

DQN으로 Shoothing Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

# DQN으로 Shoothing Airplane Game 강화학습

먼저 여러가지 설정 변수 정의

```
import numpy as np
import gym
import torch
import torch.nn as nn
import torchvision.transforms as T

# Configuration paramaters for the whole setup
seed = 42
gamma = 0.99 # Discount factor for past rewards
epsilon = 1.0 # Epsilon greedy parameter
epsilon_min = 0.1 # Minimum epsilon greedy parameter
epsilon_max = 1.0 # Maximum epsilon greedy parameter
epsilon_interval = epsilon_max - epsilon_min # Rate at which to reduce chance
                                                    # of random action being taken

batch_size = 16 # Size of batch taken from replay buffer
max_steps_per_episode = 60
max_episodes = 10000
```

## 게임 환경 설정

상태(state) 정의

- 보드판의 모양: (8 \* 8) 행렬 \* 3 채널
- 채널 0: unseen
- 채널 1: hit
- 채널 2: miss

액션 정의

- 돌의 가능한 위치 (8 \* 8 = 64)

```
env = gym.make('gym_examples:gym_examples/ShootingAirplane-v0', render_mode="text")
```

env에서 정의한 action\_space, observation\_space 의 모양 확인

- action\_space: 3개의 값의 튜플 (벡터)
- observation\_space: HWC 형태의 이미지 (마지막 축이 단일 값인 15 \* 15 \* 1 텐서) -> pytorch를 사용할 경우 적절히 1 \* 15 \* 15 텐서로 수정필요

```
env.action_space.shape
```

```
(2,)
```

```
env.observation_space.shape
```

```
(8, 8, 1)
```

## 네트워크 정의하기

참고: Conv2d 파라미터

- in\_channels (int) – Number of channels in the input image
- out\_channels (int) – Number of channels produced by the convolution
- kernel\_size (int or tuple) – Size of the convolving kernel

DQN으로 Shooting Airplane Game 강화 학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

- stride (int or tuple, optional) – Stride of the convolution. Default: 1
- padding (int, tuple or str, optional) – Padding added to all four sides of the input. Default: 0
- padding\_mode (str, optional) – 'zeros', 'reflect', 'replicate' or 'circular'. Default: 'zeros'

```
import torch
import torch.nn as nn

num_actions = 64

class QModel(nn.Module):
    def __init__(self, num_actions):
        super(QModel, self).__init__()
        self.dropout = nn.Dropout(p=0.3)
        self.conv1 = nn.Conv2d(3, 16, kernel_size=3, stride=1, padding='same')
        self.conv2 = nn.Conv2d(16, 32, kernel_size=3, stride=1, padding='same')
        self.conv3 = nn.Conv2d(32, 32, kernel_size=3, stride=1)
        self.flatten = nn.Flatten()
        self.fc1 = nn.Linear(1152, 512)
        self.fc2 = nn.Linear(512, num_actions)

    def forward(self, x):
        x = nn.functional.relu(self.conv1(x))
        x = nn.functional.relu(self.conv2(x))
        x = self.dropout(x)
        x = nn.functional.relu(self.conv3(x))
        x = self.flatten(x)
        x = nn.functional.relu(self.fc1(x))
        x = self.dropout(x)
        action = self.fc2(x)
        return action
```

## 모델 빌딩 & 로스 및 최적화 계산기 만들기

```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")

# The first model makes the predictions for Q-values which are used to
# make a action.
model = QModel(num_actions)
model.to(device)

# Build a target model for the prediction of future rewards.
# The weights of a target model get updated every 10000 steps thus when the
# loss between the Q-values is calculated the target Q-value is stable.
model_target = QModel(num_actions)
model_target.to(device)

loss_function = nn.SmoothL1Loss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.00025)
```

device

device(type='cpu')

## Replay Buffer 정의

```
# Experience replay buffers
action_history = []
action_mask_history = []
state_history = []
state_next_history = []
rewards_history = []
done_history = []
episode_reward_history = []
running_reward = 0
episode_count = 0
frame_count = 0
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
# Number of frames to take random action and observe output
epsilon_random_frames = 50000
# Number of frames for exploration
epsilon_greedy_frames = 200000.0
# Maximum replay length
# Note: The Deepmind paper suggests 1000000 however this causes memory issues
max_memory_length = 500000
# Train the model after 4 actions
update_after_actions = 4
# How often to update the target network
update_target_network = 10000
```

## 전처리

- env가 리턴하는 observation은 일단 np.array이니 torch.tensor로 캐스팅
- env가 리턴하는 상태가 (8, 8, 1)의 HWC 이미지 텐서이므로 이를 (3, 15, 15)의 CHW 이미지로 변환
- One-hot 인코딩도 필요

```
# Function to preprocess the state
# note that player 1 = env player, player 2 = agent
def preprocess_state(env_observ):
    st = torch.from_numpy(env_observ).squeeze()
    st = st.to(torch.int64)
    st = torch.nn.functional.one_hot(st, num_classes=3)
    st = st.permute(2, 0, 1)
    return st.to(torch.float32)
```

## 중간 테스트

- env.reset로 board observation를 받아서
- preprocess\_state로 input tensor로 바꾸어주고
- model로 forward computing

```
board, info = env.reset()
```

```
board
```

```
array([[0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0]],

       [[0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0]],

       [[0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0]],

       [[0],
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
[0],
[0],
[0],
[0],
[0],
[0],
[0]],
```

```
[[0],
[0],
[0],
[0],
[0],
[0],
[0],
[0]],
```

```
[[0],
[0],
[0],
[0],
[0],
[0],
[0],
[0]],
```

```
[0],
[0],
[0],
[0],
[0],
[0],
[0],
[0]],
```

```
[[0],
[0],
[0],
[0],
[0],
[0],
[0],
[0]], dtype=uint8)
```

```
info['action_mask']
```

```
array([[1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1],
       [1, 1, 1, 1, 1, 1, 1, 1]], dtype=int32)
```

```
state = preprocess_state(board)
```

```
state
```

```
tensor([[[[1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.],
          [1., 1., 1., 1., 1., 1., 1., 1.]]],
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
[[0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.]]

[[0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.],
 [0., 0., 0., 0., 0., 0., 0., 0.]]])
```

state.shape

torch.Size([3, 8, 8])

```
with torch.no_grad():
    model_output = model(state.unsqueeze(0))
```

model\_output.shape

torch.Size([1, 64])

## Epsilon-greedy 액션 선택 함수

학습시 에피소드 생성하면서 사용 (주의: 입력은 batch axis 없음)

```
# Function to select an action
# model: the torch model to compute action-state value (i.e., q-value)
# state: a torch tensor (3 x 8 x 8) of float32, which is output by preprocess_state
# mask: a 64-size array (np.array)
def get_greedy_epsilon(model, state, mask):
    global epsilon

    #if frame_count < epsilon_random_frames or np.random.rand(1)[0] < epsilon:
    if np.random.rand(1)[0] < epsilon:
        action = np.random.choice([ i for i in range(num_actions) if mask[i] == 1 ])
    else:
        with torch.no_grad():
            # add a batch axis
            state_tensor = state.unsqueeze(0)
            # compute the q-values
            q_values = model(state_tensor)
            # select the q-values of valid actions
            action = torch.argmax(
                q_values.squeeze() + torch.from_numpy(mask) * 100., # trick to select a v
                dim=0)

            #valid_q = [ (i, q_values[0][i]) for i in range(64) if mask[i] == 1 ]
            # the action of maximum q-value
            #action, _ = max(valid_q, key=lambda e: e[1])

    # decay epsilon
    epsilon -= epsilon_interval / epsilon_greedy_frames
    epsilon = max(epsilon, epsilon_min)

    return action
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
mask = np.zeros((64,), dtype=np.int64)
mask[12] = 1.
get_greedy_epsilon(model, state, mask)
```

12

## Greedy 액션 선택 함수

나중에 evaluation 시 사용

```
def get_greedy_action(model, state, mask):
    global epsilon

    with torch.no_grad():
        state_tensor = state.unsqueeze(0) # batch dimension
        q_values = model(state_tensor)

        action = torch.argmax(
            q_values.squeeze() + torch.from_numpy(mask) * 100., # trick to select a v
            dim=0)

    return action
```

## Update 파트

- Replay buffer 에서 batch하나를 샘플링하고,
- model을 update한다.

```
# sample a batch of _batch_size from replay buffers
# return numpy.ndarrays
def sample_batch(_batch_size):
    # Get indices of samples for replay buffers
    indices = np.random.choice(range(len(done_history)), size=_batch_size, replace=False)

    state_sample = np.array([state_history[i].squeeze(0).numpy() for i in indices])
    state_next_sample = np.array([state_next_history[i].squeeze(0).numpy() for i in indices])
    rewards_sample = np.array([rewards_history[i] for i in indices], dtype=np.float32)
    action_sample = np.array([action_history[i] for i in indices])

    # action mask is the mask for the valid actions at the 'next' state
    action_mask_sample = np.array([action_mask_history[i] for i in indices])
    done_sample = np.array([float(done_history[i]) for i in indices])

    return state_sample, state_next_sample, rewards_sample, action_sample, action_mask_sample, done_sample
```

```
# Function to update the Q-network
def update_network():
    # sample a batch of ...
    state_sample, state_next_sample, rewards_sample, action_sample, action_mask_sample, done_sample = sample_batch(batch_size)

    # Convert numpy arrays to PyTorch tensors
    state_sample = torch.tensor(state_sample, dtype=torch.float32).to(device)
    state_next_sample = torch.tensor(state_next_sample, dtype=torch.float32).to(device)
    action_sample = torch.tensor(action_sample, dtype=torch.int64).to(device)
    action_mask_sample = torch.tensor(action_mask_sample, dtype=torch.int64).to(device)
    rewards_sample = torch.tensor(rewards_sample, dtype=torch.float32).to(device)
    done_sample = torch.tensor(done_sample, dtype=torch.float32).to(device)

    # Compute the target Q-values for the states
    with torch.no_grad():
        future_rewards = model_target(state_next_sample)
        #future_rewards = future_rewards.cpu()

    # compute the q-value for the next state and the action maximizing the q-value
```

DQN으로 Shooting Airplane Game 강화 학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
# note: the action should be valid (i.e., mask is set to 1)
max_q_values = torch.max(
    future_rewards + action_mask_sample * 100., # trick to select a valid action
    dim=1).values.detach() - 100.

