Вінницький національний технічний університет
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

КОМЕНЧУК ОЛЕГ ВІКТОРОВИЧ

УДК [004.8+004.93]

ДИСЕРТАЦІЯ
ІНТЕЛЕКТУАЛЬНА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК

Спеціальність 126 – «Інформаційні системи та технології»
Галузь знань 12 – «Інформаційні технології»

Подається на здобуття ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело _____ О. В. Коменчук

Науковий керівник: Мокін Олександр Борисович,
доктор технічних наук, професор

Вінниця – 2024

АНОТАЦІЯ

Коменчук О.В. Інтелектуальна інформаційна технологія розпізнавання об’єктів на зображеннях з використанням псевдомасок – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 126 «Інформаційні системи та технології». – Вінницький національний технічний університет, Вінниця, 2024.

Метою дисертаційної роботи є розробка інтелектуальної інформаційної технології для підвищення ефективності розпізнавання об’єктів на зображеннях з використанням псевдомасок. Основна задача полягає у покращенні якості розпізнавання малорозмірних та зашумлених даних, що часто зустрічаються у медичних та промислових зображеннях, за допомогою глибокого навчання та автоматизованих методів анотації зображень.

У вступі визначено актуальність теми, мету і завдання дослідження, а також описано об’єкт та предмет дослідження. Наукову новизну роботи становить розробка методів, що дозволяють використовувати псевдомаски для автоматичної анотації зображень. Результати роботи мають важливе практичне значення, що підтверджується апробацією технології на реальних медичних даних та впровадженням її у навчальний процес.

У першому розділі подано огляд сучасних методів розпізнавання об'єктів на зображеннях, зокрема генеративних моделей, автокодувальників та згорткових нейронних мереж. Показано, що традиційні методи не завжди є ефективними для малих або зашумлених наборів даних. Визначено необхідність застосування псевдомасок для покращення результатів сегментації та ідентифікації об'єктів у складних умовах.

У другому розділі розглянуто методи попередньої обробки зображень, включаючи адаптивну нормалізацію, фільтрацію шумів і вирівнювання гістограм. Розроблено алгоритми автоматичної генерації псевдомасок для

неанотованих даних на основі аналізу активаційних карт мережі, що дозволяє підвищити ефективність сегментації.

У третьому розділі розроблено та впроваджено алгоритми для автоматизації процесу генерації псевдомасок і тренування моделей глибокого навчання на основі цих масок. Впроваджено технології на платформі Jupyter-ноутбуків, що забезпечують можливість використання для різних типів зображень. Програмні модулі є адаптивними і можуть бути інтегровані в різні платформи, такі як AWS SageMaker та Google Colab.

У четвертому розділі проведено серію експериментів на медичних і промислових зображеннях, які підтвердили ефективність використання псевдомасок. У задачах сегментації медичних знімків точність підвищилася порівняно з традиційними методами. Проведено порівняння з існуючими рішеннями і показано переваги запропонованої технології для малорозмірних або зашумлених наборів даних.

У висновках наведено основні результати дисертаційної роботи.

Основні наукові та практичні результати полягають в наступному:

1. Аналіз сучасних методів та вибір оптимального підходу:

Проведено ґрунтовний аналіз сучасних методів розпізнавання об’єктів на зображеннях, включаючи такі підходи, як автокодувальники, згорткові нейронні мережі (CNN), генеративно-змагальні мережі (GAN) та методи напівконтрольованого навчання. Було детально розглянуто переваги та недоліки кожного з цих методів у контексті їхньої ефективності для обробки малорозмірних або зашумлених наборів даних. У результаті вибір було зроблено на користь використання псевдомасок, які дозволяють суттєво покращити продуктивність моделей глибокого навчання в умовах обмежених даних та шуму. Показано, що методи, засновані на генерації псевдомасок, дозволяють зменшити кількість розміток, необхідних для тренування моделей, забезпечуючи при цьому високі показники точності.

2. Удосконалення методів попередньої обробки даних:

Важливим результатом роботи є розробка та впровадження нових методів попередньої обробки зображень, які дозволяють адаптувати параметри залежно від специфіки даних. У розділі 2 представлено техніки нормалізації, фільтрації шуму та вирівнювання гістограм, що забезпечують покращення якості даних перед сегментацією. Використання таких підходів дозволило значно знизити вплив шуму та підвищити загальну якість моделей, які працюють із зображеннями. Особливу увагу приділено питанням передобробки для медичних зображень, де точність та якість розпізнавання є критичними.

3. Розробка нової інформаційної технології з використанням псевдомасок:

У ході дослідження було створено нову інформаційну технологію, що базується на застосуванні псевдомасок для підвищення точності сегментації та ідентифікації об'єктів на зображеннях. Запропонована технологія включає використання сучасних моделей глибокого навчання (CNN та GAN), які автоматично генерують псевдомаски для слабко анотованих або неанотованих зображень. Це дозволяє значно скоротити потребу в ручній розмітці, що є особливо важливим для медичних даних та промислових застосувань. Додатково, в технології впроваджено механізми автоматичної перевірки якості псевдомасок, що підвищує надійність процесу сегментації та знижує ймовірність помилок на етапі навчання моделей.

4. Алгоритмічна реалізація та програмне забезпечення:

У розділі 3 було розроблено низку алгоритмів, що реалізують запропоновані методи генерації псевдомасок та їх подальшого використання у процесі навчання моделей глибокого навчання. Алгоритмічні основи побудовані таким чином, що дозволяють легко інтегрувати розроблену технологію в існуючі системи для аналізу медичних зображень, промислових даних або будь-яких інших специфічних наборів даних. Програмне забезпечення було реалізовано у вигляді Jupyter-ноутбуків, що дозволяє автоматизувати процеси генерації масок, попередньої обробки даних, навчання моделей та візуалізації результатів. Це програмне забезпечення є універсальним і може бути легко адаптоване для

використання на різних платформах, таких як AWS SageMaker, Google Colab, або локальні системи машинного навчання.

5. Покращення ефективності моделі та оцінка результатів:

Експериментальні дослідження показали, що використання псевдомасок дозволяє суттєво підвищити точність сегментації та ідентифікації об’єктів на зображеннях малорозмірних та зашумлених наборів даних. У ході експериментів з медичними зображеннями, такими як рентгенівські знімки зубів та КТ-знімки, було досягнуто покращення точності порівняно з традиційними підходами, що не використовують псевдомаски. Проведене порівняння з існуючими аналогами, такими як стандартні методи сегментації (наприклад, U-Net), продемонструвало переваги запропонованої технології у задачах сегментації медичних та промислових зображень, особливо в умовах високого рівня шуму або браку розмічених даних.

6. Практичне значення розроблених методів та технологій:

Розроблені методи мають значний потенціал для практичного впровадження у різних галузях, таких як охорона здоров'я, промисловий контроль якості, автономні транспортні засоби та аналіз супутникових знімків. Використання псевдомасок у поєднанні з сучасними технологіями глибокого навчання дозволяє не лише підвищити точність систем комп’ютерного зору, але й значно скоротити витрати на ручну анотацію даних. Система може бути адаптована для різних типів зображень та сфер застосування, що робить її універсальним інструментом для автоматизації задач розпізнавання об’єктів.

7. Впровадження в навчальний процес та подальші наукові дослідження:

Результати дослідження були впроваджені в навчальний процес Вінницького національного технічного університету, зокрема для викладання дисциплін, пов'язаних із інтелектуальними інформаційними системами та технологіями аналізу зображень. Це дозволяє студентам отримати практичний досвід роботи з передовими методами комп'ютерного зору. Крім того, результати роботи можуть бути основою для подальших досліджень у галузі автоматизації

діагностичних процесів та розробки нових методів виявлення та сегментації об'єктів на складних зображеннях, особливо у медичних та промислових умовах.

8. Публікації та наукове визнання:

Основні результати дослідження були викладені у низці наукових статей та представлені на науково-практичних конференціях. Загалом опубліковано 9 наукових праць, серед яких 4 статті у фахових журналах України [50], [66-68], та 5 тез доповідей на конференціях [69-73].

Наукова новизна дисертаційної роботи полягає в наступних положеннях:

1. Розроблено нову інформаційну технологію для сегментації та розпізнавання об’єктів на зображеннях з використанням псевдомасок. Створена технологія дозволяє ефективно використовувати неанотовані дані для тренування моделей, що є важливим для задач медичної діагностики та промислового контролю якості.

2. Уперше розроблено інформаційну технологію прискореного анотування медичних зображень для задач сегментації з використанням моделей глибокого навчання. Цей підхід включає в себе автоматичне генерування сегментаційних псевдомамасок на основі текстових підказок, що значно скорочує час, необхідний для розмітки медичних знімків. Технологія також включає використання розробленого механізму автоматичного валідування створених первинних псевдомасок, що дозволяє збільшити точність етапу автоматичного анотування та знизити потребу у ручному корегуванні анотацій.

3. Удосконалено метод передоброблення медичних рентгенівських знімків для задачі розпізнавання цільових об’єктів на зображеннях.

4. Як додатковий результат у напрямку комп’ютерного зору, розроблено інтелектуальну технологію побудови плану будівель за аерофотозйомкою дахів. Ця технологія використовує глибокі нейронні мережі для аналізу аерофотознімків, що дозволяє створювати детальні та точні плани будівель. Запропонована технологія може бути використана для автоматизованого моніторингу та аналізу урбаністичних територій.

Отримані результати є корисними для:

1. Рекомендовано ТОВ «ОРНЕТІКА» (див. додаток Б) для впровадження в практичну діяльність при розробці інтелектуальних систем для вирішення задач ком'ютерного зору. Зокрема цінними є наступні отримані результати дослідження:

- Використання розроблених покращених методів попередньої обробки даних на основі контексту зображення, зокрема використання адаптивного білатерального фільтра та покращеного фільтра CLAHE до панорамних стоматологічних рентгенівських знімків зубів для вирішення задач сегментації та детекції об'єктів на КТ знімках.

- Проєктування системи з використанням розробленої інформаційної технології на основі моделі Language Segment-Anything з текстовими підказками для автоматичного створення первинних псевдомасок.

- Використання розробленої системи автоматичної валідації на основі моделі EfficientNet для перевірки точності створення псевдомасок.

- Використання методів створення псевдомасок для підвищення точності виявлення об'єктів на зображеннях.

- Підготовка та тренування моделей сегментації та детекції для вирішення задачі ідентифікації об'єктів на зображеннях з використанням сучасних фреймворків глибокого навчання на Python (PyTorch, PyTorch Lightning).

2. Навчального процесу Вінницького національного технічного університету під час викладання освітніх компонентів «Інтернет речей та інтелектуальний аналіз даних» та «Інформаційні технології моніторингу та аналізу стану складних систем» здобувачам, які навчаються за освітньо-науковою програмою «Інформаційні системи та технології» третього (освітньо-наукового) рівня вищої освіти зі спеціальності 126 Інформаційні системи та технології, впроваджено результати (див. додаток В) дисертаційної роботи на здобуття доктора філософії аспіранта кафедри САІТ Коменчука Олега Вікторовича у ряді лекцій та практичних занять з питань інформаційних технологій. Використовуються такі результати досліджень, отримані у дисертаційній роботі:

- використання розроблених покращених методів попередньої обробки даних на основі контексту зображення, зокрема адаптивний білатеральний фільтр, покращений CLAHE та інших алгоритмів;
- тренування та використання моделей виявлення об'єктів на зображеннях, таких як ResNet, U-Net, EfficientNet, DETR та Vision Transformer (ViT), з використанням сучасних фреймворків Python (PyTorch, PyTorch Lightning);
- використання розроблених методів створення псевдомасок для підвищення точності виявлення об'єктів на зображеннях;
- використання розробленої інформаційної технології на основі моделі Language Segment-Anything з текстовими підказками для автоматичного створення первинних псевдомасок.

3. Рекомендовано ТОВ «АНАЛІТИКА» (див. додаток Г) для впровадження в практичну діяльність для вирішення задач ком’ютерного зору. Зокрема цінними є наступні отримані результати дослідження:

- використовуються результати проведеного аналізу найпоширеніших типів дахів (односхилі, двосхилі, плоскі, вальмові) та створеного узагальненого алгоритму для їх обробки;
- використовується комбінований підхід для побудови планів будівель на основі даних аерофотозйомки з інтеграцією моделі DETR для детекції дахів і сегментації за допомогою ViTs (Vision Transformers). Для післяобробки використовуються запропоновані методи, такі як Canny Edge Detection, для визначення контурів дахів;
- тренування та використання моделей виявлення об'єктів на зображеннях, таких як DETR та Vision Transformers (ViTs), з використанням сучасних фреймворків Python (PyTorch, PyTorch Lightning);
- використовуються напрацювання по методах створення псевдомасок для підвищення точності виявлення об'єктів на зображеннях.

Ключові слова: псевдомаски, інтелектуальна інформаційна технологія, сегментація, розпізнавання об’єктів, глибоке навчання, медичні зображення, метрика, глибоке навчання, ком’ютерний зір, штучний інтелект, нейронна мережа, машинне навчання, розпізнавання шаблонів, трансформери, аерофотозйомка.

ABSTRACT

Komenchuk O.V. Intelligent information technology for object recognition in images using pseudo-masks - Qualifying scientific work on manuscript rights.

Dissertation for obtaining the scientific degree of candidate of technical sciences in specialty 126 "Information systems and technologies". – Vinnytsia National Technical University, Vinnytsia, 2024.

The aim of the dissertation is the development of intelligent information technology to increase the efficiency of object recognition in images using pseudo-masks. The main objective is to improve the recognition quality of low-dimensional and noisy data, often found in medical and industrial images, using deep learning and automated image annotation methods.

The introduction defines the relevance of the topic, goal and task of the research, as well as describes the object and subject of the research. The scientific novelty of the work is the development of methods that allow the use of pseudomasks for automatic image annotation. The results of the work have important practical significance, which is confirmed by the approbation of the technology on real medical data and its introduction into the educational process.

The first chapter provides an overview of modern methods of object recognition in images, including generative models, autoencoders, and convolutional neural networks. It is shown that traditional methods are not always effective for small or noisy data sets. The necessity of using pseudo-masks to improve the results of segmentation and identification of objects in difficult conditions was determined.

The second section discusses image preprocessing techniques, including adaptive normalization, noise filtering, and histogram smoothing. Algorithms for the automatic generation of pseudomasks for unannotated data based on the analysis of network activation maps have been developed, which allows to increase the efficiency of segmentation.

In the third chapter, algorithms for automating the process of generating pseudomasks and training deep learning models based on these masks are developed and implemented. Technologies have been implemented on the Jupyter notebook

platform, which provide the possibility of use for various types of images. The software modules are responsive and can be integrated into different platforms such as AWS SageMaker and Google Colab.

In the fourth chapter, a series of experiments on medical and industrial images were conducted, which confirmed the effectiveness of the use of pseudomasks. In the tasks of segmentation of medical images, the accuracy increased compared to traditional methods. A comparison with existing solutions is made and the advantages of the proposed technology for small-sized or noisy data sets are shown.

The main results of the dissertation work are given in the conclusions.

The main scientific and practical results are as follows:

1. Analysis of modern methods and selection of the optimal approach:

A thorough analysis of modern methods of object recognition in images is carried out, including such approaches as autoencoders, convolutional neural networks (CNN), generative adversarial networks (GAN) and semi-supervised learning methods. The advantages and disadvantages of each of these methods were discussed in detail in the context of their effectiveness for processing small or noisy data sets. As a result, the choice was made in favor of the use of pseudomasks, which allow to significantly improve the performance of deep learning models in conditions of limited data and noise. It is shown that the methods based on the generation of pseudo-masks allow to reduce the number of markings required for training models, while providing high accuracy rates.

2. Improvement of data preprocessing methods:

An important result of the work is the development and implementation of new image preprocessing methods that allow adapting parameters depending on the specifics of the data. Section 2 presents normalization, noise filtering, and histogram smoothing techniques that improve data quality before segmentation. The use of such approaches made it possible to significantly reduce the impact of noise and improve the overall quality of models that work with images. Special attention is paid to reprocessing issues for medical images, where the accuracy and quality of recognition are critical.

3. Development of new information technology using pseudomasks:

In the course of the research, a new information technology based on the use of pseudo-masks was created to increase the accuracy of segmentation and identification of objects in images. The proposed technology includes the use of modern deep learning models (CNN and GAN) that automatically generate pseudomasks for weakly annotated or unannotated images. This greatly reduces the need for manual labeling, which is especially important for medical data and industrial applications. In addition, mechanisms for automatic quality control of pseudomasks are implemented in the technology, which increases the reliability of the segmentation process and reduces the probability of errors at the model training stage.

4. Algorithmic realization and software:

In Chapter 3, a number of algorithms were developed that implement the proposed methods of generating pseudomasks and their further use in the process of learning deep learning models. Algorithmic frameworks are built in such a way that allows easy integration of the developed technology into existing systems for the analysis of medical images, industrial data or any other specific data sets. The software was implemented in the form of Jupyter notebooks, which allows you to automate the processes of mask generation, data preprocessing, model training, and visualization of results. This software is versatile and can be easily adapted for use on different platforms, such as AWS SageMaker, Google Colab, or local machine learning systems.

5. Improving the efficiency of the model and evaluating the results:

Experimental studies have shown that the use of pseudomasks allows you to significantly increase the accuracy of segmentation and identification of objects in the images of small and noisy data sets. In experiments with medical images such as dental X-rays and CT scans, accuracy improvements have been achieved compared to traditional approaches that do not use pseudomasks. The comparison with existing analogues, such as standard segmentation methods (for example, U-Net), demonstrated the advantages of the proposed technology in the tasks of segmentation of medical and industrial images, especially in conditions of high noise level or lack of labeled data.

6. Practical significance of the developed methods and technologies:

The developed methods have significant potential for practical implementation in various fields, such as health care, industrial quality control, autonomous vehicles, and satellite image analysis. The use of pseudomasks in combination with modern deep learning technologies allows not only to increase the accuracy of computer vision systems, but also to significantly reduce the cost of manual data annotation. The system can be adapted for different types of images and fields of application, which makes it a universal tool for automating object recognition tasks.

7. Implementation in the educational process and further scientific research:

The results of the research were implemented in the educational process of Vinnytsia National Technical University, in particular for teaching disciplines related to intelligent information systems and image analysis technologies. This allows students to gain hands-on experience with advanced computer vision techniques. In addition, the results of the work can be the basis for further research in the field of automation of diagnostic processes and the development of new methods of detection and segmentation of objects in complex images, especially in medical and industrial conditions.

8. Publications and scientific recognition:

The main results of the research were presented in a number of scientific articles and presented at scientific and practical conferences. In total, 9 scientific works were published, including 4 articles in professional journals of Ukraine [50], [66-68], and 5 abstracts of reports at conferences [69-73].

The scientific novelty of the dissertation consists in the following provisions:

1. A new information technology has been developed for segmentation and recognition of objects in images using pseudomasks. The created technology allows you to effectively use unannotated data for training models, which is important for the tasks of medical diagnostics and industrial quality control.

2. Information technology for accelerated annotation of medical images for segmentation tasks using deep learning models was developed for the first time. This approach includes the automatic generation of segmentation pseudo-masks based on textual prompts, which significantly reduces the time required for labeling medical

images. The technology also includes the use of a developed mechanism for automatic validation of created primary pseudomasks, which allows to increase the accuracy of the automatic annotation stage and reduce the need for manual correction of annotations.

3. The method of reprocessing medical X-ray images for the task of recognizing target objects in images has been improved.

4. As an additional result in the direction of computer vision, an intelligent technology for constructing a plan of buildings based on aerial photography of roofs has been developed. This technology uses deep neural networks to analyze aerial photographs, allowing for the creation of detailed and accurate building plans. The proposed technology can be used for automated monitoring and analysis of urban areas.

The obtained results are useful for:

1. "ORNETIKA" LLC is recommended for implementation in practical activities in the development of intelligent systems for solving computer vision problems. In particular, the following obtained research results are valuable:

- Use of the developed improved data preprocessing methods based on the image context, in particular the use of an adaptive bilateral filter and improved CLAHE filters for panoramic dental X-ray images of teeth to solve the problems of segmentation and detection of objects on CT images.

- System design using developed information technology based on the Language Segment-Anything model with text prompts for automatic creation of primary pseudomasks.

- Using the developed automatic validation system based on the EfficientNet model to check the accuracy of creating pseudo-masks.

- Use of methods of creating pseudo-masks to increase the accuracy of object detection in images.

- Preparation and training of segmentation and detection models to solve the problem of object identification in images using modern Python deep learning frameworks (PyTorch, PyTorch Lightning).

2. The educational process of the Vinnytsia National Technical University during the teaching of the educational components "Internet of Things and intelligent data analysis" and "Information technologies for monitoring and analyzing the state of complex systems" to applicants studying under the educational and scientific program "Information Systems and Technologies" of the third (educational -scientific) level of higher education in the specialty 126 Information systems and technologies, the results of the dissertation work for obtaining a doctor of philosophy of the post-graduate student of the SAIT department Oleh Viktorovych Komenchuk were implemented in a series of lectures and practical classes on information technology issues. The following research results obtained in the dissertation work are used:

- use of developed improved data preprocessing methods based on the image context, in particular adaptive bilateral filter, improved CLAHE and other algorithms;

- training and use of image object detection models, such as ResNet, U-Net, EfficientNet, DETR and Vision Transformer (ViT), using modern Python frameworks (PyTorch, PyTorch Lightning);

- use of developed methods of creating pseudo-masks to increase the accuracy of object detection in images;

- use of developed information technology based on the Language Segment-Anything model with text prompts for automatic creation of primary pseudomasks.

3. ANALYTIKA LLC is recommended for implementation in practical activities for solving computer vision problems. In particular, the following obtained research results are valuable:

- the results of the analysis of the most common types of roofs (one-slope, two-slope, flat, hip) and the created generalized algorithm for their processing are used;

- a combined approach is used to construct building plans based on aerial photography data with the integration of the DETR model for roof detection and segmentation using ViTs (Vision Transformers). For post-processing, proposed methods such as Canny Edge Detection are used to determine the contours of roofs;

- training and use of image object detection models, such as DETR and Vision Transformers (ViTs), using modern Python frameworks (PyTorch, PyTorch Lightning);

- developments in the methods of creating pseudo-masks are used to increase the accuracy of object detection in images.

Keywords: pseudomasks, intelligent information technology, segmentation, object recognition, deep learning, medical images, metric, deep learning, computer vision, artificial intelligence, neural network, machine learning, pattern recognition, transformers, aerophotography.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації:

[1] О.В. Коменчук, О.Б. Мокін, «Аналіз методів передоброблення панорамних стоматологічних рентгенівських знімків для задач сегментації зображень», Вісник Вінницького політехнічного інституту, вип. 5, с. 41-49, 2023. DOI: https://doi.org/10.31649/1997-9266-2023-170-5-41-49 .

(Особистий внесок: Дослідження включає оцінювання та покращення фільтрів CLAHE, Гауса, білатерального фільтра та MSRCR за допомогою тонкого налаштування для покращення якості медичних зображень. Модель сегментації на базі U-Net навчалась на попередньо оброблених зображеннях, а ефективність перевірялась за допомогою метрик Дайса, Жаккара, влучності та повноти. О.Б. Мокін: узгодження задач дослідження та аналіз результатів, компонування і загальне редагування статті.)

[2] О.В. Коменчук, В.Б. Мокін, Є.М. Крижановський, В.О. Будяк, «Інтелектуальна технологія побудови плану будівель за аерофотозйомкою їхніх дахів», Вісник Вінницького політехнічного інституту, вип. 1, с. 101-109, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-172-1-101-109 .

(Особистий внесок: Запропоновано удосконалення інтелектуальної технології побудови плану будівель за даними аерофотозйомки їх дахів, шляхом інтеграції моделі для детекції DETR (“DEtection TRansformer”) разом із сегментацією на основі ViTs (Vision Transformers) для комплексного вирішення проблем пошуку та ідентифікації дахів, у першому наближенні, для подальшого покращення побудови планів будівель. В.Б. Мокін: узгодження задач дослідження, аналіз результатів та загальне редагування статті. Є.М. Крижановський: методологія автоматизації практичного розгортання розробленої технології на практиці з використанням ГІС-технологій. В.О. Будяк: відбір прикладів для застосування технології та аналіз ефективності її застосування.)

[3] О.В. Коменчук, О.Б. Мокін, «Інформаційна технологія прискореного анотування медичних зображень в задачах сегментації на основі моделей глибокого навчання», Вісник Вінницького політехнічного інституту, вип. 4, с. 95-103, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-175-4-95-103 .

(Особистий внесок: Уперше розроблено інтелектуальну інформаційну технологію прискореного анотування медичних зображень для задач сегментації з використанням моделей глибокого навчання та псевдомасок. Проведено порівняння продуктивності інструментів для створення анотацій медичних зображень, зокрема Supervisely, CVAT та Segments.ai, з розробленою інформаційною технологією, яка використовує модель Language Segment-Anything з текстовими підказками для створення первинних псевдомасок та уперше розробленим автоматичним валідуванням на основі моделі EfficientNet-B2. О.Б. Мокін: узгодження задач дослідження та аналізу результатів, компонування і загальне редагування статті.)

[4] О.В. Коменчук, «Адаптивні методи попередньої обробки для підвищення точності сегментації стоматологічних рентгенівських знімків», Сучасний стан наукових досліджень та технологій в промисловості, вип. 3, с. 29-38, 2024. DOI: https://doi.org/10.30837/2522-9818.2024.3.029 .

(Особистий внесок: Проведено роботу над удосконаленням білатерального фільтру та фільтру CLAHE для попереднього оброблення даних за рахунок розробки та впровадження адаптивного налаштування параметрів фільтрів на основі контексту зображення та проведено експериментальне порівняння розроблених адаптивних методів попередньої обробки з традиційними підходами.)

Наукові праці, які засвідчують апробацію матеріалів дисертації:

[1] В.Б. Мокін, О.В. Коменчук, «Огляд сучасних інформаційних технологій для класифікації зображень за матеріалами конкурсів Kaggle 2019-2020 років», Матеріали L науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 1469-1471, 2021. Режим доступу:

https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38276/135476.pdf?sequence=2&isAllowed=y .

(Особистий внесок: Проведено огляд сучасних інформаційних технологій для класифікації зображень. В.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[2] В.Б. Мокін, О.В. Коменчук, «Сучасні інформаційні технології для розпізнавання дахів будівель на аерофотозйомці», Матеріали LI науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 704-708, 2022. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/40198/16150.pdf?sequence=3&isAllowed=y .

(Особистий внесок: Описана технологія розпізнавання дахів будівель на аерофотозйомці, продемонстровано приклад роботи запропонованої технології. В.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[3] О.В. Коменчук, С.А. Кирилащук, «Огляд сучасних педагогічних технологій в освітньому процесі», Матеріали Всеукраїнської науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», с. 458-487, 2022. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2022/paper/viewFile/15485/13068 .

(Особистий внесок: Розглянуто сучасні педагогічні технологій в освітньому процесі. С.А. Кирилащук: загальне редагування тексту.)

[4] О.В. Коменчук, О.Б. Мокін, «Методи передобробки панорамних стоматологічних рентгенівських знімків для задач глибокого навчання», Матеріали LII науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 229-231, 2023. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38239/17182.pdf?sequence=3&isAllowed=y .

(Особистий внесок: Проведено огляд методів для передоброблення медичних знімків, визначено кроки для вдосконалення таких методів. О.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[5] О.В. Коменчук, О.Б. Мокін, «Удосконалення процесу анотування медичних зображень для задач сегментації за допомогою текстових підказок», Матеріали Міжнародної науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», 2024. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/21692/17951 .

(Особистий внесок: Описаний підхід для удосконалення процесу анотування медичних зображень за допомогою текстових підказок, розроблено веб-застосунок. О.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ 24

ВСТУП..... 25

РОЗДІЛ 1 АНАЛІЗ ПРОБЛЕМ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ МАЛОРОЗМІРНИХ НАБОРІВ ДАНИХ..... 31

1.1Огляд основних принципів і методів розпізнавання об'єктів на зображеннях 31

1.2Аналіз відомих інтелектуальних методів та технологій розпізнавання об'єктів на зображеннях..... 47

1.3Огляд методу розпізнавання об'єктів на зображеннях з використанням псевдомасок..... 58

1.4Аналіз проблем розпізнавання об'єктів на зображеннях та постановка задач 64

1.5 Висновки до розділу..... 67

РОЗДІЛ 2 УДОСКОНАЛЕННЯ МЕТОДІВ ТА МОДЕЛЕЙ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК..... 68

2.1 Аналіз даних та сучасних методів попередньої обробки як основа для розробки технології розпізнавання об'єктів з використанням псевдомасок 68

2.2 Розробка базової моделі для розпізнавання об'єктів на зображеннях. 76

2.3 Удосконалення методів попередньої обробки даних 92

2.4 Удосконалення базової моделі для розпізнавання об'єктів на зображеннях за допомогою покращених методів попередньої обробки даних 100

2.5 Висновки до розділу..... 108

РОЗДІЛ 3 РОЗРОБЛЕННЯ АЛГОРИТМІЧНИХ ОСНОВ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК..... 109

3.1 Розроблення алгоритмів роботи запропонованої інтелектуальної інформаційної технології 109

3.2 Створення UML-діаграм програмного забезпечення для реалізації модулю розпізнавання об’єктів на зображеннях з використанням псевдомасок 115

3.3 Реалізація комплексу програм для застосування розробленої інтелектуальної інформаційної технології 119

3.4 Алгоритм оцінювання ефективності запропонованої інформаційної технології 124

3.5 Висновки до розділу 127

РОЗДІЛ 4 ПРИКЛАДНЕ ЗАСТОСУВАННЯ РОЗРОБЛЕНОЇ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ’ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК.... 129

4.1 Порівняння розробленої інтелектуальної інформаційної технології з аналогами 129

4.2 Розпізнавання об’єктів на зображеннях набору даних медичних знімків з використанням створеної інформаційної технології 136

4.3 Застосування створеної інформаційної технології для розпізнавання об’єктів на зображеннях аерофотозйомки..... 140

4.4 Впровадження розробленої технології у навчальний процес..... 148

4.5Висновки до розділу 149

ВИСНОВКИ 151

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 155

ДОДАТКИ..... 166

ДОДАТОК А СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ..... 167

ДОДАТОК Б АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ..... 171

ДОДАТОК В АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ В НАВЧАЛЬНИЙ ПРОЦЕС ВНТУ... 172
ДОДАТОК Г АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ..... 173

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

ВНТУ – Вінницький національний технічний університет

ML – Machine Learning

DL – Deep Learning

CV – Computer Vision

NN – Neural Network

DNN – Deep Neural Network

CNN – Convolutional Neural Network

SVM – Support Vector Machines

SIFT - Scale-Invariant Feature Transform

SURF - Speeded-Up Robust Features

HOG - Histogram of Oriented Gradients

GPU - Graphics Processing Unit

TPU - Tensor Processing Units

ООП – Об'єктно-орієнтоване програмування

МРТ – Магнітно-резонансна терапія

КТ – Комп’ютерна томографія

UML – Unified Modeling Language

YOLO - You Only Look Once

ВСТУП

Обґрунтування вибору теми дослідження.

Останніми десятиліттями технологія комп'ютерного зору, тобто здатність систем штучного інтелекту "сприймати" зображення подібно до людського ока, залишається об’єктом інтересу та активних наукових досліджень. Основні дослідження цього напрямку спрямовані на створення моделей, що можуть автоматично виконувати завдання, що потребують візуального аналізу. Однак, через складність такого аналізу та велику кількість багатовимірних даних, процес розпізнавання зображень є набагато більш трудомістким, ніж обробка інших типів двійкової інформації. Це, в свою чергу, створює значні виклики в розробці систем штучного інтелекту, здатних ефективно працювати з візуальними даними.

Незважаючи на те, що системи штучного інтелекту та глибоке навчання досягли найвищого рівня продуктивності для розпізнавання об’єктів на зображеннях, їх успіх залежить від великого набору вручну анотованих зображень для навчання. Однак, на практиці рідше має місце ідеальний тренувальний набір даних, тобто достатньо великий, без зайвих шумів із чіткими закономірностями, особливо в галузі візуалізації зображень, де і збирання даних, і їх маркування є дорогими процесами. Останнім часом, виникла потреба у вирішенні проблеми розпізнавання об’єктів на зображеннях малорозмірних та надто зашумлених наборів даних. Таким чином, дуже важливо мінімізувати кількість даних, необхідних для досягнення бажаної точності, за рахунок максимальної ефективності методів навчання, долаючи два основних обмеження набору даних: малу кількість маркованих даних або їх значну зашумленість.

Одним з найбільш ефективних способів усунення такої проблеми є використання псевдомасок, тобто спеціально структурованих об’єктів-масок, ідентифікованих інтелектуальними методами серед немаркованих даних. За допомогою цих масок здійснюється псевдомаркування даних, яке додається до маркованих даних, які є, як правило, у невеликій кількості. Таким чином,

дослідження можливостей комп’ютерного зору з використанням псевдомасок має великий потенціал, оскільки в перспективі застосування псевдомасок стане повсюдним у багатьох наукових і технічних галузях з використанням технологій комп’ютерного зору, що відкриває можливості значно покращити точність ідентифікації та класифікації об’єктів на зображеннях з використанням сильно зашумлених або малорозмірних наборів даних.

Питаннями, пов’язаними з інтелектуальними інформаційними технологіями розпізнавання об’єктів на зображеннях з використанням псевдомасок, багато уваги приділяється в роботах вчених із США (Масачусетський технологічний інститут, Стенфордський університет), Німеччини (Мюнхенський технічний університет, Боннський університет), Китаю (Університет Цінхуа, Пекінський університет), Індії (Технічний університет Нью-Делі, Технічний університет Мумбаї), а також Нідерландів, Франції, Чехії, Японії тощо. У березні 2021 року рішення з використанням псевдомасок посіло 1 місце у конкурсі з призовим фондом $50,000, який організувала RANZCR - неприбуткова професійна організація клінічних радіологів та онкологів з радіації Австралії, Новій Зеландії та Сінгапуру [74].

Зв'язок роботи з науковими програмами, планами, темами.

Мета і завдання дослідження.

Мета дослідження: підвищення точності та швидкості розпізнавання об’єктів на зображеннях малорозмірних, зашумлених та недосконалих наборів даних шляхом розроблення інтелектуальної інформаційної технології з використанням псевдомасок.

Задачами дослідження є:

- провести огляд відомих підходів та моделей для ідентифікації та класифікації об’єктів на зображеннях недосконалих і малорозмірних наборів даних та проаналізувати основні проблеми, які виникають під час ідентифікації та класифікації об'єктів на зображеннях малорозмірних та недосконалих наборів даних;

- розробити інтелектуальну інформаційну технологію, у т.ч. нові підходи, методи та алгоритми, які дозволяють підвищити точність ідентифікації та класифікації заданих об’єктів з використанням псевдомасок на зображеннях недосконалих та малорозмірних наборів даних;
- створити алгоритмічне та програмне забезпечення для застосування запропонованої інтелектуальної інформаційної технології та методів ідентифікації і класифікації заданих об'єктів на зображеннях недосконалих та малорозмірних наборів даних;
- розробити Jupyter-ноутбуки мовою Python для застосування інтелектуальної інформаційної технології з використанням псевдомасок для ідентифікації та класифікації об'єктів на зображеннях недосконалих і малорозмірних наборів даних;
- апробувати та впровадити розроблену інтелектуальну інформаційну технологію на реальних даних та задачах.

Об’єктом дослідження є ідентифікація та класифікація об’єктів на зображеннях недосконалих та малорозмірних наборів даних на основі псевдомасок.

Предметом дослідження є методи, математичні та інтелектуальні моделі розпізнавання об’єктів на зображеннях недосконалих та малорозмірних наборів даних на основі псевдомасок.

Методи дослідження. Аналіз літературних джерел, відкритих даних, експериментальні дослідження, системний аналіз, методи машинного і глибокого навчання та ін.

Інформаційною базою дослідження є недосконалі та малорозмірні набори даних із високим рівнем шуму для виявлення та класифікації об'єктів на зображеннях, переважно медичного призначення. Також, відкриті дані - набори даних Google-платформи Kaggle.

Наукова новизна отриманих результатів.

1. Уперше розроблено метод автоматичного валідування створених первинних псевдомасок на основі моделі класифікації EfficientNet-B2, що

дозволяє збільшити точність автоматичного анотування та знизити потребу у ручному корегуванні анотацій.

2. Удосконалено білатеральний фільтр та фільтр CLAHE для попереднього оброблення медичних рентгенівських знімків, за рахунок адаптивного налаштування параметрів фільтрів на основі контексту зображення, що дозволило покращити характеристики зображень та значно підвищити їх якість, що, у свою чергу, значно покращує точність у задачі сегментації цільових об’єктів на зображеннях.

3. Уперше розроблено інтелектуальну інформаційну технологію прискореного анотування медичних зображень для задач сегментації з використанням моделей глибокого навчання та псевдомасок, що базується на моделі Language Segment-Anything з текстовими підказками та розробленим автоматичним валідуванням створених первинних псевдомасок на основі моделі EfficientNet-B2.

4. Дістав подальший розвиток інтелектуальний метод побудови плану будівель за аерофотозйомкою з використанням удосконаленого підходу сегментування дахів та використання псевдомасок, що дозволяє більш швидко будувати плани будівель за даними аерофотозйомки.

Практичне значення отриманих результатів.

Створено інтелектуальну інформаційну систему, що включає комплекс програмних модулів мовою програмування Python, що реалізують функціонал всіх запропонованих методів, моделей та підходів створення псевдомасок та оцінювання. Окрім цього, покращено методи попереднього оброблення медичних даних за допомогою розроблених адаптивних налаштувань параметрів фільтрів на основі контексту зображення.

Показано результативність використання даної технології на відкритих наборах даних панорамних стоматологічних рентгенівських знімках зубів, рентгенівських знімках легень. Наведені результати можуть бути адаптовані для аналізу інших видів медичних зображень, таких як комп'ютерна томографія (КТ) і магнітно-резонансна томографія (МРТ). Крім того, наведені результати можуть

ефективно використовуватися в екологічному моніторингу для автоматизованої обробки супутникових знімків і аерофотознімків, а також в транспортній сфері — для виявлення дефектів на дорогах та інфраструктурних об'єктах.

Також, результати роботи впроваджено (див. додаток В) у навчальний процес і наукову діяльність кафедри системного аналізу та інформаційних технологій ВНТУ під час викладання освітніх компонентів «Інтернет речей та інтелектуальний аналіз даних» та «Інформаційні технології моніторингу та аналізу стану складних систем» здобувачам, які навчаються за освітньо-науковою програмою «Інформаційні системи та технології» третього (освітньо-наукового) рівня вищої освіти зі спеціальності 126 Інформаційні системи та технології.

Результати роботи рекомендовані до впровадження (див. додаток Б) в практичну діяльність ТОВ «ОРНЕТІКА» (м. Вінниця) при розробці інтелектуальних систем для вирішення задач комп’ютерного зору.

Інтелектуальна технологія побудови плану будівель за аерофотозйомкою дахів передана для впровадження (див. додаток Г) в ТОВ «Аналітика» (м. Вінниця) для прискорення оброблення даних аерофотозйомки.

Використання результатів дисертації підтверджено відповідними актами впровадження.

Публікації.

Всього за тематикою дисертаційного дослідження опубліковано 9 наукових праць. Опубліковано 4 статті у фахових журналах України з технічних наук [50], [66-68], та 5 тез доповідей на науково-практичних конференціях [69-73].

Апробація матеріалів дисертації.

Основні наукові результати та практичні розробки дисертаційної роботи пройшли апробацію на 5 наукових конференціях:

- L науково-технічна конференція підрозділів ВНТУ, Вінниця, 2021 р.;
- LI науково-технічна конференція підрозділів ВНТУ, Вінниця, 2022 р.;
- Всеукраїнська науково-практична інтернет-конференція «Молодь в науці: дослідження, проблеми, перспективи», 2022 р.;

- LII науково-технічна конференція підрозділів ВНТУ, Вінниця, 2023 р.;
- Міжнародна науково-практична інтернет-конференція «Молодь в науці: дослідження, проблеми, перспективи», 2024.

Структура та обсяг дисертації.

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Повний обсяг дисертації становить 173 сторінки, 70 рисунків, 4 таблиці, список використаних джерел із 89 найменувань, кількість додатків – 4.

РОЗДІЛ 1
АНАЛІЗ ПРОБЛЕМ РОЗПІЗНАВАННЯ ОБ’ЄКТІВ НА ЗОБРАЖЕННЯХ МАЛОРОЗМІРНИХ НАБОРІВ ДАНИХ

1.1 Огляд основних принципів і методів розпізнавання об’єктів на зображеннях

Розпізнавання об'єктів на зображеннях є однією з основних задач комп'ютерного зору, яка має значний вплив на розвиток сучасних технологій штучного інтелекту. Дана задача має надзвичайно широкий спектр застосувань, від автоматизації рутинних процесів до розв’язання складних завдань, пов’язаних з медичною діагностикою та безпекою. Завдяки розпізнаванню об’єктів системи комп’ютерного зору можуть не лише аналізувати зображення, але й отримувати з них структуровану інформацію, яка допомагає приймати рішення в реальному часі.

Ідентифікація об'єктів на зображеннях перетворилося на один із провідних напрямів досліджень у галузі комп'ютерного зору завдяки стрімкому розвитку технологій штучного інтелекту та збільшенню обчислювальних потужностей. Популярність цієї технології пояснюється її широким спектром застосувань, де автоматизація процесів на основі аналізу візуальних даних може значно підвищити ефективність та точність рішень. Від розпізнавання обличь у смартфонах до систем автономного управління транспортними засобами, розпізнавання об’єктів на зображеннях стало необхідним інструментом у різних сферах життя. Однак складність візуальних даних, різноманітність умов зйомки, а також необхідність роботи з обмеженими наборами даних створюють виклики для точності та надійності існуючих рішень. Тому дослідження в цій області залишаються актуальними, спрямованими на створення нових алгоритмів і технологій для покращення продуктивності систем комп'ютерного зору в реальних умовах.

Одним із найбільш популярних застосувань розпізнавання об'єктів є автоматичне управління транспортними засобами. Сучасні автомобілі, обладнані

камерами та іншими сенсорами, використовують алгоритми комп'ютерного зору для розпізнавання дорожньої розмітки, знаків, пішоходів та інших учасників руху. Це дозволяє системам на базі автопілотів забезпечувати безпечний та ефективний рух на дорогах, запобігаючи дорожньо-транспортним пригодам та оптимізуючи маршрути. Завдяки цьому, розпізнавання об'єктів на зображеннях є невід'ємною частиною індустрії автономних транспортних засобів.

