МАТЕМАТИЧЕСКИЕ МЕТОДЫ СТЕГАНОГРАФИИ

УДК 519.85 DOI 10.17223/20710410/55/3

СТЕГОАНАЛИЗ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ И ГЕТЕРОАССОЦИАТИВНЫХ ИНТЕГРАЛЬНЫХ ПРЕОБРАЗОВАНИЙ

М. А. Дрюченко, А. А. Сирота

Воронежский государственный университет, г. Воронеж, Россия

E-mail: m_dryuchenko@mail.ru, sir@cs.vsu.ru

Рассматривается задача стегоанализа цифровых изображений. Предложен подход, предусматривающий разбиение входного изображения большого размера на небольшие неперекрывающиеся блоки и проведение последовательного стегоанализа этих блоков с помощью относительно простых свёрточных сетей с модифицированной архитектурой, включающей слои специальной обработки. После этого в ходе вторичной постобработки проводится объединение получаемой совокупности результатов классификации блоков как последовательности бинарных ответов по схеме наивного байесовского классификатора, в том числе, при необходимости, с учётом потенцильной заполняемости блоков. В качестве средств дополнительной специальной обработки изображений в свёрточных сетях предлагается использовать так называемые интегральные гетероассоциативные преобразования, обеспечивающие выделение на обрабатываемом блоке изображения оценочной и стохастической (маскирующей) составляющих на основе модели прогноза одной части блока по отношению к другой и направленные на выявление нарушения структурных и статистических свойств изображений после внедрения стегосообщения. Такие преобразования включаются в архитектуру обучаемых нейронных сетей в качестве дополнительного слоя. Рассмотрены альтернативные варианты архитектуры используемых глубоких нейронных сетей как с использованием слоя интегрального гетероассоциативного преобразования, так и без него. Исследования проведены для базы цветных изображений PPG-LIRMM-COLOR base и нескольких алгоритмов стегоскрытия, включая классические блочные и блочно-спектральные алгоритмы Куттера, Коха — Жао, более современные алгоритмы EMD, MBEP и алгоритмы адаптивной пространственной стеганографии WOW и S-UNIWARD, обладающие высокой степенью скрытности для стегоанализа. Рассмотрены также разработанные авторами алгоритмы стегоскрытия данных, основанные на использовании гетероассоциативных сжимающих преобразований. Показано, что получаемая при реализации предложенных схем обработки информации точность стегоанализа для изображений большого размера при достаточно скромных вычислительных затратах сопоставима с результатами, полученными другими авторами, а в некоторых случаях и превосходит их.

Ключевые слова: *стеганография, стегоанализ, стегосообщение, цифровые изображения, машинное обучение, глубокие нейронные сети.*

IMAGE STEGOANALYSIS USING DEEP NEURAL NETWORKS AND HETEROASSOCIATIVE INTEGRAL TRANSFORMATIONS

M. A. Dryuchenko, A. A. Sirota

Voronezh State University, Voronezh, Russia

The problem of steganalysis of digital images is considered. The proposed approach is based on the use of deep convolutional neural networks with a relatively simple architecture, distinguished by the use of additional layers of special processing. These networks are trained and used for steganalysis of small fragments of the original large images. For the analysis of full sized images, it is proposed to carry out secondary post-processing, which involves combining the obtained classification results in blocks as a sequence of binary features according to the scheme of a naive Bayesian classifier. We propose to use integral heteroassociative transformations that provide the selection of the estimated and stochastic (masking) components on the processed image fragment based on the forecast model of one part of the fragment in relation to another to identify violations of the structural and statistical image properties after message embedding. Such transformations are included in the architecture of trained neural networks as an additional layer. Alternative versions of deep neural network architectures (with and without an integral layer of heteroassociative transformation) are considered. The PPG-LIRMM-COLOR images base was used to create data sets. Experiments have been carried out for several well-known stego algorithms (including the classic block and block-spectral algorithms of Kutter, Koha — Zhao, modern algorithms EMD, MBEP and algorithms for adaptive spatial steganography WOW and S-UNIWARD) and for the stego algorithms based on the use of heteroassociative compression transformations. It is shown that the accuracy of steganalysis obtained when implementing the proposed information processing schemes for large images with relatively low computational costs is comparable to the results obtained by other authors, and in some cases even exceeds them.

Keywords: *steganography, steganalysis, machine learning, deep neural networks.*

Введение

Задача стегоанализа (СА) [1] состоит в обнаружении факта внедрения визуально незаметного стегосообщения или цифрового водяного знака (ЦВЗ) в объект цифрового контента (изображение, видео, звуковой сигнал и т. п.) и оценке параметров внедрённого сообщения. Обычно в качестве модели стеганографически скрытой информации (ССИ) рассматривается псевдослучайная двоичная последовательность. Задача СА может решаться в прямой постановке как задача анализа контейнера с неизвестным содержимым, так и как обратная задача оценки скрытности алгоритмов компьютерной стеганографии.

В последние 10–15 лет развитие стегоанализа идёт по пути использования методов и алгоритмов машинного обучения как универсального и эффективного подхода к решению любых задач анализа данных. Разнообразие применяемых решений очень велико, что требует систематизированного анализа известных результатов по отношению к вновь предлагаемым решениям с целью определения перспективных направлений исследований и разработок. Одним из таких направлений является использование глубоких нейронных сетей. Следует отметить, что основные продвижения здесь связаны с отходом от традиционной архитектуры свёрточной сети и использованием слоёв

специальной обработки, нетрадиционных функций активации, более сложных архитектур.

Целью данной работы является исследование алгоритмов стегоанализа цифровых изображений, базирующихся на анализе совокупности относительно небольших фрагментов (блоков) на изображении большого размера с помощью глубоких нейронных сетей свёрточного типа, имеющих дополнительные слои для выполнения интегральных гетероассоциативных преобразований, и реализации вторичной постобработки результатов классификации в этих блоках для принятия окончательного решения по всему изображению.

1. Анализ результатов предшествующих работ

Суть подхода, основанного на применении метода машинного обучения, состоит в построении классификаторов объектов-контейнеров для обнаружения факта наличия ССИ на основе реализации процедур обучения по представительным наборам примеров, содержащих заполненные и незаполненные контейнеры. Отличительной особенностью подхода является, прежде всего, его универсальность. Методы и алгоритмы машинного обучения в приложении к задаче стегоанализа можно разделить на две большие группы: классические «неглубокие» методы и алгоритмы машинного обучения; методы и алгоритмы, основанные на применении глубоких нейронных сетей. В дальнейшем мы остановимся на представлении авторских результатов именно в этой области.

1.1. Применение неглубоких методов машинного обучения

К неглубоким алгоритмам машинного обучения относятся: наивный байесовский классификатор; метод опорных векторов; композиционные алгоритмы на основе бэггинга («случайный лес») и бустинга и ряд других. Характерной особенностью этих алгоритмов (в отличие от алгоритмов глубокого обучения) является необходимость предварительной обработки анализируемых объектов для извлечения информативных признаков, используемых при обучении классификаторов. Обзор публикаций, иллюстрирующий широту перечня применяемых методов и алгоритмов стегонализа, включая и указанные методы машинного обучения, приведён, например, в [2]. Одними из первых были работы [3, 4]. Предложенный авторами подход заключается в применении метода опорных векторов. В качестве набора признаков используется вектор размерности 72, вычисляемый из оценок статистических характеристик распределения групп пикселей изображения: математическое ожидание, дисперсия, среднеквадратичное отклонение и т. д. Обучение проводилось по выборке из 1800 пустых контейнеров и случайного подмножества из 1800 заполненных контейнеров, при этом полученная точность классификации составила более 95 % для относительно простых алгоритмов внедрения сообщений.

В развитие этого подхода для решения задачи стегоанализа в последующем были разработаны специальные многомерные системы (пространства) признаков: SPAM (Subtractive Pixel Adjacency Matrix) [5]; SRM (Spatial Rich Model) [6]; PSRM (Projection Spatial Rich Model) [7]. Эффективность различных алгоритмов обработки информации для обнаружения скрытых сообщений по таким признакам обычно демонстрировалась на чёрно-белых изображениях размера 512×512 из базы BOSSbase 1.01 [8]. Для цветных изображений обычно используется база 512×512 PPG-LIRMM-COLOR base [9].

Проверка и сравнение алгоритмов стегоанализа и систем признаков проводилась путём создания обучающих и тестовых примеров на основе наиболее скрытных алгоритмов встраивания ЦВЗ. Для контейнеров-изображений в этой постановке скрывае-

мая информация обычно внедряется при помощи методов адаптивной стеганографии HUGO, S-UNIWARD и WOW, которые считаются наиболее трудно обнаружимыми. Именно по ним приведены лучшие из известных результатов в области стегоанализа. Особенностью этих алгоритмов является то, что они осуществляют встраивание в те области контейнера, которые в минимальной степени искажают стастистические характеристики изображения [7, 10 – 12]. Идея адаптивного внедрения заключается в том, что позиции для внедрения выбираются исходя из свойств изображения; при этом с большей вероятностью внедрение осуществляется в те области, где обнаружить информацию труднее всего (обычно это наиболее «зашумлённые» области).

В ходе сравнительного анализа различных алгоритмов установлено, что наибольшей эффективностью обладают алгоритмы, основанные на использовании метода опорных векторов и ансамблевые (композиционные) алгоритмы. Результаты [7, 13, 14] показали, что в зависимости от объёма полезной нагрузки (payload, pl), измеряемой в среднем количестве бит внедряемого стегосообщения, приходящихся на пиксель контейнера (bit per pixel, bpp), точность стегонализа изменяется в пределах 62–94 %, что говорит о достаточно высоких показателях классификации пустых и заполненных контейнеров. Одним из эффективных приёмов для стегоанализа цифровых изображений является использование алгоритмов сжатия в различных постановках. В работе [14] предложен и использован интегральный классификатор на основе сжатия данных, состоящий из набора отдельных классификаторов, каждый из которых обрабатывает только те контейнеры, которые предварительно автоматически отфильтрованы для него. Принципиальное отличие этого подхода от других, использующих сжатие данных, заключается в том, что здесь сжатие используется на предварительном этапе выбора классификатора, но не для построения самого алгоритма стегоанализа.

1.2. Применение глубоких нейронных сетей

При использовании методов глубокого обучения бинарный классификатор для выявления факта скрытия данных (стегосообщения, ЦВЗ) задаётся в виде глубокой нейронной сети. В подавляющем большинстве работ рассматриваются свёрточные нейронные сети (CNN) в различных модификациях. Информативный англоязычный обзор на эту тему представлен в [15].

