
Learning Semantic Structure from Industrial Vector CAD Floor Plans

Hatef Rahimi

Supervisors: Matthias Zinnen, Prof. Dr. Andreas Maier, Christian Petry (Siemens Healthineers AG)
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
Siemens Healthineers AG, Enterprise Services-Facility Design and Planning Services

1. Introduction

Planning a medical imaging installation begins with a 2D DWG floor plan supplied by the local planner. A Siemens Healthineers planner then re-draws that plan in HSCAD, a customized CAD software solution provided by the Siemens Healthineers AG planning department, replacing flat line work with parameterized 3D objects such as walls, doors, and windows, as well as Siemens Healthineers medical equipment, so that equipment clearances, shielding, and service access can be checked in three dimensions. The step is manual, takes time per project, and is repeated for every plan that arrives.

Over the years the department has collected several thousand drawings. Each one holds a decision an expert has already made: the original flat 2D lines and the planner's 3D objects sit in the same file, showing which lines were understood to be a wall and which were not. This means the drawings could serve as training data. In their current form, however, they cannot. Layer names differ by customer, by language, and by drafting office; the geometry is often plain lines, arcs, and polylines with nothing marking what any of them represent; some drawings hide their entire content inside block references, and others use layer names that say little. The archive was built to be read by people, not interpreted by software.

2. Ultimate goal and objectives

The long-term aim is to make this collection useful and to turn a folder of drawing files into something that can help interpret the next plan that arrives. If the individual elements of a drawing can be recognized for what they are, several applications become possible: generating 3D walls, doors, and windows in one step; recognizing rooms and zones automatically; checking drawings for consistency; and searching across past projects. Wall generation is the natural starting point, since walls are the most common element and the most time-consuming to re-draw by hand. It is a first target rather than the endpoint.

3. Research question

The central question is whether this collection can be used to train a model that recognizes what each element of a drawing represents, and which kind of method is best suited to the task.

This is not obvious in advance, because the same physical wall is drawn in different ways by different people. One planner represents a wall with two parallel lines, another with four; a wall may run in a zig-zag, and the gap between two segments may be a window, an opening, or nothing structural at all. Fixed rules and thresholds break on this kind of variation, which is what motivates a learned approach. Whether a learned model can absorb it depends on two things this thesis sets out to test:

- whether the collection contains enough clearly labeled drawings and enough variety of drawing styles;
- whether the geometry itself carries the distinction at all. A partition wall and a load-bearing wall may be indistinguishable in isolation, separable only from the surrounding context.

A second question follows from how such a system would be used. Layer names may provide useful information during training but are unreliable in practice, so the system should not depend on them for successful classification. Does the structure of a floor plan carry enough signal to classify its elements primarily from geometry and context?

4. Approach

Several directions are available, and part of the work is deciding which of them this data actually rewards. They differ mainly in how the drawing is represented before a model ever sees it.

The most direct option keeps the drawing in its original vector form and treats each element as a node in a graph, connected to the elements near it. A graph neural network can then learn relevant relationships, such as two lines running parallel a short distance apart, lines meeting at a corner, or a sequence of segments closing a room, instead of having these patterns explicitly defined by hand.

A different option abandons the vector data and draws the plan as an image, letting a vision model label it pixel by pixel. This is closer to how a person reads a floor plan, and it opens the door to models already trained on public floor-plan datasets, which could be adapted to these drawings rather than trained from scratch. Whether that transfer holds is itself worth testing, since public datasets consist mostly of ordinary buildings and these are medical facilities.

A third direction is to combine both representations in a multimodal model, using the vector structure and rasterized image as complementary inputs and fusing their representations within the network. Such a combination may preserve geometric

information from the original CAD file while also benefiting from spatial patterns that are easier to capture in image space. Which of these directions can be carried through to a full comparison will depend on the available data, early experimental results, and the scope of the thesis.

5. Data and labels

Running through all of this is the question of what the labels rest on. The drawings were never annotated or manually curated for machine learning, so labels have to be recovered: from layer names where those are meaningful, from geometry hidden inside block references, and from the position of the planner's own 3D objects where the layer names are sometimes ambiguous. These sources may therefore provide useful but noisy supervision rather than reliable ground-truth labels.

Layer names and other available metadata may also be investigated as an additional source of information during training, for example as an auxiliary input modality or as weak labels. Their usefulness can then be compared with models that rely only on geometry and spatial context.

Label quality is likely to be a limiting factor in its own right, possibly more than the choice of model, so it is treated as part of the research rather than as preparation for it.

How far the collection carries the work is therefore an open question, and one the thesis has to answer early. If too few drawings turn out to yield reliable labels, publicly available floor-plan datasets can be used to supplement them for pre-training, or to establish what the methods achieve on cleaner data before applying them here. The comparison between the two would itself be informative, since it separates the limits of the methods from the limits of this particular collection.

6. Evaluation

Primary metrics are per-element precision, recall, and F1 score for the wall class, measured on drawings held out from training so that generalization across customers and drafting conventions is tested rather than memorization within a single drawing.

Doors and windows form a secondary and harder target, provided that sufficiently reliable reference labels can be recovered or established for their evaluation. Since the available drawings were not manually annotated for machine learning, the feasibility of evaluating these classes will depend on the quality of the available labels.

A final check asks whether the detected geometry can be turned into valid 3D objects in HSCAD, which is what ultimately determines whether the result is of practical use. The exact form of this evaluation will depend on the resulting implementation and available reference data and may therefore be quantitative, qualitative through inspection of the generated objects, or a combination of both.