

실시간 온보드 End-to-end 자율주행을 위한 저해상도-고해상도 영상 특징 융합 모듈

Low and High-resolution Feature Fusion Module for Real-time Onboard End-to-end Autonomous Driving

연 주 연¹, 이 전 혁², 박 태 형^{1,*}

(Joo-Yeon Yeon¹, Jeon-Hyeok Lee², and Tae-Hyoung Park^{1,*})

¹Department of Intelligent Systems and Robotics, Chungbuk National University

²Department of Control and Robot Engineering, Chungbuk National University

Abstract: Camera-based E2E (End-to-end) autonomous driving models map camera images directly to control commands. In onboard environments, real-time inference and driving performance are required. However, their performance depends on the resolution of the input image. High-resolution images preserve visual features of objects such as traffic lights and leading vehicles but increase computation and latency, while low-resolution images reduce latency but lose important visual information. This tradeoff between inference time and driving performance is an issue in onboard environments. This paper proposes a low- and high-resolution feature fusion module. The model extracts global features from low-resolution images and selectively fuses features from high-resolution regions of interest for traffic lights and leading vehicles. The proposed method was evaluated in the MORAI simulator and on the NVIDIA Orin AGX platform. The Driving Score, which evaluates driving performance, was 22.61% higher and the inference time was 35.91% faster than those of the model with high-resolution images. The proposed method alleviates the limitation of camera-based E2E caused by the tradeoff between inference time and driving performance in onboard environments.

Keywords: end-to-end, autonomous driving, onboard, image resolution, region of interest

I. 서론

자율주행 시스템은 센서 데이터로부터 인지, 판단, 제어 명령을 수행하는 시스템이다. 자율주행 시스템은 주변 환경 인식 알고리즘을 통해, 교통법규를 고려한 안전한 주행 판단을 수행한다. 기존 자율주행 시스템[1-3]은 인지, 판단 및 제어가 각 모듈별로 구성되어 별도의 입력과 출력을 처리한다. E2E (End-to-end) [4-7] 자율주행 네트워크는 이 과정을 통합하여, 데이터로부터 주행 판단에 최적화된 특징을 학습한다.

기존 E2E 자율주행 네트워크[8-9]는 카메라 영상과 차량 상태정보를 활용하여 도심 환경에서 안전한 주행을 위한 미래 주행 궤적과 제어 명령을 예측한다. 기존 E2E 네트워크 중 TCP (Trajectory-guided Control Prediction) [10]는 단일 카메라 영상 입력과 CNN 기반의 구조로 연산 효율성을 가지며, 추출된 시각적 특징으로 주행 궤적과 제어 명령을 동시에 예측한다. CIL++ [11]는 트랜스포머 기반의 다중 카메라 융합으로 넓은 시야각(FOV)을 가지며, 주변 환경을 전역적으로 인지하여 제어 명령을 예측한다. MILE (Model-based Imitation Learning) [12]은 단일 카메라 영상을 3D 기하학적 정보가 반영된 BEV 잠재 표현으로 압축하여 공간 정보를 이해하며, 모델 기반 학습을 통해 미래의 주행 상태를 반영하여 제어 명령을 예측한다.

자율주행 네트워크는 온보드 환경에서 안전한 주행 판단을 위해 실시간 추론 성능을 요구한다. 그러나 이러한 카메라 기반 E2E 네트워크는 온보드 환경에서 카메라 영상의 해상도에 영향을 크게 받는다. 해상도에 따라 실시간 추론과 네트워크 성능 사이에 트레이드오프(tradeoff)가 발생한다. 고해상도 카메라 영상은 신호등, 전방 차량과 같이 주행 판단에 영향을 주는 시각적 특징을 효과적으로 추출한다. 그러나 입력 데이터의 증가로 추론 지연이 증가한다. 반면, 저해상도 영상은 실시간성 확보에는 유리하지만, 주행 판단에 영향을 주는 시각적 특징이 손실될 수 있다. 이는 네트워크의 환경 인지 성능이 저하되어, 안전한 주행 궤적 및 제어 명령 예측을 방해하는 원인이 된다.

본 논문에서는 온보드 환경에서 카메라 해상도에 따른 기존 방법의 한계를 개선하여 주행 성능 향상을 목표로 한다. 이를 위해 저해상도와 고해상도 영상의 특징을 융합하여, 실시간 추론과 주행 판단에 영향을 주는 특징 추출을 동시에 수행하는 방법을 제안한다. 제안된 방법은 저해상도 영상으로 전체 이미지의 전역 특징을 추출한다. 이와 동시에 고해상도 영상에서 신호등이나 전방 차량 등 주행 판단에 영향을 주는 영역만 관심영역으로 설정하여 특징을 추출한다. 이후 저해상

* Corresponding Author

Manuscript received November 7, 2025; revised December 1, 2025; accepted January 27, 2026

연주연: 충북대학교 지능로봇공학 대학원생(yjyeon@chungbuk.ac.kr, ORCID[®] 0009-0007-8151-4490)

이전혁: 충북대학교 지능로봇공학 대학원생(davian@chungbuk.ac.kr, ORCID[®] 0009-0001-7077-1949)

박태형: 충북대학교 지능로봇공학 교수(taehpark@cbnu.ac.kr, ORCID[®] 0000-0002-3695-344x)

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업임(IITP-2026-RS-2020-II201462).

