УДК 004.056.55

ПОКРАЩЕННЯ ЛОКАЛІЗАЦІЇ СТЕГОБІТІВ У ЦИФРОВИХ ЗОБРАЖЕННЯХ НА ОСНОВІ АНСАМБЛЮ SRM-ФІЛЬТРІВ

Д. С. Філонов[1], П. П. Яцура[1], Д. О. Прогонов[1]

[1] Навчально-науковий Фізико-технічний інститут

Анотація

У роботі розглянуто застосування попередньої обробки стеганограм з використанням ансамблю SRM-фільтрів для підвищення точності локалізації вбудованих повідомлень. Досліджено випадок використання моделей сегментації цифрових зображень на основі пірамідальної архітектури. В якості енкодерів моделей сегментації використано моделі: ResNet, EfficientNet, MobileNetV2; а в якості декодерів: U-Net, LinkNet, FPN та PSPNet. Стеганограми було сформовано за допомогою адаптивних методів вбудовування повідомлень до зображень-контейнерів HUGO, MG та MiPOD. За результатами дослідження встановлено, що попередня обробка стеганограм ансамблем SRM-фільтрів забезпечує вищі показники точності локалізації стегобітів порівняно з випадком використання окремих фільтрів. Для стеганографічного методу HUGO при рівні вбудування 50% попередня обробка сформованих стеганограм ансамблем SRM-фільтрів забезпечує підвищення значення коефіцієнта Жаккарда від 33% до 47%. у випадку використання ансамблю SRM-фільтрів для стеганографічних методів MG та MiPOD при рівні вбудування 50% значення коефіцієнта Жаккарда становить 47% та 44% відповідно.

Ключові слова: стеганографія, стегоаналіз, локалізація стегобітів, згорткові нейронні мережі, SRM-фільтри.

Вступ

Стрімке зростання обсягів цифрових даних, які циркулюють та зберігаються в інформаційно комунікаційних системах, підвищує необхідність за безпечення їх ефективного захисту [1]. Особлива увага фахівців з захисту інформації приділяється виявленню прихованих каналів комунікації, зокрема прихування повідомлень до мультимедійних даних, а саме цифрових зображень. Одним з найбільш складних випадків є застосування зловмисниками адаптивних методів вбудовування (HUGO, MG, MiPOD) [2, 3, 4], які концентрують стегобіти в тих ділянках зображення-контейнеру (ЗК), де зміни є найменш помітними для статистичного аналізу.

Методи глибокого навчання суттєво підвищили точність стегодетекторів. [5]. Проте дані стегодетектори забезпечують високу (більше 90%) точність виявлення лише у випадку середнього (більше 10%) та сильного (більше 20%) ступеня заповнення ЗК стегоданими. Для подолання даного обмеження становить інтерес використання методів локальної обробки зображень з метою виявлення положення пікселів ЗК, використаних для прихування окремих стегобітів.

Метою роботи є дослідження ефективності застосування методів попередньої обробки стеганограм, сформованих адаптивними методами вбудовування, за допомогою ансамблю SRM-фільтрів для підвищення точності локалізації стегобітів [6].

1. Адаптивні методи стеганографії

Особливістю адаптивних методів стеганографії для цифрових зображень є використання текстурних областей та контурів ЗК з метою надійного маскування змін яскравості пікселів, обумовлених прихованням повідомлень. Зокрема, це досягається за рахунок попереднього аналізу зображення та вибору ділянок з підвищеною варіативністю значень локальних статистичних параметрів — текстур, контурів тощо — де зміни є найменш помітними для статистичного аналізу.

Метод HUGO заснований на визначенні підгруп пікселів ЗК, модифікація значень яскравості котрих матиме мінімальних вплив на відповідні зміни статистичних параметрів. Чим візуально помітнішою є модифікація пікселя, тим менша ймовірність того, що він буде обраний для вбудовування повідомлення. Для чисельної оцінки вартості модифікації кожного пікселя використовується функція спотворення [2]:

$$\rho_{i,j} = D(\mathbf{X}, \mathbf{Y}^{i,j}) \tag{1}$$

де $\rho_{i,j}$ – вартість зміни пікселя з координатами (i, j);$D(\mathbf{X}, \mathbf{Y}^{i,j})$ – емпірична функція оцінки спотворення між вихідним зображенням $\mathbf{X}$ та зображенням $\mathbf{Y}^{i,j}$, отриманим після зміни яскравості пікселя (i, j) [2].

