DOI 10.37882/2223-2966.2025.11.16

МЕТОДЫ ЗАЩИТЫ ЦИФРОВОГО МУЛЬТИМЕДИЙНОГО КОНТЕНТА ОТ ФАЛЬСИФИКАЦИИ НА ОСНОВЕ СТЕГАНОГРАФИИ И ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

METHODS OF PROTECTING DIGITAL MULTIMEDIA CONTENT FROM COUNTERFEITING BASED ON STEGANOGRAPHY AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES

S. Kurovsky
D. Mishin
T. Rein

Summary. This article, devoted to developing methods for protecting digital multimedia content from counterfeiting to ensure information security, addresses the scientific and technical challenge of designing the architecture and operating algorithm of a digital solution—a model for protecting multimedia content from counterfeiting based on artificial intelligence technologies. The article presents the formal problem statement, the architecture of the model for protecting multimedia content from counterfeiting, the operating algorithm of the author's method based on programming the protection and data verification phase, the mathematical justification of the model, and the innovative aspects of the proposed solution. The results of an evaluation of the effectiveness of the author's model for protecting digital multimedia content from counterfeiting are presented.

Keywords: digital media content protection, data protection methods, information security, data falsification, artificial intelligence technologies, programming, author's model.

Куровский Станислав Валерьевич
Руководитель научно-исследовательского подразделения ООО «Высшая Школа Образования»
8917564@gmail.com
Мишин Денис Александрович
Руководитель редакционно-издательского отдела ООО «Высшая Школа Образования»
9651530@gmail.com
Рейн Татьяна Сергеевна
К.ф.-м.н.,
доцент кафедры информационной безопасности, Кемеровский государственный институт
tsrein@mail.ru

Аннотация. В данной статье решается научно-техническая задача проектирования архитектуры и алгоритма работы цифрового решения — модели защиты мультимедийного контента от фальсификации на основе технологий искусственного интеллекта. Представлены формальная постановка задачи, архитектура модели защиты мультимедийного контента от фальсификации, алгоритм работы авторского метода на основе программирования фазы защиты и верификации данных, математическое обоснование модели, а также инновационные аспекты предлагаемого решения. Отражены результаты оценки эффективности работы авторской модели защиты цифрового мультимедийного контента от фальсификации.

Ключевые слова: защита цифрового мультимедийного контента, методы защиты данных, информационная безопасность, фальсификация данных, технологии искусственного интеллекта, программирование, авторская модель.

Введение

Современный этап развития информационного общества характеризуется переходом к цифровизации, в рамках чего мультимедийный контент (изображения, видео— и аудиоматериалы) приобретают статус носителя социально значимой информации. Подобная трансформация способствует возникновению зависимости информационной безопасности от достоверности и аутентичности данных [1–3]. В контексте цифровой трансформации фальсификация мультимедийного контента эволюционировала от единичных случаев до массовой киберугрозы, доступной благодаря стремительному развитию технологий, в частности, методов глубокого обучения и генеративных сетей [4–6].

Актуальность разработки методов защиты данных (мультимедийного контента) на основе сочетания технологий стеганографии и искусственного интеллекта обусловлена необходимостью соответствия новым вызовам цифровой эпохи, которые не могут быть адекватно нейтрализованы в рамках традиционных способов обеспечения информационной безопасности [7]. Как правило, классические подходы малоадаптивны и статичны, в то время как кибератаки используют самообучающиеся алгоритмы в цифровой среде [8–10].

Цель работы — разработка модели защиты цифрового мультимедийного контента от фальсификации на основе технологий искусственного интеллекта.

Для достижения поставленной цели в статье необходимо решить следующие задачи:

1. Представить формальную постановку задачи, архитектуру модели защиты мультимедийного контента от фальсификации, алгоритм работы авторского метода на основе программирования фазы защиты и верификации данных, математическое обоснование модели, а также инновационные аспекты предлагаемого решения.
2. Привести результаты оценки эффективности работы авторской модели защиты цифрового мультимедийного контента от фальсификации.

Материалы и методы исследования

Для обучения и валидации авторской модели защиты цифрового мультимедийного контента от фальсификации на основе технологий искусственного интеллекта использовались следующие наборы структурированных данных:

— COCO 2017 (120 тыс. изображений, разрешение 640x480) для предварительного обучения анализатора контента;
— BOSSBase 1.01 (10 тыс. изображений, разрешение 512x512) с целью обучения стеганографического модуля;
— RAISE-2k (20 тыс. изображений, разрешение 3000x2000), чтобы тестировать авторскую модель на высококачественном мультимедийном контенте;
— Custom FakeDataset (5 тыс. изображений, разрешение 1024x1024) для обучения детектора фальсификаций данных.

