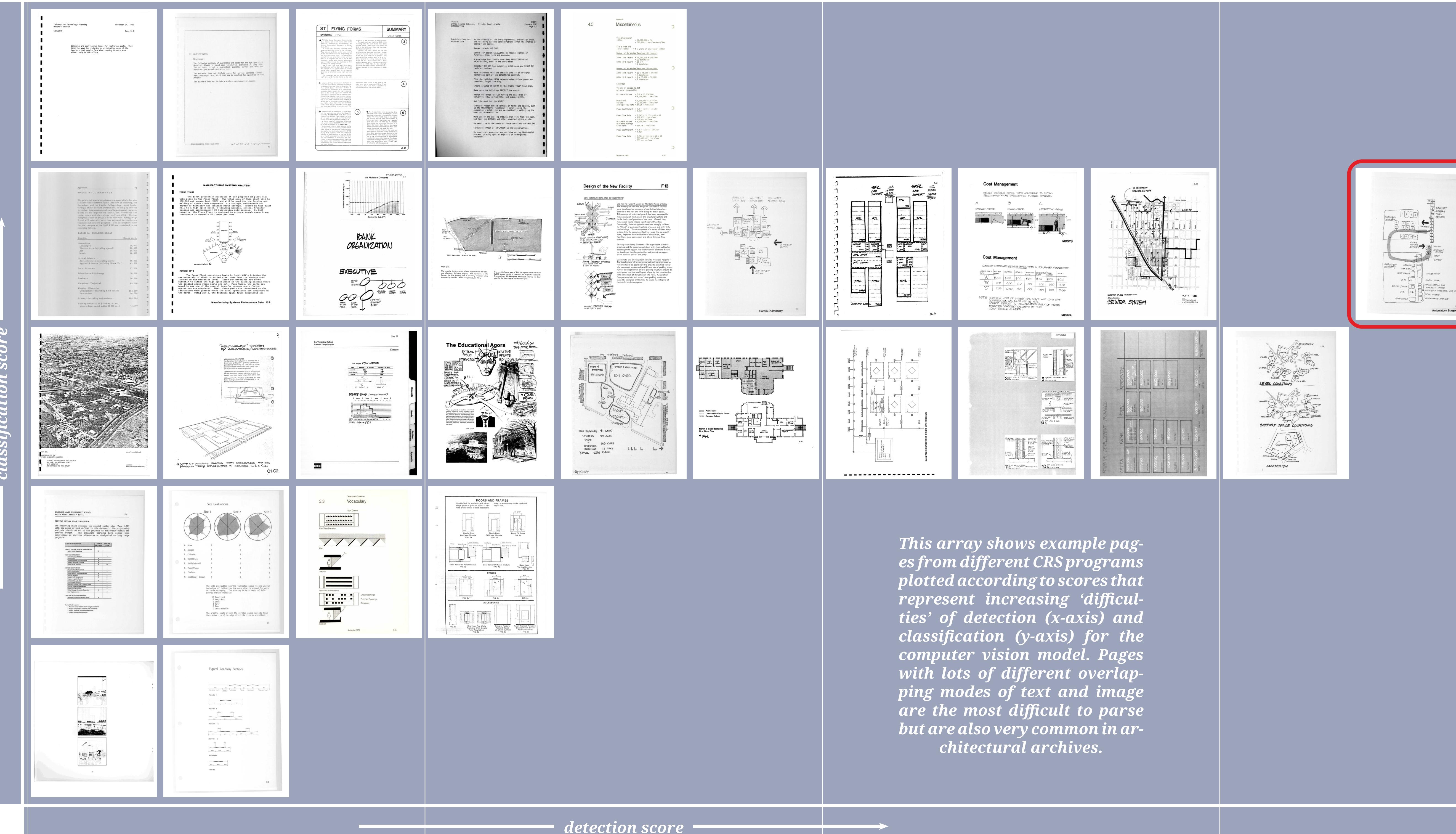
SAH 2024 Albuquerque 77th Annual International Conference of the Society of Architectural Historians

Project in Process: Machine Learning in the CRS Architectural Archive

Andrew R. Tripp, PhD and Amir Hossein Oliaee Department of Architecture Texas A&M University

Architectural histories of the recent pastare challenged by the overwhelming quantity and complexity of documentation. This is especially the case for histories of large professional practices. Researchers at Texas A&M University have addressed some of these challenges by introducing machine learning-based data practices into the processing of the CRS Archives, which holds, among other things, the largest collection of historic architectural programs from the second half of the twentieth century.



Corresponding Author: Andrew R. Tripp, atripp@arch.tamu.edu

Citation: Tripp, Andrew R., and Amir Hossein Oliaee. 2024. "Project In Process: Machine Learning in the CRS Architectural Archive." Presented at the 77th Annual International Conference of the Society of Architectural Historians, Albuquerque, New Mexico, April 17-21.

The CRS Archives was established in 1989 by the firm of Caudill Rowlett Scott (CRS, 1948-93), which at the time was the largest AEC firm in the world. CRS built its reputation largely on a methodical approach to programming defined as "problem seeking." The result of this process are historical documents that trace the complex relations established at the start of a project.

While the CRS Archives is nominally accessible to the public, most In 2022-23, researchers addressed these challenges by training a of its records, including digital version, remain practically "beyond computer vision model to detect and classify the various graphic the reach of the interested scholar." In the case of its program objects on the pages of the programs, thereby making their muldocuments, this occlusion is largely due to the collection's size timodal contents (text and image) retrievable for the first time (over 1400 historic program documents and more than 250,000 without the need for metadata entry. This research endeavors pages) and complexity (the variety of graphic modes and lay- to make a dark repository more accessible to scholars in the huouts).which altogether pose a number of challenges to existing manities, but it also intends to provoke discussion about the opportunities and risks arising from access to artificial intelligence. document processing techniques.