도 영상에서 추출된 특징과 고해상도 관심영역에서 추출된 특징을 융합하여 주행 판단에 중요한 영역을 강조한다.

본 논문은 MORAI 시뮬레이터에서 기존 카메라 기반 E2E 네트워크와 비교한다. 제안된 방법은 NVIDIA Orin AGX 환경에서 약 25ms의 낮은 추론 시간을 유지하며 고해상도 카메라 영상 입력과 비교해 35.91% 빠른 추론 시간을 달성한다. 또한, 저해상도 카메라 영상을 사용한 기존 방법 대비 22.61% 주행 점수가 향상되었다. 이를 통해, 제안된 방법이 온보드 환경에서 실시간 추론이 가능하며, 기존 방법보다 향상된 성능을 보임을 확인하였다.

II. TCP 네트워크

본 논문에서는 제한된 연산 자원을 가진 온보드 환경을 위하여, 단일 카메라 기반의 효율적인 연산 구조를 가진 TCP [10] 네트워크를 활용한다. TCP 네트워크의 구조는 그림 1과 같다. 전체 시스템은 이미지 인코더, 측정 인코더, 특징 결합 레이어와 경로-제어 모듈로 구성된다. TCP는 입력 카메라 영상 $I \in \mathbb{R}^{H \times W \times 3}$ 와 목표 지점 $\mathbf{w}_{\text{goal}} = (x, y) \in \mathbb{R}^2$ 및 차량 속도 $v_{\text{ego}} \in \mathbb{R}$ 로부터 특징을 추출한다. 경로-제어 모듈은 추출된 특징으로 제어 명령 $\mathbf{u}_{\text{ctrl}} \in \mathbb{R}^2$ 을 예측한다. 제어 명령은 조향각(steer)과 가속도(accel)로 구성된다.

이미지 인코더는 아래 식 (1)과 같이 카메라 영상 I 로부터 영상 특징 벡터 $\mathbf{f}_{\text{img}} \in \mathbb{R}^{d_{\text{img}} \times 1}$ 를 추출한다. d_{img} 는 영상 특징 벡터의 차원을 의미하며, 이미지 인코더는 CNN 기반의 ResNet34[13]를 사용한다. 측정 인코더는 아래 식 (2)와 같이 목표 지점 $\mathbf{w}_{\text{goal}}$ 와 차량 속도 v_{ego} 로부터 측정 특징 벡터 $\mathbf{f}_{\text{state}} \in \mathbb{R}^{d_{\text{state}} \times 1}$ 를 추출한다. d_{state} 는 측정 특징 벡터의 차원을 의미하며, 측정 인코더는 MLP (Multi-layer Perceptron)로 구성된다.

$$\mathbf{f}_{\text{img}} = \text{CNN}(I) \quad (1)$$

$$\mathbf{f}_{\text{state}} = \text{MLP}([\mathbf{w}_{\text{goal}}, v_{\text{ego}}]) \quad (2)$$

특징 결합 레이어는 MLP로 구성되며, 아래 식 (3)과 같이 추출된 두 특징 벡터를 입력해서 결합 특징 벡터 $\mathbf{f}_{\text{concat}} \in \mathbb{R}^{(d_{\text{img}} + d_{\text{state}}) \times 1}$ 를 생성한다. 이후 경로-제어 모듈로 전달된다.

$$\mathbf{f}_{\text{concat}} = \text{MLP}([\mathbf{f}_{\text{img}}, \mathbf{f}_{\text{state}}]) \quad (3)$$

경로-제어 모듈은 경로 예측 분기와 제어 예측 분기로 구성된다. 경로 예측 분기 GRU_{traj} 는 순환 신경망 기반 모델인 GRU (Gated Recurrent Unit) [14]로 구성된다. 아래 식 (4)와 같이 결합 특징 벡터 $\mathbf{f}_{\text{concat}}$ 와 GRU_{traj} 에서 생성한 은닉상태(hidden state) $\mathbf{h}_{\text{traj}}$ 를 입력 받아 N 개 미래 시점의 주행 경로 $\mathbf{U}_{\text{traj}} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N\}$ 를 예측한다. $\mathbf{w}_1$ 는 해당 시점의 웨이포인트 (waypoint) (x, y) 로 구성된다.

$$\mathbf{U}_{\text{traj}} = \text{GRU}_{\text{traj}}(\mathbf{f}_{\text{concat}}, \mathbf{h}_{\text{traj}}) \quad (4)$$

제어 예측 분기 GRU_{ctrl} 는 아래 식 (5)와 같이 N 개 미래 시점의 제어 명령 시퀀스 $\mathbf{U}_{\text{ctrl}} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}$ 를 예측한다. 제어 명령 $\mathbf{a}_1$ 은 조향각과 가속도로 구성된다. 이 과정에서

