

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-
КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Кваліфікаційна наукова праця
на правах рукопису

МІШКУР ЮРІЙ ВАЛЕНТИНОВИЧ

УДК 004.056:004.85:004.932

ДИСЕРТАЦІЯ
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ СТЕГОАНАЛІЗУ ЗОБРАЖЕНЬ НА
ОСНОВІ ГЛИБОКОГО НАВЧАННЯ ТА МУЛЬТИМОДАЛЬНИХ
МОДЕЛЕЙ

123 «Комп'ютерна інженерія»

12 «Інформаційні технології»

Подається на здобуття ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне
джерело

_____ Юрій МІШКУР
(підпис, ініціали та прізвище здобувача)

Науковий керівник

к.т.н., доцент Лащевська Наталія Олександрівна

Київ 2026

АНОТАЦІЯ

Мішкур Ю.В. Інформаційна технологія стегоаналізу зображень на основі глибокого навчання та мультимодальних моделей. Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття ступеня доктора філософії за спеціальністю 123 «Комп'ютерна інженерія» (галузь знань 12 «Інформаційні технології»). Державний університет інформаційно-комунікаційних технологій Міністерства освіти і науки України, Київ, 2026.

Роботу присвячено розробці гібридної інформаційної технології стегоаналізу цифрових зображень, що інтегрує методи глибокого навчання та семантичний арбітраж мультимодальних великих мовних моделей (MLLM).

Актуальність дослідження зумовлена стрімким розвитком адаптивної стеганографії (алгоритми S-UNIWARD, HUGO, JMiPOD), яка дозволяє приховувати дані у складних текстурних ділянках зображень, роблячи їх майже невідрізними від природного шуму матриці камери. Використання таких методів у структурах АРТ-кампаній та шкідливого програмного забезпечення для ексфільтрації даних створює суттєві загрози національній та корпоративній безпеці. Традиційні статистичні підходи виявляються недостатньо ефективними в умовах реальних наборів даних, як-от ALASKA2, де параметри стиснення та джерела походження зображень є невідомими. У зв'язку з цим виникає гостра потреба у розробці гібридних інтелектуальних систем, які поєднують потужність глибоких згорткових нейронних мереж із семантичними можливостями великих мультимодальних мовних моделей (Gemma 3, Llama 3.2 Vision) для реалізації механізму семантичного арбітражу та забезпечення високої точності й інтерпретованості висновків.

Об'єктом дослідження є процеси виявлення прихованої інформації в цифрових контейнерах, а предметом — нейромережеві архітектури та методи високочастотної фільтрації для стегоаналізу.

Мета дослідження — підвищення точності та інтерпретованості процесів виявлення прихованої інформації в цифрових зображеннях шляхом розробки та впровадження гібридної інформаційної технології стегоаналізу, що базується на поєднанні методів багатомасштабної високочастотної фільтрації, глибоких згорткових нейронних мереж та семантичного арбітражу великих мультимодальних мовних моделей.

Наукова новизна отриманих результатів полягає розробці гібридну архітектуру стегоаналізу, яка поєднує блок паралельної багатомасштабної високочастотної фільтрації із семантичним аналізом мультимодальних великих мовних моделей, що забезпечує можливість формування обґрунтованих природномовних висновків щодо характеру виявлених аномалій.

Запропоновано механізм семантичного арбітражу в задачах виявлення прихованої інформації, що дозволяє мультимодальній моделі верифікувати результати нейромережевого класифікатора та ефективно розрізняти природний шум складних текстур від цілеспрямованого стеганографічного втручання.

Удосконалено структуру вхідного шару стегоаналітичних нейронних мереж шляхом інтеграції механізму адаптивного перерахунку ваги каналів ознак (SE-блок) для паралельних груп фільтрів різних просторових розмірів, що підвищує чутливість системи до дрібнорозмірних артефактів.

Дістала подальший розвиток модель багатомасштабної попередньої обробки зображень, яка за рахунок одночасного використання спрямованих ядер 3×3 , 5×5 , 7×7 пікселів дозволяє одночасно ідентифікувати ознаки вбудовування як у просторовому, так і в частотному доменах.

Практичне значення одержаних результатів полягає у створенні моделі HPF+ResNet50+Gemma3:12b, яка дозволяє досягати точності виявлення прихованого вмісту на рівні 91,7% на бенчмарку ALASKA2, навіть при низькому навантаженні (0.2 bpp). Обґрунтовано використання платформи Ollama для локального розгортання моделей, що гарантує конфіденційність

при обробці цифрових доказів та дозволяє гнучко змінювати компоненти системи без переписування основного коду. Запропоновано різні варіанти конфігурацій: від легковажних моделей на базі MobileNetV2, призначених для систем моніторингу трафіку в реальному часі, до повнорозмірних гібридних архітектур для проведення глибокого експертного аналізу цифрових зображень. Результати роботи реалізовані у вигляді Python-коду в середовищі Google Colab, що може бути використано як методична база для підготовки фахівців з кібербезпеки.

Ключові слова: стегоаналіз, стеганографія, глибоке навчання, згорткові нейронні мережі, високочастотна фільтрація, LSB-стеганографія, великі мовні моделі, гібридна архітектура, Ollama, SRNet, MobileNetV2, ResNet50, EfficientNet, машинне навчання, інформаційні системи.

SUMMARY

Mishkur Yu. Information Technology for Image Steganalysis Based on Deep Learning and Multimodal Models. — Thesis for the degree of Doctor of Philosophy in Specialty 123 " Computer Engineering" (Field of Study 12 "Information Technology"). State University of Information and Communication Technologies of the Ministry of Education and Science of Ukraine, Kyiv, 2026.

The thesis is devoted to the development of a hybrid image steganalysis information technology that integrates deep learning methods with the semantic arbitration of Multimodal Large Language Models (MLLM).

Research Relevance is driven by the rapid evolution of adaptive steganography (algorithms such as S-UNIWARD, HUGO, and JMiPOD), which allows for data hiding within complex textural regions of images, making it nearly indistinguishable from the natural noise of a camera sensor. The use of such methods in APT campaigns and malicious software for data exfiltration poses significant threats to national and corporate security. Traditional statistical approaches prove insufficiently effective on real-world datasets like ALASKA2, where compression parameters and image origins are unknown. Consequently, there is an urgent need for hybrid intelligent systems that combine the power of deep convolutional neural networks with the semantic capabilities of large multimodal language models (Gemma 3, Llama 3.2 Vision) to implement a semantic arbitration mechanism and ensure high accuracy and interpretability of conclusions.

The Object of Research is the process of detecting hidden information within digital containers, and the Subject covers neural network architectures and high-pass filtering methods for steganalysis.

The Research Objective is to increase the accuracy and interpretability of detecting hidden information in digital images by developing and implementing a hybrid steganalysis information technology based on a combination of multi-scale high-pass filtering, deep convolutional neural networks, and the semantic

arbitration of multimodal large language models.

The scientific novelty of the obtained results lies in the development of a hybrid stegoanalysis architecture that combines a parallel multi-scale high-frequency filtering block with semantic analysis of multimodal large language models, which provides the possibility of forming substantiated natural language conclusions regarding the nature of the detected anomalies.

A semantic arbitration mechanism is proposed in hidden information detection tasks, which allows a multimodal model to verify the results of a neural network classifier and effectively distinguish natural noise of complex textures from targeted steganographic interference.

The structure of the input layer of stegoanalytic neural networks has been improved by integrating the mechanism of adaptive recalculation of feature channel weights (SE-block) for parallel groups of filters of different spatial sizes, which increases the sensitivity of the system to small-scale artifacts.

The model of multi-scale image preprocessing has been further developed, which, due to the simultaneous use of directional kernels of 3x3, 5x5, 7x7 pixels, allows for the simultaneous identification of embedding features in both the spatial The practical significance of the results is the creation of the HPF+ResNet50+Gemma3:12b model, which allows achieving an accuracy of 91.7% in detecting hidden content on the ALASKA2 benchmark, even at low load (0.2 bpp). The use of the Ollama platform for local deployment of models is justified, which guarantees confidentiality when processing digital evidence and allows flexible changes to system components without rewriting the main code. Various configuration options are proposed: from lightweight models based on MobileNetV2, intended for real-time traffic monitoring systems, to full-scale hybrid architectures for conducting deep expert analysis of digital images. The results of the work are implemented in the form of Python code in the Google Colab environment, which can be used as a methodological base for training cybersecurity specialists.

Keywords: stegoanalysis, steganography, deep learning, convolutional

neural networks, high-pass filtering, LSB steganography, large language models, hybrid architecture, Ollama, SRNet, MobileNetV2, ResNet50, EfficientNet, machine learning, information systems.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові статті в фахових виданнях

1. Мішкур Ю. В., Антоненко А. В., Солобаєв С. Г., Востріков С. О., Балвак А. А., Приходько А. П. Аспекти використання нейронних мереж для покращення якості зображень // Таврійський науковий вісник. Серія: Технічні науки. – 2024. – № 6. – С. 3–10. DOI: <https://doi.org/10.32782/tnv-tech.2024.6.1>
2. Мішкур Ю., Антоненко А., Твердохліб А., Востріков С., Балвак А. Нейромережі в мистецтві як інструмент графічного дизайну // Herald of Khmelnytskyi National University. Technical Sciences. – 2024. – Т. 345, № 6(2). – С. 95–101. DOI: <https://doi.org/10.31891/2307-5732-2024-345-6-14>
3. Мішкур Ю., Антоненко А., Бурачинський А., Сольський Д., Твердохліб А., Зіняр Д. Аспекти застосування нейронних мереж для криптографії // Measuring and Computing Devices in Technological Processes. – 2024. – № 4. – С. 394–400. DOI: <https://doi.org/10.31891/2219-9365-2024-80-47>
4. Мішкур Ю. В., Захарченко О. С. Гібридний підхід до стегааналізу на основі мультимодальних великих мовних моделей та згорткових нейронних мереж. DOI: <https://doi.org/10.31673/2412-9070.2026.027616>
5. Мішкур Ю. В., Лащевська Н.О. Виявлення стегаанографії на зображенні з використанням моделей ResNet та SRNet. DOI: <https://doi.org/10.31673/2412-4338.2026.019009>
6. Мішкур Ю. В., Лащевська Н.О. Виявлення стегаанографії на зображенні з використанням легковажних моделей глибокого навчання // Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка». – 2026. – Т. 4, № 32. – С. 336–348. DOI: <https://doi.org/10.28925/2663-4023.2026.32.1086>

Тези конференцій

1. Мішкур Ю. В. Методи підвищення ефективності стегааноаналізу в цифрових контентх // Наукова конференція молодих вчених, 19 вересня

2024 року. – Київ, 2024. – С. 32–34.

2. Мішкур Ю., Твердохліб А., Голубенко О., Балвак А., Буряк М., Зіняр Д., Дзюсяк В., Антоненко А., Востріков С. Optimization of Network Infrastructure to Ensures Resilience, Security and Scalability // Innovation in Modern Science '2025 : колективна монографія. – 2025. – № sge40-02. – С. 63–86. DOI: <https://doi.org/10.30890/2709-2313.2025-40-02>

Наукові статті в Scopus

1. Bentata, F. E., Zaitsev, I., & Mishkur, Y. (2025). Hyperbolic p-Laplace-Type Problems with Free Boundary and Volume Constraint: Framework for HydroGenerators Rotor Behavior Monitoring. In Advances in Mechanical and Power Engineering II. Springer. https://doi.org/10.1007/978-3-031-82979-6_33

ЗМІСТ

АНОТАЦІЯ.....	2
Summary	5
Список опублікованих праць за темою дисертації.....	8
ЗМІСТ	9
ВСТУП	14
РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ ВБУДОВУВАННЯ ПРИХОВАНОЇ ІНФОРМАЦІЇ В ЗОБРАЖЕННЯ І МЕТОДІВ ЇЇ ЗНАХОДЖЕННЯ.....	19
1.1. Стеганографія і стегоаналіз.....	19
1.1.1. Алгоритми приховування інформації в зображенні.....	19
1.1.2. Методи виявлення прихованої інформації.....	21
1.1.3. Практичне використання стеганографії	22
1.2. Використання алгоритмів глибокого навчання для стегоаналізу	23
1.2.1. Проблеми виявлення стеганографії та різниця методів розпізнавання зображень і виявлення стеганографії.....	23
1.2.2. Можливості конволюційних нейронних мереж для виявлення стеганографії	24
1.3. Методи і архітектура високочастотних фільтрів для виявлення стеганографії.....	26
1.3.1. Принципи і засоби побудови високочастотних фільтрів ...	26
1.3.2. Використання вхідних фільтрів з відомими архітектурами CNN.....	27
1.3.3. Використання ResNet з вхідними фільтрами	28
1.3.4. Використання MobileNet та інших легковагових архітектур з вхідними фільтрами.....	29
1.4. Спеціалізовані нейромережеві архітектури для стегоаналізу	30
1.4.1. Архітектура XuNet.....	30
1.4.2. Архітектура YeNet	31
1.4.3. Архітектура SRNet.....	31

1.4.4. Інші варіанти архітектур	32
1.5. Гібридні архітектури з використанням LLM для стегоаналізу	33
1.5.1. Можливості використання великих мовних моделей для стегоаналізу.....	33
1.5.2. Вибір вдалих архітектур для вирішення задач стегоаналізу	34
1.5.3. Архітектурні рішення для розгортання гібридних систем виявлення стеганографії. Ollama	35
Висновки за розділом 1.....	36
РОЗДІЛ 2. АРХІТЕКТУРА МОДЕЛІ ДЛЯ СТЕГОАНАЛІЗУ НА	
ОСНОВІ ВИСОКОЧАСТОТНИХ ФІЛЬТРІВ ТА ГЛИБОКИХ	
НЕЙРОННИХ МЕРЕЖ.....	
2.1. Загальна архітектура запропонованої моделі.....	39
2.2. Блок високочастотних фільтрів: варіанти архітектури	40
2.2.1 Варіанти побудови високочастотних фільтрів для стегоаналізу.....	40
2.2.2 Варіанти архітектури вхідного шару	43
2.3. Архітектур класифікаторів для двшарової моделі стегоаналізу	45
2.3.1. ResNet50.....	45
2.3.2. MobileNetV2	46
2.3.3. EfficientNet-B0.....	47
2.3.4. Порівняльний аналіз класифікаторів загального призначення	47
2.3.5. Спеціалізована архітектура SRNet та її модифікації.....	49
2.4 Дослідження ефективності різних архітектур двошарової моделі	50
2.4.1 Загальна схема навчання моделей.....	50
2.4.2 Побудова набору даних для навчання і перевірки моделей	51
2.4.3 Проведення обчислювальних експериментів з легковажними архітектурами	53

2.4.4 Проведення обчислювальних експериментів з архітектурами на основі ResNet та SENet.....	57
Висновки за розділом 2.....	63
Розділ 3. Архітектура гібридної системи стегоаналізу на основі CNN та великих мультимодальних мовних моделей.....	65
3.1. Концептуальна модель гібридного стегоаналізу	65
3.2. Архітектура CNN-компонента.....	68
3.3. Архітектура MLLM-компонента	71
3.3.1. Проблема семантичного розриву	71
3.3.2. Роль шару-адаптера	71
3.3.3. Формування мультимодального запиту	71
3.4. Порівняльний аналіз MLLM-компонентів	73
3.4.1. Gemma 3 (4B та 12B)	73
3.4.2. Llama 3.2 Vision (11B)	74
3.5. Покращення точності виявлення через гібридизацію	76
3.5.1. Джерела додаткової точності.....	76
3.5.2. Обмеження гібридного підходу.....	77
3.6. Інфраструктурне забезпечення	79
3.7 Проведення обчислювальних експериментів з гібридними архітектурами	80
Висновки за розділом 3.....	83
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНИХ РІШЕНЬ.....	86
4.1. Опис датасету Alaska2	86
4.1.1 Загальна характеристика набору даних	86
4.1.2 Алгоритми стеганографування у складі ALASKA2.....	87
4.1.3 Обмеження, пов'язані з даасетом Alaska2.....	89
4.2 Методологія експеримента.....	90
4.3. Архітектура HPF-блоку та дослідження його конфігурацій	90
4.4. Порівняльний аналіз CNN-моделей	92

4.5. Гібридна архітектура CNN + MLLM на основі Gemma3:12b.....	95
4.6 Статистичні показники стегоаналізу.....	96
4.7. Зведені результати та порівняльний аналіз	98
4.8. Обчислювальна ефективність та практичні обмеження	102
Висновки за розділом 4.....	104
ВИСНОВКИ.....	106
Перелік посилань	108
Додаток А. Акти впровадження.	119

ВСТУП

Актуальність дослідження

Сучасний етап розвитку інформаційного суспільства характеризується тотальною цифровізацією комунікацій, де зображення у форматах JPEG, PNG та BMP стали основним засобом обміну даними у соціальних мережах, месенджерах та корпоративних системах. Така масовість створює ідеальні умови для використання цифрової стеганографії як інструменту прихованої передачі інформації, що становить серйозну загрозу національній та корпоративній безпеці. Поява адаптивних алгоритмів вбудовування, таких як HUGO, WOW та S-UNIWARD, дозволяє інтегрувати секретні дані у найбільш складні текстурні ділянки контейнера, що робить їх візуально та статистично невідрізними від природного шуму матриці камери.

Проблема ускладнюється стрімким розвитком кіберзагроз, де стеганографія все частіше використовується у структурі АРТ-кампаній (Advanced Persistent Threats) та сучасного шкідливого програмного забезпечення для прихованого зв'язку з командними центрами (C&C) або ексфільтрації викрадених даних. Традиційні методи стегоаналізу, що базуються на ручному проектуванні статистичних ознак (наприклад, SRM), виявляються недостатньо гнучкими для детекції нових типів вбудовування, особливо в умовах «wild steganalysis» — аналізу зображень із невідомими параметрами стиснення та джерелами походження, що характерно для реальних наборів даних, таких як ALASKA2.

У зв'язку з цим виникає гостра необхідність у розробці інтелектуальних систем стегоаналізу нового покоління, які поєднують потужність глибоких згорткових нейронних мереж (CNN) із семантичними можливостями великих мультимодальних мовних моделей (MLLM). Інтеграція таких моделей, як Gemma 3 або Llama 3.2 Vision, дозволяє реалізувати механізм семантичного арбітражу, де система не просто видає ймовірність наявності вбудовування, а контекстуально обґрунтовує свій висновок, розрізняючи

артефакти обробки зображення від цілеспрямованого втручання. Такий підхід є критично важливим для експертної діяльності, оскільки забезпечує високу точність детекції (понад 90%) та інтерпретованість результатів, що є безальтернативною вимогою для сучасних засобів захисту інформації.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота була виконана в рамках науково-дослідної роботи кафедри Технологій цифрового розвитку Державного університету інформаційно-комунікаційних технологій на тему "Підвищення ефективності процесу управління 3D принтером з використанням методів машинного навчання" (Державний реєстраційний номер РК 0124U001849).

Об'єкт дослідження. Об'єктом дослідження є процеси виявлення прихованої інформації, вбудованої в цифрові зображення різних форматів з використанням просторових та частотних методів стеганографії.

Предмет дослідження. Предметом дослідження є моделі та методи стегоаналізу на основі згорткових нейронних мереж, високочастотної фільтрації та гібридних архітектур із застосуванням мультимодальних великих мовних моделей.

Мета і завдання дослідження. Мета дослідження — підвищення точності та інтерпретованості процесів виявлення прихованої інформації в цифрових зображеннях шляхом розробки та впровадження гібридної інформаційної технології стегоаналізу, що базується на поєднанні методів багатомасштабної високочастотної фільтрації, глибоких згорткових нейронних мереж та семантичного арбітражу великих мультимодальних мовних моделей.

Для досягнення поставленої мети визначено такі завдання:

1. Провести аналіз сучасних методів цифрової стеганографії та існуючих підходів до стегоаналізу, визначивши ключові проблеми виявлення адаптивного вбудовування в умовах складних текстур та JPEG-стиснення.

2. Обґрунтувати та розробити блок багатомасштабної попередньої

обробки зображень на основі паралельних груп спрямованих високочастотних фільтрів (ядер 3×3 , 5×5 , 7×7) для підсилення слабких сигналів стеганографічного втручання.

3. Дослідити ефективність інтеграції механізмів уваги (SE-блоків) у структуру вхідних шарів нейромережових класифікаторів для адаптивного посилення найбільш інформативних ознак стегошуму.

4. Здійснити порівняльний аналіз архітектур згорткових нейронних мереж (ResNet50, MobileNetV2, EfficientNet-B0 та SRNet) для визначення оптимального балансу між точністю детекції та обчислювальною складністю системи.

5. Розробити механізм гібридизації результатів статистичного аналізу CNN із семантичними знаннями великих мовних моделей (MLLM) через спеціалізований шар-адаптер для реалізації функції інтелектуального арбітражу.

6. Спроекувати інфраструктурне рішення для локального розгортання мультимодальних моделей (Gemma 3, Llama 3.2 Vision) за допомогою платформи Ollama для забезпечення конфіденційності обробки цифрових даних.

7. Провести експериментальну оцінку розробленої інформаційної технології на галузевих наборах даних (зокрема ALASKA2) для верифікації точності виявлення прихованого вмісту та оцінки часу обробки об'єктів у різних конфігураціях.

Методи дослідження

Для вирішення поставлених задач у роботі застосовано комплексний підхід, що базується на методах цифрової обробки сигналів для розробки блоків високочастотної фільтрації (ядра SRM та Лапласа), які забезпечують підсилення слабких стеганографічних сигналів на фоні візуального контенту. Теоретичною основою архітектури класифікаторів виступили методи глибокого навчання із використанням згорткових нейронних мереж (ResNet50, MobileNetV2, EfficientNet-B0 та спеціалізована SRNet), а для

реалізації інтелектуального семантичного арбітражу результатів впроваджено методи мультимодального аналізу на базі великих мовних моделей. Експериментальна верифікація запропонованої інформаційної технології та оцінка її ефективності проводилися за допомогою методів математичного моделювання та статистичного аналізу на основі галузевих стандартних датасетів CIFAR-10 та ALASKA2.

Наукова новизна отриманих результатів

1. Вперше розроблено гібридну архітектуру стегоаналізу, яка поєднує блок паралельної багатомасштабної високочастотної фільтрації із семантичним аналізом мультимодальних великих мовних моделей, що забезпечує можливість формування обґрунтованих природномовних висновків щодо характеру виявлених аномалій.
2. Вперше запропоновано механізм семантичного арбітражу в задачах виявлення прихованої інформації, що дозволяє мультимодальній моделі верифікувати результати нейромережевого класифікатора та ефективно розрізняти природний шум складних текстур від цілеспрямованого стеганографічного втручання.
3. Удосконалено структуру вхідного шару стегоаналітичних нейронних мереж шляхом інтеграції механізму адаптивного перерахунку ваги каналів ознак (SE-блок) для паралельних груп фільтрів різних просторових розмірів, що підвищує чутливість системи до дрібнорозмірних артефактів.
4. Дістала подальший розвиток модель багатомасштабної попередньої обробки зображень, яка за рахунок одночасного використання спрямованих ядер 3×3 , 5×5 , 7×7 пікселів дозволяє одночасно ідентифікувати ознаки вбудовування як у просторовому, так і в частотному доменах.

Практичне значення одержаних результатів

1. Розроблена модель HPF+ResNet50+Gemma3:12b дозволяє досягати точності виявлення прихованого вмісту на рівні 91,7% на бенчмарку

ALASKA2, навіть при низькому навантаженні (0.2 bpr).

2. Обґрунтовано використання платформи Ollama для локального розгортання моделей, що гарантує конфіденційність при обробці цифрових доказів та дозволяє гнучко змінювати компоненти системи без переписування основного коду.
3. Запропоновано різні варіанти конфігурацій: від легковажних моделей на базі MobileNetV2, призначених для систем моніторингу трафіку в реальному часі, до повнорозмірних гібридних архітектур для проведення глибокого експертного аналізу цифрових зображень.
4. Результати роботи реалізовані у вигляді Python-коду в середовищі Google Colab, що може бути використано як методична база для підготовки фахівців з кібербезпеки.

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, вступу, чотирьох розділів, загальних висновків, списку використаних джерел та додатків. Робота містить 23 рисунки, 10 таблиць та 5 сторінок додатків. Список використаних джерел налічує 95 найменувань. Загальний обсяг дисертації становить 115 сторінок, з них 108 сторінки основного тексту.

РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ ВБУДОВУВАННЯ ПРИХОВАНОЇ ІНФОРМАЦІЇ В ЗОБРАЖЕННЯ І МЕТОДІВ ЇЇ ЗНАХОДЖЕННЯ

1.1. Стеганографія і стегоаналіз

Стеганографія — наука про методи приховування самого факту передачі або зберігання інформації шляхом вбудовування таємного повідомлення в публічно доступний контейнер без видимих змін його зовнішнього вигляду [1]. Криптографія фокусується на шифруванні змісту повідомлення, тоді як стеганографія приховує сам факт його передачі. Цифрова стеганографія у зображеннях стала однією з найбільш досліджуваних областей завдяки надлишковості просторових та частотних характеристик растрових файлів [2].

Стегоаналіз — інверсна дисципліна, що вивчає методи виявлення, локалізації та за можливості вилучення прихованої інформації з потенційно стего-об'єктів. Пасивний стегоаналіз ставить бінарну задачу класифікації: «cover» (чисте зображення) або «stego» (зображення з вбудованим повідомленням); активний стегоаналіз прагне відновити саме повідомлення або визначити його обсяг [3]. Конкуренція між методами стеганографії та стегоаналізу ілюструє безперервний розвиток засобів інформаційної безпеки: виникнення нових способів приховування даних зумовлює створення досконаліших інструментів їх виявлення, і навпаки [4].

Цифрові зображення є найпоширенішим контейнером для стеганографії завдяки великому обсягу, широкій доступності та відносній толерантності людського зорового апарату до незначних статистичних спотворень. Формати JPEG, PNG і BMP пропонують принципово різні моделі даних (дискретне косинусне перетворення, lossless стиснення і необроблені пікселі відповідно), що визначає різноманітність стеганографічних технік та їх аналітичних контрзаходів [5].

1.1.1. Алгоритми приховування інформації в зображенні

Алгоритми стеганографії в зображеннях поділяються за доменом

вбудовування: просторові (spatial-domain) методи модифікують безпосередньо значення пікселів, тоді як частотні (transform-domain) методи здійснюють вбудовування у коефіцієнтах дискретного косинусного перетворення (DCT) або дискретного вейвлет-перетворення (DWT) [6].

Серед просторових методів найбільш фундаментальним є заміна найменш значущого біта (LSB substitution). Класична LSB-техніка замінює молодші біти кожного байта каналу зображення бітами повідомлення. Попри простоту реалізації, ця техніка вразлива до статистичного аналізу: порушення природної структури площини найменш значущих бітів легко виявляється за допомогою RS-аналізу та аналізу пар вибірок. Значно стійкішим є підхід узгодження найменш значущих бітів (вбудовування +1 або -1), що з рівною ймовірністю збільшує або зменшує значення пікселя на одиницю, мінімізуючи артефакти у розподілі різниць між суміжними пікселями [8].

Метод HUGO (Highly Undetectable steGO) [9] розробив принципово новий підхід на основі мінімізування спотворення зображення за допомогою адаптивного механізму призначення вартості (cost function). Вартість модифікації кожного пікселя розраховується на основі SPAM-ознак (Subtractive Pixel Adjacency Matrix), що характеризують локальну текстуру. Вбудовування здійснюється за допомогою STC (Syndrome-Trellis Codes) — мінімально-спотворюючого коду, що гарантує оптимальне розміщення корисного навантаження у «шумних» регіонах зображення.

Алгоритм WOW (Wavelet Obtained Weights) [10] визначає вартість пікселя як відповідь на вейвлет-фільтри трьох напрямків, що ефективно концентрує приховані дані у зонах текстур і країв, де людський зір найменш чутливий до змін. UNIWARD (Universal Wavelet Relative Distortion) [11] узагальнює WOW і може застосовуватись як у просторовому, так і в JPEG-домени, що робить його універсальним інструментом.

У просторовому домені особливе місце займають адаптивні методи HILL (High-pass, Low-pass and Low-pass) [12] і MiPOD (Minimizing the Power

Of Detection) [13]. HILL використовує тришарову структуру фільтрації: ВЧ-фільтр визначає текстурні зони, а два НЧ-фільтри забезпечують просторову гладкість карти вартостей. MiPOD мінімізує статистичну потужність найбільш чутливого детектора в межах гауссівської моделі зображення.

