PROCEEDINGS OF SCIENCE

Visual clustering of marine sediment particles using a combination of unsupervised machine learning methods

Viktor A. Golikov,[a,*] **Mikhail A. Krinitskiy**[b,a] **and Dmitrii G. Borisov**[b,a]

[a] *Moscow Institute of Physics and Technology,*
Dolgoprudny, Russia

[b] *Shirshov Institute of Oceanology, Russian Academy of Sciences,*
Moscow, Russia

E-mail: golikov.va@phystech.edu, krinitsky@sail.msk.ru

The information on the past climates or environments is preserved in natural archives, such as, for example, marine sediments covering the sea-floor. The study of sediment composition in coarse fraction (>0.063 mm) is widely used, yet time-consuming technique useful for recognizing ancient environments. The coarse fraction analysis is generally performed visually under binocular microscope and requires the high qualification of the observer. In this study, we propose a method to automate and accelerate this kind of work using a combination of classic computer vision and machine learning algorithms. Using an optical digital microscope with precise automatic positioning system, we photographed sieved and dried sediment samples composed of particles over 0.1 mm in size. We then applied a clustering pipeline including classical and neural machine learning techniques. We demonstrate that the proposed method is capable of dividing visual representations of marine sediment grains into homogeneous groups suitable for further accurate classification by an experienced specialist. Our method may significantly reduce the time costs of an expert conducting a study of marine sediments. This will allow further evaluation of sediment composition, main sediment sources and some important characteristics (proxies/indicators) marking a particular environmental setting in the past. The clustering results obtained using our algorithm may be used to train a more accurate classification algorithm.

**** The 6th International Workshop on Deep Learning in Computational Physics (DLCP2022) ****
**** 6-8 July 2022 ****
**** JINR, Dubna, Russia ****

[*] Speaker

1. Introduction

The ocean floor is mantled with a thick sediment cover. Scientists use the material accumulated on the sea-floor during thousands and millions of years as a timeline to study the changes of oceanographic and climatic conditions in the past. Similar records preserved in the sedimentary rock layers on continents are much more incomplete due to severe activity of the wind, water, high-amplitude temperature variations etc. Researchers have a variety of methods for deciphering sedimentary records and reconstruction of the geological past. The study of sediment composition in coarse fraction (particles over 0.063 mm in size) is a quite old but still widely used technique for recognizing ancient environments (e.g., Lankford and Shepard, 1960; [14] Brookfield, 1978 [3]). This approach provides among others information on mineral composition, sediment sources and deep ocean chemistry and circulation (i.e. variations of foraminiferal lysocline and calcite compensation depth). Researchers use specific indicators (proxies) marking a particular environmental setting in the past (Gornitz, V., 2009 [5]): specific mineral associations, the ratio between biogenic and terrigenous paticles, the ratio between tests of planktic foraminifera (single-celled organisms) and their fragments etc.

Before the analysis, sediment samples collected from the sea-floor using various sampling devices (gravity corers, grabs, box- and multi-corers) should be sieved (mesh size 0.063mm or 0.1 mm) and dried. Then the specialist manually observes and classifies particles under binocular microscope and usually cannot consider all the particles in the sample. A small quartered subsample of grains is examined (usually not less than 300 grains). It is obvious though that as the particle amount decreases, the representativeness decreases correspondingly. This work is highly-time consuming and negatively affects the health of a specialist due to high visual load (work with an optical microscope is acknowledged as harmful working conditions). Finally, the work requires a highly qualified specialist, which is always lacking, and it takes a huge amount of time to train new experts.

For all the above reasons, the aim of this paper is to simplify the work of scientists by creating an automatic classification algorithm. Unfortunately, we were unable to solve a supervised machine learning classification problem due to lack of a labeled training dataset. Instead, in this work, we tackle the problem using clustering approach, as it does not require labels. Later, having unnamed clusters containing objects of the same types, an experienced specialist may label them, thus, creating labeled dataset of particales of marine bottom sediments. This dataset may later become the basis for creating a fully automated classification machine learning algorithm.

In this study, we propose the algorithm for processing digital microphotographs of particles of marine sediments, which makes it possible to automate the partition of these particles into semantically homogeneous groups for further study by an expert. Our approach consists of the following steps:

1. we collected the microphotographs of marine sediment particles (> 0.1 mm in size) using contrastive background with 80x automated microscope;

2. we processed the microphotographs of marine sediment particles in order to isolate visual representations of individual grains;

3. we developed and implemented an algorithm based on deep learning techniques for reducing the dimensionality of a feature description of particles with the selection of semantically significant features;

4. we trained the encoder-decoder-type neural network we developed using the dataset of visual representations of individual particles;

5. we applied the encoder of the trained neural network to visual representations of individual particles resulting in hidden representation of these objects;

6. we applied K-means clustering algorithm to the dataset of hidden features of visual representations of marine sediment particles.

