



УДОСКОНАЛЕННЯ МЕТОДУ СПЕКТРАЛЬНО-ВІЗУАЛЬНОГО ТА СТРУКТУРНОГО АНАЛІЗУ ФОТО- І ВІДЕООБ'ЄКТІВ МЕРЕЖЕВИХ РЕСУРСІВ ШЛЯХОМ ЗБЕРЕЖЕННЯ ПОХОДЖЕННЯ ОЗНАК

Удосконалено метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів, які опрацьовують мережеві ресурси, шляхом збереження походження кожної ознаки до етапу їх числового поєднання. Встановлено, що рання інтеграція контейнерних і декодованих даних ускладнює відновлення байтової, просторової або часової підстави рішення моделі логістичної регресії та може прирівняти недоступну перевірку до перевірки з нульовим результатом. Сформовано два незалежні подання одного й того самого незмінного об'єкта: структурне подання ознак відображає сигнатуру, граматику контейнера, фізичні й логічні межі, вкладеність елементів, службові поля та невіднесені байтові ділянки, а спектрально-візуальне подання ознак охоплює локальні статистики, високочастотні залишки, дискретно-косинусні, хвилькові й часові характеристики. Розроблено простежуваний опис, який поєднує числові подання ознак з байтовими та просторово-часовими локалізаціями, маскою доступності, контрольною сумою, версіями правил і параметрами декодування. Проведено групо-незалежне оцінювання 200 джерельних груп, розподілених між навчальною, налаштувальною та перевіркою частинами до створення похідних об'єктів. Встановлено, що на внутрішній перевірній частині з 80 об'єктів спільне подання ознак забезпечило Accuracy 0,9750, Recall 0,9500, F1 0,9744 і ROC-AUC 0,9869. Оцінено статистичну відмінність парних рішень: точний критерій Мак-Немара підтвердив перевагу над спектрально-візуальним поданням ознак, але не підтвердив перевагу над структурним за рівня значущості 0,05. З'ясовано, що у зовнішній серії з 56 MP4-об'єктів Balanced Accuracy становила 0,8333 – для структурного, 0,7857 – для спільного та 0,5000 – для спектрально-візуального подання ознак, що вказує на залежність реалізованих зорових ознак від пристрою, кодека й попереднього опрацювання. Визначено, що 95 % часу дорівнював 2,136 мс – для структурного і 853,971 мс – для спільного варіанта. Отримані результати підтверджують придатність методу до формування відтворюваного опису контрольовано змінених об'єктів, але не доводять виявлення природно поширеного шкідливого коду або готовності до промислового застосування.

Ключові слова: цілісність мультимедійного контейнера; мультимедійна криміналістика; маска доступності перевірок; групо-незалежне оцінювання методу статичного аналізу; локалізація структурних і спектрально-візуальних відхилень; відтворюваність результатів.

Вступ / Introduction

Мультимедійний об'єкт, що надходить до мережевого ресурсу, має щонайменше два взаємопов'язані подання ознак: структурне та спектрально-візуальне. Початкова послідовність байтів визначає тип контейнера, порядок і межі його елементів, службові поля та невіднесені ділянки. Після декодування той самий об'єкт набуває просторових, частотних і, для відео, часових властивостей. Додані після логічного завершення контейнера байти можуть не змінювати зоровий вміст, тоді як локальна зміна пікселів може не порушувати граматику формату. Отже, перевірка тільки одного подання ознак не охоплює всіх можливих відхилень.

У роботі розмежовано аномалію, аномальний код і шкідливий код. Аномалією вважають спостережуване

відхилення від визначеного структурного правила або звичайних числових властивостей. Аномальний код означає навмисно сформовану послідовність даних, команду чи структурне значення, яке не відповідає заявленому призначенню носія. Висновок про шкідливий код додатково потребує підтвердження програмної семантики та небезпечної функції. Тому позитивна позначка у проведених експериментах характеризує наперед відому зміну похідного об'єкта, а не природно поширену шкідливу програму.

Спектральні та статистичні засоби виявляють слабкі зміни у декодованому вмісті мультимедійного об'єкта, тоді як структурні засоби перевіряють контейнерні правила й байтові межі. Раннє об'єднання таких ознак у єдиний вектор приховує їх джерело. Додаткова неоднозначність виникає, коли недоступну перевірку пода-

© Copyright 2026 by the author(s)

Денисюк Дмитро Олександрович, ст. викладач, кафедра комп'ютерної інженерії та інформаційних систем.

Email: denysiuk@khmnu.edu.ua; <https://orcid.org/0000-0002-7345-8341>

Боярчук Артем Володимирович, доктор PhD, доцент, наук. співробітник, Інститут права.

Email: a.boiarчук@taltech.ee; <https://orcid.org/0000-0001-7349-1371>

Цитування за ДСТУ: Денисюк Д. О., Боярчук А. В. Удосконалення методу спектрально-візуального та структурного аналізу фото- і відео-об'єктів мережевих ресурсів шляхом збереження походження ознак. Науковий вісник НЛТУ України. 2026, т. 36, № 4. С. 130–137.

Citation APA: Denysiuk, D. O., & Boyarchuk, A. V. (2026). Improvement of the method for spectral-visual and structural analysis of image and video objects in web systems through feature-provenance preservation. *Scientific Bulletin of UNFU*, 36(4), 130–137. <https://doi.org/10.36930/40360415>

ють нульовим значенням і цим самим прирівнюють до виконаної перевірки без установленого порушення. Для відтворюваного висновку разом із числовим значенням ознаки потрібно зберігати її тип, місце формування, стан доступності та умови отримання.

