INFORMATION TECHNOLOGY
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

UDC 519.6

Original Empirical Research

https://doi.org/10.23947/2587-8999-2025-9-1-52-60

Automatic Depth Value Recognition on Pilot Charts Using Deep Learning Methods

Elena O. Rakhimbaeva, Tajaddin A. Alyshov, Yulia V. Belova
Don State Technical University, Rostov-on-Don, Russian Federation
✉ lena_rahimbaeva@mail.ru

Abstract

Introduction. This study addresses the problem of automatic text recognition in images, specifically the extraction of depth information from pilot charts. The relevance of this task is driven by the need to automate the processing of large volumes of cartographic data to create depth maps suitable for mathematical modelling of hydrodynamic and hydrobiological processes. The objective of this work is to develop the software tool LocMap, designed for the automatic detection and recognition of depth values represented as numbers on pilot chart images.

Materials and Methods. The study employs deep learning methods, including convolutional neural networks (ResNet) for feature extraction, the Differentiable Binarization (DB) algorithm for text detection, and the Scene Text Recognition with a Single Visual Model (SVTR) architecture for text recognition.

Results. The developed software allows users to upload pilot chart images, perform preprocessing, detect and recognize depth values, highlight them in the image, and save the results in a text file. Testing results demonstrated that the system ensures high accuracy in recognizing depth values on pilot charts.

Discussion and Conclusion. The obtained results highlight the practical significance of the developed solution for automating the processing of pilot charts.

Keywords: text recognition, pilot charts, depth, deep learning, convolutional neural networks, differentiable binarization algorithm, Single Visual Model

Funding. This research was supported by the Russian Science Foundation, grant No. 22–71–10102, https://rscf.ru/project/22-71-10102/

For Citation. Rakhimbaeva E.O., Alyshov T.A., Belova Yu.V. Automatic Depth Value Recognition on Pilot Charts Using Deep Learning Methods. *Computational Mathematics and Information Technologies*. 2025;9(1):52–60. https://doi.org/10.23947/2587-8999-2025-9-1-52-60

Оригинальное эмпирическое исследование

Автоматическое распознавание значений глубины на лоцманских картах с использованием методов глубокого обучения

Е.О. Рахимбаева ✉, Т.А. Алышов, Ю.В. Белова
Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация
✉ lena_rahimbaeva@mail.ru

Аннотация

Введение. Рассматривается проблема автоматического распознавания текста на изображениях, в частности задача извлечения информации о глубинах с лоцманских карт. Актуальность данной задачи обусловлена необходимостью автоматизации обработки больших объемов картографических данных для построения карты глубин, пригодной для математического моделирования гидродинамических и гидробиологических процессов.

Целью работы является разработка программного средства (ПС) LocMap, предназначенного для автоматического обнаружения и распознавания значений глубин, представленных в виде чисел на изображениях лоцманских карт.
Материалы и методы. В работе использованы методы глубокого обучения, а именно сверточные нейронные сети ResNet для извлечения признаков, алгоритм дифференцируемой бинаризации DB для обнаружения текста и архитектура Scene Text Recognition with a Single Visual Model (SVTR) для распознавания текста.
Результаты исследования. Разработанное ПС позволяет загружать изображения лоцманских карт, выполнять предобработку, обнаруживать и распознавать значения глубин, выделять их на изображении и сохранять результаты в текстовый файл. Результаты тестирования показали, что разработанная система обеспечивает высокую точность распознавания значений глубин на лоцманских картах.
Обсуждение и заключение. Полученные результаты демонстрируют практическую значимость разработанного решения для автоматизации обработки лоцманских карт.

Ключевые слова: распознавание текста, лоцманские карты, глубина, глубокое обучение, сверточные нейронные сети, алгоритм дифференцируемой бинаризации, Single Visual Model

Финансирование. Исследование выполнено за счет гранта Российского научного фонда № 22–71–10102, https://rscf.ru/project/22-71-10102/

Для цитирования. Рахимбаева Е.О., Алышов Т.А., Белова Ю.В. Автоматическое распознавание значений глубины на лоцманских картах с использованием методов глубокого обучения. *Computational Mathematics and Information Technologies*. 2025;9(1):52–60. https://doi.org/10.23947/2587-8999-2025-9-1-52-60

Introduction. In today›s world, there is a rapid increase in the volume of information presented in the form of images. This drives the need for the development of efficient methods for automated data extraction and analysis from images. One of the key challenges in this field is Optical Character Recognition (OCR), which has broad applications in various areas, including document digitization, automatic license plate recognition, and cartographic data analysis.

