Leveraging Large Language Models for Architectural Style Prediction: A Multi-Modal Approach Using Location and Streetview Imagery in the Georgetown Historic District

Finn Mokrzycki

In Development as of 12/4/2024

Introduction and Rationale

The Georgetown Historic District in Washington, DC, is an architectural heritage area characterized by Federal, Victorian, Italianate, and other historically significant styles. Traditionally, architectural classification has relied on expert observation. However, advancements in large language models (LLMs) and computer vision now offer new opportunities to automate and scale this process.

This project explores how LLMs, in combination with spatial data and visual imagery, can classify architectural styles in Georgetown. It examines a multimodal approach that integrates location data, raw imagery, and structured descriptions of architectural features to determine the most effective methodology for accurate and interpretable architectural classification.

Research Questions

1. **Prediction Accuracy:** How accurately can LLMs predict architectural styles using geographic location data alone compared to visual data (Google Street View images) or textual descriptions of architectural features?
2. **Feature Importance and Interpretability:** Which features—geospatial context, visual cues, or explicitly extracted architectural elements—play the most significant role in determining architectural style?
3. **Methodological Contribution:** What advantages do multi-step, multi-modal approaches (location → image → textual feature extraction → classification) offer over direct classification from location or raw images alone?

Related Work

This study builds upon research in three key areas:

- **Geospatial Deep Learning:** Previous studies have used machine learning models to infer building functions, property values, and socio-economic indicators based on location data (Atwal et al., 2020).
- **Computer Vision for Architecture Classification:** Convolutional neural networks (CNNs) have been applied to classify buildings by architectural style or historical period (Wu et al., 2023).
- **Deep Learning for Architectural Feature Extraction:** Vision Transformers (ViT), both alone and in combination with LLMs, have been used for residential properties to determine usage and perform classification (Ramalingam & Kumar, 2024).
- **LLMs for Multimodal Reasoning:** Recent models, such as GPT-4, have demonstrated potential for combining text-based and image-based inputs for domain-specific classification tasks (Pierdicca & Paolanti, 2022).

Data Sources and Preparation

- **Location Data:** Geographic coordinates of properties in the Georgetown Historic District, sampled to ensure a variety of architectural styles.
- **Streetview Imagery:** Google Street View API will be used to obtain consistent images of properties, minimizing variability due to angle or resolution differences.
- **Architectural Style Ground Truth:** A dataset of known architectural styles will be compiled from historical records from the United States Commission of Fine Arts and Historical Surveys to validate model predictions.
- **Architectural Style Categories:**
    - Georgian
    - Federal
    - Greek Revival
    - Italianate
    - Gothic Revival
    - Second Empire
    - Queen Anne (Victorian styles)
    - Colonial Revival
    - Beaux-Arts and Classical Revival
    - Tudor Revival
    - Craftsman/Bungalow
    - Mid-Century Modern/Contemporary

Methodology

This study proposes a three-phase prediction workflow:

A. Location-Only Prediction (Baseline)

1. **Input:** The LLM receives only the property's latitude and longitude, with context about its location in Georgetown.
2. **Prompt:** "Based on the given coordinates in Georgetown, DC, what is the likely architectural style of this home?"
3. **Analysis:** The LLM will be queried to determine which contextual features influenced its decision.

B. Location + Image-Based Prediction

1. **Image Retrieval:** A street-level image of the property's facade will be obtained.
2. **Image-to-Text Conversion:** A vision-to-text model will generate a natural language description of the building's key architectural features.
3. **Prompt:** "Based on this textual description (derived from an image at these coordinates), what is the architectural style?"
4. **Analysis:** The LLM will highlight which features most influenced the classification.

C. Location + Image + Structured Feature Extraction

1. **Feature Extraction:** A vision-to-text model will identify architectural elements (e.g., window type, facade material, roof shape, decorative details) (Nikparvar & Thill, 2021).
2. **Classification:** The extracted features will be fed into the LLM for style prediction.
3. **Analysis:** The LLM will rank the importance of each feature in its prediction.

Experimental Design

- **Sample Selection:** A balanced dataset of properties spanning multiple styles will be curated.
- **Evaluation Metrics:**
    1. **Accuracy and F1-score:** Comparison of predicted vs. known styles.
    2. **Survey Comparison:** The model explanations will be compared to the reasoning from the United States Commission of Fine Arts and the Historical Building surveys.
    3. **Qualitative Analysis:** Patterns in the LLM's feature attribution will be examined.
- **Ablation Studies:**
    1. Assess predictions with and without historical context.
    2. Compare performance across different LLMs (e.g., GPT-4 vs. alternatives).

Expected Outcomes

- **Accuracy Insights:** Predictions based solely on location will be less accurate than those incorporating visual data. The structured feature extraction approach should yield the highest accuracy and interpretability.
- **Feature Importance:** The study expects LLMs to emphasize rooflines, decorative elements, and facade materials as key style indicators.
- **Contribution:** This research will demonstrate the benefits of a multi-modal approach for architectural classification, with implications for historic preservation, urban planning, and real estate analytics.

Uniqueness in the Field

This study is innovative in several ways:

- **Integration of Geographic, Visual, and Textual Data:** The study advances architectural informatics by combining multiple data streams for classification.
- **Enhancing AI Interpretability in Cultural Heritage:** The findings will contribute to AI applications in humanities and historic preservation.
- **Application to Residential Architecture**: A critical fact about the Georgetown Historic District is that the area is significant for residential properties. This means the architecture is more conservative in style, so examining the reasoning for each architectural movement will require attention to minute details.

Limitations and Future Work

- **Limitations:**
    - Image descriptions depend on Street View quality and vision-to-text accuracy (Liu et al., 2022).
    - Some styles are hybrid and may be ambiguous. For example, Georgian and Federal are similar and possess the same elements, deriving from each other.
    - Some style classifications depend on the interior to determine the difference.
- **Future Directions:**
    - Expand methodology to other historic districts.
    - To track architectural modifications, incorporate historical time-series data (e.g., older Street View images).
    - Develop interactive tools where historians can refine prompts or provide supplementary guidance.

By leveraging the power of LLMs and multimodal data, this research provides a novel architectural classification framework that enhances accuracy and interpretability, with broad applications in urban studies, historic preservation, and AI-driven design analytics.