Dmitry I. Ignatov · Michael Khachay ·
Andrey Kutuzov · Habet Madoyan ·
Ilya Makarov · Irina Nikishina ·
Alexander Panchenko et al. (Eds.)

LNCS 14486

Analysis of Images, Social Networks and Texts

11th International Conference, AIST 2023
Yerevan, Armenia, September 28–30, 2023
Revised Selected Papers

AIST

Springer

Lecture Notes in Computer Science **14486**

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.

LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Dmitry I. Ignatov · Michael Khachay · Andrey Kutuzov · Habet Madoyan · Ilya Makarov · Irina Nikishina · Alexander Panchenko · Maxim Panov · Panos M. Pardalos · Andrey V. Savchenko · Evgenii Tsymbalov · Elena Tutubalina · Sergey Zagoruyko
Editors

Analysis of Images, Social Networks and Texts

11th International Conference, AIST 2023
Yerevan, Armenia, September 28–30, 2023
Revised Selected Papers

Springer

Editors
Dmitry I. Ignatov
National Research University Higher School of Economics
Moscow, Russia

Andrey Kutuzov
University of Oslo
Oslo, Norway

Ilya Makarov
Artificial Intelligence Research Institute
Moscow, Russia

Alexander Panchenko
Skolkovo Institute of Science and Technology
Moscow, Russia

Panos M. Pardalos
University of Florida
Gainesville, FL, USA

Evgenii Tsymbalov
Apptek
Aachen, Germany

Sergey Zagoruyko
MTS AI
Moscow, Russia

Michael Khachay
Krasovskii Institute of Mathematics and Mechanics of Russian Academy of Sciences
Yekaterinburg, Russia

Habet Madoyan
American University of Armenia
Yerevan, Armenia

Irina Nikishina
University of Hamburg
Hamburg, Germany

Maxim Panov
Mohamed bin Zayed University of Artificial Intelligence
Abu Dhabi, United Arab Emirates

Technology Innovation Institute, UAE
Abu Dhabi, United Arab Emirates

Andrey V. Savchenko
National Research University Higher School of Economics
Nizhny Novgorod, Russia

Elena Tutubalina
Kazan Federal University
Kazan, Russia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-54533-7 ISBN 978-3-031-54534-4 (eBook)
https://doi.org/10.1007/978-3-031-54534-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This volume contains the refereed proceedings of the 11th International Conference on Analysis of Images, Social Networks and Texts (AIST 2023)[1]. The previous conferences (during 2012–2021) attracted a significant number of data scientists, students, researchers, academics, and engineers working on interdisciplinary data analysis of images, texts, and networks. The broad scope of AIST makes it an event where researchers from different domains, such as computer vision and natural language processing, exploiting various data analysis techniques, can meet and exchange ideas. As the test of time has shown, this leads to the cross-fertilisation of ideas between researchers relying on modern data analysis machinery.

Therefore, AIST 2023 brought together all kinds of applications of data mining and machine learning techniques. The conference allowed specialists from different fields to meet each other, present their work, and discuss both theoretical and practical aspects of their data analysis problems. Another important aim of the conference was to stimulate scientists and people from industry to benefit from knowledge exchange and identify possible grounds for fruitful collaboration.

The conference was held during September 28–30, 2023. The conference was organized with the support of Zaven and Sonia Akian College of Science and Engineering, American University of Armenia.

This year, the key topics of AIST were grouped into five tracks:

1. Data Analysis and Machine Learning chaired by Evgenii Tsymbalov (Apptek, Germany) and Maxim Panov (Mohamed bin Zayed University of Artificial Intelligence and Technology Innovation Institute, UAE)
2. Natural Language Processing chaired by Andrey Kutuzov (University of Oslo, Norway) and Elena Tutubalina (Kazan Federal University, Russia)
3. Network Analysis chaired by Irina Nikishina (Universität Hamburg, Germany) and Ilya Makarov (HSE University and AIRI, Russia)
4. Computer Vision chaired by Sergei Zagoruyko (MTS AI, Russia), and Andrey Savchenko (HSE University, Russia)
5. Theoretical Machine Learning and Optimization chaired by Panos Pardalos (University of Florida, USA) and Michael Khachay (IMM UB RAS and Ural Federal University, Russia)

The Program Committee and the reviewers of the conference included 137 well-known experts in data mining and machine learning, natural language processing, image processing, social network analysis, and related areas from leading institutions of many countries including Armenia, Austria, the Czech Republic, Finland, France, Georgia, Germany, Greece, India, Ireland, Italy, Mexico, Montenegro, the Netherlands, Norway, Portugal, Russia, Slovenia, Spain, the UAE, the UK, and the USA. This year, we received 106 submissions, mostly from Russia but also from Armenia, China, Finland, France,

[1] https://aistconf.org.

Georgia, Germany, India, Iraq, Kyrgyzstan, Saudi Arabia, Singapore, Switzerland, the UAE, and the USA.

Out of the 106 technical submissions, 13 were poster submissions, and 76 submissions remained after desk rejects and withdrawals. For the remaining papers, only 24 were accepted into this main volume. In order to encourage young practitioners and researchers, we included 21 papers in the companion volume published in Springer's Communications in Computer and Information Science (CCIS) series. Thus, the acceptance rate of this LNCS volume is 32%. Each submission was double-blind reviewed by at least three reviewers, experts in their fields, in order to supply detailed and helpful comments.

The conference featured several invited talks and tutorials dedicated to current trends and challenges in the respective areas.

