V-OCBF: Learning Safety Filters from Offline Data via Value-Guided Offline Control Barrier Functions

Mumuksh Tayal[1*], **Manan Tayal**[1*], **Ravi Prakash**[1]

[1]Centre for Cyber Physical System, IISc Bangalore
mumukshatayal@iisc.ac.in, manantayal@iisc.ac.in, ravipr@iisc.ac.in

Abstract

Ensuring safety in autonomous systems requires controllers that satisfy hard, state-wise constraints without relying on online interaction. While existing Safe Offline RL methods typically enforce soft expected-cost constraints, they do not guarantee forward invariance. Conversely, Control Barrier Functions (CBFs) provide rigorous safety guarantees but usually depend on expert-designed barrier functions or full knowledge of the system dynamics. We introduce Value-Guided Offline Control Barrier Functions (V-OCBF), a framework that learns a neural CBF entirely from offline demonstrations. Unlike prior approaches, V-OCBF does not assume access to the dynamics model; instead, it derives a recursive finite-difference barrier update, enabling model-free learning of a barrier that propagates safety information over time. Moreover, V-OCBF incorporates an expectile-based objective that avoids querying the barrier on out-of-distribution actions and restricts updates to the dataset-supported action set. The learned barrier is then used with a Quadratic Program (QP) formulation to synthesize real-time safe control. Across multiple case studies, V-OCBF yields substantially fewer safety violations than baseline methods while maintaining strong task performance, highlighting its scalability for offline synthesis of safety-critical controllers without online interaction or hand-engineered barriers.

Introduction

Ensuring the safety of autonomous systems is essential for their reliable and widespread deployment. From household service robots to autonomous vehicles and aerial drones, these systems increasingly operate in complex and unstructured environments where unsafe behavior can lead to irreversible consequences. As autonomy becomes deeply integrated into transportation, manufacturing, and healthcare, guaranteeing that such systems operate within well-defined safety boundaries is critical for reliability, and long-term adoption.

Reinforcement learning (RL) has emerged as a powerful paradigm for enabling autonomous systems to acquire sophisticated control behaviors. However, in safety-critical domains, naïve RL exploration can be hazardous. Although constrained RL (CRL) (Achiam et al. 2017; Altman 2021; Alshiekh et al. 2018; Zhao et al. 2023) methods attempt to incorporate safety constraints during learning, they typically require extensive online interaction with the environment. However, most previous studies focus on online RL setting (Liu et al. 2024), which suffers from serious safety issues in both training and deployment phases, especially for scenarios that lack high-fidelity simulators and require real system interaction for policy learning. As a result, there is a growing interest in synthesizing safe policies using offline RL or imitation learning (Levine et al. 2020; Kumar et al. 2020). Nevertheless, most online and offline safe RL approaches (Xu, Zhan, and Zhu 2022; Ciftci et al. 2024; Stooke, Achiam, and Abbeel 2020) treat safety as a *soft constraint* and regulate only the expected cumulative constraint violations. Such probabilistic constraints are insufficient for applications that demand strict state-wise safety, where even a single violation is unacceptable. Furthermore, jointly optimizing performance and safety from static datasets often leads to unstable training dynamics and overly conservative behavior, particularly when safety-critical transitions are sparsely represented(Lee et al. 2022).

Control-theoretic tools provide an alternative and more rigorous foundation for safety. In particular, Control Barrier Functions (CBFs) (Ames, Grizzle, and Tabuada 2014) offer a principled mechanism to enforce hard, instantaneous safety constraints by guaranteeing the forward invariance of a prescribed safe set. When combined with learning-based controllers, CBFs serve as minimally invasive safety filters that adjust nominal actions only when necessary to prevent constraint violations. Their integration with Quadratic Program (QP) based controllers enables real-time implementation with modern optimization solvers. Consequently, CBF-based controllers have been successfully applied to a wide range of safety-critical tasks, including adaptive cruise control (Ames, Grizzle, and Tabuada 2014), aerial robotics (Wu and Sreenath 2016; Tayal et al. 2024a), and legged locomotion (Nguyen and Sreenath 2015). In all of these applications, the performance and safety guarantees fundamentally depend on the quality of the underlying CBF.

Constructing valid CBFs, however, is a challenging problem. Hand-crafting barrier functions requires deep system knowledge and does not scale well to high-dimensional or partially known dynamical systems. This has motivated significant interest in Neural Control Barrier Functions

*These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

(NCBFs), which leverage the expressive power of neural networks to approximate complex safe sets. A variety of techniques have been proposed for learning NCBFs, including SMT-based synthesis (Abate et al. 2021, 2020), mixed-integer programming (Zhao et al. 2022), nonlinear optimization (Zhang et al. 2023), and loss-based training methods (Dawson et al. 2022; Dawson, Gao, and Fan 2023; Tayal et al. 2024b, 2025). Other recent approaches learn CBFs from value functions associated with nominal policies (So et al. 2024). However, most of these methods rely on online interaction to collect informative samples or refine the barrier, which is often infeasible in safety-critical settings.

Recent work has explored learning Control Barrier Functions (CBFs) from offline demonstrations (Robey et al. 2020; Castañeda et al. 2023; Tabbara and Sibai 2025). Existing methods either fit CBFs only on expert trajectories or rely on data-likelihood measures to filter unsafe samples, which limits their ability to generalize beyond the demonstrated states. Uncertainty-aware approaches address distributional mismatch but often become overly conservative. Overall, current offline CBF learning methods are closely tied to the empirical data distribution and do not explicitly reason about future system evolution, resulting in conservative safety guarantees.