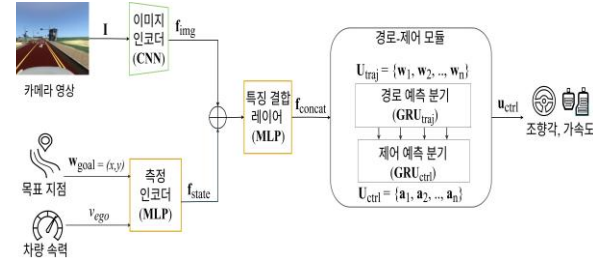


그림 1. TCP 네트워크 구조.

Fig. 1. TCP network architecture.

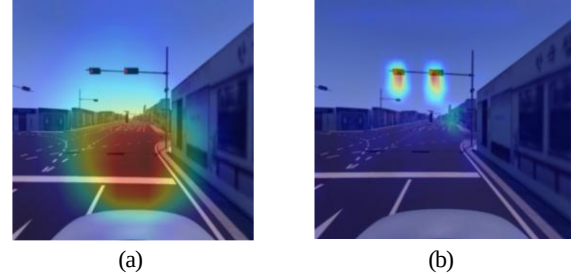


그림 2. 저해상도 영상(256 × 256) (a)와 고해상도 영상(1280 × 720)(b)에서 추출된 특징($\mathbf{f}_{\text{img}}$) 시각화.

Fig. 2. Feature visualization under low image resolution (a) and high image resolution (b).

GRU_{ctrl} 에서 생성한 은닉상태 $\mathbf{h}_{\text{ctrl}}$ 와 경로 예측 분기가 생성한 은닉상태 $\mathbf{h}_{\text{traj}}$ 를 모두 활용한다. 최종 제어 명령 $\mathbf{u}_{\text{ctrl}}$ 은 $\mathbf{U}_{\text{ctrl}}$ 의 첫 번째 제어 명령을 사용한다.

$$\mathbf{U}_{\text{ctrl}} = \text{GRU}_{\text{ctrl}}(\mathbf{f}_{\text{concat}}, \mathbf{h}_{\text{ctrl}}, \mathbf{h}_{\text{traj}}) \quad (5)$$

TCP와 같은 카메라 기반 E2E 네트워크는 안전한 제어 명령을 예측하기 위해, 입력 영상으로부터 주행 판단에 영향을 주는 특징을 추출하는 것이 중요하다. 그림 2는 TCP에서 추출된 영상 특징 $\mathbf{f}_{\text{img}}$ 을 시각화한 결과이다. 그림 2(a)와 같이 256 × 256 크기의 저해상도 영상 입력은 신호등과 같은 주행 판단에 영향을 주는 특징을 추출하기 어렵다. 반면, 그림 2(b)와 같이 1280 × 720 크기의 고해상도 영상 입력은 신호등 특징에 집중하는 것을 볼 수 있다. 그러나 입력 해상도가 높을수록 그에 따른 연산량은 증가한다. 이로 인해 온보드 환경에서 높은 주행 성능과 실시간 추론을 동시에 수행하기 어렵다.

III. 제안 방법

본 논문은 온보드 환경에서 주행과 추론 성능 향상을 목표로 저해상도-고해상도 특징 융합 모듈을 제안한다. 제안된 방법은 저해상도 영상 전체에서 전역 특징을 추출한다. 또한, 고해상도 영상에서 주행 판단에 영향을 주는 특징만 선택적으로 추출하여 융합한다. 제안된 방법의 네트워크 구조는 그림 3과 같다. 저해상도-고해상도 특징 융합 모듈은 관심영역 생성 모듈, 이미지 인코더와 특징 융합 모듈로 구성된다. 관심영역 생성 모듈은 주행 데이터를 활용하여 주행 판단에 영향을 주는 영역을 사전 정의한다. 이미지 인코더는 저해상도 영상 $\mathbf{I}_{\text{low}} \in \mathbb{R}^{H_{\text{low}} \times W_{\text{low}} \times 3}$ 와 주행 성능 향상을 위한 고해상도의

