

AI 기반 이미지 텍스트화를 활용한 건축 이미지 데이터 추이 분석

- 건축물 이미지 기반 트렌드 분석 사례를 중심으로

Trend Analysis of Architectural Image Data Using AI-Based Image Data Using
AI-Based Image-to-Text Conversion

- Focusing on Architectural Image-Based Trend Analysis Case Studies

○이 우 섭*

Yi, Woo-Sop

Abstract

Traditional architectural image analysis has relied heavily on subjective expert interpretations, resulting in significant time consumption and inconsistency. Recent advancements in Vision-Language Models (VLMs) and Large Language Models (LLMs) have addressed these limitations by automatically generating meaningful natural language descriptions of architectural images without additional labeling. Through a case study of trend analysis based on architectural images, this research proposes an efficient methodology that integrates pre-trained VLMs, LLMs, and BERTopic to automatically analyze architectural imagery, enabling structured and quantitative assessments of trends inherent in architectural image data.

키워드 : VLM, LLM, 건축 이미지 분석, 인공지능, 트렌드 분석, 데이터 기반 분석, BERTopic

Keywords : VLM, LLM, Architectural Image Analysis, AI, Trend Analysis, Data-Driven Analysis, BERTopic

1. 서론

1.1 연구의 배경과 목적

기존 건축 이미지 분석은 전문가의 시각적 해석과 경험에 의존해 심층적 통찰을 제공하지만, 대규모 이미지 분석 시 상당한 시간과 비용이 소요된다는 한계가 있었다. 이에 따라 컴퓨터 비전을 활용한 접근이 시도되었고, 특히 CNN(Convolutional Neural Network) 기반 모델들이 이미지 분류나 객체 인식 등의 과업에서 유의미한 성과를 보여왔다. (An, 2023) 그러나 CNN 모델은 높은 정확도를 유지하기 위해서 방대한 라벨링 데이터가 필수적이며, 건축 분야에서 요구되는 세분화된 맥락 정보(재료, 공간적 관계 등)를 온전히 반영하기에 제약이 따른다. 반면, 최근 등장한 Vision Language Model(VLM)은 대규모 사전 학습을 통해 개별 이미지마다 별도의 라벨링을 하지 않아도 이미지 속 건축 요소들을 자연어로 자동 추출할 수 있다. 즉, 정확한 라벨 데이터를 일일이 준비하지 않아도 건축 이미지의 다양한 요소(재료, 형태, 조형성 등)를 자연어로 추출하기가 용이해지면서, 시간 및 비용 절감과 분석 범위 확장이라는 두 가지 측면에서 큰 이점을 제공한다.

이에 본 연구는 사전 학습된 VLM과 Large Language Model(LLM), BERTopic을 결합하여, 최소한의 전처리만으로도 건축 이미지를 자동 분석해 비용 효율적이면서도 확장성이 높은 방법론을 제시한다.*