The invited talks from academia covered a wide range of machine learning and artificial intelligence areas:

– Narine Sarvazyan (George Washington University and American University of Armenia): "Decoding Hyperspectral Imaging: From Basic Principles to Medical Application"
– Hakim Hacid (Technology Innovation Institute): "Towards Edge AI: Principles, current state, and perspectives"
– Samuel Horvath (MBZUAI): "Towards Real-World Federated Learning: Addressing Client Heterogeneity and Model Size"
– Artem Shelmanov (MBZUAI): "Safety of Deploying NLP Models: Uncertainty Quantification of Generative LLMs"
– Muhammad Shahid Iqbal Malik (HSE University): "Threatening Content and Target Identification in low-resource languages using NLP Techniques"

We would like to thank the authors for submitting their papers and the members of the Program Committee for their efforts in providing exhaustive reviews.

According to the track chairs, and taking into account the reviews and presentation quality, the Best Paper Awards were granted to the following papers:

– Data Analysis and Machine Learning: "Ensemble Clustering with Heterogeneous Transfer Learning" by Vladimir Berikov;
– Natural Language Processing: "Benchmarking Multi-Label Topic Classification in Kyrgyz Language" by Anton Alekseev, Sergey Nikolenko, and Gulnara Kabaeva;
– Network Analysis: "Limit Distributions of Friendship Index in Scale-Free Networks" by Sergei Sidorov, Sergei Mironov, and Alexey Grigoriev;
– Computer Vision: "DeepLOC: Deep Learning-based Bone Pathology Localization and Classification in Wrist X-ray Images" by Razan Dibo, Andrey Galichin, Pavel Astashev, Dmitry V. Dylov, and Oleg Y. Rogov;
– Theoretical Machine Learning and Optimization: "Is Canfield Right? On the Asymptotic Coefficients for the Maximum Antichain of Partitions and Related Counting Inequalities" by Dmitry Ignatov.

We would also like to express our special gratitude to all the invited speakers and industry representatives. We deeply thank all the partners and sponsors, especially the hosting organization: Zaven and Sonia Akian College of Science and Engineering, American University of Armenia, with special thanks to Habet Madoyan and Amalya Hambardzumyan, as local organizers. Our special thanks go to Springer for their help, starting from the first conference call to the final version of the proceedings. Last but not least, we are grateful to the volunteers, whose endless energy saved us at the most critical stages of the conference preparation.

Here, we would like to mention that the Russian word "aist" is more than just a simple abbreviation as (in Russian) it means "stork". Since it is a wonderful free bird, a symbol of happiness and peace, this stork gave us the inspiration to organize the AIST conference series. So we believe that this conference will likewise bring inspiration to data scientists around the world!

December 2023

Dmitry Ignatov
Michael Khachay
Andrey Kutuzov
Habet Madoyan
Ilya Makarov
Irina Nikishina
Alexander Panchenko
Maxim Panov
Panos Pardalos
Andrey Savchenko
Evgenii Tsymbalov
Elena Tutubalina
Sergey Zagoruyko

Application of Dynamic Graph CNN* and FICP for Detection and Research Archaeology Sites

Aleksandr Vokhmintcev[1,2](✉), Olga Khristodulo[3], Andrey Melnikov[2], and Matvei Romanov[1,4]

[1] Chelyabinsk State University, Chelyabinsk, Russia
vav@csu.ru
[2] Ugra State University, Khanty-Mansiysk, Russia
melnikovav@uriit.ru
[3] Ufa University of Science and Technology, Ufa, Russia
hristodulo.oi@ugatu.su
[4] South Ural State University, Chelyabinsk, Russia

Abstract. The paper proposes a methodology for solving the task of accurate semantic classification of 3D data using a combination of 2D and 3D methods based on the YOLO detector and the modified DGCNN network. The methodology is tested on the example of the problem of classification of large-scale geospatial objects, such as digital relief models of archaeological sites. A method for accurate registration of objects (FCIP) in the class of affine transformations using geometric and color features was proposed. The results of computer modeling of the proposed methodology based on FICP+DGCNN*+YOLO were presented and discussed. The methodology has theoretical and applied significance not only for the decryption and research of archaeological sites, but also for many applications of digital information processing and robotics in general.

Keywords: 3D semantic segmentation and classification methods · object detector · DTM · DGCNN · ICP

1 Introduction

Currently, remote research methods are becoming more important in archaeology, which allow obtaining information about archaeological sites without resorting to excavation methods. Remote research methods are very diverse, for example, scanning of the archaeology sites using magnetometers (SQUID, MOKE, Torque) and georadars is often used, low-altitude aeromagnetic photography using UAVs is used, recently more and more attention has been paid to satellite remote sensing, which allows obtaining satellite images with high spatial resolution. Remote research methods make it possible to create extensive sets of archaeological data in the form of a digital terrain model (DTM) or a digital

D. I. Ignatov et al. (Eds.): AIST 2023, LNCS 14486, pp. 294–308, 2024.
https://doi.org/10.1007/978-3-031-54534-4_21

surface model (DSM), with the help of which it is possible to study the spatial organization of various types of archaeological sites. To date, the following approaches for solving the problem of semantic classifying objects are known: the first approach uses feature points and other of object properties; the second approach uses statistical methods; the third approach is based on the use of machine learning methods and, above all, neural networks. Methods based on the search for object characteristics [1,2] have high computational complexity, are poorly formally described and don't control the detection quality. These methods show poor results when detecting objects in an image with a complex spatially inhomogeneous background distorted by interference of various nature, for example, uneven illumination [3,4]. For statistical pattern recognition, object recognition methods based on ordinal statistics using linear or nonlinear filtering are used [5,6]. The main disadvantage of these methods is the erroneous identification of the object of interest with objects in the background for contextually complex scenes. To overcome this disadvantage, various methods and algorithms have been proposed, which are based on spatial differentiation, distortion of the shape of the object in the image, adaptive and locally adaptive linear and nonlinear filters [7]. Recently, when deciphering archaeological sites, researchers often use 3D mathematical models and machine learning methods to analyze geospatial data. To solve the problem of high computational resource intensity in the analysis of archaeological data, neural network models are used on sparse data structures [8] and variational autoencoders [9]. Moreover, there are certain achievements in reducing the size of the training sample: for example, neural networks are known that are robust to changing the spatial position of 3D objects [10]; models trained with partial (semi-supervised) markup [7].

