

О. В. Коменчук<sup>1</sup>  
О. Б. Мокін<sup>1</sup>

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРИСКОРЕНОГО АНОТУВАННЯ МЕДИЧНИХ ЗОБРАЖЕНЬ В ЗАДАЧАХ СЕГМЕНТАЦІЇ НА ОСНОВІ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

<sup>1</sup>Вінницький національний технічний університет

Проаналізовано інструменти для створення анотацій медичних зображень у задачах сегментації зображень. Проведено порівняння продуктивності відомих інструментів *Supervisely*, *CVAT* та *Segments.ai* із запропонованою у роботі інформаційною технологією, яка використовує модель *Language Segment-Anything* з релевантними текстовими підказками і механізмом автоматичного валідування на основі моделі класифікації *EfficientNet-B2*.

Основною метою дослідження було визначення оптимального підходу до автоматизації процесу анотування зображень для забезпечення максимальної швидкості зі збереженням експертної точності. Результати показали, що використання інструмента *Supervisely* дозволило скоротити час на первинну анотацію до 39,7 секунд, але вимагало додаткових 59,5 секунд на корегування масок. *CVAT*, зі своїми напівавтоматичними інструментами, забезпечував створення масок за 64,8 с, але потребував ще 85,1 с на корегування. У порівнянні з ними, *Segments.ai* вимагав повної ручної анотації, що тривала 130,2 с. Водночас, розроблена інформаційна технологія, яка використовує модель *Language Segment-Anything* з налаштованими під задачу текстовими підказками та додатковий механізм автоматичного валідування, значно знизилася час на створення анотацій — до близько 29,6 с на зображення, а також знизилася час на ручне корегування до 45,4 с.

Розроблена інформаційна технологія продемонструвала високу швидкість та точність у створенні псевдомасок, підтверджену експериментальними результатами. Основні переваги цього підходу полягають у зменшенні необхідності в ручному корегуванні та підвищенні ефективності процесу анотування медичних зображень.

Це дослідження вказує на значний потенціал застосування автоматизованих методів для прискорення анотування у сфері комп'ютерного зору, сприяючи покращенню швидкості виконання завдань аналізу медичних даних зі збереженням потрібної якості.

**Ключові слова:** інформаційна технологія, штучний інтелект, глибоке навчання, сегментація, анотування даних, псевдомаски, автоматичне валідування.

## Вступ

У вирішенні сучасних завдань комп'ютерного зору та задач з вчителем (*supervised learning*) у машинному навчанні процес анотування медичних зображень є важливим етапом, особливо у контексті завдань сегментації — точність та швидкість анотування мають критичне значення для досягнення високоякісних результатів. Проте, ручне анотування великих наборів даних є витратним за часом та зусиллями, що ставить під сумнів його ефективність в умовах роботи з великим обсягом даних [1]. Тому використання автоматизованих методів анотування стає критично важливим завданням у сучасних дослідженнях.

У цьому контексті розроблено інформаційну технологію, яка інтегрує передові моделі та методи для автоматичного анотування медичних зображень. Основним компонентом є модель *Language Segment-Anything* [2], розроблена компанією Meta, яка здатна генерувати псевдомаски за допомогою текстових підказок, що значно прискорює процес анотування. В цьому дослідженні проведено порівняльний аналіз ефективності цієї технології порівняно з традиційними, такими як

Supervisely [3], CVAT [4] та Segments.ai [5]. Дослідження включало оцінку середнього часу, необхідного для анотування одного зображення. До того ж, оцінено, як різні текстові підказки (prompts) впливають на точність створення псевдомасок за допомогою розробленої інформаційної технології у порівнянні з експертним анотуванням.

*Метою дослідження є визначення найефективнішого підходу до автоматичного анотування медичних зображень з використанням псевдомасок, що може бути застосований на практиці для підвищення продуктивності та точності процесів анотування. Особливу увагу приділено розробці технології, яка дозволить ефективно використовувати обмежені ресурси та мінімізувати потребу у ручному корегуванні анотацій, забезпечуючи при цьому високу якість результатів.*

### Аналіз інструментів для анотування даних

Анотування даних є критично важливим етапом у підготовці даних для машинного навчання, особливо для задач комп'ютерного зору, таких як сегментація зображень. Існує декілька популярних інструментів, які допомагають автоматизувати та пришвидшити цей процес. Розглянемо основні з них: Supervisely, CVAT та Segments.ai.