1.2 연구의 방법

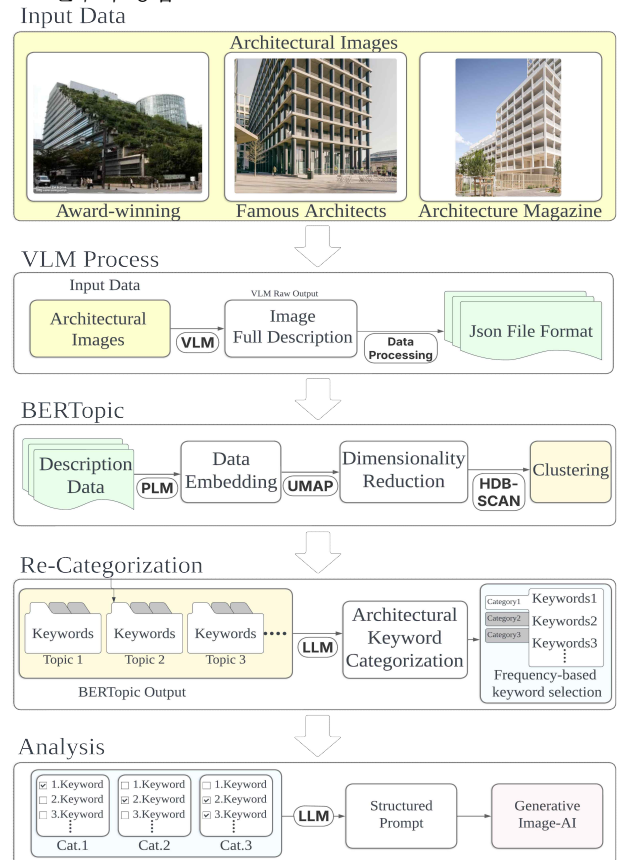


그림1. 연구 방법 개요

* 세종대학교 건축학과

이 방법론은 다음과 같은 단계를 통해 진행되었다.

1) 데이터 수집 및 전처리: 프리츠커 수상 건축가의 작품 이미지와 Archdaily 플랫폼의 건축 이미지를 수집하였다.

2) VLM 기반 이미지 분석: 사전 학습된 대규모 VLM을 활용해 각 건축 이미지에서 텍스트 설명을 추출하였다.

3) 텍스트 군집화 및 키워드 추출: 이미지 설명 문서를 임베딩한 뒤, UMAP과 HDBSCAN을 통해 군집화하고, c-TF-IDF 기법을 사용해 각 군집의 대표 키워드를 산출하였다.

4) 키워드 카테고리 자동 분류: 추출된 키워드를 LLM을 통해 건축 분야의 카테고리로 자동 분류하였다.

5) 결과기반 LLM을 통한 건축 트렌드 분석: 토픽별, 카테고리별 핵심 키워드를 기반으로 건축 이미지에서 나타나는 경향성을 파악하고, 이를 LLM을 통해 요약, 해석하였다.

6) 생성형 AI모델을 통한 시각적 대표 이미지 생성: 분석 결과를 바탕으로 이미지 생성 AI에 prompt를 제공하여, 트렌드 특성을 시각적으로 표현할 수 있는 예시 이미지를 제작하였다.

이상의 방법론을 통해 본 연구는 기존의 주관적, 수작업 중심의 이미지 분석 방식을 보완하고, 대규모 건축 이미지에 내재된 패턴과 트렌드를 보다 정량적으로 규명하는 효과적인 절차를 제안한다.

2. 연구의 차별성

2.1 라벨링 의존도 최소화

앞서 서론에서 언급했듯, CNN 접근은 대규모 라벨링에 대한 의존도가 높아 실무 적용 시 어려움이 크다. 반면, 본 연구는 사전 학습된 VLM을 활용함으로써 추가적인 레이블 부착 과정 없이 이미지로부터 풍부한 자연어 설명을 자동으로 추출한다. 이를 통해 전처리와 데이터 구축에 드는 시간과 비용을 크게 줄이고, 연구자가 라벨링에 매몰되지 않고도 광범위한 건축 이미지를 분석할 수 있다는 점에서 차별적이다.

2.2 맥락적·세분화된 정보 추출 가능

기존의 전통적 건축 이미지 분석은 연구자의 주관적 해석이나 특정 시각 요소(예: 창문 형태, 외피 재료 등)에 초점을 두는 경우가 많았고, 이를 CNN 모델에 적용할 때도 ‘정해진 라벨’만 분류하는 데 그치는 경우가 대부분이었다. 이에 비해 본 연구는 VLM을 통해 건축 이미지 속 공간적 맥락, 조형성, 공간적 배치 등 다층적인 맥락을 자연어로 포착한다. 다시 말해, CNN 접근으로는 놓치기 쉬운 빛, 주변 환경과 공간적 분위기 등도 텍스트 분석 대상에 포함되어 보다 세밀하고 종합적인 해석을 가능케 한다.