Актуальність дослідження зумовлена потребою у відтворюваному статичному аналізі фото- і відеооб'єктів мережевих ресурсів, оскільки наявні засоби переважно опрацьовують структуру контейнера та декодований сигнал окремо, а їх раннє числове поєднання може призводити до втрати відомостей про походження ознак. У проаналізованих джерелах [1, 2, 3, 4, 6, 7, 11, 15] не виявлено підходу, який до етапу інтеграції одночасно зберігав би тип ознаки, її байтову, просторову або часову локалізацію, стан доступності та умови формування. Це визначає потребу в удосконаленні методу спектрально-візуального та структурного аналізу зі збереженням походження ознак.

Об'єкт дослідження – статичний аналіз фото- і відеооб'єктів мережевих ресурсів за байтовим і зоровим поданнями.

Предмет дослідження – правила окремого формування структурних і спектрально-візуальних ознак та їх простежуваного поєднання, які дадуть змогу зберегти походження, локалізацію, стан доступності й умови формування кожної ознаки до етапу числового поєднання та забезпечити відтворюваність результатів аналізу.

Мета роботи – удосконалити метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів шляхом збереження походження, локалізації й доступності ознак до етапу їх числового поєднання, що дасть можливість здійснити перевірку сформованих подань ознак на групово-незалежних внутрішніх і зовнішніх вибірках.

Для досягнення поставленої мети визначено такі основні завдання дослідження:

- проаналізувати засоби структурного та спектрально-візуального дослідження мультимедійних об'єктів, що дасть змогу визначити обмеження їх окремого застосування;
- формалізувати вхідні дані, вихідний опис і правила збереження походження ознак, забезпечивши зв'язок між числовою оцінкою, місцем її формування та умовами отримання;
- удосконалити метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів шляхом окремого формування структурного і спектрально-візуального подань ознак із подальшим узгодженням поєднанням, що дасть змогу зберегти байтову, просторову й часову локалізацію ознак;
- розробити групово-незалежний протокол експериментальної перевірки удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів, щоб унеможливити потрапляння споріднених версій одного джерела до різних частин вибірки;
- провести внутрішню та зовнішню перевірки сформованих подань ознак, що дасть змогу визначити межі їх застосування до різних джерел, форматів і способів попереднього опрацювання.

Аналіз останніх досліджень та публікацій. У дослідженні [1] запропоновано ентропійно орієнтовану глибоку нейронну мережу для стеганоаналізу цифрових зображень. Розроблене рішення підсилює розпізнавання слабких статистичних відхилень у декодованому поданні, однак не встановлює байтову причину сформова-

ної ознаки та не описує доступність контейнерної перевірки.

У роботі [2] досліджено глибоку залишкову мережу, що навчається на фільтрованих поданнях зображення та виявляє просторові стеганографічні зміни. Такий підхід формує ознаки на рівні декодованого сигналу, проте не пов'язує їх із конкретними структурними елементами мультимедійного контейнера.

У роботі [3] проаналізовано, систематизовано просторові, часові, частотні та научувані підходи до відеостеганоаналізу. Автори показали залежність результатів від кодера, структури груп кадрів, повторного кодування і складу набору даних, що обґрунтовує потребу у фіксації правил декодування та відбору кадрів для кожної ознаки.

У роботі [4] розроблено згорткову трансформерну архітектуру для просторового стеганоаналізу. Поєднання локальних і глобальних залежностей розширює можливості дослідження зорового подання цифрового зображення, однак походження ознаки з погляду правила контейнера, службового поля або байтового інтервалу залишається поза межами моделі.

У дослідженні [6] запропоновано MalJPEG, де структурні характеристики JPEG поєднано з машинним навчанням для виявлення шкідливих зображень. Робота підтверджує значення контейнерних ознак, але орієнтується на один формат і не поєднує структурного результату із просторово-часовою локалізацією зорових відхилень.

У роботі [7] підтверджено придатність статистичних ознак для виявлення контрольованих стеганографічних змін у зображеннях. Водночас, статистичне відхилення не є тотожним порушенню контейнерної будови, а додані байти не обов'язково змінюють декодоване подання цифрового зображення.

У дослідженні [11] проаналізовано зловживання багатиформатними об'єктами, один набір байтів яких допускає кілька тлумачень. Результати показують обмеженість перевірки за розширенням, однією сигнатурою або одним декодером, але не формують опису, що зберігає взаємозв'язок між структурними та зоровими підставами рішення.

У роботі [15] досліджено форматно залежні структурні ознаки MP4 для виявлення невідомих шкідливих об'єктів. Отримані результати вказують на придатність контейнерних перевірок, однак правила одного формату не можна переносити на інший без окремої реалізації, перевірки граматики та фіксації стану доступності.

Проведений аналіз останніх досліджень та публікацій засвідчив, що наявні засоби переважно досліджують декодований сигнал або структуру контейнера окремо, а під час раннього числового поєднання можуть втрачати зв'язок ознаки з її байтовою, просторовою чи часовою підставою. Це визначає потребу в удосконаленні методу, який до інтеграції зберігає тип ознаки, локалізацію, стан доступності та умови формування.

Матеріали та методи дослідження. Матеріалом дослідження стали початкові та контрольовано змінені фото- і відеооб'єкти форматів PNG, JPEG, WebP, MP4 і WebM. Джерельні групи розподілили на навчальну, налаштувальну й перевірку частини до створення похідних копій. Структурне, спектрально-візуальне та спільне подання ознак оцінювали однаковою логістичною регресією. Результати визначали за матрицями рішень,

Accuracy, Balanced Accuracy, Precision, Recall, F1, ROC-AUC, точним критерієм Мак-Немара, часом оброблення та піковим використанням оперативної пам'яті.

