

# 딥러닝 영상분석 시스템의 성능평가 산정식 개발

## Development of Performance Evaluation Formula for Deep Learning Image Analysis System

손 현 호\* · 김 윤 상\*\* · 이 철 기\*\*\*

\* 주저자 : 도로교통공단 경기지방 안전연구원

\*\* 교신저자 : 평택경찰서 교통과 경위

\*\*\* 교신저자 : 아주대학교 교통시스템공학과 교수

Hyun Ho Son\* · Yun Sang Kim\*\* · Choul Ki Lee\*\*\*

\* Senior Researcher, Dept. of Traffic Safety & Facility, Korea Road Traffic Authority

\*\* Division of Traffic, Pyeongtack police station

\*\*\* Professor, Dept of Transportation Eng, Ajou University

† Corresponding author : Yun Sang Kim, khid3639@police.go.kr

Vol. 22 No.4(2023)  
August, 2023  
pp.78~96

pISSN 1738-0774  
eISSN 2384-1729  
<https://doi.org/10.12815/kits.2023.22.4.78>

Received 16 June 2023  
Revised 1 July 2023  
Accepted 6 July 2023

© 2023. The Korea Institute of  
Intelligent Transport Systems. All  
rights reserved.

### 요 약

도시부 교통정보 수집은 VDS, DSRC, 레이더 등 다양한 시스템에 의해 수집되고 있다. 최근 딥러닝 기술의 발전으로 스마트교차로시스템이 확대 보급되고 있으며 교통량, 속도, 차종 등 다양한 정보수집이 가능하다. 그러나 관련 문헌을 고찰한 결과 지금까지의 성능평가 기준은 딥러닝 영역을 고려하지 않은 RBS기반 평가체계로 '기준값-측정값'의 퍼센트 오차만 고려하고 있어 기존 평가방식으로는 딥러닝 부분의 평가를 수행할 수 없어 새로운 성능평가 방법이 필요하다. 따라서, 본 연구에서는 데이터 비율 및 가중치를 고려하여 Precision과 Recall 등 딥러닝 성능지표를 고려한 오차산정식을 개발하여 개별오차와 구간 오차, 전체오차를 산정하였다. 연구결과, 측정값 1의 오차율은 3.99와 3.54, 측정값 2는 5.34와 5.07로 기존 산정식과 오차율에 차이가 있는 것으로 나타났으며, 반복측정 분석결과 개발 산정식이 우수한 것으로 나타났다.

핵심어 : 딥러닝, 영상분석, 성능지표, 오차율

### ABSTRACT

Urban traffic information is collected by various systems such as VDS, DSRC, and radar. Recently, with the development of deep learning technology, smart intersection systems are expanding, are more widely distributed, and it is possible to collect a variety of information such as traffic volume, and vehicle type and speed. However, as a result of reviewing related literature, the performance evaluation criteria so far are rbs-based evaluation systems that do not consider the deep learning area, and only consider the percent error of 'reference value-measured value'. Therefore, a new performance evaluation method is needed. Therefore, in this study, individual error, interval error, and overall error are calculated by using a formula that considers deep learning performance indicators such as precision and recall based on data ratio and weight. As a result, error rates for measurement value 1 were 3.99 and 3.54, and rates for measurement value 2 were 5.34 and 5.07.

Key words : Deep learning, Traffic image analysis, Performance index, Error rate

## I. 서 론

### 1. 연구의 배경 및 목적

#### 1) 연구의 배경

우리나라는 복잡한 도시부 교통체증을 해결하기 위해 1997년 ‘국가ITS기본계획’ 수립을 시작으로 지금까지 다양한 ITS 관련 시스템을 구축 운영하고 있다. 특히 차량검지 관련 시스템은 컴퓨터 비전기술을 이용하면서 비약적 발전을 이루었다. 최근 딥러닝 알고리즘을 이용해 교통량 수집뿐 아니라 차종 분류 및 각종 이벤트 분류까지 가능한 시스템으로 발전하고 있으며, 가장 대표적인 시스템은 지능형교통체계(ITS, Intelligent Transport System)에 인공지능(AI)·첨단센서 등 신기술을 도입한 ‘스마트신호 운영 시스템’이라 할 수 있다. 스마트신호 운영은 감응 신호와 긴급차량 우선신호, 스마트교차로로 구분된다. 특히 딥러닝 기술이 접목된 스마트교차로는 방향별 교통량 및 차종 구분이 가능한 시스템으로 동시에 다방면의 정보를 처리할 수 있는 효율적인 시스템으로 국토교통부는 스마트교차로 사업을 지속 확대하고, 신기술을 적극적으로 도입할 계획이다. 스마트교차로 시스템은 정해진 명령 조건을 수행하는 Rule base system과 차종을 분석하는 Deep learning system의 조합으로 기존 평가방식은 ‘기준값-측정값’의 퍼센트 오차만 고려하고 있어 딥러닝 부분의 평가를 수행할 수 없어 Deep learning 영역에 대한 성능평가 방법이 필요하다. 따라서, 딥러닝 품질지표 및 데이터 특성을 고려한 오차 산정식 개발과 성능평가 지표를 개발하고자 한다.

#### 2) 연구의 필요성

최근 영상을 이용한 교통정보 수집시스템은 가상의 검지선을 생성하고 객체 통과 시 이벤트를 발생하는 규칙기반 하드코딩 명령인 RBS(Rule-Based System)와 학습데이터 셋 기반의 딥러닝 알고리즘을 이용해 차종을 분류하는 체계이다. RBS는 약속에 의한 분류로, ‘주어진 조건’ 입력값 결과를 발생하는 것으로 미리 명령된 Rule에 따라 동작과 데이터를 생성하는 명시적 프로그래밍 기반이지만, 딥러닝 알고리즘은 주어진 명령이 아닌 ‘예측된 측정값’을 전제로 한 Computer Vision(CV) 영역으로 검지(Object Detection), 추적(Object Tracking), 분류(Action Classification)를 기본으로 실제와 예측에 관한 결과를 생성한다.

지금까지 성능평가는 딥러닝 영역을 고려하지 않은 RBS 기반 기준평가로 ‘기준값-측정값’의 퍼센트 오차만 고려하고 있어 기존 평가방식으로는 딥러닝 부분의 평가를 수행할 수 없다. 딥러닝은 주어진 조건이 아닌 예측에 대한 평가가 필요하며, Action Classification을 평가할 수 있는 정확도(IOUS), 정밀도(Precision), 재현율(Recall)을 고려한 성능평가 방안이 필요함을 확인하였다. 최근 영상을 이용한 교통정보 성능평가 관련 지침이 개정되어 ‘스마트교차로 시스템(SIS)성능평가 기준’이 신설되었으며 차종, 대기행렬이 평가항목에 추가되었다. 평가척도 및 분석방법은 APE(Absolute Percentage Error)와 PE(Percentage Error)로 개정된 지침에도 전과 같은 ‘기준값-측정값’의 퍼센트 오차로 딥러닝 자체가 가지고 있는 품질지표에 대한 검사대상 데이터 및 평가방법은 고려하지 않았다. 즉, 품질 요구조건 및 성능지표 조건이 다른 시스템에 대하여 일관된 평가방식으로 품질을 측정하고 있다. 국내에서도 차량 검지기 및 교통량 분석을 위한 딥러닝과 관련한 많은 연구가 있었지만, 평가 기준은 과거 기준체계를 따랐으며, 딥러닝 기반 관련 연구에서는 차종 분류 세분화, 딥러닝 알고리즘 간 교통량 수집 결과율 등을 다루고 있을 뿐 성능평가 기준 및 방안에 관련한 제시는 하지 않고 있어 평가체계 및 정량적 기준 제시가 필요하며, 기술발전과 더불어 평가방안도 개선되어야 한다.

### 3) 연구의 목적

영상 검지를 이용한 교통정보 수집장치의 성능은 산술평균 관점의 PE 오차가 아닌 딤러닝 관점의 단계별 오차를 고려해야 한다. 딤러닝에서는 차량이 검출되기까지 일련의 과정을 통해 예측과 실제의 조합으로 분류한 차량을 결괏값으로 만들어 낸다. 산술오차는 기준값이 정해 있지만, 딤러닝 성능지표는 기준값 변경도 가능하며 지표별로 존재하기 때문에 단계별 오차율을 비교할 수 있다. 또한, 정확도와 정밀도의 관점에서 반복측정에 의한 결과로 알고리즘의 성향 및 문제점을 분석할 수 있다. 이러한 단계별 오차와 반복측정의 결과는 기계적 관점에서 100%의 검지율이 딤러닝 관점에서는 100%가 아닐 수 있다는 것을 전제로 한다. 또한, 기존 오차율 산정식은 각 Class별로 같은 비중을 둔 평균을 구하기 때문에 Class별 데이터 수에 영향을 받지 않지만 Class별 데이터 수가 다른 경우에는 가중치를 고려하는 것이 더 의미 있는 평가지표가 될 수 있다.