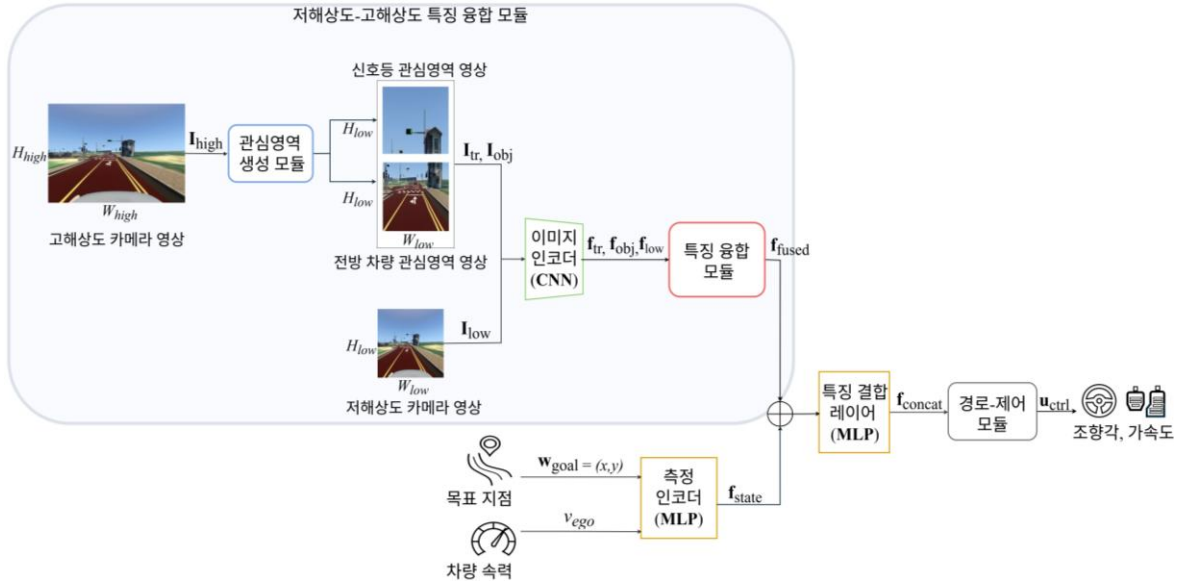


그림 3. 제안 방법 네트워크 구조.

Fig. 3. Proposed method network architecture.

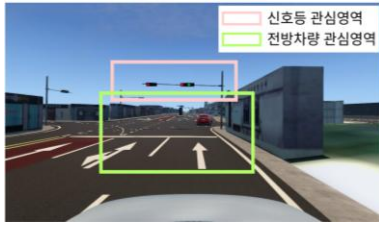
그림 4. 고해상도(I_{high}) 관심영역 예시.

Fig. 4. High-resolution ROI example.

관심영역 영상 $I_{tr} \in \mathbb{R}^{H_{low} \times W_{low} \times 3}$, $I_{obj} \in \mathbb{R}^{H_{low} \times W_{low} \times 3}$ 에서 특징을 추출한다. I_{tr} 와 I_{obj} 는 신호등과 전방 차량에 대한 고해상도 관심영역 영상을 의미하며, 저해상도 영상의 크기 $H_{low} \times W_{low}$ 로 변환되어 사용된다. 마지막으로 특징 융합 모듈은 추출된 저해상도 영상 특징과 고해상도 관심영역 특징을 융합한다. 최종 융합된 특징 벡터 $f_{fused} \in \mathbb{R}^{d_{img} \times 1}$ 는 측정 특징 벡터 f_{state} 와 결합되어 경로-제어 모듈을 통해 제어 명령 u_{ctrl} 을 예측한다.

관심영역 생성 모듈은 그림 4와 같이 고해상도 영상 $I_{high} \in \mathbb{R}^{H_{high} \times W_{high} \times 3}$ 에서 신호등과 전방 차량 관심영역을 사전 정의한다. 관심영역은 군집화 알고리즘을 통해 생성된다. 이를 위해, 학습 데이터셋에서 에고(ego) 차량의 진행 방향에 있는 신호등과 전방 차량의 경계상자(bounding box)를 수집한다. 이후에 K-means 군집화 알고리즘을 적용하여, 다양한 위치와 크기를 가진 경계상자를 군집화한다.

경계상자 정보는 아래 식 (6)과 같다. $\mathcal{B}$ 는 주행 데이터에서 수집된 경계상자 정보를 의미하며, (c_x, c_y) 와 (w, h) 는 경계상자의 중심점과 크기를 의미한다. K-means 군집화 알고리즘은 식 (7)과 같이 관심영역 후보의 중심점을 계산한다. K 는 군집의 개수, S 는 전체 군집의 집합, S_k 는 k 번째 군집에 속한 경계상자의 정보, μ_k 는 k 번째 군집의 중심점을 의미한다. 본 논문에서는 군집 개수 K 는 50을 사용한다.

$$\mathcal{B} = \{\mathbf{b}_i\}_{i=1}^N$$

$$\mathbf{b} = (c_x, c_y, w, h)$$
(6)

$$S^* = \underset{S}{\operatorname{argmin}} \sum_{k=1}^K \sum_{\mathbf{b} \in S_k} \|\mathbf{b} - \mu_k\|^2$$
(7)

식 (7)을 통해 계산된 군집 중심을 관심영역 후보로 설정한다. 이후 신호등과 전방 차량 각각에 대하여, 생성된 모든 후보 중심점 좌표를 포함하는 하나의 대표 관심영역을 생성한다. 학습 데이터셋을 기반으로 생성한 영역을 고해상도 관심영역으로 사용함으로써, 해당 객체가 빈번하게 나타나는 부분을 강조하고 주행 판단에 영향을 주는 특징을 효과적으로 반영한다.