This paper introduces Value-Guided Offline Control Barrier Functions (V-OCBF), a novel framework designed to overcome key limitations of existing offline RL and CBF-based approaches. We derive a model-free finite-difference recursion for updating the barrier function, and we show that satisfying this update provides a formal one-step safety guarantee for any control-affine system under the resulting policy. In addition, we propose an expectile-based learning objective that allows the synthesized safe policy to improve over the behavior policy in the dataset while never querying the barrier on out-of-distribution actions, ensuring stable and reliable offline learning. To summarise, the main contributions of this work are as follows:

1. We propose V-OCBF, a framework for learning formally safe controllers entirely from offline demonstrations.
2. We derive a model-free finite-difference barrier recursion and prove that adherence to this update guarantees one-step forward invariance for any control-affine system.
3. We introduce an expectile-based objective that improves upon the behavior policy without evaluating the barrier outside the dataset action support.
4. Across diverse systems, including high-dimensional Safety Gymnasium (Ji et al. 2023) tasks, V-OCBF consistently outperforms constrained offline RL and neural CBF baselines in both safety and reward.

Problem Formulation

We consider a control-affine nonlinear dynamical system defined by the state $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$, the control input $u(t) \in \mathbb{U} \subseteq \mathbb{R}^m$, and governed by the dynamics:

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are locally Lipschitz continuous functions. We are given a set $\mathcal{C} \subseteq \mathcal{X}$ that represents the *safe states* for the system and a failure set $\mathcal{F} \subseteq \mathcal{X}$ that represents the set of unsafe states for the system (e.g., obstacles for an autonomous ground robot). Furthermore, the system is controlled by a Lipschitz continuous control policy $\pi : \mathbb{R}^n \to \mathbb{R}^m$. Our focus lies in ensuring the safety of this dynamical system, which is formally defined as follows:

Definition 1 (Safety)**.** *A dynamical system is considered safe if the set, $\mathcal{C} \subseteq \mathcal{X} \subseteq \mathbb{R}^n$, is positively invariant under the control policy, π, i.e, $x(0) \in \mathcal{C}, u(t) = \pi(x(t)) \implies x(t) \in \mathcal{C},\ \forall\, t \geq 0$.*

Since, $\mathcal{F} \subseteq \mathcal{X} \setminus \mathcal{C}$, it can be trivially shown that $x(t) \in \mathcal{C} \implies x(t) \notin \mathcal{F}\ \forall\, t \geq 0$. Using this premise, we define the main objective of this paper:

Objective 1. *Our objective is to synthesize a safe policy $\pi_{\text{safe}} : [t, T) \times \mathcal{X} \to \mathbb{U}$ such that the resulting closed-loop system satisfies the positive invariance property specified in Definition* (1)*.*

Control Barrier Functions

Control Barrier Functions (Ames, Grizzle, and Tabuada 2014; Ames et al. 2017) are widely used to synthesize control policies with positive invariance guarantees, thereby ensuring system safety. The initial step in constructing a Control Barrier Function (CBF) involves defining a continuously differentiable function $B : \mathcal{X} \to \mathbb{R}$, where the *super-level set* of B corresponds to the safe region $\mathcal{C}$. This leads to the following representation:

$$\mathcal{C} = \{x \in \mathcal{X} : B(x) \geq 0\}, \quad \mathcal{X} \setminus \mathcal{C} = \{x \in \mathcal{X} : B(x) < 0\}. \tag{2}$$

The interior and boundary of $\mathcal{C}$ are further specified as:

$$\text{Int}(\mathcal{C}) = \{x \in \mathcal{X} : B(x) > 0\}, \quad \partial\mathcal{C} = \{x \in \mathcal{X} : B(x) = 0\}. \tag{3}$$

The function h qualifies as a valid Control Barrier Function if it satisfies the following definition:

Definition 2 ((Ames et al. 2017))**.** *Given a control-affine system $\dot{x} = f(x) + g(x)u$, the set $\mathcal{C}$ defined by* (2)*, with $\frac{\partial B}{\partial x}(x) \neq 0$ for all $x \in \partial\mathcal{C}$, the function B is called the Control Barrier Function (CBF) defined on the set $\mathcal{X}$, if there exists an extended class-$\mathcal{K}$ function κ such that for all $x \in \mathcal{X}$:*

$$\max_{u\in\mathbb{U}} \Big[\underbrace{\mathcal{L}_f B(x) + \mathcal{L}_g B(x)u}_{\dot{B}(x,u)} + \kappa\left(B(x)\right) \Big] \geq 0, \tag{4}$$

where $\mathcal{L}_f B(x) = \frac{\partial B}{\partial x} f(x)$ and $\mathcal{L}_g B(x) = \frac{\partial B}{\partial x} g(x)$ are the Lie derivatives and n is the dimension of the system.

