UDC: 004.032.26:902 https://doi.org/10.15407/jai2024.03.066

**O. Melnykov[1], A. Kapeleshchuk[2]**
[1,2]Donbas State Engineering Academy, Ukraine
72, Akademichna st., Kramatorsk, 84313
[1]aymelnikov1973@gmail.com;
[2]kapeleshchuk.k@gmail.com
[1] https://orcid.org/0000-0003-2701-8051
[2] https://orcid.org/0009-0003-7346-0235

# ESTIMATING THE PARAMETERS OF TRYPILLIA PROTO-CITIES USING PLANS AND MAGNETIC SURVEYS WITH THE HELP OF NEURAL NETWORKS

**Abstract.** The Tryplian culture was considered the largest of all other archaeological cultures that existed on the territory of modern Ukraine. The method for analyzing magnetic images has been described, which allows archaeologists to assess the scale of settlements without excavating them. It is noted that one of the tasks during the analysis of the settlement is to find out its characteristics: counting the number of buildings, calculating the area, etc. However, in the majority of cases, counting the number of structures is currently unfeasible, as Trypillian proto-cities are situated within the cultural layer relatively close to the surface, and any economic activities disturb this cultural layer. The capabilities of the existing system (application) are described, which solved the problem by the method of average values and had differences from the commonly accepted method. It was concluded that a more automated version of this system could be an option where the number of sites in the image will be calculated by the average number of pixels per site, that is, the number of black and gray pixels in the image, divided by the average number of pixels in the site.

It was decided to use neural network models. As an example, the largest of the famous Trypil proto-cities in Ukraine - Talyanka with an area of 450 hectares - is considered. Pictures taken between 1971 and 1974 were used because they are in the public domain. A list of actions for image preprocessing is described. A decision was made to train the model to input square images of 15 by 15 pixels, for this the entire image was divided into 748 square images, the number of buildings in each of them was determined manually. A four-point work algorithm is formulated, and it is also presented in the form of a UML class diagram and an activity diagram. The algorithm for creating a training sample for the second neural network from four points is also formulated (for the case when the zone of squares with lost information will run along the entire length of the picture, as it happens in the photo of Talyanok), presented in the form of a UML activity diagram. The neural network will accept 10 values as input, and will output one - the number of sites in the square, information about which is lost.

Tensorflow and keras frameworks were used to create all models. The most successful model has almost 2.5 million parameters, the model requires 9.36 MB of RAM. During the tests, it was found that increasing the number of convolution layers does not increase the result.

For testing the first model, a picture of the settlement of Maidanetske was submitted, for training a picture of the settlement of Talyanka was used. For the training set, the recognition accuracy was 92%, for testing - 86.5%. The neural network, which implements the algorithm for predicting the number of buildings on lost plots, provides an accuracy of 58% for the Talyanka settlement and 42% for the Maidanetske settlement. This accuracy is much better than a random guess, the probability of which is just over 6%.

**Keywords:** trypillian culture, CNN, keras, UML.

## Introduction

The Trypillia culture is an archaeological culture that existed in the left-bank Ukraine, central Moldova, and Romania between the 6th and 3rd millennia BCE (according to the currently accepted dating) and is the largest archaeological culture that existed on the territory of modern Ukraine. The only source of information about the archaeological culture is archaeological findings, dwellings, burials, etc. Surprisingly, the main archaeological source of the Trypillia culture is Trypillia settlements. Currently, more than a thousand Trypillia settlements of different sizes have been discovered in Ukraine.

From the second half of the 5th millennium BCE, the era of Trypillia proto-cities-giants began, during which settlements with hundreds of buildings began to appear, while previously Trypillia settlements consisted of only three to four dozen buildings. In the first half of the 3rd millennium BCE, the era of proto-cities-giants will come to an end. Currently, archaeologists are aware of 5 proto-cities with an area of more than 200 hectares, 15 with an area exceeding 100 hectares, 20 with an area exceeding 50 hectares, and another 100 with an area of over 10 hectares. Of course, further research on the Trypillia culture may lead to the discovery of new proto-cities-giants. Here it should be noted that excavating two or three buildings per season is considered a fairly fast pace of excavation, so excavating a proto-city with hundreds of buildings is a task that cannot be solved.

