УДК 004.93:004.056 DOI: 10.31673/2412-4338.2026.019009

Лащевська Наталія Олександрівна
к.т.н., доцент, завідувач кафедри Комп'ютерної інженерії
Державний університет інформаційно-телекомунікаційних технологій, Київ Україна
ORCID ID 0000-0003-2148-115X
n.lashchevska@duikt.edu.ua

Мішкур Юрій Валентинович
аспірант кафедри Комп'ютерної інженерії
Державний університет інформаційно-телекомунікаційних технологій, Київ Україна
ORCID ID 0009-0004-6807-6914
y.mishkur@stud.duikt.edu.ua

ВИЯВЛЕННЯ СТЕГАНОГРАФІЇ НА ЗОБРАЖЕННІ З ВИКОРИСТАННЯМ МОДЕЛЕЙ RESNET ТА SRNET

Анотація: *У даній роботі проведено комплексний аналіз ефективності сучасних архітектур глибоких згорткових нейронних мереж — SRNet та ResNetV2 (ResNet50M2, ResNet101V2, ResNet152V2) — у задачах просторового стегоаналізу цифрових зображень. Основну увагу приділено дослідженню ролі блоку попередньої високочастотної фільтрації (блок HPF-фільтрів) та впливу кількості фільтрів та архітектури блоку фільтрації на точність детектування слабких стеганографічних сигналів, внесених методом LSB.*

Для формування навчальних і тестових вибірок використано датасети CIFAR-10 та LabelMe, на основі яких створено штучні набори cover/stego-зображень із застосуванням LSB-стеганографії з підтримкою UTF-8 кодування та контрольованого корисного навантаження.

Встановлено, що канонічна архітектура SRNet потребує інтеграції додаткового HPF-блоку для стабільної збіжності при невеликій кількості епох навчання, демонструючи при цьому високу чутливість до аномалій, але значне споживання оперативної пам'яті, що обмежує розмір навчального датасету в хмарних середовищах. На противагу цьому, моделі на базі ResNetV2 виявилися більш масштабованими та практичними для систем моніторингу в реальному часі, хоча їх ефективність критично залежить від конфігурації багатомасштабного блоку фільтрів (3×3, 5×5, 7×7).

Запропоновано підхід до побудови паралельного багатоканального HPF-модуля з орієнтаційною селективністю, який забезпечує оптимальний баланс між точністю виявлення та обчислювальними витратами.

Дозвіл на навчання високочастотних фільтрів забезпечує додатковий приріст точності, що підтверджує доцільність гібридного підходу, який поєднує апріорні знання цифрової обробки сигналів із глибоким навчанням. Отримані результати можуть бути використані при проєктуванні практичних систем моніторингу цифрового контенту та дозволяють сформулювати рекомендації щодо вибору AI-моделей залежно від прикладного сценарію: від прецизійних дослідницьких інструментів до промислових систем захисту інформації.

Ключові слова: *стегоаналіз, глибоке навчання, згорткові нейронні мережі, ResNetV2, SRNet, високочастотна фільтрація, LSB-стеганографія.*

Lashchevska Nataliia
PhD in Technical Sciences, Associate Professor, Head of the Department of Computer Engineering
State University of Information and Communication Technologies, Kyiv
ORCID 0000-0003-2148-115X
n.lashchevska@duikt.edu.ua

Mishkur Yurii
PhD Student, Department of Computer Engineering
State University of Information and Communication Technologies, Kyiv
ORCID 0009-0004-6807-6914
y.mishkur@stud.duikt.edu.ua

IMAGE STEGANOGRAPHY DETECTION USING RESNET AND SRNET MODELS

Abstract*: This paper presents a comprehensive analysis of the effectiveness of modern deep convolutional neural network architectures—SRNet and ResNetV2 (ResNet50V2, ResNet101V2, ResNet152V2)—in spatial digital image*

steganalysis tasks. Primary attention is devoted to investigating the role of the high-pass filtering preprocessing block (HPF-filter block) and the influence of the filter count and filtering block architecture on the detection accuracy of weak steganographic signals introduced by the LSB method.

The CIFAR-10 and LabelMe datasets were utilized to form training and testing samples, serving as the basis for creating artificial sets of cover/stego images using LSB steganography with UTF-8 encoding support and controlled payload capacity.

It was established that the canonical SRNet architecture requires the integration of an additional HPF block to achieve stable convergence within a small number of training epochs, demonstrating high sensitivity to anomalies but significant RAM consumption, which limits the training dataset size in cloud environments. In contrast, ResNetV2-based models proved to be more scalable and practical for real-time monitoring systems, although their efficiency critically depends on the configuration of a multi-scale filter block (3×3, 5×5, 7×7).

A directional multi-channel HPF module approach with orientational selectivity is proposed, providing an optimal balance between detection accuracy and computational overhead. Permitting the training of high-pass filters provides an additional increase in accuracy, confirming the feasibility of a hybrid approach that combines a priori knowledge of digital signal processing with deep learning. The results obtained can be utilized in the design of practical digital content monitoring systems and allow for the formulation of recommendations for selecting AI models depending on the application scenario: from precision research tools to industrial information security systems.

***Keywords**: steganalysis, deep learning, convolutional neural networks, ResNetV2, SRNet, high-pass filtering, LSB steganography.*

1. Вступ

Швидкий розвиток мультимедійних технологій значно покращив ефективність передачі цифрових зображень, але водночас викликав занепокоєння щодо вразливості даних, несанкціонованих маніпуляцій та порушень конфіденційності. Стегоаналіз є критично важливою областю інформаційної безпеки, що займається виявленням прихованої інформації в цифрових носіях. З розвитком стеганографічних методів, які використовують складні алгоритми вбудовування даних, традиційні методи стегоаналізу стають менш ефективними. Це викликає необхідність розробки нових підходів, здатних виявляти сучасні стеганографічні методи [1].

Глибоке навчання революціонізувало багато областей комп'ютерного зору, включаючи стегоаналіз. Зокрема, згорткові нейронні мережі (CNN) демонструють високу ефективність у виявленні тонких артефактів, що залишаються після вбудовування стеганографічних повідомлень [2]. Однак, навчання глибоких мереж для стегоаналізу є складним завданням через малий рівень сигналу стеганографічних змін.

Сучасні методології стегоаналізу в основному поділяються на дві технічні парадигми: традиційні підходи на основі статистичних ознак та сучасні архітектури, засновані на глибокому навчанні [3].

Архітектура ResNet, запропонована в [4], вирішує проблему деградації градієнтів у глибоких мережах через використання залишкових з'єднань. Це дозволяє навчати дуже глибокі мережі, які можуть виявляти складні патерни в даних. У контексті стегоаналізу, ResNet може ефективно виявляти тонкі статистичні аномалії, викликані вбудовуванням прихованої інформації.