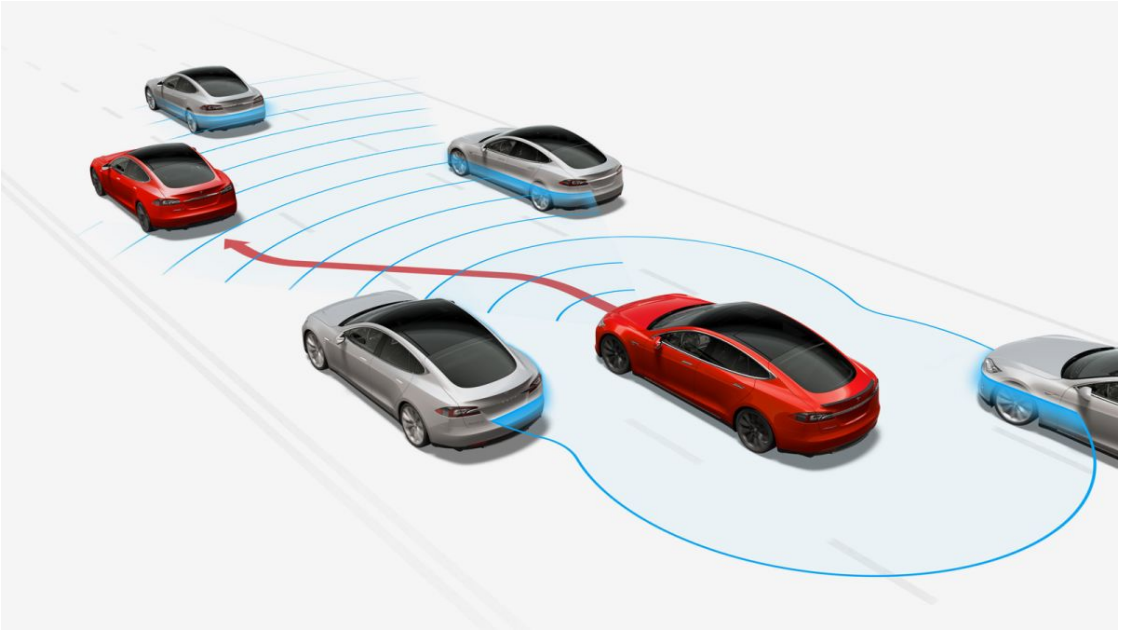


Рис. 1.1 Ілюстративне зображення роботи системи автопілоту, де камери та датчики розпізнають інших учасників руху [1]

Розпізнавання об'єктів не обмежується лише транспортною сферою. Ще одним важливим напрямком його застосування є індустріальна автоматизація. На виробничих лініях розпізнавання об’єктів дозволяє машинам і роботам ідентифікувати деталі, оцінювати їх якість та контролювати процеси складання. Завдяки використанню комп'ютерного зору, автоматизовані системи можуть своєчасно виявляти дефекти продукції, забезпечуючи постійний контроль якості. Це значно підвищує ефективність виробництва та зменшує кількість помилок, що особливо важливо в умовах масового виготовлення продукції.

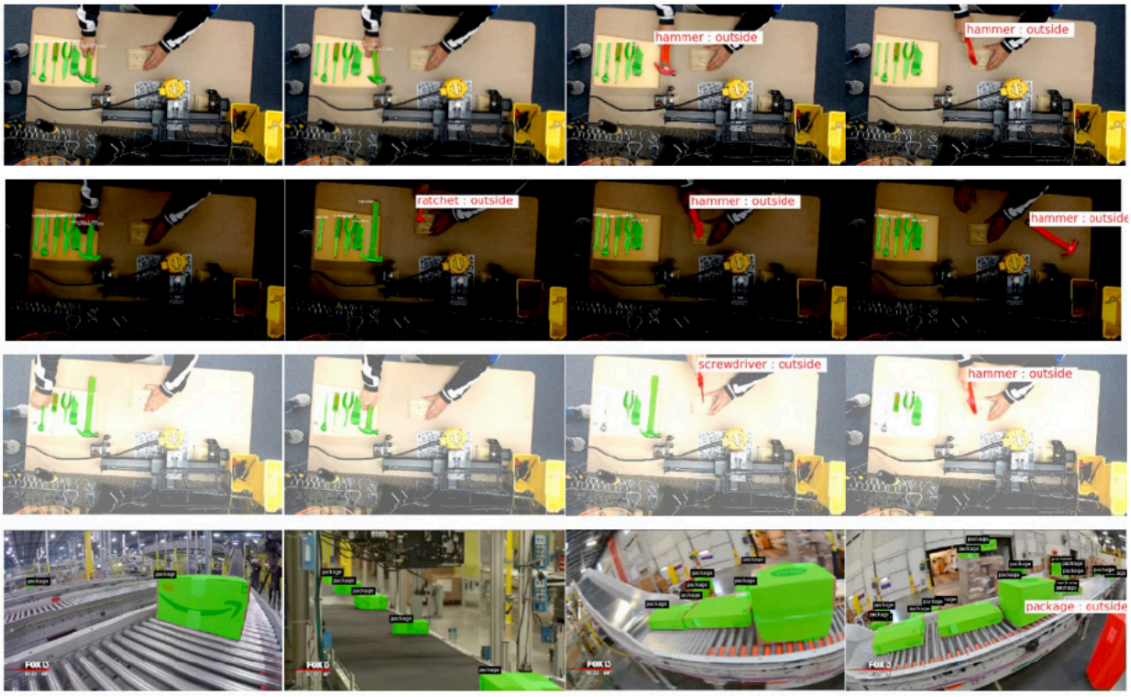


Рис. 1.2 Приклади застосування систем розпізнавання об’єктів на зображеннях на виробництвах [2]

У сфері охорони здоров'я розпізнавання об’єктів на зображеннях має досить вагоме значення для медичної діагностики. За допомогою технологій комп'ютерного зору лікарі можуть швидко і точно аналізувати рентгенівські знімки, МРТ або КТ-сканування для виявлення патологій, таких як пухлини, переломи чи інші аномалії. Це не лише підвищує точність діагностики, але й скорочує час, необхідний для аналізу зображень, що особливо важливо в умовах обмежених ресурсів або при наданні невідкладної медичної допомоги.

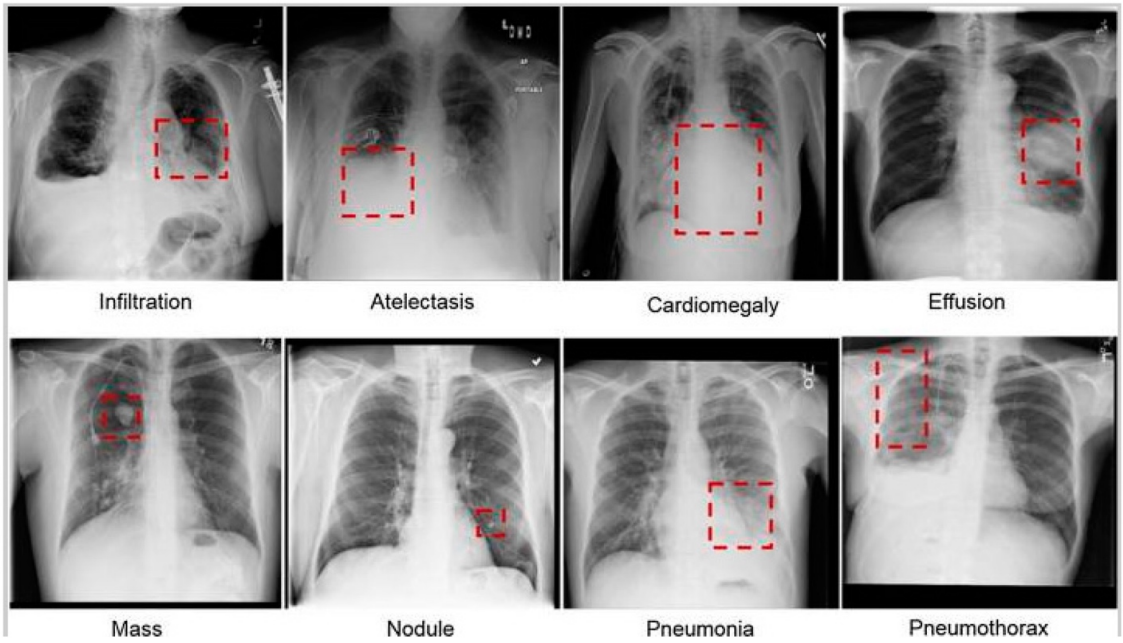


Рис. 1.3 Ілюстративний приклад роботи алгоритмів розпізнавання об’єктів на медичних знімках [3]

Розпізнавання об’єктів також широко застосовується в системах безпеки. Камери відеоспостереження, оснащені алгоритмами штучного інтелекту, можуть автоматично розпізнавати обличчя, транспортні засоби, підозрілі дії або навіть аналізувати поведінку людей у режимі реального часу. Це сприяє підвищенню ефективності функціонування систем безпеки, зменшити кількість помилкових спрацьовувань та забезпечити більш швидке реагування на можливі загрози.

Застосування технологій розпізнавання об'єктів у системах безпеки також дозволяє суттєво зменшити кількість помилкових спрацьовувань, оскільки системи комп’ютерного зору можуть відфільтровувати несуттєві події, такі як рух тварин або тіней. Це зменшує навантаження на операторів і підвищує загальну продуктивність системи. Крім того, можливість аналізу даних у реальному часі забезпечує більш швидке реагування на можливі загрози, мінімізуючи час від виявлення проблеми до прийняття відповідних дій.

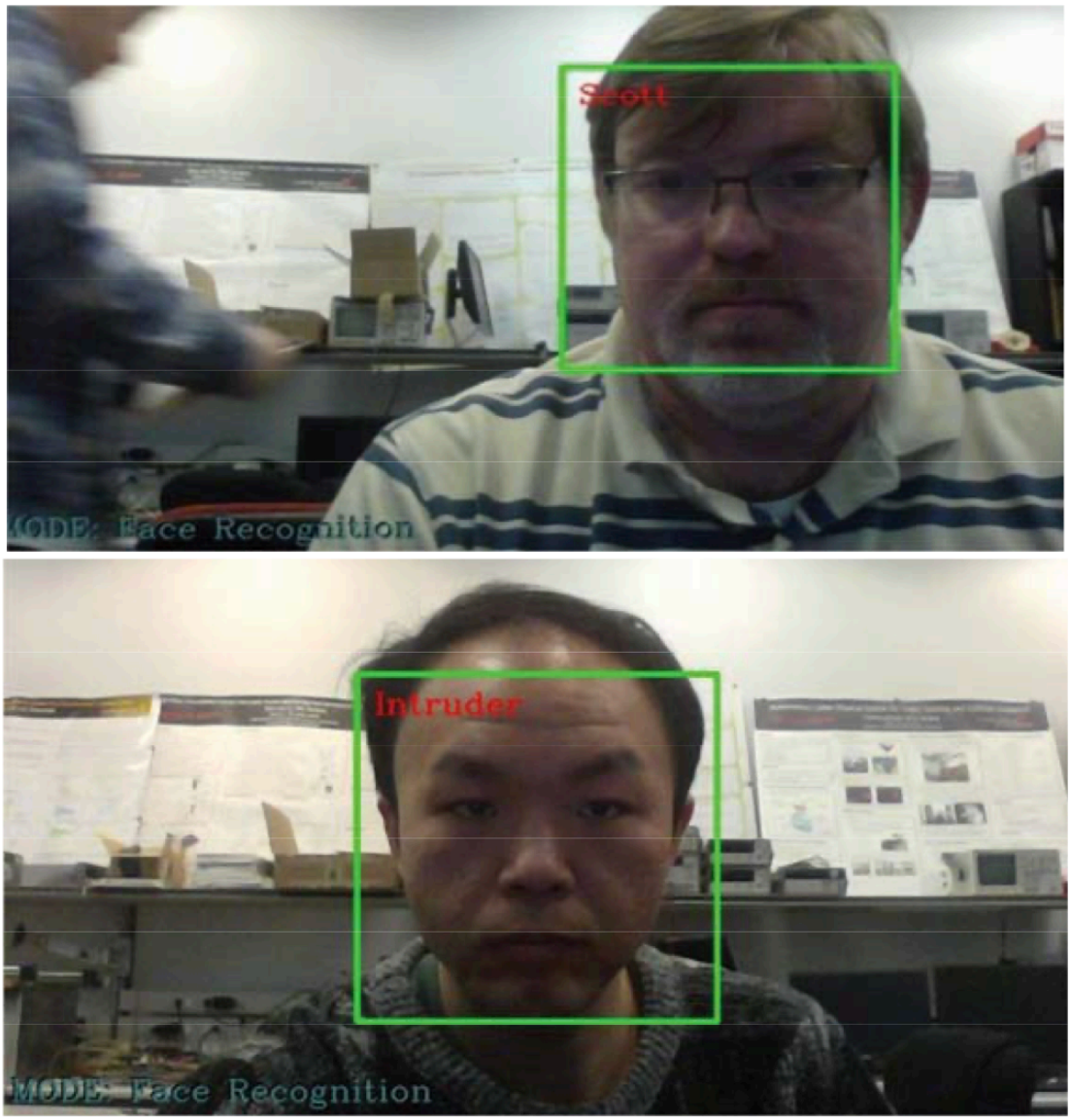


Рис. 1.4 Приклад розпізнавання обличчя людей в системах безпеки за допомогою систем комп’ютерного зору [4]

У робототехніці розпізнавання об'єктів є ключовою технологією для автоматизації завдань, пов'язаних з маніпуляцією об'єктами або навігацією в просторі. Роботи, оснащені камерами та системами комп'ютерного зору, можуть визначати форму, розмір і місце розташування об'єктів у їхньому робочому середовищі. Це дозволяє їм точно виконувати завдання, наприклад, у складських комплексах, на виробничих лініях або в домашніх умовах.

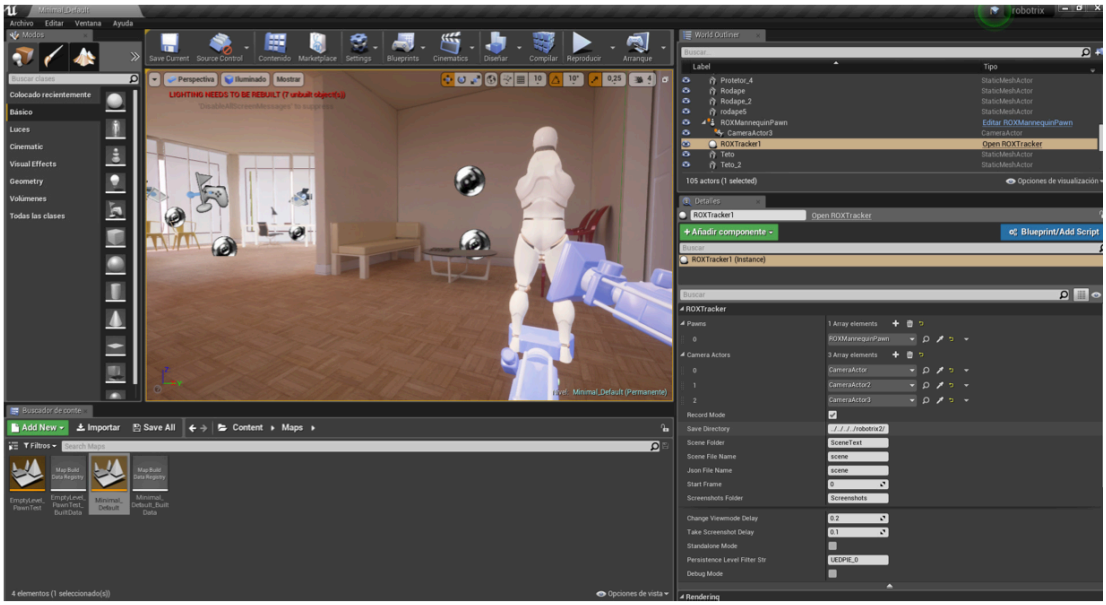


Рис. 1.5 Ілюстративний приклад інтерфейсу взаємодії робота з оточенням через системи комп’ютерного зору [5]

Ще одним яскравим прикладом застосування систем комп'ютерного зору є дрони, які використовуються в різних галузях, від сільського господарства до рятувальних операцій та військових місій. Завдяки технологіям розпізнавання об'єктів, дрони можуть автономно аналізувати навколишнє середовище, виявляти об'єкти інтересу та виконувати завдання з високою точністю. Наприклад, у сільському господарстві дрони використовуються для моніторингу стану полів: вони здатні розпізнавати ділянки з ознаками шкідників, визначати рівень вологості ґрунту чи стан росту рослин. Це дозволяє фермерам своєчасно реагувати на потенційні проблеми та оптимізувати використання ресурсів.

Рис. 1.6 Приклад опрацювання зображень сільськогосподарських угідь з камери дрону за допомогою систем комп’ютерного зору [6]

У рятувальних операціях дрони з комп'ютерним зором застосовуються для пошуку людей у важкодоступних місцях або під час стихійних лих. Вони можуть автоматично виявляти людей серед уламків будівель, у лісах або в гірській місцевості, значно скорочуючи час на пошук та підвищуючи ефективність операцій. Крім того, у поєднанні з тепловізорами дрони здатні знаходити людей уночі або в умовах поганої видимості.

Військові дрони, оснащені системами комп'ютерного зору, можуть автоматично виявляти ворожі об'єкти, транспортні засоби та підозрілі рухи, виконуючи спостереження за великими територіями. Це дозволяє збирати розвідувальні дані, не наражаючи на небезпеку людей, та проводити точні операції з мінімальними втратами.

Комп'ютерний зір також допомагає дронам у навігації, що є особливо важливим у складних середовищах, де наявні перешкоди, такі як дерева, будівлі чи інші об'єкти. Алгоритми розпізнавання дозволяють дронам самостійно уникати зіткнень та знаходити оптимальні маршрути для виконання завдань.

Таким чином, розпізнавання об'єктів на зображеннях є однією з найважливіших складових комп'ютерного зору, що має величезний потенціал для розвитку в різних галузях. Завдяки інтелектуальним системам, здатним ідентифікувати та аналізувати візуальні дані, розпізнавання об'єктів відкриває нові можливості для автоматизації процесів та покращення ефективності рішень, заснованих на аналізі зображень.

Основі принципи та методи розпізнавання об’єктів, зокрема базові та традиційні підходи, заклали фундамент для сучасних технологій комп’ютерного зору. Дані методи розвивалися з часом, від простих алгоритмів до складних систем, що використовують глибокі нейронні мережі. Розуміння цих підходів є важливим для того, щоб оцінити їхні переваги та обмеження, а також зрозуміти, як вони можуть бути інтегровані у сучасні системи комп’ютерного зору.

Розпізнавання об’єктів полягає у визначенні та класифікації об’єктів на зображеннях або відео. Це складний процес, що вимагає аналізу великої кількості даних і залучає різноманітні алгоритми та методи. Цей процес можна розділити на кілька ключових етапів.

Етап попередньої обробки зображень, що включає покращення якості зображення, нормалізацію розміру, фільтрацію шуму та інші методи, що дозволяють покращити характеристики зображення для подальшого аналізу. Попередня обробка зображення є важливою, оскільки вона забезпечує усунення різних артефактів, які можуть заважати точному розпізнаванню об'єктів. Це може включати:

- Фільтрація шуму: використання різних фільтрів (наприклад, гауссового або медіанного) для видалення шуму.

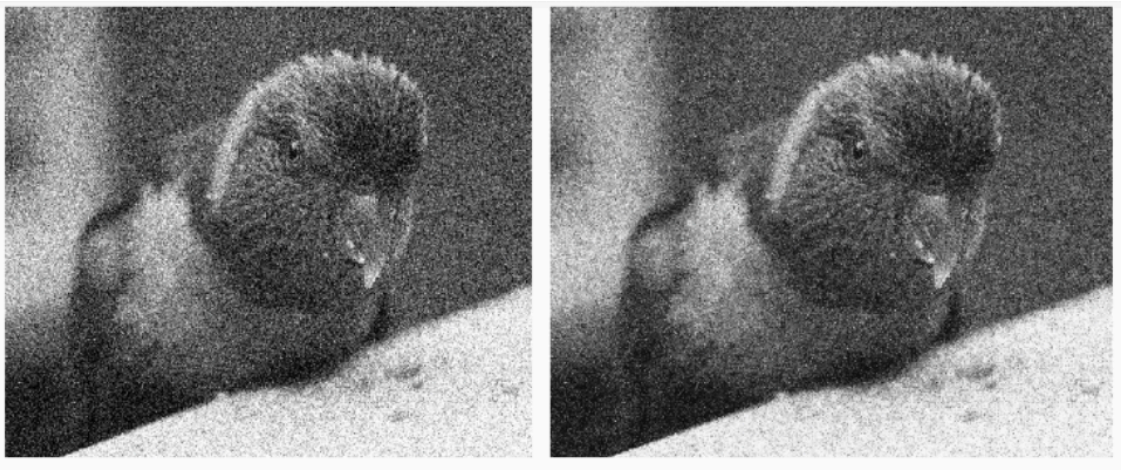


Рис. 1.7 Приклад фільтрації шуму на зображенні [7]

- Корекцію освітлення: вирівнювання яскравості та контрасту для забезпечення рівномірного освітлення на всьому зображенні.

- Нормалізацію розміру: зміна розміру зображення до стандартного, що полегшує подальшу обробку.

- Видалення фону: ізоляція об’єктів від фону для кращого фокусування на самих об’єктах.

Рис. 1.8 Приклад зображень з видаленим фоном за допомогою алгоритмів комп’ютерного зору [8]

Етап виділення ознак включає в себе визначення ключових характеристик зображення, які є інформативними для розпізнавання об’єктів. Це можуть бути краї, кути, текстури та інші важливі особливості. Виділення ознак є критичним

етапом, оскільки саме ці ознаки використовуються для подальшої класифікації. Основні методи включають:

- Детектори країв: такі як алгоритм Кенні, Собеля, що виділяють контури об'єктів.

Рис. 1.9 Застосування алгоритму Кенні для виявлення контурів зображення [9]

- Локальні ознаки: методи, що визначають ключові точки на зображенні (наприклад, SIFT та SURF).

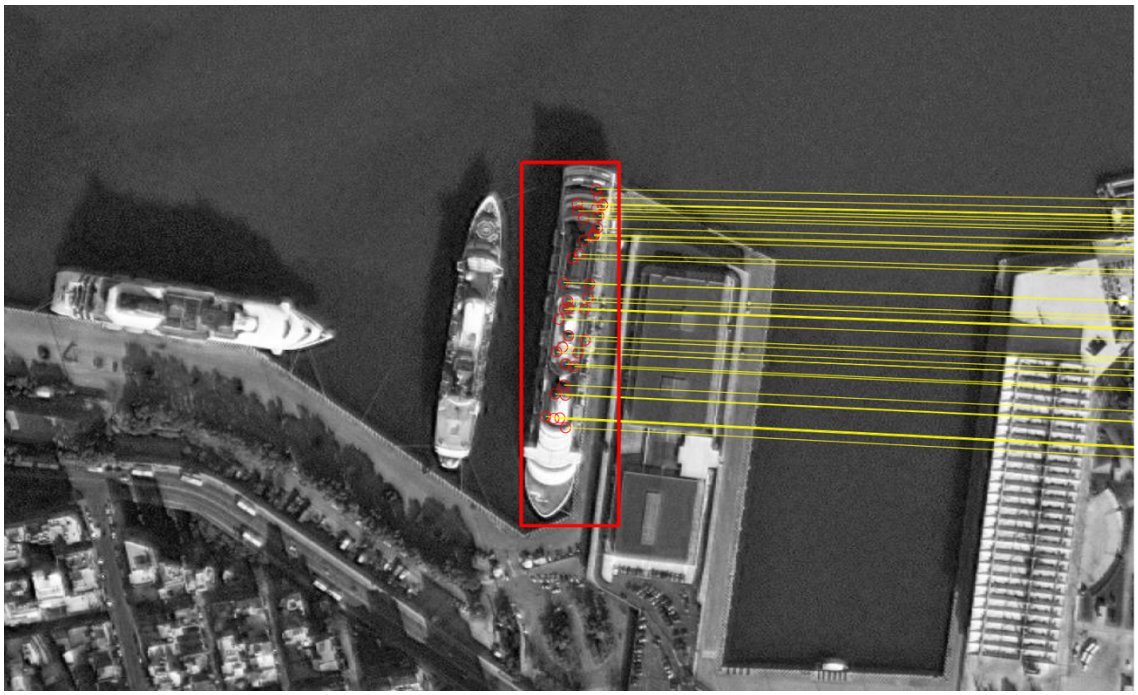


Рис. 1.10 Приклад застосування методу SIFT [10]

- Текстурні ознаки: аналіз текстур для виділення однорідних або структурованих областей.

Після виділення ознак наступним кроком є класифікація об’єктів за допомогою моделей машинного навчання або класичних алгоритмів класифікації. Це включає застосування різних технік для розпізнавання об'єктів на основі виділених ознак. Основні підходи включають:

- Моделі машинного навчання: такі як Support Vector Machine (SVM), Decision Tree, Random Forest.

- Глибокі нейронні мережі [89] (Deep Neural Networks - DNN): здатні автоматично виділяти ознаки та класифікувати об'єкти з високою точністю за допомогою штучних нейронних мереж.

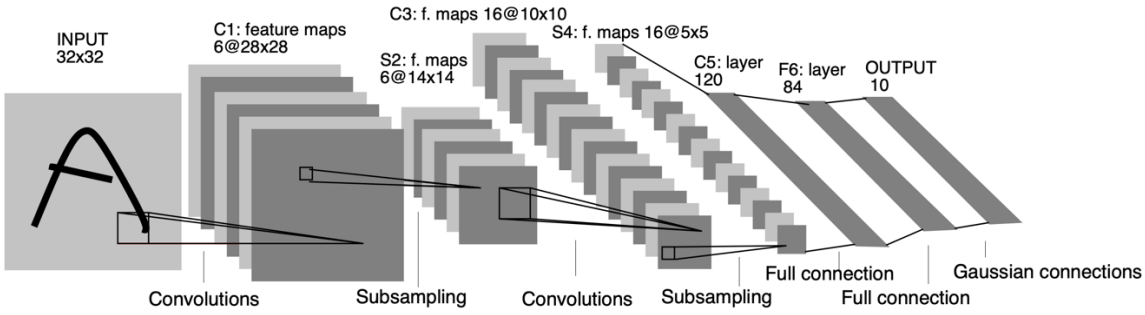


Рис. 1.11 Архітектура згорткової нейронної мережі LeNet для задачі розпізнавання рукописного тексту [11]

- Ансамблеві методи: комбінація кількох моделей для покращення точності класифікації.

Заключний етап включає фільтрацію та інтерпретацію отриманих результатів для покращення точності розпізнавання та зменшення кількості хибних позитивних результатів. Після оброблення результатів сприяє точнішій ідентифікації об'єктів та зменшенню кількості помилок. Основні кроки включають:

- Фільтрація результатів: виключення неправильно класифікованих об’єктів або об’єктів з низькою ймовірністю.

- Комбінація з іншими даними: інтеграція результатів з іншими джерелами інформації (наприклад, даними з інших сенсорів).

- Візуалізація: представлення результатів у зручному для користувача вигляді для подальшого аналізу та прийняття рішень.

Таким чином, процес розпізнавання об’єктів на зображеннях є багатоступеневим і вимагає ретельного виконання кожного етапу для досягнення високої точності та надійності результатів.

Базові методи розпізнавання об’єктів спираються на класичні алгоритми комп’ютерного зору та машинного навчання, які були розроблені ще до появи сучасних глибоких нейронних мереж. Ці методи заклали основу для багатьох сучасних технологій розпізнавання і залишаються важливими для розуміння основних принципів цієї галузі.

Один з базових методів – це шаблонний метод, що є одним з найпростіших і найперших підходів до розпізнавання об’єктів. Він полягає у порівнянні фрагментів зображення із заздалегідь встановленими шаблонами об'єктів. Цей метод працює наступним чином:

1. Підготовка шаблону: заздалегідь створюються шаблони об'єктів, які система повинна розпізнавати.

2. Порівняння: кожен фрагмент зображення порівнюється з шаблонами за допомогою певної метрики.

3. Визначення збігів: збіг вважається успішним, якщо схожість між шаблоном та фрагментом зображення перевищує певний поріг.

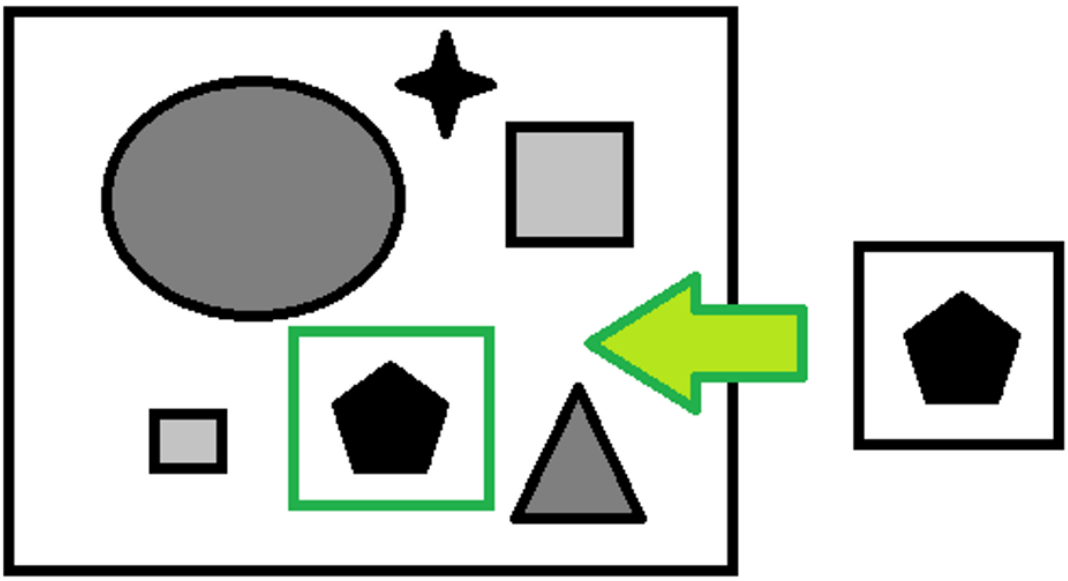


Рис. 1.12 Приклад знаходження об’єкту на зображенні за допомогою шаблонного підходу [12]

Метод шаблонів є досить простим у реалізації, але має значні обмеження. Він реагує на зміни масштабу, повороту, освітлення та інші спотворення зображення, що знижує його ефективність у складних умовах.

Наступним є метод аналізу ознак, що використовує детектори ознак для виділення ключових точок на зображенні та побудови дескрипторів об’єктів. До найвідоміших методів цього типу належать:

- SIFT (Scale-Invariant Feature Transform): виділяє ключові точки та створює їхні масштабно-незалежні дескриптори, що дозволяє розпізнавати об'єкти навіть при зміні масштабу та повороті.

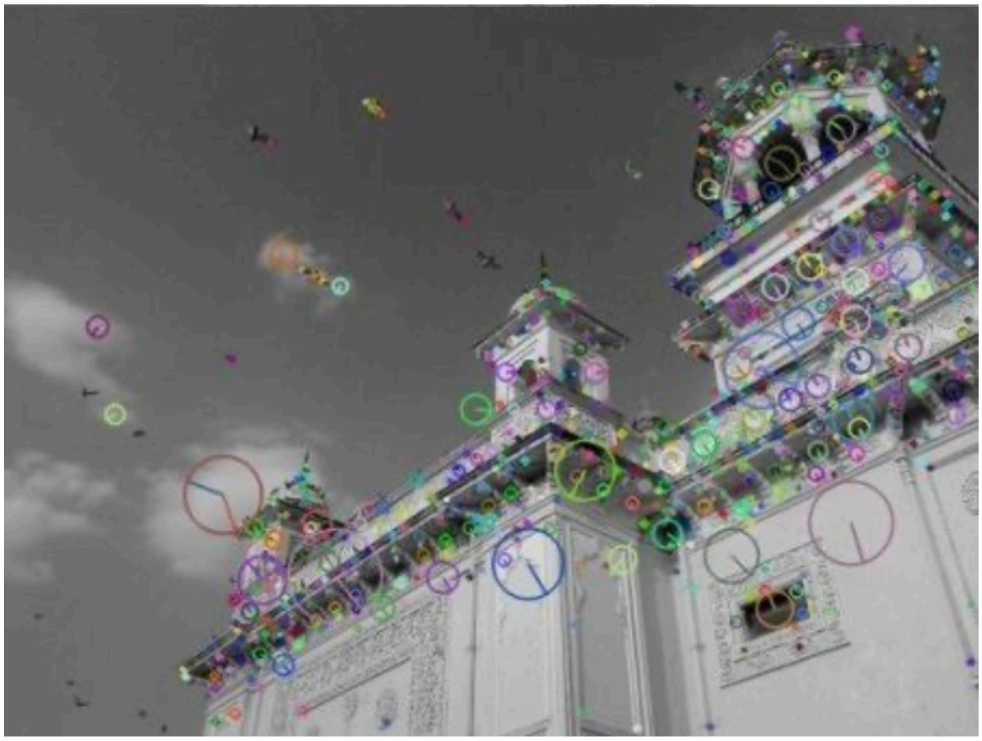


Рис. 1.13 Приклад знаходження ключових точок за допомогою SIFT алгоритму [13]

- SURF (Speeded-Up Robust Features): прискорена версія SIFT, що забезпечує високу швидкість та стійкість до різних спотворень.

Рис. 1.14 Приклад роботи SURF алгоритму [14]

Дані методи дозволяють розпізнавати об’єкти навіть при значних змінах у вигляді зображення та забезпечують достатню точність у базових задачах комп’ютерного зору.

Окрім наведених вище, існує ще один важливий базовий метод розпізнавання об'єктів – це підхід, що використовує гістограми орієнтованих градієнтів (Histogram of Oriented Gradients, HOG). Основні кроки цього методу складаються з:

1. Розрахунок градієнтів: для кожного пікселя зображення обчислюються градієнти яскравості.

2. Створення гістограм: градієнти об’єднуються у гістограми на основі їх спрямування в локальних областях зображення.

3. Нормалізація блоків: гістограми нормалізуються для зменшення впливу змін освітлення та контрасту.

4. Формування ознак: нормалізовані гістограми об'єднуються в єдиний вектор ознак, який використовується для класифікації.

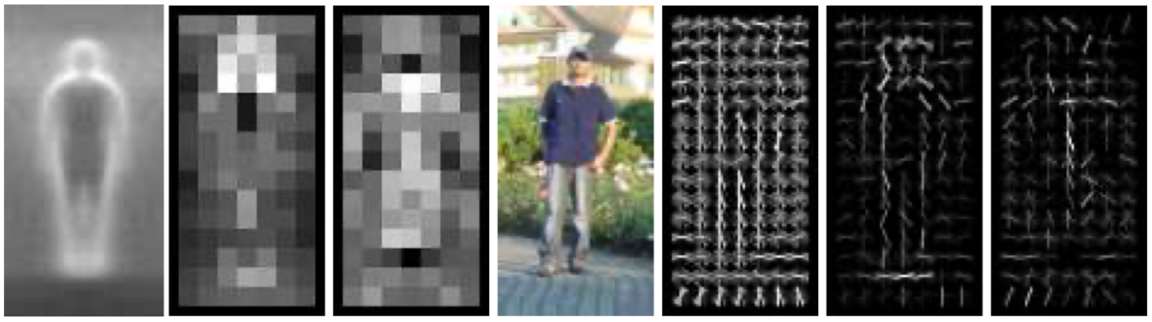


Рис. 1.15 Приклад використання методу HOG для знаходження людини на зображені [15]

Даний метод використовує гістограми градієнтів спрямування для виділення структурних характеристик об’єктів та їх контурів. HOG забезпечує високу точність при розпізнаванні об’єктів в нескладних задачах комп’ютерного зору.

Класичні методи комп’ютерного зору, які базуються на виділенні ознак та використанні традиційних алгоритмів, довгий час були основою в розпізнаванні об'єктів на зображеннях. Методи, такі як SIFT, SURF, HOG та інші, дозволяли виділяти інформативні характеристики зображень, які потім використовувалися в поєднанні з класифікаторами машинного навчання. Однак з розвитком технологій і зростанням обсягів даних ці методи почали демонструвати свої обмеження.

По-перше, обмежена точність класичних методів стає помітною при роботі з великими та складними наборами даних. Виділення ознак не завжди може захопити всі нюанси та варіації об'єктів, особливо в умовах зміни освітлення, масштабу, повороту або часткової оклюзії. Це призводить до зниження точності розпізнавання та збільшує кількість хибних спрацьовувань.

По-друге, проблеми зі швидкодією виникають при застосуванні класичних методів у реальному часі. Деякі алгоритми виділення ознак, такі як SIFT та SURF, є обчислювально складними і не підходять для застосувань, де потрібна миттєва обробка зображень, наприклад, в автономних транспортних засобах або системах відеоспостереження. Це обмежує їх практичне використання в сучасних додатках, де швидкість є критичним фактором.

Крім того, відсутність масштабованості класичних методів стає суттєвою перепоною в епоху великих даних. Зі збільшенням розмірів наборів даних та складності завдань традиційні алгоритми не можуть ефективно обробляти інформацію без значного збільшення обчислювальних ресурсів. Це робить їх менш привабливими порівняно з сучасними підходами, які можуть навчатися на великих обсягах даних більш ефективно.

Нарешті, недостатня адаптивність класичних методів обмежує їх застосування в динамічних середовищах. Ручне налаштування параметрів та вибір ознак потребує значних зусиль та експертних знань, що ускладнює адаптацію моделей до нових умов або завдань. У сучасних вимогах до систем розпізнавання об'єктів потрібна гнучкість та здатність до самонавчання, що не завжди можливе з традиційними методами.

У зв'язку з цим виникає потреба в інноваційних підходах, які можуть подолати ці обмеження. Глибоке навчання та згорткові нейронні мережі пропонують рішення, здатні автоматично вивчати складні ознаки з даних, забезпечуючи високу точність та швидкість обробки. Це відповідає актуальним вимогам до систем розпізнавання, створюючи нові перспективи для їх використання в різних сферах.

1.2 Аналіз відомих інтелектуальних методів та технологій розпізнавання об’єктів на зображеннях

Розпізнавання об’єктів на зображеннях є складною та багатогранною задачею, що вимагає використання передових інтелектуальних методів та технологій. Глибоке навчання та нейронні мережі стали ключовими інтелектуальними методами та технологіями у вирішенні сучасних завдань в комп'ютерному зорі, кардинально змінивши підхід до розпізнавання об'єктів на зображеннях. На відміну від класичних методів, які покладалися на прості алгоритми виділення ознак та прості моделі класифікації, ці інтелектуальні системи здатні автоматично вивчати багаторівневі представлення даних. Це дозволяє їм захоплювати складні шаблони та взаємозв'язки в даних, недоступні для традиційних алгоритмів.

Однією з головних переваг глибоких нейронних мереж є їх висока точність у задачах розпізнавання. Завдяки великим обсягам навчальних даних та потужним обчислювальним ресурсам, такі мережі можуть навчатися розрізняти тонкі відмінності між класами об'єктів, бути стійкими до шумів та варіацій у зображеннях. Це особливо важливо в сучасних застосуваннях, де інтелектуальні методи та технології повинні забезпечувати найвищий рівень якості розпізнавання.

Крім того, глибокі нейронні мережі забезпечують високу швидкодію, що досягається через оптимізацію архітектури та використання апаратного прискорення, такого як GPU та TPU.

Глибокі нейронні мережі є основою сучасних технологій розпізнавання об’єктів на зображеннях. Ці мережі складаються з великої кількості шарів, що дозволяє їм вивчати складні шаблони в даних та здійснювати високоточну класифікацію об’єктів. Розглянемо детально найвідоміші архітектури глибоких нейронних мереж, які значно сприяли прогресу в галузі комп'ютерного зору.

Згорткові нейронні мережі [89] (Convolutional Neural Networks - CNN) є однією з найпоширеніших і найефективніших архітектур для обробки зображень. Дані нейронні мережі використовують спеціальні шари, що дозволяють автоматично виділяти ознаки та здійснювати класифікацію. Основні компоненти CNN включають:

1. Згорткові шари: використовують фільтри (ядра), які проходять по зображенню, виділяючи різні ознаки, такі як краї, кути, текстури тощо. Ці фільтри навчаються автоматично під час процесу навчання мережі.

2. Шари підсумовування (pooling layers): зменшують розмірність даних, зберігаючи найважливіші ознаки. Найпоширеніші типи шарів підсумовування - це max pooling і average pooling.

3. Повнозв’язні шари (fully connected layers): після кількох згорткових і шарів підсумовування, дані передаються через один або більше повнозв’язних шарів для остаточної класифікації.

Приклади найбільш відомих архітектур CNN, що використовуються для розпізнавання об’єктів на зображеннях:

- AlexNet: перша глибока CNN, яка виграла змагання ImageNet у 2012 році, демонструючи значний прогрес у розпізнаванні об’єктів. Вона складається з п'яти згорткових шарів, за якими слідують три повнозв’язні шари, також використовується техніка Dropout та нормалізація за партіями [16].

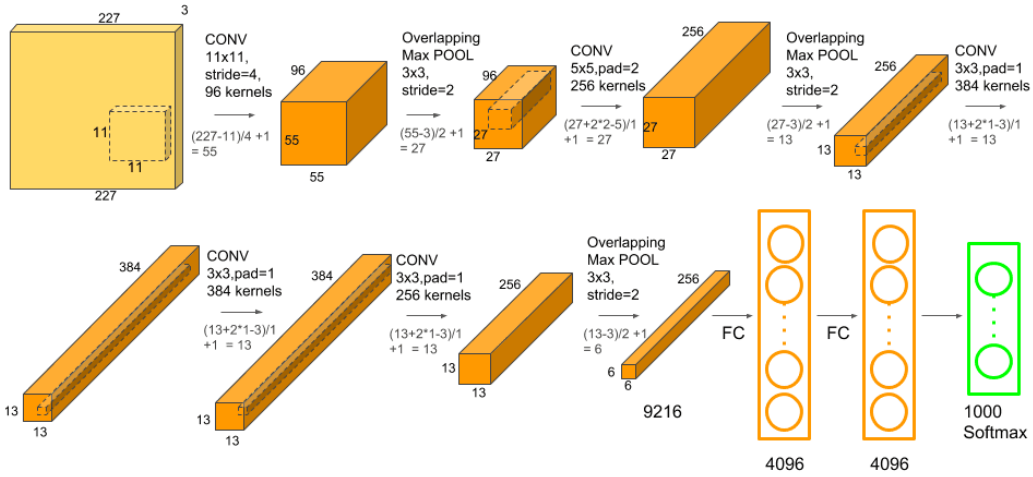


Рис. 1.16 Архітектура AlexNet

- VGG: архітектура, запропонована дослідниками з Оксфордського університету, використовує дуже глибокі мережі з простими ядрами згортки розміру 3x3. Існують різні варіанти VGG, такі як VGG16 і VGG19, які мають відповідно 16 і 19 шарів [17].

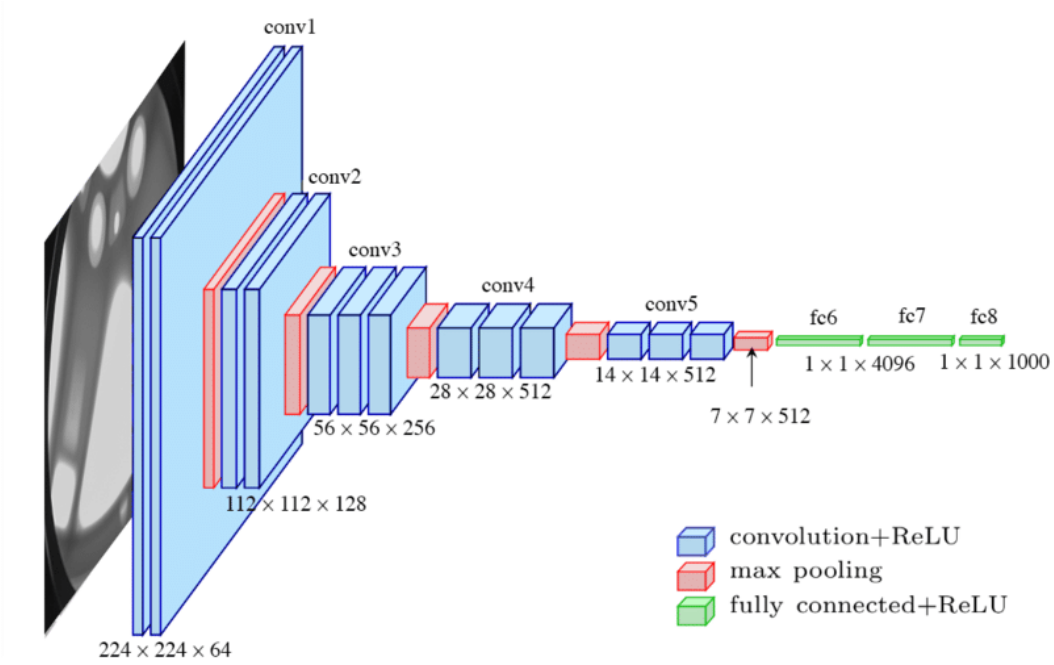


Рис. 1.17 Архітектура VGG

- ResNet: архітектура розроблена компанією Microsoft, ця архітектура використовує залишкові блоки (residual blocks), що дозволяє будувати дуже глибокі мережі без проблем із зникненням градієнтів [18]. ResNet50, ResNet101 і ResNet152 є популярними варіантами цієї архітектури.

