to erratically jump between averaged modes, compromising long-horizon stability.

1. Introduction

World models have become a central paradigm for enabling agents to reason and plan in complex domains [14, 17]. Despite notable breakthroughs in areas such as vision [4, 24, 38, 39, 54, 59, 61], medicine [10–13, 22], games, robotics, and science [53, 58], reinforcement learning (RL) remains inherently sample-inefficient, limiting its practical deployment [6, 63]. Learned simulators address this limitation by providing a controllable training environment that improves sample efficiency [20] and enables learning from limited interactions [25, 26, 62].

Despite this promise, existing world models represent environment dynamics through sequences of discrete latent variables [47, 52]. Although discretization mitigates compounding errors across long horizons, it inevitably discards certain visual or structural details that may be critical for decision making [35, 46, 55]. The limitation is particularly pronounced in safety-critical applications such as autonomous driving [9, 19], where subtle cues, for example, a traffic signal or a distant pedestrian, can substantially alter the agent's policy. Increasing the number of discrete tokens

can partially alleviate the problem but comes at the cost of significant computational overhead [46].

Diffusion-based world models offer a complementary approach [1]. By learning to reverse a gradual noising process, these models generate high-fidelity observations directly in pixel space, thereby circumventing the bottlenecks introduced by discrete latent compression [1, 29, 51]. Such models [18, 28] excel in conditioning on agent actions and capturing multimodal outcome distributions—capabilities that are particularly valuable for robust world simulation. Across various implementations, diffusion-based world models have demonstrated the ability to capture intricate visual details, maintain temporal stability over long horizons, and serve not only as training environments for RL agents, but also as standalone interactive simulators [28, 31, 48].

However, a central challenge in diffusion-based world models lies in the choice of training objective. Current approaches, such as DIAMOND [1], typically adopt a loss of ℓ_2 reconstruction preconditioned by EDM in the pixel space, which introduces a fundamental inconsistency between learned dynamics and physical constraints of the underlying environment [1, 29]. Specifically, the pixel-wise ℓ_2 loss minimizes the expected reconstruction error by averaging across multiple plausible future outcomes. In multimodal or partially observable scenarios, this averaging produces blurred object boundaries and interpolated structures that do not correspond to any physically realizable state (as shown in Figure 1) [2, 15]. Specially, such averaging can generate boundary displacements that exceed the maximum per-step motion permitted by the discrete environment dynamics, effectively producing a causal teleportation of visual elements [8, 45, 56]. When predictions are rolled out autoregressively, these violations of locality are amplified over time, leading to long-term trajectory drift, loss of fine details, and degraded consistency under low-NFE regimes [18, 34, 64].

In this paper, we propose DIAMOND-LoL, a novel training framework that augments diffusion-based world models with a Lieb-Robinson Locality Loss (LoL loss). DIAMOND-LoL enforces finite-speed propagation in pixel dynamics, thereby eliminating acausal artifacts and enhancing long-term stability under low-NFE regimes. Unlike conventional heuristic regularization techniques, our framework is grounded in the Lieb-Robinson bound, which constrains the maximum spread of local interactions in lattice systems (more details in Section 4). By translating this concept into the evolution of pixel boundaries in Atari environments [35], we ensure that predicted structures remain confined to the reachable set of the environment's discrete dynamics.

Our main contributions are summarized as follows:

- We provide a theoretical and empirical analysis showing that the standard EDM-preconditioned ℓ_2 objective induces super-local artifacts, where multimodal uncertainty is collapsed into pixel averages that violate finite propagation constraints.
- We introduce DIAMOND-LoL, a principled framework for training diffusion-based world models. DIAMOND-LoL encourages predictions that are both in the environment's reachable set and physically plausible, thereby avoiding blurred interpolations. At its core, the Lieb-Robinson Locality Loss enforces finite-speed pixel boundary propagation by penalizing acausal structural changes beyond a data-driven light-cone radius.
- We prove that minimizing LoL loss yields zero loss if and only if predicted boundaries fall within the finite propagation set, and demonstrate that LoL loss enforces linear rather than exponential error accumulation in long-horizon rollouts.
- The experimental results show that DIAMOND-LoL significantly improves upon the original DIAMOND baseline, achieving a higher mean Human Normalized Score of 1.581 on the Atari 100k benchmark.

2. Related work

2.1. Diffusion-based world model

In recent years, RL based on world models has achieved remarkable progress. Traditional world models typically adopt recurrent state-space architectures to encode high-dimensional environment observations into compact latent representations in order to improve learning efficiency [1]. However, such compression often leads to the loss of crucial visual details. To preserve richer visual fidelity, recent work has introduced diffusion models into world modeling. For example, Alonso et al. [1] proposed the DIAMOND framework, which employs a diffusion process to generate environment rollouts, producing high-quality image sequences via iterative denoising. Experiments showed that DIAMOND achieved human-level performance on the Atari 100k benchmark [35], setting a new record for purely world model-based methods. Similarly, Ding et al. [18] presented the Diffusion World Model (DWM), which leverages conditional diffusion models to jointly predict multi-step future states and rewards, thereby avoiding the compounding error inherent in traditional recursive one-step prediction. On offline RL benchmarks such as D4RL, DWM significantly outperformed conventional models and established state-of-the-art results. Moreover, diffusion models have also been applied to trajectory planning. For instance, Janner et al. [32] introduced Diffuser, which directly generates full state-action trajectories through iterative denoising, effectively mitigating error accumulation in long-horizon planning.

Although DIAMOND and similar diffusion world mod-

els improve visual fidelity, EDM and ℓ_2 pixel reconstruction tends to average multimodal futures, producing blurry, acausal boundary displacements. DIAMOND-LoL adds a Lieb-Robinson locality loss to constrain pixel boundaries within a data-driven light cone, preventing infeasible interpolation artifacts.

