Semantic Classification and Segmentation of Archaeological Sites Based on a Fusion of Object Detector and 3DEF

A. V. Vokhmintsev[1,2]
[1]*Institute of Information Technology*
Chelyabinsk State University
Chelyabinsk, Russia
[2]*Institute of Digital Economy*
Ugra State University
Khanty-Mansiysk, Russia
vav@csu.ru

O. I. Khristodulo
Geoinformation System Department
Ufa University of Science and Technology
Ufa, Russia
o-hristodulo@mail.ru

M. A. Romanov
South Ural State University
Chelyabinsk, Russia
std.romanov.ma@gmail.com

***Abstract*—In this paper, an original method of semantic objects classification and segmentation data was proposed, which uses 2d and 3d data models. This method has been tested for the task of classifying archaeological sites. The information about the semantic characteristics of archaeological objects was obtained using state-of-art methods for semantic segmentation of the environment, such as YOLO and 3DEF. These methods have been improved by observing an archaeological site from several viewing points. In order to improve the accuracy and performance of the procedure for detecting archaeological sites, the Bayes formula which allow to use collaboratively detector YOLO and 3DEF semantic segmentation method was used. A comparative analysis with various combinations of 2D and 3D detection methods was carried out, the results of computer simulation were presented and discussed.**

Keywords—3D segmentation method, semantic classification, iterative closest point, object detector

I. INTRODUCTION

Modern methods of decoding archaeological sites are based on the use of semantic segmentation and mapping, classification and recognition of 3D objects on the scene [1]. State of art methods of semantic segmentation and classification can be improved by observing the environment from several angles. The semantic characteristic of an archaeological object in the work is understood as the class name to which this object is assigned as a result of the classification of archaeological objects using neural networks. In this work, only large objects inside archaeological sites are considered, such as contours and details of external and internal architectural structures of ancient settlements, religious monuments, sites, burial mounds and ground burial grounds, sanctuaries, megaliths, dolmens and ancient mines.

Modern methods use a pipeline of methods for semantic processing of archaeological sites that solve the following tasks:

- object detection: identificate all pixels of objects of one or more classes on the scene and marking the scene using a frame that defines the boundaries for each object and contains a label with the probability of assigning the object to a certain class, for example, the YOLO detector [2-4].
- semantic segmentation of three-dimensional models (determination of all points of objects of a certain class in point cloud), for example, the 3DEF method [5].
- class segmentation: determination of all pixels of an object of a certain class separately for each object of the class on the scene using a mask, for example, the Mask R-CNN detector [6].
- image segmentation based on graph sections, for example, the GrabCut method [7].

The convolutional neural network Mask R-CNN develops the architecture of the Faster R-CNN [8] by adding another branch of analysis that predicts the position of the mask covering the found archaeological object. To improve the accuracy and performance of the procedure for detecting archaeological objects, the Bayes merger of detectors YOLO (Mask R-CNN) и 3D segmentation often was used.

When semantic marking of the space, archaeological objects of different classes on the analyzed image (point cloud) are highlighted in different colors and accompanied by an attribute – the probability of attributing the object to the corresponding class of the subject area. Each archaeological object that has been identified on a digital terrain model (DTM) is assigned a weight value. The assignment of weights is carried out on the basis of tabular data compiled by experts. Note that received semantic information on the classification of archaeological objects can be used to improve the convergence of the registration method: data on semantic characteristics are associated with the values of weights in the functional of the ICP variational problem [9, 10]. This approach makes it possible to establish high significance for some archaeological objects, and low significance for others. At the same time, it is important that data filtering on digital models of archaeological objects (the 1st stage of the ICP algorithm) is carried out on the basis of information about the semantics of objects in an archaeological monument, and not only on the basis of some metric.

The work was supported by the Russian Science Foundation, project No. 23-11-20007 and the Chelyabinsk Region, RSF-RK contract No. 584.

979-8-3503-4555-1/23/$31.00 ©2023 IEEE

In digital information processing to obtain data about the semantic characteristics of an object, the state of art methods of semantic processing of the surrounding space YOLO and 3DEF are often used, which allow solving the well-known problem of symbol binding. In applications for remote study of 3D models, including archaeological sites, semantic marking methods are often improved by observing the surrounding space from several viewing points (Multi-viewed 3D Entangled Forest) [11]. In this paper, a new accurated multi-view algorithm to register 3D models for the class of affine transformations based on closed form solution of variational problem ICP was proposed. In this paper, for semantic processing of archaeological sites, it is proposed to use a pipeline of methods that solve the following tasks:

- 3D semantic segmentation: the 3DEF 4.1 method is used.
- object detection: the YOLOv8 [4] detector is used.