따라서 본 연구의 목적은 Class별 데이터 수를 고려하여 딤러닝 성능평가 지표개발을 하고 성능지표를 선정하여 지표별 개별 오차율, 차종별 오차율, 구간별 오차율을 분석할 수 있는 단계별 산정식을 개발하고자 한다. 또한, 개발된 산정식은 반복측정에 의한 정밀도의 차이를 분석하는 기초자료로 활용하여 딤러닝 알고리즘 정밀도를 평가해보고 산정식에 의한 결과가 통계적으로 유의하고 적합한 것인지 판단해 보고자 한다.

## 2. 연구의 범위 및 절차

본 연구의 시간적 범위는 오전 침두시 08시부터 08시 30분까지를 기준으로 하여 국내 평가 기준의 최소 조건을 만족하였으며, 공간적 범위는 교차로를 진입하는 주 방향 전 차로를 기준으로 하였다. 가상의 검지선이 설치되는 공간은 정지선을 기준으로 진입과 진출에 대해 교차로 내 통과 교통류를 차로별로 구분할 수 있도록 하였으며, 화각 내 진입 및 통과 차량은 육안 판독이 가능하도록 시가 제약이 최소화되도록 하였다. 내용적 범위는 차종 구분은 승용, 버스, 화물, 이륜 4종으로 구분하였으며 주 방향 접근로는 직진, 좌회전, 우회전이 모두 가능한 차로로 정의하였다. 검지 영역 설정 시 지장물에 의해 차량 이동 및 추적이 소실되거나 영상수집에 외부 제약이 발생하지 않도록 하였다. 평가를 위한 최소차량 대수는 국내 최소차량 대수 기준인 200대 이상을 기준으로 하였다. 연구의 내용적 절차는 첫 번째로 기준값 계산을 위해 수집 영상을 16프레임 분석이 가능한 비디오 분석 프로그램을 통해 육안 분석으로 차종을 구분하고, 두 번째로 데이터 탐색을 통해 데이터의 기술통계 및 데이터 정규성을 검증한다. 세 번째는 기존 논문을 고찰하여 문제점 및 개선방안을 도출하고 네 번째는 오차율 평가에 있어 단계별 평가방안을 제시하고 마지막으로 기존 평가방식의 단점을 최소화하여 설득력 있는 산정식을 개발하여 기존평가 방식과의 차이점을 분석하고자 한다. 연구의 단계별 수행절차는 연구 범위설정, 관련 이론 및 선행연구 고찰, 연구방법론 선정, 자료수집 및 분석, 결론 순이다.

## II. 관련 이론 및 선행연구 고찰

### 1. 국내 평가 기준

교통량 영상 검지 관련 국내 평가 기준은 ‘자동차 도로교통 분야 ITS 성능평가 기준’에 명시되어 있으며, VDS (Vehicle detection system)과 SIS (Smart intersection system)에 의하면 교통량, 속도, 점유율, 차종 구분 등에 대하여 ‘분석시간 동안 기준값 - 분석시간 동안 대상 장비 측정값’에 대하여 VDS는 MAPE, SIS는 WAPE와 PE 평가식을 이용해 오차율을 계산하는 것으로 조사되었다. 평가방법에서 기준으로 하는 시간 범위

와 최소차량 대수는 30분 기준 200대 이상, 10주기 이상 수집 차량 대수 200대를 기준으로 하고 있으며 차종 구분은 소형, 버스, 대형으로 구분하고 있으나 시스템 특성에 따라 별도의 차종 분류 기준을 제시할 수 있다.

공간적 범위는 VDS는 방향별 1개 차로, SIS는 차로당 방향별 통과 차량을 기준으로 하고 있다. 기준자료 수집은 현장에서 수집하는 방식과 ITS 센터에서 영상수집 하는 두 가지 방법을 채택하고 있으며, 평가등급은 동일하게 최상급, 상급, 중급, 중사급 4단계로 구분하여 5~10%의 차등을 두어 최종 성능평가 등급을 산정하고 있다.

<Table 1> Domestic Evaluation Criteria

	VDS (Vehicle detection system)	SIS (Smart intersection system)
Evaluation standard	$MAPE(\%) = \frac{\sum_{i=1}^n \frac{ Y_i - X_i }{Y_i}}{n} \times 100$ <p>Y : Analysis time actual value X : Analysis time forecast value</p>	<p>Traffic volume per lane :</p> $APE(\%) = \frac{ Y - X }{Y} \times 100$ <p>Y : Analysis time actual value(Y&gt;0) X : Analysis time forecast value</p> <p>Vehicle type classification by direction per lane :</p> $PE(\%) = \frac{X}{Y} \times 100$ <p>Y : Analysis time actual value(Y&gt;0) X : The number of vehicles for which the evaluation target equipment misclassified or misclassified (including undetected) vehicle types during the analysis time</p>
Evaluation list	Traffic volume, speed and occupancy	Traffic volume, type of vehicle, queue
Evaluation Methods	<ul style="list-style-type: none"> <li>• 1 lane per direction</li> <li>• At least 200 passing vehicles</li> <li>• Evaluation is completed if it is performed for more than 30 minutes and the total number of passing vehicles is 500 or more</li> </ul>	<ul style="list-style-type: none"> <li>• If it is difficult to evaluate in all directions, evaluation in some directions</li> <li>• Performed over 10 cycles</li> <li>• A total of 200 or more vehicles for each evaluation period (daytime, nighttime)</li> <li>• When evaluating the vehicle type classification accuracy by direction per lane, the minimum number of vehicles subject to evaluation per lane is 20 or more by vehicle type</li> </ul>
Rating	best class : ≥95%, senior class : 95>, ≥90%, middle class : 90>, ≥80%, lower middle class : 80%>	

## 2. VDS 성능평가 연구

영상검지기가 대체검지기로 적합한지를 평가하기 위해 루프 검지기와 비교하여 서울시 주요 도로인 양재대로 교차로 접근로에서 주기별 교통량, 개별차량의 점유 및 비 점유시간에 대한 정확도, 영상검지기 이미지를 이용하여 대기행렬의 길이를 측정하였으며, 측정결과의 평가를 위해 기준값과 측정값을 비교하여 등가계수로 기준값에 대한 차이로 평가를 수행하였다(Kwon, 2004). ITS 장비·시스템 성능평가 및 검·교정 체계 실용화 연구(3단계) 최종보고서에 따르면 준공 전 성능검증과 정기검사, 수시평가로 구분하였으며, 준공 전 평가 검증항목은 교통량, 속도로 하며 평가척도는 “100(%) - 평균 절대 오차백분율”을 사용하였다(Kict, 2006). 속도 오차를 산출하는 과정에서 기준속도가 포함하고 있는 불확실성에 대한 요인들이 고려되지 않아 차량검지기 성능이 고평가될 가능성이 크며 ITS 수집정보 신뢰성이 저하될 우려가 있다고 보고 성능평가 시 발생할 수 있는 오차 및 불확실한 요인들을 포함하여 평가할 수 있는 속도 측정 성능평가방법을 연구하였다. 기

준 장비 합성표준불확도와 평가 대상 장비 표준불확도를 산출하여 오차율과 보정기준에 따라 5분 단위 MAPE를 평균하여 최종 100-MPAE%를 산출, 신뢰수준 95%를 기준으로 평가방법을 제안하였다(Lee, 2012). 환경에 따른 성능 특성 연구에서 지자기 검지기를 기준 검지기로 하여 영상검지기에 대한 평가방안을 제시 하였으며, 평가대상은 교통량과 속도로 269일간 수집한 자료에 대해 차로별 15분 평균교통량을 지자기 검지기와 비교하여 대응표본 t 검정으로 표준편차와 유의확률을 통해 신뢰수준을 구하였으며 성능평가지표는 다양한 지표 중 정확도를 즉시 확인할 수 있는 MAPE로 선정하여 연구를 수행하였다(Oh, 2021).