Одной из первых работ в этом направлении является [16]. В ней авторы предложили специализированную архитектуру свёрточной нейронной сети, которую они назвали CNN model called Gaussian-Neuron (GNCNN). Её особенностью является использование пространственного фильтра высоких частот с фиксированным ядром, специальных функций активации в виде гауссианы и слоёв субдискретизации с усреднением в пределах окна пуллинга. Использовано пять свёрточных слоёв для извлечения признаков и четыре полносвязных слоя для выполнения классификации. Главная идея такой обработки состоит в том, что пространственный высокочастотный фильтр локализует малые искажения в областях исходного контейнера, связанные с внедрением ССИ. Использование гауссовских активаций обеспечивает реакцию свёрточных слоёв сети на стегосигнал, значения которого локализованы в окрестности нуля, и подавление входных воздействий, вызванных прохождением через фильтр отдельных участков изображения. В итоге авторам по отношению к уже упоминавшимся алгоритмам адаптивной стеганографии HUGO, S-UNIWARD и WOW и чёрно-белых изображений из базы BOSSbase удалось получить меньшую ошибку обнаружения, чем при использовании SVM и признаковой системы SPAM. По сравнению с применением набора SRM и

ансамбля классификаторов, ошибка оказалось примерно на 2–5 % выше в зависимости от уровня полезной нагрузки.

В последних по времени публикациях, посвящённых использованию методологии Deep Learning в СА, используются самые разнообразные по архитектуре глубокие сети [15]. Из этих сетей, как достаточно эффективную, следует выделить Yedrouj-Net [17]. Её архитектура предполагает шесть свёрточных и три полносвязных слоя. Кроме того, в первых свёрточных слоях используются нелинейные активации в виде функций abs(...) и trunk(...) (линейная функции с ограничением снизу и сверху). Полученная с применением такой сети точность классификации при анализе данных из базы BOSSbase имеет значения порядка 72–86 % в зависимости от уровня полезной нагрузки и превосходит результаты, демонстрируемые другими архитектурами (сети Xu-Net, Ye-Net [15, 17]). Приведённые в обзоре [15] данные демонстрируют эффективность ещё одной архитектуры — сети ZhuNet [18], которая превосходит Yedrouj-Net и позволяет уменьшить ошибку на несколько процентов. Например, для алгоритма WOW с pl = 0,4 ошибка может уменьшиться с 0,14 до 0,12.

В работе отечественных авторов [19] проводится исследование собственной модели свёрточной сети и анализ известных результатов по использованию подобных алгоритмов в сравнении с алгоритмами неглубокого машинного обучения. В ходе экспериментов авторам на основе анализа сравнительно небольшого количества изображений удалось получить точность классификации порядка 85 %, что сопоставимо с ранее полученными результатами. Одновременно проводится количественный анализ различных классификаторов на основе систем признаков и нейросетевых алгоритмов. Отмечается, что существенным недостатком статистических классификаторов, отсутствующим в методах на основе нейронных сетей, является их узкая специализация на строго определённые методы формирования стегоконтейнеров.

2. Методы стегоанализа цифровых изображений с использованием глубоких нейронных сетей и гетероассоциативных преобразований

Общей идеей предлагаемого подхода является выполнение обработки информации, предусматривающей разбиение входного изображения большого размера на относительно небольшие неперекрывающиеся блоки и проведение последовательного стегоанализа этих блоков с помощью относительно простых свёрточных сетей с модифицированной архитектурой, опционально включающей дополнительный слой для специальных преобразований изображений. После этого проводится объединение результатов классификации блоков как последовательности бинарных ответов-признаков по схеме наивного байесовского классификатора в ходе вторичной постобработки. Можно ожидать, что такой подход позволит повысить точность стегоанализа, упростить процесс обучения сетей и обеспечить универсальный характер их применения, особенно при неравномерном заполнении контейнера стегосообщением, как, например, в случае, когда используются алгоритмы адаптивной пространственной стеганографии WOW, HUGO, S-UNIWARD.

В качестве средств дополнительной специальной обработки изображений в свёрточных сетях предлагается использовать так называемые интегральные гетероассоциативные преобразования (ИГП), как потенциально способные выявить нарушения структурных и статистических свойств изображений после внедрения стегосообщения (их описание дано далее). ИГП обеспечивают выделение на обрабатываемом фрагменте изображения оценочной и стохастической (маскирующей) составляющих на основе модели прогноза одной части фрагмента по отношению к другой. Такие преобразова-

ния могут включаться в архитектуру обучаемых нейронных сетей в качестве дополнительного слоя.

Как следствие блочного характера обработки в рамках реализуемого подхода, появляется возможность анализа потенциальной заполняемости блоков на этапе вторичной постобработки. При этом можно реализовать адаптивный характер стегоанализа блоков для принятия окончательного решения по всему изображению, т. е. при подсчёте «ответов» в блоках учитывать только те из них, которые могут потенциально иметь достаточно существенную заполняемость. Данный режим целесообразно использовать опционально для выявления стегосообщений, внедрённых с помощью алгоритмов адаптивной пространственной стеганографии.

Общая схема обработки информации показана на рис. 1.

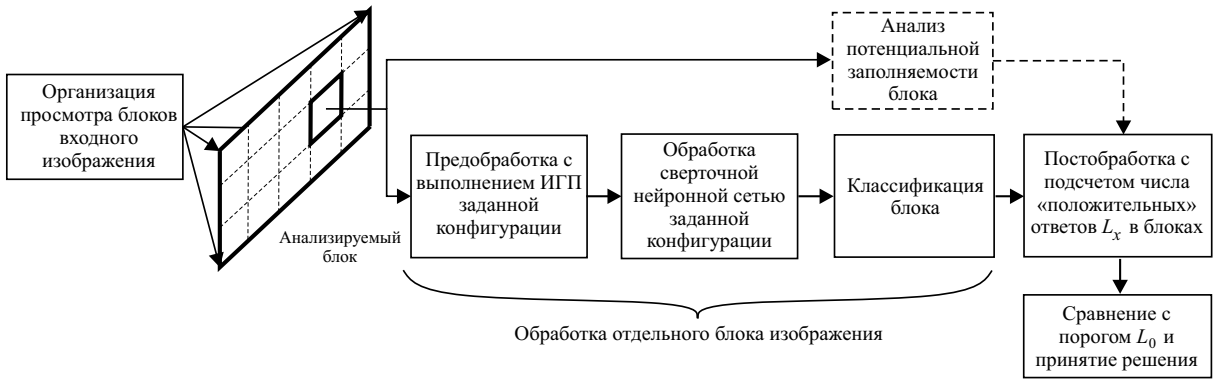


Рис. 1. Общая схема обработки информации в соответствии с предлагаемым подходом

Следует заметить, что в рамках данной схемы могут быть реализованы различные алгоритмы обработки информации, отличающиеся конкретным видом реализуемого ИГП, архитектурами нейронных сетей, способом постобработки.

В ходе проведенных исследований ставились и решались следующие задачи:

— исследование возможностей использования глубоких нейронных сетей свёрточного типа для стегоанализа цветных изображений по отношению к стандартным алгоритмам стегоскрытия в контексте общей идеи блочной обработки и их сравнение с результатами, полученными другими авторами;
— исследование возможностей повышения точности стегоанализа при использовании слоёв интегральных гетероассоциативных преобразований в глубоких нейронных сетях свёрточного типа;
— исследование возможностей повышения точности стегоанализа при использовании различных вариантов алгоритма вторичной постобработки.

2.1. Схема выполнения интегральных гетероассоциативных преобразований

В качестве исходной модели цветного изображения будем рассматривать его представление как реализацию случайного поля, заданного на дискретной сетке: $w(x,y) \in R^3$, $(x,y) \in \Psi = \{1,,\ldots,n\} \times \{1,\ldots,m\}$. Пусть $z \in R^N$ — случайный вектор, представляющий некоторую подобласть (блок) $\Omega \subset \Psi$ и формируемый путём развёртки $w(x,y)$, $(x,y) \in \Omega$, в заданном порядке $z = S_c(w)$, где $S_c(\ldots)$ — функция, определяющая порядок развёртки. Например, если Ω — прямоугольный блок размера $n_w \times m_w$, то $N = n_w \times m_w \times h$.

Вектор z может быть представлен как составной $z = (z_1^{\mathrm{T}}, z_2^{\mathrm{T}})^{\mathrm{T}}$, где $z_1 \in R^{N_1}$ представляет некоторую подобласть $\Omega_I \subset \Omega$ блока $\Omega \subset \Psi$, называемую входной частью, а

$z_2 \in R^{N_2}$ — подобласть $\Omega_O \subset \Omega$, называемую выходной частью. При этом $\Omega_I \cup \Omega_O = \Omega$, $\Omega_I \cap \Omega_O = \varnothing$. Требуется построить преобразование вида

$$F : Z_1 \to Z_{2/1} \quad z_{2/1} = F(z_1), \quad z_1 \in Z_1, \quad z_{2/1} \in Z_{2/1},$$

т. е. преобразование, выполняющее отображение данных входной части в данные выходной. Очевидно, что $z_{2/1}$ имеет определённую погрешность представления

$$z_2 = F(z_1) + V = z_{2/1} + V, \tag{1}$$

где V — стохастическая составляющая.

Указанные преобразования могут осуществляться как в полном объёме (1) с предоставлением всей информации о z_2 при известном z_1, так и со сжатием этой информации [20, 21]. Будем называть такие преобразования гетероассоциативными (ГП) или гетероассоциативными сжимающими преобразованиями (ГСП). При выполнении ГП общего вида области Ω_I, Ω_O могут быть произвольной конфигурации: прямоугольной формы, решётки случайной конфигурации внутри прямоугольных блоков и т. п. Для описания ГП случайных векторов z_1, z_2 могут использоваться как линейные, так и нелинейные модели преобразований.

Пусть некоторое количество сходных по топологии блоков Ω покрывает Ψ так, что $\bigcup\limits_{p=1}^{P} \Omega^{(p)} = \Psi$, $\Omega^{(p)} \bigcap \Omega^{(q)} = \varnothing$ для любых $p, q \in P$, таких, что $p \neq q$. Каждому блоку $\Omega^{(p)}$ соответствует реализация $z^{(p)}$ вектора z. В итоге на всём изображении может быть получена совокупность реализаций $\{z_1^{(p)}, z_2^{(p)} : p = 1, \ldots, P\}$ для входных и выходных частей $\Omega_I^{(p)}, \Omega_O^{(p)}$ блоков $\Omega^{(p)}$, $p = 1, \ldots, P$. Полученные таким образом данные служат обучающей выборкой для построения ГП с использованием прямых вычислений или нейронных сетей соответствующей архитектуры [20, 22, 23].