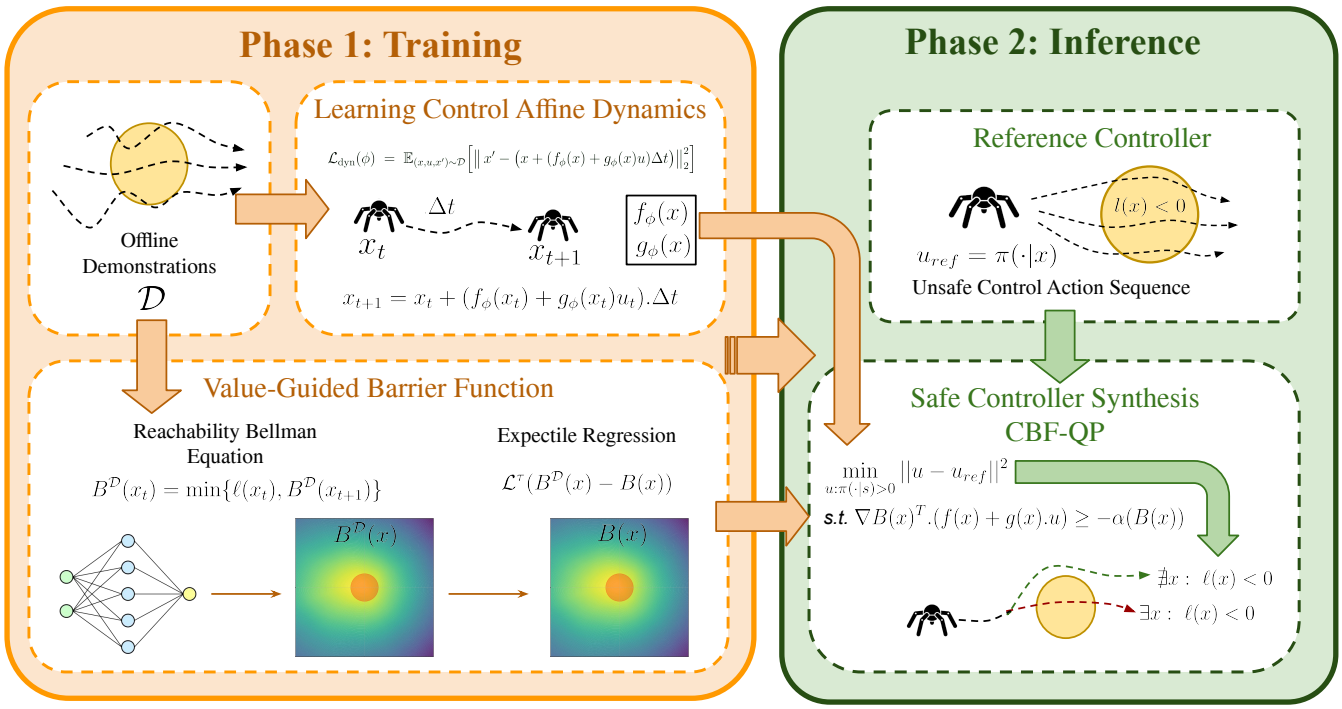

Figure 1: **Framework Overview:** *(Left)*: We learn value guided barrier function with reachability based bellman equation for learning optimal safe region and apply expectile regression for OOD case handling, when learning from Offline Dataset. *(Right)*: Inferencing with CBF-QP using learned barrier function as valid CBF to rollout safe step-wise actions, with any reference controller.

As established in (Ames et al. 2017), any Lipschitz continuous control law $\pi(x)$ that satisfies the condition $\dot{B} + \kappa(B) \geq 0$ guarantees the system's safety when $x(0) \in \mathcal{C}$. Additionally, if the initial state $x(0)$ lies outside $\mathcal{C}$, this condition ensures asymptotic convergence to the safe set $\mathcal{C}$.

While CBFs provide a principled framework to guarantee safety, their practical deployment is hindered by the lack of general methods for constructing valid barrier functions. As a result, practitioners typically resort to handcrafted or domain-specific CBFs, which can yield overly conservative safe sets. Furthermore, in the presence of control bounds, a nominal CBF may conflict with feasibility requirements, causing the corresponding CBF-QP to become infeasible.

Control Barrier Value Function (CBVF)

To overcome the limitations inherent in classical CBF formulations, (Choi et al. 2021) proposed the *Control Barrier Value Function* (CBVF), which integrates Control Barrier Functions with Hamilton–Jacobi (HJ) Reachability (Bansal et al. 2017). We begin by encoding the safety specification using a Lipschitz continuous function $\ell : \mathbb{R}^n \to \mathbb{R}$, where the failure set is defined as $\mathcal{F} := \{x \in \mathcal{X} \mid \ell(x) < 0\}$. Under this construction, a CBVF $B : \mathcal{X} \to \mathbb{R}$ is defined as the viscosity solution of the following Hamilton–Jacobi–Bellman Variational Inequality (HJB-VI):

$$\min\left\{\max_{u\in\mathbb{U}}(\nabla B(x)\cdot(f(x)+g(x)u)),\ \ell(x) - B(x)\right\} = 0, \tag{5}$$

with boundary condition $B(x)|_{t=0} = \ell(x)|_{t=0}$. The resulting value function induces a forward-invariant safe set $\mathcal{C} := \{x \in \mathcal{X} \mid B(x) \geq 0\}$ and ensures that admissible controls $u \in \mathbb{U}$ satisfy the following Lie-derivative condition:

$$\max_{u\in\mathbb{U}} \left[L_f B(x) + L_g B(x)u + \kappa(B(x))\right] \geq 0. \tag{6}$$

Safe Controller Synthesis using CBVFs: Quite often, we have a reference control policy, $\pi_{ref}(x)$, designed to meet the performance requirements of the system. However, such controllers often lack safety guarantees. To ensure the system meets its safety requirements while preserving performance, the reference controller must be minimally adjusted to incorporate safety constraints. This adjustment can be accomplished using the Control Barrier Value Function-based Quadratic Program (CBVF-QP), described as follows:

$$\begin{aligned}\pi_{\text{safe}}(x) = &\min_{u\in\mathbb{U}\subseteq\mathbb{R}^m} \|u - \pi_{ref}\|^2 \\ &\text{s.t. } \mathcal{L}_f B(x) + \mathcal{L}_g B(x)u + \kappa\left(B(x)\right) \geq 0.\end{aligned} \tag{7}$$

The CBVF-QP framework facilitates the synthesis of a provably safe control policy, $\pi_{\text{safe}}(x)$, while staying close to the reference controller to preserve system performance.

