

О. К. Юхимчук
А. А. Яровий
Л. М. Ваховська
О.О. Короленко

КОМБІНОВАНИЙ МЕТОД ВИЯВЛЕННЯ ПІДРОБЛЕНИХ ЗОБРАЖЕНЬ

Вінницький національний технічний університет

Анотація

У роботі представлені особливості створення нейромережевої технології для виявлення підроблених зображень. Обґрунтовано актуальність застосування комбінованого методу, який поєднує аналіз кольорових аномалій та спектральний аналіз зображень за допомогою швидкого перетворення Фур'є (FFT). Описано принципи створення ознак, що включають використання RGB-каналів та їх попарних різниць, а також отримання частотного представлення у формі нормалізованого спектра. Поєднавши вказані підходи, можна аналізувати зображення як в кольоровому, так і в частотному просторі, що створює передумови для підвищення надійності детекції підробок. Результати дослідження можуть слугувати основою для подальшого вдосконалення систем автоматичного аналізу та моніторингу медіа-контенту.

Ключові слова: виявлення підроблених зображень, швидке перетворення Фур'є, RGB-канали, кольорові аномалії, нейронні мережі, згенеровані зображення.

Abstract

The paper presents the features of creating neural network technology for detecting fake images. The relevance of applying a combined method that combines color anomaly analysis and spectral analysis of images using fast Fourier transform (FFT) is justified. The principles of feature creation are described, including the use of RGB channels and their pairwise differences, as well as obtaining a frequency representation in the form of a normalized spectrum. By combining these approaches, images can be analyzed in both color and frequency space, which creates the conditions for improving the reliability of counterfeit detection. The results of the study can serve as a basis for further improvement of automatic analysis and monitoring systems for media content.

Keywords: detection of fake images, fast Fourier transform, RGB channels, color anomalies, neural networks, generated images.

Вступ

У сучасному інформаційному суспільстві проблема достовірності даних набуває ключового значення. Активний розвиток технологій комп'ютерної графіки та штучного інтелекту на даному етапі забезпечив можливість створювати реалістичні підроблені зображення та відео. Одним із найяскравіших проявів цього явища є deepfake – синтетичний медіаконтент, згенерований за допомогою глибоких нейронних мереж, найчастіше архітектур на основі Generative Adversarial Networks (GAN) або автокодерів [1].

Спершу ці технології були розроблені для творчості, індустрії моделювання та кіно, але за останній час стали популярними у зловживанні: дезінформаційні кампанії, «політичні» дестабілізації, тиску та злочинності в Інтернеті [2]. Оскільки підроблені зображення стають все більш реалістичними, що створює прецеденти, коли звичайний користувач не може їх відрізнити від справжніх, тому виникає потреба у створенні надійних автоматизованих методів детекції підробок.

Науковці пропонують різні підходи до розв'язання цієї проблеми. Традиційні методи виявлення підробки ґрунтуються на метаданих фотографій і файлів; наприклад, їх структурі та історії редагування. Однак ці шляхи виявлення є вразливі до повторного збереження картинки. Інший метод зосереджений на артефактах обробки, що породжені складними алгоритмами стиснення та генерації відтінків шкіри зображень: дублювання блоків JPEG та відтінків шкіри, взаємодії кольорів між собою тощо. Однак, в наш час найбільш ефективними виявляються методи, що застосовують глибинне

навчання. Зокрема, згорткові нейронні мережі (НМ), рекурентні НМ та сучасні трансформери досягають вражаючих результатів і є домінуючими типами архітектур НМ для класифікації фейк-зображень.

Разом з тим, незважаючи на досягнуті успіхи, існує кілька проблемних аспектів. По-перше, нові генеративні моделі, такі як StyleGAN, Stable Diffusion та їх модифікації, реалізують підробки, що мають суттєво менше характерних артефактів, що ускладнює їх детекцію. По-друге, майже всі алгоритми виявлення «навчалися» на певних типах підробок і демонструють зниження точності на нових, невідомих даних. По-третє, існує проблема низької стійкості моделей до спотворень, які з'являються у реальному середовищі (повторне стиснення JPEG, зниження роздільності, накладання шуму тощо) [3].

У цьому контексті з'являється потреба у розробці комбінованих методів, що об'єднують декілька різних видів характеристик. У цьому дослідженні пропонується комбінований метод, що враховує як просторові характеристики зображень (RGB-формат), так і їхні спектральні властивості (перетворення FFT). Основна ідея полягає в тому, що хоча згенероване зображення може виглядати досить реалістично, його спектральна структура часто містить закономірності, відмінні від справжніх фотографій.

Результати дослідження

Проаналізувавши різні методи виявлення підроблених зображень, особливу увагу було приділено аналізу частотних та кольірних характеристик. Найбільш перспективним виявився комбінований підхід FFT+RGB, що об'єднує обчислення двовимірного швидкого перетворення Фур'є (Fast Fourier Transform – FFT) [4] з аналізом кольорових каналів зображення в RGB-просторі. Такий підхід дає змогу суттєво оцінити як просторову структуру зображення, так і можливі кольірні аномалії, які можуть виникати під час генерації чи маніпуляцій за допомогою штучних методів.