The YOLO, Mask-R-CNN and 3DEF methods give different accuracy estimates for different classes, so using a combination of these methods will improve the overall accuracy of the classification of archaeological objects. It is expected that the suggested combination of various semantic processing methods of RGB-D data will allow for accurate classification of archaeological objects of different classes.

It is known that for machine learning methods to work effectively, a large amount of marked-up data is needed. Earlier it was noted that for various reasons, the size of open collections of images on archaeological sites is small. Therefore, for the training of detectors, an annotated collection of digital data on archaeological sites of the Bronze Age in the Chelyabinsk region was created on the basis of:

- materials of aerial photography conducted from 1954 to 1957 by order of the branches of forestry and agriculture and the 60s-80s for the purposes of geology and geodesy. Aerial photographs were taken at a scale of 1:14,000 with high resolution and for the entire territory of the Kizilsky district of the Chelyabinsk region, a total of 246 images were selected.
- the results of remote sensing of the Earth from Sentinel-2 satellites since 2015, Landsat 4-9 since 1985, Resurs-P from 2013 to 2021, Canopus-From 2013 to 2023. A total of 689 images were selected.

In the future, it is planned to expand the training sample of archaeological objects by about 5-6 times using geometric transformations and methods of pre-training convolutional neural networks base on Mask R-CNN. This article is organized as follows: in the first section, the registration task based on ICP for 3D molels is given, in the second section, a fusion semantic segmentation algorithm, using YOLO and 3DEF, is presented, in the third section, the obtained results of computer modeling were presented and discussed.

II. An Accurated Fusion Algorithm for Registration Point Clouds

Consider the variational problem of the ICP [12] registration method for affine transformations with a point-to-point metric. Let's denote by $X=\{x_1,\ldots,x_n\}$ the 3D data about the source RGB-D frame and by $Y=\{y_1,\ldots,y_m\}$ the data about the target RGB-D frame in $\mathbb{R}^3$. Let the relationship between points in frames X and Y be such that for each point in x_i, the corresponding point in y_i can be calculated. Then the iterative nearest point algorithm (ICP) can be considered as a geometric transformation of rigid objects from X to Y of the following form:

$$Rx_i + T\,, \tag{1}$$

where R is the rotation matrix, T is the vector of parallel transfer, w_i – is the weighting factor associated with the semantic characteristic of the archaeological object, i=1,…,n. Let's imagine the functional J(R,T) as a function [13]:

$$J(R,T) = \sum_{i=1}^{n} w_i \|Rx_i + T - y_i\|^2 \tag{2}$$

We define the variational ICP problem as:

$$\underset{RT}{\operatorname{argmin}}\, J(R,T), \tag{3}$$

where

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, T = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}, x = \begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix}, y = \begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{pmatrix} \tag{4}$$

Suppose that $X=\{x_1,\ldots,x_n\}$ lies in the -Oxz plane or in the –Oxy plane. Also suppose $\sum_{i=1}^{n} x_{1i}^2 \neq 0$, $\sum_{i=1}^{n} x_{2i}^2 \neq 0$, $\sum_{i=1}^{n} x_{3i}^2 \neq 0$, then we solve the variational problem with respect to the matrix R, $k = 1,2,3, m, n \neq k$:

$$\frac{\partial J(R)}{\partial r_{1k}} = \sum_{i=1}^{n} 2\, w_i(\, r_{11}x_{1i} + r_{12}x_{2i} + r_{13}x_{3i} - y_{1i})x_{ki} = 0 \tag{5}$$

Then

$$\sum_{i=1}^{n} w_i(\, r_{1m}x_{mi} + r_{1n}x_{ni} - y_{1i})x_{ki} + r_{1k}\sum_{i=1}^{n} w_i x_{ki}^2 = 0 \tag{6}$$