Challenges in CBVF Synthesis: Traditional approaches compute Control Barrier Value Functions using grid-based HJ reachability methods (Mitchell 2005), which are fundamentally limited by the curse of dimensionality. Recent efforts have attempted to overcome these issues by learning CBVFs through online reinforcement learning (So et al. 2024); however, as discussed in Section , such approaches require extensive online interaction, rendering them unsuitable for safety-critical systems. Furthermore, existing neural CBF methodologies (Abate et al. 2020; Zhang et al. 2023; Tayal et al. 2024b) typically assume access to accurate system dynamics, an assumption that does not hold for many real-world platforms. These limitations motivate a shift toward using offline demonstrations, either sourced from public datasets (Liu et al. 2024; Sun et al. 2020) or collected in controlled settings where safety can be guaranteed. Building on this premise, we refine Objective 1 as follows:

Objective 2. *Our objective is to synthesize a Control Barrier Value Function $B : \mathcal{X} \to \mathbb{R}$ directly from an offline dataset of demonstrations $\mathcal{D}$, such that the resulting CBVF-QP controller π_{safe} in (7) satisfies the positive invariance property specified in Definition 1.*

Methodology

Having introduced the CBVF formulation in the previous section, we now describe a practical methodology for synthesizing a valid Control Barrier Value Function and thereby the safe controller from offline demonstrations $\mathcal{D}$. Our objective is to construct a data-driven approximation of the viscosity solution of the HJB-VI (5) without relying on known system dynamics or online interaction. The key idea is to reinterpret it through a finite-difference barrier recursion compatible with demonstration data, and to use this recursion to learn a value-guided barrier function that inherits forward invariance.

Finite-Difference Barrier Synthesis

To approximate the CBVF using trajectories in $\mathcal{D}$, we consider a finite difference recursive version of equation (5). Given a trajectory $\{x_t, u_t\}_{t=0}^{N}$, we define the finite-difference barrier update

$$B(x_t) = \min\big\{\ell(x_t),\, \max_{u_t} B(x_{t+1})\big\}, \qquad \forall t \in \{0, 1, 2, \dots\}, \tag{8}$$

where, $x_t = x(t)$, $u_t = u(t)$ and $x_{t+1} = x(t + \Delta t)$, along with the boundary condition $B(x_0) = \ell(x_0)$. This recursion has two significant advantages. First, it enables us to learn a barrier directly from data without requiring f and g. Second, under mild regularity assumptions, (8) preserves the forward invariance property of the CBVF. Intuitively, the recursion encodes the principle that a state is safe if and only if its immediate successor is safe or it lies outside the unsafe set as specified by $\ell(x)$.

To represent this barrier function, we parameterize a neural network $B_\psi^{\mathcal{D}}(x)$ with ψ utilizing the universal approximation property. However, directly solving for the recursion (8) can lead to degenerate solutions. For instance, setting $B_\psi^{\mathcal{D}}(x) = c$ for a sufficiently small constant c satisfies (8) but clearly does not satisfy the CBVF conditions in (5). This pathology is analogous to the non-contractive behavior of undiscounted value iteration in MDPs. Following the approach in (Fisac et al. 2019), we incorporate a discounted finite-difference loss to avoid such trivial solutions:

$$\mathcal{L}_{B^{\mathcal{D}}}(\psi) = \mathbb{E}_{(x,u,x')\sim\mathcal{D}}\Big[\Big((1-\gamma)\,\ell(x) + \gamma \min\{\ell(x),\, B_\psi^{\mathcal{D}}(x')\} - B_\psi^{\mathcal{D}}(x)\Big)^2\Big]. \tag{9}$$

where $x = x_t$, $x' = x_{t+1}$ and $u = u_t$ and $\gamma \to 1$. This discounted recursion ensures contraction, promotes stable learning, and prevents the network from collapsing to uniformly unsafe or uniformly safe solutions. To avoid evaluating barrier targets at out-of-distribution actions, we remove the maximization over actions from the loss in (9) and use only the demonstrated action in each transition. While this prevents unsupported queries, the resulting estimate reflects the safety profile of the behaviour policy that generated the data. Consequently, this produces a behaviour-induced barrier, which is typically sub-optimal because it ignores other admissible actions that could yield larger safe-set estimates.

Avoiding Out-of-Distribution Actions in Offline Learning

The naïve regression objective in (9) fits $B_\psi^{\mathcal{D}}$ to the mean of the demonstrated next-state targets, but this corresponds to the behavior-induced barrier and yields overly conservative safe sets. Ideally, if we assume unlimited capacity and no sampling error, the optimal parameters should satisfy, $B(x) \approx \mathbb{E}_u\big[\min\big\{\ell(x),\, \max_u B(x')\big\}\big]$. However, such unconstrained maximization can result in actions that are never observed in the dataset.

Since offline data only provides information about those actions selected by the behavior policy, evaluating values (or barrier targets) using unsupported actions can distort learning because the corresponding transitions are not grounded in the dataset. Subsequently, motivated by the insights of Implicit Q-Learning (IQL) (Kostrikov, Nair, and Levine 2022), we approximate the maximization over admissible actions by using expectile regression, which enables us to capture the highest admissible barrier values supported by the data while never evaluating B on unseen (x, u) pairs. This allows us to perform a principled value-style backup over the dataset control action support $\mathcal{U}_{\mathcal{D}} = \{\, u \mid (x, u, x') \in \mathcal{D} \,\}$ without extrapolating to unsafe or unobserved actions. IQL shows that expectile regression produces a value function that reflects the values induced by the behavior policy without requiring an explicit behavior model. This prevents the learning target from being influenced by actions that lie outside the dataset support, while still capturing the highest feasible values supported by the demonstrations.