Therefore, archaeologists resort to methods that allow them to estimate the scale of settlements without excavating them. One such method is magnetic surveying. Magnetic surveys record spatial changes in the Earth's magnetic field. In archaeology (both terrestrial and marine), magnetic surveying is used to detect and map archaeological artifacts and sites [1].

One of the tasks during the analysis of a settlement is to determine its characteristics: counting the number of buildings, calculating the area, etc. According to current archaeological data, Trypillia proto-cities consisted of two-story buildings made of adobe and wood. Most of the buildings were residential, but there were also buildings for economic and cult purposes. Such cities were built quite quickly and disappeared quickly. On average, a Trypillia proto-city existed for about 70 years (some current research suggests terms of 300-350 years for some proto-cities), after which it was deliberately burned down. After destruction, each building looks like a pile of clay and fragments of ceramic pottery, which can be seen on the processed magnetometer image as black spots. Geodesists call such spots anomalies, while archaeologists call them sites. The latter term will be encountered in the text further.

For a better understanding of the Trypillians' economy, we need data on the population that lived in the proto-cities - this will allow us to estimate the volume of production that each proto-city had to produce for its existence and the area of cultivated land. In addition, this is simply useful for a more detailed understanding of the Trypillian world.

Thus, to determine the approximate population, we need to calculate the number of buildings and multiply this number by the number of inhabitants in each of them. It is difficult, if not impossible, to say how many people could live in one building at the moment. Archaeologists, for equal counting, speak of 10 people in one house. Such a value is not devoid of sense, as if we assume that several generations of a family lived in one house (which is most likely the case), then by counting its members, we can easily obtain a value of more than 10 people per building. 100-150 years ago, a family with 8-10 children was not a rare occurrence. It should also be noted that in most excavated houses, there was one hearth, indicating that one house belonged to one family [2].

Trypillia proto-cities lie in the cultural layer quite close to the surface (0.4-0.5 meters), so any agricultural activity (plowing the land, building roads) destroys the cultural layer, which is why some large proto-cities have lost up to 40% of their area. Therefore, it is impossible to count the number of buildings, but by observing the logic of the layout, we can roughly estimate the scale of the construction. Here, researchers of the Trypillia culture were lucky: the Trypillians built their proto-cities not chaotically, but following a clear logic of construction. Most large Trypillia proto-cities consisted of several concentric circles or ellipses.

Despite more than 100 years of experience in studying the Trypillia culture, it still requires many more research.

Counting thousands of spots on the image is a tedious and time-consuming task, but at the moment we have only a small number of magnetic scanning images, and some of them were made almost 50 years ago. Due to the limited amount of information available, specialists currently do not have a pressing need for the development of an automatic system that would count the sites. Moreover, the analysis of magnetic scanning images involves not only counting the sites but also searching for interesting areas for archaeological excavations. Therefore, the task of analyzing Trypillia proto-cities has more scientific than practical interest.

**Analysis of recent research and publications**

Several years ago, the authors proposed and developed [3, 4] a system that solved the problem using the method of average values, but differed from the commonly accepted method only in that it suggested calculating not one average value for the entire settlement, but dividing the image into small squares, then dividing the squares into groups based on the average value, and finally counting the number of such squares in each group.

It should be noted that in situations where solving the problem does not require high accuracy, using the method of average values is a quite successful solution.

A more automated version of the above-proposed system can be a variant where the number of plots on the image will be calculated based on the average number of pixels per plot. That is, the number of black and gray pixels on the image, divided by the average number of pixels in the plot.

Also, any ready-made neural network models, such as Faster R-CNN, Alex-Net, and others, can be used for analysis.

**Presentation of the main material**

To describe the algorithm of actions that the system needs to perform, we propose to take one settlement and examine the sequence of actions on it. The analysis of other settlements should be carried out according to the same algorithm. For example, let's take the largest known Trypillia proto-city in Ukraine - Talianky, with an area of 450 hectares (Fig. 1). The proto-city has been partially preserved and has several damaged areas, partly because of this, it was chosen as an example.