2.2. Physics-Informed Learning for World Models

The core of Physics-informed neural networks (PINNs)[5] is to explicitly incorporate physical residuals, conservation laws, and boundary conditions into the loss function, making the model simultaneously subject to both data and physical constraints, thereby enhancing extrapolation performance and physical consistency [36, 49]. DeepONet provides a functional approximation framework for mapping input functions to the solution space, while FNO learns resolution-invariant solution operators through Fourier domain kernel parameterization. Based on this, PINO combines multi-resolution data supervision with high-resolution PDE constraints to alleviate convergence and extrapolation issues in multi-scale dynamics [40–42]. These methods have been introduced into tasks closely related to world models, using the inclusion of dynamic structures or physical priors in differentiable models for system identification and model-based control, to achieve more robust long-term rolling and extrapolation of unseen trajectories [21, 30, 43]. In high-dimensional visual scenarios, physical consistency in a weak form is used to constrain world modeling, such as PhyDNet that decouples PDE-like physical units from appearance branches, or gradSim that places differentiable rendering and multi-physics simulation in a pixel supervision loop, to infer dynamics and latent variables from the observation end [23, 33]. These papers demonstrate that if the prior information can be correctly and specifically incorporated and matched with the observational scale, then the physical deviations can significantly reduce model errors and enhance long-term consistency, thereby providing a feasible approach for injecting structural inductive biases into the world model [60].

The pixel-domain diffusion-based world model typically employs the pixel ℓ_2 objective under the EDM precondition (such as DIAMOND). In multi-modal or partially observable scenarios, it will conditionally average multiple feasible futures, resulting in a blurred boundary and interpolation structure. More importantly, this averaging can induce boundary displacements beyond the single-step reachable motion radius, violating the finite propagation property of discrete dynamics, and being amplified in autoregressive rolling [1, 3, 7, 16, 37, 44, 57]. Therefore, in this paper, a local prior based on the LoL loss is introduced at the pixel boundary evolution level. Through data-driven cone radius penalties to punish causal violations across step sizes, the prediction always remains within the reachable set of the environmental discrete dynamics, thereby suppressing blurry interpolation and stepwise teleportation, and strengthening long-term stability under the low-NFE condition.

3. Preliminaries

3.1. Reinforcement learning for world models

We consider a partially observable Markov decision process (POMDP). At each timestep t, the agent receives an observation x_t from the observation space $\mathcal{O}$ and executes an action a_t from the action space $\mathcal{A}$. In return, it receives a scalar reward r_t. Future rewards are discounted by a factor γ. The agent's decisions depend on the history of interactions, which we define as $h_t = (x^0_{\leq t}, a_{\leq t})$, representing the sequence of all past clean (original) observations and actions up to timestep t. A world model, with parameters θ, is trained to learn the environment's dynamics by modeling the conditional probability $p_\theta(x_{t+1} \mid h_t)$. Following the DIAMOND framework, reward and termination predictions are decoupled from the visual dynamics and handled by a separate prediction head, R_ψ, which is parameterized by ψ.

Given a dataset of trajectories collected from the environment, we train the world model p_θ by maximizing the likelihood of the true next observation x_{t+1} given the history h_t. This learned model is then used as a neural simulator to generate imagined trajectories, or rollouts. An agent policy $\pi_\phi(a_t \mid h_t)$ and a corresponding value function $V_\phi(h_t)$, both sharing parameters ϕ, are then optimized exclusively on these imagined trajectories (e.g., using an actor-critic algorithm with generalized λ-returns). The overall pipeline alternates in a closed loop between data collection in the real environment, world model updates, and agent training in imagination.

3.2. Score-based models

Diffusion (or score-based) models are generative models that operate by first corrupting data with gradually increasing Gaussian noise, and then learning the reverse-time dynamics to restore the original data distribution. This forward noising process, denoted by $\{x^\tau\}$, evolves along a continuous time index $\tau \in [0, T]$, with boundary conditions defined by the data distribution $p_0 = p_{\text{data}}$ and a tractable prior distribution $p_T = p_{\text{prior}}$ (e.g., a standard normal distribution). The process can be formally described by an Itô stochastic differential equation (SDE):

$$\mathrm{d}x = f(x, \tau)\,\mathrm{d}\tau + g(\tau)\,\mathrm{d}w, \tag{1}$$

where $f(x, \tau)$ is the drift coefficient, $g(\tau)$ is the diffusion coefficient, and w represents the standard Wiener process. Anderson (1982) showed that the reverse of this process is

also a diffusion process, described by the following reverse-time SDE:

$$\mathrm{d}x = \Big[f(x,\tau) - g(\tau)^2 \nabla_x \log p_\tau(x)\Big]\mathrm{d}\tau + g(\tau)\,\mathrm{d}\bar{w}, \quad (2)$$

where $p_\tau(x)$ is the marginal distribution of the data at diffusion time τ, the term $\nabla_x \log p_\tau(x)$ is the (Stein) score function, and $\bar{w}$ is a reverse-time Wiener process.

To generate data, one must estimate the unknown score function. Denoising score matching (DSM) trains a neural network, a score model $S_\theta(x,\tau)$, for this purpose. The model is trained by sampling a noised datapoint x^τ from the perturbation kernel $p_{0\tau}(x^\tau \mid x^0)$ and minimizing the objective:

$$\ell_{\text{DSM}}(\theta) = \mathbb{E}\big[\|S_\theta(x^\tau,\tau) - \nabla_{x^\tau} \log p_{0\tau}(x^\tau \mid x^0)\|_2^2\big]\,. \quad (3)$$

For a Gaussian perturbation kernel, this objective is equivalent to a denoising regression task. In this formulation, a denoiser network $D_\theta(x^\tau,\tau)$ is trained to predict the original clean data x^0 from its noised version x^τ by minimizing a reconstruction loss:

$$\ell_{\text{recon}}(\theta) = \mathbb{E}_\tau\big[w(\tau)\,\|D_\theta(x^\tau,\tau) - x^0\|_2^2\big]\,, \quad (4)$$

where $w(\tau)$ is the perturbation level (i.e., the standard deviation of the Gaussian noise) at diffusion time τ, and $w(\tau)$ is a positive, τ-dependent weighting function. This weight is often chosen to de-emphasize noise levels where the loss variance is high, thereby stabilizing training.