From the expression (6) we obtain the value of the parameters of the rotation matrix:

$$r_{1k} = -\frac{\sum_{i=1}^{n} w_i(\, r_{1m}x_{mi} + r_{1n}x_{ni} - y_{1i})x_{ki}}{\sum_{i=1}^{n} w_i x_{ki}^2} \tag{7}$$

Given the expression (7), we will make a number of transformations and rewrite the solution of the variational problem in the following form:

$$J(R) = \sum_{i=1}^{n} w_i \left(\begin{array}{c} r_{1m}\left(x_{mi} - x_{ki}\frac{\sum_{j=1}^{n} x_{mj}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2}\right) \\ +r_{1n}\left(x_{ni} - x_{ki}\frac{\sum_{j=1}^{n} x_{nj}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2}\right) - \left(y_{1i} - x_{ki}\frac{\sum_{j=1}^{n} y_{1j}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2}\right) \end{array} \right)^2 + w_i(\, r_{21}x_{1i} + r_{22}x_{2i} + r_{23}x_{3i} - y_{2i})^2 + + w_i(\, r_{31}x_{1i} + r_{32}x_{2i} + r_{33}x_{3i} - y_{3i})^2. \tag{8}$$

We introduce the following definitions:

$$G_{mi} = x_{mi} - x_{ki}\frac{\sum_{j=1}^{n} x_{mi}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2},\ G_{pi} = x_{ni} - x_{ki}\frac{\sum_{j=1}^{n} x_{ni}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2},$$

$$G_{ki} = y_{1i} - x_{ki}\frac{\sum_{j=1}^{n} y_{mi}x_{kj}}{\sum_{j=1}^{n} x_{kj}^2}. \quad (9)$$

Using the definitions introduced in formulas (9), we present the functional J(R) in formula (8) in the following form

$$J(R) = \sum_{i=1}^{n} w_i\left(r_{1m}G_{mi} + r_{1n}G_{ni} - G_{ki}\right)^2 + w_i\,(\,r_{21}x_{1i} + r_{22}x_{2i} + r_{23}x_{3i} - y_{2i})^2 + w_i(\,r_{31}x_{1i} + r_{32}x_{2i} + r_{33}x_{3i} - y_{3i})^2 \quad (10)$$

Let's define the partial derivative of J(R):

$$\frac{\partial J(R)}{\partial r_{1m}} = 2\sum_{i=1}^{n} w_i\left(r_{1m}G_{mi} + r_{1n}G_{pi} - G_{ki}\right)G_{mi} = 0 \quad (11)$$

Next , we get the formula for calculating r_{1m}:

$$r_{1m} = -\frac{r_{1n}\sum_{i=1}^{n} w_i G_{mi}G_{pi} - \sum_{i=1}^{n} w_i G_{mi}G_{ki}}{\sum_{i=1}^{n} w_i {G_{mi}}^2} \quad (12)$$

We introduce the following definitions:

$$q_1 = G_{pi} - \frac{G_{mi}\sum_{j=1}^{n} G_{mj}G_{pj}}{\sum_{j=1}^{n} {G_{mj}}^2},\ q_2 = G_{ki} - \frac{G_{mi}\sum_{j=1}^{n} G_{mj}G_{kj}}{\sum_{j=1}^{n} {G_{mj}}^2}. \quad (13)$$

Now we can define the partial derivative of J(R) by r_{1n}:

$$\frac{\partial J(R)}{\partial r_{1n}} = 2\sum_{j=1}^{n} w_i(r_{1n}q_1 - q_2)\,q_1 = 0 \quad (14)$$

Then

$$r_{1n} = \frac{\sum_{k=1}^{n} w_i q_1 q_2}{\sum_{k=1}^{n} w_i q_1^2}. \quad (15)$$

We define the elements of the vector of parallel transfer T through the elements of the rotation matrix:

$$t_k = \frac{1}{n}\sum_{i=1}^{n} w_i\left(y_{ki} - (r_{k1}x_{1i} + r_{k2}x_{2i} + r_{k3}x_{3i})\right) = 0 \quad (16)$$

In this we investigate the accuracy and convergence rate of the proposed algorithm and state of art algorithms based on the ICP for registering point clouds in controlled conditions (Fig. 1) and uncontrolled conditions (Fig. 2).