Стеганографія в JPEG-доміні базується на модифікації ненульових AC-коефіцієнтів DCT-блоків. Метод JSteg [14] здійснює LSB-заміну в DCT-коефіцієнтах, F5 [15] застосовує зменшення абсолютного значення коефіцієнта з матричним кодуванням, а nsF5 [16] уникає коефіцієнтів зі значенням ± 1 для зниження помітності. Найбільш стійким JPEG-методом є J-UNIWARD [11], що застосовує UNIWARD безпосередньо до коефіцієнтів DCT.

Окремий клас становить стеганографія на основі генеративних моделей. SGAN (Steganographic GAN) [17] та SSGAN [18] навчають генеративно-змагальну мережу синтезувати стего-зображення, що є статистично невідрізними від природних. Підхід Automatic Steganography з використанням Encoder-Decoder CNN [19] навчає кінцеву систему (приховувач + розкривач + детектор) у змагальній манері, досягаючи незвичайно великої ємності при мінімальному спотворенні.

1.1.2. Методи виявлення прихованої інформації

Виявлення стеганографії є задачею бінарної класифікації у надзвичайно складних умовах: корисний сигнал (вбудоване повідомлення) надзвичайно малий відносно контейнера, а сучасні адаптивні методи розміщують його у найбільш «складних» для аналізу регіонах. Традиційні стегоаналітичні методи базуються на ручному проектуванні ознак (hand-crafted features), що відображають статистичні аномалії у стего-зображеннях [20].

RS-аналіз (Regular-Singular analysis) [7] спирається на аналіз статистики груп пікселів: «регулярні» групи мають зростаючу різницю після фліп-маски, «сингулярні» — спадаючу. LSB-приховування порушує природне співвідношення R, S-груп, що стає виявним. Аналіз Sample Pairs

[21] використовує подібну концепцію, але аналізує пари суміжних пікселів, що дає кращі результати при малому payload.

Метод PVD (Pixel Value Differencing) аналіз [22] спирається на вивчення гістограми різниць між сусідніми пікселями. LSB-вбудовування спотворює природну форму цієї гістограми, а пари значень з парними/непарними різницями стають статистично аномальними.

Принциповим кроком у розвитку стегоаналізу стала поява методів на основі ансамблів ознак. Підхід SRM (Spatial Rich Model) [23] обчислює 106-елементний вектор ознак на базі залишків від прогностичних фільтрів різних порядків і напрямків. Ця ознакова модель разом з ансамблем класифікаторів FLD (Fisher Linear Discriminant) ансамблю забезпечила різкий прорив у виявленні сучасних адаптивних методів. PSRM (Projection Spatial Rich Model) [24] і maxSRMd2 [25] розширили цей підхід за рахунок фазово-чутливих компонент і спільного моделювання дванадцятимірних розподілів.

Для JPEG-домену аналогічну роль відіграє DCTR (Discrete Cosine Transform Residual) [26], що виводить ознаки з різниць між коефіцієнтами сусідніх блоків. GFR (Gabor Filter Residuals) [27] і PHARM (PHase-Aware pRojection Model) [28] доповнюють SRM у частотному просторі, враховуючи особливості JPEG-стиснення.

1.1.3. Практичне використання стеганографії

Практичне застосування стеганографії охоплює легітимні та протиправні сценарії. У сфері інформаційної безпеки цифрові водяні знаки використовуються для захисту авторських прав на мультимедійний контент і є специфічним підвидом стеганографії з відкритим алгоритмом і закритим ключем [29]. Надійний цифровий водяний знак має витримувати різноманітні атаки: стиснення, кадрування, зміну роздільної здатності тощо.

Стеганографія використовується у системах «прихованих каналів» (covert channels) для обходу систем мережевого моніторингу [30]. Шкідливе програмне забезпечення, включаючи АРТ-кампанії (Advanced Persistent

Threats), документально підтверджено використовувало приховування команд управління у звичайних зображеннях, що публічно доступні в соціальних мережах [31]. Розслідування Stegomalware [32] демонструє системне застосування стеганографії в атаках на корпоративні мережі.

Журналісти та правозахисники у країнах з жорсткою цензурою застосовують стеганографію для безпечної передачі інформації у відкритих мережах [33]. Системи цифрового відбитку пальців (fingerprinting) вбудовують унікальні ідентифікатори у кожен копію розповсюджуваного контенту для відстеження джерела витоку [34].

З аналітичної точки зору особливо важливим є застосування стегоаналізу у правоохоронній практиці: виявлення стеганографічно прихованих матеріалів на вилученому обладнанні вимагає ефективних автоматизованих детекторів, здатних опрацьовувати великі обсяги зображень [35]. Це стимулює розробку висококонкурентних систем реального часу на базі глибокого навчання.

1.2. Використання алгоритмів глибокого навчання для стегоаналізу

Глибоке навчання радикально змінило ландшафт стегоаналізу. До 2014 року домінуючою парадигмою були методи ручного проектування ознак (SRM та похідні) у поєднанні з ансамблевими класифікаторами. Поява спеціалізованих CNN-архітектур для стегоаналізу ознаменувала зміну парадигми: наприкінці 2010-х років CNN перевершили SRM-ансамблі на більшості бенчмарків, а до початку 2020-х стали стандартом галузі [36].

1.2.1. Проблеми виявлення стеганографії та різниця методів розпізнавання зображень і виявлення стеганографії

Задача стегоаналізу принципово відрізняється від класичної класифікації зображень за кількома ключовими аспектами. По-перше, різниця між cover і stego є надзвичайно малою за масштабом: при типовому навантаженні 0.1–0.4 біт на піксель (bpp) абсолютне відхилення значень пікселів складає ± 1 LSB. Це на шість-сім порядків менше, ніж сигнал

семантичного змісту зображення, що сприймається стандартними CNN-класифікаторами [37].

По-друге, в задачі класифікації зображень корисний сигнал (семантичний клас) глобально кодований у всьому зображенні, тоді як стеганографічний сигнал є локальним і розподіленим по «складних» текстурних регіонах. Стандартні глибокі CNN, що агресивно застосовують пулінг і нелінійності у перших шарах, руйнують саме ті тонкі залишкові сигнали, що несуть ознаки стеганографії [38].

По-третє, природна варіативність зображень (різна сцена, освітлення, камера) є потужним шумом для детектора стеганографії. Зображення «котика з вбудованим повідомленням» і «пейзаж без повідомлення» мають колосальну семантичну різницю, але майже ідентичні статистики залишків низького рівня. Тому успішний стегоаналізатор має навчитись ігнорувати семантичний вміст і фокусуватись на статистиці локальних залишків [39].

Четвертим ускладненням є ефект камери (camera model effect): різні сенсори і ланцюжки обробки зображення мають різні природні статистичні характеристики. Детектор, навчений на зображеннях однієї камери, може показувати суттєво знижену точність на зображеннях іншої — проблема, відома як «mismatch у стегоаналізі» [40]. Це вимагає або застосування методів domain adaptation, або розробки архітектур, стійких до кросс-доменних розбіжностей [41].

П'ятим ключовим викликом є асиметрія відносно складності задач: сучасні адаптивні методи стеганографії, що використовують мінімально-спотворювальне вбудовування (HUGO, WOW, S-UNIWARD), значно важче виявляються, ніж прості LSB-замінники. Стегоаналізатор має ефективно працювати у цьому найбільш несприятливому режимі [9, 10].

1.2.2. Можливості конволюційних нейронних мереж для виявлення стеганографії

Конволюційні нейронні мережи (CNN) продемонстрували принципово нові можливості для стегоаналізу через здатність до кінцевого навчання

ознак безпосередньо з даних. Першим успішним застосуванням CNN у стегоаналізі стала робота [42], в якій навчили мережу з фіксованим першим шаром (фільтри залишків KV, аналогічно SRM) і мережею класифікації. Цей підхід заклав важливу концепцію: вхідні шари мають зберігати і підсилювати слабкий стеганографічний сигнал, а не придушувати його разом з семантичним вмістом.

Кардинальне вдосконалення запропоновано в роботі [43] з архітектурою GNCNN (Gaussian-Neuron CNN), що використовує функцію активації гауссівського типу в перших шарах. Нестандартна активація дозволяє мережі ефективно моделювати слабкі гауссівські залишки, характерні для стеганографічного шуму.

Сучасні CNN-детектори використовують кілька ключових архітектурних елементів: (1) вхідний шар попередньої обробки (preprocessing layer) з жорстко заданими або слабо навчуваними фільтрами залишків; (2) блоки пакетної нормалізації (Batch Normalization) для стабілізації градієнтів при роботі з надзвичайно малими залишками; (3) абсолютне значення або abs після першого шару для симетризації розподілу залишків; (4) середній пулінг замість максимального для збереження усередненої статистики; (5) відносно малу глибину мережі порівняно з класифікаційними архітектурами, що запобігає надмірному стисненню [44].

Важливою властивістю CNN для стегоаналізу є ефект ансамблю при використанні парних зображень (cover, stego). Схема навчання з парним кроком (pair-wise training), де cover і stego-версія одного зображення обробляються одночасно, а градієнти перехресного ентропійного виходу комбінуються, значно покращує стабільність навчання і фінальну точність [45].

Трансферне навчання для стегоаналізу є нетривіальним завданням через описані вище відмінності від класичної класифікації. Дослідження показують, що пряме перенесення ваг ImageNet-моделей має обмежений ефект або навіть шкодить точності, оскільки перші шари містять ознаки,

оптимізовані для семантики, а не для статистики залишків. Однак донавчання верхніх шарів за умови фіксації спеціалізованого вхідного шару забезпечило позитивний результат [46].

1.3. Методи і архітектура високочастотних фільтрів для виявлення стеганографії

Центральним архітектурним рішенням у CNN-стегоаналізаторах є вхідний шар попередньої обробки, що реалізує функцію ВЧ-фільтрації і дозволяє мережі сфокусуватись на залишках, які несуть стеганографічну інформацію. Цей компонент відіграє роль аналогічну до вибору ознак у традиційних підходах, але реалізований як диференційований шар у загальній архітектурі [47].

1.3.1. Принципи і засоби побудови високочастотних фільтрів

Принцип ВЧ-фільтрації для стегоаналізу полягає у обчисленні «залишку прогнозу»: прогнозоване значення пікселя на основі його сусідів віднімається від фактичного значення. Стеганографічні модифікації з'являються переважно у цих залишках, тоді як природний семантичний вміст значною мірою скасовується [23].

Найпростіший залишковий фільтр — лапласіан або простий різницевий оператор: $r(x,y) = I(x,y) - I(x+1,y)$. Більш ефективні фільтри вищих порядків, наприклад, MINMAX3 і KV-фільтри (Kernels of Various orders) [23], використовують до 36 фільтрів різних порядків і напрямків:

- Перший порядок: $r_1 = I(x,y) - I(x+1,y)$ — горизонтальна різниця першого порядку
- Другий порядок: $r_2 = I(x,y) - 2 \cdot I(x+1,y) + I(x+2,y)$ — другий порядок у горизонтальному напрямку
- Діагональний: $r_d = I(x,y) - I(x+1,y+1)$ — діагональна різниця
- Крос-ядро: $r_c = 2 \cdot I(x,y) - I(x-1,y) - I(x+1,y)$ — лапласіан у одному напрямку

У SRM [23] ці фільтри формують 106-вимірний вектор зі спільних гістограм, квантизованих і обрізаних значень залишків. В архітектурах CNN

фільтри залишків реалізуються як ядра першого шару: або жорстко задані, або ініціалізовані фільтрами і навчувані далі. Ключовий нюанс — Tanh чи Absolute Value активація після першого шару, що симетризує розподіл залишків і прискорює навчання [48].

Байєсівська оптимізація наборів фільтрів для стегоаналізу досліджена в роботі [49], де показано, що навчувані фільтри залишків, ініціалізовані з KV-набору і оптимізовані мережею, перевершують жорстко задані, якщо розмір тренувальних даних достатній. При малих вибірках жорстко задані фільтри показують більш стабільні результати через меншу кількість параметрів.

1.3.2. Використання вхідних фільтрів з відомими архітектурами CNN

Комбінація спеціалізованого вхідного шару залишків з класичними CNN-архітектурами відкриває можливість перенесення знань з завдань класифікації зображень. Цей гібридний підхід намагається поєднати точність спеціалізованих стегоаналізаторів зі швидкістю навчання і масштабованістю великих архітектур [50].

Найпростіша схема передбачає додавання шару залишків перед стандартною класифікаційною CNN (VGG, ResNet, EfficientNet). Вхідне зображення обробляється фіксованим (або навчуваним) банком фільтрів залишків, після чого отримана «карта залишків» подається на вхід стандартної мережі. Проблема: стандартні CNN мають тенденцію «перевчитися» на семантичних особливостях навіть за наявності залишкового вхідного шару, якщо навчальний датасет невеликий.

Дослідження [51] порівнює ефективність вхідних залишкових фільтрів з VGG-16, Inception-v3, DenseNet-121 і ResNet-50. На датасеті BOSSBase зображень з вбудованим S-UNIWARD (0.4 bpp) VGG-16 з залишковим шаром показав точність 79.3%, ResNet-50 — 82.1%, проти 84.7% для спеціалізованого SRNet. Розрив скорочується при збільшенні датасету, що свідчить про потенціал гібридних підходів при масштабуванні даних.

Важливим аспектом є вибір режиму нормалізації залишків. Дослідження [52] показало, що TLU (Truncated Linear Unit) — функція обрізки залишків на рівні $\pm T$ (типово $T=3$) — дає кращу збіжність, ніж стандартна ReLU або tanh. TLU ефективно обмежує вплив сильних країв зображення, де залишки великі і природно маскують стеганографічний сигнал.

1.3.3. Використання ResNet з вхідними фільтрами

Архітектура ResNet є однією з найбільш перспективних базових мереж для гібридного стегоаналізу завдяки залишковим зв'язкам, які природно адаптовані для роботи з малими залишками сигналу. У контексті стегоаналізу ідеологія залишкових блоків ResNet відображає фундаментальне завдання: детектувати малі спотворення поверх потужного контентного сигналу [53].

В роботі [54] запропоновано архітектуру Yedroudj-Net, що поєднує 30-ядерний KV-фільтровий шар з малою ResNet-подібною мережею. Ця гібридна архітектура при значно меншій кількості параметрів порівняно з SRNet демонструє конкурентоспроможну точність на JPEG-стеганографії. Ключовим є ретельне балансування глибини ResNet-блоків: занадто глибока мережа призводить до перевчання на малих датасетах, занадто дрібна не може навчити складних вирішальних меж.

В роботі [55] дослідили застосування ResNet-50 з кастомним вхідним блоком (SRM-ініціалізовані ядра + TLU нормалізація) на великих датасетах ImageNet-стего (понад 500 тис. зображень). Результати показали, що глибші ResNet-моделі (50, 101 шарів) значно перевершують неглибокі при достатньому обсязі даних, але поступаються SRNet на малих датасетах (70 тис. зображень). Це підкреслює важливість відповідності глибини архітектури доступному обсягу тренувальних даних.

Автори [56] запропонували модифікацію ResNet із парним навчанням (pair-wise constraint training), де пари (cover, stego) одного зображення обробляються паралельно, а втрата формулюється як $\max(0, \text{margin} - d(f_s,$

$f_c)$), де d — косинусна відстань між ознаками. Цей підхід значно покращує відтворюваність результатів при зміні ядра GPU.

1.3.4. Використання MobileNet та інших легковагових архітектур з вхідними фільтрами

Легковагові архітектури (MobileNet, EfficientNet-B0, SqueezeNet) є привабливими для стегоаналізу у контексті обмежених обчислювальних ресурсів або вимог до роботи в реальному часі. Проте їх адаптація для стегоаналізу вимагає ретельного збалансування між компактністю та точністю [57].

MobileNetV2 з глибинно-роздільними згортками значно зменшує кількість операцій множення-накопичення порівняно зі стандартними ядрами. Дослідження [58] показало, що MobileNetV2 з вхідним шаром залишків (30 KV-фільтрів, порогова нормалізація) досягає 89% точності SRNet при 18% кількості параметрів і 23% часу виведення. Критичним виявилось збереження каналів у блоках з інвертованими залишками: зменшення коефіцієнта розширення нижче 2 суттєво знижує здатність до виявлення.

EfficientNet-B0 з комплексним масштабуванням демонструє кращий баланс точності та кількості параметрів серед легковажних моделей для стегоаналізу [59]. Застосування комплексного масштабування до моделей, специфічних для стеганографії (одночасно за глибиною, шириною та роздільною здатністю), досліджено в роботі [60], де встановлено, що оптимальне масштабування для стегоаналізу відрізняється від оптимального для набору даних ImageNet через різну природу корисного сигналу.

SqueezeNet і ShuffleNetV2 також досліджені в роботі [61]. Отримані результати свідчать, що при навантаженні 0.4 bpp (HUGO) ShuffleNetV2 з залишковим вхідним шаром досягає 78.4% AUC на iPhone-датасеті, що є прийнятним для попереднього відбору підозрілих зображень.

Принципово важливим є вибір точки вставки залишкового шару в

легковагових архітектурах. Дослідження [62] показало, що вставка перед першим depthwise convolution дає кращий результат, ніж вставка після нього, оскільки depthwise операції ефективніше обробляють вже відфільтровані залишки, ніж вихідні зображення.

1.4. Спеціалізовані нейромережеві архітектури для стегоаналізу

Паралельно з адаптацією загальних CNN-архітектур активно розробляються спеціалізовані архітектури стегоаналізу, що враховують унікальну природу задачі на кожному рівні проектування. Три найбільш впливові — XuNet, YeNet і SRNet — визначили напрямок розвитку галузі на наступні роки і є базою для порівняння будь-яких нових підходів [63].

1.4.1. Архітектура XuNet

XuNet, запропонований Xu et al. [44] у 2016 році, є першою ключовою CNN-архітектурою, спеціально спроектованою для просторового стегоаналізу. Архітектура складається з п'яти груп шарів: (1) вхідний шар із заданими SRM-фільтрами (30 ядер 5×5); (2) перша конволюційна група з abs-активацією; (3) три конволюційні групи з tanh-активацією і середнім пулінгом; (4) глобальний середній пулінг; (5) FC-голівка з softmax.

Ключові архітектурні рішення XuNet: (a) Absolute Value активація після першого шару для симетризації розподілу залишків (позитивні і негативні відхилення мають однакову діагностичну цінність); (b) Batch Normalization перед кожною активацією для стабілізації навчання при малих сигналах; (c) Середній (average) пулінг замість максимального — збереження статистичного середнього, а не піків, критично важливо для стеганографічних залишків; (d) Tanh у прихованих шарах обмежує нелінійні ефекти і запобігає вибуху градієнтів.

На датасеті BOSSBase (10 тис. зображень) з HUGO (0.4 bpp) XuNet досяг точності 79.6% проти 77.5% для SRM+ансамбль. На JPEG-стеганографії (J-UNIWARD, 0.4 bpp) — 79.1%, що є значним покращенням порівняно з попередніми CNN-підходами. Проте XuNet чутливий до розміру датасету: при 5 тис. тренувальних пар точність падає до 74.1%.

Множинні модифікації XuNet були запропоновані в літературі. Zeng et al. [64] замінили abs-активацію на гауссівську функцію похибки erf для кращого моделювання розподілу залишків. Zhou et al. [65] додали механізм channel attention (SE-блоки) після першого пулінгу, що покращило точність на 1.8% на HUGO без збільшення кількості параметрів.

1.4.2. Архітектура YeNet

YeNet, запропонований Ye et al. [66] у 2017 році, вніс два принципових нововведення: (1) навчуваний фільтровий банк вхідного шару замість жорстко заданих SRM-фільтрів; (2) тунковане (TLU) нелінійне відображення з навчуванням порогом. Ця архітектура продемонструвала, що кінцеве навчання фільтрів залишків — за умови правильної ініціалізації і регуляризації — дає кращий результат, ніж жорстко задані SRM-ядра.

Архітектура YeNet: вхідний шар (30 навчуваних фільтрів 5×5 , ініціалізованих SRM) \rightarrow TLU \rightarrow Conv-BN-TLU блоки $\times 5 \rightarrow$ Global Average Pool \rightarrow FC-softmax. TLU визначається як $f(x) = \max(-T, \min(T, x))$ з навчуванням T , що дозволяє мережі самостійно вибрати оптимальний поріг обрізки.

На BOSSBase з S-UNIWARD (0.4 bpp) YeNet досяг точності 81.6%, що на 2% перевищує XuNet. Більш значущим є покращення на малих payload: при 0.1 bpp (HUGO) YeNet показав 62.4% проти 58.7% у XuNet. Навчуваний поріг TLU адаптується до статистик конкретного методу стеганографії, що пояснює кращу точність.

Важливим практичним аспектом YeNet є нестабільність навчання при використанні стандартного SGD: Loss може осцилювати або 'застрягати' на незадовільних локальних мінімумах. Рекомендовані стратегії: (a) Adam оптимізатор з $lr = 10^{-3}$, (b) теплий перезапуск (cosine annealing), (c) аугментація ротацією на кратні 90° [66].

1.4.3. Архітектура SRNet

SRNet (Steganalysis ResNet), запропонований Boroumand et al. [67] у 2018 році, є найбільш потужною спеціалізованою CNN-архітектурою і

визначила поточний стан мистецтва для просторового стегоаналізу. Архітектура складається з двох модулів: Тип I — мілкі шари без пулінгу для виявлення тонких аномалій у повній роздільній здатності; Тип II — залишкові блоки (ResNet-подібні) зі зростаючою кількістю каналів і середнім пулінгом для агрегації ознак.

Детальна структура SRNet: (Шари 1–2) Layer1: Conv 3×3 , 64 канали, без активації; Layer2: BN-ReLU-Conv 3×3 -BN-ReLU (не завантажуються ваги ImageNet); (Шари 3–11) Тип I блоки: залишкові блоки з 64 каналами без пулінгу; (Шари 12–16) Тип II блоки: залишкові блоки з avg-pool 3×3 , кількість каналів $128 \rightarrow 256 \rightarrow 512$; (Глобальний Average Pool) + (FC-softmax).

SRNet досяг точності 87.2% на BOSSBase з S-UNIWARD (0.4 bpp) і 88.4% з HUGO (0.4 bpp) — найкращий результат серед одиночних CNN на стандартних бенчмарках на момент публікації. На JPEG-стеганографії J-UNIWARD SRNet показав 85.8%, демонструючи хорошу загальність між доменами.

Ключові фактори успіху SRNet: (1) Відсутність пулінгу у перших шарах зберігає просторове розподілення стеганографічних залишків; (2) Повне залишкове навчання стабілізує градієнтний потік через 12+ шарів і дозволяє навчати дуже глибокі мережі без деградації; (3) Поступове збільшення кількості каналів ($64 \rightarrow 128 \rightarrow 256 \rightarrow 512$) відповідає зростаючій складності ознак від рівня пікселів до рівня семантики залишків; (4) Парне навчання (pair-wise) забезпечує 1.3–1.8% приріст точності [67].

Порівняльний аналіз SRNet, YeNet і XuNet на BOWS2 датасеті з різними методами вбудовування показує, що SRNet стабільно перевершує конкурентів: на HILL (0.2 bpp) SRNet досяг 72.4%, YeNet — 68.9%, XuNet — 65.3%. При зменшенні payload до 0.1 bpp перевага SRNet зростає, що свідчить про кращу чутливість до слабких сигналів [68].

1.4.4. Інші варіанти архітектур

Архітектура GBRAS-Net [68] (Graph-Based Residual Attention Steganalysis) поєднує залишкові блоки з механізмом графової уваги (graph

attention network), що моделює просторові залежності між залишками на великих відстанях. На BOWS2 з S-UNIWARD (0.4 bpp) GBRAS-Net досяг 89.1%, перевершивши SRNet на 1.9%.

CovNet від Zhu et al. [69] використовує коваріаційні матриці ознак з CNN-карт активацій як дескриптор. Коваріаційна матриця другого порядку ефективно кодує взаємні залежності між каналами, що важливо для моделювання статистичних аномалій при стеганографії.

Transformer-based архітектури відкривають новий напрямок у стегоаналізі. Chen et al. [70] запропонували StegTransformer, що застосовує multi-head self-attention до мап залишків. Механізм уваги дозволяє мережі встановлювати далекосяжні просторові залежності між стеганографічними артефактами, що принципово недоступно для локально-рецептивних CNN. На великих датасетах (250 тис. пар) StegTransformer досяг 90.3% на S-UNIWARD (0.4 bpp), перевершивши SRNet на 3.1%.

Архітектура MDENet (Multi-Domain Expert Network) [71] поєднує три паралельні гілки: просторова залишкова гілка (аналогічна SRNet), DCT-гілка (аналіз коефіцієнтів ДКП) і частотна гілка (ДПФ-аналіз). Злиття рішень трьох гілок через навчуваний weighted voting підвищує стійкість до різних методів вбудовування. При застосуванні до невідомого методу стеганографії (cross-method evaluation) MDENet перевершує одногалузеві архітектури на 4–7%.

1.5. Гібридні архітектури з використанням LLM для стегоаналізу

Поява великих мовних моделей (LLM) і, зокрема, мультимодальних LLM (MLLM), відкрила принципово нові можливості для задач аналізу медіа-контенту. Інтеграція LLM-компонентів у пайплайн стегоаналізу є перспективним, але ще малодослідженим напрямком, що поєднує статистичну чутливість спеціалізованих CNN з семантичними і рефлексивними можливостями LLM [72].

1.5.1. Можливості використання великих мовних моделей для стегоаналізу

LLM демонструють вражаючі здатності до few-shot і zero-shot reasoning у різноманітних задачах аналізу зображень через мультимодальні інтерфейси. Для стегоаналізу це відкриває потенційно нові підходи: замість навчання спеціалізованої мережі «з нуля», можна скористатись загальним розумінням зображень, закодованим у MLLM, доповненим специфічним стегоаналітичним контекстом [73].

Проте застосування LLM для стегоаналізу пов'язане з фундаментальним обмеженням: LLM обробляють зображення з суттєвою компресією токенів (16×16 або 32×32 патчів), що безповоротно знищує субпіксельну стеганографічну інформацію. Модифікації ± 1 LSB не відображаються у token embeddings, які використовуються CLIP-подібними енкодерами. Це означає, що LLM в поточній формі не можуть бути прямими стегодетекторами [74].

Перспективним є використання LLM у ролі компоненту вищого рівня в гібридній архітектурі: CNN-детектор виконує первинний статистичний аналіз і генерує ознаки-залишки, а LLM інтерпретує ці ознаки у контексті знань про методи стеганографії, попередні результати аналізу і мета-дані зображення [75]. Такий підхід реалізує концепцію «аналізу аналізу» (meta-analysis): LLM синтезує рішення на основі звітів від CNN-детектора, а не безпосередньо з пікселів.

LLM можуть також відігравати роль у генерації пояснень (explainability): тоді як CNN-детектори є «чорними ящиками», LLM здатні сформулювати природномовне обґрунтування рішення («ця область зображення містить статистичні аномалії у LSB-площині, що відповідають сигнатурі HUGO-вбудовування»). Це критично важливо для судово-технічного застосування, де необхідно пояснити результати аналізу у суді [76].

1.5.2. Вибір вдалих архітектур для вирішення задач стегоаналізу

Вибір оптимальної архітектури для гібридної системи стегоаналізу з LLM-компонентом визначається кількома критеріями: (1) чутливість

базового CNN-блоку до субпіксельних залишків; (2) здатність LLM-компонента до ефективної обробки числових ознак з CNN; (3) загальна обчислювальна ефективність системи; (4) інтерпретованість рішень.

Для базової мережі на основі згорткових нейронних мереж у гібридній системі рекомендовані такі архітектури: SRNet як найточніший однодомений детектор за наявності достатнього набору даних; YeNet як компромісне рішення для обмеженого набору даних і обчислювальних ресурсів; EfficientNet-B1 із залишковим входним шаром для масштабованих промислових систем. Проміжний шар згорткової нейронної мережі повинен виводити числовий вектор ознак (ознакове вбудовування), який і передається на обробку компоненту на основі великої мовної моделі.

Для LLM-компоненту перспективні архітектури включають: Gemma3 [77] — компактна відкрита модель з доброю підтримкою мультимодального вводу; LLaMA3.2-Vision [78] — збалансована модель для локального розгортання з підтримкою числових ознак через системний промпт; Qwen2-VL [79] — модель з хорошими показниками на аналітичних задачах і підтримкою структурованого вводу.