### 3. 딥러닝 영상분석 성능평가 연구

UAV를 활용한 실시간 교통량 분석을 위한 딥러닝 기법의 적용에서 실시간 교통량 분석을 위해 UAV (Unmanned aerial vehicle) 동영상을 활용한 딥러닝(Deep learning) 기반의 차량탐지기법을 제안하였으며, 객체 탐지 알고리즘 중의 하나인 YOLOv3를 이용하여 차량 자체에 대한 검출은 90.21%의 결과를 얻었으며, Confusion matrix를 이용하여 전체 차량에 대한 차량 정확도 평가결과 95.15%의 정확도와 85.79%의 재현율을 산출하였다(Park, 2020). 성능평가의 시간 범위를 일출 30분 후부터 일몰 30분 전, 일몰 30분 후부터 일출 30분 전으로 정의하여 교통량 정확도를 산출하였으며, 평가 기준값은 실제 수집한 영상자료의 수동 측정값을 기준으로 평가지표는 교통량 정확도=100%-평균 절대 오차백분율(MAPE)로 평가등급은 ITS 성능평가 기준에 의거 분류하여 적용하였다(Kim et al., 2018). 도로 CCTV 영상기반 이동 차량 검출 차량의 검출을 위해서 Redmon의 YOLO v3를 적용하여 합성곱 신경망을 이용하여 특징을 추출하고, 3단계에 거친 업 스케일링 단계를 통하여 다양한 도로 환경에서 나타나는 여러 가지 크기의 차량을 클래스를 3가지로 설정하여 버스, 트럭, 승용차로 구분하여 학습을 진행하였으며, 2차로는 200대 이상의 데이터를 분석하여 기존 시스템에서 수집할 수 없는 교통 밀도, 대기행렬의 길이 등 다양한 교통 DB 정보를 수집하였으며, 평가식으로는 MAPE를 사용하여 정확도를 산출하였다(Park, 2020). 비전 시스템을 이용하여 FHWA의 체계에 따라 보다 미세한 차량 분류를 할 수 있는 비전 시스템을 개발했다. 제안된 시스템은 객체 인식을 Localization and classification의 두 가지로 분류했으며, 각 비디오 프레임에 대해 Class-independent region을 제안하고 생성하여 Deep convolutional neural networks를 사용하여 제안된 각 영역에 대한 특징을 추출하였으며, FHWA에서 요구하는 분류 체계인 13개 클래스에서 기존 3개 또는 4개만 분류가 가능하던 시스템을 7개 등급의 차량에 대해 분류가 가능하도록 하였다. 개발된 비전 시스템은 7가지 등급의 차량에 대해 82%~100%의 평균 정밀도(Precision)와 89%~99%의 평균 재현율(Recall)을 달성했다고 논술하고 있다(Yaw, 2017).

<Table 2> Performance evaluations : average precision and recall rates of system

Class	Total Images	TP	FP	FN	Precision(%)	Recall(%)
Class 1	18	16	0	2	1.00	0.89
Class 2	8542	7250	410	885	0.95	0.89
Class 3	1348	1007	223	118	0.82	0.90
Class 4	109	106	0	3	1.00	0.97
Class 5	567	528	22	17	0.96	0.97
Class 6	3976	3600	323	53	0.92	0.99

#### 4. 딥러닝 데이터 품질지표 관련 이론

인공지능(AI, Artificial Intelligence)산업은 다양한 분야에서 활용되고 있다. 과기부와 지능정보원은 2017년부터 매년 데이터 구축사업을 통해 인공지능과 관련한 데이터를 민간에 개방하고 있으며, 품질 관련 가이드라인을 개발하였다. 가이드라인은 데이터 정확성을 의미 정확성과 구문 정확성으로 정의하고 있으며, 의미 정확성의 성능평가지표로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score, ROC AUC 등 다양한 평가지표를 제시하고 있다. 의미 정확성이란 데이터를 라벨링하고 실제 참값과 일치하도록 데이터의 의미적인 정확성을 확보하는 것으로, 딥러닝 알고리즘의 성능을 좌우하는 매우 중요한 요소이다. <Table 3>은 의미 정확성 성능지표로 사용 가능한 것을 정리한 것으로 객체의 종류에 따라 적용 가능한 지표를 정리한 것이다.

<Table 3> Accuracy and Validity Performance Indicators

Indicators		Object		Assessment Methods
		Video	Image	Quantitative evaluation
Semantic precision	IOU	O	O	√
	Precision	O	O	√
	Recall	O	O	√
Effectiveness	Classification performance	O	O	√
	Detection performance	O	O	√
	Recognition performance	O	O	√

#### 5. 관련이론 및 문헌 고찰을 통한 시사점 도출

교통량 검지 관련 기존 연구결과를 종합해보면 VDS 기반과 딥러닝 기반으로 구분할 수 있다. 먼저 VDS 오차율 관련 연구들은 주로 측정 시점의 측정값과 기준값 차이를 회귀계열 퍼센트 오차율을 구하는 방법인 PE, APE, MAPE 등을 이용한 것을 확인할 수 있었으며, 최근 개정된 SIS 평가기준도 PE와 APE에 의한 오차율을 기준으로 하고 있다. 딥러닝 기반 오차율 관련 논문은 분류계열 오차평가 방법인 정밀도와 재현율을 주로 사용하여 정확도를 산출하였으며, MAPE와 병행한 평가방법도 있었다. VDS와는 다르게 차종 분류에 있어 다양한 차종에 대해 클래스를 나누고 각각의 차종에 대해 정확도를 산출하기도 하였다. 그러나 기존 문헌을 고찰한 결과 몇 가지 문제점이 필요할 것으로 파악되었다. 먼저, 백분율 오차를 기준으로 하는 경우, 산정 결과는 실제 값에 대해 Underestimates or Overestimates 인지 파악하기 힘든 단점이 있으며, 데이터의 대푯값으로 사용하기에 이상치가 포함된 경우나 데이터값이 존재하지 않는 구간은 적용이 부적절한 것으로 분석되었다. 이런 이유로 비율 해석에 의미 있는 값을 적용하기 위해 국내 평가 기준에서는 ‘100(%)–오차백분율(PE)이 음수(-)인 경우는 평가결과를 “0”으로 산정한다.’라는 조건을 두고 있다. 또한, 분류계열 오차평가 방법은 클래스 전체에 대한 평균으로 개별차량에 대한 오차는 고려하지 않았으며, 분류 임계값(Threshold)의 기준을 고려한 평가방안을 제시하고 있지 않았으며 두 가지 방법 모두 데이터 비율에 따른 가중치는 고려하지 않았다. 즉, 측정값이 기준값보다 큰 경우, 극단적 이상치가 존재하는 경우, 결측 구간이 있는 경우, 데이터 비율 차이가 큰 경우 기존 산정방식에 의한 오차율 결과가 정확성이 있는지 고려할 필요성이 있다. <Table 4>는 PE와 APE로 구한 정확도로 국내 기준과 같이 차종 분류에 PE를 이용한 경우 승용과 버스는 기준값보다 측정값이 작게 나타나 오차 계산이 가능하지만, 화물과 이륜의 경우 과검지로 정확도는 0%로 해석되어 전체 정확도는 60%로 매우 낮게 나타난다. 교통량 정확도에 차종 오차는 고려하지 않으므로 ‘분석시간 동안

기준값-분석시간 동안 평가대상 장비 측정값' 기준에 의하면 정확도는 100%로 차종 오차가 고려되지 않은 오류가 발생한다.

<Table 4> Problems with existing evaluation formulas

	Passenger car	Bus	Truck	Motorcycle	Sum	Accuracy	Formula
Actual value	252	3	57	2	314		
Forecast value	253	5	55	1	314		
Target	Car type	1대, 99.99%	2, 60%	-2→'0%'	-1→'0%'	60%	PE
	Traffic volume					100%	APE

이처럼 백분율 오차는 차이의 정도를 백분율 값으로 나타내기 때문에 장비나 모델의 성능을 직관적으로 이해하기 쉬워 많이 사용하고 있지만, 기준값이 측정값보다 작은 경우 과검지로 분류되어 검지율이 '0'이 되므로 정확한 평가라 할 수 없다. 반면, 분류계열 오차는 백분율 오차와 다르게 객체별 스코어에 의한 조화평균으로 정확도를 평가하며, 알고리즘의 성능과 목적에 따라 지표적용이 가능하므로 합리적이다. 그러나 의미 정확성에 관련한 성능지표는 탐지와 분류의 점수만 평가하며 개별오차와 기준값 변경에 따른 구간오차 등은 고려하지 않았다. 관련 기준 및 문헌고찰 결과 딤러닝 알고리즘을 이용한 영상분석 시스템의 성능평가 기준이 필요한 것으로 나타났다.

### Ⅲ. 연구 방법론 개발

본 연구는 회귀계열 오차를 평가방식과 분류계열 정확도 평가방식의 단점을 보완하여 평가방안을 개발하고자 하며, 선행된 연구 다수가 이상치를 배제한 방법으로 수행된 점, 수집 차종의 비율을 고려하지 않고 총량에 대한 오차를 산정한 점, 차종 구분에 대해 개별오차를 고려하지 않은 점, 반복측정에 의한 오차율 변화 등을 고려하지 않은 점 등에 착안하여 딤러닝을 이용한 차종 분류 오차율 산정에 있어 분류계열 성능지표의 특성을 고려하여 비율에 따른 가중치를 적용한 차량별 개별오차를 이용한 구간 오차, 성능지표별 오차, 반복측정에 의한 오차 변화를 반영한 산정식과 성능평가 정확도 평가방법을 개발하고자 한다.