Uncontrolled conditions are associated with the effect of various noises (poor lighting, shooting at night, clouds, etc.). The algorithms of solving the variational ICP problem, using various options for minimizing the functional J(R,T), were considered: based on point-to-point metrics with and without extrapolation, using point-plane metrics [14]. The computational experiments have shown a significant increase in the accuracy of the fusion ICP method for different class of archaeological sites. As a result of computer modeling, it was found that the use of visually related characteristics to solve the variational problem of the ICP algorithm makes it possible to improve the convergence of the method under uneven illumination.

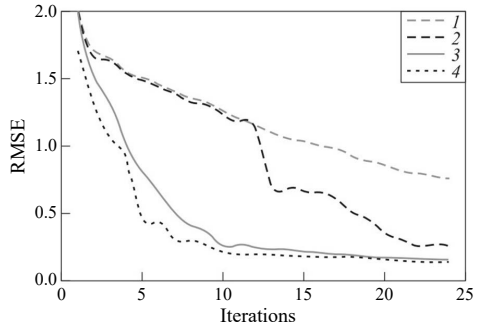

Fig. 1. Comparative analysis of registration methods in terms of convergence in controlled conditions (1- point-to-point ; 2- point-to-point with extrapolation; 3-point-to-plane; 4- suggested algorithm).

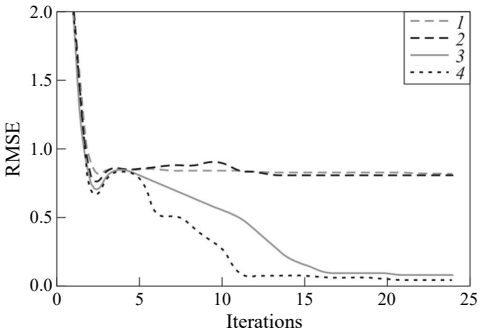

Fig. 2. Comparative analysis of registration methods in terms of convergence in uncontrolled conditions (1- point-to-point; 2- point-to-point with extrapolation; 3-point-to-plane; 4- suggested algorithm).

The proposed method improves the quality of the work of two key steps of the ICP method: determining the corresponding points between a pair of point clouds and solving the variational problem and differs from the state of art methods in that: it allows solving the problem of the dependence of the result of solving the variational problem ICP on the correctness of the choice of initial values; it is used to register point clouds with arbitrary spatial resolution and scale relative to each other; give accurate estimates for complex large-scale archaeological sites [14].

III. A Fuision Semantic Segmentation Algorithm Based on the Yolo Detector and the 3DEF Method

In the proposed approach to semantic segmentation of archaeological sites, a point cloud is divided into supervoxels using the Voxel Cloud Connectivity (VCCS) segmentation method [15]. At the next stage we use an “area growth” algorithm which recursively integrates two bordering 3D blocks into one big block using the distance function. This function is presented as a linear combination of next features: RGB, coordinates corresponding points in point clouds (point-to-plane distance), coordinates of normal drawn to the tangent plane at point p_i of the point cloud. Then the semantic classification procedure is performed, which consist of three steps:

Step 1. For each block in point cloud an algorithm selects a list of closest blocks. Algorithm uses three parameters [16]:

- the distance from the point to the plane.
- the closed angles.
- the Euclidean distance.

Step 2. Machine learning based on ResNet 50.

Step 3. Output procedure.

During step 2 and step 3, the algorithm uses this list of blocks to evaluate in series of tests that allow to detect "entangled objects". These objects allow to describe complex geometric relationships between blocks located in the neighborhood in a point cloud.

The disadvantage of 3DEF classifier is the dependence of the classification accuracy on the viewing angle for observing a 3D scene, to eliminate this disadvantage, several representations of the scene (Multi-viewed 3D Entangled Forest) are used in the work [11]. A semantic 3D segmentation method such as 3DEF can accurately segment large structural elements of a scene such as ditches, mounds and tracks. This method does not classify terrain details well, since semantic segmentation of such objects is based on color information. To eliminate this shortcoming in the classification of these objects, the YOLOv8 object detector is used.

