2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

doi 10.26089/NumMet.v27r218 УДК 004.93'1

Распознавание фальшивых видео с подменой лица с помощью учета направления взгляда человека

М. Д. Красильников
Московский государственный университет имени М. В. Ломоносова,
факультет вычислительной математики и кибернетики,
Москва, Российская Федерация
ООО "Технологии Видеоанализа",
Москва, Российская Федерация
ORCID: 0009-0009-3136-488X, e-mail: maximkras505@gmail.com

М. Ю. Никитин
ООО "Технологии Видеоанализа",
Москва, Российская Федерация
ORCID: 0009-0001-5676-5569, e-mail: mikhail.nikitin@tevian.ru

А. С. Конушин
Московский государственный университет имени М. В. Ломоносова,
факультет вычислительной математики и кибернетики,
Москва, Российская Федерация
ORCID: 0000-0002-6152-0021, e-mail: konushinas@my.msu.ru

Аннотация: Актуальность задачи распознавания поддельных видеозаписей лица обусловлена быстрым ростом качества средств синтеза и замены лица, вследствие чего традиционные признаки подделки становятся менее устойчивыми при переносе на новые методы генерации и условия съемки. Распространенные подходы, основанные на анализе отдельных кадров, как правило, опираются на локальные визуальные артефакты, которые могут исчезать после сжатия, повторного кодирования и постобработки, а также требуют совпадения распределений обучающих и прикладных данных. Такие методы не используют информацию о поведении человека во времени и не задействуют сценарии взаимодействия с экраном. В настоящей работе предлагается кооперативный подход, в котором признаком подделки служат параметры калибровки модели оценки направления взгляда, вычисляемые по видеотреку, полученному при последовательной фиксации взгляда на предъявляемых на экране точках. На основе этих параметров обучается простой классификатор, а также рассматривается объединение этих признаков с выходами стандартных детекторов фальшивых видео. Показано, что одних параметров калибровки достаточно для уверенного разделения подлинных и синтезированных треков и в ряде настроек они превосходят детекторы, обученные только на визуальных признаках, тогда как их объединение с выходами таких детекторов обеспечивает дополнительный прирост качества. Анализ значимости признаков указывает на преобладающий вклад параметров, связанных с вертикальной составляющей направления взгляда, что согласуется с предположением о повышенной уязвимости генераторов при синтезе реалистичного вида глаз на экстремальных отклонениях вверх и вниз.

Ключевые слова: фальшивые видео, распознавание фальшивых видео, подмена лица, оценка направления взгляда, кооперативная проверка подлинности, компьютерное зрение, машинное обучение.

Для цитирования: Красильников М.Д., Никитин М.Ю., Конушин А.С. Распознавание фальшивых видео с подменой лица с помощью учета направления взгляда человека // Вычислительные методы и программирование. 2026. **27**, № 2. 274–289. doi 10.26089/NumMet.v27r218.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

Detection of face-swapped fake videos using human gaze information

Maksim D. Krasilnikov
Lomonosov Moscow State University,
Faculty of Computational Mathematics and Cybernetics,
Moscow, Russia
LLC Tevian,
Moscow, Russia
ORCID: 0009-0009-3136-488X, e-mail: maximkras505@gmail.com

Mikhail Yu. Nikitin
LLC Tevian,
Moscow, Russia
ORCID: 0009-0001-5676-5569, e-mail: mikhail.nikitin@tevian.ru

Anton S. Konushin
Lomonosov Moscow State University,
Faculty of Computational Mathematics and Cybernetics,
Moscow, Russia
ORCID: 0000-0002-6152-0021, e-mail: konushinas@my.msu.ru

Abstract: The relevance of fake face video detection is driven by the rapid improvement of face synthesis and face swapping techniques, which makes traditional forgery cues less robust when transferred to new generation methods and recording conditions. Common approaches based on the analysis of individual frames usually rely on local visual artifacts that may disappear after compression, re-encoding, and post-processing, and they also require a close match between the distributions of training and deployment data. Such methods do not exploit information about human behavior over time and do not make use of screen interaction scenarios. In this paper, we propose a cooperative approach in which the forgery cue is given by the calibration parameters of a gaze estimation model computed from a video track obtained while a subject sequentially fixates on points displayed on the screen. A simple classifier is trained using these parameters, and their fusion with the outputs of standard fake video detectors is also considered. It is shown that the calibration parameters alone are sufficient for reliable separation of authentic and synthesized tracks and, in several settings, outperform detectors trained only on visual features, while their fusion with the outputs of such detectors provides an additional gain in accuracy. Feature importance analysis indicates a dominant contribution of the parameters associated with the vertical component of gaze direction, which is consistent with the assumption that current generators are more vulnerable when synthesizing realistic eye appearance at extreme upward and downward gaze angles.

Keywords: fake videos, fake video detection, face swapping, gaze estimation, cooperative authenticity verification, computer vision, machine learning.

For citation: M. D. Krasilnikov, M. Yu. Nikitin, A. S. Konushin, “Detection of face-swapped fake videos using human gaze information,” Numerical Methods and Programming. **27** (2), 274–289 (2026). doi 10.26089/NumMet.v27r218.

