КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 004.056, 004.415.24 DOI:10.34229/2707-451X.21.1.8

Н.В. КОШКІНА

ПРО ВПЛИВ ПАРАМЕТРІВ ЗОБРАЖЕНЬ У ФОРМАТІ JPEG НА ТОЧНІСТЬ ЇХ СТЕГАНОАНАЛІЗУ

Досліджено вплив параметрів та вмісту зображень у форматі JPEG на точність їх стеганоаналізу за методами на базі машинного навчання. Сформульовано закономірності, які слід враховувати під час інтерпретації результатів стеганоаналізу. Надано рекомендації щодо вибору контейнерів зображень.

Ключові слова: *інформаційна безпека, стеганографія, стеганоаналіз, інтелектуальні комп'ютерні системи, машинне навчання, точність детектування.*

Вступ. Стеганографія – це наука про непомітне приховування повідомлень у типових для даного середовища контейнерах-носіях. Вона може бути застосована у військовій або комерційній сферах, зокрема, для прихованої передачі важливих документів, підвищення секретності зашифрованих повідомлень, захисту інформації від змін, організації мультимедійних баз даних тощо. Разом з тим стеганографія може бути використана і в зловмисних цілях. Наприклад, для терористичних чи шпигунських комунікацій, поширення незаконних матеріалів, приховання програм-шкідників, викрадення комерційних таємниць і т. і.

Необхідність протидіяти неправовому застосуванню стеганографічних методів та програм привела до бурхливого розвитку методів стеганоаналізу – науки про виявлення прихованих повідомлень у відкритих джерелах та каналах.

Практична оцінка ефективності методів стеганографії і стеганоаналізу здійснюється на деякому наборі типових цифрових контейнерів, як то аудіосигнали, зображення, відеоконтейнери тощо. Проте обґрунтування вибору того чи іншого тестового набору в багатьох наукових публікаціях залишається поза дужками. Зокрема, не існує єдиних правил чи методики формування вихідного набору контейнерів, з використанням якого обчислюються оцінки точності існуючого, вдосконаленого чи нового методу стеганоаналізу зображень. Різні підходи до вибору тестових зображень можуть спричиняти проблеми при відтворенні, порівнянні стеганоаналітичних методів, їх застосуванні до зображень з іншими параметрами і т. п.

Деякі дослідники в свої роботах зауважують, що точність стеганоаналізу залежить від джерела та параметрів зображень [1, 2]. Невраховування цього факту, зокрема, тягне за собою проблему зниження точності стеганоаналізу при переході з лабораторії до реального світу. Тож, щоб мати змогу передбачити як поведе себе той чи інший стеганоаналітичний метод на зображеннях інших розмірів, іншого степеня гладкості, текстурованості чи, наприклад, стиснутих

з іншим коефіцієнтом якості, потрібно досліджувати існуючі залежності точності стеганоаналізу від параметрів зображень. Такі дослідження важливі для правильного застосування методів на практиці та коректної інтерпретації результатів стеганоаналітичних атак.

Способи формування тестових наборів зображень. У науковій літературі в основному використовується три варіанти отримання тестових зображень: 1) загальні («еталонні») набори даних, 2) завантаження зображень із різних інтернет-джерел, 3) формування наборів зображень із власних архівів. Кожен із варіантів має свої переваги та недоліки.

Загальні набори даних фігурують у різних областях, зокрема, цифровій обробці зображень, проблемах штучного інтелекту, кібербезпеці. Вони дають можливість перевіряти, порівнювати, відтворювати різні методи та алгоритми, що сприяє прозорості та цілісності академічних досліджень. Вивчення наукових публікацій дозволило виділити ряд загальних наборів зображень, що застосовуються при дослідженні якості тих чи інших методів стеганоаналізу.

1. Фотогалерея NRCS (Natural Resources Conservation Service) Міністерства сільського господарства США (Режим доступу: https://photogallery.sc.egov.usda.gov/). Містить *.tif та *.jpeg кольорові зображення порівняно великих розмірів (2100×1500 пікселів, 2700×1800, 4288×2848 та інші). Наприклад, у роботі [3] використано 3000 зображень з цієї галереї, які приведені до розмірів 640×418 та конвертовані у відтінки сірого. В роботі [4], як частина тестового набору з 5000 графічних контейнерів, застосовано 1543 зображення, що були обрізані від центру до розмірів 512×512 пікселів. У роботі [5] використано 1576 зображень з NRCS (конвертованих у відтінки сірого) розмірами 2100×1500 пікселів, які склали один із чотирьох тестових наборів, на яких досліджувалася якість запропонованого методу. В роботі [6] використано 3161 зображення розмірами 2100×1500 пікселів, кожне з них було поділене на 4 окремі контейнери, тобто загальний тестовий набір містив 12644 зображення розмірами 525×375 пікселів.

2. ImageNet (Режим доступу: http://image-net.org/) – база даних зображень, що створювалася для дослідження алгоритмів пошуку, індексації, анотації та організації даних. Вона містить понад 14 мільйонів *.jpeg зображень, розділених за більш ніж 20-ти тисячами категорій. Зображення різних (порівняно невеликих) розмірів та стиснуті з різними коефіцієнтами якості. У роботі [7] використовувалося три набори, сформованих на основі ImageNet – набір на 50 тисяч, 500 тисяч та 5 мільйонів контейнерів. Зображення обиралися таким чином, щоб їх розміри перевищували 256×256 пікселів, а потім обрізалися до згаданих розмірів, конвертувалися у відтінки сірого та повторно стискалися з коефіцієнтом якості 75. Для формування однієї з контрольних вибірок у [8] використовувалося сто тисяч випадково обраних з ImageNet зображень, конвертованих у відтінки сірого та обрізаних до розмірів 256×256. У роботі [9] використовувалася версія ImageNet CLS-LOC, що містить 1281167 зображень, відсортованих за тисячею категорій. Кожне зображення, сторони якого більші за 256 пікселів, обрізалося до розмірів 256×256 та перестискалося з коефіцієнтом 75.