Внутрішній набір охоплював 200 джерельних груп і 400 об'єктів форматів PNG та MP4. До створення похідних версій групи розподілили між навчальною, налаштувальною та перевіркою частинами у співвідношенні 60:20:20, що відповідало 240, 80 і 80 об'єктам. Усі версії одного джерела залишали в межах однієї частини даних, унеможливаючи потрапляння споріднених об'єктів до навчальної та перевірки вибірок. Контрольовані зміни охоплювали порушення структури контейнера, додавання невіднесених байтових ділянок і частотні зміни декодованого сигналу. Позитивна позначка вказувала на наперед визначену зміну, а не на наявність шкідливої програмної функції.

Після завершення налаштування методу з наявних 14 природних відеозаписів сформовано першу зовнішню серію із 56 MP4-об'єктів, серед яких 14 незмінених і 42 контрольовано змінених. Під час зовнішньої перевірки параметри моделі та пороги класифікації повторно не налаштовували.

Структурне Z_s , спектрально-візуальне Z_v і спільне Z_c подання ознак оцінювали однаковою моделлю логістичної регресії з коефіцієнтом L2-регуляризації 0,05. За налаштувальною частиною зафіксовано пороги 0,328738, 0,481738 і 0,456468 відповідно. Нормування ознак, оцінювання коефіцієнтів і вибір порогів виконували без використання перевірних даних. Основним показником була Balanced Accuracy; додатково обчислювали Accuracy, Precision, Recall, F1 і ROC-AUC. Парні рішення порівнювали точним двобічним критерієм Мак-Немара з поправкою Голма за рівня значущості 0,05.

Програмну реалізацію виконано мовою Python 3.13.3 із використанням NumPy 2.3.0 та OpenCV 4.10.0. Зображення й відеозаписи декодували засобами OpenCV за однакових параметрів для всіх подань. Для формування зовнішньої серії з кожного джерельного MP4-запису відбирали 12 послідовних кадрів і приводили їх до розміру 320×240 пікселів зі збереженням пропорцій. Ресурсні вимірювання проводили послідовно у CPU-режимі в Windows 11 Pro на комп'ютері з процесором AMD Ryzen 7 5700X і 32 ГіБ оперативної пам'яті. Початкове значення генератора для зовнішньої серії становило 20260719.

Результати дослідження та їх обговорення / Research results and their discussion

Принцип збереження походження ознак. Початковим входом методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів є незмінна послідовність байтів мультимедійного об'єкта X_{mm} . Структурний напрям формує карту контейнера X_s і подання службових даних X_m до будь-якого виправлення виявлених відхилень. Після контрольованого декодування формуються зорове або покадрове подання X_v , частотне подання X_f і, для відео, часове подання X_t . Числові структурні ознаки визначають за формулою, яка описує їх формування на основі будови контейнера та службових даних:

$$Z_s = F_s(X_s, X_m), \quad (1)$$

де Z_s є структурним числовим поданням ознак мультимедійного об'єкта, а $F_s(\cdot)$ є функцією формування структурних ознак.

Спектрально-візуальні ознаки визначають за формулою, яка описує поєднання зорових, частотних і часових даних без залучення контейнерної оцінки:

$$Z_v = F_v(X_v, X_f, X_t), \quad (2)$$

де Z_v є спектрально-візуальним числовим поданням ознак мультимедійного об'єкта, а $F_v(\cdot)$ є функцією формування відповідних ознак.

Для однакового кількісного оцінювання два числові подання ознак поєднуються тільки після збереження їх локалізацій і станів доступності. Спільне подання ознак задається рівністю, яка описує послідовне об'єднання структурного та спектрально-візуального векторів:

$$Z_c = \{Z_s, Z_v\}, \quad (3)$$

де $Z_s = \{z_{s,j}, j = \overline{1,9}\}$ – числове подання структурних і спектрально-візуальних $Z_v = \{z_{v,j}, j = \overline{1,19}\}$ ознак мультимедійного об'єкта, які використовує оцінювальна модель. У реалізованій конфігурації Z_s містить 9, Z_v – 16, а Z_c – 25 координат. Маска доступності передається окремо і не змінює розмірність числового вектора.

Повний результат роботи методу подамо кортежем, який описуватиме простежуваний зв'язок між числовими поданнями, локалізаціями та умовами їх формування, а саме:

$$H = \langle Z_s, Z_v, L_s, L_v, m, C \rangle, \quad (4)$$

де: H – повний простежуваний опис результату статичного аналізу мультимедійного об'єкта; L_s – містить байтові інтервали структурних ознак об'єкта; L_v – просторові ділянки кадрів і часові відрізки; m – маска доступності; C – контрольна сума об'єкта, ідентифікатор джерела, версії правил, засіб декодування, параметри перетворень і порядок відбору кадрів. Значення $m_j = 1$ означає, що j -ту перевірку виконано, а $m_j = 0$ – що вона недоступна для відповідного формату або умов декодування. Нульове значення ознаки допускається тільки як результат виконаної перевірки, що не встановила відхилення.

Умови збереження походження та доступності ознак на окремих етапах методу наведено в табл. 1.