# compute the target q-value
# if the step was final, max_q_values should not be added
# we assume that the negative return of the opposite player is the return of next
# that is, G(t) = r(t+1) - g*r(t+2) + g^2*r(t+3) - g^3*r(t+4) + ...
target_q_values = rewards_sample + gamma * max_q_values * (1. - done_sample)

# It's forward propagation! Compute the Q-values for the taken actions
q_values = model(state_sample)
#q_values = q_values.cpu()
q_values_action = q_values.gather(dim=1, index=action_sample.unsqueeze(1)).squeeze(1)

# Compute the loss
loss = loss_function(q_values_action, target_q_values)

# Perform the optimization step
optimizer.zero_grad()
loss.backward()
optimizer.step()
```

## Test A Single Training Step

에피소드를 초기화 하고 첫 스텝 상태를 가져오기

```
board, info = env.reset()
board.shape
```

```
(8, 8, 1)
```

```
info['action_mask'].shape
```

```
(8, 8)
```

```
state = preprocess_state(board)
state[0] # the positions of stones of the agent player
```

```
tensor([[1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1.]])
```

```
action_mask = info['action_mask'].reshape((-1,))
action_mask.shape
```

```
(64,)
```

```
action = get_greedy_epsilon(model, state, action_mask)
print(action)
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

7

Get the next step from environment.

```
board, reward, done, _, info = env.step((action // 8, action % 8))
state_next = preprocess_state(board)
action_mask = info['action_mask'].reshape((-1,))
state_next.shape
```

```
/home/yhkim/miniconda3/envs/torch/lib/python3.10/site-packages/gym/utils/passive_env_checker.py:10: in isinstance
    if not isinstance(terminated, (bool, np.bool8)):
```

```
torch.Size([3, 8, 8])
```

```
action_mask
```

```
array([1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1],
      dtype=int32)
```

```
reward
```

```
-1
```

Put the two step into replay buffer.

```
# Save actions and states in replay buffer
action_history.append(action)
action_mask_history.append(action_mask)
state_history.append(state)
state_next_history.append(state_next)
done_history.append(done)
rewards_history.append(reward)

state = state_next
```

Generate one more step.

```
# Get an action by epsilon-greedy policy
action = get_greedy_epsilon(model, state, action_mask)
action
```

```
63
```

```
board, reward, done, _, info = env.step((action // 8, action % 8))
state_next = preprocess_state(board)
action_mask = info['action_mask'].reshape((-1,))
state_next.shape
```

```
torch.Size([3, 8, 8])
```

```
action_history.append(action)
action_mask_history.append(action_mask)
```



DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
state_history.append(state)
state_next_history.append(state_next)
done_history.append(done)
rewards_history.append(reward)
```

Test batch sampling.

```
state_sample, state_next_sample, rewards_sample, action_sample, action_mask_sample, done_sample = sample_batch(2)
```

```
state_sample.shape
```

```
(2, 3, 8, 8)
```

```
state_next_sample.shape
```

```
(2, 3, 8, 8)
```

```
rewards_sample.shape
```

```
(2,)
```

```
action_sample.shape
```

```
(2,)
```

```
action_mask_sample.shape
```

```
(2, 64)
```

```
# Convert numpy arrays to PyTorch tensors
state_sample = torch.tensor(state_sample, dtype=torch.float32)
state_next_sample = torch.tensor(state_next_sample, dtype=torch.float32)
action_sample = torch.tensor(action_sample, dtype=torch.int64)
action_mask_sample = torch.tensor(action_mask_sample, dtype=torch.int64)
rewards_sample = torch.tensor(rewards_sample, dtype=torch.float32)
done_sample = torch.tensor(done_sample, dtype=torch.float32)
```

Test model prediction with the batch.

```
with torch.no_grad():
    future_rewards = model_target(state_next_sample.to(device))

future_rewards.shape
```

```
torch.Size([2, 64])
```

```
action_sample
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
tensor([63, 7])
```

```
torch.max(future_rewards, dim=1)
```

```
torch.return_types.max(
  values=tensor([0.0501, 0.0485]),
  indices=tensor([18, 54]))
```

```
torch.max(
  future_rewards + action_mask_sample.to(device) * 100.,
  dim=1)
```

```
torch.return_types.max(
  values=tensor([100.0501, 100.0485]),
  indices=tensor([18, 54]))
```

```
torch.max(
  future_rewards + action_mask_sample.to(device) * 100.,
  dim=1).values.detach() - 100.
```

```
tensor([0.0501, 0.0485])
```

```
done_sample.unsqueeze(1)
```

```
tensor([[0.],
        [0.]])
```

```
with torch.no_grad():
    future_rewards = model_target(state_next_sample.to(device))

    # compute the q-value for the next state and the action maximizing the q-value
    # note: the action should be valid (i.e., mask is set to 1)
    max_q_values = torch.max(
        future_rewards + action_mask_sample.to(device) * 100., # trick to select a valid
        dim=1).values.detach() - 100.

    # compute the target q-value
    # if the step was final, max_q_values should not be added
    target_q_values = rewards_sample.to(device) + max_q_values * (1. - done_sample.to(device))

    target_q_values.shape
```

```
torch.Size([2])
```

```
# It's forward propagation! Compute the Q-values for the taken actions
q_values = model(state_sample.to(device))
q_values_action = q_values.gather(dim=1, index=action_sample.to(device).unsqueeze(1)).squeeze(1)
q_values_action.shape
```

```
torch.Size([2])
```

```
# Compute the loss
```

DQN으로 Shooting Airplane Game 강화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
loss = loss_function(q_values_action, target_q_values)
```

```
# Perform the optimization step
optimizer.zero_grad()
loss.backward()
optimizer.step()
```

## Run DQN Training

```
for _ in range(max_episodes):
    state, info = env.reset()
    state = preprocess_state(state)
    action_mask = info['action_mask'].reshape((-1,))
    episode_reward = 0

    for timestep in range(1, max_steps_per_episode):
        frame_count += 1

        # Select an action
        #state_cuda = state.to(device)
        action = get_greedy_epsilon(model, state, action_mask)
        if action < 0:
            print(action_mask)

        # Take the selected action
        state_next, reward, done, _, info = env.step((action // 8, action % 8))
        state_next = preprocess_state(state_next)
        action_mask = info['action_mask'].reshape((-1,))

        episode_reward += reward

        # Store the transition in the replay buffer
        action_history.append(action)
        action_mask_history.append(action_mask)
        state_history.append(state)
        state_next_history.append(state_next)
        rewards_history.append(reward)
        done_history.append(done)

        state = state_next

        # Update every fourth frame and once batch size is over 32
        if frame_count % update_after_actions == 0 and len(done_history) > batch_size:
            update_network()

        if frame_count % update_target_network == 0:
            model_target.load_state_dict(model.state_dict())

        # Limit the state and reward history
        if len(rewards_history) > max_memory_length:
            del rewards_history[:1]
            del state_history[:1]
            del state_next_history[:1]
            del action_history[:1]
            del action_mask_history[:1]
            del done_history[:1]

        if done:
            break

    episode_count += 1
    episode_reward_history.append(episode_reward)

    # Update running reward to check condition for solving
    if len(episode_reward_history) > 100:
        del episode_reward_history[:1]
    running_reward = np.mean(episode_reward_history)

    if episode_count % 10 == 0:
        print(f"Episode: {episode_count}, Frame count: {frame_count}, Running reward: {running_reward}")
```

DQN으로 Shooting Airplane Game 강화 학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
if episode_count % 5000 == 0:
    torch.save(model, 'model.{}.format(episode_count)')
    #if running_reward > 20:
    #    print(f"Solved at episode {episode_count}!")
    #    break
```

