

ДУАЛЬНА АРХІТЕКТУРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АВТОМАТИЗАЦІЇ РОЗМІТКИ ДАНИХ ДЛЯ ЗАДАЧ КОМП'ЮТЕРНОГО ЗОРУ

Анотація: У цій роботі представлено дуальну архітектуру для автоматизації розмітки відеоданих, яка поєднує підходи з нульовим навчанням та активного навчання. Метою дослідження було підвищення продуктивності та точності розмітки даних, зменшення витрат часу та ресурсів. Виконані експерименти показали, що запропонована архітектура скорочує медіанний час розмітки відео на 125% у порівнянні з ручною розміткою, на 25% у порівнянні з підходом нульового навчання та на 34% у порівнянні з активним навчанням. Результати свідчать про значне підвищення ефективності процесу розмітки, що робить запропоновану дуальну архітектуру перспективним рішенням для автоматизації розмітки відеоданих.

Ключові слова: розмітка даних, автоматизації розмітки, алгоритми комп'ютерного зору, алгоритми нульового навчання, алгоритми активного навчання.

Вступ

Розмітка відеоданих є ключовим етапом у розробці та тренуванні алгоритмів комп'ютерного зору, проте вона залишається трудомістким та затратним процесом [1]. Традиційні методи розмітки базуються на ручному втручанні, що призводить до значних витрат часу і ресурсів, особливо у випадках великих наборів даних [2]. З іншого боку, автоматизовані підходи, такі як використання попередньо навчених моделей або активного навчання, мають свої обмеження. Попередньо навчені моделі ефективні для загальних категорій об'єктів, але можуть бути неточними для специфічних випадків, тоді як активне навчання потребує великої кількості попередньо розмічених даних для досягнення високої точності.

Запропонована дуальна архітектура автоматизації розмітки відеоданих об'єднує обидва підходи для забезпечення гнучкості та ефективності на різних етапах розмітки даних для задач комп'ютерного зору. Така інтеграція дозволяє максимізувати переваги кожного підходу, мінімізуючи їх недоліки, що робить процес розмітки більш швидшим. Враховуючи зростаючі обсяги відеоданих та потребу в їхній високоякісній розмітці, розробка та впровадження таких систем є надзвичайно актуальною задачею [3].

Огляд існуючих методів автоматизації розмітки даних для задач комп'ютерного зору

Існуючі підходи до розмітки даних для задач комп'ютерного зору можна класифікувати на такі категорії [4]:

- методи мануальної розмітки даних, який характеризується відсутністю автоматизації;

- методи з використання попередньо навчених алгоритмів комп'ютерного зору для автоматизації процесу розмітки даних, які охоплюють широко поширені категорії або застосування алгоритмів з нульовим навчанням.

- методи з використанням підходів активного донавчання моделей комп'ютерного зору на основі результатів розмітки даних.

Мануальний метод розмітки даних для задач комп'ютерного зору, представлений на рисунку 1, передбачає перегляд усіх зображень в наборі даних людиною, яка вручну відмічає всі наявні категорії на кожному зображенні.

