Application of Semantic Technologies in a Model of Automatic Recognition and Analysis of Elements in Images of the Earth's Surface

Elizaveta Kovsher
Department of Electronic Computing Machines
Belarusian State University of
Informatics and Radioelectronics
Minsk, Belarus
elizavetakovsher@mail.ru

Maksim Mikhalkov
Department of Information Technology Software
Belarusian State University of
Informatics and Radioelectronics
Minsk, Belarus
solution.m@mail.ru

***Abstract*—This research investigates the impact of neural network model training parameters in the context of object classification within the framework of semantic technologies, specifically for the analysis of Earth's surface imagery.**

It is shown how semantic technologies and their application to the task of classifying objects in earth surface images improve the ability of the model to process and classify objects.

A detailed analysis of the influence of key training parameters on the accuracy and stability of the model is carried out. These factors were assessed within the framework of the semantic technologies approach.

The findings highlight the effectiveness of the neural network model in semantic technology applications for Earth imagery analysis.

These results can be applied to improve the performance of neural network-based object classification systems within the domain of semantic technologies.

***Keywords*—semantic technologies, machine learning, neural networks, DARKNET, YOLOv3, geoinformation data processing**

I. Introduction

Classification of objects in images of the Earth's surface is an important issue that finds wide application in various fields of science, technology and economics [1]. These images represent unique sources of information that allow to extract valuable data on the state of the environment [2].

Semantic technologies play a key role in interpreting Earth observation data, as they allow not only to identify objects, but also to extract their semantic characteristics, identify logical relationships, and analyze the context of the scene [3]. Unlike simple object recognition, which operates at the level of individual pixel areas, semantic analysis includes higher-level abstractions such as ontological modeling, logical-computational methods, and the extraction of hidden patterns.

One of the key issues of semantic technologies is the formation of knowledge based on multispectral and hyperspectral remote sensing data. This is achieved by integrating spectral, spatial, and textural features with semantic models, which allows distinguishing similar objects based on their physical and contextual characteristics [4]. For example, distinguishing natural and anthropogenic objects can be performed taking into account their shape, spatial distribution, and spectral response.

In addition, semantic technologies provide the ability to automatically perform logical inference based on rules and predicates. This is especially important in natural disaster monitoring, urban planning and agriculture, where not only detection of objects is required, but also analysis of their condition, dynamics of change and potential threats [5].

The aim of this research is to evaluate the impact of training parameters on the performance of a neural network model, using a semantic technology approach to classify objects in images of the Earth's surface.

II. Formalization of the system model

The issue of automatic recognition and analysis of elements in images of the earth's surface using the YOLOv3 neural network architecture implemented in the Darknet framework is in the context of semantic technologies, which are actively used to extract, classify and analyze objects in images. Semantic technologies are aimed at processing and understanding data in terms of their meaning, and not only based on structural features.

In this context, using YOLOv3 as a model for detecting objects in images helps to effectively perform the issue of recognizing various objects, such as buildings, roads, trees and other elements in images of the earth's surface. The YOLOv3 architecture uses convolutional neural networks (CNN) to solve the problem of classification and localization of objects in images in real time.

Semantic technologies are closely related to machine learning and artificial intelligence approaches. In the issue of image recognition, semantics is the ability of a neural network to understand objects and their contexts [6].

The YOLOv3 model also actively uses semantic approaches to improve prediction accuracy. The essence of the algorithm is that the network divides the image into a grid and for each cell determines bounding boxes with possible objects and their classes, predicting coordinates and probabilities. This makes it possible to quickly and accurately find and classify objects in the image.

One of the important aspects of using YOLOv3 is the ability to process images with different resolutions. The YOLOv3 architecture uses several levels of convolutional layers, which allows the model to effectively detect objects both at the global and local levels, maintaining rich semantic information about various objects.

Further the main aspects of semantic technologies implemented by the YOLOv3 neural network based on the DARKNET framework were presented.