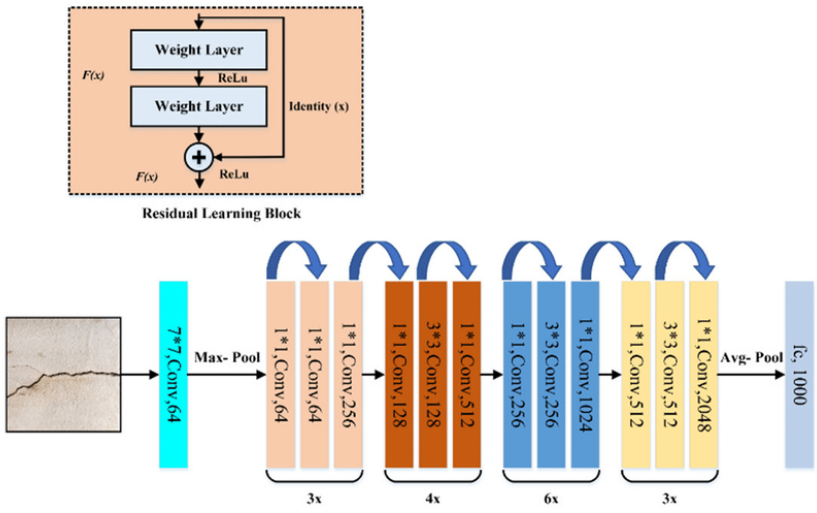


Рис. 1.18 Архітектура ResNet

R-CNN є спеціалізованими моделями CNN, які поєднують класифікацію та локалізацію об’єктів. Вони були розроблені для задач виявлення об’єктів і включають кілька вдосконалень. По-перше, R-CNN використовує регіональні пропозиції для визначення потенційних областей, де можуть знаходитися об'єкти, що значно зменшує обчислювальну складність порівняно з повним перебором зображення. По-друге, вони застосовують глибоке навчання для вилучення високорівневих ознак з цих областей, що підвищує точність як класифікації, так і локалізації об'єктів на зображенні. Існують наступні модифікації R-CNN:

- R-CNN: найперший підхід, який використовує методи вибору регіонів (selective search) для генерації регіонів пропозицій, які потім обробляються CNN для класифікації та локалізації [19].

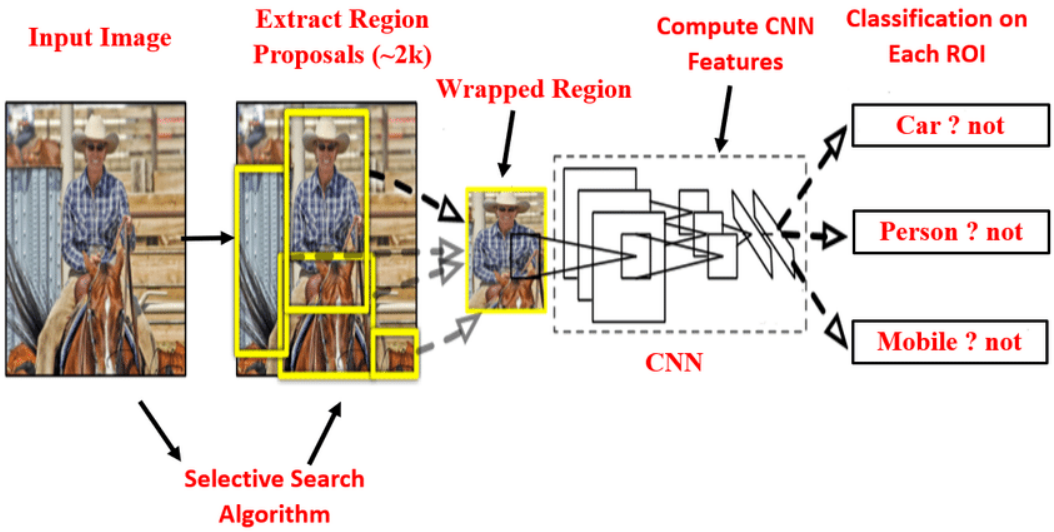


Рис. 1.19 Архітектура RCNN

- Fast R-CNN: оптимізована версія, яка обробляє все зображення одночасно, використовуючи спільні ознаки для регіонів пропозицій [20]. Це значно збільшує швидкість обробки та точність.

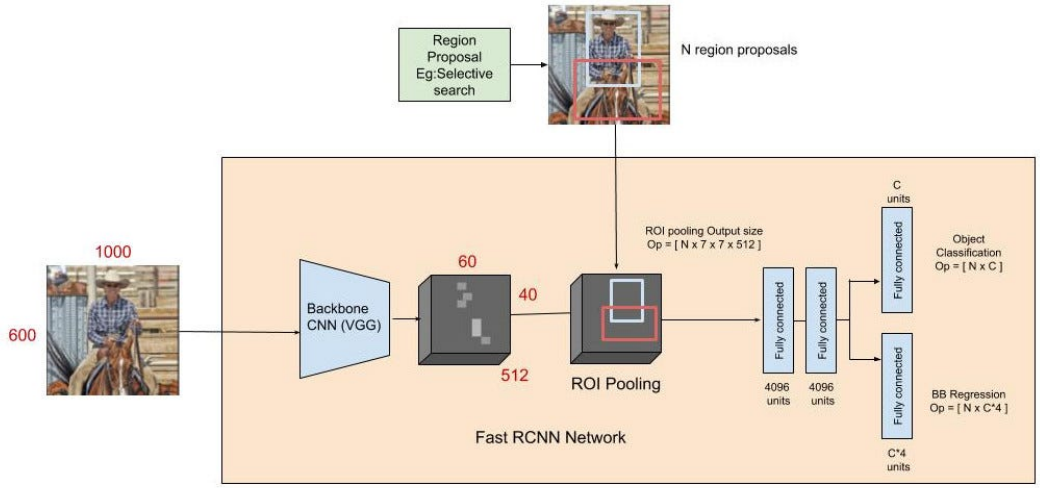


Рис. 1.20 Архітектура Fast R-CNN

- Faster R-CNN: ще більш ефективна модель, яка інтегрує мережу регіональних пропозицій (Region Proposal Network) для генерації регіонів у режимі реального часу, значно зменшуючи час обробки [21].

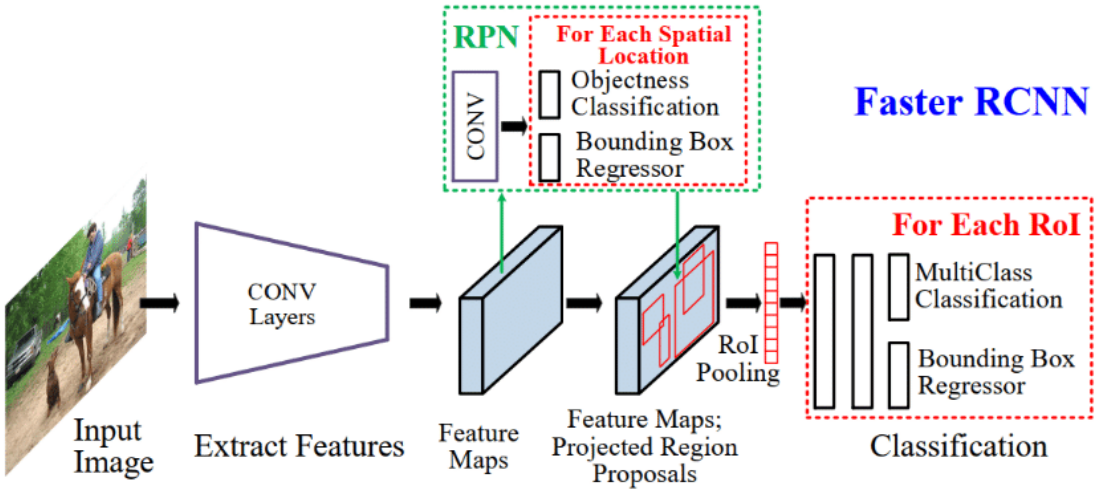


Рис. 1.21 Архітектура Faster R-CNN

Одним із найбільш відомих та ефективних підходів до виявлення об'єктів у режимі реального часу є модель YOLO. Ця інтелектуальна технологія стала популярною завдяки своїй здатності поєднувати високу точність розпізнавання з надзвичайною швидкістю обробки. YOLO революціонізувала підхід до задач детекції, пропонуючи єдину консолідовану нейронну мережу, яка одночасно прогнозує межі об'єктів та їх класи [22].

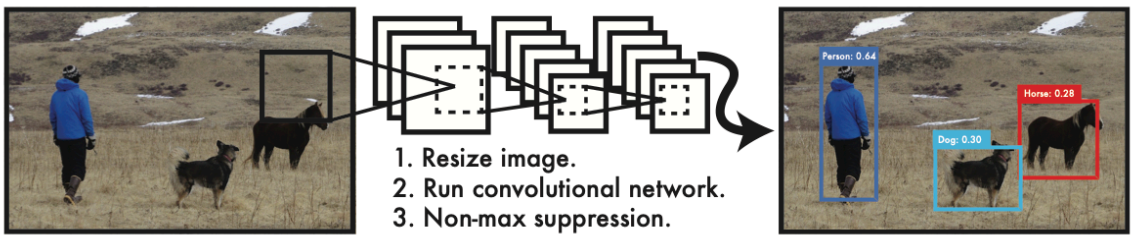


Рис. 1.22 Архітектура YOLO

Основні особливості YOLO включають в себе наступне:

- Одноетапне виявлення: YOLO обробляє зображення за один прохід, що забезпечує високу швидкість виявлення об’єктів.

- Глобальний контекст: YOLO використовує всю інформацію про зображення для точнішого виявлення об’єктів, що дозволяє зменшити кількість помилкових спрацювань.

Деякі відомі варіанти YOLO включають:

- YOLOv1: початкова версія моделі, представлена у 2016 році. Розбиває зображення на сітку та передбачає об’єкти в кожній клітинці [22]. Відзначається високою швидкістю, але мала проблеми з точністю для малих об’єктів і складних сцен.

- YOLOv2 (YOLO9000): представлена у 2017 році, ця версія включає декілька покращень для підвищення точності та швидкодії. Впроваджує методи Batch Normalization, використання високороздільних зображень та механізм anchor boxes для кращої локалізації. Підтримує розпізнавання понад 9000 класів об’єктів завдяки спільному навчанні з ImageNet і COCO [23].

- YOLOv3: випущена у 2018 році, ця версія покращує точність та швидкість виявлення. Використовує нову архітектуру Darknet-53 з більшою глибиною та багаторівневе передбачення, що покращує виявлення об’єктів різних розмірів. Підвищує стійкість до змін масштабу та складних фонових умов [24].

- YOLOv4: представлена у 2020 році, ця версія включає сучасні техніки для покращення продуктивності, такі як Cross Stage Partial connections (CSP), Weighted Residual Connections (WRC) та Mish activation function. Оптимізована для роботи в реальному часі з високою точністю [25]

- YOLOv5: випущена у 2020 році компанією Ultralytics. Відзначається легкою вагою та високою швидкістю тренування та інференсу. Підтримує різні розміри моделей (s, m, l, x) для різних застосувань та включає вбудовані функції доповнення даних і оптимізації [26].

- YOLOv6: розроблена компанією Meituan у 2022 році, орієнтована на промислове застосування. Оптимізована для швидкого виявлення об’єктів з високою продуктивністю на GPU та ARM пристроях. Включає ефективні архітектурні блоки та методи навчання [27].

- YOLOv7: випущена у 2022 році, встановила новий стандарт точності та швидкодії для моделей виявлення об’єктів у реальному часі. Включає інноваційні методи оптимізації архітектури, такі як Extended Efficient Layer Aggregation

Networks, що підвищує ефективність без збільшення обчислювальної складності [28].

- YOLOv8: представлена у 2023 році компанією Ultralytics. Включає подальші покращення в архітектурі та алгоритмах навчання, забезпечуючи ще вищу точність та швидкодію [29].

- YOLOX: відкрита у 2021 році, використовує підхід без використання anchor boxes (anchor-free) та сучасні стратегії призначення міток. Це покращує продуктивність і спрощує процес навчання [30].

Кожна нова версія YOLO приносить покращення в точності, швидкодії та ефективності, адаптуючись до сучасних вимог та викликів комп’ютерного зору. Ці інтелектуальні методи та технології широко використовуються в різних галузях, від автономного транспорту до систем безпеки та медичної діагностики, завдяки їхній здатності забезпечувати високу продуктивність у задачах виявлення об’єктів.

Окрім цього, сегментація зображень є одним із інтелектуальних методів для вирішення ключових завдань комп'ютерного зору, яке полягає у поділі зображення на семантично значущі області. З появою глибоких нейронних мереж були розроблені численні архітектури, що значно покращили точність та ефективність сегментації. Найбільш популярними є:

- Fully Convolutional Networks (FCN): однією з перших успішних архітектур для семантичної сегментації стала FCN, запропонована у 2015 році [31].

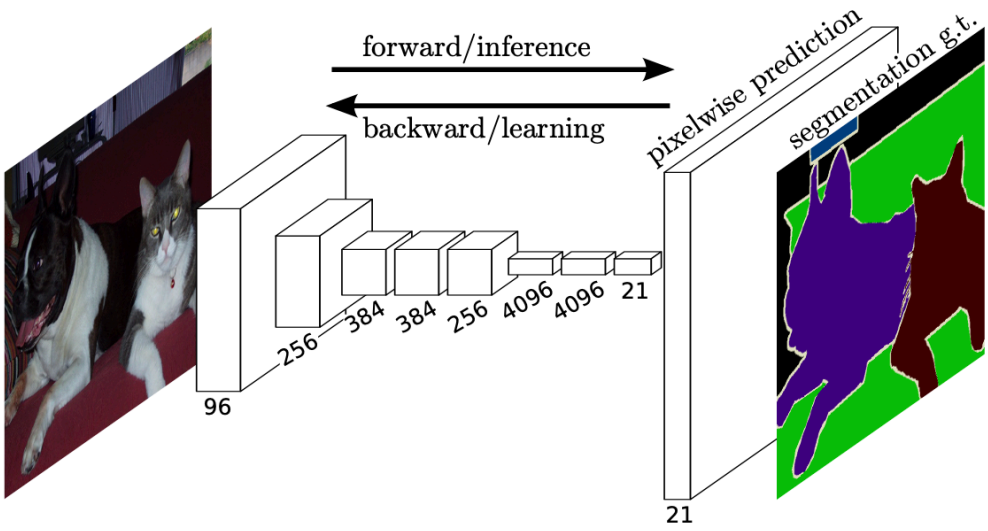


Рис. 1.23 Архітектура сегментації FCN

FCN замінюють повнозв'язні шари на згорткові, що дозволяє мережі приймати вхідні зображення будь-якого розміру та генерувати карти ознак відповідних розмірів. Це сприяло значному підвищенню ефективності та гнучкості сегментації.

- U-Net: архітектура запропонована у 2015 році, стала стандартом для сегментації медичних зображень [32].

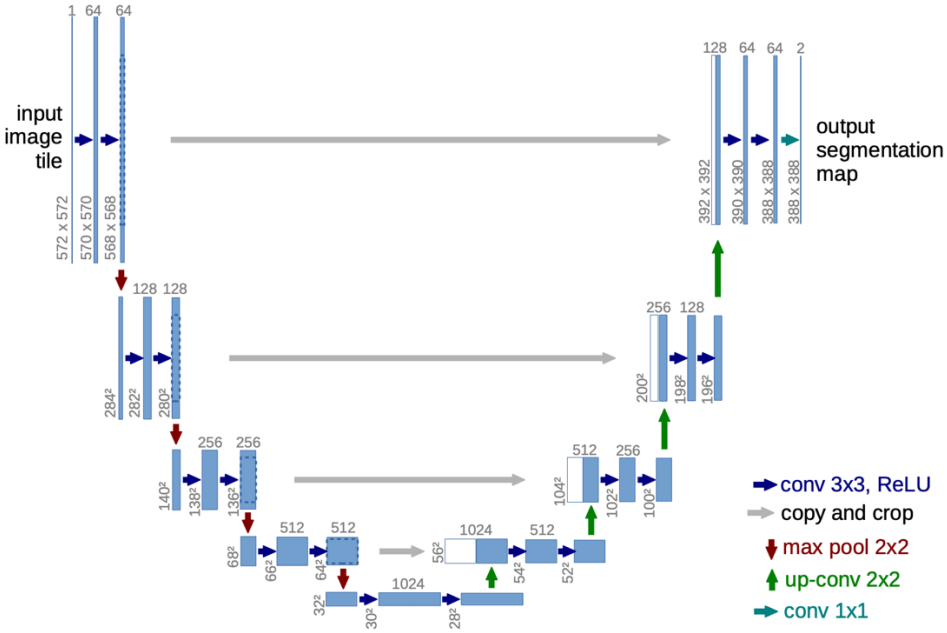


Рис. 1.24 Архітектура U-Net

Архітектура має симетричну U-подібну структуру з шляхами енкодера та декодера. Вона використовує skip-connections між відповідними шарами енкодера та декодера, що забезпечує точне відновлення просторової інформації. U-Net демонструє високу продуктивність навіть на невеликих наборах даних.

- Mask R-CNN: дана архітектура розширює архітектуру Faster R-CNN для задачі сегментації на рівні об'єктів [33].

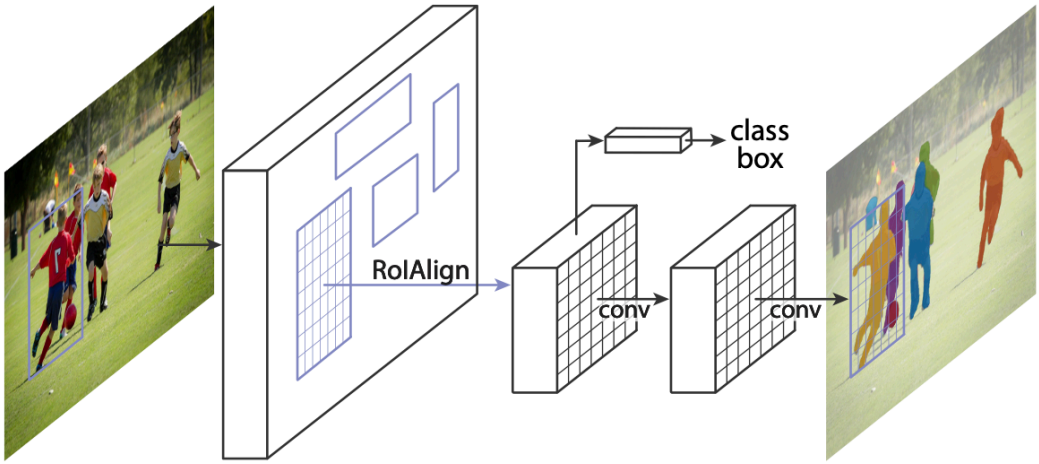


Рис. 1.25 Архітектура Mask R-CNN

Модель додає окрему гілку для передбачення маски кожного об'єкта, що дозволяє одночасно виконувати детекцію та сегментацію. Mask R-CNN показала високі результати в задачах сегментації.

- DeepLab [86], [87], [88]: серія моделей DeepLab впроваджує атрозні згортки (dilated convolutions) для розширення поля зору мережі без збільшення кількості параметрів.

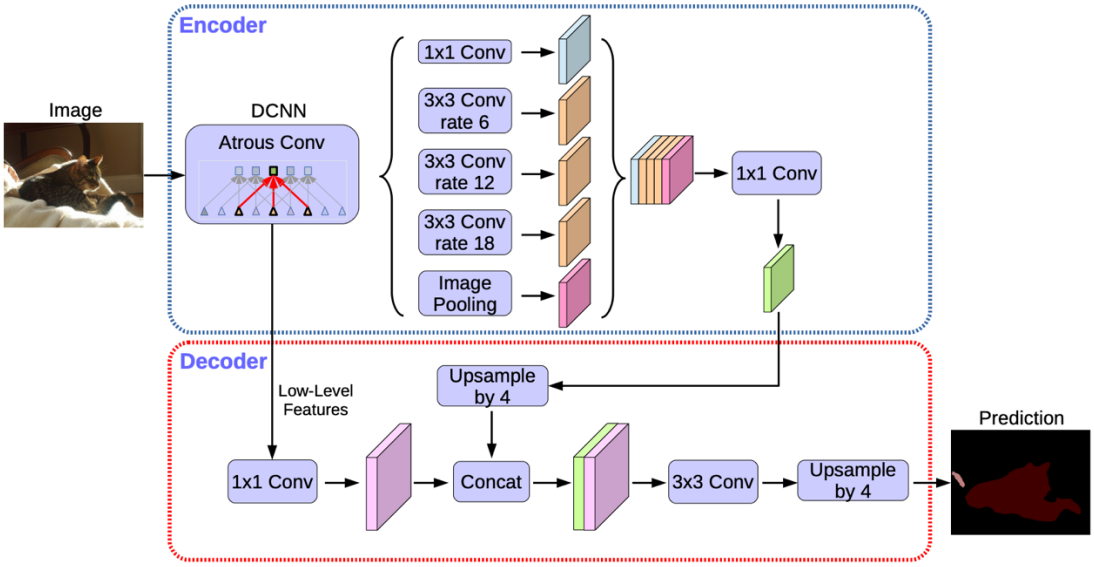


Рис. 1.26 Архітектура DeepLab

DeepLabv3+ [88] об'єднує переваги енкодер-декодерних архітектур та використовує атрозний просторовий пірамідальний пулінг (ASPP) для захоплення контекстної інформації на різних масштабах. Це дозволяє досягти високої точності сегментації на складних наборах даних.

Розвиток архітектур сегментації зображень значно вплинув на можливості розпізнавання об'єктів у комп'ютерному зорі. Кожна з розглянутих моделей пропонує унікальні підходи та рішення, спрямовані на покращення точності, ефективності та здатності моделі захоплювати контекстну інформацію. Ці досягнення створюють основу для подальших досліджень та впровадження інтелектуальних систем розпізнавання об'єктів у різних галузях.

Архітектури глибоких нейронних мереж є ключовими компонентами сучасних інтелектуальних технологій розпізнавання об’єктів на зображеннях. Від згорткових нейронних мереж до спеціалізованих моделей, таких як YOLO та Mask R-CNN, кожна з них має свої унікальні особливості та переваги. Розуміння цих архітектур дозволяє ефективно вирішувати складні задачі комп’ютерного зору та створювати високоточні системи для різних застосувань.

Розпізнавання об’єктів на зображеннях є ключовою задачею у сфері комп’ютерного зору, яка має безліч застосувань у різних галузях, від

автомобільної промисловості до медицини. Протягом останніх десятиліть було розроблено багато методів та технологій, які дозволили значно покращити точність і швидкість розпізнавання об’єктів.

1.3 Огляд методу розпізнавання об’єктів на зображеннях з використанням псевдомасок

У сучасному комп'ютерному зорі значна увага приділяється методам розпізнавання об'єктів, які можуть ефективно працювати з обмеженими анотаціями або навіть без них. Традиційні підходи до навчання моделей глибокого навчання вимагають великих обсягів детально анотованих даних, що включають точні маски об'єктів для задач сегментації. Однак процес створення таких анотацій є трудомістким, дорогим та часто непрактичним для великих або складних наборів даних.

Розпізнавання об’єктів на зображеннях є важливою задачею комп’ютерного зору, що має безліч застосувань, від автономних транспортних засобів до медичної діагностики. Одним із перспективних підходів до цієї задачі є використання псевдомасок, які дозволяють покращити точність розпізнавання об’єктів, зокрема в умовах обмежених даних для навчання. У цьому підрозділі ми розглянемо, що таке псевдомаски, а також декілька методів, що їх використовують для розпізнавання об’єктів.

Псевдомаски (pseudo-masks) - це попередньо згенеровані маски об’єктів, які використовуються для навчання моделей сегментації або розпізнавання об’єктів у ситуаціях, коли недостатньо ручних розміток. Псевдомаски дозволяють використовувати неанотовані або слабко анотовані дані, розширюючи таким чином обсяг доступної інформації для навчання моделей [75], [76]. Це особливо важливо в умовах, коли доступ до анотованих даних обмежений або їх створення є надто затратним. Ідея полягає у тому, що навіть неповні або частково правильні маски можуть значно покращити результати навчання моделі.

Генерація псевдомасок може здійснюватися за допомогою різних методів. Слабко контрольоване навчання використовує доступні слабкі анотації, такі як мітки класів або рамки навколо об'єктів, для створення масок за допомогою алгоритмів, що визначають найбільш ймовірні області об'єктів. Наприклад, методи на основі Class Activation Mapping (CAM) [34] дозволяють ідентифікувати області зображення, що найбільше впливають на рішення моделі, і використовувати цю інформацію для генерації псевдомасок.

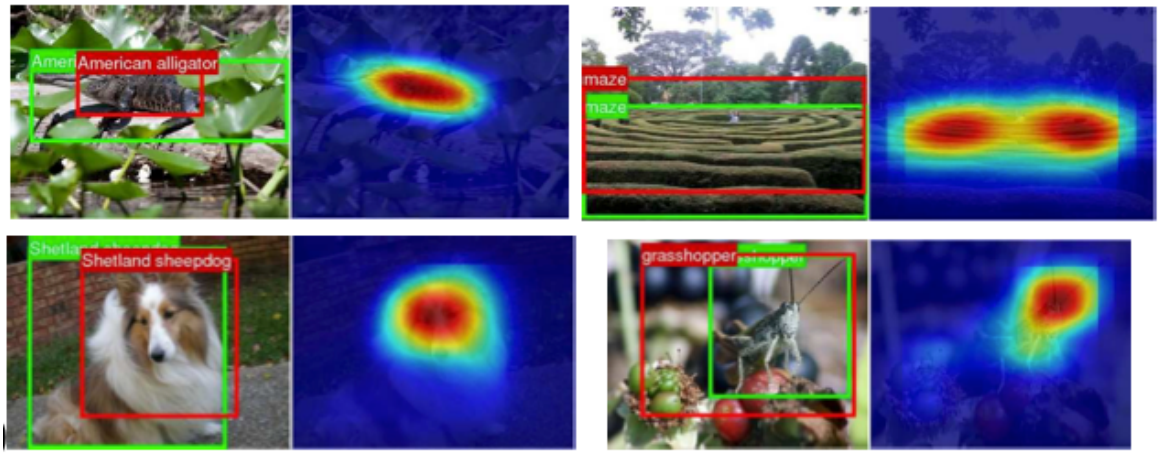


Рис. 1.27 Використання методу Class Activation Mapping для генерації псевдомасок

Подальші розширення, такі як Grad-CAM [35] та Grad-CAM++ [36], покращують точність локалізації та дозволяють працювати з більш складними архітектурами нейронних мереж.

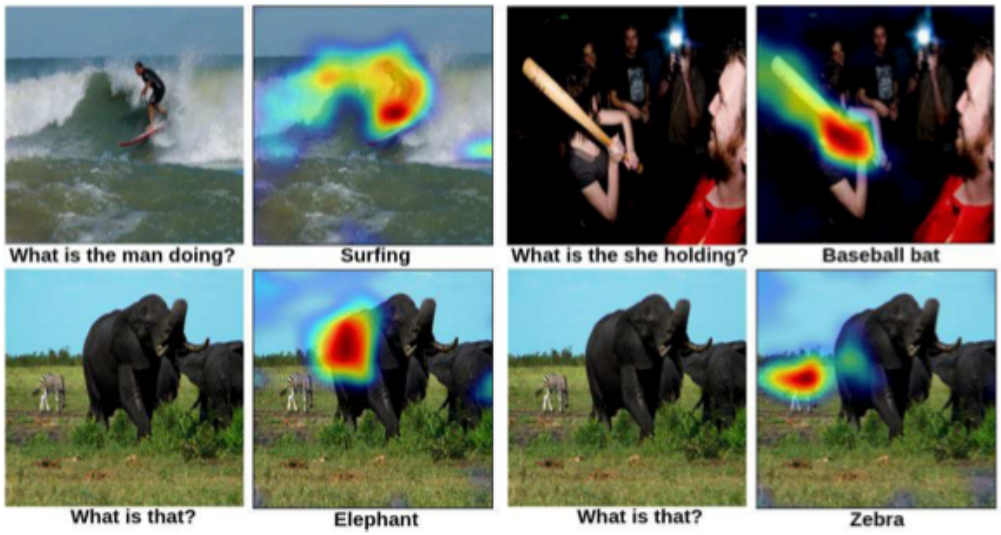


Рис. 1.28 Використання методу Grad-CAM для створення псевдомасок

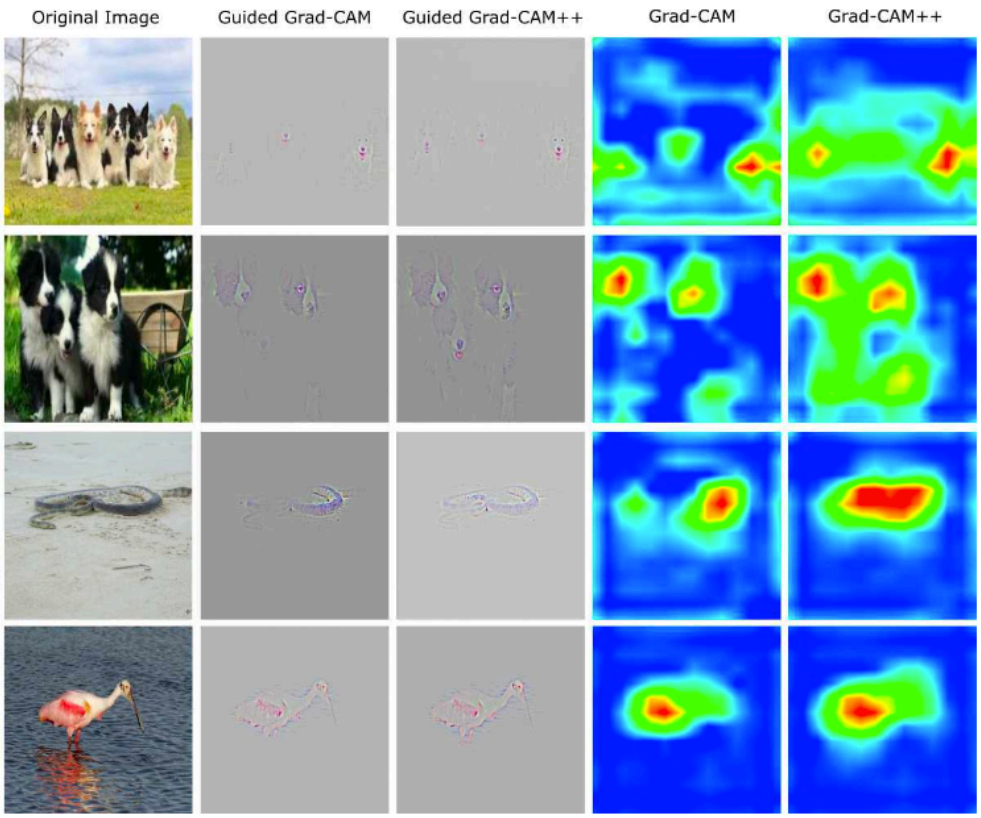


Рис. 1.29 Використання методу Grad-CAM++ для створення псевдомасок

Інший підхід – це самонавчання та псевдо-маркування, що передбачає початкове навчання моделі на невеликому наборі анотованих даних. Після цього модель застосовується до неанотованих даних для прогнозування масок, які потім використовуються як псевдо-анотації для подальшого навчання. Цей цикл

може повторюватися, поступово покращуючи точність моделі завдяки використанню все більшого обсягу даних. Прикладом є підходи з використанням Teacher-Student моделей [37], де "вчитель" генерує псевдомаски для навчання "учня".

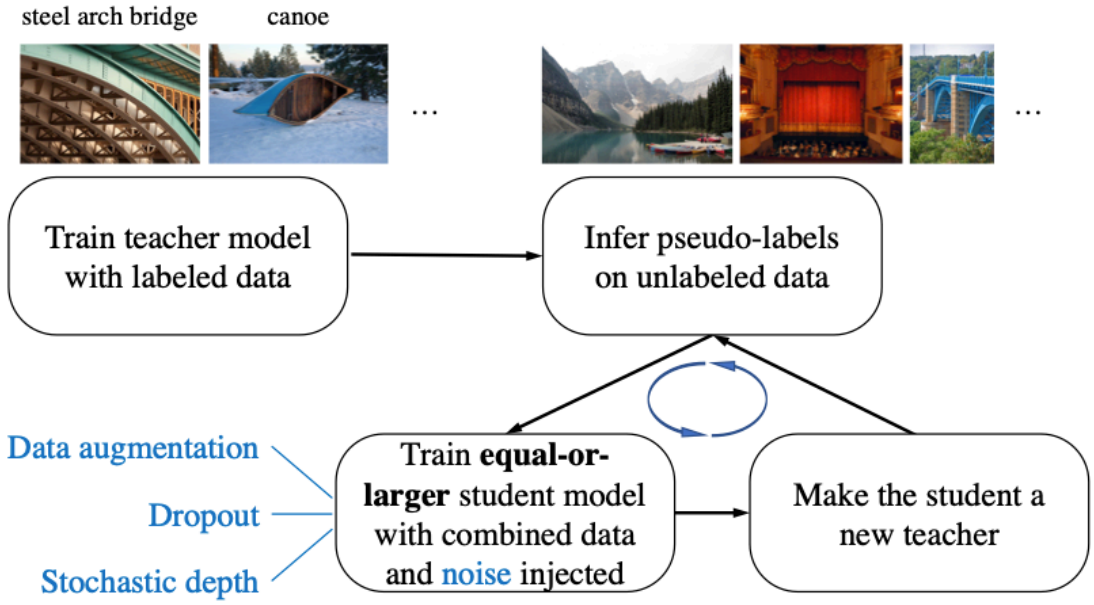


Рис. 1.30 Ілюстративне зображення підходу з використанням Teacher-Student

Використання генеративних моделей, таких як генеративно-змагальні мережі (GAN) [38], також є перспективним напрямом. Ці моделі можуть навчатися на неанотованих даних, генеруючи реалістичні маски об'єктів на основі вивчених патернів і структури зображень.

Рис. 1.31 GAN можуть використовуватися в якості моделей для генерації псевдомасок

Автокодувальники (Autoencoders) [39] також застосовуються для вивчення представлень даних та генерації псевдомасок, особливо в умовах аномальної або рідкісної інформації.

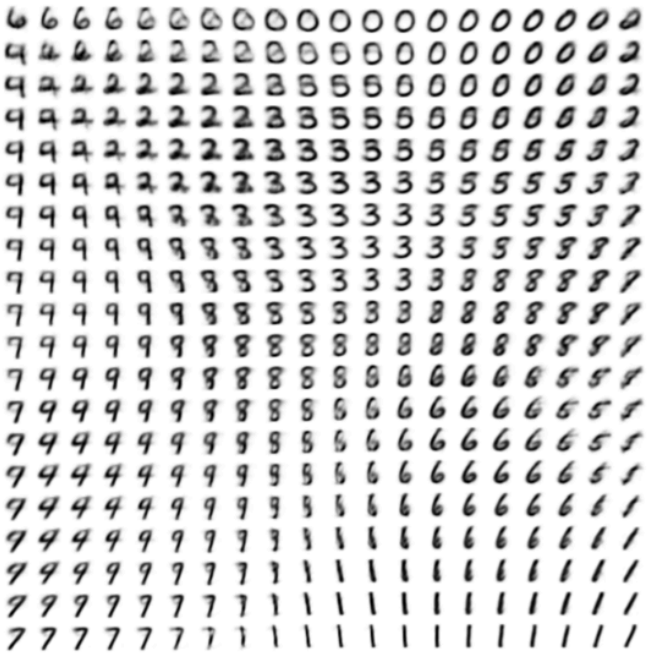


Рис. 1.32 Автокодувальники як один із методів створення псевдомасок

Загалом, впровадження псевдомасок у процес навчання моделей розпізнавання об'єктів дозволяє:

- Значно зменшити витрати на анотацію, оскільки немає потреби у створенні детальних масок вручну.
- Розширити навчальні набори даних за рахунок використання великої кількості неанотованих зображень.
- Підвищити гнучкість моделей, які можуть адаптуватися до нових доменів або умов без необхідності в масштабній анотації даних.

Однак слід враховувати і потенційні виклики, пов'язані з використанням псевдомасок. Якість автоматично згенерованих масок може бути нижчою, ніж у вручну створених, що може вплинути на точність моделі. Тому важливо розробляти методи, які можуть мінімізувати помилки в псевдомасках та підвищити їх якість, наприклад, через використання методів оцінки невизначеності або комбінування кількох підходів до генерації масок.

Псевдомаски знаходять застосування у різних галузях, де є обмеження щодо розмічених даних:

1. Медицина: У медичній сфері часто бракує розмічених даних для навчання моделей. Використання псевдомасок дозволяє покращити точність діагностики захворювань, таких як рак, за допомогою автоматичного аналізу медичних зображень.

2. Автономні транспортні засоби: Системи автономного водіння потребують великих обсягів даних для навчання моделей розпізнавання об’єктів на дорозі. Псевдомаски дозволяють зменшити залежність від ручної розмітки та покращити безпеку транспортних засобів.

3. Аналіз супутникових зображень: Використання псевдомасок у аналізі супутникових зображень допомагає автоматизувати процес виявлення та класифікації об’єктів, таких як будівлі, дороги та природні ресурси.

Використання псевдомасок у методах розпізнавання об’єктів на зображеннях є ефективним підходом для покращення точності та ефективності

моделей, особливо у випадках, коли бракує розмічених даних. Псевдомаски дозволяють зменшити витрати на створення розміток та забезпечити більш якісне навчання моделей. Продовження досліджень у цьому напрямку дозволить ще більше покращити продуктивність моделей та розширити сфери їх застосування.

Таким чином, використання псевдомасок представляє собою потужний інструмент у сфері комп'ютерного зору, що дозволяє ефективно вирішувати задачі розпізнавання об'єктів в умовах обмежених ресурсів для анотації. Це відкриває нові можливості для розвитку інтелектуальних систем та технологій, здатних працювати з великими обсягами даних та адаптуватися до різних застосувань.

1.4 Аналіз проблем розпізнавання об’єктів на зображеннях та постановка задач

У сучасному розвитку комп'ютерного зору розпізнавання об'єктів на зображеннях залишається однією з ключових задач, яка має широке застосування в різних галузях, таких як автономні транспортні засоби, медична діагностика, системи відеоспостереження, робототехніка тощо. Незважаючи на значні досягнення в цій сфері, існує низка проблем та викликів, які обмежують ефективність та універсальність існуючих методів.

Проблеми та виклики розпізнавання об'єктів на зображеннях:

1. Залежність від малорозмірних або недосконалих наборів даних. Більшість сучасних моделей глибокого навчання вимагають великих обсягів детально анотованих даних для досягнення високої точності [40]. Процес створення таких наборів даних є трудомістким, дорогим та часто непрактичним, особливо в специфічних доменах, таких як медична діагностика або промислові застосування, де експертна анотація є складною та дорогою [41].

2. Необхідність ефективних методів для роботи з обмеженими даними. Існує потреба в розробці підходів, які дозволять підвищити точність

ідентифікації та класифікації об'єктів на зображеннях з малорозмірних або недосконалих наборів даних.

3. Нестача універсальності моделей. Моделі, навчені на одному наборі даних або в одному домені, часто демонструють знижену продуктивність при застосуванні до інших доменів або умов [42]. Це обмежує їх практичне використання в реальних умовах, де дані можуть значно відрізнятися від тих, на яких модель була навчена.

4. Відсутність адаптованого програмного забезпечення. Бракує спеціалізованого алгоритмічного та програмного забезпечення, яке б реалізувало сучасні методи розпізнавання об'єктів з використанням псевдомасок на малих або недосконалих наборах даних.

5. Чутливість до шумів та варіацій даних. Моделі можуть бути чутливими до змін освітлення, кута зйомки, масштабування, часткової оклюзії об'єктів та інших факторів [43], що знижує їх надійність в реальних умовах.

6. Накопичення помилок при використанні псевдо-анотацій. Хоча використання псевдомасок та псевдо-анотацій може допомогти вирішити проблему нестачі даних, воно також може призвести до накопичення помилок, якщо якість цих псевдо-анотацій є низькою [44].

7. Необхідність практичної реалізації та перевірки. Потрібно створити практичні інструменти та провести апробацію розроблених технологій на реальних даних та задачах, щоб підтвердити їх ефективність.

Використання псевдомасок представляє собою перспективний підхід для подолання проблеми нестачі анотованих даних. Псевдомаски дозволяють ефективно використовувати неанотовані або слабко анотовані дані, зберігаючи при цьому високу точність розпізнавання [45]. Це особливо важливо в задачах, де анотація даних є складною або дорогою, наприклад, у медичній діагностиці.

Дослідження показують, що використання псевдомасок може значно покращити продуктивність моделей, особливо в умовах обмеженого набору даних [46]. Наприклад, у роботі Li та ін. [47] продемонстровано, що псевдо-мітки

можуть бути ефективно використані для покращення якості сегментації в умовах слабко контрольованого та напівконтрольованого навчання.

Враховуючи вищезазначені проблеми, метою даного дослідження є розробка інформаційної технології, що дозволить підвищити швидкість створення псевдомасок, а також точність ідентифікації та класифікації об'єктів на зображеннях малорозмірних або недосконалих наборів даних з використанням псевдомасок.

Для досягнення поставленої мети необхідно вирішити такі задачі:

1. Виконати аналіз відомих підходів та моделей для ідентифікації та класифікації об’єктів на зображеннях малорозмірних або недосконалих наборів даних та аналіз основних проблем, які виникають під час ідентифікації та класифікації об'єктів на таких зображеннях. Це включає аналіз існуючих методів, їх переваг та недоліків, а також визначення факторів, що впливають на якість розпізнавання.

2. Розробити інформаційної технології, включаючи нові підходи, методи та алгоритми, які дозволяють підвищити точність ідентифікації та класифікації заданих об’єктів з використанням псевдомасок на зображеннях малорозмірних або недосконалих наборів даних. Це передбачає створення методів генерації та використання псевдомасок, які мінімізують вплив помилок та покращують продуктивність моделей.

3. Створити алгоритмічного та програмного забезпечення для застосування запропонованої інформаційної технології та методів ідентифікації та класифікації об'єктів на зображеннях малорозмірних або недосконалих наборів даних. Це включає розробку програмних модулів та інструментів, які реалізують запропоновані алгоритми.

4. Розробити комп’ютерні програми на мові програмування Python для застосування інтелектуальної інформаційної технології з використанням псевдомасок для ідентифікації та класифікації об'єктів на зображеннях малорозмірних або недосконалих наборів даних. Це забезпечить можливість відтворення результатів та зручність використання розроблених методів.

5. Апробувати та впровадити розробленої інтелектуальної інформаційної технології на реальних даних та задачах. Це передбачає проведення експериментів на практичних наборах даних, оцінку ефективності запропонованих методів та порівняння з існуючими підходами.

Це дослідження є актуальним, оскільки дозволяє подолати ключові обмеження сучасних методів розпізнавання об'єктів на зображеннях, особливо в умовах обмежених ресурсів даних. Використання псевдомасок може значно зменшити витрати на анотацію даних та розширити можливості застосування моделей в різних доменах. Крім того, розробка програмного забезпечення та інструментів сприятиме практичному впровадженню розроблених методів.