1. Введение. Бурное развитие генеративных нейросетевых технологий привело к широкому распространению реалистичных фальшивых видеозаписей лица. Такие подделки применяются не только в развлекательных целях, но и в задачах, связанных с социальной инженерией, дискредитацией и обходом процедур удаленной идентификации. В связи с этим возрастает потребность в надежных методах распознавания поддельного видео, устойчивых к различиям в устройствах съемки, освещении, сжатии и способах генерации.

Большинство существующих методов основано на поиске артефактов синтеза, анализе текстур, мимики и пространственно-временны́х закономерностей. Однако по мере совершенствования генераторов

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

такие признаки становятся менее выраженными, а переносимость методов на новые типы подделок и новые условия съемки остается ограниченной. Особенно трудной является ситуация, когда обучающие данные заметно отличаются от целевых видеозаписей по сцене, камере, обработке и поведению человека. Поэтому представляет интерес использование более устойчивых признаков, связанных не только с изображением, но и с реакцией человека на контролируемое воздействие.

В кооперативных и активных схемах проверки подлинности системе разрешается предъявлять пользователю заранее заданное воздействие и анализировать отклик. В ряде работ показано, что управляемое изменение цветовых характеристик изображения на экране вызывает измеримую реакцию в отражении от лица и позволяет проверять согласованность между ожидаемой и наблюдаемой динамикой. Такой подход переносит акцент с поиска дефектов синтеза на анализ связи между заданным воздействием и наблюдаемым ответом.

В настоящей работе рассматривается активная схема, основанная на контролируемом движении взгляда. Пользователю предлагается последовательно фиксировать взгляд на точках на экране, после чего по кадрам лица оцениваются углы направления взгляда. Для перевода этих углов в координаты экрана выполняется калибровка под конкретную сцену, в результате которой вычисляется компактный набор параметров, отражающих геометрию наблюдения и характер ошибок оценки направления взгляда в данной записи. Выдвигается гипотеза, что для поддельных видеозаписей распределения таких параметров систематически отличаются от распределений для подлинных записей вследствие менее точного воспроизведения области глаз, особенно в кадрах, соответствующих взгляду вниз. Иллюстрация, мотивирующая данную гипотезу, приведена на рис. 1. Можно заметить, что в синтезированных последовательностях при

Рис. 1. Иллюстрация гипотезы о том, что синтезированные видеозаписи менее точно воспроизводят область глаз при переводе взгляда, особенно в кадрах, соответствующих взгляду вниз. Для одной и той же последовательности шагов калибровки показаны исходные кадры и синтезированные варианты, полученные двумя методами подмены лица. Можно заметить, что в нижних положениях взгляда синтезированные изображения передают направление взгляда иначе, чем исходная запись

Fig. 1. Illustration of the hypothesis that synthesized videos reproduce the eye region less accurately during gaze shifts, especially in frames corresponding to downward gaze. For the same sequence of calibration steps, the original frames and two face-swapped variants are shown. In the lower gaze positions, the synthesized images reproduce the gaze direction differently from the original recording

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

переходе к кадрам со взглядом вниз направление взгляда воспроизводится иначе, чем в исходной записи, т.е. нижние направления фиксации передаются менее естественно и менее согласованно.

Целью работы является разработка и экспериментальная проверка метода распознавания поддельных видеозаписей лица в кооперативном сценарии взаимодействия пользователя с экраном на основе параметров калибровки направления взгляда, вычисляемых по последовательностям фиксаций на экранных метках, а также исследование совместимости этого метода с существующими детекторами поддельного видео. Задачами данной работы являются:

1) предложить метод кооперативной проверки подлинности видеозаписи лица в кооперативном сценарии, использующий параметры калибровки отображения предсказанных углов направления взгляда в экранные координаты как трековые признаки подлинности;
2) экспериментально показать, что одних параметров калибровки достаточно для разделения подлинных и поддельных видеотреков при различных сочетаниях обучающих данных и моделей, а их объединение с выходами пассивных классификаторов повышает итоговое качество;
3) проверить гипотезу, что наибольший вклад в распознавание вносят параметры, связанные с вертикальным отклонением взгляда, что указывает на характерную слабость генераторов при воспроизведении области глаз для крайних движений вверх-вниз.

2. Обзор литературы.

2.1. Определение направления взгляда.

§ 2.1.1. Существующие наборы данных для задачи определения направления взгляда. Задача определения направления взгляда существенно зависит от способа сбора и разметки данных, поскольку именно они во многом определяют точность и переносимость методов. Для получения двумерной точки фиксации на экране обычно используют сценарии взаимодействия с интерфейсом, когда испытуемому предъявляют метки на дисплее и сохраняют координаты вместе с изображением лица. Такой протокол позволяет собирать крупные выборки при сравнительно низких затратах, что показано, например, для GazeCapture в работе по методу ITracker [1]. В наборе GazeT [2] дополнительно предусмотрены калибровочные сессии с парами “изображение лица–экранная точка”, используемыми для настройки преобразования от углов взгляда к координатам экрана.

Отдельные наборы данных ориентированы на разнообразие условий съемки или более точную разметку. В MPIIGaze [3] данные собирались в течение многих дней и в разных условиях, что позволяет изучать влияние освещения и контекста. В EYEDIAP [4] используются синхронизированные RGB и RGB-D сенсоры и контролируемое освещение, что обеспечивает высокую точность, но ограничивает разнообразие сценариев. В ETH-XGaze [5] акцент сделан на широкий диапазон поз головы и направлений взгляда, а Gaze360 [6] собран в естественных условиях и ориентирован на устойчивую оценку взгляда вне лабораторной среды.