Following this principle, we estimate a CBVF, B_θ, with θ as the Neural Network parameters, that reflects the safety values implied by demonstrated actions. Formally, we minimize the expectile loss

$$\mathcal{L}_B(\theta) \;=\; \mathbb{E}_{(x,u,x')\sim\mathcal{D}}\big[\mathcal{L}^\tau\big(B_\psi^{\mathcal{D}}(x) - B_\theta(x)\big)\big], \tag{10}$$

where $\mathcal{L}^\tau(y) = |\tau - \mathbf{1}(y < 0)|\, y^2$ is the τ-expectile loss used in (Kostrikov, Nair, and Levine 2022). Intuitively, a higher expectile level τ places greater weight on underestimation errors than overestimation errors, pushing B_θ toward the upper envelope of safety values supported by the dataset. Thus, τ controls how aggressively the learned barrier emphasizes high, data-supported safety values without extrapolating to unseen actions. The barrier function thus obtained, B_θ, is our proposed *Value-guided Offline Control Barrier Function* (V-OCBF).

Controller Synthesis via Learned Dynamics

The learned barrier function B is subsequently employed to fulfill the primary goal of synthesizing a safe policy 1 using the CBVF-QP formulation in (7). Solving this QP necessitates the evaluation of the Lie derivatives $\mathcal{L}_f B$ and $\mathcal{L}_g B$, both of which rely on the underlying control-affine system dynamics. In our offline-only setting, the true dynamics are unavailable; therefore, we construct a neural network–based surrogate model to approximate the underlying transition dynamics of the form:

$$x_{t+1} = x_t + \big(f_\phi(x_t) + g_\phi(x_t)u_t\big)\,\Delta t, \tag{11}$$

which enables computation of the required derivatives and supports safe policy synthesis. The model parameters ϕ are trained using one-step transitions from the offline dataset $\mathcal{D}$ by minimizing the prediction loss:

$$\mathcal{L}_{\text{dyn}}(\phi) = \mathbb{E}_{(x,u,x')\sim\mathcal{D}}\Big[\| x' - \big(x + (f_\phi(x) + g_\phi(x)u)\Delta t\big) \|^2\Big], \tag{12}$$

implemented as a minibatch MSE objective.

Importantly, the learned dynamics model is *not* used when learning the barrier function. Incorporating it into the CBVF learning stage would require evaluating terms involving (f_ϕ, g_ϕ) under actions outside the dataset-supported set $\mathcal{U}_\mathcal{D}$, thereby violating the action constraints critical for preventing value underestimation in the offline regime. Hence, using learned dynamics during CBVF training would allow the network to extrapolate into unsupported regions of the action space, defeating the purpose of the OOD-aware barrier learning objective described earlier.

In contrast, at *inference* time, the learned dynamics serve a different role: they enable the evaluation of Lie derivatives needed to solve the CBVF-QP ((7)). Specifically, for any query state x, we compute

$$\mathcal{L}_f B(x) = \nabla_x B_\theta(x)^\top f_\phi(x), \quad \mathcal{L}_g B(x) = \nabla_x B_\theta(x)^\top g_\phi(x). \tag{13}$$

These quantities allow the QP in (7) to be solved for the safe control action u_{safe}, completing the pipeline for constructing a safety-certified controller purely from offline demonstrations.

Experiments

The experiments are designed to evaluate: (*i*) the safety and performance of V-OCBF relative to constrained offline RL and neural CBF baselines on systems with unknown dynamics, (*ii*) the advantages of value-guided barriers over behavior-policy–induced barriers, (*iii*) the robustness of the resulting QP controller under external disturbances, and (*iv*) the effectiveness of V-OCBF compared to a CBVF synthesized using learned dynamics.

Baselines: We compare V-OCBF against a diverse set of constrained offline learning and CBF-based methods. For constrained offline learning, we include **Behavior Cloning (BC)**, **BEAR-Lag** (Lagrangian constraint version of (Kumar et al. 2019)), **COptiDICE** (Lee et al. 2022), and **FISOR** (Zheng et al. 2024) which enforce safety indirectly via soft constraints on policy optimization or behavior imitation. For CBF-based approaches, we evaluate **Neural Control Barrier Function (NCBF)** (Robey et al. 2020), *Conservative Control Barrier Function (CCBF)* (Tabbara and Sibai 2025), and *In-Distribution Barrier Function (iDBF)* (Castañeda et al. 2023), which synthesize explicit safety filters from offline data but often yield conservative safe sets. In contrast, V-OCBF learns a *value-guided* barrier function from offline demonstrations that accounts for future unsafe interactions, producing a hard, state-wise safety filter with larger safe set coverage.

Evaluation Metrics: We evaluate all methods based on (i) *safety*, measured as the total number of safety violations incurred before episode termination, and (ii) *performance*, measured via the cumulative episode rewards. These metrics allow us to assess the trade-off between strict safety enforcement and task performance across different offline RL and CBF-based approaches.

