Article no. 20

THE CIVIL ENGINEERING JOURNAL 2-2024

DIGITALIZATION OF IRRIGATION SYSTEMS FROM WATER MANAGEMENT MAPS

Adam Tejkl and Petr Kavka

Czech Technical University in Prague, Faculty of Civil Engineering, Thákurova 7/2077, 166 29 Praha 6 - Dejvice, Czech Republic, email: adam.tejkl@fsv.cvut.cz

ABSTRACT

With the increasing intensity of evapotranspiration caused by the changing climate, there is a growing need for water. This is especially true in locations where water-intensive vegetables are grown in intensive agriculture. Historically, irrigation systems were built in many intensive agriculture areas in Czechia, but they fell out of use, and evidence of their location was lost. However, Water Management Maps, which were only issued in paper form and have never been fully digitized, can provide evidence about the location of these large-scale irrigation systems. In this paper, we present a method for digitizing irrigation systems using the segmentation and classification of individual segments in the ArcGIS environment. The resulting raster is converted to polygons and is blended with the Land Parcel Identification System layer, resulting in a layer of irrigated land. Two statistical analyses were performed on this layer: statistics of the areas corresponding to the individual source watercourses, and statistics of the type of source.

KEYWORDS

Python, ArcGIS, Segmentation, Machine learning, Water Management Map, Climate change

INTRODUCTION

The social-economic changes following the end of the communist period in Eastern Europe led to the privatization of agriculture and to a decline in vegetable production, resulting in significant loss of data related to irrigation systems [1], [2]. The transfer of the irrigation agenda to various new authorities in the early 1990s and the abolition of the State Reclamation Administration resulted in the loss of entire archives containing project and operational documents for irrigation systems [3]. Moreover, the generational change of workers in agriculture and in government offices has further increased the unavailability of information sources [4], [5].

The extensive and planned construction of large-scale irrigation systems also came to an end in the early 1990s. The best remaining record of the irrigation systems at their greatest extent is preserved in the Water Management Maps, which were last updated in 1997 [6].

The prolonged drought between 2015 and 2017 highlighted the insufficient state of the information resources related to irrigation systems [7], [8]. Reanalysis of the irrigation systems from the Water Management Maps offers a way to obtain topographical, hydrological, and pedological data on implemented irrigation systems, even if they are no longer in operation [9].

Information on the location of the irrigation systems can help with planning, and with allocating resources when irrigation structures are restored or reconstructed. Large-scale irrigation systems were often implemented in areas with a historical need for irrigation, or in areas with a proven recurrence of dry years [10], [11]. Integrated irrigation solutions are more efficient than separate solutions for locations of limited size [12], [13], [14]. Large-scale irrigation systems also allow agricultural enterprises specializing in growing vegetables to rotate their crops without the need to build new irrigation systems [15], [16].

DOI 10.14311/CEJ.2024.02.0020

Information about the location of large-scale irrigation systems can be gathered from historical maps. These maps are in raster form, and automatic image classification is a valuable tool for speeding up the identification process.

Classification of images into groups with multiple labels is a fundamental but challenging task in computer vision. Remarkable progress has been made recently by predicting image labels with deep convolutional neural networks [17]. Traditional machine learning techniques have achieved good performance for terrain perception; however, most of the techniques require manually designed classifiers [18]. Many studies and research efforts follow the conventional pattern recognition paradigm, and some methods exploit a Convolutional Neural Network to conduct the classification task [19]. Data from RGB sensors classified using artificial, deep neural networks are often used in terrain recognition tasks. The results of this earlier work demonstrate the performance of artificial neural networks in terrain recognition tasks, and provide some hints on how to improve classification in the future [20].

Information extraction from historical maps using a Convolutional Neural Network for the recognition task was applied to extract human settlement symbols in the United States Geological Survey historical topographic maps [21]. Convolutional Neural Networks were used for digitizing historical maps to extract vector shapes of the objects of interest from raster images of maps. Convolutional Neural Networks provided efficient edge detection and filtering [22]. Techniques of deep learning were applicated for extraction of information from historical maps in an automated manner on the Early Twentieth Century Map series. These maps utilize standardized symbology and conventions, which greatly enhance the effectiveness of the method. The results obtained demonstrate that deep learning serves as an efficient tool for recovering georeferenced information represented in the form of conventional signs or lines. The method has consistently yielded excellent outcomes, particularly when sufficient training data is available. It performs optimally when applied to large map series that can furnish ample information for training. Deep learning approach offer the potential to map features across entire map series with significantly improved speed and coherence compared to other available methods [23].

The aim of our work is to find a way to usefully digitalize the irrigation systems depicted on the Water Management Maps. The irrigation system should be identified at the level of areas of agricultural land that receive water through these irrigation systems, such as irrigated fields. Furthermore, it should identify the source watercourses and the source types of these systems for example: river, reservoir or well [24]. Digitalized maps of irrigation pipes and pumping stations will later be used as a source for further analyses of systems, their position or irrigated soils.