이미지 인코더는 실시간 온보드 주행을 위해 저해상도 영상 I_{low} 를 입력으로 하며, 동시에 주행 판단에 영향을 주는 특징에 집중하기 위해 고해상도의 관심영역 영상 I_{tr} , I_{obj} 를 입력으로 사용한다. 입력은 아래 식 (8)과 같이 CNN 기반의 단일 이미지 인코더를 통해 저해상도 영상 특징 행렬 $f_{low} \in \mathbb{R}^{H_f \times W_f \times C_f}$ 과 고해상도 영상의 관심영역 특징 행렬 $f_{tr} \in \mathbb{R}^{H_f \times W_f \times C_f}$, $f_{obj} \in \mathbb{R}^{H_f \times W_f \times C_f}$ 을 동시에 추출한다. 이미지 인코더는 ResNet34를 사용한다. $H_f \times W_f \times C_f$ 는 $H_{low} \times W_{low} \times 3$ 크기의 영상에서 추출된 특징 행렬의 크기를 의미한다.

$$f_{low}, f_{tr}, f_{obj} = \text{CNN}([I_{low}, I_{tr}, I_{obj}])$$
(8)

다중 영상 입력을 배치 차원으로 결합하여 단일 인코더에서 병렬 처리함으로써, 추론 시간 증가를 최소화하였다. 또한, 추출된 관심영역 특징은 주행 판단에 필요한 신호등, 전방 차량에 집중하여 주행 성능을 향상시킨다. 이러한 방법은 그림 5(a), 5(b)와 같이 관심영역 특징 f_{tr} , f_{obj} 가 각각 신호등과 전방 차량에 집중하는 것을 시각적으로 볼 수 있다.

특징 융합 모듈의 구성은 그림 6과 같다. 추출된 관심영역 특징 행렬 f_{tr} , f_{obj} 은 1×1 합성곱(convolution) 연산과 Sigmoid

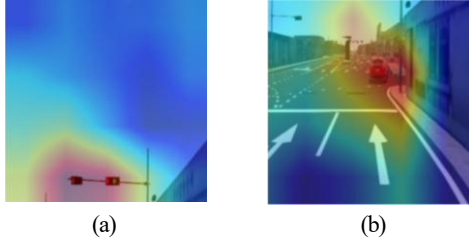


그림 5. 신호등 관심영역 특징(f_{tr}) (a)와 전방 차량 관심영역 특징(f_{obj}) (b) 시각화.

Fig. 5. ROI feature visualization of traffic light (a) and leading vehicle (b).

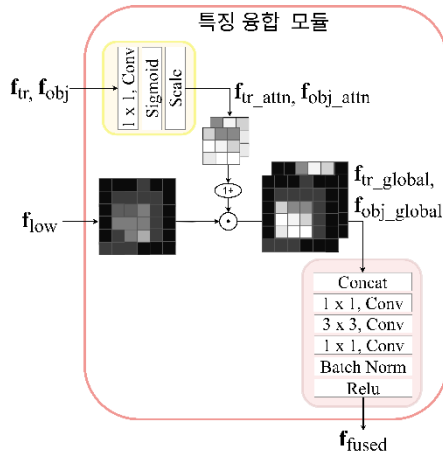


그림 6. 특징 융합 모듈 구조

Fig. 6. Structure of the feature fusion module

함수를 통해 0과 1 사이의 값을 가진 어텐션 맵(attention map)으로 변환된다. 어텐션 맵은 주행 판단에 영향을 주는 신호등과 전방 차량 특징에 집중하도록 학습된다.

생성된 어텐션 맵은 저해상도 영상에서 추출한 특징과 융합하기 위해, 쌍선형 보간법(bilinear interpolation)으로 저해상도 특징 행렬과 동일한 공간 크기로 재구성한다. 재구성된 어텐션 맵 $f_{tr_attn} \in \mathbb{R}^{H_r \times W_r \times 1}$, $f_{obj_attn} \in \mathbb{R}^{H_r \times W_r \times 1}$ 은 저해상도 영상 특징 행렬에서 고해상도 관심영역에 대응되는 위치를 나타낸다. $H_r \times W_r$ 는 저해상도 영상 특징 행렬에서 관심영역에 해당하는 공간 크기를 의미한다. 재구성된 어텐션 맵은 아래 식 (9), (10)과 같이 저해상도 영상 특징 행렬 f_{low} 에 가중치로 적용한다. 이는 관심영역 외의 특징은 유지하면서 어텐션 맵을 통해 주행 판단에 필요한 특징 영역만을 선택적으로 강조한다. 각 관심영역 기반으로 강조된 특징 행렬은 $f_{tr_global} \in \mathbb{R}^{H_f \times W_f \times C_f}$, $f_{obj_global} \in \mathbb{R}^{H_f \times W_f \times C_f}$ 이다.

