# Інформаційні технології

УДК 7.061:343.533(084.14) **doi: 10.26906/SUNZ.2025.3.55-58**

Ю. М. Главчева, М. І. Главчев

Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

## ОСОБЛИВОСТІ ВИЯВЛЕННЯ АКАДЕМІЧНОГО ПЛАГІАТУ ЗОБРАЖЕНЬ

**Анотація.** Стрімке зростання обсягів візуального контенту в академічних документах актуалізує проблему виявлення академічного плагіату зображень. На відміну від текстового плагіату, який ефективно виявляється сучасними системами, візуальний плагіат залишається складним завданням через різноманіття форм його прояву та можливість маніпуляцій. **Метою** даної роботи є аналіз існуючих підходів до розпізнавання та порівняння зображень для виявлення академічного плагіату та виявлення особливостей. **Розглянуто** алгоритми комп’ютерного зору, моделі глибокого навчання, інших технологій для аналізу текстових елементів на зображеннях. **Запропоновано класифікацію** форм візуального плагіату, включаючи точне копіювання, афінні трансформації, семантичне дублювання та генерацію зображень. **Визначено основні проблеми**: відсутність глобальної бази зображень, складність вилучення нетекстового контенту, правові обмеження доступу до повних текстів. **Висновки**: аналіз зображень для виявлення академічного плагіату лишається актуальним та є перспективним напрямом досліджень, який може реалізовуватися за напрямами: удосконалення алгоритмів виявлення модифікацій та інтеграція візуального аналізу в системи перевірки текстового плагіату, створення репозиторіїв зображень, доступність інформаційних інструментів для експертів та науковців.

**Ключові слова:** академічна доброчесність, академічний плагіат зображень, візуальний контент, маніпуляції з зображеннями, академічні документи.

### Вступ

Академічна доброчесність є фундаментом освітньої та наукової діяльності. Це питання не втрачає актуальності через появу нових викликів. Найвпливовішим є легка доступність до використання користувачами систем на основі штучного інтелекту (ШІ). ШІ може бути застосований для генерації та обробки текстів, зображень, відеоматеріалів. Стрімко змінюється та розширюється ландшафт академічного плагіату. Хоча текстовий плагіат активно виявляється за допомогою спеціалізованих систем, візуальний академічний плагіат зображень (графіки, схеми, діаграми та інше) є окремим напрямом для дослідження [1–3]. Сучасні програми для виявлення плагіату (StrikePlagiarism.com, Turnitin, Grammarly, тощо) ефективно виявляють ознаки академічного плагіату у тексті, проте це не відноситься до роботи з зображеннями. Разом з цим, відбувається зростання кількості мультимедійного контенту в кваліфікаційних та наукових роботах.

Отже, виникає потреба у створенні підходів виявлення ознак академічного плагіату зображень. Це завдання має як наукову, так і практичну значущість, оскільки сприяє підвищенню рівня академічної доброчесності та допомагає науково-педагогічним працівникам та рецензентам при перевірці академічних робіт.

### Постановка проблеми

В наукових публікаціях описуються два підходи до виявлення ознак академічного плагіату:

– визначення абсолютної ідентичності [4];

– визначення ступеню подібності зображень [5–9].

На основі публікацій за темою обробки зображень визначимо вказані терміни.

Абсолютна ідентичність – повну відповідність двох зображень на рівні піксельної структури, геометричних параметрів, кольорової палітри та метаданих, що свідчить про їхню тотожність без жодних змін чи трансформацій. У контексті виявлення академічного плагіату абсолютна ідентичність означає, що одне зображення є прямою копією іншого, без редагування, обрізання, масштабування або застосування фільтрів. Автори [3] вважають, що абсолютна ідентичність має бути критерієм для виявлення академічного плагіату зображень.