Сучасні методи цифрової стеганографії постійно еволюціонують, що вимагає розробки ефективних стегоаналітичних систем для виявлення прихованої інформації. Згорткові нейронні мережі (CNN) продемонстрували високу ефективність у задачах стегоаналізу, проте вибір оптимальної архітектури залишається актуальною проблемою.

Існуючі дослідження зосереджуються переважно на максимізації точності виявлення, часто ігноруючи практичні аспекти впровадження, такі як швидкість навчання, обчислювальні витрати та стабільність роботи в реальних умовах. Критично важливим є знаходження оптимального балансу між кількістю високочастотних фільтрів у блоці попередньої обробки, розміром навчального датасету та вибором архітектури для конкретних практичних застосувань.

2. Аналіз літературних даних і постановка проблеми

Традиційний стегоаналіз тривалий час базувався на вилученні статистичних ознак вручну [3]. Для вилучення статистичних ознак (залишків шуму) із цифрових зображень в [3] було створено набір спеціалізованих ядер згортки - SRM-фільтри (Spatial Rich Model filters). Вони є фундаментом класичного стегоаналізу та широко використовуються як початкові шари у вигляді блок високочастотних фільтрів (High-Pass Filter - HPF) при побудові глибоких нейронних мереж.

Прорив у глибокому навчанні (DL) дозволив автоматизувати цей процес, проте стандартні архітектури комп'ютерного зору виявилися малоефективними для стегоаналізу без спеціальної адаптації. Але є й проблема, яка полягає в тому, що стандартні CNN налаштовані на придушення шуму для розпізнавання об’єктів (контенту), тоді як у стегоаналізі шум і є корисним сигналом. Тому

стегоаналіз на основі CNN вимагає ретельного проектування шарів попередньої обробки з ініціалізацією фільтрів для отримання хорошої продуктивності.

Однією з перших успішних моделей стала мережа XuNet [5]. Архітектура YeNet [6] впровадила використання 30 різних фільтрів SRM як початкового шару. Вона продемонструвала, що поєднання експертних знань про цифрову обробку сигналів із глибокими згортковими мережами дозволяє ефективно виявляти складні алгоритми вбудовування типу WOW та S-UNIWARD. Інші сучасні моделі, як-от Zhu-Net [7], використовують ідеї роздільних згорток (Depthwise Separable Convolutions) для зменшення кількості параметрів та підвищення ефективності.

Архітектура SRNet [8] вважається "золотим стандартом" сучасного стегоаналізу. Це повністю наскрізна (end-to-end) архітектура, яка не вимагає фіксованих фільтрів. Її основні характеристики:

- складається з 12 блоків. Перші два шари не мають пулінгу, щоб зберегти просторові характеристики стегосигналу в піксельному просторі;
- використовує пропускні з'єднання (як у ResNet), що дозволяє мережі вивчати залишкові функції високого порядку.

На думку авторів [8], архітектура SRNet забезпечує автоматичне вивчення ознак без ручного ініціювання вхідних високочастотних фільтрів, вона ефективна для виявлення стеганографії, вбудованої за різними алгоритмами (WOW, HILL, S-UNIWARD тощо). Але через відсутність пулінгу в перших 7 шарах, мережа змушена обробляти тензори великої розмірності на значній глибині. Це призводить до величезного споживання відеопам'яті і низької швидкості обробки одного зображення порівняно з ResNet чи MobileNet. Крім того, мережа SRNet демонструє чудові результати в межах одного датасету (наприклад, BOSSBase), але її точність помітно падає при тестуванні на зображеннях з іншими характеристиками шуму сенсора або іншою ISO-чутливістю [9].

Можливим варіантом побудови надійного і потужного стегоаналізатору є гібридний підхід, коли комбінують попередню обробку через набір фіксованих високочастотних фільтрів зі стандартними класифікаторами сучасних CNN, таких як ResNetv2, які зазвичай застосовуються в задачах класифікації зображень. Варіанти архітектури ResNet відрізняються кількістю залишкових блоків: ResNet50V2 (50 шарів, 25.6 млн параметрів), ResNet101V2 (101 шар, 44.6 млн параметрів), ResNet152V2 (152 шари, 60.3 млн параметрів) [10].

Модель на основі ResNet демонструє високу продуктивність у виявленні прихованої інформації в мультимедійних даних [11]. Дослідження [12] продемонструвало успіх попередньо навчених моделей ResNet, DenseNet та Inception у сценарії невідповідності cover-stego для кожного методу приховування з різними корисними навантаженнями.

Для адаптації ResNetV2 до задачі стегоаналізу критично важливим є інтеграція вхідного блоку високочастотних фільтрів перед основною архітектурою. Зазвичай використовується набір з 30 фіксованих SRM фільтрів розміром 5×5, що виділяють високочастотні артефакти стеганографії, з наступним застосуванням Truncated Linear Unit (TLU) для обмеження динамічного діапазону активацій. Архітектура ResNetV2 з HPF-фільтрами демонструє конкурентоспроможні результати для виявлення сучасних адаптивних методів стеганографії (WOW, S-UNIWARD, HILL), досягаючи точність 94-96% залежно від архітектури моделі та складності задачі [13].

Без такої препроцесингової фільтрації, мережа схильна навчатися на семантичному змісті зображень замість на слабких стеганографічних сигналах [14-15]. Критично важливою є регуляризація для запобігання переобладнанню на специфічних артефактах тренувальних даних [16].

Але питання побудови найкращої архітектури блоку фільтрів попередньої обробки залишається дискусійним.

Кількість високочастотних фільтрів у блоці попередньої обробки CNN-стегоаналізатора істотно впливає на точність, потенціал перенавчання та обчислювальні витрати [17]. Немає теоретичного обґрунтування чому саме 30 SRM фільтрів розміру 5×5 є оптимальними. Деякі дослідження показують, що 45-60 фільтрів можуть дати невелике покращення (+0.5-1%), але це емпіричні результати без глибокого розуміння [18]. Залишається відкритим питання: чи можна систематично визначити мінімальний достатній набір фільтрів?

За даними [19], використання кількох фільтрів може генерувати надлишкові залишкові зображення та надлишкові ознаки. Надлишкові ознаки не лише споживають обчислювальні ресурси та час, але й призводять до переналаштування моделі, що знижує точність виявлення моделі.

Крім того, залишається невирішеним питання вибору фіксованих або фільтрів, що навчаються, в блоці попередньої обробки. Гібридні підходи (частина фіксованих + частина навчуваних) теоретично привабливі, але практичні результати неоднозначні [20].

Таким чином, кількість високочастотних фільтрів у блоці попередньої обробки CNN-стегоаналізатора істотно впливає на точність, потенціал перенавчання та обчислювальні витрати. Більший набір високочастотних фільтрів зазвичай покращує розпізнавання залишкових шумових ознак та підвищує точність, але може збільшувати ризик перенавчання та вимагає значно більших ресурсів. Оптимальний баланс між кількістю фільтрів, розміром датасету та архітектурою (SRNet vs ResNetv2) дозволяє досягати високих показників з мінімальними витратами. Архітектура SRNet є потужним інструментом для досліджень адаптивної стеганографії, проте для реальних систем моніторингу архітектури на базі ResNetV2 з блоком високочастотних HPF-фільтрів може виявитись більш практичним рішенням через швидкість та стабільність.