3. BOSSbase 1.01 (Режим доступу: http://agents.fel.cvut.cz/boss/), Break Our Steganography System – база, яка розроблялася саме для досліджень ефективності методів стеганоаналізу та містить 1000 *.pgm зображень. Зображення були відзняті у форматі RAW сімома різними камерами, перетворені у сірі полутонові, зменшені з використанням алгоритму Ланшоца з виключеним згладжуванням таким чином, щоб менша сторона складала 512 пікселів, а потім обрізані до розмірів 512×512. Ця база та її похідні застосовується чи не найчастіше. Наприклад, у роботах [10, 11] використовувалася версія BOSSbase 0.92, що містить 9074 зображення, а в [12] безпосередньо BOSSbase 1.01. У дослідженні [13] фігурують зображення з BOSSbase 1.01, стиснуті з коефіцієнтами якості 75 та 95. В роботі [14] – зображення з BOSSbase, перетворені у формат JPEG з 50, 75 та 95 коефіцієнтами якості, для деяких експериментів розмір зображень було зменшено до 256×256 пікселів.

4. MIRFickr (Режим доступу: http://press.liacs.nl/mirflickr/). Існує дві бази даних зображень MIRFlickr: MIRFlickr 1M, що містить 1 мільйон *.jpeg зображень, та більш ранній варіант MIRFlickr 25k, що містить 25 тисяч. Зображення з урахуванням індексу цікавості були завантажені з сайту соціальних фото Flickr, куди їх додавали різні користувачі. Зображення в цій базі в основному стиснуті з коефіцієнтом якості 96, але є деяка кількість з іншими коефіцієнтами (наприклад, 92 та 95 в MIRFlickr 25k). Розміри різні, але не перевищують 500×500 пікселів (для MIRFlickr 25k 399×462 пікселі у середньому). Наприклад, у роботі [15] використовувалося 9000 випадково обраних зображень з бази MIRFlickr 25k, а в [16] – 1300 випадково обраних зображень з MIRFlickr 1M.

Та все ж такі набори містять зображення з фіксованими параметрами та/або зображення з обмеженої кількості джерел, тому оцінки точності стеганоаналізу, отримані з їх використанням відображають тільки частину можливих випадків. Крім того, не завжди наявної кількості зображень достатньо для запланованих експериментів. Та, як правило, «еталонні» зображення мають малі розміри і не зрозуміло як буде співвідноситися отримана точність з точністю стеганоаналізу більших за розміром зображень, особливо в частотній області.

Більш наближеними до практики є оцінки, отримані шляхом завантаження великої кількості зображень з усіх куточків мережі Інтернет, тобто потенційно реальних контейнерів. Так, наприклад, в роботі [17] використовувалося 49678 зображень з мережі Інтернет. В роботі [16] окрім MIRFlickr 1M використовувалися також 5000 зображень товарів з сайту https://www.amazon.cn і 2,4 мільйони зображень, завантажених з веб-сайту Flickr та за допомогою пошукової системи Google Images.

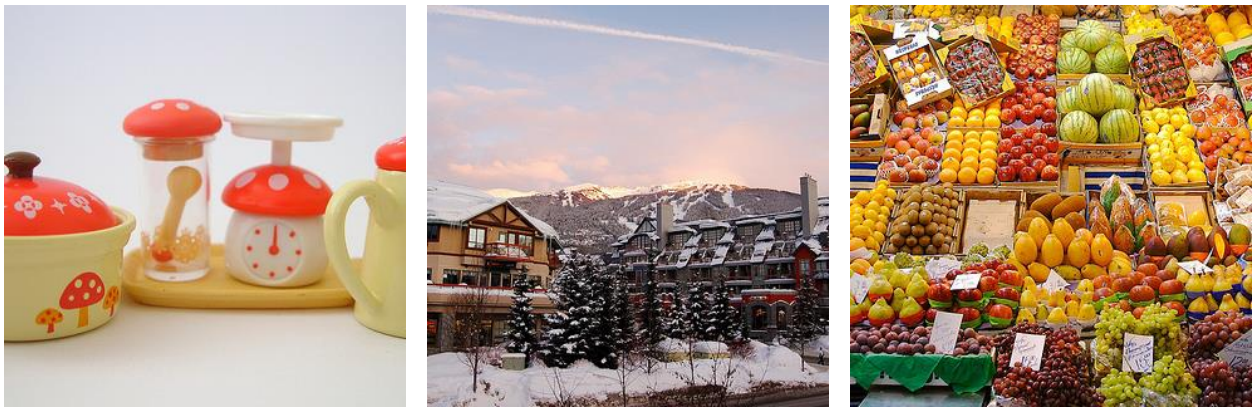
Щоб виявити існуючі закономірності в [18] з різних інтернет-ресурсів було зібрано базу з 1,1 мільйону унікальних зображень. За допомогою них автори досліджували ефективність різних існуючих на той час універсальних методів стеганоаналізу. Так як робота виконувалася у 2006 році, то атаки здійснювалися на стеганографічні методи, які на сьогодні вважаються нестійкими до пасивного стеганоаналізу, зокрема на Outguess та F5. У роботі [18] показано, що точність стеганоаналізу різна для зображень, що безпосередньо створені у форматах без втрат даних (*.bmp, *.pgm тощо), та тих, що створені у форматі *.jpeg, а потім конвертовані до одного з форматів без втрат. Та багато питань із цього дослідження його автори залишили для майбутнього, зокрема це залежність точності стеганоаналізу від розмірів контейнера.

Недоліком варіанту завантаження зображень із різних інтернет-джерел є в першу чергу неконтрольована обробка, тобто невідомо яким процедурам, що змінюють природню статистику, вони підлягали, перед тим як потрапити в мережу. Не можна також відкидати імовірність, що деякі з них вже містять приховані стеганографічними методами таємні повідомлення або цифрові водяні знаки для захисту авторських прав. Саме щоб запобігти вказаним недолікам науковці звертаються до самостійного формування тестових наборів зображень згідно з потребами дослідження.