2.3 토픽 모델링을 통한 구조적 요약

전통적 연구 방식은 이미지마다 특성을 일일이 기술하거나 빈도 기반 단순 키워드 분석에 그치는 경우가 많았다. 반면 본 접근법은 BERTopic 기법을 활용하여 대규모 텍스트(캡션)에서 ‘토픽(군집)’ 단위로 구조화하고, 이를 통해 상호 유사한 이미지·설명을 묶어내는 정량적 분석 방식을 제시한다. 이는 단순 빈도 통계와 달리, 의미적으로 밀접한 키워드들이 함께 묶여 나타나는 경향을 파악하게 해주어 기존 방식보다 한 단계 높은 수준의 요약과 분류가 가능하다.

2.4 LLM을 통한 건축 맥락 재구조화

기존에는 건축적 전문성을 반영하기 위해 연구자가 추출된 단어를 다시 카테고리화하거나 용어 정리를 해야 했다. 그러나 본 연구에서는 대규모 언어 모델을 활용하여 추출된 토픽 키워드를 자동으로 건축 분야의 주요 범주(예: Façade Composition, Form & Mass, Materials & Textures 등)로 재분류한다. 이를 통해 사용자(연구자, 건축가)가 적은 개입으로도 도메인 특화 카테고리를 일관되게 적용할 수 있어, 기존 연구보다 한층 더 효율적이고 체계적인 분류 체계를 구축할 수 있다는 점에서 차별적이다.

3. 이미지 분석 및 데이터 분석

3.1 VLM을 통한 건축 입면 이미지 분석

VLM은 일반적으로 이미지 입력에 대해 자연어 설명을 생성하는 AI모델을 의미한다. VLM이 자동으로 생성한 이미지에 대한 자세한 설명 텍스트는 이후 텍스트 클러스터링을 수행하기 위한 원시 데이터(raw text)로 사용된다. 연구에서는 자체 수집한 건축 사진을 혼합하여 755장의 이미지를 활용했다.

3.2 텍스트 데이터의 전처리 및 의미적 클러스터링

VLM으로부터 나온 캡션이나 원본 데이터셋 내에 이미 존재하는 건축 이미지 설명문 등을 종합하여 텍스트 분석에 활용한다. 이때 불필요한 단어(‘Building’, ‘Structure’, ‘Architectural’) 등 건축 분야에서 너무 일반적으로 쓰이는 단어를 도메인 불용어(Domain Stopwords)로 사전에 정의하고, spaCy 파이프라인을 활용한 기본 불용어, 구두점 제거와 형태소 분석 과정을 수행한다.

이후 문서(문장) 단위 임베딩을 위해 사전 학습된 문장 임베딩 모델(Sentence Transformer-BAAI/bge-m3)을 사용한다. 임베딩된 벡터는 1024 차원의 고차원 공간에 위치하게 되는데, 이를 UMAP 기법을 활용해 2차원으로 축소한 뒤, HDBSCAN을 적용해 밀도기반 클러스터링을 수행한다. 결과적으로 비슷한 의미를 가진 문서들(예: “브루탈리즘 콘크리트 건물” 관련 설명들)이 같은 군집으로 묶이고, 의미적으로 다른 문서들과는 군집이 분리된다. 이는 단순 빈도 기반 통계 기법보다 훨씬 유연하며, 다양한 건축 스타일이나 재료에 대한 맥락적 차이를 효과적으로 포착하는 장점이 있다.

4. BERTopic 기반 토픽 레이블링

4.1 클래스 기반 c-TF-IDF를 활용한 토픽 표현 생성

BERTopic의 핵심 아이디어 중 하나는 ‘클래스 기반 c-TF-IDF’ 다. 이는 각 토픽(클러스터)에 속한 문서들을 하나의 클래스로 묶고, 그 클래스에서 자주 등장하는 단어가 전체 코퍼스 대비 중요도를 측정한다. 이를 통해, 해당 토픽을 대표하는 핵심 키워드를 자동으로 선정할 수 있다.