Подібність **зображень** – це ступінь відповідності між двома зображеннями, що може бути визначений на основі їхніх візуальних, структурних або семантичних характеристик. Подібність може враховувати як точну відповідність пікселів, так і схожість форм, кольорів, текстур, просторових відношень або змісту, навіть у разі наявності трансформацій, таких як масштабування, обертання, обрізання чи зміна кольорової палітри. Низькорівневі методи порівняння базуються на піксельному порівнянні (MSE, Mean Squared Error – середньоквадратична похибка, SSIM, Structural Similarity Index Measure – структурна подібність). Високорівневі методи порівняння визначається за допомогою ознак (SIFT, Scale-Invariant Feature Transform – масштабоінваріантне ознакове перетворення, SURF, Speeded-Up Robust Features – прискорені стійкі ознаки) або нейронних мереж. Подібність не означає ідентичність: два зображення можуть бути подібними за змістом, але відрізнятися за виглядом. Абсолютна ідентичність є одним з випадків подібності та означає 100% подібність.

### Основна частина роботи

Згідно з дослідженнями Meuschke N. [3], візуальний плагіат може проявлятися у багатьох різноманітних формах: точне копіювання, обрізання або

масштабування версій, обертання, дзеркальність, а також повторного використання графіків із незначними змінами. Детальніше класифікацію за Meuschke N. [3] описана в табл. 1.

*Таблиця 1* – **Класифікація візуального плагіату за Meuschke N. [3]**

| № | Форма | Форма англ. | Особливості |
|---|---|---|---|
| 1. | Точне копіювання | Exact Copy | – Повна ідентичність зображення.<br>– Без змін у розмірі, кольорі, структурі.<br>– Виявляється за допомогою хешування або піксельного порівняння |
| 2. | Обрізане копіювання | Cropped Copy | – Частина оригінального зображення використана без змін.<br>– Часто застосовується для приховування джерела. |
| 3. | Афінне перетворення | Affine Transformations | – Зображення змінено шляхом масштабування, обертання, зсуву.<br>– Виявляється за допомогою методів ключових точок (SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features)). |
| 4. | Зміна якості | Quality Modification | – Зображення розмиті, змінено контраст, освітлення.<br>– Мета — ускладнити автоматичне виявлення. |
| 5. | Семантичне копіювання | Semantics-Preserving Plagiarism | – Зображення перероблене, але передає ту саму ідею або дані.<br>– Наприклад, графік з іншими кольорами, але з тими самими значеннями. |
| 6. | Копіювання ідеї | Idea-Preserving Plagiarism | – Візуалізація створено нову, але вона базується на чужій ідеї або концепції |

Для виявлення випадків повторного використання зображень у наукових роботах застосовуються різноманітні алгоритми комп'ютерного зору, які адаптовані до типів подібності, характерних для академічного контенту.

Алгоритм pHash (perceptual hashing) — це метод, який дозволяє порівнювати зображення не за точним збігом пікселів, а за візуальною схожістю, тобто тим, як зображення сприймається людським оком.

SIFT (Scale-Invariant Feature Transform) та ORB (Oriented FAST and Rotated BRIEF) виявляють стійкі ознаки на зображенні, які не змінюються при масштабуванні, обертанні чи зміні освітлення.

Моделі глибокого навчання здатні виявляти семантичну подібність, навіть якщо зображення було перероблене:

CNN (Convolutional Neural Networks) навчаються розпізнавати складні візуальні патерни;

Siamese Networks порівнюють пари зображень і визначають ступінь їх подібності на основі високорівневих ознак.

Існуючі алгоритми комп'ютерного зору можуть виконувати завдання подібності зображень з великою точністю. Особливого аналізу потребують діаграми, блок-схеми, графіки тощо. Вони можуть демонструвати значну ступінь подібності між собою та до вже опублікованого матеріалу, і не мати ознак академічного плагіату.

OCR (Optical Character Recognition) — це технологія, яка дозволяє розпізнавати текст на зображеннях та перетворювати його у машинозчитуваний формат. У контексті аналізу візуального контенту, автором Meuschke N. [3] описано, OCR використовується для:

– розпізнавання текстових міток на графіках, діаграмах, таблицях,

– зіставлення підписів, назв осей, легенд між різними зображеннями,

– виявлення повторного використання даних, навіть якщо візуальне оформлення змінено.

