ICML International Conference On Machine Learning

Spotlight

LipsNet++: Unifying Filter and Controller into a Policy Network

X. Song[1,2], L. Chen[3], T. Liu[1], W. Wang[1], Y. Wang[1], S. Qin[1], Y. Ma[4], J. Duan[1,3], S. E. Li[1]

[1]Tsinghua University, [2]University of Michigan, [3]USTB, [4]Johns Hopkins University

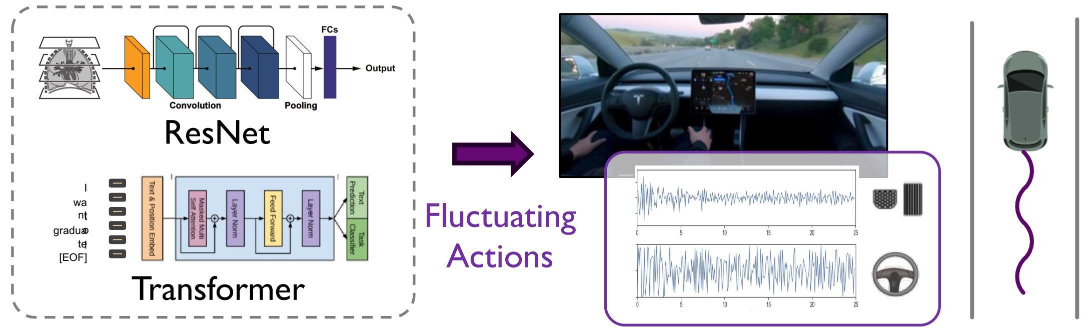
- Website : xjsong99.github.io/LipsNet_v2
- Contact : xjsong@umich.edu

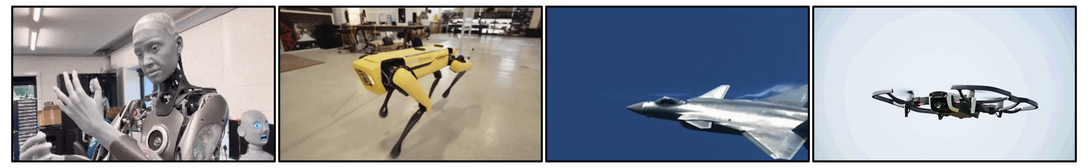
Tsinghua University · University of Michigan · Johns Hopkins University

1. Background

Deep **reinforcement learning (RL)** is effective for decision-making and control tasks like autonomous driving and embodied AI.

However, RL policies often suffer from the **action fluctuation problem** in real-world applications, resulting in severe actuator wear, safety risk, and performance degradation.