§ 2.1.2. Методы определения направления взгляда. Классические методы оценки направления взгляда опирались на геометрические модели глаза и лица [7]. Сходные подходы применялись и в прикладных задачах, например при анализе поведения водителя, где заранее задаются допустимые конфигурации глаза и головы [8].

С развитием нейросетевых методов задача сначала рассматривалась как классификация дискретных направлений [9], а затем получили распространение модели, напрямую оценивающие координаты точки фиксации или углы отклонения взгляда. В ITracker [1] для этого используются изображения глаз, лица и маска положения лица в кадре, а в ряде работ применяются раздельные ветви обработки глаз и сцены [10].

Для оценки трехмерного вектора взгляда используются более сложные архитектуры, учитывающие связь между позой головы и направлением фиксации [11]. Отдельные работы направлены на повышение устойчивости к сложному освещению [12], неравномерности ошибок по диапазону углов [13] и различиям между пользователями и устройствами за счет персонализации [14].

2.2. Распознавание фальшивых видео.

§ 2.2.1. Наборы данных для распознавания поддельных видео. По мере роста качества подделок возникла потребность в репрезентативных наборах данных для сравнения методов обнаружения и проверки их переносимости на новые способы генерации. Одним из первых стал FaceForensics++ [15], задавший базовый протокол обучения и сравнения методов на нескольких типах манипуляций. Более высокое качество подделок отражено в Celeb-DF [16], где особое внимание уделено правдоподобности и уменьшению

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

заметных артефактов. Масштабный набор данных DFDC [17] расширил разнообразие сцен, устройств и условий съемки, приблизив задачу к реальным применениям. Для оценки обобщающей способности в менее контролируемых условиях используются видео из интернета, например Deepfakes in the Wild [18], а более широкий набор видов синтеза и подмены представлен в ForgeryNet [19].

§ 2.2.2. Методы распознавания поддельных видео. Большинство современных методов обнаружения поддельных видео основано на нейросетевых классификаторах, распознающих следы генерации. Ранние работы, включая MesoNet [20], опирались на локальные текстурные и структурные искажения, а позднее появились подходы с капсульными сетями, чувствительными к нарушениям геометрии [21].

В F3Net [22] используются частотное разложение и статистики локальных компонентов, в FFD [23] — локализация наиболее вероятных областей подделки, в ряде работ — приемы стеганоанализа и анализа шумовых следов, включая фильтры SRM [24]. В SPSL [25] акцент сделан на фазовом спектре и локальной структуре.

Для повышения переносимости предлагаются методы, отделяющие общие признаки подделки от особенностей конкретного источника данных. В UCF [26] это достигается разделением универсальных и доменных признаков, а в Recce [27] — сравнением с распределением подлинных лиц, восстанавливаемым реконструктивной сетью. Однако и такие методы нередко теряют качество при появлении новых способов синтеза и в условиях видеосвязи со сжатием и нестабильным освещением.

§ 2.2.3. Кооперативные методы распознавания поддельных видео. Пассивные методы обнаружения подделок опираются на следы генерации и потому часто теряют надежность при смене условий съемки, сжатии и появлении новых способов синтеза, особенно в видеосвязи и при подмене лица в реальном времени. Поэтому развивается направление кооперативных методов, в которых система задает контролируемый стимул и проверяет согласованность между воздействием и откликом.

К таким методам относятся активная подсветка со стороны экрана с анализом реакции лица [28], а также подходы, использующие отражения в роговице глаза [29]. Другая линия работ основана на интерактивных проверках с непредсказуемыми заданиями, которые позволяют выявлять ошибки подделки в реальном времени, как показано в GOTCHA [30]. В целом кооперативные методы вводят дополнительный канал проверки, потенциально более устойчивый к смене домена, чем пассивный анализ видеоряда.

3. Предлагаемый метод.

3.1. Постановка задачи и обозначения. Рассматривается кооперативный сценарий проверки подлинности, в котором пользователь находится перед экраном монитора и выполняет калибровочную процедуру, последовательно фиксируя взгляд на предъявляемых системой точках. Видеозапись $\mathcal{V}$ представляет собой последовательность кадров $I_1, \ldots, I_N$. Моменты, в которые известно, на какую точку $(x_t^{\mathrm{gt}}, y_t^{\mathrm{gt}})$ на экране смотрит пользователь, задают набор индексов $\mathcal{M} = \{i_1, \ldots, i_T\}$. Назовем калибровочным треком $\mathcal{T}$ подпоследовательность кадров $I_{i_1}, \ldots, I_{i_T}$. Требуется по калибровочному треку $\mathcal{T}$ принять решение о подлинности соответствующей видеозаписи в рамках указанного кооперативного сценария. Подчеркнем, что в данной постановке не рассматривается пассивная детекция по произвольному видеоряду без дополнительных условий. Предполагаются наличие взаимодействия пользователя с системой, выполнение инструкции по переводу взгляда и достаточное число кадров, позволяющее оценить параметры калибровки.