METHODS

The 1:50,000 Water Management Map of the Czech Republic (hereafter referred to as ZVM) was published by the Czech Land Surveying and Cadastral Office. The map was processed and printed by the Land Surveying Office on 58x47 cm paper in a set of medium-scale basic map sheets. Around 20 map sheets were updated every year.

The map shows the network of watercourses, the distribution points and the hydrological division of the basin, structures in the state observation networks, structures and measures for the use of surface and underground water, protection zones of water structures, structures and equipment of the main water users (e.g., water pipes and sewers, use of water energy, water transport, industry) and other information [6].

Scans of ZVM sheets in tif format were imported into ArcGIS. These are rasters with a pixel size of 3.18 m. and a range of values in one pixel between 0 and 255.

For the purposes of the reanalysis, the most important lines were the lines symbolizing the underground irrigation pipes and the symbols of the irrigation pumping stations. These lines consist of a blue dashed line, with every 3 to 5 lines containing a bump. Due to the manual creation of these maps, the lengths of the spaces and dashes vary, there are different widths, and the shade of the blue color used varies slightly (See Figure 1). However, the lines on any single map sheet are always very similar.

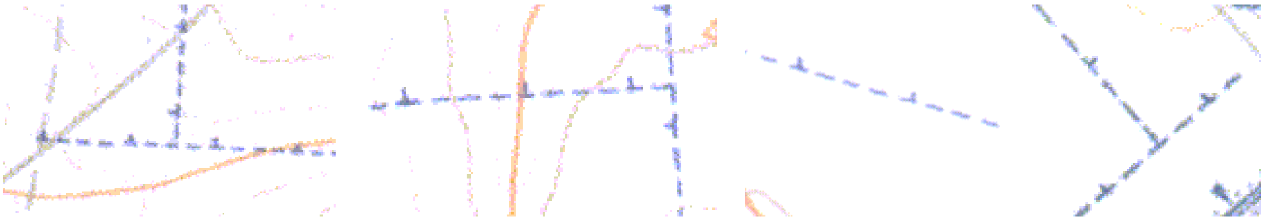

Fig. 1 – Examples of different variants of the irrigation piping symbol

Convolutional Neural Network method.

Our model is based on the winning model of the Kaggle machine learning competition [25] held in 2013, commonly known as Kaggle Cats vs. Dogs [26]. This model takes its name from the competition dataset that consists of photos of dogs and cats. The dataset was originally created as the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). Specifically, the task was known as Animal Species Image Recognition for Restricting Access (Asirra) [27].

The model was specifically designed to handle 25,000 labeled photos [28]. Due to its simplicity and the fact that it can be easily stored in memory, this model has become popular among beginners in the field of computer vision as a starting point for learning about convolutional neural networks. Additionally, this model's architecture is straightforward and adaptable, making it suitable for our purposes. While we considered other models, such as more complex deep learning architectures, the simplicity and robustness of the Kaggle Cats vs. Dogs model made it an ideal choice for the initial stages of our project. As we progress, we may explore more advanced models to further refine our approach.