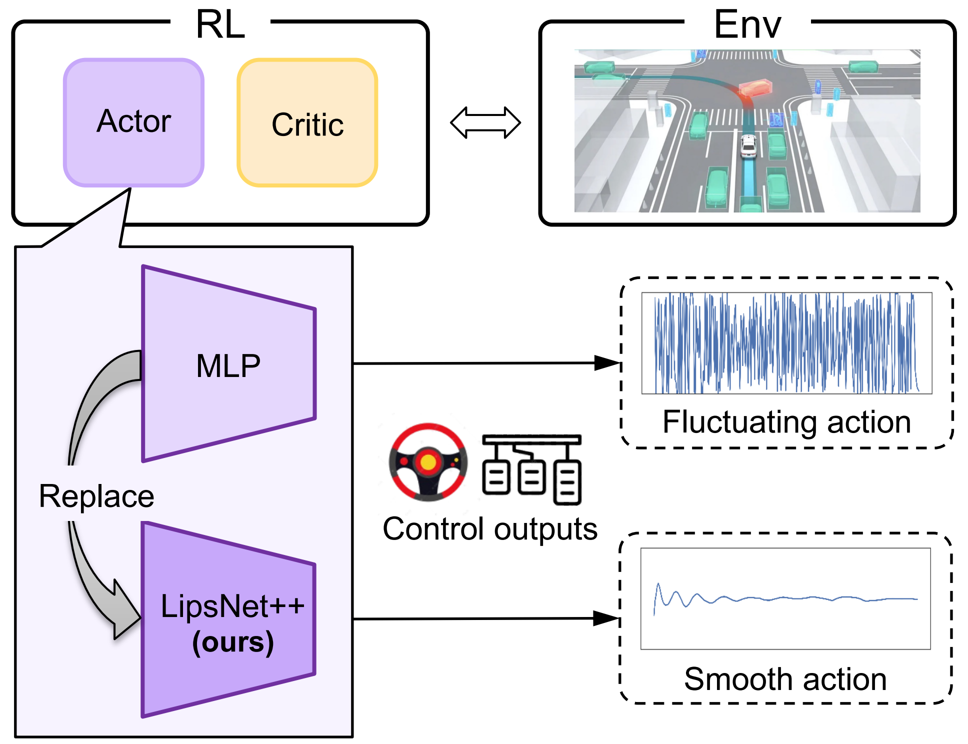




Action fluctuation problem prevalent in various tasks.

2. Objective

Our objective is to **smooth the action trajectory** in RL by designing the actor network, without complicating the RL algorithms.

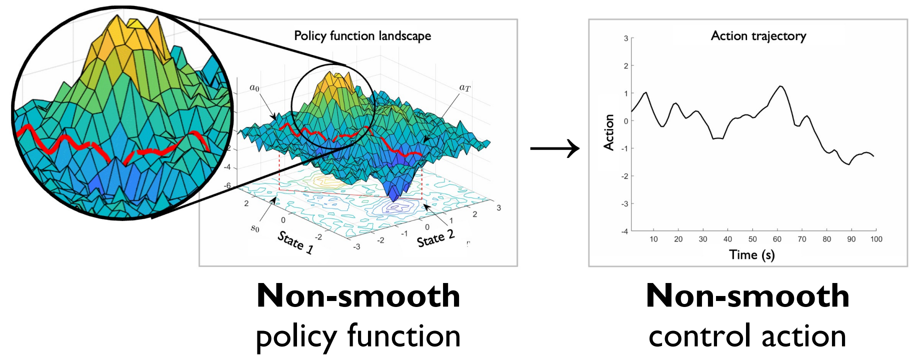


3. Reasons Identification of Action Fluctuation

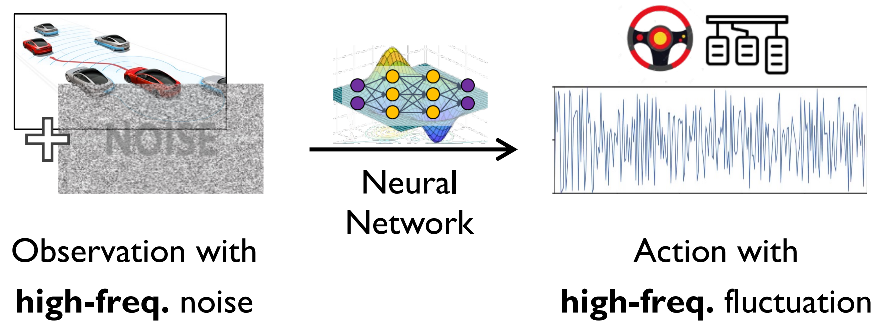
Observation: $o_t = s_t + \xi_t$ Action: $a_t = \pi(o_t)$ Action change rate: $\frac{\mathrm{d}a_t}{\mathrm{d}t} = \frac{\mathrm{d}\pi(o_t)}{\mathrm{d}o_t} \cdot \frac{\mathrm{d}o_t}{\mathrm{d}t}$

$$\left\|\frac{\mathrm{d}a_t}{\mathrm{d}t}\right\| \leq \left\|\frac{\mathrm{d}\pi(o_t)}{\mathrm{d}o_t}\right\| \cdot \left(\left\|\frac{\mathrm{d}s_t}{\mathrm{d}t}\right\| + \left\|\frac{\mathrm{d}\xi_t}{\mathrm{d}t}\right\|\right)$$