Метод MG заснований на моделюванні ЗК як послідовності незалежних випадкових величин з гаусовим розподілом [3]. На основі даної моделі для кожного

пікселя обчислюється ймовірність його модифікації β_i таким чином, щоб мінімізувати відхилення статистики зображення від оригіналу при заданому відносному обсязі вбудовування стегобітів до ЗК α:

$$\alpha_n = \sum \left(-2\beta_i \ln \beta_i - (1-2\beta_i)\ln(1-2\beta_i)\right) \quad (2)$$

де α – відносний обсяг вбудовування у бітах на піксель; β_i – ймовірність модифікації i-го пікселя.

У методі MiPOD кожне значення яскравості пікселю ЗК розглядається як випадкова величина з гаусовим розподілом, та обчислюється ймовірність зміни кожного пікселя таким чином, щоб вбудоване повідомлення було складніше виявити. З огляду на необхідність мінімізації зміни локальних статистичних параметрів зображення в процесі вбудовування стегоданих, значна кількість пікселів, використаних для вбудовування повідомлення, розташована у областях з текстурами та краями об'єктів, де модифікація значень яскравості пікселів є найменш помітною [4]:

$$\beta_n = \frac{e^{-\lambda\rho_n}}{1+2e^{-\lambda\rho_n}} \quad (3)$$

де β_n – ймовірність зміни n-го пікселя; ρ_n – вартість його модифікації; λ – параметр, що контролює обсяг вбудованих даних.

2. Методи попередньої обробки стеганограм

SRM-фільтри дозволяють виділити слабкі відхилення значень яскравості пікселів від локальних статистичних параметрів напівтонового зображення, обумовлені вбудовуванням стегобітів. На етапі попередньої обробки до вхідного напівтонового зображення $I \in \mathbb{R}^{H\times W}$ застосовується операція двовимірної згортки з фільтром k розміром 5×5 [6], у результаті чого отримується зображення R:

$$R = I * k \quad (4)$$

У випадку ансамблю з N фільтрів операція виконується для кожного фільтру k_n окремо, а результати об'єднуються у багатоканальне представлення:

$$\mathbf{R} = [R_1, R_2, \dots, R_N] \in \mathbb{R}^{H\times W\times N} \quad (5)$$

$$k_1 = \begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix} \quad k_2 = \begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$k_3 = \begin{pmatrix} 0 & 0 & -2 & 2 & -1 \\ 0 & 0 & 8 & -6 & 2 \\ 0 & 0 & -12 & 8 & -2 \\ 0 & 0 & 8 & -6 & 2 \\ 0 & 0 & -2 & 2 & -1 \end{pmatrix} \quad k_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -3 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$k_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Рис. 1. Приклади SRM-фільтрів, використаних в роботі [7]

У роботі запропоновано проводити попередню обробку стеганограм з використанням ансамблю SRM-фільтрів.Приклади використаних фільтрів наведено на рис. 1. Обрані фільтри відповідають різним просторовим напрямкам – ізотропному, горизонтальному, вертикальному та діагональному, що дозволяє враховувати просторові відхилення значень яскравості пікселів у різних напрямках при локалізації стегобітів. Отримані зображення R подаються до моделі сегментації з метою визначення положення пікселів ЗК, використаних для приховання стегобітів.

3. Пірамідальна архітектура моделей сегментації зображень в задачах локалізації стегобітів

У роботі розглянуто архітектуру типу encoder-decoder, що складається з енкодера для вилучення ієрархічних ознак зображення та декодера для формування попіксельної маски вбудованих стегобітів.