1. Supervisely — потужний інструмент для анотування зображень, який пропонує різноманітні функції для підвищення ефективності роботи. Однією з головних особливостей є використання Segment Anything Model (SAM) [6], моделі з відкритим кодом від компанії Meta, для автоматичного генерування сегментаційних масок в межах визначених координат обмежувальних рамок.

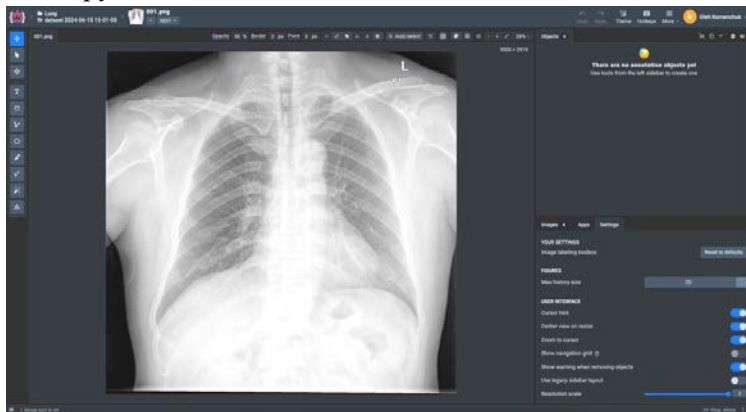


Рис. 1. Користувачський інтерфейс для анотування даних у Supervisely

додатковий час на корегування автоматично згенерованих масок.

2. CVAT — ще один популярний інструмент для анотування даних, створений і підтримуваний компанією Intel. Цей інструмент призначений для анотування зображень для задач комп'ютерного зору.

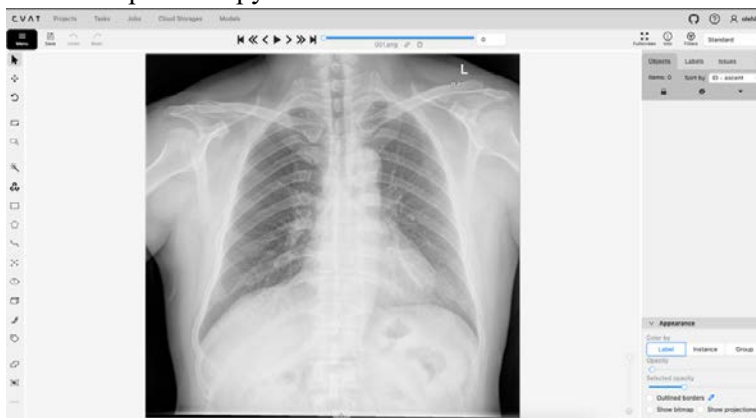


Рис. 2. Користувачський інтерфейс для анотування даних у CVAT

3. Segments.ai — інструмент, який дозволяє анотувати дані для задач комп'ютерного зору.

Основні функції:

– простий інтерфейс для анотацій зображень (рис. 3).

Основні функції:

– інтуїтивний інтерфейс для анотування зображень (рис. 1);

– інтеграція допоміжних моделей глибокого навчання для автоматичного розпізнавання та сегментування, наприклад, SAM.

Час анотування:

– використання SAM дозволяє значно скоротити час, необхідний для створення первинної анотації, тобто сегментаційних масок;

– проте, для забезпечення точнішої розмітки, необхідно витратити

Основні функції:

– зрозумілий користувачський інтерфейс для анотування зображень (рис. 2);

– підтримка різних типів анотацій (сегментація, класифікація, детекція);

– можливість інтеграції з іншими інструментами (через API).

Час анотування:

– створення сегментаційних анотацій може тривати значний час, оскільки CVAT має лише вбудовані прості методи автоматизації анотування, які базуються на бібліотеці OpenCV [7].

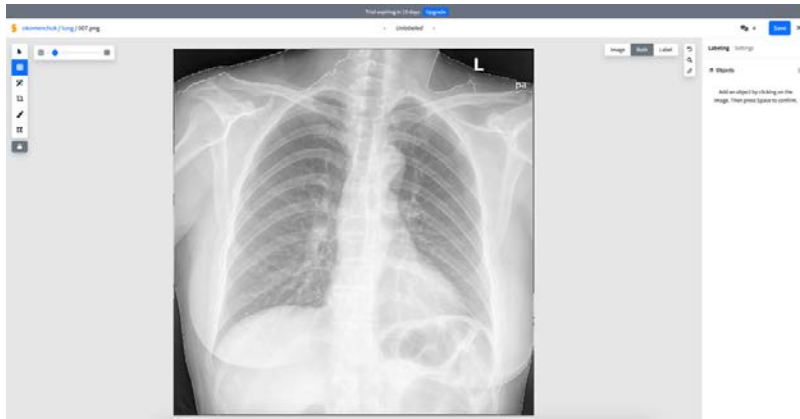


Рис. 3. Користувачський інтерфейс для анотації даних у Segments.ai

Час анотації:

– відсутність інструментів для автоматизації процесу анотування потребує значного часу на створення масок сегментації вручну.