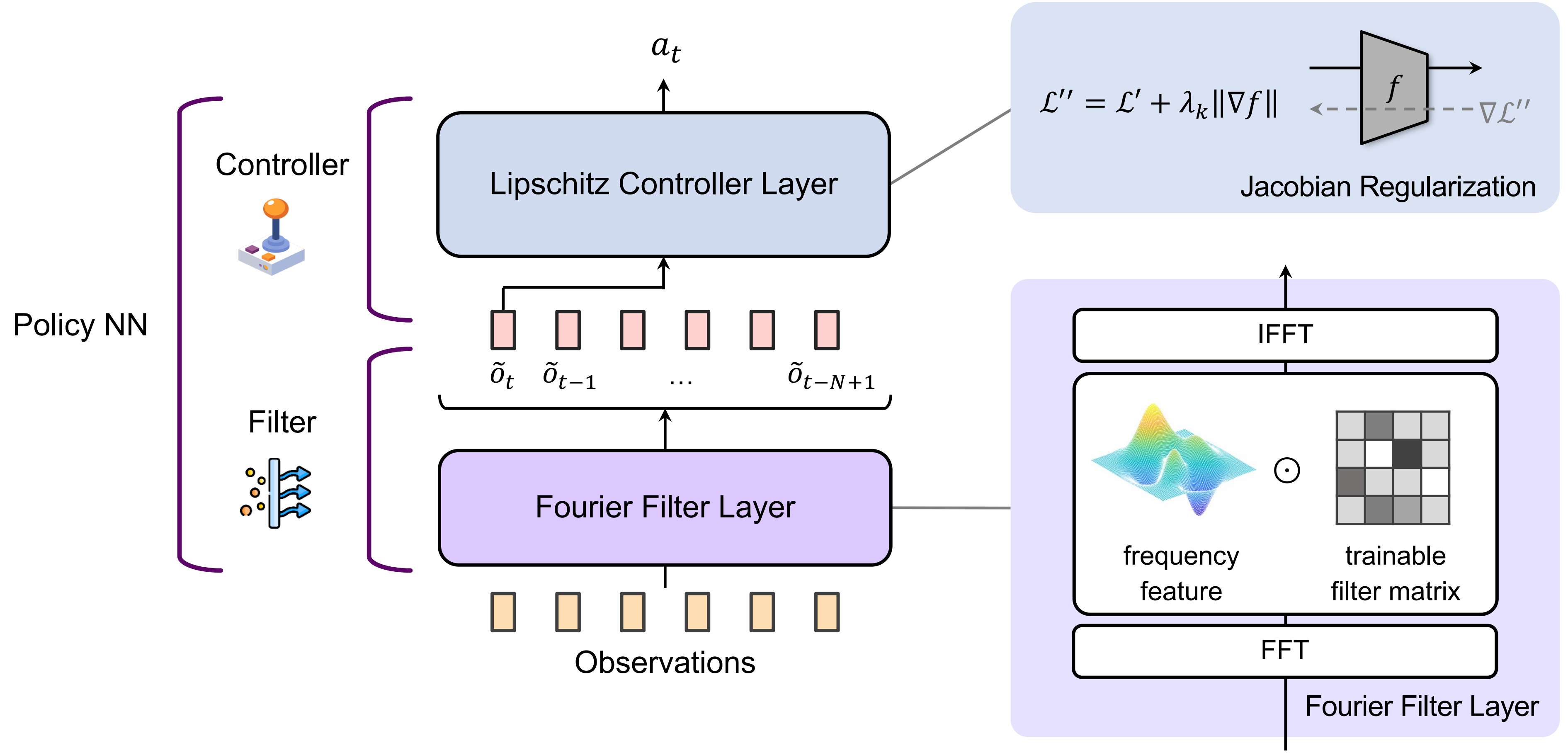
Reason 1: policy function's **non-smooth landscape**



Reason 2: existence of **observation noise**



4. Overall Structure of LipsNet++



5. Theorem & Learning Mechanism

- **Fourier Filter Layer**

  Tailor the policy improvement (PIM) loss as

$$\mathcal{L}' = \mathcal{L} + \lambda_h \|H\|_F$$

For policy improvement

For learning the filtering strength of each frequency

- **Lipschitz Controller Layer**

  In this layer, we propose Jacobian regularization to constrain the Lipschitz constant of policy network.