3. Мета і задачі дослідження

Метою даного дослідження є комплексний аналіз впливу архітектурних рішень та параметрів попередньої високочастотної фільтрації на ефективність глибокого стегоаналізу, а саме — встановлення оптимального балансу між кількістю HPF-фільтрів, розміром навчального датасету та типом архітектури (SRNet або ResNetv2) з точки зору точності виявлення, узагальнювальної здатності та обчислювальних витрат.

Особливу увагу приділено порівнянню спеціалізованої архітектури SRNet, орієнтованої на дослідження адаптивної стеганографії, та модифікованих архітектур на базі ResNetv2 з інтегрованим HPF-блоком, які розглядаються як потенційно більш практичне рішення для реальних систем моніторингу цифрового контенту, де критичними є швидкодія та стабільність.

Для досягнення поставленої мети у роботі необхідно розв’язати такі науково-практичні завдання:

1. Проаналізувати роль високочастотної попередньої фільтрації (HPF/SRM) у задачах просторового стегоаналізу та визначити її вплив на виділення слабких стеганографічних ознак у глибоких згорткових мережах.
2. Дослідити вплив кількості HPF-фільтрів на точність класифікації cover/stego зображень для архітектур SRNet і ResNetv2, зокрема в умовах різного обсягу навчальних даних.
3. Порівняти узагальнювальну здатність моделей SRNet та ResNetv2+HPF, оцінюючи ризик перенавчання при зміні складності архітектури та розміру датасету.
4. Оцінити обчислювальні витрати (кількість параметрів, час навчання, швидкість надання висновку, споживання пам’яті) для обох архітектур при різній конфігурації HPF-блоку.
5. Проаналізувати стабільність навчання моделей (збіжність, чутливість до ініціалізації та параметрів оптимізації) залежно від архітектури та кількості фільтрів у попередньому шарі.
6. Визначити доцільність використання SRNet як інструменту для досліджень адаптивної стеганографії, де пріоритетом є максимальна чутливість до слабких змін у зображеннях.
7. Оцінити практичність архітектур на базі ResNetv2 з HPF-блоком для задач реального часу та систем моніторингу цифрового контенту з обмеженими обчислювальними ресурсами.
8. Сформулювати рекомендації щодо вибору архітектури та конфігурації HPF-блоку залежно від прикладного сценарію (дослідницькі експерименти vs промислові системи виявлення стеганографії).

Запропонований підхід дозволяє не лише кількісно оцінити ефективність сучасних архітектур глибокого стегоаналізу, але й сформувати практичні рекомендації щодо їх застосування. Отримані результати сприяють кращому розумінню компромісу між точністю виявлення та обчислювальними витратами, що є критично важливим для впровадження систем стегоаналізу в реальних умовах.

4 Методика дослідження

4.1 Побудова набору даних для навчання і перевірки моделі

Для вбудовування текстових повідомлень було використано відомий набір даних CIFAR10 [21]. CIFAR-10 – це широко використовуваний датасет у комп'ютерному зорі, який містить 60 000 кольорових зображень розміром 32x32 пікселів у 10 класах.

На думку [22] CIFAR-10 є хорошим джерелом зображень для побудови великого навчального набору (120 000 фотографій), який було використано для порівняння ефективності різних AI/ML моделей у виявленні прихованих повідомлень.

Інший варіант джерела зображень - набір даних LabelMe-12-50k, який складається з 50 000 зображень JPEG (40 000 для навчання та 10 000 для тестування) [23]. Кожне зображення має розмір 256x256 пікселів. 50% зображень у навчальному та тестовому наборі показують центрований об'єкт, кожне з яких належить до одного з 12 класів об'єктів. Для використання ображення перетворювались на розмір 96x96.

Для додавання прихованого напису було використано техніку LSB. Стеганографія LSB – це підхід до приховування повідомлень, який безпосередньо змінює біти, найменш значущі для кольору пікселя: останній(і) біт(и) [24]. Точніше, він замінює значення існуючих бітів двійковим значенням повідомлення. Підхід LSB є найбільш традиційним та найпростішим у реалізації стеганографічним підходом. Хоча простий LSB-метод є легко виявним, він має і деякі переваги для створення датасетів і перевірки роботи моделей.

4.2 Побудова блоку фільтрів

Високочастотні фільтри (HPF) відіграють центральну роль у більшості сучасних систем стегоаналізу, заснованих на глибокому навчанні (CNN). Вони є необхідною передумовою для успішного виявлення стегосигналу. Основна мета HPF – виділення залишкового шуму зображення, в якому містяться слабкі статистичні аномалії, внесені стеганографічним вбудовуванням (наприклад, LSB).

Стеганографічні зміни в зображеннях часто проявляються у високочастотних компонентах. Блок HPF-фільтрів виконує попередню обробку зображення для підсилення цих компонентів перед подачею в ResNet.

HPF блок складається з набору фільтрів, включаючи [3, 14, 17]:

- фільтр SRM (Spatial Rich Model): Набір з 30 високочастотних фільтрів, що виявляють різні типи локальних змін;
- фільтр Лапласа: Виявляє різкі зміни інтенсивності пікселів;
- високочастотні фільтри Габора: Виявляють орієнтовані високочастотні компоненти.

Математична операція фільтрації для кожного фільтра k може бути представлена як [3, 14, 17]:

$$F_k(x,y) = \sum \quad \sum \quad I(x+i, y+j) \cdot H_k(i,j) \tag{1}$$

де I - вхідне зображення, H_k - ядро k-го фільтра, F_k - результат фільтрації.

Ключова вимога до HPF-ядра H у стегоаналізі полягає в тому, що сума його коефіцієнтів має дорівнювати нулю $\left(\sum_{i,j} \quad H_k(i,j) = 0\right)$. Це гарантує, що фільтр усуває низькочастотні, контентні компоненти (плавні області), максимізуючи при цьому контрастність між природним шумом та стего-шумом [1, 3, 8].

В перших роботах, присвячених побудові систем стегоаналізу, було використано фіксовані (нетреновані) HPF-шари, що передують основній CNN-архітектурі. Це забезпечує стабільний і стандартизований вхід у вигляді залишкового зображення [3, 8].

Сучасні архітектури для посилення виділення стегосигналу використовують багатошаровий вхідний HPF-блок [20, 25-26]. Перший шар може містити декілька спрямованих ядер (наприклад, горизонтальні, вертикальні, діагональні SRM-фільтри). Наступні шари можуть бути як фіксованими, так і навчальними.

Перший варіант (досить простий) побудови вхідного блока фільтрів містив декілька фільтрів 5x5 з орієнтаційною селективністю та фільтр Лапласа.

Другий варіант - багатомасштабний високочастотний блок з декількома шарами фільтрів.