```
torch.save(model, 'model.final')
```

```
Episode: 20, Frame count: 1262, Running reward: -38.25
Episode: 30, Frame count: 1832, Running reward: -38.36666666666667
Episode: 40, Frame count: 2386, Running reward: -38.075
Episode: 50, Frame count: 2971, Running reward: -38.52
Episode: 60, Frame count: 3554, Running reward: -38.75
Episode: 70, Frame count: 4121, Running reward: -38.628571428571426
Episode: 80, Frame count: 4693, Running reward: -38.625
Episode: 90, Frame count: 5253, Running reward: -38.422222222222224
Episode: 100, Frame count: 5831, Running reward: -38.46
Episode: 110, Frame count: 6379, Running reward: -38.22
Episode: 120, Frame count: 6936, Running reward: -38.14
Episode: 130, Frame count: 7491, Running reward: -37.95
Episode: 140, Frame count: 8054, Running reward: -37.98
Episode: 150, Frame count: 8630, Running reward: -37.81
Episode: 160, Frame count: 9165, Running reward: -37.27
Episode: 170, Frame count: 9747, Running reward: -37.48
Episode: 180, Frame count: 10310, Running reward: -37.41
Episode: 190, Frame count: 10854, Running reward: -37.23
Episode: 200, Frame count: 11423, Running reward: -37.18
Episode: 210, Frame count: 12007, Running reward: -37.5
Episode: 220, Frame count: 12556, Running reward: -37.4
Episode: 230, Frame count: 13102, Running reward: -37.31
Episode: 240, Frame count: 13652, Running reward: -37.16
Episode: 250, Frame count: 14161, Running reward: -36.47
Episode: 260, Frame count: 14674, Running reward: -36.19
Episode: 270, Frame count: 15216, Running reward: -35.69
Episode: 280, Frame count: 15761, Running reward: -35.41
Episode: 290, Frame count: 16290, Running reward: -35.22
Episode: 300, Frame count: 16856, Running reward: -35.13
Episode: 310, Frame count: 17374, Running reward: -34.33
Episode: 320, Frame count: 17917, Running reward: -34.25
Episode: 330, Frame count: 18404, Running reward: -33.6
Episode: 340, Frame count: 18946, Running reward: -33.46
Episode: 350, Frame count: 19440, Running reward: -33.23
Episode: 360, Frame count: 19971, Running reward: -33.37
Episode: 370, Frame count: 20500, Running reward: -33.22
Episode: 380, Frame count: 21050, Running reward: -33.29
Episode: 390, Frame count: 21553, Running reward: -33.01
Episode: 400, Frame count: 22058, Running reward: -32.34
Episode: 410, Frame count: 22615, Running reward: -32.85
Episode: 420, Frame count: 23123, Running reward: -32.46
Episode: 430, Frame count: 23583, Running reward: -32.15
Episode: 440, Frame count: 24096, Running reward: -31.86
Episode: 450, Frame count: 24591, Running reward: -31.87
Episode: 460, Frame count: 25084, Running reward: -31.49
Episode: 470, Frame count: 25548, Running reward: -30.8
Episode: 480, Frame count: 26035, Running reward: -30.11
Episode: 490, Frame count: 26532, Running reward: -30.09
Episode: 500, Frame count: 27055, Running reward: -30.27
Episode: 510, Frame count: 27570, Running reward: -29.77
Episode: 520, Frame count: 28079, Running reward: -29.8
Episode: 530, Frame count: 28603, Running reward: -30.44
Episode: 540, Frame count: 29071, Running reward: -29.95
Episode: 550, Frame count: 29538, Running reward: -29.69
Episode: 560, Frame count: 30027, Running reward: -29.69
Episode: 570, Frame count: 30497, Running reward: -29.73
Episode: 580, Frame count: 31038, Running reward: -30.29
Episode: 590, Frame count: 31535, Running reward: -30.25
Episode: 600, Frame count: 32025, Running reward: -29.94
Episode: 610, Frame count: 32523, Running reward: -29.73
Episode: 620, Frame count: 33042, Running reward: -29.81
Episode: 630, Frame count: 33522, Running reward: -29.37
Episode: 640, Frame count: 33939, Running reward: -28.86
Episode: 650, Frame count: 34428, Running reward: -29.08
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 660, Frame count: 34875, Running reward: -28.62
Episode: 670, Frame count: 35372, Running reward: -28.91
Episode: 680, Frame count: 35862, Running reward: -28.4
Episode: 690, Frame count: 36293, Running reward: -27.74
Episode: 700, Frame count: 36770, Running reward: -27.63
Episode: 710, Frame count: 37208, Running reward: -27.07
Episode: 720, Frame count: 37653, Running reward: -26.33
Episode: 730, Frame count: 38138, Running reward: -26.36
Episode: 740, Frame count: 38552, Running reward: -26.33
Episode: 750, Frame count: 38960, Running reward: -25.5
Episode: 760, Frame count: 39454, Running reward: -25.97
Episode: 770, Frame count: 39887, Running reward: -25.33
Episode: 780, Frame count: 40310, Running reward: -24.62
Episode: 790, Frame count: 40788, Running reward: -25.13
Episode: 800, Frame count: 41271, Running reward: -25.13
Episode: 810, Frame count: 41717, Running reward: -25.17
Episode: 820, Frame count: 42169, Running reward: -25.22
Episode: 830, Frame count: 42640, Running reward: -25.08
Episode: 840, Frame count: 43093, Running reward: -25.49
Episode: 850, Frame count: 43557, Running reward: -26.11
Episode: 860, Frame count: 44028, Running reward: -25.88
Episode: 870, Frame count: 44491, Running reward: -26.18
Episode: 880, Frame count: 44913, Running reward: -26.17
Episode: 890, Frame count: 45325, Running reward: -25.47
Episode: 900, Frame count: 45738, Running reward: -24.77
Episode: 910, Frame count: 46197, Running reward: -24.92
Episode: 920, Frame count: 46593, Running reward: -24.36
Episode: 930, Frame count: 47069, Running reward: -24.43
Episode: 940, Frame count: 47503, Running reward: -24.24
Episode: 950, Frame count: 47972, Running reward: -24.23
Episode: 960, Frame count: 48419, Running reward: -23.99
Episode: 970, Frame count: 48876, Running reward: -23.91
Episode: 980, Frame count: 49341, Running reward: -24.36
Episode: 990, Frame count: 49762, Running reward: -24.47
Episode: 1000, Frame count: 50242, Running reward: -25.14
Episode: 1010, Frame count: 50660, Running reward: -24.71
Episode: 1020, Frame count: 51117, Running reward: -25.34
Episode: 1030, Frame count: 51581, Running reward: -25.26
Episode: 1040, Frame count: 52060, Running reward: -25.75
Episode: 1050, Frame count: 52492, Running reward: -25.38
Episode: 1060, Frame count: 52869, Running reward: -24.68
Episode: 1070, Frame count: 53311, Running reward: -24.53
Episode: 1080, Frame count: 53732, Running reward: -24.07
Episode: 1090, Frame count: 54156, Running reward: -24.1
Episode: 1100, Frame count: 54564, Running reward: -23.38
Episode: 1110, Frame count: 54991, Running reward: -23.47
Episode: 1120, Frame count: 55372, Running reward: -22.69
Episode: 1130, Frame count: 55786, Running reward: -22.13
Episode: 1140, Frame count: 56169, Running reward: -21.11
Episode: 1150, Frame count: 56580, Running reward: -20.92
Episode: 1160, Frame count: 56988, Running reward: -21.23
Episode: 1170, Frame count: 57360, Running reward: -20.53
Episode: 1180, Frame count: 57782, Running reward: -20.54
Episode: 1190, Frame count: 58211, Running reward: -20.57
Episode: 1200, Frame count: 58602, Running reward: -20.4
Episode: 1210, Frame count: 59009, Running reward: -20.2
Episode: 1220, Frame count: 59436, Running reward: -20.68
Episode: 1230, Frame count: 59848, Running reward: -20.66
Episode: 1240, Frame count: 60309, Running reward: -21.44
Episode: 1250, Frame count: 60731, Running reward: -21.53
Episode: 1260, Frame count: 61117, Running reward: -21.31
Episode: 1270, Frame count: 61531, Running reward: -21.73
Episode: 1280, Frame count: 61932, Running reward: -21.52
Episode: 1290, Frame count: 62335, Running reward: -21.26
Episode: 1300, Frame count: 62726, Running reward: -21.26
Episode: 1310, Frame count: 63145, Running reward: -21.4
Episode: 1320, Frame count: 63535, Running reward: -21.01
Episode: 1330, Frame count: 63911, Running reward: -20.65
Episode: 1340, Frame count: 64313, Running reward: -20.06
Episode: 1350, Frame count: 64714, Running reward: -19.85
Episode: 1360, Frame count: 65096, Running reward: -19.81
Episode: 1370, Frame count: 65502, Running reward: -19.73
Episode: 1380, Frame count: 65924, Running reward: -19.94
Episode: 1390, Frame count: 66340, Running reward: -20.07
Episode: 1400, Frame count: 66749, Running reward: -20.27
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 1410, Frame count: 67159, Running reward: -20.16
Episode: 1420, Frame count: 67530, Running reward: -19.97
Episode: 1430, Frame count: 67886, Running reward: -19.77
Episode: 1440, Frame count: 68260, Running reward: -19.51
Episode: 1450, Frame count: 68687, Running reward: -19.79
Episode: 1460, Frame count: 69131, Running reward: -20.41
Episode: 1470, Frame count: 69518, Running reward: -20.22
Episode: 1480, Frame count: 69897, Running reward: -19.79
Episode: 1490, Frame count: 70292, Running reward: -19.6
Episode: 1500, Frame count: 70666, Running reward: -19.23
Episode: 1510, Frame count: 71069, Running reward: -19.16
Episode: 1520, Frame count: 71418, Running reward: -18.94
Episode: 1530, Frame count: 71782, Running reward: -19.02
Episode: 1540, Frame count: 72155, Running reward: -18.99
Episode: 1550, Frame count: 72551, Running reward: -18.66
Episode: 1560, Frame count: 72944, Running reward: -18.15
Episode: 1570, Frame count: 73361, Running reward: -18.45
Episode: 1580, Frame count: 73743, Running reward: -18.5
Episode: 1590, Frame count: 74118, Running reward: -18.28
Episode: 1600, Frame count: 74467, Running reward: -18.03
Episode: 1610, Frame count: 74833, Running reward: -17.66
Episode: 1620, Frame count: 75210, Running reward: -17.94
Episode: 1630, Frame count: 75579, Running reward: -17.99
Episode: 1640, Frame count: 75959, Running reward: -18.06
Episode: 1650, Frame count: 76353, Running reward: -18.04
Episode: 1660, Frame count: 76695, Running reward: -17.53
Episode: 1670, Frame count: 77061, Running reward: -17.02
Episode: 1680, Frame count: 77395, Running reward: -16.52
Episode: 1690, Frame count: 77744, Running reward: -16.26
Episode: 1700, Frame count: 78137, Running reward: -16.7
Episode: 1710, Frame count: 78527, Running reward: -16.94
Episode: 1720, Frame count: 78847, Running reward: -16.37
Episode: 1730, Frame count: 79182, Running reward: -16.03
Episode: 1740, Frame count: 79549, Running reward: -15.9
Episode: 1750, Frame count: 79910, Running reward: -15.57
Episode: 1760, Frame count: 80239, Running reward: -15.44
Episode: 1770, Frame count: 80600, Running reward: -15.39
Episode: 1780, Frame count: 80971, Running reward: -15.76
Episode: 1790, Frame count: 81359, Running reward: -16.15
Episode: 1800, Frame count: 81660, Running reward: -15.23
Episode: 1810, Frame count: 81994, Running reward: -14.67
Episode: 1820, Frame count: 82328, Running reward: -14.81
Episode: 1830, Frame count: 82653, Running reward: -14.71