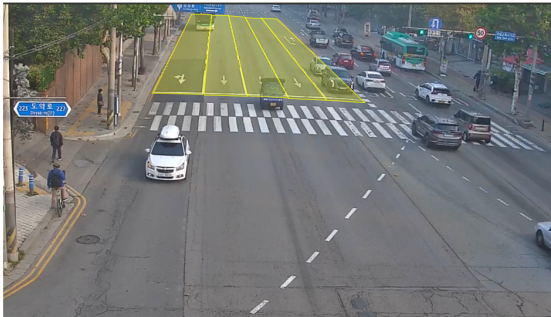
성능지표 개발 연구수행 과정은 다섯 단계로 구분할 수 있다. 첫 번째는 자료수집 단계로 기초수집 데이터는 영상수집 및 판독의 과정을 통해 결정되며, 국내 평가 기준에서 제시한 도로 인프라와 기반인프라 요구 사항을 충족한 환경의 카메라로부터 수집한 데이터를 육안 검측으로 승용, 버스, 화물, 이륜 4종으로 구분하여 기준값을 산정하는 것이다. 두 번째는 예측된 측정값을 수집하는 단계로 오차율 평가지표를 준용하여 정밀도(precision), 재현율(recall), IoU 값을 구분하고 반복측정을 통한 결과 자료를 수집하는 단계이다.

세 번째는 기준값 및 측정값의 통계분석 및 정규성 검증하는 단계로 연구 기초자료로서의 가치가 있는지 판단하는 부분이다. 네 번째는 평가 기준 및 문헌 고찰을 통해 제기한 문제점을 해결하기 위한 분류문제의 다양한 성능평가 방법을 제안하는 단계로 오차 산정식을 개발하여 오차율을 산출하고 개발 산정식의 결과가 통계적으로 설득력 있는 산정식인지 검증하는 단계이다. 마지막은 평가결과표 개발로 오차율에 따른 등급을 구분하고 성능지표별 개별 오차율 구간을 기록하여 지표별 해석이 가능하도록 정량화하는 단계이다.

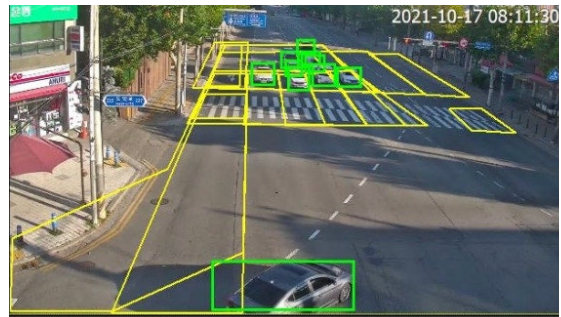
## IV. 자료수집 및 분석

### 1. 자료수집

교차로 선정 전제 조건은 검지 영역 확인 시 외부요인으로 평가영역이 손실되지 않아야 하며, 접근로는 직진, 좌회전, 우회전, 유턴이 모두 가능해야 한다. 카메라 설치 높이와 화각은 정지선을 기준으로 교차로 통과 차량의 육안식별이 가능해야 한다. 자료수집은 경기도 부천시 소재 5개 교차로이며, 기하구조는 정형·비정형 교차로 형태로 검지 영역 설정 시 물리적 제약이 없는 지점으로 선정하였다. <Fig. 1>과 <Fig. 2>는 딥러닝 기반 차량 검출 및 객체 추적기능을 수행하기 위해 CCTV 영상 검지 영역 내 설정한 관심 영역 ROI(Region of interest)를 나타낸 것이다. 관심 영역은 설정 기준은 전 차로 대상이며, 횡단보도와 일부 교차로 면적을 포함하였다. 차량 검지 후 최종 이벤트 발생 시점은 정지선 통과를 기준으로 하였다. 직진과 우회전 검출 차량의 경우 방향성을 구분하기 위해 교차로 안쪽까지 검지 영역을 확대하여 자료를 수집하였다. ROI 설정 시 진입과 진출 선을 구분하였으며, 차로 운영 특성을 반영하기 위해 접근 차로별 교통량을 수집할 수 있도록 하였다. 차종 구분은 국내평가기준을 준용하여 승용, 버스, 화물, 이륜차로 구분하였으며, 수집 대수는 총 314대로 국내평가기준 최소 조건인 200대 이상을 확보하였다. 수집자료의 참값 분석은 기준 장비에 의한 방법이 아닌 인력계수 방법을 사용하여 원본 영상을 프레임 단위로 분석 계측 하였다. 계측시 신호 위반, 화각 외 주행 등 비정상적 주행은 제외 처리하였다.



<Fig. 1> Region of interest



<Fig. 2> Setting lanes and lane boundaries

### 2. 자료분석

수집자료 분석은 데이터 탐색과 신뢰도 분석, 빈도분석을 수행하였으며, 수집데이터 N=314, 표준편차는 0.78로 나타났다. 정규성 검정은 왜도와 첨도를 사용하여 분석하였으며, 왜도 통계량은 1.65, 첨도 통계량은 0.84이며, 왜도와 첨도의 신뢰구간 계산 결과 '0'을 포함하고 있으므로 정규성을 확보한 것으로 분석되었다. 왜도와 첨도로 본 정규성 검정 근거는 '표본의 크기가 큰 경우 근사적으로 정규분포를 따른다는 이론으로 왜도와 첨도값이 [-2~2] 사이에 존재하는 경우 정규분포를 따른다'라고 보고 있다(Snedecor and Cochran, 1980). 또한 왜도는 절댓값 3을 초과하지 않고, 첨도는 절댓값 8또는 10을 초과하지 않으면 정규분포로 볼 수 있다(Kline, 2015).



<Table 5> Data analysis result

	Statistic	
Average	1.38	• [Skewness 95% Confidence interval] = Skewness $\pm$ 1.96 $\times$ Skewness's Standard deviation = 1.646 $\pm$ 1.96 $\times$ 0.138 = -0.04 ~ 0.49 ※ Calculation result includes '0', the normal distribution • [Kurtosis 95% Confidence interval] = Kurtosis $\pm$ 1.96 $\times$ Kurtosis's Standard deviation = 0.838 $\pm$ 1.96 $\times$ 0.274 = -0.31 ~ 0.76 ※ Calculation result includes '0', the normal distribution
Variance	0.606	
Standard deviation	0.778	
Skewness	1.646	
Kurtosis	0.838	

<Table 6>은 기준값과 측정값의 빈도분석 결과와 데이터 신뢰도 분석결과로 빈도분석을 통해 알아본 차종별 구성비로 승용 253대(80.6%), 버스 5대(1.6%), 화물 55대(17.5%), 이륜 1대(0.3%)로 승용의 비율이 상대적으로 매우 높게 나타났다. 신뢰도 분석에서 내적 일관성 신뢰도(Internal consistency reliability) Cronbach's alpha는 if Alpha  $\geq$  0.707로 매우 높은 신뢰도를 보였다.

<Table 6> Frequency analysis result and Cronbach's alpha

	Frequency						Reliability statistic		
	Actual value		Forecast value 1		Forecast value 2		Cronbach's alpha	Cronbach's alpha base on standardized items	Number of items
Passenger car	253	80.6	252	80.3	254	80.9			
Bus	5	1.6	3	1.0	3	1.0			
Truck	55	17.5	57	18.2	55	17.5			
Motorcycle	1	0.3	2	0.6	2	0.6			
Sum	314	100	314	100	314	100			

다음으로 딥러닝 기반 성능지표에 따라 수집자료의 IoU, Precision, Recall의 차종에 대한 개별결과에 따른 분석을 하였다. 구간별 데이터 점수 및 비율은 <Table 7~9>까지이며, IoU는 ③구간에서, Precision은 ⑦구간, Recall ⑤구간에서 비율이 가장 높은 것으로 분석되었다. 구간에 따른 차종별 분포를 보면 승용의 경우 Precision④를 제외한 모든 구간에 데이터가 분포하고 있었으며, 상대적으로 비율이 적은 버스와 이륜은 데이터 구간이 한정되어 있음을 알 수 있다. 지표별 개별데이터 분포는 측정값 1과 2에서 동일한 비율로 나타난 차종은 없었으며, 딥러닝 알고리즘 특성상 반복측정시 동일 수치가 발생할 수 있는 빈도는 거의 없는 것으로 볼 수 있다. 이런 결과는 딥러닝 알고리즘이 가진 데이터 특성으로 기존 산정방식으로는 해석할 수 없는 부분이다.