Завдання зі з'ясування, чи зображення раніше зустрічалося у статті, не є просто проблемою подібності зображень, оскільки рисунки та діаграми утворюють лише частину сторінки рукопису. Також криві (залежності) для різних змінних можуть мати високу ступінь подібності. Ця складність академічних діаграм ще більше ускладнює проблему виявлення плагіату зображень. Автори [4] вважають це однією з причин того, що для прийняття рішення про академічний плагіат зображень, необхідно встановити його абсолютну ідентичність з існуючим зображенням.

Науковці створюють авторські підходи до вирішення проблеми визначення академічного плагіату візуального контенту.

Алгоритм «VIBRANT-WALK» [4] складається з 2 етапів: попередня обробка зображень виконується шляхом створення «Матриці яскравості» діаграми/малюнка; здійснюється попіксельне порівняння зображення із зображеннями сторінок раніше опублікованих рукописів. Для оптимізації процесу порівняння, процес розпочинається з пікселя на зображенні, який має найвище значення у Матриці яскравості. Для реалізації алгоритму необхідно мати набір даних із зображеннями рисунків і діаграм, які будуть використані для порівняння.

Загальний алгоритм [5] складається з трьох етапів: попередньої обробки, вилучення ознак та порівняння. Точність становить 100% у випадку невідредагованих зображень та змінюється в інших операціях, таких як перевертання, поворот, відтінки сірого та обрізання.

Виявлення плагіату зображень є складною галуззю досліджень, оскільки можуть аналізуватися не лише текстова аналітика, а й графічні особливості. У статті [6] досліджується питання плагіату зображень через аналіз пошуку подібних семантичних значень між зображеннями шляхом застосування методів обробки зображень та семантичного картування.

Запропоновані у [7, 8] методи використовують розширене вилучення текстових ознак та методи обчислення подібності: подібність на основі текстових посилань на рисунки. На основі аналізу текстових посилань на зображення, метод дозволив класифіку-

вати даний рисунок як «академічний плагіат» або «ні» залежно від певного порогового значення. Результати [7] показали, що запропонована методика досягла результату точності = 0,78 та повноти = 0,67 з точки зору міри оцінки.

Сформулюємо основні проблеми у визначенні академічного плагіату зображень. На основі публікацій можемо виокремити особливості застосування підходів виявлення ознак академічного плагіату зображень, які впливають на ефективність застосування та на точність результатів:

– широкі можливості маніпуляцій з зображеннями;

– відсутність глобальної бази для порівняння зображень з академічних документів;

– подібні форми представлення візуального контенту в академічних документах (графіки, схеми, діаграми, тощо).

В табл. 2 представлено широкий перелік типів маніпуляцій з зображеннями. Кожен тип маніпуляції має власні ознаки, які слід враховувати при створенні технічних рішень для його виявлення.

*Таблиця 2* – **Типи маніпуляцій із зображеннями в академічному контексті**

| № | Тип маніпуляції | Опис | Джерело |
|---|---|---|---|
| 1. | Точне дублювання | Повторне використання зображення без змін | [9] |
| 2. | Дублювання з переміщенням | Копія фрагмента вставлена в інше місце | [10] |
| 3. | Дублювання з модифікацією | Копія змінена (розмиття, контраст, обертання) | [9] |
| 4. | Копіювання та вставка | Об’єкти дублюються в межах одного зображення | [11] |
| 5. | Видалення об’єктів | Видалення частин зображення з відновленням фону | [11] |
| 6. | Композиція зображень | Об’єднання елементів з різних джерел | [11] |
| 7. | Ретушування | Зміна кольору, яскравості, контрасту | [10] |
| 8. | AI-генерація / Deepfakes | Створення зображень за допомогою генеративних моделей | [11] |

Ключовим елементом успішного виявлення візуального академічного плагіату є наявність бази зображень, яка містить оригінальні графіки, діаграми, фотографії та інші візуальні матеріали з академічних документів. Ця база дозволяє здійснювати порівняння нових зображень із вже наявними, виявляючи повторне використання або модифікації. Вона також є має бути основою для навчання моделей глибокого навчання, що потребують великої кількості прикладів для досягнення високої точності.