Episode: 1840, Frame count: 83009, Running reward: -14.6
Episode: 1850, Frame count: 83333, Running reward: -14.23
Episode: 1860, Frame count: 83641, Running reward: -14.02
Episode: 1870, Frame count: 83945, Running reward: -13.45
Episode: 1880, Frame count: 84279, Running reward: -13.08
Episode: 1890, Frame count: 84641, Running reward: -12.82
Episode: 1900, Frame count: 84965, Running reward: -13.05
Episode: 1910, Frame count: 85323, Running reward: -13.29
Episode: 1920, Frame count: 85626, Running reward: -12.98
Episode: 1930, Frame count: 85965, Running reward: -13.12
Episode: 1940, Frame count: 86255, Running reward: -12.46
Episode: 1950, Frame count: 86632, Running reward: -12.99
Episode: 1960, Frame count: 86941, Running reward: -13.0
Episode: 1970, Frame count: 87280, Running reward: -13.35
Episode: 1980, Frame count: 87633, Running reward: -13.54
Episode: 1990, Frame count: 87947, Running reward: -13.06
Episode: 2000, Frame count: 88294, Running reward: -13.29
Episode: 2010, Frame count: 88596, Running reward: -12.73
Episode: 2020, Frame count: 88893, Running reward: -12.67
Episode: 2030, Frame count: 89210, Running reward: -12.45
Episode: 2040, Frame count: 89555, Running reward: -13.0
Episode: 2050, Frame count: 89879, Running reward: -12.47
Episode: 2060, Frame count: 90214, Running reward: -12.73
Episode: 2070, Frame count: 90520, Running reward: -12.4
Episode: 2080, Frame count: 90839, Running reward: -12.06
Episode: 2090, Frame count: 91134, Running reward: -11.87
Episode: 2100, Frame count: 91440, Running reward: -11.46
Episode: 2110, Frame count: 91737, Running reward: -11.41
Episode: 2120, Frame count: 92027, Running reward: -11.34
Episode: 2130, Frame count: 92316, Running reward: -11.06
Episode: 2140, Frame count: 92619, Running reward: -10.64
Episode: 2150, Frame count: 92936, Running reward: -10.57
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 2160, Frame count: 93285, Running reward: -10.71
Episode: 2170, Frame count: 93564, Running reward: -10.44
Episode: 2180, Frame count: 93859, Running reward: -10.2
Episode: 2190, Frame count: 94216, Running reward: -10.82
Episode: 2200, Frame count: 94562, Running reward: -11.22
Episode: 2210, Frame count: 94856, Running reward: -11.19
Episode: 2220, Frame count: 95195, Running reward: -11.68
Episode: 2230, Frame count: 95515, Running reward: -11.99
Episode: 2240, Frame count: 95829, Running reward: -12.1
Episode: 2250, Frame count: 96108, Running reward: -11.72
Episode: 2260, Frame count: 96410, Running reward: -11.25
Episode: 2270, Frame count: 96731, Running reward: -11.67
Episode: 2280, Frame count: 97037, Running reward: -11.78
Episode: 2290, Frame count: 97309, Running reward: -10.93
Episode: 2300, Frame count: 97619, Running reward: -10.57
Episode: 2310, Frame count: 97889, Running reward: -10.33
Episode: 2320, Frame count: 98167, Running reward: -9.72
Episode: 2330, Frame count: 98445, Running reward: -9.3
Episode: 2340, Frame count: 98783, Running reward: -9.54
Episode: 2350, Frame count: 99071, Running reward: -9.63
Episode: 2360, Frame count: 99402, Running reward: -9.92
Episode: 2370, Frame count: 99701, Running reward: -9.7
Episode: 2380, Frame count: 99991, Running reward: -9.54
Episode: 2390, Frame count: 100279, Running reward: -9.7
Episode: 2400, Frame count: 100572, Running reward: -9.53
Episode: 2410, Frame count: 100864, Running reward: -9.75
Episode: 2420, Frame count: 101148, Running reward: -9.81
Episode: 2430, Frame count: 101443, Running reward: -9.98
Episode: 2440, Frame count: 101742, Running reward: -9.59
Episode: 2450, Frame count: 102061, Running reward: -9.9
Episode: 2460, Frame count: 102348, Running reward: -9.46
Episode: 2470, Frame count: 102614, Running reward: -9.13
Episode: 2480, Frame count: 102904, Running reward: -9.13
Episode: 2490, Frame count: 103171, Running reward: -8.92
Episode: 2500, Frame count: 103419, Running reward: -8.47
Episode: 2510, Frame count: 103717, Running reward: -8.53
Episode: 2520, Frame count: 103999, Running reward: -8.51
Episode: 2530, Frame count: 104273, Running reward: -8.3
Episode: 2540, Frame count: 104541, Running reward: -7.99
Episode: 2550, Frame count: 104834, Running reward: -7.73
Episode: 2560, Frame count: 105111, Running reward: -7.63
Episode: 2570, Frame count: 105420, Running reward: -8.06
Episode: 2580, Frame count: 105667, Running reward: -7.63
Episode: 2590, Frame count: 105937, Running reward: -7.66
Episode: 2600, Frame count: 106210, Running reward: -7.91
Episode: 2610, Frame count: 106470, Running reward: -7.53
Episode: 2620, Frame count: 106766, Running reward: -7.67
Episode: 2630, Frame count: 107036, Running reward: -7.63
Episode: 2640, Frame count: 107289, Running reward: -7.48
Episode: 2650, Frame count: 107529, Running reward: -6.95
Episode: 2660, Frame count: 107795, Running reward: -6.84
Episode: 2670, Frame count: 108038, Running reward: -6.18
Episode: 2680, Frame count: 108329, Running reward: -6.62
Episode: 2690, Frame count: 108583, Running reward: -6.46
Episode: 2700, Frame count: 108850, Running reward: -6.4
Episode: 2710, Frame count: 109115, Running reward: -6.45
Episode: 2720, Frame count: 109370, Running reward: -6.04
Episode: 2730, Frame count: 109650, Running reward: -6.14
Episode: 2740, Frame count: 109915, Running reward: -6.26
Episode: 2750, Frame count: 110152, Running reward: -6.23
Episode: 2760, Frame count: 110446, Running reward: -6.51
Episode: 2770, Frame count: 110698, Running reward: -6.6
Episode: 2780, Frame count: 110966, Running reward: -6.37
Episode: 2790, Frame count: 111269, Running reward: -6.86
Episode: 2800, Frame count: 111526, Running reward: -6.76
Episode: 2810, Frame count: 111809, Running reward: -6.94
Episode: 2820, Frame count: 112084, Running reward: -7.14
Episode: 2830, Frame count: 112349, Running reward: -6.99
Episode: 2840, Frame count: 112612, Running reward: -6.97
Episode: 2850, Frame count: 112916, Running reward: -7.64
Episode: 2860, Frame count: 113184, Running reward: -7.38
Episode: 2870, Frame count: 113427, Running reward: -7.29
Episode: 2880, Frame count: 113674, Running reward: -7.08
Episode: 2890, Frame count: 113958, Running reward: -6.89
Episode: 2900, Frame count: 114227, Running reward: -7.01
```



DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 2910, Frame count: 114535, Running reward: -7.26
Episode: 2920, Frame count: 114805, Running reward: -7.21
Episode: 2930, Frame count: 115057, Running reward: -7.08
Episode: 2940, Frame count: 115321, Running reward: -7.09
Episode: 2950, Frame count: 115596, Running reward: -6.8
Episode: 2960, Frame count: 115846, Running reward: -6.62
Episode: 2970, Frame count: 116116, Running reward: -6.89
Episode: 2980, Frame count: 116392, Running reward: -7.18
Episode: 2990, Frame count: 116647, Running reward: -6.89
Episode: 3000, Frame count: 116916, Running reward: -6.89
Episode: 3010, Frame count: 117170, Running reward: -6.35
Episode: 3020, Frame count: 117415, Running reward: -6.1
Episode: 3030, Frame count: 117690, Running reward: -6.33
Episode: 3040, Frame count: 117941, Running reward: -6.2
Episode: 3050, Frame count: 118215, Running reward: -6.19
Episode: 3060, Frame count: 118468, Running reward: -6.22
Episode: 3070, Frame count: 118771, Running reward: -6.55
Episode: 3080, Frame count: 119054, Running reward: -6.62
Episode: 3090, Frame count: 119305, Running reward: -6.58
Episode: 3100, Frame count: 119575, Running reward: -6.59
Episode: 3110, Frame count: 119835, Running reward: -6.65
Episode: 3120, Frame count: 120073, Running reward: -6.58
Episode: 3130, Frame count: 120301, Running reward: -6.11
Episode: 3140, Frame count: 120562, Running reward: -6.21
Episode: 3150, Frame count: 120834, Running reward: -6.19
Episode: 3160, Frame count: 121090, Running reward: -6.22
Episode: 3170, Frame count: 121332, Running reward: -5.61
Episode: 3180, Frame count: 121593, Running reward: -5.39
Episode: 3190, Frame count: 121863, Running reward: -5.58
Episode: 3200, Frame count: 122093, Running reward: -5.18
Episode: 3210, Frame count: 122349, Running reward: -5.14
Episode: 3220, Frame count: 122598, Running reward: -5.25
Episode: 3230, Frame count: 122842, Running reward: -5.41
Episode: 3240, Frame count: 123103, Running reward: -5.41
Episode: 3250, Frame count: 123337, Running reward: -5.03
Episode: 3260, Frame count: 123574, Running reward: -4.84
Episode: 3270, Frame count: 123847, Running reward: -5.15
Episode: 3280, Frame count: 124083, Running reward: -4.9
Episode: 3290, Frame count: 124344, Running reward: -4.81
Episode: 3300, Frame count: 124566, Running reward: -4.73
Episode: 3310, Frame count: 124780, Running reward: -4.31
Episode: 3320, Frame count: 125026, Running reward: -4.28
Episode: 3330, Frame count: 125276, Running reward: -4.34
Episode: 3340, Frame count: 125519, Running reward: -4.16
Episode: 3350, Frame count: 125752, Running reward: -4.15
Episode: 3360, Frame count: 126007, Running reward: -4.33
Episode: 3370, Frame count: 126240, Running reward: -3.93
Episode: 3380, Frame count: 126477, Running reward: -3.94
Episode: 3390, Frame count: 126694, Running reward: -3.5
Episode: 3400, Frame count: 126931, Running reward: -3.65
Episode: 3410, Frame count: 127168, Running reward: -3.88
Episode: 3420, Frame count: 127443, Running reward: -4.17
Episode: 3430, Frame count: 127683, Running reward: -4.07
Episode: 3440, Frame count: 127890, Running reward: -3.71
Episode: 3450, Frame count: 128111, Running reward: -3.59
Episode: 3460, Frame count: 128361, Running reward: -3.54
Episode: 3470, Frame count: 128601, Running reward: -3.61
Episode: 3480, Frame count: 128852, Running reward: -3.75
Episode: 3490, Frame count: 129086, Running reward: -3.92
Episode: 3500, Frame count: 129319, Running reward: -3.88
Episode: 3510, Frame count: 129545, Running reward: -3.77
Episode: 3520, Frame count: 129788, Running reward: -3.45
Episode: 3530, Frame count: 130043, Running reward: -3.6
Episode: 3540, Frame count: 130270, Running reward: -3.8
Episode: 3550, Frame count: 130500, Running reward: -3.89
Episode: 3560, Frame count: 130712, Running reward: -3.51
Episode: 3570, Frame count: 130946, Running reward: -3.45
Episode: 3580, Frame count: 131203, Running reward: -3.51
Episode: 3590, Frame count: 131429, Running reward: -3.43
Episode: 3600, Frame count: 131656, Running reward: -3.37
Episode: 3610, Frame count: 131866, Running reward: -3.21
Episode: 3620, Frame count: 132125, Running reward: -3.37
Episode: 3630, Frame count: 132392, Running reward: -3.49
Episode: 3640, Frame count: 132598, Running reward: -3.28
Episode: 3650, Frame count: 132840, Running reward: -3.4
```