Опис стартових умов стеганоаналізу. Отже, мета даної роботи – дослідження впливу параметрів та вмісту зображень на точність їх стеганоаналізу. Це можна здійснити шляхом проведення серій експериментів, у яких однакові всі параметри, окрім одного (досліджуваного). Висновки про наявність чи відсутність закономірностей будуть виведені за результатами аналізу поведінки показників точності за умов зміни досліджуваного параметру.

Серед наявних форматів контейнерів було обрано JPEG, зважаючи на його більшу поширеність у комп’ютерних системах та мережах у порівнянні з форматами без втрат. З вищезгаданих загальних наборів вибір зупинено на MIRFlickr 25k. В даному випадку керувалися по-перше тим, що це кольорові зображення, а саме такі в більшій мірі використовуються у реальних застосуваннях, а по-друге, простотою їх завантаження (один архів на 2.85 Гб).

Щоб зафіксувати однакові параметри контейнерів, було здійснено аналіз набору MIRFlickr 25k та відбраковування зображень, коефіцієнт якості яких не рівний 96, а потім тих, хоч одна зі сторін яких менша за 320 пікселів. Всі залишені зображення шляхом обрізування від центру були приведені до розмірів 320×320 пікселів. Далі, щоб мати можливість відслідкувати вплив вмісту, від якого напряму залежить кількість придатних для вкраплення коефіцієнтів дискретного косинусного перетворення (ДКП), зображення, що залишилися, були згруповані у три субнабори за кількістю придатних коефіцієнтів: MIRFlickr-min, MIRFlickr-midi та MIRFlickr-max. Для рівної кількості в кожному них залишено по 7000 зображень. Типові представники даних субнаборів показані на рис. 1.



а б в

РИС. 1. Типові представники: а – MIRFlickr-min, б – MIRFlickr-midi, в – MIRFlickr-max

Розміри вихідних зображень MIRFlickr 25k недостатні для дослідження впливу розмірів контейнерів на точність їх стеганоаналізу (найбільші зображення – 500×500 пікселів і їх 870 у базі, інші менші, а найменше – 256×192 пікселя). Тому на основі власних фотоархівів створено додатковий тестовий набір OWNSET95 з 7000 кольорових зображень розмірами 1200×1600 пікселів, отриманих з понад 10 джерел та стиснених з коефіцієнтом якості 95.

Наступний підготовчий крок полягає у створенні наборів стеганозображень. На цьому кроці можливе застосування декількох підходів до визначення довжини прихованих повідомлень.

1. Використання повідомлень сталої довжини. Тобто коли розмір повідомлення однаковий для всіх тестових контейнерів.
2. Використання повідомлень, розмір яких складає певний відсоток від розміру зображення-контейнера.
3. Використання повідомлень, розмір яких складає певний відсоток від кількості придатних для вкраплення коефіцієнтів ДКП зображення.

Серед цих трьох обрано останній підхід, оскільки він гарантує рівний відсоток змін для всіх зображень незалежно від розмірів та вмісту.

Стеганографічний метод, який буде ціллю атаки, – J-UNIWARD [19], – один з найбільш безпечних на сьогодні. Для створення стеганозображень використовувався J-UNIWARD з 50 % навантаженням на придатні для приховування ДКП коефіцієнти, тобто не DC (постійна складова) та не нульові.

Подальший стеганоаналіз здійснювався за універсальною схемою з навчанням та класифікацією на базі ансамблевого класифікатора з лінійним дискримінантом Фішера [20]. Детальніше ця схема описана у роботі [21]. Для отримання характеристичних векторів зображень застосовувалися дві чутливі до J-UNIWARD високорозмірні статистичні моделі.

1. GFR – модель характеристичних векторів, побудованих як гістограми квантованих залишків, отриманих з використанням двовимірних фільтрів Габора (Gabor Filter Residual). 2D-фільтри Габора описують особливості текстури зображення з позицій різних масштабів та орієнтацій. Модель запропонована в [22]. Вона розраховується для просторового представлення. Спочатку зображення декомпресується без заокруглення значень. Генерується двовимірний банк фільтрів Габора, що включає фільтри з двофазним зміщенням $(\varphi = 0, \pi)$, чотирма масштабами ($\sigma = 0.5$, 0.75, 1, 1.25) та 32-ма орієнтаціями ($\theta = 0, \pi/32, \ldots 31\pi/32$). Далі розпаковане зображення згортається з 8×8 2D-фільтром Габора $G^{\varphi,\sigma,\theta}$, щоб отримати залишкове зображення $U^{\varphi,\sigma,\theta}$. Відповідно до фази $(a,b),\ 0 \le a,b \le 7$ залишки поділяються на 64 підмножини $U_{a,b}^{\varphi,\sigma,\theta}$, для кожної з яких обчислюються гістограми

$$h_{a,b}^{\varphi,\sigma,\theta}(r) = \frac{1}{\left|U_{a,b}^{\varphi,\sigma,\theta}\right|} \sum\nolimits_{u \in U_{a,b}^{\varphi,\sigma,\theta}} \left[Q_T\left(|u|/q\right) = r \right],$$

де Q_T – квантувач з цілочисельними центроїдами $\{0,1,\ldots T\}$, q – крок квантування, а [P] – дужка Іверсона, що дорівнює 0, коли твердження P хибне, і 1, коли P істинне. Завдяки симетрії отримані 64 гістограми $h_{a,b}^{\varphi,\sigma,\theta}$ зливаються до 25-ти: разом складаються гістограми, індекси яких (a,b), $(a,8-b)$, $(8-a,b)$, $(8-a,8-b)$, за умови що ці показники залишаються у межах $\{0,1,\ldots 7\} \times \{0,1,\ldots 7\}$. Потім 25 гістограм з'єднуються в гістограму $h^{\varphi,\sigma,\theta}$ залишку $U^{\varphi,\sigma,\theta}$. Гістограми $h^{\varphi,\sigma,\pi-\theta}$ та $h^{\varphi,\sigma,\theta}$ об'єднуються згідно симетричним орієнтаціям. Наостанок результуючі гістограми стають елементами характеристичного вектора.