Запропонований блок поєднує фільтри різного просторового масштабу (3×3, 5×5, 7×7) з орієнтаційною селективністю, що дозволяє одночасно враховувати локальні, контекстні та багатомасштабні стеганографічні ознаки. Така структура забезпечує підвищену чутливість до слабких змін яскравості, характерних для сучасних адаптивних методів вбудовування.

Високочастотний HPF-блок, що розглянуто, було побудовано як паралельний багатоканальний фільтраційний модуль з орієнтаційною (напрямною та кутовою) селективністю:

$$R \;=\; [H_{3x3}(X) \,||\, H_{5x5}(X) \,||\, H_{7x7}(X)] \tag{2}$$

де $\boldsymbol{X}$ — вхідне зображення, H_{kxk} — банк високочастотних фільтрів розміру (k×k), || — конкатенація по каналах.

Узагальнена композиція одного HPF-блоку містила 16–32 фільтри (від 4 до 8 фільтрів 3x3, від 8 до 12 фільтрів 5x5, від 4 до 8 фільтрів 7x7)

Для обох варіантів вхідних блоків була передбачена можливість обирати фіксовані фільтри або фільтри, які навчаються.

4.3 Проведення обчислювальних експериментів

Для проведення експериментів були використані моделі наступної структури (рис. 1):

- Блок попередньої обробки (один з 2 варіантів);
- Конволюційна нейронна мережа (SRNet, ResNet50v2, ResNet101v2, ResNet152v2);
- Шар GlobalAveragePooling і щільний шар з активацією «сігмоїд»;
- Блок виводу ілюстрацій і перевірки відновлення вбудованого тексту.

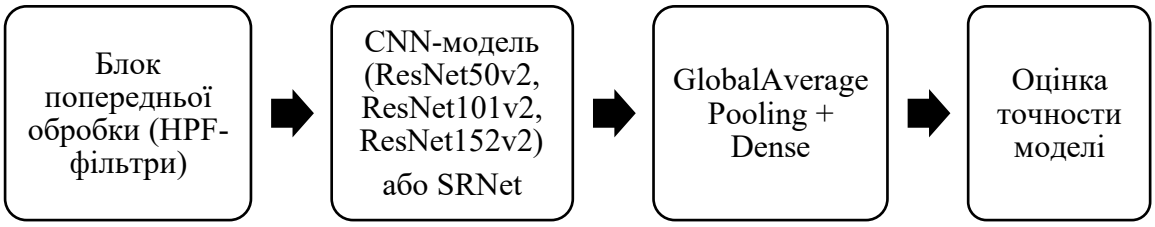

Рис. 1 Архітектура моделі стегоаналізу з використанням глибоких CNN із залишковими блоками

Всі експерименти виконувались в середовищі Google Collaboratory з використанням графічного прискорювача T4. Використовувалась мова програмування python, для побудови нейромережевих моделей був використаний пакет tensorflow з інтерфейсом keras. Також було використано деякі модулі пакету scikit-learn.

Зображення з розглянутих датасетів (Cifar10 32x32x3 або інши варианти) перед вбудовуванням прихованого тексту перетворювались на зображення 96x96x3. Для забезпечення контрольованих умов дослідження було сформовано декілька варіантів штучного датасета, які містили від 3000 до 60000 зображень, з яких: 50% — cover-зображення (без змін), 50% — stego-зображення (з вбудованим повідомленням). Розглядалися різні значення обсягу вбудовування (payload), що вимірювався в бітах на піксель (bpp).

Для вбудовування використовувались текстові повідомлення як англійською, так і українською мовою, використовувалось кодування utf-8 (це було враховано при побудові послідовності бітів).

Навчання проводилося з використанням оптимізатора Adam з регульованою початковою швидкістю навчання (в більшості експериментів 0.0001)та функції втрат binary cross-entropy.

5 Результати дослідження

У цьому дослідженні представлені результати, отримані в результаті тестування різних комбінацій фільтрів у блоці попередньої обробки, декількох варіантів архітектури глибоких конволюційних мереж з наявністю залишкових блоків (SRNet або ResNetv2), застосованих до стегоаналізу зображень у просторовій області.

У роботі реалізовано канонічну архітектуру SRNet із 11 згорткових. Вбудовування інформації здійснювалося методом LSB із підтримкою UTF-8 кодування та керованою пропускною здатністю (bpp). Якість стегоаналізу оцінюється за допомогою ROC-кривої та AUC, а коректність стеганографічного каналу підтверджується відновленням вбудованого тексту.

Спроби побудувати модель для виявлення прихованого тексту на зображенні (використовувалось LSB-вбудовування) на основі лише архітектури SRNet виявилася невдалими при кількості епох навчання в інтервалі 10-12. Але після додавання до архітектури моделі блоку HPF-фільтрів з чотирьох шарів (три універсальні фільтри 3x3 і один 5x5) швидкість і якість навчання виявилися досить високими в широкому інтервалі відносних ємностей вбудовування (payload змінювалось від 0,002 до 0,4). Ілюстрації ходу навчання та кривої ROC/AUC наведена на рис. 2 і рис.3.

Встановлено, що SRNet з невеличким блоком попередньої обробки досить швидко навчається і забезпечує високу точність моделі. Але кількість зображень в наборі даних суттєво обмежена жорсткими вимогами до об'єму оперативної пам'яті. В безкоштовній версії Google Colab модель з архітектурою SRNet вдалося навчити за набором даних, який містив до 9000 зображень. Збільшення кількості зображень в навчальній вибірці до 10000 призводило до аварийного припинення обчислень внаслідок цілковитого заповнення оперативної пам’яти.

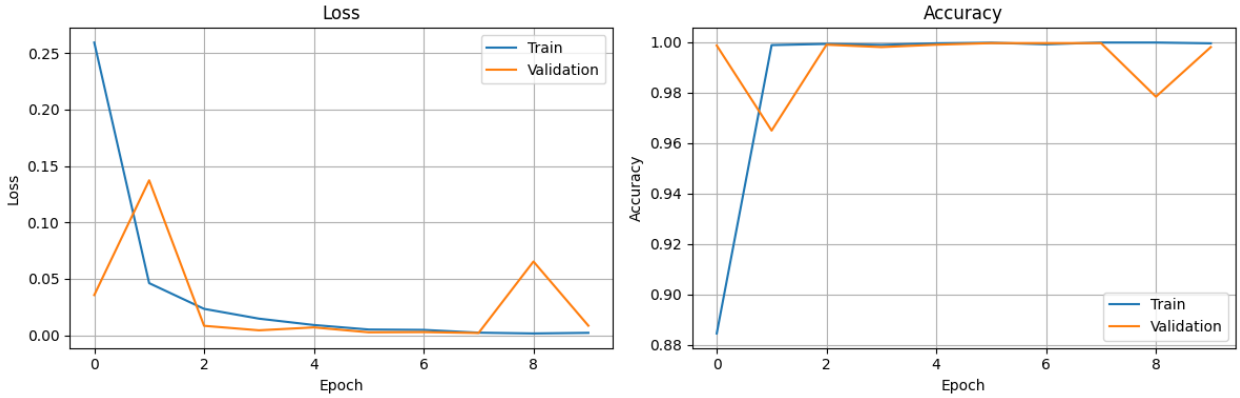

Рис. 2 Криві навчання моделі стегоаналізу з архітектурою SRNet за даними з payload=0,002.