3.3. Diffusion-based world models and DIAMOND

Diffusion-based world models adapt the generative framework described in Section 3.2 to learn the dynamics of an environment for RL. To achieve this, the model must learn from the agent's history of interactions, which we define as $h_t = (x^0_{\le t}, a_{\le t})$, representing the sequence of past clean observations and actions. The primary objective is to train a model, parameterized by θ, to approximate the conditional probability distribution of the next observation x_{t+1} given this history. This can be formally expressed as learning a model for:

$$p_\theta(x_{t+1} \mid h_t). \quad (5)$$

This is achieved by conditioning the denoising process on the history h_t. The model is trained to reconstruct the clean next observation x^0_{t+1} from a noised version $x^\tau_{t+1} \sim \mathcal{N}(x^0_{t+1}, \sigma(\tau)^2 I)$, yielding the following conditional denoising objective:

$$\ell_{\text{recon}}(\theta) = \mathbb{E}\Big[\big\|D_\theta(x^\tau_{t+1},\tau,x^0_{\le t},a_{\le t}) - x^0_{t+1}\big\|_2^2\Big]\,. \quad (6)$$

Following the approach in DIAMOND, we adopt the Elucidated Diffusion Model (EDM) parameterization for its enhanced stability, especially during sampling with a low number of function evaluations (NFE). In the EDM framework, the denoiser D_θ is parameterized as a function of a core neural network F_θ and several preconditioning coefficients :

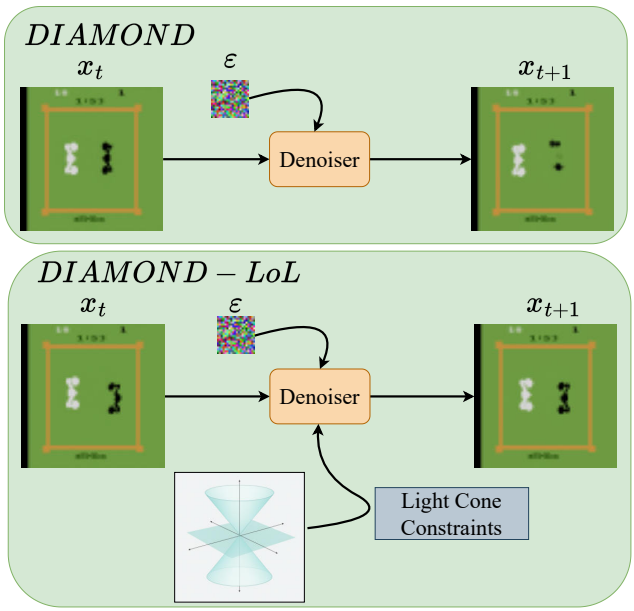


Figure 2. Overview of the DIAMOND-LoL method. Compared with the baseline DIAMOND model, the DIAMOND-LoL we proposed introduces a crucial light cone constraint module. This module implements our LoL loss, by imposing a physical constraint of finite propagation speed, to ensure that the time series generated by the model has better physical consistency.

$$D_\theta(x^\tau_{t+1}, y^\tau_t) = c^\tau_{\text{skip}}\, x^\tau_{t+1} + c^\tau_{\text{out}}\, F_\theta\big(c^\tau_{\text{in}} x^\tau_{t+1},\, y^\tau_t\big), \quad (7)$$

where $y^\tau_t = \big(c^\tau_{\text{noise}}, x^0_{\le t}, a_{\le t}\big)$ encapsulates all conditioning variables, including an embedding of the noise level $\sigma(\tau)$ via c^τ_{noise}. The preconditioners c^τ_{in} and c^τ_{out} are designed to maintain the network's input and output at unit variance across all noise levels. This leads to the final EDM training objective, which we will build upon later:

$$\ell_{\text{EDM}}(\theta) = \mathbb{E}\Big[\Big\|F_\theta(c^\tau_{\text{in}} x^\tau_{t+1}, y^\tau_t) - \tfrac{1}{c^\tau_{\text{out}}}\big(x^0_{t+1} - c^\tau_{\text{skip}} x^\tau_{t+1}\big)\Big\|_2^2\Big]\,. \quad (8)$$

To form a complete agent, the diffusion dynamics model p_θ serves as the core visual simulator. As established previously, it is augmented with a separate reward and termination prediction head R_ψ, while the agent's policy π_ϕ and value function V_ϕ are trained on imagined trajectories generated by this world model.

4. Method

4.1. Limitations of Standard EDM Training

We train a diffusion-based world model on POMDP trajectories to approximate the conditional dynamics $p_\theta(x_{t+1} \mid$

$x^0_{\leq t}, a_{\leq t})$. In practice, most systems adopt the EDM objective 8, which stabilizes sampling under a low number of denoising steps. However, this approach leads to two fundamental inconsistencies with the physical constraints of the environment. Mode averaging arises in multimodal or partially observable settings; minimizing the pixelwise ℓ_2 distance encourages the model to generate interpolated boundaries that do not correspond to any physically realizable state. Second, acausal artifacts appear when these averaged boundaries fall outside the one-step reachable set of the environment's dynamics, effectively teleporting structures. When rolled out autoregressively, these violations of locality are amplified, causing long-horizon trajectory drift and a loss of visual fidelity. In Figure 2, we give the framework of DIAMOND-LoL.

We can formalize this inconsistency as follows. Let the true distribution of the next state, x_{t+1}, be a mixture of discrete modes, where each mode corresponds to a transformation T_{δ_k} (e.g., a translation by a displacement vector δ_k) of the current state x_t with probability π_k. The true data-generating process is thus $x_{t+1} \sim \sum_k \pi_k T_{\delta_k}(x_t)$. A model trained to minimize the expected ℓ_2 error will learn to predict the conditional mean of this distribution:

$$\hat{x}_{t+1} = \mathbb{E}[x_{t+1} \mid x_t] = \sum_k \pi_k T_{\delta_k}(x_t). \tag{9}$$

If the displacement between distinct modes is larger than the maximum physically allowable single-step motion, this expected value $\hat{x}_{t+1}$ will be an interpolation—a pixel-space average that synthesizes new boundaries in locations that no single mode occupies. This formally demonstrates how the ℓ_2 objective compels the model to generate acausal structures that violate the environment's finite propagation speed. These issues motivate a locality constraint that encodes this principle of finite-speed propagation without requiring architectural changes.

4.2. Lieb–Robinson Locality Loss (LoL loss)

Our proposed Lieb-Robinson Locality Loss (LoL loss) enforces a finite-speed propagation constraint on the learned dynamics. Its formulation relies on a set of geometric operators and data-driven parameters. We first define a boundary extractor, $E(x) \in [0,1]^{H\times W\times C}$, using a normalized gradient operator, and a morphological dilation operator, $\mathrm{Dilate}_r(\cdot)$, which expands a boundary set by a given radius r. We also define a static-source mask, M, derived from training data statistics to down-weight exogenous events like HUD updates. More details about r and M are defined in Appendix A. Using these, we estimate a data-driven light-cone radius, r_t, for each timestep by finding the minimal radius that satisfies the locality condition on ground-truth trajectories:

$$\mathrm{supp}\big(E(x_{t+1})\big) \subseteq \mathrm{Dilate}_r\big(\mathrm{supp}(E(x_t))\big),$$

where $\mathrm{supp}(A)$ denotes the support of a tensor A (i.e., the set of indices of its non-zero elements).

