Интеллектуальное обнаружение стеганографического преобразования изображений, основанное на классификации контейнеров

И.В. Аникин, А.В. Ягина

Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, Казань

Аннотация: Исследована возможность обнаружения стеганографического преобразования цифровых изображений, основанного на классификации контейнеров**.** Полученные результаты демонстрируют эффективность применения для решения данной задачи глубоких нейронных сетей. Применение метода LSB может быть обнаружено с помощью нейросетевого классификатора, построенного на архитектуре EfficientNet b3. Достигаемая точность классификации при этом – выше 97%. Применение более сложных, частотных методов стеганографии, может быть эффективно обнаружено при классификации их представления в виде цифровой модели YCrBr, с аугментацией в виде вертикального и горизонтального поворотов. Достигаемая при этом точность классификации с помощью EfficientNet b3 – выше 77%.

Ключевые слова: стеганография, стегоконтейнер, машинное обучение, классификация, цифровое изображение, глубокое обучение, сверточная нейронная сеть, EfficientNet b3, конфиденциальность, защита информации.

Введение

Стеганография активно используется для защиты конфиденциальности информации, передаваемой по каналам связи [1]. По сравнению с методами криптографии, преобразующими сообщения в закрытую форму и требующими наличие секретного ключа для осуществления обратного преобразования [2], методы стеганографии скрывают сам факт передачи сообщения, встраивая его в стеганографический контейнер (изображение, видео- либо аудиофайл, текстовый документ и т.д.), исключая возможность восприятия человеком факта его изменения. При этом стеганографические контейнеры не привлекают внимания при передаче по открытому каналу [3]. Довольно часто методы стеганографии применяются дополнительно к криптографическим методам, тем самым усиливая защиту путем скрытия факта передачи шифртекста.

Одним из основных требований к стеганографическим методам является такая модификация контейнера, которую сложно или невозможно

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

обнаружить существующими методами стегоанализа, применение которых направлено на обнаружение заполненных стеганографических контейнеров, а в некоторых случаях – на извлечение из них полезной нагрузки. В связи с этим, актуальным вопросом является исследование существующих методов стеганографии на предмет устойчивости к подобным атакам на стегоконтейнер. В данной статье проведено такое исследование некоторых алгоритмов стеганографии для цифровых изображений.

Базовые сведения

Общая схема организации стеганографической системы при использовании цифрового изображения в качестве стегоконтейнера представлена на рис. 1.

Рисунок 1. – Общая схема организации стеганографической системы

Для стегоконтейнеров в виде цифровых изображений различают пространственные и частотные методы стеганографии [4]. В первом случае для внедрения сообщений используются сами пиксели изображения. Во втором случае выполняется преобразование изображения в частотную область, как правило, с помощью дискретного косинусного либо дискретного вейвлет-преобразования. Далее полезная нагрузка внедряется путем работы с коэффициентами в частотной области.

Параллельно с методами стеганографии, в последнее время активно развиваются методы стегоанализа цифровых изображений [5,6]. Их

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

применение нацелено на обнаружение заполненных стеганографических контейнеров при передаче по открытому каналу, а в некоторых случаях – на извлечение из них полезной нагрузки. Ряд таких методов основан на исследовании структурных характеристик изображений для обнаружения аномалий, вносимых стеганографическим преобразованием. Другие методы основаны определении ключевых признаков, моделирующих стегоконтейнер и дальнейшем выявлении их изменения. При решении задачи обнаружения стеганографического преобразования цифровых изображений, активно используются различные модели классификации контейнеров, основанные на методах машинного обучения: KNN, SVM, Random Forest, Naïve Bayes и другие [7-9]. Кроме этого, бурное развитие технологий глубокого обучения, актуализировало применение глубоких сверточных нейронных сетей для решения задачи классификации стеганографических контейнеров [10]. В данной статье исследуются возможности их применения для решения данной задачи.

Методология исследовании

Реализованная методология исследований возможности обнаружения стеганографического преобразования изображений, основанного на классификации контейнеров, представлена на рис. 2.

В качестве основы исследуемого датасета цифровых изображений был выбран ALASKA2 Image Steganalysis [11]. Указанный набор данных включает в себя пустые контейнеры, а также заполненные контейнеры с применением стеганографических алгоритмов JMiPOD, JUNIWARD либо UERD. Данные классы контейнеров являются выравненными, каждый из них включает по 75000 изображений. Размер изображений - 512×512 пикселов. Также из пустых контейнеров датасета ALASKA2 Image Steganalysis, был сформирован дополнительный класс заполненных контейнеров с помощью

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

метода LSB. Для заполнения контейнеров использовалась библиотека Stegano.

