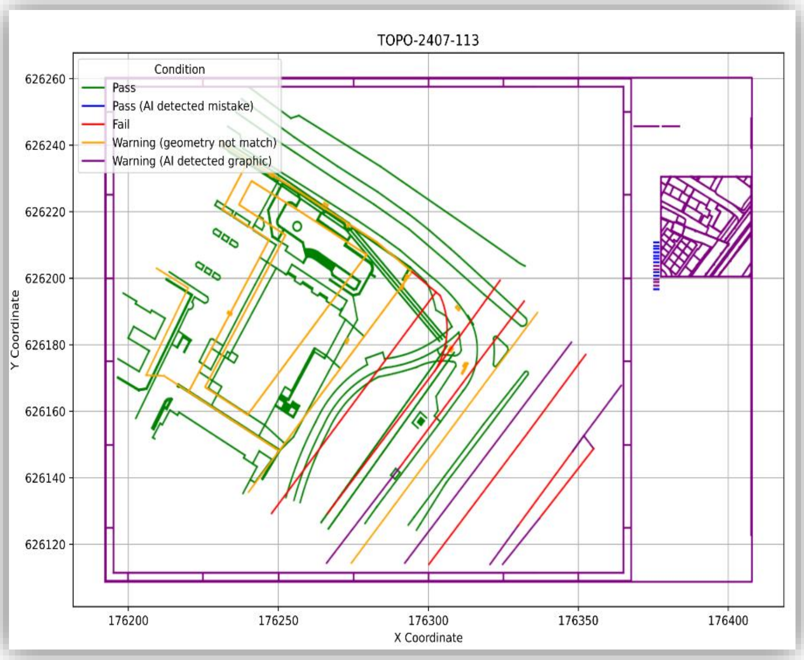
AI-Driven Conversion of DWG Files to GIS Data: Differentiating Graphical Elements from Survey Data



Medad Hoze

1/2/2025

Abstract:

The conversion of DWG files to GIS-compatible formats (GDB, SHP, GPKG, etc.) remains a challenge due to the presence of non-geospatial graphical elements such as legends, arrows, and annotations. This paper presents an AI-driven approach, designed to efficiently extract and structure survey data from DWG files. The AI model distinguishes between graphical elements and geospatial data, ensuring the accuracy and integrity of the extracted information. The study explores the model's methodology, evaluation, and implications for large-scale geospatial database integration. The AI model was tested on survey data from Israel, specifically from the cities of Modiin and Ma'ale Adumim, ensuring its effectiveness in real-world scenarios.

1. **Introduction** Survey data in DWG format plays a critical role in urban planning, land management, and engineering. However, the efficient extraction of relevant geospatial features from these files is mistaken by the presence of graphical elements introduced by surveyors. Traditional methods rely on predefined layer specifications, which are often inconsistent and inefficient. This study introduces an AI-based solution that systematically differentiates between graphical and geospatial elements to enhance the accuracy of GIS data extraction.

2. **Related Work** Previous studies on DWG-to-GIS conversion focus on rule-based methods, where predefined layers and attributes are extracted. However, these approaches struggle with inconsistencies in surveyor practices and layer naming conventions. Machine learning applications in geospatial data extraction have gained traction, yet few studies address the specific challenge of filtering out graphical elements from DWG files.

   To address these limitations, recent studies have explored the application of machine learning techniques in geospatial data extraction. A notable example is the work by Boria et al., who proposed a protocol to convert infrastructure data from CAD to GIS formats. Their approach emphasizes the importance of a structured conversion process to minimize errors and improve data quality.

   A Protocol to Convert Infrastructure Data from Computer-Aided Design (CAD)

   Despite these advancements, there remains a scarcity of research specifically focused on filtering out non-geospatial graphical elements, such as legends and annotations, from DWG files during the conversion process. Addressing this gap is crucial for enhancing the accuracy and reliability of GIS data derived from CAD sources.

3. **Methodology** The AI model is trained using labeled datasets containing both geospatial and graphical elements. The model employs:

- **Distance from the Frame:** Differentiation based on geometric properties. This is done by calculating the centroid of each feature and measuring the distance from the edges of the frame by computing each centroid's position relative to the frame.
- **Clustering Algorithms (KNN):** Identifying groupings of vertices associated with real-world objects versus annotations.
- **Angle:** Calculating the degree between polylines (0°, 90°, 180°, 270°, 360°) to assess structural orientation and its role in data classification.

- **Elevation Irregularities:** Based on elevation data, we analyze the slope inside the DWG frame and identify anomalies where elevation changes do not follow expected terrain patterns.