Given a predicted frame $\hat{x}_{t+1}$, the LoL loss is composed of two penalties. The emerge penalty, ℓ_{emerge}, quantifies new boundaries that appear acausally outside the light-cone of the previous frame. The vanish penalty, ℓ_{vanish}, quantifies boundaries that disappear without a successor inside the predicted light-cone. Their mathematical forms are:

$$\ell_{\text{emerge}} = \left\|\mathrm{vec}\left(E(\hat{x}_{t+1}) \odot M \odot (\mathbf{1} - \mathrm{Dilate}_{r_t}(E(x_t)))\right)\right\|_1, \tag{10}$$

$$\ell_{\text{vanish}} = \left\|\mathrm{vec}\left(E(x_t) \odot M \odot (\mathbf{1} - \mathrm{Dilate}_{r_t}(E(\hat{x}_{t+1})))\right)\right\|_1, \tag{11}$$

where $\odot$ is the Hadamard (element-wise) product and $\mathrm{vec}(\cdot)$ is the vectorization operator. These two components sum to the total locality loss:

$$\ell_{\text{LoL}} = \ell_{\text{emerge}} + \ell_{\text{vanish}}. \tag{12}$$

To integrate this geometric penalty into the stochastic training of the diffusion model, we introduce a noise-time gating function, $w_{\text{loc}}(\tau) \in [0,1]$. This function is monotone nonincreasing with the noise level $\sigma(\tau)$ and applies the LoL loss penalty predominantly in near-clean regimes. For our experiments, we define it as a simple linear ramp:

$$w_{\text{loc}}(\tau) = \max(0, 1 - \sigma(\tau)/\sigma_{\text{gate}}), \tag{13}$$

where σ_{gate} is a hyperparameter defining the noise level threshold. The overall training objective for our model combines the EDM objective from Equation 8 with our gated LoL loss penalty, weighted by a hyperparameter $\lambda > 0$:

$$\mathcal{L}_{\text{EDM+LoL}}(\theta) = \ell_{\text{EDM}}(\theta) + \lambda\, \mathbb{E}_\tau\big[w_{\text{loc}}(\tau)\, \ell_{\text{LoL}}(\hat{x}_{t+1}; x_t)\big], \tag{14}$$

where the predicted frame $\hat{x}_{t+1} = D_\theta(x^\tau_{t+1}, \tau, x^0_{\leq t}, a_{\leq t})$ used to compute ℓ_{LoL} is the output of the full denoiser defined in Equation 7. This objective enforces finite-speed locality without requiring auxiliary labels or architectural changes.

4.3. Theoretical Guarantees

Our proposed LoL loss is not merely a heuristic; it induces three key theoretical properties that directly address the limitations of the standard ℓ_2 objective. These guarantees, stated here and proven in Appendices B, C and D, provide a formal basis for its effectiveness in promoting physically plausible and stable dynamics.

4.3.1. Zero-Loss Condition and the Finite-Propagation Set.

We establish the exact condition under which the LoL loss vanishes. We define the finite-propagation set, $\mathcal{S}_{r_t}(x_t)$, as the set of all possible next frames whose boundaries are mutually reachable from the boundaries of the current frame x_t

within the light-cone radius r_t:

$$\mathcal{S}_{r_t}(x_t) = \Big\{ z : \operatorname{supp}(E(z)) \subseteq \operatorname{Dilate}_{r_t}(\operatorname{supp}(E(x_t))) \wedge \operatorname{supp}(E(x_t)) \subseteq \operatorname{Dilate}_{r_t}(\operatorname{supp}(E(z))) \Big\}.$$

The theorem states that the LoL loss is zero if and only if the predicted frame $\hat{x}_{t+1}$ belongs to this set, i.e., $\ell_{\text{LoL}}(\hat{x}_{t+1}; x_t) = 0 \iff \hat{x}_{t+1} \in \mathcal{S}_{r_t}(x_t)$. This provides a precise geometric characterization of a physically plausible one-step transition.

4.3.2. Provable Selection of Modes over Averaging.

We prove that the LoL loss resolves the mode-averaging problem. Consider a scenario where the true next state x_{t+1} is drawn from a mixture of discrete modes $\{T_{\delta_k}(x_t)\}_k$, where T_{δ_k} is a transformation by a displacement vector δ_k. If these modes are sufficiently separated (i.e., $\|\delta_i - \delta_j\|_\infty > r_t$), the standard ℓ_2 objective will optimally predict their pixel-space average, resulting in a physically unrealizable state. In contrast, we show that any such average incurs a strictly positive ℓ_{LoL}. The loss is minimized only by selecting a single, valid mode from the mixture, thereby enforcing the generation of crisp and physically plausible outcomes.

4.3.3. Linear Bound on Long-Horizon Error Accumulation.

We demonstrate that enforcing local consistency leads to global, long-horizon stability. If the LoL loss is bounded by a small value ϵ at each step, we prove that the error between the predicted and ground-truth boundary sets, measured by the Hausdorff distance, grows at most linearly with the time horizon τ:

$$d_{\text{H}}\big(\operatorname{supp}(E(\hat{x}_{t+\tau})), \operatorname{supp}(E(x_{t+\tau}))\big) \leq C_0 + C_1\, \tau\, \epsilon, \tag{15}$$

where C_0 and C_1 are constants determined by the dilation geometry. This linear bound contrasts sharply with the potential for exponential error accumulation in models that permit acausal single-step transitions, formally guaranteeing that our method prevents the catastrophic trajectory drift observed in less constrained models.

4.4. Training Pipeline

Our training procedure follows a standard closed-loop, model-based RL pipeline that alternates between three core phases: data collection, world model learning, and policy optimization in imagination. First, the agent interacts with the real environment to collect a dataset of experience trajectories, $\{(x_t, a_t, r_t, \text{done}_t)\}_{t=0}^{T}$. This data is then used to update our diffusion-based world model by minimizing the LoL loss in Equation 14. During this phase, the EDM term ensures high-fidelity reconstruction, while the LoL loss term enforces the physical constraint of finite-speed propagation. The updated world model is employed as a neural simulator to generate vast quantities of imagined rollouts. The agent's policy, π_ϕ, and value function, V_ϕ, are then trained to optimality on this simulated data using an actor-critic algorithm with generalized λ-returns. These phases of interaction, model learning, and imagination-based training are repeated cyclically, yielding sample-efficient learning with high visual fidelity and temporal consistency, even in low-NFE regimes.