<Table 7> IOU Score and ratio analysis by interval

IOU		Forecast value 1					Forecast value 2				
		Passenger car	Bus	Truck	Motorcycle	Ratio	Passenger car	Bus	Truck	Motorcycle	Ratio
①	Score avg	0.45									
	Data ratio	0.40				0.40					
②	Score avg	0.55	0.57	0.56			0.54		0.53		
	Data ratio	3.97	33.33	10.53		47.83	3.15		9.09		12.24
③	Score avg	0.66	0.65	0.64	0.69		0.66	0.68	0.66	0.67	
	Data ratio	12.70	66.67	10.53	50	139.90	10.63	33.33	20	50	113.96
④	Score avg	0.75		0.76			0.76	0.72	0.76	0.71	

IOU		Forecast value 1					Forecast value 2				
		Passenger car	Bus	Truck	Motorcycle	Ratio	Passenger car	Bus	Truck	Motorcycle	Ratio
	Data ratio	20.24		21.05		41.29	16.93	33.33	23.64	50	123.90
⑤	Score avg	0.85		0.86			0.85	0.85	0.84		
	Data ratio	30.56		28.07		58.63	24.02	33.33	16.36		73.71
⑥	Score avg	0.92		0.93			0.93		0.92		
	Data ratio	14.68		15.79		30.47	13.78		10.91		24.69
⑦	Score avg	0.99		0.99	1		1		1		
	Data ratio	17.46		14.04	50	81.50	31.50		20		51.50

<Table 8> Precision Score and ratio analysis by interval

Precision		Forecast value 1					Forecast value 2				
		Passenger car	Bus	Truck	Motorcycle	Ratio	Passenger car	Bus	Truck	Motorcycle	Ratio
④	Score avg									0.8	
	Data ratio									100	100
⑤	Score avg	0.83	0.86	0.83			0.86		0.85		
	Data ratio	13.49	100	26.32		139.81	21.65		30.91		52.56
⑥	Score avg	0.93		0.92	0.91		0.93		0.93		
	Data ratio	15.87		33.33	100	149.20	32.68		40		72.68
⑦	Score avg	1		1			0.98	0.99	0.98		
	Data ratio	70.63		40.35		110.98	45.67	100	29.09		174.76

<Table 9> Recall Score and ratio analysis by interval

Recall		Forecast value 1					Forecast value 2				
		Passenger car	Bus	Truck	Motorcycle	Ratio	Passenger car	Bus	Truck	Motorcycle	Ratio
①	Score avg	0.27					0.50				
	Data ratio	1.59				1.59	0.39				1.98
②	Score avg	0.55					0.57				
	Data ratio	1.59				1.59	1.18				2.77
③	Score avg	0.66	0.68	0.67			0.65		0.68		
	Data ratio	4.37	33.33	1.75		39.45	3.54		3.64		46.63
④	Score avg	0.76	0.75	0.75			0.77	0.79	0.76		
	Data ratio	14.29	33.33	8.77		56.39	11.42	66.67	5.45		139.93
⑤	Score avg	0.86	0.86	0.83	0.83		0.86		0.86	0.89	
	Data ratio	24.21	33.33	8.77	100	166.31	30.31		32.73	100	329.35
⑥	Score avg	0.92		0.91			0.93		0.92		
	Data ratio	9.92		1.75		11.67	24.02		21.82		57.51
⑦	Score avg	1		1			0.98	0.97	0.98		
	Data ratio	44.05		78.95		123	29.13	33.33	36.36		221.83

### 3. 성능평가 산정식 개발

교통량에 따른 차종 패턴은 항상 같은 값이 나오는 결정론적 데이터라 볼 수 없으며, 예측할 수 없고 같은 값이 나오지 않는 확률적 데이터의 형태에 가깝다. 이러한 특성은 앞서 빈도분석 결과에서도 확인할 수 있으며 데이터의 상대적 비율을 고려하여 오차율을 산정하는 방법이 데이터 특성에 적합한 평가방법이다. 따라서 산정식 개발은 비율과 가중치에 기초를 두고 있으며, 개별데이터 각각의 중요도, 영향도(빈도) 등에 따라 가중치를 적용하여 구하는 평균의 방식으로 계산식이 복잡하게 보이지만 극단적 데이터의 영향을 덜 받게 되며, 비율의 평균법을 사용함으로써 퍼센트 오차보다 합리적인 방법이라 할 수 있다. 퍼센트 오차는 표본의 중요도를 같게 보지만, 개발 산정식은 중요도를 구분하여 가중치가 고려된 것으로 더욱 정확한 평가결과를 산정할 수 있다. 기존 성능평가는 주로 회귀계열 퍼센트 오차 산정식으로 모든 데이터와 클래스가 같은 것으로 보고 상대적 중요성을 고려하지 않고 정확도를 산출하는 방법으로 딥러닝 성능지표 특성에 대한 오차를 산정할 수 없다. 반면, 분류계열 평가방법을 이용하면 각각의 Class에 대한 분석이 가능하며, Micro average와 Macro average, Weighted average 등의 방법이 있다. 이 중 Weighted average는 각 Class에 해당하는 data의 개수에 가중치를 주어 평균을 구하는 것으로 최종 계산이 이루어지기 전에 Class에 해당하는 Data에 가중치를 부여하여 계산하므로 퍼센트오차보다 정확한 계산 값을 산정한다.

평균값에 차이가 발생하는 이유는 집단을 구성하고 있는 항목들의비율과 집단의 크기가 다르다는 사실을 무시한 결과로 집단의 크기를 고려한 산정식 개발을 통해 이러한 오류를 해결하고자 한다. 평균에 대한 오류는 데이터의 세부 그룹별로 일정한 추세나 경향성이 나타나지만, 전체적으로 보면 그 추세가 사라지거나 반대 방향의 경향성을 나타내는 현상을 의미한다고 하였다(Simson, 1951). 전체집단에서 한 그룹이 차지하는 비중에 따라 집단 각각의 평균은 A 집단이 높았으나 비중을 고려하지 않고 평균을 계산하였을 때 B 집단이 높게 나오는 경우를 Simpson's Paradox(심슨의 역설)이라고 정의하였다(Colin, 1972).

빈도분석 결과에서도 딥러닝 Raw 데이터는 Class와 구간에 따라 다른 비율로 데이터가 분포함을 알 수 있었다. 즉, 상대적 기준과 데이터의 집중도와 중요도를 고려할 필요성이 있다고 결론 내릴 수 있다. 따라서, 딥러닝을 이용한 영상분석 오차율 산정은 집단의 크기가 고려대상이 되어야 하며, 평가지표도 분류계열 성능지표를 고려하여 정밀도(Precision), 재현율(Recall), IOU의 데이터 분석이 가능한 오차 산정식과 세분화 평가지표를 개발하고자 한다. 문헌 고찰을 통해 알아본 분류계열 오차 산정식 중 Weighted average는 산정식은 다음과 같다.

$$Sensitivity = \sum_{i=1}^n p(i) Sensitivity_i = \sum_{i=1}^n p(i) \frac{TP_i}{TP_i + FN_i} \dots\dots\dots (1)$$

Weighted average는 Macro average의 확장으로 그냥 평균을 계산하는 것이 아니라 각 Class에 해당하는 Data의 개수에 가중치를 주어 평균을 구한 것으로 전체 Data 중 Class i에 속하는 Data 개수의 비율을 p(i)로 정의하여 평균을 구하는 방식으로 연구에서 수행하고자 하는 개별가중치와 Class를 구별한 오차율 산정식 개발에 대한 이론적 배경은 확보하였다. 따라서, 최종 산정식 개발은 우선순위 배정과 가중치, 전체 데이터 중 Class i에 속하는 데이터 개수의 비율을 고려하여 오류를 편향시키는 항목에 비율을 통한 가중치를 부여함으로써 클래스 분류와 단계별 구분에 의한 오차율 분석이 가능한 산정식을 개발하고자 한다. 데이터 수의 비율을 고려한 가중치를 이용했을 때 통계 결과는 가중치가 낮은 데이터는 계산 결과에 큰 영향을 주지 않고 가중치가 높은 데이터는 상대적으로 더 큰 영향을 주게 되어 합리적인 오차 분석결과를 제시할 수 있다.

단, 가중치 부여를 주관적 방식에 따라 임의 부여하는 경우 이상치에 민감할 수 있으므로 각 Class의 데이터 수 및 구간별 데이터 수를 고려하여 비율 배정에 의한 가중치를 부여하고자 한다.

