

ABSTRACT

Topic: «Algorithmic software for big data annotation for object detection task using computer vision methods»

Master's degree thesis: 119 pages, 33 figures, 6 tables, 2 attachments, 91 references.

Relevance of the topic. AI algorithms are increasingly penetrating the lives of ordinary people, automating such common areas of life as driving a car, shopping, disease diagnosis, and much more. To simplify the development process of artificial intelligence algorithms, artificial intelligence software platforms such as Microsoft Azure AI Platform, Amazon AWS SageMaker, Google Cloud AI Platform, IBM Watson, and many others have been created. The major drawback of these platforms is that they focus on simplifying the process of creating artificial intelligence algorithms, but according to Cognilytica analysts, only 20% of the time on artificial intelligence projects is spent working with the algorithm, and the other 80% - working with data. The longest stage in working with data is data markup for further use in the learning process of machine learning algorithms, and in modern artificial intelligence platforms there is no automation of the markup process, or it available only for typical tasks, which leads to inefficient (slow) data markup process. That is why the task of developing data markup software, which will increase the speed of data markup, is relevant.

Research objective. Improve efficiency of data markup process for object detection tasks by auto-labeling data using computer vision techniques that perform pre-markup of data on which a person only needs to correct errors of an algorithm. To achieve this goal, it is necessary to organize the process of continuous learning of the algorithm for detecting objects on new data and the correct order of image markup, which will allow as soon as possible to mark the entire available data set.

The following objectives have been formulated to achieve this research objective:

- explore existing approaches to automating data markup in software platforms to solve AI problems;

- explore existing approaches for incremental learning of machine learning algorithms;
- improve the auto-labeling process for object detection tasks to speed up the markup of large data sets;
- develop data markup software for object detection tasks;
- study the effectiveness of the algorithm.

The object of research. The process of auto-markup data for computer vision tasks, including object detection tasks.

The subject of research. Approaches for auto-labeling of data for object detection tasks.

The scientific innovation. Proposed mechanism of prioritization of data markup process as improvement of the mechanism of self-diagnostics object detection models which establishes the labeling order by selecting the most difficult images based on predictions of neural network. Training on such images is much faster compared to using random images from the same data set.

The algorithm for auto-labeling of large volumes of data for object detection tasks has been improved by utilizing proposed mechanism of prioritization and by adaptating approaches for continuous learning, which allows to increase data labeling speed and can be used for labeling both typical and atypical objects.

The practical value. Developed auto-labeling software within the AI software platform for object detection problems, which differs from existing analogues by the increased speed of the markup process at very large volumes.

Approbation. The results of the master's degree thesis were reported in The Fourth International Conference on Computer Science, Engineering and Education Applications (ICCSEEA2021) and a collection of materials from the student conference held in Kiev, Ukraine in 2021.

Keywords: AI SOFTWARE PLATFORMS, COMPUTER VISION, NEURAL NETWORKS, OBJECT DETECTION, AUTOLABELING, BIG DATA