Describe the Deep Q-Network

主要在完成 act, learn 及 process 三個部分，架構分別說明如下：

Act

分為 training 跟 evaluation，在 training 過程按照 epsilon-greedy 策略，部分時間採隨機動作，其他的時間則按照 q_value.argmax(dim=1).item() 選擇 action。在 evaluation 時則都以 q_value.argmax(dim=1).item() 作為選擇

Learn

learn 的部分主要部分進行學習，進行以下動作：

1. 透過 q_value 跟取得 td_target，並且透過 mse_loss function 計算 loss
2. 透過 self.optimizer.zero_grad 清除梯度將梯度歸零（不清除可能會導致 backpropagation 的時候梯度不斷累積）
3. 透過 loss.backward() 進行 backpropagation，並且進行梯度修剪以避免梯度爆炸的可能: torch.nn.utils.clip_grad_norm_(self.network.parameters(), 1.0)
4. 更新 self.optimizer.step()
5. 回傳 loss 資訊以進行 logging

Process

1. 將 transition 的資訊更新至 buffer
2. 當 total steps 超過 warmup_steps 時，進行 learn（warmup_steps 的用意在於避免過早的 overfitting，幫助學習效能）
3. 每隔一定的 step 更新 network

Describe the architecture of your PacmanActionCNN

CNN architecture

```
self.conv1 = nn.Conv2d(state_dim, 64, kernel_size=8, stride=4)
self.bn1 = nn.BatchNorm2d(64)
self.conv2 = nn.Conv2d(64, 128, kernel_size=4, stride=2)
self.bn2 = nn.BatchNorm2d(128)
self.conv3 = nn.Conv2d(128, 128, kernel_size=3, stride=1)
self.bn3 = nn.BatchNorm2d(128)
self.fc1 = nn.Linear(128 * 7 * 7, 1024)
self.fc2 = nn.Linear(1024, action_dim)
```

設計三層的 convolution layer 用來從圖像中提取特徵，第一層輸入為預設的 state_dim，並且在每一層的 conv 後加入一層 BatchNorm2d 進行 normalization，嘗試提升訓練的穩定性跟效能。在第三層 後將 conv 攤平成 full connected layer，輸出 1024 個 hidden features。最後再用一個 full connected layer 作為 action_dim (Q) 的輸出結果

Forward

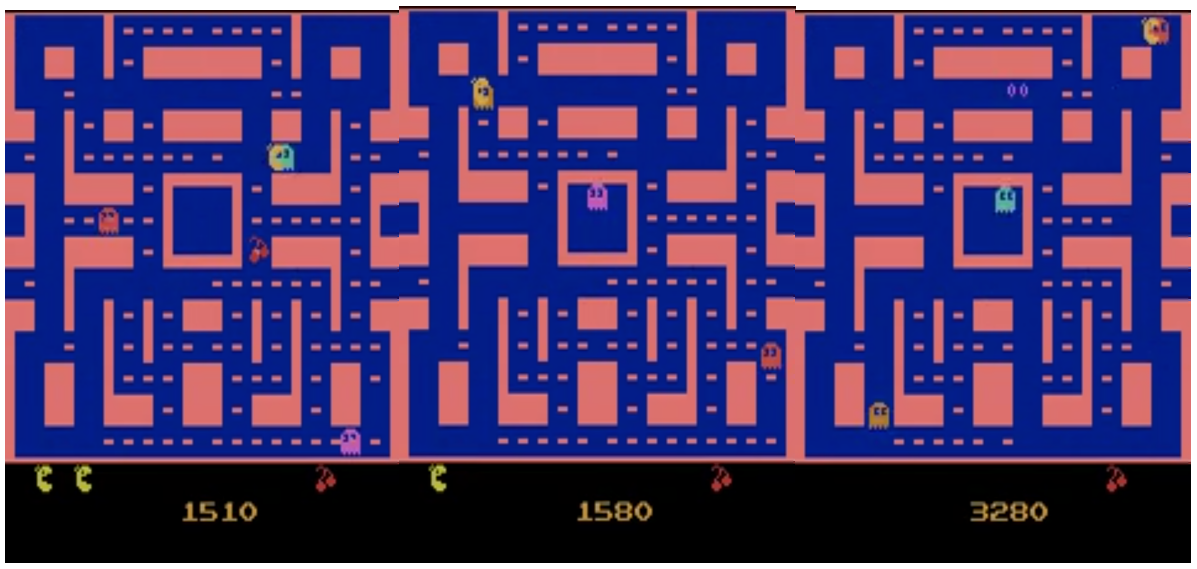
```
    def forward(self, x):
        # x = F.relu(self.conv1(x))
        "*** YOUR CODE HERE ***"
        x = F.relu(self.conv1(x))
        x = F.relu(self.conv2(x))
        x = F.relu(self.conv3(x))
        x = torch.flatten(x, start_dim=1)
        x = F.relu(self.fc1(x))
        x = self.fc2(x)
        # utils.raiseNotDefined()

        return x
```

1. **卷積層**：
    - 每個卷積層都使用 ReLU 激活函數。
    - 卷積後的輸出會通過批量歸一化。
2. **展平**：
    - 將卷積層的三維輸出展平成二維張量，作為全連接層的輸入。
3. **全連接層**：
    - 使用 ReLU 激活函數處理隱藏單元。
    - 最後通過 fc2 輸出每個動作的 Q 值。

Plot your training curve, including both loss and rewards.

Show screenshots from your evaluation video



最終 Evaluation 的分數為 3280