3.2. Оценка направления взгляда и калибровка. Оценка направления взгляда включает два этапа — калибровку и последующее определение точки фиксации. На этапе калибровки испытуемому предъявляется набор экранных точек с известными координатами. Пары “изображение лица–координата точки на экране” при этом формируются аналогично протоколу сбора данных GazeT [2]. После сбора этих пар модель предсказывает соответствующие им углы направления взгляда человека на изображении. По парам “углы–координаты” подбираются параметры преобразования, связывающего угловые предсказания с положением точки на экране и учитывающего индивидуальные особенности и геометрию сцены. На этапе определения направления взгляда модель по изображению лица вычисляет углы отклонения, которые затем переводятся в экранные координаты с помощью найденных калибровочных параметров. Предполагается, что оба этапа выполняются в одинаковых условиях съемки.

Поскольку цель работы состоит не в разработке нового метода оценки направления взгляда, а в исследовании диагностической ценности параметров калибровки, модуль оценки направления взгляда рассматривается как черный ящик. Формально его можно записать в виде

$$\left(\psi_t^{\mathrm{pred}}, \theta_t^{\mathrm{pred}}\right) = G(I_t),$$

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

где $G(\cdot)$ — нейросетевая модель, возвращающая горизонтальный и вертикальный углы отклонения зрительной оси. Для кадров калибровочного трека $\mathcal{T}$ предсказанные углы обозначим как $\left(\psi_t^{\text{pred}}, \theta_t^{\text{pred}}\right)$.

Калибровка строит соответствие между предсказанными углами (ψ, θ) и экранными координатами (x, y). Для каждой оси используется параметрическая модель на основе тангенса:

$$x_t = k_x \operatorname{tg}(\psi_t + a_x) + b_x, \tag{1}$$

$$y_t = k_y \operatorname{tg}(\theta_t + a_y) + b_y, \tag{2}$$

где $k_x, a_x, b_x, k_y, a_y, b_y$ — параметры калибровки. Здесь k и b отвечают за масштаб и сдвиг, а a задает угловое смещение.

Оценивание параметров выполняется по множеству калибровочных кадров $\mathcal{M}$:

$$\arg\min_{a_x} \min_{k_x, b_x} \sum_{t \in \mathcal{M}} \left(x_t^{\text{gt}} - k_x \operatorname{tg}(\psi_t^{\text{pred}} + a_x) - b_x \right)^2,$$

$$\arg\min_{a_y} \min_{k_y, b_y} \sum_{t \in \mathcal{M}} \left(y_t^{\text{gt}} - k_y \operatorname{tg}(\theta_t^{\text{pred}} + a_y) - b_y \right)^2.$$

При фиксированных a_x и a_y задачи по каждой оси сводятся к методу наименьших квадратов, а сами параметры смещения выбираются перебором по допустимому диапазону с минимизацией остаточной ошибки.

В результате для каждого трека получается вектор калибровочных параметров

$$\boldsymbol{c} = \left(k_x^*, a_x^*, b_x^*, k_y^*, a_y^*, b_y^*\right).$$

Обычно эти параметры используются для перевода предсказанных углов в координаты на экране при помощи преобразований (1) и (2). В данной работе они рассматриваются как самостоятельные признаки для распознавания поддельных видеотреков.

3.3. Трековые признаки и классификатор. Предлагается использовать калибровочные параметры $\boldsymbol{c}$ для оценки подлинности соответствующей видеозаписи $\mathcal{V}$. Пусть каждой видеозаписи сопоставлен бинарный признак класса $z \in \{0, 1\}$, где $z = 1$ соответствует поддельному треку, а $z = 0$ — подлинному. Требуется построить отображение

$$\mathbb{R}^d \to [0, 1], \qquad \hat{p} = f(\boldsymbol{c}),$$

где $\hat{p}$ интерпретируется как оценка вероятности того, что видеозапись является поддельной, d — длина вектора признаков.

Обучение итогового классификатора выполняется по набору размеченных треков

$$\left\{(\boldsymbol{c}_j, z_j)\right\}_{j=1}^{M},$$

где $\boldsymbol{c}_j$ — вектор калибровочных параметров для j-го трека, а $z_j \in \{0, 1\}$ — метка класса. По этой выборке строится отображение

$$\hat{p} = f(\boldsymbol{c}),$$

где $\hat{p}$ интерпретируется как оценка вероятности того, что соответствующий трек является поддельным. В силу малой размерности и интерпретируемости признакового вектора $\boldsymbol{c}$ такой подход допускает использование широкого класса классификаторов для табличных данных.

3.4. Объединение с предсказаниями пассивных классификаторов. Практически важен сценарий, в котором активная кооперативная проверка используется не изолированно, а совместно с уже имеющимся пассивным классификатором поддельных изображений лица. В отличие от предлагаемого подхода, использующего параметры калибровки и потому опирающегося на согласованность движения взгляда в пределах всего калибровочного трека, такой классификатор работает независимо на отдельных кадрах и извлекает главным образом визуальные признаки подделки. Поскольку эти два источника информации имеют различную природу, их совместное использование может повысить надежность итогового решения.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

Пусть для каждого кадра I_{i_t} калибровочного трека $\mathcal{T}$ пассивный классификатор H с параметрами ϕ вычисляет оценку вероятности подделки

$$q_t = H_\phi(I_{i_t}), \qquad t = 1, \ldots, T.$$

Тогда итоговая покадровая оценка для всего калибровочного трека определяется усреднением

$$s = \frac{1}{T}\sum_{t=1}^{T} q_t.$$

Тем самым обучение пассивного классификатора может выполняться на уровне отдельных кадров, тогда как в предлагаемой схеме объединения используется уже агрегированная трековая величина s, сопоставимая с вектором калибровочных параметров $\boldsymbol{c}$.