2. DCTR – модель характеристичних векторів із ДКП залишків (Discrete Cosine Transform Residual), що враховує фази. Була запропонована у роботі [23] і також будується для просторового представлення. Елементи характеристичних векторів у даній моделі – це гістограми залишків, отримані з використанням базових шаблонів ДКП, що мають вигляд

$$B_{mn}^{(k,l)} = \frac{w_k w_l}{4} \cos\frac{\pi k(2m+1)}{16} \cos\frac{\pi l(2n+1)}{16}, \quad w_0 = \frac{1}{\sqrt{2}}, \ w_k = 1(k>0), \ 0 \le m,n \le 7.$$

Для створення характеристичних векторів потрібно обчислити 64 згортки розпакованого в просторову зону без заокруглення до цілих чисел JPEG зображення з 64 ядрами 8×8 та сформувати нормалізовані гістограми, аналогічні тим, що описані для попередньої моделі. З метою подальшої компактизації векторів використовується симетрія шаблонів – гістограми об'єднуються за тим же принципом, що й у моделі GFR. Загальна розмірність симетризованого характеристичного вектора становить 64×(36/4+24/2+4)×(T+1) = 1600×(T+1). Як компроміс між швидкодією та точністю пропонується використовувати T = 4.

Зауважимо, що ці моделі здатні з високою точністю виявляти стеганографічні приховування за різними варіантами методу найменшого значущого біту, про що, зокрема, свідчать результати досліджень [21, 24], та є кращим вибором для детектування J-UNIWARD у порівнянні зі статистичними моделями, що будуються у частотній зоні [21].

Вплив вмісту зображень на точність їх стеганоаналізу. На першому етапі експериментів випадково обрані 6000 пустих та парні їм 6000 стеганозображень кожного набору MIRFlickr-min, MIRFlickr-midi та MIRFlickr-max по черзі використовувалися для навчання відповідного класифі-

катора, а 1000 пустих та 1000 заповнених, що залишилися, – для контролю точності. Крім того були здійснені всі можливі серії перехресних тестів – коли класифікатор навчався на контейнерах одного субнабору, а потім класифікував контейнери іншого. Серія тут та в подальшому складається з 10 експериментів, кожний із яких відрізняється від інших тільки стартовим числом генератора випадкових чисел, що ініціює поділ зображень на навчальну та контрольну вибірки. Результуюча оцінка точності – її середнє значення в серії. Крім того для кожної серії буде зазначатися також середньо квадратичне відхилення точності (в дужках після середнього значення). Точність стеганоаналізу, отримана на цьому етапі, наведена у табл. 1.

Зауважимо, що зменшення навчальної вибірки та, відповідно, збільшення контрольної у межах тестових наборів тягне за собою зменшення середньої точності та середньоквадратичного відхилення. Поділ 6000:1000 здійснено з огляду на прагнення максимально проявити чутливість моделей до стеганоперетворення зображень із заданими параметрами та не вийти за межі відсотку в середньому відхиленні.

ТАБЛИЦЯ 1. Точність стеганоаналізу при формуванні навчальної та контрольної вибірок з урахуванням вмісту зображень

Навчання: / Контроль:	MIRFlickr-min	MIRFlickr-midi	MIRFlickr-max
Модель GFR			
MIRFlickr-min	**70.4 (0.7)**	68.8 (0.3)	62.9 (0.8)
MIRFlickr-midi	60.3 (0.6)	62.5 (0.5)	60.1 (0.7)
MIRFlickr-max	52.0 (0.4)	55.9 (0.3)	56.0 (0.7)
Модель DCTR			
MIRFlickr-min	68.5 (0.6)	66.9 (0.7)	62.8 (0.4)
MIRFlickr-midi	60.2 (0.5)	**63.9 (**0.8)	63.2 (0.8)
MIRFlickr-max	53.5 (0.7)	59.6 (0.7)	**63.3 (**0.8)

З чисельних оцінок табл. 1 відстежується дві закономірності.

1. Точність краща тоді, коли для навчання та контролю використовуються зображення з одного субнабору. Так як стеганоаналітик завжди може оцінити у досліджуваному зображенні кількість придатних для приховування ДКП коефіцієнтів, урахування цієї закономірності дозволить покращити точність стеганоаналізу шляхом використання класифікатора, навченого на зображеннях з подібною кількістю придатних ДКП коефіцієнтів, які контрольовано пусті або заповнені.

2. Краще за інші виявляються зображення з субнабору MIRFlickr-min, які мають відносно невелику кількість придатних для приховування ДКП коефіцієнтів. Цікаво, що стеганозображення, сформовані з MIRFlickr-min, містять найкоротші повідомлення та, незважаючи на це, вони виявляються з найкращою точністю. А з найгіршою точністю виявляються найдовші повідомлення, що містяться в стеганозображеннях із субнабору MIRFlickr-max.

Також з даних цієї таблиці можна зробити наступне спостереження: зображення з набору MIRFlickr-min точніше класифікуються за характеристичними векторами моделі GFR, водночас як наборів MIRFlickr-midi та MIRFlickr-max – моделі DCTR.

Друга закономірність була додатково досліджена для випадку, коли навчання класифікатора здійснюється без урахування кількості придатних коефіцієнтів. Тобто класифікатор був навчений на 6000 пустих та 6000 стеганозображеннях, випадковим чином обраних з усіх трьох субнаборів та відповідаючих їм стеганонаборів (по 2000 з кожного). Для контролю використовувалися тільки пусті та стеганозображення одного з субнаборів, які раніше не були обрані для навчання. Результати наведені у табл. 2.

