

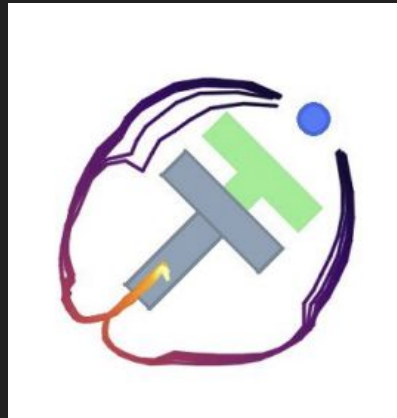
Diffusion Policy Evaluation

Franklin Wang and Ram Goel

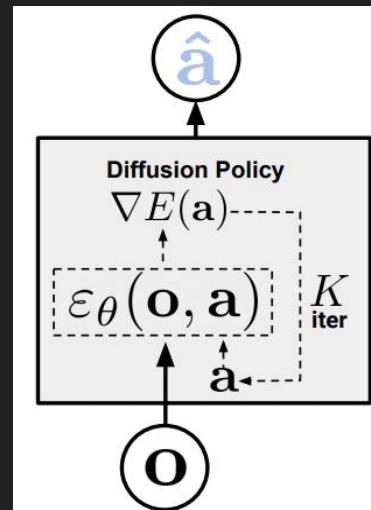
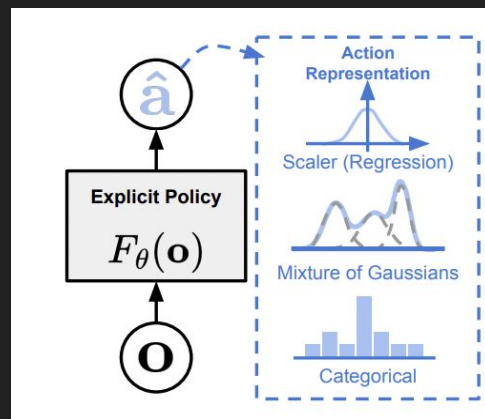
6.8200 Final Presentation

Multimodality in Policy Learning for Offline RL Problems

- Offline RL: Clone expert's policy given expert data
- Main challenge: **multimodal modeling capabilities**
- Approaches:
 - Explicit (Mixture of Gaussian, Binning)
 - Implicit/Diffusion



(From Diffusion Policy paper)



Problem Formulation: Push-T Task

- Observation space: 96 x 96 pixel
image of top-down 2D view
 - Action space: (x,y) coordinates for
hand to move to
-
- Success Criteria: Intersection over
Union (IoU) > 90%
 - No inherent reward
 - Dataset of expert examples for
behavior cloning



*Our trained **Diffusion Policy**
learning algorithm performing
Push-T*

Diffusion Policy Methodology Overview

Framework for training the noise-prediction network

(U-Net Framework)

$$\varepsilon_{\theta}(\mathbf{O}_t, \mathbf{A}_t^k, k)$$

Update Parameters & Repeat

$$\mathcal{L} = \text{MSE}(\epsilon^k, \varepsilon_{\theta}(\mathbf{x}^0 + \epsilon^k, k))$$

$$\mathbf{x}_0 + \epsilon_k \quad (\text{add noise})$$

$$\mathbf{x}_0 \quad (\text{sampled from observation})$$

Iterative Denoising Process

$$\mathbf{x}^{k-1} = \alpha(\mathbf{x}^k - \gamma \varepsilon_{\theta}(\mathbf{x}^k, k) + \mathcal{N}(0, \sigma^2 I))$$

$$\text{OR} \quad \mathbf{x}' = \mathbf{x} - \gamma \nabla E(\mathbf{x})$$

$$\mathbf{x}_k$$

(Gaussian noise)