To structure information, an ontological representation of knowledge is used, formalized in the form of triplets (O_i, R_{ij}, O_j), where O_i, O_j are objects detected by the neural network, and R_{ij} is the relation between objects, for example be near, cross, include. Formally, the semantic network is represented by a graph G = (V, E), where V is the set of detected objects, and E are the relations between them, determined using spatial proximity metrics, for example, the intersection coefficient:

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|}. \quad (1)$$

This approach allows for automatic analysis of spatial relationships between objects and the formation of meaningful conclusions about the structure of the scene.

Semantic interpretation of objects is performed using fuzzy logic and first-order predicates, which allows for formalization of relations between objects and logical inference based on rules. For example, if an object of the road class is adjacent to an unidentified object, the model can conclude that there is a high probability that this object is a car. This is formalized as a predicate:

$$(near(x,y) \wedge class(x) = road) \Rightarrow class(y) = car. \quad (2)$$

This approach enhances the capabilities of traditional recognition by introducing elements of logical analysis.

For detailed analysis of spatial structures, segmentation using deep neural networks such as U-Net and DeepLabV3 is used. Unlike classical object detection, segmentation allows you to determine the exact boundaries of objects, not just bounding boxes. This is especially important when analyzing complex scenes, such as recognizing agricultural fields, water bodies, and urban structures.

Texture analysis plays an important role in semantic recognition. For example, the GLCM (Gray Level Co-occurrence Matrix) method allows you to describe the statistical properties of image texture structures. The main GLCM metrics include energy, which reflects the homogeneity of the texture, and entropy, which characterizes the degree of randomness of the pixel distribution.

Vegetation indices are used to analyze objects with characteristic spectral properties (vegetation, water bodies, snow cover). For example, NDVI (Normalized Difference Vegetation Index) is defined as:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (3)$$

where

- NIR – the value of the near infrared channel of the image,
- RED – the value of the near red channel of the image.

Another important index, NDSI (Normalized Difference Snow Index), is used to detect snow cover and is calculated as:

$$NDSI = \frac{GREEN - SWIR}{GREEN + SWIR}. \quad (4)$$

These characteristics are integrated into the semantic analysis process, allowing to increase the accuracy of detection and interpretation of objects.

The aim of the analysis of experimental data is to determine estimates of unknown parameters b in a certain given region of the factor space X. The statistical model of the system is presented in Fig. 1.

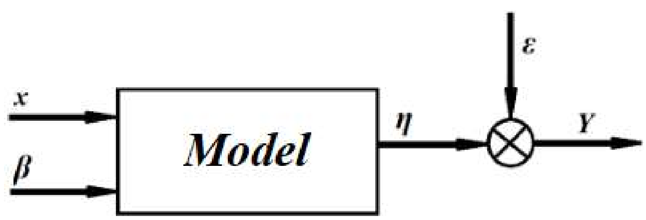

Figure 1. Statistical model of the system.

In real conditions, due to the presence of interference ϵ, instead of the true value of the output quantity η, it is necessary to measure the quantity Y. Therefore, based on the measurement results, it is impossible to obtain absolutely accurate values of b. Instead of the true parameters b, it is necessary to obtain estimates of the parameters β [8]. Then, the estimated equation for the model will have the form:

$$Y = Y(x, \beta), \quad (5)$$

where

- x – factors (input measurements),
- β – estimation of unknown parameters (coefficients),
- Y – system response taking into account the interference ϵ (output value).

For the effective use of semantic technologies in the issues of analyzing images of the earth's surface, it is necessary to select the optimal parameters of the neural network model. Their correct configuration allows to improve the results of semantic analysis, providing more accurate recognition and interpretation of objects in images.

III. Experimental design

A. General structure of experimental design

The main steps of experimental design are: defining the aim of the experiment, selecting factors and levels of their variation, defining output variables (responses), defining the type of experiment, determining the training sample size, randomization, taking into account and monitoring external conditions, conducting a pilot (test) experiment, data processing.