Modern methods for classification and segmentation of 3D data use semantic signals and sequences in data to increase accuracy, and the development of appropriate algorithms for processing point clouds is an active area of scientific research. Methods for classifying and segmenting 3D objects in the form of a point cloud can be divided into indirect and direct. Indirect methods use a sequence of 2D images of the original point cloud obtained from different viewpoints in 3D space. Then this sequence is sent to the CNN input, and after all, the CNN output, which is a pixel-by-pixel semantic markup map, is projected back into 3D space. Indirect methods include MVCNN (Multi-view Convolutional Neural Networks), SnapNet, SnapNet-R methods [11], which use RGB data and depth data, and VoxNET, SEGCloud, PointGrid methods [12], which use voxel representation. The main disadvantage of indirect methods is that they work well only with polygonal models, for processing contextually complex models, including those containing non-convex point clouds, their use is extremely limited due to the poor quality of segmentation. Also, the process of data registration in multiview methods requires large computational costs when taking pictures from the required angles, and application using voxel models require a significant amount of memory to store the results. Recently, when developing methods for semantic processing of 3D data, researchers have focused on direct methods that use various neural network architectures and extract features automatically, for example, based on Kohonen maps SO-Net [13], RNN (Recurrent Neural Networks), ResNet CNN (Residual Neural Networks) [14]. The disadvantages of these methods are the complexity of the model and low quality values in terms of accuracy and completeness when processing local features of the object, in our case, details of the microrelief. To overcome this disadvantage, methods have been proposed that can be divided into two groups: the PointNet [15] and PointNet++ group [16] and a group of methods based on GCNN (Graph Convolutional Neural Networks) [17] and its variants DGCNN (Dynamic Graph CNN) [18], RGGNN (Regularized Graph CNN) [19]. The methods of the first group in the 3D semantic processing of the point cloud use local objects properties and don't use data about a complex geometric relationships between points. These methods are invariant to permutations. The methods of the second group use information about the surface of a 3D object and are based on the use of convolution operations on spatial graphs when solving the problem of semantic 3D classification and segmentation. As an example, 3D classification method based on RGCNN uses the Laplace matrix to construct an undirected graph, which causes one of its main disadvantages - low performance, the computational complexity of graph construction in the method can be estimated as $O(n)^3$. To improve the performance of methods of this group, spectral filtering approximation methods (Cayley and Chebyshev polynomials, the Lanczos method, etc.) are often used. Despite numerous modifications, the GCNN method and its variants are not local, which is due to the fact that these methods usually use only one modality - geometric relationships between points in the cloud. The most promising method of the second group is DGCNN, which has a special EdgeConv layer that allows to achieve invariance to rotation, parallel transfer or scaling. This paper proposes a new method of semantic data processing based on a dynamic weighted graph, which combines the advantages of the well-known architectures of DGCNN and RGCNN, but is devoid of their known disadvantages, such as:

– the dependence of the accuracy of the method on the shape of the point cloud and the method of obtaining the analyzed point cloud;
– limiting the dimension of the point cloud when performing the 3D segmentation and classification procedure;
– the use of one modality when performing the convolution operation on spatial graphs and, as a consequence, the insufficient quality of semantic processing of 3D data.

In our research we use following data sources: materials of aerial photography (from the 50 s and the 60 s-80 s last century for the purposes of agriculture and geodesy respectively, the aerial frames were taken at a scale of 1:14 000 with high resolution for the entire territory of the Kizilsky district of the Chelyabinsk region; results of remote sensing of the Resource-P (from 2013 to 2021), Canopus-B (from 2013 to 2023) satellites; total station survey data obtained using the Trimble 3300 (Elta R55); DTM and DSM created as a result of archeology expeditions to the settlements of Stepnoye and Levoberezhnoye from 2006 to the present. These materials form a data set on archaeological sites and objects of the Bronze Age on the territory along the river Sintashta. Data sources were divided into two groups: 2D in the form of RGB frames $I - RGB = \{F_1, \ldots, F_n\}$ and 3D data in the form of point clouds $I - D = \{d_1, \ldots, d_k\}$. Let's consider the algorithms for processing these two groups of data: Algorithm 1 for 2D data and Algorithm 2 for 3D data. Algorithm 1 is represented as the following sequence of steps:

- Step 1. Trimming the edges (bottom = 1.5%, top = 1.5%, left (right) sides = 1%);
- Step 2. Unification of the direction of the images;
- Step 3. Image restoration and noise removal;
- Step 4. Increase the contrast of images;
- Step 5. Normalization of the snapshot size.

Algorithm 2 is represented as the following sequence of steps:

- Step 1. Division of a 3D model into semantic blocks (step = 0.01);
- Step 2. Sampling of point clouds (Down Sampling);
- Step 3. Extraction of singular points;
- Step 4. Calculations of normals in the point cloud.

This article is organized as follows: in the second chapter an accurate algorithm FICP for registering 3D data and the results of a comparative analysis of the proposed algorithm with known solutions are proposed, in the third chapter, a multimodal modified neural network architecture based on DGCNN (DGCNN*) for classifying archaeological objects is proposed, the fourth chapter presents and discusses the results of computer simulation for the proposed methodology of 3D semantic classification of archaeological objects in comparison with known modern approaches of solving this problem.

