

## **РЕЦЕНЗІЯ**

на дисертаційну роботу

Зарічкового Олександра Анатолійовича

на тему «Методи та програмне забезпечення розмітки відеоданих для задач

комп'ютерного зору»,

представлену на здобуття ступеня доктора філософії

в галузі знань 12 Інформаційні технології

за спеціальністю 121 Інженерія програмного забезпечення

### **Актуальність теми дисертації.**

Тема дисертації «Методи та програмне забезпечення розмітки відеоданих для задач комп'ютерного зору» є надзвичайно актуальною в умовах стрімкого зростання ролі штучного інтелекту у вирішенні складних завдань у багатьох галузях та особливо важлива у сфері збільшення обороноздатності України. В той же час застосування технологій машинного навчання значною мірою залежить від якості даних, які використовуються для навчання моделей. Особливо важливим фактором це стає для задач детекції та класифікації об'єктів у відеоданих, де якісна розмітка є ключовою умовою створення точних і надійних алгоритмів.

Актуальним є низка проблем, пов'язаних із розміткою даних, зокрема тривалість, висока вартість процесу та низька точність, що обмежують можливості для масштабування програмних рішень. Таким чином, актуальність тематики дисертаційного дослідження спрямована на подолання цих викликів шляхом розробки інноваційних методів, зокрема дуальної архітектури автоматизації розмітки відеоданих, що дозволяє знизити витрати на етапах обробки даних і підвищити ефективність роботи систем.

**Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.**



Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1. Вперше запропоновано дуальну архітектуру програмного забезпечення для автоматизованої розмітки даних, яка, за рахунок методу адаптивно-агрегованого навчання нейромережі, забезпечує пришвидшення процесу розмітки та, на відміну від існуючих аналогів, дає змогу ефективного застосування нульового та активного навчання нейромережі для розмітки даних та більш гнучкого використання програмного забезпечення для різноманітних задач комп'ютерного зору;
2. Вперше запропоновано метод пріоритезації складних зразків для навчання нейронної мережі, який, за рахунок відбору найскладніших зразків для навчання, підвищує якість набору даних без проведення попередньої розмітки відео, внаслідок чого збільшується точність детекції об'єктів на відео, та, на відміну від існуючих підходів, базується виключно на автоматично згенерованій репрезентації даних;
3. Вперше запропоновано ітеративний метод вибору ключових кадрів на довгих відео, що дає змогу визначати ключові кадри та сегменти відео з поступовим підвищенням точності, та, на відміну від існуючих методів, враховувати динамічно зміни контенту відео для вибору ключових кадрів, підвищуючи точність сегментації та зменшуючи обсяг відеоданих для обробки;
4. Вперше запропоновано метод агрегації знань між текстовою та візуальною частинами у візуально-мовній моделі (VLM) для обробки складних мультимодальних взаємодій, що забезпечує більш високу точність розпізнавання складних сцен на відео та їх опису у порівнянні з існуючими аналогами.

Наукові дослідження були виконані здобувачем на кафедрі інформатики та програмної інженерії КПІ ім. Ігоря Сікорського в рамках НДР «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних» (державний реєстраційний номер 0117U000924) та відповідає напряму



«Інформаційні та комунікаційні технології» з переліку пріоритетних напрямів науково-технічної діяльності.

Отже, в дисертаційній роботі поставлене наукове завдання підвищення швидкості розмітки відеоданих у контексті завдань детекції об'єктів виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

### **Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.**

За своїм змістом дисертаційна робота здобувача Зарічкового О. А. повністю відповідає Стандарту вищої освіти зі спеціальності 121 Інженерія програмного забезпечення та напрямкам досліджень відповідно до освітньої програми «Інженерія програмного забезпечення».

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям інженерії програмного забезпечення.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Зарічкового Олександра Анатолійовича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Збіги в тексті обмежуються співпадінням з власними публікаціями здобувача, що були створені з метою попереднього висвітлення наукових результатів дисертації. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

### **Мова та стиль викладення результатів.**

Дисертаційна робота написана українською мовою, зі вступу, складається з шести розділів, висновків, списку літератури та додатків. Загальний обсяг дисертації 261 сторінок.

Текст дисертації викладений логічно та послідовно, з використанням достатньої кількості сторонніх джерел та ресурсів, які належать до теми дослідження. Всі розділи містять відповідний матеріал.



У вступі детально наведено актуальність роботи в різних напрямках діяльності та важливість розмітки відеоданих для задач комп'ютерного зору, розкрито наукову новизну, зв'язок роботи з науковими програмами та планами, мету, завдання, об'єкт, предмет і методи дисертаційного дослідження.

Перший розділ містить огляд сучасного стану методів комп'ютерного зору та огляд методів для вирішення задачі детекції об'єктів, аналіз сучасних методів розмітки, включаючи YOLO і R-CNN, і основи програмного забезпечення для задач детекції об'єктів. Виконано постановку завдань дисертаційного дослідження.

