

ВІДГУК

Офіційного опонента на дисертаційну роботу

Здора Костянтина Андрійовича

На тему: «Моделі та програмні засоби підвищення швидкодії визначення відеоатрибутів за допомогою розбиття на сцени»
представлену на здобуття наукового ступеня доктора філософії
в галузі знань 12 – Інформаційні технології
за спеціальністю 121 – Інженерія програмного забезпечення

Актуальність теми дисертації.

Сучасний інформаційний простір характеризується невідомим зростанням обсягів відеоконтенту, що створює гостру необхідність у розробці та вдосконаленні ефективних інструментів для його автоматизованого аналізу. Задачі структурування відеоматеріалів, зокрема шляхом їх розбиття на семантичні одиниці (сцени та плани) та подальшого визначення атрибутів, є критично важливими для ефективної індексації, швидкого пошуку та узагальнення відеоданих. Класичні підходи до аналізу відео демонструють обмежену ефективність, особливо при обробці динамічних відеопотоків зі складними переходами між сценами. Застосування передових методів машинного навчання, зокрема архітектур глибоких нейронних мереж та візуальних трансформерів, відкриває значні перспективи для якісного покращення аналізу відео. Однак, практичне впровадження цих моделей ускладнюється їхньою високою обчислювальною складністю, що є суттєвою перешкодою для використання в системах реального часу та на платформах з обмеженими апаратними ресурсами. У зв'язку з цим, науково-технічне завдання, спрямоване на розробку моделей та програмних засобів для підвищення швидкодії процесів визначення відеоатрибутів шляхом оптимізації розбиття відео на сцени та самих

моделей, є беззаперечно актуальним і має важливе наукове та практичне значення.

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності і новизни.

Ключовим науковим досягненням представленої роботи є розробка здобувачем комплексу моделей та програмних інструментів, призначених для підвищення показників точності та швидкодії процесів сегментації відео на сцени та подальшого визначення їх атрибутів.

Вперше розроблено архітектуру розподіленого програмного забезпечення для визначення атрибутів на відео, характерною особливістю якої є оперування відеопотоками для їхнього розбиття відео на плани та сцени, що дозволило збільшити швидкість аналізу відеоконтенту мінімум в 2.5-3 рази.

Вперше розроблено метод для виявлення переходів планів у відеоконтенті на основі поєднання математичних підходів та рекурентних нейронних мереж, який на відміну від існуючих методів швидко та ефективно виділяє просторові та часові ознаки кадрів, що дозволило збільшити точність влучання та F1-оцінку для знаходження зміни планів досягаючи інноваційних результатів.

Вперше розроблено метод виявлення зміни сцени для відеоконтенту з використанням нейронної мережі на основі архітектури візуального трансформера для відео з застосуванням методу прунінгу перед навчанням, що на відміну від існуючих методів виділяє контекстуальні особливості сцен, що дозволило збільшити F1-оцінку на 5.1% та пришвидшити час виконання на 10%.

Набув подальшого розвитку метод прунінгу перед навчанням для моделей архітектури візуальних трансформерів для відео, який на відміну від існуючих методів враховує важливість механізму «уваги» та дозволяє пришвидшити час виконання моделі на 10%.

Обґрунтованість та достовірність наукових результатів і висновків, представлених у дисертації, підтверджуються коректним застосуванням сучасних наукових методів дослідження, використанням адекватного математичного апарату та результатами проведених експериментальних досліджень. Отримані результати узгоджуються з актуальними теоретичними концепціями у сферах машинного навчання та комп'ютерного зору, зокрема аналізу відеоданих. Таким чином, наукове завдання, поставлене в дисертаційній роботі, вирішене здобувачем у повному обсязі, що свідчить про його високий рівень володіння методологією наукового пошуку.

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

Зміст дисертаційної роботи Здора К.А. повністю відповідає вимогам Стандарту вищої освіти за спеціальністю 121 – «Інженерія програмного забезпечення» та узгоджується з напрямками наукових досліджень освітньої програми «Інженерія програмного забезпечення». Робота є завершеним науковим дослідженням, яке демонструє особистий внесок автора у розвиток методів аналізу відеоконтенту та підвищення їх ефективності.