B. Defining the aim of the experiment

The following hypothesis is proposed to be tested in this research: the combination of various factors, such as training sample size, number of batches, learning rate, and number of training epochs, significantly affects the accuracy, performance, and robustness of the model for automatic recognition and analysis of features in earth surface images. This hypothesis is hereinafter referred to as H_0.

C. Selecting factors and levels of their variation

Each of the factors can significantly affect the results, and by changing their levels, it is possible to find the optimal settings for the issue at hand. Below, how each of these factors will change is presented.

Table I
Factors and levels of their variation

Factor	*Low level*	*High level*
Training sample size	100	800
Number of epochs	50	300
Number of batches	8	16
Learning rate	0.001	0.01

Each of the proposed 4 factors will vary between 2 levels, which will give 16 possible combinations for experiments (2^4=16). This will allow to explore how each factor affects the performance of the YOLOv3 model in recognizing objects in images of the earth's surface.

D. Defining output variables (responses)

The output variables (or responses) in this experiment are metrics that measure the performance of the YOLOv3 model in recognizing objects in images of the Earth's surface. These metrics reflect the accuracy of the model and its ability to correctly classify objects [9]. The geometric representation of the response surface is shown in Fig. 2.

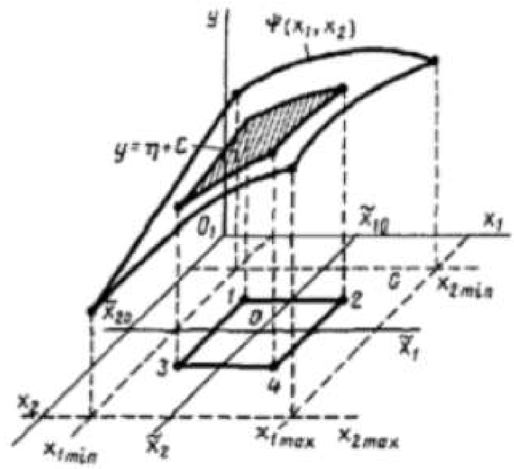

Figure 2. Geometric representation of the reaction surface.

Average Precision (mAP) is the average Precision value at different confidence threshold levels for each class. For the issue of object recognition, where the model makes predictions on the location of objects and their classes, mAP takes into account both the classification accuracy and the correctness of localization [10].

Next, it is necessary to formalize the algorithm for calculating mAP.

Precision-Recall Curve is a curve that reflects how the accuracy changes depending on what confidence threshold is used to recognize the object.

Average Precision (AP) for a class is calculated as the integral of the Precision–Recall curve. In practice, it can be calculated as the area under this curve, or the average precision at different recall levels.

The equation for calculating AP for one class:

$$AP = \frac{1}{N}\sum_{i=1}^{N} Precision(i) * \Delta Recall(i), \quad (6)$$

where

- N – number of points on the Precision–Recall curve,
- Precision(i) – precision at the i-th point of the curve,
- Recall(i) – difference in recall at the i-th point.

After each class has its Average Precision (AP), calculate mAP as the average of all APs for all classes. The equation for mAP:

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i, \quad (7)$$

where

- C – number of classes in the issue (in the model under consideration 60),
- AP_i – accuracy for the i-th class.

A high mAP value (e.g. close to 1) indicates that the model is effective at recognizing objects with high accuracy and minimal errors. A low mAP value (e.g. close to 0) indicates that the model is poor at recognizing objects, or the localization and classification error is too high.

E. Defining the type of experiment

The type of experiment that was chosen to achieve the stated aim in the issue of recognizing objects in images of the earth's surface is a full-factorial experiment.

A full-factorial experiment allows to explore all possible combinations of factors, which gives a complete picture of how each factor affects the result, and also allows to evaluate their interaction.

F. Determining the training sample size

The training sample size of an experiment is the number of observations or repetitions of the experiment that must be conducted for each combination of factors so that the results are statistically significant and provide an accurate estimate of the model's responses.