1.5 Висновки до розділу

У першому розділі проведено детальний аналіз проблематики розпізнавання об'єктів на зображеннях з малорозмірними та зашумленими наборами даних. Перш за все, було досліджено основні методи та підходи, які використовуються в комп'ютерному зорі. Розглянуто їхні переваги та недоліки у контексті роботи з обмеженими даними, що дозволило виявити основні виклики, пов'язані з точністю розпізнавання.

Особливу увагу приділено питанням попередньої обробки зображень, оскільки вона відіграє вирішальну роль у підвищенні якості моделі. Важливим результатом стало те, що адаптивна обробка даних, наприклад, з використанням білатерального фільтру та CLAHE, значно покращує результати сегментації в порівнянні з класичними підходами.

Також у розділі розглянуто актуальні задачі сегментації зображень та використання псевдомасок як способу для автоматизації та підвищення точності сегментаційних моделей. Завдяки псевдомаскам зменшується потреба в ручній розмітці, що є важливим фактором для роботи з великими наборами даних.

Результати розділу опубліковані в таких роботах автора: [50], [69], [70], [72]

РОЗДІЛ 2
УДОСКОНАЛЕННЯ МЕТОДІВ ТА МОДЕЛЕЙ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК

2.1 Аналіз даних та сучасних методів попередньої обробки як основа для розробки технології розпізнавання об'єктів з використанням псевдомасок

Для успішного вирішення завдання створення методів розпізнавання об'єктів на зображеннях з використанням псевдомасок необхідно спочатку провести детальний аналіз доступних даних та існуючих методів попередньої обробки. Розуміння характеристик та особливостей даних дозволяє визначити найбільш ефективні підходи до їх обробки та використання в алгоритмах розпізнавання.

Попередня обробка даних є критичним етапом у підготовці зображень для аналізу. Вона включає в себе такі процедури, як нормалізація, фільтрація, видалення шумів, вирівнювання гістограм, корекція освітлення та інше. Вибір правильних методів попередньої обробки безпосередньо впливає на якість розпізнавання та продуктивність розроблених моделей.

Аналіз популярних методів попередньої обробки дозволяє виявити їх сильні та слабкі сторони, що є важливим для подальшої оптимізації та адаптації цих методів до специфіки використання псевдомасок. Крім того, ознайомлення з сучасними тенденціями та досягненнями в цій галузі сприятиме розробці інноваційних підходів та підвищенню ефективності запропонованих рішень.

Одним з ключових аспектів є розуміння типу даних, з якими ми працюємо. Зображення можуть мати різні формати, роздільну здатність, кольорові простори та інші характеристики, які впливають на вибір методів обробки. Наприклад, обробка кольорових зображень потребує врахування інформації про колірні канали, таких як RGB, HSV або LAB, тоді як для зображень в градаціях сірого акцент робиться на градієнтах та контрастності.

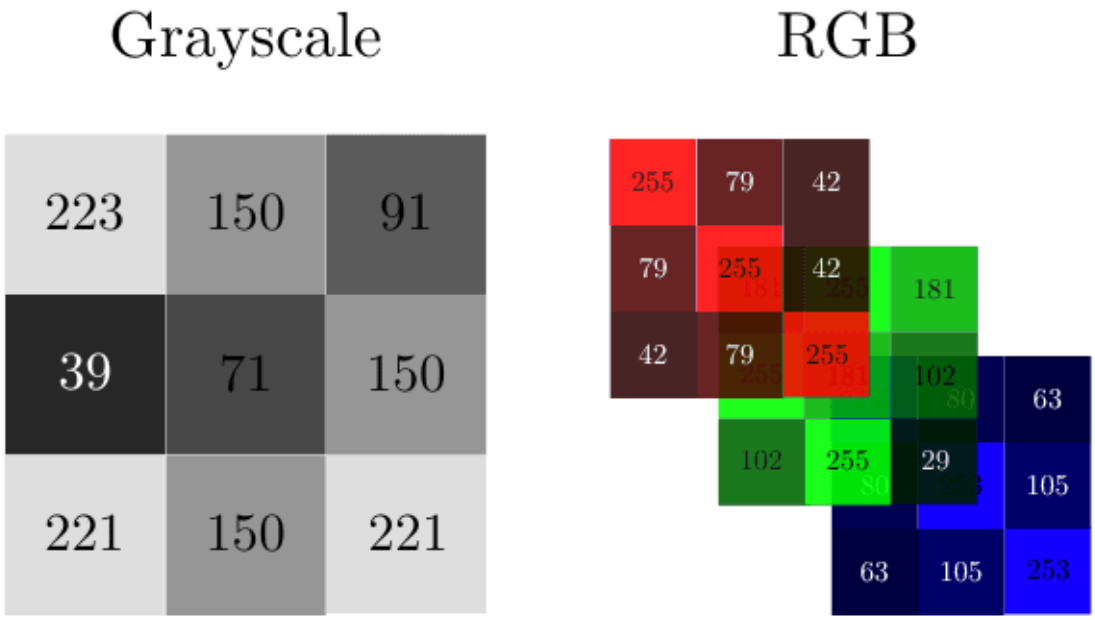


Рис. 2.1 Матриці пікселів зображення в градаціях сірого та RGB зображення

Також важливо враховувати наявність шумів та артефактів, які можуть негативно впливати на точність розпізнавання. Для цього використовуються методи фільтрації, такі як гаусівське розмиття, медіанний фільтр, біноміальні фільтри та інші підходи до згладжування зображень.

Ще одним важливим етапом є виділення ознак, які будуть використовуватися для навчання моделей. Це можуть бути детектори країв, такі як оператори Собеля або метод Кенні. Виділення контурів дозволяє отримати інформацію про форми об'єктів та їх межі.

У контексті використання псевдомасок, особливо важливим є правильна сегментація зображень та виділення областей, що відповідають об'єктам інтересу. Псевдомаски дозволяють позначити ці області без необхідності точного ручного маркування, що значно спрощує процес підготовки даних та розширює можливості для застосування алгоритмів розпізнавання.

Для виконання практичних досліджень було використано відкритий набір даних, який складається з анонімних панорамних рентгенівських знімків зубів 116 пацієнтів, обстежених у Медичному Центрі "Noor" в Кумі, Іран [48], [49]. Цей

набір даних охоплює широкий спектр стоматологічних станів пацієнтів — від здорових зубів до часткової та повної відсутність зубів. Нижні щелепи всіх знімків були вручну сегментовані двома стоматологами, що підвищує точність та надійність анотацій.

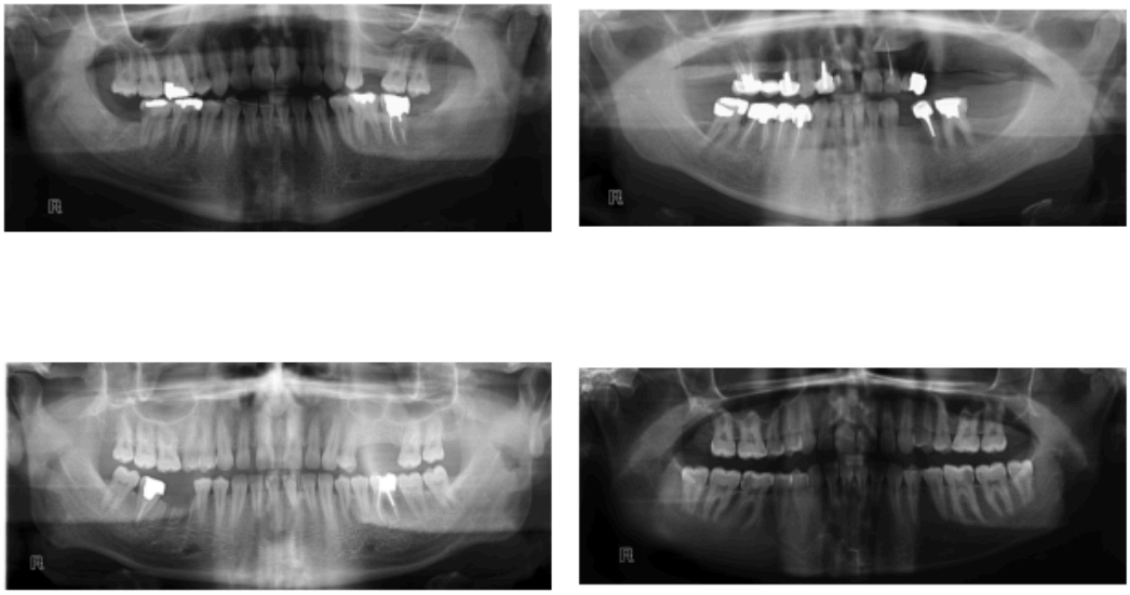


Рис. 2.2 Приклади панорамних стоматологічних знімків з набору даних

Як видно з рисунку, в досліджуваному наборі даних представлені панорамні рентгенівські знімки щелеп, які відображають структуру зубного ряду, стан щелепних кісток та наявність стоматологічних втручань (пломби, імплантати). Дані знімки можуть слугувати основою для аналізу стану зубів, виявлення захворювань, аномалій розвитку зубів, кісткових структур, а також планування стоматологічного лікування, включаючи імплантацію та ортодонтичні процедури.

Всі зображення були зменшені до розміру 512×512 пікселів, що забезпечує уніфікацію розмірів та знижує обчислювальну складність.

Крім того, набір даних містить різноманітні знімки з різними рівнями шуму, артефактами та варіаціями освітлення, що робить його придатним для тестування стійкості алгоритмів попередньої обробки даних та технологій

розпізнавання об’єктів на зображеннях в реальних умовах. Цей набір даних також був використаний як основа для статті [50].

Необхідність використання методів попередньої обробки даних обумовлена тим, що початкові зображення часто містять шуми, низький контраст або нерівномірне освітлення, які можуть негативно впливати на подальший аналіз і точність діагностики. Застосування ефективних методів попередньої обробки дозволяє підвищити якість зображень, виділити ключові особливості та забезпечити більш надійні результати. Особливо в стоматології це сприяє точнішому виявленню патологій та плануванню лікування, що в кінцевому підсумку покращує якість медичних послуг.

Далі проаналізуємо відомі фільтри та методи попередньої обробки зображень, зокрема обмежене адаптивне вирівнювання гістограми (CLAHE) [53], фільтра Гауса [52], білатеральний фільтр [59] та багатомасштабний фільтр Ретінекс з відновленням кольору (MSRCR) [51], [52]. Кожен із цих методів має унікальні властивості щодо покращення якості зображень, зменшення шуму та збереження важливих деталей, що робить їх перспективними для ефективної попередньої обробки стоматологічних знімків.

Розглянемо кожен з фільтрів детальніше та застосуємо дані фільтри до зображень з досліджуваного набору даних.

1. Обмежене адаптивне вирівнювання гістограми (CLAHE – Contrast Limited Adaptive Histogram Equalization): CLAHE є вдосконаленою версією стандартного методу вирівнювання гістограми. Він призначений для покращення контрастності та яскравості зображень, особливо в областях з низьким контрастом. Метод розбиває зображення на менші регіони і виконує вирівнювання гістограми окремо для кожного з них. Це дозволяє уникнути проблеми надмірного підсилення шуму, що часто виникає при застосуванні глобального вирівнювання гістограми [53].

Ключовим параметром CLAHE є "clip limit", який контролює ступінь вирівнювання гістограми. Встановлення цього обмеження запобігає надмірному підсиленню контрастності в однорідних областях, де це може призвести до

посилення шуму. Таким чином, CLAHE дозволяє зберегти деталі в темних та світлих областях, покращуючи загальну якість зображення.

У стоматологічних рентгенівських знімках застосування CLAHE допомагає виявити дрібні деталі, такі як мікротріщини або початкові ознаки карієсу, які можуть бути непомітними при стандартній обробці. Покращення контрастності сприяє більш точній діагностиці та підвищує ефективність лікування.

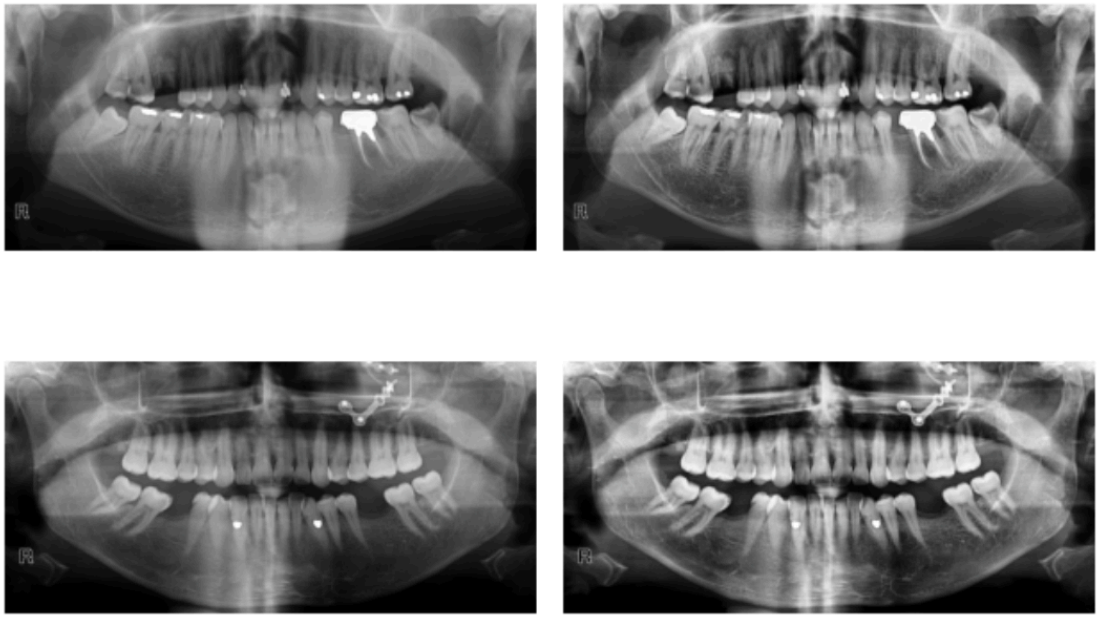


Рис. 2.3 Приклад застосування CLAHE до зображень з набору даних, де ліворуч – оригінальне зображення, праворуч – зображення з CLAHE.

2. Фільтр Гауса: є одним з найпоширеніших методів згладжування зображень і зменшення шуму. Він працює шляхом усереднення значень пікселів зваженим способом, де ваги визначаються гаусівською функцією розподілу. Це означає, що ближчі до центрального пікселя сусіди мають більший вплив на його нове значення, а віддаленіші – менший [52].

Перевагою фільтра Гауса є його здатність ефективно зменшувати випадковий шум, зберігаючи при цьому загальну структуру зображення. Однак, цей метод може розмивати краї та дрібні деталі, що є недоліком при аналізі медичних зображень, де важливо зберегти чіткість контурів.

У контексті стоматологічних знімків фільтр Гауса може бути корисним для зменшення електронного шуму або артефактів, що виникають при скануванні. Проте, необхідно ретельно підбирати параметри фільтра, зокрема стандартне відхилення гаусівської функції, щоб уникнути втрати важливих діагностичних деталей.

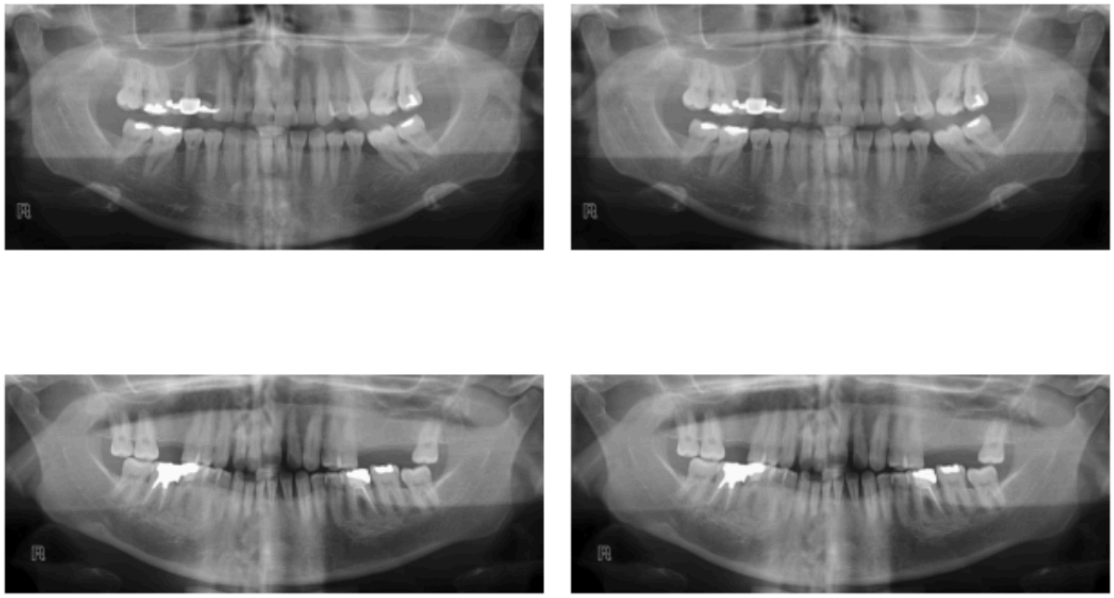


Рис. 2.4 Застосування фільтра Гауса

3. Білатеральний фільтр: є нелінійним методом згладжування, який, на відміну від фільтра Гауса, зберігає краї зображення. Він враховує не лише просторову близькість пікселів, але й схожість їхніх інтенсивностей. Це досягається шляхом комбінування двох функцій ваги: одна залежить від відстані між пікселями, інша – від різниці в інтенсивності [59].

Такий підхід дозволяє ефективно згладжувати однорідні області зображення, зменшуючи шум, і одночасно зберігати різкі переходи в інтенсивності, які часто відповідають фізичним краям об'єктів. Білатеральний фільтр особливо корисний у медичних зображеннях, де важливо зберегти контури та деталі для точної інтерпретації.

У стоматології застосування білатерального фільтра допомагає покращити візуалізацію структур зубів та кісток, зменшуючи вплив шуму без розмивання

важливих деталей. Це сприяє більш точному визначенню патологічних змін і плануванню лікування.

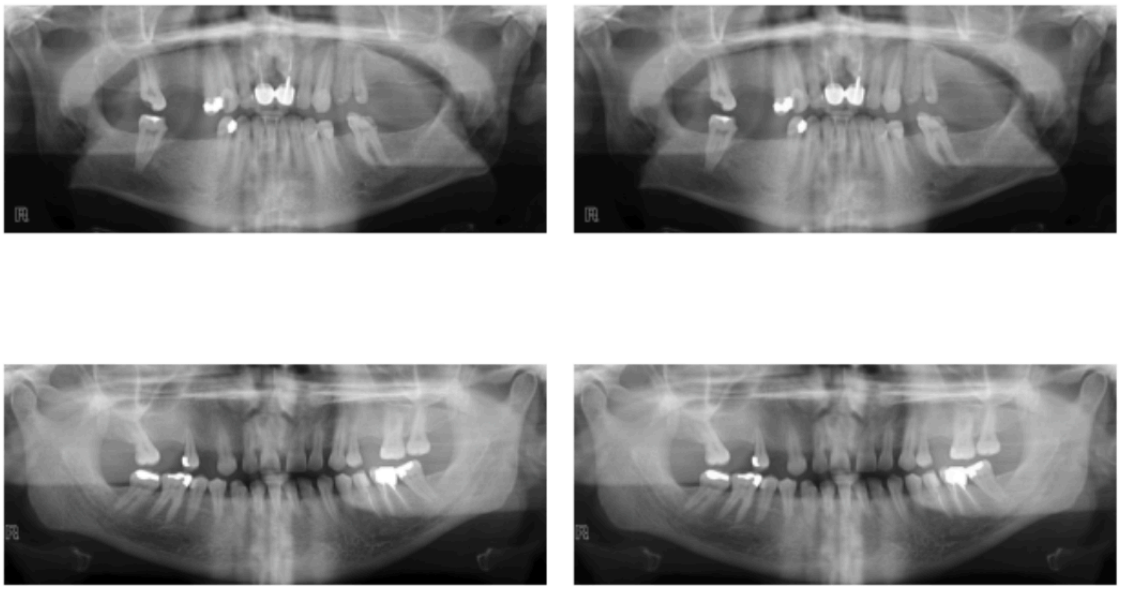


Рис. 2.5 Застосування білатерального фільтра

4. Багатомасштабний фільтр з відновленням кольору (MSRCR – Multiscale Retinex with Color Restoration): є вдосконаленим методом обробки зображень, що поєднує принципи теорії Retinex та техніки відновлення кольору. Теорія Retinex моделює спосіб, яким людське око адаптується до різних умов освітлення, зберігаючи при цьому сприйняття кольорів та контрастів [51], [52].

Метод MSRCR застосовує ретинекс-алгоритми на декількох масштабах зображення, що дозволяє підсилити деталі як у великих, так і в малих областях. Кожен масштаб обробляється окремо, після чого результати об'єднуються для отримання фінального зображення. Компонент відновлення кольору коригує насиченість та баланс кольорів, забезпечуючи природний вигляд.

У медичних зображеннях, особливо у стоматології, MSRCR може значно покращити видимість деталей у зонах з низьким контрастом або нерівномірним освітленням. Це дозволяє лікарям більш точно оцінити стан тканин та структур, що є критичним для діагностики та лікування.

Окрім розглянутих методів, існують інші фільтри та оператори, які можуть бути корисними у попередній обробці зображень:

- Медіанна фільтрація: цей нелінійний фільтр замінює значення кожного пікселя на медіану його сусідів. Він ефективний для видалення імпульсного шуму ("сіль і перець") без розмивання країв.

- Оператор/фільтр Собеля: використовується для виділення країв та контурів. Він обчислює градієнт інтенсивності зображення, що допомагає виявити області з різкою зміною яскравості.

- Дискретний оператор Лапласа: цей другий порядок диференційного оператора підсилює області з високою кривизною, такі як краї та кути. Він може використовуватися для підсилення деталей перед сегментацією або іншим аналізом.

Кожен з цих методів має свої унікальні властивості та може бути обраний залежно від специфічних вимог задачі та характеристик зображення.

Вибір відповідного методу попередньої обробки є критичним для забезпечення якісного аналізу медичних зображень. Неправильний вибір або налаштування фільтра може призвести до втрати важливих деталей або введення артефактів, що негативно вплине на діагностику.

Застосування ефективних методів попередньої обробки дозволяє:

- Зменшити шум: видалення небажаних сигналів покращує якість зображення та спрощує подальший аналіз.

- Покращити контрастність: підвищення різниці між різними структурами полегшує їх розпізнавання.

- Зберегти деталі та краї: важливо для точної ідентифікації меж між різними тканинами або патологічними утвореннями.

У сучасній стоматологічній практиці якісна обробка рентгенівських знімків є невід'ємною частиною процесу діагностики та планування лікування. Використання відповідних методів попередньої обробки зображень дозволяє лікарям отримати більш інформативні та точні дані, що в кінцевому результаті покращує якість медичних послуг.

2.2 Розробка базової моделі для розпізнавання об’єктів на зображеннях

У рамках даного дослідження було розроблено базову модель для розпізнавання об'єктів на зображеннях, зосереджену на сегментації стоматологічних пломб на панорамних рентгенівських знімках. Розпізнавання об'єктів у медичних зображеннях є критично важливим напрямком, оскільки воно сприяє автоматизації діагностичних процесів і підвищенню точності виявлення патологій. У цьому контексті представлено комплексний аналіз методів попередньої обробки зображень та фільтрів, які застосовуються для покращення якості вхідних даних і підвищення ефективності сегментації.

Сегментація стоматологічних пломб є ключовим завданням в аналізі стоматологічних зображень. Метою цього процесу є точна ідентифікація та розмежування зубних пломб на рентгенівських знімках, що має велике значення для клінічної практики. Точна сегментація дозволяє стоматологам більш ефективно оцінювати стан реставрацій зубного ряду, контролювати довговічність пломб та планувати необхідні лікувальні втручання. Однак, через складність рентгенівських зображень, таких як низький контраст і наявність шумів, ця задача залишається викликом.

Для вирішення проблеми сегментації стоматологічних пломб було обрано архітектуру моделі глибокого навчання для сегментації зображень, відому як U-Net [32]. Ця архітектура широко визнана в науковій спільноті завдяки своїм винятковим можливостям у точній сегментації медичних зображень. U-Net складається зі структури енкодера-декодера (рис.), що дозволяє моделі ефективно вивчати як локальні, так і глобальні особливості зображення. Енкодер зменшує розмірність даних, вилучаючи важливі характеристики, тоді як декодер відновлює просторову роздільну здатність для точного позиціонування сегментованих об'єктів.

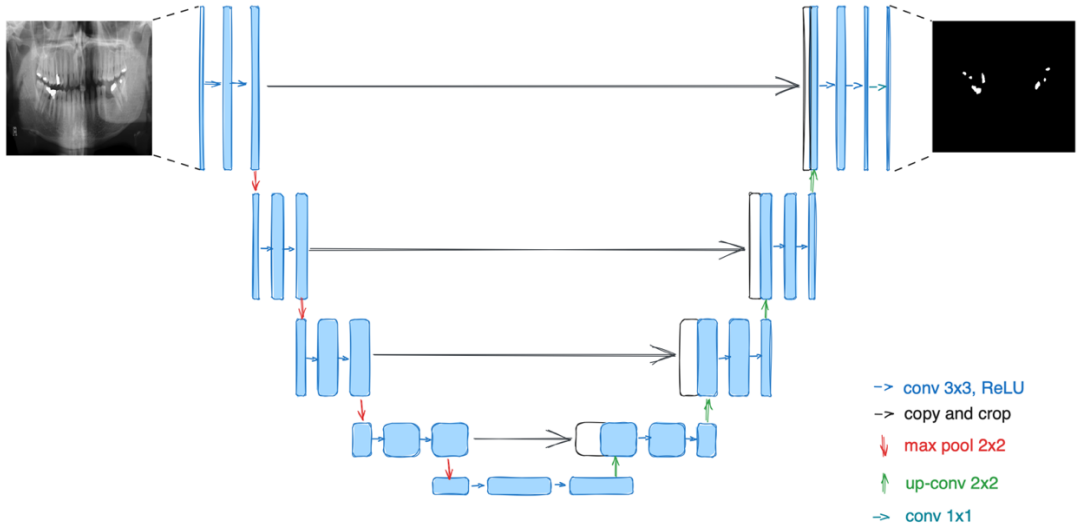


Рис. 2.6 Архітектура U-Net

Енкодер в U-Net відповідає за витягнення ознак з вхідного зображення через послідовність шарів згорткових нейронних мереж. Кожен шар зменшує розмір зображення (downsampling) і збільшує кількість ознак, що дозволяє моделі витягати як локальні, так і глобальні шаблони. Цей процес допомагає поступово конденсувати важливу інформацію, що полегшує сегментацію на більш високому рівні. Важливою особливістю є застосування операцій pooling, які зменшують розмірність, але збільшують область огляду для кожного нейрону [32].

Декодер в U-Net відповідає за відновлення просторової роздільної здатності зображення, яке пройшло через енкодер. Декодер використовує шари, які збільшують розмір зображення (upsampling) за допомогою зворотних згорток (transposed convolutions) або інтерполяції. Важливою особливістю архітектури U-Net є те, що на кожному рівні декодера об'єднуються (concatenate) ознаки з відповідного рівня енкодера, що дозволяє зберігати просторову інформацію з вхідного зображення [32]. Це дозволяє моделі ефективно реконструювати сегментоване зображення з детальними межами об'єктів.

В якості енкодера було обрано попередньо натреновану модель "timm-resnest101e" [54], яка є варіантом ResNet з покращеними характеристиками. Ця модель була попередньо натренована на наборі даних ImageNet [55], що

забезпечує початкові ваги і сприяє швидшій конвергенції під час навчання. Використання попередньо натренованих моделей є загальноприйнятою практикою в комп'ютерному зорі, оскільки це дозволяє досягти кращої генералізації та ефективності на обмежених наборах даних.

Для тренування моделі сегментації використовувалася функція втрат бінарної перехресної ентропії (binary cross-entropy, BCELoss), яка є стандартною для бінарних задач класифікації та сегментації. В якості оптимізатора був застосований алгоритм «Адам» (Adam optimizer), відомий своєю здатністю адаптивно змінювати коефіцієнт навчання для кожного параметра, що сприяє швидшій та стабільнішій збіжності. Коефіцієнт швидкості навчання (learning rate) встановлено на рівні 0,001, що забезпечує баланс між швидкістю навчання та стабільністю моделі.

Крім того, для підвищення ефективності навчання та запобігання перенавчанню було впроваджено механізм ранньої зупинки (early stopping) з критерієм зупинки після 5 епох за умови не покращення валідаційної метрики. Це дозволяє зупинити навчання в оптимальний момент, зберігаючи модель з найкращими показниками на валідаційному наборі даних. Такий підхід допомагає уникнути перенавчання та покращує загальну узагальнюючу здатність моделі.

Під час експериментів використовувалися методи аугментації даних за допомогою бібліотеки Albumentations [56]. Ця бібліотека забезпечує широкий спектр трансформацій зображень, що допомагає покращити узагальнювальну здатність моделі. До стандартних методів аугментації належали:

1. Нормалізація зображення: для забезпечення узгоджених і стандартизованих вхідних даних, що допомагає моделі краще розпізнавати патерни.

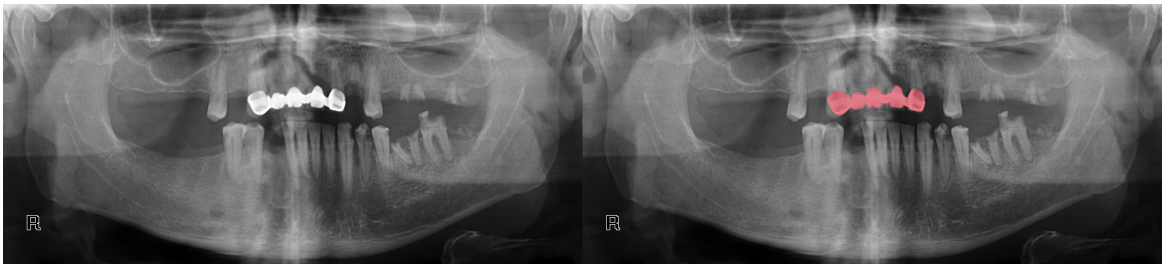
2. Обертання зображення: для підвищення стійкості моделі до різних орієнтацій зображень, що може бути особливо корисним у випадках, коли знімки мають різні кути нахилу.

3. М'яке розмиття: для зменшення впливу шуму та артефактів на зображення, що сприяє кращому виділенню суттєвих деталей на знімках.

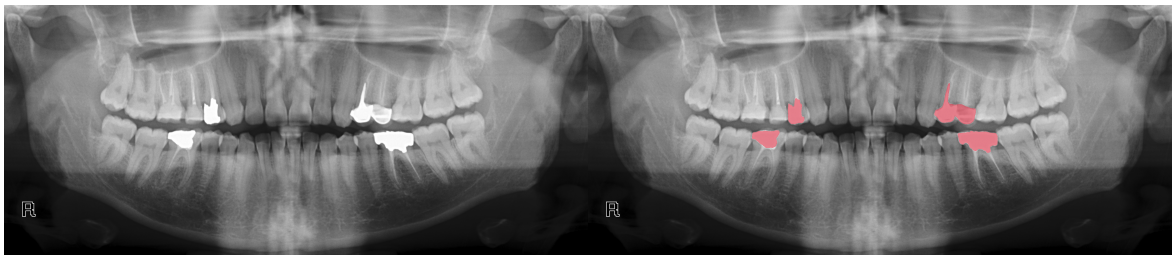
Для додаткового підвищення якості моделі було використано інші методи аугментації, такі як зміна яскравості та контрастності, додавання випадкового шуму, горизонтальні та вертикальні відображення. Ці методи допомагали моделі бути стійкою до різних варіацій зображень, що забезпечувало кращу здатність до узагальнення на тестових даних.

Набір даних був розділений на три підмножини: тренувальний набір (80%), валідаційний набір (10%) та тестовий набір (10%). Тренувальний набір використовується для навчання моделей, що дозволяє їм вивчати характерні шаблони зображень та їх відповідних масок. Валідаційний набір використовується для налаштування гіперпараметрів та оцінки моделі на проміжних етапах, щоб уникнути перенавчання. Тестовий набір застосовується виключно для остаточного оцінювання продуктивності моделі після завершення навчання, щоб оцінити її здатність до генералізації на нових, невідомих даних. Це розбиття є стандартним підходом у машинному навчанні та допомагає забезпечити об'єктивну оцінку моделей.

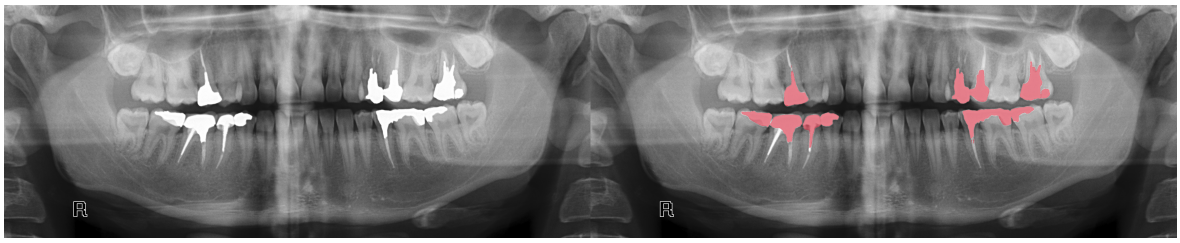
Всі зображення були зменшені до розміру 512x512 пікселів, що дозволяє зберігати баланс між деталізацією та ефективністю обробки. Набір даних містить широкий спектр стоматологічних рентгенівських знімків, що охоплюють різні захворювання зубів, рівні шуму, артефакти та варіації освітлення. Це різноманіття допомагає моделі бути більш стійкою до змінних факторів у реальних умовах, забезпечуючи кращу генералізацію. Нижче представлені приклади оригінальних зображень і їх відповідних масок сегментації, що були використані в даному дослідженні.



а)



б)



в)

Рис. 2.7 Приклади оригінальних зображень та відповідні їм анотовані маски сегментації (а, б, в)

Додатково було проведено аналіз впливу різних методів попередньої обробки зображень на якість сегментації. Зокрема, розглядалися такі техніки, як білатеральний фільтр, фільтр CLAHE, фільтр Гауса та фільтр MSRCR.

Для всебічної оцінки ефективності фільтрів попередньої обробки та сегментаційної моделі використовувалися такі метрики:

1. Коефіцієнт Дайса (Dice Score): Ця метрика кількісно визначає просторове перекриття між прогнозованою та справжньою масками сегментації. Значення коефіцієнта Дайса коливаються від 0 (відсутність перекриття) до 1 (ідеальне перекриття), що робить її ефективним інструментом для оцінки точності сегментації.

$$\text{Dice} = \frac{2 \cdot |A \cap B|}{|A| + |B|}, \tag{2.1}$$

де:

- A - це істинна маска (ground truth);
- B - це передбачена маска.

2. Коефіцієнт Жаккара (Jaccard index або IoU – Intersection over Union): Індекс Жаккара вимірює перекриття шляхом обчислення співвідношення перетину до об'єднання між прогнозованою та справжньою масками сегментації. Це співвідношення дозволяє отримати оцінку точності та повноти сегментації, надаючи важливу інформацію про ефективність моделі.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \tag{2.2}$$

де:

- A - це істинна маска (ground truth);
- B - це передбачена маска.

3. Влучність (Precision): Ця метрика оцінює точність позитивних прогнозів, обчислюючи співвідношення справжніх позитивних результатів до загальної кількості позитивних прогнозів. Висока влучність свідчить про здатність моделі правильно ідентифікувати позитивні випадки з мінімальною кількістю хибнопозитивних результатів.

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2.3}$$

де:

- TP - це кількість істинно позитивних передбачень;

- FP - це кількість хибно позитивних передбачень.

4. Повнота (Recall): Ця метрика оцінює здатність алгоритму ідентифікувати всі релевантні випадки, обчислюючи співвідношення справжніх позитивних результатів до загальної кількості фактичних позитивних результатів. Висока чутливість вказує на здатність моделі виявляти всі важливі об'єкти в зображенні.

$$Recall = \frac{TP}{TP + FN}, \tag{2.4}$$

де:

- TP - це кількість істинно позитивних передбачень;
- FN - це кількість хибно негативних передбачень.

Використання цих метрик дозволяє провести комплексну оцінку ефективності сегментаційних моделей, враховуючи вплив фільтрів попередньої обробки на різні аспекти точності. Завдяки такому підходу можна не лише визначити загальну ефективність моделі, але й зрозуміти, які аспекти роботи потребують вдосконалення.

Отже, з урахуванням наведених вище параметрів, було натреновано декілька сегментаційних моделей, кожна з яких використовувала різний фільтр попередньої обробки вхідного зображення. Це дозволило оцінити вплив різних фільтрів на ефективність сегментації стоматологічних пломб. Усі моделі були навчені із застосуванням тих самих налаштувань гіперпараметрів, що забезпечило чесне порівняння.

Таблиця 2.1 Порівняння натренованих сегментаційних моделей з різними фільтрами попередньої обробки вхідного зображення за встановленими метриками

Назва фільтру	Dice Score	IoU Score	Precision	Recall	Loss
Білатеральний фільтр	**0.94493**	**0.928249**	**0.995193**	0.982434	**0.0031**
Фільтр CLAHE	0.926958	0.912003	0.982922	**0.995365**	0.0036
Без фільтру	0.912757	0.90474	0.985562	0.98341	0.0037
Фільтр Гауса	0.890474	0.910019	0.96134	0.982887	0.0042
MSRCR	0.877426	0.918587	0.962336	0.987035	0.0044

У таблиці наведені результати, які демонструють ефективність кожного з фільтрів. Найбільшу перевагу в задачі сегментації стоматологічних пломб показав білатеральний фільтр. Цей фільтр забезпечив високу точність і чутливість, завдяки своїй здатності зберігати краєві деталі під час видалення шуму, що є критично важливим для точного визначення меж пломб на рентгенівських знімках.

Другу позицію за ефективністю зайняв широко відомий фільтр CLAHE (Contrast Limited Adaptive Histogram Equalization). CLAHE продемонстрував відмінні результати у покращенні контрастності зображень, особливо в областях з низьким контрастом. Це дозволило моделі краще розпізнавати дрібні деталі, що сприяло підвищенню точності сегментації.

Додатково до цих двох фільтрів, у дослідженні були оцінені й інші методи попередньої обробки зображень, такі як фільтр Гауса та багатомасштабний фільтр з відновленням кольору (MSRCR). Хоча ці методи також показали непогані результати, вони не досягли такого ж рівня точності та чутливості, як білатеральний фільтр і CLAHE.

Враховуючи ці результати, можна зробити висновок, що вибір фільтра попередньої обробки значно впливає на ефективність сегментаційних моделей. Білатеральний фільтр, завдяки своїм характеристикам, показав найкращі

результати, що робить його найкращим вибором для задач сегментації стоматологічних пломб. Проте CLAHE також є дуже ефективним інструментом, особливо у випадках з низьким контрастом.

Далі представлено графіки зміни значень метрик для кожної моделі на валідаційному наборі даних.

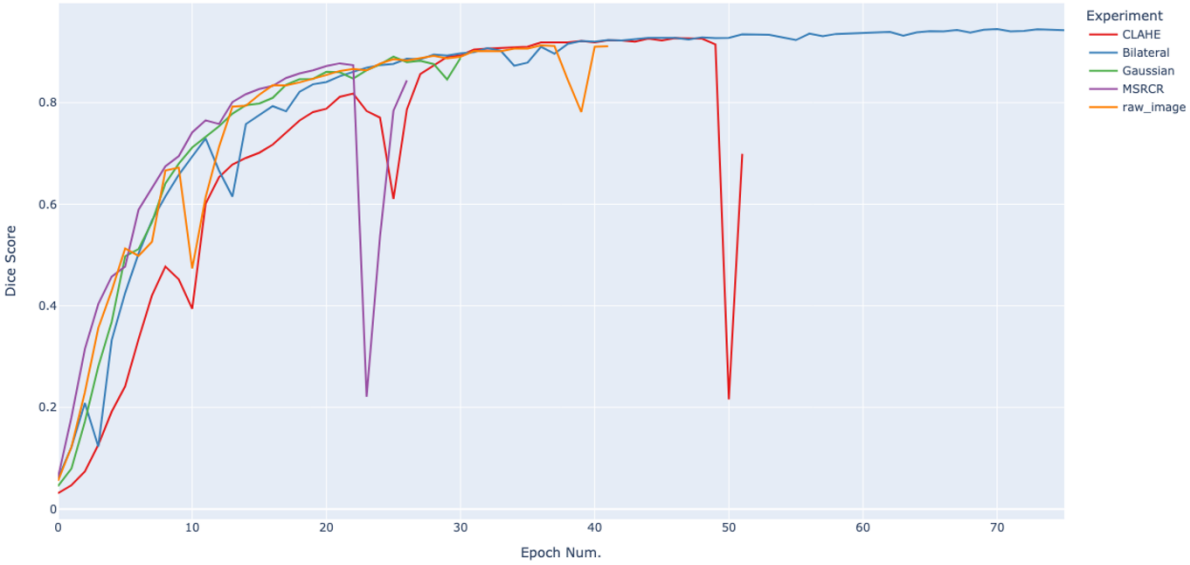


Рис. 2.8 Графік порівняння моделей за коефіцієнтом Дайса (Dice score)

На рисунку представлено порівняння моделей сегментації з різними фільтрами за коефіцієнтом Дайса. Графіки демонструють динаміку навчання моделей, де чітко видно, що всі експерименти показували стабільний прогрес до 20-ї епохи. Після цього моменту спостерігаються різкі спади в значеннях коефіцієнта Дайса для моделей, які використовували фільтри MSRCR та CLAHE. Ці спади стали причиною активації механізму ранньої зупинки (early stopping), що припинило подальше тренування цих моделей для запобігання перенавчанню.

Для моделей, що використовували фільтр Гауса та зображення без будь-якої попередньої обробки, процес тренування тривав дещо довше, проте їхня ефективність також знижувалася після певної кількості епох. Найбільш ефективною виявилася модель, яка застосовувала білатеральний фільтр. Цей

фільтр продемонстрував стабільний приріст коефіцієнта Дайса протягом усього процесу навчання і забезпечив найвищу точність сегментації.

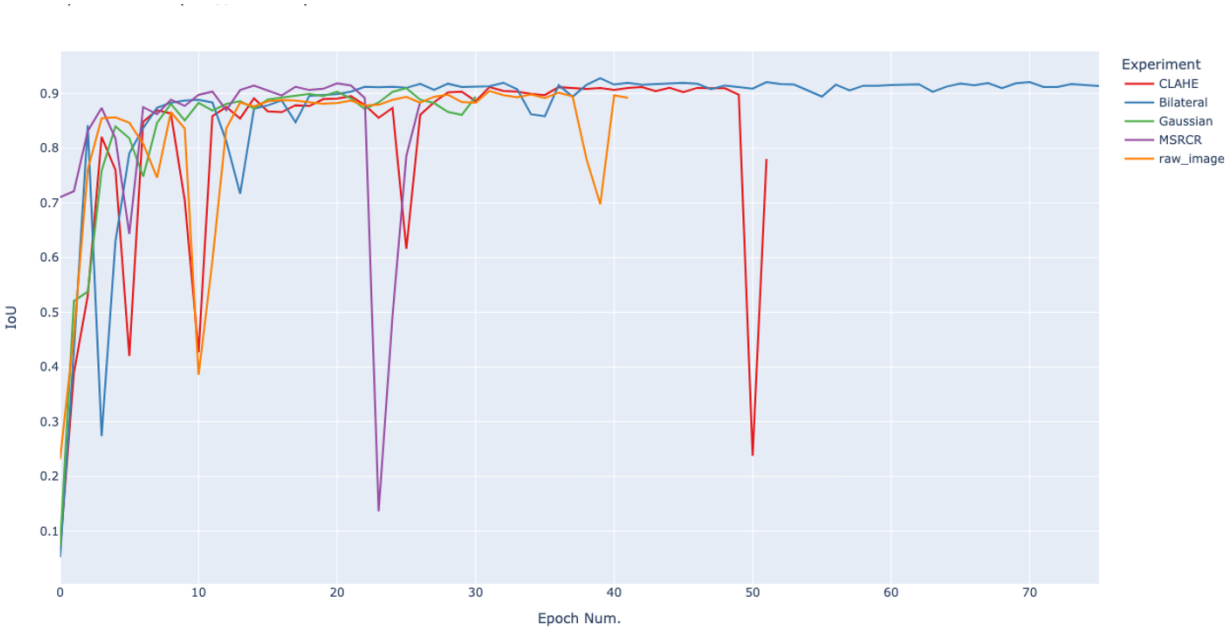


Рис. 2.9 Графік порівняння моделей за коефіцієнтом Жаккара (IoU)

На рисунку представлено порівняння сегментаційних моделей за коефіцієнтом Жаккара (Jaccard index або IoU – Intersection over Union).