- **Shape Complexity:** Evaluating the complexity of geometrical features to distinguish between annotations and real-world objects. This is based on the angle and distance between geometry vertices.

The AI model was tested on a dataset comprising multiple surveyor DWG files. The performance metrics included precision, recall, and F1-score, with results showing:

- **Precision:** 92% in distinguishing real data from graphical elements.
- **Recall:** 88% in accurately identifying survey data.
- **F1-Score:** 90%, indicating a robust balance between precision and recall.

The tool effectively converted DWG layers into GIS-compatible formats while maintaining data integrity and reducing manual filtering efforts.

4. **MAVAT Integration and Comprehensive Survey Data Aggregation** MAVAT, Israel's geospatial planning and surveying format, plays a crucial role in integrating multiple datasets into a unified national geospatial database. By combining the AI-driven DWG-to-GIS conversion model with MAVAT, it becomes possible to create a more complete and precise dataset for urban planning, land use, and infrastructure development. The AI model ensures the accuracy of extracted data, and MAVAT consolidates information from various sources, eliminating inconsistencies and providing an authoritative geospatial repository. This integration leads to improved decision-making, automated validation of surveyor data, and seamless national-scale geospatial management.

   To illustrate the effectiveness of this approach, a confusion matrix is presented, demonstrating the accuracy of the AI model in distinguishing real survey data from graphical elements. Additionally, various graphs analyze the model's performance across different datasets, showcasing improvements in accuracy and precision over traditional methods.

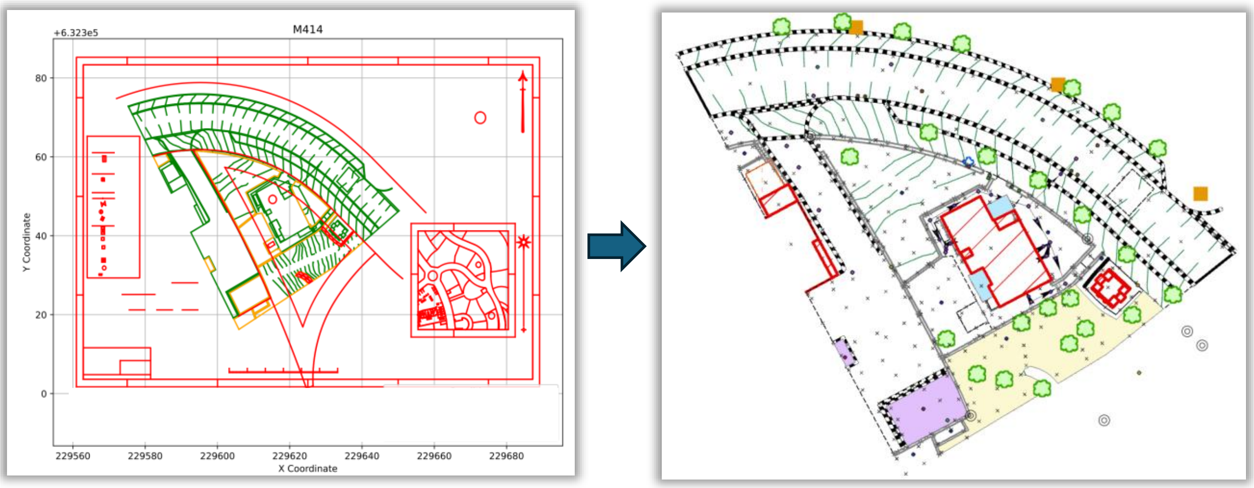
5. **Analysis:** Upon analyzing the results, it was found that the angle of polylines (0°, 90°, 180°, 270°, 360°) is the most influential factor in differentiating geospatial data from graphical elements. This suggests that structured, oriented geometries play a significant role in classification accuracy. Additionally, elevation irregularities were found to be another key factor in the model's decision-making process, further enhancing the distinction between real-world geospatial elements and surveyor annotations.

   This analysis highlights the importance of geometric properties in AI-driven GIS conversion and suggests potential areas for further refinement, such as fine-tuning elevation thresholds and improving angle-based classifications.