Табл. 1. Умови збереження походження ознак у структурному та спектрально-візуальному поданнях / Conditions for preserving feature provenance in structural and spectral-visual representations

Код	Умова	Збережені відомості
P1	Обидва напрями аналізують той самий незмінний об'єкт	Контрольна сума та ідентифікатор джерела
P2	Структурна ознака пов'язується з перевіреним правилом	Версія правила, байтовий інтервал і фактичне значення поля
P3	Спектрально-візуальна ознака має просторову або часову підставу	Номер кадру, час, локальна ділянка та параметри декодування
P4	Недоступність не замінюється числовим нулем	Окрема маска доступності та причина невиконання
P5	Числове поєднання виконується після реєстрації походження	Окремі структурне й спектрально-візуальне подання мультимедійного об'єкта, локалізації та контекст формування

Під походженням ознаки в цій роботі розуміють сукупність відомостей про початковий об'єкт, правило її формування, версію правила, байтову або просторово-часову локалізацію, стан доступності та параметри де-

кодування. Опис H вважають повним тільки за умови збереження всіх цих відомостей. Якщо незмінність об'єкта не підтверджено або зв'язок похідного подання ознак з початковою послідовністю байтів втрачено, подальше поєднання ознак припиняється. Послідовність етапів формування структурних і спектрально-візуальних ознак, реєстрації їх походження та подальшого числового поєднання наведено на рисунку.



Рисунк. Послідовність етапів формування структурного та спектрально-візуального подань зі збереженням походження ознак / The sequence of stages in the formation of structural and spectral-visual representations whilst preserving the origin of features

Структурне та спектрально-візуальне подання мультимедійного об'єкта. Структурний напрям зіставляє розширення, сигнатуру, контейнерну граматику й тип, установлений декодером; перевіряє фізичні та логічні межі, порядок і вкладеність елементів, заявлену довжину, службові значення, контрольні суми, невіднесені байти, дані після логічного завершення та можливість неоднозначного тлумачення. Виявлену розбіжність не трактують як самостійний доказ шкідливого впливу, а вона зберігається разом із правилом і конкретно байтовою ділянкою.

Спектрально-візуальний напрям використовує тільки дані, декодовані з того самого незмінного об'єкта. Зображення або кадр поділяють на локальні ділянки за сталими правилами. У них обчислюють локальні статистики, відгуки високочастотних фільтрів, властивості молодших бітів, міжканальні відмінності, сусідні переходи та залишкові характеристики. Частотне подання ознак утворюють дискретне косинусне й хвилькове перетворення, а часовий опис узагальнює зміни між відібраними кадрами. Статистичне відхилення не ототожнюється з контейнерним порушенням, а початкову структуру не виправляють до реєстрації фактичних меж і джерел ознак.

У межах цієї роботи спектрально-візуальний аналіз виконують над цифровим зображенням або декодованими відеокадрами. Текстуру складову цифрового зображення описують за допомогою локальних статистик, високочастотних залишків, міжканальних відмінностей, сусідніх переходів, дискретно-косинусних і хвилькових характеристик. Ці ознаки відображають просторово-часотну організацію значень пікселів та використовують для виявлення локальних цифрових відхилень і контрольованих змін. Отже, текстурний аналіз зображення у цьому дослідженні означає аналіз цифрової його текстури, а не визначення властивостей або походження об'єктів, зображених у кадрі.

Для відео кадри відбирають за чотирма підставами: ключовий кадр, рівномірне покриття тривалості, межа сцени або близькість до часової ділянки, визначеної структурним напрямом. Якщо кадр відповідає кільком умовам, його не дублюють, але зберігають усі причини відбору. Непереверений часовий проміжок не позначається як перевірений, а структурна підстава може ініціювати відбір кадру, проте сама не збільшує спектрально-візуальну оцінку.

Табл. 2. Джерела, розмірність і склад реалізованих подань ознак / Sources, dimensionality, and composition of the implemented feature representations

Подання ознак	Розмірність	Джерело	Основні групи ознак
Z_s	9	Послідовність байтів, карта контейнера, службові дані	Узгодженість типу, межі, порядок, вкладеність, довжини, контрольні поля, невіднесені та завершальні ділянки
Z_v	16	Декодоване зображення або відеокадри	Локальні статистики, високочастотні залишки, дискретно-косинусні, хвилькові та часові характеристики
Z_c	25	Послідовне поєднання Z_s і Z_v	Спільне числове подання ознак без втрати окремих локалізацій і маски доступності

Метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів виконується в такій послідовності:

- 1) фіксується контрольна сума та ідентифікатор початкового мультимедійного об'єкта;
- 2) без виправлення відхилень формується карта контейнера і реєструються фактичні байтові межі;
- 3) обчислюється структурне подання ознак Z_s та відповідні байтові локалізації L_s ;
- 4) об'єкт декодується за зафіксованими правилами, а для відео визначається набір кадрів;
- 5) формуються спектрально-візуальне подання ознак Z_v і просторово-часові локалізації L_v ;
- 6) для кожної перевірки встановлюється стан доступності та формується маска m ;
- 7) після реєстрації походження утворюється числове подання ознак Z_c ;
- 8) зберігається повний опис H , що використовується для повторного аналізу та тлумачення рішення.

Для оцінювання обчислювальної складності методу позначимо через B кількість байтів початкового мультимедійного об'єкта, а через P — загальну кількість пікселів у зображеннях або відеокадрах, відібраних для опрацювання. Структурний напрям виконує один послідовний прохід байтовим поданням контейнера, тому його часова складність методу становить $T_s(B)=O(B)$. За сталої кількості фільтрів і перетворень та фіксованих

параметрів опрацювання спектрально-візуальний напрям має складність $T_v(P)=O(P)$. Отже, загальну часову складність спільного опрацювання визначають співвідношенням:

$$T_c(B, P) = T_s(B) + T_v(P) = O(B) + O(P) = O(B + P), \quad (5)$$

де: $T_s(B)$ – часова складність структурного напрямку; $T_v(P)$ – часова складність спектрально-візуального напрямку; $T_c(B, P)$ – загальна часова складність спільного формування ознак. Оскільки оцінювальна модель опрацює вектор фіксованої розмірності, її застосування не змінює наведеної асимптотичної оцінки. Ця оцінка не враховує передавання мультимедійних об'єктів мережею, очікування в чергах і одночасне обслуговування кількох запитів.