Pipeline scheme

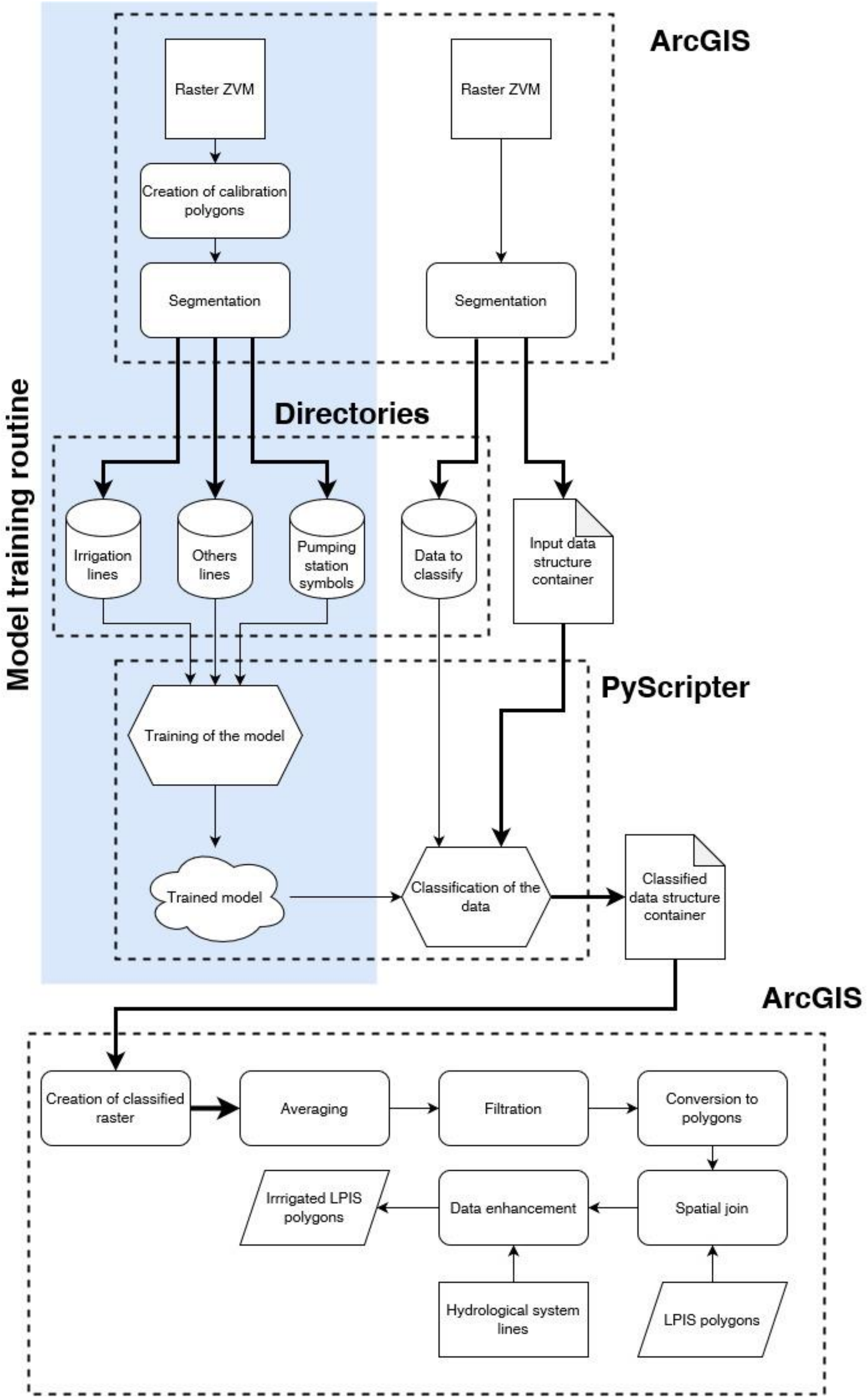

Fig. 2 – Analysis scheme

Input data treatment

Creation of training polygons

During the model training process, polygons were manually created on the georeferenced ZVM list of maps. Three distinct training categories were established: Irrigation, Other area, and Pumping station. The Irrigation Polygons closely replicated the line symbolizing the lines of irrigation pipes, with a slight offset (Figure 3). Creating these polygons proved to be a time-intensive task. Polygons belonging to the Other area class were predominantly rectangular in shape and encompassed areas where the irrigation lines were not visually identifiable. Pumping station polygons were delineated around the corresponding pumping station symbols, ensuring accurate representation within the dataset.

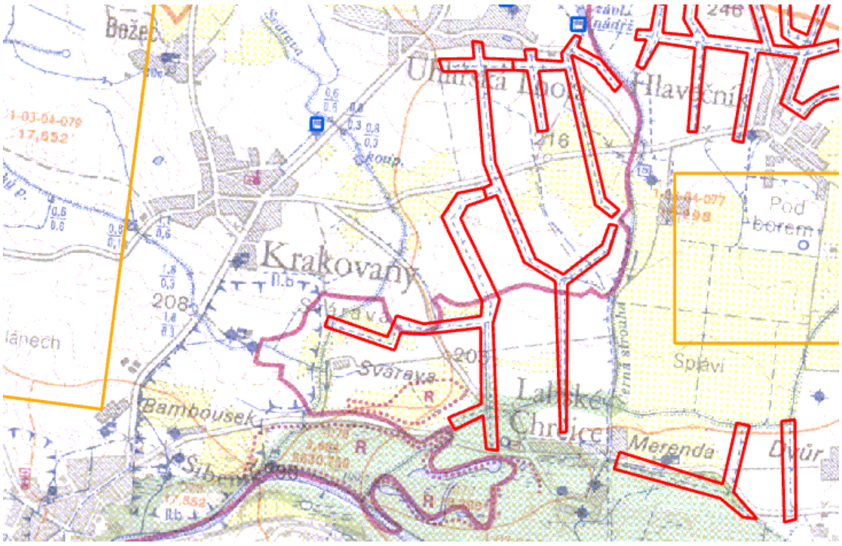

Fig. 3 – Example of training polygons at ZVM, red - Irrigation, orange - Other areas, blue - Pumping stations

Segmentation

Each map sheet is split into segments. The selected segment edge length was set to 31 pixels, equivalent to approximately 98.58 meters. By implementing the Segment shift technique, the resulting accuracy was effectively doubled, resulting in a segment size of 49.29 meters. This size was chosen to accommodate the dimensions of the pumping station symbol, ensuring that the entire symbol could be contained within a single segment. Opting for a larger segment size would have compromised the accuracy of the digitization process, increasing the risk of misclassifying non-irrigated fields as irrigated.

Next, the raster was systematically traversed in steps corresponding to the size of the edges of the individual segments. If training polygons were present within a segment, the segment was marked accordingly.