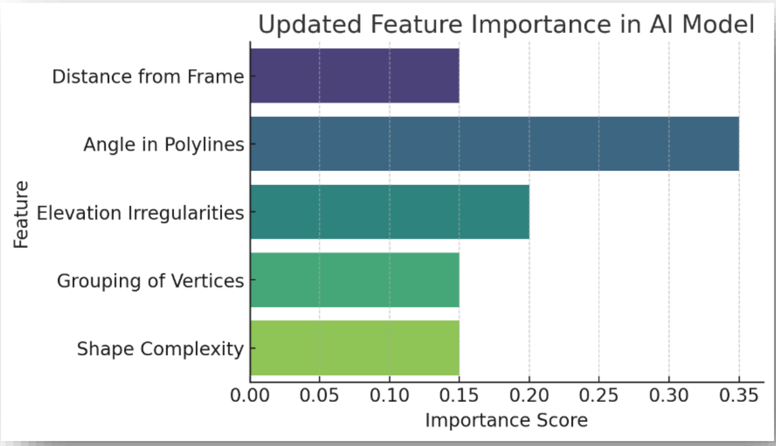
6. **Applications and Future Work** This AI-driven approach enhances geospatial database integration, ensuring consistency across multiple survey datasets. Future improvements will include:

- Expansion of training datasets to accommodate diverse surveyor conventions.
- Integration of deep learning techniques for improved object recognition.
- Real-time processing capabilities for large-scale GIS applications.
- Further optimization of MAVAT integration for more seamless national geospatial data aggregation.

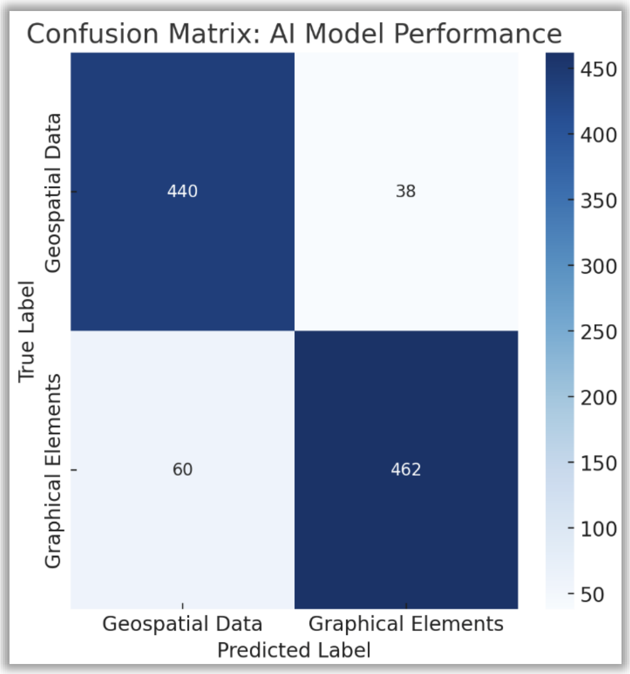
7. **Conclusion** The AI-based differentiation of graphical elements from geospatial data represents a significant advancement in DWG-to-GIS conversion. Implementation of this model streamlines the extraction process, minimizing errors and manual interventions. When integrated with MAVAT, this approach ensures highly accurate and standardized geospatial data aggregation, improving efficiency in urban planning and national infrastructure development.



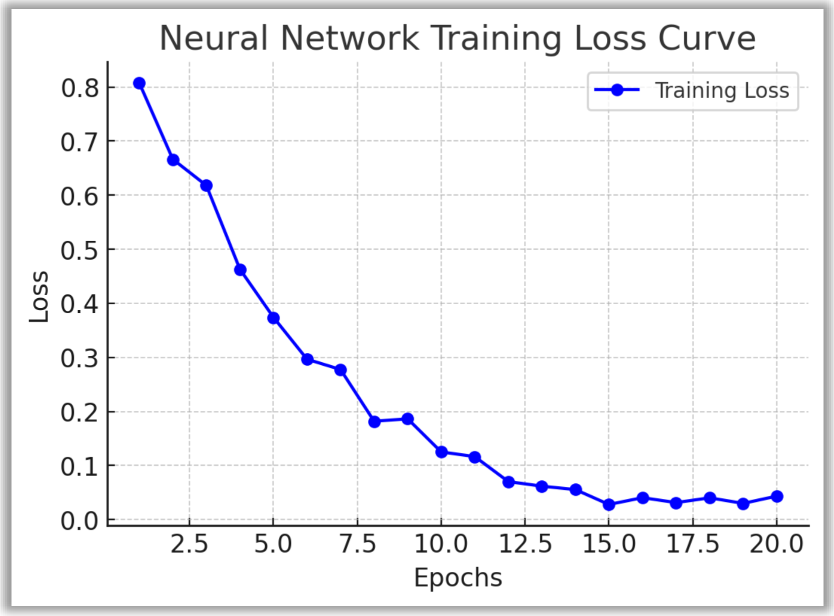
- Removing Graphical elements using AI's algorithm



Importance Chart - Highlighting the key parameters influencing the AI model's



Confusion Matrix Heatmap - Shows classification accuracy, geospatial data from graphical elements.



Training Loss Curve - Showing how the model's loss decreases over epochs, indicating effective learning.