В якості енкодерів розглядались нейронні мережі, що реалізують різні підходи до оцінки статистичних параметрів оброблюваних зображень.

ResNet — мережа, побудована на основі залишкових з'єднань, що дозволяє ефективно тренувати глибокі мережі без втрати точності [8].

EfficientNet — мережа, яка масштабує одночасно глибину, ширину та роздільну здатність, досягаючи вищої точності та меншої кількості параметрів порівняно з поширеними згортковими мережами [9].

MobileNetV2 — мережа на основі інвертованих залишкових блоків, розроблена для ефективної роботи в умовах обмежених обчислювальних ресурсах [10].

Як декодери розглядались архітектури моделей сегментації зображень, орієнтовані на вирішення задачі попіксельної сегментації.

U-Net – архітектура з симетричним шляхом відновлення просторової роздільної здатності для точної локалізації пікселів ЗК [11].

LinkNet – архітектура для точної та швидкої сегментації з обмеженою кількістю параметрів [12].

FPN – архітектура з багаторівневим аналізом ознак на різних масштабах [13].

PSPNet – архітектура на основі пірамідального об'єднання ознак для сегментації пікселів в ЗК [14].

3.1. Функція втрат на основі коефіцієнту Дайса

У задачі локалізації стегобітів в цифрових зображеннях виникає проблема значного переважання фону (немодифікованих пікселів ЗК) над модифікованими пікселями. Для зменшення негативного впливу «фонових» пікселів на точність локалізації пікселів, використаних для вбудовування окремих стегобітів, в роботі запропоновано використовувати функцію втрат на основі коефіцієнту Дайса [15]:

$$L = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (6)$$

де L – значення функції втрат на основі коефіцієнту Дайса; N – загальна кількість пікселів у зображенні;

p_i – спрогнозована моделлю ймовірність того, що i-й піксель є модифікованим; g_i – фактичний стан пікселя (де 1 означає наявність прихованих даних, а $\emptyset$ — їх відсутність).

4. Результати експериментального дослідження

Для оцінювання точності локалізації стегобітів використано датасет ALASKA v2, що містить 10 000 зображень розміром 512 × 512 пікселів [16]. Стеганограми сформовано за допомогою трьох адаптивних стеганографічних методів — HUGO, MG та MiPOD — при ступені заповнення ЗК стегоданими $\phi = 3\%$, 20% та 50%. Як основну метрику якості локалізації обрано коефіцієнт Жаккара.

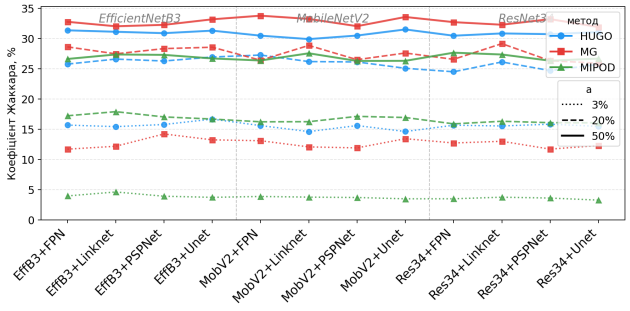

Рис. 2. Точність локалізації прихованих даних (коефіцієнт Жаккара) для різних комбінацій нейронних мереж при використанні фільтра k_2

4.1. Отримані результати у випадку використання окремих SRM-фільтрів

На рис. 2 та рис. 3 представлено результати для фільтрів k2 та k5, для яких було отримано найбільші та найменші значення коефіцієнту Жаккара серед усіх проаналізованих фільтрів.

За результатами аналізу для методу HUGO при використанні фільтра k_2 встановлено, що при ступені заповнення ЗК $\phi = 50\%$ значення коефіцієнту Жаккара досягає 33%, при $\phi = 20\%$ – 26%, а при $\phi = 3\%$ – 15%. Показано, що аналогічна тенденція до зниження точності локалізації при зменшенні ступеня заповнення ЗК характерна і для методів MG та MiPOD.