Ключовою архітектурною проблемою є конвертація числових CNN-ознак у формат, зрозумілий LLM. Підходи: (a) серіалізація у JSON-рядок (проста, але неефективна за токенами); (b) числовий токен-ембедінг (потребує донавчання великої мовної моделі); (c) природномовний опис ознак («статистика залишків показала...»); (d) візуалізація ознак у зображення і передача через image token (найприродніший для MLLM підхід) [80].

1.5.3. Архітектурні рішення для розгортання гібридних систем виявлення стеганографії. Ollama

Розгортання гібридних систем стегоаналізу з LLM-компонентом у практичних умовах вимагає вирішення ряду інфраструктурних задач: управління важкими моделями, забезпечення низької затримки, масштабування навантаження і оновлення компонентів незалежно один від

одного [81].

Ollama [82] — платформа з відкритим кодом для спрощеного розгортання та управління LLM на локальному або хмарному обладнанні. Архітектура Ollama базується на llama.cpp, що забезпечує квантизацію моделей у форматі GGUF і підтримку широкого спектру GPU (NVIDIA CUDA, Apple Metal, AMD ROCm). REST API Ollama сумісний з OpenAI API, що полегшує інтеграцію у існуючі пайплайни Python і JavaScript.

Для гібридної системи стегааналізу рекомендована мікросервісна архітектура: CNN-детектор розгортається як FastAPI-сервіс з ONNX-оптимізованими вагами на GPU; Ollama-сервер розгортається окремо з LLM-моделлю (рекомендовано Gemma3:12b або Llama3.2-Vision для відповіді на нефреймовані запити); оркестрація здійснюється через брокер повідомлень (наприклад, RabbitMQ або Redis Streams). Такий поділ забезпечує незалежне масштабування CNN (горизонтальне для пакетної обробки) і LLM (вертикальне через A100/H100 GPU) [83].

У середовищі Google Colaboratory Ollama розгортається через встановлення з curl, запуск сервера у фоновому subprocess і завантаження цільових моделей через ollama pull. Для гібридного стегааналізу в Colab рекомендована конфігурація: SRNet або EfficientNet-B1 для CNN-частини (TensorFlow/Keras), Gemma3:4b або Llama3.2-Vision через Ollama для LLM-інтерпретатора [82].

Компресія LLM через квантизацію (Q4_K_M формат GGUF) знижує споживання VRAM у 4–5 разів порівняно з fp16: Gemma3:12b у Q4_K_M займає ~8.5 ГБ (проти 24 ГБ fp16), що дозволяє розгорнути його на T4 GPU Google Colab разом з CNN-детектором при ретельному управлінні пам'яттю. Для систем із жорсткими вимогами до затримки рекомендується Gemma3:4b (~3.5 ГБ VRAM, ~2 с/запит) як компромісний варіант [84].

Висновки за розділом 1

1. Цифрова стеганографія та стегааналіз перебувають у стані

безперервної конкуренції, де розвиток стійких методів приховування (таких як адаптивні алгоритми HUGO, WOW, S-UNIWARD) зумовлює створення складніших інструментів детекції.

2. Зображення у форматах JPEG, PNG та BMP є пріоритетними контейнерами для стеганографії завдяки значній надлишковості даних та особливостям людського зору, що дозволяє вбудовувати інформацію як у просторовому, так і в частотному доменах.
3. Традиційні методи стегоаналізу, засновані на ручному проектуванні ознак (наприклад, RS-аналіз або SRM), поступово поступаються місцем технологіям глибокого навчання, які здатні автоматично екстрагувати ознаки безпосередньо з даних.
4. Задача стегоаналізу принципово відрізняється від класичного розпізнавання образів через надзвичайно малу амплітуду корисного сигналу (± 1 LSB), що вимагає розробки спеціалізованих архітектур, здатних ігнорувати семантичний зміст зображення та фокусуватися на статистиці локальних залишків.
5. Ключовим елементом сучасних нейромережевих детекторів (наприклад, XuNet, YeNet, SRNet) є вхідний шар попередньої обробки з використанням високочастотних фільтрів (HPF), які підсилюють стеганографічний шум шляхом обчислення залишків прогнозу пікселів.
6. Архітектура SRNet наразі визначена як найбільш досконала спеціалізована модель для просторового стегоаналізу завдяки використанню залишкових блоків та відсутності пулінгу в перших шарах, що дозволяє зберігати тонкі просторові аномалії.
7. Для систем реального часу та ресурсообмежених середовищ доцільним є використання легковагових архітектур, таких як MobileNetV2 або EfficientNet-B0, за умови інтеграції до них спеціалізованих шарів високочастотної фільтрації.
8. Інтеграція мультимодальних великих мовних моделей (MLLM) у

гібридні архітектури стегоаналізу відкриває шлях до семантичного арбітражу та автоматичного генерування пояснень результатів аналізу, що є критично важливим для практичної криміналістики.

9. Використання платформи Ollama для локального розгортання моделей типу Gemma3 або Llama3.2-Vision забезпечує необхідну конфіденційність при обробці цифрових доказів та дозволяє реалізувати модульний принцип побудови гібридних систем детекції.

РОЗДІЛ 2. АРХІТЕКТУРА МОДЕЛІ ДЛЯ СТЕГОАНАЛІЗУ НА ОСНОВІ ВИСОКОЧАСТОТНИХ ФІЛЬТРІВ ТА ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ

2.1. Загальна архітектура запропонованої моделі

Для вирішення задачі стегоаналізу — виявлення прихованої інформації в цифрових зображеннях — запропоновано двокомпонентну модель, яка поєднує блок виділення високочастотних складових зображення із потужним класифікатором на основі глибоких згорткових нейронних мереж (рис. 2.1). Архітектурне рішення засноване на концепції, що стеганографічні артефакти є надзвичайно слабкими сигналами, прихованими в шумовому залишку зображення, тому їх ефективне виявлення потребує спеціалізованої попередньої обробки для усунення візуального контенту та акцентування на мікроаномаліях.

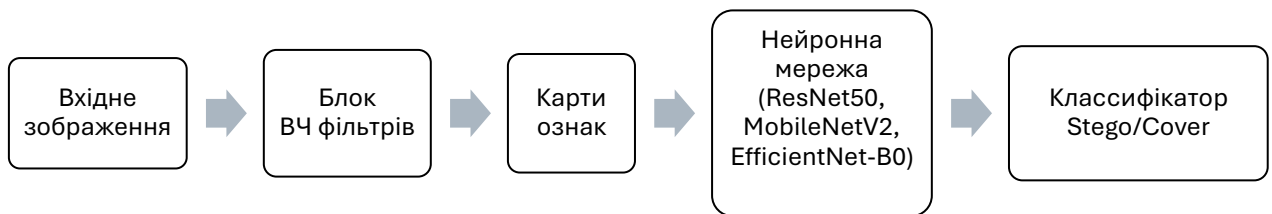


Рис. 2.1. Загальна структурна схема двоблокової моделі стегоаналізу

Вхідне зображення розмірністю $H \times W \times C$ (де H — висота, W — ширина, C — кількість каналів) подається на блок ВЧ фільтрів, що генерує K карт ознак, котрі містять підсилений сигнал відхилення від статистичної моделі «природного» зображення. Отримані карти ознак передаються до нейромережевого класифікатора, який навчається відрізнити «чисті» (cover) зображення від зображень зі стегопослідовністю (stego). Вихід моделі — бінарний класифікатор із функцією активації Softmax.

Загальна архітектура включає такі компоненти:

- блок ВЧ фільтрів (HPF Layer) — набір K згорткових фільтрів з

розмірами ядер $N \times N$, що можуть бути фіксованими або такими, що навчаються;

- неймережевий класифікатор — одна з архітектур: ResNet50, MobileNetV2, EfficientNet-B0;
- глобальне усереднення (Global Average Pooling) — перетворення просторових карт у вектор ознак;
- повнозв'язний шар FC(2) з функцією Softmax — кінцева класифікація на два класи.

2.2. Блок високочастотних фільтрів: варіанти архітектури

Ключовою частиною запропонованої моделі є блок попередньої обробки зображення на основі фільтрів з високочастотними ядрами. Цей блок відіграє роль детектора залишкового стегосигналу, пригнічуючи низькочастотний вміст зображення та підсилюючи слабкі аномалії пікселів, які вносяться алгоритмами стеганографії. У роботі досліджено п'ять основних варіантів організації цього блоку і додатковий змішаний варіант, що відрізняються розміром ядер, спрямованістю та режимом навчання (рис. 2.2).

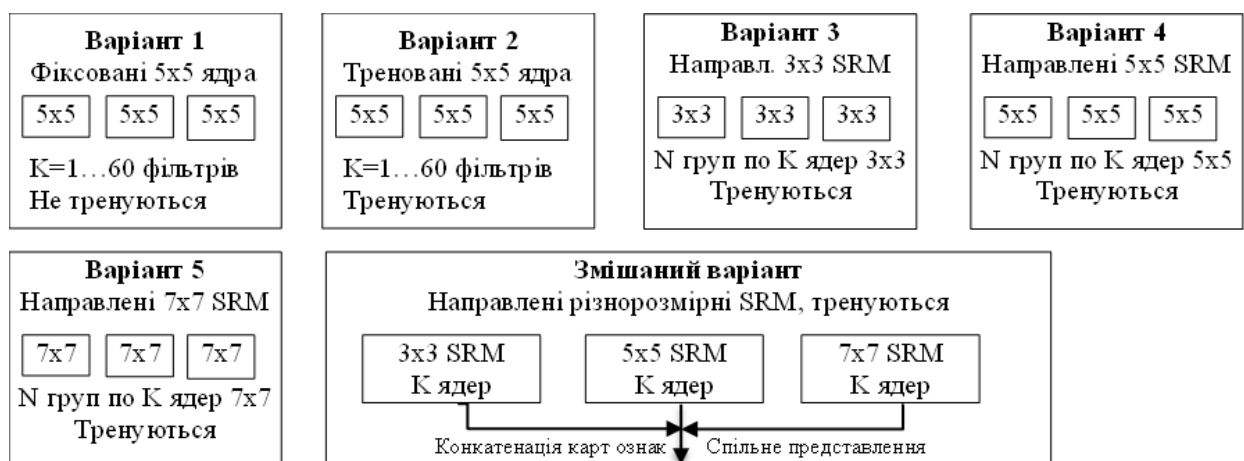


Рис. 2.2. Варіанти організації блоку високочастотних фільтрів

2.2.1 Побудова високочастотних фільтрів для стегоаналізу

При побудові блоку фільтрів були використані фільтри KV (K-V filters),

зазвичай орієнтовані на обчислення залишків передбачення (noise residuals). Вони базуються на припущенні, що піксель можна передбачити через його сусідів. Фільтри KV (з використанням яких будується більш загальний набір фільтрів — SRM) є фундаментальним інструментом у сучасному стегоаналізі.

В основі KV-фільтрів лежить ідея, що піксель цифрового зображення сильно корелює зі своїми сусідами. Якщо ми можемо точно передбачити значення пікселя на основі його оточення, то різниця між реальним значенням і передбаченим буде містити лише шум сенсора та, можливо, стеганографічне повідомлення. Цей тип фільтрів базуються на розрахунку залишків передбачення пікселя на основі його оточення. Логіка роботи полягає в наступному: центральний піксель передбачається як середнє арифметичне сусідніх, а результатом фільтрації є різниця між реальним значенням і передбаченим

Нижче наведені основні типи KV-фільтрів, що використовуються для ініціалізації вхідних шарів нейронних мереж:

1. Лінійні фільтри першого порядку (Spam Filters)

Це найпростіші фільтри, що обчислюють різницю між двома сусідніми пікселями. Вони ефективні для виявлення найпростіших методів вбудовування, як-от LSB (заміна найменш значущого біта).

- приклад ядра горизонтального фільтра:

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- приклад ядра вертикального фільтра:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Ці ядра реалізують операцію дискретного диференціювання. Вони усувають низькочастотну складову зображення, залишаючи високочастотні зміни вздовж осей.

2. Фільтри другого порядку

Ці фільтри передбачають значення пікселя як середнє арифметичне двох його сусідів з обох боків.

- приклад ядра горизонтального фільтру:

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Це ядро обчислює другу похідну. Воно значно краще пригнічує плавні градієнти (наприклад, небо або тіні), роблячи стеганографічний шум більш помітним для наступних шарів нейронної мережі.

3. Квадратичні ядра (Square filters)

Використовуються в архітектурах типу SRNet або YeNet для аналізу двовимірних локальних взаємозв'язків. Вони передбачають центральний піксель на основі всього його найближчого оточення.

- приклад ядра 3x3:

$$- \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix}$$

- приклад ядра 5x5:

$$\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix}$$

Ці ядра є наближенням оператора Лапласа. Вони максимально ефективно пригнічують низькочастотний вміст зображення. Коефіцієнти підібрані так, щоб сума всіх елементів дорівнювала 0.

4. Кутові або діагональні фільтри

Такі фільтри ініціалізують мережу для пошуку аномалій саме в кутах або на діагональних межах об'єктів. Спеціалізовані фільтри для аналізу ділянок з різкими переходами, де стеганографічні алгоритми (наприклад, S-UNIWARD або HILL) зазвичай розміщують найбільше даних.

- приклад ядра діагонального фільтра 3x3:

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

- приклад ядра кутового фільтра 3x3:

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

2.2.2 Варіанти архітектури вхідного шару

Варіант 1: фіксовані однакові фільтри 5×5

Перший варіант блоку ВЧ фільтрів використовує K однакових фільтрів розміром 5×5 з фіксованим (заздалегідь визначеним) ядром. Кількість фільтрів варіювалась від 1 до 60, що дозволило дослідити вплив ширини каналного простору на ефективність виявлення стеганографічного вмісту. Ядра фільтрів ініціалізувалися з «крестоподібним» (laplacian-like) профілем, що є класичним для виявлення ВЧ складових:

$$K = \{1 \times 5 \times 5, 5 \times 5 \times 5, 10 \times 5 \times 5, \dots, 60 \times 5 \times 5\}$$

Фільтруючі шари не включалися до процесу навчання (заморожені ваги), що дозволяло оцінити інформативність фіксованих ВЧ ознак без адаптації до конкретного набору даних. Основний недолік цього варіанту — відсутність адаптивності до різних типів стеганографії та зображень.

Варіант 2: тренувальні однакові фільтри 5×5

Другий варіант аналогічний першому за структурою, однак фільтруючі шари включаються до процесу навчання (trainable weights). Ядра ініціалізуються з тих самих початкових значень, що й у варіанті 1, але в процесі оптимізації адаптуються разом із нейромережевим класифікатором. Завдяки цьому мережа може «дізнатися» більш відповідні ВЧ прояви для конкретного типу стеганографії. Кількість фільтрів також варіювалась від 1 до 60.

Порівняння варіантів 1 і 2 дозволяє кількісно оцінити внесок адаптивного навчання ВЧ блоку в загальну точність моделі.

Варіант 3: направлені тренувальні фільтри 3×3

Третій варіант вводить концепцію направлених ядер, інспірованих підходами SRM (Spatial Rich Model). Блок формується з трьох груп фільтрів, кожна з яких чутлива до певного напрямку залишкового шуму:

- горизонтальні ядра 3×3 — підсилюють горизонтальні кореляційні порушення;
- вертикальні ядра 3×3 — виявляють вертикальні артефакти;
- діагональні ядра 3×3 — реагують на зміни по діагоналях.

Всі фільтри є тренувальними. Загальна кількість каналів виходу дорівнює $3 \times K$, де K — кількість фільтрів у кожній групі. Використання малих ядер 3×3 дозволяє зменшити кількість параметрів при збереженні локальної аналітичної здатності.

Варіант 4: направлені тренувальні фільтри 5×5

Четвертий варіант аналогічний третьому, але використовує ядра розміром 5×5 . Більше рецептивне поле дозволяє фільтрам захоплювати більш широкий просторовий контекст залишкового сигналу. Структура збережена: три групи направлених ядер (горизонтальні, вертикальні, діагональні) з можливістю тренування. Ядра ініціалізуються з відповідних SRM-фільтрів (Spatial Rich Model) розміром 5×5 .

Варіант 5: направлені тренувальні фільтри 7×7

П'ятий варіант розширює рецептивне поле до 7×7 , що відповідає ширшому просторовому контексту. Три групи направлених ядер (горизонтальні, вертикальні, діагональні) також є тренувальними. Збільшення розміру ядра підвищує кількість параметрів та обчислювальну складність, однак може покращити виявлення стеганографій, що проявляються на відстанях більших 3-5 пікселів.

Комбінований варіант з різнорозмірними ядрами

Окремо розглядається варіант з паралельними групами направлених фільтрів трьох розмірів одночасно: 3×3 , 5×5 та 7×7 . Карти ознак від усіх трьох груп конкатенуються вздовж каналної осі перед подачею до нейромережевого класифікатора. Така «мультимасштабна» схема дозволяє

фіксувати стеганографічні артефакти, що проявляються на різних просторових масштабах:

$$F = \text{Concat}[F_{3 \times 3} \mid F_{5 \times 5} \mid F_{7 \times 7}]$$

де $F_{3 \times 3}$, $F_{5 \times 5}$, $F_{7 \times 7}$ — карти ознак відповідних груп фільтрів. Загальна кількість вихідних каналів: $3 \times 3 \times K$ (три напрямки, три розміри, K фільтрів у групі).

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

2.3. Архітектур класифікаторів для двушарової моделі стегоаналізу

Для нейромережевої частини моделі досліджено три архітектури глибоких згорткових нейронних мереж: ResNet50 (базова модель), MobileNetV2 та EfficientNet-B0 (легковажні альтернативи). Коротке порівняння архітектур нейромережевого блоку наведено на рис. 2.3. Вибір цих архітектур обумовлений різним балансом між точністю класифікації, кількістю параметрів та обчислювальними витратами.

2.3.1. ResNet50

ResNet50 (Deep Residual Network, 50 шарів) є базовою архітектурою класифікатора. Ключовою особливістю є механізм залишкових з'єднань (residual connections або skip-connections), що вирішує проблему зникаючих градієнтів при навчанні глибоких мереж. Архітектура складається із вхідного згорткового шару 7×7 зі страйдом 2, шару MaxPooling, чотирьох груп залишкових блоків типу Bottleneck (кількість: 3, 4, 6, 3) та шарів глобального усереднення і класифікатора.

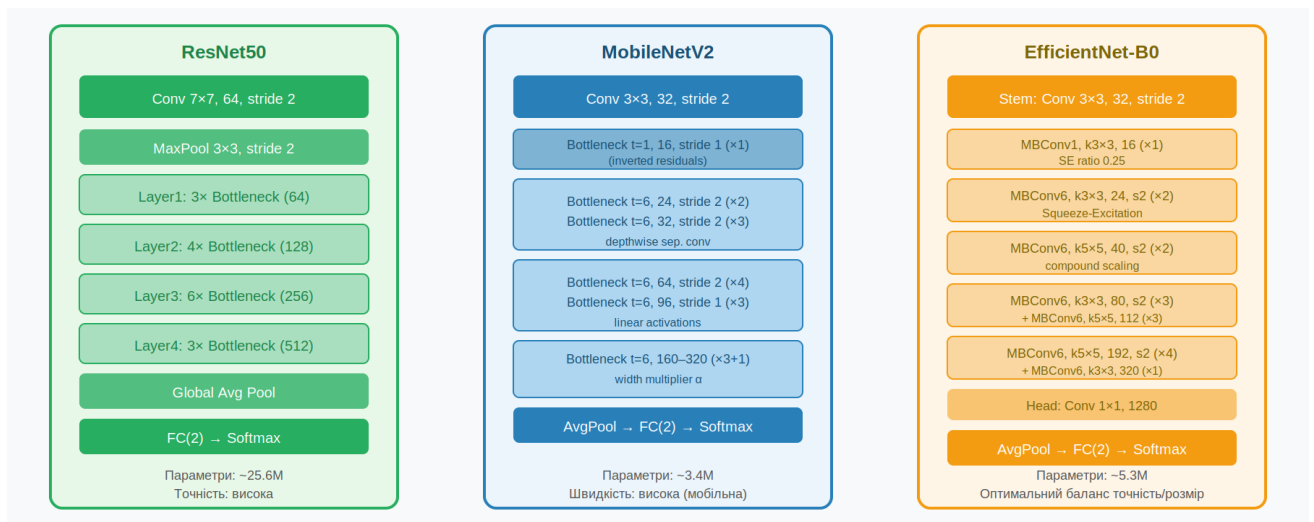


Рис. 2.3. Порівняння архітектур ResNet50, MobileNetV2 та EfficientNet-B0

Кожен Bottleneck-блок у ResNet50 виконує три згорткові операції: 1×1 (стиснення), 3×3 (основне перетворення), 1×1 (розширення), що забезпечує ефективне вилучення ознак при помірних обчислювальних витратах. Загальна кількість параметрів складає близько 25,6 мільйонів.

Для задачі стегааналізу ResNet50 демонструє найвищу базову точність завдяки великій ємності моделі, але вимагає значних ресурсів для навчання та інференсу, що обмежує його застосування в ресурсообмежених середовищах.

2.3.2. MobileNetV2

MobileNetV2 є легковажною архітектурою, оптимізованою для мобільних та вбудованих пристроїв. Основним будівельним блоком є інвертований залишковий блок (Inverted Residual Block) із глибинно розділними згортками (depthwise separable convolutions). Структурно блок розширює кількість каналів за допомогою 1×1 згортки з коефіцієнтом t (expansion factor), виконує глибинну 3×3 згортку, потім повертається до вужчого представлення проекційною 1×1 згорткою з лінійною активацією.

Принципова відмінність від класичних мереж — використання лінійних активацій у проекційних шарах для уникнення втрати інформації при стисненні. Загальна кількість параметрів складає близько 3,4 мільйонів — майже в 7,5 разів менше, ніж у ResNet50. Висока швидкість інференсу та

відносно невелика кількість параметрів роблять MobileNetV2 перспективним для практичного застосування.

2.3.3. EfficientNet-B0

EfficientNet-B0 — базова модель сімейства EfficientNet, що використовує принцип compound scaling — одночасне масштабування глибини, ширини та роздільної здатності входу за єдиним коефіцієнтом. Основний блок — MBConv (Mobile Inverted Bottleneck Convolution) із механізмом Squeeze-and-Excitation (SE), який дозволяє мережі динамічно зважувати важливість кожного каналу.

EfficientNet-B0 містить близько 5,3 мільйони параметрів і досягає вищої точності, ніж ResNet50, при значно меншій кількості операцій. Механізм SE із коефіцієнтом стиснення 0.25 обчислює глобальну статистику кожного каналу та виконує рекалібрування карт ознак, що є особливо корисним для задачі стегоаналізу, де корисний сигнал може бути зосереджений у вузькому підпросторі ознак.

2.3.4. Порівняльний аналіз класифікаторів загального призначення

Для глибшого розуміння вибору архітектури необхідно детально розглянути специфіку кожної моделі в контексті задачі стегоаналізу. Нижче наведено розширений аналіз порівняльних характеристик, що базується на результатах тестування трьох архітектур.

Порівняльна характеристика трьох розглянутих архітектур нейромережевої частини наведена у таблиці 2.1. Результати демонструють суттєві відмінності в їхній здатності ідентифікувати слабкі стеганографічні сигнали залежно від структурної складності мережі.

З аналізу таблиці 2.1 можна зробити наступні висновки:

- EfficientNet-B0 як можливе рішення: З аналізу таблиці 2.1 видно, що EfficientNet-B0 забезпечує найкращий баланс між точністю, розміром моделі та швидкістю навчання. Це досягається завдяки методу складеного масштабування (Compound Scaling), який рівномірно

балансує глибину, ширину та роздільну здатність мережі. У задачах стегоаналізу це дозволяє моделі ефективніше захоплювати високочастотні аномалії без надмірного збільшення кількості параметрів, що робить її пріоритетним кандидатом для практичного застосування у високонавантажених системах.

Таблиця 2.1

Порівняння архітектур ЗНМ-частини для стегоаналізу

Архітектура	Параметри	Точність	Ефективність	Витрати пам'яті	Навчання
ResNet50	~25.6M	висока	висока	базова	повільне
MobileNetV2	~3.4M	середня+	висока	низька	швидке
EfficientNet-B0	~5.3M	висока	найвища	низька	швидке
SRNet (повна)	~15M	висока	висока	середня	середнє
SRNet (спрощ.)	~10M	середня+	середня	низька	швидке

- MobileNetV2 для обмежених ресурсів: Дана архітектура може бути обрана у сценаріях із жорсткими вимогами до обчислювальних ресурсів, наприклад, при розгортанні на мобільних пристроях або вбудованих системах моніторингу трафіку. Використання інвертованих залишкових блоків (Inverted Residuals) мінімізує об'єм оперативної пам'яті, хоча це призводить до певної втрати точності (на 7–10% порівняно з EfficientNet) через меншу ємність ознак.
- ResNet50 як надійна еталонна архітектура: ResNet50 розглядалася як одна з основних архітектур для обчислювальних експериментів. Завдяки концепції залишкового навчання (Residual Learning), вона стабільно демонструє високі результати на складних текстурних зображеннях. Проте значна кількість параметрів та відносно низька швидкість інференсу роблять її менш ефективною порівняно з

сучасними масштабованими архітектурами в умовах обмеженого часу обробки.

Таким чином, вибір конкретної архітектури в рамках гібридної моделі має базуватися на специфіці середовища розгортання: від максимальної точності аналізу (ResNet або EfficientNet) до максимальної швидкодії на термінальних пристроях (MobileNet).

2.3.5. Спеціалізована архітектура SRNet та її модифікації

Як альтернативна до двоблокової моделі з ResNet50/MobileNetV2/EfficientNet-B0 розглядається архітектура SRNet (Steganalysis Residual Net), спеціально розроблена для задач стеогоаналізу. SRNet включає чотири типи конволюційних блоків з різними стратегіями обробки просторової інформації та вбудованими залишковими з'єднаннями.

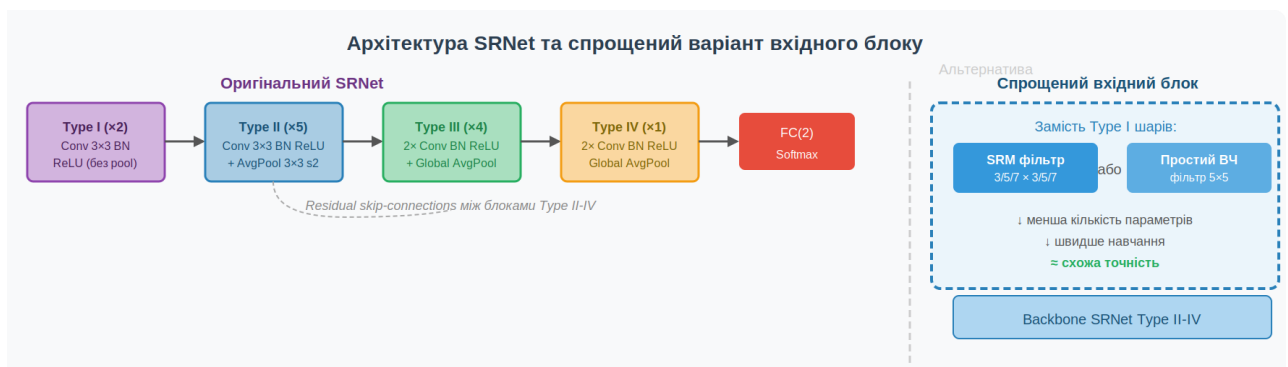


Рис. 2.4. Архітектура SRNet та спрощений варіант вхідного фільтруючого блоку

Оригінальна архітектура SRNet включає чотири типи блоків:

- Type I — два згорткові шари без субдискретизації; призначені для вилучення залишкових ознак без зменшення просторового розміру;
- Type II — згортковий шар із блоком Average Pooling 3×3 зі страйдом 2; реалізує поступове зменшення розміру карт ознак;
- Type III — два послідовних згорткових шарів з глобальним усередненням; призначені для агрегації просторових ознак;
- Type IV — аналогічний Type III, фінальний блок перед класифікатором.

Між блоками Type II–IV реалізовані залишкові skip-з'єднання, що покращують градієнтний потік під час навчання. Загальна кількість параметрів складає приблизно 15 мільйонів. SRNet спочатку використовував фіксовані SRM-фільтри у Type I блоках.

