

도시 가로경관 평가를 위한 로컬 AI 에이전트 구축 연구[†]

- 부산광역시를 대상으로 -

Development of a Local AI Agent for Urban Streetscape Evaluation[†]

- A Case Study of Busan Metropolitan City -

이청훈*, 심지수**

*부산대학교 조경학과 석사과정, **부산대학교 조경학과 조교수

Lee, Cheonghoon*, Sim, Jisoo**

*Master's Student, Dept. of Landscape Architecture, Pusan National University

**Assistant Professor, Dept. of Landscape Architecture, Pusan National University

Received: May 25, 2026

Revised: June 26, 2026 (1st)

August 13, 2026 (2nd)

Accepted: August 13, 2026

3인익명 심사필

Corresponding author :

Jisoo Sim

Assistant Professor, Dept. of

Landscape Architecture, Pusan

National University, Miryang

50463, Korea

Tel.: +82-55-350-5407

E-mail: jsoosim@pusan.ac.kr

국문초록

최근 구글 스트리트뷰를 통해 가로경관을 해석하고자 하는 시도가 증가하고 있다. 하지만 많은 양의 가로경관 데이터를 분석하고 해석하기 위해서는 프로그래밍 언어 등 전문성이 필요하다. 거대언어모델의 도입은 누구나 쉽게 접근할 수 있는 환경을 제공할 수 있으나 데이터 보안이나 환각 문제 등이 발생할 수 있다. 이에 이 연구는 Vision AI 기술을 활용하여 가로경관 데이터베이스를 구축하고, 이를 로컬환경에서 분석하여 데이터 보안과 답변의 신뢰성을 높인 가로경관 평가 에이전트 구현을 위한 기초 연구이다. 먼저 부산광역시 도로망을 기준으로 평균 40m 간격으로 구글 스트리트뷰 이미지를 수집하였다. 수집한 이미지는 SegFormer-B0, YOLOv10, CLIP 모델을 통해 가로경관의 물리적 환경, 보행환경, 인지적 특성을 평가하고 데이터베이스를 구축하였다. 다음으로 Qwen3-14b 언어모델에 양자화를 적용해 가로경관 평가 에이전트를 설계하였다. 분석 결과, 가로경관 평가 에이전트는 이용자가 질문하면 데이터베이스를 기반으로 답변을 도출하였으며, 간단한 답변뿐만 아니라 지역 간 비교, 변수 간 상관관계 분석과 같은 심층 분석을 통한 답변도 할 수 있었다. 이를 통해 외부 서버와 데이터를 공유하지 않는 독립적인 환경 구축과 데이터베이스 기반 응답 구조가 가로경관 분석의 보안성과 신뢰성을 높일 수 있음을 확인하였다. 이 연구는 전문적인 기술이 없는 이용자도 가로경관 데이터를 탐색하고 분석할 수 있는 환경을 제시하였으며, 조정 및 경관계획 분야의 데이터 기반 의사결정을 지원할 수 있는 기반을 마련했다는 점에서 의의가 있다.

주제어: 가로경관 분석, 비전 AI, 거대언어모델, ReAct 프레임워크

ABSTRACT

Recently, attempts to analyze streetscapes using Google Street View have increased. However, analyzing and interpreting large-scale streetscape data generally requires expertise in programming language and data analysis. Although the adoption of large language models (LLMs) has the potential to provide environmental data access to a wide range of users, concerns remain regarding data security and hallucination. Therefore, this study presents a preliminary framework for implementing a streetscape evaluation agent that enhances data security and response reliability by constructing a Vision AI-based streetscape database and performing analyses in a local environment. Google Street View images were collected at approximately 40 m intervals along the road network of Busan Metropolitan City. The collected images were analyzed using SegFormer-B0, YOLOv10, and CLIP models to evaluate the physical environment, pedestrian environment, and perceptual characteristics of streetscapes, and the results were stored in a structured database. For the agent, a quantized Qwen3-14B model was deployed locally and integrated into the evaluation pipeline. The results showed that the proposed agent generated responses based on the constructed database when users submitted queries. Beyond providing simple answers, the agent was also capable of performing in-depth analyses, including regional comparisons and correlation analyses among streetscape variables. These findings demonstrate that a standalone local environment without external data sharing, combined with a database-driven response architecture, can improve the security and reliability of streetscape analysis. This study provides an

[†]이 논문은 2024년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2024S1A5A8021160).

environment in which users without specialized technical expertise can explore and analyze streetscape data and establishes a foundation for supporting data-driven decision-making in landscape architecture and landscape planning.

Keywords: Streetscape Analysis, Vision AI, Large Language Model, ReAct Framework

1. 서론

가로경관은 시민들이 일상에서 접하는 도시환경이자 도시의 이미지를 형성하는 중요한 공간이다. 조정계획과 경관계획에서는 가로공간의 물리적 환경을 어떻게 조성할 것인지뿐만 아니라, 이러한 환경이 이용자의 인식과 행동, 심리적 건강에 미치는 영향을 고려하는 것이 중요한 과제로 다루어지고 있다. 실제로 잘 조성된 가로경관은 우울감 완화와 감정 조절뿐만 아니라 공중보건 향상에도 긍정적인 영향을 미치는 것으로 보고되고 있다(김연우 등, 2025; Huang et al., 2025). 최근 이러한 가로경관에 대한 관심에 따라 이를 객관적으로 평가하기 위한 여러 가지 기법이 제안되어 왔다. 최근에는 지도 서비스와 기술의 발전으로 가로경관 데이터를 정량적으로 분석하는 연구가 활발하게 이루어져 있다. 특히 AI 기술의 발달로 방대한 양의 가로경관 이미지를 자동으로 분석하고 정량화할 수 있게 되었다(Biljecki and Ito, 2021). 가로경관 이미지에서 녹지와 같은 경관의 구성요소를 Vision AI 기술인 의미론적 분할(Semantic Segmentation) 등을 통해 자세하게 분석하는 것이 가능해졌다(Biljecki and Ito, 2021). 이를 통해 녹지율(Yang et al., 2009; Aikoh et al., 2023), 천공률(Oke, 1981) 등 가로경관 지표를 정량화하고 분석하는 연구가 축적되고 있다(이다연 등, 2024). 경관의 구성요소뿐만 아니라 가로경관에 대한 이용자의 주관적 인식도 분석하는 연구도 시도되고 있다. 유승재 등(2021)은 구글스트리트뷰(Google Street View, 이하 GSV)를 활용해 가로경관에 대한 주관적 인식을 분석할 수 있음을 보였다.

국내에서도 GSV와 같은 가로경관 이미지는 녹지 형평성이나 보행환경을 평가하는 데 활용되고 있다(기동환 등, 2021; 이다연 등, 2024). 그러나 가로경관 데이터를 실제 의사결정에 활용하는 데에는 한계가 있다. 기존 가로경관 관련 연구는 분석결과를 주로 지도와 그래프 형태로 제시하고 있으며 이용자가 이를 활용하기 위해서는 필요한 데이터를 직접 분석하고 지표의 의미를 해석해야 한다. 예를 들어, “가로환경이 쾌적한 곳은 어디인가?”와 같은 의문이 있어도, 기존의 데이터와 결과만으로 답을 도출하기 어렵다. 결국, 가로경관 데이터를 활용하기 위해서는 이용자의 질문을 데이터 분석과 해석으로 연결해 주는 에이전트 도입이 필요하다.

문제 해결 방안으로 ChatGPT, Gemini 등과 같은 범용 거대언어모델(large language models, LLMs)을 사용하는 방법도 있다. 하지만 범용 거대언어모델을 조정 및 경관 계획에 적용하기에는 몇 가지 문제점이 존재한다. 먼저, 데이터 보안 문제이다. 중요한 데이터를 외부 서버에 올리는 것은 보안 문제가 발생할 수 있다(우상근, 2023). 다음으로 범용 거대언어모델은 최신 데이터까지 학습하지 않았기 때문에, 사실이 아닌 정보를 왜곡하여서 전달하는 환각(hallucination) 문제를 야기시켜 데이터의 신뢰성 문제를 일으킨다(Ji et al., 2023). 또한 초기 단계의 거대언어모델은 언어 이해와 텍스트 생성을 중심으로 활용되어, 도시 데이터를 처리하는 데 한계가 있었다(Liu et al., 2025). 최근 김이정과 이수기(2025)가 멀티모달 모델을 활용해 가로경관의 식별 요인을 분석한 연구가 있다. 이와 달리 이 연구는 가로경관 데이터와 AI 에이전트를 연계했다는 점에서 차별성을 갖는다.