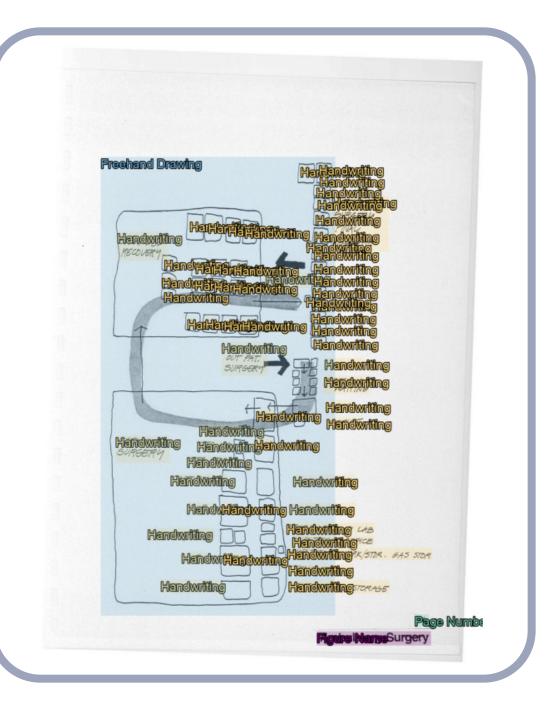
For technical information, see our publication:

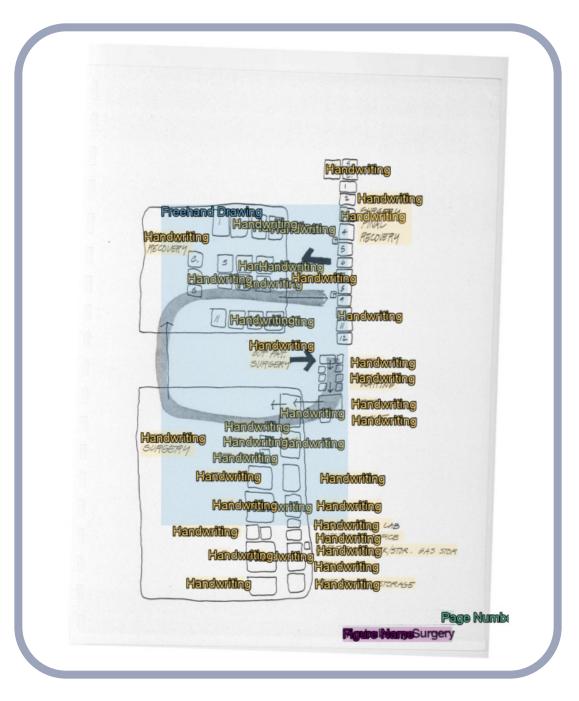
Oliaee, A. H., and Tripp, A. R. 2023. "Layout Analysis of Historic Architectural Program Documents." In *Proceed*ings of the ACM Symposium on Document Engineering 2023 (DocEng'23), August 22-25, 2023, Limerick, Ireland. ACM, New York, NY. https://doi.org/10.1145/3573128.3609339

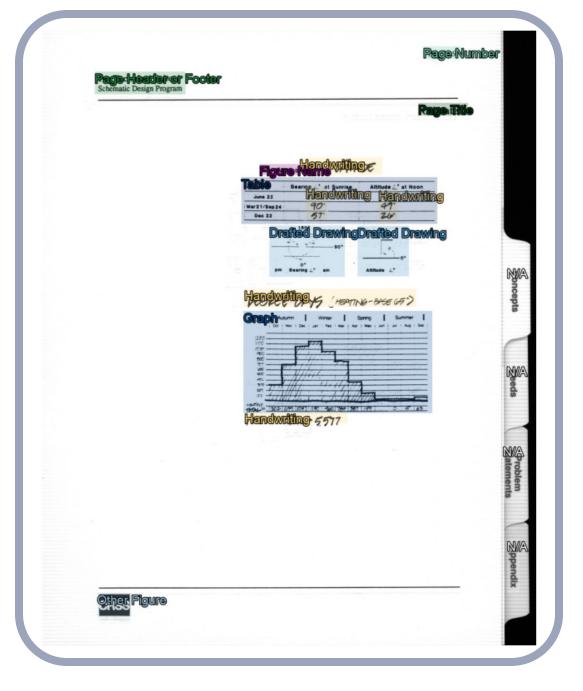
For a demonstration and comparison of the sev-eral computer vision models tested, use the QR code to see our Streamlit app:



Predicted Labels







Pigene dering _* at Sunrise Anne 22 Handwitting Handwitting Des 22 57 2/0 Dreafteid Drewing 00 Pr Baaring _* an Atbuck 2* Hendwitting

jects into 16 categories, and ensurng quality control

the page ob- tuned for document sis. The model predicts the page jects, which makes them sepa retrievable in real time without need for inputting metadata. More 'difficult' pages show errors.



Acknowledgements

Grant. Portions of this research & Society 37, 819–822. were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing. We are grateful to Lincoln Clark-Bateman and Ava Carlson for their hard work.

Notes

This research was supported by a
Texas A&M Triads for Transforma-
tion (T3) Grant and a CRS Probes1. Stapleton, L. and Jaillant, L. 2022.
"Born digital' shedding light into
the darkness of digital culture," AI