Ці тести підтвердили другу закономірність. Також бачимо, що загальний класифікатор дав гіршу точність, аніж той, що враховує кількість придатних коефіцієнтів.

Отримана точність свідчить про чутливість розглянутих моделей до J-UNIWARD, але вона недостатньо висока, щоб можна було вести мову про доцільність використання зазначених моделей та параметрів стеганоаналізу на практиці. Один з шляхів покращення точності (до певної межі) – це збільшення кількості контейнерів для навчання класифікатора. В даному випадку в цілому маємо 21000 пустих та 21000 заповнених контейнерів, з яких по 1000 тих та інших залишаємо для контролю. За умови використання загального класифікатора, що буде вчитися на всіх наявних контейнерах, крім випадково обраних з певного класу контрольних, максимальна кількість зображень для навчання – це 40000. Точність, обчислена за цієї умови, наведена у табл. 3.

Бачимо, що порівняно з даними табл. 2 точність зросла, проте щоб дослідити межу цього зростання потрібні вже додаткові зображення з такими ж параметрами та досить значні обчислювальні ресурси, тому дане питання залишиться за рамками роботи.

ТАБЛИЦЯ 2. Точність стеганоаналізу зображень з різних субнаборів при умові використання одного загального класифікатора

MIRFlickr-min	MIRFlickr-midi	MIRFlickr-max
Модель GFR		
69.6 (0.9)	62.2 (0.4)	56.0 (0.5)
Модель DCTR		
68.3 (0.8)	**63.2** (0.8)	**60.3** (1.0)

ТАБЛИЦЯ 3. Точність стеганоаналізу зображень з різних субнаборів при збільшенні кількості контейнерів у навчальній вибірці

MIRFlickr-min	MIRFlickr-midi	MIRFlickr-max
Модель GFR		
70.5 (0.7)	62.8 (0.7)	56.4 (0.5)
Модель DCTR		
69.9 (0.7)	**64.7** (0.6)	**62.5** (0.7)

Вплив коефіцієнта якості зображень на точність їх стеганоаналізу. При розміщенні зображень у веб-фотогалереях чи на інших інтернет-ресурсах, як правило, їх оптимізують згідно певним правилам. Зокрема є обмеження на обсяг файлів: чим він менший, тим краще (проте якість зображення має залишатися прийнятною). Коли стоїть задача залишити розміри зображення без змін, для зменшення обсягу файла застосовують стиснення з меншими коефіцієнтами. Тому варто дослідити як зменшення коефіцієнта якості вплине на результати стеганоаналізу. Для цього було взято три вихідні субнабори MIRFlickr та на їх основі сформовано набори порожніх контейнерів, стиснених з коефіцієнтами якості від 55 до 95 з кроком 5 (стиснення здійснювалося з використанням програми IrfanView – поширеного графічного переглядача та редактора). В ці нові порожні контейнери вкраплялися повідомлення, тобто створювалися відповідні стеганонабори. А далі здійснювався стеганоаналіз з тими ж параметрами, що в першій серії експериментів (субнабори навчання та контролю співпадали, 6000+6000 контейнерів для навчання, 1000+1000 – для контролю). Результати даного дослідження наведені у табл. 4.

З даних цієї таблиці бачимо, що зі зменшенням коефіцієнта якості, точність стеганоаналізу зростає. Для наборів MIRFlickr-midi та MIRFlickr-max стиснених з коефіцієнтами 95 (табл. 4) і 96 (табл. 1 – 3) перевагу в точності має модель DCTR, для всіх інших випадків точність краща при використанні моделі GFR.

Також були проведені перехресні тести, в який класифікатор навчався на зображеннях, стиснених з одним коефіцієнтом якості, а потім використовувався для класифікації зображень з іншим коефіцієнтом. Результати цих тестів показали, що коли коефіцієнти якості зображень навчальної та контрольної вибірок не співпадають, точність стеганоаналізу суттєво погіршується, при чому в деяких випадках класифікатор стає повністю нечуттєвим (точність ≈ 50 %). Також у деяких експериментах спостерігалося збільшення середньоквадратичного відхилення, тобто погіршувалася стабільність оцінок точності.

ТАБЛИЦЯ 4. Точність стеганоаналізу зображень, стиснених з різним коефіцієнтом якості

Коефіцієнт якості: / Субнабір:	55	65	75	85	95
Модель GFR					
MIRFlickr-min	91.6 (0.6)	90.1 (0.5)	88.8 (0.7)	84.3 (0.6)	71.3 (1.0)
MIRFlickr-midi	89.5 (0.6)	87.9 (0.9)	86.4 (0.7)	79.2 (0.8)	63.7 (0.6)
MIRFlickr-max	85.9 (0.8)	83.1 (0.9)	78.7 (0.9)	74.3 (0.8)	58.2 (0.6)
Модель DCTR					
MIRFlickr-min	87.0 (0.6)	85.5 (0.7)	83.6 (0.7)	79.3 (1.0)	70.3 (0.5)
MIRFlickr-midi	83.7 (0.6)	81.5 (0.6)	78.8 (0.6)	72.2 (0.6)	64.1 (0.3)
MIRFlickr-max	78.0 (0.7)	74.6 (0.9)	69.1 (0.9)	65.2 (0.5)	62.9 (0.9)

Якщо ж формувати загальну навчальну вибірку, що містить пари контейнерів із всіма досліджуваними коефіцієнтами стиснення, ансамблевий класифікатор залишається чуттєвим, проте його точність погіршується. Як приклад цього, у табл. 5 наведено точність стеганоаналізу, отриману з використанням загального класифікатора, що навчений на 6000 пар контейнерів, де кожна тисяча пар контейнерів має свій коефіцієнт якості: 96, 95, 85, 75, 65 або 55. Також зазначимо, що у комбінації з загальним класифікатором модель GFR показала кращі результати, ніж DCTR.