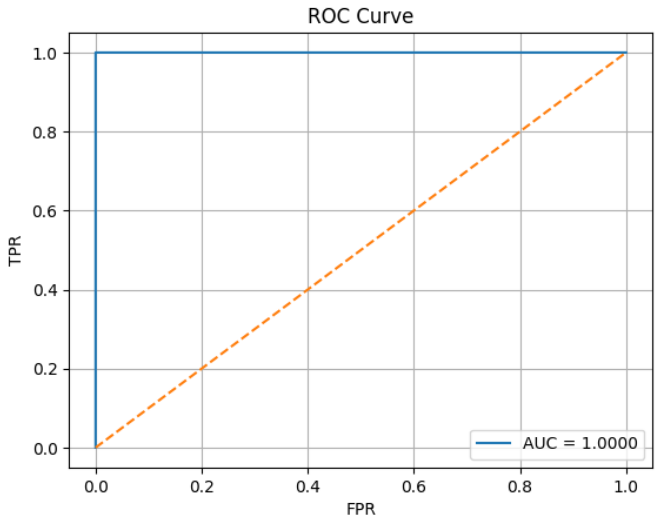

Рис.3 Крива ROC/AUC моделі стегоаналізу з архітектурою SRNet, яку було навчено за даними з payload=0,002

Моделі стегоаналізу з архітектурою ResNetv2 значно більш чутливі до характеристик блоку попередньої фільтрації, ніж моделі з архітектурою SRNet. Найкращі результати було отримано з використанням багатомасштабного фільтру.

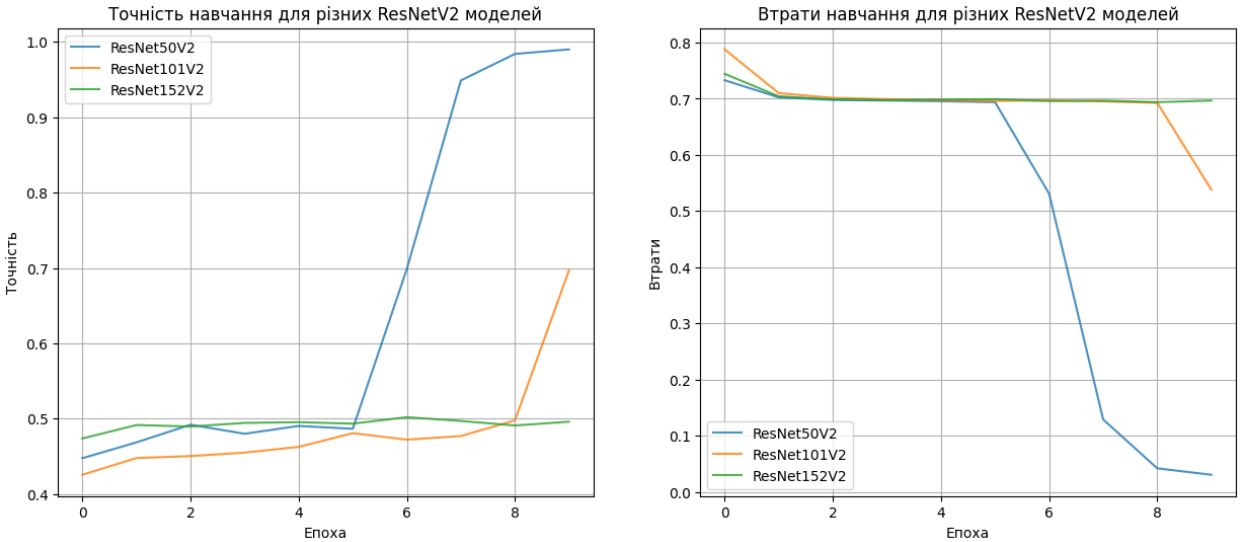

Рис. 4 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 5 шарами фільтрів 5x5 при навчанні за набором даних з payload=0,002

Попередне дослідження було виконано з використанням декількох груп HPF-фільтрів. Кожна група містила 9 орієнтованих фільтрів 5x5, кількість груп змінювалась від 1 до 9. Отримані результати досить неоднозначні, тому що не вдалося виявити систематичний вплив кількості груп фільтрів на

точність навчання моделі. Для досить високих значень payload (більш або дорівнює 0,2) одного блока фільтрів досить для стійкого виявлення прихованого вмісту незалежно від глибини використаної архітектури. На низьких payload, які зазвичай ускладнюють виявлення стеганографії, результат виявився неоднозначним (див. рис. 4-5).

Як видно з рис.4, збільшення глибини моделі не надає переваги в точності і надійності виявлення стеганографії. Збільшення кількості блоків фільтрів (9 фільтрів в блоці, які розраховано на виявлення особливостей за геометричними напрямками) з 1 до 10 не надало систематичного покращення точності навчання моделі. Але час навчання моделей послідовно збільшувався при переході від ResNet50v2 до ResNet101v2 і потім ResNet152v2. Більш високі значення payload (bpp=0,2 або bpp=0,4) значно спрощують виявлення стеганографії і зменшують вимого до вхідних фільтрів.

При навчанні моделі за набором даних з однаковою кількістю зображень встановлено, що час навчання моделі з архітектурою ResNet50v2 виявився таким же, як і для моделі SRNet. Для моделі з архітектурою ResNet101v2 час навчання моделі на чверть перевищував час навчання SRNet. Для моделі з архітектурою ResNet152v2 час навчання моделі більш ніж вдвічі перевищував час навчання SRNet.

Значно більш систематичний результат отримано з використанням багатомасштабних фільтрів. Але при випробуванні цих комплексних фільтрів встановлено, що один блок фільтрів не за безпечує повного виявлення ознак стеганографічного вбудовування і точність моделі не перевищувала 60-65% незалежно від архітектури моделі.

Послідовне використання двох мультимасштабних фільтрів забезпечило успішне навчання моделей з усіма варіантами архітектури ResNet (рис. 6). Потрійне використання блоку мультімасштабних фільтрів забезпечило швидке і надійне навчання моделі (рис. 7).