Extracting data from image processing, including satellite imagery, is becoming increasingly significant for modelling processes in complex natural systems. A pressing issue is obtaining initial information for mathematical models of hydrodynamics and hydrobiology [1] and refining the parameters of these models [2]. The development of satellite image processing methods enables the acquisition of input data for predictive modelling of processes occurring in water bodies, particularly in the Azov and Black Seas [3].

Pilot charts are a special type of map containing detailed information about water basins, designed to ensure safe navigation for vessels. One of the key tasks when working with pilot charts is determining depth values, which are typically represented as standard numbers and subscripted numbers on the maps. Traditional methods of processing pilot charts, based on manual analysis, are extremely labor-intensive and prone to errors. Therefore, the development of automated methods for recognizing depth values on pilot charts is an important and relevant task.

The aim of this study is to develop a software tool for the automatic detection and recognition of depth values on pilot charts using deep learning methods. To achieve this goal, the following tasks were set:

- analyze existing text recognition methods for images;
- collect and prepare a training dataset of pilot chart images;
- develop a data augmentation algorithm to enhance model robustness against various distortions;
- develop and train a model for detecting and recognizing depth values;
- create a software tool with a user-friendly interface;
- conduct testing and evaluate the performance of the developed software.

Materials and Methods

Dataset Description. In this study, a dataset consisting of 1,590 images of pilot charts of the Azov and Black Seas was used. The images were obtained from open sources on the Internet. The images have a resolution of 400×300 pixels and depict sections of the seas with depth markings, fairways, coastlines, and other navigational objects. Figure 1 shows an example of a pilot chart image from the dataset.

For training the model, the following elements were selected for recognition: numerical depth values represented by Arabic numerals, numerical values with a subscript indicating tenths of a meter. Elements that do not represent depth values, such as coastline markings, object names, and kilometer markers, were not subject to recognition.

Data Annotation. The data annotation was performed manually using the PPOCRLabel software. The elements to be recognized were identified and assigned corresponding labels in the form of numbers, such as “10” or “12.4”. A total of 1,590 images were annotated, containing approximately 12,500 depth values. Figure 2 shows an example of annotated pilot chart data.

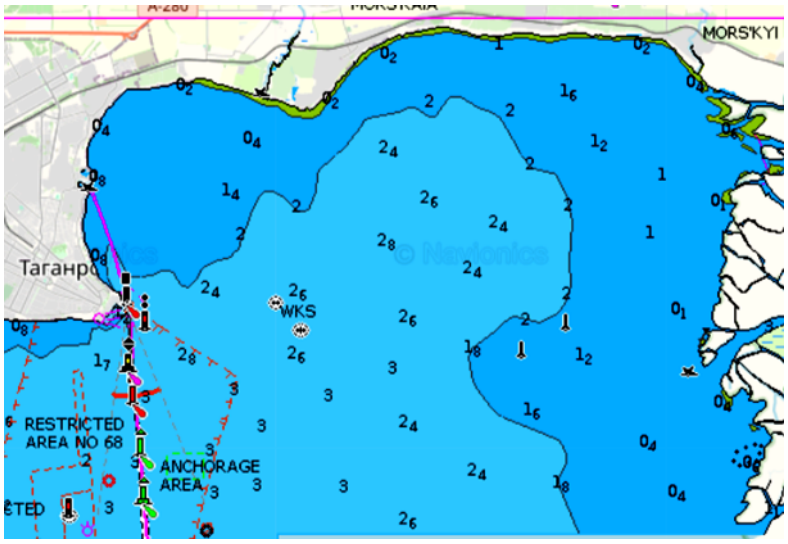

Fig. 1. Pilot chart (depth map)