Segment shift

The shift was applied by half the length of the segment edge either in the horizontal direction, or in the vertical direction, or in both simultaneously (Figure 4). This adjustment generated additional mosaics that could be classified or utilized for model training purposes.

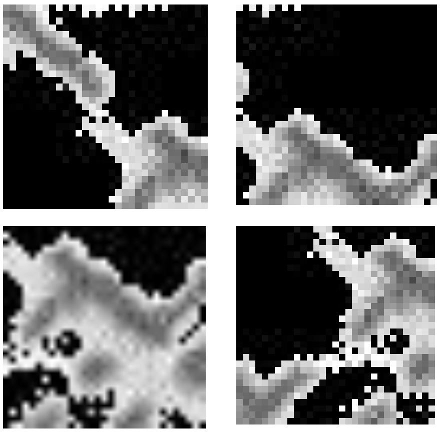

Fig. 4 – Example of segment shift. Segment side length is 31 pixels.

Building and training of the model

Model architecture is a Visual Geometry Group type and consists of blocks, where each block is composed of a 2D Convolution layer and a Max Pooling layer. As the number of layers increases in the Convolution Neural Network, the ability of the model to fit more complex functions also increases. Each layer will use the Rectified Linear Activation Function and He weight initialization, which are generally best practices. In a neural network, the activation function is responsible for transforming the summed weighted input from the node into the activation of the node or the output for that input.

Training the model

The original classification model categories were changed from Cats and Dogs to Irrigation, Other area and Pumping stations and model was trained on the relevant training data. The training data is loaded and split into a calibration part and a validation part for model calibration purposes. This split ratio was chosen to be 20%, which is a commonly used value.

Classification of the data

Individual map sheets not used for model training are classified by the trained model. The probability value with which the segments are classified as Irrigation, Pumping station or Other area is then saved as Classified Data. Training segments are omitted. The sum of these probabilities is equal to one. This process repeats for each classified segment, and for several thousands of segments per map sheet. This is the most time-consuming part.

Treatment of classified data

Averaging

By shifting the segments from each other, four classified rasters are created from a single map sheet. By calculating the arithmetic mean, a raster of average probabilities is created. During this step, the raster is also resampled. The edge length of the pixel of the new raster is 49.12 m, i.e., half the edge length of the classification segment.

Filtration

Filtering was done using Python script. This script first created a new raster with a value of 0. Subsequently, the filtered irrigation raster was scanned pixel by pixel, and the pixels or their immediate surroundings (a 3x3 pixel window) were analyzed. If the conditions for considering a pixel as a pixel containing irrigation were met, a value of 1 was written into the new raster.

The first filtering step was to zero out pixels with a value smaller than 0.5. Next, isolated pixels were zeroed. If a pixel had a value greater than 0.5 and was also in the neighbourhood of a pixel with a value greater than 0.5, a value of 1 was written into the resulting raster. For pixels with a value greater than 0.75 and their immediate surroundings, the resulting raster value 1 was written to the filtered raster (Figure 5).

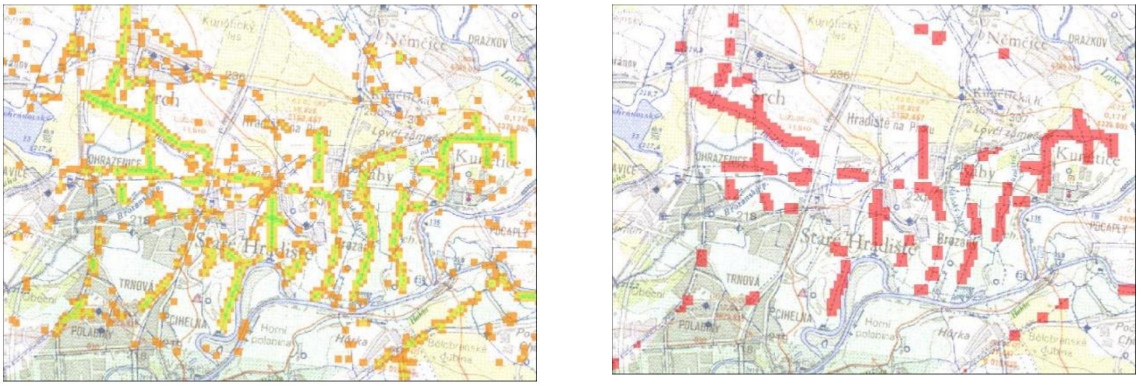

Fig. 5 – Example of filtration. The picture on the left shows pixels with high probability (green), and with lower probability (orange), and the picture on the right shows filtered data (red)

Creation of polygons

The filtered rasters were converted to a polygon. Subsequently, circular buffers with a radius of 1,000 m were created for all polygons with an area greater than 30,000 m2. Polygons with an area smaller than 30,000 m2 and not lying inside the buffers were deleted.