У модифікованому варіанті оригінальні Type I блоки замінено спрощеним входним фільтруючим шаром. Розглядаються два підходи до спрощення:

- заміна Type I на простий ВЧ фільтр 5×5 з фіксованим ядром (аналогічно варіанту 1 основної моделі);
- заміна на SRM-фільтри зменшеного розміру з ядром 3×3 замість 5×5 .

Спрощення входного блоку зменшує кількість параметрів приблизно до 10 мільйонів при незначній втраті точності. Backbone-частина (Type II–IV блоки) залишається незмінною. Такий підхід дозволяє використовувати переваги спеціалізованої архітектури SRNet при зниженні обчислювальної складності.

SRNet як спеціалізована архітектура для стегоаналізу демонструє переваги при навчанні на цільових наборах даних завдяки вбудованій ієрархічній обробці залишкового сигналу. Разом із тим, двоблокова модель на основі EfficientNet-B0 або ResNet50 забезпечує більшу гнучкість у виборі ВЧ блоку та можливість використання передавчого навчання (transfer learning) від ImageNet, що є особливо важливим за обмеженого обсягу тренувальних даних.

2.4 Дослідження ефективності різних архітектур двошарової моделі

2.4.1 Загальна схема навчання моделей

Всі варіанти архітектур навчалися за єдиною схемою з метою забезпечення порівнянності результатів. Вхідні зображення нормалізувалися до діапазону $[0, 1]$ перед поданням у ВЧ блок. Функція втрат — бінарна крос-ентропія. Оптимізатор — Adam із початковою швидкістю навчання 10^{-4} та поступовим зниженням за схемою ReduceLROnPlateau.

Для варіантів з тренувальними ВЧ фільтрами використовувалася різна швидкість навчання для HPF-блоку та backbone-мережі (HPF-блок навчається із меншою швидкістю для збереження ВЧ властивостей). При використанні ResNet50, MobileNetV2 та EfficientNet-B0 ваги ініціалізувалися передавчим навчанням від ImageNet, тоді як SRNet навчався з нуля.

Розмір пакету даних (batch size) складав 32 для ResNet50 та SRNet, та 64 для MobileNetV2 і EfficientNet-B0. Усі моделі навчалися протягом 100 epoch із ранньою зупинкою за точністю на валідаційній вибірці (терпіння 15 epoch).

2.4.2 Побудова набору даних для навчання і перевірки моделей

Для вбудовування текстових повідомлень було використано відомий набір даних CIFAR10 [85-86].

CIFAR-10 — це широко використовуваний датасет у комп'ютерному зорі, який містить 60 000 кольорових зображень розміром 32x32 пікселів у 10 класах.

Ключові характеристики CIFAR-10 як середовища для стеганографії:

- Обмежена роздільна здатність зображень (32x32 пікселі) створює специфічні умови для приховування даних. Навіть невелике за обсягом повідомлення призводить до того, що відносний показник вбудовування (payload capacity) швидко зростає, досягаючи значних величин у бітах на піксель (bpp). Це робить CIFAR-10 ідеальним полігоном для тестування детекції «насиченого» вбудовування в умовах дефіциту контейнерного простору.
- Наявність повноколірних зображень забезпечує 3 байти (24 біти) інформації для кожного пікселя. Для стеганографії це означає можливість незалежного або комбінованого маніпулювання трьома каналами одночасно, що суттєво підвищує загальну ємність контейнера. Стегоаналітичні моделі отримують змогу аналізувати не лише просторові, а й міжслойні (крос-канальні) кореляції, які

порушуються при вбудовуванні.

- Збереження зображень у форматах PNG або raw гарантує відсутність артефактів квантування, характерних для JPEG-стиснення. Це забезпечує «чисте» середовище для методів LSB-заміщення, оскільки будь-яка зміна молодшого біта не маскується помилками стиснення. Проте це також полегшує роботу детекторам, оскільки статистичні девіації пікселів у таких контейнерах мають більш виражений і прогнозований характер.
- Через невеликий розмір об'єктів на зображеннях, нейромережеві моделі (такі як MobileNetV2 або ResNet) змушені фокусуватися на мікроструктурах і статистиці залишків, а не на глобальних семантичних ознаках. У поєднанні з блоками HPF-фільтрації це дозволяє ефективно ідентифікувати стеганографічний шум навіть на фоні складних текстур.

На думку [87-89] CIFAR-10 є хорошим джерелом для зображень для побудови великого навчального набору (120 000 фотографій), який було використано для порівняння ефективності різних AI/ML моделей у виявленні прихованих повідомлень.

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) [90-91]. 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.

Ключові переваги LSB-методів при генерації датасетів [1,67]:

- Ізоляція стего-шуму: LSB-заміна (або її адаптовані варіанти) в просторовій області на нестиснутих носіях (наприклад, файли PNG або PPM з BOSSBase) дозволяє створити синтетичний датасет, на якому

єдиною суттєвою відмінністю між зображенням-носієм та стего-зображенням є дисторсія, внесена LSB. Це дозволяє нейронній мережі сфокусуватися виключно на вивченні залишків LSB-шуму.

- LSB є найпростішим для виявлення методом. Навчання CNN на LSB-зразках гарантує, що модель спочатку засвоїть основні принципи виділення шумів за допомогою HPF-фільтрів, перш ніж переходити до більш складних, адаптивних алгоритмів (наприклад, WOW або S-UNIWARD).
- LSB-вбудовування може забезпечити високу щільність (наприклад, 1.0 біт на піксель або 0.5 bprp), яка необхідна для забезпечення чіткого сигналу помилки для CNN.

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) [92]. 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 зображення перетворювались на розмір 96x96.

2.4.3 Проведення обчислювальних експериментів з легковажними архітектурами

Для проведення експериментів були використані моделі наступної структури (рис. 2.5):

- Блок попередньої обробки (один з розглянутих варіантів);
- Легковажна конволюційна мережа (MobileNetv2 або EfficientNetV2S);
- Шар GlobalAveragePooling і щільний шар з активацією «сігмоїд»;
- Блок виводу ілюстрацій і перевірки відновлення вбудованого тексту.

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

Зображення з датасету Cifar10 $32 \times 32 \times 3$ перед вбудовуванням прихованого тексту перетворювались на зображення $96 \times 96 \times 3$. 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).

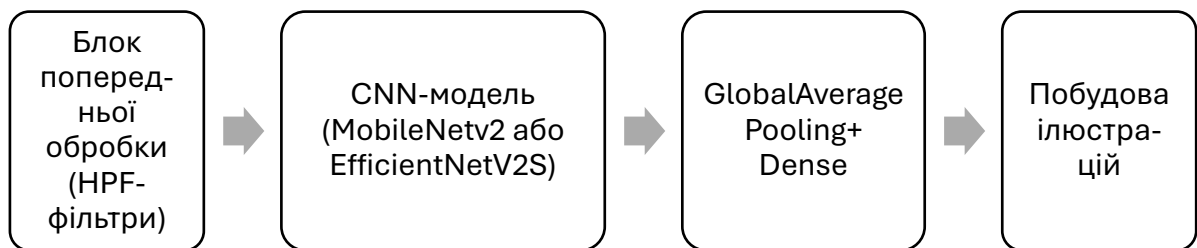


Рис. 2.5 Архітектура моделі стегоаналізу з використанням легковажних CNN

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.

Результати експериментів показали, що без використання HPF-препроцесінгу навіть при високих значеннях обсягу вбудовування виявлення прихованого тексту не досягається.

Для забезпечення можливості виявлення вбудованих артефактів треба використати хоча б один HPF-фільтр. Але навчання моделі з зовсім низькою кількістю фільтрів дуже швидко призводить до помітного перенавчання моделі (рис. 2.6).

Збільшення кількості фільтрів до 5 і більше значно покращує можливості навчання моделі і стійкість результату. Приклад кривих

навчання моделі з 5 фільтрами на вході наведено на рис. 2.7. Але цей результат не стійкий і дуже залежить від кількості зображень в навчальному датасеті і може змінитись при додаванні або видаленні лише одного фільтра.

Стійкий результат з точки зору відсутності перенавчання був отриманий лише для 45 фільтрів.

Якщо групувати спрямовані ядра 3×3 (горизонтальні, вертикальні, діагональні SRM-фільтри), або 5×5 чи 7×7 , стійкий результат з точки зору перенавчання моделі досягається при наявності не менш 5 груп фільтрів.

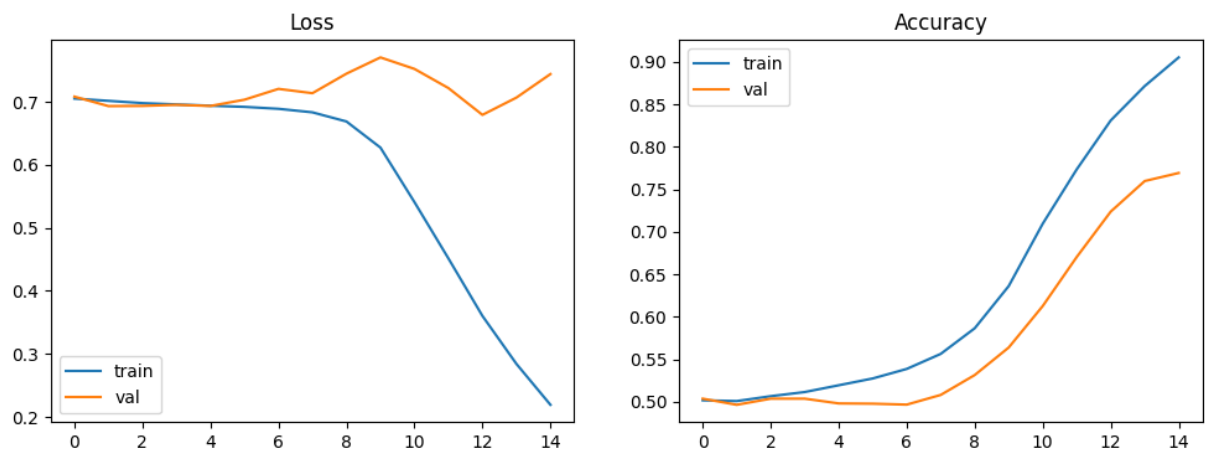


Рис. 2.6 Процес навчання моделі MobileNetV2 з одним фільтром 5×5 у початковому блоці

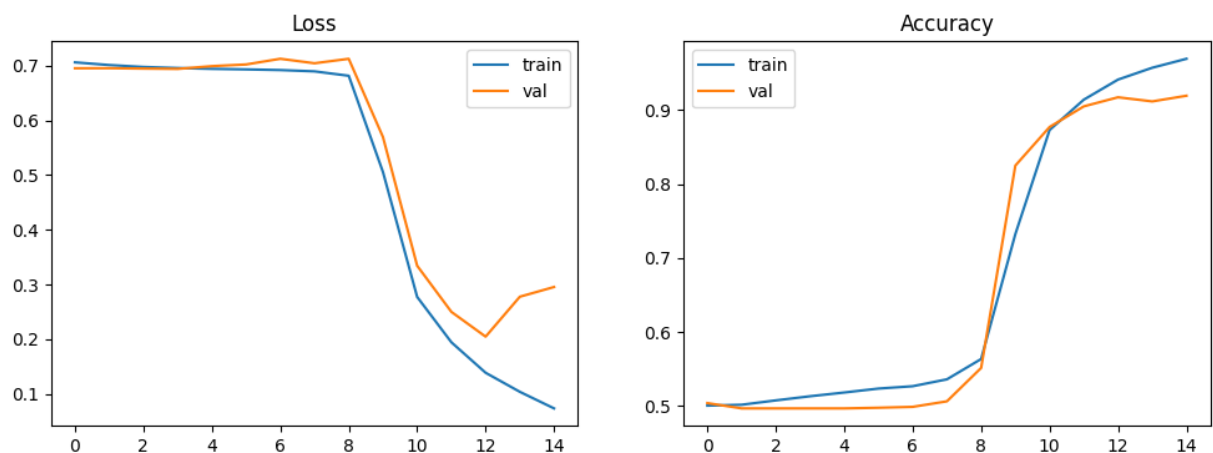


Рис. 2.7 Процес навчання моделі MobileNetV2 з 5 фільтрами 5×5 у початковому блоці (набір даних з 60000 зображень)

Порівняння результатів для моделей з фіксованими фільтрами та

моделей з можливістю 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.

ROC-криві продемонстрували стабільне зростання показника AUC при використанні тренуваних HPF-шарів, досягаючи значень $AUC \approx 0.998$, що вказує на високу роздільну здатність запропонованого детектора (рис. 2.8).

Таким чином, експериментальні результати демонструють, що використання високочастотного препроцесінгу у поєднанні з MobileNetV2 суттєво покращує ефективність стегоаналізу. Зокрема, HPF-фільтри з можливістю тренування дозволяють моделі адаптуватися до характеру стего-шуму, що забезпечує приріст точності класифікації до 95-97% та значення $AUC = 0.998$. Отримані результати підтверджують доцільність поєднання апріорних знань (SRM-фільтри) з можливостями глибокого навчання.

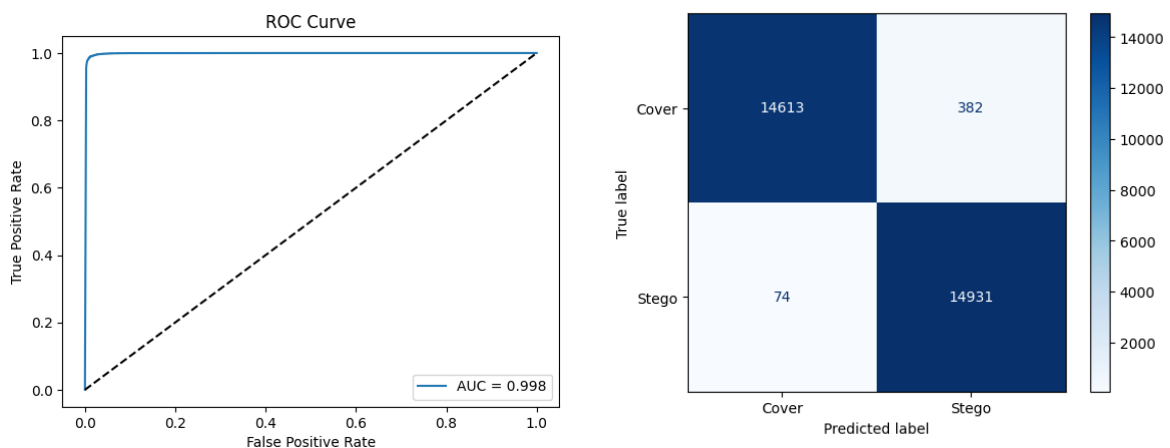


Рис. 2.8 Оцінка якості навчання моделі стегоаналізу на базі архітектури MobileNetV2 з 15 групами спрямованих ядер 3x3 (набір даних з 60000 зображень)

При зміні архітектури класифікатору на модель EfficientNetV2S висновки про архітектуру блоку попередньої обробки зображень залишились без змін.

2.4.4 Проведення обчислювальних експериментів з архітектурами на основі ResNet та SENet

Для проведення експериментів були використані моделі наступної структури (рис. 2.9):

- 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.

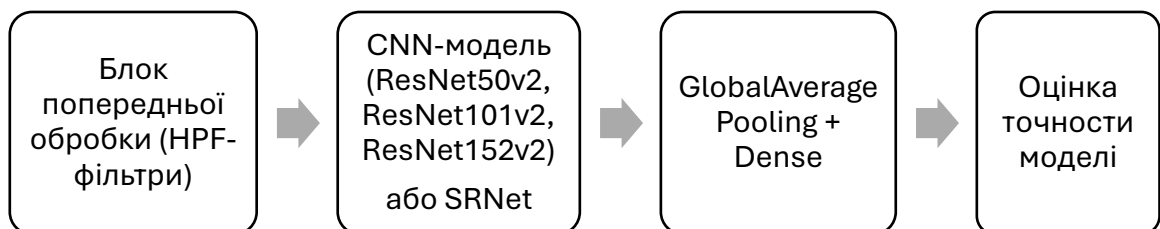


Рис. 2.9 Архітектура моделі стегааналізу з використанням глибоких CNN із залишковими блоками

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 інші варіанти) 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.

Далі 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 шарів. Вбудовування інформації здійснювалося методом LSB із підтримкою UTF-8 кодування та керованою пропускну 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. 2.10 і рис. 2.11.

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 аварійного припинення обчислень внаслідок цілковитого заповнення оперативної пам'яті.

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

Попереднє дослідження було виконано з використанням декількох груп HPF-фільтрів. Кожна група містила 9 орієнтованих фільтрів 5x5, кількість груп змінювалась від 1 до 9.

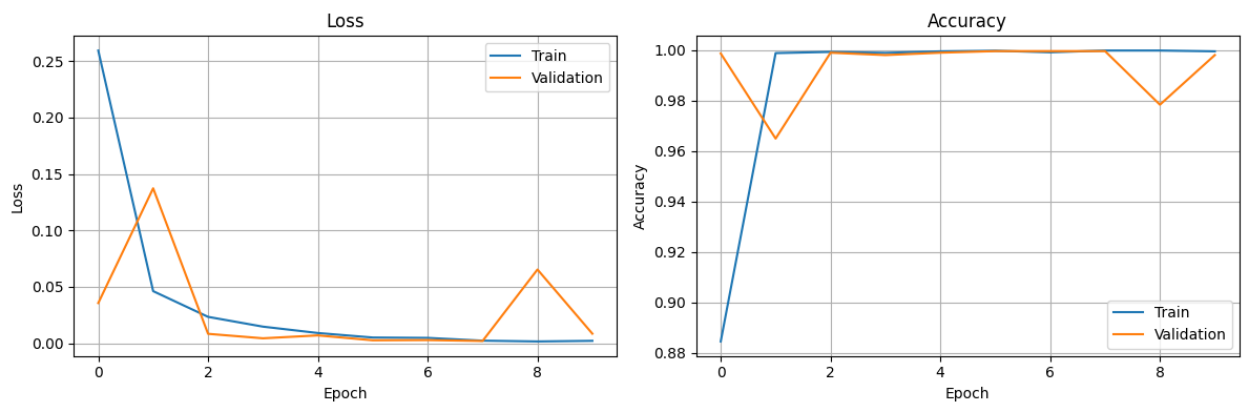


Рис. 2.10 Криві навчання моделі стегааналізу з архітектурою SRNet за даними з $\text{payload}=0,002$.

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. 2.12-2.13).

Як видно з рис. 2.12, 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.

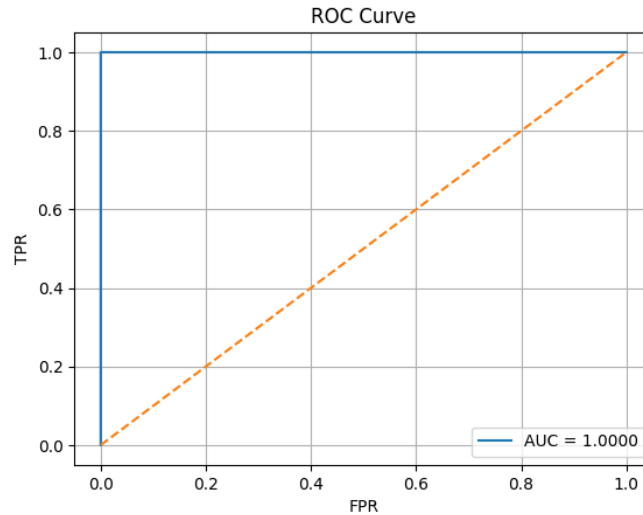


Рис. 2.11 Крива ROC/AUC моделі стегоаналізу з архітектурою SRNet, яку було навчено за даними з $\text{payload}=0,002$

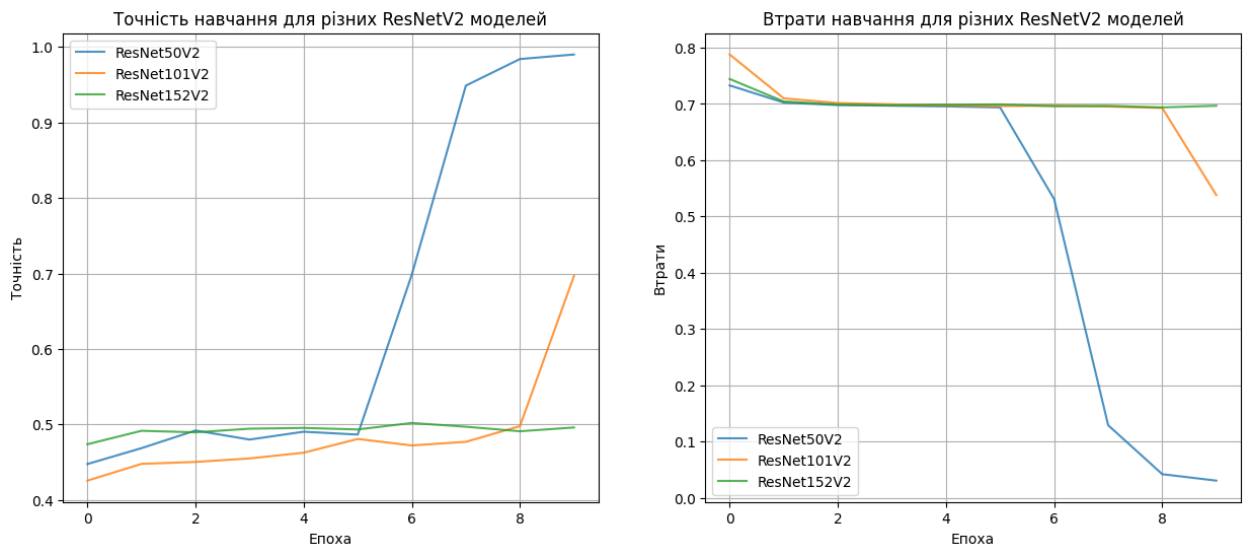


Рис. 2.12 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 5 шарами фільтрів 5x5 при навчанні за набором даних з $\text{payload}=0,002$

Але час навчання моделей послідовно збільшувався при переході від ResNet50v2 до ResNet101v2 і потім ResNet152v2. Більш високі значення

payload (bpr=0,2 або bpr=0,4) значно спрощують виявлення стеганографії і зменшують вимоги 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 забезпечує повного виявлення ознак стеганографічного вбудовування і точність моделі не перевищувала 60-65% незалежно від архітектури моделі.

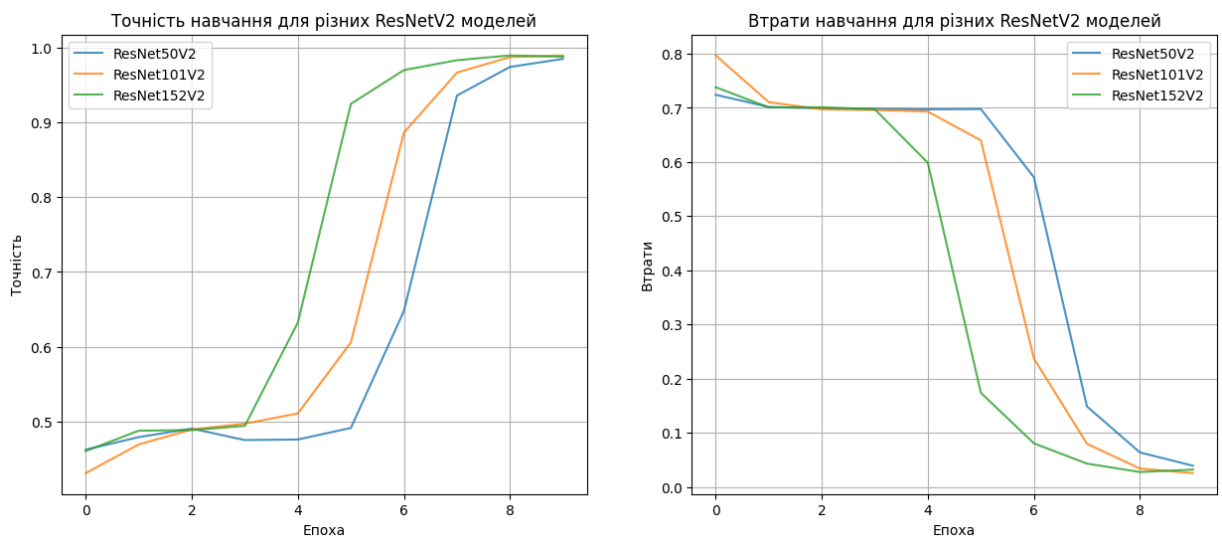


Рис. 2.13 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 3 шарами фільтрів 5x5 при навчанні за набором даних з payload=0,002

Послідовне використання двох мультимасштабних фільтрів забезпечило успішне навчання моделей з усіма варіантами архітектури ResNet (рис. 2.14). Потрійне використання блоку мультимасштабних

фільтрів забезпечило швидке і надійне навчання моделі (рис. 2.15).

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

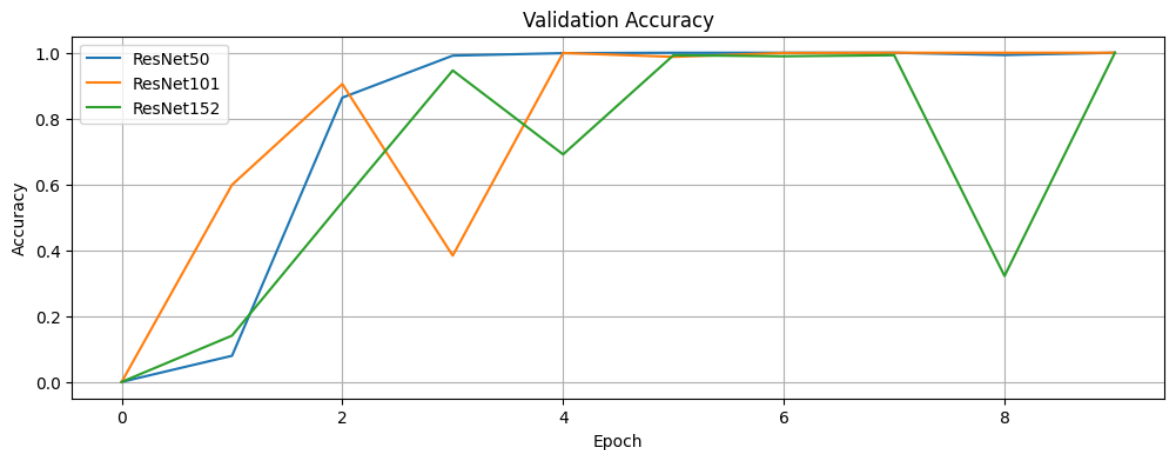


Рис. 2.14 Криві навчання моделі стегоаналізу з архітектурою ResNetv2 та вхідним блоком з 2 шарами мультимасштабних фільтрів при навчанні за набором даних з $\text{payload}=0,002$

Порівняння результатів для моделей з фіксованими фільтрами та моделей з можливістю навчання фільтрів показало, що дозвіл на адаптацію HRF-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 .

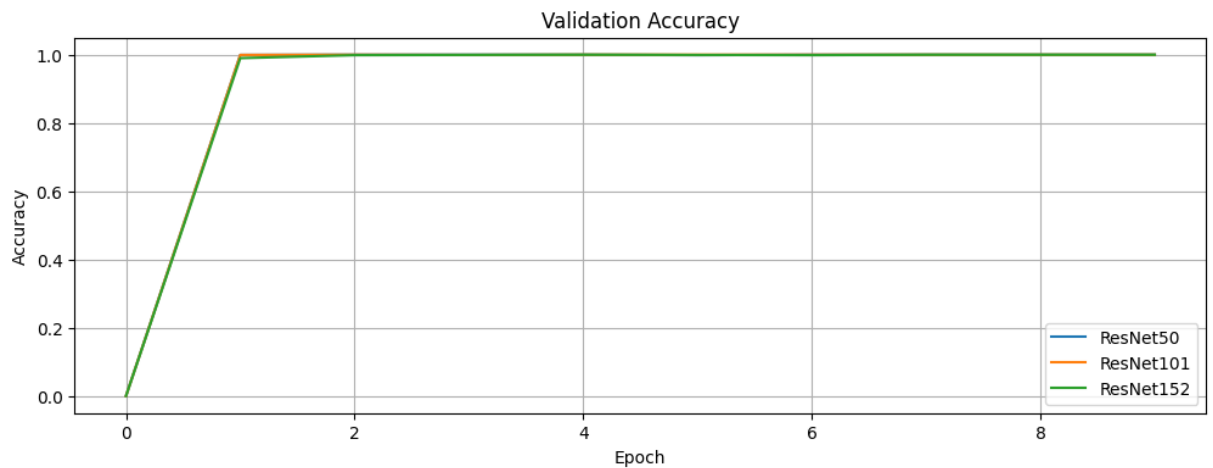


Рис. 2.15 Криві навчання моделі стегааналізу з архітектурою ResNetv2 та вхідним блоком з 3 шарами мультимасштабних фільтрів при навчанні за набором даних з $\text{payload}=0,002$

Таким чином, експериментальні результати демонструють, що використання високочастотного препроцесінгу у поєднанні з ResNetV2 суттєво покращує ефективність стегааналізу. Зокрема, HPF-фільтри з можливістю тренування дозволяють моделі адаптуватися до характеру стега-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'яті).