"Let's note that from here on out, all tests will use images taken between 1971 and 1974, while all other data consider research results only up to 2010. This selectivity is due to the fact that from 2011 to 2016, a repeat magnetic scan of parts of the settlements was conducted with more precise modern equipment, but the original files of this study could not be found in open access.

At the first stage, we will prepare our image: remove all inscriptions and rotate it so that the areas outside the proto-city boundaries are minimal. We also remove the background (shown in gray in Fig. 2) so that the algorithm can distinguish between undeveloped territories and territories where information has been lost.

Next, it was decided to feed 15x15 pixel square images into the model for training. For this, the entire image was divided into 748 square images, and the number of buildings in each was manually determined. These data became the training set for the future model.

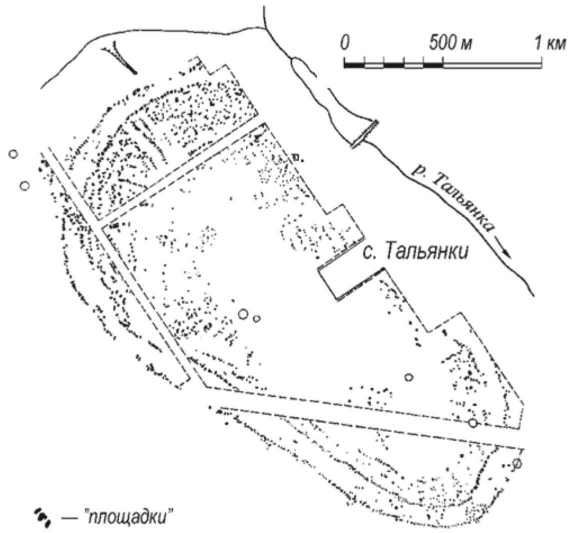


Fig. 1. Snapshot of the Trypillian figurine Talianka

Fig. 2. Processed image of the Talianki settlement

Splitting the image into parts is necessary because any model has a fixed number of inputs, and the size of the images for each protocity is different due to their varying shapes. The size of the squares was chosen as 15x15 after considering the option of 30x30 pixels. However, with the larger size, more than 20 buildings would fall into one square in densely populated areas, increasing the number of classification categories. For the 15x15 size, the maximum number of buildings in one square is 15, resulting in 16 classification categories (from one to 15 buildings + unpopulated territory, i.e., 0). Another problem with using larger squares is that a significant portion of the squares would include parts of lost zones, leading to less accurate estimation of the number of settlements (as lost zones would be considered as white pixels, i.e., unpopulated territory). In the final version, only squares with at least 80% information retained were included. However, image splitting has a negative aspect: some border cuts fall on building sites (ruins), splitting them into two halves and potentially being counted as two separate buildings instead of one. But this error is quite acceptable, considering that it can be partially compensated for by training the model based on the fact that in most cases, the plot will be divided into a larger and smaller part, and the probability of splitting into two equal parts is low.

A much bigger problem for estimating protocity parameters is that a significant portion of the protocity is lost, so we need to find a way to estimate the number of lost buildings. By examining the protocity plan, we can understand the logic of the development, so one way to estimate the probable number of buildings on lost areas would be to use a neural network, with data from intact areas of the same protocity used for training. However, since each protocity has its own development logic, this network would need to be trained separately for each protocity.

In summary, the system's algorithm can be described as follows:

1. The preprocessed image is fed into an algorithm that divides it into separate squares and presents the data as 15x15 two-dimensional arrays to the first (convolutional) network.

2. After the number of buildings in each square is calculated, a two-dimensional array is created, with dimensions corresponding to the number of squares the input file was divided into, in width and height. Squares with lost information are not fed into the building count network, and instead, a corresponding text label is recorded in the two-dimensional array.

3. For each square, a 3x3 mask is applied along the direction from the edge of the damaged area, ensuring that at least four corner squares are intact or already calculated. The central cell of the 3x3 field is then calculated.

4. Once the building count for all squares is calculated, they are summed, and the resulting value is considered the number of buildings in the settlement. Knowing the number of buildings, it is easy to calculate density and population.

The described algorithm is presented in the form of UML class diagram (Figure 3) and activity diagram (Figure 4). The Images class is responsible for image division, the CNN class feeds the results to the first (convolutional) neural network, the Learn class is responsible for training the perceptron for calculating lost areas, and the Dans class is responsible for calculating lost areas and presenting the results.