Experimental Case Studies

To perform a holistic performance analysis of our proposed approach, we apply V-OCBF in conjunction with Behavior Cloning (BC) as the nominal (reference) controller for all the different environments which are supposed to assess varying objectives. Below we list all the environments that we use:

- **Autonomous Ground Vehicle (AGV) Collision Avoidance:** In our first experiment, we examine a 3-dimensional collision avoidance problem involving an autonomous ground vehicle governed by Dubins' car dynamics (Dubins 1957). The objective is to ensure safety by avoiding a static obstacle while navigating through a bounded environment.
- **MuJoCo Safety Gymnasium:** We next evaluate our framework on Safety Gymnasium environments (Ji et al. 2023). Specifically, we evaluate the V-OCBF-based QP (equation 7) on high-dimensional MuJoCo tasks like Hopper, Swimmer, Half Cheetah, Walker2D and Ant. The objective in each environment is to maximize reward while keeping the agent velocity below the velocity thresholds. We keep the reward and safety-violation metrics identical to the Safety-Gymnasium definitions and use the standard DSRL dataset for safe offline RL (Liu et al. 2024). To evaluate our method against baselines, we randomly sampled 500 initial states for each environment, respectively, the results for which can be referred to from Figure 2.

Method	Safe Episodes (%)	Episode Reward	Safe Set Volume (%)
BC	48.92 ± 1.69	20.45 ± 1.84	42.51
BEAR-Lag (Kumar et al. 2019)	65.12 ± 0.24	13.85 ± 0.81	58.21
COptiDICE (Lee et al. 2022)	68.91 ± 0.32	15.33 ± 0.67	62.32
BC+NCBF (Robey et al. 2020)	92.48 ± 0.60	44.61 ± 2.58	81.92
BC+iDBF (Castañeda et al. 2023)	92.87 ± 0.73	48.23 ± 2.01	83.32
BC+CCBF (Tabbara and Sibai 2025)	93.56 ± 0.56	49.66 ± 2.34	90.94
FISOR (Zheng et al. 2024)	95.78 ± 0.2	52.33 ± 0.93	90.14
BC+**V-OCBF** (Ours)	$\mathbf{98.28 \pm 0.54}$	$\mathbf{54.93 \pm 0.46}$	**92.57**

Table 1: AGV Collision Avoidance Experiment: Percentage Safe Episodes, Mean Episode Reward and Safe Set Volume across different methods. Evaluated over 500 episodes and 5 seed values.

Results

We begin by evaluating all methods on the AGV Collision Avoidance task, which provides a clear setting to study how different approaches balance safety and performance. The results in Table 1 highlight notable differences in how offline RL and CBF-based methods handle this trade-off.

Offline RL baselines such as BC, BEAR-Lag, and COptiDICE achieve relatively low safety rates. BC tends to reproduce unsafe behaviors from the dataset, while BEAR-Lag and COptiDICE try to account for safety but remain limited because they operate with soft constraint formulations. Their lower reward and safety scores indicate that they struggle to balance both safety and performance objectives.

In contrast, methods that incorporate a CBF-QP layer, such as BC+NCBF, BC+iDBF, and BC+CCBF, achieve much higher safety rates. The QP ensures that unsafe actions are filtered out, even if the nominal controller is imperfect. However, the performance of these approaches still depends heavily on the quality of the learned barrier. FISOR performs better than the other offline RL baselines because it explicitly expands the feasible safe region before optimizing for performance. However, due to lack of explicit safety filtering, it leads to lesser safety rates than our proposed method, due to the impending learning errors in the computation of feasible region. This also highlights the importance of QP based safety filtering scheme for achieving better safety. Overall, *V-OCBF achieves the strongest results across all metrics*.

To further analyze scalability, we extend the evaluation to MuJoCo Safety Gymnasium environments (Hopper, HalfCheetah, Ant, Swimmer, and Walker2D), with unknown dynamic models. The results in Figure 2 demonstrate that V-OCBF again achieves the lowest safety violation rates across all tasks. Notably, the method maintains near-zero violations on while preserving satisfactory reward levels compared to BC and outperforming iDBF, NCBF, and CCBF. These neural CBF baselines degrade sharply in higher dimensions: NCBF suffers from optimization difficulties, while iDBF often enforces overly restrictive boundaries that suppress task performance. FISOR again remains competitive but does not match the safety consistency of V-OCBF.

Overall, the experiments provide strong empirical evidence that V-OCBF effectively co-optimizes safety and performance, scaling from low-dimensional AGV dynamics to complex MuJoCo systems. The method consistently outperforms existing offline RL and neural CBF baselines in terms of safety while maintaining competitive reward, highlighting its suitability for offline settings where both strict safety and reliable performance are required.

Conclusion

The experimental results demonstrate that V-OCBF consistently learns barrier functions that are practical for safety-critical control while learning from offline data. Across a diverse set of environments, V-OCBF produces more precise feasible safe sets compared to traditional neural CBF baselines. Its reachability-inspired formulation allows it to effectively encode actuation constraints, leading to a significant reduction in safety violations, particularly with tight control limits. These findings highlight V-OCBF as a scalable, offline-capable framework for learning control barrier functions that combine theoretical soundness with practical applicability.

References

Abate, A.; Ahmed, D.; Edwards, A.; Giacobbe, M.; and Peruffo, A. 2021. Fossil: A software tool for the formal synthesis of Lyapunov functions and barrier certificates using neural networks. In *Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control*, 1–11.

Abate, A.; Ahmed, D.; Giacobbe, M.; and Peruffo, A. 2020. Formal synthesis of Lyapunov neural networks. *IEEE Control Systems Letters*, 5(3): 773–778.

Achiam, J.; Held, D.; Tamar, A.; and Abbeel, P. 2017. Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, 22–31. JMLR.org.

