УДК 004.056.5

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ИСПОЛЬЗОВАНИЯ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ В СТЕГОСИСТЕМАХ ДЛЯ СИНТЕЗА КОНТЕЙНЕРОВ

студентка 4 курса, 5КБ группы Борисюк Д. С.
Научный руководитель - канд. техн. наук Садов В. С.
Белорусский государственный университет
Минск, Беларусь

В последние годы искусственный интеллект и машинное обучение нашли широкое применение в различных областях, включая стеганографию - науку о скрытой передаче информации. Стегосистемы являются одним из инструментов защиты конфиденциальной информации, позволяя ее скрыть в неподозрительном контейнере, например, изображении. В последнее время генеративно-состязательные сети (GANs) привлекают все большее внимание исследователей стеганографии благодаря их возможностям в синтезе изображений и подделке контента. В данной статье мы исследуем потенциал использования GANs в стегосистемах для синтеза контейнеров и обсуждаем их преимущества и недостатки в этой области.

Структурная схема стегосистемы представлена на риснке 1.

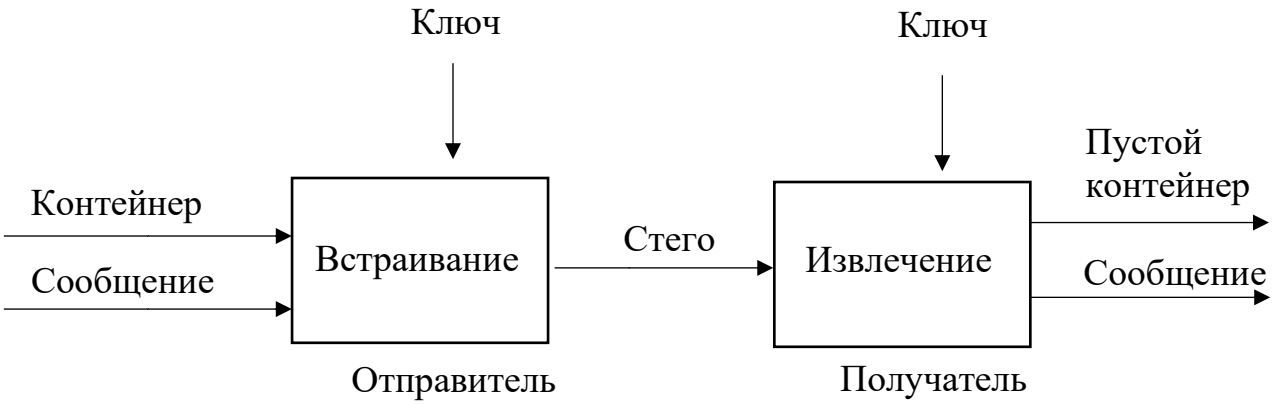

Рисунок 1. Структурная схема стегосистемы

Для создания безопасных стеганографических систем, использующих синтез изображений, необходимо создать новый образ, в котором будет

содержаться секретная информация. Одним из ключевых критериев успеха такой системы является достаточная реалистичность созданного изображения, чтобы оно было неразличимо от обычного. В настоящее время в стеганографии существуют два типа методов, основанных на синтезе изображений.

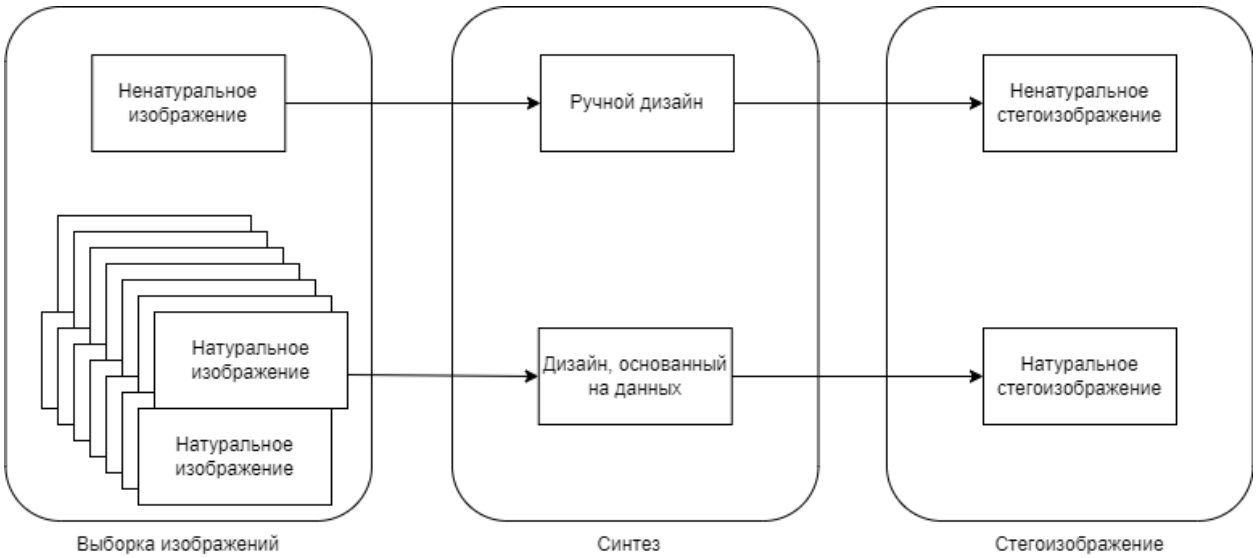

Рисунок 2. Создание стегоизображения методом синтеза

Из-за сложности задачи синтеза реалистичных изображений традиционные методы синтеза стегоизображений использовали различные подходы для решения задачи стеганографии, включая:

создание неестественных изображений, текстурных изображений [1/22] и изображений отпечатков пальцев [2/23];

обратимый синтез текстурных изображений для скрытия секретных сообщений;

сокрытие информации в процессе синтеза текстурных изображений;

использование текстур, основанных на деформации, для сокрытия информации.