За результатами аналізу для методу HUGO при використанні фільтра k_5 встановлено, що при ступені заповнення ЗК $\phi = 50\%$ значення коефіцієнту Жаккара становить – 21%, при $\phi = 20\%$ – 14%, а при $\phi = 3\%$ – 7%. Показано, що аналогічна тенденція до зниження точності локалізації при зменшенні ступеня заповнення ЗК характерна і для методів MG та MiPOD.

У випадку використання фільтру k_2 та k_5 незначні зміни значення коефіцієнту Жаккара (у межах 1%–2%) між різними комбінаціями енкодерів та декодерів свідчить про незначний вплив архітектури нейромережі на точність локалізації.

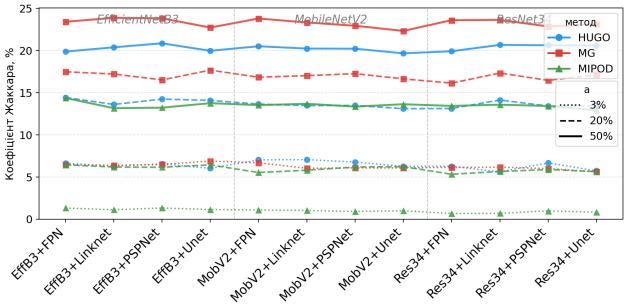

Рис. 3. Точність локалізації прихованих даних (коефіцієнт Жаккара) для різних комбінацій нейронних мереж при використанні фільтра k_5

4.2. Отримані результати у випадку використання ансамблю SRM-фільтрів

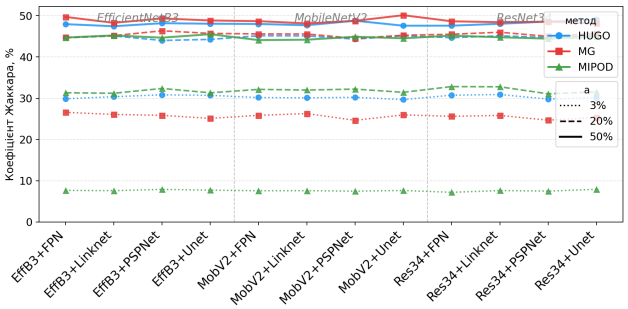

Рис. 4. Точність локалізації прихованих даних (коефіцієнт Жаккара) для різних комбінацій нейронних мереж при використанні ансамблю фільтрів

На рис. 4 представлено результати оцінювання якості локалізації стегобітів при використанні ансамблю SRM-фільтрів, який за підсумками дослідження продемонстрував найкращі показники IoU порівняно з окремими фільтрами.

За результатами аналізу ансамблю для методу HUGO встановлено, що при $\phi = 50\%$ значення коефіцієнту Жаккара становить 48%, при $\phi = 20\%$ – 46%, а при $\phi = 3\%$ – 31%. Показано, що аналогічна тенденція характерна і для методу MG. Для методу MiPOD виявлено більш суттєве зниження точності локалізації при зменшенні ступеня заповнення ЗК: при $\phi = 50\%$ значення коефіцієнту Жаккара складає 45%, при $\phi = 20\%$ – 33%, а при $\phi = 3\%$ – 8%.

Незначні зміни значення коефіцієнту Жаккара (у межах 1%–2%) між різними комбінаціями енкодерів та декодерів свідчить про незначний вплив архітектури нейромережі на точність локалізації.

З порівняння результатів для випадку використання окремих SRM-фільтрів та ансамблю SRM-фільтрів показано, що застосування ансамблю фільтів забезпечує підвищення точності локалізації стегобітів у ЗК. Для методу HUGO при $\phi = 50\%$ значення коефіцієнту Жаккара зростає з 33% у випадку використання окремих фільтрів до 48% у випадку використання ансамблю фільтрів.

