

This paper considers the challenging problem of predicting relative future locations and scales of nearby vehicles with respect to an ego-vehicle equipped with an egocentric camera.

By a multi-stream RNN encoder-decoder(RNN-ED) architecture to effectively encode past observations from different domains and generate future bounding box.

**Purpose:** Predicting the future location of vehicles. Introducing a novel approach to simultaneously predict both the location and scale of target vehicles in the first-person(egocentric) view of an ego-vehicle.

**Method:** Proposing **a multi-stream recurrent neural network (RNN) encoder-decoder model** (that separately capture both object location and scale and pixel-level observation for future vehicle localization)

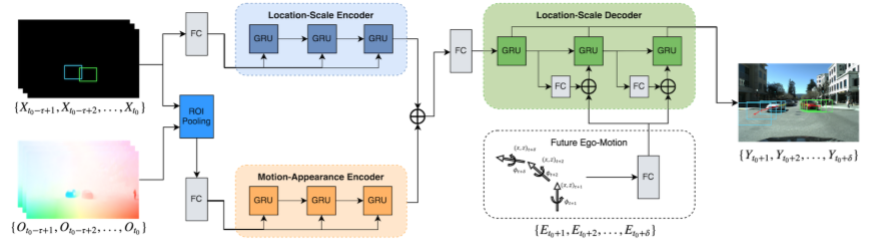
**Problem:** Extensive research has been conducted on predicting vehicles’ future action and trajectories using overhead(bird’s eye view) observations. But obtaining overhead views requires additional equipment like an externally-mounted camera(or LiDAR).

**Suggestion:** natural approach is to use forward-facing cameras that record the driver’s egocentric perspective

1. 지능적 주행시스템이 자기중심의 관점에서 교차로와 같은 까다로운 주행 시나리오에서 차량의 future location을 예측할 수 있는 새로운 관점을 제시함

2. 조밀한 광학 흐름(dense optical flow)과 future ego-motion을 입력으로 사용해 시간 모델링을 개선하고 차량의 움직임과 외관 정보를 명시적으로 캡처할 수 있는 multi-stream RNN-ED 아키텍처 제안함.

3. 도로 교차로와 관련된 다양한 시나리오에서 수집된 새로운 1인칭 video dataset(HONDA Egocentric View-Intersection(HEV-I)dataset)를 게시함 (dataset에는 230개 비디오에 2400개 이상의 차량(after filtering)이 포함)



차량의 특징들은 optical flow map으로부터 이중선형보간을(bilinear interpolation) 사용하여 관심영역 풀링동작(ROIPooling)에 의해 추출됨

The resulting relative motion vector is represented as Ot = [u1,v1,u2,v2,...un,vn]t, where n is the size of the pooled region.

hX t = GRUX(φX(Xt−1),hX t−1;θX)

hO t = GRUO(φO(Ot−1),hO t−1;θO)

H = φH(Average(hX t0,hO t0))

(GRU는 매개 변수 θ - gated recurrent unit

φ (·)는 ReLU 활성화를 사용한 선형 투영

hx t 및 ho t는 시간 t에서 GRU 모델의 숨겨진 상태 벡터)