

РЕЦЕНЗІЯ

на дисертаційну роботу

Мельниченко Артема Васильовича

на тему «Методи та програмні засоби підвищення швидкодії моделей розпізнавання образів на основі машинного навчання»,
представлену на здобуття ступеня доктора філософії
в галузі знань 12 - Інформаційні технології
за спеціальністю 121 - Інженерія програмного забезпечення

Актуальність теми дисертації.

Актуальність досліджень направлених на розробку та удосконалення методів підвищення швидкодії моделей розпізнавання образів на основі машинного інтелекту обумовлена широким спектром застосування таких досліджень. Розпізнавання образів є ключовим етапом процесу ідентифікації об'єкту, що в свою чергу є однією з проблем, що виникає під час створення математичної моделі динамічного процесу, також при створенні комп'ютерних моделей та обробки цифрових даних.

За останні декілька років необхідність в дослідженнях спрямованих на розпізнавання образів стала однією з проблем, що потребують швидкого та результативного рішення. Це пов'язано зі зміною тактики ведення бою, ключовим моментом якої стало застосування безпілотних літальних апаратів(БПЛА). Використання безпілотних літальних апаратів в розвідувальних цілях потребує подальшої обробки отриманих даних та ідентифікації на них ворожих об'єктів. Ці графічні дані є об'ємними, мають потоковий характер і потребують швидкої обробки та класифікації. Від швидкості розпізнання образу на даних отриманих з БПЛА часто залежать результативність військових дій, і як наслідок - бойові втрати. Розроблене програмне забезпечення для військової апаратури спрямоване на автоматизований пошук та ураження цілей, потребує удосконалення алгоритмів пошуку цілі. Швидкодія цих алгоритмів та точність відображення цілі на екрані перебуває в прямій залежності з ймовірністю влучання. Одним з можливих підходів до вирішення даної проблеми є застосування штучного інтелекту. Проте, нейронні мережі мають велику обчислювальну складність та потребують великих затрат електроенергії. І, як наслідок, постає задача оптимізації нейронних мереж з метою зменшення точності та збільшення швидкодії методів.

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

Удосконалено модель нейронної мережі для виявлення облич RetinaFace, яка на відміну від існуючих використовує метод прунінгу SNIP для оптимізації, що дозволяє використовувати розріджені матриці для зберігання і виконання мережі з метою подальшого удосконалення та збільшення швидкодії.

Удосконалено метод прунінгу SNIP для моделі виявлення облич RetinaFace, який на відміну від існуючих передбачає можливість виключення контекстних модулів з процесу прунінгу. Вдосконалений метод дозволяє досягти більшої точності при незмінній кількості виключених параметрів.

Вперше розроблено метод прунінгу перед навчанням для моделей архітектури трансформер, який на відміну від існуючих враховує важливість механізму «уваги». Використання розробленого методу дозволяє значно збільшити точність класифікації кінцевої моделі в порівнянні з методом SNIP.

Вперше розроблено архітектуру програмного забезпечення для моделювання та дослідження методів прунінгу перед навчанням нейронних мереж, яка на відміну від існуючих дозволяє приводити матриці вагових коефіцієнтів мережі до розрідженого формату, використовуючи запропонований механізм оцінки важливості вагів.

Оптимізована мережа RetinaFace містить на 68% параметрів менше ніж початкова мережа при втраті точності на лише 1.4%. Вдосконалений метод дозволив зменшити втрати точності з 1.4% до 0.7% порівняно з методом SNIP при порівнянні з необрізаною моделлю, при скороченні параметрів на 68%.

Реалізація методу прунінгу для архітектури трансформер дозволила натренувати мережу з покращенням точності до 37% порівняно з методом SNIP при порівнянні з необрізаною моделлю, при скороченні кількості параметрів на 90%.

Встановлено, що результати визначення критеріїв важливості вагів, отриманих розробленим алгоритмом, можуть бути використані для підвищення швидкодії нейронних мереж від 20% до 65% шляхом використання розріджених матриць формату 2:4, в залежності від графічного процесора.

Встановлено, що додаткові виходи для сіамських нейронних мереж, призначених для встановлення схожості двох зображень, не дають приросту в швидкості сходження і точності моделі.

Викладені в дисертаційній роботі наукові положення, висновки та рекомендації є достатньо обґрунтованими і достовірними, так як вони базуються

на загальнонаукових, фундаментальних положеннях сучасної науки, результатах наукових досліджень вітчизняних та закордонних вчених з інформаційних технологій. Інформаційною базою роботи стали 117 літературних джерел, які включають праці провідних науковців з досліджуваної проблематики.

Вірогідність та обґрунтованість результатів дисертаційного дослідження підтверджується зіставленням результатів тестових прикладів із близькими за змістом, а також впровадженнями в навчальний процес. Основні положення роботи опубліковані в наукових фахових виданнях України, де здійснено рецензування коректності результатів з позицій інженерії програмного забезпечення. Основні положення дисертації проілюстровано засобами комп'ютерних анімацій.

Наукові дослідження були виконані здобувачем на кафедрі інженерії програмного забезпечення в енергетиці КПІ ім. Ігоря Сікорського в рамках науково-дослідницької роботи «Методи і алгоритми оптимізації розпізнавання образів на основі методів машинного навчання» (Державний реєстраційний номер №0121U109207, м. Київ) під керівництвом доцента, кандидата технічних наук Шалденко Олексія Вікторовича та Недашківського Олексія Леонідовича

Отже, в дисертаційній роботі поставлене наукове завдання збільшення ефективності моделей нейронних мереж, а саме зменшення втрати точності при збільшенні швидкодії, після застосування методів оптимізації моделей глибинного навчання, створених для вирішення задач комп'ютерного зору 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 Мельниченка Артема Васильовича 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 оптимізації нейронних мереж.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Мельниченка Артема Васильовича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

Мова та стиль викладення результатів.