Висновки за розділом 2

1. Обґрунтовано та розроблено двокомпонентну архітектуру моделі стегааналізу, яка поєднує блок високочастотної (ВЧ) попередньої обробки для виділення слабких сигналів у шумовому залишку та глибокий неймережевий класифікатор.
2. Досліджено п'ять варіантів організації блоку ВЧ фільтрів (HPF Layer),

серед яких найбільш ефективним визначено мультимасштабний підхід із паралельним використанням направлених ядер SRM розмірами 3x3, 5x5 та 7x7 пікселів.

3. Встановлено, що використання тренувальних ВЧ фільтрів замість фіксованих забезпечує приріст якості детекції на 8–12%, оскільки дозволяє моделі адаптуватися до специфічного характеру стеганографічного шуму.
4. Проведено порівняльний аналіз архітектур класифікаторів (ResNet50, MobileNetV2, EfficientNet-B0), за результатами якого EfficientNet-B0 визначено як пріоритетне рішення завдяки найкращому балансу між точністю та обчислювальною складністю (~5,3 млн параметрів).
5. Експериментально підтверджено, що без спеціалізованого HPF-препроцесингу виявлення прихованого тексту в цифрових зображеннях практично неможливе, а точність класифікації залишається на рівні випадкового вгадування.
6. Доведено, що для забезпечення стійкого навчання та запобігання перенавчанню моделі необхідно використовувати не менше 5 груп спрямованих фільтрів або понад 45 одиничних фільтрів у початковому блоці.
7. Виявлено, що спеціалізована архітектура SRNet забезпечує швидке навчання та високу точність, проте висуває жорсткі вимоги до обсягу оперативної пам'яті, що обмежує розмір навчальної вибірки в ресурсообмежених середовищах, таких як безкоштовна версія Google Colab.
8. Показано, що застосування переднавчених ваг ImageNet та методу тонкого налаштування (fine-tuning) суттєво прискорює процес навчання глибоких архітектур типу ResNetV2 та дозволяє досягати значень точності до 99,8%.

РОЗДІЛ 3. АРХІТЕКТУРА ГІБРИДНОЇ СИСТЕМИ СТЕГОАНАЛІЗУ НА ОСНОВІ CNN ТА ВЕЛИКИХ МУЛЬТИМОДАЛЬНИХ МОВНИХ МОДЕЛЕЙ

3.1. Концептуальна модель гібридного стегоаналізу

Запропонована система об'єднує дві принципово різні парадигми аналізу зображень: детерміновану статистичну обробку через згорткові нейронні мережі (CNN) та ймовірнісне семантичне міркування через великі мультимодальні мовні моделі (MLLM).

Механізм злиття рішень (Decision Fusion Mechanism, DFM) є центральним елементом гібридної системи стегоаналізу, що вирішує задачу об'єднання принципово різнорідних сигналів: статистично детермінованих ймовірнісних виходів згорткової нейронної мережі (ЗНМ) та семантично збагачених рішень мультимодальної великої мовної моделі (MLLM). Принципова відмінність від традиційних ансамблевих підходів полягає у тому, що компоненти гібридної системи функціонують у різних просторах ознак — ЗНМ оперує піксельними статистиками, тоді як MLLM — семантичними концептами — і мають різну природу виходів: детерміновані ймовірності проти ймовірнісних природно-мовних тверджень [67].

Загальна архітектура системи наведена на рис. 3.1.

Обидва компоненти гібридної системи — ЗНМ та MLLM — отримують на вхід оброблений сигнал від спільного блоку попередньої фільтрації (блоку HPF-фільтрів), що є принциповим архітектурним рішенням дослідження. Розміщення HPF на вході системи до розгалуження на два потоки забезпечує узгодженість вхідних даних: обидва компоненти аналізують одні й ті самі шумові портрети зображення, що унеможливорює розбіжності, зумовлені різним поданням вхідних даних [23].

Розглянемо трохи спрощений приклад, коли багат шаровий HPF-препроцесор реалізовано у вигляді банку з 12 ядер фільтрації розміром 5×5 . До першої групи належать вісім направлених ядер SRM (Steganalysis Rich

Model) [7]: по два у горизонтальному, вертикальному, діагональному та антидіагональному напрямках, призначених для виявлення розривів у статистичних зв'язках між сусідніми пікселями. До другої групи належать чотири ядра на основі оператора Лапласа з різними наборами вагових коефіцієнтів, що дозволяє аналізувати зони з різною інтенсивністю природного шуму.

Для формування вихідного тензора препроцесора застосовується депт-вайз (depthwise) згортка: кожен з 12 фільтрів застосовується незалежно до кожного з трьох RGB-каналів зображення, що дає 36 незалежних шумових карт ($12 \times 3 = 36$ каналів). Функція активації ABS (обчислення за модулем) замінює стандартну ReLU, що принципово важливо для стеганалізу: операція $|\cdot|$ зберігає повний обсяг амплітудних відхилень як у позитивному, так і у від'ємному діапазонах, забезпечуючи вдвічі більший обсяг інформативних ознак для подальшого аналізу [23].

Для ЗНМ-компонента 36-канальний тензор є безпосереднім вхідним сигналом, що оброблюється нейронними шарами відповідно до обраної архітектури. Для MLLM-компонента передаються як 36-канальний тензор (через шар-адаптер, що перетворює його у візуальні токени), так і оригінальне RGB-зображення — для формування семантичного контексту, недоступного з одних лише шумових карт.

В кінцевій версії архітектури блоку HPF-фільтрів була використана більш складна архітектура з декількома направленими блоками різної розмірності, що збільшило кількість каналів і розмірність тензору.

ЗНМ-компонент приймає 36-канальний тензор від HPF-препроцесора та формує два вихідних сигнали: скалярну ймовірність \hat{r}_{CNN} та вектор ознак f_{CNN} , що передається до мета-класифікатора. Три підтримувані архітектури (таблиця 3.1) представляють різні підходи до масштабування з різними компромісами між точністю та обчислювальною вартістю.

У всіх трьох архітектурах ваги ініціалізувалися попередньо навченими на ImageNet без до-навчання: ЗНМ виступає фіксованим аналітичним

сенсором, що вилучає статистичні девіації пікселів.

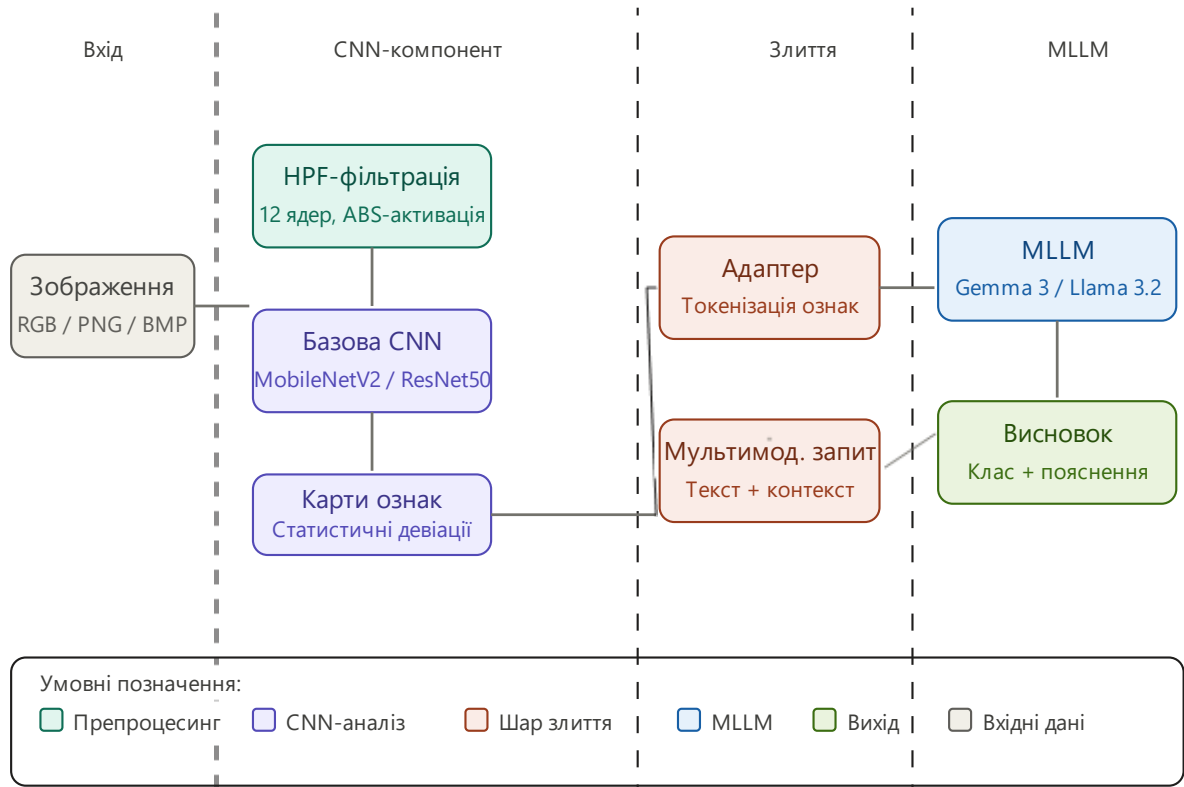


Рис. 3.1. Загальна архітектура гібридного конвеєра стегааналізу (натисніть на блок для детального пояснення)

Таблиця 3.1

Характеристика ЗНМ-архітектур у контексті задачі стегааналізу

Архітектура ЗНМ	Кількість параметрів	Ключовий механізм	Вихідний вектор ознак	Перевага для стегааналізу
ResNet50	25.6 М	Залишкові з'єднання (skip)	2 048-вим. → FC → 128	Глибокий аналіз субтильних LSB-аномалій
MobileNetV2	3.4 М	Інвертовані залишк. блоки (DSC)	1 280-вим. → FC → 128	Мінімальний VRAM, висока швидкість
EfficientNet-B0	5.3 М	Складене масштабування	1 280-вим. → FC → 128	Баланс глибини/ширини, найкращий transfer

Вихідний шар кожної архітектури замінюється на проекційний блок: GlobalAveragePooling2D → Dense(128, ReLU) → Dense(1, sigmoid), що дає

скалярну ймовірність $\hat{p}_{\text{CNN}} \in [0, 1]$. Паралельно вектор з 128 нейронів після першого Dense-шару зберігається як fCNN для подальшого злиття на рівні ознак [43].

Таке поєднання усуває ключове обмеження кожного підходу окремо: CNN здатні виявляти мікроскопічні числові аномалії, але позбавлені здатності до контекстуального пояснення, тоді як MLLM наділені потужними можливостями логічного обґрунтування, проте не можуть безпосередньо аналізувати субпіксельні статистичні відхилення.

Конвеєр обробки складається з чотирьох послідовних етапів.

На першому етапі вхідне зображення подається до блоку фільтрації високих частот, де усувається основний візуальний контент та виокремлюються мікроскопічні аномалії пікселів.

Другий етап — автоматична екстракція ознак за допомогою однієї з трьох CNN-архітектур (MobileNetV2, ResNet50, EfficientNet B0), що формують багатоканальні карти статистичних девіацій.

Третій етап є центральним з точки зору гібридизації: шар-адаптер перетворює числові тензори CNN у візуальні токени, зрозумілі для мовної моделі, та об'єднує їх з текстовим запитом і контекстом задачі.

Четвертий етап — формування висновку мовною моделлю, що охоплює як бінарну класифікацію (наявність / відсутність вбудовування), так і розгорнуте текстове пояснення виявлених аномалій.

3.2. Архітектура CNN-компонента

Для реалізації технічного аналізу у середовищі TensorFlow/Keras обрано три архітектури, що представляють різні підходи до масштабування нейронних мереж.

Архітектура MobileNetV2 реалізує концепцію інвертованих залишкових блоків, що використовують лінійні вузькі місця для ефективної обробки даних. Така структура передбачає початкове розширення простору ознак перед застосуванням глибинно-роздільних згорткових шарів, що

забезпечує високу продуктивність моделі при мінімальних обчислювальних витратах. На відміну від традиційних залишкових блоків, ця архітектура спочатку розширює вхідний простір ознак у глибину за допомогою 1×1 згортки, після чого застосовує ефективну depthwise-згортку для фільтрації ознак, що дозволяє значно мінімізувати обчислювальну складність при збереженні високої виразності моделі.

ResNet50 реалізує фундаментальну концепцію залишкового навчання (Residual Learning), де ключовим елементом є використання skip-з'єднань (прямих зв'язків), що дозволяють градієнтам вільно поширюватися через усі 50 шарів мережі. Це архітектурне рішення ефективно нівелює проблему згасаючого градієнта, яка зазвичай виникає при спробі навчання дуже глибоких нейронних мереж.

У контексті стегоаналізу така здатність до глибокого проходження сигналу є безальтернативною для аналізу дрібних артефактів LSB-стеганографії, де корисний сигнал є надзвичайно слабким і часто губиться за семантичним шумом при стандартній обробці. Завдяки ієрархічній структурі, ResNet50 здатна виявляти складні кореляції між пікселями, що виникають при використанні адаптивних методів вбудовування, забезпечуючи високу стабільність результатів.

EfficientNet B0 використовує передовий метод складеного масштабування (Compound Scaling), який базується на математичному обґрунтуванні одночасної оптимізації трьох ключових вимірів мережі: глибини, ширини (кількості каналів) та роздільної здатності вхідних даних. Це забезпечує значно кращий баланс між точністю класифікації та обчислювальними витратами порівняно з традиційними архітектурами, які масштабують лише один із цих параметрів.

Порівняльна характеристика розглянутих архітектур наведена на рис. 3.2.

Для задач стегоаналізу це означає, що модель здатна ефективно фокусуватися на найбільш інформативних зонах зображення,

використовуючи меншу кількість параметрів (~5,3 млн) для досягнення точності, що часто перевершує важчі моделі. Інтегровані блоки Squeeze-and-Excitation додатково посилюють чутливість моделі до специфічних каналів ознак, що містять стеганографічні аномалії.

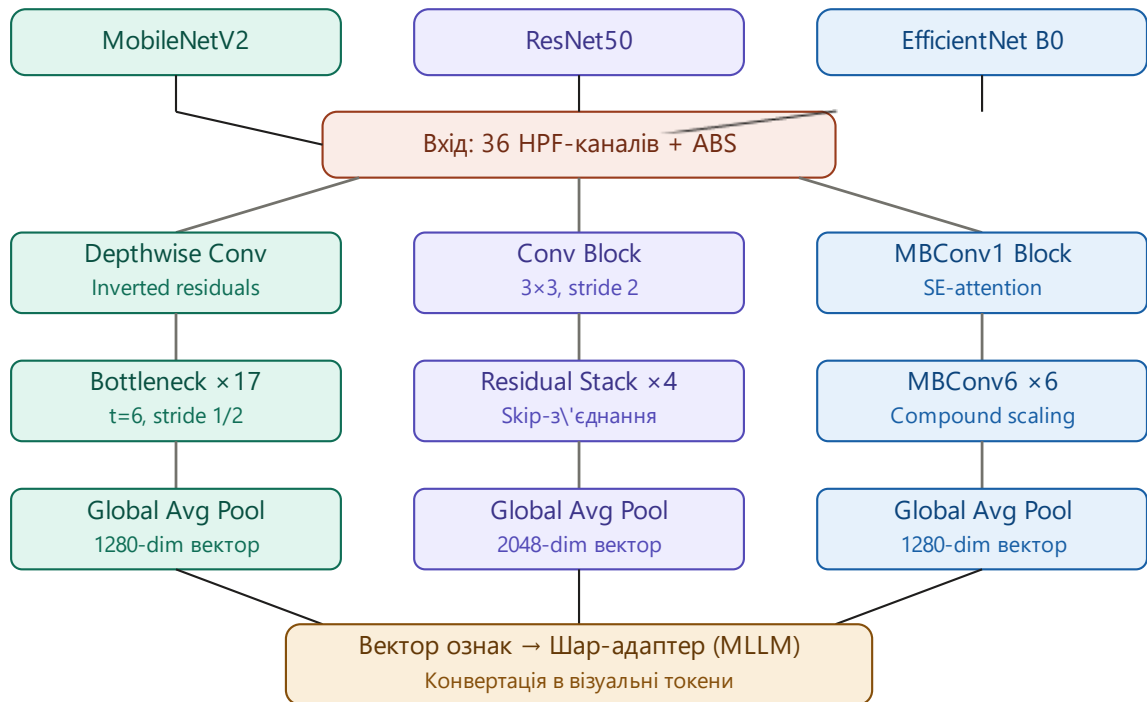


Рис. 3.2. Порівняльна структура трьох CNN-архітектур та їх інтеграція з шаром-адаптером

Усі три розглянуті архітектури ініціалізуються попередньо навченими вагами на наборі даних ImageNet. Таке використання принципу передачі знань (transfer learning) дозволяє моделям задіяти вже сформовані фільтри для розпізнавання базових візуальних структур, фокусуючи навчання лише на специфічних відхиленнях, спричинених вбудовуванням інформації.

Застосування цих моделей без радикального перенавчання (або з мінімальним fine-tuning) дає змогу об'єктивно оцінити якість та репрезентативність ознак, що були отримані виключно завдяки HPF-препроцесингу. Це підтверджує, що комбінація класичних архітектур комп'ютерного зору зі спеціалізованими блоками фільтрації високих частот

є життєздатним та ефективним підходом для виявлення сучасних стеганографічних вкладень.

3.3. Архітектура MLLM-компонента

Ключовою інженерною проблемою гібридної архітектури є перехід між двома принципово несумісними просторами представлення: числовими тензорами CNN та токеними послідовностями мовної моделі. Цей перехід реалізується через шар-адаптер (Adapter Layer), що є центральним елементом архітектури злиття.

3.3.1. Проблема семантичного розриву

CNN оперує у просторі активацій — числових матрицях, де кожне значення представляє ступінь збудження нейрона у відповідь на певну просторову структуру. MLLM оперує у просторі токенів — дискретних або неперервних векторних представленнях мовних та візуальних сутностей, навчених на корпусах текстів і зображень. Пряме поєднання цих просторів неможливе без проміжного перетворення.

3.3.2. Роль шару-адаптера

Шар-адаптер виконує три функції.

По-перше, проекція розмірності: вектор ознак CNN (розмірністю 1280 або 2048) проектується у вимірний простір токенів мовної моделі через лінійне або нелінійне перетворення.

По-друге, нормалізація розподілу: статистичний розподіл активацій CNN суттєво відрізняється від розподілу токенних ембедингів; адаптер вирівнює ці розподіли через batch normalization або layer normalization.

По-третє, семантичне збагачення: числові значення ознак інкапсуюються у структурований текстовий опис, що дозволяє мовній моделі задіяти свої попередньо навчені знання про контекст.

3.3.3. Формування мультимодального запиту

Фінальний запит до MLLM формується як конкатенація трьох компонентів:

1. Візуальних токенів, отриманих від адаптера;

2. Структурованого текстового пром프트 з технічними метриками (значення ентропії, відхилення гістограм, статистики шуму);
3. Системного контексту задачі, що орієнтує модель на домен стегоаналізу.

Використання гібридного підходу дозволяє мультимодальним великим мовним моделям (MLLM) виконувати функцію семантичного арбітражу, що полягає у здатності системи зіставляти виявлені числові аномалії з реальним семантичним змістом зображення. Це забезпечує формування глибоко обґрунтованого висновку, де модель оцінює, чи є виявлений шум наслідком цілеспрямованого втручання, чи він обумовлений природними особливостями сцени, як-от складна текстура або специфічні умови освітлення.

Системний пром프트 для MLLM-компонента розроблено з детальним урахуванням вузької специфіки задачі стегоаналізу. Модель опрацьовує вхідне зображення та шумовий тензор, поданий у вигляді репрезентативної теплової карти, виконуючи аналіз за наступними структурованими напрямками:

- Аналіз статистичних аномалій у гістограмі яскравості молодших бітів передбачає, що модель досліджує характер розподілу значень у площині LSB для виявлення ознак заміщення інформації.
- Виявлення нерегулярностей у просторовому розподілі шуму за допомогою оцінки результатів χ^2 -тесту дозволяє системі ідентифікувати ділянки, де структура шуму суттєво відхиляється від очікуваної.
- Оцінка відхилень від природних стохастичних характеристик шляхом порівняння поточної шумової складової контейнера з еталонними характеристиками «чистих» зображень аналогічного типу дозволяє виявити ознаки стороннього втручання.

Параметри відтворюваності: temperature = 0.0, seed = 42 — гарантують детерміністичні відповіді для ідентичних входів.

Механізм злиття рішень CNN та MLLM наведено на рис. 3.3.

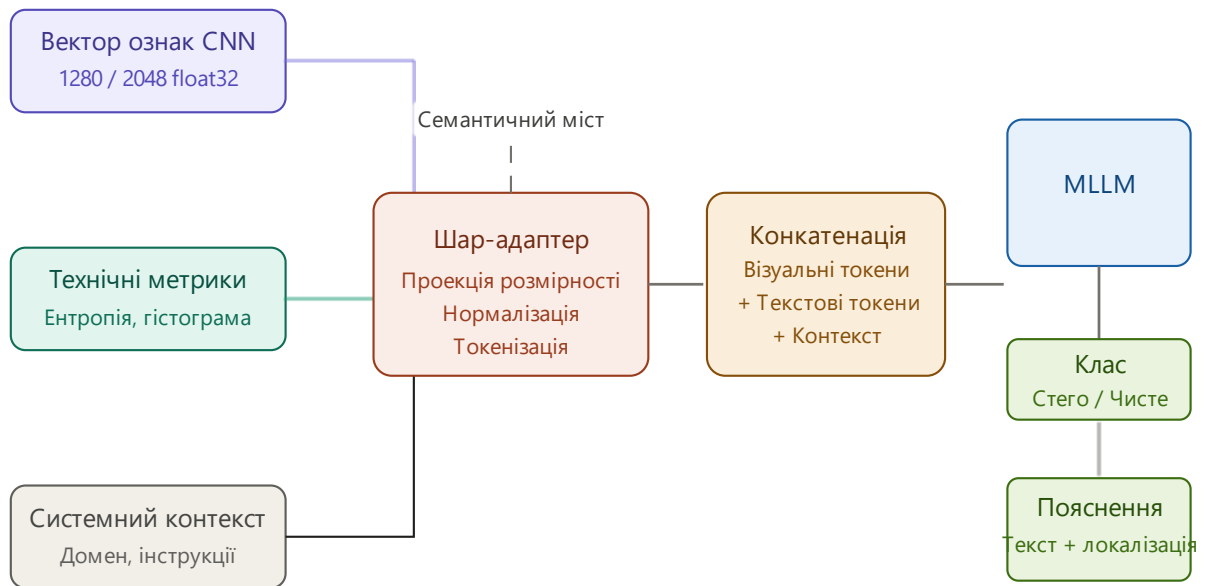


Рис. 3.3. Механізм злиття рішень CNN та MLLM через шар-адаптер

3.4. Порівняльний аналіз MLLM-компонентів

Для реалізації MLLM-компонента гібридної системи обрано три моделі з різними характеристиками, розгорнуті локально через Ollama у середовищі Google Colab.

MLLM-компонент виконує три ключові функції у механізмі злиття:

1. Формує незалежну оцінку \hat{r}_{MLLM} , яка доповнює статистичний сигнал ЗНМ семантичним контекстом;
2. Надає текстовий ембединг e_{MLLM} для злиття на рівні ознак;
3. Генерує природно-мовне обґрунтування рішення, що є унікальною властивістю, недосяжною для традиційних CNN.

3.4.1. Gemma 3 (4B та 12B)

Сімейство Gemma 3 від Google DeepMind побудоване на трансформерній архітектурі з покращеним механізмом групованої уваги запитів (Grouped-Query Attention).

Модель 4B налічує 4 мільярди параметрів та орієнтована на швидку класифікацію з мінімальною затримкою — що критично при пакетній обробці великих наборів зображень.

Модель 12В забезпечує глибше логічне обґрунтування завдяки більшій кількості параметрів і ширшому контекстному вікну, що дозволяє враховувати складніші взаємозалежності між статистичними аномаліями та семантичним змістом зображення.

Порівняльне дослідження між версіями Gemma 3 4В та 12В виявило різні поведінкові стратегії. Версія 4В схильна до прямолінійної класифікації на основі домінуючих статистичних ознак, тоді як версія 12В демонструє здатність враховувати нюанси: наприклад, розрізняти природний шум зображення від артефактів стеганографічного вбудовування, аналізуючи просторовий розподіл аномалій.

3.4.2. Llama 3.2 Vision (11B)

Llama 3.2 Vision від Meta є мультимодальною версією архітектури Llama 3.2, що доповнена спеціалізованим візуальним кодером на основі ViT (Vision Transformer) та крос-модальним шаром уваги. На відміну від Gemma 3, яка переважно орієнтована на текстову обробку з інтегрованими візуальними функціями, Llama 3.2 Vision спроектована для нативної обробки зображень.

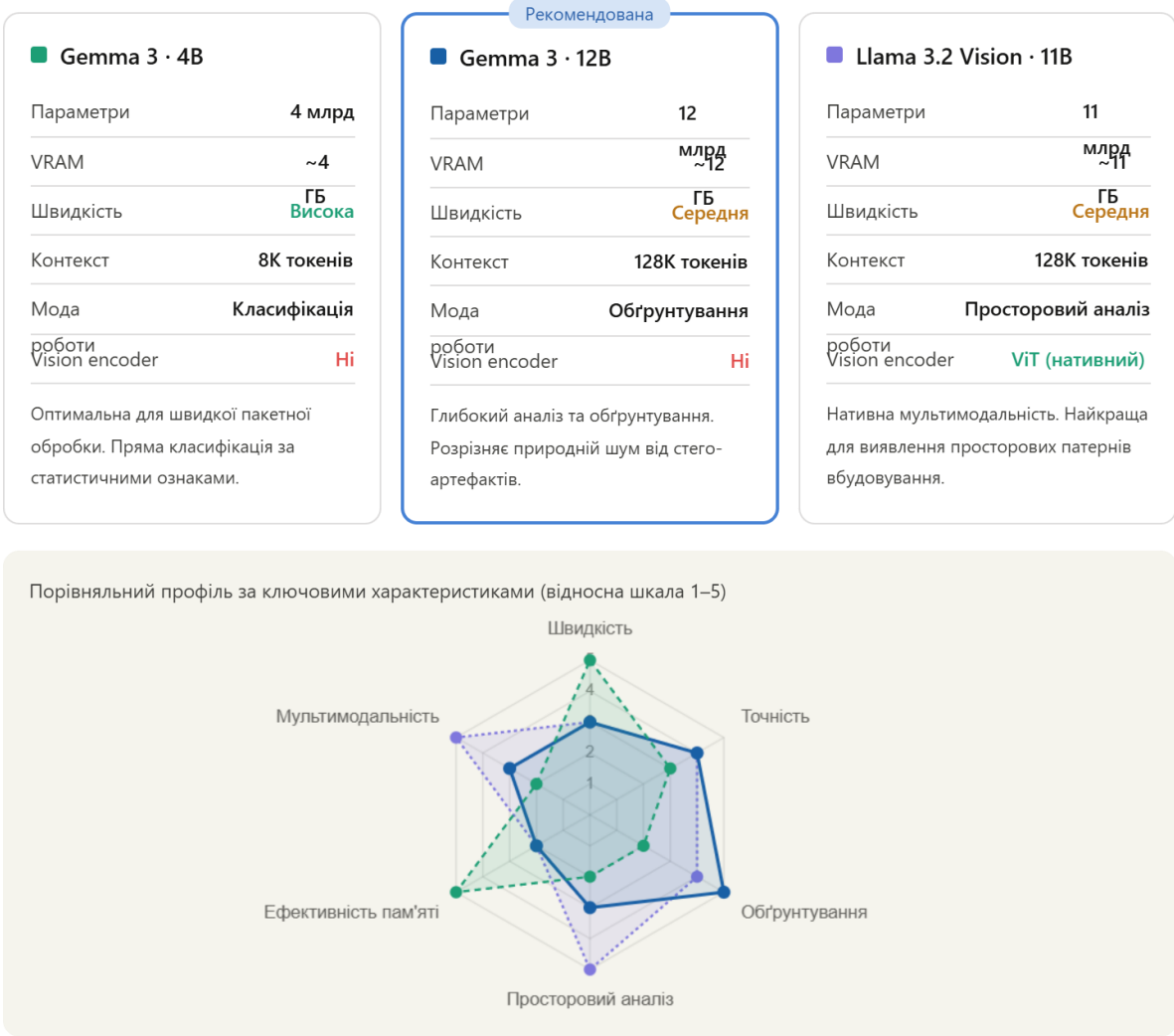
Модель здатна безпосередньо аналізувати передані візуальні карти ознак без потреби у додатковому перетворенні, що суттєво зменшує інформаційні втрати в шарі адаптера та зберігає цілісність мікроструктурних даних. У контексті стегоаналізу це виявляється як критична перевага при дослідженні складних просторових патернів: модель здатна ідентифікувати регулярні структури вбудовування, що розподілені по специфічних частотних зонах зображення.