После этого формируется расширенный вектор признаков

$$\tilde{\boldsymbol{c}} = (\boldsymbol{c}, s)\,,$$

по которому обучается итоговый классификатор

$$\hat{p} = f\big(\tilde{\boldsymbol{c}}\big).$$

Полная схема работы предложенного метода представлена на рис. 2. Такая схема соответствует позднему объединению признаков, в котором усредненный выход пассивного покадрового классификатора дополняется параметрами калибровки, отражающими геометрическую согласованность наблюдаемого движения глаз. Предполагается, что даже при высоком качестве пассивного классификатора добавление вектора $\boldsymbol{c}$ улучшает разделение классов, поскольку ошибки синтеза на экстремальных вертикальных отклонениях взгляда могут проявляться не только в локальной текстуре изображения, но и в характере калибровочного преобразования.

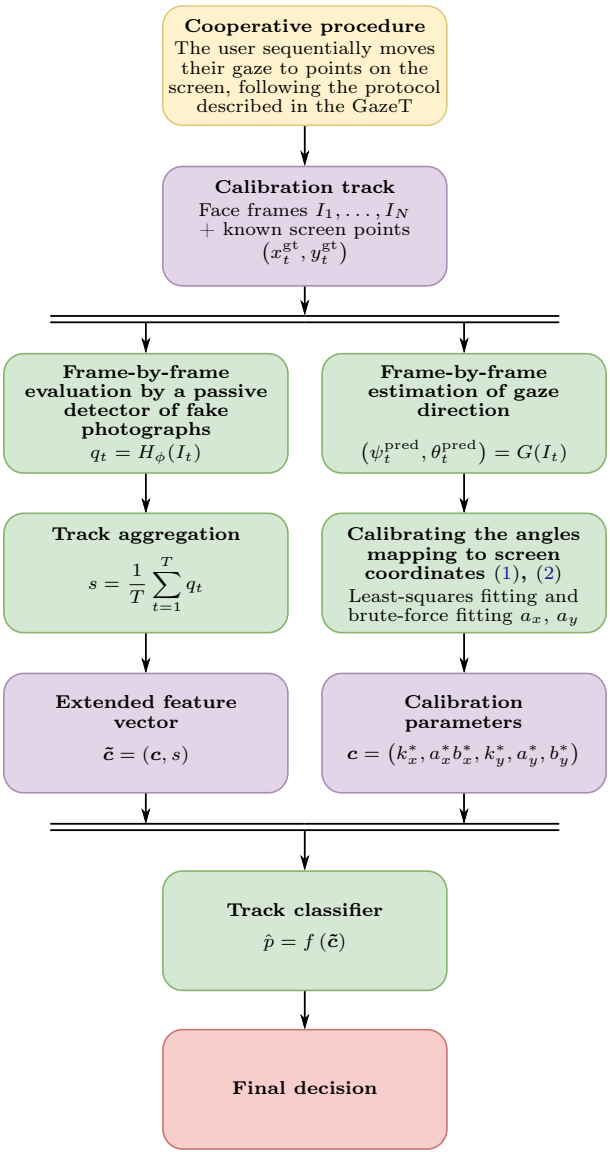

Рис. 2. Полная схема работы метода

Fig. 2. The full scheme of the method's pipeline

3.5. Вычислительная сложность и практическая применимость. Предлагаемый метод ориентирован на интерактивную постановку, в которой пользователь в ходе короткой калибровочной процедуры последовательно фиксирует взгляд на заданных точках экрана. Во время этой процедуры по кадрам параллельно применяются модель оценки направления взгляда и базовый пассивный детектор поддельных фотографий, а после ее завершения выполняются лишь усреднение выходов пассивного детектора, вычисление параметров калибровки и финальная классификация по вектору признаков $\boldsymbol{c}$. При потоковой реализации значительная часть кадров может быть обработана непосредственно в промежутках между последовательными фиксациями взгляда, так что после завершения калибровки, как правило, остается дождаться обработки последних кадров. Таким образом, итоговое решение о подлинности может быть получено через несколько секунд после завершения процедуры. Следует отметить, что в рассматриваемом сценарии строгая покадровая обработка в режиме реального времени не требуется, допустима небольшая задержка, которая не будет сильно влиять на пользовательский опыт.

4. Экспериментальная постановка.

4.1. Данные и формирование поддельных видеозаписей. В качестве основы использован набор видеозаписей GazeT, созданный для задачи оценки направления взгляда. Он содержит сессии, в

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

которых участники последовательно переводят взгляд на различные точки на экране, так что для кадров видеозаписи доступны интересующие нас пары вида “изображение лица–координата точки на экране”. Кроме того, в наборе предусмотрены отдельные калибровочные сессии, предназначенные для настройки преобразования между предсказанными углами взгляда и координатами на экране. В работе использована тестовая часть GazeT, что важно для демонстрации переносимости. Тестовая часть GazeT включает 1161 подлинных треков от разных участников, состоящих из 26152 фотографий.