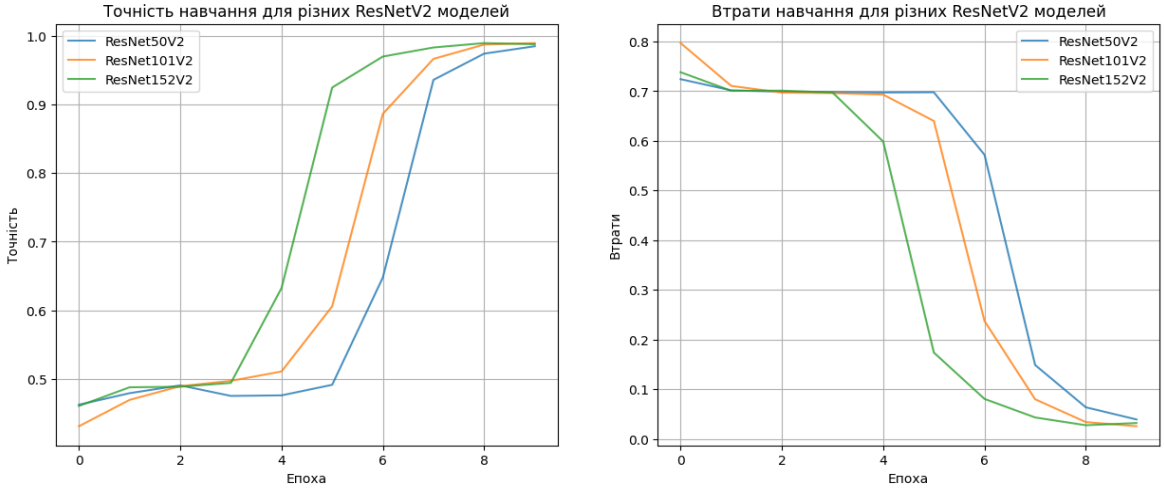

Рис. 5 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 3 шарами фільтрів 5x5 при навчанні за набором даних з payload=0,002

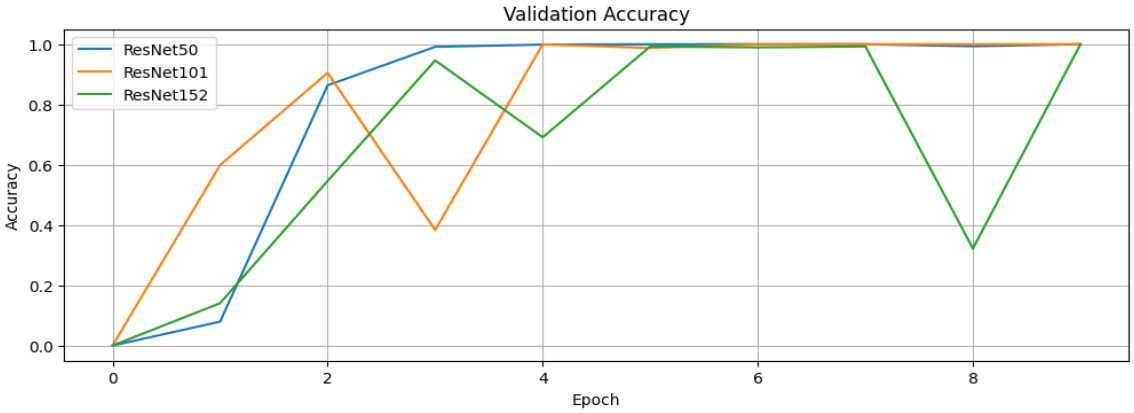

Рис. 6 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 2 шарами мультимасштабних фільтрів при навчанні за набором даних з payload=0,002

Час навчання моделі з архітектурою ResNet50v2 і подвійним мультимасштабним фільтром з тренуванням його шарів виявився на 10% менше у порівняння з SRNet, час навчання моделі з потрійним мультимасштабним фільтром виявився приблизно на 25% віще. Для інших варіантов архітектури ResNet з більшою глибиною час навчання виявився помітно виявився більшим для усіх варіантів мультимасштабного фільтру.

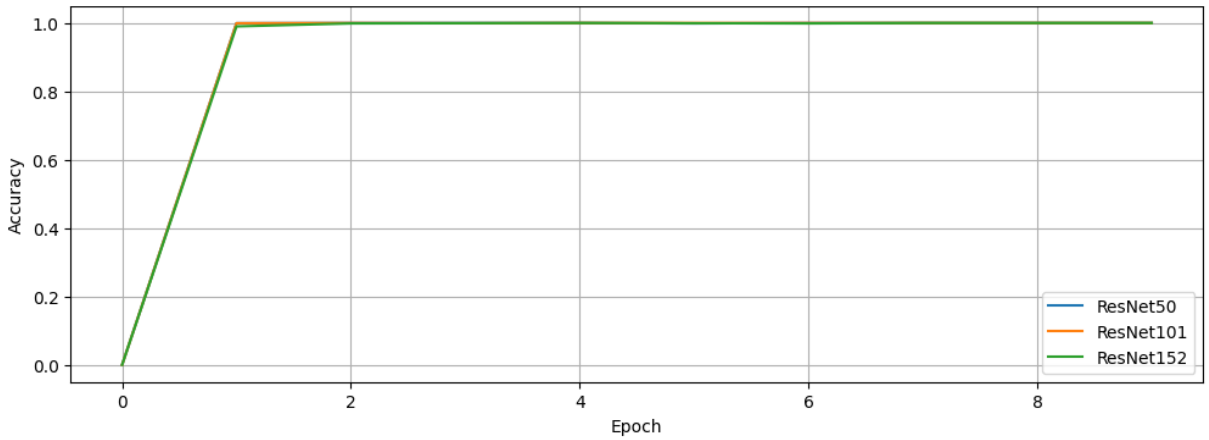

Рис. 7 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 3 шарами мультимасштабних фільтрів при навчанні за набором даних з payload=0,002

Порівняння результатів для моделей з фіксованими фільтрами та моделей з можливостю навчання фільтрів показало, що дозвіл на адаптацію HPF-ядер забезпечує приріст якості на 8-12%, що свідчить про доцільність поєднання апріорних знань із глибоким навчанням.

Використання переднавчених ваг ImageNet дозволило стабілізувати і прискорити процес навчання глибоких архітектур ResNetV2. Для адаптації ваг, орієнтованих на розпізнавання об'єктів, до задачі стегоаналізу, було застосовано повне розморожування шарів (fine-tuning) разом із попередньою високочастотною фільтрацією вхідних даних. Навчання моделі без попередньо наданих ваг могло взагалі не дати позитивного результату виявлення стеганографії при низьких payload.

Таким чином, експериментальні результати демонструють, що використання високочастотного препроцесінгу у поєднанні з ResNetV2 суттєво покращує ефективність стегоаналізу. Зокрема, HPF-фільтри з можливостю тренування дозволяють моделі адаптуватися до характеру стего-шуму, що забезпечує приріст точності класифікації до 99,5-99,8% та значення AUC біля 1,0.

При зміні архітектури класифікатору на модель SRNet встановлено, що наявність блоку попередньої обробки забезпечує надійне виявлення прихованого тексту. Ця спеціалізована архітектура забезпечує досить швидке навчання моделі, але пред'являє жорсткі вимоги до обчислювальних ресурсів (в першу чергу пам’яти).