Процес вилучення зображень для формування бази є окремим складним завданням. З моменту появи академічних видань було опубліковано мільйони статей та різних академічних текстів, отже, неможливо витягти окремі зображення з усіх цих документів. Дослідження з вилучення зображень з документів постійно розвиваються. Одним із таких прикладів є дослідження [12], яке мало 83% точності вилучення зображень.

Розділення «тексту» та «не тексту» є важливим кроком обробки для будь-якої системи аналізу документів. Тому важливо мати чітке розуміння сучасного стану розділення «тексту» та «не тексту», щоб сприяти розробці ефективних систем обробки документів. У статті [13] підсумовано технічні проблеми розділення тексту, питання класифікації зображень, висвітлено майбутні проблеми та напрямки досліджень у цій галузі.

Окремим напрямом для впровадження є створення репозиторію всіх сторінок статей, які були опубліковані. Однаково критично на це завдання впливають правовий та технологічний аспекти. Авторське право регулює доступ до повних текстів документів та «закриває» доступ до значної частини контенту. Навіть за наявності 5001 репозиторіїв в світі (за даними Registry of Open Access Repositories), це завдання ще далеке до вирішення: Африка (201); Азія (1118); Європа (1798); Північна Америка (1071); Океанія (102); Південна Америка (711). В Україні відповідно до вказаного реєстру зареєстровано 138 інституційних репозитаріїв.

В електронному репозитарії Національного технічного університету «Харківський політехнічний інститут» (eNTUKhPIIR, ISSN 2409-5982) міститься близько 80 тис. повних тексті академічних документів.

Без наявності бази зображень для порівнянння підходи до визначення академічного плагіату зображень не зможуть досягти своїх цілей.

## Висновки і перспективи подальших розвідок

Аналіз зображень для виявлення академічного плагіату, особливо в умовах доступності ШІ, лишається актуальним та є перспективним напрямом, що потребує подальших досліджень. Сучасні методи комп’ютерного зору демонструють високу точність, особливо при використанні глибоких нейронних мереж. Впровадження таких систем у освітні та наукові установи сприятиме підвищенню рівня академічної доброчесності. Подальші дослідження мають бути спрямовані на:

– створення підходів до визначення подібності візуального контенту;

– розвиток алгоритмів вилучення нетекстового контенту з академічних документів;

– формування глобальної бази зображень для проведення порівнянь; удосконалення алгоритмів виявлення модифікованих зображень для всіх типів маніпуляцій;

– інтеграцію процесів виявлення академічного плагіату візуального контенту в системи виявлення текстових запозичень (подібностей у тексті);

– включення спеціальних інструментів до наукових платформи та у редакційні процеси для допомоги експертам та науковцям.