첫 번째 단계는 개별데이터 가중치를 계산하는 산정식으로 Equation 2과 같다.

$$W_{(iw)} = D_{ra} \times D_{i,p,r} = \frac{D_{i,p,r}}{D_T} \times D_{i,p,r} \quad \dots\dots\dots (2)$$

여기서,  $W_{(iw)}$  : 개별 데이터 가중치,  $D_{ra}$  : 데이터 비율  
 $D_{i,p,r}$  : IoU, Precision, Recall의 개별데이터 수,  $D_T$  :  $D_{i,p,r}$ 의 전체 데이터 수

개별데이터 가중치(Individual data weight)는 데이터 성능지표별 관측 데이터를 전체 데이터 수로 나눈 비율에 개별 관측 데이터를 곱하여 산정한다. Di,p,r의 경우, 중복 값이면 중복된 수를 개별데이터 수로 정의한다. 예를 들면, 첫 번째 승용차의 Recall 값이 0.667, 두 번째와 세 번째 Recall 값이 0.667인 경우 Di,p,r은 3이 된다.

두 번째 단계는 구간별 가중치를 산정하는 것으로 식은 Equation 3와 같다.

$$W_{(sw)} = \sum_{i=1}^n (W_{iw}) \quad \dots\dots\dots (3)$$

여기서,  $W_{(sw)}$  : 구간별 가중치

구간별 가중치(Section weight average)는 성능지표별 구간 정의에 따라 개별데이터 가중치를 합한 것이다. 세 번째는 개별데이터에 대한 가중오차를 산정하는 것으로 식은 Equation 4와 같으며,

$$W_{(iwe)} = \frac{W_{(iw)}(D_a - D_f)}{W_{(iw)} \times D_f} \quad \dots\dots\dots (4)$$

여기서,  $W_{(iwe)}$  : 개별 데이터 가중오차,  $D_a$  : 기준값,  $D_f$  : 관측값

개별데이터 가중 오차(Individual data weight error)는 기준값-개별데이터 측정값에 개별데이터 가중치를 곱하여 준 뒤 이를 개별데이터 가중치와 개별데이터 측정값의 곱으로 나누어 구한다.

마지막으로 구간 데이터 오차를 산정하는 식은 Equation 5와 같다.

$$W_{(se)} = \sum_{i=1}^n (W_{(iwe)}) / S_{dn} \quad \dots\dots\dots (5)$$

여기서,  $W_{se}$  : 구간 데이터오차  
 $S_{dn}$  : 구간 데이터 수

구간 데이터 오차(Section data error average)는 개별데이터 가중오차의 합을 구간 데이터 수로 나누어 구한다.

<Table 10>는 구간데이터 오차 산정 기준을 정하기 위해 분류계열 성능지표를 임계값 기준에 따라 구별한 것으로 IoU와 Recall은 7단계로, Precision은 4단계로 정의하였다.

<Table 10> Section definition by evaluation index

Section	Error	IoU	Precision	Recall
①	~ 35	IoU score $\leq 0.5$		Recall score $\leq 0.5$
②	~ 30	$0.5 < \text{IoU score} \leq 0.6$		$0.5 < \text{Recall score} \leq 0.6$
③	~ 25	$0.6 < \text{IoU score} \leq 0.7$		$0.6 < \text{Recall score} \leq 0.7$
④	~ 20	$0.7 < \text{IoU score} \leq 0.8$	$0.7 < \text{Precision score} \leq 0.8$	$0.7 < \text{Recall score} \leq 0.8$
⑤	~ 15	$0.8 < \text{IoU score} \leq 0.9$	$0.8 < \text{Precision score} \leq 0.9$	$0.8 < \text{Recall score} \leq 0.9$
⑥	~ 10	$0.9 < \text{IoU score} \leq 0.95$	$0.9 < \text{Precision score} \leq 0.95$	$0.9 < \text{Recall score} \leq 0.95$
⑦	~ 5	$0.95 < \text{IoU score} \leq 1.0$	$0.95 < \text{Precision score} \leq 1.0$	$0.95 < \text{Recall score} \leq 1.0$

## V. 적용분석 및 평가

### 1. 가중치 산정 결과

<Table 11>는 Equation 2와 Equation 3의 산정식을 사용하여 계산한 결과로 측정값에 대해 데이터가 존재하는 구간의 성능지표별 가중치 결과를 보여준다. 결과표를 해석하면 구간별 가중치는 ⑦구간이 105.28과 18.34로 가장 높았으며, 가장 낮은 구간은 ①구간으로 0.004로 나타났다.

<Table 11> Weight section

		Forecast value 1				Forecast value 2			
		Passenger car	Bus	Truck	Motorcycle	Passenger car	Bus	Truck	Motorcycle
①	IoU	0.004							
	Precision	0.004							
	Recall	0.02				0.004			
②	IoU	0.04	0.33	0.11		0.03		0.09	0.5
	Precision								
	Recall	0.02				0.01			
③	IoU	0.14	0.67	0.11	0.50	0.13	0.33	0.24	
	Precision							0.02	
	Recall	0.13	0.33	0.02		0.04		0.04	
④	IoU	0.28		0.28		0.23	0.33	0.31	0.5
	Precision	0.04		0.32		0.03		0.04	
	Recall	0.83	0.33	0.12		0.24	0.33	0.06	
⑤	IoU	0.57		0.32		0.40	0.33	0.22	
	Precision	0.33	1.00	0.30		0.46		0.53	
	Recall	1.23	0.33	0.19	1.00	0.85	0.33	0.62	1
⑥	IoU	0.27		0.19		0.22		0.11	
	Precision	0.60		1.04	1.00	1.54	0.33	0.76	
	Recall	0.24		0.02		0.86		0.58	
⑦	IoU	5.47		0.67	0.50	20.46		1.82	
	Precision	104.25		9.28		15.53	1.33	1.13	2
	Recall	46.30				5.25	0.33	1.02	

## 2. 오차율 산정결과

성능지표별 구간에 따른 오차율 계산 결과는 <Table 11>와 같다. 오차는 측정값 1의 경우, 구간①에서 32.58로 가장 크게 나타났으며, 구간⑦에서 0.32로 가장 낮게 나타난 것을 알 수 있다. 가중치는 구간①이 가장 낮았지만 반대로 오차율은 가장 높았으며, 데이터 비율은 ⑤와⑥ 구간이 가장 높았으나 오차율은 세 번째로 낮게 나타나 가중치를 적용한 오차 결과에서는 데이터 비율과 오차율이 상대적 중요도에 따라 정(+)의 관계를 갖지 않는다는 것을 알 수 있다. 가중치를 고려한 오차율 산정결과에서는 구간⑦이 가장 낮게 나타난 것을 알 수 있다. 측정값 2의 경우 오차율은 구간 ②와③에서 8.30과 7.13으로 가장 크게 나타났으며, 데이터 비율은 ⑤와⑦이 가장 높았으나 오차율은 구간⑦이 0.59로 가장 낮게 나타났다. ‘차종별+지표별’ 오차율은 측정값 1에서 구간①의 승용 Recall이 37.46으로 가장 높았으며, 구간⑦의 Recall 승용이 0.09로 가장 낮았다.

‘지표별+차종별 Avg error’는 측정값 2의 이륜의 Recall이 12.5로 가장 높았으며, 차종 비율이 높은 승용의 경우 측정값 1의 Precision이 11.07로 나타났다. Total vehicle avg error에서는 측정값 2의 IoU가 6.84로 가장 높았으며 Precision은 1.97로 가장 낮은 오차율을 보였다. 측정값 1과 2는 동일영상에서 반복 측정시 발생하는 오차율 결과로 딥러닝을 이용한 영상분석은 기준값이 존재하는 구조화된 영상에서 개별 차종에 대해 성능지표별 정량적 측정 수치 결과를 제공하지만, 반복측정 시 같은 수치 결과를 장담할 수 없다.

