Physics-Aware Reinforcement Learning for Object Localization

Prithvi Raj Singh[1] , Anthony S Maida[2]
[1]McNeese State University [2]UL Lafayette

Abstract

Deep object detectors such as Faster R-CNN and EfficientDet achieve strong performance on static images but often produce spatially inconsistent or unstable localizations under moderate motion or viewpoint change. To address this limitations, we propose a **Physics-Aware Reinforcement Learning (RL)** framework for object localization that embeds a kinematic motion prior directly into the learning process. The key insight is that physical motion—expressed as constant-velocity dynamics or its learnable extension—can serve as a guiding constraint for bounding-box refinement, enabling the detector to produce more stable and temporally consistent localizations.

Keywords: Reinforcement Learning, Object Localization, Physics-Aware Model, PPO, Visual Detection & Tracking, Motion Consistency

1 Problem Formulation

We formulate bounding-box refinement as a Markov Decision Process (MDP), where an RL agent iteratively adjusts the box parameters to maximize an IoU-based reward. A physics module predicts the next box position using constant-velocity dynamics, while a policy network learns residual corrections to this prediction. The agent is trained using Proximal Policy Optimization (PPO), balancing the objective of improved IoU with a penalty for deviating from physically plausible motion. This combination produces a hybrid model that fuses perception and motion reasoning in a unified learning loop.

The **state** s_t encodes visual features extracted from the current bounding box region, positional parameters (x_t, y_t, w_t, h_t), and optionally the previous motion vector. The **action** a_t predicts a parameter adjustment values $(\Delta x, \Delta y, \Delta w, \Delta h)$ applied to the current box, while the **reward** r_t balances localization accuracy and motion plausibility:

$$r_t = \alpha\,(\mathrm{IoU}_t - \mathrm{IoU}_{t-1}) - \lambda_{\text{phys}} \|b_t - \hat{b}_t^{\text{phys}}\|_2,$$

where $\hat{b}_t^{\text{phys}}$ is the box predicted by a constant-velocity kinematic model, and λ_{phys} controls adherence to the physics prior. The RL agent is optimized using Proximal Policy Optimization (PPO), jointly learning to refine detections and respect motion constraints.

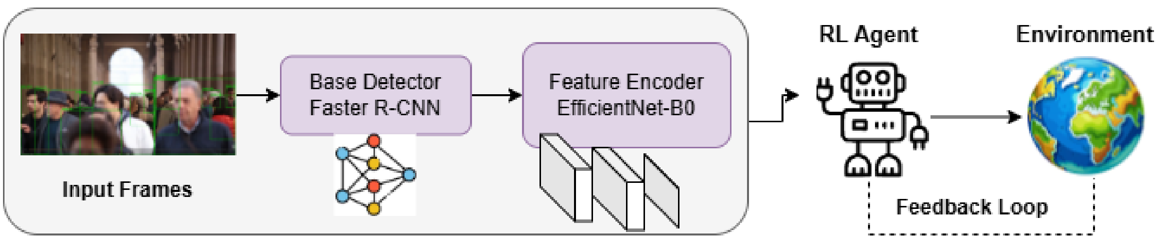

Figure 1: Overview of our proposed framework for Object Localization. The RL agent refines bounding-box predictions from a pre-trained detector by integrating a physics module that enforces motion consistency. The policy network learns to adjust box parameters while adhering to physical dynamics, resulting in improved localization accuracy and temporal stability.

We implement the framework using a pre-trained Faster R-CNN as a base object detector and an EfficientNet-B0 backbone as a feature encoder. A constant-velocity model provides the physical prior. We evaluate the method on the Pascal VOC and VisDrone dataset, comparing four variants: (1) Detector-only baseline, (2) Heuristic refinement policy, (3) PPO without physics, and (4) Physics-aware PPO. Preliminary experiments show that the physics-aware model improves mean IoU and success rate (SR@0.5/0.7) over both PPO and detector-only baselines, while maintaining stable bounding-box transitions and higher average rewards. The framework operates efficiently on existing detectors without retraining them end-to-end, demonstrating its plug-and-play flexibility.

2 Initial Results and Analysis

Table 1 shows the result of using Faster R-CNN, our base detector, and Heuristic Policy on PASCAL VOC test set. The Heuristic Policy shows significant improvements in Mean IoU, Success Rates, and mAP over the baseline detector. This shows that full PPO and Physics-Aware PPO methods are likely to yield further gains - accurate object localization, better average reward and temporal stability. We also believe our proposed physics-aware PPO will exhibit faster convergence and higher reward stability during training. We intent to also present ablation studies on the effect of varying λ_{phys} on localization performance and motion consistency.

Method	**Mean IoU**	**∆IoU**	**SR@0.5**	**SR@0.7**	**mAP(0.5:0.95)**
Detector Only	0.487	–	50.6%	39%	0.425
Heuristic Policy	0.541	+0.109	53.6%	41.3%	0.482

Table 1: Initial results on Pascal VOC dataset comparing Detector Only and Heuristic Policy methods.

3 Evaluation Metrics

We use standard object detection metrics including mean Intersection-over-Union (IoU), Success Rate (SR) at thresholds 0.5 and 0.7, and mean Average Precision (mAP) over IoU thresholds from 0.5 to 0.95. These metrics quantify both localization accuracy and temporal consistency of bounding-box predictions across video sequences. Additionally, we monitor average reward during training to assess learning progress and policy stability. ∆IoU indicates improvement in IoU between refined and detector-only bounding boxes. It is expressed as: ΔIoU = IoU$_{final}$ - IoU$_{init}$.

4 Expected Contributions and Future Work

Beyond its immediate gains, this study also aims to highlights the role of physical consistency as an inductive bias in vision-based reinforcement learning. By coupling motion models with learned policies, we seek to bridge the gap between data-driven detection and physically grounded prediction. Our immediate future work will extend this formulation to a fully differentiable physics model, enabling the system to adaptively infer motion dynamics from data rather than relying on fixed priors. This research paves the way toward physics-consistent perception for robust visual tracking and dynamic scene understanding. We also plan to extend the framework to multi-object scenarios, where inter-object dynamics and occlusions present additional challenges for localization. We intend on releasing a smaller models capabale of running on edge devices to facilitate real-world applications such as drone-based surveillance and autonomous navigation.

Code and Framework. The complete implementation includes modular components for environment simulation, PPO training, heuristic baselines, and evaluation tools (`eval_map_voc.py`) supporting COCO-style metrics. The source code and supplementary materials will be made publicly available upon acceptance.

Note: Our hypothesis and experimental setup as well as co-authorship might change.