DQN으로 Shooting Airplane Game 강화 학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 3660, Frame count: 133082, Running reward: -3.7
Episode: 3670, Frame count: 133300, Running reward: -3.54
Episode: 3680, Frame count: 133564, Running reward: -3.61
Episode: 3690, Frame count: 133799, Running reward: -3.7
Episode: 3700, Frame count: 134025, Running reward: -3.69
Episode: 3710, Frame count: 134286, Running reward: -4.2
Episode: 3720, Frame count: 134508, Running reward: -3.83
Episode: 3730, Frame count: 134757, Running reward: -3.65
Episode: 3740, Frame count: 134974, Running reward: -3.76
Episode: 3750, Frame count: 135232, Running reward: -3.92
Episode: 3760, Frame count: 135487, Running reward: -4.05
Episode: 3770, Frame count: 135716, Running reward: -4.16
Episode: 3780, Frame count: 135968, Running reward: -4.04
Episode: 3790, Frame count: 136219, Running reward: -4.2
Episode: 3800, Frame count: 136443, Running reward: -4.18
Episode: 3810, Frame count: 136656, Running reward: -3.7
Episode: 3820, Frame count: 136921, Running reward: -4.13
Episode: 3830, Frame count: 137147, Running reward: -3.9
Episode: 3840, Frame count: 137387, Running reward: -4.13
Episode: 3850, Frame count: 137616, Running reward: -3.84
Episode: 3860, Frame count: 137858, Running reward: -3.71
Episode: 3870, Frame count: 138092, Running reward: -3.76
Episode: 3880, Frame count: 138321, Running reward: -3.53
Episode: 3890, Frame count: 138507, Running reward: -2.88
Episode: 3900, Frame count: 138735, Running reward: -2.92
Episode: 3910, Frame count: 138945, Running reward: -2.89
Episode: 3920, Frame count: 139197, Running reward: -2.76
Episode: 3930, Frame count: 139408, Running reward: -2.61
Episode: 3940, Frame count: 139646, Running reward: -2.59
Episode: 3950, Frame count: 139883, Running reward: -2.67
Episode: 3960, Frame count: 140115, Running reward: -2.57
Episode: 3970, Frame count: 140329, Running reward: -2.37
Episode: 3980, Frame count: 140549, Running reward: -2.28
Episode: 3990, Frame count: 140758, Running reward: -2.51
Episode: 4000, Frame count: 140954, Running reward: -2.19
Episode: 4010, Frame count: 141162, Running reward: -2.17
Episode: 4020, Frame count: 141374, Running reward: -1.77
Episode: 4030, Frame count: 141612, Running reward: -2.04
Episode: 4040, Frame count: 141801, Running reward: -1.55
Episode: 4050, Frame count: 142024, Running reward: -1.41
Episode: 4060, Frame count: 142241, Running reward: -1.26
Episode: 4070, Frame count: 142467, Running reward: -1.38
Episode: 4080, Frame count: 142657, Running reward: -1.08
Episode: 4090, Frame count: 142884, Running reward: -1.26
Episode: 4100, Frame count: 143109, Running reward: -1.55
Episode: 4110, Frame count: 143326, Running reward: -1.64
Episode: 4120, Frame count: 143551, Running reward: -1.77
Episode: 4130, Frame count: 143776, Running reward: -1.64
Episode: 4140, Frame count: 144014, Running reward: -2.13
Episode: 4150, Frame count: 144215, Running reward: -1.91
Episode: 4160, Frame count: 144442, Running reward: -2.01
Episode: 4170, Frame count: 144649, Running reward: -1.82
Episode: 4180, Frame count: 144874, Running reward: -2.17
Episode: 4190, Frame count: 145099, Running reward: -2.15
Episode: 4200, Frame count: 145300, Running reward: -1.91
Episode: 4210, Frame count: 145514, Running reward: -1.88
Episode: 4220, Frame count: 145733, Running reward: -1.82
Episode: 4230, Frame count: 145956, Running reward: -1.8
Episode: 4240, Frame count: 146202, Running reward: -1.88
Episode: 4250, Frame count: 146432, Running reward: -2.17
Episode: 4260, Frame count: 146649, Running reward: -2.07
Episode: 4270, Frame count: 146858, Running reward: -2.09
Episode: 4280, Frame count: 147076, Running reward: -2.02
Episode: 4290, Frame count: 147303, Running reward: -2.04
Episode: 4300, Frame count: 147515, Running reward: -2.15
Episode: 4310, Frame count: 147702, Running reward: -1.88
Episode: 4320, Frame count: 147925, Running reward: -1.92
Episode: 4330, Frame count: 148143, Running reward: -1.87
Episode: 4340, Frame count: 148344, Running reward: -1.42
Episode: 4350, Frame count: 148553, Running reward: -1.21
Episode: 4360, Frame count: 148758, Running reward: -1.09
Episode: 4370, Frame count: 148976, Running reward: -1.18
Episode: 4380, Frame count: 149185, Running reward: -1.09
Episode: 4390, Frame count: 149390, Running reward: -0.87
Episode: 4400, Frame count: 149597, Running reward: -0.82
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 4410, Frame count: 149809, Running reward: -1.07
Episode: 4420, Frame count: 150038, Running reward: -1.13
Episode: 4430, Frame count: 150245, Running reward: -1.02
Episode: 4440, Frame count: 150449, Running reward: -1.05
Episode: 4450, Frame count: 150662, Running reward: -1.09
Episode: 4460, Frame count: 150869, Running reward: -1.11
Episode: 4470, Frame count: 151104, Running reward: -1.28
Episode: 4480, Frame count: 151301, Running reward: -1.16
Episode: 4490, Frame count: 151493, Running reward: -1.03
Episode: 4500, Frame count: 151716, Running reward: -1.19
Episode: 4510, Frame count: 151934, Running reward: -1.25
Episode: 4520, Frame count: 152135, Running reward: -0.97
Episode: 4530, Frame count: 152345, Running reward: -1.0
Episode: 4540, Frame count: 152557, Running reward: -1.08
Episode: 4550, Frame count: 152772, Running reward: -1.1
Episode: 4560, Frame count: 153010, Running reward: -1.41
Episode: 4570, Frame count: 153214, Running reward: -1.1
Episode: 4580, Frame count: 153450, Running reward: -1.49
Episode: 4590, Frame count: 153636, Running reward: -1.43
Episode: 4600, Frame count: 153848, Running reward: -1.32
Episode: 4610, Frame count: 154085, Running reward: -1.51
Episode: 4620, Frame count: 154313, Running reward: -1.78
Episode: 4630, Frame count: 154505, Running reward: -1.6
Episode: 4640, Frame count: 154731, Running reward: -1.74
Episode: 4650, Frame count: 154933, Running reward: -1.61
Episode: 4660, Frame count: 155139, Running reward: -1.29
Episode: 4670, Frame count: 155363, Running reward: -1.49
Episode: 4680, Frame count: 155566, Running reward: -1.16
Episode: 4690, Frame count: 155739, Running reward: -1.03
Episode: 4700, Frame count: 155961, Running reward: -1.13
Episode: 4710, Frame count: 156175, Running reward: -0.9
Episode: 4720, Frame count: 156374, Running reward: -0.61
Episode: 4730, Frame count: 156575, Running reward: -0.7
Episode: 4740, Frame count: 156803, Running reward: -0.72
Episode: 4750, Frame count: 157017, Running reward: -0.84
Episode: 4760, Frame count: 157240, Running reward: -1.01
Episode: 4770, Frame count: 157442, Running reward: -0.79
Episode: 4780, Frame count: 157664, Running reward: -0.98
Episode: 4790, Frame count: 157873, Running reward: -1.34
Episode: 4800, Frame count: 158057, Running reward: -0.96
Episode: 4810, Frame count: 158235, Running reward: -0.6
Episode: 4820, Frame count: 158421, Running reward: -0.47
Episode: 4830, Frame count: 158617, Running reward: -0.42
Episode: 4840, Frame count: 158818, Running reward: -0.15
Episode: 4850, Frame count: 159018, Running reward: -0.01
Episode: 4860, Frame count: 159203, Running reward: 0.37
Episode: 4870, Frame count: 159412, Running reward: 0.3
Episode: 4880, Frame count: 159627, Running reward: 0.37
Episode: 4890, Frame count: 159814, Running reward: 0.59
Episode: 4900, Frame count: 160000, Running reward: 0.57
Episode: 4910, Frame count: 160199, Running reward: 0.36
Episode: 4920, Frame count: 160392, Running reward: 0.29
Episode: 4930, Frame count: 160564, Running reward: 0.53
Episode: 4940, Frame count: 160775, Running reward: 0.43
Episode: 4950, Frame count: 160982, Running reward: 0.36
Episode: 4960, Frame count: 161185, Running reward: 0.18
Episode: 4970, Frame count: 161387, Running reward: 0.25
Episode: 4980, Frame count: 161574, Running reward: 0.53
Episode: 4990, Frame count: 161751, Running reward: 0.63
Episode: 5000, Frame count: 161977, Running reward: 0.23
Episode: 5010, Frame count: 162174, Running reward: 0.25
Episode: 5020, Frame count: 162351, Running reward: 0.41
Episode: 5030, Frame count: 162557, Running reward: 0.07
Episode: 5040, Frame count: 162746, Running reward: 0.29
Episode: 5050, Frame count: 162944, Running reward: 0.38
Episode: 5060, Frame count: 163155, Running reward: 0.3
Episode: 5070, Frame count: 163355, Running reward: 0.32
Episode: 5080, Frame count: 163544, Running reward: 0.3
Episode: 5090, Frame count: 163740, Running reward: 0.11
Episode: 5100, Frame count: 163951, Running reward: 0.26
Episode: 5110, Frame count: 164147, Running reward: 0.27
Episode: 5120, Frame count: 164356, Running reward: -0.05
Episode: 5130, Frame count: 164544, Running reward: 0.13
Episode: 5140, Frame count: 164741, Running reward: 0.05
Episode: 5150, Frame count: 164918, Running reward: 0.26
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 5160, Frame count: 165130, Running reward: 0.25
Episode: 5170, Frame count: 165339, Running reward: 0.16
Episode: 5180, Frame count: 165530, Running reward: 0.14
Episode: 5190, Frame count: 165720, Running reward: 0.2
Episode: 5200, Frame count: 165924, Running reward: 0.27
Episode: 5210, Frame count: 166113, Running reward: 0.34
Episode: 5220, Frame count: 166295, Running reward: 0.61
Episode: 5230, Frame count: 166504, Running reward: 0.4
Episode: 5240, Frame count: 166690, Running reward: 0.51
Episode: 5250, Frame count: 166889, Running reward: 0.29
Episode: 5260, Frame count: 167070, Running reward: 0.6
Episode: 5270, Frame count: 167255, Running reward: 0.84
Episode: 5280, Frame count: 167430, Running reward: 1.0
Episode: 5290, Frame count: 167602, Running reward: 1.18
Episode: 5300, Frame count: 167788, Running reward: 1.36
Episode: 5310, Frame count: 167966, Running reward: 1.47
Episode: 5320, Frame count: 168186, Running reward: 1.09
Episode: 5330, Frame count: 168391, Running reward: 1.13
Episode: 5340, Frame count: 168583, Running reward: 1.07
Episode: 5350, Frame count: 168774, Running reward: 1.15
Episode: 5360, Frame count: 168966, Running reward: 1.04
Episode: 5370, Frame count: 169190, Running reward: 0.65
Episode: 5380, Frame count: 169390, Running reward: 0.4