ТАБЛИЦЯ 5. Точність стеганоаналізу зображень, стиснених з різним коефіцієнтом якості, при умові використання одного загального класифікатора

Коефіцієнт якості: / Субнабір:	55	65	75	85	95	96
Модель GFR						
MIRFlickr-min	89.6 (0.7)	88.9 (0.6)	87.4 (0.9)	82.7 (0.9)	67.93 (0.8)	65.9 (1.0)
MIRFlickr-midi	88.4 (0.5)	86.6 (0.7)	84.6 (0.7)	78.7 (0.5)	62.7 (0.7)	61.4 (0.6)
MIRFlickr-max	84.2 (1.0)	81.9 (1.1)	77.2 (0.7)	73.0 (0.6)	57.2 (0.5)	55.6 (0.4)
Модель DCTR						
MIRFlickr-min	85.3 (0.6)	83.6 (0.7)	82.0 (0.6)	77.8 (0.9)	65.4 (0.8)	63.6 (0.6)
MIRFlickr-midi	83.2 (0.7)	80.7 (0.5)	77.3 (0.8)	71.8 (0.6)	59.6 (0.6)	59.0 (0.7)
MIRFlickr-max	76.6 (0.9)	73.9 (0.7)	68.6 (0.5)	65.2 (0.9)	55.8 (0.8)	55.6 (0.5)

Вплив розмірів зображень на точність їх стеганоаналізу. Стеганоаналітик не може бути впевненим, що jpeg-зображення до його можливого використання як стеганографічного контейнера не підлягало якійсь обробці, зокрема таким типовим операціям, як обрізування, масштабування, перестиснення тощо. Так як характеристичні вектори в обох моделях GFR та DCTR будуються з використанням згортки з блоками даних розміром 8×8, то на якомусь з етапів обробки, наприклад, при зсуві, обрізуванні, обертанні можливе виникнення десинхронізації блоків відносно вихідного зображення. Це приводить до критичних розбіжностей у значеннях елементів характеристичних векторів до та після обробки з десинхронізацією, при яких класифікатор, навчений тільки на контейнерах без обробки, не зможе правильно класифікувати десинхронізовані відносно них.

Зазначимо, що десинхронізація не відбувається, коли зображення зсувається на кількість пікселів кратну 8 по горизонталі та по вертикалі. Крім того в обох моделях передбачене злиття гістограм з індексами (a,b), $(a,8-b)$, $(8-a,b)$, $(8-a,8-b)$, а тому характеристичний вектор зображення, початок якого був зсунутий з точки (0,0) до точки (a,b) буде подібним до характеристичних векторів зображень, початок яких був зсунутий з (0,0) до $(a,8-b)$, $(8-a,b)$ чи $(8-a,\ 8-b)$.

При дослідженні впливу розмірів зображень на точність стеганоаналізу тестування здійснювалося на базі набору OWNSET95, який був сформований з власних фотоархівів. У порівнянні з MIRFlickr він містить більш подібні між собою зображення та відносно невелику кількість високотекстурованих контейнерів.

Щоб дослідити влив розмірів з OWNSET95 було створено ряд похідних наборів шляхом обрізування зображень від лівого краю до розмірів вказаних у верхньому рядку табл. 6. При цьому використовувалася програма IrfanView, десинхронізації вихідних блоків не відбувалося. Далі здійснювався стеганоаналіз OWNSET95 та похідних наборів з тими ж параметрами, що й раніше з MIRFlickr. Отримана для моделі GFR точність зазначена у другому рядку цієї ж таблиці, для DCTR – у третьому. Загалом із даних табл. 6 бачимо, що зі збільшенням розмірів зображень, точність їх стеганоаналізу зростає.

ТАБЛИЦЯ 6. Точність стеганоаналізу зображень різних розмірів

Розміри, пікс.	320×320	512×512	600×800	1000×1000	1200×1200	1200×1600
Модель GFR	64.7 (0.7)	68.6 (0.8)	72.5 (0.9)	77.5 (0.9)	80.4 (0.9)	83.8 (0.9)
Модель DCTR	66.0 (0.9)	73.8 (0.7)	79.1 (0.8)	85.2 (0.7)	87.8 (0.7)	88.0 (0.5)

Також зазначимо, що J-UNIWARD з 50 % корисним навантаженням стає дуже вразливим на зображеннях розмірами 1200×1600, що були перетиснуті з меншим коефіцієнтом якості. Так, точність класифікації зображень набору OWNSET75, що отриманий з OWNSET95 шляхом перестиснення зображень з коефіцієнтом 75, для моделі GFR склала 99.2 (0.2), для DCTR – 98.3 (0.3). Точність стеганоаналізу набору OWNSET55 (з коефіцієнтом якості 55) – 99.6 (0.1) та 99.4 (0.2), відповідно.

Вплив обрізування з десинхронізацією блоків зображень на точність стеганоаналізу. Вплив десинхронізації на точність детектування досліджувався додатково. Виявилося, що зображення, які перед використанням для стеганографічного приховування обрізалися зі зсувом, не кратним 8 пікселям, більш вразливі до стеганоаналізу. Так, точність виявлення зображень розмірами 320×320 для таких контейнерів зросла в середньому з 64.7 до 69.6 для моделі GFR та з 66.0 до 74.1 для DCTR. Вказана точність отримана для зображень зі зсувами (0,1), (1,0), (1,1), (1,2), (2,2), (3,3), (4,4), (0,4), (4,0). Модель DCTR сильніше за GFR реагувала на зсув, при чому найбільше на точність стеганоаналізу в її випадку вплинув зсув зображень на 4 пікселя по горизонталі.