### Інформаційна технологія для прискорення анотування

У цій роботі запропоновано інформаційну технологію для прискореного анотування медичних зображень в задачах сегментації. Ця технологія включає використання моделей глибокого навчання, а саме Language Segment-Anything з використанням текстових підказок (prompts) для створення початкових псевдомасок з їхнім подальшим автоматичним валідуванням для підвищення точності анотування.

Для автоматичної валідації цільового класу в межах обмежувальної рамки псевдомаски використовується передтренована модель класифікації на базі архітектури EfficientNet-B2 [8]. Для тренування моделі використовувалися такі гіперпараметри:

- кількість епох: 100;
- розмір міні-батча: 8;
- оптимізатор: Adam;
- коефіцієнт швидкості навчання (learning rate): 0,001.

Структурна схема запропонованого підходу показана на рис. 4.

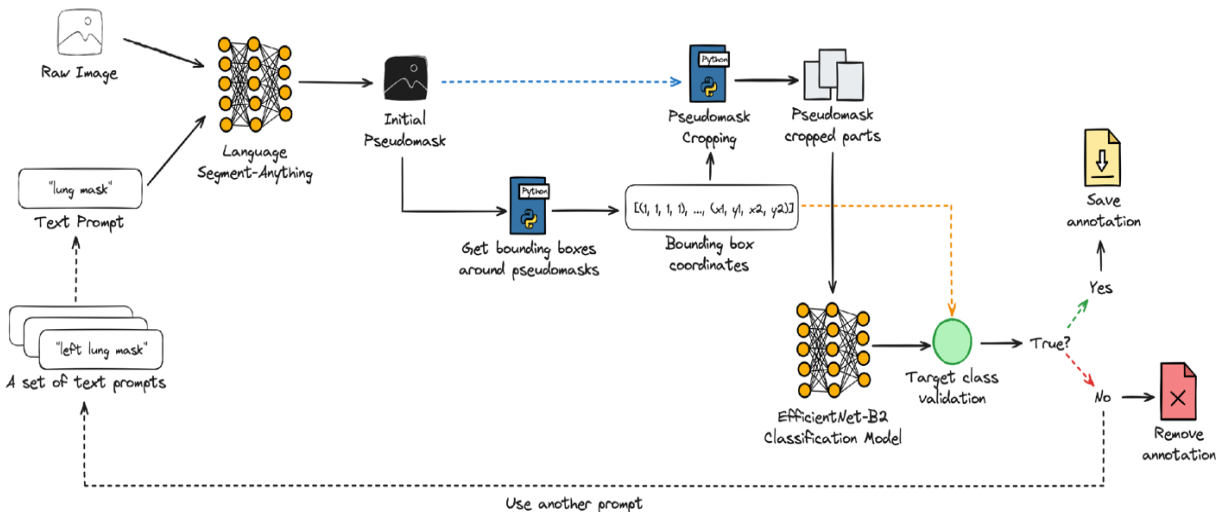


Рис. 4. Структура інформаційної технології прискореного анотування медичних зображень

Як видно з рис. 4, спершу застосовується модель Language Segment-Anything для створення первинних псевдомасок сегментації на основі текстових підказок. Ця модель здатна інтерпретувати текстові команди та генерувати відповідні маски для об'єктів на зображеннях. Оптимальні текстові підказки підбираються експериментальним шляхом для забезпечення максимальної відповідності псевдомасок до об'єктів, що анотуються, що дозволяє моделі точно виділяти потрібні області на зображеннях. Під час підбору підказок потрібно насамперед ґрунтуватися на метриці повноти

(recall) [9], щоб охопити якомога більше цільових об'єктів, навіть якщо результати будуть містити деякі зайві маски, які лише схожі на цільові.

Наступним кроком є визначення обмежувальних рамок навколо кожної псевдомаски, що дозволяє виділити області, які будуть перевірені далі. Частина зображення всередині обмежувальних рамок передаються до моделі класифікації на базі архітектури EfficientNet-B2, яка визначає, чи належить об'єкт всередині рамки до цільового класу. Якщо об'єкт не належить до цільового класу, такі анотації видаляються. Такий підхід дозволяє автоматично валідувати та видаляти зайві псевдомаски, тим самим підвищуючи метрику влучності (precision) [9] сегментації.