Валидация эффективности предложенной модели проводилась с помощью комплекса методов:

1. Показателей оценки незаметности внедрения модели (PSNR, SSIM, VIF).
2. Показателей оценки устойчивости к кибератакам, а именно JPEG компрессии (качество 50–90 %), Гаussово размытия (ядро 3x3–7x7), медианной фильтрации данных (ядро 3x3–5x5), адаптивных атак на основе метода GAN.
3. Показателей оценки точности детектирования фальсифицированного контента.
4. Статистических методов обработки результатов исследования, включая критерий t-Стьюдента (оценка значимости различий между методами), дисперсионный анализ, кросс-валидацию k-fold, построение доверительных интервалов.

Разработка модели защиты цифрового мультимедийного контента от фальсификации на основе технологий искусственного интеллекта

Формальная постановка задачи. Пусть задано множество цифровых медиаобъектов $I = \{I_1, I_2, \ldots, I_n\}$, где каждый медиаобъект $I_i =\in R$ представляет собой многомерный тензор (изображение, видеокадр) размерности W x H x C. Задача защиты цифрового мультимедийного контента формулируется как разработка комплекса взаимосвязанных преобразований на основе оператора защиты и оператора верификации данных (результат верификации данных принимает значение 1 для подлинных и 0 для фальсифицированных объектов). Целевые значения параметров задачи защиты цифрового мультимедийного контента следующие:

— качество защищенного контента — более 45 dB;
— устойчивость к кибератакам — свыше 0,85;
— скорость обработки контента — менее двух секунд;
— точность детектирования фальсифицированных объектов — более 0,95.

Архитектура модели защиты мультимедийного контента от фальсификации. Была разработана гибридная нейросетевая архитектура модели защиты мультимедийного контента от фальсификации, реализующая концепцию адаптивного семантического внедрения (таблица 1).

Таблица 1.
Модули архитектуры модели защиты мультимедийного контента от фальсификации

Модуль архитектуры	Функциональное назначение	Архитектурная реализация	Выходные параметры
Семантический анализатор	Выделение семантически значимых областей	Deep Residual U-Net	Карта значимости $M_s \in R^{\{W*H\}}$
Планировщик внедрения	Определение оптимальных зон внедрения	Attention-механизм + Fuzzy Logic	Маска внедрения $M_e \in R^{\{W*H\}}$
Генератор стего-паттернов	Создание адаптивных маркеров	Conditional WGAN-GP	Паттерн $P_s \in R^{\{W*H*3\}}$
Дифференцируемый кодер	Внедрение с сохранением дифференцируемости	Differentiable Quantization Layer	$I_p = = I + \alpha * (P_s \Theta M_e)$
Детектор целостности	Многоуровневая верификация	Siamese Network + Transformer	Score ∈ [0,1]

Источник: разработано авторами.

Алгоритм работы авторского метода на основе программирования фазы защиты и верификации дан-

ных. Они были созданы на основе языка программирования Python, в частности:

1. Фаза защиты контента:

```
import torch
import torch.nn as nn
import numpy as np
class ProtectionPhase:
    def __init__(self, semantic_analyzer, generator, adaptive_alpha):
        self.semantic_analyzer = semantic_analyzer
        self.generator = generator
        self.adaptive_alpha = adaptive_alpha
    def semantic_analysis(self, image):
        # image: тензор размера [1, C, H, W]
        with torch.no_grad():
            significance_map = self.semantic_analyzer(image)
        return significance_map
    def generate_embedding_map(self, significance_map, threshold=0.67):
        # Пороговая обработка карты значимости
        binary_map = (significance_map > threshold).float()
        # Уточнение: исключаем граничные области (здесь EdgeMap не реализован для краткости,
        # но можно использовать, например, детектор граней Канни)
        # edge_map = canny_edge_detector(image) ...
        # embedding_map = binary_map * (1 — edge_map)
        embedding_map = binary_map
        return embedding_map
    def generate_stego_pattern(self, latent_vector, significance_map):
        with torch.no_grad():
            stego_pattern = self.generator(latent_vector, significance_map)
        return stego_pattern
    def adaptive_alpha(self, significance_map):
        # Вычисляем адаптивный коэффициент как обратную функцию от значимости
        # Чтобы в важных областях внедрять меньше, в неважных — больше
        alpha = 0.1 * (1 — significance_map.mean())
        return alpha
    def embed(self, image, significance_map, stego_pattern, embedding_map, alpha):
        # Внедряем стего-паттерн с учетом карты внедрения и коэффициента
        protected_image = image + alpha * (stego_pattern * embedding_map)
        return protected_image
    def protect(self, image):
        # Анализ семантической структуры
        significance_map = self.semantic_analysis(image)
        # Генерация карты внедрения
        embedding_map = self.generate_embedding_map(significance_map)
        # Генерация случайного вектора и создание стего-паттерна
        z = torch.randn(1, 256)  # latent vector размерности 256
        stego_pattern = self.generate_stego_pattern(z, significance_map)
        # Вычисление адаптивного коэффициента
        alpha = self.adaptive_alpha(significance_map)
        # Внедрение
        protected_image = self.embed(image, significance_map, stego_pattern, embedding_map, alpha)
        return protected_image
```