To integrate the YOLOv8 object detector into the proposed approach, it is necessary to: select an object in the frame using a detector; within the detector's bounding box, carry out the segmentation procedure of the object along its contours based on the Grabcut method. GrabCut is an image segmentation method based on graph sections, in which a weighted graph is constructed based on neighboring pixels and labeled masks. Next, background or foreground segmentation is created, combining rigid segmentation by iterative optimization of the graph with matting of the boundaries in order to work with blurred and mixed pixels on the boundaries of the object. The Grabcut detector can be is used for both color data and RGB-D frame depth data. Segmentation obtained from RGB data and depth data can be combined using the logical OR operation. For each detected object, the Grabcut method is initialized by the YOLOv8 detector's bounding box for reliability, since Grabcut cannot always find segmentation, the number of iterations in the method is 5. If the Grabcut method cannot find segmentation, then there are two alternatives: skipping detection or considering the entire bounding field as the foreground. After this step, YOLOv8 labels and Grabcut labels can be saved, while the bounding boxes of the detector and the Grabcut method in the frame may overlap, so the order of processing the bounding boxes may negatively affect the results of the pipeline of methods as a whole. For reliable semantic segmentation, the YOLOv8 bounding boxes are sorted in decreasing order of size [17].

Combining the results of semantic segmentation of the 3DEF method and a fusion based on YOLOv8 + Grabcut methods is carried out using a merger based on the Bayes formula. Let's present the above in the form of two algorithms: algorithm 1 is a semantic segmentation algorithm based on clarifying the position of the detecting frame; Algorithm 2 is a semantic segmentation algorithm combining a 3DEF segmentation algorithm and a segmentation algorithm based on refined restrictive frameworks.

A. Algorithm 1. Semantic segmentation algorithm based on clarifying the position of the detecting frame.

Input data: a sequence of RGB-D frames – an image (I-RGB) and depth map (I-D).

Output: A segmented image marked with labels and probabilities.

Step 1. Detecting objects in an I-RGB image using the YOLOv8 method – determining the bounding boxes of images.

Step 2. Sort the list of bounding boxes (priors, anchors) in order of decreasing the size of the area of the frames. This stage solves the problem of overlapping frames during the detection procedure.

Step 3. Perform the following actions in a loop for each bounding box in the list [16]:

- perform semantic segmentation of data on an I-RGB image using the Grabcut method (GM).
- perform semantic segmentation of data on the I-D depth map using the GM. Let's perform a graph section optimization method for 5 iterations.
- for robust segmentation, we use fusion of data obtained from RGB and depth frames using the logical OR pixel-by-pixel operation.
- we check the condition: if the GM can return any segmentation, the bounding box falls into the list that are subject to markup. Otherwise the initial frames of the detected object is the foreground.

Step 4. Image markup based on the list of bounding boxes obtained in Step 3. For each element of this list, the corresponding image is selected and each object of this image is assigned a label with the name of the object class and the degree of probability of being assigned to this class.

B. Algorithm 2. A combined semantic segmentation algorithm using Algorithm 1 and the 3DEF method.

Input data: a frame containing 2D and 3D data – an image, a depth map, a three-dimensional point cloud.

Output data: segmented marked smoothed 3D point cloud with labels and probability values.

Step 1. Execute Algorithm 1 for semantic segmentation using color and depth map information. Get a marked-up image with class labels and the probability value of assigning a segmented object to a class.

Step 2. Apply the 3DEF semantic segmentation method to the data in the form of a point cloud. At the output of the method, we get a list of classes and a marked-up point cloud. The 3DEF method consists of the following steps: super-segmentation of super-pixels in 3D patches; combining similar neighboring segments into larger, mostly flat segments; segment classification.

Step 3. For each frame, merge the output data obtained in Step 1 (labeled image) and Step 2 (labeled point cloud) based on the Bayes formula. The merge is applied to the values of the probability of objects belonging to the class. At the output of this step, we get a new marked-up three-dimensional point cloud.

Step 4. Smoothing the data. This step takes into account the pixel context using Moore neighborhood. This step allows to increase robustness algorithm against noise in locally inaccurate registration based on ICP and errors from forward projection process [16]. At the output we get a marked smoothed 3D point cloud.

IV. COMPUTER SIMULATION

In this section, the results of computer modeling are presented and discussed. An Intel Core i7-based computer with a graphics processor was used to conduct the simulation. Let's evaluate the accuracy and convergence of the proposed method using the example of an annotated collection of digital data on archaeological monuments of the Bronze Age in the Chelyabinsk region.