For each set of factors, several repetitions are planned (2-3 repetitions for each combination). This is necessary to account for random fluctuations associated with the initialization of the model weights and the random learning process.

G. Randomization

Randomization in an experiment is the process of randomly assigning different conditions and variants of an experiment to eliminate possible systematic errors and improve the reliability of the results.

For 16 factor combinations (a full-factorial experiment with 4 factors, each with 2 levels), the order of treatment of each combination in different rounds of the experiment will be randomly assigned. Thus, if 2 repetitions are planned for each combination, the order of execution of each of the 32 experiments will be randomly shuffled.

H. Taking into account and monitoring external conditions

The consideration and control of external conditions in the experiment are aimed at minimizing the influence of factors that are not part of the variables under research, but can have a significant impact on the results. In this case, this includes conditions that can affect the training and performance of the YOLOv3 model in recognizing objects in images of the earth's surface, such as hardware characteristics, execution environment, random fluctuations, and other unchangeable factors.

I. Conducting a pilot (test) experiment

A pilot experiment is a small preliminary experiment that is conducted before the main research to check the correctness of the selected factors, conditions, and equipment.

Table II
Pilot Experiment Variant

Factor	*Value 1*	*Value 2*
Training sample size	100	100
Number of epochs	50	50
Number of batches	8	16
Learning rate	0.01	0.001

Table below summarizes the factors and their values that will be used for the pilot experiment.

After training the model with these factors combinations, it is necessary to calculate the average mAP values for each configuration and select the most appropriate factors for the main experiment.

J. Data processing

Since the experiment is full-factorial, each factor is coded using binary values (0 and 1) for ease of subsequent analysis. Low level is value 0, high level is value 1.

All collected data are entered into a table (the matrix of a full-factorial experiment), where each row corresponds to one combination of factors.

Analysis of Variance (ANOVA) is used to assess the significance of the influence of factors on the output variable - in this case, mAP.

For ANOVA, the ratio of intergroup and intragroup variances is calculated:

$$F = \frac{MS_{factor}}{MS_{error}}, \tag{8}$$

where

- $MS_{factor}=\frac{SS_{factor}}{df_{factor}}$ – average squared deviations of changes in factors,
- $MS_{error}=\frac{SS_{error}}{df_{error}}$ – average squared deviations of random errors,
- SS_{factor} – sum of squared deviations due to factor influence,
- df_{factor} – sum of squared deviations due to factor influence,
- SS_{error} – sum of squared deviations due to factor influence,
- df_{error} – sum of squared deviations due to factor influence.

The total sum of squared deviations is the sum of squared deviations of all observations from the overall mean response value:

$$SS_{all} = \sum_{i=1}^{N}(Y_i - \overline{Y})^2, \tag{9}$$

where

- Y_i – response value (mAP) for the i-th combination of factors,
- $\overline{Y}$ – overall average mAP value across all experiments,

- N – total number of experiments.

Next, for each factor, its contribution to the variability of the response is calculated:

$$SS_{factor} = n\sum_{j=1}^{2}(\overline{Y}_{factor_j} - \overline{Y})^2, \quad (10)$$

where

- $\overline{Y}_{factor_j}$ – average response value for the j-th level of the factor,
- n – number of repetitions for each level.

Finally, the sum of the squares of the errors is calculated:

$$SS_{error} = SS_{all} - \sum SS_{factor}. \quad (11)$$

Regression analysis is used to build a mathematical model that describes the dependence of the output variable (mAP) on the values of the factors. The aim is to determine which factors have a significant effect on the result and to obtain a regression equation that can be used to predict the mAP value for new combinations of factors. The least squares method is used to estimate the coefficients of the model.

Multiple linear regression assumes that the response (Y, in this case – mAP) is a linear function of the factors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon, \quad (12)$$

where

- Y – output variable (mAP),
- $X_1, X_2, ..., X_n$ – factors,
- β_0 – free term,
- $\beta_1, \beta_2, ..., \beta_n$ – regression coefficients,
- ϵ – random error.