Для получения поддельных примеров для каждого исходного видеотрека были синтезированы версии с подменой лица с помощью Simswap [31] и Deep-live-cam [32]. При генерации сохранялись фон, движения головы и временна́я структура исходного ролика, а изменения затрагивали главным образом область лица. Это позволило корректно сопоставлять подлинные и поддельные версии одного трека и исключить влияние посторонних факторов, связанных со сменой сцены, монтажом и различиями в сценарии предъявления точек. При синтезе поддельных роликов использовались рекомендуемые авторами базовые параметры соответствующих работ, без дополнительного подбора под рассматриваемый набор данных.

4.2. Модели оценки направления взгляда. В экспериментах применялись две независимо обученные модели, одна из которых обучалась на MPIIGaze, а другая на Gaze360. Такая пара выбрана для проверки устойчивости предлагаемого признакового описания к смене обучающих данных и, соответственно, к изменению статистики входных изображений и диапазонов наблюдаемых поз. Важно, что обучение моделей направления взгляда проводилось на внешних наборах данных, отличных от используемого в работе материала GazeT, что позволяет рассматривать постановку как перенос между различными условиями съемки и разметки. Набор MPIIGaze содержит 213659 изображений от 15 участников, собранных в ходе повседневного использования ноутбуков в течение более чем трех месяцев, и характеризуется высокой вариативностью освещения и внешнего вида. Набор Gaze360 включает 238 участников и 172 тыс. размеченных изображений, записанных в помещениях и на улице. Особенностью этого набора данных является широкий диапазон поз головы, расстояний до камеры и направлений 3D-взгляда.

В обоих случаях использовалась одна и та же архитектура ResNet-10 из работы GazeT [2]. Обучение выполнялось в PyTorch в течение 100 эпох с оптимизатором Adam с параметрами $\beta_1 = 0.9$, $\beta_2 = 0.999$, начальной скоростью обучения 10^{-4} и weight decay 10^{-4}. На вход сети подавались изображения лицевой области разрешения 160×160. Для стабилизации обучения использовался планировщик скорости обучения с линейным warmup на первых 5 эпохах и последующим косинусным убыванием. На этапе обучения применялись стандартные аугментации изображений, включавшие случайное масштабирование и кадрирование, небольшие геометрические возмущения, а также случайные изменения яркости, контраста и насыщенности.

4.3. Классификаторы калибровочных признаков. Для классификации треков по вектору калибровочных параметров $\boldsymbol{c}$ рассматривались три метода машинного обучения: линейная регрессия, метод k ближайших соседей и градиентный бустинг над решающими деревьями. Такой выбор позволяет сопоставить линейную модель, локальный непараметрический метод и нелинейный ансамблевый подход на одном и том же наборе калибровочных признаков. Все рассматриваемые методы применимы к малоразмерным табличным данным и не требуют сложного признакового описания. Во всех сериях экспериментов использовалась одна и та же схема обучения на фиксированном разбиении треков на обучающую и тестовую части, что обеспечивало корректное сопоставление результатов для разных сочетаний моделей оценки направления взгляда, пассивных детекторов поддельных видео и итоговых классификаторов.

4.4. Детекторы поддельных видео. Для получения базовых оценок подлинности видеозаписей в работе использовались три нейросетевые архитектуры различной вычислительной сложности: ResNet-18, ResNet-50 и TinyViT. В качестве обучающих выборок были выбраны общедоступные коллекции поддельных лицевых видео FF++ и DFDC, различающиеся способами синтеза, условиями съемки, степенью сжатия и качеством исходного материала. Набор FF++ содержит 1000 оригинальных видеозаписей и 4000 поддельных видеозаписей, полученных четырьмя различными методами манипуляции лица [15]. Набор DFDC представляет собой крупномасштабную коллекцию из более чем 100000 видеоклипов, записанных с участием 3426 актеров и дополненных большим числом синтезированных подделок [17]. Отдельные версии детекторов обучались независимо на каждой из этих коллекций, что позволяло оценивать переносимость моделей при применении к независимым видеотрекам.

Выбор архитектур ResNet-18, ResNet-50 и TinyViT обусловлен тем, что они различаются числом обучаемых параметров, вычислительной сложностью и архитектурным типом. ResNet-18 представляет легкую сверточную модель, ResNet-50 — более емкую сверточную архитектуру, а TinyViT — компакт-

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

ную трансформерную модель. Такой набор не претендует на исчерпывающий перебор всех возможных пассивных детекторов, однако позволяет проверить, сохраняется ли прирост от включения предлагаемых калибровочных признаков при переходе от сравнительно легких к более мощным и архитектурно отличным базовым моделям. Для использованных базовых пассивных детекторов в табл. 1 приведены число обучаемых параметров и среднее время инференса на CPU.

Обучение детекторов проводилось на уровне отдельных кадров лицевой области в постановке двоичной классификации с целевой переменной $y \in \{0, 1\}$, где $y = 1$ соответствует кадру из поддельного видео, а $y = 0$ — кадру из подлинного видео. На вход модели подавался кадр лицевой области фиксированного размера x, а на выходе формировалась оценка вероятности подделки $p_\theta(y = 1 \,|\, x)$.