Если применить ГП к реализации случайной функции (к изображению) в целом, то можно выделить её полную оценочную составляющую $E(x, y) = \tilde{w}(x, y)$ на области определения, которую следует рассматривать как интерполированную функцию, полученную на основе известных входных частей всех блоков, покрывающих Ψ. Такая интерполяция является индивидуальной для каждого изображения, так как преобразование (1) формируется на основе обучающих данных. Точно так же можно выделить полную стохастическую составляющую $V(x, y) = w(x, y) - \tilde{w}(x, y)$, т. е. разложить изображение на две независимые части. Общей идеей такого разделения на составляющие, очевидно различающиеся по корреляционным и частотным свойствам, на основе блоков, покрывающих всё изображение, является использование прямого и обратного ИГП.

Под прямым ИГП понимается преобразование входной части в выходную, под обратным ИГП — выходной части во входную. Для обеспечения статистической однородности оценочной и стохастической маскирующей составляющих желательно, чтобы входная и выходная части были одинакового размера и конфигурации. С учётом использования прямого и обратного преобразований переобозначим введённое отображение входной части в выходную как $F_{\text{io}} : Z_1 \to Z_{2/1}$, $z_{2/1} = F_{\text{io}}(z_1)$. Аналогично введём обратное отображение $F_{\text{oi}} : Z_2 \to Z_{1/2}$, $z_{1/2} = F_{\text{oi}}(z_2)$. Тогда прямым ИГП будем называть отображение

$$G_{\text{io}} : W_1 \to W_{2/1}, \quad w_{2/1} = G_{\text{io}}(w_1), \quad w_1 \in W_1, \quad w_{2/1} \in W_{2/1},$$

где $w_1(x,y)$; $(x,y) \in \Psi_1 = \bigcup_{p=1}^{P} \Omega_1^{(p)}$ — случайное поле, заданное на объединённой области всех входных частей; если $(x,y) \in \Omega_1^{(p)}$, то $w_1^{(p)}(x,y) = S_c^{-1}(z_1^{(p)})$ формируется путём обратной развёртки из вектора $z_1^{(p)}$; $w_{2/1}(x,y)$, $(x,y) \in \Psi_2 = \bigcup_{p=1}^{P} \Omega_2^{(p)}$ — случайное поле, заданное на объединённой области всех выходных частей; если $(x,y) \in \Omega_2^{(p)}$, то $w_{2/1}^{(p)}(x,y) = S_c^{-1}(z_{2/1}^{(p)})$ формируется путём обратной развёртки вектора $z_{2/1}^{(p)} = F_{\text{io}}(z_1^{(p)})$.

Обратным ИГП будем называть отображение

$$G_{\text{oi}} : W_2 \to W_{1/2}, \quad w_{1/2} = G_{\text{oi}}(w_2), \quad w_{1/2} \in W_{1/2}, \quad w_2 \in W_2,$$

где $w_{1/2}(x,y)$; $(x,y) \in \Psi_1 = \bigcup_{p=1}^{P} \Omega_1^{(p)}$ — случайное поле, заданное на объединённой области всех входных частей; если $(x,y) \in \Omega_2^{(p)}$, то $w_{1/2}^{(p)}(x,y) = S_c^{-1}(z_{1/2}^{(p)})$ формируется путём обратной развёртки из вектора $z_{1/2}^{(p)} = F_{\text{oi}}(z_2^{(p)})$; $w_2(x,y)$, $(x,y) \in \Psi_2 = \bigcup_{p=1}^{P} \Omega_2^{(p)}$ — случайное поле, заданное на объединённой области всех выходных частей; если $(x,y) \in \Omega_2^{(p)}$, то $w_2^{(p)}(x,y) = S_c^{-1}(z_2^{(p)})$ формируется путём обратной развёртки из вектора $z_2^{(p)}$.

Для заданных $w_2 \in W_2$, $w_{2/1} \in W_{2/1}$ и $w_1 \in W_1$, $w_{1/2} \in W_{1/2}$ можно получить случайные поля стохастических составляющих как

$$v_2 = w_2 - G_{\text{io}}(w_1) = w_2 - w_{2/1}, \ v_2 \in V_2, \ v_1 = w_1 - G_{\text{oi}}(w_2) = w_1 - w_{1/2}, \ v_1 \in V_1,$$

где $v_2(x,y)$, $(x,y) \in \Psi_2$; если $(x,y) \in \Omega_2^{(p)}$, то $v_2(x,y) = S_c^{-1}(v_2^{(p)})$ формируется путём обратной развёртки из вектора $v_2^{(p)} = z_2^{(p)} - z_{2/1}^{(p)} = z_2^{(p)} - F_{\text{io}}(z_1^{(p)})$; $v_1(x,y)$, $(x,y) \in \Psi_1$; если $(x,y) \in \Omega_1^{(p)}$, то $v(x,y) = S_c^{-1}(v_1^{(p)})$ формируется путём обратной развёртки из вектора $v_1^{(p)} = z_1^{(p)} - z_{1/2}^{(p)} = z_1^{(p)} - F_{\text{oi}}(z_2^{(p)})$.

В итоге можно получить общее интегральное отображение для формирования оценочной и стохастической составляющих всего изображения в виде

$$G_e : W \to \tilde{W}, \ \tilde{w} = G_e(w), \ w = \{w_1, w_2\} \in W = \{W_1, W_2\},$$
$$\tilde{w} = \{w_{1/2}, w_{2/1}\} \in \tilde{W} = \{W_{1/2}, W_{2/1}\}, \ G_m : W \to V, \ v = G_m(w),$$
$$v = \{v_1, v_2\} \in V = \{V_1, V_2\}.$$

Общая схема получения ИГП с заданной конфигурацией входной и выходной частей представлена на рис. 2. Для управления уровнями оценочной и стохастической составляющих в схеме используется зависимость остаточной ошибки обучения нейросетевых преобразователей от числа циклов обучения.

В работах [20, 22 – 24] проведены теоретические исследования, которые позволили проанализировать свойства ГП и выявить наличие определённых преимуществ при их применении в различных задачах обработки изображений. В задачах стегоанализа подобные преобразования предлагается использовать как средства дополнительной обработки, способные выявить нарушения структурных и статистических свойств изображений после внедрения стегосообщения.

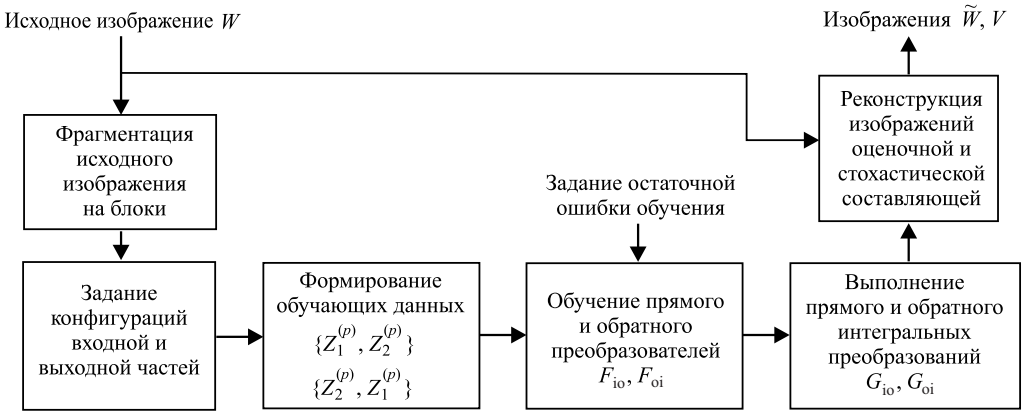


Рис. 2. Схема получения интегральных гетероассоциативных преобразований

2.2. Предлагаемые архитектуры глубоких нейронных сетей

В рамках предлагаемого подхода предложены и исследованы несколько архитектур глубоких нейронных сетей свёрточного типа. Эти сети обучались и тестировались на фрагментах изображений малого размера для последующего стегоанализа изображений большого размера на основе вторичной обработки результатов классификации, полученных на совокупности непересекающихся блоков, покрывающих изображение. При задании архитектуры сетей рассматривались возможности использования ИГП с выделением оценочной и маскирующей составляющих, которые подробно описаны в [20]. Далее представлены две глубокие нейронные сети, применение которых заслуживает особого внимания.

Первая из этих сетей имеет стандартную архитектуру, но содержит всего три обучаемых свёрточных слоя и три обучаемых полносвязных слоя. Её отличие от известных сетей, используемых для СА (например, GNCNN), состоит в том, что после входного слоя введён слой предобработки входного изображения для всех трёх каналов, в котором реализован пространственный высокочастотный фильтр с возможностью гибкой перестройки параметров [25]. Ядро фильтра является симметричным и описывается следующим выражением:

$$h(x,y) = \begin{cases} (-1)^{|x|+|y|+1} \exp\left[-0{,}5\alpha\sqrt{x^2+y^2}\right] \operatorname{sinc}(x/3)\operatorname{sinc}(y/3), & |x| \leqslant d/2, |y| \leqslant d/2, \\ 0, & |x| > d/2, |y| > d/2, \end{cases}$$

где x, y — целочисленные значения аргумента, принимающие положительные и отрицательные значения в пределах размера ядра d. Параметры функции ядра: $d = 5$, $\alpha = 0{,}1$.

Что касается используемых в свёрточных слоях функций активации, то в первом свёрточном слое после слоя высокочастотной фильтрации, свёртки и батч-нормализации реализована (по аналогии с GNCNN) гауссовская функция активации с настраиваемым в процессе обучения параметром влияния σ (среднеквадратичным отклонением). Начальная инициализация σ проводится датчиком случайных чисел в диапазоне 0,01–0,5. Во всех остальных свёрточных и полносвязных слоях применяются активации Relu, за исключением выходного полносвязного слоя, где используется стандартная активация Softmax. Архитектура сети представлена на рис. 3. Далее будем использовать для этой сети обозначение CNN-HF-GN (подчеркивая особенности первого свёрточного слоя).