Definition 3.1 (Local Lipschitz Constant) *Suppose $f: \mathbb{R}^n \to \mathbb{R}^m$ is a continuous neural network. The $K(x)$ is defined as the local Lipschitz constant of f on the neighborhood $\mathcal{B}(x,\rho) = \{x': \|x' - x\| < \rho\}$:*

$$K(x) = \max_{x_1, x_2 \in \mathcal{B}(x,\rho)} \frac{\|f(x_1) - f(x_2)\|}{\|x_1 - x_2\|}.$$

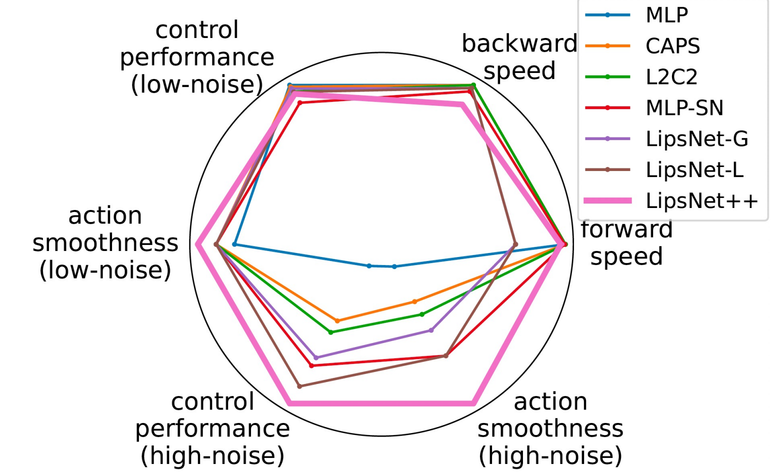
Theorem 3.2 (Lipschitz's Jacobian Approximation) *Let $f: \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable network. The Jacobian norm $\|\nabla_x f\|$ is an approximation of $K(x)$ within $\mathcal{B}(x,\rho)$. The approximation error is*

$$\max_{\delta \in \mathcal{B}(0,\rho)} \left[\left(\nabla_x \|\nabla_x f(x)\|\right)^\top \delta + o(\delta)\right].$$

Moreover, as $\rho \to 0$, the Jacobian norm converges to the exact local Lipschitz constant, i.e. $\lim_{\rho\to 0} \|\nabla_x f\| = K(x)$.

6. Overall Performance

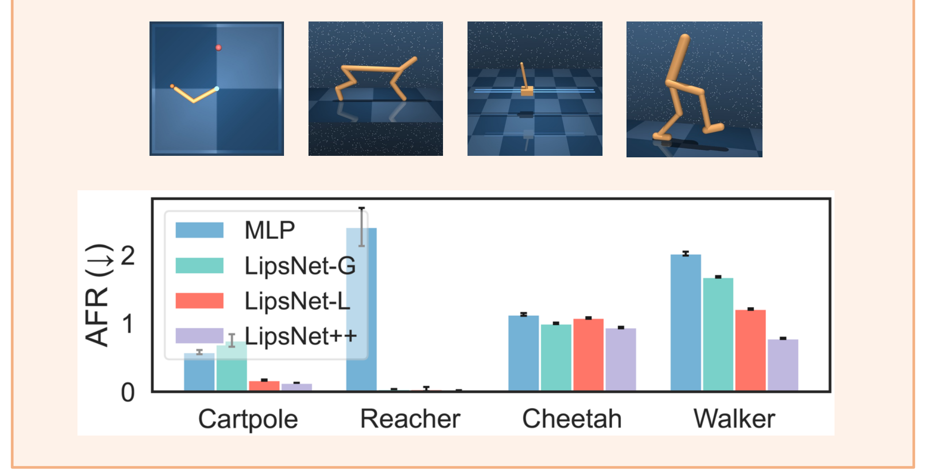
We evaluate the overall performance of LipsNet++ with 6 baselines across 6 metrics. It shows that **LipsNet++ achieves the SOTA overall performance.**



7. Experiment Results

- **DeepMind Control Suit**

  The results show that LipsNet++ has the **lowest** action fluctuation ratio (AFR) with the same level of total average return (TAR). E.g., LipsNet++ reduces the AFR by 35.5% in Walker env. compared to LipsNet *(Song, ICML 2023)*.



- **Mini-Vehicle Driving**

  We evaluated LipsNet++ on the trajectory tracking and obstacle avoidance task in 4 scenarios with varies noise levels. The result implies that LipsNet++ has much better **action smoothness** and **noise robustness**.