Оптимизация выполнялась по бинарной кросс-энтропии:

$$\mathcal{L}(\theta) = -\frac{1}{M}\sum_{i=1}^{M}\Big(y_i \log p_\theta(y_i = 1 \,|\, x_i) + (1 - y_i)\log\big(1 - p_\theta(y_i = 1 \,|\, x_i)\big)\Big),$$

где $\{(x_i, y_i)\}_{i=1}^{M}$ — обучающие примеры, соответствующие отдельным кадрам.

Все пассивные детекторы обучались в среде PyTorch в течение 60 эпох с оптимизатором Adam с параметрами $\beta_1 = 0.9$, $\beta_2 = 0.999$, начальной скоростью обучения 10^{-4} и weight decay 10^{-4}. На вход сети подавались изображения лицевой области разрешения 256×256. Для повышения устойчивости применялись стандартные аугментации, включавшие случайное масштабирование и кадрирование, небольшие повороты и сдвиги, случайные изменения яркости и контраста, а также имитацию компрессии и слабого размытия. Скорость обучения изменялась с помощью планировщика с коротким линейным warmup на начальном этапе и последующим косинусным убыванием. Во всех сериях экспериментов использовалась одна и та же схема обучения, что обеспечивало корректное сопоставление результатов для разных архитектур и обучающих выборок.

4.5. Сравниваемые метрики. Качество решений оценивается на уровне треков с помощью метрики ROC-AUC. Для каждого трека классификатор формирует значение $\hat{p}$, после чего при заданном пороге $\tau \in [0, 1]$ принимается решение $\hat{y}(\tau) = \mathbb{I}[\hat{p} \geqslant \tau]$, где $\mathbb{I}[\cdot]$ — индикатор. На множестве тестовых треков вычисляются величины $TP(\tau)$, $FP(\tau)$, $TN(\tau)$ и $FN(\tau)$, соответствующие числам истинно положительных, ложно положительных, истинно отрицательных и ложно отрицательных решений.

Основными характеристиками служат доля верно обнаруженных подделок и доля ложных срабатываний, определяемые как

$$TPR(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)}, \qquad FPR(\tau) = \frac{FP(\tau)}{FP(\tau) + TN(\tau)}.$$

Зависимость $TPR(\tau)$ от $FPR(\tau)$ при изменении порога τ образует кривую рабочих характеристик, используемую для сравнения методов в условиях неизвестного рабочего порога. Для компактного представления качества применяется площадь под этой кривой

$$\text{ROC-AUC} = \int_0^1 TPR\big(FPR\big)\, dFPR,$$

Таблица 1. Число параметров и среднее время инференса базовых пассивных детекторов на CPU и GPU. Измерения на CPU проводились на Intel(R) Xeon(R) Gold 6230 CPU @ 2.10 GHz с использованием 4 потоков, а измерения на GPU — на NVIDIA GeForce RTX 2080 Ti

Table 1. Number of parameters and average inference time of the baseline passive detectors on CPU and GPU. CPU measurements were performed on an Intel(R) Xeon(R) Gold 6230 CPU @ 2.10 GHz using 4 threads, and GPU measurements were performed on an NVIDIA GeForce RTX 2080 Ti

Модель Model	Параметры, $\times 10^6$ Parameters, $\times 10^6$	CPU, мс/кадр CPU, ms/frame	GPU, мс/кадр GPU, ms/frame
ResNet-18	11.3	11.5	1.69
ResNet-50	24.1	27.9	2.56
TinyViT	20.8	34.9	4.03

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

которая принимает значения от 0 до 1 и тем выше, чем лучше метод разделяет классы при вариации порога.

5. Результаты и анализ. Результаты, приведенные в табл. 2, подтверждают основную гипотезу работы: параметры калибровки, связывающие предсказанные углы направления взгляда с координатами точек на экране, сами по себе содержат информацию, достаточную для различения подлинных и синтезированных видеотреков. Без использования внешнего детектора поддельного видео классификация с помощью градиентного бустинга только по этим параметрам дает значение ROC-AUC 0.675 для модели оценки направления взгляда, обученной на MPIIGaze, и 0.717 для модели, обученной на Gaze360. Для линейной регрессии и метода k ближайших соседей классификация только по параметрам калибровки оказывается существенно слабее, но сохраняется: для модели, обученной на MPIIGaze, значения ROC-AUC составляют 0.587 и 0.609 соответственно, а для модели, обученной на Gaze360, — 0.563 и 0.628. Это показывает, что дискриминативный сигнал присутствует именно на уровне трека и не сводится к локальным артефактам отдельных кадров.