Використання запропонованого підходу до автоматичного валідування дозволяє значно скоротити час на ручну перевірку анотацій та забезпечити високу точність результатів. Потенційна інтеграція з наявними системами управління даними дозволить забезпечувати безперервний процес анотування та валідування.

### Порівняльний аналіз технологій сегментаційного анотування

У рамках цього дослідження проведено порівняльний аналіз швидкості анотування даних за допомогою трьох наведених вище інструментів та запропонованої технології. Використано відкритий набір даних (рис. 5), що містить близько 800 анонімізованих рентгенівських знімків з розміткою, створеною професійними радіологами [10]. Анотована частина набору даних містить високоякісні анотації для задачі сегментації легень, які створені радіологами ручним способом. В ході дослідження ця анотація використовувалася в якості істинних міток (ground truth). Для кожного інструмента вимірювався час, необхідний для створення анотацій, та оцінено ефективність відповідних інтегрованих засобів автоматизації.

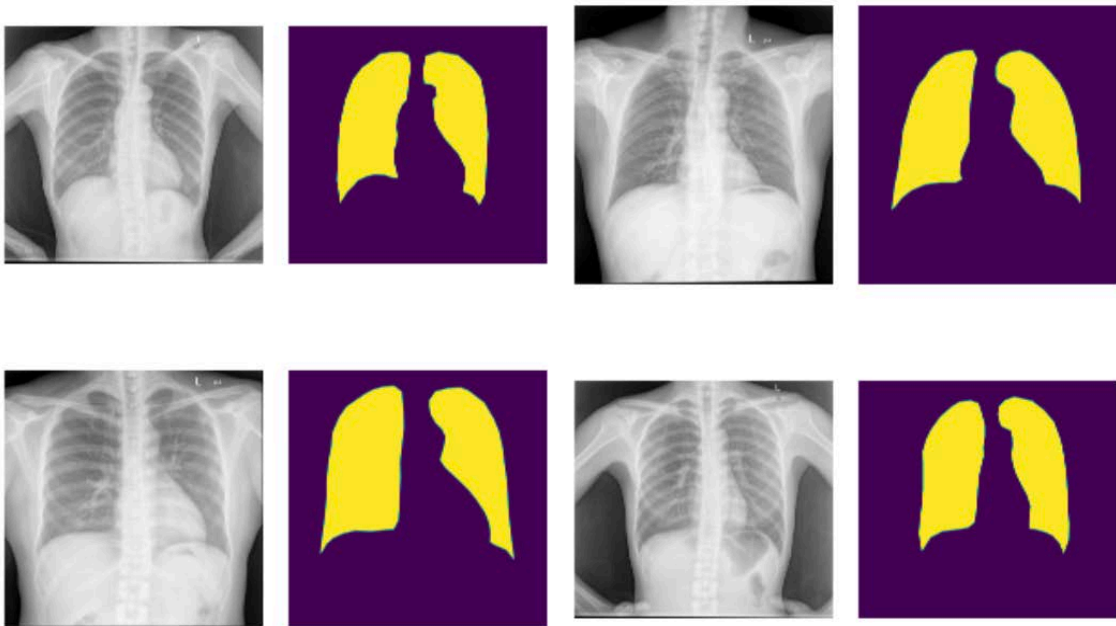


Рис. 5. Приклади зображень та масок сегментації, створених експертною групою радіологів

Швидкість анотування оцінювалася за часом, необхідним для анотування фіксованої кількості зображень, а точність — за допомогою метрик, таких як влучність (precision), повнота (recall) та коефіцієнт Жаккара (відомий також як Intersection over Union або IoU) [11] між створеними сегментаційними псевдомасками та істинними мітками.

Влучність визначає частку правильно сегментованих пікселів серед усіх пікселів, які модель визначила як частину об'єкта. Виражається формулою:

$$precision = \frac{|A \cap B|}{|B|}, \quad (1)$$

де  $A$  — це набір пікселів у «ground truth» масці, а  $B$  — набір пікселів у прогнозованій масці.

Повнота показує частку правильно сегментованих пікселів серед усіх пікселів, які належать до об'єкта в «ground truth» масці, і виражається формулою

$$recall = \frac{|A \cap B|}{|A|}. \quad (2)$$

Коефіцієнт Жаккара є стандартною метрикою для оцінки якості сегментації зображень, який визначається так:

$$Io = \frac{|A \cap B|}{|A \cup B|}. \quad (3)$$

Чим ближче значення  $IoU$  до 1, тим точнішим є перетин масок.