1.1 Related works

The main problem we tackle in this study from methodological point of view, is the clustering of visual representations of objects of similar origin. The main issue of processing the visual objects using classic machine learning methods such as K-means clustering is the so-called "dimensionality curse" [2] which is still in action when one tackle the images of marine sedimentary particles due to their size (280x280 px. on average, RGB channels, 6 images per object — see "Data and methods" section). Also, the size of the images of individual objects may differ due to sampling method. In order to tackle these issues, in state-of-the-art methods of machine learning and deep learning, several approaches were developed for the clustering of visual objects. One of the most popular methods is the two-step approach which includes dimensionality reduction with the consequent application of clustering algorithm. The first step allows one to reduce the dimensionality of a feature space of a homogeneous (meaning the sampling nature of the examples) sample taking into account its statistical patterns.

One of the most studied dimensionality reduction methods is principal components analysis (PCA) [15]. In case of simple clustering problems with linearly separable data, PCA may provide meaningful feature space with high quality of corresponding clustering results. However, PCA does not deliver semantically meaningful features. Also, the images of marine sedimentary particles may occur having different spatial sizes which prevents one from exploiting PCA.

Among deep learning approaches for the dimensionality reduction, there are also several methods proposed recently for solving the task of feature learning such as autoencoders [10] or contrastive learning models [6], e.g. MoCo [8]. In this study, we chose to use autoencoders as the ones with most straight-forward training procedure. Autoencoders are also known for the tractable modifications one may apply to restrict the distributions of hidden features, e.g. Sparse Autoencoder [17], Variational Autoencoder (VAE) [13], β-VAE [9], *etc*.

The main contributions of our study are the following:

- we propose the pipeline for extracting the imagery of marine sedimentary particles representing them in several focal distances, using SLIC [1] spatial superpixel-based segmentation;

- we propose the architecture of a variational autoencoder with ResNet-like [7] encoder which we train in order to perform the dimensionality reduction with the constraints imposed on the hidden feature space;

- we propose the pipeline including the previous two methods for clustering the particles resulting in semantically homogeneous groups which may be further either annotated by an experienced marine paleooceanologist or subjected to the similar processing in order to subdivide them.

2. Data and methods

The studied sediment samples were obtained in research cruises organized by the Shirshov Institute of Oceanology (e.g. [11]). Sediment samples collectected using gravity corers and grabs, were sieved (mesh size 0.1 mm), dried and disassembled. The resulting sediment particles look like sand, but contain valuable information in the form of the distribution of types and sizes of grains. We placed the studied particles on a bright contrasting green plate of a Levenhuk DTX 90 microscope and evenly distributed them. The microscope (see fig. 1) is able to take pictures on any part of the plate under 80x magnification with varying focal distances ("focal layers" hereafter). The resulting images are of 1280x1024 px resolution. To take the most volumetric information from the images, we took the photographs in each position of the plate six times varying the focal distance. The camera moves along the substrate and takes pictures automatically. The only action required of the researcher operating the microscope is to change the contents of the substrate as the shooting progresses. This way, a dataset of 58302 photographs with certain amount of marine sedimentary particles is obtained.

In order to cluster the grains, it is necessary to select individual particles in each photograph. For this reason, we applied the traditional machine learning algorithm SLIC (Simple Linear Iterative Clustering) [1]. In this method, a special distance metric is proposed that considers color and image plane space distances. Using SLIC, we divided each image into many homogeneous segments called superpixels, with sufficient accuracy and low computational costs. Some superpixels frame the background, while others frame particles visual representation. At the same time, some particles may be covered by several superpixels. Thus, the problem of extracting the imagery of individual particles is decomposed into two tasks: to separate particle-related superpixels from the background ones, and to enclose the adjoint particle superpixels by a rectangle which is then used to cut the image of a particle from the photograph.

A significant feature of the bright green pixels of the backgroud is their high saturation value in HSV color space. This fact allows one to separate the segments with studied particles of marine sediment from the background segments that are not of research interest. We did it by dropping the segments with the highest pixel average saturation (see fig. 1). We then united the contacting particle-related superpixels, and also eclosed the resulting segments by rectangles using minimum-eclosing bounding box fitting. In order to be able to synchronize the bounding boxes in several focal layers, we used non-rotated eclosing bounding boxes fitting. Since there are six photographs in each position at the plate with varying focal distances, the segmentation described here was performed six times. The resulting bounding boxes for each particle differ slightly between the images taken with different focal distances. We consider the widest bounding box of the six as the one enclosing the particle image. Using these bounding boxes, we cut out the individual particles from all the six focal layers. Thus, the features of objects (marine sediment particles) are the six layers of RGB images with varying spatial sizes.