5. Experiments

5.1. Experimental Setup

We conduct our primary evaluation on the Atari 100k benchmark [35], a standard for assessing sample efficiency in RL. For each of its 26 games, an agent is permitted only 100,000 environmental steps, roughly two hours of human gameplay, to learn its policy before evaluation. All of our results are averaged over 5 random seeds per game. All models are trained and evaluated on a single NVIDIA RTX 5090D GPU.

To ensure that any observed performance improvements are attributable solely to our proposed loss, the implementation of DIAMOND-LoL is identical to the original DIAMOND in all other aspects. We employ the same core network architectures, RL algorithm, and closed-loop training paradigm. The training procedure runs for 1000 epochs, each epoch consists of 100 environmental steps for data collection (using an ϵ-greedy policy with $\epsilon = 0.01$) followed by 400 training updates with a batch size of 32. For the agent, we use an imagination horizon of 15 steps, a discount factor $\gamma = 0.985$, a λ-return coefficient of 0.95, and an entropy weight of 0.001. The U-Net based DWM conditions on the 4 most recent frames and actions and generates rollouts using the Euler sampler with 3 denoising steps (NFE=3). All model components are trained using the AdamW optimizer with a learning rate of 1e-4; we apply a weight decay of 1e-2 to the world model and reward model, and no weight decay to the actor-critic network. All Atari environments use a frameskip of 4, provide $64 \times 64 \times 3$ pixel observations, and have rewards clipped to the set $\{-1, 0, 1\}$. The total loss function is $\mathcal{L}_{\text{EDM+LoL}} = \ell_{\text{EDM}} + \lambda \ell_{\text{LoL}}$, where the locality weight λ is the only new hyperparameter, set to 10^{-2} based on our sensitivity analysis.

5.2. Comparative Analysis

In the Atari 100k benchmark test [35], we compared and evaluated the performance of DIAMOND-LoL with the previously most advanced world model. Table 1 shows the specific scores and comprehensive indicators for each game and we report the results of DIAMOND in [1]. The results indicate that our method outperformed DIAMOND. It is notable that the average human normalized score (HNS) of DIAMOND-LoL was 1.581, and the interquartile mean

Table 1. Returns on the 26 games of the Atari 100k benchmark after 2 hours of real-time experience, and human-normalized aggregate metrics. Bold numbers indicate the best performing methods. DIAMOND-LoL notably achieves the highest mean score over 5 seeds.

Game	Random	Human	SimPLe [35]	TWM [50]	IRIS [46]	DreamerV3 [27]	STORM [65]	DIAMOND [1]	DIAMOND-LoL (ours)
Alien	227.8	7127.7	616.9	674.6	420.0	959.0	983.6	744.1	**1021.5**
Amidar	5.8	1719.5	74.3	121.8	143.0	139.0	204.8	225.8	**231.4**
Assault	222.4	742.0	527.2	682.6	1524.4	706.0	801.0	1526.4	**1598.2**
Asterix	210.0	8503.3	1128.3	1116.6	853.6	932.0	1028.0	3698.5	**4102.1**
BankHeist	14.2	753.1	34.2	466.7	53.1	**649.0**	641.2	19.7	20.5
BattleZone	2360.0	37187.5	4031.2	5068.0	13074.0	12250.0	**13540.0**	4702.0	4688.0
Boxing	0.1	12.1	7.8	77.5	70.1	78.0	79.7	86.9	**89.2**
Breakout	1.7	30.5	16.4	20.0	83.7	31.0	15.9	132.5	**165.3**
ChopperCommand	811.0	7387.8	979.4	1697.4	1565.0	420.0	**1888.0**	1369.8	1402.7
CrazyClimber	10780.5	35829.4	62583.6	71820.4	59324.2	97190.0	66776.0	99167.8	**101450.3**
Demon Attack	152.1	1971.0	208.1	350.2	**2034.4**	303.0	164.6	288.1	295.6
Freeway	0.0	29.6	16.7	24.3	31.1	0.0	33.5	33.3	**33.6**
Frostbite	65.2	4334.7	236.9	**1475.6**	259.1	909.0	1316.0	274.1	280.9
Gopher	257.6	2412.5	596.8	1674.8	2236.1	3730.0	**8239.6**	5897.9	6015.4
Hero	1027.0	30826.4	2656.6	7254.0	7037.4	**11161.0**	11044.3	5621.8	5590.1
Jamesbond	29.0	302.8	100.5	362.4	462.7	445.0	**509.0**	427.4	433.8
Kangaroo	52.0	3035.0	51.2	1240.0	838.2	4098.0	4208.0	5382.2	**5421.7**
Krull	1598.0	2665.5	2204.8	6349.2	6616.4	7782.0	8412.6	8610.1	**8695.3**
KungFuMaster	258.5	22736.3	14862.5	24554.6	21759.8	21420.0	**26182.0**	18713.6	19004.2
MsPacman	307.3	6951.6	1480.0	1588.4	999.1	1327.0	**2673.5**	1958.2	2011.6
Pong	-20.7	14.6	12.8	18.8	14.6	18.0	11.3	20.4	**20.5**
Private Eye	24.9	69571.3	35.0	86.6	100.0	882.0	**7781.0**	114.3	119.8
Qbert	163.9	13455.0	1288.8	3330.8	745.7	3405.0	**4522.5**	4499.3	4520.7
RoadRunner	11.5	7845.0	5640.6	9109.0	9614.6	15565.0	17564.0	20673.2	**22541.6**
Seaquest	68.4	42054.7	683.3	**774.4**	661.3	618.0	525.2	551.2	560.3
UpNDown	533.4	11693.2	3350.3	**15981.7**	3546.2	9234.0	7985.0	3856.3	3888.1
#Superhuman (↑)	0	N/A	1	8	10	9	10	11	**13**
Mean (↑)	0.000	1.000	0.332	0.956	1.046	1.097	1.266	1.459	**1.581**
IQM (↑)	0.000	1.000	0.130	0.459	0.501	0.497	0.636	0.641	**0.695**

(IQM) was 0.695, surpassing all benchmarks, including the direct predecessor DIAMOND and the previous leading method STORM. Moreover, our agent achieved superhuman performance in 13 out of 26 games, outperforming any other method.