Висновки

1. Підтверджено, що високочастотна попередня фільтрація є критично важливим компонентом CNN-стегоаналізаторів, оскільки дозволяє ефективно пригнічувати семантичний вміст зображень та підсилювати слабкі стеганографічні ознаки.
2. Встановлено, що спеціалізована архітектура SRNet демонструє високу точність та швидку збіжність навчання, однак характеризується значними вимогами до оперативної пам’яті, що обмежує її практичне застосування на великих наборах даних.
3. Показано, що архітектури ResNetv2 без HPF-блоку є малоефективними для просторового стегоаналізу, проте інтеграція високочастотної попередньої обробки суттєво підвищує їхню здатність до виявлення прихованих повідомлень.
4. Експериментально доведено, що просте збільшення кількості орієнтованих фільтрів або глибини ResNetv2 не гарантує покращення точності, особливо при малих значеннях payload, тоді як багатомасштабні HPF-блоки забезпечують більш стабільний і відтворюваний результат.
5. Виявлено, що послідовне використання двох або трьох мультимасштабних HPF-блоків дозволяє досягти високої точності класифікації (до 99,5–99,8%) та значень AUC, близьких до 1,0, навіть для складних сценаріїв з низьким рівнем вбудовування.
6. Показано, що навчувані HPF-фільтри забезпечують приріст якості порівняно з повністю фіксованими фільтрами, що підтверджує доцільність гібридних підходів у блоці попередньої обробки.
7. Встановлено, що ResNet50v2 з оптимально сконфігурованим HPF-блоком є найбільш збалансованим варіантом з точки зору точності, швидкості навчання та обчислювальних витрат, тоді як глибші моделі (ResNet101v2, ResNet152v2) не демонструють суттєвих переваг.
8. Отримані результати дозволяють сформулювати практичні рекомендації щодо вибору архітектури та конфігурації HPF-блоку залежно від прикладного сценарію: SRNet доцільно

використовувати для дослідницьких задач, тоді як ResNetv2 з багатомасштабною попередньою фільтрацією є більш придатним для реальних систем виявлення стеганографії.

Внесок авторів:

Наталія Лащевська – концептуалізація, методика дослідження, формулювання наукової проблеми, аналіз результатів, наукове редагування тексту;

Юрій Мішкур – програмне забезпечення, збір і перевірка емпіричних даних, проведення обчислювальних експериментів, емпіричне дослідження, аналіз джерел, підготовка огляду літератури, підготовка первинної версії рукопису.

Декларація про використання штучного інтелекту.

Під час підготовки цієї статті інструменти штучного інтелекту не використовувалися ані для отримання наукових результатів, ані для аналізу даних, написання основного змісту статті чи формулювання висновків.

Конфлікт інтересів.

Автори заявляють про відсутність конфлікту інтересів. Під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б вплинути на результати дослідження або їх інтерпретацію. Робота виконана з дотриманням принципів академічної доброчесності, етичних норм проведення наукових досліджень і вимог редакційної політики щодо запобігання конфлікту інтересів.

8. Список використаної літератури

1. Fridrich J. Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, 2009. 437 p.

2. Xu G., Wu H.Z., Shi Y.Q. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 2016, vol. 23, no. 5, pp. 708-712. DOI: 10.1109/LSP.2016.2548421

3. Fridrich J., Kodovsky J. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 2012, vol. 7, no. 3, pp. 868-882. DOI: 10.1109/TIFS.2012.2190402

4. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90

5. Xu G. Deep convolutional neural network to detect J-UNIWARD. In: *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, 2017, pp. 67-73. DOI: 10.1145/3082031.3083236

6. Ye J., Ni J., Yi Y. Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2017, vol. 12, no. 11, pp. 2545-2557. DOI: 10.1109/TIFS.2017.2710946

7. Zhang R., Zhu F., Liu J., Liu G. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. *IEEE Transactions on Information Forensics and Security*, 2020, vol. 15, pp. 1138-1150. DOI: 10.1109/TIFS.2019.2936913

8. Boroumand M., Chen M., Fridrich J. Deep residual network for steganalysis of digital images. *IEEE Transactions on* Information *Forensics and Security*, 2019, vol. 14, no. 5, pp. 1181-1193. DOI: 10.1109/TIFS.2018.2871749

9. Tabares-Soto R., Arteaga-Arteaga H.B., Buritica O.M.A., et al. Deep learning for steganalysis: evaluating model robustness against image transformations. *Frontiers in Artificial Intelligence*, 2025, vol. 8, article 1532895. DOI: 10.3389/frai.2025.1532895

10. Veit A., Wilber M.J., Belongie S. Residual networks behave like ensembles of relatively shallow networks. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, vol. 29, pp. 550-558.

11. Ntivuguruzwa J.-P., Kurundayev M., Ullah M., et al. A convolutional neural network to detect possible hidden data in spatial domain images. *Cybersecurity*, 2023, vol. 6, article 32. DOI: 10.1186/s42400-023-00156-x

12. Ozcan S., Mustacoglu A.F. Transfer learning effects on image steganalysis with pre-trained deep residual neural network model. In: *IEEE International Conference on Big Data*, 2018, pp. 2280-2287. DOI: 10.1109/BigData.2018.8622437

13. Yedroudj M., Comby F., Chaumont M. Yedroudj-Net: An efficient CNN for spatial steganalysis. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2092-2096. DOI: 10.1109/ICASSP.2018.8461438

14. Dwaik A., Nandi A.K., Naous T., Alani S., Alsarhan A. Enhancing the performance of convolutional neural network image-based steganalysis in spatial domain using Spatial Rich Model and 2D Gabor filters. *Journal of Information Security and Applications*, 2024, vol. 85, article 103862. DOI: 10.1016/j.jisa.2024.103862

15. Wang Z., Gao N., Wang X., Qu X., Li L., Zhang X. Deep learning for steganalysis of diverse data types: A review of methods, taxonomy, challenges and future directions. *arXiv preprint arXiv:2308.04522*, 2024.

16. Tabares-Soto R., Arteaga-Arteaga H.B., Orozco-Arias S., et al. Strategy to improve the accuracy of convolutional neural network architectures applied to digital image steganalysis in the spatial domain. *PeerJ Computer Science*, 2021, vol. 7, e451. DOI: 10.7717/peerj-cs.451

17. Liu F., Zhou X., Yan X., Peng J., Hu Y., Chen Q. Preprocessing enhancement method for spatial domain steganalysis. *Mathematics*, 2022, vol. 10, no. 21, article 3936. DOI: 10.3390/math10213936

18. Wang X., Liao D., Dai Y., Li H. A steganalysis framework based on CNN using the filter subset selection method. *Multimedia Tools and Applications*, 2020, vol. 79, pp. 21307-21326. DOI: 10.1007/s11042-020-08831-8

19. Jin Z., Yang Y., Chen Y., Chen Y. IAS-CNN: Image adaptive steganalysis via convolutional neural network combined with selection channel. *Journal of Sensors*, 2020, article 1550147720911002. DOI: 10.1177/1550147720911002