따라서 이 연구는 Vision AI 기술을 바탕으로 가로경관 데이터베이스를 구축하여, 보안 및 신뢰성이 있는 가로경관 평가 에이전트를 만들고자 한다. 앞서 언급된 범용 거대언어모델의 문제점을 해결하기 위해 외부 서버 전송이 없는 로컬 환경에 구축한다. 또한 실제 가로경관 데이터베이스를 조회하고 연산하도록 구축한다. 구체적으로 1) 이용자의 질문을 데이터 연산으로 변환하고, 2) 해당 조건에 맞는 공간을 찾고, 3) 데이터를 바탕으로 답변하는 가로경관 평가 에이전트를 구축하는 데 목표로 둔다. 가로경관 평가 에이전트는 “쾌적한 녹지공간을 찾아줘”라는 질문을 “녹지율 수치를 내림차순 정렬해서 상위 지점을 추출”이라는 정량적 분석으로 변환할 수 있다. 이것은 가로경관 데이터를 분석하고 해석하는 과정에서 누구나 데이터 기반으로 공간을 평가할 수 있도록 한다. 무엇보다 거대언어모델이 기존에 학습된 지식에 의존하여 답변하는 것이 아닌 가로경관 데이터베이스를 바탕으로 답변하여 신뢰성을 확보할 수 있다는 점에서 기존 연구와 구별된다. 이 연구의 목적은 조정 및 경관계획에서 복잡한 분석 과정을 거치지 않고 누구나 쉽게 가로경관 데이터를 이용할 수 있는 환경을 만드는 데 있다. 이를 위해 이용자가 자연어로 질문을 하면 가로경관 정보를 분석 및 제고를 하는 시스템을 구축하였다. 이 시스템은 별도의 GIS나 프로그래밍 언어 등에 대한 전문성 없이도 실무자나 연구자에게 가로경관 정보를 제공한다. 이를 통해 이용자가 가로경관을 보다

깊이 이해하고 활용할 수 있는 가로경관 시스템이자 플랫폼을 제안하는 데 목적이 있다.

연구는 다음과 같이 구성된다. 2장에서는 가로경관의 구성요소, Vision AI, 거대언어모델에 관한 이론적 논의를 검토한다. 3장에서는 Vision AI를 활용한 가로경관 변수의 측정 및 가로경관 평가 에이전트 구축 방법을 설명한다. 4장에서는 가로경관 에이전트를 통한 평가 시나리오 결과를 보여주며, 마지막 5장에서는 연구의 결론 및 향후 과제를 논의한다.

2. 선행연구 고찰

2.1 가로경관의 구성요소

선행연구에서는 가로경관을 물리적 요소와 보행 및 이동 활동의 수준, 이용자의 인지 등으로 구분하여 평가해 왔다. 여기서 물리적 환경은 가로의 공간적 특성을 의미하고, 보행 및 이동 활동의 수준은 사람이나 이동수단으로 공간의 이용과 활력을 측정하는 변수를 의미한다. 이용자의 인지는 이용자가 해당 가로 경관을 이용하면서 느끼는 주관적 인식을 의미한다. Ewing and Handy(2009)는 가로경관을 물리적 특성과 인지적 특성으로 구분하여 평가해 왔다. 최근 연구에서는 Street View 이미지와 의미론적 분할을 활용하여 보행환경과 이용자의 인지를 정량적으로 평가하고, 이를 통해 주요 영향 요인을 해석하고 있다(Choi and Kang, 2025).

이 중 물리적 환경은 가로경관 이미지에서 가장 기초적인 요소이다. 가로경관의 물리적 요소에는 녹지, 건물, 도로, 보도, 하늘, 차량 및 도시시설물 등이 있으며, 이러한 요소들은 스트리트뷰 기반의 가로경관 분석에서 주요하게 활용되는 구성요소이다(Liu and Sevtsuk, 2024). 선행연구에서는 녹지 가로경관이 보행자에게 긍정적인 영향을 주는 지표이며 녹지가 방문객의 선호도에 중요한 영향을 준다고 하였다(양유선과 손용훈, 2019). Lu et al.(2018)은 GSV를 활용하여 가로녹지가 이용자의 보행 활동을 유도하는 요소임을 확인하였다. 가로수의 수관 폭, 수형 등과 같은 특성이 보행자 시점의 시각적 녹사율에 중요한 영향을 미친다(Zhu et al., 2024). 하늘, 수목, 건물의 가시 비율인 천공률, 수목률, 건물 비율 또한 가로경관을 나타내는 물리적 지표이다(Gong et al., 2018).

보행 및 이동 활동의 수준은 가로 공간의 활력을 나타내는 지표이다. De Nadai et al.(2016)은 이탈리아 6개 도시에서 도시 활력을 평가하면서 도시 활력에 중요한 요소가 보행자의 보행 활동임을 확인하였다. Ewing et al.(2016)은 가로경관의 물리적 특성과 보행자 수의 관계를 분석하여, 가로경관 특성이 보행 활동에 미치는 영향을 확인하였다. 보행자만이 아니라 자전거, 버스, 자동차, 오토바이 등 이동수단을 포함해 도시활력을 분석하는 연구도 이루어지고 있다(Doan et al., 2025). 또한 시계열 스트리트뷰 영상을 활용하여 보행자와 이동수단을 포함한 도로 교통을 분석하고, 이를 바탕으로 도시 활력을 정량적으로 평가하기도 한다(Nathvani et al., 2025).

인지는 가로경관이 사람에게 주는 감정이나 분위기를 의미한다. 잘 조성된 가로경관 요소는 시민들의 우울감 완화에 긍정적인 영향을 주는 것으로 나타난다(김연우 등, 2025). 또한, 가로경관의 개선이 시민들의 감정 및 공중보건에 긍정적인 효과를 준다(Huang et al., 2025). 최근에는 가로경관에 대한 이용자의 인지를 활기찬 분위기와 같은 주관적 인지 차원으로 정량화하는 연구가 수행되고 있다(Zhu et al., 2024). 이러한 인지 특성을 추론하기 위해 고안된 CLIP(contrastive language-image pre-training)¹⁾은 이미지와 텍스트를 동일한 의미 공간에서 학습하여 이미지의 의미적 특성을 이해하는 Vision AI로 이미지를 통해서 그 공간의 분위기를 파악하는 것을 돕는다.

종합하면 가로경관은 물리적 환경, 보행 및 이동 활동의 수준, 이용자의 인지 측면으로 구분하여 볼 수 있다. 물리적 환경은 가로공간의 시각적 특성을 정량화하여 분석하는 데 활용되며, 보행 및 이동 활동의 수준은 사람과 이동수단을 중심으로 도시의 활력을 평가하는 지표로 활용된다. 또한 이용자의 인지는 가로경관에 대한 주관적인 인식을 평가하는데 활용되고 있다. 그러나 선행연구들은 각 요소를 개별적으로 분석하는 데 초점을 두었으며 이를 종합하여 가로경관을 평가한 연구는 부족한 실정이다. 따라서 이 연구에서는 물리적 환경, 보행 및 이동 활동의 수준, 이용자의 인지를 종합적으로 고려하여 가로경관 분석 체계를 제안하고자 한다.

2.2 Vision AI 기반의 분석 기법

Vision AI²⁾는 인공지능이 시각적 데이터를 분석하는 분야로 이미지 내의 객체를 탐지하거나 영역을 나누는 기술을 포함한다. 가로경관 평가에서 사용되는 Vision AI의 기술은 크게 의미론적 분할(semantic segmentation), 객체 탐지(object detection), 멀티모달(multimodal) 분석으로 나눌 수 있다(Biljecki and Ito, 2021).

의미론적 분할은 이미지 내의 모든 픽셀을 녹지, 하늘, 도로 등과 같이 특정 클래스를 통해 분류하는 방법이다(그림 1 참조). 의미론적 분할을 통해 가로경관이 가진 물리적 환경을 픽셀 단위로 측정하고 정량화할 수 있다. Xie

et al.(2021)이 제안한 SegFormer 모델은 CNN 모델과 비교했을 때 도로 환경에서의 인식률을 개선했다. 이 기술을 바탕으로 유승재 등(2021), 이다연 등(2024)은 가로환경 요소를 측정하고, 보행 환경을 평가하는 연구를 진행했다. 객체 탐지는 사람, 자동차, 자전거와 같은 개별적으로 있는 객체를 탐지하여 객체의 위치랑 개수를 탐지할 수 있다. 김영우 등(2023)은 YOLO 객체 탐지 모델을 활용해 GSV 내의 보행자를 식별하고 분석하는 방법을 개발하였다. Nathvani et al.(2025)은 GSV에서 보행자를 측정하고 이를 통해 도시 활력을 평가하였다. 이들의 연구는 보행 인구 및 보행 수단 분석을 통해 보행 활동의 정도와 도시의 활력을 의미론적 분할과 객체 탐지로 측정할 수 있음을 시사한다.

멀티모달은 텍스트와 이미지를 연관 지어 학습시키는 기법을 의미한다. 이 기법은 기존의 의미론적 분할 및 객체 탐지를 넘어 인간이 이미지에서 느끼는 감정이나 분위기 같은 추상적인 정보를 정량화할 수 있다. Yuan et al.(2026)은 비전 및 언어 모델과 거대언어모델을 결합하여 가로경관이 주는 활기참과 같은 분석하는 방법을 제안하였다. 국내에서도 김이정과 이수기(2025)가 멀티모달 모델을 활용해 가로경관의 쇠퇴 요인을 분석한 연구가 있다. 이는 멀티모달 모델을 통해 조경 및 경관계획 분야의 가로경관 분석에 활용될 수 있음을 시사한다.

현재 Vision AI 기술은 이미지의 물리적 환경, 보행 활동의 정도, 이용자의 인지 특성까지 분석할 수 있다. 그러나 이러한 분석을 활용하기 위해서는 GIS, 프로그래밍, Vision AI 모델에 대한 전문성이 요구되어 실무자나 연구자가 활용하기에는 어려움이 있다. 따라서 분석결과를 쉽게 활용하기 위해서는 사용자 중심의 접근이 필요하다. 따라서 이 연구는 이러한 개별적인 Vision AI 기술을 통합하여 가로경관 데이터베이스를 구축하고, 이를 통해 도시 공간을 평가할 수 있는 에이전트를 구축하고자 한다.