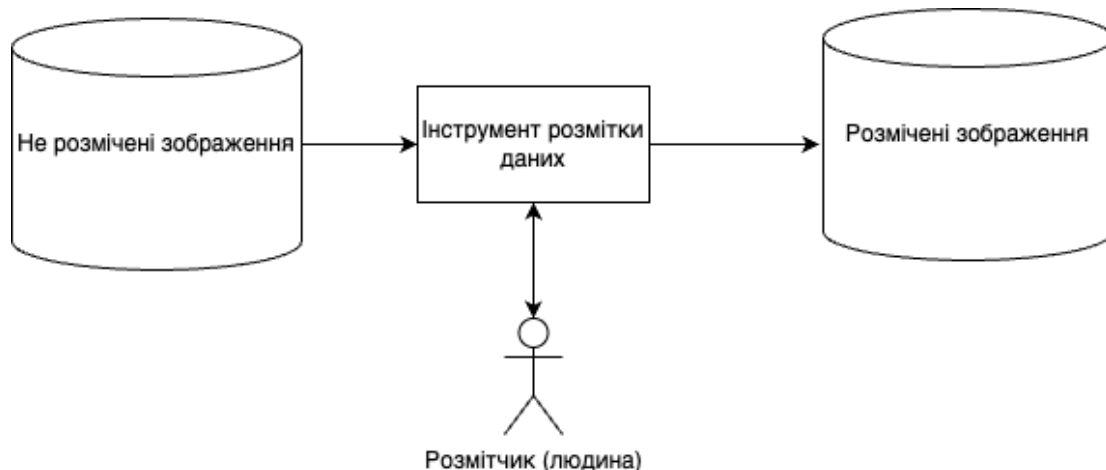


Рисунок 1. Мануальний метод розмітки

Оптимізація мануального підходу можлива шляхом застосування алгоритмів комп'ютерного зору, які навчаються на вже розмічених даних (процес донавчання). Ці алгоритми поступово набувають здатності надавати рекомендації щодо виявлення об'єктів на зображеннях, дозволяючи користувачеві лише підтверджувати або відхиляти припущення алгоритму, а не самостійно розмічати об'єкти. Альтернативно, можуть використовуватися алгоритми з нульовим навчанням, які здатні застосовуватися до даних, на яких вони не були попередньо навчені. Проте такі алгоритми зазвичай демонструють нижчу точність порівняно з донавченими моделями. Схематичне зображення цього підходу наведено на рисунку 2.

Застосування активного навчання моделей на основі результатів розмічених даних дозволяє системі визначати ті дані, які потребують додаткової уваги та

розмітки, з метою підвищення точності моделей. Під час активного навчання моделі ідентифікують неоднозначні чи складні приклади, що можуть збагатити варіативність об'єктів та сцен у розміченому наборі даних. Процес активного навчання може бути побудований на основі різноманітних стратегій, таких як вибір прикладів, що викликають високий рівень невизначеності для моделі, або врахування прикладів, що знаходяться на межі між різними категоріями. Це дозволяє моделі фокусуватися на тих областях, де її точність може бути покращена.

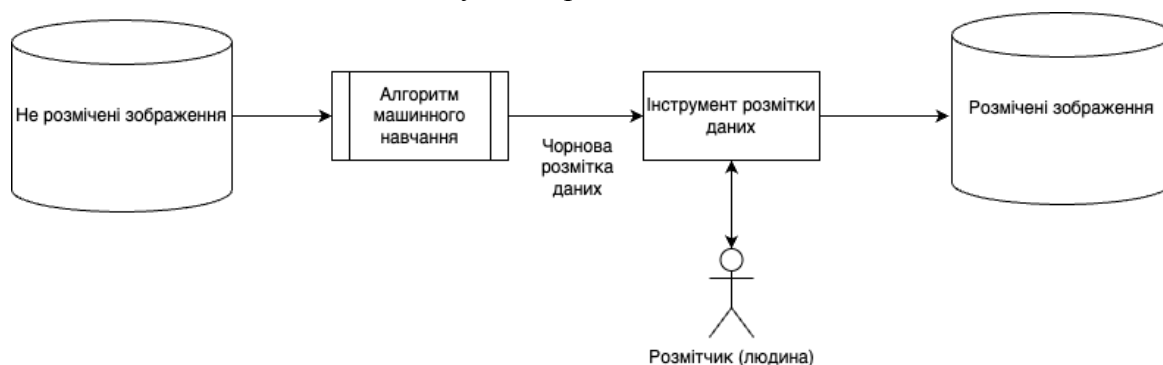


Рисунок 2. Підхід з використанням попередньо навчених алгоритмів машинного навчання для пришвидшення процесу розмітки даних

Навчання на зразках, вибраних активним чином, сприяє прискоренню процесу розмітки даних, оскільки модель самостійно визначає пріоритетні області для подальшої розмітки. Такий підхід дозволяє зменшити зусилля та час, необхідний для створення високоякісних розмічених наборів даних. Деталізацію цього процесу можна знайти на схематичному зображенні, поданому на рисунку 3.