This significant performance improvement is entirely attributed to the introduction of LoL loss. All other aspects of the experimental setup remained unchanged. The standard ℓ_2 objective would compress the multimodal uncertainty into a single, physically unreasonable average value. Our LoL loss counteracts this phenomenon by implementing finite propagation constraints, forcing the model to select a physically coherent future from a series of possibilities. This advantage is particularly evident in games such as "Breakout", "Asterix", and "Racer", where physically coherent modeling of fast-moving objects provides a more stable and reliable imagination environment for the agents. By preventing non-causal instantiation errors and ensuring that the evolution of all game elements is physically reasonable, the LoL loss promotes the learning of better strategies, ultimately leading to higher game scores. Moveover, we provide a comparison with DIAMOND and DIAMOND-LoL in Figure 3.

5.3. Sensitivity Analysis

A key component of our method is the locality weight hyperparameter, λ, which balances the influence of our proposed geometric loss against the standard EDM reconstruction loss. To evaluate the robustness of DIAMOND-LoL to the choice of this hyperparameter, we conduct a sensitivity analysis. We select "Boxing" and "breakout", high determinism with fast objects games. We train our agent from scratch on these games while varying λ across several orders of magnitude, including the baseline case: $\{0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. A λ value of 0 is equivalent to the original DIAMOND model. All other hyperparameters are held constant as described in the experimental setup.

Figure 4 shows a consistent and informative trend. When λ is too small (10^{-4}), the locality constraint is too weak to have a significant effect, and performance is comparable to the original DIAMOND baseline ($\lambda = 0$). As λ increases to 10^{-3} and 10^{-2}, we observe a notable improvement in scores, indicating that the LoL loss is effectively regularizing the model. However, when λ becomes too large (10^{-1}), performance begins to decline. This suggests that an overly strong locality penalty can compromise the generative fidelity of the EDM term, making the model too rigid. This analysis demonstrates that the benefits of our LoL loss are robust across a reasonable range for λ, with an

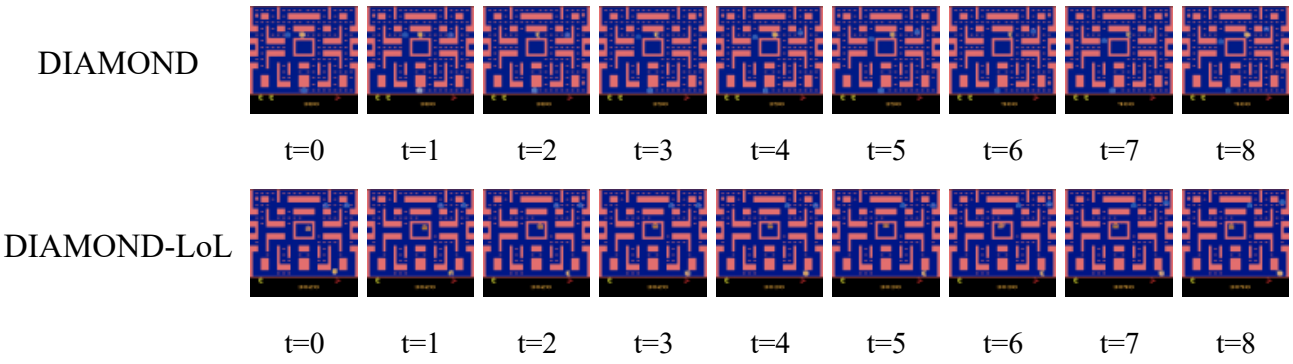
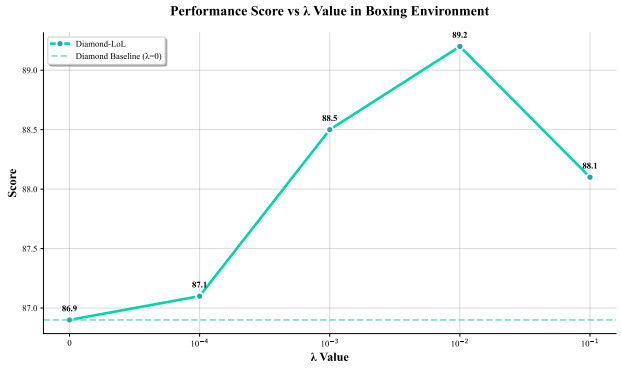
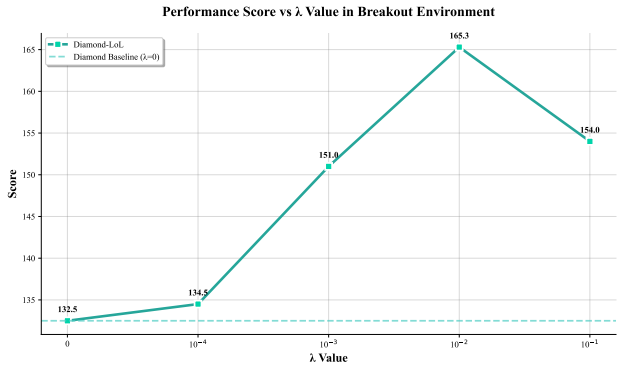


Figure 3. Qualitative comparison of long-horizon rollouts generated by DIAMOND and our DIAMOND-LoL. The figure displays two 9-step imagined trajectories (t=0 to t=8) from the Breakout environment.



(a) Boxing



(b) Breakout

Figure 4. DIAMOND-LoL scores in the two games change with the variation of the λ value.

optimal value that provides a clear advantage over the baseline. Based on these findings, we use a value of $\lambda = 10^{-2}$ for all other experiments in this paper.

6. Complexity Analysis

Our proposed LoL loss is designed for computational efficiency, introducing zero overhead at inference time. Since the loss is only computed during training and the network architecture is unchanged, the per-frame sampling cost is identical to the baseline. During training, our method adds a marginal overhead from boundary extraction, morphological dilation, and loss aggregation, which scales linearly with the number of pixels, $\mathcal{O}(B{\cdot}H{\cdot}W{\cdot}C)$. This cost is negligible compared to the backpropagation through the main denoiser network and is further reduced on average by a noise-gating function that applies the loss selectively. Similarly, the additional space complexity for storing transient tensors is also a negligible $\mathcal{O}(B \cdot H \cdot W \cdot C)$, and no new learnable parameters are introduced. Thus, in exchange for a minimal increase in training cost, DIAMOND-LoL gains a significant, theoretically-backed improvement in long-horizon stability, providing a highly efficient method for enforcing physical consistency in diffusion world models.