또한, BERTopic은 LLM을 활용하면 토픽 내 문서들을 간략히 요약한 뒤, 해당 토픽을 가장 잘 표현하는 라벨을 자동으로 제안해 준다. 예를 들어 “Topic0”이 “Residential, Wooden” 등의 키워드를 갖는다면, LLM이 이를 종합하여 “0_residential_natural_wooden”와 같은 레이블을 제안할 수 있다. 본 연구에서는 Llama 모델(openhermes-2.5-mistral-7b.Q4_K_M.gguf)을 통해 레이블링 프롬프트를 구성하여 레이블을 추출하였다.

4.2 건축 입면 이미지 분석 결과

분석 결과 26개의 토픽이 발견되었으며, 각 토픽은 서로 다른 건축 요소를 반영하고 있었다. 토픽별 시각화를 수행한 결과인 (그림2)는 다음과 같이 요약할 수 있다.

1) Topic clusters in 2D space: UMAP 2차원 공간상에서 유사 토픽끼리 밀집하여 나타났다.

2) Topic Word Scores: 각 토픽 내에서 c-TF-IDF 점수가 높은 단어일수록, 해당 토픽을 설명하는 핵심 키워드로 선정되었다.

이처럼 BERTopic으로부터 토픽별 핵심 단어, 문서를 파악함으로써, 중복되는 표현을 일일이 사람이 정하지 않아도, 방대한 양의 건축 이미지를 요약할 수 있다.

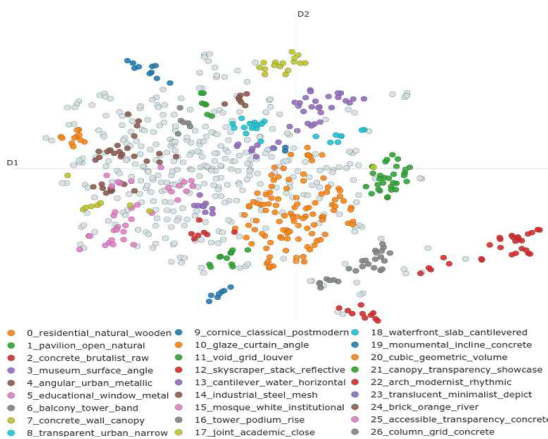


그림2. UMAP 기반 데이터 텍스트의 토픽 클러스터링 결과

5. 건축 관련 카테고리 재분류

5.1 토픽 키워드 추출

Bertopic 결과로부터 각 토픽에 포함된 핵심 키워드를 수집한 뒤, 해당 키워드의 c-TF-IDF 점수를 기반으로 중요도를 측정한다. 토픽별로 단어와 점수 정보를 병합하여, 최종적으로는 (topic_id, keyword, score) 형태의 리스트를 생성한다.

5.2 LLM을 활용한 건축 카테고리 재분류

토픽별 키워드 집합을 다시 LLM에 입력하여, 각 키워

드가 “Facade composition”, “Form and Mass”, “Materials and Textures”, “Architectural program” 등으로 자동 분류되도록 한다. 연구에서는 GPT4.5 모델을 활용했으며, 프롬프트에 “해당 단어가 건축에서 어떤 범주에 속하는지”를 질의해 응답받는 구조다. 이러한 일괄 분류 과정을 통해 토픽 키워드들이 건축 분야에서 어떠한 의미 영역에 해당하는지 일관성 있게 재구조화할 수 있다.

5.3 분류 결과와 c-TF-IDF 점수 결합

키워드별 카테고리 분류가 완료된 후, 각 카테고리 내에서 키워드들의 c-TF-IDF 점수를 합산한다. 예를 들어, “Facade composition” 카테고리에 속한 모든 키워드 점수를 더하면 그 카테고리의 총점(total_score)이 된다. 재분류 및 분포의 결과인 (그림3)에 따르면, “Facade composition”이 가장 높은 점수를 차지했으며, 그 뒤를 “Form and Mass”, “Materials and Textures”, “Design Expression” 등이 따르는 양상을 보였다. 이는 현대 건축 이미지에서 입면 구성이 매우 다양한 형태로 표현되고 있음을 시사한다.