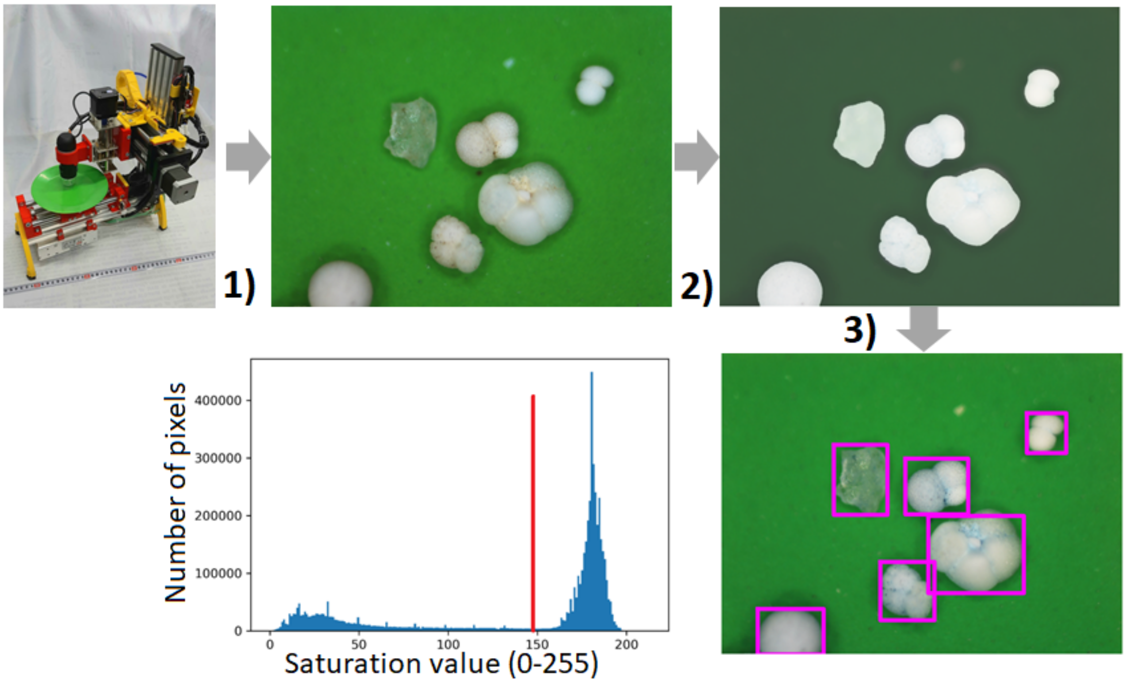

Figure 1: The pipeline of image processing resulting in bounding boxes enclosing individual particles: (1) we capture the microphotographs of a region of the substrate distributed at the green plate using Levenhuk DTX 90 automated microscope; (2) we apply SLIC method resulting in segments of individual particles; (3) we apply bounding boxes fitting crop the particles from the image. In the histogram, we demonstrate the distribution of the saturation values in the image; the red line indicates the threshold which we used for distinguishing between background and particle-related superpixels delivered by SLIC at step (2).

After processing all 58′302 RGB photographs that are 9′717 sets of six focal layers, 28′801 particles were identified. We then measured the average grain sizes in pixels. This is the lengths of bounding boxes along both spatial axes. The average transverse length in our dataset is close to 280 px. For the correct operation of the neural network algorithm, we applied resize transformation to the images of individual particles to make them equally sized. In order to minimize the particle distortion in average, we considered 280x280 px. size as the target during the resize operation. Because of this, some semantic features related to the sizes and proportions of the grains may be irrelevant. However, this resize transformation simplifies further processing significantly.

As a result of the abovementioned procedure involving SLIC algorithm along with the subsequent transformations, we collected a dataset consisting of a series of six RGB images 280x280 px. in size. These images sets are equivalent to vectors in a 1411200-dimensional real-valued feature space. Under the conditions of such a high-dimensional space, clustering algorithms expected to produce the results of poor quality due to the so-called "curse of dimensionality" effect [2]. This effect manifests itself in the exponentially increasing in volume associated with adding extra dimensions to Euclidean space. This effect results in weak coordinate dependency of Euclidean distance measure between a test vector and all other sample vectors. Consequently, this often results in statistical indistinguishability of groups of examples for the algorithms that rely on Euclidean metric similarity measures and similar ones. There are several ways to deal with the curse of dimensionality in clustering problems. One of them is to modify the similarity measure. For example, one may use higher-order Minkowski metrics. However, the choice of the right order of this metric in clustering problems is a difficult task, since clustering itself is an ill-posed problem. Therefore, it is difficult to determine the clustering quality measure in advance. The second way to combat the effects of the curse of dimensionality is to nonlinearly reduce the dimensionality of examples. As we pointed out in Section 1.1, there are both linear methods for dimensionality reduction (e.g., PCA [15] or linear neural autoencoder [23]) and nonlinear ones, including Stacked Autoencoders [10] or contrastive learning models [6], e.g. MoCo [8]. Nonlinear methods is the preferred approach, as they allow one to extract semantically meaningful features. In the original high-dimensional space, semantic features are not expressed explicitly. At the same time, clustering should be carried out guided by high-level features such as color, shape, presence of protrusions, their number and location. These visual features are not expressed at the pixel level in the visual representation of the grains. Therefore, the task is to compress the high-dimensional vector of an object into a low-dimensional one, which is suitable for clustering while preserving and, if possible, highlighting important semantic features.

In our study, we decided to apply variational autoencoder (VAE) [13] at the feature learning step as it is one of the recommended and most commonly used dimensionality reduction methods in clustering problems. VAE also delivers satisfactory results when one applies it to imagery datasets characterized by simple visual content [22]. The content of our images may be considered simple since they have nothing but a solid background and a clustered particle.