2.3 거대언어모델과 AI 에이전트의 활용

거대언어모델은 많은 양의 텍스트 데이터를 학습하여 인간의 언어를 해석하고 생성할 수 있는 인공지능 모델을 의미한다(Brown, 2020). 초기에는 간단한 텍스트 완성이나 번역의 용도로 사용되었지만, 기술의 발전을 통해 복잡한 추론을 할 수 있을 정도로 발전하였다. 최근에는 스스로 문제를 확인하고 해결 방법을 찾고 분석 기능을 통해서 이용자의 요구를 수행하는 에이전트로 진화하고 있다(임준호, 2025).

공간계획 분야에서도 데이터 분석과 디자인 보조까지 거대언어모델의 활용 범위가 확대되고 있다. Xu et al.(2023)은 도시 문제를 해결하기 위해 거대언어모델을 활용한 플랫폼을 제안하고 언어모델이 도시 환경에서 자율적으로 의사결정을 수행할 가능성을 제시하였다. 박영호(2025)는 거대언어모델과 디자이너 간의 협업을 통해 디자인 분야에서 활용 가능성을 보여주었다. Liu et al.(2025)의 연구에서는 거대언어모델이 자연어를 이해하고 도시계획 업무를 지원하는 인터페이스로 활용될 수 있다고 설명하였다. 이를 통해 거대언어모델은 전문적인 기술이 없어도 가로경관 데이터에 기반한 의사결정에 도움을 주고, 조경 및 경관계획 정책을 수립하는 보조적 도구로 활용될

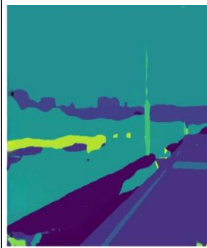
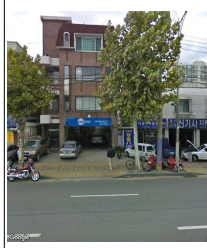
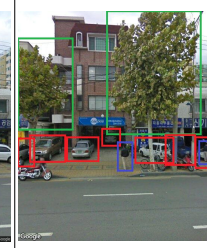
구분	의미론적 분할		객체 탐지	
핵심 질문	이 픽셀은 어떤 카테고리에 속하는가?		이미지 내에 어떤 객체가 어디에 몇 개 있는가?	
분석 단위	픽셀 단위		객체 단위	
출력 형태	각 픽셀에 클래스 부여		객체를 둘러싸는 경계 상자	
주요 특징	이미지 내 모든 영역을 분류, 같은 클래스의 객체들은 구분하지 않고 하나의 덩어리로 표시		목표 객체의 위치와 크기를 파악, 같은 클래스라도 개별 객체로 인식하여 각각의 경계 상자를 생성	
이미지 예시	원본 이미지	의미론적 분할 예시	원본 이미지	객체 탐지 예시
				

그림 1. 의미론적 분할과 객체 탐지 비교

수 있다. 하지만 실제로 활용하기 위해서는 몇 가지 문제점이 존재한다. 우선, 데이터 보안 문제이다. 도시 공간 데이터의 경우 비공개 데이터나 민감한 데이터를 포함하는 경우가 많다. 우상근(2023)은 외부 서버에 데이터를 전송해야 하므로 보안 문제가 발생할 수 있음을 지적하였다. 따라서 외부 서버와 독립된 로컬 시스템 구축이 필요하다. 다음으로 신뢰성 및 환각 문제이다. 거대언어모델은 실제 정보가 아닌 가짜 정보를 생성하는 환각 현상이 발생할 수 있다. 이를 방지하기 위해 데이터를 실시간으로 검색해 답변하는 기법 등이 연구되고 있다(김동국 등, 2025). 이 연구에서는 거대언어모델을 Python 분석 기능을 통해 데이터베이스를 조회하고 계산하도록 설계하여 신뢰성과 정확성을 확보하고자 한다.

종합하면 거대언어모델은 복잡한 데이터분석과 해석 과정을 이용자와의 질의와 답변으로 쉽게 만드는 기술이다. 그러나 거대언어모델과 관련된 기존 연구는 대부분 클라우드 기반 서비스를 활용하여 데이터 보안 문제에 노출되어 있으며, 거대언어모델의 환각으로 인해 분석 결과의 신뢰성을 확보하는 데 한계가 있다.

3. 연구방법

3.1 연구의 공간적 범위 및 GSV 이미지 수집

이 연구는 부산광역시를 공간적 범위로 한다. 부산은 바다와 산지, 신도시와 자연 발생한 주거지역이 공존하고 있어 다양한 가로경관 조건에서 로컬 AI 에이전트의 적용 가능성을 검증하기에 적합하다. 가로경관 수집을 위한 기준인 도로는 국토지리정보원에서 제공하는 도로중심선 데이터(2025년 8월 1일 기준)를 활용하였다. 해당 데이터는 전국의 도로중심선을 선형 공간정보로 구축한 자료로 도로의 총연장은 7,864km이며, 도로 중 폭 2m 이상의 도로를 대상으로 했으며 GSV 수집에 용이하지 않은 고속국도, 면리간도로, 소로는 제외하였다. 또한 도로길이 2m 이하의 도로 구간은 제외하였다. 이후 총연장 6,051km를 최종 분석 대상으로 선정하였다. 부산광역시 전역의 가로경관을 균등하게 반영하기 위해 행정구별 약 2,500개의 분석 지점을 선정하여 총 약 37,500개의 후보 지점을 도출하였다. 이후 각 분석 지점에서 보행자의 주변 시야를 반영하기 위해 수평 방향(heading)은 4방향(0° , 90° , 180° , 270°)으로 이미지를 수집하였고, 시야각(field of view, FOV)은 135° 로 설정하였다(그림 2 참조). 카메라의 상하 기울기(pitch)는 0° 로 설정하여 정면에서 바라보는 이미지를 수집하였다.

약 37,500개의 지점에서 총 150,000장의 이미지를 수집하였으며, 이미지 다운로드 오류(1~10KB 크기의 손상 파일) 및 중복으로 수집된 이미지를 제거하였다. 최종적으로 GSV 이미지가 없거나 오류가 발생하는 지점(총 333개)을 제외하고 이미지 수집이 가능한 37,167개의 지점을 대상으로 148,668장의 이미지를 수집하였다. Zhu et al.(2025)은 가로경관 분석에서 100m 간격의 표본 지점은 공간의 변화를 충분히 반영하기 어렵고, 50m 미만의 간격이 세밀한 분석에는 적합하다고 보았다. 이 연구의 최종 분석 지점 간 평균 간격은 약 40m로 선행연구에서 제시한 기준에 부합하고, 부산 지역의 가로경관 변화를 반영하기에 적절하다고 판단하였다(그림 3 참조).

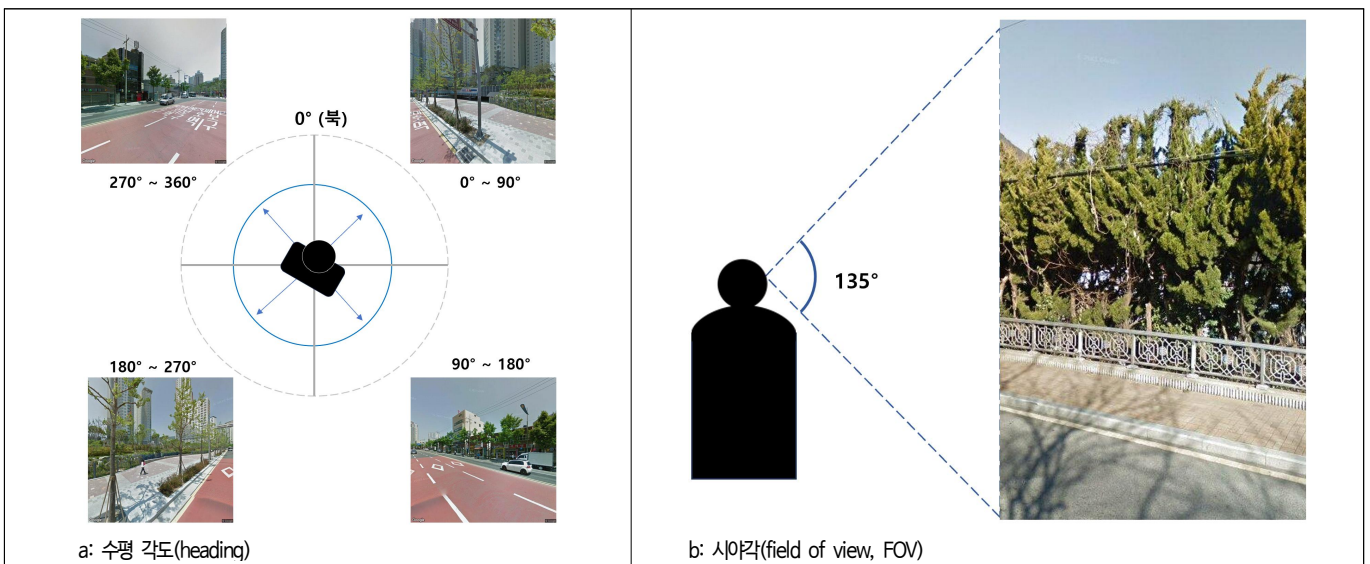


그림 2. 구글 스트리트뷰 수집 방향 예시

수집한 GSV 이미지는 Vision AI 기반 모델을 통해 가로경관 특성을 물리적 환경, 보행 및 이동활동의 수준, 인지적 특성으로 구분하여 측정하였다. 물리적 환경지표로는 녹사율, 천공률, 수목 수, 보행 및 이동활동 지표에서는 사람 수, 이동수단 수, 인지적 특성 지표로는 활기찬 분위기와 조용한 분위기 특성을 측정하였다. 이러한 변수들은 가로의 물리적 환경과 보행 및 이동 활동의 수준, 인지적 특성을 종합적으로 반영하기 위해 선정하였다. 각 이미지 별로 산출한 지표를 가로경관 데이터베이스로 구축하고, 이후 가로경관 평가 에이전트의 분석 자료로 활용하였다. 각 지표의 산출 과정 및 방법은 다음 절에서 설명한다.

