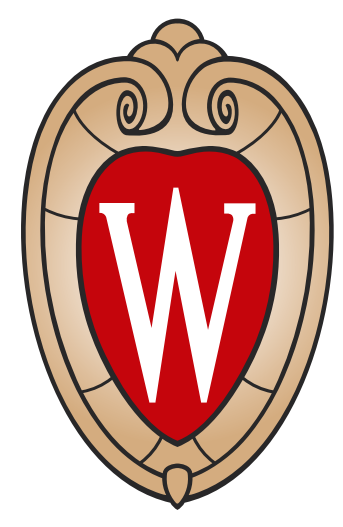
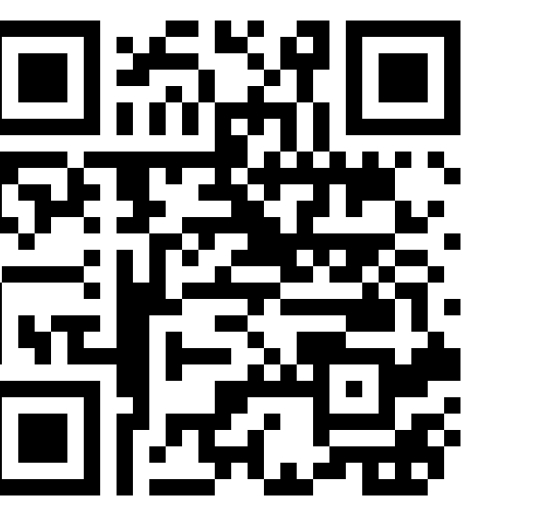




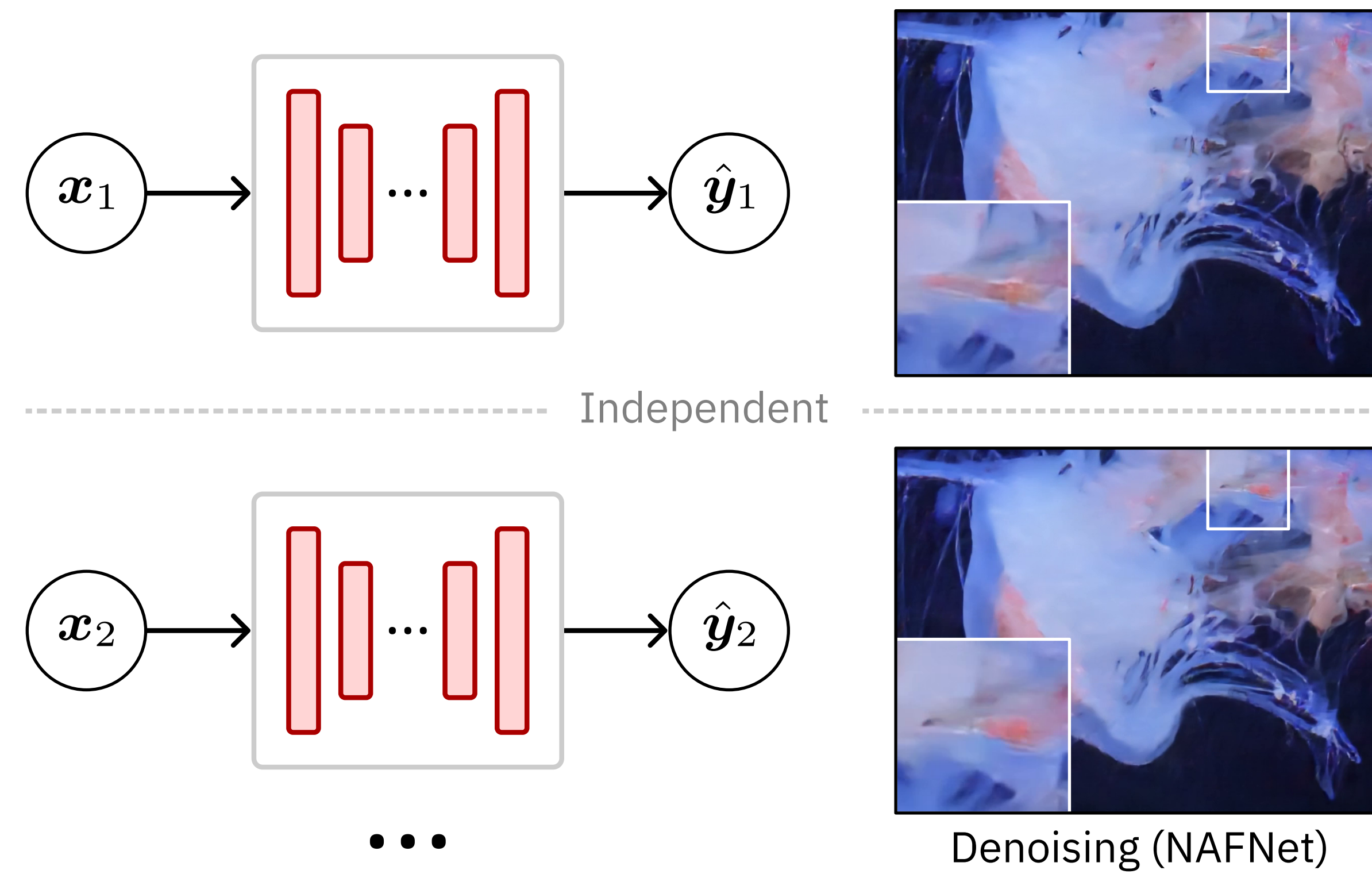
Instant Video Models: Universal Adapters for Stabilizing Image-Based Networks

