**Motivation**

Global context is essential in complex scenarios, e.g., relation between traffic lights and vehicles.

Fusion-based methods can capture geometric and semantic information of the 3D scene using multiple sensors (e.g., camera and LiDAR).

**Our Method**

Our key idea is to use attention-based feature fusion to incorporate global context of the 3D scene.

\[
\mathcal{L} = \sum_{t=1}^{T} | \mathbf{w}_t - \mathbf{w}_t^{\theta_t} |_1
\]

**Driving Results**

Generalization to New Town (GF: Geometric Fusion, TF: TransFuser)

GF: Stop at intersection, Stopping at signal, Collision with vehicle

TF: Stopping for vehicle, Stopping for signal, Success

Generalization to New Weather

GF: Collision with vehicle

TF: Stopping at intersection, Stopping at intersection, Success

Infraction Analysis

TransFuser focuses on vehicles and traffic lights at intersections and can safely navigate difficult scenarios.