Організація експерименту. Внутрішній набір містив 200 початкових груп і 400 PNG- та MP4-об'єктів. Спочатку групи поділено у співвідношенні 60:20:20, і тільки після цього створено похідні версії. Навчальна частина містила 240, налаштувальна – 80, перевірна – 80 об'єктів; жодна похідна копія не могла опинитися в іншій частині, ніж її джерело. Контрольовані зміни охоплювали контейнерні порушення, додані або невіднесені байтові ділянки та частотні зміни. Позитивний клас означав наявність наперед заданої зміни (табл. 3).

Першу зовнішню серію сформовано після фіксації параметрів: із 14 природних записів створено 56 MP4-об'єктів, серед яких 14 не мали змін, а 42 містили контрольовані похідні зміни. Дві додаткові серії перевіряли перенесення на п'ять форматів і нові пристрої. Споріднені фрагменти в усіх серіях утримували в одній джерельній групі.

Табл. 3. Склад експериментальних серій, кількість об'єктів і призначення даних / Composition of experimental series, number of objects, and purpose of the data

Серія	Джерела / групи	Об'єкти	Призначення
Внутрішня	200 джерельних груп	400 PNG і MP4: 240 навчальних, 80 налаштувальних, 80 перевірних	Налаштування та внутрішня групова-незалежна перевірка
Перша зовнішня	14 природних записів	56 MP4: 14 незмінених і 42 контрольовано змінених	Перша перевірка без повторного налаштування
Кілька форматів	8 природних записів	10 000 об'єктів форматів PNG, JPEG, WebP, MP4 і WebM	Допоміжна перевірка перенесення між форматами
Нові пристрої	4 нові пристрої, 8 записів	2 000 об'єктів п'яти форматів	Допоміжна перевірка перенесення між джерелами

Для кількісного зіставлення застосовано одну модель логістичної регресії, яка виконувала роль однакового оцінювального засобу і не входила до визначення методу. Коефіцієнт L_2 -регуляризації становив 0,05. За налаштувальною частиною зафіксовано пороги 0,328738 – для Z_s , 0,481738 – для Z_v і 0,456468 – для Z_c ; у зовнішніх серіях їх не змінювали. Нормування, оцінювання коефіцієнтів і вибір порога виконували без використання перевірних даних.

Основним показником була *Balanced Accuracy*, додатковими – Accuracy, Precision, Recall, F_1 і ROC-AUC. Парні внутрішні рішення порівнювали точним двобічним критерієм Мак-Немара за рівня значущості 0,05. Для двох запланованих порівнянь застосовано поправку Голма; скориговані значення становили $5,96 \cdot 10^{-6}$ – для

зіставлення Z_c із Z_v і 0,0625 – для зіставлення Z_c із Z_s , тому змістовні висновки не змінилися.

Внутрішні та зовнішні результати. На внутрішній частині Z_c правильно класифікував 78 із 80 об'єктів; обидва пропуски припали на MP4 із частотними змінами. Порівняно з Z_v спільне подання ознак виправило 23 помилки й погіршило одне рішення. Порівняно з Z_s воно виправило п'ять помилок без нових помилок, але точне значення $p = 0,0625$ не дає підстав стверджувати перевагу за рівня 0,05 (табл. 4).

Табл. 4. Матриці рішень і показники внутрішньої ($n = 80$) та першої зовнішньої ($n = 56$) перевірок / Confusion matrices and metrics for the internal ($n = 80$) and first external ($n = 56$) evaluations

Серія	Подання ознак	TN	FP	FN	TP	Accuracy	BA
Внутрішня	Z_s	39	1	6	34	0,9125	0,9125
	Z_v	33	7	17	23	0,7000	0,7000
	Z_c	40	0	2	38	0,9750	0,9750
Зовнішня	Z_s	14	0	14	28	0,7500	0,8333
	Z_v	9	5	27	15	0,4286	0,5000
	Z_c	12	2	12	30	0,7500	0,7857

Структурне подання ознак пропустило шість із десяти суто частотних змін, що відповідає визначеній області його спостереження. Такий результат не є помилкою байтової локалізації: коректна контейнерна будова не містить достатньої інформації про кожну зміну декодованого сигналу. Водночас, два пропуски Z_c показують, що додавання спектрально-візуального напрямку не гарантує виявлення кожної частотної зміни (табл. 5).

Табл. 5. Додаткові показники внутрішньої ($n = 80$) та першої зовнішньої ($n = 56$) перевірок / Additional metrics for the internal ($n = 80$) and first external ($n = 56$) evaluations

Серія	Подання ознак	Precision	Recall	F_1	ROC-AUC
Внутрішня	Z_s	0,9714	0,8500	0,9067	0,9337
	Z_v	0,7667	0,5750	0,6571	0,7500
	Z_c	1,0000	0,9500	0,9744	0,9869
Зовнішня	Z_s	1,0000	0,6667	0,8000	0,8333
	Z_v	0,7500	0,3571	0,4839	0,5034
	Z_c	0,9375	0,7143	0,8108	0,8316

Перша зовнішня серія не відтворила внутрішньої переваги Z_c над Z_s . Структурне подання ознак правильно визначило всі 14 незмінених об'єктів і мало *Balanced Accuracy* 0,8333. Для Z_c цей показник становив 0,7857, а для Z_v – 0,5000. Отже, реалізовані спектрально-візуальні ознаки виявилися чутливими до джерела, кодека й попереднього опрацювання. Для багатоформатної серії та серії з новими пристроями збережено тільки сукупні значення *Balanced Accuracy* 0,7916 і 0,7827 для Z_c , але не повні матриці рішень. Тому ці результати характеризують напрям перенесення, проте не використовуються для статистичного порівняння конфігурацій.