Середній час анотування одного зображення будемо визначати як

$$T_{avg} = \frac{1}{N} \sum_{i=1}^N T_i, \quad (4)$$

де  $N$  — кількість зображень,  $T_i$  — час анотування для  $i$ -го зображення.

Експерименти цього дослідження спрямовано на оцінку швидкості кожного підходу, включно з розробленою інформаційною технологією. Під час проведення експериментів спочатку використовувалася методи автоматизації анотування відповідного інструмента, а далі, використовуючи розмітку експертів, автори вручну доводили анотації до експертного рівня, фіксуючи час, витрачений на це додаткове ручне корегування.

Таким чином процес складався з двох етапів:

1. Автоматичне анотування зображень за допомогою всіх чотирьох засобів: Supervisely, CVAT, Segments.ai, а також інформаційної технології, що пропонується. Автоматизація допомагала знизити час на первинну анотацію зображень, використовуючи відповідні, описані вище, методи для створення первинних псевдомасок.

2. Ручне корегування анотацій для досягнення рівня експертної розмітки, тобто після автоматичного анотування отримані маски перевірялися та корегувалися вручну, використовуючи наявну експертну розмітку.

Результати, спостереження та висновки:

1. Supervisely завдяки використанню моделі Segment Anything дозволяв значно скоротити час на первинну анотацію, проте вимагав додаткового часу на ручне корегування масок.

2. CVAT, використовуючи напівавтоматичні інструменти анотування, також дозволяв швидко створювати маски, але потребував ще більшого часу на ручне корегування.

3. Segments.ai, не маючи вбудованих алгоритмів автоматизації, вимагав найбільше часу на анотування.

4. Запропонована інформаційна технологія завдяки використанню моделі Language Segment Anything з релевантними текстовими підказками та автоматичного процесу валідування згенерованих сегментаційних псевдомасок показала найкращий час.

У табл. 1 наведено узагальнені результати швидкості анотування для всіх чотирьох підходів.

Таблиця 1

Порівняння швидкості анотування з використанням різних підходів

Інструмент/підхід	Загальний час створення анотації, с/зображення	Середній час автоматичного анотування, с/зображення	Середній час ручного анотування, с/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

З використанням інструмента Supervisely та вбудованої моделі Segment Anything в середньому необхідно близько 39,7 с для отримання автоматичних анотацій та в середньому 59,5 с на корегування контурів масок. Водночас інструмент CVAT має вбудовані методи бібліотеки OpenCV для отримання автоматичних анотацій, які тривають в середньому 64,8 с на зображення, та потребує додаткових 85,1 с на корегування масок. Інструмент Segments.ai немає вбудованих методів для автоматизації отримання анотацій, тому цей інструмент потребує повністю ручного анотування,

на яке витрачається в середньому приблизно 130,2 с. Розроблена інформаційна технологія скорочує час на ручне корегування, оскільки автоматичне валідування фільтрує те, що раніше потребувало б перевірки та видалення людиною, заощаджуючи час, а також суттєво прискорює автоматичне створення псевдомасок за допомогою Language Segment-Anything з текстовими підказками. Завдяки цьому середній час на створення анотацій становить 75,0 с на зображення, що значно швидше порівняно з іншими досліджуваними підходами.

Для наочності результати також подано в графічному вигляді на рис. 6.

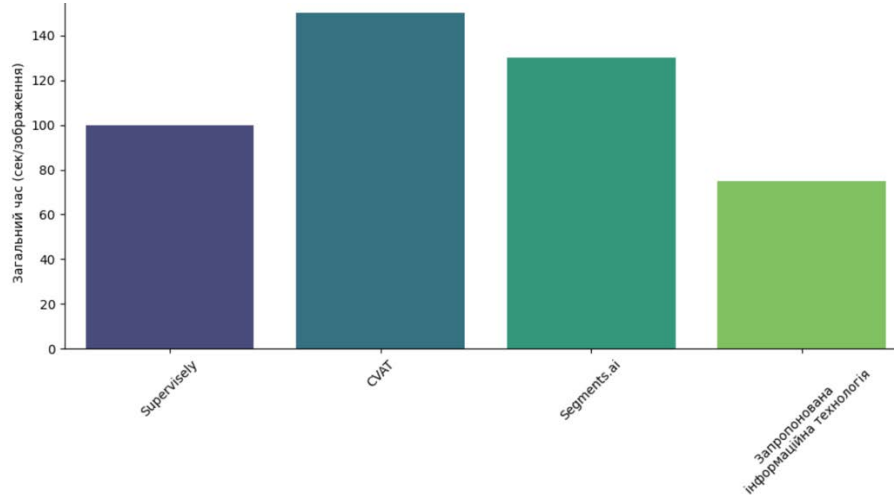


Рис. 6. Середній час анотування з урахуванням корегування для чотирьох порівнюваних підходів