Рисунок 3. Підхід з використанням методів активного донавчання моделей на результатах розмічених даних

Дуальна архітектура програмного забезпечення для автоматизації розмітки даних для задач комп'ютерного зору

Більшість інструментів для розмітки даних реалізують автоматизацію розмітки даних для задач комп'ютерного зору за допомогою або попередньо навчених алгоритмів, або алгоритмів активного навчання на вже розмічених даних. Проте, жоден з них не поєднує ці підходи, що є значним недоліком, оскільки кожен має свої переваги та обмеження на різних етапах проєкту з розмітки даних.

Підходи, що базуються на попередньо навчених моделях, використовують моделі з нульовим навчанням (zero-shot learning), такі як SAM [5], рисунок 4. Ці моделі навчені на обширних наборах даних та здатні розпізнавати об'єкти в різноманітних сценаріях. Моделі з нульовим навчанням зазвичай мають великі розміри, що коливаються від кількох десятків до сотень мільйонів параметрів, і демонструють посередні результати при роботі з рідкісними типами даних, наприклад, розмітка ядер клітин людських тканин. Водночас, ці моделі можуть показувати високі результати якості для поширених категорій об'єктів, таких як людина. Таким чином, зазначений підхід є оптимальним на початкових етапах розмітки даних, коли ще не накопичено достатньо даних для активного навчання.

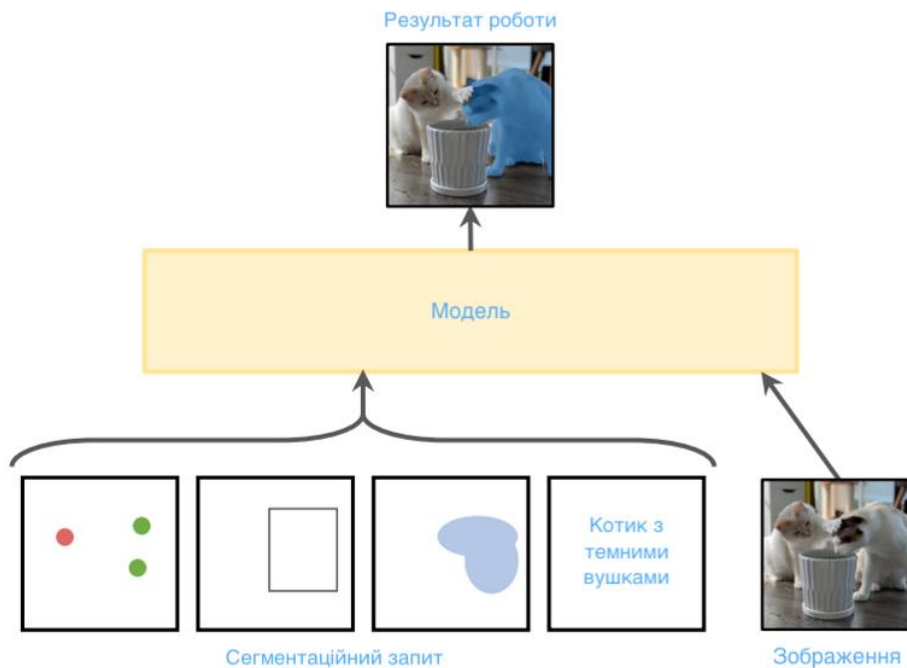


Рисунок 4. SAM – підхід з нульовим навчанням (zero-shot learning) [6]

На противагу, підходи активного навчання можуть використовувати менші за кількістю параметрів моделі, що містять від сотень тисяч до декількох мільйонів параметрів. Це робить їх застосування економічно доцільним у проєктах, що

вимагають розмітки великих об'ємів даних, але вони мають значний недолік – потребують часу для накопичення достатньої кількості даних для навчання моделей та досягнення рівня моделей з нульовим навчанням (zero-shot learning). В контексті цієї роботи використовується модель VideoIntern [7], тому підхід з активним навчанням підходить для середньої та пізньої стадії процесу розмітки даних, коли є вже певна кількість розмічених даних.

Ці підходи мають різні сильні та слабкі сторони і підходять для різних етапів розмітки даних. Тому в рамках даної дисертаційної роботи запропоновано підхід, який об'єднує обидва підходи в єдину систему шляхом моніторингу якості розмітки даних обома підходами та прийняттям рішення щодо того, з якої моделі буде використовуватися розмітка.