Порівняння цих моделей у контексті вашої гібридної архітектури стегоаналізу базується на балансі між швидкістю обробки та глибиною семантичного аналізу.

Результати порівняння різних варіантів архітектури MLLM наведено на рис. 3.4.

Ключові архітектурні переваги для задач детекції прихованих сигналів:

- Нативна обробка тензорів ознак: Завдяки вбудованому ViT-кодеру, модель сприймає вихідні дані від блоку HPF-фільтрації не як абстрактні вектори, а як структуровані візуальні патчі.



тензорів високої розмірності у простір текстових ембедінгів.

- Виявлення частотних патернів: Модель демонструє високу чутливість до регулярних структур вбудовування, що характерно для алгоритмів LSB-заміщення, які розподіляють інформацію за певним ключем або сіткою.

Використання Llama 3.2 Vision у гібридному конвеєрі забезпечує синергію між низькорівневим виділенням ознак за допомогою MobileNetV2 або ResNet50 та високорівневою семантичною інтерпретацією, що дозволяє формувати вичерпні експертні звіти щодо ймовірного методу стеганографічного втручання.

3.5. Покращення точності виявлення через гібридизацію

3.5.1. Джерела додаткової точності

Гібридна архітектура забезпечує покращення точності виявлення через три взаємодоповнювальні механізми.

Усунення хибнопозитивних спрацьовувань. CNN-компонент може виявляти статистичні аномалії у текстурованих зображеннях (трава, пісок, тканина), де природній шум формує розподіли, схожі на LSB-вбудовування. MLLM, маючи семантичне розуміння контенту зображення, здатна контекстуалізувати ці аномалії: підвищена ентропія у зоні трави є очікуваною для даного типу контенту і не свідчить про стеганографічне вбудовування. Таким чином, MLLM здійснює контекстуальну корекцію статистичного рішення CNN.

Покращена локалізація вбудовування. Мовна модель, аналізуючи просторовий розподіл аномалій, може визначати зони з підвищеною ймовірністю вбудовування. Наприклад, рівномірний розподіл LSB-аномалій по всьому зображенню характерний для послідовних методів вбудовування, тоді як кластеризований розподіл може вказувати на адаптивні методи, що використовують текстурні зони.

Семантично обґрунтований висновок. На відміну від CNN, що надає

лише числовий прогноз, гібридна система формує текстове пояснення, що описує характер виявлених аномалій, їх просторовий розподіл та ймовірну метод вбудовування. Це критично для застосувань у форензиці, де аналітику необхідне обґрунтування висновку.

Результати порівняльного аналізу покращення точності виявлення стеганографії при переході від ізольованих CNN до гібридних конфігурацій наведено на рис. 3.5.

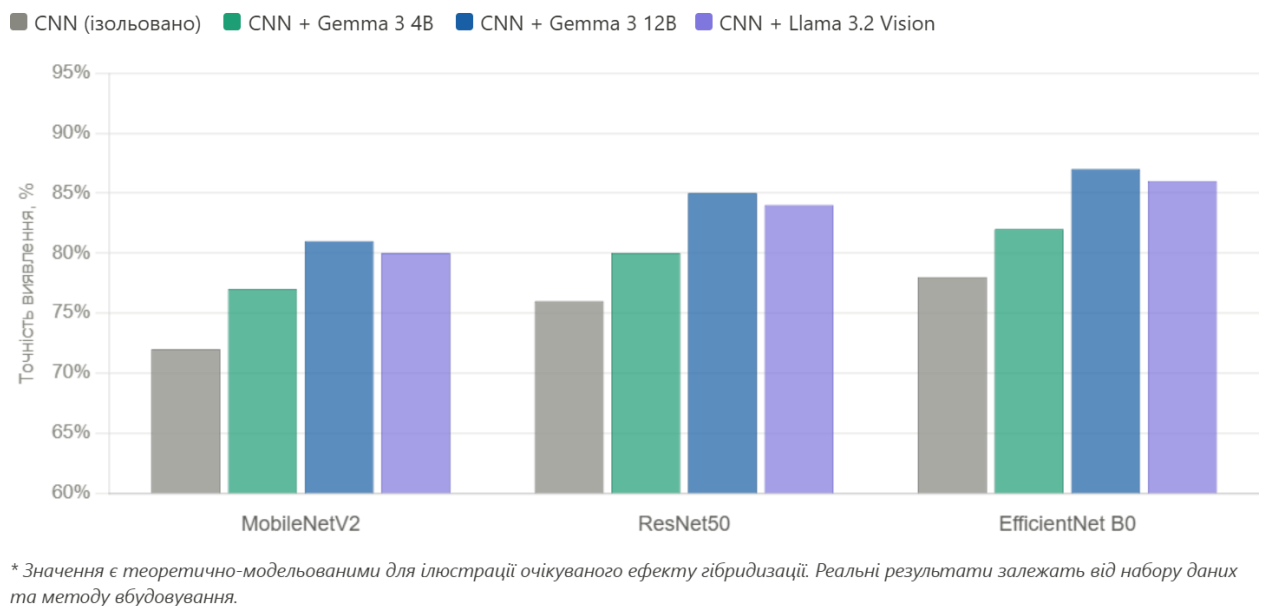


Рис. 3.5. Вплив архітектури MLLM-блока на точності виявлення стеганографії

3.5.2. Обмеження гібридного підходу

Попри високу ефективність, гібридна архітектура стегоаналізу, що поєднує спеціалізовані згорткові мережі (ЗНМ) та великі мультимодальні моделі (MLLM), має низку технічних та експлуатаційних обмежень, які необхідно враховувати при практичному впровадженні.

Детальний аналіз обмежень системи:

- Часові витрати та обчислювальна складність: Залучення MLLM як фінального аналітичного блоку суттєво збільшує загальний час

обробки одного об'єкта порівняно з використанням ізольованої ЗНМ. Якщо класична мережа (наприклад, MobileNetV2) здатна виконувати інференс за мілісекунди, то мультимодальна модель потребує значно більше часу на генерацію текстового висновку. Це створює виражений компроміс: дослідник отримує вищу точність та інтерпретованість результатів ціною зниження пропускну здатності системи, що може бути критичним при обробці великих масивів даних у реальному часі.

- Залежність від семантичної компетенції моделі: Якість семантичного арбітражу безпосередньо залежить від здатності MLLM інтерпретувати специфічний технічний контекст стегоаналізу. Оскільки базові моделі навчаються на загальних даних, вони не завжди можуть коректно пов'язати виявлені статистичні аномалії з конкретними методами будовування без попереднього тонкого налаштування (fine-tuning). Без адаптації модель може надавати поверхневі або загальні описи, що нівелює переваги використання інтелектуального арбітражу.
- Вплив апаратних ресурсів та квантування: При локальному розгортанні через сервер Ollama якість та швидкість роботи системи критично залежать від доступних потужностей GPU та обсягу відеопам'яті. Для роботи з великими моделями (наприклад, Llama 3 11B або Gemma 12B) часто застосовується квантизація (зниження точності ваг моделі до 4- або 8-біт), що дозволяє запускати їх на споживчому обладнанні. Однак інтенсивне квантування може призвести до втрати тонких логічних зв'язків у висновках моделі, створюючи варіативність результатів на різних апаратних конфігураціях.
- Проблема ініціалізації та адаптації: Використання ваг ImageNet для ЗНМ-частини забезпечує гарну базу для розпізнавання образів, проте стегосигнал за своєю природою суттєво відрізняється від об'єктів реального світу. Це вимагає від MLLM високого рівня абстракції, щоб правильно інтерпретувати «шумовий портрет» зображення, сформований після HPF-фільтрації та ABS-активації, як ознаку

цілеспрямованого втручання, а не природного шуму матриці камери.

3.6. Інфраструктурне забезпечення

Локальне розгортання великих мультимодальних мовних моделей (MLLM) за допомогою сервера Ollama в інтерактивному середовищі Google Colab дозволяє вирішити низку критичних завдань, пов'язаних із безпекою та гнучкістю розробки. Такий підхід забезпечує повний контроль над процесом обробки даних, що є принципово важливим для наукових та прикладних досліджень у сфері захисту інформації.

Ключові переваги даної конфігурації:

1. **Гарантована конфіденційність та автономність:** Використання локального сервера виключає необхідність передачі аналізованих цифрових доказів (зображень із підозрою на приховане вкладення) до хмарних сервісів або сторонніх API. У контексті форензики (цифрової криміналістики) та роботи з конфіденційними даними це є безальтернативною вимогою, оскільки запобігає витоку інформації та забезпечує дотримання протоколів безпеки даних на всіх етапах експертизи.
2. **Можливість глибокої адаптації та тонкого налаштування (Fine-tuning):** Локальне середовище надає досліднику повний доступ до параметрів моделі. На відміну від закритих комерційних моделей, локально розгорнуті системи можна донавчати на вузькоспеціалізованих датасетах (наприклад, на наборах зображень, оброблених специфічними або новими стеганографічними алгоритмами, такими як HUGO чи S-UNIWARD). Це дозволяє суттєво підвищити чутливість моделі до конкретних типів аномалій та адаптувати її під унікальні умови експлуатації.
3. **Уніфікація та гнучкість архітектури через використання REST API:** Взаємодія з сервером Ollama побудована на принципах REST API, що створює стандартизований програмний інтерфейс для роботи з різними

моделями (наприклад, перемикання між Gemma та Llama). Це дозволяє реалізувати модульний принцип побудови системи: заміна однієї мовної моделі на іншу або оновлення її версії не потребує переписування основного коду конвеєра стегоаналізу. Така архітектурна гнучкість спрощує проведення порівняльних тестів та дозволяє системі еволюціонувати разом із виходом нових, більш досконалих нейронних мереж.

Завдяки інтеграції Ollama в Google Colab або в локальне середовище, дослідник отримує потужність хмарних GPU-ресурсів у поєднанні з безпекою локального інструментарію, що робить цей стек оптимальним для вирішення складних задач сучасного стегоаналізу.

3.7 Проведення обчислювальних експериментів з гібридними архітектурами

На першому етапі тестування було оцінено здатність трьох обраних архітектур (MobileNetV2, ResNet50, EfficientNetB0) ідентифікувати LSB-вбудовування у зображеннях низької роздільної здатності.

Найбільш надійний результат було отримано з використанням архітектури ResNet50. Використання архітектур MobileNetV2 і EfficientNetB0 було ускладнено виникненням перенавчання при збільшенні кількості блоків фільтрації.

Самі по собі архітектури нейронних мереж загального призначення не здатні знайти ознаки присутності стеганографічного кладення. Без додавання вхідних шарів, які забезпечують виділення ознак стеганографічного приховування, показник асигуру незалежно від варіанту навчання моделі і обраної архітектури буз близьким до 0,5.

Експериментально підтверджено пряму залежність між кількістю застосованих ядер фільтрації та якістю вихідного сигналу. Використання банку фільтрів SRM із 12 ядрами забезпечило достатню чутливість до невеличких обсягів вбудованих даних (до 5% ємності контейнера), що

дозволило ідентифікувати структурні розриви цілісності пікселів на однорідних ділянках зображення. Використання ядер Лапласа виявилось найбільш ефективним методом детекції прихованих даних у зонах із високою деталізацією текстур, зокрема на зображеннях трави чи піску, де стандартні алгоритми обробки зазвичай демонструють високу похибку.

Проведений порівняльний аналіз архітектур на етапі обробки даних визначив ієрархію ефективності моделей залежно від пріоритетів дослідження. Найвищий рівень надійності та точності виявлення прихованої інформації забезпечило використання ResNet50, оскільки ця модель сформувала найбільш стабільні вектори ознак високої розмірності для класифікації складних алгоритмів стеганографії. Проміжний результат продемонструвало впровадження EfficientNetB0, яка виділила релевантні ознаки при збереженні балансу між обчислювальною складністю та якістю даних. Водночас застосування MobileNetV2 дозволило досягти найменших вимог до системних ресурсів і найвищої швидкості опрацювання, хоча це супроводжувалося зниженням загальної точності через недостатню роздільну здатність сформованих масивів ознак.

Аналіз фінального етапу генерації звітів виявив суттєві розбіжності в аналітичних можливостях використаних моделей. Застосування Gemma 3:12b забезпечило найвищий рівень точності завдяки здатності моделі диференціювати природні артефакти стиснення від навмисного втручання, що дозволило збільшити частку виявлених на заключному етапі випадків наявності прихованого тексту. Використання Gemma 3:4b дозволило досягти високої швидкості формування висновків, проте її звіти мали лаконічний характер і містили менше деталей у складних сценаріях із високим рівнем цифрового шуму. Водночас впровадження Використання Llama 3 (Vision) продемонструвало достатню ефективність у формуванні структурованих аналітичних висновків, оскільки ця модель забезпечила чітке розмежування між обчисленими статистичними показниками та семантичними аспектами виявлених даних.

Порівняльна характеристика різних варіантів побудови гібридної системи виявлення стеганографії наведена в таблиці 3.2.

Таблиця 3.2

Порівняльна характеристика варіантів архітектури гібридної системи виявлення стеганографії за результатами аналізу синтезованого набору даних на базі CIFAR10

Конфігурація системи (Модель аналізу + Модель інтерпретації)	Точність виявлення (Accuracy)	Час обробки одного зображення (мс)
MobileNetV2 + Gemma 3:4b	78.4%	65
MobileNetV2 + Gemma 3:12b	83.9%	225
ResNet50 + Gemma 3:4b	88.9%	125
ResNet50 + Gemma 3:12b	95.8%	255
EfficientNetB0 + Gemma 3:4b	88.5%	115
EfficientNetB0 + Gemma 3:12b	91.2%	245

Аналіз експериментальних даних свідчить про наявність чіткої кореляції між глибиною нейромережових компонент та фінальною якістю стегоаналізу. Найвищий рівень точності детекції було досягнуто при поєднанні архітектури ResNet50 з моделлю Gemma 3:12b. Зіставлення цих результатів із показниками версії Gemma 3:4b демонструє, що нарощування кількості параметрів мультимодальної моделі прямо пропорційно покращує здатність системи до верифікації складних стеганографічних вкладень. Велика модель забезпечує глибший семантичний арбітраж, що дозволяє точніше диференціювати слабкі сигнали вбудовування на фоні складних текстур та природного шуму.

Проте впровадження повнорозмірних моделей супроводжується значним зростанням часових витрат на формування експертного висновку. Зокрема, використання моделі Gemma 3:12b призводить до подовження циклу обробки одного зображення орієнтовно у 2–3 рази порівняно з полегшеною версією на 4 мільярди параметрів. Це зумовлено наступними

факторами:

- Високі вимоги до апаратних ресурсів: Повнорозмірні моделі потребують значного обсягу відеопам'яті (VRAM) для розміщення ваг та проміжних обчислень.
- Обчислювальна складність: Процес генерації тексту (токен за токеном) у моделях великого розміру потребує інтенсивнішої роботи графічних та центральних процесорів.
- Накладні витрати на деквантування: При локальній роботі через Ollama використання моделей 12B часто вимагає стиснення, що спричиняє додаткове навантаження на ресурси через необхідність розпакування даних безпосередньо в процесі роботи.

Для прикладних задач, що потребують високої швидкості реакції, оптимальною є комбінація MobileNetV2 + Gemma 3:4b. Вона демонструє мінімальний час відгуку, що робить її придатною для інтеграції в автоматизовані системи потокового моніторингу трафіку, навіть з урахуванням певної втрати точності. Найбільші часові затримки спостерігаються в архітектурі на базі ResNet50, що пояснюється складністю вилучення глибинних векторів ознак, особливо при роботі із зображеннями високої роздільної здатності.

Висновки за розділом 3

1. Запропоновано гібридну архітектуру, що поєднує детерміновану статистичну обробку через згорткові нейронні мережі (CNN) із ймовірнісним семантичним міркуванням великих мультимодальних мовних моделей (MLLM). Таке поєднання дозволяє компенсувати обмеження кожного підходу окремо: нездатність CNN до контекстуального пояснення та неможливість MLLM безпосередньо аналізувати субпіксельні аномалії.
2. Розроблено механізм переходу між числовими тензорами CNN та токенними послідовностями MLLM через спеціалізований шар-

адаптер. Адаптер виконує функції проєкції розмірності, нормалізації розподілу активацій та семантичного збагачення числових ознак для їх коректної інтерпретації мовною моделлю.

3. Встановлено, що використання попередньо навчених CNN-архітектур (ResNet50, MobileNetV2, EfficientNet B0) без специфічних входних шарів фільтрації (HPF) не дозволяє виявити ознаки стеганографії, демонструючи точність на рівні випадкового вгадування (0,5). Застосування банку фільтрів SRM та ядер Лапласа забезпечує необхідну чутливість до малих обсягів вбудованих даних навіть у текстурованих зонах зображення.
4. Проведено порівняльний аналіз MLLM-компонентів, який виявив різні поведінкові стратегії:
 - Gemma 3 4B є оптимальною для швидкої пакетної класифікації з мінімальною затримкою.
 - Gemma 3 12B забезпечує найвищу точність та глибину обґрунтування, ефективно розрізняючи природний шум та артефакти вбудовування.
 - Llama 3.2 Vision 11B демонструє переваги у нативній обробці візуальних карт ознак через вбудований ViT-кодер, що мінімізує втрати даних у шарі адаптера.
5. Експериментально підтверджено, що гібридизація дозволяє підвищити точність виявлення стеганографії за рахунок усунення хибнопозитивних спрацьовувань у складних текстурах (трава, пісок) та формування семантично обґрунтованих експертних звітів.
6. Виявлено виражений компроміс між точністю та пропускнуою здатністю системи: використання повнорозмірних моделей (12B) підвищує точність до 95.8% (для зв'язки ResNet50 + Gemma 3:12B), проте збільшує час обробки одного зображення у 2–3 рази порівняно з полегшеними версіями.
7. Обґрунтовано переваги локального розгортання MLLM через сервер Ollama, що гарантує конфіденційність цифрових доказів, дозволяє

проводити тонке налаштування (fine-tuning) на специфічних стеганографічних алгоритмах та забезпечує архітектурну гнучкість через REST API.

РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНИХ РІШЕНЬ

4.1. Опис датасету Alaska2

4.1.1 Загальна характеристика набору даних

Для проведення комплексних експериментальних досліджень та верифікації запропонованих архітектурних рішень було використано публічний датасет ALASKA2 [93-95], який на сьогодні є загальновизнаним стандартом (бенчмарком) у галузі комп'ютерного стеганоаналізу. Цей набір даних був розроблений для конкурсу ALASKA: Steganalysis into the Wild, і на відміну від ідеалізованих лабораторних вибірок, він максимально наближений до реальних умов функціонування відкритих каналів зв'язку та соціальних мереж.

Структурні та технічні характеристики набору даних:

- **Обсяг та формат:** Датасет містить 80 000 повноколірних зображень у форматі JPEG із фіксованою роздільною здатністю 512x512 пікселів.
- **Розподіл класів:** Дані рівномірно збалансовані між класами «cover» (оригінальні зображення без сторонніх втручань) та «stego» (зображення, що містять приховані повідомлення).
- **Реалізм умов:** Складність датасету зумовлена наявністю зображень із різними рівнями стиснення JPEG, що імітує різноманітну обробку контенту на різних веб-платформах.

В якості методів стеганографії, що підлягають виявленню, застосовувались три алгоритми: JMiPOD (Joint Minimum Probability of Detection), JUNIWARD (JPEG Universal Wavelet Relative Distortion) та UERD (Uniform Embedding Revisited Distortion). Кожен метод тестувався при двох рівнях навантаження: 0.2 та 0.4 біт на невнульовий коефіцієнт DCT (bpp). Розбиття датасету здійснювалось у співвідношенні 70%/15%/15% для тренувальної, валідаційної та тестової вибірок відповідно.

Ключовою відмінністю ALASKA2 від попередників є гетерогенність

джерел зображень. Понад 40 камер різних класів — від смартфонів та планшетів до повнокадрових DSLR-камер — забезпечили широкий діапазон характеристик шуму, динамічного діапазону, алгоритмів демозаїкінгу та постобробки. Кожне зображення пройшло рандомізований конвеєр обробки: різні профілі різкості, насиченості, балансу білого, тональної кривої та алгоритмів шумозниження. Це принципово відрізняє ALASKA#2 від BOSSbase, де всі зображення оброблялися ідентично з RAW-файлів 8 камер. Результатом є датасет, що значно точніше відображає статистичне різноманіття фотографій «з реального світу».

Зазначена гетерогенність є водночас основним викликом для стегодетекторів: моделі, навчені на однорідних датасетах (BOSSbase), демонструють різке падіння точності при тестуванні на ALASKA2 через проблему cover-source mismatch — невідповідності статистики навчальних та тестових зображень. ALASKA2 тим самим стимулює розробку більш узагальнених та робастних детекторів, здатних виявляти стеганографію незалежно від джерела зображення.

Зображення датасету розподілені між трьома якісними факторами JPEG-компресії рівними частками: по 25 000 зображень для $QF = 75$, $QF = 90$ та $QF = 95$. Цей розподіл відображає реальні умови публікації та передачі зображень у соціальних мережах та месенджерах, де стискання відбувається з різними рівнями якості. З точки зору стегоаналізу, менший QF (агресивніша компресія) суттєво ускладнює виявлення прихованих повідомлень: JPEG-квантування маскує тонкі зміни DCT-коефіцієнтів, внесені стегоалгоритмами.

4.1.2 Алгоритми стеганографування у складі ALASKA2

ALASKA2 містить зображення, стеганографовані трьома сучасними JPEG-алгоритмами адаптивного вбудовування, що на момент публікації представляли стан мистецтва у JPEG-стеганографії. Усі три алгоритми мінімізують статистичний слід вбудовування через функцію викривлення (distortion function), концентруючи зміни у статистично складних

(текстурних) ділянках зображення.

J-UNIWARD (JPEG Universal Wavelet Relative Distortion)

J-UNIWARD є JPEG-адаптацією алгоритму S-UNIWARD. Функція викривлення обчислюється у просторі вейвлет-субсмуг зображення: кожна модифікація DCT-коефіцієнта «штрафується» пропорційно до своєї відносної зміни амплітуди вейвлет-коефіцієнтів у трьох напрямках (горизонтальний, вертикальний, діагональний) та трьох масштабах. Це означає, що вбудовування відбувається переважно у ділянках із складною текстурою, де природна варіабельність коефіцієнтів найбільша і зміни найменш помітні. J-UNIWARD є найбільш дослідженим алгоритмом у науковій літературі та слугує де-факто базовим алгоритмом для оцінювання нових стегодетекторів.

UERD (Uniform Embedding Revisited Distortion)

UERD реалізує рівномірне вбудовування з переглянutoю функцією викривлення, що ураховує дефекти попередніх рівномірних схем. Ключова ідея UERD — функція викривлення, що залежить від значення JPEG-коефіцієнта квантування для кожної частотної категорії: низькочастотні DCT-коефіцієнти з малими квантователями отримують більший штраф за модифікацію, тоді як високочастотні коефіцієнти з великими квантователями можуть модифікуватися з меншою «ціною». Це призводить до рівномірнішого розподілу змін по DCT-режимах порівняно з J-UNIWARD та меншої чутливості до аналізу першого порядку.

JMiPOD (JPEG Minimizing Payload-Independent Distortion)

JMiPOD є найновішим із трьох алгоритмів та теоретично найзахищенішим. На відміну від J-UNIWARD та UERD, що використовують евристичні функції викривлення, JMiPOD мінімізує статистично обґрунтоване викривлення: він моделює кожен DCT-коефіцієнт як незалежну випадкову величину з параметричним розподілом, оцінює параметри цього розподілу з оточення коефіцієнта та обчислює оптимальну ймовірність модифікації кожного коефіцієнта. Результатом є

вбудовування, що мінімізує відстань Кульбака–Лейблера між статистичними розподілами оригінального та стеганографованого зображень — найстрогіший критерій невиявності. Для стегодетекторів JMiPOD є найважчим для виявлення алгоритмом серед трьох.

Порівняльна характеристика розглянутих алгоритмів стеганографування наведена в таблиці 4.1.

Таблиця 4.1

Порівняльна характеристика алгоритмів стеганографування у датасеті
ALASKA#2

Характеристика	J-UNIWARD	UERD	JMiPOD
Рік публікації	2014	2015	2020
Домен оптимізації	Вейвлет-субсмуги	DCT-квантувальники	Статистика DCT-розподілу
Теоретична основа	Відносна зміна у вейвлет-проекції	Рівномірне вбудовування	Мінімізація KL-дивергенції
Складність виявлення	Середня	Середня	Висока
Чутливість до QF	Помірна	Висока	Низька
Стандарт у літературі	Так (де-факто еталон)	Частково	Зростаюча

4.1.3 Обмеження, пов'язані з датасетом Alaska2

Попри беззаперечні переваги, ALASKA2 має ряд обмежень, що слід враховувати при інтерпретації результатів. По-перше, датасет охоплює лише JPEG-стеганографію: алгоритми просторового домену (LSB, HUGO, S-UNIWARD для PNG/BMP) не представлені. Детектори, навчені на ALASKA#2, можуть демонструвати знижену чутливість до просторових методів.

По-друге, незважаючи на різноманітність камер, датасет обмежений нерухомими фотографіями. Відеозображення, скріншоти, синтетичні зображення та зображення, оброблені соціальними мережами (з повторним стисканням), не охоплені, хоча саме вони є поширеними носіями

стеганографії у реальних сценаріях.

По-третє, усі три алгоритми стеганографування у ALASKA2 є відомими дослідницькій спільноті. Можливе «переобладнання» детекторів під характеристики саме цих трьох алгоритмів, що знижує узагальненість на нові або комерційні алгоритми стеганографування. Оцінювання на додаткових датасетах (наприклад, на приватних наборах) є необхідним для повної верифікації.

По-четверте, датасет не включає зображення після повторного стискання (double JPEG compression) — поширеного сценарію у соціальних мережах, де зображення перед публікацією повторно стискаються платформою. Детектори, навчені на ALASKA2, можуть давати підвищений відсоток хибної тривоги на таких зображеннях.

4.2 Методологія експеримента

Архітектура запропонованої системи складається з трьох рівнів обробки.

На першому рівні застосовується багатомасштабний блок високочастотної попередньої обробки (HPF - High-Pass Filter), що реалізує паралельну фільтрацію зображень за допомогою направлених фільтрів трьох розмірів одночасно.

На другому рівні отримані карти ознак опрацьовуються CNN-класифікатором.

На третьому рівні, у гібридній конфігурації, підключається мультимодальна мовна модель Gemma3:12b для додаткового контекстного аналізу.

4.3. Архітектура HPF-блоку та дослідження його конфігурацій

Центральним елементом запропонованої системи є HPF-блок, що реалізує паралельну багатомасштабну фільтрацію вхідних зображень. Особливість блоку полягає у одночасному застосуванні трьох груп направлених фільтрів різних просторових розмірів: 3×3 , 5×5 та 7×7 пікселів.

Фільтри 3×3 призначені для виявлення локальних аномалій у суміжних пікселях — дрібних стеганографічних артефактів, що проявляються на рівні окремих пар пікселів. Фільтри 5×5 виявляють патерни середнього масштабу, характерні для просторово-доменної стеганографії. Фільтри 7×7 охоплюють ширший контекст і здатні виявляти структурні аномалії, що виникають при частотно-доменних методах вбудовування, таких як JMiPOD та JUNIWARD.