Matthew Dutson, Nathan Labiosa, Yin Li, and Mohit Gupta

University of Wisconsin–Madison



Unstable Frame-Based Processing



Objective

Adapt frame-based networks for *stability* and *robustness*

Avoid modifying the original architecture

Use an online (streaming) approach

Definitions

Input x	Output y	Duration τ	Model f_ϕ	Corruption ε
--------------	---------------	--------------------	-------------------	-----------------------------

Robustness under corruptions

$$\mathcal{R}_c = -\mathbb{E}_{\varepsilon \sim \mathcal{E}, (x, y, \tau) \sim \mathcal{D}} \left[\sum_{t=1}^{\tau} \delta(f_\phi(\varepsilon_t(x)), y_t) \right]$$

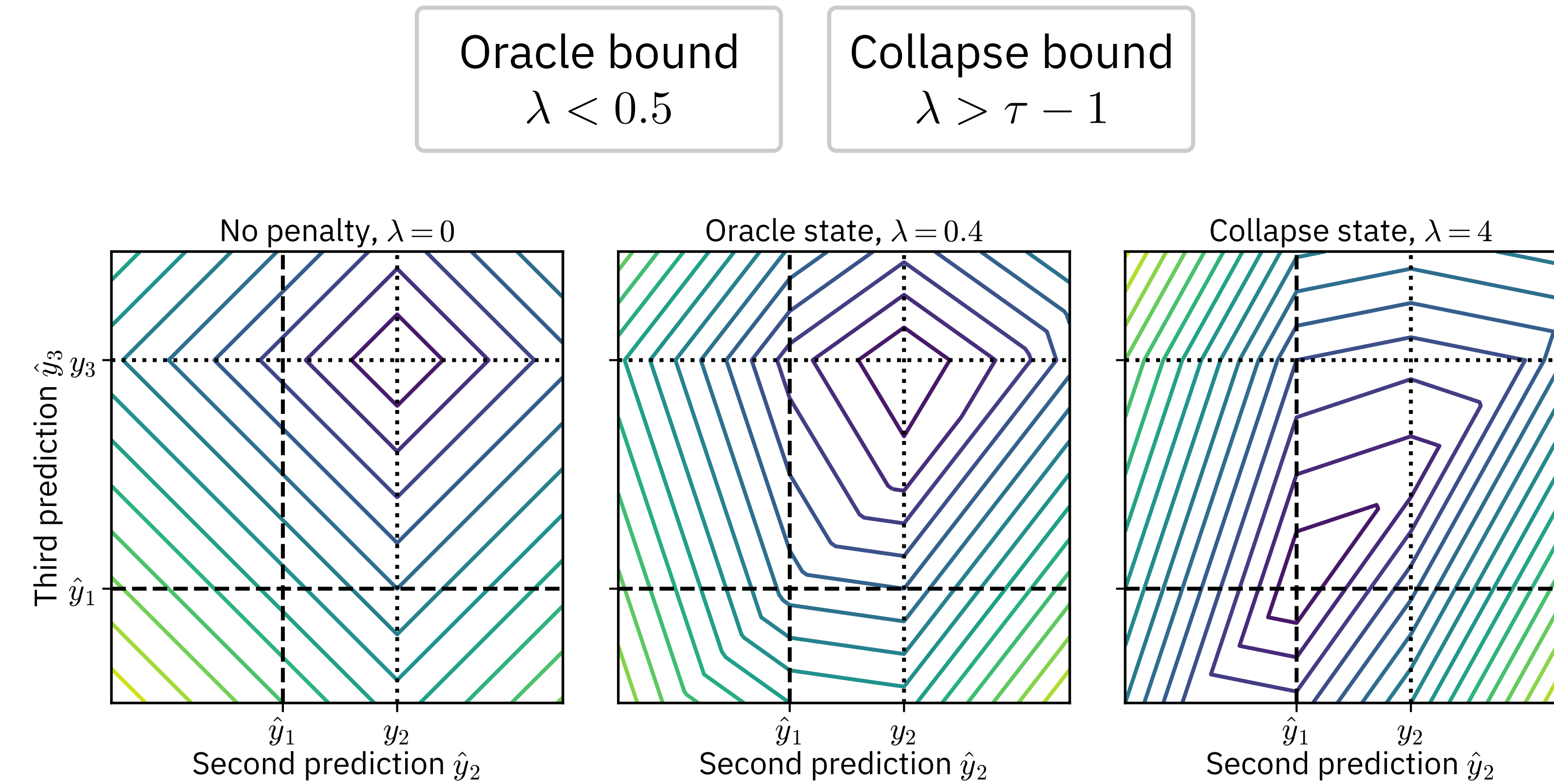
Stability under corruptions

$$\mathcal{S}_c = -\mathbb{E}_{\varepsilon \sim \mathcal{E}, (x, y, \tau) \sim \mathcal{D}} \left[\sum_{t=1}^{\tau-1} \delta(f_\phi(\varepsilon_t(x)), f_\phi(\varepsilon_{t+1}(x_{t+1}))) \right]$$