Аналіз дисертації свідчить про те, що вона є самостійною науковою працею здобувача. Усі запозичені ідеї, результати досліджень та текстові фрагменти інших авторів супроводжуються відповідними посиланнями на джерела, що підтверджує дотримання принципів академічної доброчесності. Ознак плагіату, фабрикації чи фальсифікації даних не виявлено.

Мова та стиль викладення результатів.

Дисертаційна праця написана державною українською мовою, характеризується чіткою структурою та логічною послідовністю викладу матеріалу. Оформлення роботи відповідає встановленим стандартам, що сприяє належному сприйняттю ключових результатів дослідження.

Робота складається зі вступу, чотирьох логічно пов'язаних розділів, загальних висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 178 сторінок, у тому числі 154 сторінки основного тексту, 3 додатки на 8 сторінках. Робота містить 24 рисунки, 34 формули та 9 таблиць, які наочно ілюструють теоретичні положення та результати експериментів.

У вступі аргументовано актуальність обраної теми, розкрито зв'язок роботи з науковими програмами та планами. Чітко сформульовано мету й завдання дослідження, визначено його об'єкт та предмет, а також висвітлено наукову новизну та практичну цінність отриманих результатів. Наведено інформацію про особистий внесок дисертанта та апробацію результатів на наукових форумах.

Перший розділ присвячено глибокому аналізу сучасного стану проблеми визначення відеоатрибутів, включаючи огляд традиційних та нейромережових підходів (зокрема, візуальних трансформерів) до аналізу просторово-часових особливостей відео. Розглянуто ключові методи оптимізації моделей, такі як прунінг, квантування та дистиляція знань, з особливим акцентом на прунінгу на етапі навчання. За результатами аналізу виявлено проблемні аспекти існуючих рішень, зокрема недостатню точність детекції планів у сучасному відеоконтенті, обмежені можливості врахування контексту сцен та відсутність ефективних методів прунінгу відеотрансформерів перед навчанням.

Другий розділ деталізує запропоновані оптимальні умови для аналізу відеосцен, зокрема використання чотирьох планів з одним ключовим кадром для ефективного визначення переходів та контекстуальної інформації. Охарактеризовано виклики, що виникають при обробці великих обсягів даних, такі як збільшення розміру моделей та ризик перенавчання, а також специфіку роботи з короткими сценами (наприклад, у рекламних роликах), та запропоновано можливі шляхи їх подолання. Експериментально підтверджено, що розроблений підхід до сегментації відео на плани та сцени

дозволяє суттєво зменшити кількість операцій для визначення атрибутів, зберігаючи при цьому високу точність. Обґрунтовано перспективи подальшого вдосконалення запропонованих рішень, зокрема шляхом застосування методу прунінгу візуальних трансформерів перед навчанням.

Третій розділ містить опис вдосконаленого методу прунінгу, який продемонстрував високу ефективність при оптимізації моделей візуальних трансформерів для обробки відео, забезпечуючи значне скорочення обчислювальних ресурсів при збереженні рівня точності. Викладено результати дослідження різних стратегій прунінгу для трансформерних архітектур, що дозволило розробити модифікований метод. Цей метод включає покращену процедуру визначення важливості вагових коефіцієнтів у методі SNIP за рахунок урахування активацій механізму уваги та застосування розрідженості матриць за шаблоном 2:4. Представлено результати експериментів, які свідчать про прискорення роботи моделі приблизно на 10% при використанні стандартного апаратного забезпечення, з перспективою значно більшого прискорення на спеціалізованих графічних процесорах.

Четвертий розділ присвячено опису розробленої архітектури програмного комплексу, який ефективно використовує запропоновані методи розбиття відео на сцени та плани для істотного підвищення швидкодії процесу визначення відеоатрибутів. Представлено гнучкі архітектурні рішення, що забезпечують можливість експлуатації системи як на окремих обчислювальних вузлах, так і в розподілених хмарних інфраструктурах, з використанням технології Docker для спрощення інтеграції нових аналітичних модулів.

