AI-Pipeline Integration for Historic Plan Digitisation

Custom desktop app for digitising landscaping plans using deep learning models

Graduate

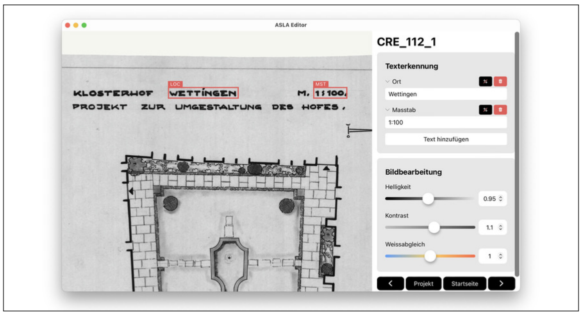
Kevin Löffler

Initial Situation: The Swiss Archive for Landscape Architecture at OST in Rapperswil houses over 100'000 historic plans. This collection includes works from prominent Swiss and European architects of the 20th century. The archive serves as a vital reference for contemporary architects, in addition to supporting teaching at OST. Currently, the archive manually digitises its documents, a process that is highly time-consuming. This thesis automates the digitisation process. The solution involves two main components: A transformer-based deep learning and image processing pipeline to locate and extract text, identify relevant entities such as client, date, or scale of a plan, and a desktop app to manually edit and refine the AI-generated output.

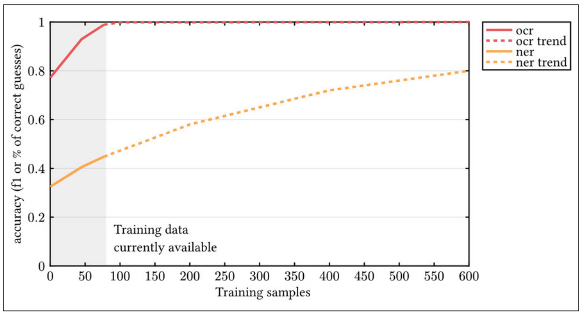
Approach / Technology: The desktop app is developed using Tauri, a cross-platform framework that integrates SvelteKit for the GUI and Rust for the backend. This setup allows computationally intensive tasks like image white balance adjustments to be handled asynchronously by the efficient Rust backend. Users can manage estates (projects), upload images, and visually inspect and override AI predictions. A global environment file holds all configuration settings, enabling simultaneous work by different employees.
The AI pipeline is dockerized and exposed via a FastAPI web server. Upon image upload, the plan's size and position are determined, followed by preprocessing steps such as contrast normalization, noise reduction, adaptive thresholding, and morphological operations. Three deep-learning models are then applied: a pretrained layout model (LayoutLMv3) to detect all text occurrences, K-means clustering to group text boxes into logical blocks, and a transformer-based OCR model (TrOCR) to extract text. Relevant entities are then identified using a custom-trained German BERT model. The output undergoes post-processing for formatting and normalization, with project-specific keywords like the architect's name filtered out. The predicted metadata is sent back to the client app, which saves it into the corresponding working directory. Metadata files track all changes to the image, ensuring non-destructive editing. Only when a project is uploaded to the archive website is a copy exported, keeping the original image as well as all changes available. Every weekend, images with manually corrected predictions are automatically uploaded to the server, and the models are retrained on the corrected data and evaluated on a test set. If a model's performance improves, it is automatically updated. The thesis focuses more on implementing a robust pipeline and continuous retraining than on improving the models because the continuous retraining is expected to enhance the AI pipeline's performance over time.

Result: The app significantly speeds up the digitisation process for the archive and is an improvement over the old Excel-based workflow. The AI pipeline's prediction accuracy varies by model. Marker detection is 100% reliable, and the OCR model reaches 98% accuracy after retraining on only 77 images. The NER model, currently at 46% accuracy, is about 10% better than the model from the SA project. If the accuracy continues to increase with additional training data, an F1-score of over 80% can be foreseen within 600 images. Thanks to the new app, these images can be collected in less than a month of archival work.

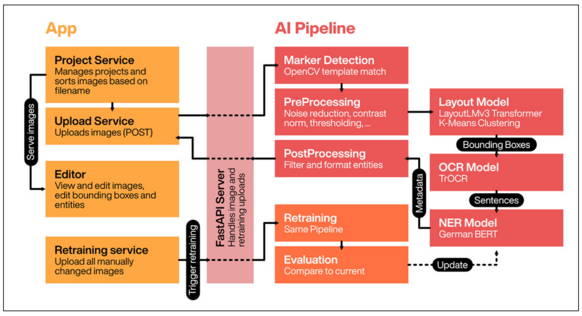
ASLA-Editor: The desktop app
Own presentment



Evolution of model accuracy with more training data
Own presentment



Architecture Diagram
Own presentment



Advisor
Prof. Dr. Mitra Purandare

Co-Examiner
Dr. Marc Röösli, Mettler-Toledo International Inc., Greifensee, ZH

Subject Area
Artificial Intelligence, Software, Application Design