2 Fusion 3D Registration Algorithm (FICP)

It is known that the convergence and accuracy of the ICP (Iterative Closest Point) algorithm proposed in the works [20] can be significantly improved. Known ICP methods are characterized by the following disadvantages: firstly, classical ICP methods do not take into account the local shape of the surface around each point in a 3D point cloud when analyzing and processing information, and secondly, known ICP-based data logging methods have great computational complexity, while the most expensive operation is the search for the nearest points; thirdly, the result of solving the variational problem depends on the correctness of the choice of the initial approximation. The last drawback of the ICP method can be eliminated, for this purpose special points are used in the study, which, as is known, match data frames without specifying initialization parameters. Horn proposed a solution to the conditional variational problem in closed form for affine and orthogonal transformations, in this paper a closed form solution for affine transformations is obtained, which allows: to register non-convex objects on a 3D data; to obtain a solution of the ICP variational problem in a closed form for various degenerate cases associated with the location of points in a 3D data on the same straight line (plane). The data matching algorithm proposed in the articles is used to process characteristic points on an image in an RGB-D frame. The process of processing feature points is represented as the following sequence of steps (Algorithm 3):

- Step 1. Determination of feature points in an RGB-D frame using DHNG (descriptor based on histograms of directional gradients) [21];
- Step 2. Matching feature points in two consecutive RGB-D frames;
- Step 3. Elimination of outliers in the data by estimating the parameters of a 2D image model based on random samples (RANSAC);
- Step 4. Solving the variational problem of 2D data registration with respect to singular points in frames.

Let's define the values of the normalized centered DHNG

$$\overline{HOG_i^R}(\alpha) = (HOG_i^R(\alpha) - Mean^R)/\sqrt{(Var^R)}, \tag{1}$$

where $HOG_i^R(\alpha)$ is GNG value in each position of the i-th round sliding window, $Mean^R$ is the mean HNG value, Var^R is the HNG variance. Let's take a closer look at the step. 4. The solution of the variational problem with respect to singular points in the frame is represented as

$$J(R^F) = \frac{1}{|A_f|}\sum_{i\in A_f}^{n} w_i \parallel M(R^F Hog_x^i) - M(Hog_y^i) \parallel^2, \tag{2}$$

where R^F is the affine transformation matrix for feature points for color data; w_i are the weight characteristics; Hog_x^i and Hog_y^i are singular points in two consecutive frames, respectively: $Hog_x^i = (x_{1f}^i, x_{2f}^i, x_{3f}^i)^T$, $Hog_y^i = (y_{1f}^i, y_{2f}^i, y_{3f}^i)^T$, where M is the function of converting the coordinates of points of a 3D scene Hog_x^i and $Hog_y^i \in R^3$ into the coordinate system of the camera $C^i = (C_x^i, C_y^i, D^i) \in R^3$, where C_x^i, C_y^i are the corresponding coordinates of points in pixel space, D^i is the depth value in pixel space.

$$C_x^i = \frac{f}{x_{3f}^i} x_{1f}^i + O_x, C_y^i = \frac{f}{x_{3f}^i} x_{2f}^i + O_y, D^i = \sqrt{{x_{1f}^i}^2 + {x_{2f}^i}^2 + {x_{3f}^i}^2}, \tag{3}$$

where O_x and O_y are the coordinates of the image center in pixel space, f is the camera focus. The coordinates in the frame Hog_y^i can be obtained in a similar manner. The process of decryption of the archaeology sites is carried out on the basis of digital DTM relief models obtained both from single view point and from different viewing points. In the second case, it is necessary to solve the problem of 3D data registration. The paper considers 3D models in the form of a point cloud. Let's $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ – the pair of point cloud in R^3, respectively. In digital information processing, an ICP algorithm [22,23] is often used to solve 3D data registration task. This algorithm searches for a geometric transformation between X and Y in the following form $Rx_i + T$, where R is the rotation matrix, T is the vector of parallel transfer, $i = 1, \ldots, n$ and uses an incremental an approach to calculating a sparse 3D model of a scene and a bundle method to refine the camera parameters and coordinates of points. In this paper we suggested an accurate algorithm Fusion ICP for registering 3D data with a point-to-point metric for affine transformations class, in which the closed solution of the variational problem is

$$J(R^F, R^D) = \frac{\alpha \sum_{i \in A_f}^{m} \| M(R^F Hog_x^i) - M(Hog_y^i) \|^2}{|S_f|} + \frac{(1-\alpha) \sum_{j \in A_d}^{n} \| R^D x_j + T - y_j \|^2}{|S_d|}, \quad (4)$$

where R^D is affine rotation matrices for depth data; T is a parallel transfer vector; α is hyperparameter, the value equal to 0.3 is used in the work; S_f a set which contains the correspondences between features points; S_d a set which contains the correspondences between points x_j and y_j in X and Y. Let $RT^*(i)$ be the best geometric transformation for the i-th step. This algorithm has been adapted to the task solution of the registration of archaeological sites, uses two channels: data on the frame color and data on the depth in the form of a point cloud and can be presented as procedure (Algorithm 4, FICP):

- Step 1. Determine the feature points using an image matching algorithm based on DHNG (Algorithm 3);
- Step 2. Establish a correspondence between the special points Hog_x^i and Hog_y^i in frames using a k-d tree to improve performance. We use the obtained result of matching points as the initial values of the geometric transformation $RT^*(0)$ of the ICP;
- Step 3. Establish the correspondence between the S_d points in the clouds X and Y using the KNN nearest neighbor method. Eliminate outliers based on the RANSAC;
- Step. 4. Solve the variational problem (4) with respect to the transformation $RT^*(i)$.