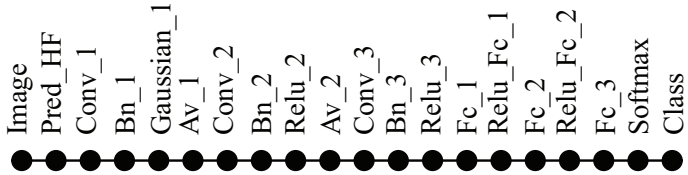


Рис. 3. Архитектура сверточной нейронной сети CNN-HF-GN

На рис. 3 использованы следующие обозначения: Image — входной слой $n_x \times m_y \times 3$, принимающий блок изображения в трёх каналах цветности; Pred_Hf — слой, отвечающий за выполнение пространственной высокочастотной фильтрации; Conv_1 — первый свёрточный слой с ядром свертки 5×5, шагом 1 и 32 каналами; Conv_2, Conv_3 — второй и третий слои с ядром свёртки 3×3, шагом 1, 64 и 128 каналами соответственно; Bn_1, Bn_2, Bn_3 — слои стандартной батч-нормализации; Gaussing_1 — гауссовская функция активации для первого слоя свёртки $f(x) = \exp(-x^2/2\sigma^2)$; Relu_2, Relu_3 — стандартные функции активации для второго и третьего слоёв типа «линейная с ограничением снизу» $f(x) = \max(0, x)$; Av-1, Av-2, Av-3 — слои субдискретизации (пуллинга) на основе вычисления среднего (average pooling) в окне 3×3 с шагом 2; Fc_1, Fc_2, Fc_3 — полносвязные слои, имеющие соответственно 128, 128 и 2 выхода; Relu_Fc_1, Relu_Fc_2 — стандартные функции активации для первого и второго полносвязных слоёв типа «линейная с ограничением снизу» $f(x) = \max(0, x)$; Softmax — стандартная функция активации для классификации на два класса с использованием в качестве функции потерь кросс-энтропии; Class — слой классификации, отвечающий за вычисление функции потерь кросс-энтропии при классификации на несколько взаимоисключающих классов образов.

Как показали многочисленные эксперименты, увеличение количества свёрточных слоев, а также использование слоев dropout не даёт ощутимого прироста точности классификации.

Вторая сеть имеет схожую с предыдущей архитектуру и содержит три обучаемых свёрточных слоя и три обучаемых полносвязных слоя. Её отличие от сети CNN-HF-GN состоит в том, что сразу после входного слоя введён слой обработки входного изображения, основанный на выполнении ИГП с выделением маскирующей составляющей (Masking Component, MC). Можно ожидать, что выделение MC, в которой в основном локализуются искажения, возникающие при внедрении стегосообщения, позволит повысить информативность обработки в целом.

Такие свёрточные сети будем обозначать CNN-MC-HF-GN. Может быть использована также архитектура CNN-MC-GN с исключением высокочастотного фильтра из схемы обработки. На рис. 4 приведена сеть CNN-MC-HF-GN, где обозначение Pred_MC определяет авторский слой обработки с выделением маскирующией составляющей. Все остальные обозначения соответствуют введённым для CNN-HF-GN.

Слой предобработки для выделения MC основан на выполнении ИГП, описанных в п. 2.1. Для построения операторов связи входной и выходной частей (как «вперёд», так и «назад») в качестве типовой использована конфигурация «шахматы», формируемая параллельно во всех трёх каналах цветности в блоках размера 8×8 с размером элементарного квадрата 2×2 (рис. 5).

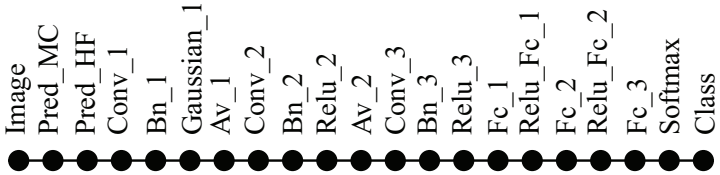
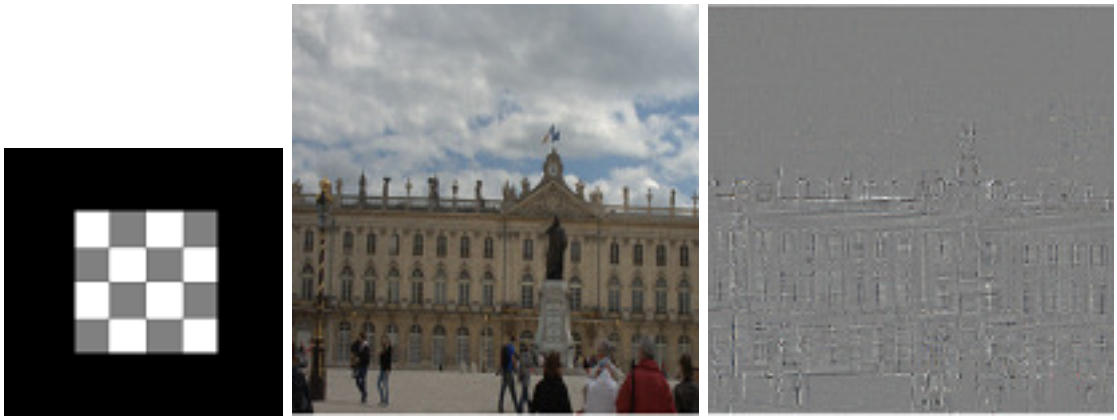


Рис. 4. Архитектура сверточной нейронной сети CNN-MC-HF-GN с использованием дополнительного слоя Pred_MC



а *б* *в*

Рис. 5. Типовая конфигурация входной и выходной частей в проекции на каждый цветовой канал (*а*) и пример исходного изображения (*б*) и его маскирующей составляющей (*в*)

Построение оператора связи осуществлялось на «пустых» изображениях той же выборки, что и при обучении глубокой нейронной сети. При построении оператора по обучающим примерам может использоваться как линейная, так и нелинейная оценка, в соответствии с процедурой, подробно описанной в [20].

2.3. Алгоритмы вторичной обработки результатов классификации

Как уже упоминалось, идея предлагаемого подхода состоит в обучении сетей относительно простой архитектуры на небольших фрагментах (блоках) размером 32×32, 64×64, 128×128 и последовательной вторичной обработке (постобработке) результатов классификации, выполненной с использованием ранее обученного нейросетевого классификатора на блоках целостного изображения большого размера, с целью принятия окончательного решения. При анализе цветных изображений размера $N_x \times M_y$ в ходе постобработки результатов классификации глубокой сетью совокупности блоков размера $n_x \times m_y$, размещаемых без перекрытия, получим $P = (N_x \cdot M_y)/(n_x \cdot m_y)$ бинарных ответов (считаем, что блоки по осям укладываются в исходном изображении кратное число раз). В статистическом смысле можно считать получаемые по каждому блоку решения независимыми. Обозначим эти решения x_k, $x_k \in \{1, 0\}$, $k = 1, \ldots, P$, где $x_k = 1$ обозначает решение в пользу наличия стегосообщения в анализируемом блоке, а $x_k = 0$ — в пользу его отсутствия.

Для синтеза окончательного решающего правила по результатам вторичной обработки введём обозначения для вероятностей значений бинарных признаков двух

классов ω_1 и ω_2, соответствующих пустому и заполненному контейнерам:

$$p_k = \mathsf{P}[x_k = 1 \,|\, \omega_1], \quad 1 - p_k = \mathsf{P}[x_k = 0 \,|\, \omega_1], \quad k = 1, \ldots, P,$$
$$q_k = \mathsf{P}[x_k = 1 \,|\, \omega_2], \quad 1 - q_k = \mathsf{P}[x_k = 0 \,|\, \omega_2], \quad k = 1, \ldots, P.$$

Тогда выражения для функций правдоподобия классов и логарифма отношения правдоподобия можно записать в следующем виде:

$$p(x \,|\, \omega_1) = \prod_{k=1}^{P} p_k^{x_k}(1 - p_k)^{1-x_k}, \quad p(x \,|\, \omega_2) = \prod_{k=1}^{P} q_k^{x_k}(1 - q_k)^{1-x_k},$$
$$g(x) = \sum_{k=1}^{P} \ln \frac{p(x_k \,|\, \omega_1)}{p(x_k \,|\, \omega_2)} = \sum_{k=1}^{P} x_k \ln \frac{p_k\,(1 - q_k)}{q_k\,(1 - p_k)} + P \ln \frac{(1 - p_k)}{(1 - q_k)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} l_0 = \ln \frac{p(\omega_2)}{p(\omega_1)}. \qquad (2)$$

Таким образом, структура классификатора предполагает взвешенное суммирование значений признаков и сравнение этой суммы с вычисляемым порогом. Такой классификатор может быть получен с помощью любого алгоритма машинного обучения по соответствующей выборке ответов первичного классификатора по всем используемым изображениям.

Однако представляет интерес получение подобного классификатора в простой форме. Предположим, что для всех ответов каждого класса вероятности единиц и нулей одинаковы: $p_k = p \neq 0$, $q_k = q \neq 0$, $k = 1, \ldots, P$, причём $p < q$. Тогда $\ln(p(1 - q)/q(1 - p)) < 0$, поэтому решающее правило преобразуется к виду

$$L_x = \sum_{k=1}^{P} x_k \underset{\omega_1}{\overset{\omega_2}{\gtrless}} L_0 = \left(l_0 - P \ln \frac{1 - p}{1 - q} \right) \Big/ \ln \frac{p(1 - q)}{q(1 - p)}, \qquad (3)$$

где L_x — количество единиц (ответов «да»), полученных в ходе наблюдения. Данная ситуация означает, что фактически проводится «опрос» P равноценных признаков и сравнение общего количества полученных единиц с пороговым значением, зависящим от априорных вероятностей классов и соотношения вероятностей ошибок первого и второго рода ответов первичной обработки $p \simeq er_{12}$, $1 - q \simeq er_{21}$, получаемых от нейросетевого классификатора в ходе анализа блоков. Определим порог принятия решения в (3) в виде целочисленной величины

$$L_0 = \operatorname{round} \left[\left(l_0 - P \ln \frac{1 - p}{1 - q} \right) \Big/ \ln \frac{p(1 - q)}{q(1 - p)} \right].$$

Модель принятия решений (3) является приближённой, особенно в ситуациях неравномерного заполнения контейнера стегосообщением, что характерно, например, для алгоритмов WOW, HUGO, S-UNIWARD. В подобных случаях вероятности ошибок принятия решений в блоках нельзя считать одинаковыми, что не позволяет точно рассчитать порог для L_0. Для алгоритма (3) можно спрогнозировать значения порога. Анализ выражения для L_0 при $l_0 = 0$ (стратегия максимального правдоподобия) показывает, что при одинаковых ошибках первого и второго рода ($p = 1 - q$, $q = 1 - p$) выполняется $L_0/P = 0{,}5$. При увеличении $er_{12} = p$, очевидно, пороговое отношение должно увеличиваться, а при увеличении $er_{21} = 1 - q$ — уменьшаться. Расчёты L_0/P показывают, что для значений соотношения $p/1 - q = er_{12}/er_{21}$, лежащих в интервале 0,5–2, значения порога находятся в пределах 0,4–0,6. В ходе тестирования нейросетевого классификатора на блоках из обучающей выборки можно провести оценку

оптимального значения порога, заменив тем самым обучение классификатора (2) обучением порога.