This algorithm may not work when the zone of squares with lost information spans the entire length of the image, as in the case of the Talianky image (Figure 2). In such cases, there will be no more than three out of the required four cells for the data recovery network for the outermost squares. Therefore, in such cases, it is proposed to consider all lost zone squares on the image border as zero. This assumption will not affect the accuracy of the count for most settlements because these squares usually correspond to the outskirts of the protocity, where there are few buildings.

To handle cases like the Talianky image, an additional step can be added to the algorithm:

If the zone of squares with lost information spans the entire length of the image, consider all lost zone squares on the image border as zero.

By incorporating this additional step, the algorithm can better handle situations where the lost information zone affects the entire image border, improving the overall accuracy of the building count and subsequent calculations for settlements.

In conclusion, the proposed algorithm,

with its additional step to handle edge cases, offers a robust approach to estimating the number of buildings, population density, and population for ancient settlements using satellite images and neural networks. The UML class diagrams and activity diagrams provide a clear visual representation of the algorithm's structure and flow, facilitating its implementation and further development.

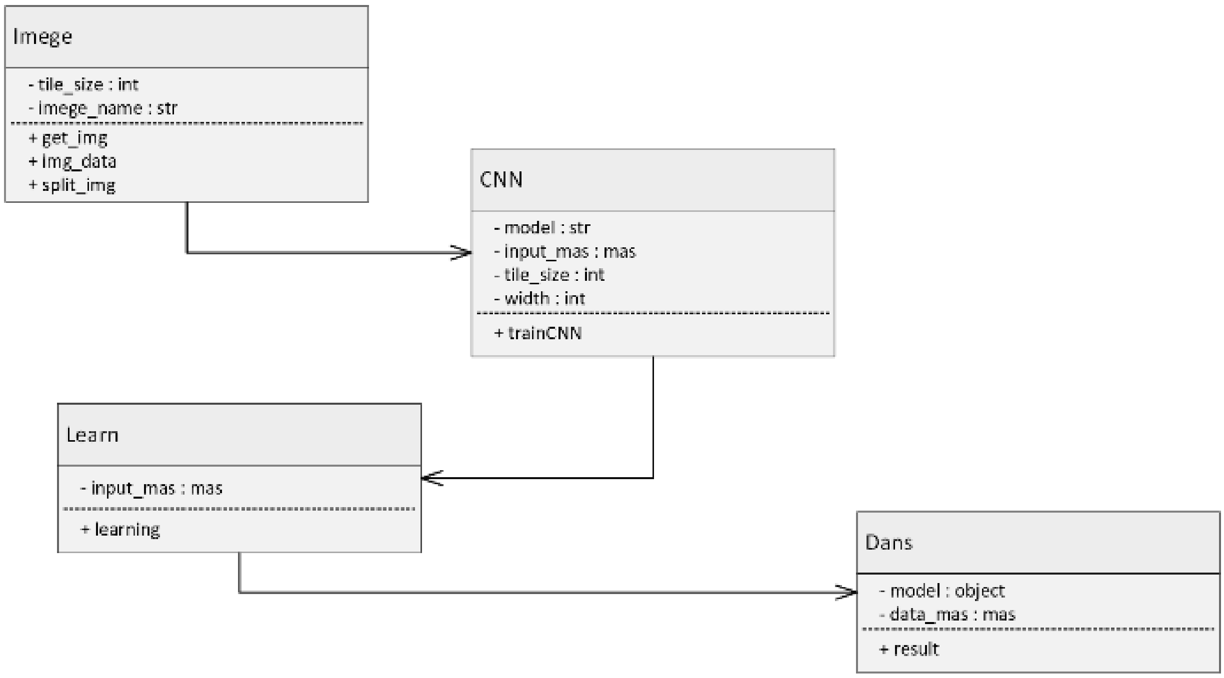


Fig. 3. Class diagram

Algorithm for creating a training set for the second neural network can look like this:

1. By placing a 3x3 square, we check if there is a value in the central cell. If there is, we continue, otherwise we shift the 3x3 square by one value along one of the axes and repeat the check.