$$f_{tr_global} = f_{low} \odot (1 + f_{tr_attn}) \quad (9)$$

$$f_{obj_global} = f_{low} \odot (1 + f_{obj_attn}) \quad (10)$$

이와 같이 생성된 특징 행렬 f_{tr_global} , f_{obj_global} 은 융합 레이어를 통해 최종 영상 특징 벡터 f_{fused} 을 생성한다. 융합 레이어는 1×1 합성곱 연산을 통해 채널을 감소시켜 연산량 증가를 최소화한다. 융합된 특징 벡터 f_{fused} 는 경로-제어 모듈의 입력으로

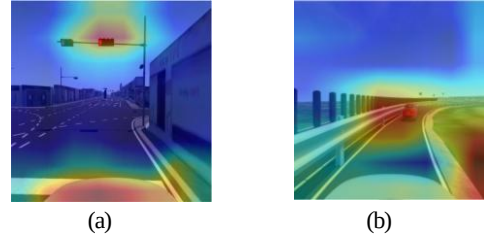


그림 7. 신호등 (a) 및 전방 차량 (b) 영상에서 융합된 특징 (f_{fused}) 시각화.

Fig. 7. Fused feature visualization of traffic light (a) and leading vehicle (b) images.

사용된다. 또한, 그림 7(a), 7(b)와 같이 저해상도 영상 특징과 고해상도 관심영역 특징을 융합하여, 주행에 필요한 신호등과 전방 차량 영역에 집중하는 것을 볼 수 있다.

IV. 실험

1. 실험 환경 구성

본 논문은 도심 주행 성능 검증을 위해 MORAI 시뮬레이터에서 페루프(closed-loop) 평가를 수행한다. E2E 네트워크의 추론은 온보드 주행을 위해 NVIDIA Orin AGX에서 실행된다. 학습 데이터는 MORAI 시뮬레이터의 KATRI 맵에서 수집된 약 300,000 프레임의 주행 데이터를 사용하였다. 입력 카메라의 해상도는 256×256 이며, 화각은 90° 를 사용한다. 주행 성능은 카메라 기반 E2E 네트워크인 TCP [10], CIL++ [11], MILE [12]과 비교한다. TCP-RoI는 제안된 방법을 의미한다. 그림 8(a)

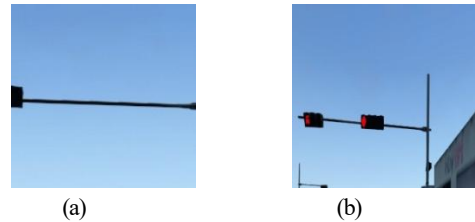


그림 8. 군집 개수 $K=10$ (a)와 $K=50$ (b)에 대한 관심 영역 영상 시각화.

Fig. 8. ROI visualization for different numbers of clusters K .

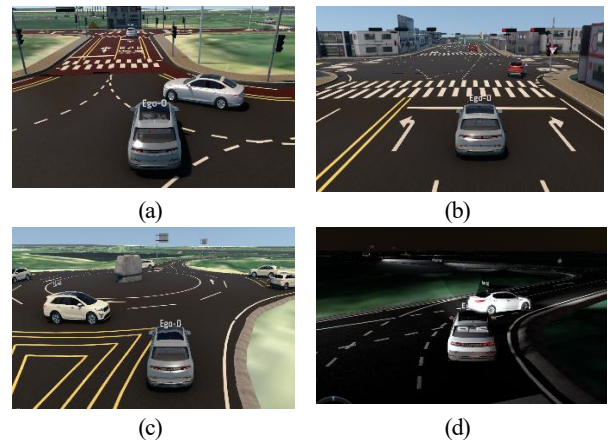


그림 9. MORAI KATRI 맵 도심 주행 환경.

Fig. 9. Urban driving scenarios in the MORAI KATRI.

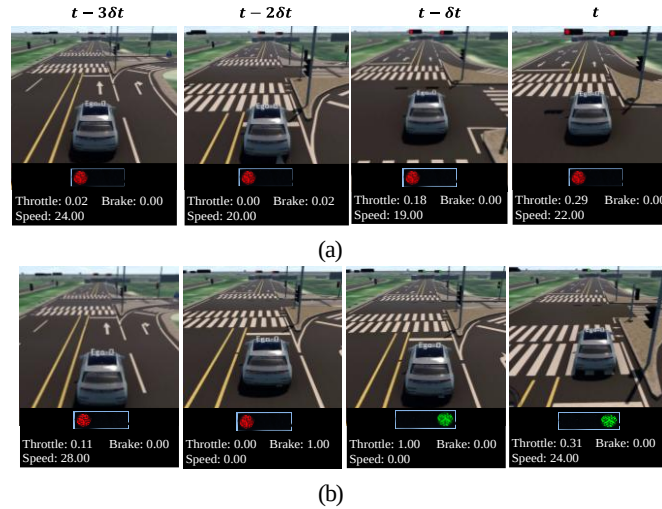


그림 10. MORAI 시뮬레이터에서 신호등 상태 변화에 따른 TCP (a) 및 TCP-RoI (b) 주행 성능 비교.