В качестве одной из возможных модификаций алгоритма с применением правила (3) в случае неравномерного заполнения контейнера стегосообщением можно предложить адаптивный алгоритм, предусматривающий подсчёт числа решений не по всем блокам анализируемого изображения, а только по тем из них, которые могут содержать определённое количество модифицированных пикселей. Для этого проводится анализ каждого блока с точки зрения его зашумлённости и выбираются только те блоки, в которых число потенциально модифицированных пикселей может быть больше заданного порога. Простейшим способом подобного отбора блоков является независимое встраивание в него произвольной псевдослучайной последовательности с помощью одного из алгоритмов WOW, HUGO, S-UNIWARD и подсчёта числа изменённых пикселей. Полученный результат может быть близок к реальному числу модифицируемых пикселей. Примерно аналогичные выводы даёт дисперсионный анализ блоков. Тогда правило (3) преобразуется к виду

$$L_x = \sum_{k=1}^{P_{\text{ch}}} x_{i_k} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} L_{0,\text{ch}}, \quad R_{i_k} > \rho\, m_{\text{stego}}, \quad k = 1, \ldots, P, \quad m_{\text{stego}} = \frac{1}{P} \sum_{k=1}^{P} R_k, \qquad (4)$$

где P_{ch} — число блоков анализируемого изображения, которые могут содержать определённое количество модифицированных пикселей; $L_{0,\text{ch}}$ — порог принятия решения для таких блоков; R_{i_k} — количество потенциально модифицированных пикселей в блоке с индексом i_k, $k = 1, \ldots, P$; m_{stego} — среднее арифметическое потенциального заполнения блоков в данном изображении; ρ — коэффициент, определяющий порог, по которому блоки отбираются для использования в процессе принятия решения. Экспериментально установлено, что значения ρ должны устанавливаться в диапазоне 0,25–0,5 в зависимости от P. Подбор оптимального порога может проводиться в ходе эксперимента с учётом ошибок первого и второго рода, полученных на этапе обучения и тестирования нейронной сети, анализирующей блоки. Выявлено, что чем меньше величина pl, определяющая потенциально возможную полезную нагрузку изображения при встраивании в него ССИ, тем больше должен быть коэффициент ρ. Можно ожидать, что эффект от применения адаптивного алгоритма с правилом (4) будет проявляться при использовании блоков малых размеров и анализе изображений с минимальной полезной нагрузкой.

Обобщённая схема алгоритма вторичной обработки представлена на рис. 6, где показан адаптивный вариант алгоритма с правилом (4) с анализом заполняемости блоков. В случае правила (3) соответствующие блоки в этой схеме опускаются.

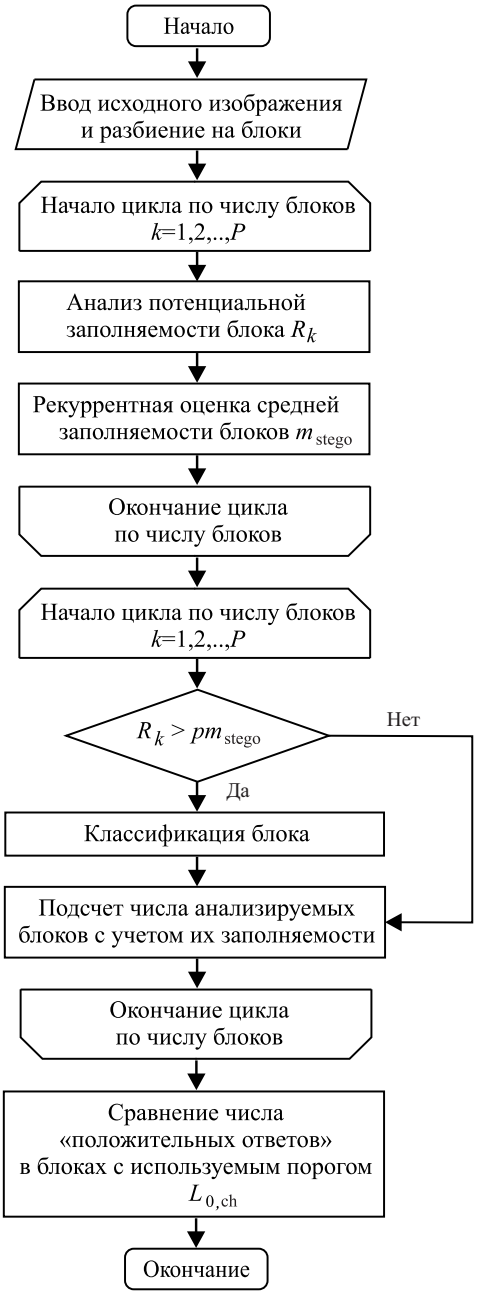


Рис. 6. Схема алгоритма вторичной обработки (постобработки)

2.4. Описание исследуемых алгоритмов ССИ и методики обучения глубоких сетей

В качестве объектов исследования выбрано четыре группы алгоритмов стегоскрытия, в каждой из которых рассматрено по два типовых представителя. Эти группы можно определить следующим образом:

— классические блочные алгоритмы стегоскрытия с внедрением элементов скрываемого сообщения в неперекрывающиеся блоки фиксированного размера — алгоритмы Куттера [26] и Коха — Жао [27];
— современные алгоритмы внедрения в непересекающиеся группы выбранных в псевдослучайном порядке элементов контейнера и адаптивные алгоритмы с использова-

нием различных по степени зашумлённости областей для встраивания — алгоритмы EMD [28] и MBEP [29];

— блочные алгоритмы стегоскрытия на основе ГСП с внедрением элементов сообщения путём модификации спектральных составляющих ГП при случайной конфигурации входной и выходной частей — алгоритмы ГСП+ и ГСП++ [20];

— современные алгоритмы адаптивной пространственной стеганографии, принципиально ориентированные на обеспечение высокой скрытности по отношению к стегоанализу — WOW [10] и S-UNIWARD [11].

Рассмотрим особенности применения этих алгоритмов в представленных далее экспериментах.

В алгоритме Куттера — Джордана — Боссена осуществлялось скрытие данных в пространственной области, разбитой на блоки размера 8×8, путём внесения относительных изменений яркости в синей цветовой компоненте. Это изменение пропорционально значению пикселя и может принимать как положительные, так и отрицательные значения в зависимости от значения встраиваемого бита ЦВЗ. В исследованиях параметр вносимых искажений λ задавался равным 0,05 (уровень модификации 5 %).

В алгоритме Коха — Жао осуществлялось скрытие элементов битовой последовательности в блоки 8×8, подвергнутые дискретному косинусному преобразованию, путём изменения соотношения уровней спектральных коэффициентов среднечастотных составляющих синей цветовой составляющей с позициями (5,4) и (4,5). Параметр изменения ε задавался равным 3, что соответствует среднему уровню вносимых искажений при минимальном значении $\varepsilon = 1$.

Алгоритм EMD (Exploiting Modification Direction) реализует встраивание элементов исходного сообщения, представленных значениями, заданными в позиционной системе счисления с основанием $(2n + 1)$, в выбранные в пседослучайном порядке группы из n элементов изображения. Вносимые в контейнер искажения не превышают одного уровня яркости/цветности. При проведении исследований параметры алгоритма (число элементов n в группах, процент используемых для наполнения групп пикселей контейнера) определялись таким образом, чтобы объём полезной нагрузки составлял примерно 0,4 bpp. Встраивание реализовывалось во все цветовые каналы (RGB) тестовых контейнеров.

Алгоритмы ГСП+ и ГСП++ относятся к блочным алгоритмам, осуществляющим скрытие данных в частотной области ГП [22, 23]. Для извлечения стегосообщения требуется иметь ключ в виде нейронной сети, обученной для классификации блоков изображений. Алгоритм ГСП+ основан на выполнении ГСП в прямом направлении с модификацией пикселей только выходной части блока и встраивании одного бита сообщения в каждый блок. Алгоритм ГСП++ основан на выполнении ГСП как в прямом, так и в обратном направлениях с модификацией пикселей как выходной, так и входной части блока и встраивании двух бит сообщения в каждый блок. В исследованиях использовалась реализация алгоритмов со случайной конфигурацией входной и выходной частей в «пироге трёх цветов» [24] для блоков размера 8×8. Амплитуда встраиваемой последовательности задавалась равной 3/255, что обеспечивает близкую к нулю вероятность ошибки при восстановлении стегосообщения.

В алгоритмах WOW и S-UNIWARD для восстановления стегосообщения требуется знание исходного немодифицированного изображения. Это позволяет обеспечить минимальный уровень искажения каждого пикселя, равного по модулю одному уровню квантованного представления (один бит), и высокую пропускную способность, сопоставимую с общим числом пикселей изображения. В экспериментах мы применяли

алгоритмы с двумя вариантами значений полезной нагрузки pl = 0,2 bpp и 0,4 bpp. Для каждого из этих значений использовалась процедура встраивания только в один цветовой канал (синий) или во все три цветовых канала (RGB).

Алгоритм MBEP (Keyless dynamic optimal multi-bit image steganography using energetic pixels), подобно алгоритмам WOW и S-UNIWARD, реализует адаптивное встраивание информации в пространственном представлении, выбирая для модификации лишь наиболее шумные участки. Однако, в отличие от WOW и S-UNIWARD, он не требует знания исходного контейнера при извлечении скрытых данных. Значения яркости/цветности пикселей контейнера на выбранных участках модифицируются путём перезаписи их младших разрядов битами встраиваемого сообщения (не более четырёх разрядов), что позволяет обеспечить высокую пропускную способность при относительно малых визуальных искажениях носителя. При проведении исследований объём полезной нагрузки для заполнения контейнера составлял примерно 0,4 bpp. Встраивание реализовывалось во все цветовые каналы (RGB) тестовых контейнеров.