2. We generate a set of 8 lists (the number of lists can be different, but the author believes that more than 8 lists is excessive), each consisting of numbers from 0 to 7 randomly chosen so that no number repeats within the list.

3. We check if all squares with these numbers have values that differ from a special marker indicating the absence of a value.

4. We create a training array consisting of: four values of squares with numbers specified in the array of random values, four values of square addresses, and coordinates of the sought square, the value of which goes to the output set of the network. We normalize all input data, convert the output value to a vector of 16 categorical values.

Thus, we need to develop a neural network architecture that takes 10 input values and outputs one - the number of squares in the square, information about which has been lost. The algorithm is described in more detail in the activity diagram in Figure 6.

Creating a system based on a neural network begins with the development of an application for creating this network. Various variations of perceptrons and convolutional neural networks were tested as the neural network model. The TensorFlow and Keras frameworks were used to create all models. The most successful model has almost 2.5 million parameters and the following structure (Table 1). The model requires 9.36 MB of RAM to operate. During testing, it was found that increasing the number of convolutional layers did not improve the results. The learning curve is shown in Figure 7.

To test the model, an image of the settlement of Maidanetske was used, while the image of the settlement of Talianky was used for training. In Maidanetske, the model was able to recognize 1363 buildings, according to V.P. Dudkin's calculations, there were 1575 buildings in the preserved part of the settlement [5, p. 128]. Thus, the recognition accuracy is 86.5%. It should be noted that in open sources, only low-resolution images were found, which reduced the recognition quality. For the Talianky settlement, where the model was trained, the accuracy was 92%, according to the author's calculations, there are 1149 buildings in the preserved part of the Talianky proto-city, while the model recognized 1054.