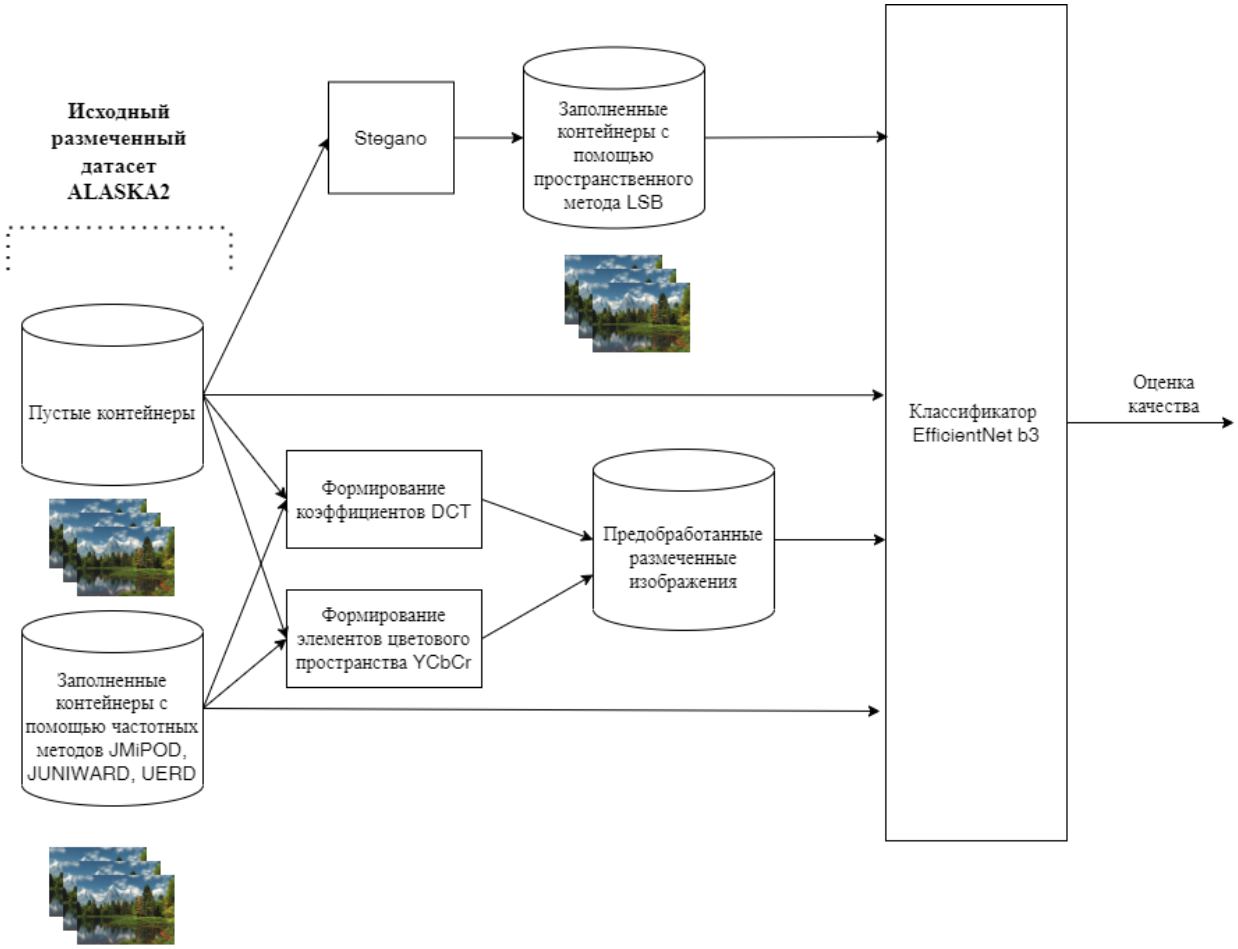

Рисунок 2 – Методология исследований

Для классификации стеганографических контейнеров использовалась нейросетевая архитектура EfficientNet b3 (рис. 3), обладающая высокой точностью и сравнительно небольшим временем обучения модели [12]. EfficientNet является моделью, основанной на комбинировании сверточных фильтров с различными размерами ядер, что обеспечивает более эффективное использование ресурсов памяти и вычислительной мощности.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