7. Conclusion

In this work, we addressed the physical inconsistency of the standard ℓ_2 objective in diffusion-based world models, which leads to acausal artifacts and mode-averaging failures. We introduced DIAMOND-LoL, a novel training framework that incorporates a LoL loss to enforce a finite-speed propagation constraint. Our theoretically-grounded approach compels the model to select physically plausible futures rather than generating unrealizable averages. This results in a great performance on the Atari 100k benchmark by providing a more stable and reliable imagined environment for the agent. Our research highlights the value of integrating fundamental physical priors directly into the training objective of generative models, a principle that could be extended to more complex domains like robotics to create more robust autonomous agents.

References

[1] Eloi Alonso, Adam Jelley, Vincent Micheli, Anssi Kanervisto, Amos J Storkey, Tim Pearce, and François Fleuret.

Diffusion for world modeling: Visual details matter in atari. *Advances in Neural Information Processing Systems*, 37: 58757–58791, 2024. 2, 3, 6, 7

[2] Mohammad Babaeizadeh, Chelsea Finn, Dumitru Erhan, Roy H Campbell, and Sergey Levine. Stochastic variational video prediction. *arXiv preprint arXiv:1710.11252*, 2017. 2

[3] Mohammad Babaeizadeh, Chelsea Finn, Dumitru Erhan, Roy H. Campbell, and Sergey Levine. Stochastic variational video prediction, 2017. 3

[4] Jiesong Bai, Yuhao Yin, Yihang Dong, Xiaofeng Zhang, Chiman Pun, and Xuhang Chen. Lensnet: An end-to-end learning framework for empirical point spread function modeling and lensless imaging reconstruction. In *IJCAI*, pages 684–692, 2025. 1

[5] Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738, 2021. 3

[6] Miguel Calvo-Fullana, Santiago Paternain, Luiz FO Chamon, and Alejandro Ribeiro. State augmented constrained reinforcement learning: Overcoming the limitations of learning with rewards. *IEEE Transactions on Automatic Control*, 69(7):4275–4290, 2023. 1

[7] Lluis Castrejon, Nicolas Ballas, and Aaron Courville. Improved conditional vrnns for video prediction, 2019. 3

[8] Lluis Castrejon, Nicolas Ballas, and Aaron Courville. Improved conditional vrnns for video prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7608–7617, 2019. 2

[9] Haoqiang Chen, Yadong Liu, and Dewen Hu. Representation learning for vision-based autonomous driving via probabilistic world modeling. *Machines*, 13(3):231, 2025. 1

[10] Xuhang Chen, Baiying Lei, Chi-Man Pun, and Shuqiang Wang. Brain diffuser: An end-to-end brain image to brain network pipeline. In *PRCV*, pages 16–26, 2023. 1

[11] Xuhang Chen, Shenghong Luo, Chi-Man Pun, and Shuqiang Wang. Medprompt: Cross-modal prompting for multi-task medical image translation. In *PRCV*, pages 61–75, 2024.

[12] Xuhang Chen, Zhuo Li, Yanyan Shen, Mufti Mahmud, Hieu Pham, Chi-Man Pun, and Shuqiang Wang. High-fidelity functional ultrasound reconstruction via a visual auto-regressive framework. *arxiv*, 2025.

[13] Xuhang Chen, Michael Kwok-Po Ng, Kim-Fung Tsang, Chi-Man Pun, and Shuqiang Wang. Connectomediffuser: Generative ai enables brain network construction from diffusion tensor imaging. *IEEE Transactions on Consumer Electronics*, 2025. 1

[14] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *Advances in neural information processing systems*, 31, 2018. 1

[15] Emily Denton and Rob Fergus. Stochastic video generation with a learned prior. In *International conference on machine learning*, pages 1174–1183. PMLR, 2018. 2

[16] Remi Denton and Rob Fergus. Stochastic video generation with a learned prior, 2018. 3

[17] Jingtao Ding, Yunke Zhang, Yu Shang, Yuheng Zhang, Zefang Zong, Jie Feng, Yuan Yuan, Hongyuan Su, Nian Li, Nicholas Sukiennik, et al. Understanding world or predicting future? a comprehensive survey of world models. *ACM Computing Surveys*, 2024. 1

[18] Zihan Ding, Amy Zhang, Yuandong Tian, and Qinqing Zheng. Diffusion world model: Future modeling beyond step-by-step rollout for offline reinforcement learning. *arXiv preprint arXiv:2402.03570*, 2024. 2

[19] Tuo Feng, Wenguan Wang, and Yi Yang. A survey of world models for autonomous driving. *arXiv preprint arXiv:2501.11260*, 2025. 1

[20] Yunhai Feng, Nicklas Hansen, Ziyan Xiong, Chandramouli Rajagopalan, and Xiaolong Wang. Finetuning offline world models in the real world. *arXiv preprint arXiv:2310.16029*, 2023. 1

[21] C. Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax – a differentiable physics engine for large scale rigid body simulation, 2021. 3

[22] Changwei Gong, Changhong Jing, Xuhang Chen, Chi Man Pun, Guoli Huang, Ashirbani Saha, Martin Nieuwoudt, Han-Xiong Li, Yong Hu, and Shuqiang Wang. Generative ai for brain image computing and brain network computing: a review. *Frontiers in Neuroscience*, 17:1203104, 2023. 1

[23] Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction, 2020. 3

[24] Xiaojiao Guo, Shenghong Luo, Yihang Dong, Zexiao Liang, Zimeng Li, Xiujun Zhang, and Xuhang Chen. An asymmetric calibrated transformer network for underwater image restoration: X. guo et al. *The Visual Computer*, pages 1–13, 2025. 1

[25] David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2(3), 2018. 1

[26] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International conference on machine learning*, pages 2555–2565. PMLR, 2019. 1

[27] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, pages 1–7, 2025. 7

[28] Haoran He, Yang Zhang, Liang Lin, Zhongwen Xu, and Ling Pan. Pre-trained video generative models as world simulators. *arXiv preprint arXiv:2502.07825*, 2025. 2