이 연구의 전체 수행 절차는 다음과 같다. 1단계는 데이터 구축 단계로 도로망을 대상으로 분석 지점을 도출하고, 각 지점에서 구글스트리트뷰 API를 활용하여 가로경관 이미지를 수집하였다. 2단계에서는 수집한 이미지를 바탕으로 물리적 환경, 보행 및 이동활동 수준, 인지 항목을 정량화하여 지표를 통합하고 가로경관 데이터베이스를 구축하였다. 마지막 3단계에서는 가로경관 평가 에이전트를 구축하고, 사용자 자연어 질의에 대해서 데이터베이스를 바탕으로 데이터 조회 및 분석을 통해서 가로경관 평가 결과를 구할 수 있도록 구축하였다(그림 4 참조).

3.2 Vision AI 기반 변수 측정

분석을 위해 GSV를 대상으로 가로경관을 구성하는 요소 중 물리적 환경(녹사율, 천공률, 수목 수), 보행 및 이동 활동의 수준(수목 수, 이동수단 수), 인지 특성(활기찬 및 조용한 분위기)을 정량적으로 측정하였다. 이를 위해 각 요소에 적합한 Vision AI 모델을 선정하여 분석을 진행하였다. 각 이미지에서 녹사율, 천공률, 수목 수, 사람 수,



그림 3. 연구 대상지 및 구글 스트리트뷰 수집 지점

범례: ● 분석지점, ■ 도로현황

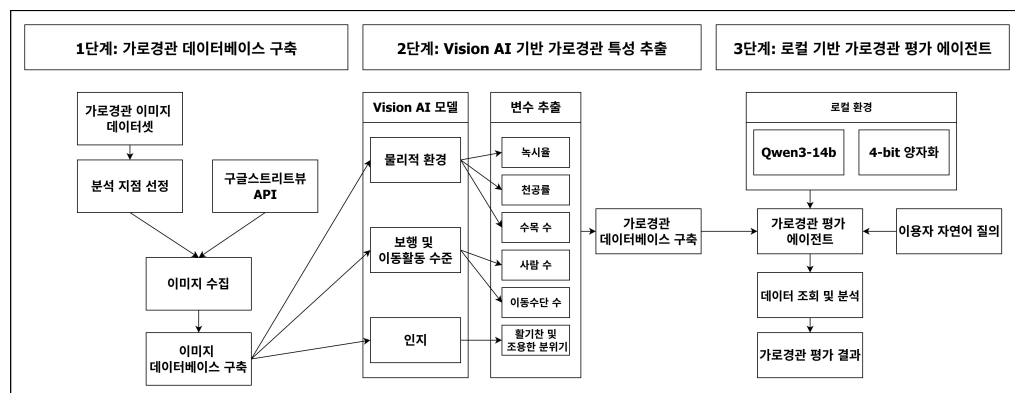


그림 4. 연구흐름도

이동수단 수, 활기찬 및 조용한 분위기 특성을 각각 산출한 후, 동일 지점에서 수집된 4장의 분석 결과를 평균하여 지점별 대표값을 생성하였다. 이를 통해 보행자가 한 지점에서 경험하는 주변 가로경관 특성을 종합적으로 반영하고자 하였다. 표 1은 이 연구에서 활용한 Vision AI 기반 분석 변수와 측정 방법을 나타낸다. 가로경관 데이터베이스 구축 과정의 구체적인 측정 방법은 다음과 같다.

가로 경관의 물리적 환경(그림 5 참조)을 정량적으로 분석하기 위해 의미론적 분할 모델인 SegFormer-B0를 적용하였다. 분석에는 ADE20K 데이터 세트로 사전 학습된 모델을 사용하였으며 모든 GSV 이미지는 512 × 512 픽셀의 해상도로 변환한 후 분석하였다. 수집된 GSV를 대상으로 이미지 내의 클래스별 픽셀 비율을 통해 녹지의 비율인 녹지율과 하늘의 비율인 천공률을 산출하였다. 그리고 객체 탐지 기법을 통해 이미지 내 수목의 개수를 산출하였다. 녹지율은 SegFormer-B0 모델에서 식생과 관련된 9개 클래스(tree, grass, plant, bush, shrub, flower, forest, hedge, vegetation)에 해당하는 픽셀을 합산하고, 이를 전체 이미지의 픽셀 수(P_{total})로 나누어 산출하였다. 산출식은 다음과 같다.

$$GVI = \frac{\sum_{i \in C_{veg}} P_i}{P_{total}} \times 100$$

(식 1)

여기서 GVI 는 녹지율, C_{veg} 는 식생으로 분류된 클래스의 집합, P_i 에서 해당 클래스 i 에 속하는 픽셀의 수, P_{total} 은 이미지 전체 픽셀 수를 의미한다.

표 1. 가로경관 평가를 위한 변수 설정

평가 요소	개념	사용 모델	측정 변수	내용
물리적 환경	가로의 물리적 공간 특성	SegFormer-B0	녹지율(Lu et al., 2018)	식생 픽셀의 비율
			천공률(Gong et al., 2018)	하늘 픽셀의 비율
			수목 수(Zhu et al., 2024)	수목 영역을 구분한 것의 수
보행 및 이동 활동의 수준	공간 이용 및 활력	YOLOv10	사람 수(Nathvani et al., 2025)	사람 객체의 수
			이동수단 수(Nathvani et al., 2025)	이동수단(car, bus, truck, motorcycle) 객체의 수
인지	이용자의 심리적 인식	CLIP	활기찬 및 조용한 분위기 (Ogawa et al., 2024)	해당 의미(활기찬, 조용한)의 비율

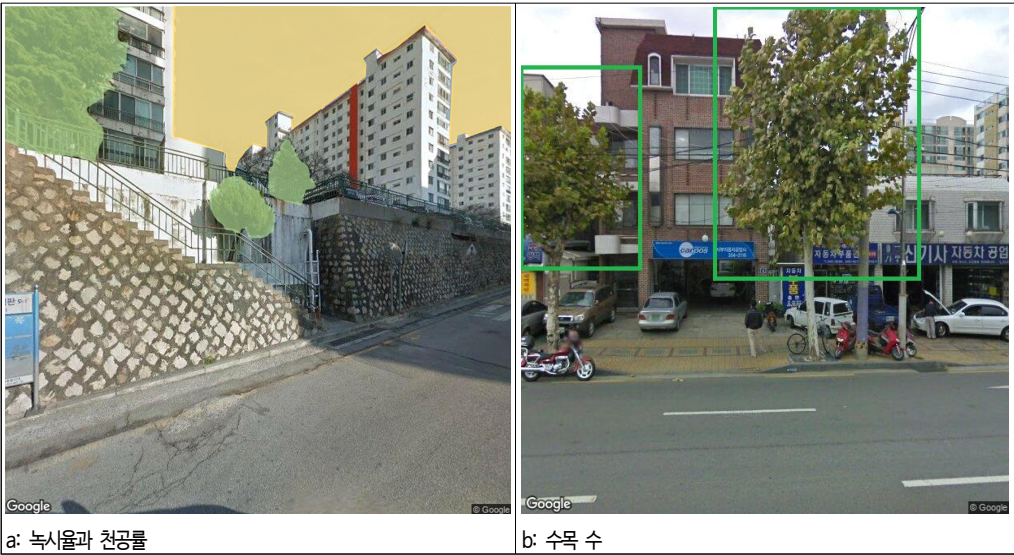


그림 5. 물리적 환경 측정 예시 이미지
범례: 녹지율, 천공률, 수목 수

다음으로 SegFormer-B0의 분할 결과에서 하늘(sky, P_{sky})에 해당하는 픽셀의 비율을 통해 천공률($SkyIndex$)을 산출하였다(식 2).

$$SkyIndex = \frac{P_{sky}}{P_{total}} \times 100 \quad (\text{식 2})$$

마지막으로 수목의 수는 해당 모델에 별도의 클래스가 없어 이를 산출하기 위해 별도의 알고리즘을 적용하였다(식 3). 우선 이미지에서 나무 클래스에 해당하는 이진 마스크(M_{tree})를 추출하였다. 그 후 3×3 커널을 이용한 형태적 닫힘 연산(morphological closing)을 수행하여 수목 영역 내부에 존재하는 미세한 노이즈를 교정하고 서로 근접한 영역을 연결하였다. 이후 연결 성분 분석(connected component analysis, CCA)을 적용하여 분리된 픽셀 영역을 개별 객체로 식별하였다. 이때 80 픽셀 미만의 객체는 노이즈로 판단하여 제외하였다. 최종적으로 독립된 객체로 식별되는 객체의 수를 수목의 수(N_{tree})로 산출하였다. 수목 수 산출 결과를 검증하기 위해 전체 이미지 중 100장을 무작위로 추출하고, 위의 알고리즘을 통해 산출한 값과 연구자가 직접 계수한 값을 비교하였다. 비교 결과, 알고리즘으로 산출한 값과 연구자가 집계한 값의 평균 차이는 1.62그루였으며, 피어슨 상관계수는 0.932, 결정계수(R^2)는 0.869로 나타나 비교적 높은 수준의 일치성을 확인하였다.

