Problem Statement:

Traffic congestion is a significant issue in urban areas, leading to wasted time, fuel, and environmental pollution. Conventional traffic signal timing strategies often rely on fixed schedules or pre-defined algorithms. These strategies may not adapt well to changing traffic patterns, leading to inefficiencies and congestion. This project aims to develop a Deep Q-Learning (DQL) agent for traffic signal control, with the goal of maximizing traffic efficiency at intersections. By leveraging reinforcement learning, the project seeks to alleviate congestion, decrease travel time, and enhance overall traffic flow.

Approach:

The project involves controlling traffic signals at a single intersection where each phase lasts 10 seconds, including a 4-second yellow phase before a phase change.

The intersection has four arms with four lanes each, spanning 750 meters. Lanes are dedicated to specific turn directions: left, right, or straight. Traffic lights are placed at the left-most lane and shared by the other three lanes.

Every episode spawn 1000 cars with arrival times following a Weibull distribution. Most cars (75%) go straight, while the rest turn left or right. Car arrivals are randomly generated for each episode.

Choice of the traffic light phase from 4 possible predetermined phases, described below. Every phase has a duration.

North lane vehicles are given green signal and at that time vehicles in south are dedicated to go right or to go straight.

North-South Left Advance: green signal is given for north lanes and vehicles in south lane are dedicated to go left.

East-West Advance: east lane vehicles are given green signal and at that time vehicles in the west are dedicated to go right or to go straight.

East-West Left Advance: green signal is given for east lane vehicles and west lane vehicles are dedicated to go left.

The agent receives a reward based on the change in cumulative waiting time of cars at the intersection. Waiting time is measured as the duration a car spends with zero speed since its spawn. The cumulative waiting time includes all cars in incoming lanes, with the waiting time of cars no longer counted once they cross the intersection. This rewards the agent for reducing overall waiting time and improving traffic flow.

Learning-Mechanism: The agent employs Q-learning to update action values, using the equation $Q(s,a) = reward + gamma \cdot \max Q'(s',a')$. A deep neural network learns the state-action function, with 80 input neurons representing the state, 5 hidden layers of 400 neurons each, and 4 output neurons for possible actions. Additionally, an experience replays mechanism stores agent experiences in memory.

At the end of each episode, batches of randomized samples are extracted from memory to train the neural network, after updating action values with the Q-learning equation. Reinforcement Learning (RL) enables traffic signals to adapt dynamically to changing traffic conditions and learn optimal control policies. The goal is to improve traffic flow, reduce congestion, and enhance transportation efficiency by using RL techniques. DQN is a deep learning extension of Q-learning, leveraging neural networks to approximate the Q-function. The Q-function estimates the expected cumulative rewards for each action taken in a given state.

By using DQN, the agent can efficiently learn and update its policy for traffic signal control based on the observed rewards. The first step involves defining how the environment is represented to the agent, including current traffic flow and queue lengths at intersections. The next step is to specify the possible actions the agent can take, such as changing signal timings or adjusting phase durations. This allows the agent to understand the state of the traffic system and choose appropriate actions to improve traffic flow.

A deep neural network (the Q-network) is designed to approximate the Q-function. The Q-values generated by the network represent the expected cumulative rewards for each possible action in the current state. The Q-network is trained using a variant of Q-learning, where it learns to minimize the discrepancy between predicted Q-values and observed rewards. During training, the agent interacts with the environment, collecting experiences in the form of state-action-reward-next state tuples. After training, the performance of the trained DQN agent is evaluated on unseen traffic scenarios or simulation environments like SUMO.

The Q-network gradually improves as the agent learns from its experiences, leading to better policy decisions aimed at maximizing cumulative rewards over time. The reward function focuses on minimizing the total delay experienced by vehicles at the intersection. By minimizing delay, the traffic signal control system aims to reduce congestion and improve overall travel time for vehicles passing through the intersection. The agent aims to maximize the number of vehicles passing through the intersection within a given time frame.

Maximizing throughput helps ensure efficient utilization of road capacity and reduces the likelihood of traffic queues and congestion. The reward function can also prioritize reducing vehicle emissions, contributing to environmental sustainability.