Alshiekh, M.; Bloem, R.; Ehlers, R.; Könighofer, B.; Niekum, S.; and Topcu, U. 2018. Safe Reinforcement Learning via Shielding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Altman, E. 2021. *Constrained Markov decision processes*. Routledge.

Ames, A. D.; Grizzle, J. W.; and Tabuada, P. 2014. Control barrier function based quadratic programs with application to adaptive cruise control. In *53rd IEEE Conference on Decision and Control*, 6271–6278. IEEE.

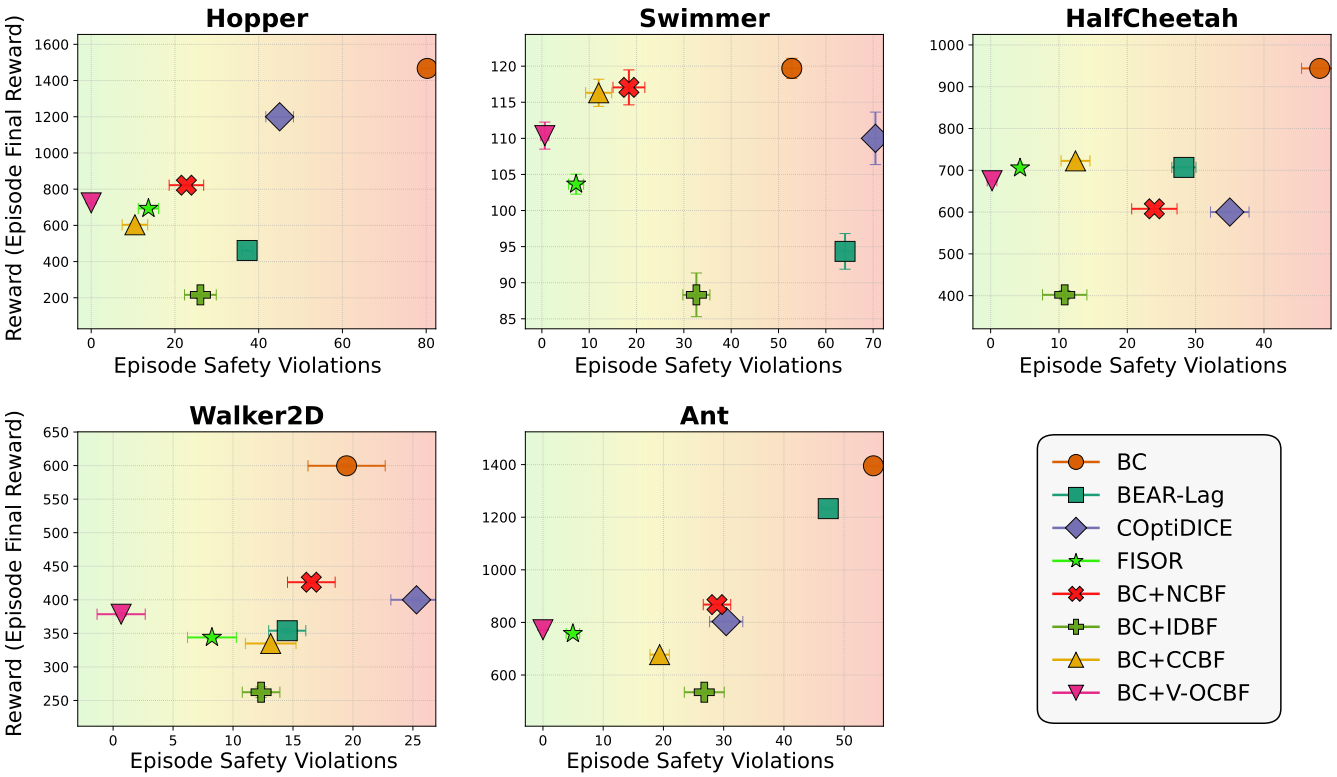

Figure 2: Results plot for all the Mujoco based Safety Gymnasium (Ji et al. 2023) environments. Points towards left (←) are more safe than those on right (→). Evaluated over 500 episodes and 5 seed values.

Ames, A. D.; Xu, X.; Grizzle, J. W.; and Tabuada, P. 2017. Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control*, 62(8): 3861–3876.

Bansal, S.; Chen, M.; Herbert, S.; and Tomlin, C. J. 2017. Hamilton-jacobi reachability: A brief overview and recent advances. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2242–2253. IEEE.

Castañeda, F.; Nishimura, H.; McAllister, R. T.; Sreenath, K.; and Gaidon, A. 2023. In-Distribution Barrier Functions: Self-Supervised Policy Filters that Avoid Out-of-Distribution States. In Matni, N.; Morari, M.; and Pappas, G. J., eds., *Proceedings of The 5th Annual Learning for Dynamics and Control Conference*, volume 211 of *Proceedings of Machine Learning Research*, 286–299. PMLR.

Choi, J. J.; Lee, D.; Sreenath, K.; Tomlin, C. J.; and Herbert, S. L. 2021. Robust control barrier–value functions for safety-critical control. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 6814–6821. IEEE.

Ciftci, Y. U.; Chiu, D.; Feng, Z.; Sukhatme, G. S.; and Bansal, S. 2024. SAFE-GIL: SAFEty Guided Imitation Learning for Robotic Systems. *arXiv preprint arXiv:2404.05249*.

Dawson, C.; Gao, S.; and Fan, C. 2023. Safe Control With Learned Certificates: A Survey of Neural Lyapunov, Barrier, and Contraction Methods for Robotics and Control. *IEEE Transactions on Robotics*.