СПИСОК ЛІТЕРАТУРИ

1. Eisa T. A. E. Plagiarism detection of figure images in scientific publications. International Journal of Data Mining, Modelling and Management. 2022. Vol. 14, no. 1. P. 15. DOI: https://doi.org/10.1504/ijdmmm.2022.10046101
2. Saliba, T., Rotzinger, D. Figure plagiarism and manipulation, an under-recognised problem in academia. Eur Radiol. 2025. Vol. 35, 4518–4521. DOI: https://doi.org/10.1007/s00330-025-11426-2
3. Meuschke N. Analyzing Non-Textual Content Elements to Detect Academic Plagiarism. Springer Fachmedien Wiesbaden GmbH, 2023.
4. Parmar S., Jain B. VIBRANT-WALK: An algorithm to detect plagiarism of figures in academic papers. Expert Systems with Applications. 2024. P. 124251. DOI: https://doi.org/10.1016/j.eswa.2024.124251
5. Ibrahin A. S. B., Khalifa O. O., Ahmed D. E. M. Plagiarism Detection of Images. 2020 IEEE Student Conference on Research and Development (SCOReD), Batu Pahat, Johor, Malaysia, 27–29 September 2020. 2020. DOI: https://doi.org/10.1109/scored50371.2020.9250940
6. Eisa T. A. E., Salim N., Abdelmaboud A. Content-Based Scientific Figure Plagiarism Detection Using Semantic Mapping. Advances in Intelligent Systems and Computing. Cham, 2019. P. 420–427. DOI: https://doi.org/10.1007/978-3-030-33582-3_40
7. Eisa T. A. E., Salim N., Alzahrani S. Figure plagiarism detection based on textual features representation. 2017 6th ICT International Student Project Conference (ICT-ISPC), Johor, Malaysia, 23–24 May 2017. 2017. DOI: https://doi.org/10.1109/ict-ispc.2017.8075305
8. Eisa T., Salim N., Alzahrani S. Figure Plagiarism Detection Using Content-Based Features. Advances in Intelligent Systems and Computing. Singapore, 2017. P. 17–20. DOI: https://doi.org/10.1007/978-981-10-3779-5_3
9. Mazaheri, G., Avila, K. U., & Roy-Chowdhury, A. K. Learning to identify image manipulations in scientific publications. 2021. arXiv preprint arXiv:2102.01874.
10. Mijatović, A., Žuljević, M. F., Ursić, L., Bralić, N., Vuković, M., Roguljić, M., & Marušić, A. How good are medical students and researchers in detecting duplications in digital images from research articles: a cross-sectional survey. Research Integrity and Peer Review. 2025. Vol. 10(1), 14.
11. Duszejko, P., Walczyna, T., & Piotrowski, Z. (2025). Detection of Manipulations in Digital Images: A Review of Passive and Active Methods Utilizing Deep Learning. Applied Sciences. 2025. Vol. 15(2), 881. DOI: https://doi.org/10.3390/app15020881.
12. JOSHI, Parag Mulendra; LIU, Sam. Web document text and images extraction using DOM analysis and natural language processing. In: Proceedings of the 9th ACM symposium on Document engineering. 2009. p. 218-221.
13. Bhowmik, S., Sarkar, R., Nasipuri, M., & Doermann, D. Text and non-text separation in offline document images: a survey. International Journal on Document Analysis and Recognition (IJDAR). 2018. 21(1-2), 1–20.

Received (Надійшла) 11.06.2025
Accepted for publication (Прийнята до друку) 27.08.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

**Главчева Юлія Миколаївна** – PhD, комп'ютерні науки, директор науково-технічної бібліотеки, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
**Yuliia Hlavcheva** – PhD, Computer Science, Director of the Scientific and Technical Library, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: Yuliia.Hlavcheva@khpi.edu.ua; ORCID Author ID: https://orcid.org/0000-0001-7991-541 ;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57845103200.

**Главчев Максим Ігорович** – кандидат економічних наук, професор кафедри комп'ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
**Maksym Glavchev** – Candidate of Economic Sciences, Associate Professor, Professor of Department Computer Engineering and Programming, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: Maksym.Glavchev@khpi.edu.ua; ORCID Author ID: https://orcid.org/0000-0001-9670-9118;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57222569081.

**Features of detecting academic image plagiarism**

Yuliia Hlavcheva, Maksym Glavchev

**Abstract.** The rapid growth of visual content in academic documents highlights the issue of detecting academic image plagiarism. Unlike textual plagiarism, which is effectively identified by modern systems, visual plagiarism remains a challenging task due to the diversity of its manifestations and the potential for manipulation. **The aim of this work** is to analyze existing approaches to image recognition and comparison for detecting academic plagiarism and to identify key features. The study examines computer vision algorithms, deep learning models, and other technologies for analyzing textual elements within images. A classification of visual plagiarism forms is proposed, including exact copying, affine transformations, semantic duplication, and image generation. **The main features are identified:** extensive possibilities for manipulation and specific forms of visual content representation, lack of a global image database, difficulty in extracting non-textual content, and legal restrictions on access to full texts. **Conclusions**: Image analysis for detecting academic plagiarism remains a relevant and promising research direction, which can be implemented through the following avenues: improving modification detection algorithms and integrating visual analysis into textual plagiarism detection systems, creating image repositories, and ensuring the availability of informational tools for experts and researchers.

**Keywords:** academic integrity, academic image plagiarism, visual content, image manipulation, academic documents.