Оскільки моделі мають різну вартість використання, це вплинуло на алгоритм, який мінімізує кількість використання моделей з нульовим навчанням (zero-shot learning) шляхом пропуску обробки вхідних даних з частотою, що контролюється послідовністю Фібоначчі:

$$F_1 = 1, F_2 = 1, F_{n+2} = F_n + F_{n+1}, n = 1, 2, 3, \dots \quad (1)$$

Послідовність Фібоначчі обрана для експонентного зменшення частоти викликів алгоритму з нульовим навчанням, що підтверджується формулою Біне:

$$F_n = \frac{\phi^n - (-\phi)^{-n}}{\phi - (-\phi)^{-1}} \approx \frac{\phi^n}{\sqrt{5}}, \quad (2)$$

де ϕ – приймає значення золотого перетину $\phi = \frac{1 + \sqrt{5}}{2}$.

Алгоритм починається з ініціалізації керуючих значень a , b , та $control$, що виконується за константний час $O(1)$. Алгоритм обробляє датасет частинами, яка рівна K , що обмежена кількістю даних n в датасеті. На кожній ітерації, для кожної частини з датасету D , виконується обчислення моделі активного навчання A , що має складність $O(a)$. Раз на $control$ кроків, що контролюється послідовністю Фібоначчі, виконується порівняння якості алгоритмів активного A та нульового навчання Z . Для цього додатково виконується обрахунок алгоритмом нульового навчання Z , що має складність $O(z)$. Далі відбувається порівняння якості алгоритмів, що відбувається за допомогою метрики mAP, складність обчислення якої складає $O(C^2 \log_2 C)$, де C – кількість об'єктів на відео. Формалізоване представлення алгоритму у вигляді блок схеми наведено на рисунку 5.

Загальна складність алгоритму залежить від якості моделей активного та нульового навчання, що буде змінювати кількість викликів до моделей, але в найгіршому випадку, якщо алгоритм нульового навчання буде завжди точнішим за алгоритм активного навчання, то складність можна оцінити як $O(K \times (a + z + C^2 \log C))$.