Manual cleaning

The resulting polygons were marked with the cadaster of the municipality where they were located and were subsequently combined into multipart polygons according to the cadaster designation. This was followed by going through this layer manually, from the smallest polygons according to size and erasing of them in cases when the polygon did not correctly mark the symbols of the irrigation pipe.

This is where the problem of the interchangeability of the irrigation pipes and the waste pipes became apparent. The waste pipe is symbolized by the same type of line, except that every fifth piece or so has the opposite orientation. In addition, the drinking water feeder from the Želivka water treatment plant was incorrectly marked, using the same symbology, only doubled (Figure 6).

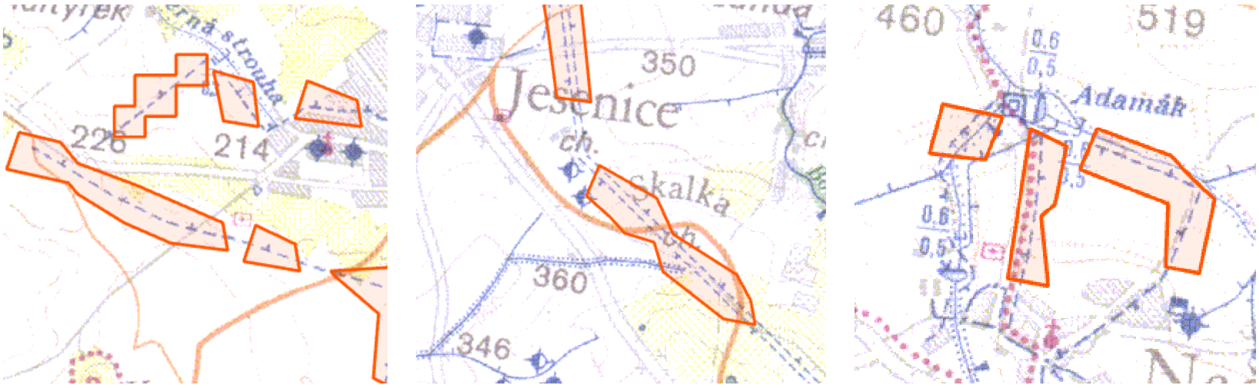

Fig. 6 – Example of resulting polygons, correctly classified irrigation on the left; part of the feeder from the Želivka water treatment plant in the middle; and waste pipe on the right

The result was a layer with more than 2200 polygons, which was converted to a single-part polygon layer and was further edited and refined. The result is 220 polygons.

Information enhancement

First, pumping stations were located and were signed with a point. These pumping stations were then given a name according to the particular irrigation system. The water source for these pumping stations was also found. The identified water sources are river, well, artificial channel, dam, storage tank, and others (Table 1).

Tab. 1. - Attribute table

Column name	Description
System_name	The unique identifier or name of the specific irrigation system polygon.
Source_type	The classification of the irrigation system based on a source type. River, lake, reservoir, well, or other.
Source_name	The name of the specific water source supplying the irrigation system. This could be the name of a river, artificial channel, dam, storage tank, and others.
River_name	The name of the river associated with the irrigation system, if applicable. This identifies the primary watercourse that supplies or is connected to the irrigation network.
Importance	The significance or priority level of the irrigation system.
RAD_III	Name of the third-class watershed in which is the system source located.

Referencing with LPIS data

The polygon layer is then merged with the LPIS layer, showing the agricultural land in use as of 2021. This layer was obtained from the Ministry of Agriculture website (https://eagri.cz). This creates a polygon layer of agricultural land that sits on top of the large-scale irrigation system. Final map is shown on Figure 9.

RESULTS

Efficiency of the model

During the run of the model, a total of 23 658 174 segments were generated, including the segment shift. 136 309 of these segments were used for training. The training dataset consisted of 106 942 Others class segments, 6 835 Pumping Station class segments and 22 532 Irrigation class segments (Fig. 7). This means less than 1% of the total number of segments was used for training the model. The trained model was then used to classify 23 521 865 segments.

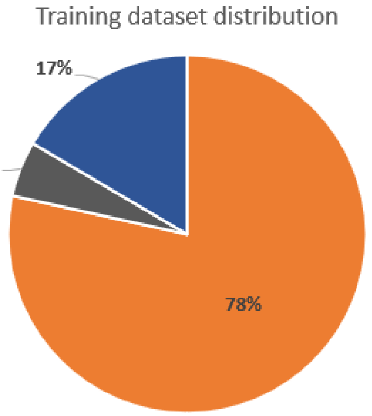

Fig. 7 – Graph showing statistics of the training dataset distribution classes. Class Others (Orange), class Irrigation (Blue) and class Pumping station (Grey).