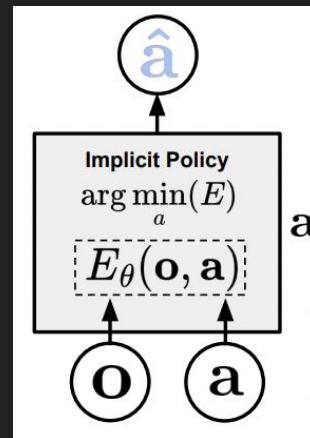
$$\mathbf{x}_{k-1}$$

$$\mathbf{x}_0$$

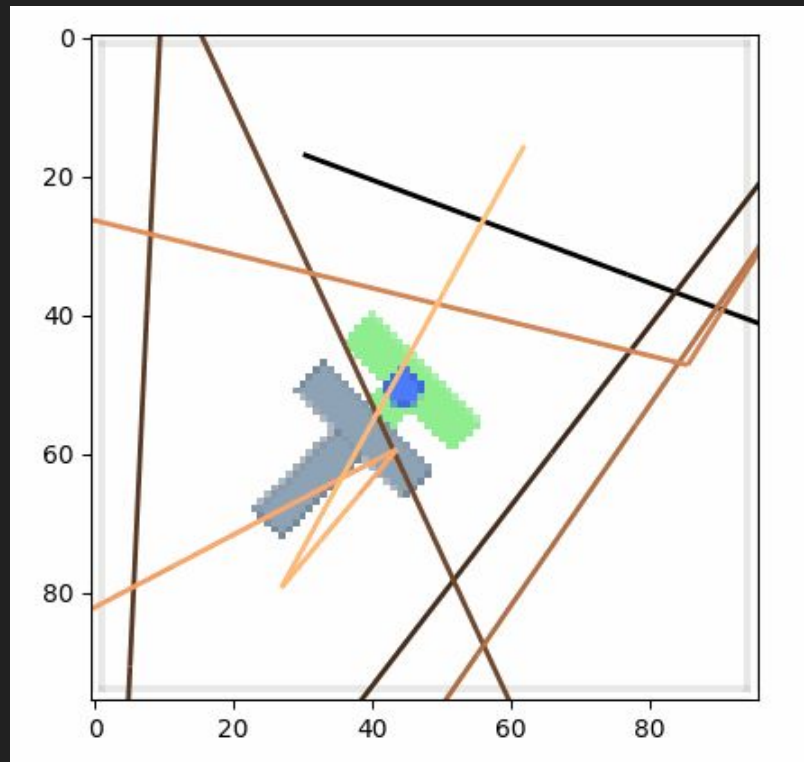
(generated data similar to expert policy)

- Encode observation images and actions into **latent space**
- Use **clipped** observations/actions for each timestep

Baseline: Implicit Behavior Cloning (IBC)



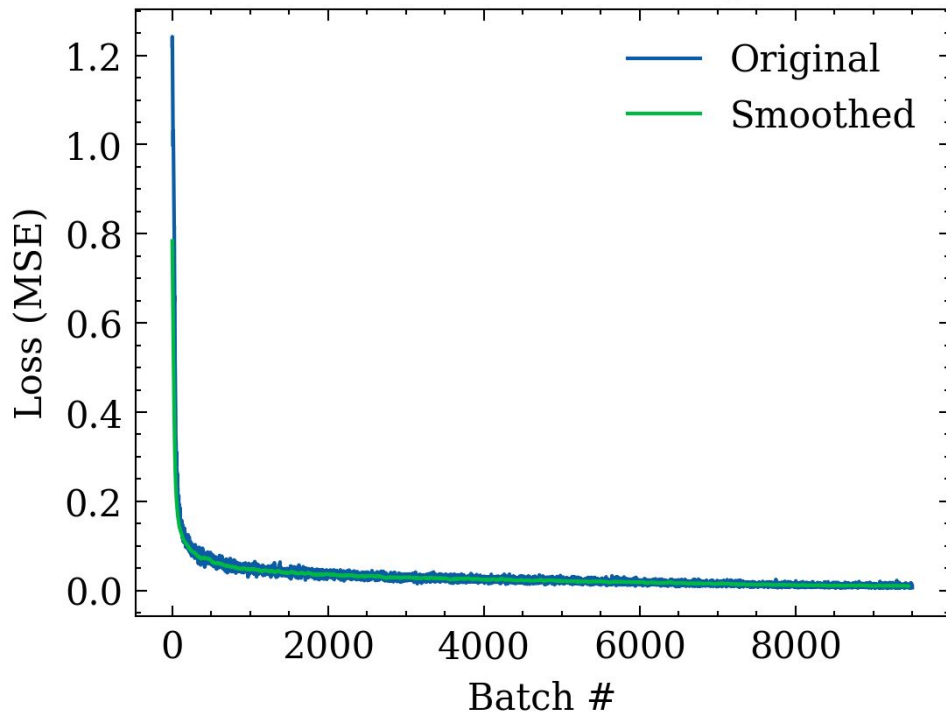
- Minimize an energy function for observation to select action
- Potential for **multimodal** modeling