20. Chaumont, Marc. (2020). Deep learning in steganography and steganalysis. 10.1016/B978-0-12-819438-6.00022-0. In book: Digital Media Steganography (pp.321-349)

21. Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

22. Stefanek G., Gulbransen L., Spink G., Morawski J., Filla D., Rabello De Castro R. A comparison of ai models to detect hidden messages in images. (2024). Issues In Information Systems. 119-132. DOI: 10.48009/3_iis_2024_110

23. The LabelMe-12-50k dataset. URL: https://www.ais.uni-bonn.de/download/datasets.html

24. Meike Helena Kombrink, Zeno Jean Marius Hubert Geradts, and Marcel Worring. 2024. Image Steganography Approaches and Their Detection Strategies: A Survey. ACM Comput. Surv. 57, 2, Article 33 (February 2025), 40 pages. DOI: 10.1145/3694965

25. S. Wu, S. -h. Zhong and Y. Liu, "A Novel Convolutional Neural Network for Image Steganalysis With Shared Normalization," in IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 256-270, Jan. 2020, DOI: 10.1109/TMM.2019.2920605.

26. Dwaik, A., & Belkhouche, Y. (2024). Enhancing the performance of convolutional neural network image-based steganalysis in spatial domain using Spatial Rich Model and 2D Gabor filters. Journal of Information Security and Applications, 85, 103864. DOI: 10.1016/j.jisa.2024.103864

9. References

1. Fridrich J. Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, 2009. 437 p.

2. Xu G., Wu H.Z., Shi Y.Q. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 2016, vol. 23, no. 5, pp. 708-712. DOI: 10.1109/LSP.2016.2548421

3. Fridrich J., Kodovsky J. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 2012, vol. 7, no. 3, pp. 868-882. DOI: 10.1109/TIFS.2012.2190402

4. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90

5. Xu G. Deep convolutional neural network to detect J-UNIWARD. In: *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, 2017, pp. 67-73. DOI: 10.1145/3082031.3083236

6. Ye J., Ni J., Yi Y. Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2017, vol. 12, no. 11, pp. 2545-2557. DOI: 10.1109/TIFS.2017.2710946

7. Zhang R., Zhu F., Liu J., Liu G. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. *IEEE Transactions on Information Forensics and Security*, 2020, vol. 15, pp. 1138-1150. DOI: 10.1109/TIFS.2019.2936913

8. Boroumand M., Chen M., Fridrich J. Deep residual network for steganalysis of digital images. *IEEE Transactions on* Information *Forensics and Security*, 2019, vol. 14, no. 5, pp. 1181-1193. DOI: 10.1109/TIFS.2018.2871749

9. Tabares-Soto R., Arteaga-Arteaga H.B., Buritica O.M.A., et al. Deep learning for steganalysis: evaluating model robustness against image transformations. *Frontiers in Artificial Intelligence*, 2025, vol. 8, article 1532895. DOI: 10.3389/frai.2025.1532895

10. Veit A., Wilber M.J., Belongie S. Residual networks behave like ensembles of relatively shallow networks. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, vol. 29, pp. 550-558.

11. Ntivuguruzwa J.-P., Kurundayev M., Ullah M., et al. A convolutional neural network to detect possible hidden data in spatial domain images. *Cybersecurity*, 2023, vol. 6, article 32. DOI: 10.1186/s42400-023-00156-x

12. Ozcan S., Mustacoglu A.F. Transfer learning effects on image steganalysis with pre-trained deep residual neural network model. In: *IEEE International Conference on Big Data*, 2018, pp. 2280-2287. DOI: 10.1109/BigData.2018.8622437

13. Yedroudj M., Comby F., Chaumont M. Yedroudj-Net: An efficient CNN for spatial steganalysis. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2092-2096. DOI: 10.1109/ICASSP.2018.8461438

14. Dwaik A., Nandi A.K., Naous T., Alani S., Alsarhan A. Enhancing the performance of convolutional neural network image-based steganalysis in spatial domain using Spatial Rich Model and 2D Gabor filters. *Journal of Information Security and Applications*, 2024, vol. 85, article 103862. DOI: 10.1016/j.jisa.2024.103862

15. Wang Z., Gao N., Wang X., Qu X., Li L., Zhang X. Deep learning for steganalysis of diverse data types: A review of methods, taxonomy, challenges and future directions. *arXiv preprint arXiv:2308.04522*, 2024.

16. Tabares-Soto R., Arteaga-Arteaga H.B., Orozco-Arias S., et al. Strategy to improve the accuracy of convolutional neural network architectures applied to digital image steganalysis in the spatial domain. *PeerJ Computer Science*, 2021, vol. 7, e451. DOI: 10.7717/peerj-cs.451

17. Liu F., Zhou X., Yan X., Peng J., Hu Y., Chen Q. Preprocessing enhancement method for spatial domain steganalysis. *Mathematics*, 2022, vol. 10, no. 21, article 3936. DOI: 10.3390/math10213936

18. Wang X., Liao D., Dai Y., Li H. A steganalysis framework based on CNN using the filter subset selection method. *Multimedia Tools and Applications*, 2020, vol. 79, pp. 21307-21326. DOI: 10.1007/s11042-020-08831-8

19. Jin Z., Yang Y., Chen Y., Chen Y. IAS-CNN: Image adaptive steganalysis via convolutional neural network combined with selection channel. *Journal of Sensors*, 2020, article 1550147720911002. DOI: 10.1177/1550147720911002

20. Chaumont, Marc. (2020). Deep learning in steganography and steganalysis. 10.1016/B978-0-12-819438-6.00022-0. In book: Digital Media Steganography (pp.321-349)

21. Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

22. Stefanek G., Gulbransen L., Spink G., Morawski J., Filla D., Rabello De Castro R. A comparison of ai models to detect hidden messages in images. (2024). Issues In Information Systems. 119-132. DOI: 10.48009/3_iis_2024_110

23. The LabelMe-12-50k dataset. URL: https://www.ais.uni-bonn.de/download/datasets.html

24. Meike Helena Kombrink, Zeno Jean Marius Hubert Geradts, and Marcel Worring. 2024. Image Steganography Approaches and Their Detection Strategies: A Survey. ACM Comput. Surv. 57, 2, Article 33 (February 2025), 40 pages. DOI: 10.1145/3694965

25. S. Wu, S. -h. Zhong and Y. Liu, "A Novel Convolutional Neural Network for Image Steganalysis With Shared Normalization," in IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 256-270, Jan. 2020, DOI: 10.1109/TMM.2019.2920605.

26. Dwaik, A., & Belkhouche, Y. (2024). Enhancing the performance of convolutional neural network image-based steganalysis in spatial domain using Spatial Rich Model and 2D Gabor filters. Journal of Information Security and Applications, 85, 103864. DOI: 10.1016/j.jisa.2024.103864

Надійшла до редакції: 28.11.25
Прийнята до друку: 17.03.26
Опубліковано: 30.03.26