$$N_{tree} = Count(CCA(Closing(M_{tree}))) \quad (\text{식 3})$$

보행 및 이동 활동의 수준(그림 6 참조)을 정량화하기 위해서는 객체 탐지 모델인 YOLOv10n을 적용하였다. 이 이미지에서 관찰되는 보행자 수는 가로의 보행 및 이동 활동의 수준을 추정하는 지표로 활용하였다. 모델이 탐지한 객체 중 사람(person) 클래스(c_{person})에 해당하는 경계 박스의 수를 합산하여 이미지별 보행자 수를 다음과 같이 산출하였다(식 4).

$$N_{pop} = \sum_{k=1}^K I(c_{ls_k} = c_{person}) \quad (\text{식 4})$$

여기서 N_{pop} 은 이미지에서 식별된 보행자의 수, K 는 이미지 내에서 탐지된 전체 객체의 수를 의미하며, c_{ls_k} 는 k 번째로 탐지된 객체의 클래스, c_{person} 은 사람(person) 클래스를 의미한다. I 는 지시함수로, 탐지된 객체가 사람 클래스로 분류된 경우 1, 그렇지 않은 경우에 0의 값을 도출한다. 이를 통해 탐지된 모든 객체에서 지시함수의 값을 합산하여 이미지별 보행자 수를 산출하였다.

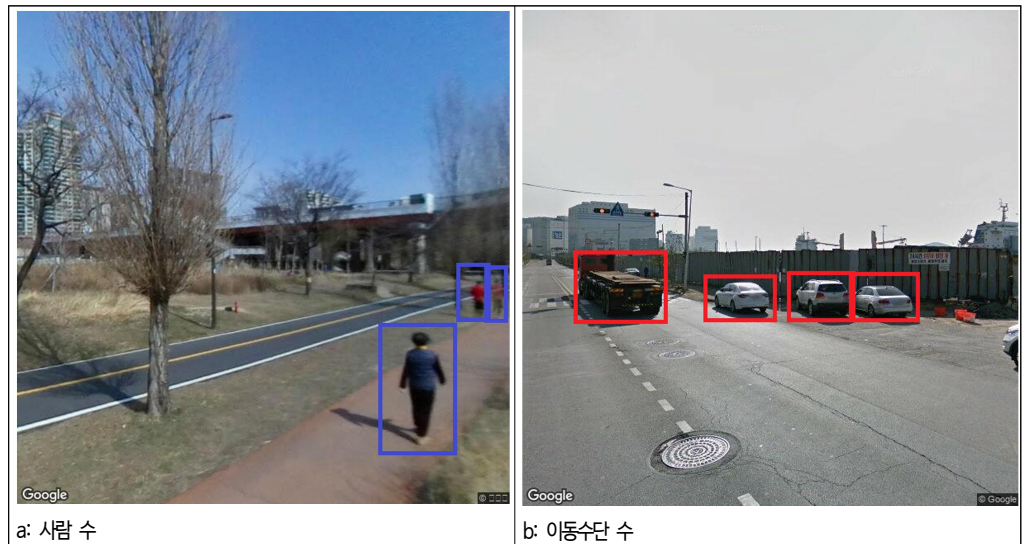


그림 6. 보행 및 이동 활동의 수준 측정 예시 이미지
범례: 사람 수, 이동수단 수

또한 가로에서 보행 외 이동 수단의 통행을 산출하기 위해 탐지된 객체 중 이동 수단에 해당하는 4개 클래스($c_{vehicle}$: car, bus, truck, motorcycle)에 해당하는 객체의 수를 합산하여 이미지별 이동 수단의 수를 추출하였다(식 5).

$$N_{veh} = \sum_{k=1}^K I(c_{cls_k} \in c_{vehicle}) \quad (\text{식 } 5)$$

여기서 N_{veh} 는 이미지에서 식별된 이동수단의 수, $c_{vehicle}$ 은 이동수단에 해당하는 클래스를 의미한다.

가로경관의 인지적 특성(그림 7 참조)을 정량화하기 위해 CLIP을 사용하였다. CLIP은 이미지와 텍스트를 동일한 의미 공간에 학습시켜서 두 모달 간의 유사성을 추론하는 멀티모달 모델이다(Radford et al., 2021). 이 연구에서는 가로경관의 인지적 특성을 서로 상반된 활기찬 및 조용한 분위기로 구분하고 각 이미지가 둘 중 어느 분위기와 더 높은 연관성을 갖는지를 평가하였다. 평가를 위해 설정한 프롬프트는 활기찬 분위기(T_{lively})를 의미하는 “Lively”와 조용한 분위기(T_{quiet})를 의미하는 “Quiet”이다. 여기서 입력된 이미지 I 의 특징 벡터(V_I)와 텍스트 프롬프트 T 의 특징 벡터(V_T) 간의 코사인 유사도($S(I, T)$)를 다음과 같이 산출하였다(식 6).

$$\sim(I, T) = \frac{V_I \times V_T}{\|V_I\| \|V_T\|} \quad (\text{식 } 6)$$

이후 두 프롬프트의 유사도 점수를 비교하고, 더 높은 유사도를 보인 인지를 해당 이미지의 인지적 분위기로 다음과 같이 분류하였다(식 7).

$$Label_I = \arg \max_{t \in T_{lively}, T_{quiet}} S(I, t) \quad (\text{식 } 7)$$

여기서 $Label_I$ 는 입력 이미지 I 의 최종 분위기를 의미한다. ($t \in T_{lively}, T_{quiet}$)는 활기찬 및 조용한 분위기 프롬프트 중 하나를 의미한다. argmax는 코사인 유사도가 가장 큰 프롬프트를 선택하는 연산을 의미한다.

이상의 과정을 통해 Vision AI의 분석 기법을 사용하여 각 GSV 이미지를 대상으로 가로경관의 구성요소를 정량화하였다. 구체적으로 물리적 환경으로는 녹지율, 천공률, 수목 수를 통해 측정하고, 가로경관에서 발생하는 보행 및 이동 활동의 수준은 보행자 수와 이동수단 수를 통해 파악하였다. 이후 CLIP을 활용하여 물리적 요소나 활동 수준으로 설명하기 어려운 가로의 인지적 분위기를 활기참과 조용함의 두 관점으로 분류하였다. 이러한 분석 결과는 이후 평가 에이전트를 구축하는 데, 기초 자료로 활용하였다.

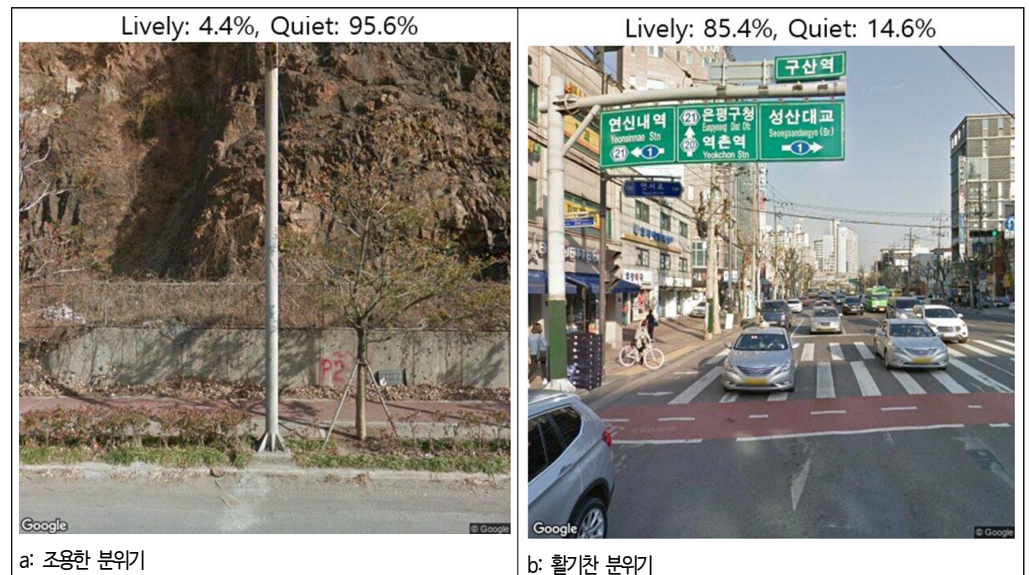


그림 7. 인지 측정 예시 이미지

3.3 가로경관 평가 에이전트 구축

연구의 3단계인 가로경관 평가 에이전트 구축 단계에서는 앞서 구축한 가로경관 평가 지표와 데이터베이스를 활용하여 이용자가 자연어 질의에 따라 가로경관 평가 정보를 확인할 수 있는 가로경관 평가 에이전트를 구축하였다. 특히 이 에이전트는 가로경관 데이터를 외부 서버로 전송하지 않고 분석할 수 있도록 로컬 실행 환경 기반의 가로경관 평가 에이전트를 구축하였다. 분석 환경은 GPU는 NVIDIA GeForce RTX 5070 Ti(VRAM 16GB), CPU는 AMD Ryzen 9 9950X, 메모리(RAM)는 64GB로 구성하였다.