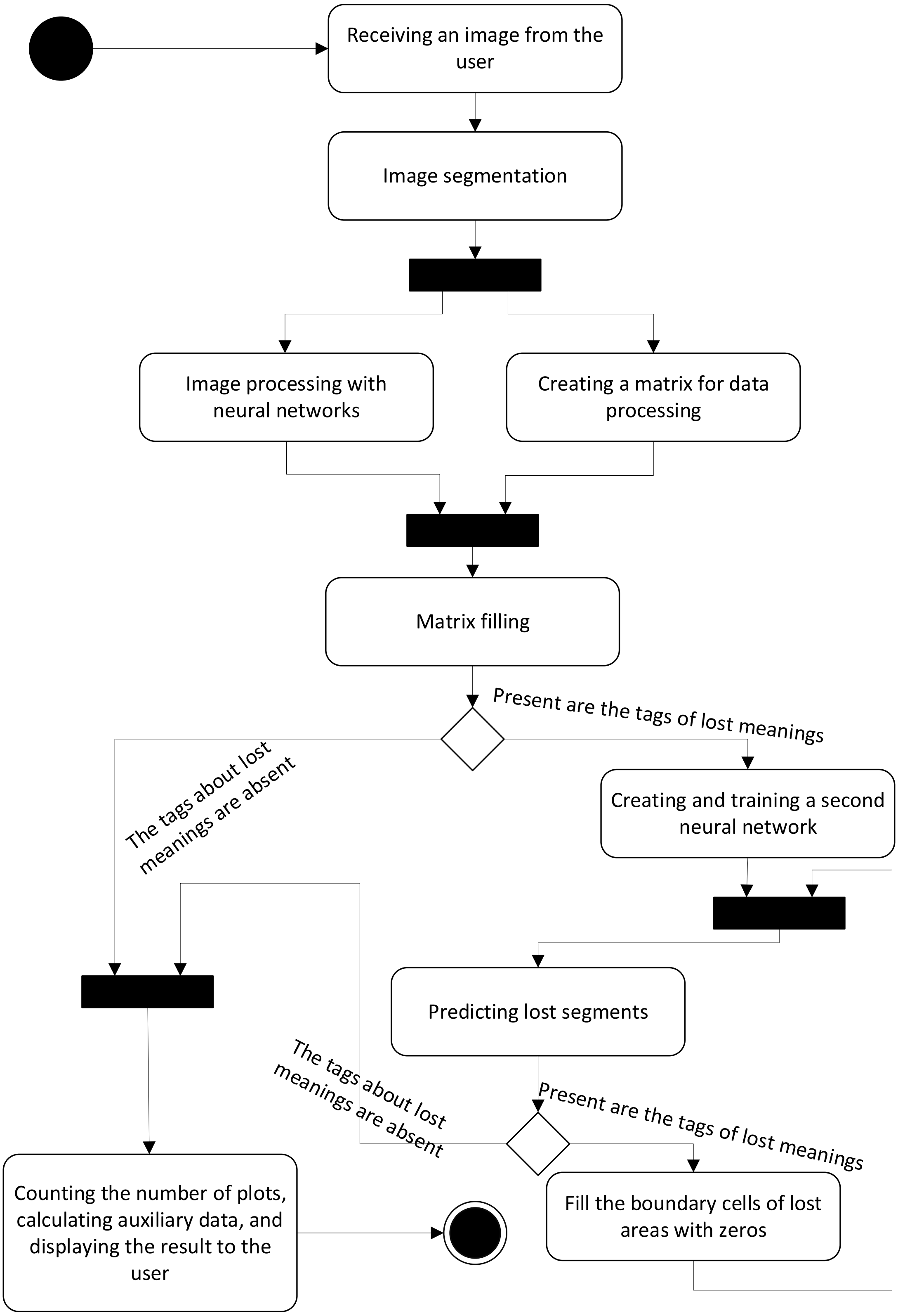


Fig. 4. Activity diagram

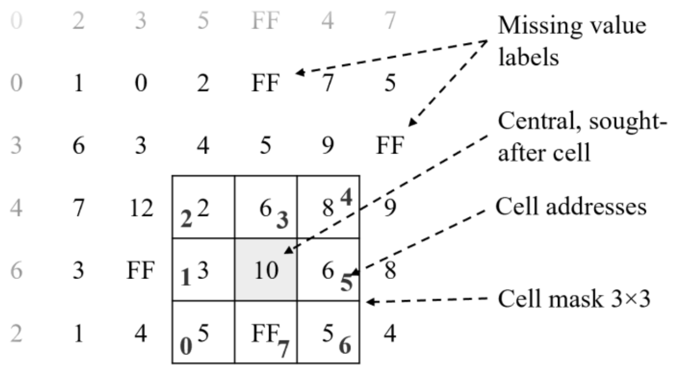


Fig. 5. Schematic representation of a portion of the data array

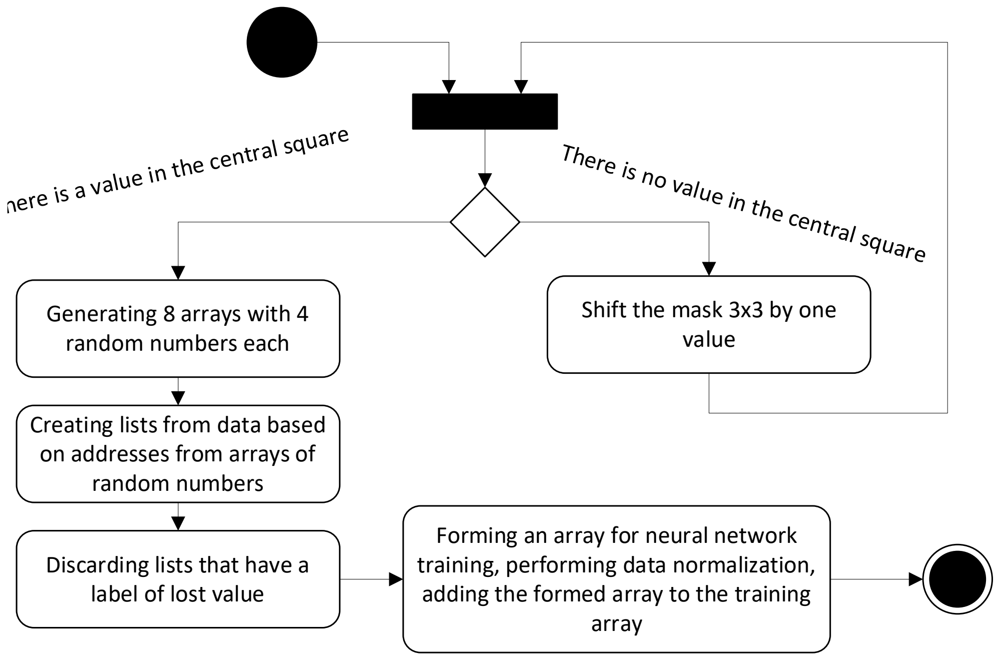


Fig. 6. Activity diagram algorithm

Table 1. Neural Network Structure

| *Layer (type)* | *Output Shape* | *Param #* |
|---|---|---|
| conv2d (Conv2D) | (None, 15, 15, 32) | 320 |
| max_pooling2d (MaxPooling2D) | (None, 14, 14, 32) | 0 |
| dropout (Dropout) | (None, 14, 14, 32) | 0 |
| conv2d_1 (Conv2D) | (None, 14, 14, 64) | 18496 |
| max_pooling2d_1 (MaxPoolin2D) | (None, 13, 13, 64) | 0 |
| dropout_1 (Dropout) | (None, 13, 13, 64) | 0 |
| conv2d_2 (Conv2D) | (None, 13, 13, 128) | 73856 |
| max_pooling2d_2 (MaxPooling2D) | (None, 12, 12, 128) | 0 |
| dropout_2 (Dropout) | (None, 12, 12, 128) | 0 |
| flatten (Flatten) | (None, 18432) | 0 |