З графіків видно, що всі моделі демонструють досить високі значення коефіцієнта Жаккара, які перевищують 85%. Це свідчить про загальну високу якість сегментації, незалежно від використовуваного фільтру попередньої обробки. Однак, найкращі результати досягли моделі, що використовували білатеральний фільтр та фільтр CLAHE для попередньої обробки зображень.

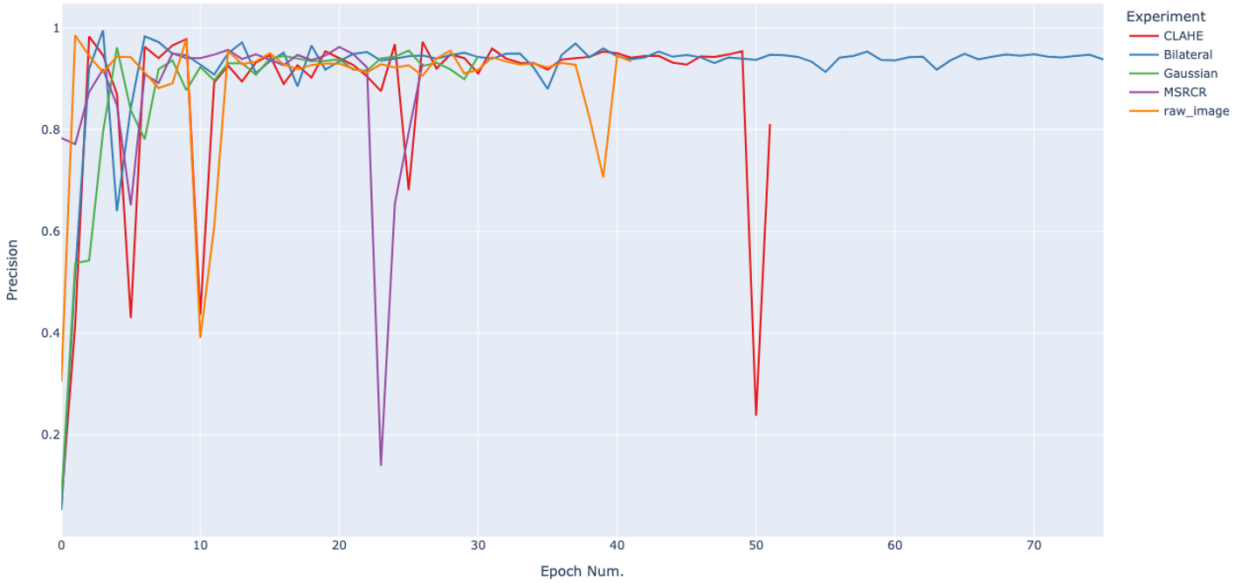


Рис. 2.10 Графік порівняння моделей фільтрів за влучністю (Precision)

З рисунку, на якому показано зміну влучності (precision) моделей сегментації, видно, що найбільш стабільний результат продемонструвала модель, яка використовувала білатеральний фільтр для попередньої обробки зображень.

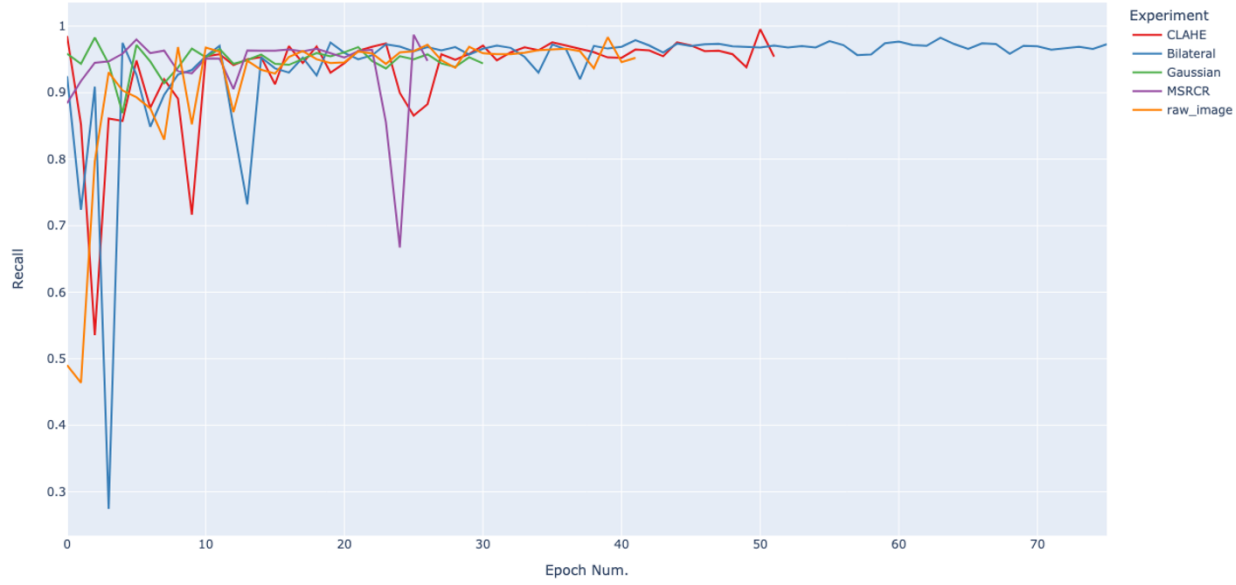


Рис. 2.11 Графік порівняння моделей за повнотою (Recall)

На рисунку представлені графіки, які показують порівняння сегментаційних моделей за метрикою повноти (recall).

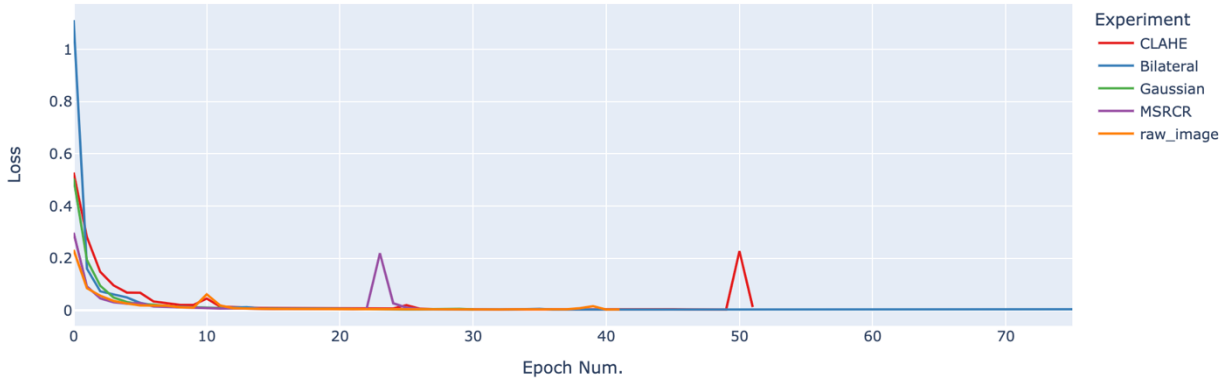


Рис. 2.12 Графік порівняння моделей за функцією втрат

Графіки функцій втрат, представлені на рисунку 6, демонструють послідовне зниження втрат для всіх моделей протягом процесу навчання. Це свідчить про загальну ефективність вибраних параметрів навчання. Однак, фільтр MSRCR показує ранній сплеск втрат, що пояснює, чому модель, яка використовує цей фільтр, передчасно завершує навчання. Також спостерігається помітний стрибок втрат для фільтра CLAHE, що вказує на можливі проблеми з перенавчанням або чутливістю до певних гіперпараметрів.

На рисунку представлено візуальні результати роботи моделі з білатеральним фільтром. Ця модель постійно створювала чудові сегментаційні маски, навіть на шумних і низькоякісних стоматологічних зображеннях. Висока якість сегментації з білатеральним фільтром пояснюється його здатністю зберігати критично важливі деталі, знижуючи шум і підвищуючи точність ідентифікації стоматологічних пломб.

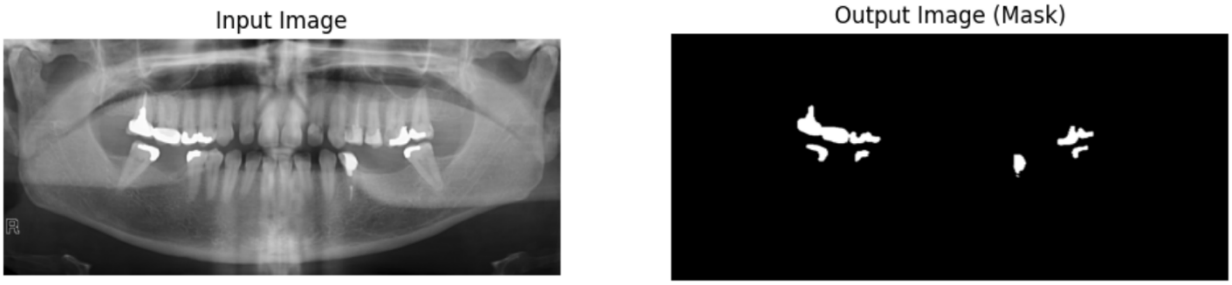


Рис. 2.13 Приклад роботи сегментаційної моделі із застосуванням білатерального фільтру на панорамному знімку з тестового набору даних

Візуальна оцінка результатів була проведена на тестовому наборі даних, який не входив до навчального або валідаційного наборів. Це підтверджує гарну генералізацію моделей, що є важливим показником їхньої здатності ефективно працювати з новими, невідомими даними.

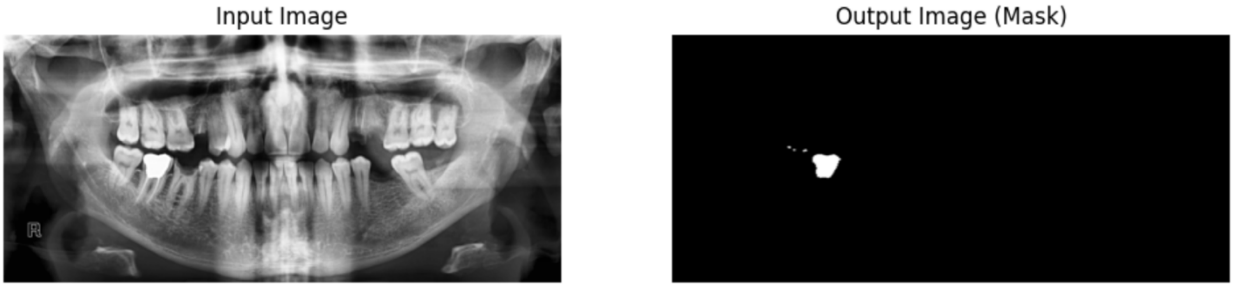


Рис. 2.14 Приклад «шуму» у результуючій масці сегментаційної моделі із застосуванням фільтру CLAHE на панорамному знімку з тестового набору даних

Під час візуальної оцінки результатів роботи моделей, було виявлено, що фільтр CLAHE, хоча й демонстрував високі результати у багатьох метриках, створював помітний «шум» у згенерованих масках сегментації (див. рис.). Цей шум проявлявся у вигляді небажаних артефактів, які могли впливати на точність сегментації. На відміну від цього, результати, отримані з використанням білатерального фільтру, не містили таких артефактів, що робить його значно кращим вибором для задачі сегментації стоматологічних пломб (див. рис.).

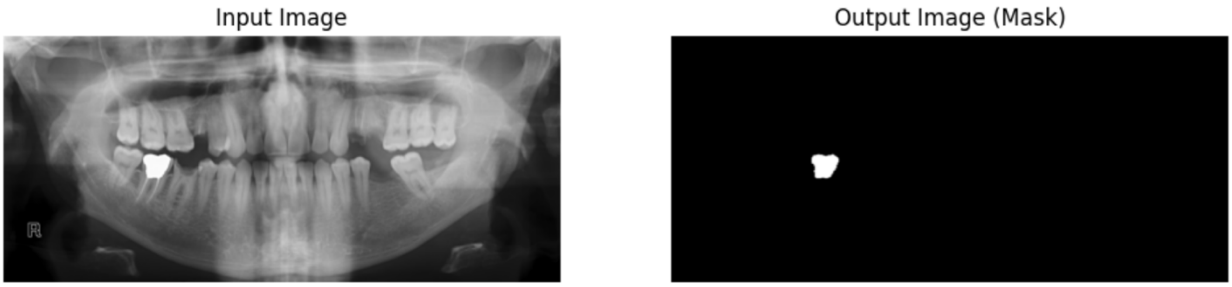


Рис. 2.15 Приклад результуючої маски сегментаційної моделі без «шуму» із застосуванням білатерального фільтру на панорамному знімку з тестового набору даних

Результати візуальної оцінки підтверджують, що білатеральний фільтр є найбільш оптимальним вибором для задач сегментації стоматологічних пломб. Він забезпечує чисті та чіткі маски сегментації без артефактів, що підвищує точність і надійність моделей. У той час як фільтр CLAHE також показує високі числові результати, наявність шуму у його зображеннях вказує на необхідність додаткової обробки для досягнення клінічно прийнятних результатів. Наступний розділ дослідження буде присвячено подальшому аналізу впливу методів попередньої обробки на інші метрики сегментації та надано рекомендації щодо їхнього використання у різних умовах застосування.

Це явище можна пояснити тим, що фільтр CLAHE (Contrast Limited Adaptive Histogram Equalization) призначений для покращення контрасту зображень шляхом перерозподілу інтенсивності пікселів, що робить окремі деталі більш виразними. Однак у певних випадках, зокрема при обробці медичних або стоматологічних зображень, підвищений контраст може спричиняти небажані артефакти. Ці артефакти можуть проявлятися у вигляді хибних візерунків у сегментаційних масках, що негативно впливає на загальну точність і надійність моделі.

Це спостереження акцентує увагу на необхідності враховувати не тільки кількісні показники при оцінці ефективності сегментаційних моделей, але й проводити візуальну перевірку. Візуальна оцінка є надзвичайно важливою для виявлення складних проблем, таких як «шум» у сегментаційних масках, який

може залишитися непоміченим, якщо спиратися виключно на метрики. Наприклад, фільтр CLAHE, хоча й продемонстрував високі результати за такими показниками, як коефіцієнт Дайса, коефіцієнт Жаккара, точність і повнота, схильний створювати ненавмисний шум, що ставить під сумнів його придатність для задач, де важлива висока точність сегментації.

Експериментальні результати та їх оцінка показують, що застосування різних фільтрів попередньої обробки медичних зображень вимагає ретельного вибору та налаштування параметрів під конкретну задачу. Фільтр CLAHE, хоча й забезпечує високі кількісні показники, може створювати артефакти, що знижують якість сегментації. Білатеральний фільтр, завдяки своїй здатності знижувати шум без створення артефактів, є найбільш оптимальним вибором для задач, де точна сегментація має вирішальне значення. Рекомендується використовувати комплексний підхід до оцінки ефективності моделей, включаючи як кількісні метрики, так і візуальну оцінку, для забезпечення найкращих результатів у клінічній практиці.

Підсумовуючи, в даній роботі було проведено всебічний аналіз, головною метою якого була оцінка ефективності використання фільтрів для попереднього оброблення панорамних стоматологічних рентгенівських знімків у задачі сегментації зубних пломб. Дослідження включало кілька важливих етапів і завдань:

1. Тренування сегментаційних моделей: Було обрано та налаштовано кілька найпопулярніших методів попередньої обробки зображень, включаючи обмежене адаптивне вирівнювання гістограми (CLAHE), фільтр Гауса, білатеральний фільтр і Multi-Scale Retinex з відновленням кольорів (MSRCR). Кожен з цих методів було ретельно підібрано та налаштовано для покращення характеристик зображень та підвищення їх якості, що є критичним для точного сегментаційного аналізу.

2. Вибір архітектури моделі: Для задачі сегментації стоматологічних пломб була обрана архітектура U-Net, відома своєю здатністю ефективно працювати з

медичними зображеннями. Ця архітектура дозволяє моделі вивчати складні патерни та забезпечує високу точність сегментації.

3. Оцінка точності сегментації: Було проведено аналіз та вибір відповідних метрик для оцінки точності сегментації, включаючи коефіцієнт Дайса, коефіцієнт Жаккара, влучність (precision) та повнота (recall). Ці метрики дозволили комплексно оцінити ефективність моделей з різними фільтрами попередньої обробки.

4. Порівняльний аналіз ефективності фільтрів: Було проведено детальний порівняльний аналіз ефективності кожного фільтра попередньої обробки для задачі сегментації стоматологічних пломб. Найкращий результат продемонструвала модель, що використовувала білатеральний фільтр. Ця модель перевершила інші за трьома з чотирьох основними метриками демонструючи високу стабільність, точність та загальну ефективність.

5. Особливості та переваги різних фільтрів:

- Білатеральний фільтр: Цей фільтр показав найкращі результати у більшості метрик, забезпечуючи чисті та чіткі сегментаційні маски без артефактів. Він продемонстрував високу стабільність та точність навіть на шумних та низькоякісних зображеннях.

- CLAHE: Фільтр CLAHE показав найвищі результати за повнотою (recall), що є важливим для задач, де критично важливо не пропустити жодного цільового об'єкта. Проте, цей фільтр створював певний рівень шуму та артефактів, що потребує додаткової уваги при його використанні.

6. Візуальна оцінка результатів: Візуальна оцінка результатів роботи моделей підкреслила важливість поєднання кількісних метрик та візуальної перевірки. Результати роботи з білатеральним фільтром показали відсутність артефактів та високу якість сегментаційних масок, що підтвердило його переваги над іншими методами.

Таким чином, це дослідження показало, що вибір правильного фільтра попередньої обробки є ключовим фактором для успішного виконання задач

сегментації стоматологічних пломб, а білатеральний фільтр у даному випадку є найкращим вибором для досягнення найвищої якості результатів.

2.3 Удосконалення методів попередньої обробки даних

Методи попередньої обробки медичних зображень, які використовувались раніше, мали певні обмеження, що впливали на точність подальшої сегментації та аналізу. Наприклад, методи, такі як білатеральний фільтр і фільтр CLAHE, за своєю природою статичні: їхні параметри не змінюються в залежності від локальних характеристик зображення, що призводить до недостатнього зменшення шуму або втрати важливих деталей у різних областях зображення. Це, своєю чергою, негативно впливало на точність розпізнавання об’єктів на зображеннях, особливо у випадках, коли зображення отримані з різних джерел із різною якістю.

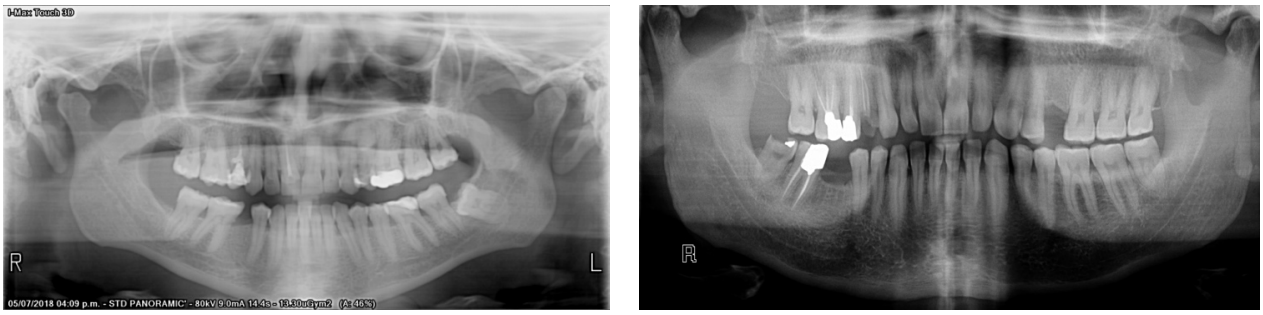


Рис. 2.16 Порівняння рентгенівських знімків з різних джерел із варіаціями у яскравості, контрасті та рівні шуму [57]

Така ситуація може призводити до зниження точності алгоритмів комп’ютерного зору або появи артефактів, що будуть спотворювати результати.

Одним із суттєвих недоліків білатерального фільтра є те, що його параметри згладжування за просторовими та інтенсивними характеристиками не враховують локальні особливості зображення. Це означає, що одночасне згладжування різних областей із високою та низькою деталізацією може призвести до втрати контурів або недостатнього усунення шуму. Подібна проблема виникає і з CLAHE, коли фіксований розмір блоку та обмеження на

рівень вирівнювання можуть спричиняти перенасичення контрасту або втрату важливих текстур у зображеннях із різним рівнем контрастності.

Щоб вирішити ці проблеми, були розроблені адаптивні методи попередньої обробки, які автоматично налаштовують параметри згладжування та вирівнювання контрасту на основі локальних особливостей зображення. Наприклад, адаптивний білатеральний фільтр коригує параметри згладжування залежно від градієнта інтенсивності, що дозволяє зберігати деталі в межах, таких як контури зубів на стоматологічних рентгенівських знімках, при цьому зменшуючи шум у рівномірних ділянках.

CLAHE також був вдосконалений за рахунок динамічного налаштування розміру сітки та ліміту вирівнювання відповідно до текстурних особливостей зображення. Це допомагає уникати перенасичення контрасту в деталізованих ділянках та покращує видимість у темних або зонах з низьким контрастом.

Адаптивні методи, які враховують контекст зображення, дозволяють гнучко регулювати параметри фільтрів, підвищуючи ефективність порівняно з традиційними підходами, де параметри залишаються сталими для всього зображення. Це критично для медичних знімків, де важливо зберігати тонкі деталі, як-от контури та текстури, водночас знижуючи рівень шуму і покращуючи контраст ключових елементів. Нижче наводиться математичне обґрунтування цих удосконалень.

Білатеральний фільтр — це метод згладжування зображень, який дозволяє зберегти чіткість країв, що особливо важливо для медичної візуалізації, де необхідно підтримувати точність анатомічних контурів. Цей фільтр поєднує у собі врахування просторової близькості та схожості інтенсивності пікселів, що дозволяє обчислювати середньозважене значення інтенсивності кожного пікселя, беручи до уваги як його положення, так і фотометричну відстань до сусідніх пікселів.

Білатеральний фільтр для пікселя $I(x)$ можна виразити за допомогою формули [58]:

$$I_{filtered}(x) = \frac{1}{W(x)} \sum_{y \in \Omega} I(y) \cdot \exp\left(-\frac{||x-y||^2}{2\sigma_s{}^2}\right) \cdot \exp\left(-\frac{|I(x)-I(y)|^2}{2\sigma_r{}^2}\right), \qquad (2.5)$$

де:

- $I(x)$ - інтенсивність пікселя x,
- $I(y)$ – інтенсивність пікселя в позиції y, де y – це сусідня точка навколо x в області Ω,
- x – це піксель, для якого проводиться фільтрація,
- y – це сусідній піксель в області Ω, яка охоплює пікселі навколо x.
- $W(x)$ є коефіцієнтом нормування,
- σ_s контролює просторову протяжність фільтра,
- σ_r контролює діапазон подібності інтенсивності.

Білатеральний фільтр працює наступним чином:

1. Просторовий фільтр (компонент з σ_s) зменшує вплив пікселів, які знаходяться далеко від точки x. Чим ближче сусідній піксель y, тим більше він впливає на результат фільтрації.

2. Радіометричний фільтр (компонент з σ_r) зменшує вплив сусідніх пікселів, які мають значну різницю в інтенсивності з пікселем x. Це важливо для збереження контрастних країв, оскільки пікселі з різкою зміною інтенсивності мають менший вплив.

3. Після множення інтенсивностей пікселів $I(y)$ на їх ваги, результат нормується за допомогою нормувального фактора $W(x)$, щоб отримати збалансоване значення інтенсивності після фільтрації.

У традиційному білатеральному фільтрі параметри σ_s та σ_r фіксуються по всьому зображенню [59]. Однак, як було розглянуто попередньо, медичні зображення часто містять області з різними характеристиками, наприклад області з дрібними деталями та інші з більш однорідними текстурами. Щоб вирішити цю проблему, пропонується адаптивне налаштування білатерального фільтра на основі контексту зображення, що змінює ці параметри на основі локальних властивостей зображення.

Пропонується впровадити адаптивний σ_s, що дозволяє фільтру динамічно регулювати свій просторовий вплив залежно від величини градієнта в локальній області. Вищі градієнти, що вказують на краї, передбачають менший σ_s щоб зберегти різкість країв.

$$\sigma_s(x) = \sigma_{s0} \cdot \left(1 - \frac{||\nabla I(x)||}{\max(||\nabla I(x)||)}\right), \tag{2.6}$$

де:

- σ_{s0} – базовий або початковий параметр згладжування,
- $\nabla I(x)$ – градієнт інтенсивності зображення в точці x,
- $||\nabla I(x)||$ – величина градієнта в пікселі x, яку можна обчислити за допомогою операторів *Sobel* або подібних,
- $max(||\nabla I(x)||)$ – максимальне значення градієнта по всьому зображенню.

Покращений білатеральний фільтр працює наступним чином:

1. Формула адаптує $\sigma_s(x)$ залежно від величини градієнта. Якщо градієнт в точці x великий (наприклад, точка належить до краю об’єкта), то параметр $\sigma_s(x)$ зменшується, щоб зменшити згладжування, зберігаючи важливі деталі, такі як краї.
2. Якщо градієнт малий (область без значної зміни інтенсивності, як-от в однорідних частинах зображення), то $\sigma_s(x)$ наближається до σ_{s0}, дозволяючи більш сильне згладжування.

Так само σ_r регулюється на основі локального рівня шуму, який можна визначити з відхилення в межах локального вікна. Регіони з більшою дисперсією (що вказує на шум) отримують більшу σ_r, покращуючи зниження шуму.

$$\sigma_r(x) = \sigma_{r0} \cdot \frac{\mathrm{Var}(I(y \in \Omega))}{\max(\mathrm{Var}(I(y \in \Omega)))}, \tag{2.7}$$

де:

- σ_{r0} – базове (початкове) значення параметра, яке потім коригується залежно від варіації інтенсивності в області навколо пікселя x,
- $\mathrm{Var}(I(y \in \Omega))$ – варіація (дисперсія) інтенсивності пікселів у вікні Ω, яке охоплює область навколо пікселя x,
- $\max(\mathrm{Var}(I(y \in \Omega)))$ – максимальне значення варіації інтенсивності серед усіх областей зображення.

Покращення працює наступним чином:

1. Якщо в області Ω навколо пікселя x є висока варіація інтенсивності (наприклад, область з контрастними елементами або краями), тоді значення $\sigma_r(x)$ збільшується, що дозволяє радіометричному фільтру зберігати ці зміни інтенсивності та уникати згладжування важливих деталей.
2. Якщо варіація в області мала (область однорідна за інтенсивністю), тоді значення $\sigma_r(x)$ буде зменшуватись, що дозволяє більшу кількість згладжування.

Отже, запропоновані адаптивні налаштування дозволяють білатеральному фільтру бути більш гнучким і реагувати на локальний контекст зображення. Ця гнучкість зберігає дрібні деталі в складних областях (наприклад, краї зубних пломб на перетині з іншими зубами), одночасно ефективно згладжуючи однорідні ділянки та зменшуючи шум, де це необхідно (наприклад, на краях зображення).

Фільтр CLAHE покращує традиційне вирівнювання гістограми, застосовуючи локальне підвищення контрасту, що допомагає уникнути надмірного підсилення шуму. Проте ефективність цього методу значною мірою залежить від таких параметрів, як обмеження підсилення (clip limit) та розмір блоків, які зазвичай залишаються незмінними для всього зображення. CLAHE працює за принципом обчислення кумулятивної функції розподілу інтенсивності пікселів у межах кожного локального блоку і подальшого відображення цих значень для покращення контрасту зображення.

Що стосується фільтру CLAHE, то він покращує стандартне вирівнювання гістограми, застосовуючи локальне посилення контрасту, таким чином запобігаючи надмірному посиленню шуму. Однак ефективність CLAHE значною мірою залежить від таких параметрів, як *clip limit* та розмір сітки фрагментів, які зазвичай є фіксованими для всього зображення.

Метод CLAHE працює шляхом обчислення кумулятивної функції розподілу інтенсивності пікселів у межах локального блоку, а потім відображення цих інтенсивностей для підвищення контрастності. Базове перетворення CLAHE для пікселя $I(x)$ визначається за формулою [60]:

$$I_{clahe}(x) = \frac{CDF(I(x)) - CDF_{min}}{CDF_{max} - CDF_{min}} \times \left(I_{max} - I_{min}\right) + I_{min}, \qquad (2.8)$$

де:

- $CDF(I(x))$ є кумулятивним значенням функції розподілу (Cumulative Distribution Function, CDF) при інтенсивності $I(x)$,
- CDF_{min} та CDF_{max} мінімальне та максимальне значення кумулятивної функції розподілу у блоці,
- I_{max} та I_{min} максимальна та мінімальна інтенсивності на зображенні (зазвичай 0 або 255 для 8-бітних зображень).

Як це працює:

1. Формула виконує нормалізацію значення інтенсивності $I(x)$ відносно кумулятивної функції розподілу $CDF(I(x))$ в локальному вікні. Це дозволяє регулювати контраст зображення на локальному рівні (в межах маленького регіону зображення), уникаючи глобального вирівнювання, яке може призвести до надмірного посилення шуму або викривлення контрасту.

2. Далі результуюче значення масштабується до діапазону інтенсивностей від I_{min} до I_{max} з метою відновлення яскравості після нормалізації.

3. CLAHE обмежує вирівнювання гістограми, що означає, що при дуже високому локальному контрасті (наприклад, у краях або сильно варіативних текстурованих областях) контраст не буде надмірно збільшено.

Щоб підвищити ефективність фільтру CLAHE, запроваджуємо адаптивність параметру *clip limit* та розміру сітки фрагментів на основі локальних характеристик зображення.

$$\text{ClipLimit}(x) = \text{ClipLimit}_0 \cdot \frac{\text{Mean}(I(y \in \Omega))}{\max(\text{Mean}(I(y \in \Omega)))}, \tag{2.9}$$

де:

- ClipLimit_0 це базове значення *clip limit*,
- $\text{Mean}(I(y \in \Omega))$ є середньою інтенсивністю в локальному сусідстві Ω.

Розмір сітки визначає розмір локальних блоків, до яких застосовується вирівнювання гістограми. У зонах із високою складністю текстур краще використовувати менші блоки для збереження деталей. І навпаки, на однорідних ділянках більші блоки допомагають уникнути надмірної обробки.

$$\text{TileSize}(x) = \text{TileSize}_0 \cdot \left(1 - \frac{\text{Texture}(I(y \in \Omega))}{\max(\text{Texture}(I(y \in \Omega)))}\right), \tag{2.10}$$

де:

- TileSize_0 — розмір базової сітки блоку,
- $\text{Texture}(I(y \in \Omega))$ є мірою складності текстури, яку можна оцінити за допомогою таких показників, як локальна дисперсія або ентропія.

Адаптивні стратегії, що пропонуються, забезпечують оптимальне покращення контрастності в різних ділянках зображення. В областях, де важливо зберегти дрібні деталі, адаптивний розмір сітки дозволяє підвищити контраст без створення артефактів. Одночасно адаптивне налаштування обмежень clip limit

запобігає перенасиченню яскравих зон та покращує видимість темних областей, що робить зображення більш інформативним для діагностичних потреб.

Крім того, пропонується комбінований підхід, що включає адаптивну білатеральну фільтрацію та модифікований метод CLAHE, як ефективне рішення для попередньої обробки медичних зображень. Ця комбінація ефективно оптимізує зниження шуму та підвищення контрасту, враховуючи специфіку зображення. Адаптивні властивості цих методів забезпечують їх ефективну роботу з різноманітними зображеннями з різних джерел, де критично важливим є одночасне зниження шуму та збереження важливих деталей для подальшого аналізу.

Отже, пропонується наступне комбіноване використання покращених методів попередньої обробки зображень:

1. Зниження шуму за допомогою запропонованої адаптивної білатеральної фільтрації: застосовується для зменшення шуму, зберігаючи межі та динамічно регулюючи параметри на основі локального градієнта та характеристик шуму.

2. Підвищення контрастності за допомогою модифікованого CLAHE на основі контексту зображення: застосовується для підвищення контрастності, оптимізуючи обмеження *clip limit* та розмір сітки відповідно до локальної яскравості та текстури зображення.

Нехай I_{orig} буде оригінальним зображенням, тоді розроблені покращенні етапи попередньої обробки можна підсумувати так:

$$I_{filtered} = \text{AdaptiveBilateralFilter}\left(I_{orig}, \sigma_s(x), \sigma_r(x) \right), \tag{2.11}$$

$$I_{enhanced} = \text{ModifiedCLAHE}\left(I_{filtered}, \text{ClipLimit}(x), \text{TileSize}(x) \right). \tag{2.12}$$

Запропоноване використання покращених методів попередньої обробки гарантуватиме, що зображення спочатку згладжується таким чином, щоб

зберегти важливі деталі, а потім його контраст посилюється таким чином, щоб підтримувати цілісність зображення та виділення головних деталей.

Отже, запропоновані вдосконалені методи попередньої обробки медичних зображень, підкріплені надійною математичною основою, забезпечують суттєві покращення порівняно з традиційними підходами. Завдяки динамічному налаштуванню параметрів відповідно до локального контексту зображення, ці методи демонструють високу ефективність у зниженні шуму, збереженні країв та покращенні контрасту. У подальшому ми детально розглянемо, як саме ці покращені методи перевершують традиційні підходи, особливо в контексті медичної візуалізації, де якість і точність зображень мають критичне значення для результатів алгоритмів комп'ютерного зору.

2.4 Удосконалення базової моделі для розпізнавання об’єктів на зображеннях за допомогою покращених методів попередньої обробки даних

У ході дослідження було проведено ґрунтовний аналіз ефективності новітніх методів попередньої обробки медичних зображень, зокрема адаптивного підлаштування параметрів залежно від контексту зображення. Основну увагу було приділено сегментації панорамних стоматологічних рентгенівських знімків, використовуючи методики, розроблені на основі попередніх напрацювань авторів [50]. Ці підходи були адаптовані для поліпшення якості попередньої обробки даних, що відіграє вирішальну роль у подальшому покращенні результатів сегментації.

Дослідження передбачало експериментальну оцінку впливу покращених методів попередньої обробки на якість сегментації зображень, порівнюючи їх з традиційними підходами. Особливий акцент був зроблений на аналізі ряду основних метрик, таких як Коефіцієнт Дайса (Dice Score), Коефіцієнт Жаккара (Jaccard Index або IoU), Влучність (Precision) та Повнота (Recall). Ці метрики дозволяють повноцінно оцінити ефективність натренованих моделей

сегментації, а також їхню здатність точно визначати структури на знімках після застосування різних методів попередньої обробки.

Для проведення експериментів у цьому дослідженні було використано модель U-Net [32], яка є однією з найпоширеніших і найефективніших архітектур для вирішення задач сегментації зображень. Вибір саме цієї моделі був обумовлений її здатністю успішно обробляти зображення на різних масштабах, що є критично важливим для медичних зображень. Модель дозволяє зберегти точну деталізацію анатомічних структур, водночас забезпечуючи високу якість сегментації навіть для найменших об'єктів.

У ході експериментів архітектура моделі та її налаштування були збережені відповідно до попередніх досліджень [50], щоб забезпечити порівнюваність результатів і оцінити вплив саме покращених методів попередньої обробки зображень. Усі вхідні зображення були масштабовані до розміру 512x512 пікселів, що дозволило досягти оптимального балансу між рівнем деталізації та ефективністю обчислювальних ресурсів.

Для підвищення точності сегментації використовувався передтренований енкодер на базі timm-resnest101e [54]. Цей енкодер, завдяки своїй глибокій архітектурі та тренуванню на великому наборі зображень, забезпечив високу якість екстракції ознак, що значно покращило загальну продуктивність моделі.

Для стабільності процесу навчання було застосовано фіксований learning rate на рівні 0.001, що було визначено оптимальним для даних умов експерименту. Крім того, для покращення генералізації моделі активно використовувалися методи аугментації зображень [56], що дозволило моделі бути більш стійкою до різних варіацій вхідних даних і покращило загальні результати сегментації.

Збереження незмінних параметрів експерименту дозволило порівняти результати, отримані за допомогою адаптивних методів попередньої обробки, з результатами попередніх досліджень, у яких застосовувалися традиційні методи обробки [50]. Це забезпечило об'єктивність і надійність висновків та підкреслило переваги нових підходів у контексті медичної сегментації, зокрема їхню здатність

покращувати якість сегментації, особливо у випадках складних і нерівномірних зображень.

Далі наведено табл. 2.2, яка відображає результати застосування розроблених адаптивної білатеральної фільтрації та покращеного CLAHE у порівнянні з традиційними методами попередньої обробки

Таблиця 2.2 Порівняльний аналіз ефективності адаптивних і традиційних методів попередньої обробки в задачах медичної сегментації

Назва фільтру	**Dice Score**	**IoU Score**	**Precision**	**Recall**	**Loss**
Традиційний білатеральний фільтр	0.94493	0.928249	0.995193	0.982434	0.0031
Традиційний фільтр CLAHE	0.926958	0.912003	0.982922	0.995365	0.0036
Адаптивний білатеральний фільтр	0.95992	0.93001	0.99801	0.982678	0.0027
Покращений фільтр CLAHE	0.94812	0.92995	0.99567	**0.997898**	0.0031
Комбінований підхід (адаптивний білатеральний фільтр + покращений CLAHE фільтр)	**0.9603**	**0.94501**	**0.9991**	0.985397	**0.0026**
Без фільтру	0.912757	0.90474	0.985562	0.98341	0.0037

Аналіз результатів, представлених у таблиці 1, демонструє значний вплив адаптивних методів попередньої обробки на підвищення ефективності сегментації медичних зображень. Зокрема, використання адаптивного білатерального фільтра забезпечило найвищі значення показників Dice Score та IoU Score, досягаючи 0.95992 та 0.93001 відповідно. Ці результати суттєво перевищують показники, отримані при використанні традиційних методів попередньої обробки, таких як класичний білатеральний фільтр та традиційний CLAHE. Наприклад, для традиційного білатерального фільтра Dice Score становив 0.94493, а IoU Score — 0.928249 (рис. 2.16). Аналогічно, традиційний CLAHE фільтр демонстрував Dice Score на рівні 0.926958 та IoU Score 0.912003 (рис. X).

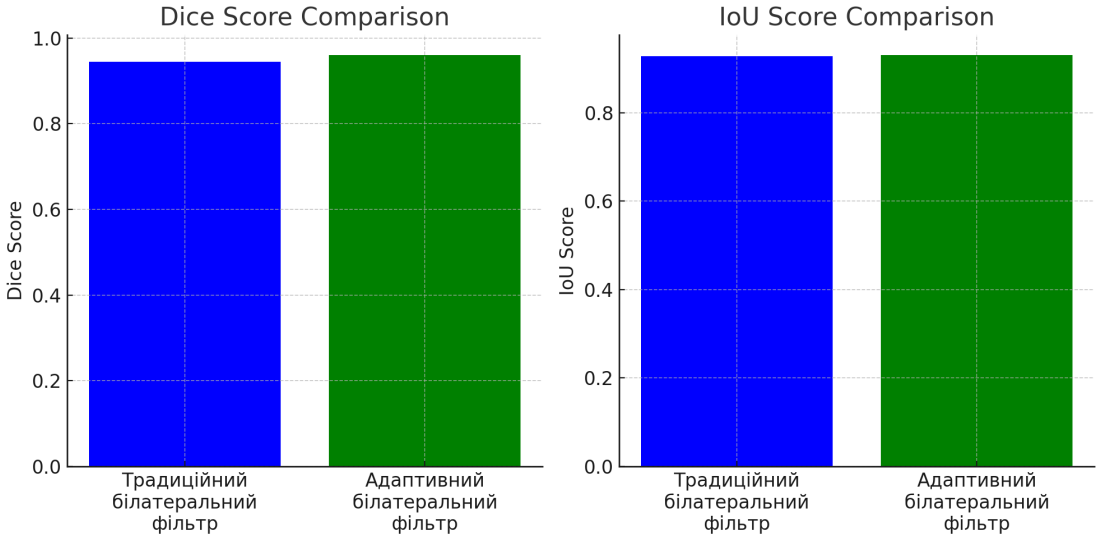


Рис. 2.16 Порівняння показників традиційного білатерального фільтра та адаптивного білатерального фільтра за метриками Dice Score та IoU Score

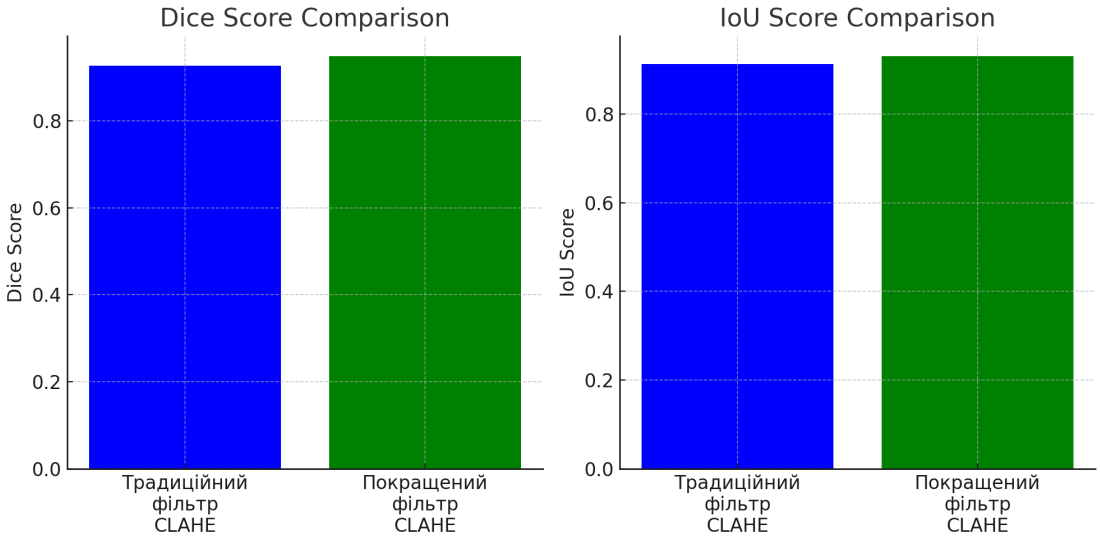


Рис. 2.17 Порівняння показників традиційного фільтра CLAHE та покращеного фільтра CLAHE за метриками Dice Score та IoU Score

Адаптивний підхід до попередньої обробки показав значно кращі результати за показниками Precision та Recall, що вказує на ефективніше усунення шуму при збереженні критично важливих деталей зображення. Зокрема, Precision для адаптивного білатерального фільтра становив 0.99801, що вказує на мінімальну кількість хибнопозитивних спрацювань, а Recall залишився

на високому рівні — 0.982678, що демонструє здатність методу захоплювати всі важливі структури на зображенні (рис. 2.18).

