

A Deep Learning Approach for Automated Vectorization of Historical Mining Maps

David Demmers¹, C.P. van Middelkoop¹, Samantha Kim¹, Joana E. Martins¹, Erik van Linden¹

¹TNO – Geological Survey of the Netherlands, Utrecht, The Netherlands

Historical mining maps hold detailed geospatial information, yet most are currently only archived as static images while holding features that can tell us much more about specific areas. [1] In the Limburg province, in the Netherlands, approximately 3500 historical mining maps document all mining activities, but they have never been vectorized in high detail, nor have depth and temporal information been systematically extracted from the maps.

To extract this data we developed a deep learning workflow for automated extraction of underground mining structures from historical coal mining maps. The core component is a semantic segmentation model based on the U-Net architecture. The model performs pixel-wise classification to delineate two primary feature classes: mining panels and galleries, together with a background class.

The dataset consists of approximately 3,500 georeferenced rasterized maps documenting about 150 years of mining activity in Southern Limburg. The collection, curated by the Geological Survey of the Netherlands - TNO, contains maps produced by multiple private mining operators with varying cartographic notations. All maps had been digitized and georeferenced prior to this study. From this corpus, 2,000 image patches (4096 × 4096 pixels) were manually annotated with semantic masks for panels, galleries, and background. In order to evaluate the trade-off between spatial detail and contextual information the annotated dataset was prepared in three resolutions: the original resolution, half resolution and quarter resolution. Experiments showed a consistently better model fit for both primary feature classes when trained on the half-resolution dataset.

We trained a custom U-Net using the Dice Similarity Coefficient loss on the training data, which consists of 70% of the data set. While monitoring the model performance on unseen validation data using both the Dice coefficient and the Jaccard index. Both metrics are used to ensure a more robust validation. The validation data consists of 21% of our dataset. In addition to supervised validation a separate self-supervised inspection of model outputs is used to identify error rates and uncertainty patterns across the corpus when compared to a rough manual vectorization of previously annotated coal layers.

Evaluation using the Jaccard index (Intersection over Union, IoU) yielded scores of 0.985 for background, 0.896 for panels, and 0.676 for galleries. Corresponding Dice Similarity Coefficient (DSC) scores were 0.993 for background, 0.942 for panels, and 0.803 for galleries. The model demonstrates high segmentation accuracy for the features, while gallery structures do remain comparatively more challenging due to their smaller spatial extent and greater visual variability across maps.

The results demonstrate that the proposed U-Net based workflow can accurately extract mining panels and galleries from heterogeneous historical coal mining maps, producing high-quality raster representations. The workflows generated geometries are suitable for further digitization and spatial analysis, performing well across our diverse corpus spanning multiple operators and conventions.

Overall, the results represent a significant step toward the automated digitization of historical mining maps. Further improvements will likely benefit from training on a broader and more varied dataset to better capture stylistic variation across the corpus. Important attributes such as excavation dates and depth measurements, currently embedded as textual annotations, will be addressed in future work where we leverage Vision Language Model (VLM) based text extraction methods.

Bekendam, R. F. (2020). A reanalysis of the collapse of the Heidegroeve: subsidence over an abandoned room and pillar mine due to previously unknown mine workings underneath. *Proceedings of the International Association of Hydrological Sciences*, 382, 269–275.
<https://doi.org/10.5194/piahs-382-269-2020>