In autoencoders, the process of computing the low-dimensional feature vector of an object with the extraction of new features is called encoding; and the process of reconstructing the original examples from the low-dimensional feature vector is called decoding. Auto-encoding is potentially lossy transformation, and the rate of quality loss is strongly dependent on the expressive power of encoder and decoder, input data distribution in original feature space, the width of the hidden representation (i.e., dimensionality of the low-dimensional feature space). In order to compose the most lossless autoencoder, we conducted hyperparameters optimization regarding the architecture of the autoencoder using pixel-wise mean squared error (MSE) metric 1 as a quality measure.

$$MSE = \frac{\sum_{l=1}^{6}\sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{c=1}^{C}\left(x_{ijcl} - \mathcal{D}\left(\mathcal{E}\left(x\right)\right)_{ijcl}\right)^{2}}{6 * W * H * C}, \qquad (1)$$

where x is the feature vector of an object (i.e., the six RGB images of a particle of marine sediment); indices i, j enumerate spatial coordiinates of these images; index c enumerates color channels (either Red, Green or Blue); index l enumerates the focal layer. Here W, H are the width and height of the images of individual particles; C is the number of color channels, which is three for RGB images; $\mathcal{D}, \mathcal{E}$ are the decoder and the encoder, thus, $\mathcal{D}\left(\mathcal{E}\left(x\right)\right)$ is the reconstructed example.

There is an issue of the regular autoencoder (i.e. Stacked Autoencoder [10]) so that it does not have any mechanism to govern the organization of the data in the hidden representation feature space. The only thing a Stacked Autoencoder (SAE) learns is to map the input data vectors into reconstructed feature space as accurately as possible in terms of pixel-wise MSE loss function, regardless of how the data in the hidden representation is adjusted. This can also lead to overfitting, that is, the algorithm will learn perfectly on training images, but the quality of the reconstruction may drop noticeably when processing new out-of-train examples. To alleviate these effects, one may impose some restrictions of either architectural matter or training procedure matter. Thus, it is possible to define a variational autoencoder (VAE, [13]) as an encoder whose learning process is ordered to avoid overfitting and ensure that the hidden representation demonstrates suitable properties, e.g., for clustering.

The architecture of VAE (fig. 2) is similar to the SAE: VAE is characterized by the two neural networks, encoder and decoder. The task of a VAE is to reconstruct the imput examples with minimal pixel-wise MSE error. In contrast with SAE, the encoder maps an input example to the distribution of its hidden representation which is encouraged to be normal using regularization term of the loss function (see eqs. (2),(3)). The decoder reconstructs the example using s low-dimensional vector in hidden representation feature space drawn from the normal distridution characterized by the parameters approximated by the encoder. From the application point of view, the regularization term, namely KL-divergence (3) between the returned distribution and the standard Gaussian, is the term one uses to regulate the properties of the hidden representation feature space. The resulting loss function of VAE is the following:

$$LOSS = MSE(x, d(e(x))) + KL(N(\mu_x, \sigma_x), N(0, 1)), \tag{2}$$

where MSE is the (1) equasion, and the KL divergence between the distribution P to the distribution Q in general is the following:

$$\mathrm{KL}(P \| Q) = \int P(x) \log \frac{dP}{dQ} dx, \tag{3}$$

where the integral is taken over the entire X space of outcomes that P and Q need to have in common.

Hidden representation feature space generated by a VAE is characterized by two properties: completeness and continuity. The latter one means that two vectors that are close in hidden representation feature space will result in semantically similar reconstructed examples being transformed by the decoder. In our study, this property is essential since it allows one to perform the clustering due to placement of semantically similar examples close to each other in compact regions of low-dimensional feature space. The task for the clustering algorithm is to patrition these areas the right way.

2.1 VAE implementation details

Convolutional neural networks are used as an encoder and decoder of the VAE we employed in our study for dimensionality reduction task. For the encoder, ResNet-152 [7] backbone is used which is pre-trained with Imagenet dataset [26]. As we have six images in each set to extract features, the encoder contains six ResNet-152 branches which outputs are contatenated. The dimension of the hidden representation feature space in our study is 512.

The adaptive momentum estimation optimizer (Adam [12]) with weight normalization and initial learning rate of 10^{-4} is used to minimize the loss function.

Using the trained encoder, we extracted low-dimensional features from the images visually representing the individual marine sediment particles. Encoder maps the examples into the mean