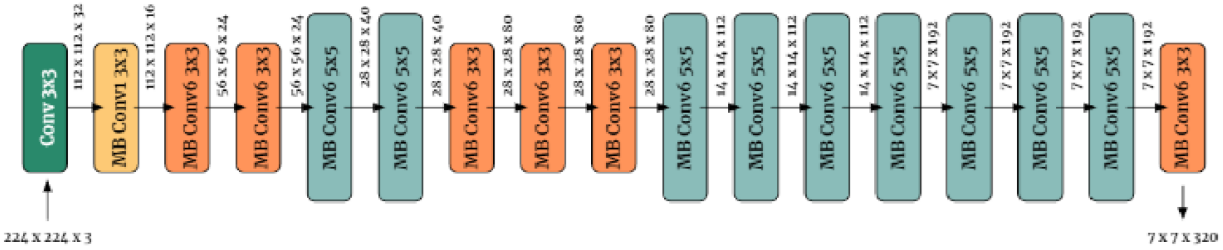

Рисунок 3 – Архитектура нейросетевой модели EfficientNet

Для классификации стеганографических контейнеров использовалось два подхода:

- непосредственная классификация контейнеров как цифровых изображений;

- классификация предобработанных изображений с формированием коэффициентов дискретного косинусного преобразования (DCT) [13] и элементов цветового пространства YCbCr (яркость, цветность синего, цветность красного). Учет элементов цифрового пространства и коэффициентов DCT позволяет учесть специфику стеганографических методов, выполняющих преобразование в частотной области (не LSB). Для формирования коэффициентов DCT использовалась библиотека JpegIO.

Эксперименты

На первом этапе исследован потенциал EfficientNet b3 для решения задачи обнаружения заполненных контейнеров при их непосредственной классификации как цифровых изображений. Использовался вариант EfficientNet b3, предобученный на ImageNet по методу Noisy Student.

Подготовленный датасет был разделен на обучающую выборку (80%) и тестовую (20%). Полученные результаты представлены в таблице 1.

Таблица № 1

Результаты классификации стеганографических контейнеров в виде цифровых изображений

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

Тип стеганографического преобразования	Точность (training) %	Точность (testing) %
Пространственный метод стеганографии (LSB)	98,5	97,1
Частотные методы стеганографии (JMiPOD, JUNIWARD, UERD)	63,05	62,1

Полученные результаты показывают, что архитектура EfficientNet b3 хорошо справляется с задачей классификации стеганографических контейнеров для пространственного метода LSB, разделяя пустые и заполненные контейнеры. При этом, на вход нейронной сети подается исходное изображение. Для частотных же методов стеганографии получен неудовлетворительный результат, что связано с иным способом встраивания сообщений.

На втором этапе исследован потенциал EfficientNet b3 для решения задачи классификации стеганографических контейнеров, представленных коэффициентами DCT и элементами цветового пространства YCbCr. Для оценки качества использовалась метрика Average Precision (AP). Полученные результаты представлены в таблице 2.

Таблица № 2

Результаты классификации стеганографических контейнеров в виде коэффициентов DCT и YCbCr

Представление изображения	AP (training) %	AP (testing) %
Коэффициентами DCT	50,6	50,4
Цветовой моделью YCbCr	76,7	74,2

Полученные результаты показывают, что коэффициенты DCT не позволяют «уловить» факт стеганографического преобразования в частотной области, в то время как цветовая модель YCbCr позволяет это сделать более эффективно.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

В выполненных ранее экспериментах не применялась аугментация изображений, так как она предполагает внесение дополнительных искажений в данные, что может частично уничтожить признаки стеганографии. Однако было решено применить аугментацию на тренировочном наборе данных в модели YCrBr в виде вертикального и горизонтального поворотов. Результаты классификации стеганографических контейнеров после аугментации данных в модели YCrBr представлены в таблице 3.

Таблица № 3

Результаты классификации стеганографических контейнеров после аугментации данных в модели YCrBr

Представление изображения	AP (training) %	AP (testing) %
Цветовой моделью YCbCr	79,9	77,1

Полученные результаты показали увеличение качества классификации контейнеров по сравнению с цветовой моделью YCbCr без аугментации. Полученное качество является приемлемым при решении задачи обнаружения стеганографического преобразования контейнеров.

Выводы

Результаты выполненных исследований демонстрируют эффективность применения глубоких нейронных сетей для обнаружения стеганографического преобразования цифровых изображений.

Применение метода LSB, относящегося к пространственной стеганографии может быть эффективно обнаружено с помощью нейросетевого классификатора, построенного на архитектуре EfficientNet b3. Достигаемая точность классификации при этом – выше 97%.

Применение более сложных, частотных методов стеганографии (JMiPOD, JUNIWARD, UERD), может быть эффективно обнаружено при классификации их представления в виде цифровой модели YCrBr, с

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

аугментацией в виде вертикального и горизонтального поворотов. Достигаемая при этом точность классификации с помощью EfficientNet b3– выше 77%, что является приемлемым для решения поставленной задачи.