### Підбір текстових підказок

У розробленій інформаційній технології одним з важливих етапів є правильний підбір релевантних текстових підказок у моделі Language Segment-Anything. Важливість цього кроку полягає в тому, що оптимальні текстові підказки дозволяють моделі повніше виділяти потрібні області на зображеннях, що покращує результати створення анотацій [12]. У таблиці 2 подано декілька прикладів різних підказок, які використано в ході експериментів, та відповідні отримані метрики точності створення первинних псевдомасок, виміряні за показниками влучності (precision), повноти (recall) та коефіцієнта Жаккара (IoU).

Таблиця 2

Приклади текстових підказок та отримані оцінки метрик на наборі даних

Текстова підказка	<i>Precision</i>	<i>Recall</i>	<i>IoU</i>
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 табл. 2, текстові підказки суттєво впливають на результати створення первинних псевдомасок. Найкращий результат продемонстрував другий варіант, зокрема і для ключової метрики повноти (для цього етапу). Для нього і варто застосовувати другий етап автоматичного валідування для підвищення влучності. Приклади псевдомасок, які виправлено за допомогою автоматичної валідування і тому не потребували додаткового часу на ручне корегування, показані на рис. 7. Як видно з рис. 7, механізм автоматичного валідування розробленої інформаційної технології дозволяє автоматично відфільтрувати псевдомаски, що не належать до цільового класу. Це значно економить час на ручні правки. Завдяки цьому підходу підвищується влучність псевдомасок. Для наочності на рис. 8 подано приклади анотації медичних зображень за допомогою розробленої інформаційної технології з урахуванням механізму автоматичного валідування.

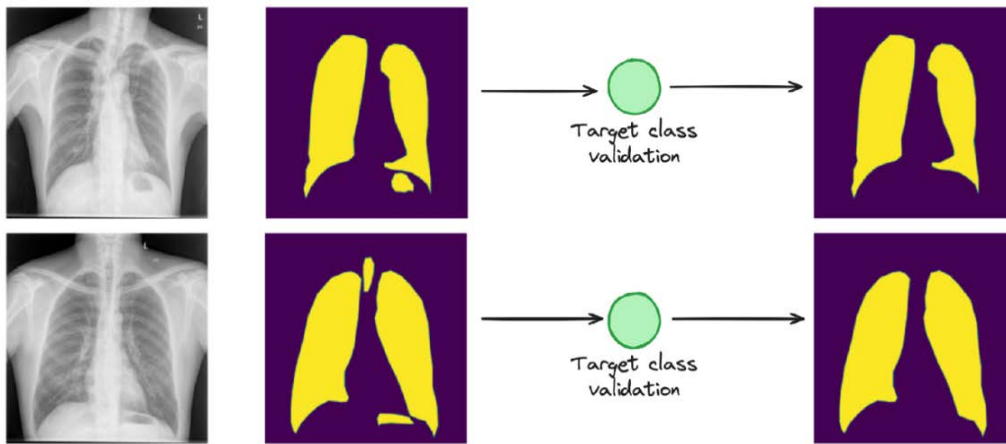


Рис. 7. Приклади роботи розробленого механізму автоматичного валідування попередньо (автоматично) згенерованих псевдомасок