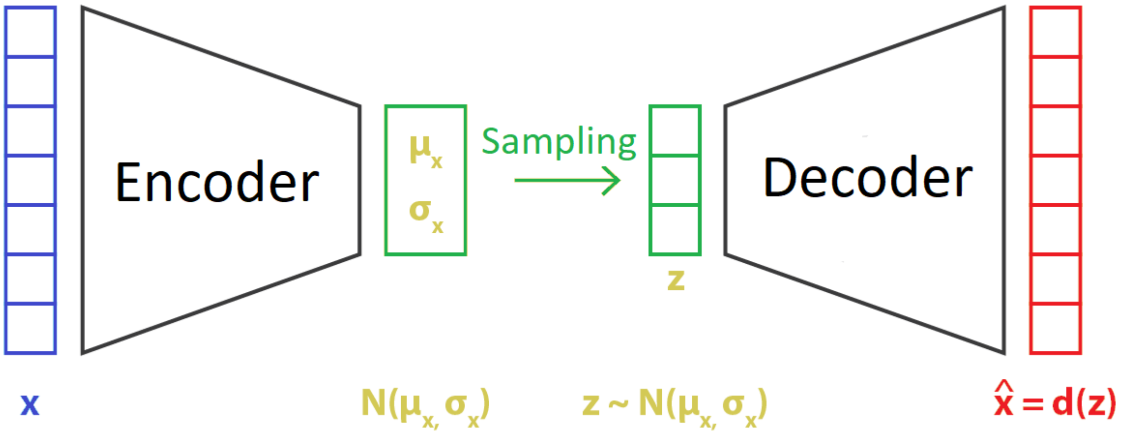

Figure 2: Archictecture of variational autoencoder used in our study for dimensionality reduction. Here, x is visual representation (six focal layers of RGB images) of an input example; μ_x and σ_x are the parameters computed by the Encoder; z is the random vector in hidden representation feature space drawn from the normal distributiion parameterized by μ_x and σ_x; $\hat{x}$ is the reconstructed example.

(μ_x in fig. 2) and variance (σ_x^2 in fig. 2) parameters of the normal distribution. Following the best practices [27], we employ the μ_x parameter as the hidden representation of the examples.

2.2 Clustering approach

Using the hidden representation of the examples, we applied K-means clustering [16] algorithm. K-means is relatively simple, and yet it is a widely used choice delivering suitable results in case of vectors distributed convexly and isotropically. K-means hypeparameter is K number of clusters into which the sample is divided.

In our study, we intentionally applied VAE as the dimensionality reduction method that is characterized by the desired property of hidden representation feature space, i.e., its continuity. At the same time, VAE is characterized by the downside which is normal distribution of the feature vectors of the dataset mapped to this space. Thus, the clear structure of the dataset is lost in this space, so as the capability of clear identification of the clusters. To overcome this issue, we chose to use the so-called overclustering approach, i.e., we divided the dataset into large ($K = 100$) number of clusters. This method is also applied for data clustering with fuzzy labels [24], which is what our data essentially is. The number of clusters $K = 100$ is chosen following the expert suggestion of expected particle groups (10), with the intention to partition the hidden representation feature space into finely defined clusters containing semantically homogeneous sets of examples.

2.3 Implemetation details

In our study, we exploited Python v.3.7 programming language [28]. For implementing our clustering approach, we used scikit-learn package [21]. For implementing the dimensionality reduction with the VAE model, we used Pytorch library [29]. For implementing the classic computer vision operations, we exploited OpenCV library [30]. The computations were performed on

NVIDIA DGX Station equipped with 256GB RAM using NVIDIA Tesla V100 graphics processing unit with 32GB video memory.

3. Results

We applied the proposed method for the clustering of marine sediment particles in a form of their visual representations taken at six fical distances as described in Section 2 "Data and Methods" to the dataset we collected and processed as described in the same section. The procedure results in clusters labels from 0 to 99 for each one of the particles. In fig. 3, we present the examples of four different clusters in order to demonstrate their inter-cluster semantic differences and also intra-cluster semantic homogeneity.

An experienced paleooceanologist then examined the groups of images fallen into different clusters. According to the examination results, our method is capable of partitioning the dataset into semantically meaningful clusters. Some groups of images are almost entirely composed of particles of terrigenous origin (i.e. derived from the continent). Furthermore, they can also be clearly divided into transparent grains, most likely quartz, and dark-colored ones that is subject to further study (probably, hornblende). Other clusters mostly consist of biogenic particles (product made by or of life forms). These are mainly represented by foraminiferal tests, which, in turn, are clearly divided into two groups: broken tests and intact ones.

In addition, few clusters were assessed as containing "chopped" particles. These are the grains which occured on the edge of the field of view during the photography, thus, they did not fit completely into the image. Such particles were recognized by our algorithm as being semantically different significantly from the rest ones. Therefore, they can be clearly separated from the rest so that they do not interfere with the calculation of paleoindicators. The same situation occurred with particles that are too close to each other on the substrate and, because of this, counted by our algorithm as a single object. Such examples were grouped into separate clusters, presumably, due to their placement at some distance from the others in the hidden representation feature space, thus, forming their own clusters.

In figure 4, we present the visualization of the distribution of dataset particles in the hidden space projected into two-dimensional plane. The projection is made using UMAP method (Uniform Manifold Approximation and Projection for Dimension Reduction, [18]) for the visualization purpose only. UMAP is characterized by the property of the mapping preserving the spatial relations between the data points. Thus, one may use UMAP to visualize the relations between objects and groups of a dataset. In figure 4, one may see two large groups of objects in the center of the diagram. The group in the lower right region of the central conglomeration contains mostly the clusters with quartz grains. The other one includes foraminiferal tests. In the far right region, one may see a small collection of clusters including dark-colored particles. Particles that are not completely captured in the frame ("chopped" ones), as well as the objects consisting of several particles stuck together, are located at the edges of the diagram. This means their distant placement in the hidden representation feature space generated by the projection of the encoder. This means that such clusters will be easy to identify and exclude from consideration.