Вплив черговості операцій обробки зображень на точність їх стеганоаналізу. Коли операцій обробки більше ніж одна, то для статистичного стеганоаналізу на базі машинного навчання їх черговість важлива. Так, наприклад, коли мова йде про операції обрізування та перестиснення з меншим коефіцієнтом, то точність стеганоаналізу зображень з комбінацією цих операцій як попередня обробка буде суттєво відрізнятися у залежності від їх черговості. Як приклад, приведемо результати стеганоаналізу для двох нових наборів, похідних від OWNSET95. Перший з них, OWNSET94-800 був отриманий шляхом стиснення зображень вихідного набору с коефіцієнтом якості 94 та подальшого обрізування до розмірів 800×800. Другий OWNSET800-94 – шляхом обрізування зображень набору OWNSET95 до розмірів 800×800 пікселів та подальшого стиснення обрізаних зображень з коефіцієнтом 94. Обрізування в обох випадках здійснювалося від центру. Точність стеганоаналізу для цих наборів зазначена в перших двох стовпцях табл. 7, в наступних двох стовпцях зазначені результати перехресних тестів, коли класифікатор навчався на зображеннях з одного набору, а класифікував зображення іншого.

ТАБЛИЦЯ 7. Точність стеганоаналізу зображень з обрізуванням та перестисненням

1. OWNSET94-800	2. OWNSET800-94	1. –> 2.	2. –> 1.
Модель GFR			
77.8 (0.8)	87.1 (0.6)	63.0 (0.8)	57.8 (0.5)
Модель DCTR			
85.7 (0.6)	93.2 (0.5)	71.0 (1.3)	53.5 (0.8)

Значне погіршення точності, отримане у перехресному тестуванні – це типовий прояв так званої проблеми невідповідності джерела контейнерів. Ця проблема виникає, коли в навчальному наборі немає контейнерів з такими ж параметрами як і ті, що класифікуються, або вони є, проте в недостатній кількості.

Висновки. Отримані в процесі дослідження чисельні оцінки показали, що підбір зображень-контейнерів грає важливу роль як для задач стеганографії, так і для задач стеганоаналізу.

З погляду стеганографії для організації прихованої передачі даних найкраще використовувати високотекстуровані зображення, стиснуті з високою якістю: вони мають перевагу над іншими як у своїй місткості (багато придатних для вкраплення ДКП коефіцієнтів), так і у стійкості до пасивного стеганоаналізу (найгірша точність класифікації). При чому краще обирати зображення відносно невеликих розмірів, оскільки великі зображення містять більше статистичної інформації, що використовується для стеганоаналізу. Також не слід обирати як контейнери, зображення невідомого походження, так як вони могли підлягати ряду операцій обробки, що не вплинули на візуальне сприйняття, проте змінити статистику, внаслідок чого контейнер став більш вразливим до стеганоаналітичних атак.

З погляду стеганоаналізу, отримані результати свідчать про залежність точності детектування від вмісту зображень, їх параметрів та можливої попередньої обробки. Найкраще виявляються приховані повідомлення у контейнерах, що містять значний відсоток гладких областей, найгірше – у високотекстурованих. Точність стеганоаналізу зростає зі зменшенням коефіцієнту якості зображення, а також зі збільшенням його розмірів. Точність зростає і коли при попередній обробці було порушено вихідну блокову структуру 8×8. Стеганоаналіз може здійснюватися двома шляхами: 1) з використанням направлених класифікаторів, що навчені на контейнерах, максимально наближених за всіма показниками, до тих, які необхідно дослідити, 2) з використанням загальних класифікаторів, що навчені на контейнерах, які охоплюють якнайбільший діапазон можливих показників зображень (при чому потрібно враховувати, що черговість можливої попередньої обробки

зображення впливає на результат). Для одного й того ж діапазону можливих показників зображень мінусом першого шляху є висока обчислювальна складність, мінусом другого – погіршення точності. Прийнятним компромісом може виявитися комбінація обох шляхів, проте дане припущення потребує більш глибокого дослідження.

