Homework 1: Imitation Learning

Qingpeng Kong

1 Action Chunking with MSE Loss

1.1 Policy and Training Setup

The policy is an MSE-based multilayer perceptron (MLP). It takes the environment state as input and outputs a fixed-length chunk of actions.

The network consists of three fully connected hidden layers with 256 units each. ReLU activations are used between layers, and the output layer is linear.

Training is performed on the provided Push-T demonstration dataset using the Adam optimizer with a learning rate of 3×10^{-4} and a batch size of 128. The model is trained for 400 epochs on a CPU. Evaluation is run periodically during training using the provided evaluation function.

1.2 Results

Figure 1 shows the training loss as a function of training steps. The loss decreases quickly at the beginning and stabilizes as training progresses.

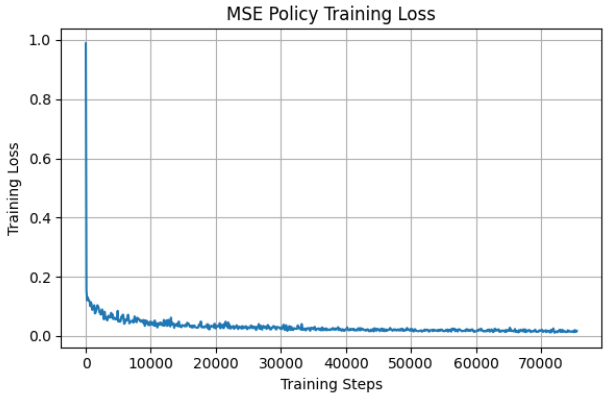

Figure 1: Training loss (MSE) versus training steps.

Figure 2 shows the evaluation mean reward over training. The policy performance improves steadily and reaches a final mean reward of approximately **0.68**.

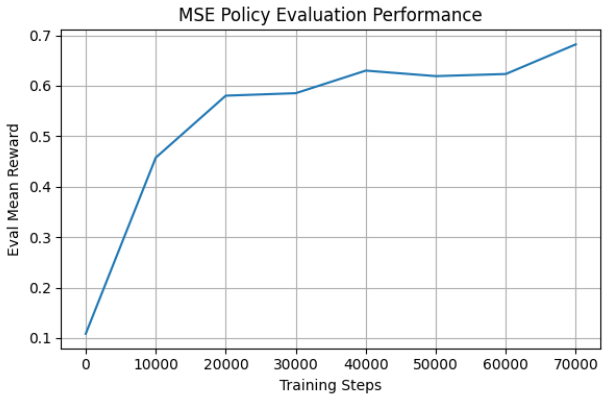

Figure 2: Evaluation mean reward versus training steps.

2 Action Chunking with Flow Matching

2.1 Policy and Training Setup

The FlowMatchingPolicy is implemented using the same MLP architecture as the MSE policy. The network takes as input the observation, a flattened action chunk, and the flow matching timestep τ, and outputs the corresponding velocity vector.

During training, the policy is optimized using the flow matching loss by interpolating between Gaussian noise and ground-truth action chunks. At inference time, action chunks are generated by sampling noise and integrating the learned vector field from $\tau = 0$ to $\tau = 1$ using Euler integration.

2.2 Results

Figure 3 shows the training loss over training steps. The loss decreases steadily and stabilizes toward the end of training.

Figure 4 shows the evaluation mean reward over training steps. The policy reaches a final mean evaluation reward of **0.77** after 75,500 training steps, satisfying the performance requirement.

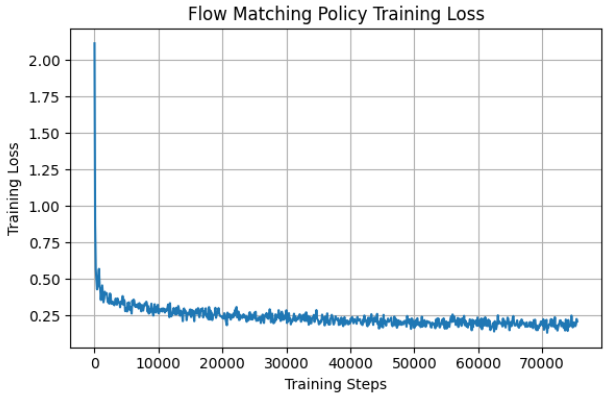

Figure 3: Training loss versus training steps for the flow matching policy.

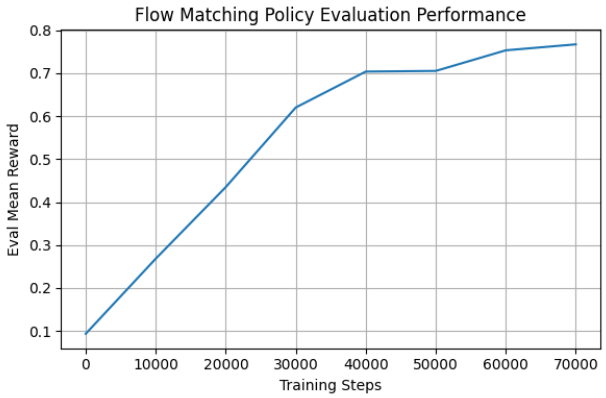

Figure 4: Evaluation mean reward versus training steps for the flow matching policy.

2.3 Comparison with MSE Policy

Compared to the MSE policy, the flow matching policy produces smoother and more consistent action chunks. From the rollout videos, the behavior appears less noisy and more stable, particularly when executing longer action chunks.