운영체제는 WSL2(windows subsystem for Linux 2)를 활성화하여 Ubuntu 22.04 LTS 환경을 구성하였다. 여기서 WSL2는 Windows 운영체제에서 Linux를 실행할 수 있도록 지원하는 가상화 기술이며, Ubuntu 22.04 LTS는 장기간 안정적인 업데이트를 제공하는 Linux 운영체제이다. 이렇게 리눅스 환경을 구축하여 AI 연구가보다 원활하게 가능하게 하였다. 또한 Anaconda³⁾ 가상환경에서 개발 언어는 Python 3.11을 사용하여 가상환경을 구축하였다. 또한 동일한 실행 환경에서 분석을 재현할 수 있도록 하드웨어 사양, 운영체제, 개발 환경 및 라이브러리 버전을 고정하였다.

에이전트의 추론 모델로는 대규모 언어모델의 오픈소스 모델인 Qwen3-14B를 선정하였다. 해당 모델은 16-bit로 구동할 경우, 약 28GB의 메모리가 요구되어 VRAM을 16GB로 사용하고 있는 본 연구에서는 그대로 실행은 어렵다. 이에 이 연구는 모델의 가중치를 4-bit로 압축하는 NF4(4-bit NormalFloat) 양자화 기술을 적용하여 모델의 성능저하는 최소화하고 GPU 메모리 사용량은 줄여서 앞서 구축한 로컬 GPU 환경에서도 Qwen3-14B를 안정적으로 구동할 수 있었다.

구축한 가로경관 평가 에이전트가 가로경관 데이터베이스를 직접 조회하고 분석할 수 있도록 LangChain⁴⁾ 프레임워크를 적용하였다. 구체적으로 langchain-experimental 라이브러리의 create_pandas_dataframe_agent 모듈을 활용하였다. 이 모듈은 자연어로 입력된 사용자의 분석 요청을 Python 코드로 자동 변환하고, 해당 코드를 실행하여 데이터 조회, 통계 분석 및 시각화를 수행한 후 그 결과를 다시 자연어로 답변하는 모듈이다. 이를 통해 사용자는 Python 코드를 직접 작성하지 않고도 자연어 질의를 통해 가로경관 데이터베이스를 탐색하고 분석할 수 있다.

가로경관 평가 에이전트는 ReAct(reasoning and acting) 프레임워크를 기반으로 작동하도록 구성하였다. ReAct 프레임워크는 추론하는 과정에서 사고와 행동을 반복해서 수행하면서 추론하고 답변을 하게 하는 방식이다. 따라서 가로경관 평가 에이전트는 이용자의 질의가 입력되면 2단계인 사고부터 4단계인 관찰까지의 과정을 반복하고, 충분한 분석 결과가 도출되면 최종 답변을 생성한다(표 2 참조).

가로경관 평가 에이전트가 연구 대상과 지표의 의미를 일관적으로 해석하도록 프롬프트를 별도로 설계하고 적용하였다. 우선 에이전트를 '부산시 전문 도시 데이터 분석가'라는 역할을 부여하고 분석의 공간적 범위를 부산광역시로 한정하였다. 다음으로 데이터베이스에 있는 각 변수의 정의와 측정 단위, 값의 범위를 인식하도록 하였다. 마지막으로 부산시 내 지역 간 비교 질문 시에는 지역별 평균값을 우선 연산하고, 순위를 질문할 때는 정렬(sort_values) 함수를 우선 사용하여 값을 정렬하도록 Python 코드 생성 규칙을 설정하였다. 이를 통해 자연어 표현이 달라지더라도 유사한 질의에 일관된 분석 절차가 적용되도록 하였다.

구축한 에이전트의 활용성을 확인하기 위해 3단계의 분석 시나리오를 설정하고 가로경관 평가 에이전트가 서로 다른 유형의 질의를 수행할 수 있는지를 확인하였다. 시나리오 1인 기초 탐색 시나리오에서는 이용자의 질문을 데이터 연산으로 변환하는 능력을 확인하였다. 시나리오 2인 비교 분석 시나리오에서는 부산시 내 지역 간의 데이터를 비교하고 그 차이를 설명하는 능력을 검증하고 확인하였다. 시나리오 3인 심층 추론 시나리오에서는 변수 간의 상관관계를 심층적으로 분석하고, 분석 결과를 이용자의 질문을 검증하거나 가로경관의 개선 방안을 제시할 수 있는지 평가하였다.

표 2. 가로경관 평가 에이전트의 분석 절차

단계	과정	구체적 내용	예시
1단계	입력	이용자가 가로경관에 대한 분석 요구사항을 입력	부산에서 녹지가 가장 쾌적한 5곳 요청
2단계	사고	질의 분석하여 어떤 데이터를 조회하고 통계적 분석이 필요한지를 추론	녹사율을 정렬하여 상위 5개를 탐색
3단계	행동	분석의 실행을 위해 데이터베이스에 접근하는 Python 코드를 작성	녹사율을 기준 상위 5개를 추출하는 코드 실행
4단계	관찰	생성된 코드를 구동하여 실제 데이터 결과 확인	[O동: 45.2% ...] 등의 실제 데이터와 매칭
5단계	최종 답변	이용자가 이해하기 쉬운 문장으로 최종 답변	OO(45.2%)를 포함한 5개 지역이라 응답

4. 연구결과

4.1 가로경관 평가 에이전트의 단계별 성능 검증

연구에서는 가로경관 평가 에이전트가 이용자의 자연어 질의를 적절한 데이터 분석 과정으로 변환하고, 데이터 베이스를 기반으로 정확한 답변을 생성하는지를 검증하였다. 검증은 기초 탐색, 비교 분석, 심층 추론의 세 가지 시나리오로 구분하여 수행하였다. 각 시나리오는 서로 다른 유형의 질의 5개를 구성하였으며, 각 질의를 10회 반복 수행하여 각 시나리오별 각각 50회 총 150회 반복 실행 시 결과의 일관성을 평가하였다(표 3 참조). 정확도는 에이전트가 생성한 Python 분석 코드의 계산 결과를 데이터베이스의 실제 계산값과 비교하여 산정하였으며, 수치는 소수점 둘째 자리까지 일치하는 경우를 정답으로 판정하였다. 시나리오 1(기초 탐색)의 평균 정확도는 94%, 시나리오 2(비교 · 분석)의 평균 정확도는 90%, 시나리오 3(심층 추론)의 평균 정확도는 72%이다(표 4 참조).

4.2 기초 탐색 시나리오: 추상적인 질문을 구체적인 데이터로 변환

시나리오 1에서는 이용자의 자연어 질의를 데이터 조회 및 연산으로 변환하는 능력을 검증하였다. 총 5개의 질의를 대상으로 각각 10회 반복 수행한 결과 평균 정확도는 94%로 나타났다. 대표 사례로, 이용자가 “부산에서 녹지가 풍부하고 활기찬 곳을 찾아줘”라고 질문하면 에이전트는 녹지율과 활기참 지표를 기준으로 동별 평균을 계산한 후 최상위 지역을 추출하였다(그림 8 참조). 결과적으로 “부산에서 녹지가 풍부하고 활기찬 곳은 중구 영주동입니다.”라는 답변을 생성하였다. 이를 통해 이용자의 자연어 질의를 적절한 데이터 연산으로 변환하여 결과를 도출할 수 있음을 확인하였다.

4.3 비교 분석 시나리오: 지역의 차이 비교 및 설명

시나리오 2에서는 지역 간 데이터를 비교하고 설명하는 능력을 검증하였다. 총 5개의 비교 질의를 대상으로 각각 10회 반복 수행한 결과 평균 정확도는 90%로 나타났다. 대표 사례로, “부산 북구와 동래구의 평균 하늘 개방감

표 3. 가로경관 평가 에이전트 검증 기준

시나리오	입력 질의	생성된 분석 규칙	산출 결과	검증 기준
기초 탐색 시나리오	녹지율이 가장 높은 동은?	동별 평균 녹지율 계산 후 최댓값 분석	선동	• 각 질의별 10회 반복 실행 • 데이터의 계산 결과와 일치 여부
	나무 수가 가장 많은 동은?	동별 평균 나무 수 계산 후 최댓값 분석	오륜동	
	이동수단 수가 가장 많은 동은?	동별 평균 나무 수 계산 후 최댓값 분석	대창동2가	
	사람 수가 가장 많은 동은?	동별 평균 나무 수 계산 후 최댓값 분석	남포동3가	
	가장 활기찬 분위기를 가진 동은?	동별 평균 나무 수 계산 후 최댓값 분석	남포동3가	
비교 분석 시나리오	중구와 서구 중 천공률이 더 높은 곳은?	두 지역의 평균 천공률 비교	서구	• 각 질의별 10회 반복 실행 • 데이터의 계산 결과와 일치 여부
	동구와 영도구 중 나무 수가 더 많은 곳은?	두 지역의 평균 나무 수 비교	영도구	
	해운대구와 사하구 중 이동수단 수가 더 많은 곳은?	두 지역의 평균 이동수단 수 비교	사하구	
	금정구와 강서구 중 사람 수가 더 많은 곳은?	두 지역의 평균 사람 수 비교	금정구	
	연제구와 수영구 중 더 조용한 분위기를 가진 곳은?	두 지역의 평균 조용한 분위기 비교	수영구	
심층 추론 시나리오	녹지율과 수목 수의 상관관계	녹지율과 수목 수의 상관관계 분석	상관관계: 0.392	• 각 질의별 10회 반복 실행 • 상관계 계산 결과와 해석의 일치 여부
	녹지율과 천공률의 상관관계	녹지율과 천공률의 상관관계 분석	상관계: -0.336	
	이동수단과 활기참의 상관관계	이동수단과 활기참의 상관관계 분석	상관계: 0.471	
	천공률과 활기참의 상관관계	천공률과 활기참의 상관관계 분석	상관계: 0.051	
	사람 수와 이동수단 수의 상관관계	사람 수와 이동수단 수의 상관관계 분석	상관계: 0.128	