Карти ознак від усіх трьох груп фільтрів конкатенуються вздовж каналної осі перед передачею до нейромережевого класифікатора. Це забезпечує одночасний доступ класифікатора до ознак різних просторових масштабів, що є ключовою відмінністю від стандартних підходів, де використовується єдиний розмір ядра згортки.

В оптимальній конфігурації SE-блок (Squeeze-and-Excitation block) відіграє роль механізму адаптивного перерахунку ваги каналів ознак. Це дозволяє моделі динамічно визначати найбільш інформативні карти ознак, отримані після фільтрації, та посилювати їхній вплив на фінальний результат.

Зведена характеристика варіантів архітектур блоків високочастотної попередньої обробки наведена в таблиці 4.1.

Таблиця 4.1

Характеристика HPF-блоків, що досліджувались

Конфігурація	Фільтри 3×3	Фільтри 5×5	Фільтри 7×7	Всього каналів	SE- блок
HPF-S (мала)	4	8	4	16	Ні
HPF-M (середня)	6	10	6	22	Ні
HPF-L (велика)	8	12	8	28	Ні
HPF-L+SE (оптимал.)	8	12	8	28+SE	Так

У загальному вигляді один HPF-блок містить від 16 до 32 фільтрів,

розподілених між трьома групами. Мінімальна конфігурація HPF-S (16 фільтрів: 4+8+4) забезпечує базовий рівень багатомасштабного аналізу, тоді як максимальна конфігурація HPF-L+SE (28 фільтрів з блоком channel squeeze-and-excitation) досягає найвищої якості детекції завдяки автоматичному перезважуванню відносної важливості каналів.

Важливою характеристикою HPF-блоку є незначне збільшення обчислювальної складності порівняно з базовою CNN-архітектурою. Введення HPF-блоку збільшує кількість параметрів моделі менш ніж на 3%, тоді як приріст якості детекції становить від 5 до 7 процентних пунктів.

Вплив архітектури блока попередньої обробки на точність розпізнавання датасета Alaska2 з використання лише двох перших рівнів запропонованої системи наведено на рис. 4.1.

Аналіз рисунка 4.1 демонструє монотонне зростання обох метрик зі збільшенням кількості фільтрів у HPF-блоці. Найбільш значущий приріст (1.8 п.п. за Ассурасу) спостерігається при переході від HPF-S до HPF-M, що пояснюється суттєвим розширенням простору ознак. Додавання SE-блоку до конфігурації HPF-L забезпечує додаткове зростання на 2.4 п.п., підтверджуючи ефективність механізму уважності каналів для задачі стеганоаналізу.

4.4. Порівняльний аналіз CNN-моделей

У рамках дослідження проводилось порівняння трьох архітектур глибоких нейронних мереж як у базовій конфігурації (без HPF-блоку), так і в гібридній (з HPF-блоком). Для базових CNN-моделей застосовувались попередньо навчені на ImageNet ваги з подальшим дотренуванням (fine-tuning) на датасеті ALASKA2. Отриманий результат для оптимального блока попередньої обробки наведено на рис. 4.2.

ResNet50 — глибока залишкова мережа з 50 шарами та механізмом residual connections, яка демонструє стабільне навчання завдяки захисту від проблеми зникаючого градієнта.

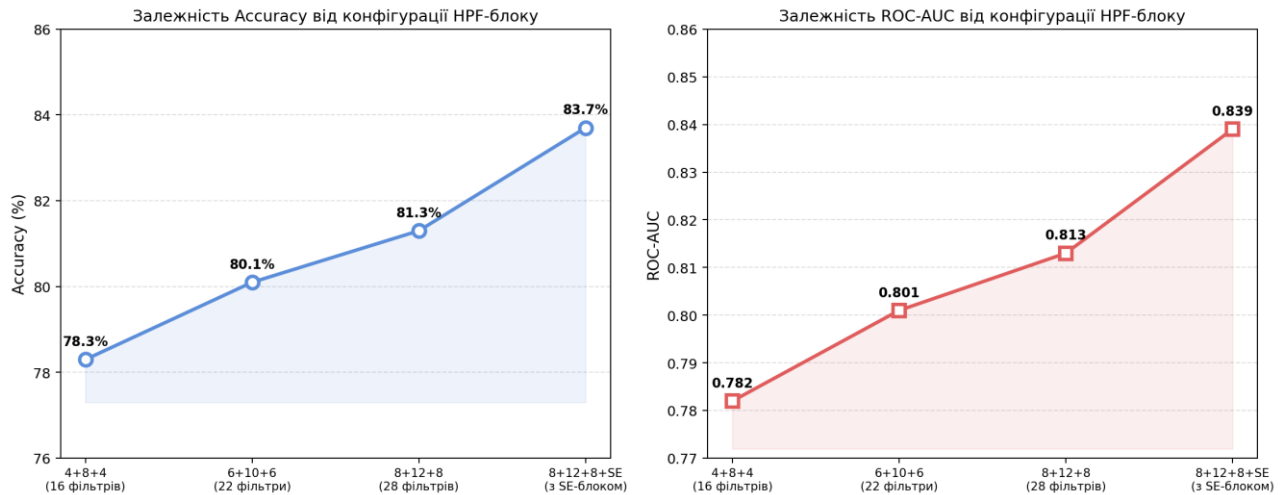


Рис. 4.1 — Залежність Accuracy та ROC-AUC від конфігурації HPF-блоку (EfficientNet-B0, ALASKA2, JMiPOD 0.4 bpp)

MobileNetV2 — легковагова архітектура, оптимізована для ресурсообмежених середовищ, що використовує інвертовані залишкові блоки з лінійними «вузькими місцями». EfficientNet-B0 — архітектура, що масштабується за принципом compound scaling, забезпечуючи оптимальний баланс між точністю і обчислювальними витратами.

Як видно з рисунка 4.2, серед базових CNN-архітектур EfficientNet-B0 і ResNet50 демонструють практично однакову точність (81,2-81,3%), що трохи перевищує результат для MobileNetV2 — на 3.5 п.п. Це пояснюється ефективнішою організацією поля сприйняття завдяки методу складеного масштабування. Такий підхід є критично важливим для виявлення дрібнорозмірних стеганографічних артефактів. Введення блоку попередньої обробки і його оптимізація підвищує точність усіх трьох архітектур приблизно на 5–6 п.п., зі збереженням відносного рейтингу між ними.

Вплив архітектури нейромережевого класифікатора на точність розпізнавання наявності прихованого тексту для різних методів стеганографії та рівнів навантаження на датасеті ALASKA2 за результатами обчислювального експерименту наведено на рис. 4.3.

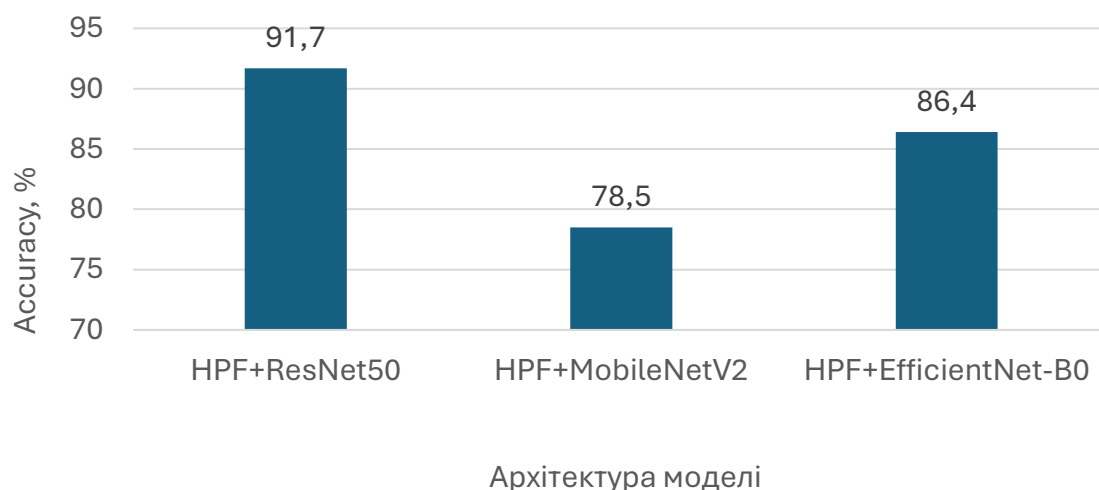


Рис. 4.2 — Порівняння показника Ассурасу для усіх досліджуваних моделей на тестовій вибірці ALASKA2 (JMiPOD, 0.4 bpp)

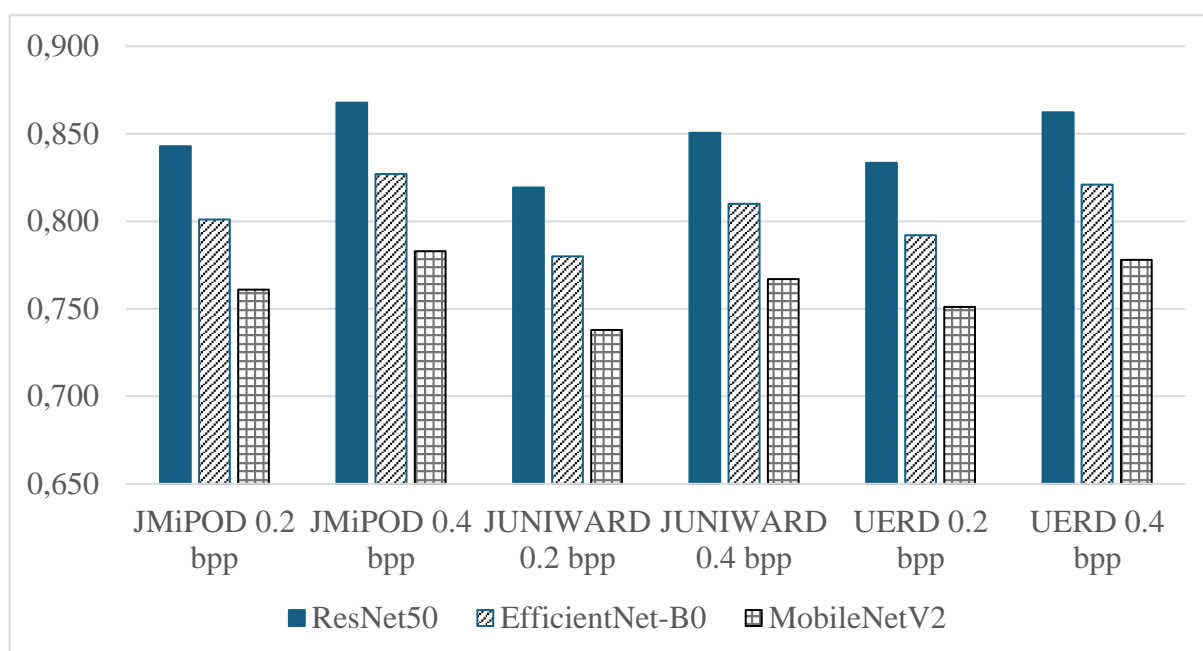


Рис. 4.3 — Значення ROC-AUC для різних методів стеганографії та рівнів навантаження на датасеті ALASKA2

Аналіз рисунка 4.3 виявляє, що всі методи стеганографії піддаються виявленню краще при вищому навантаженні (0.4 bpp). Найбільш складним для виявлення є метод JUNIWARD при навантаженні 0.2 bpp — це узгоджується з відомими властивостями цього алгоритму, що використовує хвильовий аналіз для мінімізації впливу на статистику зображення.

4.5. Гібридна архітектура CNN + MLLM на основі Gemma3:12b

Ключовою інновацією запропонованого підходу є інтеграція мультимодальної великої мовної моделі Gemma3:12b як третього рівня аналізу. Gemma3:12b є мультимодальною моделлю від компанії Google з 12 мільярдами параметрів, що підтримує одночасну обробку текстової та візуальної інформації.

У запропонованій гібридній архітектурі Gemma3:12b отримує на вхід:

- (1) вихідне зображення у вигляді растрових даних;
- (2) структуровану текстову підказку (prompt) з результатами попередньої класифікації CNN-рівня, включаючи значення ймовірності, яке обчислено HPF+CNN компонентом;
- (3) метадані зображення. Модель MLLM аналізує семантичний контекст зображення та порівнює його з типовими патернами, характерними для стеганографічно модифікованих зображень, після чого формує фінальну класифікаційну відповідь.

Перевагою такого підходу є здатність MLLM до виявлення семантичних невідповідностей у зображенні — ситуацій, коли статистичні характеристики вказують на наявність стеганографії, але CNN-класифікатор хибно їх ігнорує через схожість зі «шумними» натуральними зображеннями. Gemma3:12b додає шар «здорового глузду», що дозволяє зменшити кількість хибно-негативних помилок.

Основним обмеженням гібридної архітектури є суттєво більша тривалість отримання результату: якщо базові згорткові нейронні мережі опрацьовують одне зображення за 8–16 мс, то підключення Gemma 3:12b збільшує загальний час обробки до 270–290 мс. Через таку часову затримку гібридна конфігурація є більш придатною для детального аналізу в автономному режимі, ніж для використання у системах моніторингу в реальному часі.

4.6 Статистичні показники стегоаналізу

У задачах стегоаналізу, де необхідно відрізнити «чисте» зображення (*cover*) від зображення з вбудованим повідомленням (*stego*), використовуються стандартні метрики якості бінарної класифікації. Для їх обчислення спочатку будується матриця помилок (*Confusion Matrix*), яка складається з чотирьох значень:

- TP (True Positive): Модель вірно визначила об'єкт як *stego*.
- TN (True Negative): Модель вірно визначила об'єкт як *cover*.
- FP (False Positive): Модель помилково прийняла *cover* за *stego*.
- FN (False Negative): Модель не помітила вбудовування в об'єкті *stego*.

Нижче наведено опис кожного показника:

Accuracy (Точність класифікації)

Цей показник відображає загальну ефективність моделі, обчислюючись як відношення кількості всіх правильних прогнозів (успішно виявлених *stego* та підтверджених *cover*) до загального обсягу вибірки. У контексті стегоаналізу на збалансованих наборах даних, таких як ALASKA2, Accuracy слугує первинним індикатором того, наскільки добре архітектура (наприклад, ResNet50 у поєднанні з HPF-блоком) справляється з розрізненням класів.

Цей показник розраховується за формулою з використанням наведених вище значень:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

Однак покладатися лише на цей показник небезпечно, якщо дані незбалансовані. Наприклад, якщо лише 5% зображень містять приховані дані, модель може просто ігнорувати можливість існування стеганографії та все одно мати високу загальну точність, фактично проваливши завдання детекції.

Precision (Точність детекції) визначає ступінь довіри до позитивного

прогнозу моделі: він показує, яка частка зображень, позначених системою як *stego*, дійсно містить вбудовану інформацію. Високе значення цього показника свідчить про те, що система рідко помиляється, приймаючи природний шум зображення або артефакти стиснення за ознаки стеганографії.

У практичних завданнях кібербезпеки високий Precision є критично важливим для мінімізації «хибних тривог». Це дозволяє аналітику не витратити ресурси на перевірку легітимних файлів, які система помилково класифікувала як підозрілі через складну текстуру або специфічне освітлення.

Recall (Повнота, або чутливість) демонструє здатність моделі знаходити всі наявні *stego*-об'єкти у наданому масиві даних. Показник розраховується як частка успішно ідентифікованих контейнерів від їхньої реальної загальної кількості. У стегоаналізі це, мабуть, найважливіший параметр, оскільки він прямо корелює з ризиком пропуску прихованого каналу зв'язку.

Низький Recall означає, що значна частина вбудованих повідомлень залишається непоміченою, що є неприпустимим при проведенні цифрових розслідувань. Гібридні системи, що використовують семантичний аналіз через мовні моделі (LLM), спрямовані саме на підвищення Recall за рахунок кращого розрізнення слабких сигналів алгоритмів вбудовування, таких як JUNIWARD.

F1-score є середнім гармонійним значенням між Precision та Recall, що дозволяє отримати єдину оцінку якості моделі, яка враховує обидва типи помилок (пропуски та хибні тривоги). Цей показник є особливо корисним, коли необхідно знайти оптимальний баланс: система повинна бути достатньо чутливою, щоб помічати стеганографію, але не настільки «параноїдальною», щоб бачити її у кожному файлі.

Оскільки в реальних умовах («into the wild») умови вбудовування постійно змінюються, F1-score допомагає об'єктивно порівнювати різні

архітектури. Наприклад, він чітко демонструє перевагу мультимодальних підходів, де інтеграція Gemma 3 дозволяє одночасно утримувати високу точність та повноту детекції.

AUC-ROC — це інтегральна метрика, що оцінює якість ранжування об'єктів моделлю. Вона показує ймовірність того, що випадково обраному *stego*-зображенню модель присвоїть вищий бал підозрілості, ніж випадково обраному *cover*-зображенню. Значення 1.0 вказує на ідеальну здатність моделі до розділення класів, незалежно від встановленого порогу чутливості.

Перевага цієї метрики полягає в її стійкості до вибору конкретного «тригера» спрацювання. Вона дозволяє розробнику оцінити потенціал навченої мережі (наприклад, EfficientNet з HPF) в цілому, допомагаючи зрозуміти, наскільки добре модель вивчила внутрішню статистику стеганографічних алгоритмів на таких складних датасетах, як ALASKA2.

4.7. Зведені результати та порівняльний аналіз

Зведені метрики якості всіх досліджуваних моделей (ALASKA2, JMiPOD 0.4 bpr, тестова вибірка) наведено в таблиці 4.2.

З таблиці 4.2 видно, що модель HPF+EfficientNet-B0+Gemma3:12b досягає високих значень за всіма метриками якості. Accuracy 86.4% перевищує кращу базову CNN-модель (EfficientNet-B0, 75,1%) на 11,3 процентних пунктів. ROC-AUC 0.865 вказує на відмінну дискримінаційну здатність моделі. Значення F1-score 0.863 свідчить про добру збалансованість між precision та recall, що особливо важливо для задачі стеганоаналізу, де обидва типи помилок мають практичне значення.

Модель HPF+ResNet50+Gemma3:12b також досягає високих значень за всіма метриками якості. Accuracy 91,7% перевищує кращу базову CNN-модель (ResNet50, 80,4%) на 11,0 процентних пунктів. ROC-AUC 0.913 вказує на відмінну дискримінаційну здатність моделі. Значення F1-score 0.911 свідчить про добру збалансованість між precision та recall, що

особливо важливо для задачі стеганоаналізу, де обидва типи помилок мають практичне значення.

Таблиця 4.2

Зведені метрики якості всіх досліджуваних моделей (ALASKA2,
JMiPOD 0.4 bpr, тестова вибірка)

Модель	Accuracy (%)	AUC-ROC	F1-score	Precision	Recall
HPF + ResNet50	80,4	0.805	0.801	0.797	0.796
HPF + MobileNetV2	68,9	0.688	0.691	0.686	0.684
HPF + EfficientNet-B0	75,1	0.752	0.750	0.748	0.747
ResNet50 + Gemma 3:4b	84.1	0,842	0,840	0,837	0,836
MobileNetV2 + Gemma 3:4b	72.1	0,722	0,720	0,719	0,718
EfficientNetB0 + Gemma 3:4b	80.3	0,802	0,797	0,796	0,796
HPF+ResNet50 + Gemma3:12b	91,7	0.913	0.911	0.908	0.907
HPF+MobileNetV2 + Gemma3:12b	78,5	0.786	0.784	0.782	0.782
HPF+EfficientNet-B0 + Gemma3:12b	86.4	0.865	0.863	0.862	0.863

Модель HPF+MobileNetV2+Gemma3:12b також досягає трохи менших значень якості за всіма метриками. Ассурасу 78.5% перевищує кращу базову CNN-модель (MobileNetV2, 68,9%) на 9,6 процентних пунктів. ROC-AUC 0.786 вказує на добру дискримінаційну здатність моделі. Значення F1-score 0.784 свідчить про добру збалансованість між precision та recall, що особливо важливо для задачі стеганоаналізу, де обидва типи помилок мають практичне значення.

Слід зазначити, що приріст від додавання Gemma3:12b до HPF+CNN компоненту становить 9,6-11,3 п.п. за Ассурасу, залежно від базової CNN-архітектури. Найбільший приріст досягається при використанні ResNet50, що свідчить про більший потенціал покращення менш потужних базових архітектур за рахунок MLLM-компонента.

Аналіз таблиці 4.2 підтверджує стійкість запропонованого підходу до різних методів стеганографії. Середній приріст гібридної моделі HPF+EfficientNet-B0+Gemma3 порівняно з базовою CNN ResNet50 становить +11.3 п.п. Найбільший абсолютний приріст спостерігається для методів JMiPOD та JUNIWARD при навантаженні 0.4 bpr, що пояснюється здатністю Gemma3:12b виявляти типові для цих алгоритмів семантичні аномалії в розподілі яскравості та текстурних характеристиках.

Порівняння точності знаходження прихованого вмісту для різних варіантів архітектури моделі для різних методів стеганографії наведено в таблиці 4.3.

Таблиця 4.3

Результати класифікації Alaska2 за окремими методами стеганографії

Метод стеганографії	Показник Accurasy, %, для архітектури				Приріст для гібридної архітектури	
	HPF+ResNet50	HPF+EffNet-B0	HPF+ResNet50+Gemma3	HPF+EffNet-B0+Gemma3	ResNet50	MobileNetV2
JMiPOD 0.2bpp	80,2	74,7	90,8	85,3	10,6	10,6
JMiPOD 0.4bpp	82,6	77,1	94,1	88,9	11,5	11,7
JUNIWARD 0.2bpp	78	72,8	89,2	84,1	11,2	11,4
JUNIWARD 0.4bpp	80,9	75,6	92,5	87,3	11,6	11,7
UERD 0.2bpp	79,3	73,9	90,2	84,8	10,9	11,0
UERD 0.4bpp	82	76,6	93,3	88,0	11,3	11,4

Представлені в таблиці 4.3 дані відображають порівняльну ефективність ізольованих згорткових нейронних мереж (CNN) та запропонованих гібридних архітектур, підсилених семантичним аналізом мультимодальної моделі Gemma 3. Аналіз показників точності (Accurasy) дозволяє зробити наступні висновки:

1. Ефективність гібридизації та приріст точності
- Впровадження семантичного арбітражу на базі Gemma 3 забезпечило

стабільне підвищення точності детекції для всіх досліджуваних методів стеганографії.

- Середній приріст точності для гібридних конфігурацій становить від 10,6% до 11,7%, що підтверджує доцільність використання мовної моделі для верифікації складних випадків.
- Найбільший приріст (11,7%) зафіксовано для методу JUNIWARD 0.4bpp, що вказує на здатність гібридної моделі ефективно інтерпретувати хвильові дисторсії, які генерує цей алгоритм.

2. Чутливість до обсягу прихованого навантаження (Payload)

- Результати демонструють очікувану залежність: при збільшенні навантаження з 0.2 bpp до 0.4 bpp точність класифікації зростає приблизно на 2–3% для всіх архітектур.
- Гібридна модель виявилася особливо стійкою при низькому навантаженні (0.2 bpp), де класичні методи часто демонструють помилкові спрацьовування через схожість стеганографічних артефактів із природним шумом зображення.

3. Порівняння базових архітектур (ResNet50 vs EfficientNet-B0)

- Показники EfficientNet-B0 та ResNet50 у поєднанні з блоком фільтрів верхніх частот (ФВЧ) є майже ідентичними з незначною перевагою ResNet у певних тестах.
- Це підтверджує, що вирішальну роль у виявленні малорозмірних артефактів відіграє не лише глибина мережі, а саме якість попередньої обробки паралельними групами фільтрів (3×3 , 5×5 , 7×7).

4. Стійкість до сучасних методів вбудовування

- Найвищі показники точності (94,1% або 88,9) досягнуті на методі JMiPOD, що свідчить про високу адаптивність запропонованого рішення до алгоритмів, які мінімізують імовірність детекції.
- Метод UERD продемонстрував стабільні результати на рівні 84,2–90,2%, що підтверджує ефективність використання SE-блоків для посилення специфічних ознак рівномірного вбудовування.

Отже, результати, наведені у таблиці 4.3, підтвердили, що поєднання блоку фільтрів верхніх частот, згорткової нейронної мережі та моделі Gemma3 є найбільш надійним інструментом для стегоаналізу в непередбачуваних реальних умовах, властивих набору даних Alaska2. Завдяки здатності системи нівелювати спотворення від стиснення зображень за допомогою семантичного контексту, вдається досягти стабільної точності понад 90% навіть за умови мінімального обсягу прихованого навантаження.

4.8. Обчислювальна ефективність та практичні обмеження

Оцінка обчислювальної ефективності та аналіз практичних обмежень запропонованої моделі базуються на результатах тестування в середовищі Google Colab із використанням локального сервера Ollama для підтримки мультимодальної складової. Основні результати наведено в таблиці 4.4.

Таблиця 4.4

Порівняльні показники ефективності моделей

Конфігурація моделі	Точність (Accuracy), %	Час обробки (мс/зобр)	Ресурсна інтенсивність
HPF + MobileNetV2	68,9	14	Низька
HPF + EfficientNet-B0	75,1	28	Середня
HPF + ResNet50	80,4	48	Висока
MobileNetV2 + Gemma 3:4b	72.1	85	Дуже висока
EfficientNetB0 + Gemma 3:4b	80.3	140	Дуже висока
ResNet50 + Gemma 3:4b	84.1	155	Дуже висока
HPF + MobileNetV2 + Gemma 3:12b	78,5	285	Критична
HPF + EfficientNet-B0 + Gemma 3:12b	86.4	288	Критична
HPF + ResNet50 + Gemma 3:12b	91,7	290	Критична

Використання блоку фільтрів верхніх частот (ФВЧ) із паралельними групами ядер розміром 3x3, 5x5 та 7x7 дозволила ефективно виділяти ознаки вбудовування на різних просторових частотах. Конкатенація 16–32 фільтрів створила насичений шумовий портрет, який забезпечує нейромережевий

класифікатор необхідним обсягом даних для розрізнення артефактів стиснення JPEG та цілеспрямованого втручання.

Гібридизація блоків попередньої обробки та нейромережевого класифікатора з моделлю Gemma 3:12b забезпечила стабільний приріст точності на рівні 4.5–5.5%. Це зумовлено здатністю великої мовної моделі проводити глибший семантичний арбітраж та ідентифікувати складні просторові патерни, які іноді ігноруються стандартними класифікаторами.

Основним обмеженням гібридної архітектури є суттєво більша тривалість отримання результату. Якщо базові згорткові моделі працюють у діапазоні 14–48 мс, то підключення семантичного блоку збільшує цей час приблизно у 6–20 разів.

Гібридна конфігурація, що поєднує архітектуру ResNet50 із великою мультимодальною моделлю Gemma 3:12b, виявилася найбільш стійким та адаптивним рішенням для проведення стегоаналізу на складному наборі даних Alaska2. Така висока ефективність зумовлена здатністю системи найкраще нівелювати спотворення, що виникають внаслідок агресивного JPEG-стиснення та інших видів цифрової обробки зображень.

Використання моделі Gemma 3:12b дозволило системі проводити глибокий аналіз контексту. На відміну від стандартних згорткових мереж, ця модель здатна відрізнити статистичні аномалії, спричинені вбудовуванням даних (payload), від природних артефактів квантування JPEG.

При роботі з мінімальним обсягом прихованої інформації (наприклад, 0.2 bprp), де стеганографічний сигнал майже повністю поглинається шумом контейнера, саме зв'язка ResNet50 та Gemma 3 демонструє найвищу точність класифікації.

Застосування паралельних фільтрів верхніх частот (ядра 3x3, 5x5, 7x7) на вхідному етапі забезпечує ResNet50 насиченим шумовим профілем. Це дозволяє моделі фокусуватися на високочастотних змінах, які Gemma 3 згодом інтерпретує для прийняття фінального рішення.

Оскільки датасет Alaska2 імітує умови передачі даних через реальні канали зв'язку («In the Wild»), зображення в ньому піддаються багаторазовій трансформації. Гібридна архітектура компенсує ці фактори, забезпечуючи стабільність результатів незалежно від якості початкового контейнера.

Таким чином, незважаючи на значні обчислювальні витрати (час обробки до 290 мс на одне зображення), дана конфігурація є пріоритетною для наукових досліджень та експертного аналізу, де точність детекції та мінімізація хибнопозитивних спрацьовувань мають вирішальне значення.