본 연구의 개발 산정식이 평가식으로서 일관성이 유지되는가는 ‘지표별+Total vehicle avg error’의 결과를 보면 알 수 있다. 측정값 1의 오차율은 5.34이며 정확도는 94.66%로 산출되었고, 측정값 2의 오차율은 5.07로 정확도는 94.94%로 산출되어 ‘측정값 1의 정확도-측정값 2의 정확도= 0.28’로 오차율 산정에 있어 큰 차이가 발생하지 않는 것을 알 수 있다. 또한, 본 연구에 사용된 시스템은 반복측정 시 오차율 변화가 거의 없는 정밀도가 높은 성능을 가진 시스템이라 할 수 있다. 이처럼 개발 산정 식은 기존평가방법으로는 해석할 수 없었던 부분들이 개선되었다. 첫 번째로 Threshold 기준값에 따른 구간별 평균 오차율 산정이 가능했으며, 두 번째 ‘차종 구분에 따른 구간별 오차율 산정’이 가능하여 동일 구간 내 차종에 따른 IoU와 Precision, Recall의 관계 해석이 가능하였다. 세 번째는 지표별 차종에 따른 오차율을 분석할 수 있었으며, 네 번째는 지표에 따라 차종별 전체 오차율을 구분하여 분석할 수 있는 장점이 있어 같은 비중을 부여하여 해석하는 PE 기반 분석결과보다 상세한 분석이 가능함을 확인하였다. 이러한 구체적인 결과와 해석은 기존평가 기준에서는 확인할 수 없는 부분으로 성능평가 또는 성능 기준에 따른 등급 부여로 ‘딥러닝 지표분류에 의한 평가방법이 필요하다’는 연구의 근거를 확보하였다.

<Table 12> Section error(%)

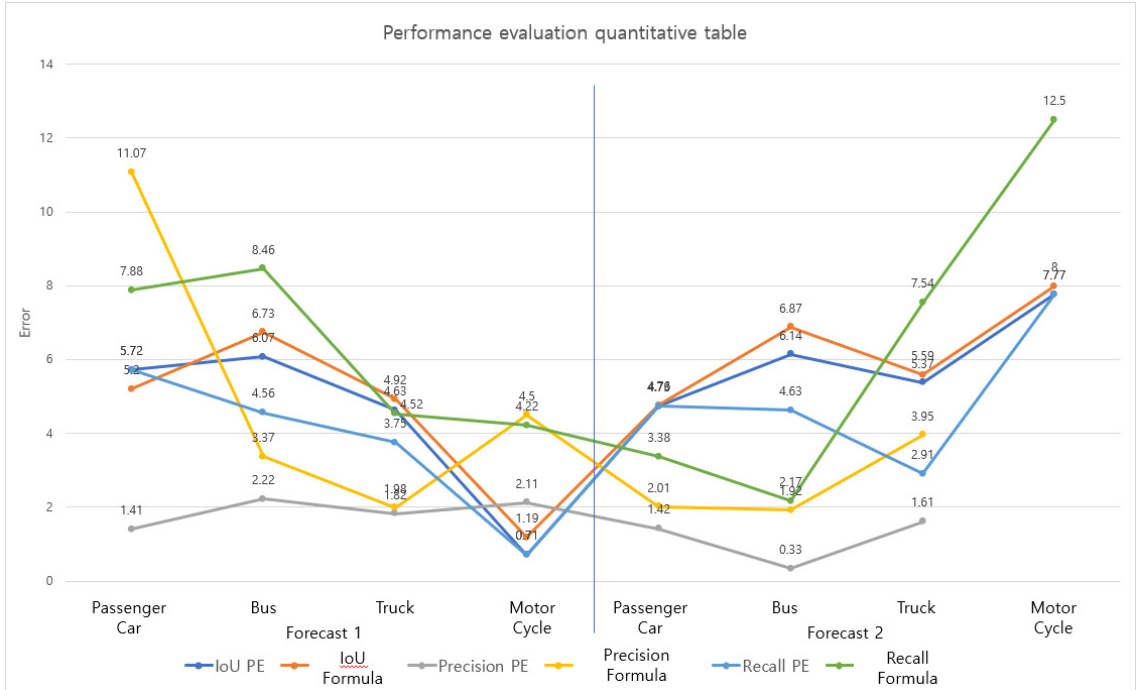
		Forecast value 1						Forecast value 2					
		Passenger car	Bus	Truck	Motor cycle	Avg	Total avg	Passenger car	Bus	Truck	Motor cycle	Avg	Total avg
①	IoU	10.28				32.58							
	Precision	50											
	Recall	37.46											
②	IoU	7.79	5.68	7.25		8.37	11.44		12.60		8.30		
	Precision												
	Recall	8.94						5.16					
③	IoU	6.63	7.77	10.24	1.19	5.39	5.98	2.61	6.27	4.04	7.13		
	Precision								8.89				

		Forecast value 1						Forecast value 2					
		Passenger car	Bus	Truck	Motor cycle	Avg	Total avg	Passenger car	Bus	Truck	Motor cycle	Avg	Total avg
	Recall	4.14	2.34	4.48				8.29		24.72			
④	IoU	5.36		3.70		3.01		4.48	11.68	2.77	11.96	3.71	
	Precision	2.29		2.78			4.07		4.55				
	Recall	1.39	9.01	4.12			2.56	0.77	5.60				
	IoU	3.68		4.92			4.50	6.31	4.51				
⑤	Precision	1.93	3.37	2.29		2.36		2.42		3.19		3.13	
	Recall	1.48	14.02	5.10	4.22		2.47		3.77	12.50			
	IoU	1.99		2.49			1.94		1.81				
⑥	Precision	0.88		0.88	4.50	1.51		0.91		1.75		1.25	
	Recall	1.67		4.40			0.92	3.57	1.60				
	IoU	0.66		0.93			0.25						
⑦	Precision	0.22				0.32		0.62	1.92	1.34		0.59	
	Recall	0.09					0.90		2.03				
	IoU avg error	5.20	6.73	4.92	1.19		4.51	4.76	9.00	5.59	8.00		
	Precision avg error	11.07	3.37	1.98	4.50	5.23	2.01	1.92	3.95			1.97	
	Recall avg error	7.88	8.46	4.52	4.22	6.27	3.38	2.17	7.54	12.50		6.40	
	Total vehicle avg error	8.05	6.18	3.81	3.30	5.34	3.38	4.36	5.69	6.83		5.07	

<Table 13> Comparison of the error rate between the PE and development formula

		Forecast value 1				Forecast value 2				Total
		Passenger car	Bus	Truck	Motorcycle	Passenger car	Bus	Truck	Motorcycle	
IoU	PE	5.72	6.07	4.63	0.71	4.73	6.14	5.37	7.77	
	W(SC)	5.20	6.73	4.92	1.19	4.76	9	5.59	8	
Precision	PE	1.41	2.22	1.82	2.11	1.42	0.33	1.61	0	
	W(SC)	11.07	3.37	1.98	4.5	2.01	1.92	3.95	0	
Recall	PE	5.72	4.56	3.75	0.71	4.73	4.63	2.91	7.77	
	W(SC)	7.88	8.46	4.52	4.22	3.38	2.17	7.54	12.5	
Avg	PE	3.99				3.54				7.53
	W(SC)	5.34				5.07				10.41
SE	PE	0.26								
	W(SC)	0.14								
SD	PE	0.32								
	W(SC)	0.19								

E



<Fig. 3> Performance evaluation quantitative table

<Table 13>은 PE 오차율에 근거하여 산출한 결과와 개발 산정식 오차를 비교한 것으로, 오차가 발생하지 않은 구간을 제외하면 모든 부분에서 차이가 있음을 알 수 있다. 전체 오차율 합은 PE 방식일 때 7.53, 개발 산정은 10.41로 산정식이 2.88 높게 나타났다. 그러나 반복측정에 대한 오차 정밀도는  $5.34-5.07=0.27$ , PE는  $3.99-3.54=0.45$ 로 산정식이 낮게 나타났다. 표준오차와 표준편차도 산정식이 PE 방식보다 작게 나타나 정밀한 오차율이 계산된 것을 알 수 있으며, 개발 산정식은 데이터 비율에 따른 가중치를 고려하여 이상치에 의한 영향을 최소화한 것으로 반복측정에 의한 오차 산정식으로 문제가 없음을 입증하였다.

평균 차이를 검증하기 위해 대응표본 t-검정을 통해 평균의 차이 검정을 수행하였다. t-검정을 수행하기 전 데이터 탐색과 정규성 검정을 통해 수행에 필요한 근거(모집단의 평균, 표본의 평균과 분산, 데이터 수, 정규성 및 신뢰도)는 수집자료 분석에서 확인하였다. 대응표본 t-검정을 선택한 이유는 사용하고자 하는 이유는 한 개의 모집단에서 2회의 결과값이 존재하며, PE 방식과 개발 산정식의 오차율은 서로 대응하기 때문에 사용하였다. 다시 말하면, 같은 데이터를 다른 산정식에 의해 반복한 계산 결과에 차이가 있는지 비교하기 위함이다. 대응표본 t-검정 결과  $t=2.217$ ,  $p=0.037$ 로 유의수준 0.05 기준으로 통계적으로 유의하게 나타나 기존 식과 산정식을 이용한 분석결과는 차이가 있음을 알 수 있다.