After the creation and filtration of classified averaged rasters, a total of 344 615 irrigation pixels was reached. That amounts to 83 724,36 ha. These pixels were converted into polygons, and a total of 4783 polygons were created. Due to the Raster to Polygon tool, due to the use of the Simplify Polygon option, the area of these polygons is only 77 524,61 ha. Buffer filtration reduced the number of polygons to 1962, with an area of 58 617,46 ha. Subsequent manual go through and edit resulted in 238 polygons with an area of 69 082,97 ha (Figure 8). This sharp decrease in the number of polygons is created by joining the continuous nets polygons and merging them into the multipart feature.

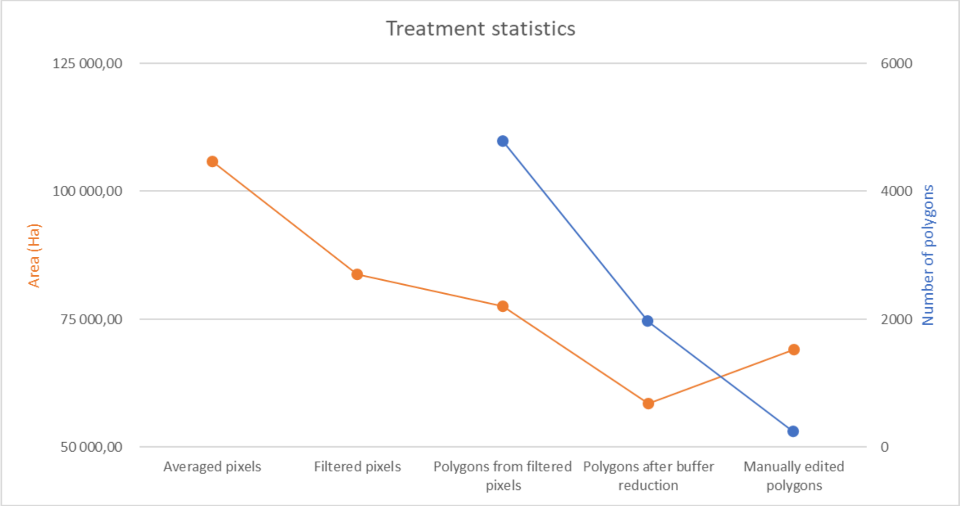

Fig. 8 – Graph showing the development of the number of pixels and hectares during the treatment of the data

Irrigated area and number of irrigation systems

Joining the manually edited polygons with the LPIS database resulted in 17 912 polygons belonging to 198 systems (Figure 9), covering an area of 189 743 ha. Most of the systems are in South Moravia. The rest of the systems follow the course of the Elbe River from East Bohemia to North Bohemia. There is one isolated irrigated region in Northwest Bohemia, in the rain shadow area. Several small systems are distributed all over the territory of Czechia. The only region without an irrigation system is West Bohemia. These results will be compared with further data sources (ISMS, SPÚ) in consecutive papers.

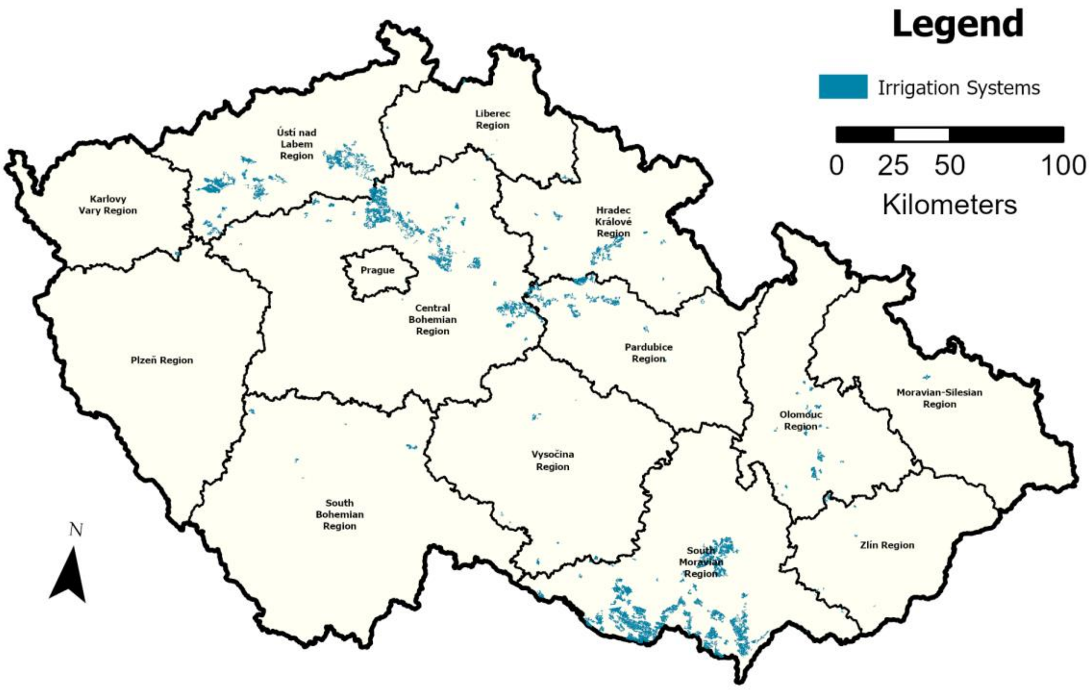

Fig. 9 – Distribution of identified irrigated fields across the territory of Czechia

The model for Digitalization of Irrigation systems is characterized by its high accuracy. This is a natural characteristic of neural networks. However, a lack of training data and the similarity between some symbols limited the use of the model for finding pumping stations. By segmenting the input data, the neural network processed smaller batches of data at a time. This allowed us to reduce the computational resources needed to train the model and enabled the use of ordinary computers instead of computational stations.