Дисертаційна робота написана українською мовою.

Дисертація являє собою завершену наукову працю, виконану на високому науковому рівні. Дисертацію написано грамотно, з використанням сучасної бібліографії та наукової термінології. Зміст та результати досліджень викладено лаконічно та аргументовано. Суть дисертації в логічній послідовності відбиває етапи дослідження – від аналізу літературних джерел до побудови моделей та методів розв'язання задачі, з наступним проведенням комп'ютерних експериментів з конкретними чисельними даними та впровадженні результатів дослідження. Дисертація містить нові науково обґрунтовані результати, одержані особисто автором, і являє собою завершене дослідження, виконане на високому науковому рівні.

Дисертація складається з вступу, 4 розділів, висновків, списку літератури та додатків. Загальний обсяг дисертації 179 сторінок.

У вступі зазначена актуальність теми дослідження, зв'язок автора з розробкою науково-дослідницької роботи «Методи і алгоритми оптимізації розпізнавання образів на основі методів машинного навчання» з зазначенням особистого внеску в дослідницьку роботу автора. Наведена мета та наукові завдання дослідження, об'єкт та предмет дослідження, методи дослідження, наукова та практична новизна, практичне значення отриманих результатів та апробація матеріалів роботи, публікації, структура та обсяг роботи.

У розділі 1 розглянуто застосування нейронних мереж для вирішення задач розпізнавання образів, а саме для сегментації та кластеризації зображень, виявлення та відстеження на зображенні об'єктів. Також в першому розділі наводяться задачі для удосконалення нейронних мереж на різних стадіях розробки.

У розділі 2 розглянуто математичні методи та підходи до вирішення задач виявлення образу та удосконалення методу Прунінгу за рахунок зменшення кількості параметрів нейронної мережі. Проведений аналіз впливу на розроблену математичну модель зовнішніх факторів.

Розділ 3 направлений на дослідження та забезпечення швидкодії методу Прунінгу для нейронної мережі встановленої архітектури. Розроблено архітектуру трансформу та протестована за допомогою задачі розпізнавання образу. Наведено обґрунтування вдосконалення архітектури, розроблено алгоритм вдосконалення методу прунінгу для моделей заданої архітектури,

підтверджено ефективність удосконаленого алгоритму шляхом порівняння експериментальних та вихідних даних.

Розділ 4 присвячено розробці програмного забезпечення для оптимізації нейронних мереж перед навчанням. Вказані засоби розробки, обґрунтовано вибір бібліотеки для реалізації нейронних мереж, забезпечено реалізацію знайденого алгоритму та архітектури програмного забезпечення перед навчанням. Приведено розробки по реалізації необхідних модулів для роботи програмного забезпечення

Висновки побудовані у відповідності до задач, що були поставлені для досягнення мети.

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 3 наукових публікаціях здобувача, серед яких: 3 статті у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України.

Також результати дисертації були апробовані на 7 наукових фахових конференціях.

Дисертаційна робота є самостійно виконаною науковою працею. Всі представлені наукові результати, приклади та експериментальні розрахунки, викладені у дисертації, одержані здобувачем одноосібно. У наукових роботах, що опубліковані у співавторстві, в дисертаційній роботі використані лише ті результати, які становлять індивідуальний внесок автора. У друкованих працях, опублікованих у співавторстві, здобувачеві належать:

Удосконалений метод прунінгу нейронних мереж для виявлення облич на зображеннях. Зазначений метод дозволяє виключити з оцінки важливі шари в мережі, при цьому зберігаючи кількість вагів, що виключаються незмінною. Завдяки цьому точність результуючої мережі збільшується на 0.7% на тестовій вибірці.

Виявлення падіння точності і неефективність оптимізації шляхом додавання додаткових виходів для вирішення задач класифікації і регресії в сіамських нейронних мережах.

Удосконалений метод прунінгу перед навчанням для застосування на моделях трансформів шляхом включення оцінок уваги в розрахунок критерію

важливості вагів. Дана модифікація дозволила натренувати мережу що має точність на 37% більше порівняно з оригінальним методом.

Вдосконалений метод прунінгу в поєднанні з використанням розріджених матриць може бути використаний для підвищення швидкодії при тренуванні і виконанні нейронних мереж від 25% до 65%.

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

Недоліки та зауваження до дисертаційної роботи.

1. Алгоритм підрахунку точності удосконаленого методу доцільно було б навести в роботі або надати посилання на джерело.

2. Графічна візуалізація точності оптимізації потребує більш ретельного оформлення згідно з вимог наказу МОН від 12.01.2017 № 40 "Про затвердження Вимог до оформлення дисертації".

3. В роботі наявні друкарські помилки. Наприклад, у вступі при формулюванні мети дослідження двічі підряд надруковане слово «зменшення» (ст.22).

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 Мельниченко Артема Васильовича на тему «Методи та програмні засоби підвищення швидкодії моделей розпізнавання образів на основі машинного навчання» 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 інформаційних технологій. 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 Мельниченко Артем Васильович заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

Рецензент:

Доцент кафедри цифрових технологій в енергетиці
Навчально-наукового інституту атомної та теплової енергетики
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»

Кандидат технічних наук





« 5 » червня 2024 року