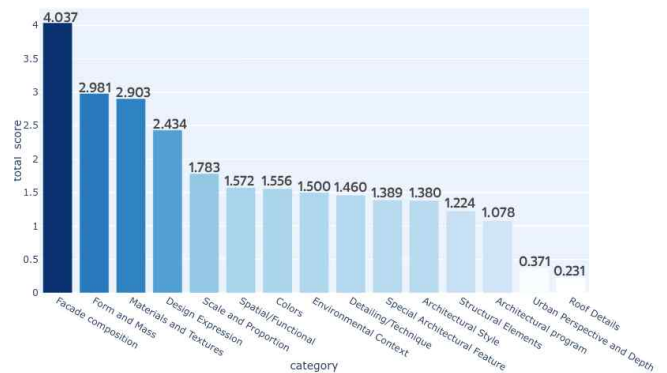


그림3. 건축 카테고리별 재정의를 한 키워드 분포(합산 점수 기준)

5.4 분석 결과를 활용한 데이터 경향성 파악

1) 카테고리별 상위 키워드 시각화 : “Facade composition”에서는 window, wall, vertical, panel 등이 높은 비중을 차지한다. 이는 외피 구성 요소가 건축적 트렌드에서 중요한 위치를 차지함을 보여준다.

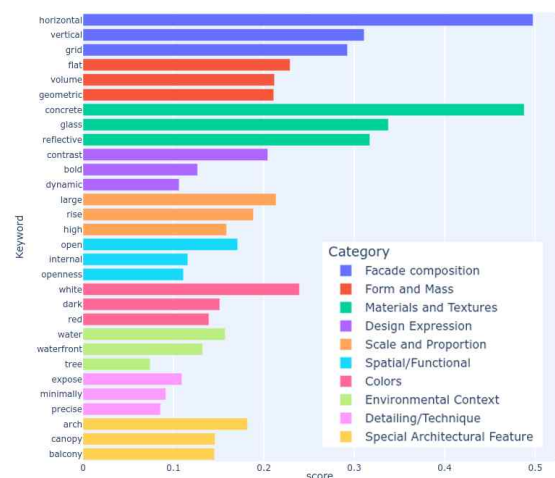


그림4. 카테고리별 주요 키워드 상위 3개의 c-TF-IDF 점수

2) 시계열 또는 세부군 비교: 만약 이미지를 연도별 혹은 지역별로 분류한다면, 특정 시점에 유리나 금속 같은 재료 키워드가 급증하는 시기를 포착할 수 있다. 본 논문에서는 시점별 분석까지는 다루지 않았지만, 확장 연구에서는 연도별 경향 파악이 가능할 것으로 기대한다.

6. 결론

6.1 최종 건축 트렌드 분석

본 연구가 도출한 트렌드 분석 결과는 다음과 같다.

“현대 건축 트렌드는 전반적으로 모던하면서도 미니멀한 조형미를 추구하며, 콘크리트와 유리 같은 노출, 반사형 재료가 많이 활용된다. 특히 대형 볼륨과 개방형 내부를 구현하는 동시에, 구조와 디테일을 정교하게 드러내는 표현 방식이 돋보인다.

도시 밀도 증가에 대응하여 건물은 수직화(‘tower’)되는 경향이 강하며, ‘arches’나 ‘canopies’, ‘클래식 요소’ 등을 현대적으로 재해석하여 전통과 혁신의 긴장감을 더하기도 한다. 또한 ‘contrast’와 ‘bold’라는 키워드가 강조되면서, 형태, 재료 간 대비를 통해 시각적 임팩트를 극대화하고 있다.

이러한 현대적 조형미와 함께 환경 및 맥락적 고려도 점차 중요해지고 있다. ‘water’, ‘tree’, ‘roof’와 같은 자연 및 외부 공간 요소를 적극적으로 통합하여 지속가능성을 높이고, 실내외 경계의 유연화를 통해 쾌적한 사용자 경험을 제공하려는 시도가 늘고 있다. 이때 ‘canopies’나 ‘발코니’, ‘테라스’ 등은 실내외 연결성을 극대화하고, 사람이 자연이 교감할 수 있는 공간을 만들어낸다.