The least squares method assumes that we minimize the sum of the squared deviations of the actual response values from those predicted by the model:

$$S = \sum_{i=1}^{N}(Y_i - \overline{Y}_i)^2 = \sum_{i=1}^{N}(Y_i - (\beta_0 + \sum_{j=1}^{n}\beta_j X_{ij}))^2. \quad (13)$$

To find the regression coefficients, it is necessary to solve the system of normal equations:

$$X^T X \beta = X^T Y, \quad (14)$$

where

- X – matrix of factors,
- β – vector of regression coefficients,
- Y – vector of responses.

Based on the results of statistical and regression analysis, the values of the factors that provide the best response value (mAP) are selected. This allows to determine the optimal combination of parameters for training the YOLOv3 model.

IV. Conducting an experiment

Immediately after compiling the matrix of the full-factorial experiment, randomization was performed, i.e. random distribution of the sequence of experiments. It was carried out by means of the Python programming language, namely using the random library to perform this issue.

To ensure stability and reproducibility of the results, the experiment was conducted on a single computing device, eliminating the influence of hardware differences on the rate and accuracy of model training. All versions of libraries and drivers were fixed to exclude incompatibilities and changes in calculation algorithms.

During the experiment, the mAP was recorded at each stage. All collected data are summarized in a table below, where each row corresponds to one combination of factors. The mAP values were calculated for two images whose terrains contain different classes of objects suitable for recognition.

Table III
Extended Matrix of Full-factorial Experiment

Experiment number	*Training sample size*	*Number of batches*	*Learning rate*	*Number of training epochs*	*mAP (1, 2)*
1	0	0	0	0	0.124400, 0.107424
2	0	0	0	1	0.210523, 0.171322
3	0	0	1	0	0.135386, 0.125329
4	0	0	1	1	0.122273, 0.107425
5	0	1	0	0	0.211587, 0.174100
6	0	1	0	1	0.240295, 0.187991
7	0	1	1	0	0.217966, 0.191696
8	0	1	1	1	0.246674, 0.205587
9	1	0	0	0	0.198827, 0.177805
10	1	0	0	1	0.233914, 0.209292
11	1	0	1	0	0.205206, 0.195401
12	1	0	1	1	0.233914, 0.209292
13	1	1	0	0	0.286014, 0.244481
14	1	1	0	1	0.316848, 0.278746
15	1	1	1	0	0.287786, 0.261768
16	1	1	1	1	0.311532, 0.272263

A. Dispersion analysis

The F-criterion values obtained for each factor are summarized in table below. In this case, the tabular value of the F-criterion at a significance level of α=0.05 is 4.60. Based on this and the previously formulated hypothesis H_0, it was concluded that the influence of such factors as the size of the training sample and the number of epochs is significant, while the influence of the factors of the learning rate and the number of batches has a smaller effect, but it also cannot be excluded in further studies.

Table IV
F-criterion Values for Each Factor

Factor	*Value of the F-criterion*
Training sample size (X_1)	10.48
Number of epochs (X_2)	11.29
Learning rate (X_3)	0.44
Number of batches (X_4)	0.83

B. Regression analysis

During this experiment, the following equation was obtained, describing the dependence of mAP on the model parameters:

$$Y = 0.07X_1 + 0.08X_2 + 0.01X_3 + 0.02X_4 + 0.12. \quad (15)$$

The t-statistic values obtained for each factor are summarized in table below. In this case, the tabular value of the t-statistic at a significance level of α=0.05 is 2.20. Based on this and the previously formulated hypothesis H_0, it was concluded that the influence of such factors as the volume of the training sample, the learning rate, the number of epochs and the number of batches is significant, while the influence of the learning speed factor has a smaller effect, but it also cannot be excluded in further research.

The analysis of the resulting equation also confirms the formulated hypothesis H_0 and indicates a significant influence of the factors put forward in the course of the research.