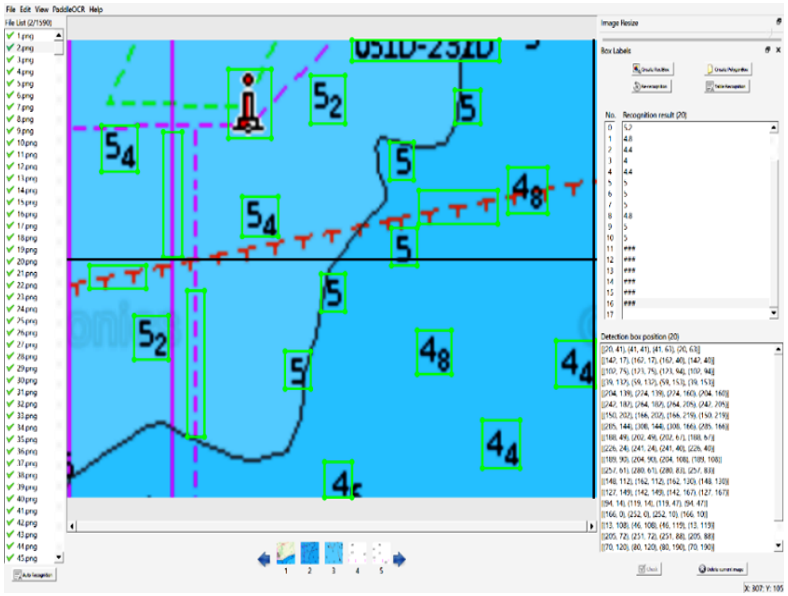

Figure 2. Data annotation on an image

During the annotation process, the following challenges were identified: low quality of some images, dense arrangement of objects on the map, truncated depth values at the edges of images.

Data Augmentation. To enhance the model's robustness against various distortions and to increase the size of the training dataset, a data augmentation algorithm was applied. The augmentation included the following methods [4]:

• scaling (image sizes were adjusted by a factor of 0.8–1.2 while maintaining proportions),

• shifting (images were shifted horizontally and vertically by a random number of pixels within the range of –50 to +50 pixels),

• applying filters (Gaussian blur and sharpening filters were used).

Figure 3 presents examples of augmented images.

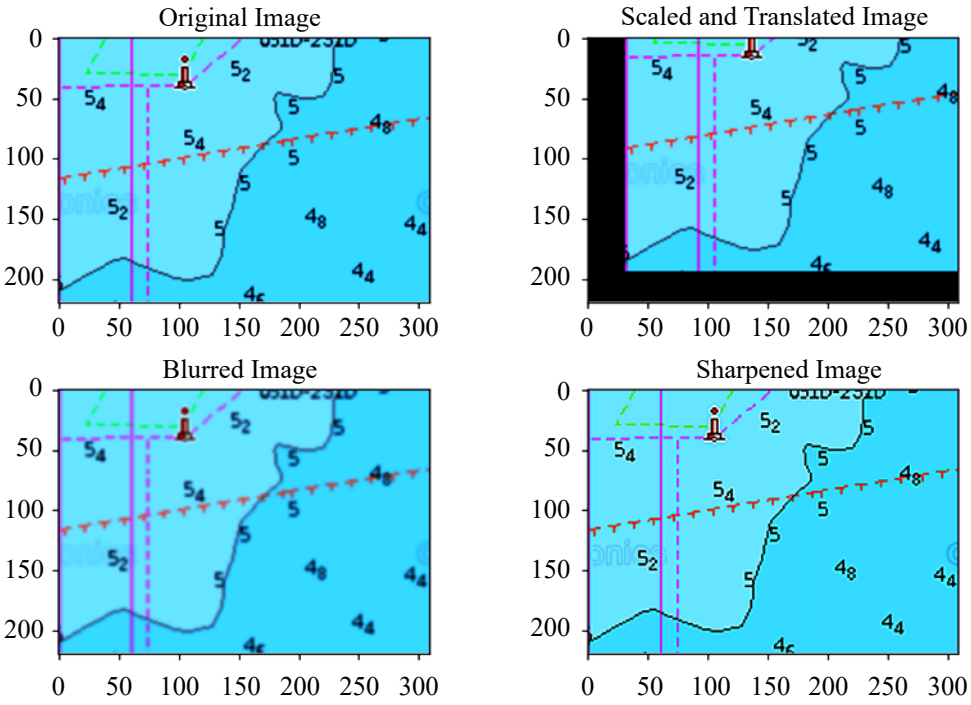

Fig. 3. Examples of image augmentation

Detection Model Architecture. For text detection in pilot charts, the Differentiable Binarization (DB) algorithm was used, as illustrated in Figure 4. DB is a state-of-the-art text detection method based on segmentation, enabling efficient text region extraction with a dynamic threshold.