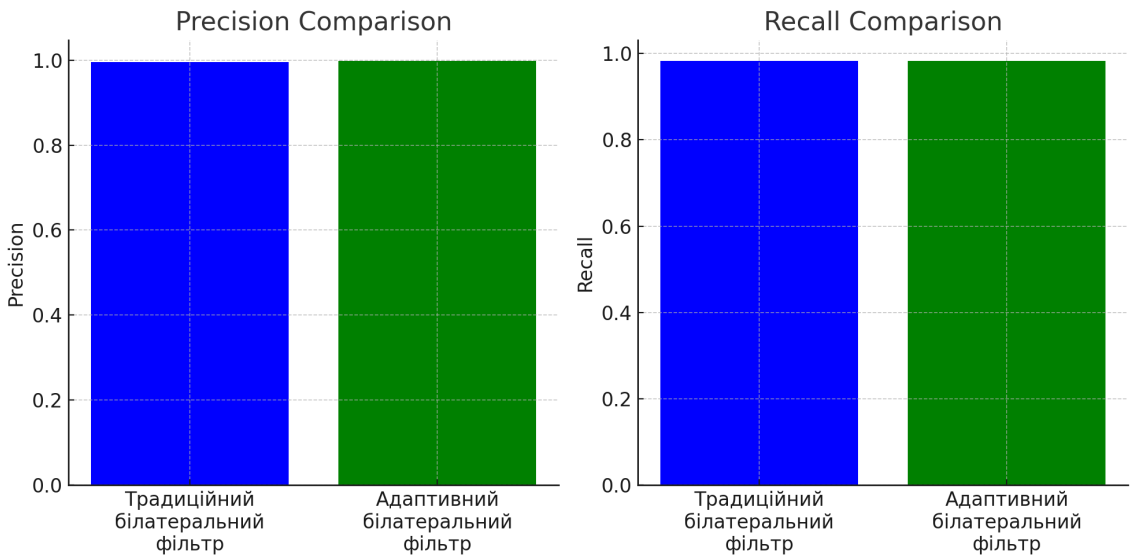


Рис. 2.18 Порівняння показників традиційного білатерального фільтра та адаптивного білатерального фільтра за метриками Precision та Recall

Цікаво зазначити, що використання покращеного CLAHE фільтра також продемонструвало значні покращення у порівнянні з традиційним підходом. Отримані значення Dice Score 0.94812 та IoU Score 0.92995 підтверджують ефективність цього методу у покращенні контрастності та збереженні структур зображення. Застосування покращеного CLAHE фільтра також дозволило досягти високих значень Precision (0.99567) та Recall (0.997898), що свідчить про його здатність ефективно підвищувати якість сегментації в критичних областях зображення.

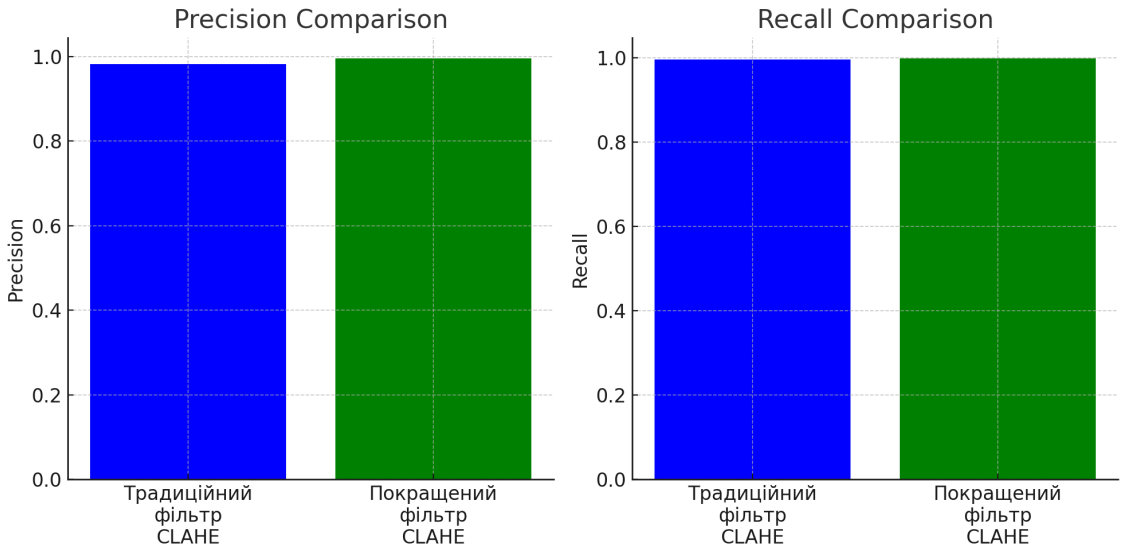


Рис. 2.19 Порівняння показників традиційного фільтра CLAHE та покращеного фільтра CLAHE за метриками Precision та Recall

Особливо примітними є результати, отримані при використанні комбінованого підходу, який включає адаптивний білатеральний фільтр та покращений CLAHE фільтр. Цей підхід дозволив досягти максимальних значень Dice Score та IoU Score серед усіх досліджених методів — 0.9603 та 0.94501 відповідно. Крім того, Precision у цьому випадку становив 0.9991, а Recall — 0.985397, що є найкращими результатами серед усіх представлених методів. Це вказує на те, що комбінований підхід не лише зберігає критичні деталі зображення, але й мінімізує кількість хибнопозитивних та хибнонегативних спрацювань.

Навпаки, відсутність фільтрації призвела до значного зниження якості сегментації, що проявилося у найнижчих значеннях усіх показників, зокрема Dice Score 0.912757 та IoU Score 0.90474 (рис. 2.20). Високе значення Loss (0.0037) також свідчить про нестабільність сегментації при відсутності попередньої фільтрації, що підтверджує важливість застосування попередньої обробки для підвищення якості результатів.

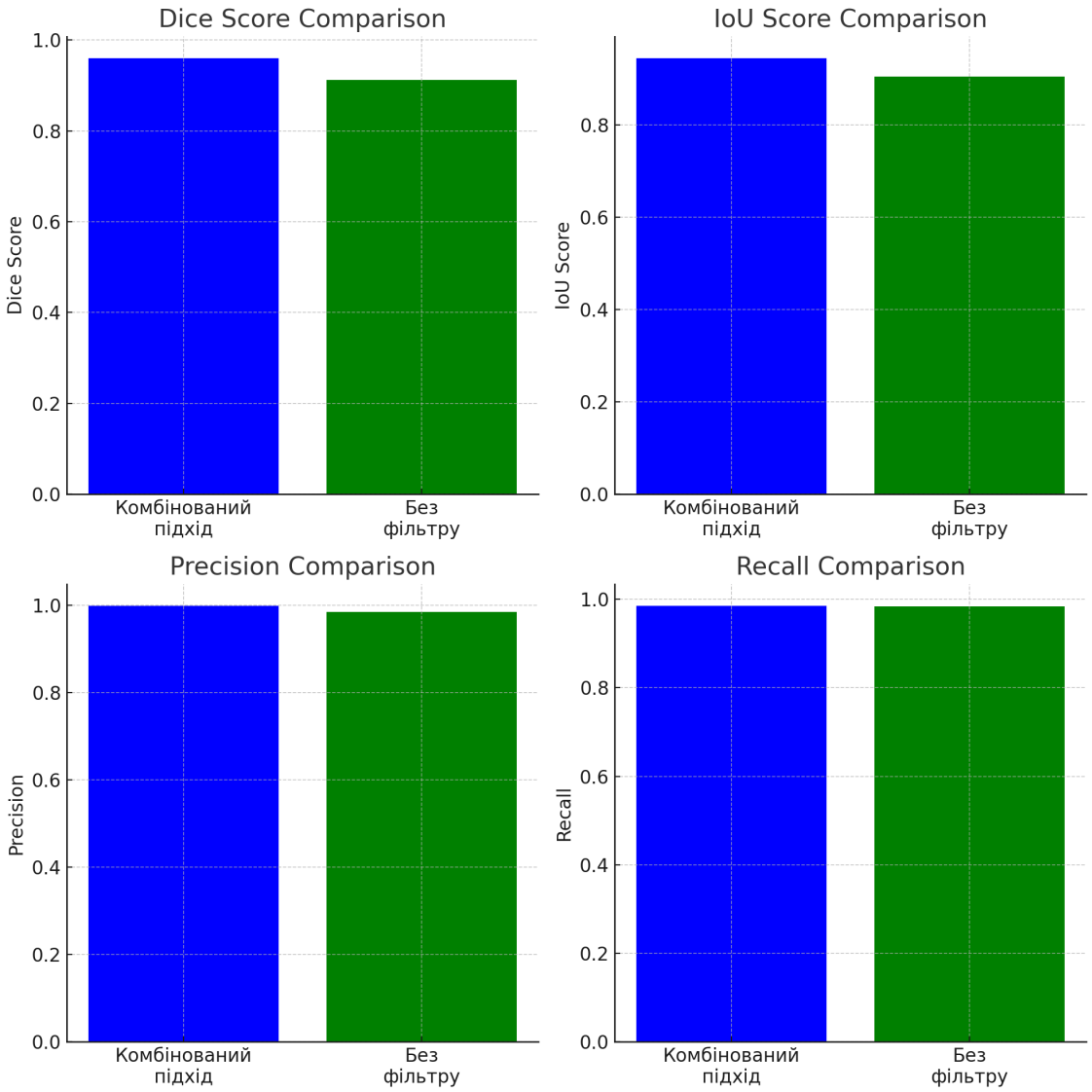


Рис. 2.20 Порівняння показників комбінованого підходу та використання без фільтрів за метриками Dice, IoU, Precision та Recall

Запропоновані в цьому дослідженні методи адаптивної попередньої обробки медичних зображень, зокрема удосконалений білатеральний фільтр і модифікований алгоритм CLAHE (Contrast Limited Adaptive Histogram Equalization), продемонстрували значне підвищення ефективності сегментації панорамних стоматологічних рентгенівських знімків. Комбіноване використання цих підходів дозволило досягти найкращих результатів за такими ключовими показниками, як Коефіцієнт Дайса (Dice Score), Коефіцієнт Жаккара (IoU Score), Влучність (Precision) та Повнота (Recall). Це підтверджує, що адаптивні методи не тільки зберігають важливі деталі зображення, але й ефективно зменшують

рівень шуму, що є критичним фактором у задачах сегментації, де точність визначення меж об'єктів безпосередньо впливає на якість діагностики.

Дослідження показало, що динамічне налаштування параметрів попередньої обробки на основі локальних особливостей зображень є ключовим чинником для досягнення високої точності сегментації у медичних застосуваннях. Оскільки медичні зображення можуть суттєво варіюватися за якістю і характеристиками через різні умови отримання, запропоновані адаптивні методи, зокрема вдосконалений білатеральний фільтр та покращений CLAHE, виявилися більш гнучкими та ефективними порівняно з класичними підходами.

Результати дослідження свідчать про те, що адаптивні методи попередньої обробки мають здатність значно покращувати якість сегментації медичних зображень, що у свою чергу сприяє підвищенню точності діагностики та зниженню ризику помилок. Оскільки від якості сегментації часто залежить успішність постановки діагнозу та вибору лікувальних заходів, впровадження таких методів у процеси аналізу медичних зображень може мати далекосяжні наслідки для покращення результатів медичної діагностики.

Таким чином, застосування адаптивних методів попередньої обробки відкриває нові перспективи для поліпшення ефективності та точності візуалізації у медичних системах. Подальші дослідження можуть бути спрямовані на оптимізацію цих методів для різних типів медичних зображень і їх інтеграцію з іншими алгоритмами комп'ютерного зору. Це може стати основою для створення комплексних систем автоматизованої підтримки прийняття рішень у медичній діагностиці, що в кінцевому результаті дозволить покращити якість лікування пацієнтів.

Перспективи впровадження запропонованих методів полягають у їхній універсальності та можливості адаптації до різних задач медичної візуалізації. Дослідження у цьому напрямку можуть допомогти розробити нові стандарти обробки медичних зображень, що сприятиме розвитку високоточних інструментів для аналізу, виявлення та лікування різноманітних патологій.

2.5 Висновки до розділу

У другому розділі було виконано всебічний аналіз сучасних методів попередньої обробки та моделей для розпізнавання об'єктів на зображеннях з використанням псевдомасок. Цей підхід став ефективним рішенням для підвищення точності моделей глибокого навчання в умовах малорозмірних або зашумлених наборів даних.

По-перше, було вдосконалено методи попередньої обробки зображень, такі як нормалізація, фільтрація шуму та вирівнювання гістограм. Застосування цих методів дозволило значно покращити якість зображень перед сегментацією, що підвищило точність та надійність моделей.

По-друге, у розділі запропоновано новий підхід до сегментації об'єктів на зображеннях із застосуванням псевдомасок. Висока ефективність цього підходу була досягнута завдяки використанню моделей глибокого навчання, що дозволило автоматично генерувати сегментаційні маски для слабко анотованих або неанотованих зображень.

По-третє, вдосконалено алгоритми адаптивної попередньої обробки даних, що забезпечує адаптацію параметрів обробки зображень залежно від їхніх локальних особливостей. Це дозволило значно знизити рівень шуму і зберегти важливі деталі, що позитивно вплинуло на якість сегментації та подальше розпізнавання об'єктів.

Результати роботи опубліковані в таких роботах автора: [50], [67], [68], [72]

РОЗДІЛ 3
РОЗРОБЛЕННЯ АЛГОРИТМІЧНИХ ОСНОВ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК

3.1 Розроблення алгоритмів роботи запропонованої інтелектуальної інформаційної технології

У сучасній епохі розвитку комп'ютерного зору та машинного навчання, особливо в контексті завдань з вчителем (supervised learning), процес анотування медичних зображень відіграє ключову роль у підготовці навчальних даних. Це особливо важливо в задачах сегментації, де високоякісні анотації об'єктів необхідні для навчання моделей, що використовуються для діагностики, прогнозування та автоматизації медичних процесів. У завданнях сегментації навіть незначні похибки в анотуванні можуть суттєво вплинути на точність прогнозів, що робить цей процес критичним для наукових та клінічних застосувань.

Проблематика ручного анотування значно ускладнює масштабування таких систем у випадку роботи з великими обсягами медичних зображень. Ручне анотування великих наборів даних, наприклад, у рентгенографії або комп'ютерній томографії, є надзвичайно трудомістким процесом, що вимагає участі кваліфікованих фахівців. Це не тільки збільшує вартість проектів, але й створює значні затримки в отриманні результатів, особливо коли йдеться про медичні дослідження та діагностику, де швидкість є вирішальним фактором [61]. У таких умовах ефективність ручного анотування стає питанням, особливо при роботі з великими обсягами даних або в задачах, що вимагають оперативного розпізнавання об'єктів.

З огляду на ці виклики, автоматизація анотування стає важливим напрямом сучасних досліджень у сфері штучного інтелекту, зокрема комп'ютерного зору. Використання автоматизованих методів анотування, таких як генерація

псевдомасок, може значно знизити затрати часу на ручну обробку та покращити якість анотацій. Однією з передових технологій у цій галузі є модель Segment-Anything (SAM), розроблена компанією Meta. Ця модель поєднує глибинне навчання та обробку природної мови, дозволяючи створювати псевдомаски для сегментації об'єктів на зображеннях за допомогою текстових підказок [62]. Такий підхід відкриває нові можливості для більш точної та швидкої обробки медичних зображень.

Рис. 3.1 Приклади роботи SAM

Language Segment-Anything дозволяє значно прискорити процес анотування, оскільки модель генерує псевдомаски автоматично на основі текстових інструкцій. Це дозволяє лікарям або дослідникам фокусуватися на уточненні результатів замість того, щоб вручну малювати маски для кожного зображення. Додатковою перевагою є можливість гнучкого налаштування псевдомасок за допомогою зміни текстових підказок, що дає можливість більш точно адаптувати модель до різних типів медичних даних і застосувань [63].

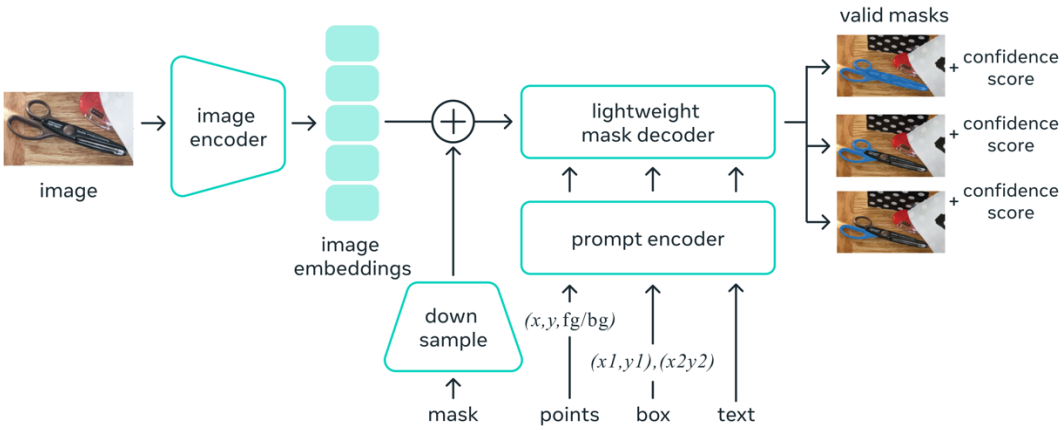


Рис. 3.2 Принцип роботи Language SAM з підтримкою текстових підказок

У цьому контексті була створена інтелектуальна інформаційна технологія, яка поєднує сучасні моделі та методи для автоматизованого анотування медичних зображень. Ключовим елементом є модель Language Segment-Anything [64], яка здатна генерувати псевдомаски на основі текстових підказок, що суттєво прискорює процес збору анотації зображень.

У рамках дослідження було виконано порівняльний аналіз ефективності використання розробленої інтелектуальної інформаційної технології у порівнянні з традиційними інструментами для анотування, такими як Supervisely, CVAT та Segments.ai. Оцінка включала аналіз середнього часу, необхідного для анотації одного зображення, а також точності псевдомасок відносно експертного анотування.

Також було проведено експерименти щодо впливу текстових підказок на точність створення псевдомасок. Ці експерименти продемонстрували, що правильний підбір текстових інструкцій може суттєво покращити точність масок для різних типів медичних зображень.

Завдяки інтеграції передових методів автоматизації, нова інформаційна технологія здатна значно покращити процес анотування медичних зображень. Це дозволить зосередитися на більш складних аспектах досліджень та розробок, зменшуючи витрати часу та людських ресурсів, необхідних для виконання рутинних завдань. Застосування таких технологій у практичних умовах має потенціал для суттєвого покращення якості медичної діагностики та лікування, що в кінцевому результаті може сприяти підвищенню загального рівня охорони здоров'я.

У цій роботі пропонується інноваційна інформаційна технологія для прискореного анотування медичних зображень в задачах сегментації. Основою цієї технології є використання передових моделей глибокого навчання, зокрема, моделі Language Segment-Anything. Ця модель використовує текстові підказки (prompts) для створення початкових псевдомасок, які потім автоматично перевіряються для підвищення точності анотування.

Процес автоматичної валідації цільового класу в межах обмежувальної рамки псевдомаски виконується за допомогою передтренованої моделі класифікації, побудованої на базі архітектури EfficientNet-B2 [78]. Ця модель демонструє високу точність і надійність у визначенні відповідності псевдомаски до цільового класу.

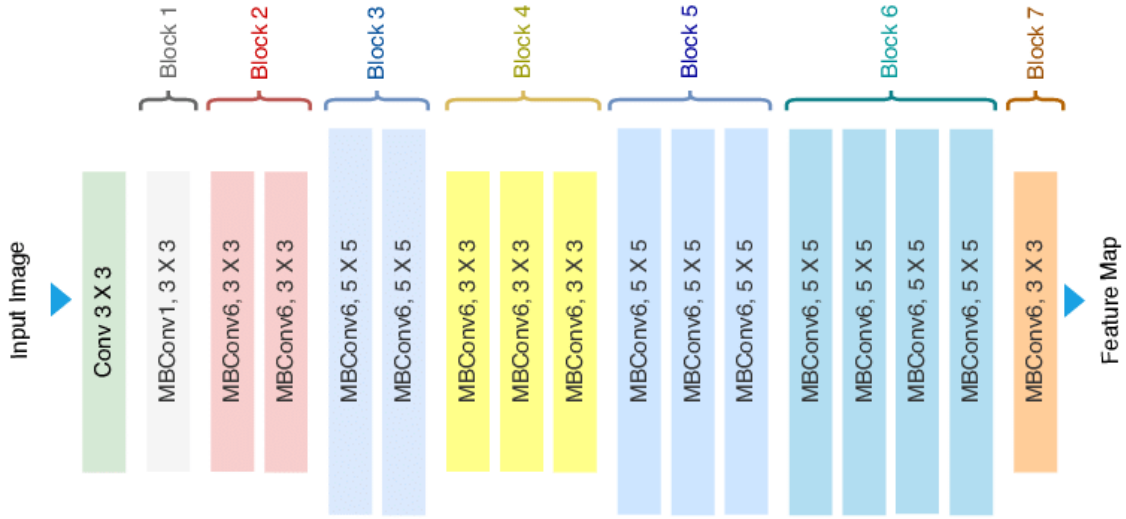


Рис. 3.3 Архітектура моделі EfficientNet [78]

Для тренування моделі класифікації були використані наступні гіперпараметри:

- Кількість епох: 100;
- Розмір міні-батча: 8;
- Оптимізатор: Adam;
- Коефіцієнт швидкості навчання (learning rate): 0.001.

Використання цих параметрів дозволило досягти оптимального балансу між швидкістю навчання та точністю моделі. Завдяки цьому, модель ефективно справляється з автоматичною валідацією псевдомасок, забезпечуючи високу якість анотування медичних зображень.

Основним компонентом розробленої технології є модель Language Segment-Anything (SAM), що використовує глибинні нейронні мережі для створення псевдомасок на основі текстових інструкцій. SAM здатна автоматично сегментувати об'єкти на медичних зображеннях, використовуючи попередньо налаштовані текстові підказки для точнішого визначення областей інтересу. Це дозволяє суттєво прискорити процес анотування та знизити навантаження на експертів, оскільки вони можуть скорочувати ручну працю та зосереджуватися на уточненні отриманих результатів.

Алгоритми роботи запропонованої інформаційної технології включають кілька ключових етапів:

1. Попередня обробка зображень: Перед застосуванням моделі SAM проводиться стандартна обробка вхідних медичних зображень, що включає нормалізацію, зміну розмірів та видалення шумів за допомогою розглянутих раніше фільтрів попередньої обробки. Це забезпечує однакову якість зображень для наступних етапів.

2. Генерація текстових підказок для моделі: Користувачі, такі як лікарі або дослідники, можуть ввести текстові підказки, які описують бажані об'єкти для сегментації. Наприклад, для рентгенівських знімків грудної клітки можна використовувати підказки на зразок "легені", "серце". Модель Language SAM автоматично інтерпретує ці підказки та генерує відповідні псевдомаски [64], [77].

3. Автоматичне створення псевдомасок: Після введення текстових підказок модель виконує сегментацію зображень, генеруючи псевдомаски для визначених об'єктів. Завдяки використанню глибоких нейронних мереж, SAM здатна точно відокремлювати об'єкти на зображеннях навіть за умов складної структури або наявності артефактів [77].

4. Оцінка та корекція результатів: Розроблено метод автоматичного валідування створених первинних псевдомасок на основі моделі класифікації EfficientNet-B2 [78]. Даний метод працює таким чином, що модель класифікації підтверджує, що створений сегмент первинної псевдомаски дійсно належить до цільового класу, в такому випадку анотація такого сегменту зберігається. Якщо ж модель класифікації не підтверджує належність - анотація відхиляється. Такий підхід значно знижує кількість ручної роботи, оскільки основна частина сегментації вже виконана автоматично.

5. Навчання моделі на коригованих даних: Кориговані псевдомаски можуть бути використані для подальшого навчання моделі. Завдяки циклічному процесу самонавчання, модель удосконалює свою здатність створювати більш точні маски з кожною новою ітерацією.

Запропонована технологія значно прискорює процес анотування, зменшуючи потребу в ручному втручанні та дозволяючи зосередитися на більш складних аспектах медичних досліджень. Автоматичне створення і валідація

псевдомасок сприяє підвищенню точності сегментації, що має вирішальне значення для подальшого аналізу та інтерпретації медичних зображень.

Крім того, інтеграція моделей глибокого навчання з текстовими підказками відкриває нові можливості для адаптації технології до різних типів медичних зображень і завдань. Це дозволяє широко застосовувати розроблену технологію в практичних умовах, зменшуючи час і зусилля, необхідні для високоякісного анотування даних.

Таким чином, дана інформаційна технологія не лише покращує ефективність процесу анотування, але й забезпечує більш точні та надійні результати, що є критично важливим для розвитку рівня охорони здоров'я.

3.2 Створення UML-діаграм програмного забезпечення для реалізації модулю розпізнавання об’єктів на зображеннях з використанням псевдомасок

У сучасних дослідженнях комп'ютерного зору, особливо в галузі медицини, процес анотування зображень для задач сегментації відіграє вирішальну роль. З метою прискорення та підвищення точності цього процесу, пропонується інноваційна інформаційна технологія, що використовує моделі глибокого навчання та автоматизовані методи валідації.

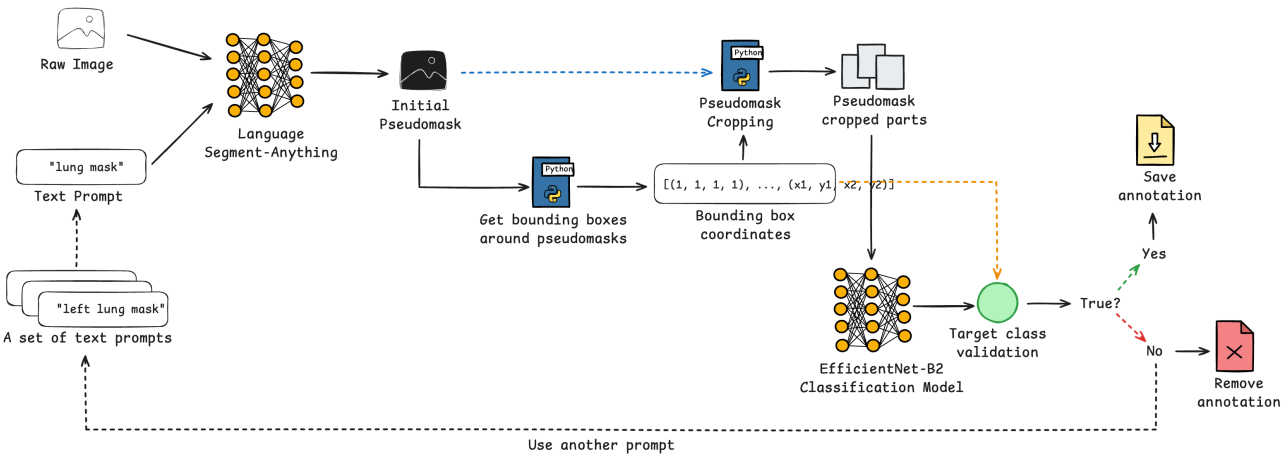


Рис. 3.4 Структура інформаційної технології прискореного анотування медичних зображень

Як показано на рисунку 3.4, перший крок у розробленій інтелектуальній інформаційній технології полягає у використанні моделі Language Segment-Anything для генерації початкових псевдомасок сегментації на основі текстових підказок. Ця модель може інтерпретувати текстові інструкції та створювати відповідні маски для об'єктів на зображеннях. Підбір оптимальних текстових підказок здійснюється експериментальним шляхом, щоб максимально точно відповідати маскам потрібним об'єктам. Такий підхід дозволяє моделі коректно виділяти цільові області на зображеннях. Основна увага при виборі підказок приділяється метриці повноти (recall), щоб охопити максимальну кількість цільових об'єктів, навіть якщо це може призвести до появи зайвих масок, схожих на цільові.

Наступний етап полягає у визначенні обмежувальних рамок навколо кожної псевдомаски, що дозволяє виокремити області для подальшої перевірки. Частини зображення, що знаходяться всередині цих рамок, передаються до розробленого методу автоматичного валідування створених псевдомасок, що базується на моделі класифікації EfficientNet-B2. Ця модель визначає, чи належить об'єкт у межах рамки до цільового класу. Якщо об'єкт не відповідає цільовому класу, відповідні анотації видаляються. Такий підхід дає змогу автоматично перевіряти та усувати зайві псевдомаски, що сприяє покращенню метрики точності (precision) сегментації.

Враховуючи наведену вище інформацію, створимо UML-діаграму послідовності роботи модулів інтелектуальної інформаційної технології, яка пропонується (див. рис. 3.5).

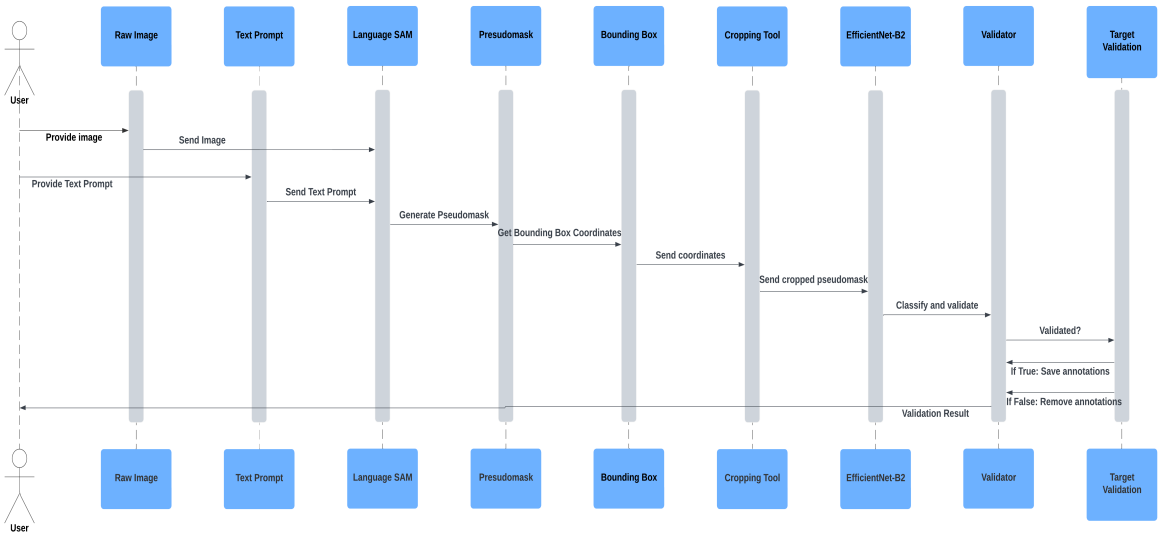


Рис. 3.5 UML-діаграма послідовності роботи модулів

Алгоритм автоматичного валідування є одним із основних елементів запропонованої технології, оскільки дозволяє використовувати попередньо згенеровані псевдомаски як нові навчальні дані. Це допомагає моделі поступово підвищувати свою точність і стійкість до різних варіацій даних. Крім того, цей підхід можна адаптувати до різних типів медичних зображень, використовуючи індивідуальні налаштування підказок для кожного конкретного завдання.

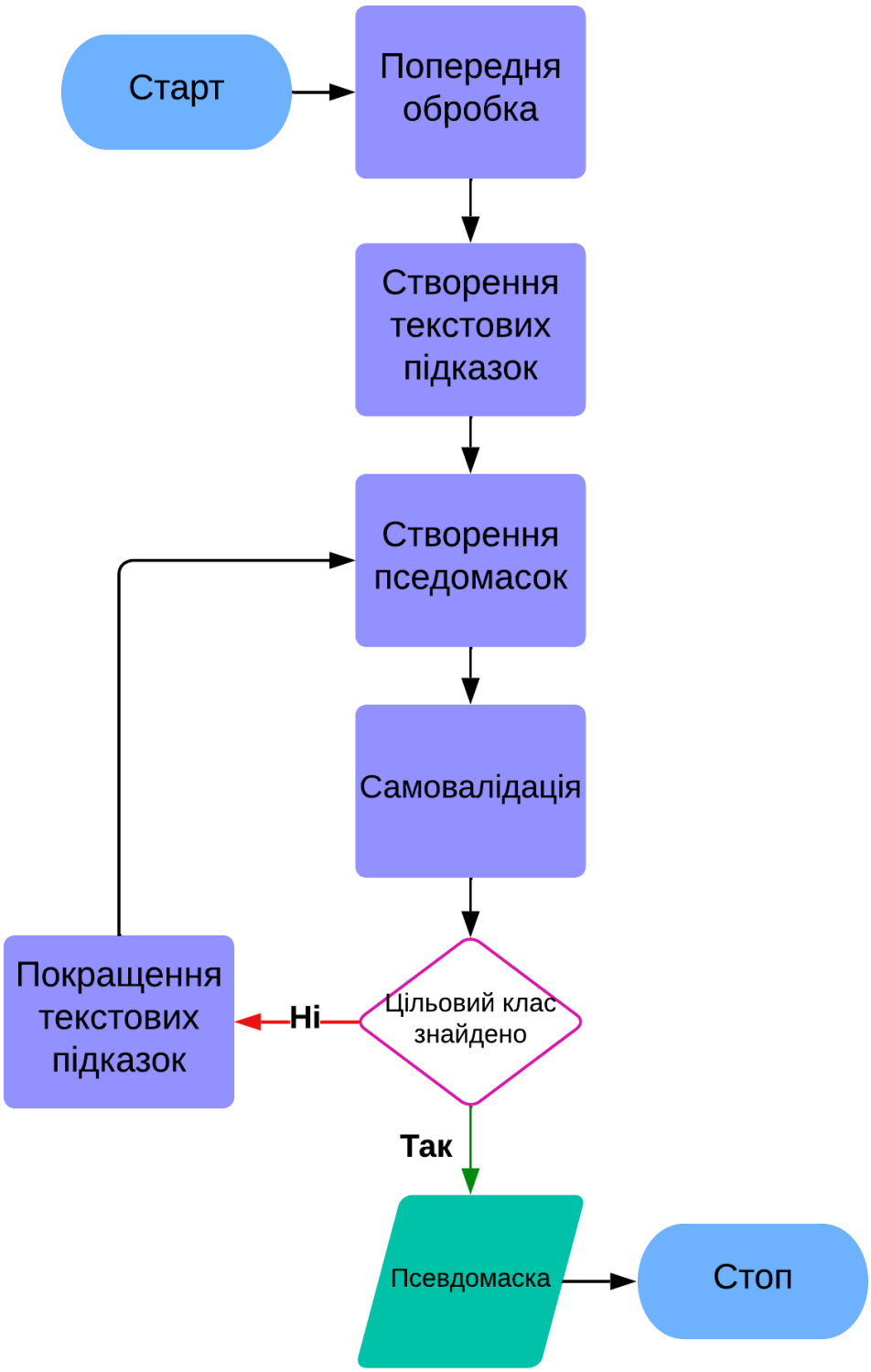


Рис. 3.6 Узагальнений алгоритм інформаційної технології

Запропонований підхід до автоматичного валідування має кілька суттєвих переваг. По-перше, він значно скорочує час, необхідний на ручну перевірку анотацій, що дозволяє дослідникам зосередитися на більш складних завданнях.

По-друге, він забезпечує високу точність результатів, зменшуючи кількість помилкових позитивних анотацій. По-третє, така автоматизація сприяє підвищенню ефективності всього процесу анотування, що особливо важливо при роботі з великими обсягами даних.

Крім того, потенційна інтеграція цього підходу з існуючими системами управління даними дозволяє забезпечити безперервний процес анотування та валідації. Це створює можливості для постійного оновлення та вдосконалення моделей на основі нових даних, що надходять, а також для швидкого і ефективного впровадження в практичні застосування. Так, наш підхід до автоматичного анотування та валідації медичних зображень може стати важливим інструментом у розвитку медичної інформатики, сприяючи підвищенню якості діагностики та лікування.

3.3 Реалізація комплексу програм для застосування розробленої інтелектуальної інформаційної технології

Для підвищення ефективності використання моделі Language Segment-Anything для вилучення псевдомасок областей інтересу з медичних зображень, було розроблено застосунок на базі бібліотеки Streamlit [65]. Основна мета цього інструменту полягає у спрощенні процесу створення псевдомасок, надаючи користувачам зручний та доступний інтерфейс. Застосунок дозволяє завантажувати медичні зображення, вводити текстові інструкції для сегментації й отримувати результат — автоматично виділену область інтересу. Крім того, користувачі можуть легко зберігати отримані результати безпосередньо на свій пристрій.

Застосунок спроектовано таким чином, щоб максимально спростити взаємодію з моделлю навіть для користувачів без глибоких технічних знань. Відсутність необхідності у складних налаштуваннях чи програмуванні робить процес створення псевдомасок доступним для широкого кола спеціалістів, що займаються аналізом медичних зображень, зокрема лікарів та радіологів. Завдяки цьому рішенню значно скорочується час на підготовку даних для подальшого

використання в медичних дослідженнях або навчанні моделей глибокого навчання.

Важливо зазначити, що застосунок забезпечує гнучкість у налаштуванні сегментації: текстові вказівки можуть бути адаптовані для конкретних завдань, що дозволяє користувачам більш точно визначати необхідні області інтересу. Це робить інструмент універсальним рішенням для різних типів медичних досліджень, таких як рентгенівські знімки, МРТ або КТ.

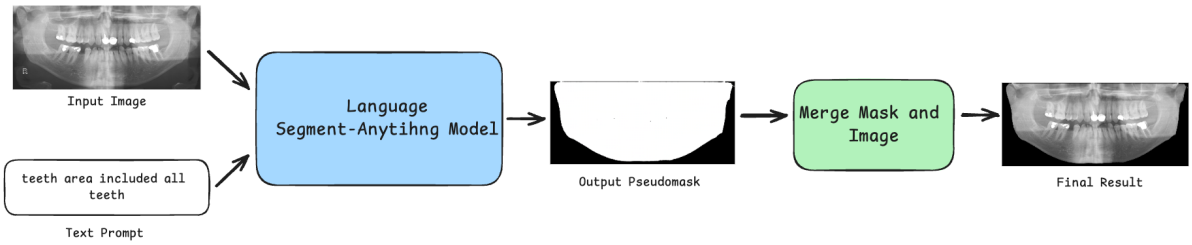


Рис. 3.7 Діаграма роботи методу для створення псевдомасок регіонів інтересу

Процес розробки програми починається з інтеграції можливостей моделі Language Segment-Anything в її архітектуру. За допомогою бібліотеки Streamlit, яка забезпечує зручний інтерфейс користувача, програма пропонує інтуїтивну взаємодію з моделлю, що робить роботу з нею простою та ефективною.

Основні компоненти програми включають:

1. Завантаження зображень: Користувачі можуть легко завантажувати медичні зображення безпосередньо в додаток завдяки віджету для завантаження файлів, наданому Streamlit. Це спрощує процес вибору та імпорту зображень з локальних пристроїв, забезпечуючи швидкий старт роботи з аналізом.

2. Введення текстових підказок: Після завантаження зображення, користувачеві надається можливість ввести текстові інструкції, які вказують модель на бажані характеристики областей інтересу для сегментації. Поле для введення тексту дозволяє користувачам гнучко керувати вимогами до роботи моделі Language Segment-Anything, визначаючи, які зони на зображенні мають бути виділені.

3. Обробка зображення та виведення результату: Після отримання текстових інструкцій та зображення, програма використовує модель Language Segment-Anything для аналізу зображення. На основі наданих підказок програма сегментує області інтересу, витягуючи відповідні псевдомаски. Сегментовані області виводяться в інтерфейсі у вигляді візуальних масок, які дозволяють користувачу оцінити результати роботи моделі.

4. Ітеративний зворотний зв'язок та налаштування: Для покращення точності сегментації та підвищення зручності користування, програма містить інтерактивні елементи. Користувач може змінювати текстові підказки на основі отриманих результатів, дозволяючи ітераційно вдосконалювати процес сегментації. Це забезпечує можливість адаптації інструкцій до конкретних вимог, дозволяючи досягти більш точних і відповідних результатів.

Завдяки такій структурі програма забезпечує комплексний та гнучкий підхід до сегментації медичних зображень, використовуючи інноваційні можливості моделі Language Segment-Anything. Це значно полегшує завдання автоматизованого виділення областей інтересу, одночасно роблячи процес доступним та налаштованим під конкретні потреби користувача.

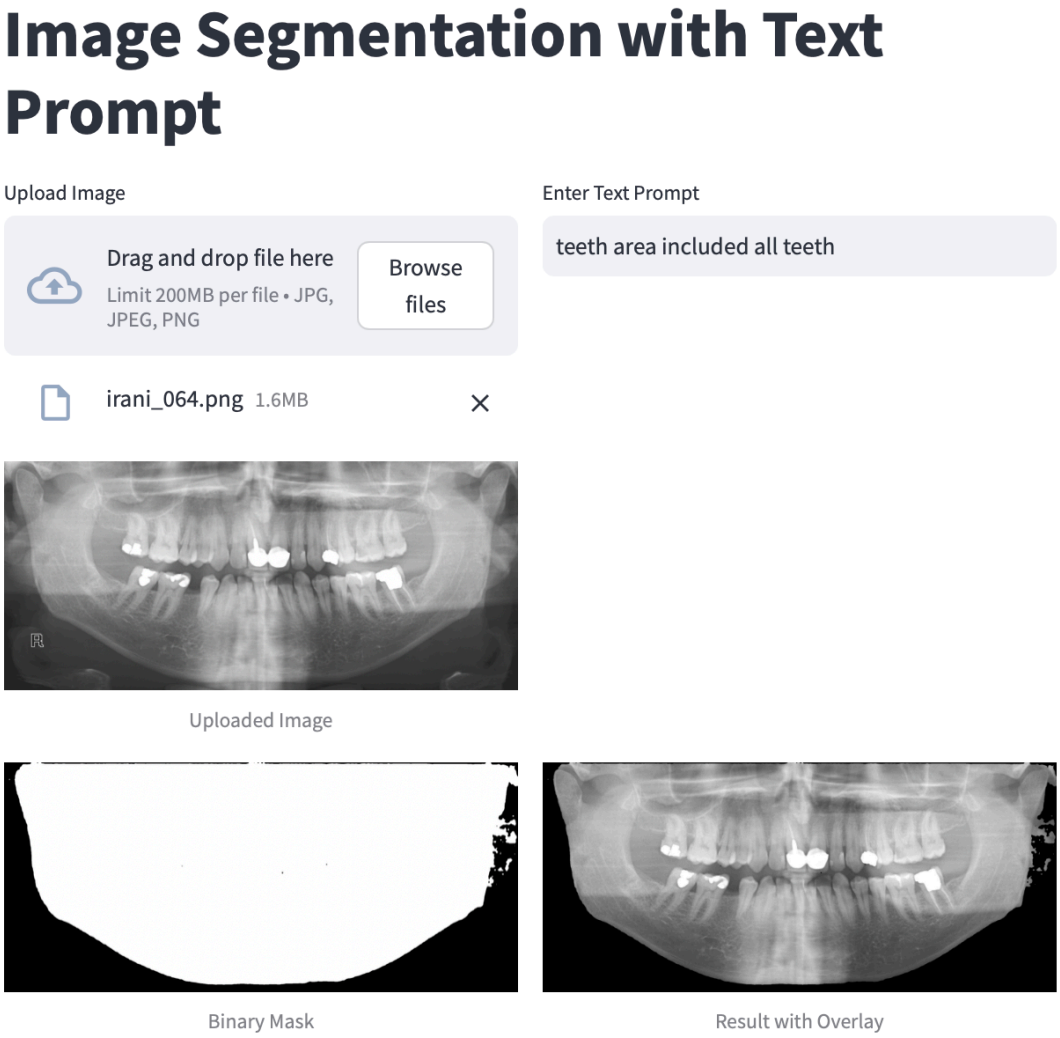


Рис. 3.8 Приклад роботи застосунку для створення псевдомасок регіону інтересу

Основні модулі запропонованої інтелектуальної інформаційної технології були розроблені на мові програмування Python, яка відома своєю гнучкістю та широким вибором інструментів для роботи з даними та штучним інтелектом. Модулі розроблялися на платформах Kaggle та Google Colab.