Episode: 5390, Frame count: 169590, Running reward: 0.12
Episode: 5400, Frame count: 169797, Running reward: -0.09
Episode: 5410, Frame count: 169964, Running reward: 0.02
Episode: 5420, Frame count: 170144, Running reward: 0.42
Episode: 5430, Frame count: 170330, Running reward: 0.61
Episode: 5440, Frame count: 170526, Running reward: 0.57
Episode: 5450, Frame count: 170694, Running reward: 0.8
Episode: 5460, Frame count: 170869, Running reward: 0.97
Episode: 5470, Frame count: 171058, Running reward: 1.32
Episode: 5480, Frame count: 171238, Running reward: 1.52
Episode: 5490, Frame count: 171440, Running reward: 1.5
Episode: 5500, Frame count: 171630, Running reward: 1.67
Episode: 5510, Frame count: 171808, Running reward: 1.56
Episode: 5520, Frame count: 172023, Running reward: 1.21
Episode: 5530, Frame count: 172247, Running reward: 0.83
Episode: 5540, Frame count: 172418, Running reward: 1.08
Episode: 5550, Frame count: 172618, Running reward: 0.76
Episode: 5560, Frame count: 172803, Running reward: 0.66
Episode: 5570, Frame count: 172991, Running reward: 0.67
Episode: 5580, Frame count: 173183, Running reward: 0.55
Episode: 5590, Frame count: 173366, Running reward: 0.74
Episode: 5600, Frame count: 173571, Running reward: 0.59
Episode: 5610, Frame count: 173768, Running reward: 0.4
Episode: 5620, Frame count: 173962, Running reward: 0.61
Episode: 5630, Frame count: 174150, Running reward: 0.97
Episode: 5640, Frame count: 174335, Running reward: 0.83
Episode: 5650, Frame count: 174518, Running reward: 1.0
Episode: 5660, Frame count: 174698, Running reward: 1.05
Episode: 5670, Frame count: 174880, Running reward: 1.11
Episode: 5680, Frame count: 175089, Running reward: 0.94
Episode: 5690, Frame count: 175283, Running reward: 0.83
Episode: 5700, Frame count: 175452, Running reward: 1.19
Episode: 5710, Frame count: 175649, Running reward: 1.19
Episode: 5720, Frame count: 175831, Running reward: 1.31
Episode: 5730, Frame count: 176013, Running reward: 1.37
Episode: 5740, Frame count: 176201, Running reward: 1.34
Episode: 5750, Frame count: 176378, Running reward: 1.4
Episode: 5760, Frame count: 176557, Running reward: 1.41
Episode: 5770, Frame count: 176738, Running reward: 1.42
Episode: 5780, Frame count: 176909, Running reward: 1.8
Episode: 5790, Frame count: 177084, Running reward: 1.99
Episode: 5800, Frame count: 177270, Running reward: 1.82
Episode: 5810, Frame count: 177441, Running reward: 2.08
Episode: 5820, Frame count: 177628, Running reward: 2.03
Episode: 5830, Frame count: 177820, Running reward: 1.93
Episode: 5840, Frame count: 178002, Running reward: 1.99
Episode: 5850, Frame count: 178189, Running reward: 1.89
Episode: 5860, Frame count: 178356, Running reward: 2.01
Episode: 5870, Frame count: 178540, Running reward: 1.98
Episode: 5880, Frame count: 178734, Running reward: 1.75
Episode: 5890, Frame count: 178907, Running reward: 1.77
Episode: 5900, Frame count: 179089, Running reward: 1.81
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 5910, Frame count: 179270, Running reward: 1.71
Episode: 5920, Frame count: 179444, Running reward: 1.84
Episode: 5930, Frame count: 179626, Running reward: 1.94
Episode: 5940, Frame count: 179796, Running reward: 2.06
Episode: 5950, Frame count: 179978, Running reward: 2.11
Episode: 5960, Frame count: 180151, Running reward: 2.05
Episode: 5970, Frame count: 180329, Running reward: 2.11
Episode: 5980, Frame count: 180506, Running reward: 2.28
Episode: 5990, Frame count: 180683, Running reward: 2.24
Episode: 6000, Frame count: 180862, Running reward: 2.27
Episode: 6010, Frame count: 181044, Running reward: 2.26
Episode: 6020, Frame count: 181224, Running reward: 2.2
Episode: 6030, Frame count: 181411, Running reward: 2.15
Episode: 6040, Frame count: 181589, Running reward: 2.07
Episode: 6050, Frame count: 181757, Running reward: 2.21
Episode: 6060, Frame count: 181933, Running reward: 2.18
Episode: 6070, Frame count: 182099, Running reward: 2.3
Episode: 6080, Frame count: 182261, Running reward: 2.45
Episode: 6090, Frame count: 182423, Running reward: 2.6
Episode: 6100, Frame count: 182586, Running reward: 2.76
Episode: 6110, Frame count: 182759, Running reward: 2.85
Episode: 6120, Frame count: 182938, Running reward: 2.86
Episode: 6130, Frame count: 183127, Running reward: 2.84
Episode: 6140, Frame count: 183294, Running reward: 2.95
Episode: 6150, Frame count: 183466, Running reward: 2.91
Episode: 6160, Frame count: 183648, Running reward: 2.85
Episode: 6170, Frame count: 183819, Running reward: 2.8
Episode: 6180, Frame count: 183999, Running reward: 2.62
Episode: 6190, Frame count: 184178, Running reward: 2.45
Episode: 6200, Frame count: 184354, Running reward: 2.32
Episode: 6210, Frame count: 184520, Running reward: 2.39
Episode: 6220, Frame count: 184709, Running reward: 2.29
Episode: 6230, Frame count: 184875, Running reward: 2.52
Episode: 6240, Frame count: 185030, Running reward: 2.64
Episode: 6250, Frame count: 185218, Running reward: 2.48
Episode: 6260, Frame count: 185385, Running reward: 2.63
Episode: 6270, Frame count: 185558, Running reward: 2.61
Episode: 6280, Frame count: 185736, Running reward: 2.63
Episode: 6290, Frame count: 185906, Running reward: 2.72
Episode: 6300, Frame count: 186073, Running reward: 2.81
Episode: 6310, Frame count: 186266, Running reward: 2.54
Episode: 6320, Frame count: 186436, Running reward: 2.73
Episode: 6330, Frame count: 186596, Running reward: 2.79
Episode: 6340, Frame count: 186758, Running reward: 2.72
Episode: 6350, Frame count: 186937, Running reward: 2.81
Episode: 6360, Frame count: 187122, Running reward: 2.63
Episode: 6370, Frame count: 187309, Running reward: 2.49
Episode: 6380, Frame count: 187467, Running reward: 2.69
Episode: 6390, Frame count: 187611, Running reward: 2.95
Episode: 6400, Frame count: 187782, Running reward: 2.91
Episode: 6410, Frame count: 187964, Running reward: 3.02
Episode: 6420, Frame count: 188147, Running reward: 2.89
Episode: 6430, Frame count: 188326, Running reward: 2.7
Episode: 6440, Frame count: 188497, Running reward: 2.61
Episode: 6450, Frame count: 188693, Running reward: 2.44
Episode: 6460, Frame count: 188864, Running reward: 2.58
Episode: 6470, Frame count: 189024, Running reward: 2.85
Episode: 6480, Frame count: 189199, Running reward: 2.68
Episode: 6490, Frame count: 189380, Running reward: 2.31
Episode: 6500, Frame count: 189553, Running reward: 2.29
Episode: 6510, Frame count: 189731, Running reward: 2.33
Episode: 6520, Frame count: 189917, Running reward: 2.3
Episode: 6530, Frame count: 190089, Running reward: 2.37
Episode: 6540, Frame count: 190269, Running reward: 2.28
Episode: 6550, Frame count: 190428, Running reward: 2.65
Episode: 6560, Frame count: 190593, Running reward: 2.71
Episode: 6570, Frame count: 190787, Running reward: 2.37
Episode: 6580, Frame count: 190945, Running reward: 2.54
Episode: 6590, Frame count: 191097, Running reward: 2.83
Episode: 6600, Frame count: 191267, Running reward: 2.86
Episode: 6610, Frame count: 191458, Running reward: 2.73
Episode: 6620, Frame count: 191634, Running reward: 2.83
Episode: 6630, Frame count: 191806, Running reward: 2.83
Episode: 6640, Frame count: 191973, Running reward: 2.96
Episode: 6650, Frame count: 192141, Running reward: 2.87
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 6660, Frame count: 192323, Running reward: 2.7
Episode: 6670, Frame count: 192480, Running reward: 3.07
Episode: 6680, Frame count: 192640, Running reward: 3.05
Episode: 6690, Frame count: 192825, Running reward: 2.72
Episode: 6700, Frame count: 192987, Running reward: 2.8
Episode: 6710, Frame count: 193157, Running reward: 3.01
Episode: 6720, Frame count: 193314, Running reward: 3.2
Episode: 6730, Frame count: 193471, Running reward: 3.35
Episode: 6740, Frame count: 193642, Running reward: 3.31
Episode: 6750, Frame count: 193813, Running reward: 3.28
Episode: 6760, Frame count: 193988, Running reward: 3.35
Episode: 6770, Frame count: 194164, Running reward: 3.16
Episode: 6780, Frame count: 194338, Running reward: 3.02
Episode: 6790, Frame count: 194494, Running reward: 3.31
Episode: 6800, Frame count: 194650, Running reward: 3.37
Episode: 6810, Frame count: 194826, Running reward: 3.31
Episode: 6820, Frame count: 194996, Running reward: 3.18
Episode: 6830, Frame count: 195144, Running reward: 3.27
Episode: 6840, Frame count: 195305, Running reward: 3.37
Episode: 6850, Frame count: 195464, Running reward: 3.49
Episode: 6860, Frame count: 195624, Running reward: 3.64
Episode: 6870, Frame count: 195776, Running reward: 3.88
Episode: 6880, Frame count: 195931, Running reward: 4.07
Episode: 6890, Frame count: 196090, Running reward: 4.04
Episode: 6900, Frame count: 196256, Running reward: 3.94
Episode: 6910, Frame count: 196424, Running reward: 4.02
Episode: 6920, Frame count: 196572, Running reward: 4.24
Episode: 6930, Frame count: 196746, Running reward: 3.98
Episode: 6940, Frame count: 196909, Running reward: 3.96
Episode: 6950, Frame count: 197053, Running reward: 4.11
Episode: 6960, Frame count: 197217, Running reward: 4.07
Episode: 6970, Frame count: 197377, Running reward: 3.99
Episode: 6980, Frame count: 197559, Running reward: 3.72
Episode: 6990, Frame count: 197723, Running reward: 3.67
Episode: 7000, Frame count: 197881, Running reward: 3.75
Episode: 7010, Frame count: 198035, Running reward: 3.89
Episode: 7020, Frame count: 198190, Running reward: 3.82
Episode: 7030, Frame count: 198348, Running reward: 3.98
Episode: 7040, Frame count: 198494, Running reward: 4.15
Episode: 7050, Frame count: 198666, Running reward: 3.87
Episode: 7060, Frame count: 198819, Running reward: 3.98
Episode: 7070, Frame count: 198972, Running reward: 4.05
Episode: 7080, Frame count: 199132, Running reward: 4.27
Episode: 7090, Frame count: 199290, Running reward: 4.33
Episode: 7100, Frame count: 199458, Running reward: 4.23
Episode: 7110, Frame count: 199637, Running reward: 3.98
Episode: 7120, Frame count: 199801, Running reward: 3.89
Episode: 7130, Frame count: 199950, Running reward: 3.98
Episode: 7140, Frame count: 200124, Running reward: 3.7