The marked-up data set contains 9 classes of objects related to the main objects in the archaeological site:

- Bronze Age dirt mound K-1 (on virgin soil, covered with turf).
- bronze Age dirt mound K-2 (plowed surface).
- early Iron Age K-3 dirt mound (on virgin soil, covered with turf).
- stone mound of the early Iron Age or the Middle Ages (with a stone shell) K-4.
- a dirt or stone mound of the early Middle Ages with a "mustache" K-5.
- funeral cult complexes with shaft-shaped horseshoe-shaped or dumbbell-shaped structures of the Middle Ages M-1.
- burial grounds with stone fences of the Middle Ages M-2.
- fortified settlement of the Bronze Age P-1.
- an undefended settlement of the Bronze Age P-2.

The results of semantic segmentation of the marked-up data set are used both as a training sample for a neural network and as a reference data set, on which the accuracy of the proposed registration method and state of art methods of 3D reconstruction is checked.

Similarly to other well-known approaches three indicators are used in the work: general accuracy (OT), class completeness (KP), class accuracy (CT). The last two indicators are less dependent on the value of the classes frequency in the analyzed scene as a whole. Let's analyze various combinations of the 3DEF segmentation method and the YOLO object detector and evaluate the impact of using multiple representations on the accuracy of the semantic segmentation method. Table 1 shows that the integration of the YOLOv8 object detector always significantly improves the segmentation accuracy compared to the basic 3DEF method or its modification MV-3DEF. The MV-3DEF+YOLO-based method works slightly better than 3DEF + YOLO, but on the other hand, the first method requires significantly more computational resources when merging data about different representations of the scene than the second method of semantic space labeling.

Table 2 and Table 3 present the results of semantic segmentation of the space for the created data set, which contain class accuracy values for the kurgan classes and for other archaeological objects.

TABLE I. ACCURACY AND COMPLETENESS OF SEMANTIC SEGMENTATION METHOD (IN %)

Semantic processing method	**Name Accuracy**		
	Class accuracy	***Overall accuracy***	***Class completeness***
3D Entangled Forest	55.8	63.4	55.2
Multi-view - 3DEF	58.1	66.4	58
3DEF and YOLOv8	65.4	68.1	56.9
Multi-view 3DEF and YOLOv8	67.1	69.5	58.2

TABLE II. ACCURACY (IN %) OF SEMANTIC SEGMENTATION METHODS FOR THE KURGAN CLASSES

Semantic processing method	**The name class of the archaeological object**				
	K-1	***K-2***	***K-3***	***K-4***	***K-5***
3D Entangled Forest	86	95.5	35.4	79.2	37.3
Multi-view 3DEF	84.2	96.8	31.2	81.1	49.6
3DEF and YOLOv8	83.2	96.7	29.1	77.1	39.8
Multi-view 3DEF and YOLOv8	84.3	98.1	40.3	83.4	56.1

As you can see, in both tables, a significant increase in class accuracy is achieved by combining the 3DEF classifier and the YOLO object detector for most classes of objects that are represented in created dataset. It is known that the YOLO, Mask-R-CNN, 3DEF methods give different accuracy estimates for different classes.

TABLE III. ACCURACY (IN %) OF SEMANTIC SEGMENTATION METHODS FOR BURIAL GROUNDS AND SETTLEMENTS

Semantic processing method	**The name class of the archaeological object**			
	P-1	***P-2***	***M-1***	***M-2***
3D Entangled Forest	72.4	67.9	27.7	44.6
Multi-view 3DEF	73.9	69.3	30.1	48
3DEF and YOLOv8	88.3	87.2	25.1	50.6
Multi-view 3DEF and YOLOv8	89.6	86.8	25.7	53.2

Table 3 shows that the best accuracy results of 3DEF+YOLO exceed the results based on 3DEF or MV-3DEF (+5.2% in CT, +2.3% in OT and +2.5% in KP), on the other hand, for classes such as M-1 и M-2 the proposed combination of methods does not give a significant increasing the accuracy of classification, for example.

This circumstance is due to the fact that the 3DEF segmentation method does not classify flat objects well: when calculating the class accuracy index, this method receives a penalty for detecting objects belonging to classes with a large number of samples in the reference dataset on the scene. This disadvantage of 3DEF cannot be compensated by the YOLOv8 object detector, since it is not trained in these classes. The solution to this problem can be based on training an object detector for these classes or by modifying the procedure for pre-segmentation of a growing area in 3DEF.