The key advantage of DB lies in its differentiable binarization function, allowing the network to be trained end-to-end, yielding more accurate results compared to traditional fixed-threshold binarization methods. The fundamental difference from other approaches is that DB includes a threshold map, predicting the threshold for each pixel point in the image using a neural network rather than assigning a fixed value. This enables better differentiation between text foreground and background.

The DB algorithm applies differentiable binarization, which approximates the step function of conventional binarization. The following formula is used:

$$\hat{B}_{i,j} = \frac{1}{1+e^{-k(P_{i,j}-T_{i,j})}}, \quad (1)$$

where $\hat{B}$ is the approximate binary map; k is the enhancement factor, equal to 50; P is the probability map; T is the threshold map obtained from the network.

This approximate binarization function is differentiable, allowing it to be optimized along with the segmentation network during the training process. Differentiable binarization with adaptive threshold values can not only help distinguish text regions from the background but also separate tightly connected text instances [5, 6].

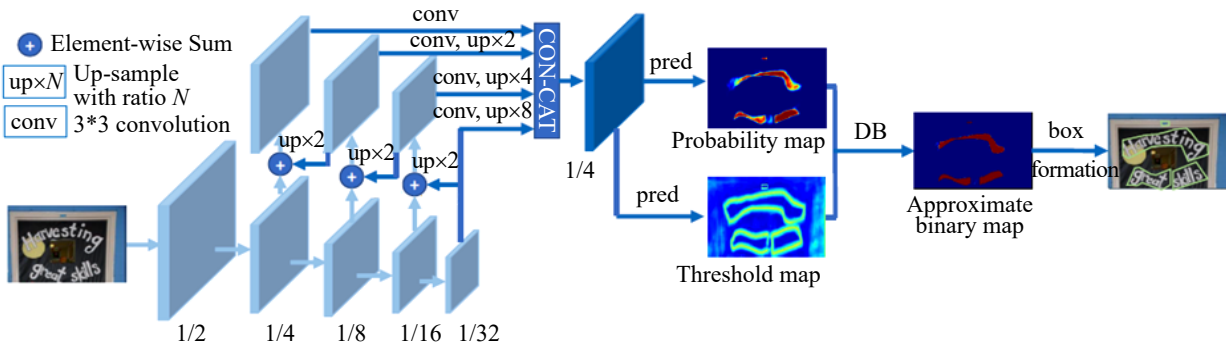

Fig. 4. Architecture of Differentiable Binarization

The ResNet networks and Differentiable Binarization Feature Pyramid Network (DBFPN) extract features from the input image, which are then combined to form a feature map with a quarter of the original image's size. A convolutional

layer is applied to generate the probability map and threshold map. Subsequently, based on formula (1), the binary map is created, and then, using DB post-processing, the contour is extracted.

The text detection algorithm using differentiable binarization can be described as follows:

Step 1. Feature extraction. The input image is fed into a network, such as ResNet, which extracts features at different levels of the pyramid (1/2, 1/4, 1/8, 1/16, 1/32) relative to the input image scale;

Step 2. Feature fusion. The extracted features are successively upscaled to a common scale and merged. After merging, they pass through 3×3 convolutional layers and additional upscaling operations to create a unified feature (F);

Step 3. Map prediction. The feature F is used to predict the probability map (P) and the threshold map (T);

Step 4. Differentiable binarization (DB). The probability (P) and threshold (T) maps are used to compute the approximate binary map ($\hat{B}$) using the differentiable binarization function. This allows the binarization process to be optimized along with the training of the segmentation network;

Step 5. Bounding box generation. During inference, text bounding boxes can be easily obtained from the approximate binary map ($\hat{B}$) or the probability map (P) using the bounding box generation module.

The use of differentiable binarization for text detection in cartographic images enables the creation of an efficient and accurate system capable of working in real-world conditions with diverse and distorted data. This approach ensures high flexibility and adaptability of the model, which is a key factor for successful recognition of text elements in cartographic images [7].

The overall dataset of 1590 images was divided into a training set (1272 images) and a validation set (318 images).

For the DB architecture (based on ResNet-34), the DB++ model with the DBFPN module for feature extraction and the DBHead module were used for text detection. The loss was calculated using the combined DBLoss function, which includes DiceLoss with weights α=5 and β=10. An online hard example mining (OHEM) mechanism with a coefficient of 3 was also applied. It selects only hard examples from the mini-batch for gradient calculation, skipping easy ones so that the model focuses on more challenging cases.