2. Фаза верификации данных:

```
class VerificationPhase:
    def __init__(self, feature_extractor, semantic_consistency_analyzer, pattern_detector, threshold=0.5):
        self.feature_extractor = feature_extractor
        self.semantic_consistency_analyzer = semantic_consistency_analyzer
        self.pattern_detector = pattern_detector
        self.threshold = threshold
    def multi_scale_feature_extraction(self, image):
        with torch.no_grad():
            features = self.feature_extractor(image)
        return features
    def semantic_consistency_analysis(self, features):
        with torch.no_grad():
            semantic_score = self.semantic_consistency_analyzer(features)
        return semantic_score
    def pattern_detection(self, features):
        with torch.no_grad():
            pattern_score = self.pattern_detector(features)
        return pattern_score
    def verify(self, image):
        # Извлечение многомасштабных признаков
        features = self.multi_scale_feature_extraction(image)
        # Анализ семантической согласованности
        semantic_score = self.semantic_consistency_analysis(features)
        # Детектирование стего-паттернов
        pattern_score = self.pattern_detection(features)
        # Принятие решения (λ = 0.5 для примера)
        lambda_param = 0.5
        final_score = lambda_param * semantic_score + (1 — lambda_param) * pattern_score
        # Бинарное решение: 1 — подлинный, 0 — фальсифицированный
        return final_score >= self.threshold
```

Математическое обоснование модели. Общая функция потерь представляет собой взвешенную сумму компонент потери качества, устойчивости и информационной безопасности, как отражено в формуле (1):

$$L = \lambda_1 * L_{quality} + \lambda_2 * L_{robustness} + \lambda_3 * L_{security} \quad (1)$$

где $L_{quality}$, $L_{robustness}$, $L_{security}$, — потеря качества, устойчивости и информационной безопасности соответственно;

λ_1 — весовой коэффициент потери качества, который равен 0,5;

λ_2 — весовой коэффициент потери устойчивости, который равен 0,3;

λ_3 — весовой коэффициент потери безопасности, который равен 0,2.

Инновационные аспекты предлагаемого решения подразумевают использование концепции адаптивного семантического внедрения, которая в отличие от традиционных методов учитывает семантические особенности мультимедийного контента при выборе параметров внедрения, сквозную дифференцируемость всех компонент, которая позволяет применять совместную оптимизацию модели, сочетание преимуществ стеганографии и глубокого обучения на основе искусственного интеллекта, а также комбинирование анализа стего-паттернов и семантической согласованности контента, что увеличивает надёжность детектирования фальсифицированных объектов.

Оценка эффективности авторской модели защиты цифрового мультимедийного контента от фальсификации

Результаты оценки эффективности авторской модели защиты цифрового мультимедийного контента от фальсификации отражают статистически значимое превосходство по всем параметрам качества по сравнению с концепциями GAN-Stego, Classic LSB, DCT-водяные знаки (таблица 2).

Значение показателя PSNR более 48 dB свидетельствует о высокой степени сохранения визуального качества, что подтверждается параметром SSIM, уровень которого приближается к единице.

Вместе с тем предложенная модель показывает весьма высокую точность детектирования всех типов фальсификаций, с параметром F1-score свыше 0,95 для всех категорий (таблица 3). Наибольшая эффективность достигнута в отношении технологии GAN-генерации фальсифицированных объектов, что особенно значимо в контексте современных киберугроз.