Taking into account the examination of the expert, and also the visual similarity of the images of particles in separate clusters (see fig. 3), one may consider the proposed approch for clustering the marine sediment particles successful.

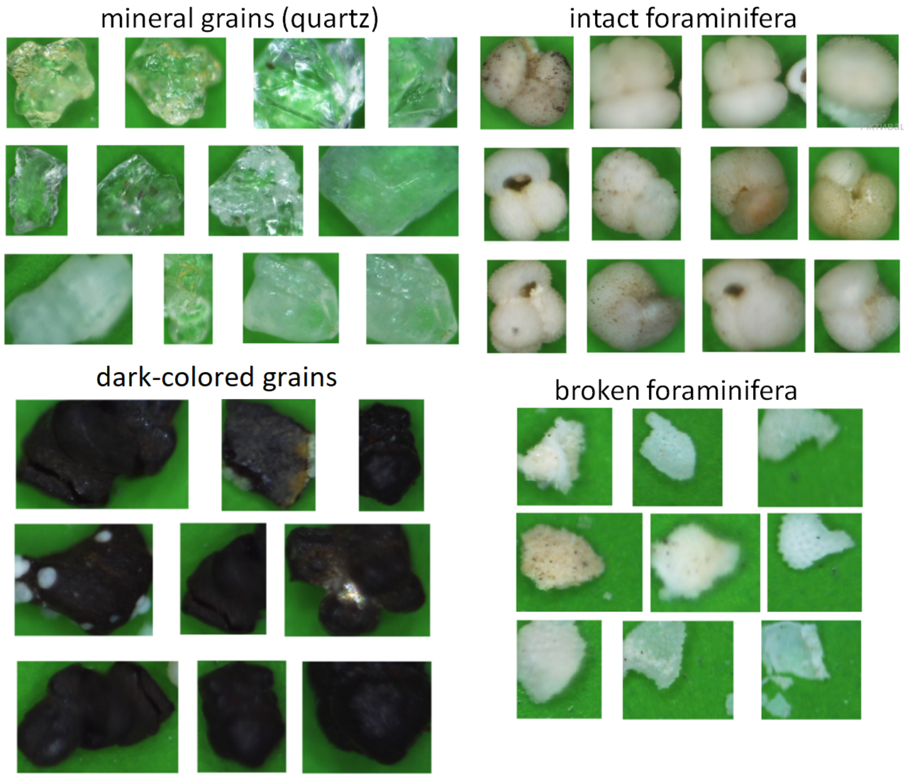

Figure 3: Here, we present the examples of particles from the most successfully clustered groups. Inorganic particles are quartz grains and undefined dark-colored grains. Particles of biogenic origin are foraminiferal test and their fragments

3.1 Discussion

There are several issues we faced in this work. First, there is an issue of "chopped" particles along with the particles that were placed too close to each other at the photography stage. These issues result in reduced quality of the segmentation of individual grains. In case of the latter issue, multiple particles are represented in the dataset as one, thus, they should not be considered at the stage of estimating the paleoindicators. This issue may introduce bias or increasing uncertainty into the estimated values. In further study, we plan to distribute the particles with lower spatial density at the plate in order to decrease the number of adjacent grains. This manner of grains spatial ditribution will also decrease the number of particles occured close to margins of the microscope field of view, thus, decrease the number of "chopped" grains.

Within the two-step clustering paradigm, both stages may be optimized. First, one needs to note that VAE lacks object-level semantics due to its pixel-wise reconstruction loss that tends to over-emphasize low-level semantics. For example, MoCo [8] representation learning algorithm may deliver better quality as it implements data augmentation which may significantly improve the performance in our study. Next, one may also notice that in our method, the traditional K-means algorithm is used, which have to be recalculated every time new examples are obtained.