The main disadvantage of the model that we found is the need for a large, labeled dataset: Supervised neural networks require a large, labeled dataset for training. However, this was time-consuming and elaborate to create, especially for segmentation tasks that require precise labeling, e.g. pumping. Although the model provides the required results, it was overfitted in some locations.

Classification success level

The digitization success rate was rated as good. However, due to the time-consuming training and running of the model, the optimal segment size was not determined iteratively, and the required accuracy value was based on a value of approx. 50 m. A lower digitization success rate can also be tolerated because it was a one-time analysis, and the digitized polygons will not change. Digitized lines were subsequently used to mark fields specified in the LPIS database, thereby significantly reducing the influence of digitization errors on the result, as missing line elements will not prevent the marking of LPIS polygons.

A comparison between the model results and the Research Institute for Soil and Water Conservation database of irrigation systems (https://meliorace.vumop.cz) shows high agreement between the datasets. However, the database of irrigation systems also stores information about newly built systems. The Water Management Maps do not contain information about systems built after 1994, or about the current working status of the systems. Further comparison with newer data is planned in consecutive papers.

CONCLUSION

The application of segmentation in a supervised neural network model for digitalizing irrigation systems depicted on Water Management Maps has demonstrated significant potential. This method allows for the precise identification of irrigation fields, source watercourses, and source types. Compared to classical methods, our approach offers a more automated and scalable solution, addressing the current lack of information on irrigation infrastructure effectively.

However, there are certain limitations to consider. The accuracy of the model heavily relies on the quality of input data and labeling. Inadequate or inaccurate data can lead to less reliable results. Additionally, while our model has shown promise in initial tests, further validation with diverse datasets is necessary to confirm its robustness and generalizability.

The potential for practical applicability is substantial, particularly within the context of the Land Parcel Identification System (LPIS). The digitalized maps of irrigation pipes and pumping stations created by our model can serve as valuable resources for agricultural planning and resource allocation. This can lead to more efficient water management, especially in regions facing water scarcity and the impacts of climate change.

In summary, our solution offers a viable and innovative approach to addressing the information gap regarding irrigation systems. With further refinement and validation, it has the potential to significantly enhance the management and optimization of irrigation resources, contributing to more sustainable agricultural practices.

ACKNOWLEDGEMENTS

We would like to thank to T.G.Masaryk Water Research Institute for access to the scans of Water Management Maps on behalf of project The potential and risks of irrigation in the Czech Republic in a changing climate (SS01020052). The research is financed by the Technology Agency of the Czech Republic (research project TH02030428) and an internal student grant of CTU (SGS20/156/OHK1/3T/11).

REFERENCES

[1] Báčová M., Krása J., 2016. Application of historical and recent aerial imagery in monitoring water erosion occurrences in Czech highlands. Soil and Water Research, vol. 11, no. 4, 267–276. doi: 10.17221/178/2015-SWR.

[2] Buchtová I., Trnka Z., 2004. Situační a výhledová zpráva Zelenina. Prague: Ministerstvo zemědělství ČR (written in Czech).

[3] SPÚ, Historická souvislost - Meliorační stavby. Státní pozemkový úřad. Available: https://www.spucr.cz/stavby-k-vodohospodarskym-melioracim-pozemku/historicka-souvislost (written in Czech).

[4] Hosnedlová P., Zemědělci stárnou. Mladší generace nemají přístup k půdě a na zemědělství pohlížejí negativně. BusinessInfo.cz. Available: https://www.businessinfo.cz/clanky/zemedelci-starnou-mladsi-generace-nemaji-pristup-k-pude-a-na-zemedelstvi-pohlizeji-negativne/ (written in Czech).

[5] Vorlíček P., Gandalovič: Chceme stimulovat generační výměnu v zemědělství. KIS Středočeského kraje. Available: https://www.kis-stredocesky.cz/2008/01/gandalovic-chceme-stimulovat-generacni-vymenu-v-zemedelstvi/ (written in Czech).

[6] HEIS, HEIS VÚV - Informační stránky a data ke stažení. VÚV. Available: https://heis.vuv.cz/data/spusteni/pgstart.asp?pg=HTML_HEIS$ZVM50LN$stazeni&pgload=1&ico=icoopenid1.png&nadpis1=Z%25E1kladn%25ED vodohospod%25E1%25F8sk%25E1 mapa %25C8R 1:50 000: mapov%25E9 listy (archiv, 1986 - 1999)&nadpis2=Informa%25E8n%25ED str%25E1nky (written in Czech).