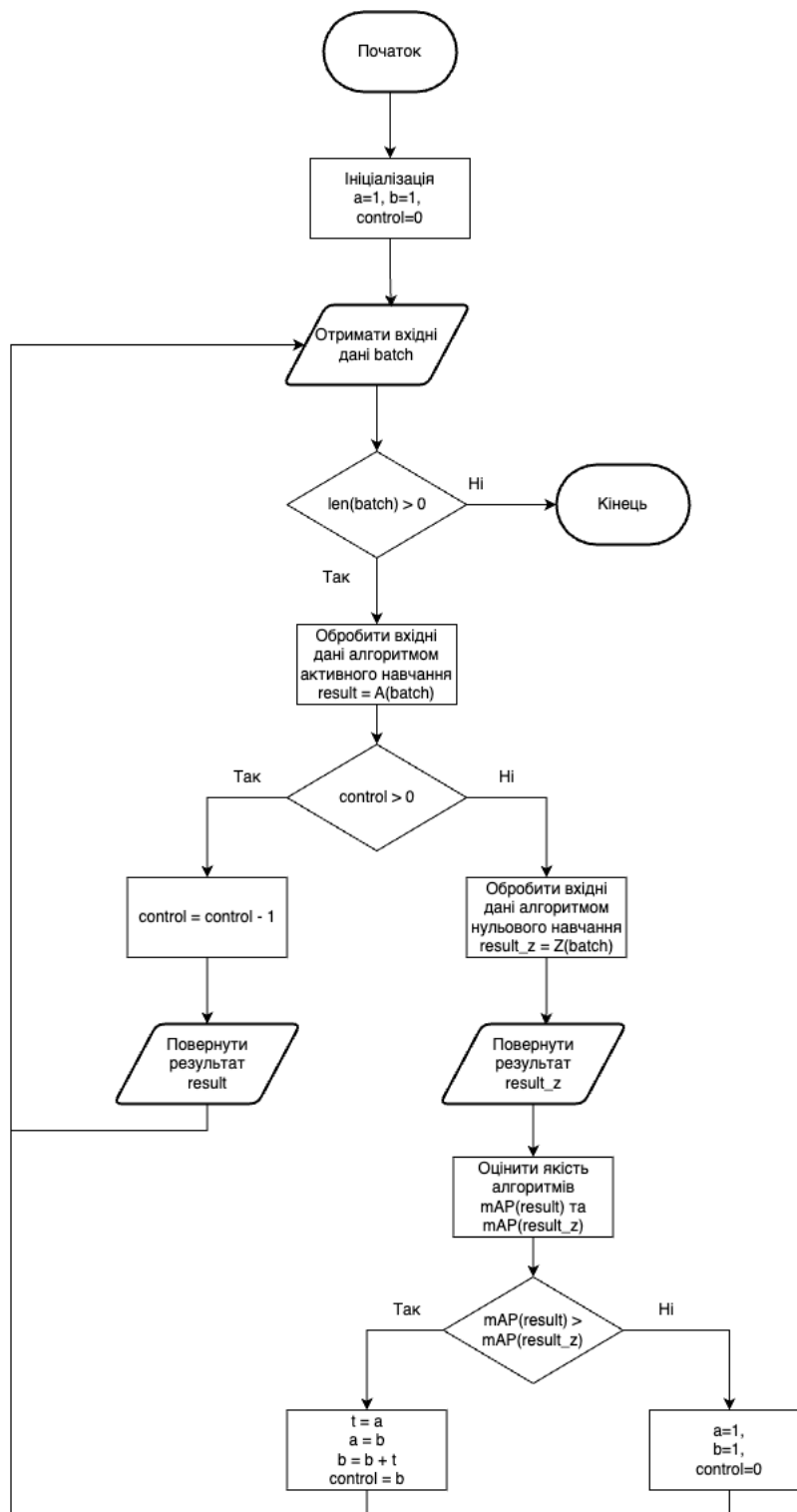


Рисунок 5. Блок-схема запропонованого алгоритму, алгоритм активного навчання позначено як A та алгоритм нульового навчання Z .

Оскільки всі операції на кожному кроці виконуються за кінцевий час і жоден цикл не є нескінченним, можна зробити висновок, що алгоритм завжди завершується

за скінченну кількість кроків. Це забезпечує його обчислювальну здійсненність, тобто відсутність безкінечних циклів або інших перешкод для його успішного завершення.

Принцип даного методу візуалізовано на рисунку 6, де наведено такі процеси:

- 1) визначення списку відео з найбільшою складністю для алгоритму детекції зображень за допомогою механізму пріоритезації;
- 2) виконання розмітки відібраних даних за допомогою алгоритмів активного навчання та нульового навчання (за необхідності);
- 3) вибір алгоритму, який має найвищу точність розмітки на даному етапі;
- 4) розмітчик даних аналізує розмітку, отриману від алгоритму автоматизації розмітки; зберігає правильну розмітку, видаляє хибну розмітку та виконує розмітку пропущених об'єктів;
- 5) інформація про правки, внесені розмітчиком даних, відправляється на алгоритми для подальшого навчання та мінімізації таких помилок у майбутньому;
- 6) інформація про реальну складність обраних зображень передається від алгоритму активного навчання до механізму пріоритезації процесу розмітки даних.