We search for the transformation until one of the stop criteria is met. We will investigate the accuracy (RMSE) and convergence rate of the proposed registration algorithm for a class of affine transformations of data in controlled (see Fig. 1) and uncontrolled conditions (see Fig. 2), which are associated with noises of various nature. To conduct tests, we will use the proposed DHNG descriptor and descriptors such as SIFT (Scale-invariant feature transform), SURF (Speeded up robust features), ORB (Oriented FAST and Rotated BRIEF) are known. From Fig. 1 it can be seen that the combination of the FCIP + DHNG registration method allows for better accuracy, but in general, under controlled conditions, the accuracy of solving the registration problem does not strongly depend on the choice of a 2D descriptor that is used to select the initial values of FCIP. In uncontrolled conditions(see Fig. 2), the FCIP + DHNG combination has significant advantages in terms of accuracy and convergence over all other combinations (FCIP + SIFT, FCIP + SURF, FCIP + ORB), our method of registering 3D data converges after the 10-th step. The processing time of algorithms based on DHNG and ORB is

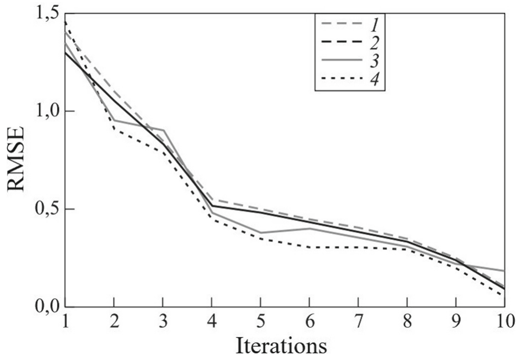

Fig. 1. Comparative analysis of registration methods in terms of convergence in controlled conditions (1 - SIFT; 2 - SURF; 3 - ORB; 4 - DHNG).

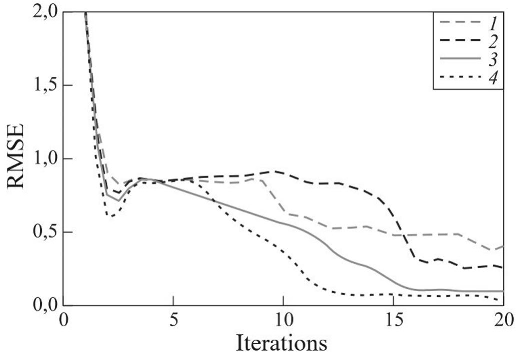

Fig. 2. Comparative analysis of registration methods in terms of convergence in uncontrolled conditions (1 - SIFT; 2 - SURF; 3 - ORB; 4 - DHNG).

about 2 s. The processing time of the SIFT algorithm is more then 15 s. The processing time of the SURF is 0.6 s. The method FCIP solves the problem of the dependence of the 3D registration result on the correctness of the selection of initial values R and T. FICP can be used for accurate registration of point clouds with arbitrary spatial resolution and scale relative to each other. DHNG descriptor allows to improve the convergence of the ICP.

3 Modified DGCNN* for Semantic Classification 3D Data

When deciphering archaeological sites, there is always a need for its 3D registration based on data obtained from different viewing points. Therefore, in order to overcome the disadvantages inherent in multi-species segmentation methods [11], a methodology (see Fig. 3) is used in which the following sequence of steps is performed:

– Step 1. 3D registration of an archaeological sites using a combined registration method based on a fusion iterative closest points algorithm (FICP);
– Step 2. 3D semantic segmentation of a point cloud associated with an archaeological site based on a modified DGCNN*;
– Step 3. Detecting objects using RGB-D data based on the YOLO detector [24];
– Step 4. Combining the results of segmentation and classification in Step 2 and Step 3 using the Bayes formula.

This methodology makes it possible to eliminate two key disadvantages of the DGCNN and RGCNN convolutional neural networks for classification and segmentation 3D data. The first disadvantage is the limitation of the dimensionality of the processed 3D data, which makes these methods inapplicable for many applied tasks related to the semantic processing of large-scale scenes, the second disadvantage is related to the dependence of these CNNs on the method of collecting data about the point cloud, for example, the quality of the registration of point clouds. The following modifications were made to the DGCNN architecture: the modality of the CNN has been expanded, the accurate solution is obtained based on a combination of various geometric and independent features of points in the cloud; the first EdgeConv layer has been replaced with a specialized layer that performs two functions: forms a multimodal feature vector consisting of point coordinates and their normalized coordinates, coordinates of normals and independent HSV color features; performs a higher discretization of the point cloud, which allows you to form a homogeneous dense point cloud and thereby get the best quality point cloud segmentation; the metric classifier that generates the output values of the DGCNN network has been replaced by a combination of two MLP network and an one RBF (Radial Basis Function) network - segmentation output. The scheme of concatenation of data from the outputs of various EdgeConv layers in the DGCNN network was also changed, which made it possible to increase the efficiency of processing local features of objects.

The introduced changes, as will be shown in Sect. 4, made it possible to obtain an accurate solution to the problem of semantic segmentation and classification for 3D large-scale DTMs, which is important when processing data on archaeological sites and objects. For some classes, the proposed 3D object classification method does not provide the required quality (at least 0.8 according to the F1-score), therefore, a combination of a classification 3D data method based on DGCNN* and a classification 2D data method based on the YOLO v.8 detector was used in the this work. The results of these methods were combined using the Bayes formula

$$p(j|m_{DGCNN*} \bigwedge m_{YOLO}) = w_j p(m_{DGCNN*}|j) \times p(m_{YOLO}|j), \quad (5)$$

where $p(m_{DGCNN*})$ is the confidence returned by DGCNN*, $p(m_{YOLO})$ is the confidence returned by YOLO detector, w_j is normalization hyperparameter, m_{DGCNN*} and m_{YOLO} are class labels for DGCNN* and YOLO respectively.