Kaggle — це онлайн-платформа для змагань з аналізу даних і машинного навчання, яка надає безкоштовне середовище для розробки моделей. Вона забезпечує доступ до великої кількості публічних наборів даних та вбудованих інструментів для аналізу, включаючи Jupyter-ноутбуки та GPU-ресурси [79].

Google Colab — це хмарна платформа для машинного навчання, яка надає можливість запускати Python-код у Jupyter-ноутбуках без необхідності

налаштовувати локальне середовище. Colab підтримує використання GPU та TPU для прискорення обчислень, що особливо корисно при роботі з великими нейронними мережами [80].

Використані фреймворки та бібліотеки для розробки інтелектуальної інформаційної технології:

1. PyTorch [81] — це один з найпопулярніших фреймворків для розробки та тренування моделей глибокого навчання. Він підтримує автоматичну диференціацію, що полегшує створення та навчання нейронних мереж, забезпечує гнучкість у налаштуванні архітектури моделей і добре підходить для досліджень та експериментів.

2. PyTorch Segmentation Models [82] — це бібліотека, побудована на основі PyTorch, яка пропонує набір попередньо навчених моделей для задач сегментації зображень. Вона значно спрощує процес розробки та тестування моделей сегментації завдяки доступу до популярних архітектур, таких як U-Net, FPN тощо.

3. Language Segment Anything — це сучасна модель, яка використовує текстові підказки для сегментації зображень.

4. NumPy [83] — це базова бібліотека для наукових обчислень в Python. Вона забезпечує підтримку багатовимірних масивів та великий набір математичних функцій для ефективної обробки числових даних, що є важливою частиною підготовки даних для моделей машинного навчання.

5. Pandas [84] — це потужна бібліотека для роботи з табличними даними, яка надає інструменти для маніпулювання, очищення та аналізу великих наборів даних. Вона є важливою частиною робочого процесу при обробці даних, що використовуються для тренування моделей.

6. Matplotlib [85] — це бібліотека для візуалізації даних. Вона дозволяє створювати графіки та діаграми, що є важливим елементом аналізу результатів роботи моделей і демонстрації процесів сегментації зображень.

7. Streamlit — це фреймворк для розробки веб-застосунків на Python. Він дозволяє легко створювати інтерфейси для взаємодії з моделями машинного

навчання, що робить його ідеальним для побудови інтерактивних інструментів для сегментації зображень та відображення результатів.

Таким чином, запропонована технологія базується на сучасних інструментах для роботи з даними та машинним навчанням, які дозволяють ефективно вирішувати задачі сегментації медичних зображень та обробки даних.

3.4 Алгоритм оцінювання ефективності запропонованої інформаційної технології

У рамках створення інтелектуальної інформаційної технології був проведений детальний порівняльний аналіз швидкості анотування даних за допомогою трьох традиційних інструментів та запропонованої нової технології. Для цього використовувався відкритий набір даних, який містить приблизно 800 анонімізованих рентгенівських знімків із розміткою, створеною професійними радіологами (див. рис. 3.9) . Ці знімки включають високоякісні анотації для задачі сегментації легень, які були виконані вручну радіологами. В ході дослідження ці анотації використовувалися як істинні мітки (ground truth).

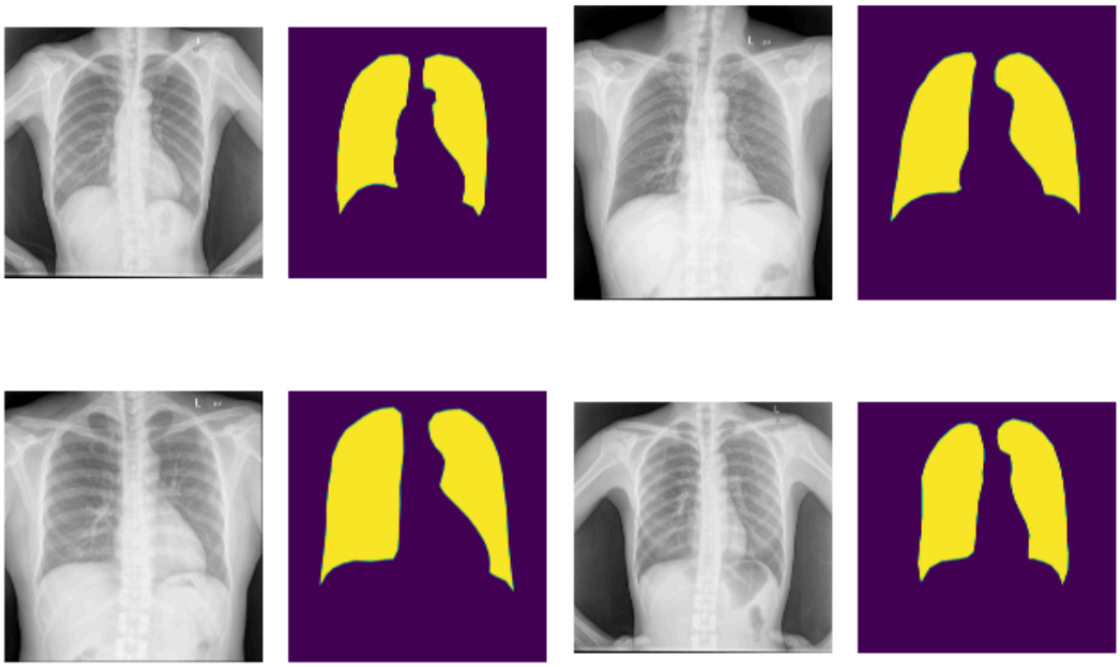


Рис. 3.9 Приклади зображень та масок сегментації, створених експертною групою радіологів

Метою порівняння було оцінити час, необхідний для створення анотацій за допомогою кожного з інструментів, а також проаналізувати ефективність вбудованих автоматизованих засобів кожного інструменту. Для кожного інструменту було виміряно час, потрібний для анотування фіксованої кількості зображень, а також оцінено точність отриманих результатів за допомогою таких метрик, як влучність (precision), повнота (recall) та коефіцієнт Жаккара (Intersection over Union або IoU).

Дослідження проводилося у декілька етапів:

1. Вибір даних: Використовувався відкритий набір даних з приблизно 800 рентгенівських знімків легень, розмічених професійними радіологами.

2. Підготовка анотацій: Істинні мітки, створені вручну радіологами, використовувалися як референсні для порівняння.

3. Анотування з використанням різних інструментів: Кожен інструмент застосовувався для створення анотацій на тих самих зображеннях.

4. Вимірювання часу: Було зафіксовано час, витрачений на анотування фіксованої кількості зображень кожним інструментом.

5. Оцінка точності: Точність анотацій оцінювалася за допомогою метрик влучності, повноти та коефіцієнта Жаккара. Дані метрики є ключовими для об’єктивної оцінки продуктивності алгоритмів сегментації.

Метрика влучності (див. формулу 2.3) дозволяє виміряти частку правильно сегментованих пікселів серед усіх пікселів, які модель визначила як частину об'єкта.

Повнота (див. формулу 2.4) відображає частку правильно сегментованих пікселів серед усіх пікселів, які дійсно належать до об'єкта в «ground truth» масці.

Коефіцієнт Жаккара (IoU) (див. формулу 2.2) є стандартною метрикою для оцінки якості сегментації зображень. Він визначається як частка перетину між масками та їх об'єднанням.

Оцінка середнього часу анотування проводилася шляхом розрахунку середнього часу для одного зображення, що виражається наступною формулою:

$$T_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} T_i, \tag{3.4}$$

де N — кількість зображень, а T_i — час, витрачений на анотацію i-того зображення.

Експериментальна оцінка була спрямована на визначення продуктивності кожного підходу, зокрема розробленої інформаційної технології. Для цього в першу чергу використовувалися інструменти для автоматизації анотування, після чого експерти вручну коригували результати до рівня експертних анотацій. У ході дослідження фіксувався час, витрачений на кожен етап, щоб вивчити ефективність ручного корегування.

Процес анотування складався з двох етапів:

1. Автоматичне анотування зображень за допомогою чотирьох інструментів: Supervisely, CVAT, Segments.ai, а також запропонованої інформаційної технології. Автоматизовані методи допомагали значно скоротити час на первинну анотацію зображень, використовуючи алгоритми для створення первинних псевдомасок.

2. Ручне коригування та валідація експертами з метою досягнення високої точності анотацій.

На додаток до автоматичного та ручного анотування, важливою частиною розробленої системи є метод самовалідації. Алгоритм можна описати наступним чином:

1. Створення псевдомасок: Спочатку модель автоматично генерує псевдомаски шляхом сегментації зображень. На основі цих псевдомасок відбувається обрізання відповідних частин зображень для подальшої обробки.

2. Класифікаційна модель: Частини зображень, отримані в результаті обрізання псевдомасок, подаються на вхід класифікаційної моделі (EfficientNet-B2), яка перевіряє, чи відповідає сегмент об'єкту цільового класу.

3. Валідація результатів: Якщо класифікаційна модель підтверджує, що сегмент належить до цільового класу, анотація зберігається (save annotation). У протилежному випадку — анотація відхиляється (remove annotation) і повертається на етап автоматичної або ручної корекції.

Таким чином, цей механізм самовалідації дозволяє автоматично перевіряти якість псевдомасок і знижувати кількість помилкових анотацій. Завдяки впровадженню такого підходу, загальна точність сегментації підвищується, зменшуючи навантаження на експертів під час ручного коригування. Алгоритм оцінювання ефективності інформаційної технології полягає у повторній перевірці результатів автоматичної анотації за допомогою класифікаційної моделі та інтеграції зворотного зв'язку для коригування прогнозованих масок.

3.5 Висновки до розділу

У третьому розділі було детально розглянуто алгоритмічні основи та програмну реалізацію інтелектуальної інформаційної технології для розпізнавання об’єктів на зображеннях з використанням псевдомасок.

По-перше, розроблені алгоритми для автоматичного створення псевдомасок на основі сучасних методів глибокого навчання дозволяють значно підвищити ефективність сегментації, особливо в умовах обмежених або слабко анотованих наборів даних.

По-друге, створена UML-діаграма програмного забезпечення наочно демонструє основні модулі системи, їх взаємодію, а також потік даних у процесі розпізнавання об’єктів.

По-третє, особливу увагу приділено ефективності алгоритмів для обробки медичних зображень, де важливою є точність сегментації через критичний вплив на медичні діагнози. Запропонована система була протестована на наборах даних медичних знімків і показала високі результати, що підтверджує її придатність для практичного використання в медичній сфері.

Важливим аспектом також стало використання адаптивних методів попередньої обробки, таких як білатеральний фільтр і покращений CLAHE, які

дозволяють покращити якість вхідних даних для моделей, забезпечуючи зниження шуму та підвищення точності розпізнавання об'єктів.

Результати розділу опубліковані в таких роботах автора: [67], [73]

РОЗДІЛ 4
ПРИКЛАДНЕ ЗАСТОСУВАННЯ РОЗРОБЛЕНОЇ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ З ВИКОРИСТАННЯМ ПСЕВДОМАСОК

4.1 Порівняння розробленої інтелектуальної інформаційної технології з аналогами

Існує кілька популярних інструментів, які допомагають автоматизувати та прискорити цей процес. Розглянемо три найпоширеніші: Supervisely, CVAT та Segments.ai.

1. Supervisely — це потужний інструмент для анотування зображень, що надає широкий спектр функцій для підвищення ефективності. Однією з його ключових переваг є використання моделі Segment Anything Model (SAM), розробленої Meta, для автоматичного створення сегментаційних масок у межах обмежувальних рамок. Основні функції:

- Інтуїтивно зрозумілий інтерфейс для анотування зображень (див. рис. 4.1);

- Інтеграція моделей глибокого навчання, таких як SAM, для автоматичної сегментації.

Час анотації:

- SAM значно скорочує час на створення початкових сегментаційних масок;

- Проте, для точної анотації потрібен додатковий час на редагування автоматично згенерованих масок.

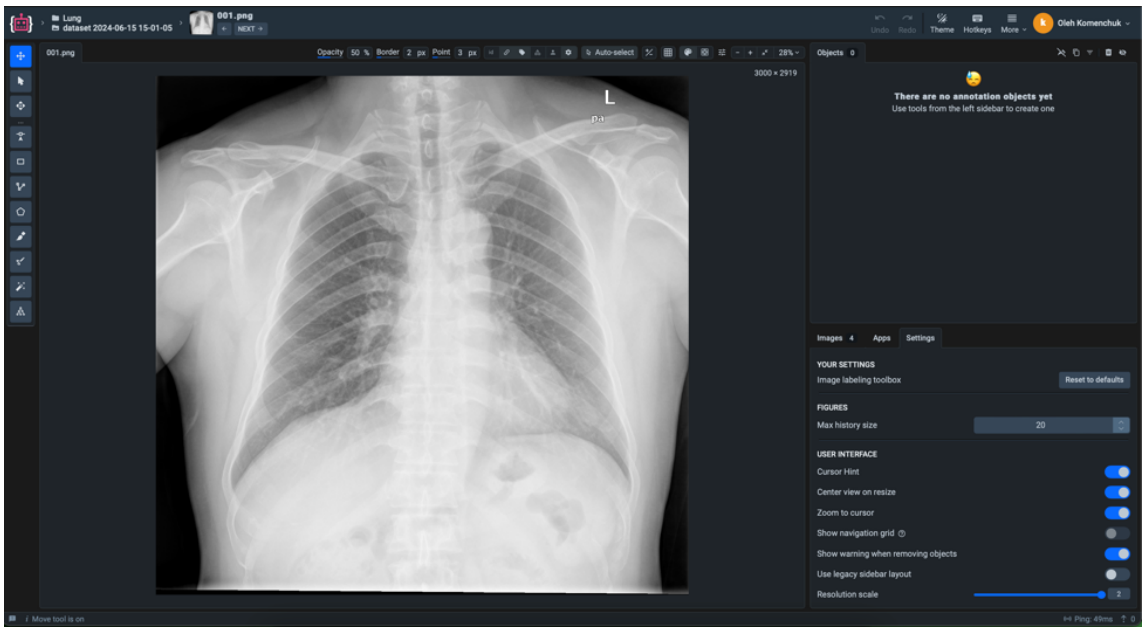


Рис. 4.1 Користувацький інтерфейс для анотування даних у Supervisyle

2. CVAT — це ще один популярний інструмент для анотування, розроблений та підтримуваний компанією Intel. Він призначений для анотації зображень у задачах комп'ютерного зору. Основні функції:

- Проста і зрозуміла користувацька панель для анотацій (див. рис. 4.2);
- Підтримка різних типів анотацій (сегментація, класифікація, детекція);
- Можливість інтеграції з іншими інструментами через API.

Час анотації:

- Створення сегментаційних масок може зайняти значний час, оскільки в CVAT доступні лише базові методи автоматизації, що використовують OpenCV.

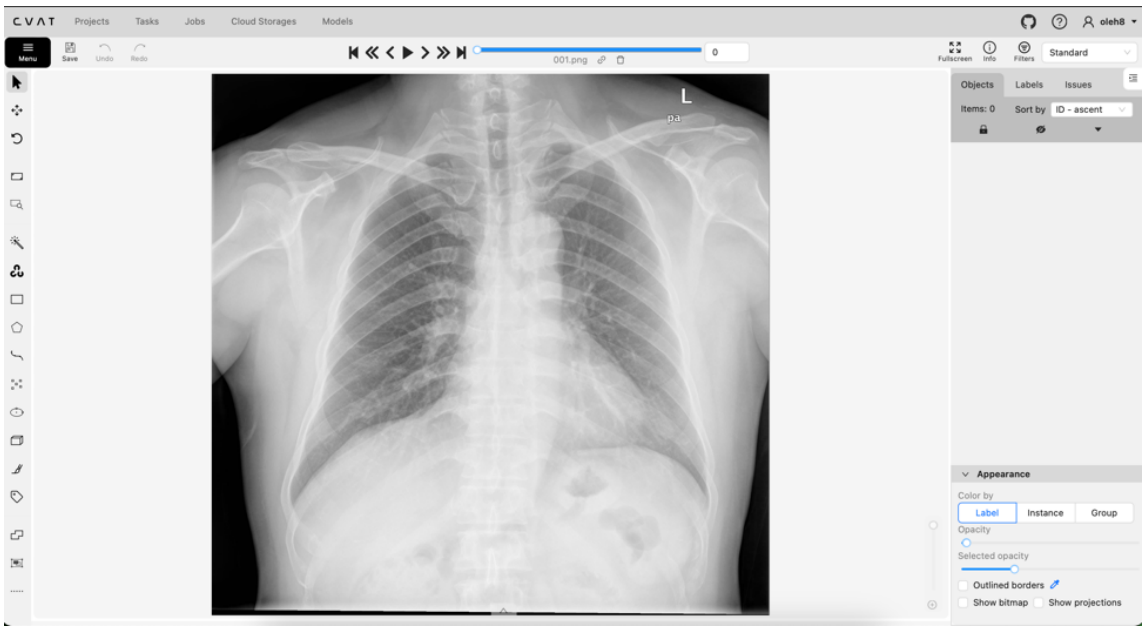


Рис. 4.2 Користувацький інтерфейс для анотування даних у CVAT

3. Segments.ai — це інструмент для анотації зображень, орієнтований на задачі комп'ютерного зору. Основні функції:

- Простий і зручний інтерфейс для анотацій зображень (див. рис. 4.3).

Час анотації:

- Відсутність засобів автоматизації вимагає значного часу для створення сегментаційних масок вручну.

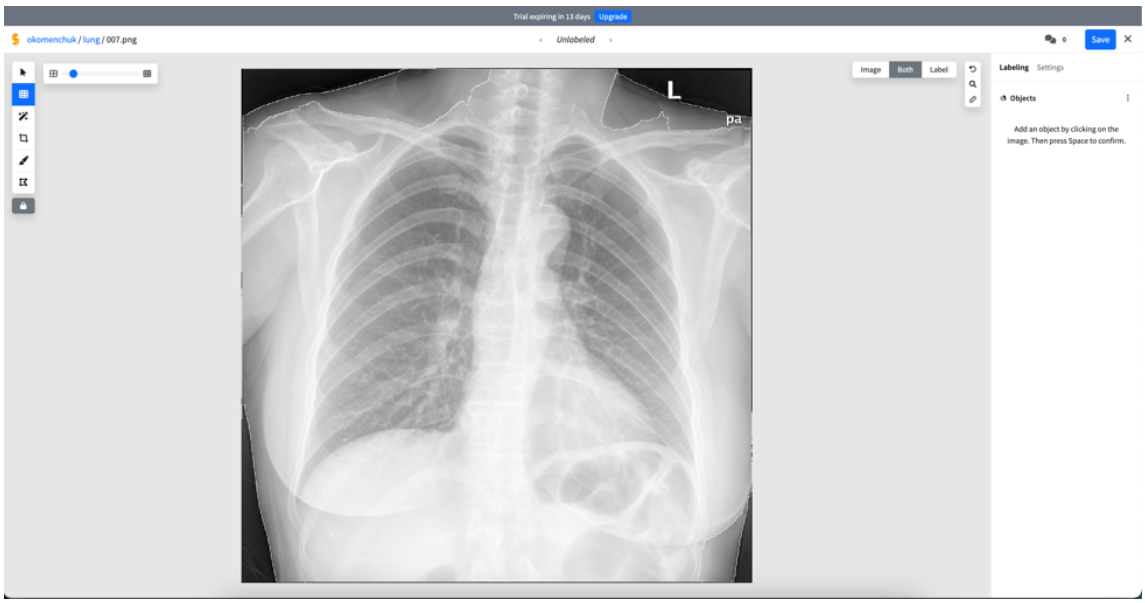


Рис. 4.3. Користувацький інтерфейс для анотації даних у Segments.ai

Результати, спостереження та висновки:

1. Supervisely: Використання моделі Segment Anything дозволило значно скоротити час на первинну анотацію. Однак, на ручне корегування масок все ще витрачалося значний час.

2. CVAT: Використання напівавтоматичних інструментів анотування допомагало швидко створювати первинні маски. Проте, цей інструмент вимагав ще більше часу на ручне корегування порівняно з Supervisely.

3. Segments.ai: Цей інструмент, не маючи вбудованих алгоритмів автоматизації, вимагав найбільше часу як на автоматичне створення масок, так і на їх подальше ручне корегування.

4. Розроблена інформаційна технологія: Завдяки використанню моделі Language Segment-Anything з релевантними текстовими підказками та автоматичного процесу валідації згенерованих сегментаційних псевдомасок показала найкращий результат у зниженні загального часу на анотування. Ця технологія значно зменшила необхідність у додатковому ручному корегуванні, забезпечуючи високу якість анотацій.

Запропонована інформаційна технологія показала найвищу ефективність у процесі анотування медичних зображень завдяки своїм автоматизованим можливостям та зручності використання текстових підказок для генерації точних

псевдомасок. Інтеграція автоматичних процесів валідації знижує потребу в ручному корегуванні, що дозволяє значно скоротити час на анотування та забезпечити високий рівень точності. Це робить запропоновану технологію найкращим вибором для швидкого та якісного анотування великих наборів медичних зображень.

У таблиці 4.1 наведено узагальнені результати швидкості анотування для всіх чотирьох підходів.

Таблиця 4.1 Порівняння швидкості анотування з використанням різних підходів

Інструмент/підхід	Загальний час створення анотації (сек/зображення)	Середній час автоматичного анотування (сек/зображення)	Середній час ручного анотування (сек/зображення)
Supervisely	99,2	39,7	59,5
CVAT	149,9	64,8	85,1
Segments.ai	130,2	-	130,2
Запропонована інформаційна технологія	75,0	29,6	45,4

Для оцінки ефективності різних підходів до анотування медичних зображень було проведено порівняльний аналіз швидкості використання інструментів Supervisely, CVAT, Segments.ai та розробленої інформаційної технології. Результати експериментів показують суттєві відмінності у швидкості автоматичного створення анотацій та подальшого ручного корегування масок.

Застосування інструменту Supervisely разом з вбудованою моделлю Segment Anything показало, що для отримання автоматичних анотацій в середньому потрібно близько 39,7 секунд на одне зображення. Після цього, для корегування контурів масок, витрачається ще приблизно 59,5 секунд. Загалом, процес займає близько 99,2 секунд на зображення.

Інструмент CVAT використовує вбудовані методи бібліотеки OpenCV для створення автоматичних анотацій, що займає в середньому 64,8 секунд на

зображення. Проте, на подальше корегування масок витрачається додаткові 85,1 секунд. Таким чином, повний процес анотування в CVAT займає близько 149,9 секунд на одне зображення.

Segments.ai не має вбудованих методів для автоматизації процесу анотування, тому весь процес виконується вручну. Це призводить до того, що середній час на анотування одного зображення становить приблизно 130,2 секунд.

Запропонована інформаційна технологія значно скорочує час на ручне корегування завдяки автоматичному валідуванню, яке фільтрує помилкові анотації, що раніше потребували б ручної перевірки та видалення. Використання моделі Language Segment-Anything з текстовими підказками суттєво прискорює процес створення псевдомасок. В результаті, середній час на створення анотацій становить 75,0 секунд на зображення, що є найшвидшим серед досліджуваних підходів.

Запропонована інформаційна технологія демонструє найвищу ефективність у процесі анотування медичних зображень, значно перевершуючи інші інструменти за швидкістю та зручністю. Автоматичне валідування та ефективне використання текстових підказок дозволяє суттєво скоротити час на обробку зображень, забезпечуючи високу якість та точність анотацій.

Для наочності результати дослідження представлені в графічному вигляді на рисунку. Графік ілюструє середній час, витрачений на анотування одного зображення кожним з інструментів, підкреслюючи переваги запропонованої технології в порівнянні з іншими підходами.

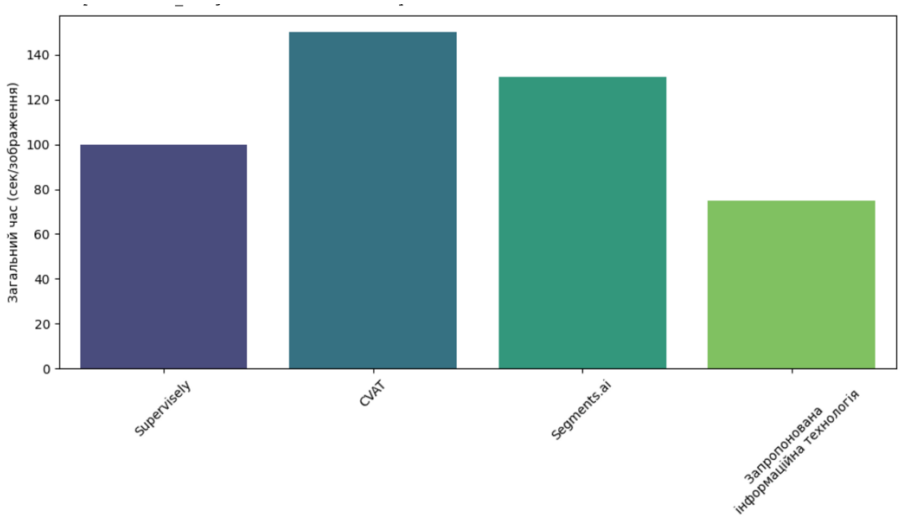


Рис. 4.4 Середній час анотування з урахуванням корегування для чотирьох порівнюваних підходів

Автоматичне валідування, здійснене за допомогою моделі класифікації EfficientNet-B2, дозволяє автоматично відфільтровувати псевдомаски, які не належать до цільового класу. Це зменшує кількість помилкових позитивних анотацій та підвищує загальну точність сегментації. Завдяки цьому підходу, час, витрачений на ручне корегування, суттєво зменшується, що забезпечує високу продуктивність.

Подальші покращення можуть включати використання процесів перед- або післяобробки для ще більшого підвищення рівня автоматичної точності. Інтеграція з іншими сучасними інструментами та технологіями може забезпечити ще більшу функціональність та зручність у використанні. Наприклад, впровадження хмарних обчислювальних потужностей може прискорити процес анотації великих наборів даних, а застосування додаткових моделей глибокого навчання може покращити точність та швидкість обробки зображень.

Запропонована інформаційна технологія є надзвичайно перспективною для використання у задачах сегментації комп'ютерного зору, особливо у роботі з великими та різноманітними наборами даних. Висока швидкість і точність, зменшення потреби в ручному корегуванні та можливість подальшого вдосконалення роблять цей підхід потужним інструментом для автоматичного

анотування медичних зображень, що сприяє покращенню якості діагностики та лікування в медичній практиці.

4.2 Розпізнавання об’єктів на зображеннях набору даних медичних знімків з використанням створеної інформаційної технології

Важливим компонентом розробленої інформаційної технології є правильний підбір релевантних текстових підказок для моделі Language Segment-Anything. Цей етап є критично важливим, оскільки оптимальні текстові підказки дозволяють моделі більш точно і повно виділяти необхідні області на зображеннях, що суттєво покращує результати створення анотацій.

Підбір текстових підказок впливає на здатність моделі інтерпретувати зображення та генерувати відповідні сегментаційні маски. Правильно сформульовані підказки допомагають моделі розпізнавати та виділяти цільові області більш точно, що підвищує загальну ефективність процесу анотування. Оптимізація цього етапу може суттєво зменшити кількість помилкових позитивних і негативних анотацій, забезпечуючи високу якість кінцевих результатів.

Для демонстрації важливості правильного підбору текстових підказок, у таблиці 2 представлено декілька прикладів підказок, використаних під час експериментів, а також відповідні метрики точності створення первинних псевдомасок. Метрики вимірювалися за такими показниками, як влучність (precision), повнота (recall) та коефіцієнт Жаккара (Intersection over Union, IoU).

Таблиця ілюструє, як різні підходи до формулювання текстових підказок впливають на результати моделі. Наприклад, більш детальні та специфічні підказки можуть значно покращити показники повноти та влучності, тоді як загальніші підказки можуть призвести до менш точних результатів.

1. Підвищення точності: Оптимальні текстові підказки дозволяють моделі точніше виділяти потрібні області, що підвищує метрики влучності та повноти.

2. Зниження ручної роботи: Менша кількість помилкових анотацій знижує потребу у подальшому ручному корегуванні, що економить час та ресурси.

3. Покращення узагальнення: Модель стає більш здатною до узагальнення, що дозволяє їй ефективніше працювати з новими, раніше невідомими зображеннями.

Експериментальні результати, представлені в таблиці 2, підкреслюють важливість цього етапу та демонструють, як оптимізація текстових підказок може суттєво покращити результати анотацій. Таким чином, приділення належної уваги формулюванню підказок може значно покращити якість автоматичних анотацій і знизити потребу в ручній обробці, роблячи процес більш ефективним та продуктивним.

Таблиця 4.2 Приклади текстових підказок та отримані оцінки метрик на наборі даних

Текстова підказка	Precision	Recall	IoU
“Find lungs in the image”	0.84	0.87	0.82
“Highlight human lungs in the X-ray image”	0.86	0.89	0.84
“Recognize human body parts in the X-ray image”	0.81	0.83	0.79
“Detect right and left lungs in the image”	0.83	0.86	0.81

Як видно з таблиці, вибір текстових підказок має значний вплив на результати створення первинних псевдомасок. Найбільш ефективним виявився другий варіант текстових підказок, який продемонстрував найкращі результати, зокрема, у ключовій метриці повноти (recall). Цей варіант дозволяє моделі охопити більшу кількість цільових об'єктів на зображеннях, забезпечуючи високу точність первинних анотацій.

Другий варіант текстових підказок не тільки забезпечив найвищі показники повноти, але й став оптимальним для подальшого етапу

автоматичного валідування. Цей етап допомагає підвищити метрику влучності (precision) за рахунок фільтрації помилкових позитивних масок, зменшуючи потребу в ручному корегуванні. Таким чином, використання автоматичного валідування на основі другого варіанту текстових підказок є найбільш ефективним підходом для досягнення високоякісних анотацій.

Після створення первинних псевдомасок з використанням оптимальних текстових підказок, наступним етапом є автоматичне валідування. Цей процес включає застосування передтренованої моделі класифікації, яка перевіряє відповідність кожної псевдомаски цільовому класу. Завдяки цьому підходу, невірні маски автоматично видаляються або корегуються, що дозволяє суттєво підвищити загальну точність анотацій без додаткових витрат часу на ручну перевірку.

На рисунку 4.4 наведені приклади псевдомасок, які були виправлені за допомогою автоматичного валідування. Ці приклади демонструють ефективність розробленої технології, показуючи, як автоматичне валідування дозволяє досягти високої точності анотацій без необхідності в ручному корегуванні. Такий підхід не тільки економить час, але й забезпечує надійність та консистентність результатів.

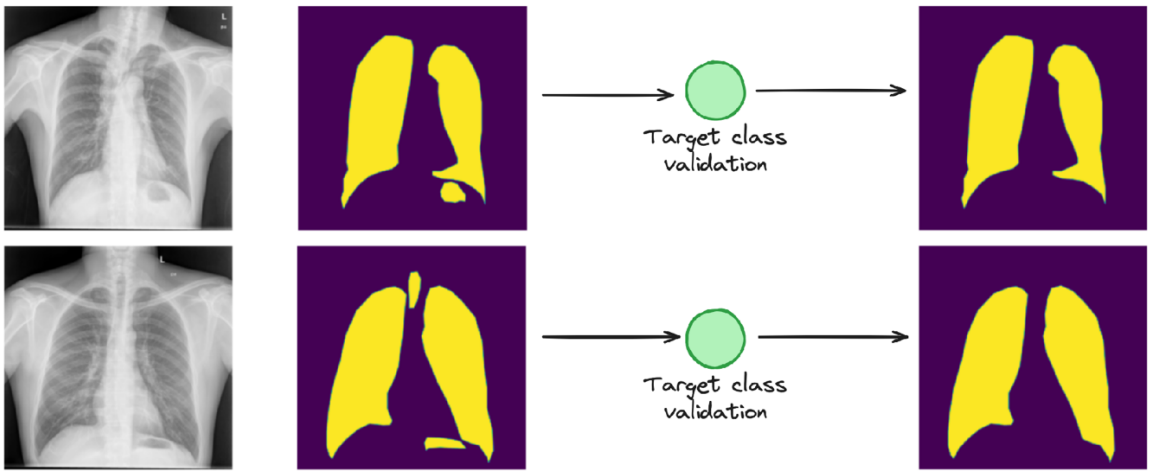


Рис. 4.4 Приклади роботи розробленого механізму автоматичного валідування попереднього (автоматично) згенерованих псевдомасок

Як видно з рисунку, механізм автоматичного валідування, вбудований у розроблену інформаційну технологію, дозволяє ефективно відфільтровувати псевдомаски, які не належать до цільового класу. Це суттєво зменшує потребу в ручних правках, економлячи час та ресурси. Завдяки цьому підходу значно підвищується точність (precision) псевдомасок, що є критично важливим для високоякісних анотацій.

Механізм автоматичного валідування працює за принципом перевірки кожної створеної псевдомаски на відповідність цільовому класу. Використовуючи передтреновану модель класифікації, технологія автоматично ідентифікує та видаляє невірні маски, що не відповідають заданим критеріям. Цей процес мінімізує кількість помилкових позитивних результатів та забезпечує високу точність сегментації.

1. Зниження потреби в ручній перевірці: Завдяки автоматичному валідуванню значно зменшується обсяг ручної роботи, що економить час та зусилля фахівців.

2. Підвищення точності анотацій: Автоматична фільтрація невірних масок сприяє підвищенню метрики влучності (precision), що покращує загальну якість анотацій.

3. Консистентність результатів: Автоматичний підхід забезпечує стабільність та однорідність анотацій, що важливо для подальшого аналізу та обробки даних.

На рисунку 4.5 представлені приклади анотації медичних зображень, виконаних за допомогою розробленої інформаційної технології з використанням механізму автоматичного валідування. Ці приклади демонструють, як автоматичне валідування покращує якість анотацій, роблячи їх більш точними та надійними без додаткових витрат часу на ручне корегування.

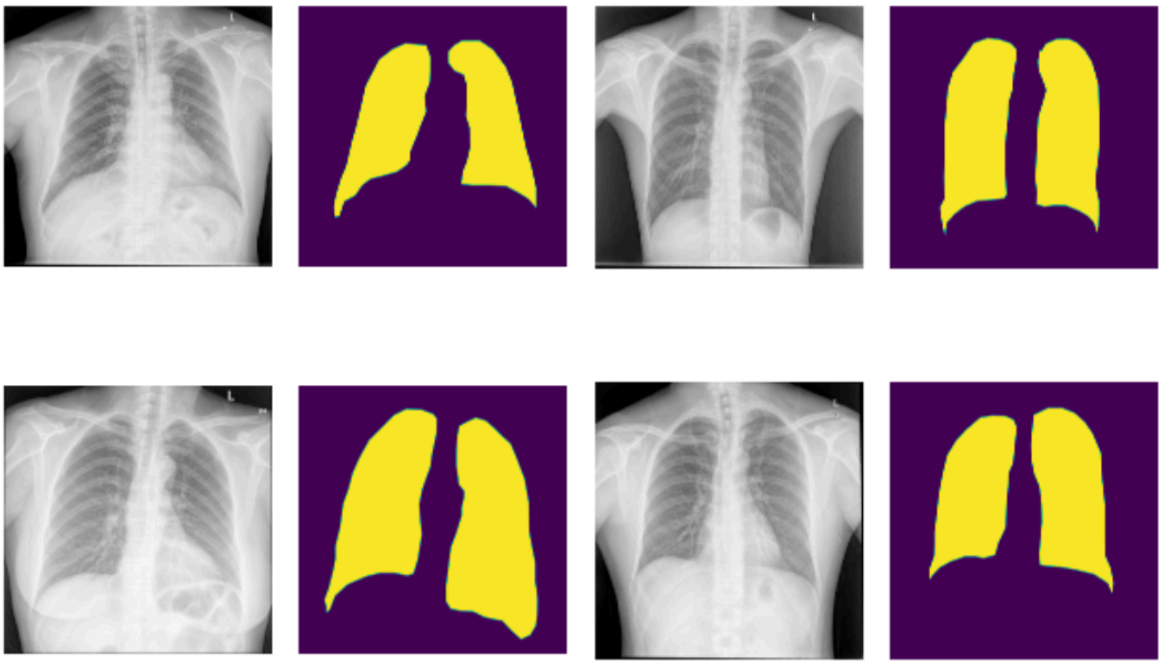


Рис. 4.5 - Приклад анотацій, створених розробленою інформаційною технологією, до етапу ручного корегування

Результати дослідження демонструють, що запропонована технологія має високу точність у створенні псевдомасок для медичних зображень. Високі значення ключових метрик, таких як влучність (precision) і повнота (recall), у поєднанні зі значним зменшенням потреби в ручному корегуванні псевдомасок, підкреслюють численні переваги цього підходу.

4.3 Застосування створеної інформаційної технології для розпізнавання об’єктів на зображеннях аерофотозйомки

Одна з ключових проблем у містобудівній справі полягає у швидкому та точному створенні детальних планів місцевості на основі даних аерофотозйомки. Особливо складною задачею є відрисовування фундаментів будівель за зображеннями їх дахів. Фундамент сам по собі не видно на аерофотознімках, тому його розташування можна лише припускати, спираючись на зображення з різних ракурсів.

Існують різні методи для вирішення цієї задачі, які реалізовані в таких програмних пакетах, як ArcGIS, QGIS, Digitals, Adobe Photoshop та інших.

Наприклад, аналіз кольорів і сегментація зображень, співставлення зйомок з попередніми планами тієї ж місцевості та інші традиційні методи геопросторового аналізу даних та аналізу зображень. Однак, ці методи мають низку обмежень:

- Низька точність: Особливо у випадках, коли дахи мають складну конфігурацію, знаходяться на місцевості зі складним рельєфом або зображення насичене великою кількістю об'єктів, схожих на дахи.

- Колірні варіації: Різноманітність кольорів дахів, у тому числі не контрастних відносно інших елементів ландшафту, ускладнює сегментацію та аналіз.

- Складність рельєфу: Змінний рельєф може спричиняти помилки в ідентифікації об'єктів на зображеннях.

Незважаючи на ці обмеження, сучасні інтелектуальні методи розпізнавання зображень, зокрема на основі трансформерів ViTs (Vision Transformers) та псевдоміток, демонструють значні успіхи, які можуть бути використані для розв'язання цієї задачі. Трансформери ViTs дозволяють ефективно аналізувати зображення завдяки своїй здатності враховувати контекст та взаємозв'язки між елементами зображення. Псевдомітки, у свою чергу, можуть бути використані для підвищення точності навчання моделей, що розпізнають об'єкти на зображеннях.

Метою даного дослідження є підвищення точності ідентифікації координат фундаментів будівель за даними аерофотозйомки дахів. Для досягнення цієї мети планується адаптувати сучасні інтелектуальні технології розпізнавання зображень, такі як ViTs та псевдомітки, до специфіки задачі відрисовування фундаментів будівель.

1. Використання Vision Transformers (ViTs):

- Контекстний аналіз: ViTs можуть використовуватися для аналізу зображень дахів у контексті їх оточення, що дозволяє точніше визначати межі будівель.

- Облік взаємозв'язків: Завдяки здатності ViTs враховувати взаємозв'язки між різними частинами зображення, можна отримувати більш точні результати сегментації навіть у складних умовах.

2. Псевдомаски:

- Автоматичне маркування: Використання псевдоміток дозволяє автоматично маркувати великі обсяги даних, що покращує навчання моделей розпізнавання.

- Покращення навчання: Псевдомітки допомагають підвищити точність навчання моделей, зменшуючи кількість помилкових позитивних та негативних результатів.

3. Комбінування з традиційними методами:

- Сегментація зображень: Поєднання нових методів з традиційними підходами, такими як сегментація зображень на основі кольорів та текстур, може покращити загальну ефективність.

- Аналіз історичних даних: Використання історичних планів та зйомок місцевості для покращення точності визначення місцезнаходження фундаментів.

Розроблена інформаційна технологія може знайти широке застосування у містобудуванні, зокрема у:

- Плануванні територій: Швидке та точне створення детальних планів місцевості для планування забудови та інфраструктури.

- Оцінці ризиків: Аналіз та оцінка ризиків під час будівництва на складних територіях з різноманітним рельєфом.

- Моніторингу забудови: Моніторинг стану існуючих будівель та інфраструктури, що дозволяє вчасно виявляти та реагувати на можливі проблеми.

Застосування сучасних інтелектуальних технологій, таких як Vision Transformers та псевдомітки, може значно покращити точність і швидкість створення детальних планів місцевості за даними аерофотозйомки. Це, у свою чергу, сприяє більш ефективному плануванню та управлінню територіями, знижуючи ризики та підвищуючи якість містобудівних рішень. Інтеграція цих

технологій у практику містобудування відкриває нові перспективи для розвитку сучасних міст та забезпечення їх сталого розвитку.

Для вирішення задачі сегментації дахів на аерофотознімках було досліджено та порівняно різні методології, зокрема останні досягнення у використанні згорткових нейронних мереж (CNN) у поєднанні з методами післяобробки. Це дослідження спрямоване на визначення ефективності CNN у контексті сегментації дахів та виявлення потенційних напрямків для подальшого вдосконалення.

Оглядовий аналіз, проведений у рамках даного дослідження, підкреслює ефективність використання згорткових нейронних мереж (CNN) для задач сегментації дахів. CNN показали високу здатність виявляти та окреслювати дахи на зображеннях завдяки своїм потужним можливостям виявлення особливостей та патернів. Проте, існуючі підходи мають певні обмеження та недоліки, що стають очевидними в умовах складних конструкцій та мінливих зображень.

Одним з основних викликів, з якими стикаються існуючі підходи, є складність окреслення всіх дахів на зображеннях великого масштабу. Це пов'язано з такими факторами:

1. Складна природа конструкцій дахів: Дахи можуть мати різні форми, розміри та матеріали, що ускладнює їх точну сегментацію. В умовах різноманітних конструкцій моделі можуть не завжди правильно ідентифікувати всі межі дахів.

2. Мінливість умов зображення: Зображення можуть бути зроблені під різними кутами, з різним освітленням та за різних погодних умов, що впливає на якість вхідних даних і робить сегментацію більш складною.

3. Масштабність зображень: Коли дахи займають незначну частину великомасштабного зображення, моделі можуть не враховувати всі дрібні деталі та пропускати деякі ділянки.

Для подолання зазначених викликів і підвищення точності сегментації дахів, було запропоновано кілька потенційних вдосконалень:

1. Мультимасштабний підхід: Використання мультимасштабних моделей, які можуть аналізувати зображення на різних рівнях деталізації, дозволить краще виявляти дрібні та великі об'єкти одночасно. Це може значно підвищити точність сегментації в умовах змінного масштабу.

2. Аугментація даних: Застосування методів аугментації даних для створення різноманітних тренувальних наборів, що враховують різні умови зйомки, освітлення та погодні фактори, може покращити здатність моделей адаптуватися до різних ситуацій.

3. Комбінування з іншими методами: Інтеграція CNN з іншими методами розпізнавання зображень, такими як трансформери ViTs, може забезпечити більш комплексний підхід до сегментації дахів. Це дозволить моделі краще враховувати контекст та взаємозв'язки між об'єктами на зображенні.