Список літератури

1. Holotyak T., Fridrich J., Voloshynovskyy, S. Blind statistical steganalysis of additive steganography using wavelet higher order statistics. *Communications and Multimedia Security, 9th IFIP TC-6 TC-11 International Conference*. 2005. P. 273–274. https://doi.org/10.1007/11552055_31
2. Pevny T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on information Forensics and Security*. 2010. **5** (2). P. 215–224. https://doi.org/10.1145/1597817.1597831
3. Ker A. Steganalysis of LSB matching in grayscale images. *IEEE Signal Processing Letters.* 2005. **12** (6). P. 441– 444. https://doi.org/10.1109/LSP.2005.847889
4. Huang F., Shi Y.Q., Huang J. New JPEG steganographic scheme with high security performance. *International Workshop on Digital Watermarking*. 2010. 6526. P. 189–201. https://doi.org/10.1007/978-3-642-18405-5_16
5. Pevny T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on information Forensics and Security*. 2010. **5** (2). P. 215–224. https://doi.org/10.1109/TIFS.2010.2045842
6. Xia Z., Wang X., Sun X., Wang B. Steganalysis of least significant bit matching using multi-order differences. *Security and Communication Networks*. 2014. **7** (8). P.1283–1291. https://doi.org/10.1002/sec.864
7. Zeng J., Tan S., Li B., Huang J. Large-Scale JPEG Image Steganalysis Using Hybrid Deep-Learning Framework. *IEEE Transactions on Information Forensics and Security.* 2018. **13** (5). P. 1200–1214. https://doi.org/10.1109/TIFS.2017.2779446
8. Mustafa E.M., Elshafey M.A., Fouad M.M. Enhancing CNN-based Image Steganalysis on GPUs. *Journal of Information Hiding and Multimedia Signal Processing.* 2020. **11** (3). P. 138–150. https://www.researchgate.net/publication/344140896_Enhancing_CNN-based_Image_Steganalysis_on_GPUs
9. Boroumand M., Chen M., Fridrich J. Deep Residual Network for Steganalysis of Digital Images. *IEEE Transactions on Information Forensics and Security*. 2019. **14** (5). P. 1181–1193. https://doi.org/10.1109/TIFS.2018.2871749
10. Kodovsky J., Fridrich J. Steganalysis in high dimensions: fusing classifiers built on random subspaces. *Proc. SPIE, Electronic Imaging, Media, Watermarking, Security and Forensics XIII*. 2011. **7880** (78800L). https://doi.org/10.1117/12.872279
11. Fridrich J., Kodovsky J. Rich Models for Steganalysis of Digital Images. *IEEE Transactions on Information Forensics and Security*. 2012. **7** (3). P. 868–882. https://doi.org/10.1109/TIFS.2012.2190402
12. Holub V., Fridrich J. Random Projections of Residuals for Digital Image Steganalysis. *IEEE Transactions on Information Forensics and Security*. 2013. **8** (12). P.1996–2006. https://doi.org/10.1109/TIFS.2013.2286682
13. Holub V., Fridrich J. Phase-Aware Projection Model for Steganalysis of JPEG Images. *Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XVII*. 2015. 9409. https://doi.org/10.1117/12.2075239
14. Li W., Zhou W., Zhang W., Qin C., Hu H., Yu N. Shortening the Cover for Fast JPEG Steganography. *IEEE Transactions on Circuits and Systems for Video Technology*. 2020. **30** (6). P. 1745–1757. https://doi.org/10.1109/TCSVT.2019.2908689
15. Progonov D. Statistical Steganalysis of Multistage Embedding Methods. *International Journal "Information Models & Analyses"*. 2016. **5** (1). P. 23 –36.
16. Yang Y., Kong X., Wang B., Ren K., Guo Y. Steganalysis on Internet images via domain adaptive classifier *Neurocomputing*. 2019. 351. P. 205 –216. https://doi.org/10.1016/J.NEUCOM.2019.04.025
17. Корольов В.Ю., Поліновський В.В., Герасименко В.А., Горинштейн М.Л. Дослідження кольорових цифрових фотографій методами RS-стегоаналізу та статистики. *Інформація і право*. 2011. **3** (3). С. 102 – 110. http://dspace.nbuv.gov.ua/handle/123456789/39035
18. Kharrazi M., Sencar H.T., Memon N.D. Performance study of common image steganography and steganalysis techniques. *Journal of Electronic Imaging*. 2006. **15** (4). P. 041104-1–16. https://doi.org/10.1117/1.2400672
19. Holub V., Fridrich J., Denemark T. Universal Distortion Function for Steganography in an Arbitrary Domain. *EURASIP Journal on Information Security*. 2014. 1. P. 1–13. https://doi.org/10.1186/1687-417X-2014-1
20. Kodovský J., Fridrich J., Holub V. Ensemble Classifiers for Steganalysis of Digital Media. *IEEE Transactions on Information Forensics and Security*. 2012. **7** (2). P. 432 – 444. https://doi.org/10.1109/TIFS.2011.2175919

21. Кошкіна Н.В. Дослідження основних компонентів систем JPEG-стеганоаналізу на базі машинного навчання. *Захист інформації*. 2020. **22** (2). С. 97–108. http://193.178.34.32/index.php/ZI/article/view/14801/21490
22. Song X., Liu F., Yang C., Luo X., Zhang Y. Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters. *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security. ACM*. 2015. P. 15–23. https://doi.org/10.1145/2756601.2756608
23. Holub V., Fridrich J. Low Complexity Features for JPEG Steganalysis Using Undecimated DCT. *IEEE Transactions on Information Forensics and Security*. 2015. **10** (2). P. 219 – 228. https://doi.org/10.1109/TIFS.2014.2364918
24. Koshkina N.V. Comparison of Efficiency of Statistical Models Used for Formation of Feature Vectors by JPEG Images Steganalysis. *Theoretical and Applied Cybersecurity*. **2** (2). P. 22–28. https://doi.org/10.20535/tacs.2664-29132020.1.209433

Одержано 04.03.2021

Кошкіна Наталія Василівна,
докторка технічних наук, старша наукова співробітниця
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.
https://orcid.org/0000-0001-5180-2255
nata.koshkina@gmail.com

MSC 68M25, 68P27

N. Koshkina

About JPEG Images Parameters Impact to Steganalys Accuracy

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv
Correspondence: nata.koshkina@gmai.com

Introduction. Existing examples of illegal use of computer steganography prove the need for the development of steganalytical methods and systems as one of the most important areas of cybersecurity. The advantage of machine learning-based steganalytical methods is their versatility: they do not rely on knowledge of the injection algorithm and can be used to detect a wide range of steganographic methods. However, before being used for detecting steganocontainers, the methods mentioned require training on containers that are determined for sure whether they contain hidden messages or not. On this stage, it is very important to understand how the parameters of containers under investigation, in particular, such a common variant as JPEG images, affect the accuracy of steganalysis. After all, the inconsistency of the source of containers is an open problem of steganalysis leading to significant decrease of accuracy of detecting hidden messages after the classifier is moved from the laboratory to the real world.

The purpose of the work is investigation of influence of the content, size and quality factor of JPEG images to the accuracy of their steganalysis performed by statistical methods based on machine learning.

Results. During the research the following patterns were revealed: 1) the accuracy is better when images with a close percentage of coefficients suitable for DCT concealment are used for training and control, 2) images are classified more accurately when they have a relatively small number of suitable DCT coefficients, 3) with using mixed training samples (by content or parameters) the accuracy of steganalysis deteriorates, 4) decreasing quality factor of JPEG-images leads to increasing the accuracy of their steganalysis, 5) increasing size of images increases the accuracy of their steganalysis, 6) images where desynchronization of blocks took place during preprocessing are classified more accurately, 7) the sequence of the image preprocessing operations affects the accuracy of its steganoanalysis.

Conclusions. For steganography tasks – the choice of JPEG containers, taking into account revealed patterns, makes steganographic hides more resistant to passive attacks. Considering them for tasks of steganalysis allows one to interpret the obtained results more accurately.

Keywords: information security, steganography, steganalysis, intelligent computer systems, machine learning, detection accuracy.