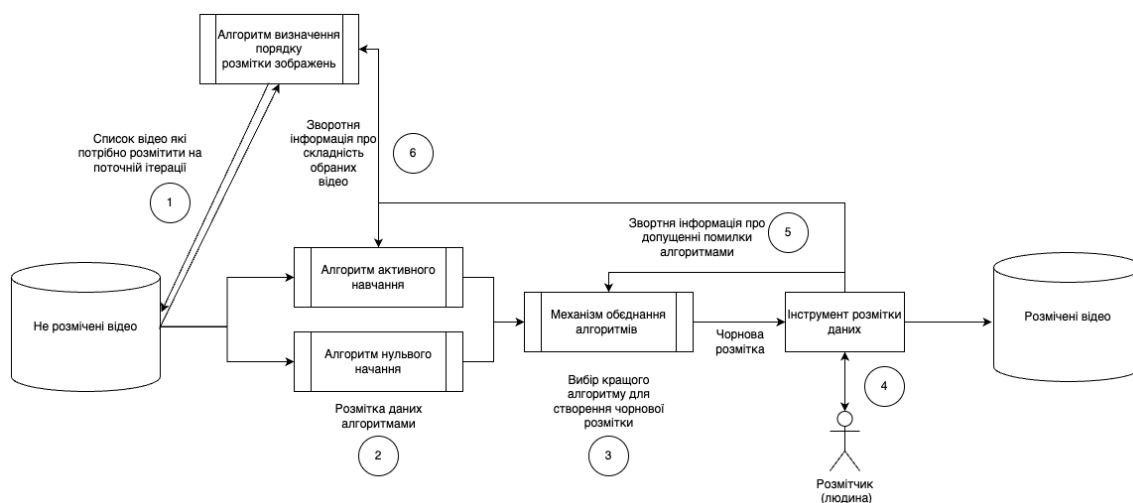


Рисунок 6. Принцип роботи дуальної архітектури автоматизації розмітки даних

Варто зазначити, що кроки 3, 5 та 6 у цьому процесі є критичними, оскільки система постійно поліпшується за рахунок зворотного зв'язку на кроці 5, а отже, ті зображення, які були раніше складними, могли стати простими для алгоритму детекції об'єктів і навпаки. Тому потрібно адаптувати механізм пріоритезації процесу розмітки даних після кожної частини розмічених даних, щоб він залишався актуальним.

Опис умов проведення експериментів

Для проведення детальних експериментів з розмітки даних та оцінки ефективності різних підходів до автоматизації, була сформована група з 10

досвідчених розмітчиків даних. Ці учасники виконували розмітку 10 відібраних відео, підданих різним умовам автоматизації. З метою забезпечення об'єктивності та точності оцінок, для кожного учасника експерименту було обчислено час, витрачений на розмітку. Результати кожного експерименту визначалися за допомогою аналізу часу розмітки, обчисленого на основі даних всіх учасників та всіх відео.

Порівняння різних методів до автоматизації розмітки даних для задач комп'ютерного зору

Порівняльний аналіз усіх методів розмітки даних для задач комп'ютерного зору представлено на рисунку 7.