마지막으로, ‘인간적 스케일’을 유지하는 섬세한 디테일(정교한 마감, 촉각적 재료 선택 등)이 중요하게 부각되고 있다. 무분별한 대형화, 수직화를 하기보다, 주변 도시 맥락과 환경을 고려한 균형점을 찾는 것이 핵심과제로 자리 잡고 있다고 볼 수 있다.”

6.2 시각화 이미지

추가로 건축 카테고리별 분석 결과를 활용하여 상위 키워드들을 조합하여 prompt(조건부 이미지 생성을 위한 텍스트 설명)를 아래의 예시와 같이 생성할 수 있다.

“Envision a minimalist, modern tower with flat geometric volumes, featuring a bold contrast between white concrete and reflective glass facades. This contemporary residential building embodies sleek design and harmonious integration with its environment”

이를 이미지 생성 AI(예: Stable Diffusion, Midjourney 등)에 입력하면, 방법론이 제시한 건축 트렌드와 일치하는 시각적 이미지를 자동으로 생성할 수 있다. 이는 건축가나 디자이너에게 새로운 아이디어를 촉진하는 참고 자료가 될 수 있다.



그림5. 이미지 생성 AI를 활용한 시각화 이미지

6.3 연구 성과 및 한계점

연구 성과로는 첫째, 대규모 건축 이미지를 신속히 처리하며 전문가의 수작업 의존도를 낮출 수 있는 방법론을 제시했다는 점. 둘째, 이미지로부터 추출된 텍스트(키워드)를 재분류해 건축 분야에 맞게 보다 정량적인 지표를 생성하는 절차를 구체화했다는 점이 있다. 나아가, 이러한 지표를 통해 현대 건축 트렌드를 제시함으로써, 실무 및 교육 현장에서 참고할 만한 구체적인 방향성을 제안했다는 점에서 의의가 있다.

그러나, 한계점도 존재한다. (1) VLM의 품질에 따라 텍스트 생성이 부정확할 수 있고, (2) 건축적 전문 용어를 완벽히 반영하기 위해서는 추가적인 Fine-tuning이 필요할 수 있으며, (3) 데이터셋이 특정 지역, 시대에 편중되어 있으면 결과 해석에 주의가 요구된다. (4) 토픽 모델링 과정에서 일부 노이즈 토픽이 여전히 존재할 수 있으며, LLM 기반 분류에서도 프롬프트 엔지니어링 및 LLM 모델 특성에 따라 분류 정확도가 달라질 수 있다.

6.4 향후 연구 방향

1) 도시 단위의 빅데이터: 도시 스케일에서 다양한 시기, 지역별 건축물을 대상으로 적용하여, 지역별 건축 경향의 시, 공간적 변화를 추적할 수 있을 것이다.

2) 실시간 데이터 연계: SNS나 건축 포털 사이트로부터 실시간으로 건축 이미지를 수집하고, 자동 분석 시스템을 구축해 최근 건축 경향의 빠른 변화를 추적할 수 있을 것이다.

3) 생성형 AI와의 결합: 본 연구의 최종 결과(카테고리별 키워드)를 기반으로, 이미지 생성형 AI를 통해 새로운 설계 아이디어나 컨셉 스케치 이미지를 제안할 수 있을 것으로 보인다.

참고문헌

1. LH 토지주택연구원, AI 기반 건설기술 현황 및 연구전략, 2020
2. 안종국, 딥러닝을 활용한 건축물 외관의 유형화 방법론 개발: Deep Learning Based Typification Method for Building Exterior 석사학위논문, 서울대학교대학원 공과대학 건축학과, 2023
3. Perez, H., Tah, J.H.M., & Mosavi, A. (2019). Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. arXiv preprint arXiv:1908.04392