Dawson, C.; Qin, Z.; Gao, S.; and Fan, C. 2022. Safe nonlinear control using robust neural Lyapunov-barrier functions. In *Conference on Robot Learning*, 1724–1735. PMLR.

Dubins, L. E. 1957. On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. *American Journal of Mathematics*, 79(3): 497–516.

Fisac, J. F.; Lugovoy, N. F.; Rubies-Royo, V.; Ghosh, S.; and Tomlin, C. J. 2019. Bridging Hamilton-Jacobi Safety Analysis and Reinforcement Learning. In *2019 International Conference on Robotics and Automation (ICRA)*, 8550–8556.

Ji, J.; Zhang, B.; Zhou, J.; Pan, X.; Huang, W.; Sun, R.; Geng, Y.; Zhong, Y.; Dai, J.; and Yang, Y. 2023. Safety Gymnasium: A Unified Safe Reinforcement Learning Benchmark. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Kostrikov, I.; Nair, A.; and Levine, S. 2022. Offline reinforcement learning with implicit q-learning. *International Conference on Learning Representations*.

Kumar, A.; Fu, J.; Soh, M.; Tucker, G.; and Levine, S. 2019. Stabilizing off-policy q-learning via bootstrapping error reduction. *Advances in neural information processing systems*, 32.

Kumar, A.; Zhou, A.; Tucker, G.; and Levine, S. 2020. Conservative Q-Learning for Offline Reinforcement Learning. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and

Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 1179–1191. Curran Associates, Inc.

Lee, J.; Paduraru, C.; Mankowitz, D. J.; Heess, N.; Precup, D.; Kim, K.-E.; and Guez, A. 2022. COptiDICE: Offline Constrained Reinforcement Learning via Stationary Distribution Correction Estimation. In *International Conference on Learning Representations*.

Levine, S.; Kumar, A.; Tucker, G.; and Fu, J. 2020. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*.

Liu, Z.; Guo, Z.; Lin, H.; Yao, Y.; Zhu, J.; Cen, Z.; Hu, H.; Yu, W.; Zhang, T.; Tan, J.; and Zhao, D. 2024. Datasets and Benchmarks for Offline Safe Reinforcement Learning. *Journal of Data-centric Machine Learning Research*.

Mitchell, I. M. 2005. A Toolbox of Level Set Methods. In *A Toolbox of Level Set Methods*.

Nguyen, Q.; and Sreenath, K. 2015. Safety-critical control for dynamical bipedal walking with precise footstep placement. *IFAC-PapersOnLine*, 48(27): 147–154.

Robey, A.; Hu, H.; Lindemann, L.; Zhang, H.; Dimarogonas, D. V.; Tu, S.; and Matni, N. 2020. Learning Control Barrier Functions from Expert Demonstrations. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 3717–3724.

So, O.; Serlin, Z.; Mann, M.; Gonzales, J.; Rutledge, K.; Roy, N.; and Fan, C. 2024. How to train your neural control barrier function: Learning safety filters for complex input-constrained systems. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 11532–11539. IEEE.

Stooke, A.; Achiam, J.; and Abbeel, P. 2020. Responsive Safety in Reinforcement Learning by PID Lagrangian Methods. In III, H. D.; and Singh, A., eds., *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 9133–9143. PMLR.

Sun, P.; Kretzschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; Vasudevan, V.; Han, W.; Ngiam, J.; Zhao, H.; Timofeev, A.; Ettinger, S.; Krivokon, M.; Gao, A.; Joshi, A.; Zhang, Y.; Shlens, J.; Chen, Z.; and Anguelov, D. 2020. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Tabbara, I.; and Sibai, H. 2025. Learning Neural Control Barrier Functions from Offline Data with Conservatism. arXiv:2505.00908.

Tayal, M.; Singh, A.; Kolathaya, S.; and Bansal, S. 2025. A Physics-Informed Machine Learning Framework for Safe and Optimal Control of Autonomous Systems. In *Forty-second International Conference on Machine Learning*.

Tayal, M.; Singh, R.; Keshavan, J.; and Kolathaya, S. 2024a. Control barrier functions in dynamic uavs for kinematic obstacle avoidance: A collision cone approach. In *2024 American Control Conference (ACC)*, 3722–3727. IEEE.

Tayal, M.; Zhang, H.; Jagtap, P.; Clark, A.; and Kolathaya, S. 2024b. Learning a Formally Verified Control Barrier Function in Stochastic Environment. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 4098–4104.

Wu, G.; and Sreenath, K. 2016. Safety-critical control of a planar quadrotor. In *2016 American Control Conference (ACC)*, 2252–2258.

Xu, H.; Zhan, X.; and Zhu, X. 2022. Constraints penalized q-learning for safe offline reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 8753–8760.

Zhang, H.; Wu, J.; Vorobeychik, Y.; and Clark, A. 2023. Exact Verification of ReLU Neural Control Barrier Functions. In Oh, A.; Neumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 5685–5705. Curran Associates, Inc.

Zhao, Q.; Chen, X.; Zhao, Z.; Zhang, Y.; Tang, E.; and Li, X. 2022. Verifying Neural Network Controlled Systems Using Neural Networks. In *25th ACM International Conference on Hybrid Systems: Computation and Control*, 1–11.

Zhao, W.; He, T.; Chen, R.; Wei, T.; and Liu, C. 2023. Safe reinforcement learning: A survey. *arXiv preprint arXiv:2302.03122*.

Zheng, Y.; Li, J.; Yu, D.; Yang, Y.; Li, S. E.; Zhan, X.; and Liu, J. 2024. Safe offline reinforcement learning with feasibility-guided diffusion model. In *International Conference on Learning Representations*.