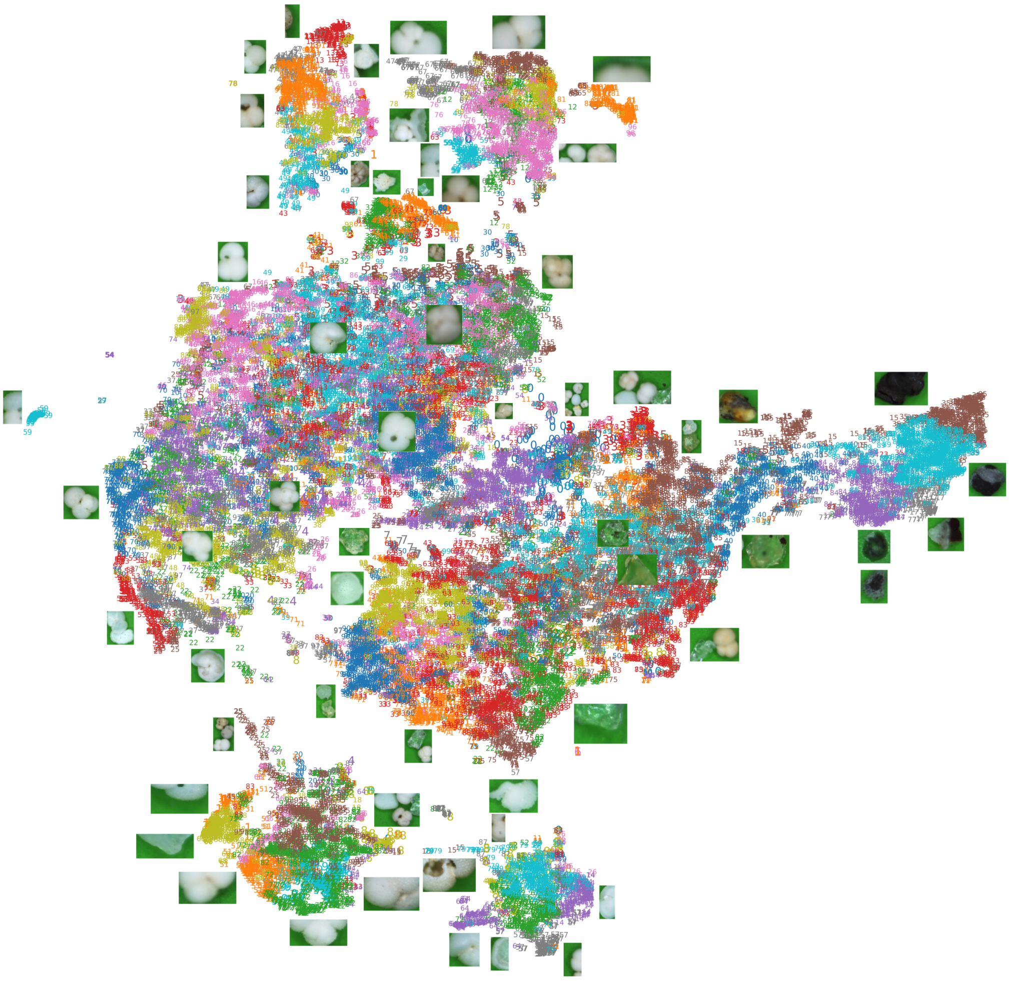

Figure 4: UMAP [18] visualization of sediment sample images in hidden representation feature space generated by the VAE employed in our study.

In our method, there is no obvious way to apply a trained K-means model. Moreover, although overclustering helps to highlight homogeneous clusters. Lots of ’marginal’ clusters contain a sufficient number of particles of a various kind. SPICE [19] algorithm proposes method to re-train neural clustering sub-network using most confident samples.

4. Conclusions and outlook

In this work, we presented an innovative approach for processing digital photographs of marine sediment particles, which makes it possible to automate the division of these particles into semantically homogeneous groups. For this, an algorithm for processing microphotographs was developed based on the approaches of classical computer vision, which makes it possible to isolate individual grains. Using this algorithm, we compiled the database of feature vectors of all sediment particles of the studied sample, based on digital photography data taken with several focus distances by a 80x automated microscope. We also developed and implemented the algorithm based on artificial neural networks, namely convolutional variational autoencoder, to reduce the dimensionality of the feature space of grains. We optimized the neural network based on the compiled database of microphotographs of individual particles. Using the coding part of the optimized neural network, we calculated semantically meaningful features for each particle of the sample. Using these hidden representations of the reduced dimensionality, we clustered the particles of the sediment of the studied sample into 100 clusters.

The results of clustering show that the proposed approach makes it possible to divide the sediment particles of the studied sample into semantically homogeneous groups. Some of them obviously represent particles of terrigenous origin; other clusters contain particles of biogenic origin, whole foraminiferal tests; still others contain fragments of such tests. However, some of the clusters contains particles of different types. This issue may be addressed with further application of the proposed clustering approach to the subsample containing the examples of these mixed clusters.

Despite the large number of groups, one may note the importance of the result: the exammination and classification of these groups by an experienced expert allows one to quickly classify all the grains of the sample with significantly less time spent compared to conventional research methods that result in the same classification. The classification obtained this way makes it possible to automatically calculate important characteristics (proxies, paleoindicators), such as the ratio between calcareous particles and siliclastic (terrigenous) grains, as well as the ratio between the number of foraminiferal test and their fragments.

The results of our study suggest the promising potential of further development of the proposed approach for semantic grouping of marine sediment particles. Based on visual representation of the particles, one may estimate their linear dimensions and shape parameters, which will make it possible to assess the distribution of these characteristics for each individual group.

As a part of the task of facilitating the work of marine geologists, sedimentologists and paleoceanographers, it is also important to be able to develop to a fully automated classification algorithm based on the labeling of a clustered sample made by a specialist. We are going to deploy the algorithms of our study as a website or mobile app.

Aknowledgements The study was supported by the Russian Science Foundation grant 22-27-00421. The development of the VAE and its training is supported by the program FMWE-2022-0002.

References

[1] R. Achanta, A. Shaji, K. Smith et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods — 2012. — Vol. 34, no. 11. — Pp. 2274–2282.

[2] Bellman Richard. Dynamic programming // Science. — 1966. — Vol. 153, no. 3731. — Pp. 34–37.