The training parameters included:

- Adam optimizer with β1=0.9 and β2=0.999;
- Cosine learning rate decay (initial value 0.0005) with two epochs for warm-up;
- L_2 regularization with a coefficient of 0;
- Batch size — 8;
- Total number of epochs — 21.

During training, evaluation metrics such as Hmean were used, which were calculated every 7 epochs. The images were resized to 960×960 pixels, according to the DB architecture property (image resolutions must be divisible by 32) [8].

Recognition Model Architecture. For recognizing the detected depth values, the SVTR (Single Visual Model for Scene Text Recognition) architecture was chosen, as shown in Figure 5. SVTR represents an innovative approach to text recognition, in which the traditional sequential model is replaced by a unified visual model, improving efficiency and processing speed [9].

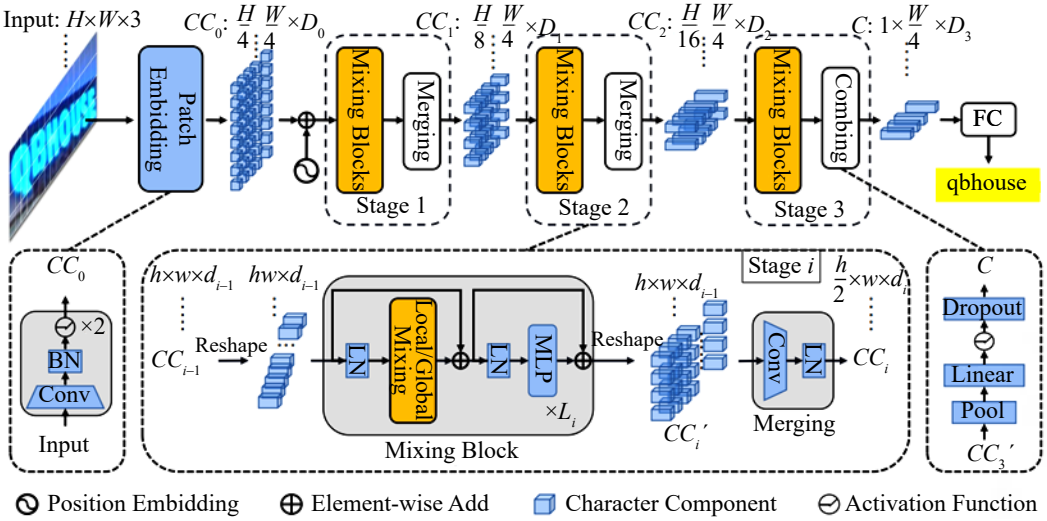

Fig. 5. SVTR Architecture

Main components and stages:

- Input (the input image with dimensions H×W×3);
- Patch Embedding (divides the input image into small patches and converts them into vector representations). A Position Embedding is added to the output of Patch Embedding to encode the position information of each patch on the image;

• Stage 1, Stage 2, Stage 3 (three processing stages, each including Mixing Blocks and Merging). In each Mixing Block, local and global information are combined. This allows the model to account for fine details (e. g., textures, edges) as well as the overall structure or context (global features) of the image. For example, to understand what is depicted in a photograph, it is important to consider the spatial arrangement of objects relative to each other;

• Fully Connected (the final fully connected layer that produces the prediction) [10].

The SVTR architecture is based on the principle of tokenizing images into parts. The depth value image is split into small 2D patches, called "character components". Hierarchical cascades recursively apply mixing, merging, and combining operations at the level of these components [11]. The architecture uses global and local mixing blocks to perceive both inter-character and intra-character patterns, enabling multi-level perception of character components. Character recognition is performed through a simple linear prediction at the end of the network. SVTR consists of a three-cascade network, with progressively reduced height, which facilitates efficient feature extraction [12].

For training the SVTR model, the authors used a dataset of 12,495 depth value images. The dataset was split into training (8,747 images) and validation (3,748 images) sets. The model was trained using the AdamW optimizer with a decay weight of 0.05. The initial learning rate was set at 0.00005 using a Cosine Learning Rate Scheduler and a linear warm-up phase for 2 epochs. The batch size was 256 images. The total number of training epochs was set to 50. The input image size for the SVTR network was 48×36 pixels [13].

The loss function used was CPPDLoss (Character Position and Pixel Distance Loss). The CPPDLoss function is specifically designed for recognition tasks and takes into account both the accuracy of character recognition and their positional alignment.