Episode: 7150, Frame count: 200288, Running reward: 3.78
Episode: 7160, Frame count: 200445, Running reward: 3.74
Episode: 7170, Frame count: 200608, Running reward: 3.64
Episode: 7180, Frame count: 200782, Running reward: 3.5
Episode: 7190, Frame count: 200932, Running reward: 3.58
Episode: 7200, Frame count: 201091, Running reward: 3.67
Episode: 7210, Frame count: 201252, Running reward: 3.85
Episode: 7220, Frame count: 201412, Running reward: 3.89
Episode: 7230, Frame count: 201567, Running reward: 3.83
Episode: 7240, Frame count: 201719, Running reward: 4.05
Episode: 7250, Frame count: 201861, Running reward: 4.27
Episode: 7260, Frame count: 202027, Running reward: 4.18
Episode: 7270, Frame count: 202201, Running reward: 4.07
Episode: 7280, Frame count: 202354, Running reward: 4.28
Episode: 7290, Frame count: 202514, Running reward: 4.18
Episode: 7300, Frame count: 202679, Running reward: 4.12
Episode: 7310, Frame count: 202836, Running reward: 4.16
Episode: 7320, Frame count: 202993, Running reward: 4.19
Episode: 7330, Frame count: 203158, Running reward: 4.09
Episode: 7340, Frame count: 203318, Running reward: 4.01
Episode: 7350, Frame count: 203482, Running reward: 3.79
Episode: 7360, Frame count: 203638, Running reward: 3.89
Episode: 7370, Frame count: 203798, Running reward: 4.03
Episode: 7380, Frame count: 203947, Running reward: 4.07
Episode: 7390, Frame count: 204114, Running reward: 4.0
Episode: 7400, Frame count: 204271, Running reward: 4.08
```



DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 7410, Frame count: 204420, Running reward: 4.16
Episode: 7420, Frame count: 204578, Running reward: 4.15
Episode: 7430, Frame count: 204765, Running reward: 3.93
Episode: 7440, Frame count: 204923, Running reward: 3.95
Episode: 7450, Frame count: 205092, Running reward: 3.9
Episode: 7460, Frame count: 205244, Running reward: 3.94
Episode: 7470, Frame count: 205404, Running reward: 3.94
Episode: 7480, Frame count: 205561, Running reward: 3.86
Episode: 7490, Frame count: 205712, Running reward: 4.02
Episode: 7500, Frame count: 205872, Running reward: 3.99
Episode: 7510, Frame count: 206044, Running reward: 3.76
Episode: 7520, Frame count: 206201, Running reward: 3.77
Episode: 7530, Frame count: 206358, Running reward: 4.07
Episode: 7540, Frame count: 206518, Running reward: 4.05
Episode: 7550, Frame count: 206676, Running reward: 4.16
Episode: 7560, Frame count: 206847, Running reward: 3.97
Episode: 7570, Frame count: 207004, Running reward: 4.0
Episode: 7580, Frame count: 207153, Running reward: 4.08
Episode: 7590, Frame count: 207314, Running reward: 3.98
Episode: 7600, Frame count: 207471, Running reward: 4.01
Episode: 7610, Frame count: 207623, Running reward: 4.21
Episode: 7620, Frame count: 207789, Running reward: 4.12
Episode: 7630, Frame count: 207952, Running reward: 4.06
Episode: 7640, Frame count: 208107, Running reward: 4.11
Episode: 7650, Frame count: 208263, Running reward: 4.13
Episode: 7660, Frame count: 208415, Running reward: 4.32
Episode: 7670, Frame count: 208575, Running reward: 4.29
Episode: 7680, Frame count: 208736, Running reward: 4.17
Episode: 7690, Frame count: 208892, Running reward: 4.22
Episode: 7700, Frame count: 209041, Running reward: 4.3
Episode: 7710, Frame count: 209198, Running reward: 4.25
Episode: 7720, Frame count: 209362, Running reward: 4.27
Episode: 7730, Frame count: 209520, Running reward: 4.32
Episode: 7740, Frame count: 209669, Running reward: 4.38
Episode: 7750, Frame count: 209826, Running reward: 4.37
Episode: 7760, Frame count: 209978, Running reward: 4.37
Episode: 7770, Frame count: 210134, Running reward: 4.41
Episode: 7780, Frame count: 210291, Running reward: 4.45
Episode: 7790, Frame count: 210447, Running reward: 4.45
Episode: 7800, Frame count: 210616, Running reward: 4.25
Episode: 7810, Frame count: 210779, Running reward: 4.19
Episode: 7820, Frame count: 210933, Running reward: 4.29
Episode: 7830, Frame count: 211098, Running reward: 4.22
Episode: 7840, Frame count: 211250, Running reward: 4.19
Episode: 7850, Frame count: 211417, Running reward: 4.09
Episode: 7860, Frame count: 211574, Running reward: 4.04
Episode: 7870, Frame count: 211739, Running reward: 3.95
Episode: 7880, Frame count: 211908, Running reward: 3.83
Episode: 7890, Frame count: 212066, Running reward: 3.81
Episode: 7900, Frame count: 212237, Running reward: 3.79
Episode: 7910, Frame count: 212389, Running reward: 3.9
Episode: 7920, Frame count: 212543, Running reward: 3.9
Episode: 7930, Frame count: 212706, Running reward: 3.92
Episode: 7940, Frame count: 212872, Running reward: 3.78
Episode: 7950, Frame count: 213029, Running reward: 3.88
Episode: 7960, Frame count: 213190, Running reward: 3.84
Episode: 7970, Frame count: 213354, Running reward: 3.85
Episode: 7980, Frame count: 213524, Running reward: 3.84
Episode: 7990, Frame count: 213695, Running reward: 3.71
Episode: 8000, Frame count: 213849, Running reward: 3.88
Episode: 8010, Frame count: 214009, Running reward: 3.8
Episode: 8020, Frame count: 214170, Running reward: 3.73
Episode: 8030, Frame count: 214347, Running reward: 3.59
Episode: 8040, Frame count: 214508, Running reward: 3.64
Episode: 8050, Frame count: 214662, Running reward: 3.67
Episode: 8060, Frame count: 214820, Running reward: 3.7
Episode: 8070, Frame count: 214986, Running reward: 3.68
Episode: 8080, Frame count: 215143, Running reward: 3.81
Episode: 8090, Frame count: 215317, Running reward: 3.78
Episode: 8100, Frame count: 215462, Running reward: 3.87
Episode: 8110, Frame count: 215630, Running reward: 3.79
Episode: 8120, Frame count: 215797, Running reward: 3.73
Episode: 8130, Frame count: 215971, Running reward: 3.76
Episode: 8140, Frame count: 216123, Running reward: 3.85
Episode: 8150, Frame count: 216276, Running reward: 3.86
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 8160, Frame count: 216430, Running reward: 3.9
Episode: 8170, Frame count: 216595, Running reward: 3.91
Episode: 8180, Frame count: 216761, Running reward: 3.82
Episode: 8190, Frame count: 216935, Running reward: 3.82
Episode: 8200, Frame count: 217082, Running reward: 3.8
Episode: 8210, Frame count: 217233, Running reward: 3.97
Episode: 8220, Frame count: 217408, Running reward: 3.89
Episode: 8230, Frame count: 217557, Running reward: 4.14
Episode: 8240, Frame count: 217708, Running reward: 4.15
Episode: 8250, Frame count: 217853, Running reward: 4.23
Episode: 8260, Frame count: 218023, Running reward: 4.07
Episode: 8270, Frame count: 218199, Running reward: 3.96
Episode: 8280, Frame count: 218355, Running reward: 4.06
Episode: 8290, Frame count: 218515, Running reward: 4.2
Episode: 8300, Frame count: 218667, Running reward: 4.15
Episode: 8310, Frame count: 218826, Running reward: 4.07
Episode: 8320, Frame count: 218970, Running reward: 4.38
Episode: 8330, Frame count: 219136, Running reward: 4.21
Episode: 8340, Frame count: 219289, Running reward: 4.19
Episode: 8350, Frame count: 219451, Running reward: 4.02
Episode: 8360, Frame count: 219598, Running reward: 4.25
Episode: 8370, Frame count: 219766, Running reward: 4.33
Episode: 8380, Frame count: 219917, Running reward: 4.38
Episode: 8390, Frame count: 220074, Running reward: 4.41
Episode: 8400, Frame count: 220236, Running reward: 4.31
Episode: 8410, Frame count: 220400, Running reward: 4.26
Episode: 8420, Frame count: 220549, Running reward: 4.21
Episode: 8430, Frame count: 220692, Running reward: 4.44
Episode: 8440, Frame count: 220849, Running reward: 4.4
Episode: 8450, Frame count: 221002, Running reward: 4.49
Episode: 8460, Frame count: 221184, Running reward: 4.14
Episode: 8470, Frame count: 221340, Running reward: 4.26
Episode: 8480, Frame count: 221495, Running reward: 4.22
Episode: 8490, Frame count: 221638, Running reward: 4.36
Episode: 8500, Frame count: 221798, Running reward: 4.38
Episode: 8510, Frame count: 221932, Running reward: 4.68
Episode: 8520, Frame count: 222090, Running reward: 4.59
Episode: 8530, Frame count: 222261, Running reward: 4.31
Episode: 8540, Frame count: 222420, Running reward: 4.29
Episode: 8550, Frame count: 222585, Running reward: 4.17
Episode: 8560, Frame count: 222737, Running reward: 4.47
Episode: 8570, Frame count: 222887, Running reward: 4.53
Episode: 8580, Frame count: 223044, Running reward: 4.51
Episode: 8590, Frame count: 223212, Running reward: 4.26
Episode: 8600, Frame count: 223383, Running reward: 4.15
Episode: 8610, Frame count: 223535, Running reward: 3.97
Episode: 8620, Frame count: 223707, Running reward: 3.83
Episode: 8630, Frame count: 223847, Running reward: 4.14
Episode: 8640, Frame count: 223986, Running reward: 4.34
Episode: 8650, Frame count: 224140, Running reward: 4.45
Episode: 8660, Frame count: 224290, Running reward: 4.47
Episode: 8670, Frame count: 224440, Running reward: 4.47
Episode: 8680, Frame count: 224602, Running reward: 4.42
Episode: 8690, Frame count: 224768, Running reward: 4.44
Episode: 8700, Frame count: 224924, Running reward: 4.59
Episode: 8710, Frame count: 225085, Running reward: 4.5
Episode: 8720, Frame count: 225239, Running reward: 4.68
Episode: 8730, Frame count: 225392, Running reward: 4.55
Episode: 8740, Frame count: 225540, Running reward: 4.46
Episode: 8750, Frame count: 225715, Running reward: 4.25
Episode: 8760, Frame count: 225863, Running reward: 4.27
Episode: 8770, Frame count: 226016, Running reward: 4.24
Episode: 8780, Frame count: 226165, Running reward: 4.37
Episode: 8790, Frame count: 226324, Running reward: 4.44
Episode: 8800, Frame count: 226479, Running reward: 4.45
Episode: 8810, Frame count: 226626, Running reward: 4.59
Episode: 8820, Frame count: 226776, Running reward: 4.63
Episode: 8830, Frame count: 226939, Running reward: 4.53
Episode: 8840, Frame count: 227097, Running reward: 4.43
Episode: 8850, Frame count: 227259, Running reward: 4.56
Episode: 8860, Frame count: 227424, Running reward: 4.39
Episode: 8870, Frame count: 227579, Running reward: 4.37
Episode: 8880, Frame count: 227731, Running reward: 4.34
Episode: 8890, Frame count: 227887, Running reward: 4.37
Episode: 8900, Frame count: 228051, Running reward: 4.28
```