Другий розділ присвячено розробці методу пріоритезації складних зразків, обробці складних негативних зразків, експериментальному дослідженню розробленого методу. Завдяки впровадженню запропонованого підходу в процес навчання для нейромереж на базі архітектури Video Swin Transformer отримано підвищення точності детекції без попередньої розмітки.

Третій розділ описує розробку та дослідження ефективності ітеративного методу вибору ключових кадрів для сегментації довгих відео. Запропонований метод зменшує кількість нерелевантних кадрів, які визначені на фазі попереднього відбору. Експериментальні дослідження методу показали підвищення якості узагальнення змісту відео на наборах даних SumMe та TVSum.

Четвертий розділ присвячений використанню мультимодальних моделей нейронних мереж для задач розпізнавання на відео. Представлений підхід Attr4Vis націлений на вдосконалення передачі знань між візуальними та текстовими модальностями візуально-мовних (VLM) нейронних мереж для підвищення точності розпізнавання на відео.

П'ятий розділ присвячений розробці дуальної архітектури програмного забезпечення, розробці вимог до програмного забезпечення та автоматизації процесу розмітки на основі розробленої архітектури.

Шостий розділ присвячений аналізу ефективності розробленої архітектури, узагальненню експериментальних досліджень і порівнянню підходів до автоматизації розмітки відеоданих.



Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

### **Оприлюднення результатів дисертаційної роботи.**

Наукові результати дисертації висвітлені у 6 наукових публікаціях здобувача, серед яких: 1 стаття у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України; 4 статті у періодичних наукових виданнях, проіндексованих у базах даних Web of Science Core Collection та/або Scopus.

Також результати дисертації були апробовані на 4 наукових фахових конференціях.

Таким чином, наукові результати описані в дисертаційній роботі повністю висвітлені у наукових публікаціях здобувача.

### **Недоліки та зауваження до дисертаційної роботи.**

#### **1. Зауваження до оформлення роботи:**

- В тексті дисертації застосовано досить багато англіцизмів.
- У п'ятому розділі на рисунку 5.14 наведено схематичне представлення процесу роботи інструменту розмітки з використанням дуальної архітектури. Водночас таке представлення сприймається складно та вимагає паралельного звернення до текстового опису. Для такого типу схем варто використати UML діаграми активності чи бізнес процесів.
- На рисунках 5.9 – 5.11 зображено підходи до розмітки даних. Водночас не зрозуміло, в якій нотації виконані дані діаграми.

#### **2. Висновки щодо ефективності запропонованих методів базуються на результатах, отриманих з використанням надвеликих масивів даних. Водночас не досліджено роботу методів на малих наборах відеоданих, які нерідко виникають у реальних умовах. Відсутність аналізу цього аспекту може обмежувати практичне застосування запропонованих методів.**



3. У роботі не проведено аналіз швидкодії запропонованих методів при використанні різних обчислювальних ресурсів, таких як: оперативна пам'ять, центральні та графічні процесори. Такий аналіз є важливим для оцінки ресурсів, необхідних для реалізації запропонованих методів, водночас його відсутність може ускладнити застосування розробок у високонавантажених середовищах і знизити їхню практичну значимість.
4. У шостому розділі дисертаційної роботи не враховано особливості фіксації відеоданих при різних часових та погодних умовах, таких як: денні або нічні записи, туман, засніженість, дощ тощо. Зважаючи на ці особливості, відеодані можуть значно відрізнятися за своїми характеристиками, тому ігнорування цього аспекту може призводити до потенційної похибки у висновках дослідження або обмеження їх застосовності.
5. У підрозділі 6.4 наведено експерименти з різними підходами до автоматизації розмітки даних. Водночас даний підрозділ не містить опису програмно-технічного комплексу, на якому проводились експерименти, що важливо для заміру різних часових показників.
6. У роботі автором не розглянуто можливість інтеграції запропонованого рішення із хмарними платформами, що є важливим фактором для забезпечення масштабованості, швидкодії та ефективної обробки великих обсягів відеоданих. Така інтеграція могла б значно підвищити практичну цінність розробки, дозволяючи застосовувати їх у сучасних високонавантажених обчислювальних середовищах.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

#### **Висновок про дисертаційну роботу.**

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Зарічкового Олександра Анатолійовича на тему «Методи та програмне забезпечення розмітки відеоданих для задач комп'ютерного зору» виконана на



високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для галузі знань 12 Інформаційні технології. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п.6 – 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Зарічковий Олександр Анатолійович заслуговує на присудження ступеня доктора філософії в галузі 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

**Рецензент:**

доцент кафедри інформатики та  
програмної інженерії

Національного технічного університету України

«Київський політехнічний інститут

імені Ігоря Сікорського»,

кандидат технічних наук



« 3 » березня 2025 року