Тривалість опрацювання і пікове використання оперативної пам'яті. Вимірювання тривалості виконання й обсягу використаної оперативної пам'яті виконано у п'яти однопотокових запусках удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережеских ресурсів на 160 об'єктах. Основні витрати створювали декодування та спектрально-візуальні перетворення. У 95 % вимірювань спільний варіант завершувався не пізніше ніж за 853,971 мс, тоді як структурний – не пізніше ніж за 2,136 мс (табл. 6).

Табл. 6. Тривалість опрацювання і пікове використання оперативної пам'яті за п'ять однопотоківих запусків удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережесих ресурсів на 160 об'єктах / Processing time and peak RAM usage for five single-threaded runs of the improved method for static spectral-visual and structural analysis of photo and video objects from network resources

Подання ознак	Медіана, мс	95 %, мс	Пам'ять, МБ
Z_s	0,304	2,136	78,77
Z_v	65,524	760,662	96,79
Z_c	68,666	853,971	101,53

Наведені значення характеризують конкретну локальну однопотоківу реалізацію без мережевого передавання, черг і конкурентних запитів. Тому їх не варто тлумачити як оцінку виробничої пропускну здатності або переносити на інше апаратно-програмне середовище без повторного вимірювання.

Загрози достовірності та межі тлумачення. Внутрішню достовірність удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережесих ресурсів підтримує груповий поділ до створення похідних об'єктів, однак контрольовані зміни формувалися за наперед визначеними правилами. Зовнішнє узагальнення обмежують 14 природних записів у першій серії та відсутність повних матриць рішень для двох допоміжних серій. Окрім цього, 56 зовнішніх об'єктів утворено з 14 джерел, тому похідні одного запису не є статистично незалежними; подальші довірчі інтервали доцільно обчислювати повторним вибиранням на рівні первинних джерел.

Цільова позначка описує відому зміну, а не намір автора чи шкідливу програмну функцію. Метод не виконує динамічного аналізу програмної поведінки та не встановлює шкідливості без додаткових доказів. Дві допоміжні серії з неповними протоколами розглядають тільки як опис напрямку перенесення. Ресурсні вимірювання характеризують п'ять локальних однопотоківих запусків удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережесих ресурсів і не визначають виробничої пропускну здатності веб-системи.

Обговорення результатів дослідження. Збереження походження ознак (англ. *Feature Provenance*) передбачає відстеження всієї хронології формування, трансформації та виділення математичних, спектральних і структурних характеристик об'єкта від моменту зйомки до його аналізу в мережі. Удосконалення методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів у мережесих ресурсах через збереження походження (провенієнсу, provenance) ознак є вкрай актуальним науково-технічним завданням. Це пов'язано з масовим поширенням дезінформації, генеративного контенту (Deepfakes), а також необхідністю точної цифрової експертизи мультимедіа.

У дослідженні [5] сформовано великий набір контрольованих змінених наукових зображень для порівняння засобів виявлення підробок. Робота підтверджує цінність точної еталонної розмітки, але також показує, що синтетичне формування змін обмежує перенесення висновків на природні випадки. Аналогічне обмеження враховано в цій роботі: внутрішні результати характеризують наперед задані похідні зміни.

У роботі [8] мережа MVSS-Net використовує кілька подань і багатомасштабний нагляд для виявлення мані-

пуляцій. Порівняння показує, що багатоподанневий аналіз зменшує залежність від одного виду сигналу, однак раннє спільне научуване подання ознак не гарантує збереження походження кожної числової координати.

У дослідженні [9] система TruFor поєднує різномірні криміналістичні підказки та формує карту локалізації з оцінкою надійності. Це узгоджується з потребою не обмежувати рішення одним типом даних. Відмінність запропонованого методу полягає у збереженні поряд із зоровими характеристиками форматно залежних правил, байтових інтервалів і маски доступності.

У роботі [10] запропоновано карту слідів для дослідження зображень зі стеганографічно прихованим програмним навантаженням і перевірено її на структурованих наборах. Такий підхід підтверджує значення простежуваного зв'язку між ознакою та джерелом, тоді як удосконалений метод додатково зберігає часову локалізацію і стан доступності перевірки.

У роботі [12] мережа CAT-Net поєднує RGB- і DCT-подання для виявлення та локалізації монтажу за артефактами стиснення. Отримані результати підтверджують доцільність одночасного використання різних подань сигналу, однак таке поєднання не відновлює фактичні байтові межі контейнера.

У розділі монографії [13] походження мультимедійних даних розглянуто як установлення історії перетворень і зв'язків між кількома об'єктами. На відміну від міжоб'єктного графа, запропонований у цій роботі опис фіксує внутрішнє походження ознак одного об'єкта: правило, локалізацію, версію та контекст формування.

У роботі [14] запропоновано паралельну архітектуру для класифікації значних обсягів мультимедійних матеріалів у цифровій криміналістиці. Порівняння ресурсних показників підтверджує доцільність каскадного застосування: швидкий структурний напрям може виконувати первинний контроль, а витратніший спектрально-візуальний аналіз – запускатися вибірково.

