A COMPARISON OF TOP-DOWN AND BOTTOM-UP APPROACHES TO RECOGNIZING COMPONENT ASSEMBLIES IN IMAGE MINING ELECTRONIC CIRCUITS

William J. Turkel & Zain Sirohey
The University of Western Ontario
Canadian Society for Digital Humanities 1 June 2020
Contact: william.j.turkel@gmail.com
The work we present here is part of an ongoing effort that involves collecting millions of pages of source documents from the history of electronics and using text and image mining to explore them (Jones-Imhotep and Turkel 2019).

The problem of image mining circuit diagrams is an example of the larger problem of automating the understanding of complex engineering diagrams of all kinds (Moreno-García, Elyan, and Jayne 2019).

Surprisingly little work has been done on the specific problem of electronic circuit schematics, although the wider problem of understanding line drawings has received significant attention.
ELECTRONIC CIRCUIT SCHEMATICS

• Line drawings of various kinds have played a central role in learning and communicating electronic ideas for a century (i.e., for the whole history of electronics).

• An electronic circuit schematic shows the interconnection of various kinds of components. Automating the recognition of these components is one step in a workflow for handling schematics.

• Circuit schematic diagrams evolved with the invention of new technologies. The introduction and adoption of the transistor, for example, was accompanied by dispute over the best way to represent the device schematically.

Transistor representations, late 1950s (Jones-Imhotep 2008)
THE BOTTOM-UP APPROACH

- Preprocess the schematic image to remove noise and unneeded details (such as labels) and binarize
- Segment the component symbols
- Reveal connections between symbols
- Identify symbols and connections
- Extract circuit topology
- Recognize meaningful assemblies of components and connections
- (Adapted and expanded from Bailey et al 1995).
PREPROCESSING

https://www.industrial-electronics.com/eed5th_7.html
Chain coding (Freeman 1974) is a method of representing line drawings by recording the direction of the step to the next pixel on a contour around a region.

In the example shown on the right, starting at the upper left corner of the region, the first step is to the right (0), as is the next (0), followed by a step to the lower right (7), and then to the right again (0) …

If we keep track of the starting position and the chain code (0070055544343420001) we can reconstruct the region exactly.

Source: Bailey 1995
Next, the chain code is vectorized to find primitive features in the image, such as lines and circles.

These chain code vectors are shown here in alternating colours.
The vectors are separated into component and connection vectors using an angle-based length threshold.

Next, a multiple-pass clustering algorithm groups vectors into components.

Once the vectors are grouped into components, the connections become immediately available.
A simple neural network (based on the LeNeT model) is trained to identify components.

This method is more effective than the rule-based identification technique of Bailey et al (1995) and makes it simpler to detect complex components with curves or complex shapes (e.g. inductors, transistors, lamps).
• Connections are in a vectorized format, which makes them easy to process. We classify each end point depending on the number of adjacent end points.
• If 0 (Red) – connects to a component
• If 1 (Green) – connects to another connection segment (called an intermediary)
• If 2 or more (Dark Blue) – connects to more than one other connection segment (called a joint)
Tracing each end point corresponding to a component or joint, until another end point belonging to a component or joint is found
Once the connections between all components and joints have been identified, a connection graph can be extracted.
THE TOP-DOWN APPROACH

• In the top-down approach, we start by analyzing the whole schematic into components and connections, then abstract away from the components to extract the connections and the nodes where they meet.

• Due to the exigencies of the time (i.e., COVID-19) we had to postpone the exploration of this in as much detail as we would have liked!

• Our preliminary work uses transfer learning with pretrained neural nets (e.g., YOLO, Redmon and Farhadi 2018) for localizing and detecting symbols.
ONGOING AND FUTURE WORK: UNSUPERVISED CLUSTERING

• Parsed, labelled and hand-corrected schematic images can be used to train machine learners for various tasks including unsupervised clustering.

• The method presented here can be readily extended to other kinds of technical line drawings such as flow charts and block diagrams.
ONGOING AND FUTURE WORK: SEARCHING FOR DESIGN IDIOMS

- Historians of electronics are typically interested in meaningful assemblies that occur above the level of individual components.
- Four resistors connected in a particular topology, for example, form a Wheatstone bridge, a circuit with many applications in instrumentation, filtering, etc.
- Automated understanding of these higher-level structures can be incorporated into semantic image search (e.g., searching for “Wheatstone bridge” instead of “resistor”).
- This is analogous to text mining, which becomes much more useful at levels of analysis above the individual character.
ONGOING AND FUTURE WORK: NON-DETERMINISTIC ANALYSIS

• Bottom-up and top-down approaches can be combined into a non-deterministic analysis system (Henderson 2014).

• Henderson describes the use of independent software processes (agents) that make autonomous analyses of the image, communicating and cooperating to produce a final understanding.

• For example, in the work presented here we eliminated schematic labels during image preprocessing to make identification of component symbols easier. This information is necessary for a complete understanding of the circuit, however. Another subsystem might start with the original image, extract those labels and apply OCR. Label information could then be fed back into the understanding of the circuit at a later stage.
REFERENCES