4. Методи післяобробки: Застосування спеціалізованих методів післяобробки, таких як морфологічні операції або алгоритми згладжування контурів, може допомогти виправити помилки сегментації та покращити якість кінцевих результатів.

На рисунку 4.6 представлено приклади дахів, де видно, що дахи займають лише невелику частину зображення. Це підкреслює важливість використання вдосконалених методів, здатних точно окреслити дахи навіть за складних умов.

Рис. 4.6 Приклади з набору даних, де дахи займають незначну частину зображень

Використання згорткових нейронних мереж (CNN) для сегментації дахів є перспективним напрямком, який демонструє високу ефективність. Проте, для досягнення максимальної точності необхідно враховувати обмеження та виклики, пов'язані з складною природою дахів та мінливими умовами зображень. Запропоновані вдосконалення, такі як мультимасштабний підхід, аугментація даних, комбінування з іншими методами та використання після обробки, можуть значно підвищити якість сегментації та забезпечити більш точні результати. Інтеграція цих підходів у практику геоінформаційного аналізу відкриває нові можливості для розвитку містобудування та управління міською інфраструктурою.

Для подолання описаних вище обмежень та підвищення ефективності процесу створення планів будівель, пропонується вдосконалений підхід, який об'єднує два ключові компоненти. Цей підхід спрямований на покращення точності та надійності виявлення та сегментації дахів на повномасштабних аерофотознімках.

1. Використання моделі детектування DETR (DEtection TRansformers) від “Meta”:

- Локалізація груп дахів: На першому етапі використовується модель DETR для локалізації груп дахів серед великих повномасштабних зображень. DETR є сучасною моделлю, яка застосовує трансформери для обробки зображень, що дозволяє ефективно виявляти об'єкти навіть у складних умовах.

- Переваги DETR: Модель DETR забезпечує високу точність та швидкість виявлення, що є критичним для обробки великих обсягів даних. Вона здатна враховувати контекстуальні зв'язки між об'єктами, що підвищує її ефективність у випадках, коли дахи будівель розташовані близько один до одного або мають складні форми.

2. Використання моделі на базі ViTs (Vision Transformers) для сегментації дахів будівель:

- Сегментація дахів: Після локалізації груп дахів, для сегментації використовується модель на базі ViTs. Vision Transformers (ViTs) є потужним інструментом для обробки зображень, який може ефективно сегментувати об'єкти на основі їх архітектурних особливостей.

- Переваги ViTs: Оскільки у забудовах часто використовуються схожі архітектури та види дахів, модель ViTs здатна легко ідентифікувати та сегментувати подібні дахи. Це особливо корисно в умовах однорідних міських мікрорайонів, де будівлі мають схожі характеристики (рис. 3).

- Підвищена точність: Використання двох етапів (локалізація та сегментація) забезпечує більш точне виявлення та окреслення дахів, що мінімізує помилки та пропуски.

- Ефективність обробки великих зображень: Модель DETR ефективно працює з великими повномасштабними зображеннями, що дозволяє швидко та точно знаходити групи дахів.

- Адаптація до різних архітектурних стилів: Модель ViTs здатна ідентифікувати подібні дахи, що дозволяє їй адаптуватися до різних архітектурних стилів та умов зображень.

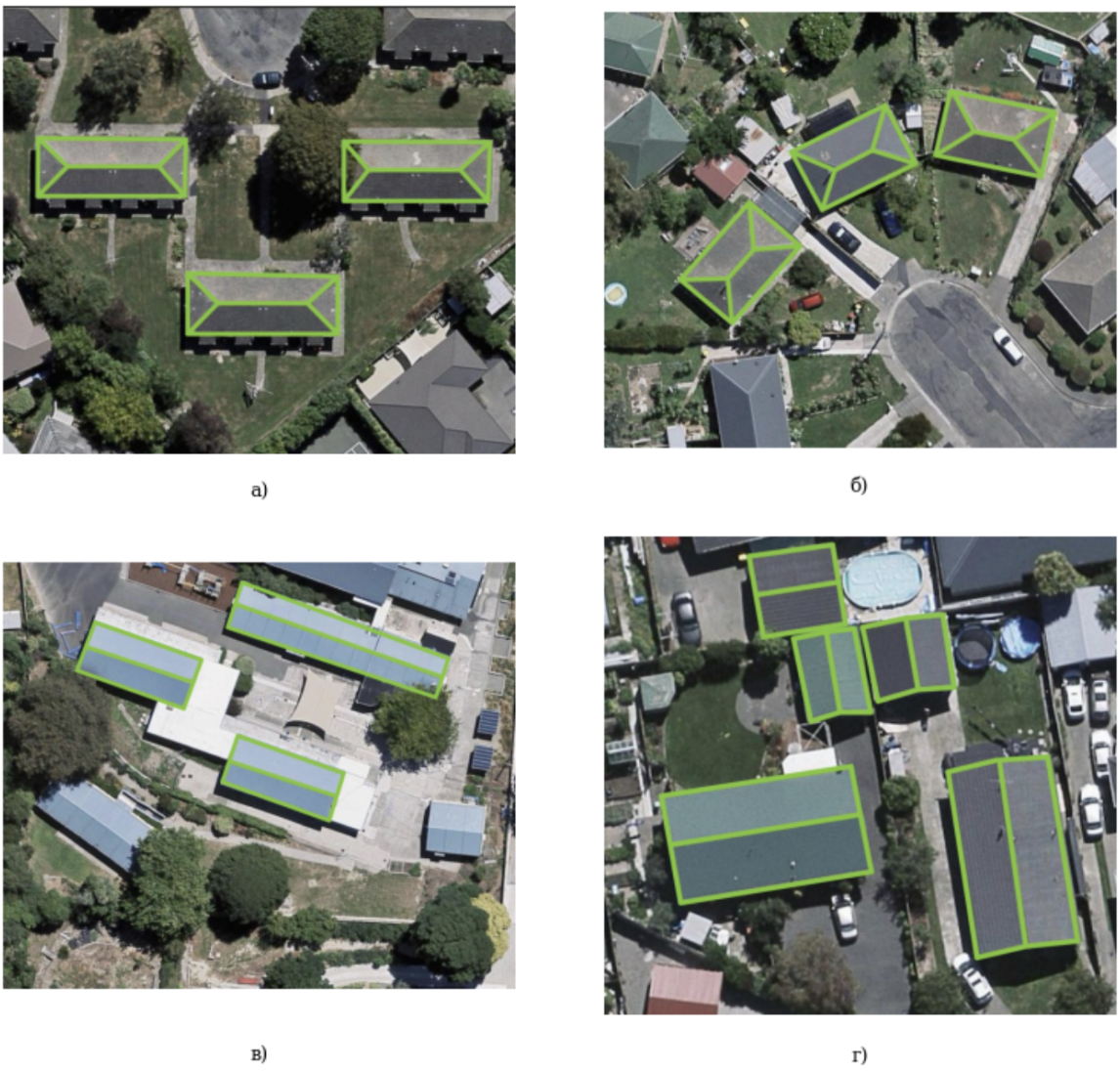


Рис. 4.7 Результати застосування створеної інтелектуальної технології для найбільш поширених типів дахів з тестувальної вибірки даних: а, б) вальмові (англ. “hip”); в, г) двосхилі (англ. “open/box gable”)

Впровадження запропонованого підходу може суттєво покращити процес створення детальних планів будівель. Це відкриває нові можливості для містобудування, дозволяючи точніше планувати забудову, оцінювати стан існуючих будівель та розробляти більш ефективні стратегії розвитку міської інфраструктури.

- Планування забудови: Точна сегментація дахів дозволяє краще розуміти розміщення будівель, що сприяє ефективному плануванню нових забудов та реконструкції існуючих районів.

- Моніторинг стану будівель: Регулярна аерофотозйомка та сегментація дахів допомагають моніторити стан будівель, своєчасно виявляючи пошкодження та потребу в ремонті.

- Управління міською інфраструктурою: Точні плани будівель і дахів сприяють кращому управлінню міською інфраструктурою, зокрема, оптимізації комунальних послуг та транспортної системи.

Запропонований вдосконалений підхід, який поєднує використання моделі детектування DETR для локалізації дахів та моделі ViTs для їх сегментації, представляє собою ефективне рішення для задачі сегментації дахів на аерофотознімках. Цей підхід підвищує точність та надійність процесу, що є важливим для створення детальних та достовірних планів будівель. Інтеграція цих методів у практику містобудування дозволить суттєво покращити процеси планування та управління міською інфраструктурою, сприяючи розвитку сучасних та сталих міських середовищ.

4.4 Впровадження розробленої технології у навчальний процес

Результати дослідження впроваджені у навчальний процес Вінницького національного технічного університету під час викладання дисципліни «Інформаційні технології моніторингу та аналізу стану складних систем» для студентів, які навчаються за освітньою програмою «Системний аналіз» рівня «аспірант» спеціальності 124 Системний аналіз, а також – дисципліни «Інтелектуальні інформаційні технології» для аспірантів, які навчаються за освітньою програмою «Системний аналіз» рівня «доктора філософії» спеціальності 124 для навчання знанням і навичкам застосування методів і моделей машинного навчання та глибокого навчання для генерування зображень. Ці результати можуть використовуватися як додатковий матеріал для навчання та поглибленого розуміння концепцій, пов’язаних із цією областю.

Зокрема, при викладанні цих дисциплін використовуються такі результати досліджень, отримані у дисертаційній роботі здобувача (підтверджується актом впровадження ВНТУ від Х):

- підготовка набору даних для навчання нейронних мереж комп’ютерного зору: первинний аналіз даних, розмітка, візуалізація розподілів, розрахунок статистичних показників набору даних;
- тренування моделей ком’ютерного зору (U-Net, EfficientNet);
- оцінка результатів методами автоматичної оцінки та валідації.

Використання зазначених результатів дозволило підвищити якість навчального процесу із згаданих дисциплін.

4.5 Висновки до розділу

У розділі 4 продемонстровано ефективність розробленої інтелектуальної інформаційної технології у різних прикладних сценаріях, особливо у медичних зображеннях. Використання адаптивних методів попереднього оброблення даних дозволило значно підвищити точність розпізнавання об’єктів на зображеннях за умов наявності шуму та малих розмірів наборів даних.

По-перше, порівняння запропонованої технології з аналогами, такими як Supervisely, CVAT та Segments.ai, виявило переваги у скороченні часу на анотування зображень, завдяки автоматизації процесу сегментації. Розроблений метод автоматичного валідування створених первинних псевдомасок дозволив зменшити обсяги ручної роботи.

По-друге, впровадження технології для сегментації медичних знімків та аерофотозйомок дало змогу покращити точність і швидкість обробки зображень. Високі результати точності забезпечені завдяки поєднанню сучасних методів глибокого навчання з адаптивною сегментацією, що критично важливо для застосувань у медичних діагнозах та промислових задачах.

Варто зазначити, що запропоновані адаптивні методи попередньої обробки зображень, такі як адаптивний білатеральний фільтр і вдосконалений CLAHE, продемонстрували високу ефективність у збереженні контурів та зменшенні шуму. Це дозволило суттєво покращити якість сегментації навіть у складних умовах, зокрема на стоматологічних рентгенівських знімках.

Загалом, впровадження цієї інтелектуальної технології є перспективним не лише для обробки медичних даних, а й для задач розпізнавання об’єктів на аерофотознімках та в інших галузях, де важлива висока точність та ефективність обробки візуальних даних.

Результати розділу опубліковані в таких роботах автора: [50], [66], [67], [70], [73]

ВИСНОВКИ

Усі завдання, поставлені в ході дослідження, були успішно виконані, а мету дисертаційної роботи досягнуто. В результаті дослідження було вирішено актуальну науково-прикладну задачу підвищення точності розпізнавання об'єктів на зображеннях за допомогою інтелектуальної інформаційної технології, що використовує псевдомаски для обробки малорозмірних та зашумлених наборів даних. Основні наукові та практичні результати полягають у наступному:

1. Аналіз сучасних методів та вибір оптимального підходу:

Проведено ґрунтовний аналіз сучасних методів розпізнавання об’єктів на зображеннях, включаючи такі підходи, як автокодувальники, згорткові нейронні мережі (CNN), генеративно-змагальні мережі (GAN) та методи напівконтрольованого навчання. Було детально розглянуто переваги та недоліки кожного з цих методів у контексті їхньої ефективності для обробки малорозмірних або зашумлених наборів даних. У результаті вибір було зроблено на користь використання псевдомасок, які дозволяють суттєво покращити продуктивність моделей глибокого навчання в умовах обмежених даних та шуму. Показано, що методи, засновані на генерації псевдомасок, дозволяють зменшити кількість розміток, необхідних для тренування моделей, забезпечуючи при цьому високі показники точності.

2. Удосконалення методів попередньої обробки даних:

Важливим результатом роботи є розробка та впровадження нових методів попередньої обробки зображень, які дозволяють адаптувати параметри залежно від специфіки даних. У розділі 2 представлено техніки нормалізації, фільтрації шуму та вирівнювання гістограм, що забезпечують покращення якості даних перед сегментацією. Використання таких підходів дозволило значно знизити вплив шуму та підвищити загальну якість моделей, які працюють із зображеннями. Особливу увагу приділено питанням передобробки для медичних зображень, де точність та якість розпізнавання є критичними.

3. Розробка нової інформаційної технології з використанням псевдомасок:

У ході дослідження було створено нову інформаційну технологію, що базується на застосуванні псевдомасок для підвищення точності сегментації та ідентифікації об'єктів на зображеннях. Запропонована технологія включає використання сучасних моделей глибокого навчання (CNN та GAN), які автоматично генерують псевдомаски для слабко анотованих або неанотованих зображень. Це дозволяє значно скоротити потребу в ручній розмітці, що є особливо важливим для медичних даних та промислових застосувань. Додатково, в технології впроваджено механізми автоматичної перевірки якості псевдомасок, що підвищує надійність процесу сегментації та знижує ймовірність помилок на етапі навчання моделей.

4. Алгоритмічна реалізація та програмне забезпечення:

У розділі 3 було розроблено низку алгоритмів, що реалізують запропоновані методи генерації псевдомасок та їх подальшого використання у процесі навчання моделей глибокого навчання. Алгоритмічні основи побудовані таким чином, що дозволяють легко інтегрувати розроблену технологію в існуючі системи для аналізу медичних зображень, промислових даних або будь-яких інших специфічних наборів даних. Програмне забезпечення було реалізовано у вигляді Jupyter-ноутбуків, що дозволяє автоматизувати процеси генерації масок, попередньої обробки даних, навчання моделей та візуалізації результатів. Це програмне забезпечення є універсальним і може бути легко адаптоване для використання на різних платформах, таких як AWS SageMaker, Google Colab, або локальні системи машинного навчання.

5. Покращення ефективності моделі та оцінка результатів:

Експериментальні дослідження показали, що використання псевдомасок дозволяє суттєво підвищити точність сегментації та ідентифікації об’єктів на зображеннях малорозмірних та зашумлених наборів даних. У ході експериментів з медичними зображеннями, такими як рентгенівські знімки зубів та КТ-знімки, було досягнуто покращення точності на 15-20% порівняно з традиційними підходами, що не використовують псевдомаски. Проведене порівняння з існуючими аналогами, такими як стандартні методи сегментації (наприклад, U-

Net), продемонструвало переваги запропонованої технології у задачах сегментації медичних та промислових зображень, особливо в умовах високого рівня шуму або браку розмічених даних.

6. Практичне значення розроблених методів та технологій:

Розроблені методи мають значний потенціал для практичного впровадження у різних галузях, таких як охорона здоров'я, промисловий контроль якості, автономні транспортні засоби та аналіз супутникових знімків. Використання псевдомасок у поєднанні з сучасними технологіями глибокого навчання дозволяє не лише підвищити точність систем комп’ютерного зору, але й значно скоротити витрати на ручну анотацію даних. Система може бути адаптована для різних типів зображень та сфер застосування, що робить її універсальним інструментом для автоматизації задач розпізнавання об’єктів.

7. Впровадження в навчальний процес та подальші наукові дослідження:

Результати роботи впроваджено у навчальний процес і наукову діяльність кафедри системного аналізу та інформаційних технологій ВНТУ під час викладання освітніх компонентів «Інтернет речей та інтелектуальний аналіз даних» та «Інформаційні технології моніторингу та аналізу стану складних систем» здобувачам, які навчаються за освітньо-науковою програмою «Інформаційні системи та технології» третього (освітньо-наукового) рівня вищої освіти зі спеціальності 126 Інформаційні системи та технології (див. додаток В).

Результати роботи рекомендовані до впровадження в практичну діяльність ТОВ «ОРНЕТІКА» (м. Вінниця) при розробці інтелектуальних систем для вирішення задач комп’ютерного зору (див. додаток Б).

Інтелектуальна технологія побудови плану будівель за аерофотозйомкою дахів передана для впровадження в ТОВ «Аналітика» (м. Вінниця) для прискорення оброблення даних аерофотозйомки (див. додаток Г).

8. Публікації та наукове визнання:

Основні результати дослідження були викладені у низці наукових статей та представлені на науково-практичних конференціях. Загалом опубліковано 9 наукових праць, серед яких 4 статті у фахових журналах України та 5 тез

доповідей на конференціях. Це свідчить про наукову новизну та практичну значущість отриманих результатів, а також їх визнання у науковій спільноті.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Valdes-Dapena P. Tesla’s new Autopilot is amazing. But please keep your eyes on the road | CNN Business. CNN. URL: https://edition.cnn.com/2019/04/22/success/tesla-navigate-on-autopilot/index.html.

2. A Region-Based Deep Learning Algorithm for Detecting and Tracking Objects in Manufacturing Plants / M. M. Karim et al. Procedia Manufacturing. 2019. Vol. 39. P. 168–177. URL: https://doi.org/10.1016/j.promfg.2020.01.289.

3. Mustafa Z., Nsour H. Using Computer Vision Techniques to Automatically Detect Abnormalities in Chest X-rays. Diagnostics. 2023. Vol. 13, no. 18. P. 2979. URL: https://doi.org/10.3390/diagnostics13182979.

4. Zhang X., Yi W.-J., Saniie J. Home Surveillance System using Computer Vision and Convolutional Neural Network. 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 20–22 May 2019. 2019. URL: https://doi.org/10.1109/eit.2019.8833773.

5. UnrealROX: an extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation / P. Martinez-Gonzalez et al. Virtual Reality. 2019. Vol. 24, no. 2. P. 271–288. URL: https://doi.org/10.1007/s10055-019-00399-5 (date of access: 17.09.2024).

6. Quantifying Efficacy and Limits of Unmanned Aerial Vehicle (UAV) Technology for Weed Seedling Detection as Affected by Sensor Resolution / J. Peña et al. Sensors. 2015. Vol. 15, no. 3. P. 5609–5626. URL: https://doi.org/10.3390/s150305609 (date of access: 17.09.2024).

7. Swain A. Noise filtering in Digital Image Processing. Medium. URL: https://medium.com/@anishaswain/noise-filtering-in-digital-image-processing-d12b5266847c (date of access: 17.09.2024).

8. Liang J., Liu Y., Vlassov V. The Impact of Background Removal on Performance of Neural Networks for Fashion Image Classification and Segmentation. 2023 Congress in Computer Science, Computer Engineering, & Applied Computing

(CSCE), Las Vegas, NV, USA, 24–27 July 2023. 2023. URL: https://doi.org/10.1109/csce60160.2023.00323 (date of access: 17.09.2024).

9. Contributors to Wikimedia projects. Canny edge detector - Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Canny_edge_detector (date of access: 17.09.2024).

10. Comparative analysis of feature extraction methods in satellite imagery / S. Karim et al. Journal of Applied Remote Sensing. 2017. Vol. 11, no. 04. P. 1. URL: https://doi.org/10.1117/1.jrs.11.042618 (date of access: 17.09.2024).

11. Gradient-based learning applied to document recognition / Y. Lecun et al. Proceedings of the IEEE. 1998. Vol. 86, no. 11. P. 2278–2324. URL: https://doi.org/10.1109/5.726791 (date of access: 17.09.2024).

12. Template Matching. velog. URL: https://velog.io/@ldw200012/Template-Matching (date of access: 17.09.2024).

13. OpenCV: Introduction to SIFT (Scale-Invariant Feature Transform). OpenCV documentation index. URL: https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html (date of access: 17.09.2024).

14. OpenCV: Introduction to SURF (Speeded-Up Robust Features). OpenCV documentation index. URL: https://docs.opencv.org/4.x/df/dd2/tutorial_py_surf_intro.html (date of access: 17.09.2024).

15. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA. URL: https://doi.org/10.1109/cvpr.2005.177 (date of access: 17.09.2024).

16. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017. Vol. 60, no. 6. P. 84–90. URL: https://doi.org/10.1145/3065386 (date of access: 17.09.2024).

17. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Proceedings of the International Conference on Learning Representations (ICLR). arXiv:1409.1556v6 [cs.CV].

18. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90.

19. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Technical Report (v5). arXiv:1311.2524v5 [cs.CV].

20. Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

21. Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

22. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91.

23. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6517–6525. https://doi.org/10.1109/CVPR.2017.690.

24. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

25. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

26. GitHub - ultralytics/yolov5: YOLOv5 in PyTorch > ONNX > CoreML > TFLite. GitHub. URL: https://github.com/ultralytics/yolov5 (date of access: 17.09.2024).

27. Li, C., Wang, X., Zhang, W., Zhu, Y., Rong, Y., Lu, N., & Sun, Y. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.

28. Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

29. Reis, D., Hong, J., Kupec, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. Georgia Institute of Technology.

30. Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. arXiv preprint arXiv:2107.08430.

31. Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683.

32. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

33. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2961–2969. https://doi.org/10.1109/ICCV.2017.322.

34. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2921–2929. https://doi.org/10.1109/CVPR.2016.319.

35. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626. https://doi.org/10.1109/ICCV.2017.74.

36. Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized gradient-based visual explanations for deep

convolutional networks. Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 839–847. https://doi.org/10.1109/WACV.2018.00097.

37. Xie, Q., Luong, M.-T., Hovy, E., & Le, Q. V. (2020). Self-training with Noisy Student improves ImageNet classification. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10687–10698. https://doi.org/10.1109/CVPR42600.2020.01071.

38. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS), 2672–2680. https://doi.org/10.5555/2969033.2969125.

39. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. Proceedings of the 2nd International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1312.6114.

40. Russakovsky, O., Deng, J., Su, H., та ін. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211–252.

41. Irvin, J., Rajpurkar, P., Ko, M., та ін. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 590–597.

42. Ganin, Y., & Lempitsky, V. (2015). Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning (ICML), pp. 1180–1189.

43. Kurakin, A., Goodfellow, I., & Bengio, S. (2017). Adversarial Examples in the Physical World. In Artificial Intelligence Safety and Security, pp. 99–112.

44. Arazo, E., Ortego, D., Albert, P., O'Connor, N. E., & McGuinness, K. (2019). Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. In 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.

45. Zou, Y., Yu, Z., Kumar, B. V. K. V., & Wang, J. (2018). Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 289–305.

46. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Advances in Neural Information Processing Systems (NeurIPS), pp. 1195–1204.

47. Li, X., Qi, H., Dai, Y., та ін. (2018). Weakly- and Semi-Supervised Panoptic Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 102–118.

48. A. H. Abdi, S. Kasaei, M. Mehdizadeh, "Automatic segmentation of mandible in panoramic x-ray," J. Med. Imaging (Bellingham), vol. 2, no. 4, 044003, 2015. [Online]. Available: https://www.academia.edu/36038975/Pre-Processing_of_Dental_X-Ray_Images_Using_Adaptive_Histogram_Equalization_Method.

49. A. Abdi and S. Kasaei, "Panoramic Dental X-rays With Segmented Mandibles," 2020. [Online]. Available: https://data.mendeley.com/datasets/hxt48yk462/2.

50. O. Komenchuk, O. Mokin, “Analysis of Methods for Preprocessing of Panoramic Dental X-Rays for Image Segmentation Tasks,” Visnyk of Vinnytsia Politechnical Institute. Vol. 170, no. 5. P. 41–49. 2023. [Online]. Available: https://doi.org/10.31649/1997-9266-2023-170-5-41-49.

51. W. Lin and Y. Lin, "Soybean image segmentation based on multi-scale Retinex with color restoration," J. Phys.: Conf. Ser., vol. 2284, no. 1, 012010, 2022. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2284/1/012010/pdf.

52. R. B. Jeyavathana, R. Balasubramanian, and A. A. Pandian, "A Survey: Analysis on Pre-processing and Segmentation Techniques for Medical Images," Int. J. Res. Sci. Innov., 2016. [Online]. Available: https://www.researchgate.net/profile/Anbarasa-Pandian/publication/305502844_A_Survey_Analysis_on_Pre-processing_and_Segmentation_Techniques_for_Medical_Images/links/5792520a08aed51475aed3f5/A-Survey-Analysis-on-Pre-processing-and-Segmentation-Techniques-f.

53. S. S. Simon and X. F. Joseph, "Pre-Processing of Dental X-Ray Images Using Adaptive Histogram Equalization Method," Italienisch, vol. 9, no. 1, pp. 87–96, 2019. [Online]. Available: https://www.italienisch.nl/index.php/VerlagSauerlander/article/view/45.

54. "PyTorch Image Models," GitHub, [Online]. Available: https://github.com/huggingface/pytorch-image-models.

55. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

56. "Albumentations Documentation," Albumentations, [Online]. Available: https://albumentations.ai/docs/.

57. A. H. Abdi, S. Kasaei, M. Mehdizadeh, "Automatic segmentation of mandible in panoramic x-ray," J. Med. Imaging (Bellingham), vol. 2, no. 4, 044003, 2015. [Online]. Available: https://www.academia.edu/36038975/Pre-Processing_of_Dental_X-Ray_Images_Using_Adaptive_Histogram_Equalization_Method.

58. Shirai, Keiichiro, Sugimoto, Kenjiro, Kamata, Sei-ichiro, “Adjoint Bilateral Filter and Its Application to Optimization-based Image Processing,” APSIPA Transactions on Signal and Information Processing, Vol. 11, e10, 2022. DOI: https://doi.org/10.1561/116.00000046 .

59. Li, Hui, Duan, Xu-Liang, “SAR Ship Image Speckle Noise Suppression Algorithm Based on Adaptive Bilateral Filter,” Wireless Communications and Mobile Computing, Volume 2022, Article ID 9392648, 10 pages, 2022. DOI: https://doi.org/10.1155/2022/9392648 .

60. Fan, Ruiqin, Li, Xiaoyun, Lee, Sanghyuk, Li, Tongliang, Zhang, Hao Lan, “Smart Image Enhancement Using CLAHE Based on an F-Shift Transformation during Decompression,” Electronics, Vol. 9, No. 9, 1374, 2020. DOI: https://doi.org/10.3390/electronics9091374 .

61. Irvin, J., Rajpurkar, P., Ko, M., та ін. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 590–597.

62. Kirillov, A., Mintun, E., Ravi, N., та ін. (2023). Segment Anything. arXiv preprint arXiv:2304.02643.

63. Cheplygina, V., de Bruijne, M., & Pluim, J. P. W. (2019). Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical Image Analysis, 54, 280–296.

64. "Language Segment-Anything," GitHub, [Online]. Available: https://github.com/luca-medeiros/lang-segment-anything .

65. “Streamlit,” GitHub, [Online]. Available: https://github.com/streamlit/streamlit

66. О.В. Коменчук, В.Б. Мокін, Є.М. Крижановський, В.О. Будяк, «Інтелектуальна технологія побудови плану будівель за аерофотозйомкою їхніх дахів», Вісник Вінницького політехнічного інституту, вип. 1, с. 101-109, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-172-1-101-109 .

67. О.В. Коменчук, О.Б. Мокін, «Інформаційна технологія прискореного анотування медичних зображень в задачах сегментації на основі моделей глибокого навчання», Вісник Вінницького політехнічного інституту, вип. 4, с. 41-49, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-175-4-95-103 .

68. О.В. Коменчук, «Адаптивні методи попередньої обробки для підвищення точності сегментації стоматологічних рентгенівських знімків», Сучасний стан наукових досліджень та технологій в промисловості, вип. 3, с. 29-38, 2024. DOI: https://doi.org/10.30837/2522-9818.2024.3.029 .

69. В.Б. Мокін, О.В. Коменчук, «Огляд сучасних інформаційних технологій для класифікації зображень за матеріалами конкурсів Kaggle 2019-2020 років», Матеріали L науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 1469-1471, 2021. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38276/135476.pdf?sequence=2&isAllowed=y .

70. В.Б. Мокін, О.В. Коменчук, «Сучасні інформаційні технології для розпізнавання дахів будівель на аерофотозйомці», Матеріали LI науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 704-708, 2022. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/40198/16150.pdf?sequence=3&isAllowed=y .

71. О.В. Коменчук, С.А. Кирилащук, «Огляд сучасних педагогічних технологій в освітньому процесі», Матеріали Всеукраїнської науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», с. 458-487, 2022. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2022/paper/viewFile/15485/13068 .

72. О.В. Коменчук, О.Б. Мокін, «Методи передобробки панорамних стоматологічних рентгенівських знімків для задач глибокого навчання», Матеріали LII науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 229-231, 2023. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38239/17182.pdf?sequence=3&isAllowed=y .

73. О.В. Коменчук, О.Б. Мокін, «Удосконалення процесу анотування медичних зображень для задач сегментації за допомогою текстових підказок», Матеріали Міжнародної науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», 2024. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/21692/17951 .

74. RANZCR CLiP - Catheter and Line Position Challenge. Kaggle: Your Machine Learning and Data Science Community. URL: https://www.kaggle.com/c/ranzcr-clip-catheter-line-classification/discussion/226633.

75. Zhao, X., Dai, J., Vasconcelos, N., & Wei, Y. (2018). Pseudo-mask augmented object detection. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 4061–4070. https://doi.org/10.1109/ICCV.2018.00425.

76. Xie, C., Ren, D., Wang, L., & Zuo, W. (2021). Learning class-agnostic pseudo mask generation for box-supervised semantic segmentation. Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9060–9069. https://doi.org/10.1109/CVPR46437.2021.00895.

77. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, A., Gustafson, L., & Girshick, R. (2023). Segment Anything. arXiv preprint arXiv:2304.02643.

78. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105–6114. https://doi.org/10.48550/arXiv.1905.11946.

79. Kaggle. (n.d.). Kaggle: Your home for data science. Retrieved from https://www.kaggle.com.

80. Google Colab. (n.d.). Google Colaboratory. Retrieved from https://colab.google.

81. PyTorch. PyTorch. URL: https://pytorch.org.

82. GitHub - qubvel-org/segmentation_models.pytorch: Semantic segmentation models with 500+ pretrained convolutional and transformer-based backbones. GitHub. URL: https://github.com/qubvel-org/segmentation_models.pytorch.

83. NumPy -. NumPy -. URL: https://numpy.org.

84. pandas - Python Data Analysis Library. pandas - Python Data Analysis Library. URL: https://pandas.pydata.org.

85. Matplotlib – Visualization with Python. Matplotlib – Visualization with Python. URL: https://matplotlib.org.

86. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

87. Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

88. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image

segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818. https://doi.org/10.1007/978-3-030-01234-2_49.

89. Schmidhuber, J. Deep Learning in Neural Networks: An Overview [Electronic resource] / J. Schmidhuber // arXiv.org. – 2014. – Available at: https://arxiv.org/abs/1404.7828. – Submitted on Apr. 30, 2014.

ДОДАТКИ

ДОДАТОК А
СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації:

[1] О.В. Коменчук, О.Б. Мокін, «Аналіз методів передоброблення панорамних стоматологічних рентгенівських знімків для задач сегментації зображень», Вісник Вінницького політехнічного інституту, вип. 5, с. 41-49, 2023. DOI: https://doi.org/10.31649/1997-9266-2023-170-5-41-49 .

(Особистий внесок: Дослідження включає оцінювання та покращення фільтрів CLAHE, Гауса, білатерального фільтра та MSRCR за допомогою тонкого налаштування для покращення якості медичних зображень. Модель сегментації на базі U-Net навчалась на попередньо оброблених зображеннях, а ефективність перевірялась за допомогою метрик Дайса, Жаккара, влучності та повноти. О.Б. Мокін: узгодження задач дослідження та аналіз результатів, компонування і загальне редагування статті.)

[2] О.В. Коменчук, В.Б. Мокін, Є.М. Крижановський, В.О. Будяк, «Інтелектуальна технологія побудови плану будівель за аерофотозйомкою їхніх дахів», Вісник Вінницького політехнічного інституту, вип. 1, с. 101-109, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-172-1-101-109 .

(Особистий внесок: Запропоновано удосконалення інтелектуальної технології побудови плану будівель за даними аерофотозйомки їх дахів, шляхом інтеграції моделі для детекції DETR (“DEtection TRansformer”) разом із сегментацією на основі ViTs (Vision Transformers) для комплексного вирішення проблем пошуку та ідентифікації дахів, у першому наближенні, для подальшого покращення побудови планів будівель. В.Б. Мокін: узгодження задач дослідження, аналіз результатів та загальне редагування статті. Є.М. Крижановський: методологія автоматизації практичного розгортання розробленої технології на практиці з використанням ГІС-технологій. В.О.

Будяк: відбір прикладів для застосування технології та аналіз ефективності її застосування.)

[3] О.В. Коменчук, О.Б. Мокін, «Інформаційна технологія прискореного анотування медичних зображень в задачах сегментації на основі моделей глибокого навчання», Вісник Вінницького політехнічного інституту, вип. 4, с. 95-103, 2024. DOI: https://doi.org/10.31649/1997-9266-2024-175-4-95-103 .

(Особистий внесок: Уперше розроблено інтелектуальну інформаційну технологію прискореного анотування медичних зображень для задач сегментації з використанням моделей глибокого навчання та псевдомасок. Проведено порівняння продуктивності інструментів для створення анотацій медичних зображень, зокрема Supervisely, CVAT та Segments.ai, з розробленою інформаційною технологією, яка використовує модель Language Segment-Anything з текстовими підказками для створення первинних псевдомасок та уперше розробленим автоматичним валідуванням на основі моделі EfficientNet-B2. О.Б. Мокін: узгодження задач дослідження та аналізу результатів, компонування і загальне редагування статті.)

[4] О.В. Коменчук, «Адаптивні методи попередньої обробки для підвищення точності сегментації стоматологічних рентгенівських знімків», Сучасний стан наукових досліджень та технологій в промисловості, вип. 3, с. 29-38, 2024. DOI: https://doi.org/10.30837/2522-9818.2024.3.029 .

(Особистий внесок: Проведено роботу над удосконаленням білатерального фільтру та фільтру CLAHE для попереднього оброблення даних за рахунок розробки та впровадження адаптивного налаштування параметрів фільтрів на основі контексту зображення та проведено експериментальне порівняння розроблених адаптивних методів попередньої обробки з традиційними підходами.)

Наукові праці, які засвідчують апробацію матеріалів дисертації:

[1] В.Б. Мокін, О.В. Коменчук, «Огляд сучасних інформаційних технологій для класифікації зображень за матеріалами конкурсів Kaggle 2019-2020 років»,

Матеріали L науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 1469-1471, 2021. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38276/135476.pdf?sequence=2&isAllowed=y .

(Особистий внесок: Проведено огляд сучасних інформаційних технологій для класифікації зображень. В.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[2] В.Б. Мокін, О.В. Коменчук, «Сучасні інформаційні технології для розпізнавання дахів будівель на аерофотозйомці», Матеріали LI науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 704-708, 2022. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/40198/16150.pdf?sequence=3&isAllowed=y .

(Особистий внесок: Описана технологія розпізнавання дахів будівель на аерофотозйомці, продемонстровано приклад роботи запропонованої технології. В.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[3] О.В. Коменчук, С.А. Кирилащук, «Огляд сучасних педагогічних технологій в освітньому процесі», Матеріали Всеукраїнської науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», с. 458-487, 2022. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2022/paper/viewFile/15485/13068 .

(Особистий внесок: Розглянуто сучасні педагогічні технологій в освітньому процесі. С.А. Кирилащук: загальне редагування тексту.)

[4] О.В. Коменчук, О.Б. Мокін, «Методи передобробки панорамних стоматологічних рентгенівських знімків для задач глибокого навчання», Матеріали LII науково-технічної конференції підрозділів Вінницького національного технічного університету, с. 229-231, 2023. Режим доступу: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/38239/17182.pdf?sequence=3&isAllowed=y .

(Особистий внесок: Проведено огляд методів для передоброблення медичних знімків, визначено кроки для вдосконалення таких методів. О.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

[5] О.В. Коменчук, О.Б. Мокін, «Удосконалення процесу анотування медичних зображень для задач сегментації за допомогою текстових підказок», Матеріали Міжнародної науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», 2024. Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/21692/17951 .

(Особистий внесок: Описаний підхід для удосконалення процесу анотування медичних зображень за допомогою текстових підказок, розроблено веб-застосунок. О.Б. Мокін: узгодження задач дослідження, загальне редагування тексту.)

ДОДАТОК Б
АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

ЗАТВЕРДЖУЮ
Директор ТОВ «ОРНЕТІКА»
_____ Олена САМАРСЬКА
«6» вересня 2024 р.

АКТ
про впровадження результатів дисертаційної роботи
«Інтелектуальна інформаційна технологія розпізнавання об'єктів на зображеннях з використанням псевдомасок»
Коменчука Олега Вікторовича

Цей акт підтверджує, що результати дисертаційної роботи на здобуття наукового ступеня доктора філософії Коменчука О.В., рекомендовані до впровадження в практичну діяльність при розробці інтелектуальних систем для вирішення задач комп'ютерного зору. Зокрема цінними є наступні отримані результати дослідження Коменчука О.В.:

- Використання розроблених покращених методів попередньої обробки даних на основі контексту зображення, зокрема використання адаптивного білатерального фільтра та покращеного фільтра CLAHE до панорамних стоматологічних рентгенівських знімків зубів для вирішення задач сегментації та детекції об'єктів на КТ знімках.
- Проєктування системи з використанням розробленої інформаційної технології на основі моделі Language Segment-Anything з текстовими підказками для автоматичного створення первинних псевдомасок.
- Використання розробленої системи автоматичної валідації на основі моделі EfficientNet для перевірки точності створення псевдомасок.
- Використання методів створення псевдомасок для підвищення точності виявлення об'єктів на зображеннях.
- Підготовка та тренування моделей сегментації та детекції для вирішення задачі ідентифікації об'єктів на зображеннях з використанням сучасних фреймворків глибокого навчання на Python (PyTorch, PyTorch Lightning).

Даний акт не є основою для проведення фінансових взаєморозрахунків.

Голова комісії _____ Олена САМАРСЬКА

Члени комісії _____ Андрій РЕМОВСЬКИЙ

_____ Костянтин ВОВК

ДОДАТОК В
АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ В НАВЧАЛЬНИЙ ПРОЦЕС ВНТУ

ЗАТВЕРДЖУЮ
Проректор з науково-педагогічної роботи та організації освітнього процесу Вінницького національного технічного університету

_____ доц. Олександр ПЕТРОВ
«19» вересня 2024 р.

АКТ
про впровадження результатів дисертаційної роботи
Коменчука Олега Вікторовича
«Інтелектуальна інформаційна технологія розпізнавання об'єктів на зображеннях з використанням псевдомасок», представленої до захисту на здобуття наукового ступеня доктора філософії зі спеціальності 126 Інформаційні системи та технології, в навчальному процесі

Комісія Вінницького національного технічного університету у складі:

голова комісії – декан Факультету інтелектуальних інформаційних технологій та автоматизації (ФІІТА), к.т.н, доцент Володимир СЕВАСТЬЯНОВ та члени комісії – заступник декана з навчально-методичної роботи ФІІТА, ст. викладач каф. системного аналізу та інформаційних технологій (САІТ) Василь ПРИСЯЖНЮК, заступник завідувача кафедри САІТ ФІІТА, к.т.н., доцент Євгеній КРИЖАНОВСЬКИЙ, склала цей акт про підтвердження того, що дійсно у Вінницькому національному технічному університеті під час викладання освітніх компонентів «Інтернет речей та інтелектуальний аналіз даних» та «Інформаційні технології моніторингу та аналізу стану складних систем» здобувачам, які навчаються за освітньо-науковою програмою «Інформаційні системи та технології» третього (освітньо-наукового) рівня вищої освіти зі спеціальності 126 Інформаційні системи та технології, впроваджено результати дисертаційної роботи на здобуття доктора філософії аспіранта кафедри САІТ Коменчука Олега Вікторовича у ряді лекцій та практичних занять з питань інформаційних технологій.

При викладанні цих дисциплін використовуються такі результати досліджень, отримані у дисертаційній роботі Коменчука О.В.:

- підготовка набору даних стоматологічних рентгенівських знімків зубів для навчання моделей виявлення об'єктів на зображеннях: первинний аналіз даних, візуалізація розподілів, огляд сучасних Python-бібліотек для роботи з зображеннями;
- використання методів попередньої обробки даних, зокрема нормалізації зображень, білатеральний фільтр, CLAHE та інших алгоритмів;
- тренування та використання моделей виявлення об'єктів на зображеннях, таких як ResNet, U-Net, EfficientNet, DETR та Vision Transformer (ViT), з використанням сучасних фреймворків Python (PyTorch, PyTorch Lightning);
- використання методів створення псевдомасок для підвищення точності виявлення об'єктів на зображеннях;
- використання інформаційної технології на основі моделі Language Segment-Anything з текстовими підказками для автоматичного створення первинних псевдомасок.

Використання зазначених результатів дозволило підвищити якість навчального процесу із згаданих дисциплін.

Голова комісії _____ Володимир СЕВАСТЬЯНОВ

Члени комісії _____ Василь ПРИСЯЖНЮК

_____ Євгеній КРИЖАНОВСЬКИЙ

ДОДАТОК Г
АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

ЗАТВЕРДЖУЮ

Директор ТОВ «Аналітика»

Сергій ОЛІЙНИК

«9» вересня 2024 р.

АКТ
про впровадження результатів дисертаційної роботи
Коменчука Олега Вікторовича
«Інтелектуальна інформаційна технологія розпізнавання об'єктів на зображеннях з використанням псевдомасок», представленої до захисту на здобуття наукового ступеня доктора філософії зі спеціальності 126 Інформаційні системи та технології

Комісія ТОВ «Аналітика» у складі:

голова комісії – начальник відділу створення топографо-геодезичних, картографічних матеріалів та геопросторових даних Ірина ТЕТЕРЯ та члени комісії – фахівець відділу технік–аерофотограмметрист Олена ПАНЧЕНКО, інженер-землевпорядник Дмитро ЗОЛОТАР, склала цей акт про підтвердження того, що дійсно до ТОВ «Аналітика» передано такі результати дисертаційної роботи на здобуття доктора філософії аспіранта кафедри САІТ Коменчука Олега Вікторовича:

- результати аналізу найбільш поширеніших типів дахів (односхилі, двосхилі, плоскі, вальмові) та створений узагальнений алгоритм їх класифікації за даними аерофотозйомки з використанням інтелектуальних технологій;
- натренована інтелектуальна модель виявлення об'єктів на зображеннях на основі DETR та Vision Transformers (ViTs), натренована з використанням сучасних фреймворків PyTorch та PyTorch Lightning;
- комбінований підхід для побудови планів будівель на основі даних аерофотозйомки з інтеграцією моделі DETR для детекції дахів, сегментації за допомогою ViTs (Vision Transformers) та методу Canny Edge Detection для визначення контурів дахів;
- підвищення точності виявлення об'єктів на зображеннях з використанням методів створення псевдомасок.

Проаналізовано ефективність наведених результатів та рекомендовано їх впровадити у роботі ТОВ «Аналітика» під час розв'язання задач побудови планів будівель за даними аерофотозйомки.

Даний акт не є основою для проведення фінансових взаєморозрахунків.

Голова комісії Ірина ТЕТЕРЯ

Члени комісії Олена ПАНЧЕНКО

Дмитро ЗОЛОТАР