Проведене обговорення результатів дослідження підтверджує доцільність поєднання кількох подань мультимедійного об'єкта, але показує, що сама класифікаційна оцінка не є достатньою для обґрунтованого висновку. Відмінною властивістю удосконаленого методу спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережесих ресурсів є збереження зв'язку між числовою ознакою, місцем її формування, станом доступності та версією правила до етапу інтеграції.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – удосконалено метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережесих ресурсів, який, на відміну від наявних, передбачає окреме формування структурного та спектрально-візуального подань зі збереженням байтової, просторової й часової локалізації, маски доступності та контексту формування до етапу числового поєднання, що дає змогу простежити походження кожної ознаки та розрізнити недоступну перевірку і перевірку без установленого відхилення.

Практична значущість результатів дослідження – можна застосувати як основу засобів статичного контролю фото- і відеооб'єктів мережесих ресурсів для відновлення джерела кожної ознаки, повторного виконан-

ня аналізу за зафіксованими правилами та розмежування відсутнього значення, недоступної перевірки й перевірки без виявленого порушення.

Висновки / Conclusions

Удосконалено метод спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів шляхом збереження походження, локалізації й доступності ознак до етапу їх числового поєднання, що дало можливість здійснити перевірку сформованих подань ознак на групово-незалежних внутрішніх і зовнішніх вибірках. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Удосконалено метод статичного спектрально-візуального та структурного аналізу фото- і відеооб'єктів мережевих ресурсів. Його відмінність полягає в окремому формуванні структурного Z_s і спектрально-візуального Z_v подань та їх поєднанні тільки після реєстрації локалізацій, станів доступності й контексту формування ознак. Формалізований опис результату H зберігає зв'язок числових значень із байтовими та просторово-часовими даними й дає змогу відрізнити недоступну перевірку від виконаної перевірки з нульовим результатом.
2. Експериментальне оцінювання проведено на 200 джерельних групах, з яких сформовано 400 PNG- та MP4-об'єктів. Розподіл джерельних груп між навчальною, налаштувальною та перевіркою частинами виконано до створення похідних версій. Це унеможливило потрапляння споріднених об'єктів до різних частин даних і забезпечило групово-незалежне порівняння структурного, спектрально-візуального та спільного подань за однакової моделі логістичної регресії.
3. На внутрішній перевірній частині з 80 об'єктів спільне подання Z_c правильно класифікувало 78 об'єктів і забезпечило Accurasy та Balanced Accurasy 0,9750, Recall 0,9500, F1 0,9744 і ROC-AUC 0,9869. Точний двобічний критерій Мак-Немара з поправкою Голма підтвердив статистично значущу перевагу Z_c над спектрально-візуальним Z_v поданням $p = 5,96 \cdot 10^{-6}$. Відмінність між Z_c і структурним поданням Z_s не досягла рівня статистичної значущості 0,05 $p = 0,0625$.
4. У першій зовнішній серії з 56 MP4-об'єктів внутрішня перевага спільного подання над структурним не відтворилася. Значення Balanced Accurasy становило 0,8333 для Z_s , 0,7857 для Z_c і 0,5000 для Z_v . Отримані результати свідчать про чутливість реалізованих спектрально-візуальних ознак до пристрою, кодека та попереднього опрацювання. Отже, підвищення внутрішньої точності внаслідок поєднання ознак не гарантує збереження цієї переваги за зміни джерела мультимедійних даних.
5. Ресурсне оцінювання показало, що 95 % тривалості опрацювання зображення становив 2,136 мс – для структурного та 853,971 мс – для спільного варіанта, а пікове використання оперативної пам'яті – відповідно 78,77 і 101,53 МБ. Результати підтверджують придатність методу до формування простежуваного й відтворюваного опису контролювано змінених об'єктів, однак не підтверджують виявлення природно поширеного шкідливого коду або готовності до промислового застосування.
6. Подальші дослідження доцільно спрямувати на розширення кількості незалежних джерел і пристроїв, формування повних пооб'єктних протоколів, розвиток формально залежних правил та перевірку повного циклу опрацювання за умов функціонування мережевого ресурсу.

Декларація про використані засоби штучного інтелекту. Під час підготовки цієї роботи автори використовували сервіси ChatGPT та Grammarly для: перевірки граматики та орфографії, перефразування та переформулювання складних речень з текстовими зворота-

ми. Після використання цих сервісів автори проаналізували та відредагували контент майбутньої статті, тобто несуть повну відповідальність за зміст публікації.