V. Conclusions

The paper presents a new robust method of semantic classification of archaeological objects based on the Baysevsky fusion of object detector and semantic segmentation and classification methods in 2D and 3D space. The use of a combination of various methods of semantic processing of RGB-D data based on convolutional neural networks made it possible to accurately classify archaeological objects of different classes. The results of the semantic classification of objects were used in the project to improve the quality indicators of the 3D registration algorithm in terms of the error value (RMSE) and the number of iterations of the method (convergence index). The paper proposes a accurate fusion method for registering (F-ICP) for a group of affine transformations to allow the reconstruction of an archaeological objects in the form of a 3D point cloud based on solving the ICP variational problem in a closed form using the point-to-point metric. The proposed method improves the quality of the work of two key steps of the ICP method: determining the corresponding points between a pair of point clouds and solving the variational problem.

Acknowledgment

The work was supported by the Russian Science Foundation, project No. 23-11-20007 and the Chelyabinsk Region, RSF-RK contract No. 584.

References

[1] K. Lambers, W. V. Verschoof-van der Vaart, and Q. P. G. Bourgeois, “Integrating remote sensing, machine learning, and citizen science in dutch archaeological prospection,” Remote Sensing, vol. 11, no. 7, pp. 794, March 2019.

[2] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” IEEE Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), USA, vol 1. pp. 7263–7271, June 2016.

[3] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 2018. [Online]. Available: arXiv:1804.02767.

[4] G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics,” 2023. [Online]. Available: https://github.com/ultralytics/

[5] D. Wolf, J. Prankl, and M. Vincze, “Enhancing semantic segmentation for robotics: the power of 3-d entangled forests,” IEEE Robotics and Automation Letters, vol. 1, no. 1, pp. 49–56, April 2016.

[6] H. Kaiming, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” 2018. [Online]. Available: arXiv:1703.06870v3.

[7] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” ACM Transactions on Graphics (TOG), vol. 23, no. 3, pp. 309–314, August 2004.

[8] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” Proceedings of 29th Conference on Neural Information Processing Systems (NIPS), Canada, vol. 4, pp. 1–9, June 2015.

[9] P. Besl and N. McKay, “A method for registration of 3-D shapes,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[10] B. K. P. Horn, “Closed-form solution of absolute orientation using unit quaternions,” Journal of the Optical Society of America A, vol. 4, no. 4, pp. 629–642, April 1987.

[11] M. Antonello, D. Wolf, J. Prankl, S. Ghidoni, E. Menegatti, and M. Vincze, “Multi-view 3d entangled forest for semantic segmentation and mapping,” IEEE Proceedings of the International Conference on Robotics and Automation (ICRA), pp. 1855–1862, May 2018.

[12] S. Rusinkiewicz and M. Levoy, “Efficient variants of the ICP algorithm,” IEEE Proceedings of International Conference on 3-D Digital Imaging and Modeling (3DIM), Canada, vol 1. pp. 145–152, May-June 2001.

[13] A. V. Vokhmintcev, A. V Melnikov, and S. A. Pachganov, “Simultaneous localization and mapping method in three-dimensional space based on the combined solution of the point-point variation problem icp for an affine transformation,” Informatics and Applications, vol. 14, no. 1, pp. 101–112, 2020.

[14] A. V. Vokhmintcev, A. V. Melnikov, K. V. Mironov, and V. V. Burlutskiy, “Reconstruction of three-dimensional map based on closed form solution of variational problem of multi- sensor data registration,” Doklady Mathematics, vol. 99, no. 1, pp. 108–112, May 2019.

[15] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, “Voxel cloud connectivity segmentation-supervoxels for point clouds,” IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2027–2034, June 2013.

[16] M. Antonello, S. Chiesurin, and S. Ghidoni, “Enhancing semantic segmentation with detection priors and iterated graph cuts for robotics,” Engineering Applications of Artificial Intelligence, vol. 90, pp. 103467, April 2020.

[17] A. Vokhmintcev and M. Timchenko, “The new combined method of the generation of a 3d dense map of environment based on history of camera positions and the robot's movements,” Acta Polytechnica Hungarica, vol. 17, no. 8, pp. 95–108, 2020.