Такие методы стеганографии, основанные на синтезе текстур, предполагают, что контейнером может быть изображение без семантической информации, что ограничивает их применение в более широких областях стеганографии.

Метод генерации стегоизображений с использованием генеративно-состязательной сети (GAN) позволяет создавать реалистичные естественные изображения, которые содержат скрытые сообщения. Обучение с учителем

включает в себя три участника: Алису, Боба и Еву, где Алиса встраивает секретное сообщение в изображение контейнер, создавая стеганографическое изображение, а Боб может восстановить сообщение. Ева же пытается определить, является ли передаваемое изображение контейнером или стеганографическим изображением. На начальных этапах обучения Ева может легко отделить изображения контейнера от стеганографических изображений, но по мере продолжения тренировок Ева начала лучше справляться со своей задачей, что заставляет Алису улучшать способы встраивания сообщений [3/24].

К селекционным методам также относится стеганография без встраивания. Секретные сообщения преобразуются в вектор шума, который затем передается генератору для создания стегоизображения. Сначала генератор обучается на наборе данных, чтобы создавать реалистичные изображения. Затем, на втором этапе, экстрактор обучается функции потерь, чтобы извлекать сообщение из стегоизображения. Цель этого этапа заключается в том, чтобы восстановить сообщение из созданного стегоизображения.

На последнем этапе отправитель устанавливает связь между вектором шума и секретным сообщением, а затем сегментирует их для создания отображения. Получатель может использовать экстрактор для восстановления вектора шума и затем получить секретное сообщение с помощью полученного отображения. Этот процесс показан на рисунке 3 [4/25]:

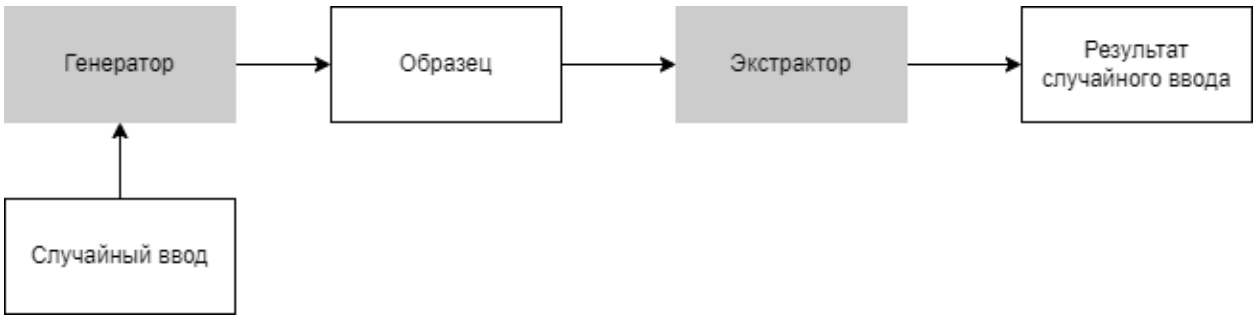

Рисунок 3. Тренировка экстрактора

Также к данной группе методов относится WGAN-GP. Это метод генеративной состязательной сети для создания стегоизображений высокого качества, который использует функцию потерь Вассерштейна и добавляет штраф за градиент. В отличие от других методов стеганографии, экстрактор и генератор обучаются одновременно. Генератор обучается в минимакс игре с дискриминатором и экстрактором, чтобы улучшить качество изображений.

Штраф за градиент вычисляется на основе расстояния между величиной градиента и идеальной нормой, равной 1 [5/26].

Таким образом, метод WGAN-GP представляет собой более эффективный и универсальный подход к генерации стегоизображений на основе генеративно-состязательных сетей. Одним из главных преимуществ WGAN-GP является то, что он решает проблему взрывающегося градиента, что может приводить к затуханию градиента в стандартных моделях GAN.

Литература

1. Deterministic texture analysis and synthesis using tree structure vector quantization / Li-Yi Wei // Stanford, CA, USA, 1999
2. Fingerprint image synthesis based on statistical feature models / Qijun Zhao, Anil K. Jain, Nicholas G. Paulter, Melissa Taylor // Arlington, VA, USA, 2012
3. Generating steganographic images via adversarial training // Jamie Hayes and George Danezis / 2017
4. A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks // Donghui Hu; Liang Wang; Wenjie Jiang; Shuli Zheng; Bin Li / School of Computer and Information, Hefei University of Technology, Hefei, China, 2018
5. A coverless steganography method based on generative adversarial network // Xintao Duan, Baoxia Li, Daidou Guo, Zhen Zhang & Yuanyuan Ma / 2020