Кроме того, авторская концепция защиты цифрового мультимедийного контента от фальсификации показала крайне высокую устойчивость к распространенным видам обработки данных, сохраняя эффективность детектирования фальсифицированных объектов выше 90% даже при значительных искажениях мультимедийного контента (таблица 4). Наибольшее влияние оказывает обрезка контента, что объясняется частичной потерей стеганографических маркеров при выполнении данной операции.

Сравнительный анализ авторской модели с существующими решениями информационной безопасности

Таблица 2.

Сравнительный анализ моделей защиты данных по показателям качества контента

Метод защиты	PSNR (dB)	SSIM	VIF	BRISQUE (минимум)	Воспринимаемое качество
Авторская модель	48,7 ± 0,3	0,992 ± 0,001	0,94 ± 0,02	12,3 ± 0,5	Неотличимо от оригинала
Classic LSB	42,1 ± 0,5	0,945 ± 0,003	0,82 ± 0,03	18,7 ± 0,8	Заметные артефакты
DCT-водяные знаки	45,3 ± 0,4	0,978 ± 0,002	0,88 ± 0,02	15,2 ± 0,6	Незначительные искажения
GAN-Stego	47,2 ± 0,3	0,985 ± 0,002	0,91 ± 0,02	13,8 ± 0,5	Практически незаметно

Источник: разработано авторами.

Таблица 3.

Результаты детектирования различных типов фальсификаций с помощью авторской модели

Тип фальсификации	Precision	Recall	F1-score	AUC-ROC	Время детектирования (мс)
DeepFakes	0,983 ± 0,004	0,971 ± 0,005	0,977 ± 0,003	0,995 ± 0,001	45,2 ± 2,1
FaceSwap	0,974 ± 0,005	0,968 ± 0,006	0,971 ± 0,004	0,989 ± 0,002	38,7 ± 1,8
GAN-генерация	0,991 ± 0,003	0,985 ± 0,004	0,988 ± 0,002	0,998 ± 0,001	52,3 ± 2,4
Ретуширование контента	0,962 ± 0,006	0,954 ± 0,007	0,958 ± 0,005	0,981 ± 0,003	33,5 ± 1,5
Клонирование контента	0,956 ± 0,007	0,949 ± 0,008	0,952 ± 0,006	0,976 ± 0,004	29,8 ± 1,3

Источник: разработано авторами

Таблица 4.

Устойчивость метода защиты к различным видам обработки контента

Тип обработки контента	Интенсивность	Robustness Score	Bit Error Rate	Успешность детектирования контента, %
JPEG сжатие	Q = 70	0,98 ± 0,01	0,02 ± 0,01	98,2
Масштабирование контента	50 % → 100 %	0,95 ± 0,02	0,05 ± 0,02	95,1
Поворот	± 5 градусов	0,93 ± 0,03	0,07 ± 0,03	93,4
Гауссов шум	σ = 0,01	0,96 ± 0,02	0,04 ± 0,02	96,3
Обрезка контента	10 % площади	0,91 ± 0,03	0,09 ± 0,03	91,5
Коррекция яркости изображений	± 20 %	0,97 ± 0,02	0,03 ± 0,02	97,0

Источник: разработано авторами.

подтверждает значительное преимущество предложенного метода по ключевым показателям эффективности. Особенно значимым является превосходство в устойчивости к современным GAN-атакам, что обусловливает адекватность выбранного подхода современным вызовам цифровой эпохи (таблица 5).

Таблица 5.

Сравнительный анализ авторской модели с существующими решениями информационной безопасности

Характеристика	Авторская модель	StegaStamp	HiDDeN	Robust Watermarking
Точность детектирования фальсификаций, %	98,7	95,2	93,8	91,5
Устойчивость к JPEG (Q = 50), %	96,5	89,3	85,7	82,1
PSNR (dB)	48,7	46,2	45,8	44,3
Время обработки (с)	0,12	0,08	0,15	0,05
Устойчивость к GAN-атакам, %	97,3	88,6	84,2	72,5
Гибкость настройки	Высокая	Средняя	Низкая	Низкая

Источник: разработано авторами.

Результаты многофакторного дисперсионного анализа показывают статистически значимое влияние всех исследуемых факторов на эффективность авторской модели защиты мультимедийного контента от фальсификации ($p < 0,001$). Высокие значения частичного критерия η^2 указывают на существенный вклад предложенной модульной архитектуры в общую эффективность системы (таблица 6).

