International Cartographic Association
Association Cartographique Internationale

Exploring the link between neural network metrics and the completeness and connectivity of predicted watercourses.

Christian Koski *, Pyry Kettunen, Anssi Jussila

Finnish Geospatial Research Institute (FGI, NLS), Christian Koski – christian.koski@nls.fi, Anssi Jussila – anssi.jussila@nls.fi, Pyry Kettunen – pyry.kettunen@nls.fi

* Corresponding author

Keywords: Neural networks, Data quality, Metrics, Watercourses, Connectivity

Abstract:

There has been a rapid increase in research applying deep learning methods in extraction of geospatial features from remote sensing data. In such applications, small watercourses, such as streams and drainage ditches, have been a particular interest in recent years (Stanislawski et al., 2021; Lidberg et al., 2023; Koski et al., 2023). During the training of the convolutional neural networks (CNNs) that are used at the core of these methods, results are often evaluated and reported with machine learning metrics, such as f1-score, recall, and precision. The metrics are calculated on a pixel level, by comparing the model predictions with manually digitised labels. However, the raw predictions from the network are often not used as such. In use cases such as maps and hydrographic network analysis the data requires vectorisation and generalisation (see Jussila, Koski and Kettunen, 2024). Currently, it is unclear how well machine learning metrics represent data quality of the derived vector datasets, for example dataset completeness (are all features in the dataset?), feature completeness (are individual features complete?) and connectivity (are networks connected correctly?). Individual false positive and false negative errors may result from phenomenon that have little impact on the quality of the vector dataset (see Koski et al., 2023). A better understanding of how well machine learning metrics represent the quality of spatial data can help to develop understanding of how CNN results that are used for spatial data extraction should be assessed.

In this abstract we present early concepting and results from investigating the relationship between machine learning metrics and spatial data quality metrics when using deep learning to extract watercourses. We vectorised predictions from a CNN and compared them to the manually digitised data that was used to create the labels. The dataset and feature completeness, and connectivity of the vectorised results for five areas were investigated through various indicator values, that were then compared to recall. The predictions were vectorised using the process described in (Jussila, Koski and Kettunen, 2024). The long-term aim of this work is to understand how results from CNN methods used for spatial data extraction should be assessed during the method development phase, and whether integration of spatial data quality metrics into the method is needed. A secondary motivation for this work is to spark discussion about how quality of data is reported from machine learning methods for extraction of geospatial features.

Our results from the initial analysis on completeness showed that changes in recall are quite consistent with changes in features found completely (Table 1). The results from the initial analysis on connectivity show that recall does not precisely predict the number of gaps per kilometre (km) of watercourses (Table 2). For example, the Suonenjoki 1 area has more gaps/km than Suonenjoki 5, despite having a significantly higher recall. Nevertheless, the results would, as expected, also suggest that there is still some association between recall and gaps/km. The presented comparisons are a first step in building an understanding of how machine learning metrics from CNN predictions reflect the spatial data quality of extracted vector features. Understanding this relation will help in understanding the relevance of different machine learning metrics when developing appropriate quality assurance for CNN-based spatial data extraction processes. For more insight, the data still needs to be tested for correlation. We will continue the analysis by comparing precision to the extracted vector dataset. In addition, mean length of missing and extra features will be considered.

https://doi.org/10.5194/ica-abs-10-153-2025

Table 1. Results from comparing recall to percentage of vector features that were found completely, the percentage of vector features that were fully missing and the percentage of vector features that were partially found.

Area	Recall	% of vector features found completely	% of vector features fully missing	% of vector features partially found and not fragmented	% of vector features partially found and fragmented
Suonenjoki 1	0,863	74,2	6,27	12,6	6,94
Suonenjoki 2	0,764	57,6	11,8	17,3	13,4
Suonenjoki 3	0,744	54,6	11,8	16,8	16,7
Suonenjoki 4	0,716	54,0	14,5	16,3	15,2
Suonenjoki 5	0,777	60,3	11,2	16,5	12,0

Table 2. Results from comparing recall to the number of gaps that result in separate watercourse sections per kilometer (km) of watercourses in the training dataset.

Area	Recall	Total length of manually digitised watercourses (km)	Sections in manually digitised data	Gaps in area that result in separate sections	Sections with gaps	Gaps per km of watercourses
Suonenjoki 1	0,863	89,5	230	160	63	1,79
Suonenjoki 2	0,764	76,2	215	160	56	2,1
Suonenjoki 3	0,744	75,1	270	213	75	2,84
Suonenjoki 4	0,716	55,5	148	144	44	2,60
Suonenjoki 5	0,777	50,7	116	90	26	1,78

The method does not account for possible improvements to the dataset after calculating machine learning metrics. For example, post-processing steps that improve connectivity and completeness of the network by filling gaps with additional data. Also, we did not consider how meaningful the gaps are for the network. In addition, other spatial data quality indicators, including positional accuracy, were not considered. The results can also be influenced by errors in the manual digitising and the types of watercourses and terrain in the areas. The method can later be refined and extended to provide more robust and precise results.

Acknowledgements

This research has been conducted as part of the development work of the Basic Spatial Data Services of the National Land Survey of Finland and using the CSC high-performance computing by Geoportti.

References

Jussila, A., Koski, C., and Kettunen, P., 2024. Vectorization of watercourse detection from neural network. In: Abstracts of the ICA, Vol 7, 65. https://doi.org/10.5194/ica-abs-7-65-2024

Koski, C., Kettunen, P., Poutanen, J., Zhu, L. and Oksanen, J., 2023. Mapping Small Watercourses from DEMs with Deep Learning—Exploring the Causes of False Predictions. In: Remote Sensing 2023, Vol 15(11), 2776, pp. 18. https://doi.org/10.3390/rs15112776

Lidberg, W., Paul, S. S., Westphal, F., Richter, K. F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A. and Ågren, A. M., 2023. Mapping Drainage Ditches in Forested Landscapes Using Deep Learning and Aerial Laser Scanning. In: Journal of Irrigation and Drainage Engineering 2023, Vol 149(3). https://doi.org/10.1061/JIDEDH.IRENG-9796

Stanislawski, L. V., Shavers, E. J., Wang, S., Jiang, Z., Usery, E. L., Moak, E., Duffy, A. and Schott, J., 2021. Extensibility of U-Net Neural Network Model for Hydrographic Feature Extraction and Implications for Hydrologic Modeling. In: Remote Sensing 2021, Vol 13(12), 2368. https://doi.org/10.3390/rs13122368

https://doi.org/10.5194/ica-abs-10-153-2025