Обґрунтування вибору послідовності дій у комбінованому методі:

1. FFT дозволяє виділити частотні елементи зображення. У реальних фотографіях спектральні характеристики слідує певним закономірностям, тоді як у згенерованих зображеннях часто спостерігаються височастотні шуми, нерівномірні текстури або артефакти згладжування.
2. RGB-аналіз допомагає виявити невідповідності в кольорових каналах. Реальні фотографії відрізняються корельованістю каналів R, G та B. Тоді як у підроблених зображеннях (наприклад, deepfake або GAN-зображеннях) можуть спостерігатися порушення у балансі кольорів чи неприродні переходи.
3. Комбінування FFT+RGB забезпечує взаємне доповнення. Частотний аналіз підсилює здатність виявляти приховані текстурні дефекти, тоді як аналіз кольірних каналів допомагає зафіксувати глобальні невідповідності в спектрально-кольірному просторі.

Метод включає такі основні кроки:

1. Аналіз кольорових аномалій

Для вивчення кольірних невідповідностей була розроблена спеціальна архітектура нейронної мережі, яка обробляє канали R, G, B та їх попарні різниці (R–G, R–B, G–B). Таким чином формується 6-канальне вхідне зображення, що дає змогу моделі виявляти приховані аномалії у кореляції кольорів.

Таким чином, модель навчається розпізнавати нетипові патерни кольорових зв'язків, що характерні згенерованим чи відредагованим зображенням, але рідко зустрічаються у реальних фотографіях.

Математично вхідні ознаки описуються так:

$$X = \{R, G, B, (R - G), (R - B), (G - B)\} \quad (1)$$

$$y = f_{CNN}(X), y \in [0,1] \quad (2)$$

де y – ймовірність підробки.

2. Спектральний аналіз на основі FFT

Для кожного зображення виконується перетворення Фур'є:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (3)$$

У результаті формується спектр амплітудних компонентів:

$$|F(u, v)| = \sqrt{\operatorname{Re}(F(u, v))^2 + \operatorname{Im}(F(u, v))^2} \quad (4)$$

Для стабільності розрахунків застосовується логарифмування:

$$S(u, v) = \log(|F(u, v)| + \epsilon), \quad \epsilon = 10^{-8} \quad (5)$$

Нормалізований спектр подається у неймережу для навчання та визначення ознак підробок.

На основі досліджень можна відзначити, що реальні зображення характеризуються більш «згладженим» спектром із високою концентрацією енергії в діапазоні низьких частот. На відміну від них, фейкові зображення містять додаткові піки та високочастотні шуми, пов'язані з алгоритмами генерації (рис. 1) [5].



Рисунок 1 – Спектр справжнього та підробленого зображення

Окрім цього, у фейкових зображеннях при візуалізації окремих каналів (R, G, B) можна помітити різницю у кольоровій насиченості та балансі, що додатково демонструє переваги використання комбінованого методу.

Висновки

Проведене дослідження показало, що комбінування аналізу кольорових аномалій та спектральних характеристик зображення є доцільним напрямом для розробки та удосконалення методів виявлення підроблених зображень. Запропонований комбінований підхід на основі поєднання аналізу за FFT та RGB дозволяє поєднувати локальні та глобальні ознаки зображень, що може стати основою для майбутніх експериментів та покращень неймережевих моделей у цій області для виявлення підроблених зображень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ian Goodfellow, et al.: Generative Adversarial Nets. Advances in Neural Information Processing Systems, 2014. URL: <https://arxiv.org/abs/1406.2661> (дата звернення: 22.11.2025).
2. Використання підробленими зображеннями. URL: <https://medialab.online/news/photoweapon/> (дата звернення: 22.11.2025).
3. StyleGAN. URL: <https://www.geeksforgeeks.org/machine-learning/stylegan-style-generative-adversarial-networks/> (дата звернення: 22.11.2025).
4. Fast Fourier transform (FFT). URL: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm> (дата звернення: 22.11.2025).
5. Convertini V. N., Impedovo D., Lopez U., Pirlo G., Sterlicchio G. Discrete Fourier Transform in Unmasking Deepfake Images: A Comparative Study of StyleGAN Creations // *Information*. – 2024. – Vol. 15, No. 11. – Art. 711. – DOI: 10.3390/info15110711 (дата звернення: 23.11.2025).

Юхимчук Олексій Костянтинович – студент групи 2КН-24м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: yukha.2019@gmail.com

Яровий Андрій Анатолійович – доктор технічних наук, професор, завідувач кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Ваховська Любов Михайлівна – асистент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Короленко Олег Олександрович – аспірант кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Oleksii K. Yukhymchuk – Student of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia, email: yukha.2019@gmail.com

Andrii A. Yarovi – Dr. Sc. (Eng.), Professor, Head of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.

Lyubov M. Vakhovska – Assistant of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.

Oleh O. Korolenko – Postgraduate Student of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.