В качестве исходного датасета при проведении исследований была взята база данных PPG-LIRMM-COLOR, содержащая 10000 цветных изображений размером 512×512 в формате *.bmp. В качестве стегосообщения генерировалась псевдослучайная последовательность с оригинальным для каждого изображения установочным ключом. Для создания и встраивания стегоообщений в изображения применялись перечисленные алгоритмы ССИ в оригинальной реализации, кроме алгоритмов адаптивной пространственной стеганографии WOW и S-UNIWARD, для которых была использована реализация симулятора [30]. Таким образом, всего проанализировано 20000 изображений: 10000 исходных и 10000 со стегосообщениями.

В табл. 1 приведены характеристики алгоритмов ССИ с точки зрения уровня вносимых искажений и возможности восстановления стегосообщения (MAE — максимальная абсолютная ошибка, MSE — средняя квадратичная ошибка, SSIM — индекс структурного сходства, P_{ex} — вероятность ошибки восстановления). Для алгоритмов WOW и S-UNIWARD данные приведены для случая использования одного канала цветности. При вычислении MAE бралось медианное значение по выборке изображений. Данные по показателям MSE и SSIM рассчитывались после приведения динамического диапазона яркостей пикселей в каналах цветности к диапазону $[0, 1]$.

Для обучения нейронных сетей использовалось 16000 изображений: 8000 исходных и 8000 со стегосообщениями. Валидация и тестирование проводились на 4000 оставшихся изображений: 2000 исходных и 2000 заполненных. При обучении и тестировании сети с различными размерами входного блока из исходных и заполненных изображений вырезались небольшие изображения размера $n_x \times m_y$ со случайным смещением и формировались подвыборки из такого же числа изображений малого размера.

Ввиду пространственной неравномерности заполнения контейнеров стегоообщениями, чтобы исключить использование при обучении детекторов алгоритмов WOW и S-UNIWARD пустых фрагментов заполненных изображений, проводился анализ каждого блока с точки зрения его зашумлённости с помощью алгоритма WOW и подсчёта потенциального числа изменённых пикселей (как для исследования по WOW, так и по S-UNIWARD). При формировании датасета в каждом изображении (как исходном, так и заполненном) выбирались только те блоки, в которых число потенциально модифицированных пикселей больше порога: $R_{i_k} > 0{,}5\, m_{\text{stego}}$, $k = 1, \ldots, P$ (см. (4)). Затем из полученного списка случайным образом выбирался один из блоков. Так получались обучающая и валидационная подвыборки из 16000 и 4000 изображений размера $n_x \times m_y$ для WOW и S-UNIWARD. Кроме того, из 4000 исходных полноразмерных

Таблица 1

Результаты сравнения алгоритмов ССИ по основным характеристикам стегоскрытия

Характеристика	Наименование алгоритма			
	Куттера	**Коха — Жао**	**EMD**	**MBEP**
MAE	12	7	1	15
MSE	9,3205e-06	9,8018e-06	1,6257e-07	6,9528e-06
SSIM	0,99973	0,99979	0,99997	0,99962
P_{ex}	∼0,05	∼0,00	∼0,00	∼0,00
	ГСП+	**ГСП++**	**WOW**	**S-UNIWARD**
MAE	2	2	1	1
MSE	6,9739e-07	1,4331e-06	1,2522e-06	1,0546e-06
SSIM	0,99998	0,99996	0,99996	0,99996
P_{ex}	∼0,01	∼0,01	Для восстановления требуется исходное изображение	Для восстановления требуется исходное изображение

изображений, не участвующих в обучающей подвыборке, формировалась тестирующая подвыборка со случайным выбором участвующих в ней блоков. Это обеспечивало практическую независимость тестирующей и валидационной подвыборок.

При обучении сети использовался оптимизатор adam на 30 эпохах с начальной скоростью 0,001, параметром L_2-регуляризации 0,001, размером минибатча 64. Затем проводилось дообучение сети на 10 эпохах с начальной скоростью 0,0001 и параметром L_2-регуляризации 0,0001.

При обучении нейронных сетей на изображениях с малым уровнем вносимых при ССИ искажений или с малой полезной нагрузкой может случиться, что сеть из-за слишком малых различий пустых и заполненных изображений не выходит в режим обучения. Для преодоления этой ситуации целесообразно проводить задание стартовых значений весовых коэффициентов с использованием сети, ранее обученной на изображениях для высоких уровней полезной нагрузки, и переобучать её на изображениях с малым уровнем полезной нагрузки. Подобная стратегия оказалась весьма эффективной и в ситуациях, когда сети удавалось обучаться самостоятельно при малой полезной нагрузке изображений.

Принципиальным моментом при проведении исследования также являлось использование, по возможности, одинаковых базовых архитектур нейронных сетей. Если есть отличия от базовой архитектуры, они обязательно оговариваются. Все алгоритмы обработки информации реализованы в среде Matlab с использованием пакета Deep Learning Toolbox.

3. Результаты и их обсуждение

Проведено обучение и тестирование предложенных глубоких сетей, а также алгоритма вторичной постобработки. Первоначально внимание уделялось исследованию наиболее скрытных для стегоанализа алгоритмов WOW и S-UNIWARD при различных объёмах полезной нагрузки и заполнении каналов цветности (подробно результаты представлены в [25]). В табл. 2 приведены результаты для базовой архитектуры глубокой сети CNN-HF-GN, полученные при тестировании на блоках изображений

небольших размеров и после постобработки совокупности блоков для изображений размера 512×512 на основе алгоритмов (3)/(4).

Таблица 2

Результаты оценки точности обнаружения в блоках и по всему изображению для нейронной сети CNN-HF-GN при различных объемах полезной нагрузки в %

Алгоритм ССИ	**WOW (grayscale)**		**WOW (RGB)**		**S-UNIWARD (grayscale)**		**S-UNIWARD (RGB)**	
Размеры и число блоков на изображении	pl=0,2	pl=0,4	pl=0,2	pl=0,4	pl=0,2	pl=0,4	pl=0,2	pl=0,4
$n_x = 32$,	54,83	60,55	61,60	72,45	53,37	60,10	61,35	73,12
$m_y = 32$,	55,97/	70,38/	81,30/	92,10/	56,40/	65,63/	77,90/	91,47/
$P = 256$	59,02	71,50	82,53	92,85	57,95	67,30	79,03	91,73
$n_x = 64$,	56,05	66,27	65,75	81,57	55,37	65,70	64,82	77,80
$m_y = 64$,	58,07/	76,57/	82,65/	93,92/	58,13/	74,28/	82,42/	92,47/
$P = 64$	62,50	77,28	82,17	93,83	59,02	75,35	82,82	92,03
$n_x = 128$,	59,78	74,05	76,18	87,90	60,62	75,55	73,62	85,62
$m_y = 128$,	65,38/	82,57/	87,30/	94,72/	66,30/	81,37/	84,42/	94,45/
$P = 16$	66,32	82,28	88,00	94,47	67,35	81,88	84,45	94,45

Анализ результатов показывает, что при заполнении всех трёх цветовых каналов точность обнаружения существенно повышается, как и при увеличении размера анализируемого фрагмента. При малой нагрузке pl=0,2 только в одном цветовом канале, что соответствует реальному заполнению контейнера ещё в 3 раза меньше, точность обработки не выше 60 %, при этом величина ошибки первого рода, когда пустой фрагмент принимается за заполненный, может превышать 50 % при существенно меньшей ошибке второго рода. Однако при достаточно большой нагрузке вероятности ошибок для оптимального значения порога отличаются не так существенно.

На рис. 7 показаны типичные гистограммы, описывающие распределение числа ошибок первого и второго рода после вторичной обработки и точности классификации в зависимости от отношения порога L_0/P при размере блока 32×32. Внедрение стегосообщения здесь осуществлялось алгоритмом WOW в трёх каналах цветности при объёме полезной нагрузки pl = 0,4. Максимальное значение точности стегоанализа 92,1 достигается при отношении $L_0/P = 0{,}65$. При использовании правила (4) максимальное значение точности 92,85 достигается при отношении $L_{0,\mathrm{ch}}/P_{\mathrm{ch}} = 0{,}5$.

Таким образом, предложенный подход и реализованные на его основе алгоритмы позволяют достичь точности обнаружения стегосообщений, сопоставимой с лучшими результатами, представленными в известных работах, а в некоторых случаях и превышающей их. Также следует отметить, что, как и ожидалось, применение адаптивного алгоритма вторичной обработки оправдано в большей степени при малых размерах анализируемых блоков и соответственно большом их количестве, а также при меньшей величине полезной нагрузки. Этот вариант алгоритма позволяет получить прирост точности от 0,5 до 4,5 %.

Результаты исследования возможностей стегоанализа цветных изображений по отношению к алгоритмам стегоскрытия на основе использования ИГП при построении глубоких нейронных сетей представлены в табл. 3. Применялись сети CNN-HF-GN и

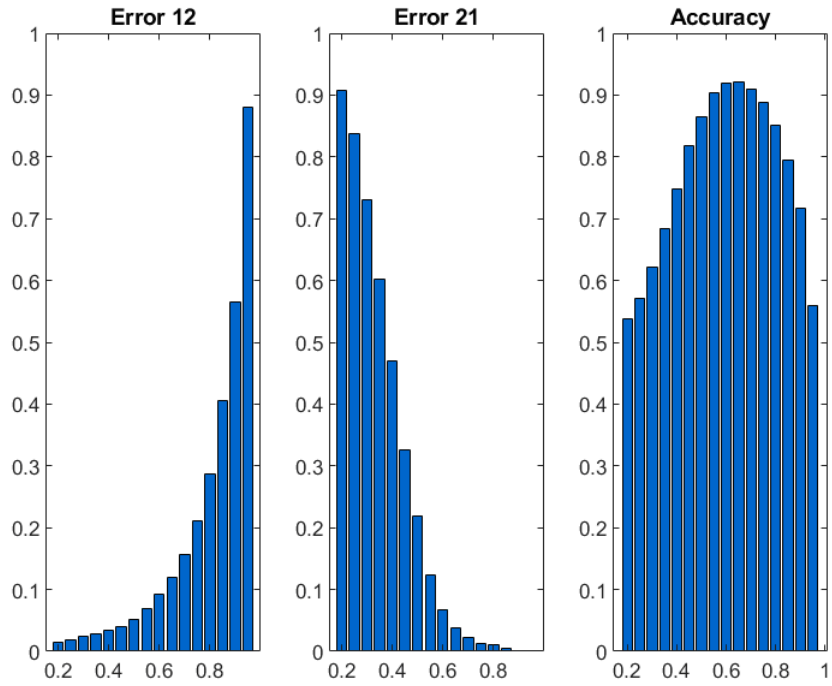


Рис. 7. Типичные гистограммы распределения числа ошибок первого и второго рода и точности классификации изображений от отношения L_0/P

CNN-MC-HF-GN, а также вариант CNN-MC-GN для блоков размера 64×64 и вторичной обработки ответов в блоках для изображений размером 512×512.