표 4. 가로경관 평가 에이전트 검증 결과(단위: %)

시나리오	검증 문항	정확도(참/반박 횟수)	오류율
기초 탐색 시나리오	녹시율이 가장 높은 동은?	90(9/10)	10
	나무 수가 가장 많은 동은?	100(10/10)	0
	이동수단 수가 가장 많은 동은?	100(10/10)	0
	사람 수가 가장 많은 동은?	100(10/10)	0
	가장 활기찬 분위기를 가진 동은?	80(8/10)	20
비교 분석 시나리오	중구와 서구 중 천공률이 더 높은 곳은?	80(8/10)	20
	동구와 영도구 중 나무 수가 더 많은 곳은?	100(10/10)	0
	해운대구와 사하구 중 이동수단 수가 더 많은 곳은?	100(10/10)	0
	금정구와 강서구 중 사람 수가 더 많은 곳은?	100(10/10)	0
	연제구와 수영구 중 더 조용한 분위기를 가진 곳은?	70(7/10)	30
심층 추론 시나리오	녹시율과 수목 수의 상관관계	80(8/10)	20
	녹시율과 천공률의 상관관계	70(7/10)	30
	이동수단과 활기참의 상관관계	60(6/10)	40
	천공률과 활기참의 상관관계	60(6/10)	40
	사람 수와 이동수단 수의 상관관계	90(9/10)	10

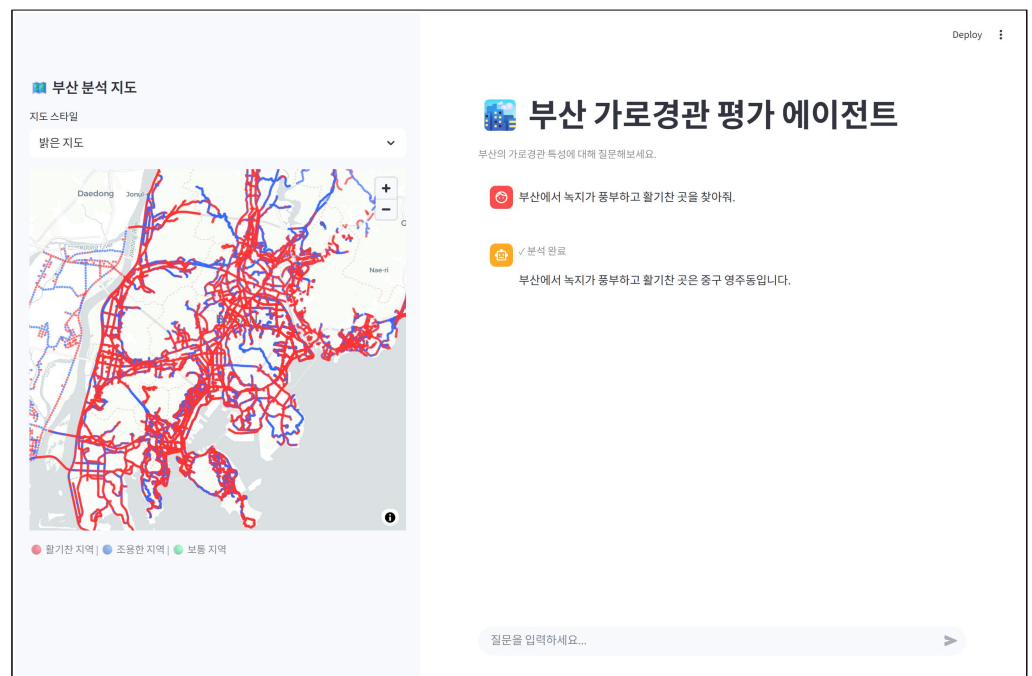


그림 8. 기초 탐색 시나리오의 질의와 답변 예

을 비교해 줄래?”라는 질문에 대해 에이전트는 행정구역별 평균 천공률을 계산한 후 “북구의 평균 하늘 개방감은 0.261이고, 동래구는 0.247입니다. 따라서 북구가 동래구보다 더 높은 하늘 개방감을 보입니다.”라고 답변하였다. 이를 통해 행정구역별 데이터를 비교 분석하고 그 결과를 설명할 수 있음을 확인하였다.

4.4 심층 추론 시나리오: 변수 간 상관관계 분석 및 설명

시나리오 3에서는 변수 간 상관관계를 분석하고 이를 해석하는 능력을 검증하였다. 검증 문항은 녹시율과 수목 수, 사람 수와 활기참, 이동수단과 활기참, 천공률과 활기참, 사람 수와 차량 수의 관계를 분석하도록 구성하였다.

대표 사례로, 이용자가 “가로수가 많으면 정말로 사람들이 활기차다고 느끼나?”라고 질문하면 에이전트는 피어슨 상관계수를 계산하여 수목 수와 활기참의 상관계수는 0.142로 나타남을 확인하였다. 이어 사람 수와 활기참의 상관계수(0.326)를 추가로 분석하여, 사람 수가 활기참과 상대적으로 높은 관련성을 보인다는 결과를 제시하였다.

다만 심층 추론 단계는 단순 정보 조회와 달리 통계적 해석과 추가적인 분석 과정이 요구되는 영역이다. 이전 시나리오와 다르게 심층 추론 질의에서는 해석 오류와 응답의 일관성 저하가 확인되었다. 이는 현재 적용한 로컬 LLM의 심층 추론 성능이 복합적인 통계 분석과 해석을 수행하기에는 아직 한계가 있음을 보여준다. 따라서 향후에는 추론 성능이 향상된 모델을 적용하고, 다양한 질의를 기반으로 심층 추론 능력을 보완할 필요가 있다.

5. 결론

이 연구는 GSV 이미지를 수집하고 Vision AI 기술을 활용하여 부산시 가로경관 데이터베이스를 구축한 후, 이를 자연어 질의를 통해 탐색하고 분석할 수 있는 로컬 환경 기반 가로경관 평가 에이전트를 개발한 연구이다. 부산시 도로망을 기준으로 37,167개 지점을 설정하고, 각 지점의 네 방향에서 수집된 총 148,668장의 GSV 이미지를 수집하였다. 수집한 이미지에 SegFormer, YOLO, CLIP 모델을 적용하여 물리적 환경(녹지율, 천공률 등), 보행 및 이동 활동의 수준, 인지적 특성(활기참과 조용함)을 측정하고 이를 통합하여 가로경관 데이터베이스를 구축하였다. 가로경관 데이터가 외부 서버로 전송되는 것을 제어하기 위해 오픈소스 언어모델인 Qwen3-14B 모델을 로컬 환경에서 구축하고, 16GB의 GPU 환경에서 안정적으로 실행되도록 NF4 4-bit 양자화 기술을 적용하였다. LangChain과 ReAct 기법을 적용하여 이용자의 질의를 데이터 탐색과 Python 기반 분석으로 변환하고 그 결과를 다시 자연어로 답변하는 가로경관 평가 에이전트를 구현하였다. 가로경관 평가 에이전트의 수행 결과 “녹지가 풍부하고 활기찬 곳”과 같은 이용자의 정성적 질의를 녹지율과 활기참 지표를 활용한 정량적 조건으로 변환하여 해당 지역을 탐색하고 특정할 수 있었다. 또한, 단순 조회뿐만 아니라 지역 간 비교 분석, 그리고 변수 간의 상관관계를 분석하는 심층 추론까지 수행함을 확인하였다.

이 연구는 크게 세 가지 점에서 기존 연구와 구별된다. 첫째, Vision AI에 기반한 SegFormer, YOLO, CLIP을 활용하여 가로경관의 물리적 환경, 가로 환경, 인지적 특성을 하나의 가로경관 데이터베이스로 종합하였다. 기존 연구도 Vision AI를 활용했지만, 개별 지표를 구축하고 분석하였다면, 이 연구는 여러 지표를 통합하여 구축하여 동시에 분석할 수 있는 구조를 마련하였다. 이를 바탕으로 언어모델을 연계하여 누구나 쉽게 자연어를 사용하여 가로경관 데이터를 분석할 수 있도록 하였다. 둘째, 에이전트가 언어모델의 자체 지식에 의존하지 않고, 연구에서 구축된 가로경관 데이터베이스를 조회하고 분석하여 답변하도록 설계하였다. 이러한 방식은 거대언어모델의 문제점인 환각 현상을 줄이고, 조정계획과 경관계획 등 관련 분야에서 데이터 분석을 지원하는 도구로 활용할 수 있는 기반을 마련하였다. 마지막으로 기존 연구가 멀티모달 모델을 활용한 가로경관 특성 분석에 중점을 두었다. 반면, 이 연구는 분석 결과를 하나의 데이터베이스로 통합하여 AI 에이전트와 연계를 통해 자연어 기반의 데이터 조회 및 분석이 가능한 활용 체계를 제시했다는 점에서 기존 연구와 차별성을 가진다.