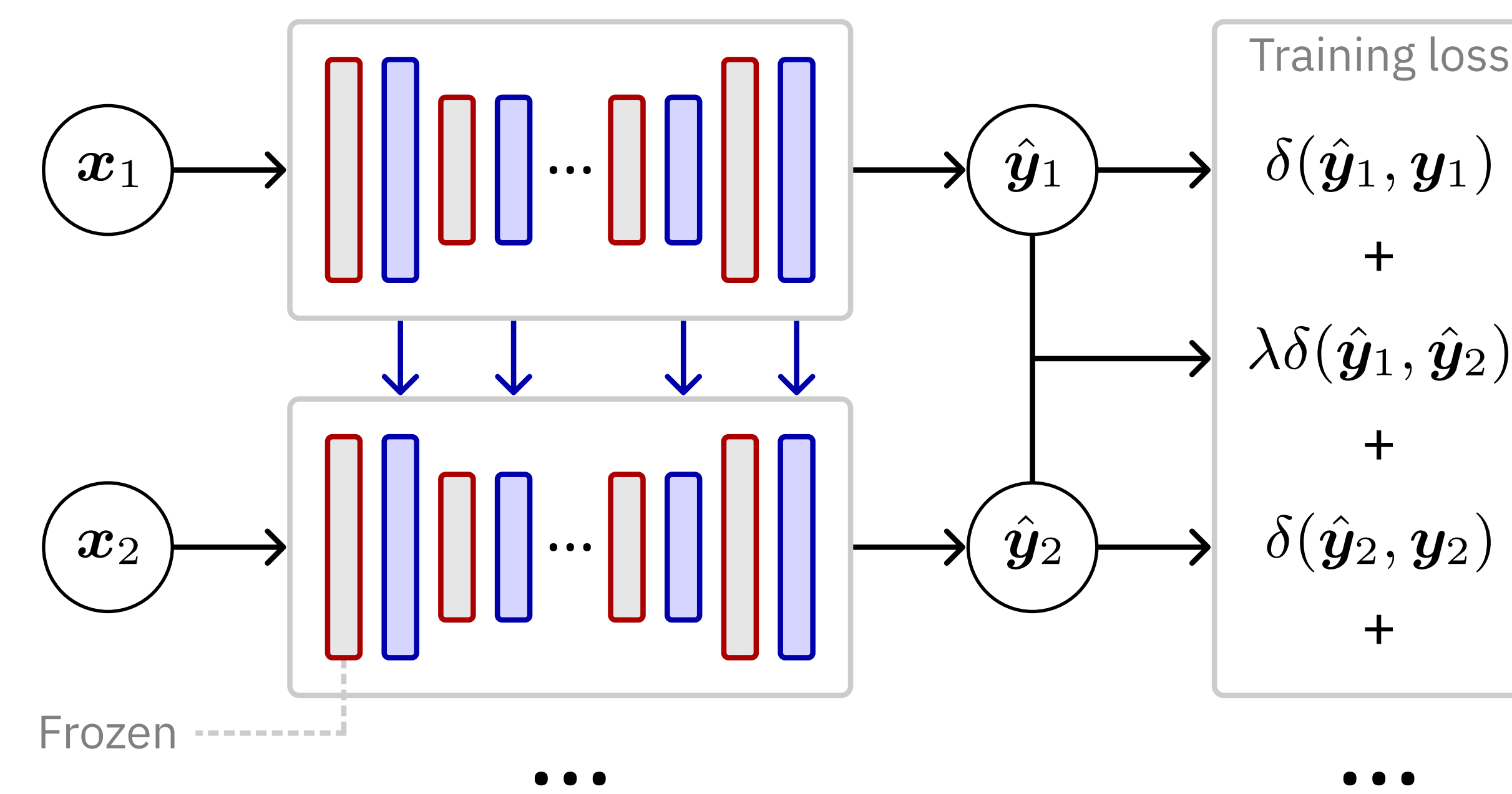
Accuracy-stability-robustness loss

$$\mathcal{U}_c = -(\mathcal{R}_c + \lambda \mathcal{S}_c)$$

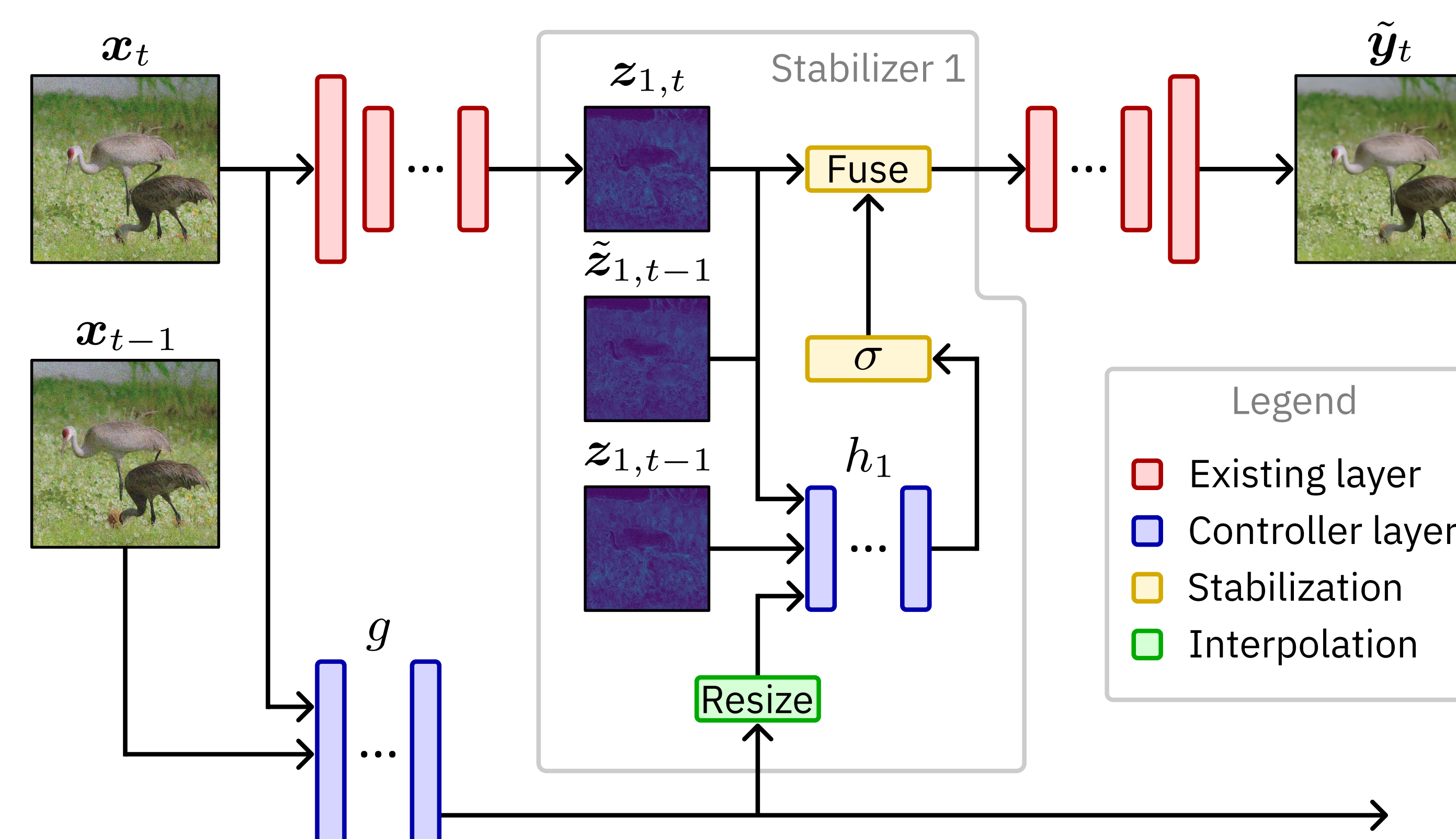
Analysis



Injecting Stabilizers



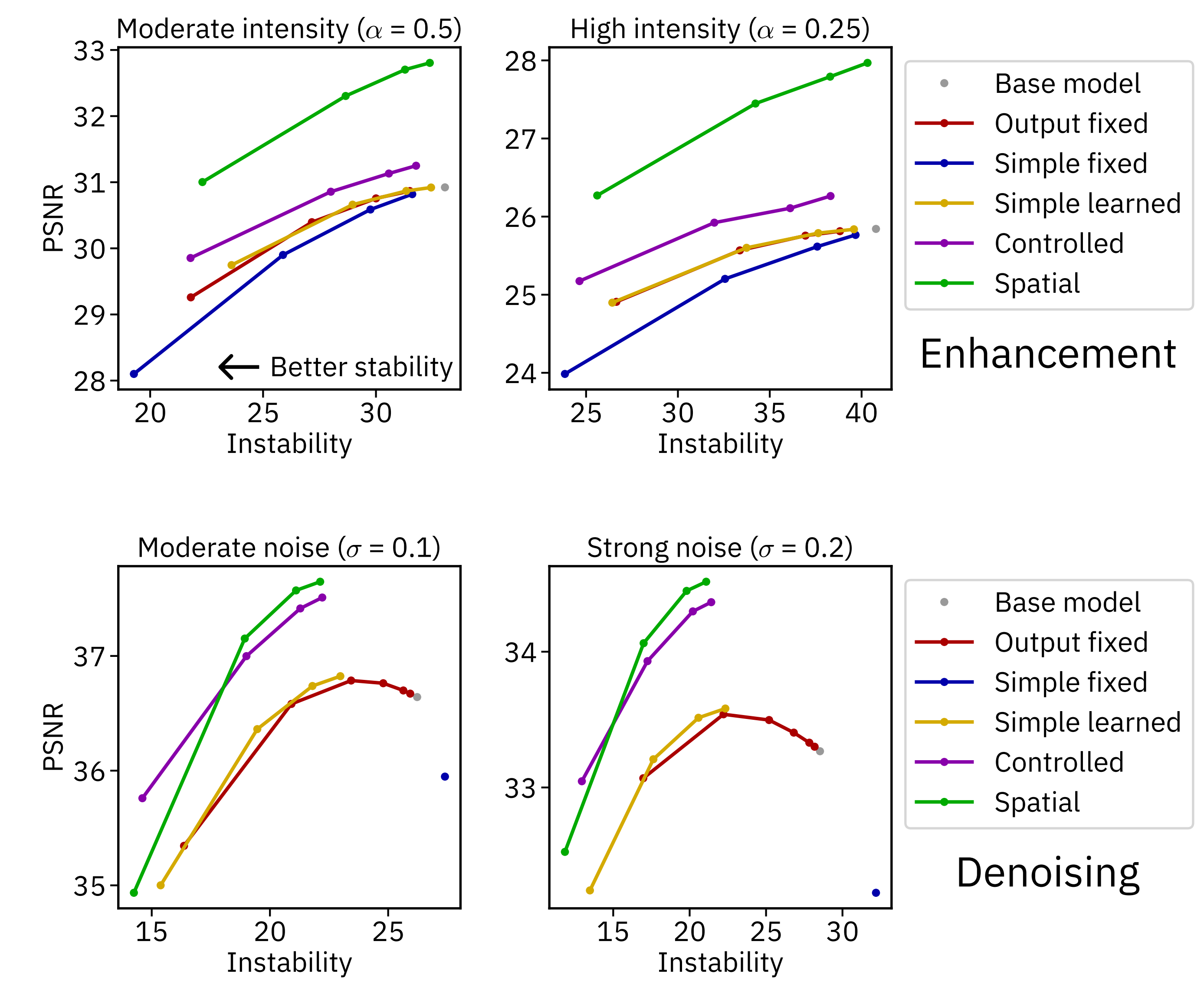
Stabilization Controllers



Versatility



Robustness-Accuracy Relationship



Corruption Robustness