Таблица 3

Результаты оценки точности обнаружения в блоках и по всему изображению для нейронной сети CNN-HF-GN и CNN-MC-HF-GN (CNN-MC-GN)

Используемая сеть и размер изображения	CNN-HF-GN		CNN-MC-HF-GN / CNN-MC-GN	
	64×64	512×512	64×64	512×512
Алгоритм ССИ				
Куттера	86,78	96,50	85,70	96,77
Коха — Жао	73,10	84,75	79,15 / 91.27	89,15 / 98,05
EMD	85,88	96,25	86,60	96,57
MBEP	94,23	98,04	93,93 / 93,53	98,72 / 98,83
ГСП+	85,30	92,20	85,78	92,33
ГСП++	93,00	97,23	93,00 / 92,80	97,25 / 97,23
WOW(rgb), pl=0,2	67,90	83,57	69,45	85,35
WOW(rgb), pl=0,4	83,30	93,98	82,97	94,35
S-UNIWARD(rgb), pl=0,2	65,48	80,73	65,60	83,33
S-UNIWARD(rgb), pl=0,4	77,80	92,03	80,37	93,33

Анализ данных в табл. 3 показывает, что эффективность обработки на изображениях большого размера повышается при использовании дополнительного слоя, обеспечивающего выделение маскирующей составляющей на основе ИГП. Выигрыш составляет в среднем 0,5–2 %, а порой и более 5 % (для алгоритма Коха — Жао). Указанный эффект наблюдается даже в случаях, когда между CNN-HF-GN и CNN-MC-HF-GN нет существенных отличий в точности анализа блоков, что объясняется лучшей обобщаю-

щей способностью второй сети, фиксируемой также и на графиках процесса обучения. Для некоторых алгоритмов стегоскрытия хорошие результаты позволяет достичь использование CNN-MC-GN без высокочастотного фильтра, причём для алгоритма Коха — Жао этот вариант даёт весьма значительное преимущество — до 9 %.

Алгоритмы ГСП+ и ГСП++ обладают сопоставимой с другими устойчивостью по отношению к стегоанализу. При этом алгоритм ГСП+ в этом плане уступает только алгоритмам WOW и S-UNIWARD при малой полезной нагрузке pl=0,2. Поскольку последние требуют исходного изображения для восстановления стегосообщения, данный результат свидетельствует о высокой степени скрытности алгоритмов ССИ, основанных на использовании гетероассоциативных сжимающих преобразований.

Для сравнения с результатами других авторов приведены данные в табл. 4. Здесь указаны характеристики архитектур (количество и параметры обучаемых слоёв) наиболее эффективных сетей Yedroudj-Net и Zhu-Net [17, 18] и достигнутая ими точность стегоанализа для алгоритмов WOW и S-UNIWARD на изображениях размера 256×256. Приведены аналогичные данные по архитектуре и точности сети CNN-MC-HF-GN, показавшей лучшие результаты в наших экспериментах, для блоков размера 64×64 и после вторичной обработки блоков на изображениях размера 512×512.

Таблица 4

Результаты оценки точности обнаружения факта стегоскрытия для сетей Yedroudj-Net, Zhu-Net и CNN-MC-HF-GN

Используемая сеть и характеристики её архитектуры	Yedroudj-Net (6 слоев свертки — ядра 5×5, 3×3; 3 полносвязных слоя)	Zhu-Net (6 слоев свертки — ядра 3×3; 2 полносвязных слоя)	CNN-MC-HF-GN (3 слоя свертки — ядра 5×5, 3×3; 3 полносвязных слоя)	
	256×256	256×256	64×64	512×512
Алгоритм ССИ				
WOW(rgb), pl=0,2	72,20	76,70	69,45	85,35
WOW(rgb), pl=0,4	85,60	88,20	82,97	94,35
S-UNIWARD(rgb), pl=0,2	63,40	71,50	65,60	83,33
S-UNIWARD(rgb), pl=0,4	77,20	84,70	80,37	93,33

Анализ представленных данных показывает, что после обучения даже в блоках размера 64×64 сеть CNN-MC-HF-GN с использованием слоя ИГП показывает сопоставимые результаты с Yedroudj-Net и несколько хуже Zhu-Net. В ходе обработки изображений большого размера её эффективность резко возрастает и даёт возможность ожидать, как минимум, не худшую точность, чем рассматриваемые аналоги. При этом архитектура сети CNN-MC-HF-GN с точки зрения количества обучаемых слоёв проще, что, с учётом существенно меньшей размерности входного изображения в блоке, позволяет снизить затраты на обучение сетей с предлагаемой архитектурой.

Заключение

В развитие известных результатов в области машинного обучения и применения глубоких нейронных сетей для стегоанализа цифровых изображений предложен подход, основанный на использовании свёрточных сетей в качестве классификаторов для последовательного анализа небольших блоков изображений с вторичной постобработкой для вынесения решения по всему изображению в целом. Рассмотрены и исследованы альтернативные варианты архитектур глубоких сетей. Предложена простая архитектура свёрточной сети, состоящей из трех свёрточных и трёх полносвязных

обучаемых слоёв. В первом свёрточном слое дополнительно реализован слой пространственной высокочастотной фильтрации с возможностью гибкой перестройки параметров фильтра и гауссовская функция активация с настраиваемым в процессе обучения параметром влияния. Альтернативный вариант архитектуры сети основан на использовании слоя, выполняющего ИГП. Последнее обеспечивает выделение в обрабатываемом блоке изображения оценочной и стохастической (маскирующей) составляющих путём построения модели прогноза одной части блока по отношению к другой. В качестве алгоритма вторичной обработки результатов классификации совокупности блоков в пределах одного изображения использован простой алгоритм сравнения общего числа «положительных» и «отрицательных» решений с экспериментально подбираемым (обучаемым) порогом.

При проведении исследований рассматривались возможности стегоанализа цветных изображений по отношению к широкому перечню алгоритмов стегоскрытия. Показано, что предложенные алгоритмы стегоанализа позволяют выявлять факт наличия стеганографически скрытой информации при использовании наиболее скрытных алгоритмов внедрения с точностью, не уступающей результатам, полученным в работах других авторов. При этом вычислительные затраты на процесс обучения снижаются за счёт более простой архитектуры и уменьшения размерности входных данных.

Установлено, что результирующая эффективность обработки на изображениях большого размера повышается при использовании дополнительного слоя, обеспечивающего выделение маскирующей составляющей на основе ИГП. Как одно из преимуществ предлагаемого подхода следует отметить независимость реализуемой схемы обработки от размера анализируемого изображения в той её части, где проводится обучение нейронных сетей, а также возможность быстрого переобучения нейронных сетей, ранее обученных на изображениях с высокой полезной нагрузкой, для обнаружения стегосообщений на изображениях с малой полезной нагрузкой.

ЛИТЕРАТУРА

1. *Шелухин О. И.* Стеганография. Алгоритмы и программная реализация. М.: Горячая линия — Телеком, 2017. 592 с.
2. *Czaplewski B.* Current trends in the field of steganalysis and guidelines for constructions of new steganalysis schemes // Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne. 2017. No. 10. P. 1121–1125.
3. *Lyu S. and Farid H.* Detecting hidden messages using higher-order statistics and support vector machines // Intern. Workshop Inform. Hiding. Berlin; Heidelberg: Springer, 2002. P. 340–354. `https://farid.berkeley.edu/downloads/publications/ih02.pdf`.
4. *Lyu S. and Farid H.* Steganalysis using color wavelet statistics and one-class support vector machines // Proc. SPIE. California, USA, 2004. P. 35–45. `https://www.researchgate.net/publication/221011180\_Steganalysis\_using\_color\_wavelet\_statistics\_and\_one-class\_support\_vector\_machines`.
5. *Pevny T., Bas P., and Fridrich J.* Steganalysis by subtractive pixel adjacency matrix // IEEE Trans. Inform. Forensics Security. 2010. V. 5. No. 2. P. 215–224.
6. *Fridrich J.* Rich models for steganalysis of digital images // IEEE Trans. Inform. Forensics Security. 2012. V. 7. No. 3. P. 868–882.
7. *Holub V. and Fridrich J.* Random projections of residuals for digital image steganalysis // IEEE Trans. Inform. Forensics Security. 2013. V. 8. No. 12. P. 1996–2006.
8. *Bas P., Filler T., and Pevny T.* Break our steganographic system the ins and outs of organizing BOSS // LNCS. 2011. V. 6958. P. 59–70.