DQN으로 Shooting Airplane Game 강  
화학습

게임 환경 설정

네트워크 정의하기

모델 빌딩 & 로스 및 최적화 계산  
기 만들기

Replay Buffer 정의

전처리

중간 테스트

Epsilon-greedy 액션 선택 함수

Greedy 액션 선택 함수

Update 파트

Test A Single Training Step

Run DQN Training

Evaluation

```
Episode: 8910, Frame count: 228212, Running reward: 4.14
Episode: 8920, Frame count: 228377, Running reward: 3.99
Episode: 8930, Frame count: 228528, Running reward: 4.11
Episode: 8940, Frame count: 228700, Running reward: 3.97
Episode: 8950, Frame count: 228855, Running reward: 4.04
Episode: 8960, Frame count: 229007, Running reward: 4.17
Episode: 8970, Frame count: 229155, Running reward: 4.24
Episode: 8980, Frame count: 229310, Running reward: 4.21
Episode: 8990, Frame count: 229448, Running reward: 4.39
Episode: 9000, Frame count: 229613, Running reward: 4.38
Episode: 9010, Frame count: 229773, Running reward: 4.39
Episode: 9020, Frame count: 229924, Running reward: 4.53
Episode: 9030, Frame count: 230075, Running reward: 4.53
Episode: 9040, Frame count: 230255, Running reward: 4.45
Episode: 9050, Frame count: 230411, Running reward: 4.44
Episode: 9060, Frame count: 230584, Running reward: 4.23
Episode: 9070, Frame count: 230737, Running reward: 4.18
Episode: 9080, Frame count: 230883, Running reward: 4.27
Episode: 9090, Frame count: 231041, Running reward: 4.07
Episode: 9100, Frame count: 231205, Running reward: 4.08
Episode: 9110, Frame count: 231372, Running reward: 4.01
Episode: 9120, Frame count: 231537, Running reward: 3.87
Episode: 9130, Frame count: 231695, Running reward: 3.8
Episode: 9140, Frame count: 231848, Running reward: 4.07
Episode: 9150, Frame count: 232000, Running reward: 4.11
Episode: 9160, Frame count: 232168, Running reward: 4.16
Episode: 9170, Frame count: 232332, Running reward: 4.05
Episode: 9180, Frame count: 232488, Running reward: 3.95
Episode: 9190, Frame count: 232644, Running reward: 3.97
Episode: 9200, Frame count: 232800, Running reward: 4.05
Episode: 9210, Frame count: 232952, Running reward: 4.2
Episode: 9220, Frame count: 233116, Running reward: 4.21
Episode: 9230, Frame count: 233274, Running reward: 4.21
Episode: 9240, Frame count: 233441, Running reward: 4.07
Episode: 9250, Frame count: 233600, Running reward: 4.0
Episode: 9260, Frame count: 233758, Running reward: 4.1
Episode: 9270, Frame count: 233914, Running reward: 4.18
Episode: 9280, Frame count: 234079, Running reward: 4.09
Episode: 9290, Frame count: 234226, Running reward: 4.18
Episode: 9300, Frame count: 234390, Running reward: 4.1
Episode: 9310, Frame count: 234558, Running reward: 3.94
Episode: 9320, Frame count: 234702, Running reward: 4.14
Episode: 9330, Frame count: 234872, Running reward: 4.02
Episode: 9340, Frame count: 235039, Running reward: 4.02
Episode: 9350, Frame count: 235214, Running reward: 3.86
Episode: 9360, Frame count: 235362, Running reward: 3.96
Episode: 9370, Frame count: 235509, Running reward: 4.05
Episode: 9380, Frame count: 235667, Running reward: 4.12
Episode: 9390, Frame count: 235810, Running reward: 4.16
Episode: 9400, Frame count: 235965, Running reward: 4.25
Episode: 9410, Frame count: 236126, Running reward: 4.32
Episode: 9420, Frame count: 236278, Running reward: 4.24
Episode: 9430, Frame count: 236424, Running reward: 4.48
Episode: 9440, Frame count: 236582, Running reward: 4.57
Episode: 9450, Frame count: 236723, Running reward: 4.91
Episode: 9460, Frame count: 236880, Running reward: 4.82
Episode: 9470, Frame count: 237028, Running reward: 4.81
Episode: 9480, Frame count: 237179, Running reward: 4.88
Episode: 9490, Frame count: 237343, Running reward: 4.67
Episode: 9500, Frame count: 237509, Running reward: 4.56
Episode: 9510, Frame count: 237667, Running reward: 4.59
Episode: 9520, Frame count: 237816, Running reward: 4.62
Episode: 9530, Frame count: 237976, Running reward: 4.48
Episode: 9540, Frame count: 238115, Running reward: 4.67
Episode: 9550, Frame count: 238269, Running reward: 4.54
Episode: 9560, Frame count: 238418, Running reward: 4.62
Episode: 9570, Frame count: 238583, Running reward: 4.45
Episode: 9580, Frame count: 238760, Running reward: 4.19
Episode: 9590, Frame count: 238899, Running reward: 4.44
Episode: 9600, Frame count: 239051, Running reward: 4.58
Episode: 9610, Frame count: 239205, Running reward: 4.62
Episode: 9620, Frame count: 239363, Running reward: 4.53
Episode: 9630, Frame count: 239524, Running reward: 4.52
Episode: 9640, Frame count: 239687, Running reward: 4.28
Episode: 9650, Frame count: 239850, Running reward: 4.19
```



## Evaluation

|   |
|---|
| Episode: 9660, Frame count: 240003, Running reward: 4.15  |
| Episode: 9670, Frame count: 240152, Running reward: 4.31  |
| Episode: 9680, Frame count: 240316, Running reward: 4.44  |
| Episode: 9690, Frame count: 240480, Running reward: 4.19  |
| Episode: 9700, Frame count: 240634, Running reward: 4.17  |
| Episode: 9710, Frame count: 240812, Running reward: 3.93  |
| Episode: 9720, Frame count: 240967, Running reward: 3.96  |
| Episode: 9730, Frame count: 241130, Running reward: 3.94  |
| Episode: 9740, Frame count: 241297, Running reward: 3.9   |
| Episode: 9750, Frame count: 241463, Running reward: 3.87  |
| Episode: 9760, Frame count: 241622, Running reward: 3.81  |
| Episode: 9770, Frame count: 241781, Running reward: 3.71  |
| Episode: 9780, Frame count: 241930, Running reward: 3.86  |
| Episode: 9790, Frame count: 242079, Running reward: 4.01  |
| Episode: 9800, Frame count: 242231, Running reward: 4.03  |
| Episode: 9810, Frame count: 242384, Running reward: 4.28  |
| Episode: 9820, Frame count: 242538, Running reward: 4.29  |
| Episode: 9830, Frame count: 242692, Running reward: 4.38  |
| Episode: 9840, Frame count: 242845, Running reward: 4.52  |
| Episode: 9850, Frame count: 243001, Running reward: 4.62  |
| Episode: 9860, Frame count: 243155, Running reward: 4.67  |
| Episode: 9870, Frame count: 243314, Running reward: 4.67  |
| Episode: 9880, Frame count: 243481, Running reward: 4.49  |
| Episode: 9890, Frame count: 243630, Running reward: 4.49  |
| Episode: 9900, Frame count: 243784, Running reward: 4.47  |
| Episode: 9910, Frame count: 243944, Running reward: 4.4   |
| Episode: 9920, Frame count: 244091, Running reward: 4.47  |
| Episode: 9930, Frame count: 244256, Running reward: 4.36  |
| Episode: 9940, Frame count: 244421, Running reward: 4.24  |
| Episode: 9950, Frame count: 244573, Running reward: 4.28  |
| Episode: 9960, Frame count: 244735, Running reward: 4.2   |
| Episode: 9970, Frame count: 244896, Running reward: 4.18  |
| Episode: 9980, Frame count: 245052, Running reward: 4.29  |
| Episode: 9990, Frame count: 245209, Running reward: 4.21  |
| Episode: 10000, Frame count: 245357, Running reward: 4.27 |
| Episode: 10010, Frame count: 245523, Running reward: 4.21 |

```
import time, sys
from IPython.display import clear_output

board, info = env.reset()
state = preprocess_state(board)
action_mask = info['action_mask'].reshape((-1,))
done = False
env.render()

while not done:
    action = get_greedy_action(model, state, action_mask)
    print("action: ({}, {})".format(action // 8, action % 8))
    sys.stdout.flush()

    time.sleep(1.0)
    clear_output(wait=False)
    board, reward, done, _, info = env.step((action // 8, action % 8))
    state = preprocess_state(board)
    action_mask = info['action_mask'].reshape((-1,))
    env.render()
```

|       |  |       |
|-------|--|-------|
|       |  |       |
|       |  |       |
| M     |  |       |
| M H   |  | H     |
| HHHHH |  | HHHHH |
| MH    |  | H     |
| HHH   |  | HHH   |