| dense (Dense) | (None, 128) | 2359424 |
| dropout_3 (Dropout) | (None, 128) | 0 |
| dense_1 (Dense) | (None, 16) | 2064 |

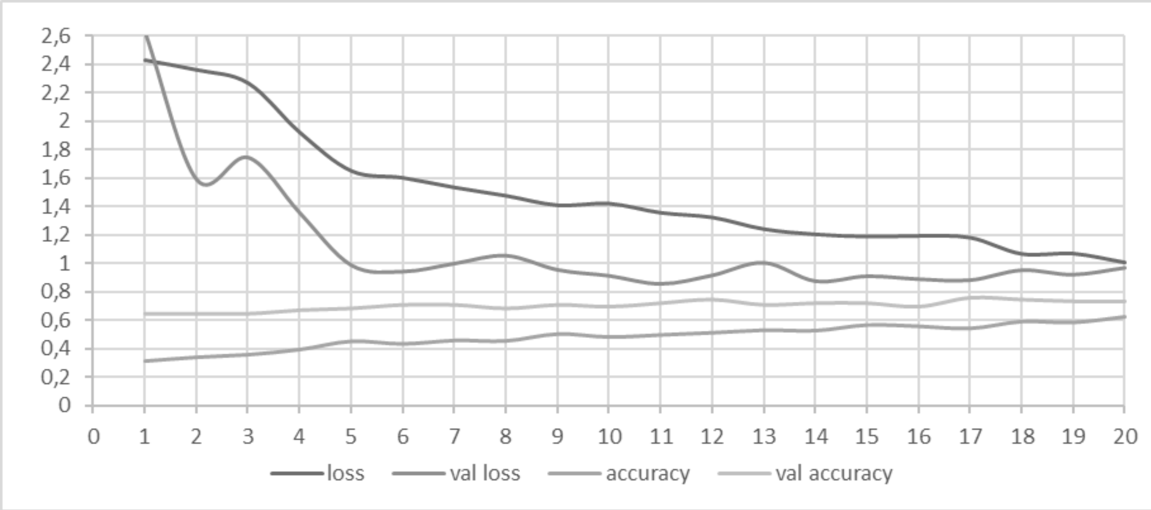


Fig. 7. Neural network learning curve

To implement the algorithm for predicting the number of buildings in lost areas, a neural network (perceptron) was created, the structure of which is described in Table 2. The network input included 10 values: the number of buildings in any four of the eight neighboring squares relative to the square where the number of buildings needs to be found, the numbers of these squares, and the coordinates of the searched square. The output of the neural network is the value of the searched number of buildings in the given square. Since the shape of all proto-cities is different, the model is trained separately for each proto-city, on its preserved part.