The developed software tool, LocMap, performs the following functions:

• image upload (the user uploads a pilot chart image in png, jpeg, or bmp formats);

• image preprocessing (converting to grayscale and binarizing the images using thresholding);

• text detection (text areas in the image are detected using the DB algorithm);

• text recognition (the detected text areas are passed to the SVTR model for recognizing depth values);

• results output (the recognized depth values are highlighted on the original image and displayed in a separate window);

• results saving (the user can save the image with highlighted depth values and a text file with the recognized values and their coordinates on the image).

Results

Evaluation of Detection and Recognition Quality. The following metrics were used to evaluate the quality of detection:

• precision — the proportion of correctly recognized depth values among all detected values;

• recall — the proportion of correctly recognized depth values among all depth values present in the image;

• harmonic mean (hmean) — the harmonic mean of precision and recall, a balanced metric that takes both characteristics into account.

As a result of training the depth detection model, the best metric values were achieved at the 18th epoch, which are presented in Table 1.

Table 1

Best Metric Values for the Detection Model

Metric	Value
precision	90.89%
recall	82.66%
hmean	86.58%

To evaluate the recognition quality, the RecMetric metric was used, with the primary indicator being accuracy. Additionally, the Norm Edit Distance (norm_edit_dis) metric was used, which measures the degree of similarity between the predicted text and the reference (labeled) text. During the training process, the model that showed the best accuracy on the validation dataset was saved for further use in inference tasks.

The recognition model achieved the best results at the 39th epoch, which are presented in Table 2.

Table 2

Best Metric Values for the Recognition Model

Metric	Value
accuracy	95.03%
norm_edit_dis	97.60%

Examples of Software Operation. For user interaction with the LocMap software, four buttons were implemented, displayed at the bottom of the window:

- "Open";
- "Save Images";
- "Save Values";
- "Re-recognize".

After opening a file with an image and performing recognition, the result of the software operation is displayed on the screen in the corresponding areas, as shown in Figure 6. This includes a list of recognized depth values and the coordinates of the points where these values were determined. The obtained values and their coordinates are saved in a file with the .txt extension.

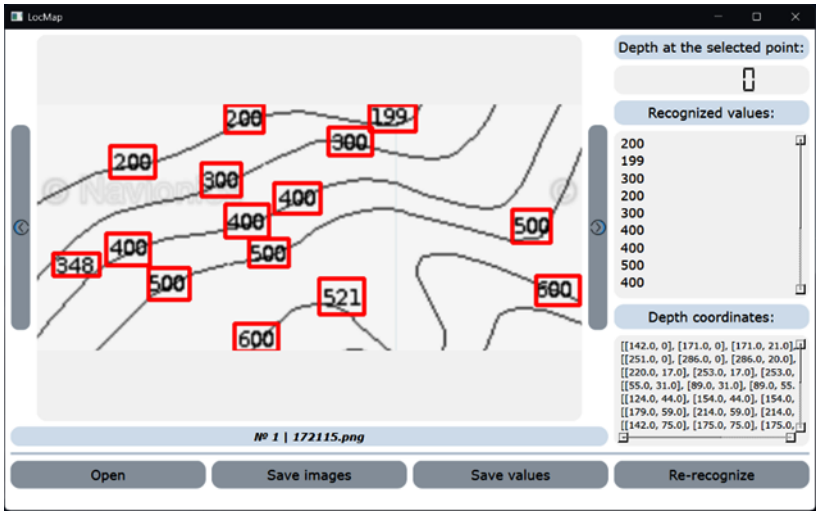

Fig. 6. Result of Software Operation

The LocMap software module allows obtaining the depth value at a selected point on the image, as demonstrated in Fig. 7.