Table V
T-statistic Values for Each Factor

Factor	*Value of the t-statistic*
Training sample size (X_1)	8.48
Number of epochs (X_2)	9.03
Learning rate (X_3)	1.22
Number of batches (X_4)	2.38

V. Conclusions

Based on the conducted experiment, it was concluded that the proposed hypothesis H_0 was confirmed through variance and regression analysis, validating the relationship between parameters and model performance.

In the context of semantic technologies, these improvements allowed the model to better interpret the meaning of objects within images. With a larger training sample, the model was able to capture a wider range of semantic features, improving its ability to detect and classify objects like roads, buildings, and vegetation. Fine-tuning the learning rate helped the model converge faster, leading to more accurate semantic segmentation and object detection.

References

[1] Wei, D.M. Research on Automatic Extraction Method of Landslide Boundary Based on Remote Sensing Image; Southwest Jiaotong University: Chengdu, China, 2013.

[2] Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sens. 2019, 11, 196.

[3] Bragagnolo, L.; Rezende, L.R.; Silva, R.V.; da Grzybowski, J.M.V. Convolutional neural networks applied to semantic segmentation of landslide scars. Catena 2021, 201, 105189.

[4] Li, Y.; Cui, P.; Ye, C.M.; Junior, J.M.; Zhang, Z.T.; Guo, J.; Li, J. Accurate Prediction of Earthquake-Induced Landslides Based on Deep Learning Considering Landslide Source Area. Remote Sens. 2021, 13, 3436.

[5] Wang, H.J.; Zhang, L.M.; Yin, K.S.; Luo, H.Y.; Li, J.H. Landslide identification using machine learning. Geosci. Front. 2021, 12, 351–364.

[6] Mohan, A.; Singh, A.K.; Kumar, B.; Dwivedi, R. Review on remote sensing methods for landslide detection using machine and deep learning. Trans. Emerg. Telecommun. Technol. 2021, 32, e3998.

[7] Xu, L.; Liu, X.H.; Zhang, J.Y.; Liu, Z. Research on landslides detection method using remote sensingimages based on Mask R-CNN. J. Shandong Jianzhu Univ. 2023, 38, 94–103.

[8] Chai, J.W.; Nan, Y.T.; Guo, R.; Lin, Y.Q.; Liu, Y.T. Recognition Method of Landslide Remote Sensing Image based on Efficient-Net. In Proceedings of the 2nd ICETCI, Changchun, China, 27–29 May 2022; pp. 1224–1228.

[9] Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. 2018, 40, 834–848.

[10] Li, C.D.; Long, J.J.; Liu, Y.; Yi, S.F.; Feng, P.F. Landslide Remote Sensing Image Recognition Based on EfficientNet: Taking Bijie City, Guizhou Province as an Example. South China Geol. 2023, 39, 403–412.

ПРИМЕНЕНИЕ СЕМАНТИЧЕСКИХ ТЕХНОЛОГИЙ В МОДЕЛИ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ И АНАЛИЗА ЭЛЕМЕНТОВ НА СНИМКАХ ЗЕМНОЙ ПОВЕРХНОСТИ

Ковшер Е. И., Михальков М. Д.

Данное исследование направлено на оценку влияния параметров обучения нейронной сети в контексте классификации объектов на изображениях земной поверхности с использованием семантических технологий.

Показано, как семантические технологии и их применение в задаче классификации объектов на изображениях земной поверхности улучшить способность модели обрабатывать и классифицировать объекты.

Проведен подробный анализ влияния ключевых параметров обучения на точность и стабильность модели. Эти факторы были оценены в рамках подхода семантических технологий.

Полученные результаты подчеркивают эффективность предложенной модели нейронной сети в приложениях семантических технологий для анализа изображений земной поверхности.

Эти результаты могут быть использованы для повышения производительности систем классификации объектов на основе нейронных сетей в области семантических технологий.

Received 13.03.2024