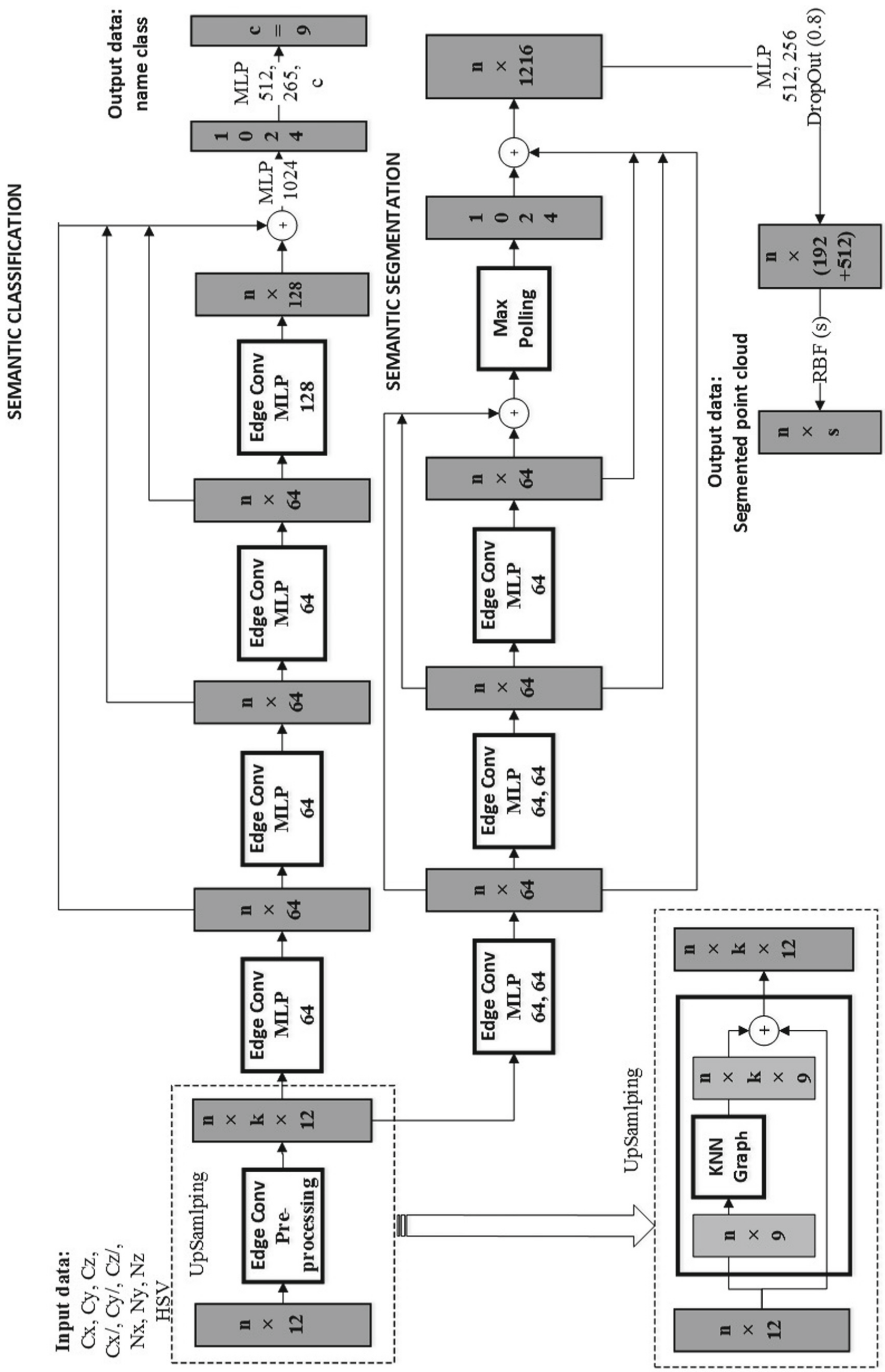

Fig. 3. The architecture of the modified DGCNN*.

Let's solve a variational problem of the following form in order to determine the class label with the highest probability value

$$J = \arg\max_{j} p(j|m_{DGCNN*} \bigwedge m_{YOLO}). \tag{6}$$

Two algorithms were proposed to solve the task: Algorithm 5. Semantic segmentation algorithm based on clarifying the position of the detecting frame and Algorithm 6. A combined semantic segmentation algorithm using Algorithm 5 and DGCNN* (see Fig. 4).

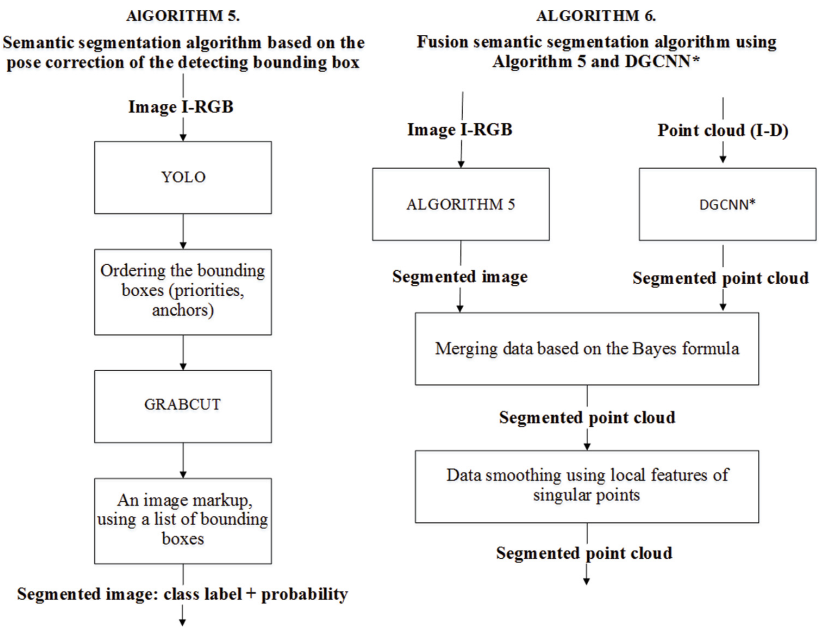

Fig. 4. The block diagrams of algorithms for 3D semantic classification and segmentation.