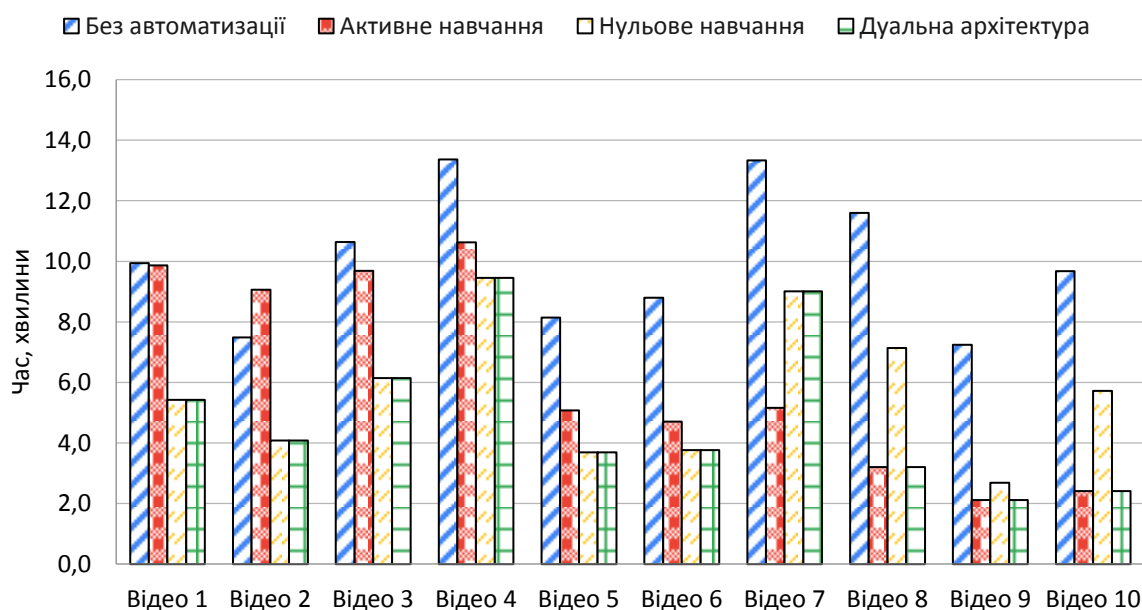


Рисунок 7. Порівняння медіанного часу розмітки кожного відео різними підходами

Результати експериментів свідчать, що медіанний час розмітки обраних відео за допомогою запропонованого дуального підходу склав 4,3 хвилини, на противагу до 9,7 хвилини методам без автоматизації, 5,8 хвилини методам з використанням алгоритмів активного навчання, 5,3 хвилини методам з використанням алгоритмів нульового навчання. Це дозволило скоротити час розмітки на 125% у порівнянні з мануальним підходом, на 25% у порівнянні з підходом нульового навчання та на 34% у порівнянні з підходом активного навчання. Виконаний статистичний аналіз показав, що впровадження автоматизації суттєво підвищує продуктивність, особливо при використанні дуального підходу, який продемонстрував найнижчий середній час

виконання завдань та статистично значущу перевагу над іншими методами автоматизації.

З наведених результатів можна зробити висновок, що запропонований підхід поводить як алгоритми з нульовим навчанням на ранніх етапах розмітки. Однак, з накопиченням більшої кількості даних, доступних для алгоритмів активного навчання, він починає функціонувати як алгоритм активного навчання, що дозволяє ще більше скоротити час розмітки. Таким чином, запропонований підхід поєднує переваги обох методів, усуваючи їхні недоліки.

Висновки

У даній роботі запропонована дуальна архітектура для автоматизації розмітки відеоданих продемонструвала свою ефективність, значно скоротивши час розмітки у порівнянні з традиційними підходами. Поєднання методів з нульовим навчанням та активного навчання дозволило отримати суттєві переваги на різних етапах процесу розмітки. Виконані експерименти показали, що медіанний час розмітки скоротився на 125% у порівнянні з ручною розміткою, на 25% у порівнянні з підходом нульового навчання та на 34% у порівнянні з активним навчанням.

Таким чином, дуальна архітектура забезпечує значне підвищення продуктивності та точності процесу розмітки, що є важливим кроком у напрямку автоматизації робочих процесів у сфері комп'ютерного зору. Подальші дослідження можуть бути спрямовані на оптимізацію алгоритмів та їх адаптацію до різних типів відеоданих, що ще більше покращить ефективність запропонованого підходу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Caruana R., Niculescu-Mizil A.* (2006). An Empirical Comparison of Supervised Learning Algorithms. Proceedings of the 23rd international conference on Machine learning (ICML '06), pp. 161-168.
2. *Hossain, M. S., Muhammad, G., Hasan, M. M.* (2020). The Challenges of Data Annotation for Machine Learning: A Review. CoRR, vol. abs/2004.03705, <https://arxiv.org/abs/2004.03705> (2020), last accessed 2024/07/13.
3. Gartner. "Gartner Forecasts Worldwide Artificial Intelligence Software Market to Reach \$62 Billion in 2022", 2021. Available at: <https://www.gartner.com/en/newsroom/press-releases/2021-11-22-gartner-forecasts-worldwide-artificial-intelligence-software-market-to-reach-62-billion-in-2022>, last accessed 2024/07/13.
4. *Зарічковий, О. А.* Алгоритмічне забезпечення для розмітки надвеликих об'ємів даних для задачі детекції об'єктів методами комп'ютерного зору : магістерська дис. : 121 Інженерія програмного забезпечення / Зарічковий Олександр

Анатолійович. - Київ, 2021. - 119 с. Available at: <https://ela.kpi.ua/items/d34f60b3-9832-40dd-a12f-035e4b71dd6>, last accessed 2024/07/13.

5. Kirillov A., Mintun E., Ravi N., Mao H., et al. (2023). Segment Anything. 3992-4003. 10.1109/ICCV51070.2023.00371.

6. Medium. "Segment Anything" - An Overview. 2023. Available at: <https://medium.com/@ghadi.alhajj/segment-anything-model-an-overview-118905735135>, last accessed 2024/07/13.

7. arXiv. "Y. Wang, K. Li, Y. Li, et al. InternVideo: General Video Foundation Models via Generative and Discriminative Learning". Available at: <https://arxiv.org/abs/2212.03191>, last accessed 2024/07/13.