Применение предложенного метода стегоанализа имеет высокую практическую ценность и позволяет решать следующие задачи:

- оценивать эффективность существующих и новых методов стеганографии, в том числе их способности скрывать факт наполнения контейнеров полезной нагрузкой;

- выявлять факт утечки конфиденциальной информации из организации за счет внедрения передаваемых сообщений в цифровые изображения, не привлекающие внимания;

- выявлять вредоносное программное обеспечение, осуществляющее проникновение в корпоративные сети организаций за счет внедрения в цифровые изображения.

Литература

1. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. М.: СОЛОН-ПРЕСС, 2009. 273 с.

2. Аникин И.В., Альнаджар Х.Х. Подход к построению потоковых шифров с применением генератора псевдослучайных последовательностей, основанного на нечеткой логике // Инженерный вестник Дона, 2023, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2023/8455.

3. Гибадуллин Р.Ф., Вершинин И.С., Глебов Е.Е. Разработка приложения для ассоциативной защиты файлов // Инженерный вестник Дона, 2023, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2023/8462.

4. Verdiyev S.G., Naghiyeva A.F. A brief overview of data hiding methods in digital images // Инфокоммуникационные технологии. 2020. № 4 (18). С. 427-437.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

5. Вильховский Д.Э. Обзор методов стеганографического анализа изображений в работах зарубежных авторов // Математические структуры и моделирование. 2020. № 4 (56). С. 75–102.

6. Грачев Я.Л., Сидоренко В.Г. Стегоанализ методов скрытия информации в графических контейнерах // Надежность. 2021. № 3 (21). С. 39-46.

7. Shankar D.D., Azhakath A.S. Minor blind feature based Steganalysis for calibrated JPEG images with cross validation and classification using SVM and SVM-PSO // Multimedia Tools and Applications. 2021. 80. pp. 4073-4092.

8. Евсютин О.О., Мещеряков Р.В., Шумская О.О. Стегоанализ цифровых изображений с использованием наивного байесовского классификатора // Вестник компьютерных и информационных технологий. 2018. № 10. С. 11-21.

9. Bhat Veena H., Krishna S., Deepa Shenoy P. SURF: Steganalysis using random forests // Proceedings of 10th International Conference on Intelligent Systems Design and Applications. 2010. pp. 373-378.

10. Chaumont M. Digital Media Steganography. Academic Press, 2020. pp. 321–349.

11. Cogranne R., Giboulot Q., Bas P. The ALASKA steganalysis challenge: A first step towards steganalysis // Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. 2019. pp. 125–137.

12. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // URL arxiv.org/pdf/1905.11946.pdf.

13. Ахмед Н., Рао К. Р. Ортогональные преобразования при обработке цифровых сигналов. М.: Связь, 1980. 248 с.

References

1. Gribunin V.G., Okov I.N., Turintsev I.V. Tsifrovaya steganografiya [Digital steganography]. M. SOLON-PRESS, 2009. 273 p.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

2. Anikin I.V., Al'nadzhar Kh.Kh. Inzhenernyj vestnik Dona, 2023, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2023/8455.

3. Gibadullin R.F., Vershinin I.S., Glebov E.E. Inzhenernyj vestnik Dona, 2023, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2023/8462.

4. Verdiyev S.G., Naghiyeva A.F. Infokommunikatsionnyye tekhnologii. 2020. № 4 (18). pp. 427-437.

5. Vil′khovskiy D.E. Matematicheskiye struktury i modelirovaniye. 2020. № 4 (56). pp. 75–102.

6. Grachev YA.L., Sidorenko V.G. Nadezhnost′. 2021. № 3 (21). pp. 39-46.

7. Shankar D.D., Azhakath A.S. Multimedia Tools and Applications. 2021. 80. pp. 4073-4092.

8. Yevsyutin O.O., Meshcheryakov R.V., Shumskaya O.O. Vestnik komp′yuternykh i informatsionnykh tekhnologiy. 2018. № 10. pp. 11-21.

9. Bhat Veena H., Krishna S., Deepa Shenoy P. Proceedings of 10th International Conference on Intelligent Systems Design and Applications. 2010. pp. 373-378.

10. Chaumont M. Digital Media Steganography. Academic Press, 2020. pp. 321–349.

11. Cogranne R., Giboulot Q., Bas P. Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. 2019. pp. 125–137.

12. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks URL arxiv.org/pdf/1905.11946.pdf.

13. Akhmed N., Rao K. R. Ortogonal′nyye preobrazovaniya pri obrabotke tsifrovykh signalov [Orthogonal transforms in digital signal processing]. M.: Svyaz′, 1980. 248 c.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023