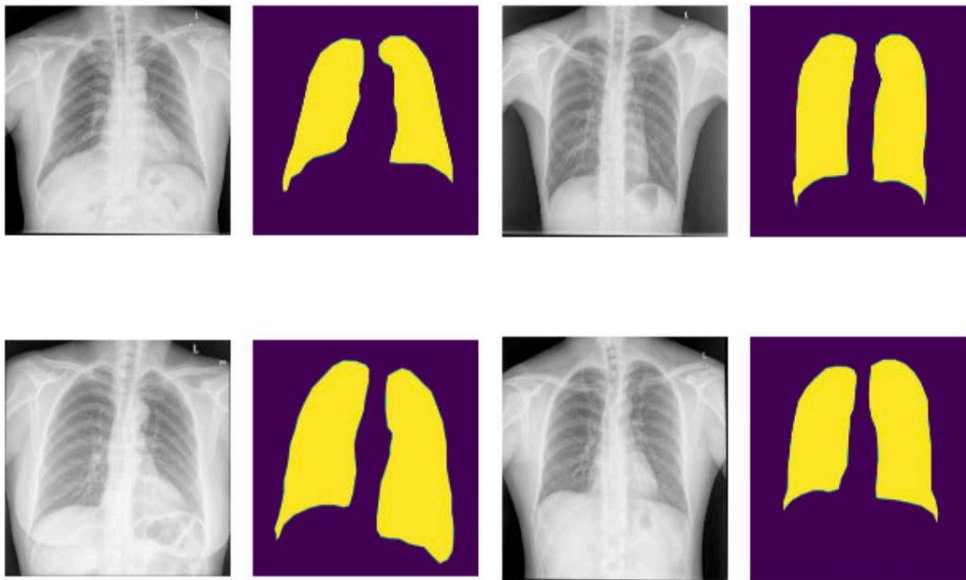


Рис. 8. Приклад анотацій, створених розробленою інформаційною технологією, до етапу ручного корегування

З цих результатів видно, що запропонована технологія демонструє високу точність створення псевдомасок для медичних зображень. Високі значення цільових метрик та зменшення потреби у ручному корегуванні зайвих псевдомасок свідчать про його переваги. Це робить запропонований підхід перспективним для використання в сегментаційних задачах комп'ютерного зору, особливо в умовах великих та різноманітних наборів даних. Подальші поліпшення можуть включати використання процесів перед- або післяобробки [13], для ще більшого рівня автоматичного підвищення точності, а також інтеграції з іншими сучасними інструментами та технологіями.

### Висновки

Досліджено швидкість та ефективність роботи різних інструментів та підходів для створення анотацій медичних зображень у задачах сегментації, включно з такими інструментами, як Supervisely, CVAT, Segments.ai, а також запропоновано розроблену інформаційну технологію для прискореного анотування медичних зображень, що поєднує модель автоматичного первинного сегментування зображення за допомогою Language Segment-Anything з релевантними текстовими підказками та механізм автоматичного валідування на основі моделі класифікації EfficientNet-B2.

Оцінено середній час, необхідний для анотування одного зображення:

1. Supervisely — використання моделі Segment Anything дозволяло скоротити час на первинну анотацію до близько 39,7 с, проте вимагало додаткових 59,5 с на корегування масок.

2. CVAT — напівавтоматичні інструменти бібліотеки OpenCV дозволяли створювати маски за 64,8 с, але потребували ще 85,1 с на корегування.

3. Segments.ai — не мав вбудованих методів автоматизації, тому вимагав повністю ручної анотації, що тривала в середньому близько 130,2 с.

4. Запропонована інформаційна технологія, використовуючи модель Language Segment-Anything з найкращими текстовими підказками та автоматичне валідування, вдалося скоротити час на первинне автоматичне створення анотацій до 29,6 с та ручне корегування до 45,4 с.

Запропонована інформаційна технологія демонструє високу швидкість та точність у створенні псевдомасок. З використанням оптимальних текстових підказок модель Language Segment-Anything дозволила досягти значних поліпшень процесу автоматичного анутовування зображень. Основною перевагою запропонованого комбінованого підходу є автоматичне валідування псевдомасок, що значно знижує потребу в ручному корегуванні розмітки.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] M. Aljabri, M. AlAmir, and M. AlGhamdi, "Towards a better understanding of annotation tools for medical imaging: a survey," *Multimed Tools Appl* 81, pp. 25877-25911, 2022. <https://doi.org/10.1007/s11042-022-12100-1>.

[2] "Language Segment-Anything," *GitHub*. [Electronic resource]. Available: <https://github.com/luca-medeiros/language-segment-anything>.

[3] "Supervisely," *GitHub*, [Electronic resource]. Available: <https://github.com/supervisely/supervisely>.

[4] "CVAT," *GitHub*. [Electronic resource]. Available: <https://github.com/cvat-ai/cvat>.

[5] "Segments.ai," *GitHub*. [Electronic resource]. Available: <https://github.com/segments-ai/segments-ai>.

[6] "Segment Anything Model (SAM)," *GitHub*. [Electronic resource]. Available: <https://github.com/facebookresearch/segment-anything>.

[7] "OpenCV: Open Source Computer Vision Library," *GitHub*. [Electronic resource]. Available: <https://github.com/opencv/opencv>.