4 Computer Simulation

This section presents and discusses the results of computer modeling. Let's evaluate the accuracy and convergence of the proposed methodology in relation to the task of classifying archaeological objects in comparison with the known approaches to 3D classification based on MVCNN and DGCNN. In a comparative analysis, we will evaluate the quality of these methods both individually and in combination with the YOLO v8. For computer modeling, we will use the created data set about archaeological sites of the Bronze Age which consist of: aerial photography; results of remote sensing of the Resource-P, Canopus-B satellites; total station survey data obtained using the Trimble 3300 total station. We are going to expand the training subsample with examples of archaeological objects for the analysis of 2D data models by about 10 times using geometric transformations and Mask R-CNN. Archaeologists of Chelyabinsk State University have identified signs of deciphering archaeological sites which are characteristic of Bronze Age monuments in the Southern Urals [25]. The following classes of archaeological objects have been identified: Bronze Age dirt mound (on virgin soil, covered with turf) K1, Bronze Age Dirt Mound (Plowed surface) K2, A dirt mound of the early Iron Age (on virgin soil, covered with turf) K3, Stone mound of the early Iron Age or the Middle Ages (with a stone shell) K4, A dirt or stone mound of the early Middle Ages with a "mustache" K5, Burial cult complexes M1, Burial grounds with stone fences of the Middle Ages M2, Fortified settlement of the Bronze Age (with linear or concentric layout) P1, The undefended settlement of the Bronze Age P2. All evaluation tests were carried out using the following hardware platform: an Intel Core i7-based computer with a GPU, training of CNN was carried out for 150 epochs, the size of the training sample was 274 frames, the test sample was 117 frames. Table 1 presents the results of the classification of objects that relate to mounds classes.

Table 1. F1-score (in 10^{-3}) of semantic classification methods for the kurgan classes

Name methodology	K1	K2	K3	K4	K5
ICP + DGCNN	853	786	428	701	588
MVCNN	813	684	401	655	567
ICP + DGCNN + YOLO	868	845	355	812	477
MVCNN + YOLO	827	745	387	847	404
FICP + DGCNN*	902	894	443	877	518
FICP + DGCNN* + YOLO	893	922	432	902	511

Analyzing the data in Table 1, it can be concluded that the proposed methodology based on the fusion of FICP + DGCNN* + YOLO methods allows for the classification of mounds of the Early Iron Age (K2) or Middle Ages (K4) and soil mounds (K1) with high accuracy, while soil or stone mounds of the early Middle Ages with "moustaches" (K5) and Iron Age soil mounds (K3) are classified with low accuracy. Also from the Table 1 it can be seen that the use of a combination of the 3D classification method and the YOLO object detector almost always leads to an increase in the accuracy of solving the problem of 3D classifying archaeological objects. The results of computer modeling confirm the well-known fact that methods of semantic data processing based on CNN have advantages in terms of accuracy over multi-view methods such as MVCNN. On the other hand, it can be seen that using a combination of FICP + DGCNN* + YOLO almost always has accuracy advantages over FICP + DGCNN*, ICP + DGCNN + YOLO and ICP + DGCNN. When classifying a dirt mounds (K3), many errors of the first and second kind occur, in this case the object of interest may be mistakenly correlated with the background in the image or the point cloud. This is due to the presence of noise of various nature in the images, outliers in 2D and 3D data, and most importantly ambiguous signs of decryption of archaeological objects of this class. Table 2 presents the results of the classification of objects that belong to settlements and burial grounds (M1, M2) of various classes. Analyzing the data in the Table 2 it can be concluded that the proposed methodology based on the fusion of FICP + DGCNN* +YOLO methods allows for the classification of fortified (P1) and non-fortified (P2) settlements of the Bronze Age with high accuracy, while burial grounds with stone fences of the Middle Ages (M2) are classified with F1-score no more than 0.748, and funeral cult complexes (M1) are in principle poorly detected, thar it is due to their shape and signs of decryption (see Fig. 5).

Table 2. F1-score (in 10^{-3}) of semantic classification methods for burial grounds and settlements

Name methology	P1	P2	M1	M2
ICP + DGCNN	804	825	388	698
MVCNN	682	626	261	414
ICP + DGCNN + YOLO	834	844	454	704
MVCNN + YOLO	718	696	291	501
FICP + DGCNN*	961	872	577	748
FICP + DGCNN* + YOLO	966	922	569	742

Fig. 5. The results of the classification of the archaeological site in the area of the village of “Levoberezhnoye according to aerial photography

For the settlement and burial grounds classes, it is also true that the accuracy of the classification process of archaeological objects is increased when using the YOLO object detector in the data processing process. The MVCNN method also showed worse accuracy than the DGCNN method and its combinations with ICP, FICP and YOLO, while the worst results were obtained for the M1 class. From the Table 2 it can be seen that using a combination of FICP + DGCNN* + YOLO methods has accuracy advantages in comparison with ICP + DGCNN, ICP + DGCNN + YOLO and FICP + DGCNN* for objects of settlement classes (P1, P2), while for objects from the burial grounds class group, accuracy was obtained similar to the FICP + DGCNN*. The results of computer simulation showed that the proposed DGCNN* architecture has advantages in F1-score compared to the classical version of DGCNN, on average, the F1-score can achieve an increase in accuracy by 0.095 points. The proposed methodology in this paper doesn't have advantages in terms of performance relative to state-of-the-art methodologies (see Table 1 and Table 2).