Дисертаційна робота оформлена з дотриманням 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 наукових публікаціях здобувача, серед яких: 4 статті у фахових наукових виданнях України. Із них одна наукова стаття опублікована у періодичному видання, що входить до наукометричної бази Scopus та відноситься до першого — третього квартилів (Q1—Q3) відповідно до класифікації SCImago Journal and Country Rank або Journal Citation Reports.

Також результати дисертації були апробовані на 6 наукових фахових конференціях.

Усі публікації здобувача мають належний науковий рівень, в них достатньо повно описані головні наукові здобутки, представлені в дисертації. Особистий внесок здобувача до всіх наукових публікацій, опублікованих у співавторстві, є вагомим.

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

Недоліки та зауваження до дисертаційної роботи.

1. У першому розділі, розглядаючи архітектури моделей для роботи з відео, зокрема гібридні архітектури, було б корисним включити більш глибокий порівняльний аналіз їхньої масштабованості та вимог до обсягів навчальних даних.

2. У першому розділі автор визнає суб'єктивність оцінки сцен (наприклад, анотації BBC Planet Earth), але не пропонує спосіб або метод для формалізації цього поняття чи розробки уніфікованої метрики.

3. У другому розділі при описі методу розбиття відео на сцени за допомогою візуального трансформера для відео, зокрема при виборі ключових кадрів з планів, доцільно було б детальніше обґрунтувати вибір саме чотирьох планів як оптимальної кількості для аналізу та дослідити вплив зміни цієї кількості на кінцеву точність та обчислювальну складність.

4. У другому розділі дисертації не приділено належної уваги тому, як запропоновані методи працюють в умовах артефактів відео, низької якості

зображення, складного освітлення чи аномалій у потоці. Це особливо критично для практичного використання в системах відеоспостереження.

5. У третьому розділі згадується значний потенціал прискорення оптимізованих моделей на спеціалізованому апаратному забезпеченні (NVIDIA A100, H100), було б корисним додати хоча б теоретичну оцінку або результати симуляції очікуваного приросту продуктивності для кращої демонстрації переваг запропонованої напівструктурованої розрідженості.

6. У третьому розділі прунінг описано на рівні логіки алгоритму, однак не подано формалізованої постановки задачі, наприклад у вигляді задачі мінімізації втрат при фіксованих ресурсних обмеженнях.

7. У четвертому розділі при описі архітектури розподіленого програмного забезпечення, було б доцільно детальніше зупинитися на механізмах забезпечення консистентності даних та управління станами між взаємодіючими модулями, особливо в умовах інтенсивної паралельної обробки декількох відеопотоків.

8. Порівняння продуктивності моделей відбувається переважно на власних реалізаціях. Тільки AutoShot використано як базовий приклад, але немає повноцінного тестування на відкритих еталонних наборах (ClipShots, RAI Dataset, SumMe, TVSum тощо).

Висловлені зауваження не носять принципового характеру, не знижують загальної наукової цінності та практичної значущості отриманих результатів і не впливають на загальну позитивну оцінку дисертаційної роботи.

Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Здора Костянтина Андрійовича на тему «Моделі та програмні засоби підвищення швидкодії визначення відеоатрибутів за допомогою розбиття на сцени» виконана на високому науковому та методологічному рівні, відповідає принципам академічної доброчесності та є завершеним науковим

дослідженням. Сукупність представлених теоретичних положень та практичних результатів успішно вирішує важливе наукове завдання, що має істотне значення для розвитку галузі знань 12 – Інформаційні технології.

Дисертаційна робота за своєю актуальністю, практичною цінністю та науковою новизною повною мірою відповідає вимогам чинного законодавства України, зокрема 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 Здор Костянтин Андрійович заслуговує на присудження наукового ступеня доктора філософії в галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення.

Офіційний опонент:

Завідувач кафедри інформаційних
технологій та кібербезпеки Приватної
установи "Університет науки,
підприємництва та технологій"
доктор технічних наук, професор

Сергій СЕМЕНОВ

М.П.



«17» 06 2025 року