9. http://www.lirmm.fr/chaumont/PPG-LIRMM-COLOR.html — PPG-LIRMM-COLOR base.
10. *Pevny T., Filler T., and Bas P.* Using high-dimensional image models to perform highly undetectable steganography // LNCS. 2010. V. 6387. P. 161–177.
11. *Holub V. and Fridrich J.* Digital image steganography using universal distortion // Proc. 1st ACM Workshop IHMMSec. ACM, 2013. P. 59–68.
12. *Holub V. and Fridrich J.* Designing steganographic distortion using directional filters // Proc. 4th IEEE Intern. Workshop WIFS. 2012. P. 234–239.
13. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media // IEEE Trans. Inform. Forensics Security. 2010. V. 7. No. 2. P. 434–444.
14. *Монарев В. А., Пестунов А. И.* Эффективное обнаружение стеганографически скрытой информации посредством интегрального классификатора на основе сжатия данных // Прикладная дискретная математика. 2018. № 40. С. 59–71.
15. *Tabares-Soto R., Ramos-Pollan R., and Isaza G* Deep learning applied to steganalysis of digital images: A systematic review // IEEE Access. 2019. V. 7. P. 68970–68990.
16. *Qian Y., Dong J., Wang W., and Tan T.* Deep learning for steganalysis via convolutional neural networks // Proc. Int. Symp. Electron. Imag. 2015. V. 9409. Art. no. 94090J.
17. *Yedroudj M., Comby F., and Chaumont M* Yedrouj-Net: An efficient CNN for spatial steganalysis // Proc. IEEE Intern. Conf. Acoustics, Speech Signal Processing. 2018. P. 2092–2096.
18. *Zhang R., Zhu F., Liu J., and Liu G* Efficient Feature Learning and Multisizeimage Steganalysis Based on CNN. https://arxiv.org/pdf/1807.11428.pdf. 2018.
19. *Полунин А. А., Яндашевская Э. А.* Использование аппарата сверточных нейронных сетей для стегоанализа цифровых изображений // Труды ИСП РАН. 2020. Т. 32. № 4. С. 155–164.
20. *Сирота А. А., Дрюченко М. А.* Обобщенные алгоритмы сжатия изображений на фрагментах произвольной формы и их реализация с использованием искусственных нейронных сетей // Компьютерная оптика. 2015. № 5. С. 751–761.
21. *Dryuchenko M. A. and Sirota A. A.* Interpolation and masking effects of heteroassociative compressive transformations // J. Phys.: Conf. Ser. 2020. V. 1902. P. 1–10. https://iopscience.iop.org/article/10.1088/1742-6596/1902/1/012058/pdf.
22. *Дрюченко М. А., Сирота А. А.* Гетероассоциативные сжимающие преобразования цифровых изображений и их интерполирующие и маскирующие свойства // Сб. трудов Междунар. науч.-техн. конф. «Актуальные проблемы прикладной математики, информатики и механики». Воронеж, 07–09 декабря 2020. С. 312–322.
23. *Сирота А. А., Дрюченко М. А., Митрофанова Е. Ю.* Метод создания цифровых водяных знаков на основе гетероассоциативных сжимающих преобразований изображений и его реализация с использованием искусственных нейронных сетей // Компьютерная оптика. 2018. № 3. С. 483–494.
24. *Сирота А. А., Дрюченко М. А., Митрофанова Е. Ю.* Нейросетевые функциональные модели и алгоритмы преобразования информации для создания цифровых водяных знаков // Изв. вузов. Радиоэлектроника. 2015. № 1. С. 3–16.
25. *Сирота А. А., Дрюченко М. А., Иванков А. Ю.* Стегоанализ цифровых изображений с использованием методов поверхностного и глубокого машинного обучения: известные подходы и новые решения // Вестник Воронежского гос. ун-та. Сер. Системный анализ и информационные технологии. 2021. № 1. С. 33–53.
26. *Kutter M., Jordan F., and Bossen F* Digital signature of color images using amplitude modulation // Proc. SPIE. 1997. P. 518–526.

27. *Zhao J. and Koch E* Embedding robust labels into images for copyright protection // Proc. Intern. Congress Intellectual Property Rights for Specialized Information, Knowledge and New Technologies. Vienna, August 1995. P. 242–251.
28. *Zhang X. P. and Wang S. Z* Efficient steganographic embedding by exploiting modification direction // IEEE Commun. Let. 2006. V. 10. No. 11. P. 781–783.
29. *Paul G., Davidson I., Mukherjee I., and Ravi S. S.* Keyless dynamic optimal multi-bit image steganography using energetic pixels // Multimedia Tools Appl. 2017. V. 76. P. 7445–7471.
30. `http://dde.binghamton.edu/download/stego\_algorithms/` — Digital Data Embedding Laboratory Department of Electrical and Computer Engineering SUNY Binghamton.

REFERENCES

1. *Sheluhin O. I.* Steganografiya. Algoritmy i programmnaya realizatsiya. [Steganography. Algorithms and Their Implementation]. Moscow, Goryachaya liniya — Telekom, 2017. 592 p. (in Russian)
2. *Czaplewski B.* Current trends in the field of steganalysis and guidelines for constructions of new steganalysis schemes. Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne, 2017, no. 10, pp. 1121–1125.
3. *Lyu S. and Farid H.* Detecting hidden messages using higher-order statistics and support vector machines. Intern. Workshop Inform. Hiding, Berlin; Heidelberg, Springer, 2002, pp. 340–354. `https://farid.berkeley.edu/downloads/publications/ih02.pdf`.
4. *Lyu S. and Farid H.* Steganalysis using color wavelet statistics and one-class support vector machines. // Proc. SPIE, California, USA, 2004, pp. 35–45. `https://www.researchgate.net/publication/221011180\_Steganalysis\_using\_color\_wavelet\_statistics\_and\_one-class\_support\_vector\_machines`.
5. *Pevny T., Bas P., and Fridrich J.* Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. Inform. Forensics Security, 2010, vol. 5, no. 2, pp. 215–224.
6. *Fridrich J.* Rich models for steganalysis of digital images. IEEE Trans. Inform. Forensics Security, 2012, vol. 7, no. 3, pp. 868–882.
7. *Holub V. and Fridrich J.* Random projections of residuals for digital image steganalysis. IEEE Trans. Inform. Forensics Security, 2013, vol. 8, no. 12, pp. 1996–2006.
8. *Bas P., Filler T., and Pevny T.* Break our steganographic system the ins and outs of organizing BOSS. LNCS, 2011, vol. 6958, pp. 59–70.
9. `http://www.lirmm.fr/chaumont/PPG-LIRMM-COLOR.html` — PPG-LIRMM-COLOR base.
10. *Pevny T., Filler T., and Bas P.* Using high-dimensional image models to perform highly undetectable steganography. LNCS, 2010, vol. 6387, pp. 161–177.
11. *Holub V. and Fridrich J.* Digital image steganography using universal distortion. Proc. 1st ACM Workshop IHMMSec, ACM, 2013, pp. 59–68.
12. *Holub V. and Fridrich J.* Designing steganographic distortion using directional filters. Proc. 4th IEEE Intern. Workshop WIFS, 2012, pp. 234–239.
13. *Kodovsky J., Fridrich J., and Holub V.* Ensemble classifiers for steganalysis of digital media. IEEE Trans. Inform. Forensics Security, 2010, vol. 7, no. 2, pp. 434–444.
14. *Monarev V. A. and Pestunov A. I.* Effektivnoye obnaruzheniye steganograficheski skrytoy informatsii posredstvom integral'nogo klassifikatora na osnove szhatiya dannykh [Efficient steganography detection by means of compression-based integral classifier]. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 59–71. (in Russian)
15. *Tabares-Soto R., Ramos-Pollan R., and Isaza G* Deep learning applied to steganalysis of digital images: A systematic review. IEEE Access, 2019, vol. 7, pp. 68970–68990.

16. *Qian Y., Dong J., Wang W., and Tan T.* Deep learning for steganalysis via convolutional neural networks. Proc. Int. Symp. Electron. Imag., 2015, vol. 9409, Art. no. 94090J.
17. *Yedroudj M., Comby F., and Chaumont M* Yedrouj-Net: An efficient CNN for spatial steganalysis. Proc. IEEE Intern. Conf. Acoustics, Speech Signal Processing, 2018, pp. 2092–2096.
18. *Zhang R., Zhu F., Liu J., and Liu G* Efficient Feature Learning and Multisizeimage Steganalysis Based on CNN. `https://arxiv.org/pdf/1807.11428.pdf`. 2018.
19. *Polunin A. A. and Yandashevskaya E. A.* Ispol'zovaniye apparata svertochnykh neyronnykh setey dlya stegoanaliza tsifrovykh izobrazheniy [Using of convolutional neural networks for steganalysis of digital images]. Proc. ISP RAS, 2020, vol. 32, no. 4, pp. 155–164. (in Russian)
20. *Sirota A. A. and Dryuchenko M. A.* Obobshchennyye algoritmy szhatiya izobrazheniy na fragmentakh proizvol'noy formy i ikh realizatsiya s ispol'zovaniyem iskusstvennykh neyronnykh setey [Generalized image compression algorithms for arbitrarily-shaped fragments and their implementation using artificial neural networks]. Computer Optics, 2015, no. 5, pp. 751–761. (in Russian)
21. *Dryuchenko M. A. and Sirota A. A.* Interpolation and masking effects of heteroassociative compressive transformations. J. Phys.: Conf. Ser., 2020, vol. 1902, pp. 1–10. `https://iopscience.iop.org/article/10.1088/1742-6596/1902/1/012058/pdf`.
22. *Dryuchenko M. A. and Sirota A. A.* Geteroassotsiativnye szhimayushchie preobrazovaniya tsifrovykh izobrazheniy i ikh interpoliruyushchie i maskiruyushchie svoystva [Interpolation and masking effects of heteroassociative compressive transformations]. Proc. Intern. Conf. AMCSM, Voronezh, 07–09 December 2020, pp. 312–322. (in Russian)
23. *Sirota A. A., Dryuchenko M. A., Mitrofanova E. Yu.* Metod sozdaniya tsifrovykh vodyanykh znakov na osnove geteroassotsiativnykh szhimayushchikh preobrazovaniy izobrazheniy i ego realizatsiya s ispol'zovaniem iskusstvennykh neyronnykh setey [Digital watermarking method based on heteroassociative image compression and its realization with artificial neural networks]. Computer Optics, 2018, no. 3, pp. 483–494. (in Russian)
24. *Sirota A. A., Dryuchenko M. A., and Mitrofanova E. Yu.* Neyrosetevye funktsional'nye modeli i algoritmy preobrazovaniya informatsii dlya sozdaniya tsifrovykh vodyanykh znakov [Neural network functional models and information transformation algorithms for creating digital watermarks]. Radioelectronics and Communications Systems, 2015, no. 1, pp. 3–16. (in Russian)
25. *Sirota A. A., Dryuchenko M. A., Ivankov A. Yu.* Stegoanaliz tsifrovykh izobrazheniy s ispol'zovaniem metodov poverkhnostnogo i glubokogo mashinnogo obucheniya: izvestnye podkhody i novye resheniya [Steganalysis of digital images by means of shallow and deep machine learning: existing approaches and new solutions]. Proc. Voronezh State University, Ser. Systems Analysis and Inform. Technol., 2021, no. 1, pp. 33–53. (in Russian)
26. *Kutter M., Jordan F., and Bossen F* Digital signature of color images using amplitude modulation. Proc. SPIE, 1997, pp. 518–526.
27. *Zhao J. and Koch E* Embedding robust labels into images for copyright protection. Proc. Intern. Congress Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, August 1995, pp. 242–251.
28. *Zhang X. P. and Wang S. Z* Efficient steganographic embedding by exploiting modification direction. IEEE Commun. Let., 2006, vol. 10, no. 11, pp. 781–783.
29. *Paul G., Davidson I., Mukherjee I., and Ravi S. S.* Keyless dynamic optimal multi-bit image steganography using energetic pixels. Multimedia Tools Appl., 2017, vol. 76, pp. 7445–7471.
30. `http://dde.binghamton.edu/download/stego\_algorithms/` — Digital Data Embedding Laboratory Department of Electrical and Computer Engineering SUNY Binghamton.