[29] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2

[30] Taylor A. Howell, Simon Le Cleac'h, Jan Brüdigam, Qianzhong Chen, Jiankai Sun, J. Zico Kolter, Mac Schwager, and Zachary Manchester. Dojo: A differentiable physics engine for robotics, 2022. 3

[31] Siqiao Huang, Jialong Wu, Qixing Zhou, Shangchen Miao, and Mingsheng Long. Vid2world: Crafting video diffusion models to interactive world models. *arXiv preprint arXiv:2505.14357*, 2025. 2

[32] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. *arXiv preprint arXiv:2205.09991*, 2022. 2

[33] Krishna Murthy Jatavallabhula, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jerome Parent-Levesque, Kevin Xie, Kenny Erleben, Liam Paull, Florian Shkurti, Derek Nowrouzezahrai, and Sanja Fidler. gradsim: Differentiable simulation for system identification and visuomotor control. *International Conference on Learning Representations (ICLR)*, 2021. 3

[34] Wonsuhk Jung, Utkarsh A Mishra, Nadun Ranawaka Arachchige, Yongxin Chen, Danfei Xu, and Shreyas Kousik. Joint model-based model-free diffusion for planning with constraints. *arXiv preprint arXiv:2509.08775*, 2025. 2

[35] Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, Roy H Campbell, Konrad Czechowski, Dumitru Erhan, Chelsea Finn, Piotr Kozakowski, Sergey Levine, et al. Model-based reinforcement learning for atari. *arXiv preprint arXiv:1903.00374*, 2019. 1, 2, 6, 7

[36] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. 3

[37] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models, 2022. 3

[38] Yingtie Lei, Fanghai Yi, Yihang Dong, Weihuang Liu, Xiaofeng Zhang, Zimeng Li, Chi-Man Pun, and Xuhang Chen. Cmamrnet: A contextual mask-aware network enhancing mural restoration through comprehensive mask guidance. *arXiv*, 2025. 1

[39] Mingxian Li, Hao Sun, Yingtie Lei, Xiaofeng Zhang, Yihang Dong, Yilin Zhou, Zimeng Li, and Xuhang Chen. High-fidelity document stain removal via a large-scale real-world dataset and a memory-augmented transformer. In *WACV*, pages 7614–7624, 2025. 1

[40] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2020. 3

[41] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations, 2021.

[42] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. 3

[43] Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning, 2019. 3

[44] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error, 2015. 3

[45] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015. 2

[46] Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are sample-efficient world models. *arXiv preprint arXiv:2209.00588*, 2022. 1, 2, 7

[47] Vincent Micheli, Eloi Alonso, and François Fleuret. Efficient world models with context-aware tokenization. *arXiv preprint arXiv:2406.19320*, 2024. 1

[48] Chaojun Ni, Guosheng Zhao, Xiaofeng Wang, Zheng Zhu, Wenkang Qin, Xinze Chen, Guanghong Jia, Guan Huang, and Wenjun Mei. Recondreamer-rl: Enhancing reinforcement learning via diffusion-based scene reconstruction. *arXiv preprint arXiv:2508.08170*, 2025. 2

[49] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019. 3

[50] Jan Robine, Marc Höftmann, Tobias Uelwer, and Stefan Harmeling. Transformer-based world models are happy with 100k interactions. *arXiv preprint arXiv:2303.07109*, 2023. 7

[51] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[52] Aidan Scannell, Mohammadreza Nakhaei, Kalle Kujanpää, Yi Zhao, Kevin Sebastian Luck, Arno Solin, and Joni Pajarinen. Discrete codebook world models for continuous control. *arXiv preprint arXiv:2503.00653*, 2025. 1

[53] Tadahiro Taniguchi, Shingo Murata, Masahiro Suzuki, Dimitri Ognibene, Pablo Lanillos, Emre Ugur, Lorenzo Jamone, Tomoaki Nakamura, Alejandra Ciria, Bruno Lara, et al. World models and predictive coding for cognitive and developmental robotics: frontiers and challenges. *Advanced Robotics*, 37(13):780–806, 2023. 1

[54] Xin Tian, Yingtie Lei, Xiujun Zhang, Zimeng Li, Chi-Man Pun, and Xuhang Chen. Sformer: Snr-guided transformer for underwater image enhancement from the frequency domain. *arXiv*, 2025. 1

[55] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 1

[56] Ruben Villegas, Jimei Yang, Yuliang Zou, Sungryull Sohn, Xunyu Lin, and Honglak Lee. Learning to generate long-term future via hierarchical prediction. In *international conference on machine learning*, pages 3560–3569. PMLR, 2017. 2

[57] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification, 2017. 3

[58] Jingchao Wang, Zhengnan Deng, Tongxu Lin, Wenyuan Li, and Shaobin Ling. A novel prompt tuning for graph transformers: Tailoring prompts to graph topologies. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3116–3127, 2024. 1

[59] Jingchao Wang, Guoheng Huang, Xiaochen Yuan, Guo Zhong, Tongxu Lin, Chi-Man Pun, and Fenfang Xie. The structure-sharing hypergraph reasoning attention module for cnns. *Expert Systems with Applications*, 259:125240, 2025. 1

[60] Yuandi Wu, Brett Sicard, and Stephen Andrew Gadsden. Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications*, 255:124678, 2024. 3

[61] Fanghai Yi, Zehong Zheng, Zexiao Liang, Yihang Dong, Xiyang Fang, Wangyu Wu, and Xuhang Chen. Mac-lookup: Multi-axis conditional lookup model for underwater image enhancement. *arXiv*, 2025. 1

[62] Dong Yin, Sridhar Thiagarajan, Nevena Lazic, Nived Rajaraman, Botao Hao, and Csaba Szepesvari. Sample efficient deep reinforcement learning via local planning. *arXiv preprint arXiv:2301.12579*, 2023. 1

[63] Yang Yu. Towards sample efficient reinforcement learning. In *IJCAI*, pages 5739–5743, 2018. 1

[64] Ye Yuan, Jiaming Song, Umar Iqbal, Arash Vahdat, and Jan Kautz. Physdiff: Physics-guided human motion diffusion model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16010–16021, 2023. 2

[65] Weipu Zhang, Gang Wang, Jian Sun, Yetian Yuan, and Gao Huang. Storm: Efficient stochastic transformer based world models for reinforcement learning. *Advances in Neural Information Processing Systems*, 36:27147–27166, 2023. 7