Fig. 10. Comparison of driving performance of TCP (a) and TCP-RoI (b) under different traffic light states in the MORAI simulator.

와 같이 군집 개수 K 가 적을 때 발생하는 객체 정보 손실을 방지하기 위하여, 그림 8(b)와 같이 K 를 50으로 설정하여 관심 영역을 생성한다.

실험은 그림 9와 같이 KATRI 맵에서 진행하며, 날씨는 Sunny로 설정하였다. 주간 13시와 야간 18시 환경에서 모든 시나리오를 각각 5번 반복 주행한다. 주행 시나리오는 최대 20분 주행하는 장거리 주행, 고속도로 주행, 그리고 10개의 단거리 주행 경로로 구성된다. 그림 9(a)와 같이 교차로 내 전방 차량 대응, 그림 9(b)와 같이 신호 변화 대응, 그림 9(c)와 같이 회전 교차로에서 차량 간 상호작용, 그리고 그림 9(d)와 같이 야간 환경에서 전방 차량 대응 등의 다양한 도심 주행 상황을 포함한다. 장거리 주행 시나리오는 맵 내 55대의 차량이 사전 정의된 경로를 따라 주행하며 예고 차량과 교차하고, 11개의 신호등의 상태가 실행마다 변화한다.

모델의 주행 성능은 CARLA 공식 리더보드 2.1 [15]의 평가 지표를 사용하여 평가한다. 아래 식 (11)~(13)과 같이 경로 완주율(route completion, RC), 위반 점수(infraction score, IS), 그리고 주행 점수(driving score, DS)를 사용하여 평가한다.

$$RC = \left(\frac{R_{completed}}{R_{total}} \right) \cdot 100 \quad (11)$$

$$IS = \frac{1}{1 + \sum_j (c_j \cdot N_j)} \quad (12)$$

$$DS = RC \cdot IS \quad (13)$$

RC는 전체 경로 R_{total} 에서 평가 모델이 완료한 주행 경로 $R_{completed}$ 의 비율을 나타낸다. IS는 초기점수 1.0을 시작으로 위반 행위가 발생할 때마다 해당 페널티 계수 c_j 를 분모에 합산하여 점수를 차감하는 방식으로 계산된다. 본 논문에서는 CARLA 공식 평가 지표와 동일하게, 신호 위반과 차량 충돌에 대한 페널티 계수를 각각 0.4와 0.7로 설정한다. DS는 RC와 IS를 곱한 최종 주행 점수를 의미한다.

2. 온보드 환경에서의 주행 성능 비교

그림 10은 네 개 시점에서 TCP와 TCP-RoI의 신호 변화 대응 과정을 시각적으로 비교한다. TCP-RoI는 적색 신호를 안정적

표 1. MORAI KATRI 맵 주행 성능 비교.

Table 1. Comparison of driving performance in the MORAI KATRI.

Model	RC [%]	IS	DS	Latency [ms]
TCP	91.51	0.57	53.91	21.72
CIL++	83.22	0.42	35.88	31.25
MILE	92.60	0.36	32.84	30.79
TCP-RoI	94.92	0.69	66.10	25.91

표 2. 제안 방법의 모듈에 따른 주행 성능 비교.

Table 2. Ablation study on the effectiveness of different components.

Model			RC [%]	IS	DS	Latency [ms]
Multi-resolution	RoI	Fusion				
			92.53	0.64	59.22	40.43
✓			84.58	0.67	56.36	48.56
✓	✓		81.98	0.62	53.13	24.23
✓	✓	✓	94.92	0.69	66.10	25.91

표 3. Bench2Drive 주행 궤적 예측 L2 거리 오차 비교.

Table 3. Comparison of trajectory prediction L2 errors on Bench2Drive.

Model	L2 [m]				
	0.5s	1.0s	1.5s	2.0s	Avg
TCP	1.16	1.67	2.31	3.07	2.05
TCP-RoI	1.00	1.28	1.76	2.42	1.62

으로 인지하여 정지선 정차 및 주행 재개를 수행하는 반면, TCP는 불안정한 신호 인지로 인하여 충분한 감속을 하지 못하여 신호를 위반한다.

표 1은 MORAI 시뮬레이터의 KATRI 맵에서 주행 성능을 비교한다. TCP-RoI는 저해상도-고해상도 특징 융합을 통해 약 25ms의 낮은 추론 지연을 달성하며, 기존 방법 대비 주행 점수 DS가 22.61% 향상되었다.

표 2는 고해상도 단일 입력을 기준으로, 제안 방법의 구성 요소를 단계적으로 적용했을 때의 성능 변화를 보여준다. 관심 영역 적용 시 추론 지연이 크게 감소하였으며, 융합 모듈을 적용하였

을 때 주행 점수 DS가 66.10으로 가장 높은 성능을 달성하였다.