We can only check the quality of the neural network's work on the control data set used for training control. Thus, we obtained an accuracy of 58% for the Talianky settlement and 42% for the Maidanetske settlement. Such results cannot be called very good, however, this accuracy is much higher than random guessing, the probability of which is slightly over 6%. As experiments have shown, increasing the number of neurons and input data did not improve the results, it was also found that using different optimization functions during training can increase or decrease the result by up to 4% for different settlements. For example, the Adam optimizer showed the best result for Talianky, while RMSprop showed the best result for Maidanetske. In the final version of the system, the Adam optimizer was chosen because it showed a more stable result. The total number of neurons in the model that performed the best is about 280 thousand.

Table 2. Neural network structure

| *Layer (type)* | *Output Shape* | *Param #* |
|---|---|---|
| dense (Dense) | (None, 512) | 5632 |
| dropout (Dropout) | (None, 512) | 0 |
| dense_1 (Dense) | (None, 512) | 262656 |
| dropout_1 (Dropout) | (None, 512) | 0 |
| dense_2 (Dense) | (None, 16) | 8208 |

Thus, it was calculated that there were at least 1603 dwellings in Maidanetske, archaeologists believe that the number could have reached up to 2000 [6]. Assuming 2000 as the correct result, we can say that the error was slightly less than 20%. The situation with the estimation of the number of Talianky settlements is complicated by the fact that a significant part of the settlement area has been lost (up to 40%), making it difficult to provide adequate values. According to the system's calculations, there could have been 2885 buildings in the proto-city. Comparing these data with the results obtained by the author using the average building density method: 2484 dwellings [4]. The difference between the calculations of both methods is approximately 15%, so both methods can provide quite correct results. If there is data on the city's volumes, one can already estimate the approximate number of inhabitants, the areas of fields needed to provide food for people and animals, the city's economic power, which will allow for a more detailed picture of the Trypillia world.

## Conclusions

The article examines the possibilities of using neural networks for the analysis of magnetic images of Trypillia settlements to estimate their parameters, such as the number of buildings, building density, and the approximate population. The developed algorithm, which combines a convolutional neural network and a perceptron, shows promise for automating the analysis of magnetic images and obtaining important information about Trypillia proto-cities. With minor modifications, it can be used to solve a whole range of tasks related to searching for or counting small contrasting elements in the image, far beyond the scope of archaeological science.

## References

1. Tite, M. S., Mullins, C. (1971) Enhancement of the magnetic susceptibility of soils on archaeological sites. Archaeometry, 13 (2), 209-219. doi: 10.1111/j.1475-4754. 1971. tb00043.x [Online]. Available: https://doi.org/10.1111/j.1475-4754.1971.tb00043.x
2. Shmaglij, N. M., Videjko, M. Ju. (2000) Izuchenie poselenija tripolskoj kultury Majdaneckoe. Zbirnyk naukovyx pracz: Drevnee obshhestva juga Vostochnoj Evropy v jepohu polimetalla. SPb: Institut istorii materialnoj kultury RAN.
3. Melnykov, O. Yu., Kapeleshhuk, A. O. (2020) Prykladna programa dlya rozraxunku kilkosti meshkanciv arxeologichnogo poselennya. Cyfrova ekonomika ta informacijni texnologiyi: zbirnyk tez mizhnarodnoyi naukovo-praktychnoyi konferenciyi, m. Kyiv, 15-16 kvitnya 2020 r. Kyiv: DUIT, 48-50.
4. Melnykov, O. Yu., Kapeleshhuk, A. O. (2020) Rozraxunok kilkosti zhyteliv v gorodyshhax arxeologichnyx kultur za dopomogoyu specialnogo programnogo zabezpechennya. Informacijni texnologiyi v kulturi, mystecztvi, osviti, nauci, ekonomici ta biznesi: materialy V Mizhnarodnoyi naukovo – praktychnoyi konferenciyi. Kyiv: Vydavnychyj centr KNUKiM, 127-129.
5. Burdo, N. B., Videjko, M. Yu. (2007) Trypilska kultura. Spogady pro zolotyj vik. Xarkiv: Folio.
6. Smolij V. A. ta inshi. (2009) Encyklopediya istoriyi Ukrayiny u 10 t. T.6: La – Mi. Kyiv: Naukova dumka.

The article has been sent to the editors 19.04.24.
After processing 24.09.24.
Submitted for printing 30.09.24.