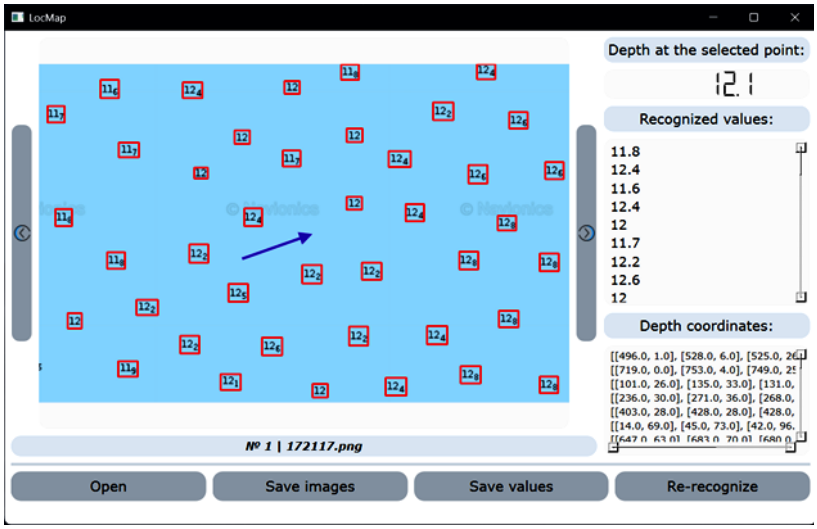

Fig. 7. Depth at the Selected Point

Discussion and Conclusion. The results obtained demonstrate that the developed software LocMap provides high accuracy in recognizing depth values on navigation charts. The best results are achieved when recognizing values located in open areas of the map with good contrast and clear typography. Difficulties arise when recognizing values placed near complex graphic elements, such as contour lines, markers, or text annotations.

The advantage of the developed method is the use of modern deep learning algorithms, such as DB, ResNet, and SVTR, which allow effective detection and recognition of text on images with various distortions. The use of data augmentation has improved the model's robustness to various numeral writing styles, changes in scale and orientation, and noise on the image.

Despite the high recognition accuracy, the developed software LocMap has several limitations. One of the main factors is the dependence on the quality of data labeling. Errors or inaccuracies in labeling can lead to incorrect model training, which is especially critical for complex text regions on navigation charts. Another limitation is the computational

complexity of the method, associated with the use of deep neural networks. Specifically, resource-intensive stages of data processing and computation hinder the application of the method in real-time on devices with limited computational power.

The next step in the research will be the construction of a depth map for the Azov and Black Seas using the algorithm proposed in [14]. This algorithm uses a solution to the equation employed to obtain high-order accuracy schemes for the Laplace equation. The use of this algorithm will allow for the interpolation of the seabed surface with sufficiently smooth functions. This will improve the accuracy of modelling hydrodynamic and hydrobiological processes by constructing a computational grid that matches current cartographic data [15].

The conducted experiments showed that the developed system ensures high recognition accuracy. The obtained results demonstrate the practical significance of the developed solution for automating the processing of navigation charts.

Potential directions for future research include expanding the dataset, improving text detection and recognition algorithms, integrating with GIS systems, recognizing other elements of navigation charts, and constructing seabed relief based on the obtained depths and their coordinates. The application area of the developed software is mathematical modelling of hydrodynamic and hydrobiological processes of water bodies. The application of the developed recognition methods will help build computational grids based on up-to-date cartographic information.

References

1. Lyashenko T.V., Chistyakov A.E., Nikitina A.V. Modelling the process of phosphate pollution spread in the aquatic ecosystem. *Computational Mathematics and Information Technologies*. 2023;6(4):47–53. (In Russ.) https://doi.org/10.23947/2587-8999-2023-7-4-47-53

2. Belova Yu.V., Filina A.A., Chistyakov A.E. Forecasting the dynamics of summer phytoplankton species based on satellite data assimilation methods. *Computational Mathematics and Information Technologies*. 2024;8(4):27–34. (In Russ.) https://doi.org/10.23947/2587-8999-2024-8-4-27-34

3. Belova Yu.V., Razveeva I.F., Rakhimbaeva E.O. Development of an Algorithm for Semantic Segmentation of Earth Remote Sensing Data to Determine Phytoplankton Populations. *Advanced Engineering Research (Rostov-on-Don)*. 2024;24(3):283–292. https://doi.org/10.23947/2687-1653-2024-24-3-283-292

4. Cubuk E. D., Zoph B., Mané D., Vasudevan V. and Le Q. V., Autoaugment: Learning augmentation strategies from data[C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA, 2019. P. 113–123. https://doi.org/10.1109/CVPR.2019.00020

5. Minghui Liao, Zhaoyi Wan, Cong Yao, Chen Kai, Xiang Bai. Real-Time Scene Text Detection with Differentiable Binarization. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(07):11474–11481. https://doi.org/10.1609/aaai.v34i07.6812