<Table 14> Performance evaluation quantitative table

	Descriptive statistics			t(p)	p
	N	평균(M)	Correlation Coefficient		
PE	24	3.62±2.34	0.612	-2.957	0.007**
W(SC)	24	5.11±3.08			

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$



<Table 15> 은 평가기준 범위에 따른 오차율을 성능지표와 측정값으로 구분하여 표기한 성능평가 결과표로, 오차율 구간은 5% 단위로 구분하여 총 6개로 세부 구분하였다. 오차율 구간에 따라 등급 적용이 가능하도록 평가 테이블을 구성하여 개발 산정식에 의한 오차율을 적용한 결과표이다.

<Table 15> Evaluation result table

Error rate(%)		Performance indicator								Standard error	Standard deviation	Pass
		Forecast value 1				Forecast value 2						
		Pc	B	T	M	Pc	B	T	M			
0~5	Superlative class			√	√	√				0.14	0.19	Accept (Senior class)
6~10	Senior class	√	√				√	√	√			
11~15	Middle class											
16~20	Lower middle class											
21~25	Low class											
26~30	Lowest class											
31~	Poor class											
Vehicle error		5.20	6.73	4.92	1.19	4.76	9	5.59	8			
Vehicle error avg		5.34				5.07						

## VI. 결론 및 향후 연구과제

본 연구는 딥러닝 기술을 이용한 스마트교차로 시스템의 성능평가방법을 개발한 것으로, 단순히 차종 분류 결과에 대한 퍼센트 오차를 산정하는 방식이 아닌 딥러닝 영상분석 통한 교통정보 수집의 핵심요소라 할 수 있는 Object Detection, Object Tracking, Action Classification 단계와 성능지표인 IOU, recall, precision의 스코어를 고려해 개별차량 분석값을 수집하였다. 또한, 기존 평가방식의 단점인 개별데이터의 중요도와 비율을 고려하지 않아 발생하는 오차를 개선하는 방법으로 분류계열 오차 산정방법인 Weighted average를 활용하여 개별차량 비율에 따른 가중치 부여와 임계값(Threshold)에 따라 달라지는 구간별 오차를 산정할 수 있는 산정식을 개발하였다. 개발 산정식을 통해 실제값과 예측값에서 발생하는 오차는 딥러닝 성능지표와 구간별 오차에 대해 분석할 수 있었으며 기존 평가방법보다 정밀하고 다양한 오차율 결과해석이 가능함을 확인하였다.

연구결과, 구간별 가중치가 가장 낮은 구간은 Threshold 기준값이  $\leq 0.5$  구간에 속하는 IOU ①구간으로 0.004였으며, 가장 높은 구간은 기준값이  $0.95 < \text{score} \leq 1.0$ 에 속하는 ⑦구간으로 105.28로 분석되었다. 오차율 분석결과 ‘측정값 1’에서는 가중치가 가장 낮은 구간①에서 32.58%로 가장 높았으며, 가중치가 가장 높은 구간⑦에서 0.32%로 가장 낮은 오차 결과를 보였다. ‘측정값 2’에서는 오차율이 가장 높은 구간은 ②와③으로 8.30%와 7.13%로 나타났지만, 가중치는 ⑤와⑦구간이 가장 높았다. 개발 산정식을 이용한 오차율 분석은 구간에 따른 성능지표와 오차율 관계, 구간별 데이터 가중치와 오차율 관계 등 딥러닝 특성을 반영한 정량적 결과해석이 가능하였다. 또한, 개발 산정식이 통계적으로 유의하고 기존 평가방식에 비해 정밀한 평가결과와 산정식으로 문제가 없음을 확인하기 위해 오차율 평균에 의한 비교와 표준오차, 표준편차를 통해 오차 범위를 확인하였다. 확인결과 차종별 오차 평균에 있어 산정식은 ‘측정값 1’에서 평균 오차율 5.34, ‘측정값 2’에

서 5.07로 0.27의 편차가 발생하였으며, 기존 평가식은 ‘측정값 1’에서 3.99 ‘측정값 2’에서 3.54로 0.45의 편차가 발생하여 개발 산정식이 오차 범위가 좁은 것으로 나타났다. 표준오차와 표준편차 결과도 산정식은 0.14와 0.19인 반면 기존 평가식은 0.26과 0.32로 산정식이 0.12와 0.13 작게 나타나 성능평가 산정식으로 문제가 없음을 검증하였다. 본 연구는 스마트교차로 특성에 맞는 평가방법을 정량적으로 제시함으로써 딥러닝을 활용한 교통정보수집 시스템의 다양화에 따른 성능평가에 유용하게 활용될 수 있을 것이다.

본 연구의 부족한 부분은 스마트교차로에서 수집 가능한 다양한 정보 중 속도정보와 관련하여 다루지 못한 점이다. 개별차량 속도정보는 차량의 운행행태와 관련한 것으로 도심지에서 일괄적용되고 있는 황색신호 시간의 문제점 분석과 개선방안을 도출하는 기초자료로 활용할 수 있으며, 차로 운영에 따른 방향별 개별차량 접근속도와 신호 운영 DB, 신호 위반 정보와 상관분석을 통해 딜레마 존 영향을 최소화할 수 있는 최적 황색신호시간을 결정에 관한 연구도 고려할 수 있다. 또한, 스마트교차로에서 수집 가능한 데이터 중 비정상적 차량 주행 행태인 역주행, 중앙선 침범, 보행자와 차량의 상충 위험 감지, 차로변경 금지 구간 내 이동행태 등의 연구를 통해 교차로별 잠재적 사고위험 지수와 관련한 기준 개발 시 교차로 운영 안정성 개선 도움이 될 것이다. 따라서, 향후 연구에서는 도로에서 발생하는 다양한 정보수집이 가능한 스마트교차로 시스템과 관련해 본 연구에서 다루지 못한 부분들의 다각적 연구가 필요할 것으로 판단된다.

## REFERENCES

- AI times, <https://www.aitimes.com>, 2023.06.14.
- Colin, R. B.(1972), “Is Pitman Closeness a Reasonable Criterion Comment”, *Journal of the American Statistical Association*, vol. 88, no. 421, pp.72-74.
- Curran, P. J., Stephen, G. W. and John, F. F.(1996), “The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis”, *American Psychological Association Psychological Methods*, vol. 1, no. 1, pp.16-29.
- Dreamsherpa, <https://m.blog.naver.com>, 2023.05.24.
- Jang, J. H., Baek, N. C. and Kang, W. U.(2004), “Study on Vehicle Performance Test Method”, *The Korea Institute of Intelligent Transport Systems Conference*, p.199.
- Kim, S. H., Jo, Y. J., Joung, Y. T. and Sin, K. W.(2018), “Deep learning-based smart intersection system evaluation”, *The 2018 Korean Institute of ITS Conference*, p.589.
- Kline, R. B.(2015), *Principles and practice of structural equation modeling*, Guilford Publications, NewYork, USA, p.534.
- Korea Institute of Civil Engineering and Building Technology(2006), *Research on practical use of ITS equipment/system performance evaluation and inspection/calibration system*, p.32.
- Kwon, C. H.(2004), “A Study On Camera-Detector System for Traffic Control in Korea”, *The Society of Digital Policy & Management 2004 Spring Conference*, pp.593-608.
- Lee, H. P.(2012), “Development of an Performance Evaluation Method for Vehicle Detector Speed Measurement Applying Uncertainty in Measurement”, *International Journal of Highway Engineering*, vol. 14, no. 2, pp.165-174.
- Oh, J. S.(2021), “A Study on the Performance Characteristics of Image Vehicle Detectors Depending on the Environment”, *The Korea Contents Association*, vol. 21, no. 1, pp.123-126.

- Park, H. L., Byun, S. H. and Lee, H. S.(2020), “Application of Deep Learning Method for Real-Time Traffic Analysis using UAV”, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 38, no. 4, pp.356-360.
- Park, Y. K.(2020), “Traffic Data Calculation Solution for Moving Vehicles using Vision Tracking”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 19, no. 5, pp.99-103.
- Simson, E. H.(1951), “The Interpretation of Interaction in Contingency Tables”, *Journal of the Royal Statistical Society Series B*, vol. 13, no. 2, pp.238-241.
- Snedecor, G. W. and Cochran, W. G.(1980), “Statistical Methods (7th ed.)”, *Iowa State University Press, Ames*, vol. 7, p.507.
- Sorower, M. S.(2010), “A literature survey on algorithms for multi-label learning”, *Oregon State University Corvallis*, vol. 18, no. 1, pp.3-8.
- Yaw, O. A.(2017), “Automated Vehicle Recognition with Deep Convolutional Neural Networks”, *Information Technology, Geospatial Information, and Advanced Computing*, vol. 2645, no. 1, pp.113-122.