Diffusion Demonstration
of our Trained Model

Diffusion Policy Results

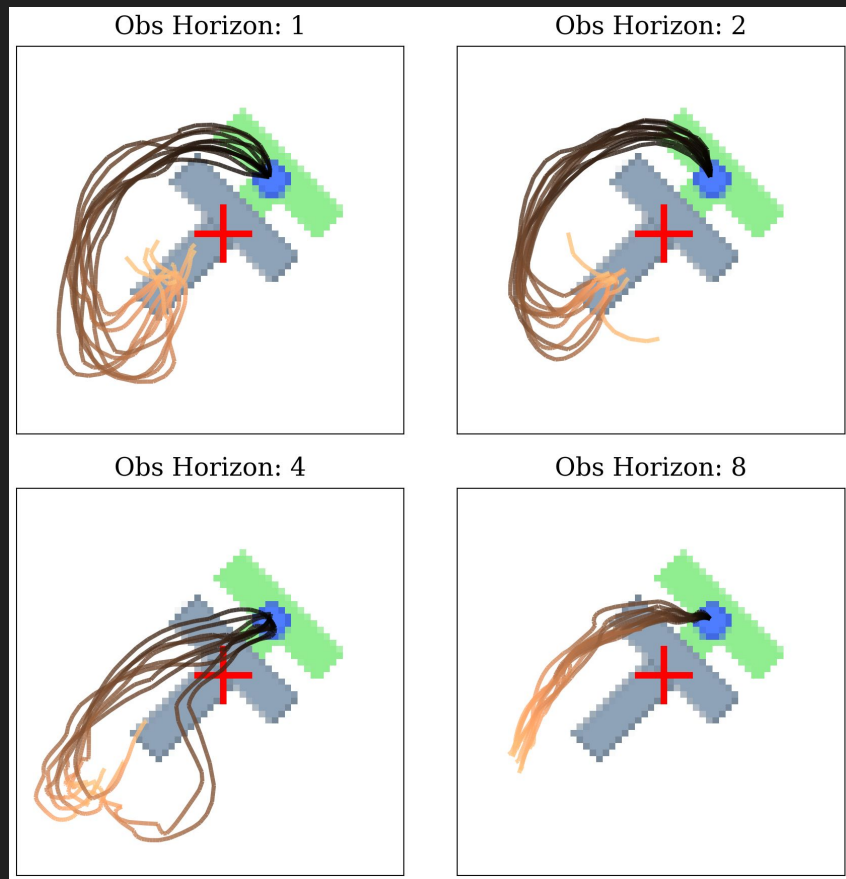
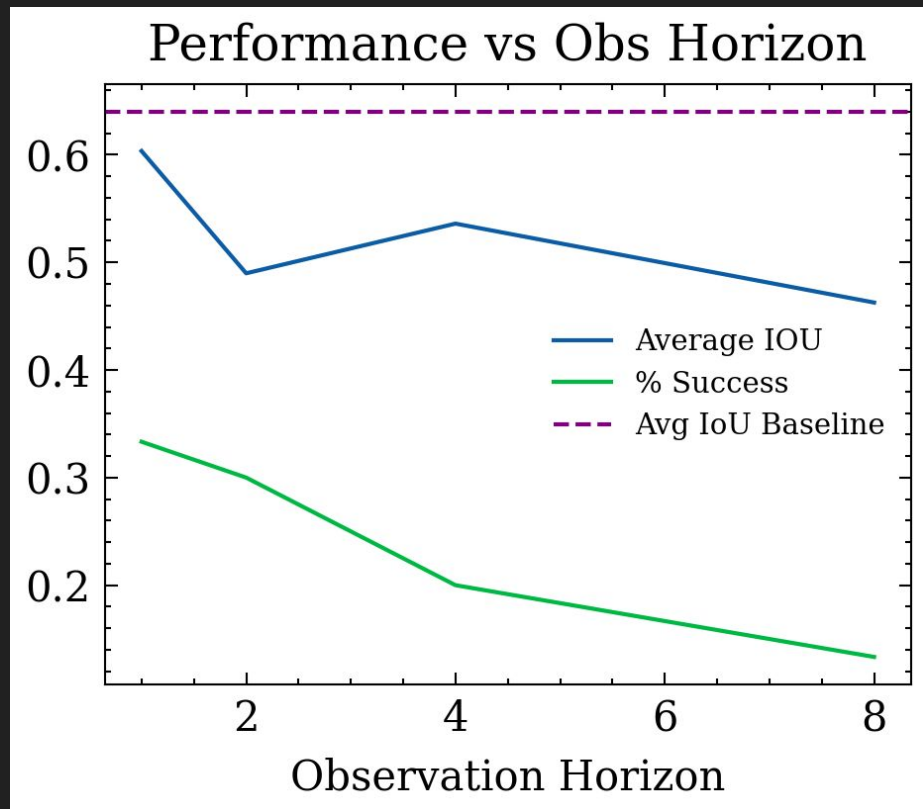
Diffusion Model Training Curve