Висновки

У роботі запропоновано підхід до підвищення точності попіксельної локалізації стегобітів, який полягає у попередній обробці стеганограм за допомогою ансамблю SRM-фільтрів. При оброці стеганограм ансамблем фільтрів отримано вищі показники точності локалізації прихованих повідомлень порівняно з випадком використанням окремих фільтрів. За результатами дослідження встановлено, що для методу HUGO при 50% ступені заповнення ЗК застосування ансамблю SRM-фільтрів дозволяє отримати значення коефіцієнту Жаккара до 48%, тоді як при обробці окремими фільтрами коефіцієнт Жаккара має значення від 19% до 33%. Отримані результати підтверджуються для методів MG та MiPOD при всіх досліджених рівнях вбудовування. Встановлено, що вибір конкретної комбінації архітектури нейромережі має незначний вплив на якість локалізації стегобітів.

Перелік використаних джерел

1. *European Union Agency for Cybersecurity (ENISA).* ENISA Threat Landscape 2024 : тех. звіт. / ENISA. — 2024. — URL: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.
2. *Pevný T., Filler T., Bas P.* Using High-Dimensional Image Models to Perform Highly Undetectable Steganography // Information Hiding. IH 2010. Lecture Notes in Computer Science. T. 6387. — 2010. — C. 161—177. — URL: https://doi.org/10.1007/978-3-642-16435-4_13.
3. *Fridrich J., Kodovský J.* Multivariate Gaussian model for designing additive distortion for steganography // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2013. — URL: https://doi.org/10.1109/ICASSP.2013.6638198.
4. *Sedighi V., Cogranne R., Fridrich J.* Content-Adaptive Steganography by Minimizing Statistical Detectability // IEEE Transactions on Information Forensics and Security. — 2016. — T. 11, № 2. — C. 221—234. — URL: https://doi.org/10.1109/TIFS.2015.2486744.
5. *Boroumand M., Chen M., Fridrich J.* Deep Residual Network for Steganalysis of Digital Images // IEEE Transactions on Information Forensics and Security. — 2019. — T. 14, № 5. — C. 1181—1193. — URL: https://doi.org/10.1109/TIFS.2018.2871749.
6. *Fridrich J., Kodovský J.* Rich Models for Steganalysis of Digital Images // IEEE Transactions on Information Forensics and Security. — 2012. — T. 7, № 3. — C. 868—882. — URL: https://doi.org/10.1109/TIFS.2012.2190402.
7. GBRAS-Net: A Convolutional Neural Network Architecture for Spatial Image Steganalysis / R. Tabares-Soto, H. B. Arteaga-Arteaga, A. Mora-Rubio [та ін.] // IEEE Access. — 2021. — T. 9. — C. 14340—14350. — URL: https://doi.org/10.1109/ACCESS.2021.3052494.
8. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren, J. Sun // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — URL: https://doi.org/10.1109/CVPR.2016.90.
9. *Tan M., Le Q. V.* EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning (ICML). — 2019. — C. 6105—6114. — URL: http://proceedings.mlr.press/v97/tan19a.html.
10. MobileNetV2: Inverted Residuals and Linear Bottlenecks / M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — URL: https://doi.org/10.1109/CVPR.2018.00474.
11. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). — 2015. — C. 234—241. — URL: https://doi.org/10.1007/978-3-319-24574-4_28.
12. *Chaurasia A., Culurciello E.* LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation // 2017 IEEE Visual Communications and Image Processing (VCIP). — 2017. — URL: https://doi.org/10.1109/VCIP.2017.8305148.
13. Feature Pyramid Networks for Object Detection / T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — URL: https://doi.org/10.1109/CVPR.2017.106.
14. Pyramid Scene Parsing Network / H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — URL: https://doi.org/10.1109/CVPR.2017.660.
15. *Milletari F., Navab N., Ahmadi S.-A.* V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation // arXiv preprint arXiv:1606.04797. — 2016. — URL: https://arxiv.org/abs/1606.04797.
16. *Cogranne R., Giboulot Q., Bas P.* ALASKA#2: Challenging Academic Research on Steganalysis with Realistic Images // 2020 IEEE International Workshop on Information Forensics and Security (WIFS). — 2020. — URL: https://ieeexplore.ieee.org/document/9360896.