이 연구는 Vision AI와 거대언어모델을 결합하여 도시 가로경관 평가에 새로운 가능성을 제시하였으나, 다음과 같은 부분에서 한계가 있다. 이 연구에서 활용한 변수는 실제 시민의 인식이나 행동을 통해 측정된 값이 아닌 이미지에 기반해서 측정한 대리 지표라는 점에서 해석상의 한계를 가지고 있다. 특히 CLIP 기반의 인지 지표는 모델이 학습한 이미지와 텍스트의 관계에 의존하기 때문에 향후 설문조사나 현장 조사 등의 추가적인 교차 검증이 필요하다. 또한 이 연구는 부산광역시 단일 도시에 대해서 분석을 진행했으므로 다른 도시 유형에 대해서 전체적으로 일반화하는데 한계가 있다. 향후에는 도시 형태 및 도시 규모가 다른 지역을 대상으로 비교와 분석을 해 검증할 필요가 있다. 마지막으로 에이전트의 성능 평가는 일부 질의를 중심으로 분석이 되었으며, 다양한 질의에 대한 체계적인 분석 평가 체계는 부족한 실정이다. 향후 표준화된 분석 체계를 구축하여 에이전트의 신뢰성을 확보하는 후속 연구가 필요하다. 이러한 한계에도 불구하고 이 연구는 조정 데이터 분석에 AI 기술을 접목하여 활용 가능성을 제시하였다는 점에서 의미가 있다. 향후 현장 및 전문가 데이터를 연계한 검증을 통해 조정 및 경관계획을 지원하는 AI 기반 분석 방법으로 확장할 수 있을 것이다.

주 1. CLIP(contrastive language-image pre-training)란 Open AI에서 개발한 언어-이미지 멀티모달 모델로, 이미지와 여러 텍스트 설명과의 유사성을 계산하는 모델이다. 예를 들어, 하나의 이미지에 “a diagram”, “a dog”, “a cat”을 입력하면 각각의 확률을 99.28%, 0.42%, 0.30%와 같이 계산하는 모델이다.

주 2. Vision AI란 이미지와 영상을 분석하여 객체, 텍스트 등의 정보를 인식하고 계산하는 인공지능 기능이다. 예를 들어, 하나

의 이미지에서 사람이나 자동차와 같은 객체의 비율이나 숫자를 보여주거나 텍스트로 추출하는 기능이 있다.

주 3. Anaconda는 프로젝트별로 독립적인 Python 실행 환경과 라이브러리를 관리하여 개발 환경의 충돌을 방지하고 재현성을 높이는 플랫폼이다. 예를 들어, A 프로젝트에서 필요한 라이브러리를 설치해도 B 프로젝트에 영향을 주지 않아 오류나 충돌을 줄일 수 있다.

주 4. LangChain이란 거대언어모델이 필요한 프로그램이나 데이터를 활용할 수 있게 연결하는 오픈소스 프레임워크다. 예를 들어, 사용자의 질문을 받아 데이터베이스를 검색하고 이 내용을 바탕으로 거대언어모델이 답변을 생성하도록 하는 역할을 한다.

References

1. 기동환, 김선재, 이수기(2021) Google Street View와 딥러닝을 활용한 서울시 녹지 형평성 분석: NDVI와 가로 이미지 기반 녹지 산출방법과의 비교를 중심으로. *대한국토·도시계획학회* 56(4): 194-211.
2. 김동국, 최연준, 장백철(2025) RAG 기반 뉴스 정보와 LLM을 활용한 시계열 예측: COVID-19 데이터를 중심으로. *인터넷정보학회논문지* 26(4): 79.
3. 김연우, 구분유, 이종현, 윤희연(2025) 보행 가로 경관 요소가 도시민의 우울감에 미치는 영향 분석: 거시적 및 미시적 경관 요소와 도보 이용률을 고려하여. *한국조경학회지* 53(1): 77-93.
4. 김영우, 황용하, 정은석, 강범준(2023) 구글 스트리트 뷰 데이터와 인공지능을 활용한 위치기반 가로공간 보행자 데이터 수집 방법 개발. *대한건축학회논문집* 39(9): 57-68.
5. 김이정, 이수기(2025) 멀티모달 대규모 언어모델과 기계학습을 활용한 도시 가로 경관 쇠퇴 영향요인 분석. *대한국토·도시계획학회* 60(3): 65-83.
6. 박영호(2025) 생성형 AI를 활용한 실내디자인 협업 방안 연구. *한국실내디자인학회 논문집* 34(1): 23-34.
7. 양유선, 손용훈(2019) 덕수궁 돌담길의 역사문화가로 장소 인식 특성에 관한 연구. *한국조경학회지* 47(3): 60-70.
8. 우상근(2023) LLM(거대 언어모델) 활용 방식 및 주요 이슈 분석. *한국지능정보사회진흥원보고서 IT & Future Strategy*.
9. 유승재, 하정원, 김해준, 기동환, 이수기(2021) 서울시 가로경관 이미지에 대한 주관적 인지에 영향을 미치는 가로 환경 요인 분석. *국토계획* 56(2): 79-93.
10. 이다연, 이지운, 이재호(2024) 구글 스트리트 뷰와 딥러닝을 활용한 보행 친화적 환경이 여가보행에 미치는 영향 평가: 서울특별시 용산구를 대상으로. *한국조경학회지* 52(4): 45-55.
11. 임준호(2025) 거대언어모델(LLM) 응용 기술 최근 동향. *한국통신학회지* 42(4): 24-31.
12. Aikoh, T., R. Homma and Y. Abe(2023) Comparing conventional manual measurement of the green view index with modern automatic methods using Google Street View and semantic segmentation. *Urban Forestry & Urban Greening* 80: 127845.
13. Biljecki, F. and K. Ito(2021) Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning* 215: 104217.
14. Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal and D. Amodi(2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33: 1877-1901.
15. Choi, J. and Y. Kang(2025) Modeling and explaining perceived walkability in urban environments using street view images and explainable AI. *Spatial Information Research* 33(6): 57.
16. De Nadai, M., J. Staiano, R. Larcher, N. Sebe, D. Quercia and B. Lepri(2016) The death and life of great Italian cities: A mobile phone data perspective. In *Proceedings of the 25th International Conference on World Wide Web*: 413-423.
17. Doan, Q. C., J. Ma, S. Chen and X. Zhang(2025) Nonlinear and threshold effects of the built environment, road vehicles and air pollution on urban vitality. *Landscape and Urban Planning* 253: 105204.
18. Ewing, R. and S. Handy(2009) Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design* 14(1): 65-84.
19. Ewing, R., A. Hajrasouliha, K. M. Neckerman, M. Purciel-Hill and W. Greene(2016) Streetscape features related to pedestrian activity. *Journal of Planning Education and Research* 36(1): 5-15.

20. Gong, F. Y., Z. C. Zeng, F. Zhang, X. Li, E. Ng and L. K. Norford(2018) Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment* 134: 155–167.
21. Huang, Y., C. Zhong, T. He and Y. Jiang(2025) Dynamics of street environmental features and emotional responses in urban areas: Implications for public health and sustainable development. *Frontiers in Public Health* 13: 1589183.
22. Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto and P. Fung(2023) Survey of hallucination in natural language generation. *ACM Computing Surveys* 55(12): 1–38.
23. Liu, L. and A. Sevtsuk(2024) Clarity or confusion: A review of computer vision street attributes in urban studies and planning. *Cities* 150: 105022.
24. Liu, K., T. Yigitcanlar, R. Mehmood, J. Corchado and X. Fu(2025) Large language models in urban planning: A systematic review and conceptual framework. *Journal of Urban Technology* 2025: 1–44.
25. Lu, Y., C. Sarkar and Y. Xiao(2018) The effect of street-level greenery on walking behavior: Evidence from Hong Kong. *Social Science & Medicine* 208: 41–49.
26. Nathvani, R., A. Cavanaugh, E. Suel, H. Bixby, S. N. Clark, A. B. Metzler and M. Ezzati(2025) Measurement of urban vitality with time-lapsed street-view images and object-detection for scalable assessment of pedestrian-sidewalk dynamics. *ISPRS Journal of Photogrammetry and Remote Sensing* 221: 251–264.
27. Ogawa, Y., T. Oki, C. Zhao, Y. Sekimoto and C. Shimizu(2024) Evaluating the subjective perceptions of streetscapes using street-view images. *Landscape and Urban Planning* 247: 105073.
28. Oke, T. R.(1981) Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. *Journal of Climatology* 1(3): 237–254.
29. Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal and I. Sutskever(2021) Learning transferable visual models from natural language supervision. *Proceeding of the International Conference on Machine Learning*, pp. 8748–8763.
30. Xie, E., W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez and P. Luo(2021) SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems* 34: 12077–12090.
31. Xu, F., J. Zhang, C. Gao, J. Feng and Y. Li(2023) Urban generative intelligence (UGI): A foundational platform for agents in embodied city environment. *arXiv preprint arXiv:2312.11813*.
32. Yang, J., L. Zhao, J. McBride and P. Gong(2009) Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning* 91(2): 97–104.
33. Yuan, K., H. Lan, Y. Gao and K. Wang(2026) Interpretable multimodal framework for human-centered street assessment: Integrating visual-language models for perceptual urban diagnostics. *Land* 15(3): 449.
34. Zhu, H., X. Nan, N. Kang and S. Li(2024) How much visual greenery can street trees generate from a humanistic perspective? An attempt to quantify the canopy green view index based on tree morphology. *Forests* 15(1): 88.
35. Zhu, H., X. Nan, D. Chi and S. Li(2025) Validity of online street view imagery in urban studies: Evidence from on-site mobile acquisition. *Urban Forestry & Urban Greening*: 129173.