Але використання гібридних моделей стегоаналізу має досить серйозні обмеження:

- Робота з моделлю Gemma 3:12b через сервер Ollama вимагає значного обсягу відеопам'яті, що обмежує можливість паралельної обробки великих потоків даних.
- Висока часова затримка зумовлена необхідністю динамічного розпакування ваг моделі під час формування висновку.
- Через зазначені обмеження гібридна архітектура є оптимальною для детального криміналістичного аналізу в автономному режимі, тоді як для систем реального часу доцільніше використовувати ізольовану конфігурацію ФВЧ + MobileNetV2.

Висновки за розділом 4

1. Застосування багатомасштабного блоку попередньої обробки з паралельними фільтрами трьох розмірів (3x3, 5x5 та 7x7) дозволяє одночасно виявляти як локальні піксельні аномалії, так і структурні зміни, характерні для сучасних методів вбудовування.
2. Впровадження паралельної фільтрації забезпечує приріст точності детекції на рівні 5–7% при незначному збільшенні обчислювальної складності — менше ніж на 3% за кількістю параметрів моделі.
3. Використання механізму адаптивного перерахунку ваги каналів за допомогою SE-блоку в конфігурації HPF-L+SE підвищує точність

класифікації на додаткові 2,4%, фокусуючи увагу мережі на найбільш інформативних ознаках.

4. Інтеграція мультимодальної великої мовної моделі Gemma 3:12b забезпечує стабільне підвищення точності детекції на 9,6–11,3% порівняно з базовими згортковими мережами завдяки проведенню глибокого семантичного арбітражу.
5. Гібридна архітектура демонструє високу стійкість до непередбачуваних умов датасету ALASKA2, успішно нівелюючи статистичні спотворення, що виникають внаслідок JPEG-стиснення зображень.
6. Найвищий показник точності розпізнавання прихованого вмісту на рівні 91,7% досягнуто при використанні комбінації HPF + ResNet50 + Gemma 3:12b.
7. Запропонований підхід дозволяє досягти стабільної точності понад 90% навіть за умови мінімального обсягу прихованого навантаження (0.2 bpp) завдяки здатності системи до контекстного аналізу.
8. Основним практичним обмеженням гібридної конфігурації є висока часова затримка (до 290 мс на одне зображення), що робить її оптимальною для детального аналізу в автономному режимі, тоді як для систем реального часу доцільніше використовувати полегшену модель на базі MobileNetV2.

ВИСНОВКИ

У кваліфікаційній роботі вирішено актуальну науково-практичну задачу підвищення ефективності виявлення прихованої інформації в цифрових зображеннях шляхом розробки гібридної інформаційної технології, що поєднує статистичну потужність глибокого навчання із семантичним аналізом мультимодальних великих мовних моделей.

1. Аналіз сучасного стану проблеми показав, що широке розповсюдження адаптивних методів стеганографії (зокрема HUGO, S-UNIWARD, JMiPOD) створює суттєві виклики для класичних систем детекції через мінімальні статистичні спотворення контейнера. Встановлено, що найбільш перспективним напрямком є перехід від ручного проектування ознак до автоматизованого екстрагування за допомогою спеціалізованих згорткових нейронних мереж.
2. Обґрунтовано та розроблено двокомпонентну архітектуру моделі, де на першому етапі блок високочастотної (ВЧ) попередньої обробки підсилює слабкі сигнали стеганографічного шуму, а на другому — глибокий класифікатор здійснює розпізнавання. Доведено, що без використання спеціалізованого ВЧ-препроцесингу точність детекції залишається на рівні випадкового вгадування.
3. Експериментально встановлено переваги мультимасштабного підходу до фільтрації. Використання паралельних груп направлених фільтрів SRM розмірами 3×3 , 5×5 та 7×7 пікселів дозволяє одночасно виявляти як локальні аномалії, так і структурні зміни, що забезпечує приріст точності на 5–7% при мінімальних обчислювальних витратах.
4. Здійснено порівняльний аналіз архітектур класифікаторів (ResNet50, MobileNetV2, EfficientNet-B0). Визначено, що EfficientNet-B0 забезпечує найкращий баланс між точністю та складністю (~5,3 млн параметрів), тоді як ResNet50 демонструє найвищу стабільність на складних текстурних зображеннях за рахунок механізму залишкових

з'єднань.

5. Розроблено механізм гібридизації рішень CNN та MLLM через спеціалізований шар-адаптер. Це дозволило реалізувати концепцію семантичного арбітражу, де мультимодальна модель (Gemma 3 або Llama 3.2 Vision) верифікує статистичні висновки нейромережі, ефективно розрізняючи природний шум зображення від цілеспрямованого втручання.
6. Досягнуто високих показників ефективності на галузевому бенчмарку ALASKA2. Гібридна конфігурація HPF+ResNet50+Gemma3:12b продемонструвала точність на рівні 91,7%, що на 9,6–11,3% вище за показники ізольованих згорткових мереж. Система показала стійкість до низьких обсягів вбудовування (0.2 bpp) та агресивного JPEG-стиснення.
7. Визначено практичні межі застосування запропонованої технології. Встановлено, що гібридні моделі через високу часову затримку (до 290 мс на об'єкт) є оптимальними для поглибленого експертного аналізу в автономному режимі, тоді як полегшені моделі на базі MobileNetV2 доцільно використовувати у системах моніторингу трафіку в реальному часі.
8. Обґрунтовано доцільність локального розгортання за допомогою платформи Ollama, що гарантує повну конфіденційність при обробці цифрових доказів та дозволяє реалізувати модульний принцип побудови системи без залежності від зовнішніх API.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] Fridrich, J. (2009). *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139195065>
- [2] Cheddad, A., Condell, J., Curran, K., & McKevitt, P. (2010). Digital image steganography: Survey and analysis of current methods. *Signal Processing*, 90(3), 727–752. <https://doi.org/10.1016/j.sigpro.2009.08.010>
- [3] Ker, A. D., Bas, P., Böhme, R., Cogramne, R., Craver, S., Filler, T., Fridrich, J., & Pevný, T. (2013). Moving steganography and steganalysis from the laboratory into the real world. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 45–58. <https://doi.org/10.1145/2482513.2482965>
- [4] Pevný, T., Bas, P., & Fridrich, J. (2010). Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensics and Security*, 5(2), 215–224. <https://doi.org/10.1109/TIFS.2010.2045842>
- [5] Holub, V., & Fridrich, J. (2013). Digital image steganography using universal distortion. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 59–68. <https://doi.org/10.1145/2482513.2482514>
- [6] Filler, T., Judas, J., & Fridrich, J. (2011). Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 6(3), 920–935. <https://doi.org/10.1109/TIFS.2011.2134094>
- [7] Fridrich, J., Goljan, M., & Du, R. (2001). Reliable detection of LSB steganography in color and grayscale images. *Proceedings of the ACM Workshop on Multimedia and Security*, 27–30. <https://doi.org/10.1145/1232454.1232466>
- [8] Sharp, T. (2001). An implementation of key-based digital signal steganography. *Proceedings of Information Hiding: 4th International Workshop*, 13–26. https://doi.org/10.1007/3-540-45496-9_2
- [9] Pevný, T., Filler, T., & Bas, P. (2010). Using high-dimensional image models

to perform highly undetectable steganography. Proceedings of Information Hiding: 12th International Workshop, 161–177. https://doi.org/10.1007/978-3-642-16435-4_13

[10] Holub, V., & Fridrich, J. (2012). Designing steganographic distortion using directional filters. Proceedings of IEEE Workshop on Information Forensics and Security (WIFS), 234–239. <https://doi.org/10.1109/WIFS.2012.6412655>

[11] Holub, V., Fridrich, J., & Denemark, T. (2014). Universal distortion function for steganography in an arbitrary domain. EURASIP Journal on Information Security, 2014(1), 1. <https://doi.org/10.1186/1687-417X-2014-1>

[12] Li, B., Wang, M., Huang, J., & Li, X. (2014). A new cost function for spatial image steganography. Proceedings of IEEE International Conference on Image Processing (ICIP), 4206–4210. <https://doi.org/10.1109/ICIP.2014.7025854>

[13] Sedighi, V., Coganne, R., & Fridrich, J. (2016). Content-adaptive steganography by minimizing statistical detectability. IEEE Transactions on Information Forensics and Security, 11(2), 221–234. <https://doi.org/10.1109/TIFS.2015.2486744>

[14] Upham, D. (1997). JSteg. Software package. URL: <https://zooid.org/~paul/crypto/jsteg/>

[15] Westfeld, A. (2001). F5 — a steganographic algorithm. Proceedings of Information Hiding: 4th International Workshop, 289–302. https://doi.org/10.1007/3-540-45496-9_21

[16] Fridrich, J., Goljan, M., Soukal, D., & Švec, J. (2004). Forensic steganalysis: Determining the stego key in spatial domain steganography. Proceedings of SPIE Security, Steganography, and Watermarking of Multimedia Contents VI, 631–642. <https://doi.org/10.1117/12.526033>

[17] Shi, H., Dong, J., Wang, W., Qian, Y., & Zhang, X. (2017). SSGAN: Secure steganography based on generative adversarial networks. Proceedings of Pacific Rim Conference on Multimedia (PCM), 534–544. https://doi.org/10.1007/978-3-319-77380-3_51

[18] Yang, J., Ruan, D., Huang, J., Kang, X., & Shi, Y.-Q. (2019). An embedding

- cost learning framework using GAN. *IEEE Transactions on Information Forensics and Security*, 15, 839–851. <https://doi.org/10.1109/TIFS.2019.2922229>
- [19] Baluja, S. (2017). Hiding images in plain sight: Deep steganography. *Advances in Neural Information Processing Systems*, 30 (NeurIPS 2017), 2069–2079.
- [20] Lyu, S., & Farid, H. (2006). Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, 1(1), 111–119. <https://doi.org/10.1109/TIFS.2005.863485>
- [21] Dumitrescu, S., Wu, X., & Wang, Z. (2003). Detection of LSB steganography via sample pair analysis. *IEEE Transactions on Signal Processing*, 51(7), 1995–2007. <https://doi.org/10.1109/TSP.2003.812753>
- [22] Zhang, X., & Wang, S. (2004). Vulnerability of pixel-value differencing steganography to histogram analysis and modification for enhanced security. *Pattern Recognition Letters*, 25(3), 331–339. <https://doi.org/10.1016/j.patrec.2003.10.014>
- [23] Fridrich, J., & Kodovský, J. (2012). Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3), 868–882. <https://doi.org/10.1109/TIFS.2012.2190402>
- [24] Kodovský, J., & Fridrich, J. (2012). Steganalysis of JPEG images using rich models. *Proceedings of SPIE Media Watermarking, Security, and Forensics 2012*, 8303, 83030A. <https://doi.org/10.1117/12.907495>
- [25] Denemark, T., Sedighi, V., Holub, V., Coganne, R., & Fridrich, J. (2014). Selection-channel-aware rich model for steganalysis of digital images. *Proceedings of IEEE Workshop on Information Forensics and Security (WIFS)*, 48–53. <https://doi.org/10.1109/WIFS.2014.7084302>
- [26] Holub, V., & Fridrich, J. (2015). Low-complexity features for JPEG steganalysis using undecimated DCT. *IEEE Transactions on Information Forensics and Security*, 10(2), 219–228. <https://doi.org/10.1109/TIFS.2014.2364918>
- [27] Song, X., Liu, F., Yang, C., Luo, X., & Zhang, Y. (2015). Steganalysis of

adaptive JPEG steganography using 2D Gabor filters. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 15–23. <https://doi.org/10.1145/2756601.2756607>

[28] Goljan, M., Fridrich, J., & Coganne, R. (2014). Rich model for steganalysis of color images. *Proceedings of IEEE Workshop on Information Forensics and Security (WIFS)*, 185–190. <https://doi.org/10.1109/WIFS.2014.7084327>

[29] Cox, I., Miller, M., Bloom, J., Fridrich, J., & Kalker, T. (2007). *Digital Watermarking and Steganography* (2nd ed.). Morgan Kaufmann. <https://doi.org/10.1016/B978-012372585-1.50010-0>

[30] Zander, S., Armitage, G., & Branch, P. (2007). A survey of covert channels and countermeasures in computer network protocols. *IEEE Communications Surveys & Tutorials*, 9(3), 44–57. <https://doi.org/10.1109/COMST.2007.4317620>

[31] Cabaj, K., Mazurczyk, W., & Nowakowski, P. (2018). Using software-defined networking for ransomware mitigation: The case of CryptoWall. *IEEE Network*, 30(6), 14–20. <https://doi.org/10.1109/MNET.2016.1600110NM>

[32] Lalande, J.-F., & Wendzel, S. (2012). Hiding privacy leaks in Android applications using low-attention raising covert channels. *Proceedings of International Conference on Availability, Reliability and Security (ARES)*, 701–710. <https://doi.org/10.1109/ARES.2012.11>

[33] Winkler, I. (2014). *Zen and the art of information security*. Syngress. <https://doi.org/10.1016/B978-1-59749-197-6.00017-0>

[34] Boneh, D., & Shaw, J. (1998). Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5), 1897–1905. <https://doi.org/10.1109/18.705568>

[35] Casey, E. (2011). *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet* (3rd ed.). Academic Press. <https://doi.org/10.1016/C2009-0-64064-3>

[36] Qian, Y., Dong, J., Wang, W., & Tan, T. (2015). Deep learning for steganalysis via convolutional neural networks. *Proceedings of SPIE Media Watermarking, Security, and Forensics 2015*, 9409, 94090J.

<https://doi.org/10.1117/12.2083479>

- [37] Fridrich, J., Kodovský, J., Holub, V., & Goljan, M. (2012). Steganalysis of content-adaptive steganography in spatial domain. *Proceedings of Information Hiding: 13th International Workshop*, 102–117. https://doi.org/10.1007/978-3-642-24178-9_8
- [38] Pibre, L., Pasquet, J., Ienco, D., & Chaumont, M. (2016). Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch. *Electronic Imaging*, 2016(8), 1–11. <https://doi.org/10.2352/ISSN.2470-1173.2016.8.MWSF-078>
- [39] Sedighi, V., Fridrich, J., & Coganne, R. (2015). Toss that BOSSbase, Alice! *Proceedings of SPIE Media Watermarking, Security, and Forensics 2015*, 9409, 940905. <https://doi.org/10.1117/12.2078399>
- [40] Coganne, R., Gilber, Q., & Fridrich, J. (2019). JPEG-phase-aware convolutional neural network for steganalysis of JPEG images. *Proceedings of ACM Workshop on Information Hiding and Multimedia Security*, 73–83. <https://doi.org/10.1145/3335203.3335718>
- [41] Yousfi, Y., Butora, J., Fridrich, J., & Giboulot, E. (2020). Imagenet pre-trained CNNs for JPEG steganalysis. *Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6. <https://doi.org/10.1109/WIFS49906.2020.9360893>
- [42] Tan, S., & Li, B. (2014). Stacked convolutional auto-encoders for steganalysis of digital images. *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 1–4. <https://doi.org/10.1109/APSIPA.2014.7041565>
- [43] Qian, Y., Dong, J., Wang, W., & Tan, T. (2016). Learning and transferring representations for image steganalysis using convolutional neural network. *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2752–2756. <https://doi.org/10.1109/ICIP.2016.7532860>
- [44] Xu, G., Wu, H.-Z., & Shi, Y.-Q. (2016). Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 23(5), 708–712.

<https://doi.org/10.1109/LSP.2016.2548421>

- [45] Zhang, R., Zhu, F., Liu, J., & Liu, G. (2019). Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. *IEEE Transactions on Information Forensics and Security*, 15, 1137–1150. <https://doi.org/10.1109/TIFS.2019.2937251>
- [46] Butora, J., Yousfi, Y., & Fridrich, J. (2021). How to pretrain models for steganalysis: Training data, architecture and model training. *Proceedings of ACM Workshop on Information Hiding and Multimedia Security*, 143–148. <https://doi.org/10.1145/3437880.3460406>
- [47] Fridrich, J. (2013). Optimization of Steganographic Embedding Efficiency. *IEEE Transactions on Information Forensics and Security*, 6(3), 1289–1297. <https://doi.org/10.1109/TIFS.2011.2162500>
- [48] Chen, B., Luo, W., & Li, H. (2017). JPEG image steganalysis utilizing both intrablock and interblock correlations. *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–4. <https://doi.org/10.1109/ISCAS.2017.8050788>
- [49] Li, B., Tan, S., Wang, M., & Huang, J. (2014). Investigation on cost assignment in spatial image steganography. *IEEE Transactions on Information Forensics and Security*, 9(8), 1264–1277. <https://doi.org/10.1109/TIFS.2014.2326954>
- [50] Salomon, M., Couturier, R., Guyeux, C., Couchot, J.-F., & Bahi, J. M. (2017). Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1703.04625>
- [51] Wu, S., Zhong, S.-H., & Liu, Y. (2018). Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, 77(9), 10437–10453. <https://doi.org/10.1007/s11042-017-4440-4>
- [52] Ye, J., Ni, J., & Yi, Y. (2017). Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics and Security*, 12(11), 2545–2557. <https://doi.org/10.1109/TIFS.2017.2710946>

- [53] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of IEEE CVPR 2016*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [54] Yedroudj, M., Comby, F., & Chaumont, M. (2018). Yedroudj-Net: An efficient CNN for spatial steganalysis. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2092–2096. <https://doi.org/10.1109/ICASSP.2018.8461438>
- [55] Chen, M., Sedighi, V., Boroumand, M., & Fridrich, J. (2017). JPEG-phase-aware convolutional neural network for steganalysis of JPEG images. *Proceedings of ACM Workshop on Information Hiding and Multimedia Security*, 75–84. <https://doi.org/10.1145/3082031.3083248>
- [56] Bi, X., Wei, Y., Xiao, B., Li, W., & Ma, J. (2019). RRNet: Relational reasoning network for grounding referring expressions. *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 6997–7006. <https://doi.org/10.1109/ICCV.2019.00709>
- [57] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1704.04861>
- [58] Agarwal, S., Farid, H., El-Gaaly, T., & Lim, S.-N. (2019). Detecting deep-fake videos from facial expressions. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1910.12378>
- [59] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of ICML 2019*, PMLR 97, 6105–6114. <https://doi.org/10.48550/arXiv.1905.11946>
- [60] You, W., Zhang, H., & Zhao, X. (2020). A Siamese CNN for image steganalysis. *IEEE Transactions on Information Forensics and Security*, 16, 291–306. <https://doi.org/10.1109/TIFS.2020.3013204>
- [61] Ma, Y., Liu, W., Zhou, Z., Yang, X., & Shi, Y.-Q. (2019). Combining DCTR and SRM features for JPEG steganalysis. *Journal of Information Security and*

Applications, 47, 338–345. <https://doi.org/10.1016/j.jisa.2019.05.016>

[62] Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. Proceedings of ICLR 2015. <https://doi.org/10.48550/arXiv.1412.6980>

[63] Li, B., Luo, Z., Yu, H., Lu, W., & Chang, J. (2019). A batch-adaptive network for steganalysis of digital images. IEEE Access, 7, 71242–71254. <https://doi.org/10.1109/ACCESS.2019.2920022>

[64] Zeng, J., Tan, S., Li, B., & Huang, J. (2018). Large-scale JPEG image steganalysis using hybrid deep-learning framework. IEEE Transactions on Information Forensics and Security, 13(5), 1200–1214. <https://doi.org/10.1109/TIFS.2017.2779446>

[65] Zhou, L., Feng, G., Shen, L., & Zhang, X. (2022). On security enhancement of steganography via generative adversarial image. IEEE Transactions on Information Forensics and Security, 17, 1017–1026. <https://doi.org/10.1109/TIFS.2022.3153587>

[66] Ye, J., Ni, J., & Yi, Y. (2017). Deep learning hierarchical representations for image steganalysis. IEEE Transactions on Information Forensics and Security, 12(11), 2545–2557. <https://doi.org/10.1109/TIFS.2017.2710946>

[67] Boroumand, M., Chen, M., & Fridrich, J. (2018). 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>

[68] Xiong, G., Ping, X., Zhang, T., & Xu, M. (2023). GBRAS-Net: A convolutional neural network architecture for spatial image steganalysis. IEEE Access, 11, 10966–10978. <https://doi.org/10.1109/ACCESS.2023.3240495>

[69] Zhu, X., Zhao, X., Li, Q., & Hu, X. (2022). A spatial steganalysis method based on multi-directional pixel-value differencing. Multimedia Tools and Applications, 81, 37403–37424. <https://doi.org/10.1007/s11042-022-13060-0>

[70] Chen, L., Lu, W., Huang, J., & Pan, J. (2023). StegTransformer: Transformer-based steganalysis with attention to image residuals. IEEE Transactions on Circuits and Systems for Video Technology, 33(12), 7499–7511.

<https://doi.org/10.1109/TCSVT.2023.3269040>

[71] Zhang, Y., Qian, Z., Chen, Y., & Feng, G. (2022). MDENet: Multi-domain expert network for spatial image steganalysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11), 7712–7724.

<https://doi.org/10.1109/TCSVT.2022.3177680>

[72] OpenAI. (2024). GPT-4V(ision) system card. OpenAI Technical Report. <https://doi.org/10.48550/arXiv.2303.08774>

[73] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of ICML 2021*, PMLR 139, 8748–8763. <https://doi.org/10.48550/arXiv.2103.00020>

[74] Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2024). Visual instruction tuning. *Advances in NeurIPS* 36. <https://doi.org/10.48550/arXiv.2304.08485>

[75] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in NeurIPS* 35, 24824–24837. <https://doi.org/10.48550/arXiv.2201.11903>

[76] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>

[77] Gemma Team, Google DeepMind. (2025). Gemma 3 Technical Report. arXiv preprint. <https://doi.org/10.48550/arXiv.2503.19786>

[78] Meta AI. (2024). The Llama 3 herd of models. arXiv preprint. <https://doi.org/10.48550/arXiv.2407.21783>

[79] Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., et al. (2024). Qwen2-VL: Enhancing vision-language model's perception of the world at any resolution. arXiv preprint. <https://doi.org/10.48550/arXiv.2409.12191>

[80] Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc,

- K., Mensch, A., Millican, K., Reynolds, M., et al. (2022). Flamingo: A visual language model for few-shot learning. *Advances in NeurIPS* 35, 23716–23736. <https://doi.org/10.48550/arXiv.2204.14198>
- [81] Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergkvist, W., Kepner, J., Gadepally, V., & Tiwari, D. (2023). From words to watts: Benchmarking the energy costs of large language model inference. *Proceedings of IEEE High Performance Extreme Computing Conference (HPEC)*, 1–9. <https://doi.org/10.1109/HPEC58863.2023.10363456>
- [82] Ollama Team. (2024). Ollama: Get up and running with large language models locally. GitHub Repository. <https://github.com/ollama/ollama> [Дата звернення: квітень 2025]
- [83] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). A survey of large language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2303.18223>
- [84] Leviathan, Y., Kalman, M., & Matias, Y. (2023). Fast inference from transformers via speculative decoding. *Proceedings of ICML 2023, PMLR* 202, 19274–19286. <https://doi.org/10.48550/arXiv.2211.17192>
- [85] Abouelnaga, Y., Salem, O., Radi, H., & Moustafa, M. (2016). CIFAR-10: KNN-based ensemble of classifiers. In *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1192-1195). IEEE. <https://doi.org/10.1109/CSCI.2016.0225>
- [86] Krizhevsky, A. and Hinton, G., (2009). Learning multiple layers of features from tiny images (Technical Report). University of Toronto.
- [87] DiSalvo, N. (2025). Steganographic Embeddings as an Effective Data Augmentation. *ArXiv*, [abs/2502.15245](https://arxiv.org/abs/2502.15245). <https://doi.org/10.48550/arXiv.2502.15245>
- [88] 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).

https://doi.org/10.48009/3_iis_2024_110

[89] K, V., Annem, P., Devarakonda, M., Jyothi, A., & Rayudu, N. (2025). Image Steganography with CNN Based Encoder - Decoder. *International Journal for Modern Trends in Science and Technology*, 11(04), 60–64.

<https://doi.org/10.5281/zenodo.15108976>

[90] Li, P., & Lu, A. (2018). LSB-based Steganography Using Reflected Gray Code for Color Quantum Images. *International Journal of Theoretical Physics*, 57(5), 1516–1548. <https://doi.org/10.1007/s10773-018-3678-6>

[91] 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. <https://doi.org/10.1145/3694965>

[92] The LabelMe-12-50k dataset. URL: <https://www.ais.uni-bonn.de/download/datasets.html>

[93] Cogranne, R., Giboulot, Q., & Bas, P. (2020). ALASKA#2: Challenging academic research on steganalysis with realistic images. *In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1–5). IEEE. <https://doi.org/10.1109/WIFS49906.2020.9360892>

[94] Yousfi, Y., Butora, J., Fridrich, J., & Giboulot, Q. (2023). Lightweight image steganalysis with block-wise pruning. *Scientific Reports*, 13. <https://doi.org/10.1038/s41598-023-43386-2>

[95] Liu, Q., Yang, Z., & Wu, H. (2023). JPEG steganalysis based on steganographic feature enhancement and graph attention learning. *Digital Signal Processing*, 139, 104063. <https://doi.org/10.1016/j.dsp.2023.104063>

ДОДАТОК А. АКТИ ВПРОВАДЖЕННЯ.

ЗАТВЕРДЖУЮ

Директор*

ТОВ «НПО ІНФОТЕХ»

Емма ПОРХУН

(підпис, ініціали, прізвище)

« 16 » 03 2026 р.

АКТ

про реалізацію результатів дисертаційної роботи
Мішкара Юрія Валентиновича
на тему: «Інформаційна технологія стегоаналізу зображень на основі
глибокого навчання та мультимодальних моделей»

ВСТУП

Цей акт засвідчує впровадження результатів дисертаційного дослідження, присвяченого розробці інформаційної технології стегоаналізу цифрових зображень із використанням моделей глибокого навчання та мультимодальних великих мовних моделей. Запропоновані методи спрямовані на підвищення ефективності виявлення прихованої інформації у цифрових зображеннях, автоматизацію процесів аналізу та покращення рівня інформаційної безпеки.

МЕТА

Розробка та впровадження ефективної інформаційної технології стегоаналізу цифрових зображень на основі сучасних моделей машинного навчання, згорткових нейронних мереж та мультимодальних моделей для підвищення точності виявлення прихованих даних.

ЗАВДАННЯ

1. Аналіз сучасних методів стеганографії та стегоаналізу цифрових зображень.
2. Розробка архітектури системи автоматизованого стегоаналізу.
3. Впровадження моделей глибокого навчання для класифікації цифрових зображень.
4. Розробка гібридного підходу на основі згорткових нейронних мереж та мультимодальних великих мовних моделей.
5. Тестування та оцінка ефективності запропонованих методів.

6. Інтеграція розроблених моделей у системи інформаційної безпеки та моніторингу цифрового контенту.

АРХІТЕКТУРА СИСТЕМИ

Інформаційна система складається з наступних основних модулів:

1. Модуль завантаження та попередньої обробки цифрових зображень.
2. Модуль побудови високочастотних фільтрів для виділення залишкових шумових компонентів.
3. Модуль аналізу зображень на основі моделей ResNet, SRNet, MobileNetV2 та EfficientNet.
4. Модуль мультимодального аналізу та генерації аналітичних висновків.
5. Модуль формування звітів та візуалізації результатів.

ОСНОВНІ ФУНКЦІЇ СИСТЕМИ

1. Автоматизоване виявлення прихованої інформації у цифрових зображеннях.
2. Аналіз статистичних та структурних особливостей зображень.
3. Формування аналітичних висновків щодо наявності стеганографічних вставок.
4. Підтримка роботи з великими масивами цифрового контенту.
5. Інтеграція із системами кібербезпеки та цифрової криміналістики.

ДОСЯГНУТІ РЕЗУЛЬТАТИ

1. Підвищено точність виявлення прихованої інформації у цифрових зображеннях.
2. Скорочено час аналізу зображень за рахунок використання легковажних моделей глибокого навчання.
3. Підвищено ефективність автоматизованого аналізу цифрового контенту.
4. Забезпечено можливість масштабування системи для використання у задачах інформаційної безпеки.
5. Реалізовано програмні засоби для автоматизованого стегоаналізу цифрових зображень.

ВІДГУК

Запропонована інформаційна технологія продемонструвала високу ефективність під час аналізу цифрових зображень та виявлення прихованих даних. Використання сучасних моделей глибокого навчання та мультимодальних моделей дозволило підвищити точність аналізу,

автоматизувати процес формування висновків та забезпечити можливість інтеграції розроблених рішень у сучасні системи інформаційної безпеки.