[3] Brookfield M.E. The use of coarse fraction analysis in paleoenvironmental interpretation: Upper Oxfordian and lower Kimmeridgian (Jurassic) sediments in Dorset, England // Sedimentary Geology. — 1978. — jan. — Vol. 20. — Pp. 249–280. — URL: https://linkinghub.elsevier.com/retrieve/pii/003707387890057X.

[4] Gansbeke Wouter Van, Vandenhende Simon, Georgoulis Stamatios et al. SCAN: Learning to Classify Images without Labels. — 2020.

[5] Gornitz Vivien. Paleoclimate Proxies, An Introduction // Encyclopedia of Paleoclimatology and Ancient Environments. — Dordrecht: Springer Netherlands. — Pp. 716–721. — URL: http://link.springer.com/10.1007/978-1-4020-4411-3$_1$71.

[6] Hadsell R., Chopra S., LeCun Y. Dimensionality Reduction by Learning an Invariant Mapping // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). — Vol. 2. — 2006. — Pp. 1735–1742.

[7] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 770-778 (2016)

[8] He Kaiming, Fan Haoqi, Wu Yuxin et al. Momentum Contrast for Unsupervised Visual Representation Learning. — 2019. — URL: https://arxiv.org/abs/1911.05722.

[9] beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework / Irina Higgins, Loic Matthey, Arka Pal et al. // International Conference on Learning Representations. — 2017.

[10] Hinton G. E., Salakhutdinov R. R. Reducing the Dimensionality of Data with Neural Networks. — 2006. — Vol. 313, no. 5786. — Pp. 504–507. — URL: https://science.sciencemag.org/content/313/5786/504 (online; accessed: 2019-06-18).

[11] E. V. Ivanova, I. O. Murdmaa, D. G. Borisov et al. Study of the Contourite Systems of the West Atlantic in the 50th Cruise of Research Vessel Akademik Ioffe // Oceanology. — 2016. — Vol. 56, no. 6. — Pp. 888–889.

[12] Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. — 2014.

[13] Kingma Diederik P, Welling Max. Auto-Encoding Variational Bayes. — 2014.

[14] Lankford Robert R., Shepard Francis P. Facies Interpretations in Mississippi Delta Borings // The Journal of Geology. — 1960. — Vol. 68, no. 4. — Pp. 408–426.

[15] Maćkiewicz Andrzej, Ratajczak Waldemar. Principal components analysis (PCA). — 1993. — Vol. 19, no. 3. — Pp. 303–342. — URL: https://www.sciencedirect.com/science/article/pii/009830049390090R (online; accessed: 2021-06-07).

[16] Macqueen J. Some methods for classification and analysis of multivariate observations // In 5-th Berkeley Symposium on Mathematical Statistics and Probability. — 1967. — Pp. 281–297.

[17] Makhzani Alireza, Frey Brendan. K-sparse autoencoders // arXiv preprint arXiv:1312.5663. — 2013.

[18] McInnes Leland, Healy John, Melville James. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. — 2020.

[19] Niu Chuang, Shan Hongming, Wang Ge. SPICE: Semantic Pseudo-labeling for Image Clustering. — 2022.

[20] O'Shea, K. & Nash, R. An introduction to convolutional neural networks. *ArXiv Preprint ArXiv:1511.08458*. (2015)

[21] F. Pedregosa, G. Varoquaux, A. Gramfort et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. — 2011. — Vol. 12. — Pp. 2825–2830.

[22] Peihao Huang, Yan Huang, Wei Wang, Liang Wang Deep Embedding Network for Clustering // 2014 22nd International Conference on Pattern Recognition. — 2014. — Pp. 1532–1537.

[23] Plaut Elad. From principal subspaces to principal components with linear autoencoders // arXiv preprint arXiv:1804.10253. — 2018.

[24] Lars Schmarje, Johannes Brünger, Monty Santarossa et al. Fuzzy Overclustering: Semi-Supervised Classification of Fuzzy Labels with Overclustering and Inverse Cross-Entropy // Sensors. — 2021. — Oct. — Vol. 21, no. 19. — P. 6661. — URL: http://dx.doi.org/10.3390/s21196661.

[25] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C. A survey on deep transfer learning. *International Conference On Artificial Neural Networks*. pp. 270-279 (2018)

[26] Jia Deng, Wei Dong, Richard Socher et al. Imagenet: A large-scale hierarchical image database // 2009 IEEE conference on computer vision and pattern recognition / Ieee. — 2009. — Pp. 248–255.

[27] Doersch Carl. Tutorial on variational autoencoders // arXiv preprint arXiv:1606.05908. — 2016.

[28] Van Rossum Guido, Drake Fred L. Python 3 Reference Manual. — Scotts Valley, CA: CreateSpace, 2009.

[29] Adam Paszke, Sam Gross, Francisco Massa et al. / PyTorch: An Imperative Style, High-Performance Deep Learning Library // Advances in Neural Information Processing Systems 32. — Curran Associates, Inc., 2019. — Pp. 8024–8035.

[30] Bradski G. The OpenCV Library // Dr. Dobb's Journal of Software Tools. — 2000.