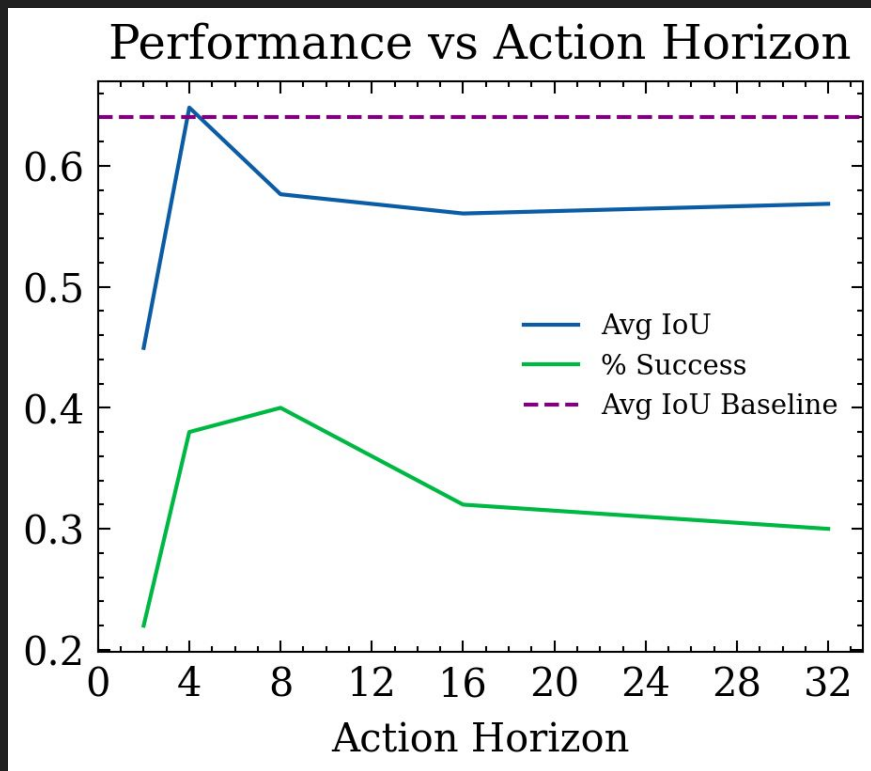
	Best Model	IBC
Average Best IoU	75%	64%
Success Rate (IoU > 0.9)	47%	42%



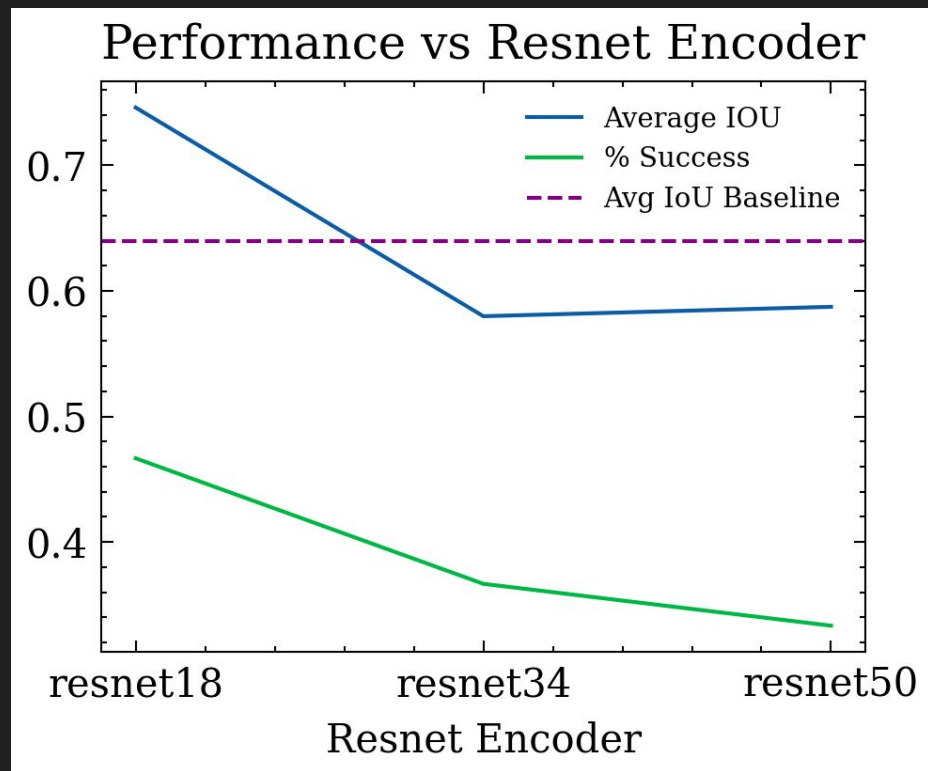
Observation Horizon: # of observations fed into the model per timestep



Action Horizon: # of actions we act upon at a time



Resnet Encoder is used to encode the image observations



Issues, Changes, and Learnings

Analysis:

- Main important hyperparameters
 - Action horizon
 - Observation Horizon
- Discrepancies between our findings vs original paper
 - Different “optimal” hyperparameters
 - Performance: ~90% of original
 - 75% vs 84% IoU Average

Challenges:

- Multimodality not as apparent as it should be
 - Possible overfitting to one mode due to longer training epochs
 - Size of validation/test set
- Difficulty training diffusion models