

РЕФЕРАТ

Розмір пояснювальної записки – 116 аркушів, містить 3 ілюстрації, 30 таблиць, 5 додатків, 36 посилань на джерела.

Актуальність теми. У роботі розглянуто проблему створення наборів даних зображень, показано основні особливості існуючих рішень описаних проблем, їх переваги та недоліки. Виявлено потребу в удосконаленні методів автоматизації створення наборів даних зображень, генерації синтетичних зображень та покращення якості зображень, а також в розробці програмного забезпечення для реалізації цих методів для широкого кола задач машинного навчання.

Мета дослідження. Основною метою є підвищення ефективності створення високоякісних наборів даних зображень, придатних для широкого кола задач машинного навчання.

Об'єкт дослідження: програмне забезпечення для інтелектуального створення наборів даних зображень.

Предмет дослідження: методи та засоби створення програмного забезпечення для інтелектуального створення наборів даних зображень.

Для реалізації поставленої мети **сформульовані наступні завдання:**

- аналіз існуючих методів та проблем автоматизованого створення наборів даних зображень;
- розробка методів автоматизованого збору, генерації та аугментації зображень;
- удосконалення методів автоматизованого аотування та оцінки якості зображень;
- інтеграція розроблених методів у єдиний конвеєр для створення наборів даних;
- експериментальна перевірка ефективності запропонованих методів.

Наукова новизна результатів магістерської дисертації полягає в тому, що розроблено програмне рішення, яке на відміну від інших надає користувачеві можливості для створення наборів даних зображень для різних задач машинного навчання з можливістю повної автоматизації процесу. Результат досягнутий шляхом розробки методів для аугментації та створення синтетичних зображень з використанням генеративної моделі Stable Diffusion, методу покращення зображення за допомогою Stable Diffusion та методу фільтрації нерелевантних зображень за допомогою моделей автоматичного тегування та LLM ChatGPT-4o.

Практичне значення отриманих результатів полягає в тому, що завдяки автоматизації процесів збору, обробки та анотування даних значно скорочуються часові та людські ресурси, необхідні для створення наборів даних зображень. Універсальність системи забезпечується підтримкою різних задач машинного навчання, що робить продукт доступним для широкого кола користувачів. Інтеграція сучасних технологій, таких як генеративні моделі та великі мовні моделі, забезпечують створення більш різноманітних та якісних наборів даних. Економічна ефективність досягається через зменшення витрат на ручну обробку.

Зв'язок з науковими програмами, планами, темами. Робота виконувалась на кафедрі інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського".

Апробація. Наукові положення дисертації пройшли апробацію на VII Міжнародній науково-практичній конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології SoftTech-2024»

Публікації. Наукові положення дисертації опубліковані в:

- 1) Українець Д.Р. Бібліотека для інтелектуального створення наборів даних зображень // Інженерія програмного забезпечення і передові інформаційні технології (SoftTech-2024) : матеріали тез доповідей VII Всеукраїнської

наук.-практ. конф. молодих вчених та студентів (м. Київ, 19-22 листопада 2024). – К. : КПІ ім. Ігоря Сікорського, 2024.

Ключові слова: НАБІР ДАНИХ ЗОБРАЖЕНЬ, МАШИННЕ НАВЧАННЯ, STABLE DIFFUSION, ГЕНЕРАТИВНІ МОДЕЛІ, АНОТАЦІЯ, ОЦІНКА ЯКОСТІ ЗОБРАЖЕНЬ, СИНТЕТИЧНІ ДАНІ, PYTHON.