References

1. Agarwal, S., & Jung, K.-H. (2024). Digital image steganalysis using entropy driven deep neural network. *Journal of Information Security and Applications*, 84, article ID 103799. <https://doi.org/10.1016/j.jisa.2024.103799>
2. Boroumand, M., Chen, M., & Fridrich, J. (2019). Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5), 1181–1193. <https://doi.org/10.1109/TIFS.2018.2871749>
3. Bouzegza, M., Belatreche, A., Bouridane, A., & Tounsi, M. (2022). A comprehensive review of video steganalysis. *IET Image Processing*, 16(13), 3407–3425. <https://doi.org/10.1049/ipr2.12573>
4. Bravo-Ortiz, M. A., Mercado-Ruiz, E., Villa-Pulgarin, J. P., Hormaza-Cardona, C. A., Quiñones-Arredondo, S., Arteaga-Arteaga, H. B., Orozco-Arias, S., Cardona-Morales, O., & Tabares-Soto, R. (2024). CVTStego-Net: A convolutional vision transformer architecture for spatial image steganalysis. *Journal of Information Security and Applications*, 81, article ID 103695. <https://doi.org/10.1016/j.jisa.2023.103695>
5. Cardenuto, J. P., & Rocha, A. (2022). Benchmarking scientific image forgery detectors. *Science and Engineering Ethics*, 28, article number 35. <https://doi.org/10.1007/s11948-022-00391-4>
6. Cohen, A., Nissim, N., & Elovici, Y. (2020). MalJPEG: Machine learning based solution for the detection of malicious JPEG images. *IEEE Access*, 8, 19997–20011. <https://doi.org/10.1109/ACCESS.2020.2969022>
7. Denysiuk, D., Savenko, O., Lysenko, S., Savenko, B., & Kashtalian, A. (2023). Method for detecting steganographic changes in images using machine learning. In *2023 IEEE 13th International Conference on Dependable Systems, Services and Technologies (DESSERT)* (pp. 1–6). IEEE. <https://doi.org/10.1109/DESSERT61349.2023.10416453>
8. Dong, C., Chen, X., Hu, R., Cao, J., & Li, X. (2023). MVSS-Net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 3539–3553. <https://doi.org/10.1109/TPAMI.2022.3180556>
9. Guilaro, F., Cozzolino, D., Sud, A., Dufour, N., & Verdoliva, L. (2023). TruFor: Leveraging all-round clues for trustworthy image forgery detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20606–20615). IEEE. <https://doi.org/10.1109/CVPR52729.2023.01974>
10. Kiltz, S., Dittmann, J., Loewe, F., Heidecke, C., John, M., Mädel, J., & Preißler, F. (2024). Forensic image trace map for image-stego-malware analysis: Validation of the effectiveness with structured image sets. In *Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security* (pp. 125–130). Association for Computing Machinery. <https://doi.org/10.1145/3658664.3659659>
11. Koch, L., Oesch, S., Sadovnik, A., Weber, B., Chaulagain, A., Dixon, M., Dixon, J., Huettel, M., Watson, C., Hartman, J., & Patulski, R. (2025). On the abuse and detection of polyglot files. In *Proceedings of the ACM Web Conference 2025* (pp. 4810–4822). Association for Computing Machinery. <https://doi.org/10.1145/3696410.3714814>
12. Kwon, M.-J., Yu, I.-J., Nam, S.-H., & Lee, H.-K. (2021). CAT-Net: Compression artifact tracing network for detection and localization of image splicing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 375–384). IEEE. <https://doi.org/10.1109/WACV48630.2021.00042>
13. Moreira, D., Theisen, W., Scheirer, W., Bharati, A., Brogan, J., & Rocha, A. (2022). Image provenance analysis. In H. T. Sencar, L. Verdoliva, & N. Memon (Eds.). *Multimedia Forensics* (pp. 389–432). Springer. https://doi.org/10.1007/978-981-16-7621-5_15

14. Spalazzi, L., Paolanti, M., & Frontoni, E. (2022). An offline parallel architecture for forensic multimedia classification. *Multimedia Tools and Applications*, 81, 22715–22730. <https://doi.org/10.1007/s11042-021-10819-x>
15. Tsafirir, T., Cohen, A., Nir, E., & Nissim, N. (2023). Efficient feature extraction methodologies for unknown MP4-malware detection using machine learning algorithms. *Expert Systems with Applications*, 219, article ID 119615. <https://doi.org/10.1016/j.eswa.2023.119615>

D. O. Denysiuk¹, A. V. Boyarchuk²

¹ Khmelnytskyi National University, Khmelnytskyi, Ukraine

² Tallinn University of Technology, Tallinn, Estonia

IMPROVEMENT OF THE METHOD FOR SPECTRAL-VISUAL AND STRUCTURAL ANALYSIS OF IMAGE AND VIDEO OBJECTS IN WEB SYSTEMS THROUGH FEATURE-PROVENANCE PRESERVATION

A method for the static spectral-visual and structural analysis of image and video objects processed by web systems was improved by preserving the provenance of each feature until the numerical fusion stage. It was established that early fusion of container-level and decoded data complicates the recovery of the byte-level, spatial, or temporal evidence underlying a logistic regression models decision and may cause an unavailable check to be treated as a check with a zero-valued result. Two independent representations of the same immutable object were formed: the structural representation captures the signature, container grammar, physical and logical boundaries, element nesting, service fields, and unassigned byte regions, whereas the spectral-visual representation comprises local statistics, high-frequency residuals, discrete-cosine, wavelet, and temporal features. A traceable description was developed that links numerical representations to byte-level and spatiotemporal locations, an availability mask, a checksum, rule versions, and decoding parameters. A group-disjoint evaluation was conducted using 200 source groups assigned to the training, tuning, and test subsets before derivative objects were generated. On the internal test subset of 80 objects, the combined representation achieved an accuracy of 0.9750, a recall of 0.9500, an F1 score of 0.9744, and a ROC-AUC of 0.9869. The statistical difference between paired decisions was assessed: the exact McNemar test confirmed the advantage of the combined representation over the spectral-visual feature representation but did not confirm its advantage over the structural representation at the 0.05 significance level. In the external series of 56 MP4 objects, balanced accuracy was 0.8333 for the structural representation, 0.7857 for the combined representation, and 0.5000 for the spectral-visual feature representation, indicating that the implemented visual features are sensitive to the source device, codec, and preprocessing. The 95th percentile of processing time was 2.136 ms for the structural variant and 853.971 ms for the combined variant. The obtained results confirm the suitability of the method for producing a reproducible description of objects subjected to controlled modifications but do not establish the detection of naturally occurring malicious code or readiness for industrial deployment.

Keywords: multimedia container integrity; multimedia forensics; check availability mask; group-disjoint evaluation of the static analysis method; localization of structural and spectral-visual anomalies; result reproducibility.