6. Sheeba M.C., Christopher Ch. S. Adaptive deep residual network for image denoising across multiple noise levels in medical, nature, and satellite images. *Ain Shams Engineering Journal*. 2025;16(1):103188. https://doi.org/10.1016/j.asej.2024.103188

7. Bradley D., Roth G. Adaptive thresholding using the integral image. *Journal of graphics tools*. 2007;12(2):13–21. https://doi.org/10.1080/2151237X.2007.10129236

8. Tensmeyer C., Martinez T. Document image binarization with fully convolutional neural networks. In: *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 2017;1:99–104. https://doi.org/10.1109/ICDAR.2017.25

9. Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, Wenyu Liu. Textboxes: A fast text detector with a single deep neural network. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017;31(1). https://doi.org/10.1609/aaai.v31i1.11196

10. Yongkun Du, Zhineng Chen, Caiyan Jia, et al. SVTR: Scene Text Recognition with a Single Visual Model. *International Joint Conference on Artificial Intelligence (2022)*. https://doi.org/10.48550/arXiv.2205.00159

11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 770–778. https://doi.org/10.1109/CVPR.2016.90

12. Paschalis Tsirtsakis, Georgios Zacharis, George S. Maraslidis, George F. Fragulis. Deep learning for object recognition: A comprehensive review of models and algorithms. *International Journal of Cognitive Computing in Engineering*. 2025;6:298–312. https://doi.org/10.1016/j.ijcce.2025.01.004

13. Miao Tian, Kai Ma, Zhihao Liu, Qinjun Qiu, Yongjian Tan, Zhong Xie. Recognition of geological legends on a geological profile via an improved deep learning method with augmented data using transfer learning strategies. *Ore Geology Reviews*. 2023;153:105270. https://doi.org/10.1016/j.oregeorev.2022.105270

14. Sukhinov A.I., Belova Y.V., Filina A.A. Parallel implementation of substance transport problems for restoration the salinity field based on schemes of high order of accuracy. *CEUR Workshop Proceedings (2019)*, 2500.

15. Nikitina A., Belova Y., Atayan A. Mathematical modeling of the distribution of nutrients and the dynamics of phytoplankton populations in the Azov Sea, taking into account the influence of salinity and temperature. *AIP Conference Proceedings*. 2019; 2188(1): 050027. https://doi.org/10.1063/1.5138454

About the Authors:

Elena O. Rakhimbaeva, Postgraduate student, Assistant lecturer of the Department of "Computer Engineering and Automated Systems Software", Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), ORCID, SPIN-code, lena_rahimbaeva@mail.ru

Tadjaddin A. Alyshov, Master's Degree student of the Department of "Mathematics and Computer Science", Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), ORCID, taci2002@mail.ru

Yulia V. Belova, Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of "Mathematics and Computer Science", Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), ORCID, SPIN-code, yvbelova@yandex.ru

Contributions of the authors:

E.O. Rakhimbayeva: concept development; data curation; formal analysis; conducting research; visualization; writing a draft manuscript.

T.A. Alyshov: conducting research; software development; validation of results; writing a draft manuscript.

Yu.V. Belova: scientific guidance and writing the manuscript — reviewing and editing.

Conflict of Interest Statement: **the authors declare no conflict of interest.**

All authors have read and approved the final manuscript.

Об авторах:

Елена Олеговна Рахимбаева, аспирант, ассистент кафедры программного обеспечения вычислительной техники и автоматизированных систем Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), ORCID, SPIN-код, lena_rahimbaeva@mail.ru

Таджаддин Аледдин оглы Алышов, магистрант кафедры математики и информатики Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), ORCID, taci2002@mail.ru

Юлия Валериевна Белова, кандидат физико-математических наук, доцент кафедры математики и информатики Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), ORCID, SPIN-код, yvbelova@yandex.ru

Заявленный вклад авторов:

Е.О. Рахимбаева: разработка концепции; курирование данных; формальный анализ; проведение исследования; визуализация; написание черновика рукописи.

Т.А. Алышов: проведение исследования; разработка программного обеспечения; валидация результатов; написание черновика рукописи.

Ю.В. Белова: научное руководство и написание рукописи — рецензирование и редактирование.

Конфликт интересов: ***авторы заявляют об отсутствии конфликта интересов.***

Все авторы прочитали и одобрили окончательный вариант рукописи.

Received / Поступила в редакцию 03.02.2025
Revised / Поступила после рецензирования 24.02.2025
Accepted / Принята к публикации 17.03.2025