표 3은 CARLA 시뮬레이터의 다양한 날씨, 시간, 교통 상황을 포함한 Bench2Drive 데이터셋 기반의 open-loop 실험 결과로, 50개 시나리오에 대한 궤적 예측 L2 오차를 비교한다. 제안 방법 TCP-RoI는 MORAI 데이터 기반의 관심 영역을 그대로 사용했음에도, 기존 방법 TCP 대비 모든 예측 구간에서 낮은 주행 궤적 오차를 기록하며, 평균 오차는 약 21% 감소하였다.

V. 결론

본 논문은 온보드 환경에서 실시간 추론과 주행 성능을 동시에 달성하기 위해, 저해상도-고해상도 영상 특징 융합 모듈을 제안하였다. 제안된 방법은 저해상도 영상에서 추출된 특징과 고해상도 영상의 관심영역에서 추출된 특징을 융합하였다. 신호등과 전방 차량에 대한 특징만을 선택적으로 융합함으로써, 해상도에 따라 발생하는 추론 시간과 주행 성능 간 트레이드오프를 완화하였다. NVIDIA Orin AGX 환경에서 실험 결과, 제안된 방법은 25ms의 실시간 추론을 달성하며, 기존 방법 대비 주행 점수가 약 22.61% 이상 향상되었다. 제안된 방법을 통해 온보드 환경에서 입력 카메라 영상의 해상도로 인한 기존 카메라 기반 E2E 자율주행 네트워크의 성능 한계를 효과적으로 개선함을 보였다.

REFERENCES

- [1] Y. Dai and S.-G. Lee, "Perception, planning and control for self-driving system based on on-board sensors," *Advances in Mechanical Engineering*, vol. 12, no. 9, 2020.
- [2] H. J. Sim, J. Kim, J. Hong, S. W. Nam, Y. H. Kim, J. Heo, H. C. Hwang, and K. K. Kim, "Visual navigation in unstructured environments and the development of a multi-purpose mobile robot platform," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 9, pp. 913-923, Sep. 2024.
- [3] N. Ludlow, Y. Lyu, and J. Dolan, "Hierarchical learned risk-aware planning framework for human driving modeling," *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2024.
- [4] B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, and X. Wang, "VAD: Vectorized scene representation for efficient autonomous driving," *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 8340-8350, 2023.
- [5] J. Lee, J. Yeon, and T. Park, "End-to-end network for autonomous vehicles based on a bird's eye view perception module," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 31, no. 8, pp. 921-926, Aug. 2025.
- [6] W. Zheng, R. Song, X. Guo, C. Zhang, and L. Chen, "GenAD: Generative end-to-end autonomous driving," *Proc. of the European Conference on Computer Vision (ECCV)*, pp. 87-104, 2024.
- [7] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, L. Lu, X. Jia, Q. Liu, J. Dai, Y. Qiao, and H. Li, "Planning-oriented autonomous driving," *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17853-17862, 2023.
- [8] M. Mirzaie and B. Rosenhahn, "Interpretable decision-making for end-to-end autonomous driving," *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 794-804, 2025.
- [9] D. Porres, Y. Xiao, G. Villalonga, A. Levy, and A. M. López, "Guiding attention in end-to-end driving models," *Proc. of the IEEE Intelligent Vehicles Symposium (IV)*, pp. 2353-2360, 2024.
- [10] P. Wu, X. Jia, L. Chen, J. Yan, H. Li, and Y. Qiao, "Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline," *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [11] Y. Xiao, F. Codevilla, D. Porres, and A. M. Lopez, "Scaling vision-based end-to-end autonomous driving with multi-view attention learning," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1586-1593, 2023.
- [12] A. Hu, G. Corrado, N. Griffiths, Z. Murez, C. Gurau, H. Yeo, A. Kendall, R. Cipolla, and J. Shotton, "Model-based imitation learning for urban driving," *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 20703-20716, 2022.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
- [14] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724-1734, Oct. 2014.
- [15] CARLA team. Carla autonomous driving leaderboard. <https://leaderboard.carla.org/>, 2025. Accessed: 2025-12-29.



연 주 연

2024년 충북대학교 전자공학부 졸업.
2024년~현재 충북대학교 지능로봇공학 석사과정. 관심분야는 자율주행 환경 인지, End-to-end 자율주행.



이 전 혁

2021년 충북대학교 전자공학부 졸업. 2022년 충북대학교 제어로봇공학 석사 졸업. 2022년~현재 충북대학교 제어로봇공학 박사과정. 관심분야는 자율주행 환경 인지, End-to-end 네트워크 학습 방법 등.



박 태 형

1988년 서울대 제어계측공학과 졸업. 1990년 동 대학원 석사. 1994년 동 대학원 박사. 1994년~1997년 삼성테크윈(주) 선임연구원 1997년~현재 충북대학교 지능로봇공학과 교수. 2000년~2001년 토론토 대학교 방문교수. 2012년~2017년 대한전기학회 로봇틱스 및 자동화 연구회 위원장. 2020년~현재 과기정통부 지정 산업인공지능 연구센터 (Grand ICT) 센터장. 관심분야는 로봇틱스, 자율주행, 산업인공지능 등.