Conclusions

This work is devoted to the study of methods for solving the task of segmentation and classification of 3D data of archaeological sites of the Bronze Age using deep machine learning methods. The paper proposes a new methodology for solving the 3D classification task based on a combination of the FICP registration method, a modified DGCNN* architecture and the YOLO object detector. The study suggests a new neural network architecture based on DGCNN with EdgeConv convolutional layers, which allows to obtain an accurate solution to the problem of classification and segmentation of 3D data based on a combination of geometric and color features of archaeological objects. The results of computer modeling based on data obtained as a result of the geometric survey of Bronze Age monuments in the area of the Sintashta River in the Chelyabinsk region and remote sensing data of the Earth showed advantages in accuracy of the proposed methodology in comparison with well-known approaches to classification and segmentation of 3D data for various classes of archaeological objects. The point-to-point problem in a closed form for affine transformations without initial initialization (FICP) based on a joint solution of the variational problem was solved using fusion depth data and features. The theoretical results were obtained, which allow us to evaluate the impact of features with using suggested descriptor based on histograms of oriented gradients on the accuracy and convergence of the solution of the variational ICP problem with point-to-point metric. The results obtained were used to solve the problem of constructing a digital mathematical model of an archaeological site.

Acknolwledgments. The work was supported by the Russian Science Foundation, project no. 23-11-20007.

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. Comput. Vision Pattern Recogn. **1**, 886–893 (2005)
2. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision. IEEE, Kerkyra (1999)
3. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
4. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56
5. Manzurv, T., Zeller, J., Serati, S.: Optical correlator based target detection, recognition, classification, and tracking. Appl. Opt. **51**, 4976–4983 (2012)
6. Ouerhani, Y., Jridi, M., Alfalou, A., Brosseau, C.: Optimized preprocessing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter. Opt. Commun. **2013**(289), 33–44 (2013)
7. Kumar, B.V.-K.V., Mahalanobis, A., Juday, R.D.: Correlation Pattern Recognition. Cambridge University Press, Cambridge (2005)
8. Wang, P.S., Sun, C.Y., Liu, Y.: Adaptive O-CNN: a patch-based deep representation of 3D shapes. ACM Trans. Graphics **37**(6), 1–11 (2018)
9. Brock, A., Lim, T., Ritchie, J.M.: Generative and discriminative voxel modeling with convolutional neural networks. http://arxiv.org/abs/1608.04236. Accessed 08 June 2023
10. You, Y., Lou, Y., Qi, L., Tai, Y.W., Wang, W., Ma, L.: PRIN: pointwise rotation-invariant network. http://arxiv.org/abs/1811.09361. Accessed 08 June 2023
11. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of International Conference on Computer Vision (ICCV). IEEE, Santiago (2015)
12. Maturana, D., Scherer, S.: VoxNet: a 3D convolutional neural network for real-time object recognition. In: Proceedings of International Conference on Intelligent Robots and Systems (IROS). IEEE, Hamburg (2015)
13. Li, J., Chen, B.M., Lee, G.H.: SO-net: self-organizing network for point cloud analysis. In: Proceedings Computer Vision and Pattern Recognition. IEEE, Salt Lake City (2018)
14. Lambers, K., Verschoof-van der Vaart, W.V., Bourgeois, Q.P.G.: Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. Remote Sens. **11**(7), 794 (2019)
15. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of Conference on Computer Vision and Pattern Recognition. IEEE, Honolulu (2017)
16. Charles, R.Q., Li, Y., Hao, S., Leonidas, J. G.: PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of 31st Conference on Neural Information Processing Systems (NIPS). NeurIPS Media Kit, Long Beach (2017)
17. Zhang, Y., Rabbat, M.: A graph-CNN for 3D point cloud classification. In: Proceedings of International Conf. on Acoustics, Speech and Signal Processing. IEEE, Calgary (2018)
18. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graphics. **38**(5), 146, 1–12 (2018)

19. Te, G., Hu, W., Guo, Z., Zheng, A., Guo, Z.: RGCNN: regularized graph CNN for point cloud segmentation MM. In: Proceedings of the 26th ACM International Conference on Multimedia. ACM Digital Library, Seoul (2018)
20. Horn, B.K.P.: Closed form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A **4**(4), 629–642 (1987)
21. Vokhmintcev, A.V., Sochenkov, I.V., Kuznetsov, V.V., Tikhonkikh, D.V.: Face recognition based on matching algorithm with recursive calculation of local oriented gradient histogram. Dokl. Math. **466**(3), 453–459 (2016)
22. Vokhmintcev, A., Timchenko, M.: 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 Polytech. Hung. **17**(8), 95–108 (2020)
23. Vokhmintcev, A.V., Melnikov, A.V., Pachganov, S.A.: Simultaneous localization and mapping method in 3D space based on the combined solution of the point-point variation problem ICP for an affine transformation. Inform. Appl. **14**(1), 101–112 (2020)
24. YOLO by Ultralytics 2023. https://github.com/ultralytics/. Accessed 08 June 2023
25. Zdanovich, G.B., Batanina, I.M., Levit, N.V., Batanin, S.A.: Step'-lesostep'. Kizil'skij rajon. Arheologicheskij atlas Chelyabinskoj oblasti 2003. https://search.rsl.ru/ru/record/01002755616?ysclid=lm3sa1zfkp696194850. Accessed 09 Jan 2023

Author Index

D. I. Ignatov et al. (Eds.): AIST 2023, LNCS 14486, pp. 363–364, 2024.
https://doi.org/10.1007/978-3-031-54534-4