[7] MŽP, 2015. Národní akční plán adaptace na změnu klimatu. Prague (written in Czech).

[8] Fischer E. M., Sippel S., Knutti R., 2021. Increasing probability of record-shattering climate extremes. Nat Clim Chang, vol. 11, no. 8, 689–695. doi: 10.1038/s41558-021-01092-9.

[9] Daňhelka J., 2015. Vyhodnocení sucha na území České republiky v roce 2015. Český Hydrometeorologický ústav. Prague (written in Czech).

[10] Rozkošný M., Závlahy (VÚV TGM, v.v.i.). Accessed: Mar. 09, 2023. Available: https://heis.vuv.cz/data/webmap/datovesady/projekty/zavlahy/default.asp?lang=&tab=1&wmap= (written in Czech).

[11] Eslamian S., Eslamian F., 2023. Handbook of Irrigation Hydrology and Management. Boca Raton: CRC Press. doi: 10.1201/9780429290114.

[12] Kibret E. A., Abera A., Ayele W. T., Alemie N. A., 2021. Performance Evaluation of Surface Irrigation System in the Case of Dirma Small-Scale Irrigation Scheme at Kalu Woreda, Northern Ethiopia. Water Conservation Science and Engineering, vol. 6, no. 4, 263–274. doi: 10.1007/s41101-021-00119-8.

[13] Pardo M. Á., Riquelme A. J., Jodar-Abellan A., Melgarejo J., 2020. Water and Energy Demand Management in Pressurized Irrigation Networks. Water (Basel), vol. 12, no. 7, 1878. doi: 10.3390/w12071878.

[14] Trivedi A., Nandeha N., 2018. Small Scale Irrigation Development. Irrigation & Drainage Systems Engineering, vol. 07, no. 01. doi: 10.4172/2168-9768.1000206.

[15] Conrad C., Lamers J. P. A., Ibragimov N., Löw F., Martius C., 2016. Analysing irrigated crop rotation patterns in arid Uzbekistan by the means of remote sensing: A case study on post-Soviet agricultural land use. J Arid Environ, vol. 124, 150–159. doi: 10.1016/j.jaridenv.2015.08.008.

[16] Larney F. J., 2018. Irrigated Crop Rotations. Agriculture & Agri-Food Canada, Lethbridge.

[17] Chen T., Wang Z., Li G., Lin L., 2018. Recurrent Attentional Reinforcement Learning for Multi-label Image Recognition. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 6730–6737. doi: 10.1609/aaai.v32i1.12281.

[18] Zhang W., Chen Q., Zhang W., He X., 2018. Long-range terrain perception using convolutional neural networks. Neurocomputing, vol. 275, 781–787. doi: 10.1016/j.neucom.2017.09.012.

[19] Makantasis K., Karantzalos K., Doulamis A., Doulamis N., 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 4959–4962. doi: 10.1109/IGARSS.2015.7326945.

[20] Kozlowski P., Walas K., 2018. Deep neural networks for terrain recognition task. 2018 Baltic URSI Symposium (URSI), IEEE, 283–286. doi: 10.23919/URSI.2018.8406736.

[21] Uhl J. H., Leyk S., Chiang Y. Y., Duan W., Knoblock C. A., 2020. Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. IEEE Access, vol. 8, 6978–6996. doi: 10.1109/ACCESS.2019.2963213.

[22] Chen Y., Carlinet E., Chazalon J., Mallet C., Duménieu B., Perret J., 2021. Combining Deep Learning and Mathematical Morphology for Historical Map Segmentation. Lecture Notes in Computer Science, 79–92. doi: 10.1007/978-3-030-76657-3_5.

[23] Garcia-Molsosa A., Orengo H. A., Lawrence D., Philip G., Hopper K., Petrie C. A., 2021. Potential of deep learning segmentation for the extraction of archaeological features from historical map series. Archaeol Prospect, vol. 28, no. 2, 187–199. doi: 10.1002/ARP.1807.

[24] Kulhavý Z., Kulhavý F., 2008. Navrhování hydromelioračních staveb. Praha: Informační centrum ČKAIT.

[25] Jajodia T., Garg P., 2019. Image Classification-Cat and Dog Images. International Research Journal of Engineering and Technology. [Online]. Available: www.irjet.net

[26] Cukierski W., Dogs vs. Cats. Kaggle. [Online]. Available: https://www.kaggle.com/c/dogs-vs-cats

[27] Elson J., Douceur J. R., Howell J., Saul J., 2007. Asirra: A CAPTCHA that exploits interest-aligned manual image categorization. Proceedings of the ACM Conference on Computer and Communications Security, 366–374. doi: 10.1145/1315245.1315291.

[28] Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y., 2013. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings. doi: https://doi.org/10.48550/arXiv.1312.6229.