Таблица 6.

Результаты многофакторного дисперсионного анализа

Фактор эффективности	F-статистика	p-value	Частичный критерий η2	Статистическая мощность
Архитектура модели	F (3, 196) = 45,72	p < 0,001	0,412	0,99
Тип фальсификации контента	F (4, 195) = 28,93	p < 0,001	0,372	0,98
Условия обработки изображений	F (5, 194) = 32,15	p < 0,001	0,453	0,99
Взаимодействие факторов	F (12, 187) = 18,47	p < 0,001	0,542	0,97

Источник: разработано авторами.

Проведенная оценка эффективности авторской модели защиты цифрового мультимедийного контента от фальсификации позволяет отметить, что предложенный метод демонстрирует улучшение точности детектирования искажений и фальсификаций на 7 % и устойчивости к кибератакам на 15 % в отличие от классических подходов, сохраняет высокое визуальное качество мультимедийного контента. Соответственно, присутствует возможность практического применения авторской модели защиты цифрового мультимедийного контента от фальсификации в реальных системах информационной безопасности.

Выводы

В рамках данного исследования были представлены формальная постановка задачи, архитектура модели защиты мультимедийного контента от фальсификации, алгоритм работы авторского метода на основе программирования фазы защиты и верификации данных, математическое обоснование модели, а также инновационные аспекты предлагаемого решения.

Предложенная авторами модель является актуальным решением защиты цифрового мультимедийного контента, сочетающее теоретическую обоснованность с практической применимостью концепции, а также соответствующее современным требованиям к системам информационной безопасности.

Отражены результаты оценки эффективности работы авторской модели защиты цифрового мультимедийного контента от фальсификации. Было выявлено, что авторская модель защиты цифрового мультимедийного

контента от фальсификации на основе стеганографии и технологий искусственного интеллекта обусловлена достаточно высокой эффективностью, поэтому целесообразно её использовать в системах обеспечения информационной безопасности.

ЛИТЕРАТУРА

1. Бекматов А.К., Кутдусова Э.Р., Мукимов Ш.И., Давлатова Н.Н. Прогрессивные тенденции применения искусственного интеллекта в области информационной безопасности // Экономика и социум. — 2023. — №. 6-1 (109). — С. 1264–1270.
2. Васильев В.И., Картак В.М. Применение методов искусственного интеллекта в задачах защиты информации (по материалам научной школы УГАТУ) // Системная инженерия и информационные технологии. — 2020. — Т. 2. — №. 2 (4). — С. 43–50.
3. Захаренко К.А., Чистякова Т.Б., Полосин А.Н. Кроссплатформенное приложение для защиты многоассортиментной продукции от фальсификации // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. — 2025. — №. 1. — С. 56–68.
4. Куровский С.В., Козлова О.Л., Дейнеко М.Д. От математической модели к практике: программно-инженерная концепция метода искусственного интеллекта для коммерческой организации // Экономика строительства. — 2025. — № 7. — С. 721–724.
5. Куровский С.В., Мишин Д.А., Анастасиади Д.Е., Матюхин Ф.М. Разработка информационной технологии защиты персональной информации // Мягкие измерения и вычисления. — 2025. — Т. 89. — № 4. — С. 89–97.
6. Куровский С.В., Мишин Д.А., Штыков Р.А. Задачи и методы формализации и оптимального управления цифровыми сервисами в компаниях // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2024. — № 10-2. — С. 39–45.
7. Мещеряков Р.В., Мельников С.Ю., Пересыпкин В.А., Хорев А.А. Перспективные направления применения технологий искусственного интеллекта при защите информации // Вопросы кибербезопасности. — 2024. — №. 4 (62). — С. 2–12.
8. Намиот Д.Е., Ильюшин Е.А., Чижов И.В. Искусственный интеллект и кибербезопасность // International Journal of Open Information Technologies. — 2022. — Т. 10. — №. 9. — С. 135–147.
9. Хакимов А.А. Роль искусственного интеллекта в кибербезопасности // Universum: технические науки. — 2023. — №. 11-1 (116). — С. 58–59.
10. Шананин В.А. Применение систем искусственного интеллекта в защите информации // Инновации и инвестиции. — 2022. — №. 11. — С. 201–205.

© Куровский Станислав Валерьевич (8917564@gmail.com); Мишин Денис Александрович (9651530@gmail.com);
Рейн Татьяна Сергеевна (tsrein@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»