[8] M. Tan, and Q.V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 2019*, pp. 6105-6114. [Electronic resource]. Available: <https://arxiv.org/pdf/1905.11946>.

[9] Malhotra Priyanka, Gupta Sheifali, Koundal Deepika, Zaguia Atef, and Enbeyle Wegayehu, "Deep Neural Networks for Medical Image Segmentation," *Journal of Healthcare Engineering*, 9580991, 15 pp., 2022. <https://doi.org/10.1155/2022/9580991>.

[10] "Pulmonary Chest X-Ray Defect Detection," *Kaggle*, [Electronic resource]. Available: <https://www.kaggle.com/datasets/nikhilpandey360/chest-xray-masks-and-labels/data>.

[11] F. van Beers, A. Lindström, E. Okafor and M. Wiering, "Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, vol. 1 ICPRAM, 2019, pp. 438-445. SciTePress. [Electronic resource]. Available: [https://pure.rug.nl/ws/portalfiles/portal/87088047/ICPRAM\\_2019\\_35.pdf](https://pure.rug.nl/ws/portalfiles/portal/87088047/ICPRAM_2019_35.pdf).

[12] Feng Li, et. al, "Visual In-Context Prompting," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 12861-12871. [Electronic resource]. Available: [https://openaccess.thecvf.com/content/CVPR2024/papers/Li\\_Visual\\_In-Context\\_Prompting\\_CVPR\\_2024\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2024/papers/Li_Visual_In-Context_Prompting_CVPR_2024_paper.pdf).

[13] О. В. Коменчук, і О. Б. Мокін, «Аналіз методів передоброблення панорамних стоматологічних рентгенівських знімків для задач сегментації зображень», *Вісник Вінницького політехнічного інституту*, № 5, с. 41-49, 2023. <https://doi.org/10.31649/1997-9266-2023-170-5-41-49>.

Рекомендована кафедрою системного аналізу та інформаційних технологій ВНТУ

Стаття надійшла до редакції 24.06.2024

**Коменчук Олег Вікторович** — аспірант кафедри системного аналізу та інформаційних технологій, e-mail: [komenchuk77@gmail.com](mailto:komenchuk77@gmail.com) ;

**Мокін Олександр Борисович** — д-р техн. наук, професор, професор кафедри системного аналізу та інформаційних технологій, e-mail: [abmokin@gmail.com](mailto:abmokin@gmail.com).

Вінницький національний технічний університет, Вінниця



O. V. Komenchuk<sup>1</sup>O. B. Mokin<sup>1</sup>

## Information Technology for Accelerated Annotation of Medical Images in Segmentation Tasks Based on Deep Learning Models

<sup>1</sup>Vinnitsia National Technical University

*The paper analyzed tools for creating annotations of medical images in image segmentation tasks. The performance of the well-known tools Supervisely, CVAT, and Segments.ai is compared with the information technology proposed in the work, which uses the Language Segment-Anything model with relevant text prompts and an automatic validation mechanism, based on the EfficientNet-B2 classification model.*

*The main objective of the study was to determine the optimal approach to the automation of the image annotation process to ensure maximum speed, maintaining expert accuracy. The results showed that usage of the Supervisely tool reduced the initial annotation time to 39.7 seconds, but required additional 59.5 seconds to adjust the masks. CVAT, with its semi-automated tools, produced masks in 64.8 seconds, but required another 85.1 seconds for adjustments. In comparison, Segments.ai required a full manual annotation, which took 130.2 seconds. At the same time, the developed information technology, which uses the Language Segment-Anything model with task-specific text prompts and an additional automatic validation mechanism, significantly reduced the time for creating annotations to about 29.6 seconds per image, and also reduced the time for manual correction to 45.4 seconds.*

*The developed information technology demonstrated high speed and accuracy in creating pseudo-masks, confirmed by experimental results. The main advantages of this approach are the decrease of time, needed for manual correction and increase the efficiency of the medical image annotation process.*

*This work points out to the significant potential of using automated methods to accelerate annotation in the field of computer vision, improving the speed of performing medical data analysis tasks while maintaining the desired quality.*

**Keywords:** information technology, artificial intelligence, deep learning, image segmentation, data annotation, pseudo-masks, automatic validation.

**Komenchuk Oleh V.** — Post-Graduate Student with the Chair of System Analysis and Information Technologies, e-mail: komenchuk77@gmail.com ;

**Mokin Oleksandr B.** — Dr. Sc. (Eng.), Professor, Professor with the Chair of System Analysis and Information Technologies, e-mail: abmokin@gmail.com