ABSTRACT

Explanatory note size – 116 pages, contains 3 illustrations, 30 tables, 5 applications, 36 references.

Topicality. The work examines the problem of creating image datasets, shows the main features of existing solutions to the described problems, their advantages and disadvantages. It has been concluded that there is a need to improve methods for automating the creation of image datasets, generating synthetic images and improving image quality, and to develop software to implement these methods for a wide range of machine learning tasks.

The aim of the study. The main goal is to increase the efficiency of creating high-quality image datasets suitable for a wide range of machine learning tasks.

The object of research: software for intelligent creation of image datasets.

The subject of research: software for intelligent creation of image datasets.

To achieve this goal, the **following tasks** were formulated:

- analysis of existing methods and problems of automated creation of image datasets;
- development of methods for automated image collection, generation and augmentation;
- improvement of methods for automated annotation and image quality assessment;
- integration of the developed methods into a single pipeline for creating datasets;
- experimental verification of the effectiveness of the proposed methods.

The scientific novelty of the results of the master's dissertation is that a software solution has been developed that, unlike others, provides the user with the ability to create image datasets for various machine learning tasks with the possibility of full automation of the process. The result was achieved by developing methods for augmentation and creation of synthetic images using the Stable Diffusion generative model, a method for

image enhancement using Stable Diffusion, and a method for filtering irrelevant images using automatic tagging models and LLM ChatGPT-4o.

The practical value of the obtained results is that the automation of data collection, processing and annotation significantly reduces the time and human resources required to create image datasets. The versatility of the system is ensured by the support of various machine learning tasks, which makes the product available to a wide range of users. Integration of modern technologies, such as generative models and LLM, ensures the creation of more diverse and high-quality datasets. Cost-effectiveness is achieved by reducing the cost of manual processing.

Relationship with working with scientific programs, plans, topics. Work was performed at the Department of Informatics and Software Engineering of the National Technical University of Ukraine «Kyiv Polytechnic Institute. Igor Sikorsky».

Approbation. The scientific provisions of the dissertation were tested at the VII International Scientific and Practical Conference of Young Scientists and Students "Software Engineering and Advanced Information Technologies SoftTech-2024".

Publications. The scientific provisions of the dissertation were published in:

1) Ukrainets D.R. Library for Intelligent Image Dataset Creation // Software engineering and advanced information technologies (SoftTech-2024): materials of abstracts of the VII All-Ukrainian scientific and practical conference of young scientists and students (Kyiv, 19-22 November 2024) - K. : Igor Sikorsky Kyiv Polytechnic Institute, 2024.

Keywords: IMAGE DATASET, MACHINE LEARNING, STABLE DIFFUSION, GENERATIVE MODELS, ANNOTATION, IMAGE QUALITY ASSESSMENT, SYNTHETIC DATA, PYTHON.