Добавление параметров калибровки улучшает качество во всех сериях экспериментов при использовании градиентного бустинга. Для ResNet-18, обученной на FF++, значение ROC-AUC возрастает с 0.635 до 0.705 при использовании модели оценки направления взгляда, обученной на MPIIGaze, и до 0.737 при использовании модели, обученной на Gaze360. Для ResNet-50, обученной на DFDC, показатель увеличивается с 0.628 до 0.706 и 0.737 соответственно. В целом прирост составляет от 0.010 до 0.109 по ROC-AUC, что указывает на независимую информативность калибровочных параметров по отношению к обычным визуальным детекторам. Эффект сохраняется как при смене обучающего набора данных детектора поддельного видео, так и при смене обучающего набора данных модели оценки направления взгляда, что говорит о хорошей переносимости подхода. В большинстве случаев лучшие результаты достигаются

Таблица 2. Влияние параметров калибровки направления взгляда на качество распознавания поддельных видеотреков для различных классификаторов. LR — логистическая регрессия, KNN — метод k ближайших соседей, GB — градиентный бустинг

Table 2. The effect of gaze direction calibration parameters on the recognition quality of fake video tracks for different classifiers. LR — logisict regression, KNN — k nearest neighbours, GB — gradient boosing

Архитектура детектора Detector architecture	Обучающий набор данных Training dataset	Классификатор Classifier	Без параметров калибровки Without calibration parameters	С параметрами калибровки (MPIIGaze) With calibration parameters (MPIIGaze)	Δ	С параметрами калибровки (Gaze360) With calibration parameters (Gaze360)	Δ
Классификация только по параметрам калибровки Classification by calibration parameters only							
—	—	KNN	—	0.609	—	0.628	—
—	—	LR	—	0.587	—	0.563	—
—	—	GB	—	0.675	—	0.717	—
Объединение с детекторами поддельного видео Integration with fake video detectors							
ResNet-18	FF++	KNN	0.635	0.623	-0.012	0.634	-0.001
ResNet-18	FF++	LR	0.635	0.653	+0.018	0.645	+0.010
ResNet-18	FF++	GB	0.635	0.705	**+0.070**	0.737	**+0.102**
ResNet-18	DFDC	KNN	0.647	0.620	-0.027	0.634	-0.013
ResNet-18	DFDC	LR	0.647	0.664	+0.017	0.656	+0.009
ResNet-18	DFDC	GB	0.647	0.706	**+0.059**	0.728	**+0.081**
ResNet-50	FF++	KNN	0.670	0.631	-0.039	0.645	-0.025
ResNet-50	FF++	LR	0.670	0.679	+0.009	0.673	+0.003
ResNet-50	FF++	GB	0.670	0.731	**+0.061**	0.759	**+0.089**
ResNet-50	DFDC	KNN	0.628	0.621	-0.007	0.627	-0.001
ResNet-50	DFDC	LR	0.628	0.649	+0.021	0.639	+0.011
ResNet-50	DFDC	GB	0.628	0.706	**+0.078**	0.737	**+0.109**
TinyViT	FF++	KNN	0.812	0.677	-0.135	0.684	-0.128
TinyViT	FF++	LR	0.812	0.812	0.000	0.810	-0.002
TinyViT	FF++	GB	0.812	0.824	**+0.012**	0.822	**+0.010**
TinyViT	DFDC	KNN	0.703	0.624	-0.079	0.643	-0.060
TinyViT	DFDC	LR	0.703	0.712	+0.009	0.711	+0.008
TinyViT	DFDC	GB	0.703	0.749	**+0.046**	0.771	**+0.068**

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

при использовании модели, обученной на Gaze360, что, вероятно, связано с большим разнообразием поз и направлений взгляда в этом наборе данных. Особенно заметен выигрыш для более слабых исходных детекторов, однако и для наиболее сильного в рассматриваемой таблице варианта, а именно TinyViT, обученного на FF++, добавление параметров калибровки также улучшает результат: значение ROC-AUC возрастает с 0.812 до 0.824.

В то же время для линейной регрессии выигрыш от добавления параметров калибровки оказывается небольшим и, как правило, не превышает 0.022 по ROC-AUC, а для TinyViT, обученного на FF++, практически исчезает. Для метода k ближайших соседей добавление калибровочных параметров не приводит к улучшению качества и во всех рассмотренных сериях дает отрицательное изменение ROC-AUC. Это можно объяснить тем, что зависимость между параметрами калибровки и меткой поддельности имеет нелинейный характер и определяется взаимодействием нескольких признаков, которое линейная регрессия описывает лишь приближенно. Кроме того, метод k ближайших соседей чувствителен к локальной структуре признакового пространства: при частичном перекрытии распределений подлинных и поддельных треков соседние объекты часто оказываются смешанными по классам, из-за чего слабый дополнительный сигнал от калибровочных параметров не усиливается, а, напротив, размывается. Тем самым полученные результаты указывают на то, что калибровочные параметры являются информативными, но для их эффективного использования требуется достаточно гибкий нелинейный классификатор, способный учитывать пороговые эффекты и взаимодействия между признаками.

Таблица 3. Значения важности признаков для параметров калибровки, вычисленные по обученному классификатору градиентного бустинга, использующему только калибровочные признаки
Table 3. The values of the importance of features for calibration parameters, calculated using a trained gradient boosting classifier that uses only calibration features

Параметр калибровки Calibration parameter	MPIIGaze	Gaze360
Вертикальная составляющая Vertical component		
k_y	0.201	0.248
b_y	0.194	0.231
a_y	0.185	0.218
Суммарная важность **Total importance**	**0.581**	**0.697**
Горизонтальная составляющая Horizontal component		
k_x	0.159	0.137
b_x	0.190	0.113
a_x	0.070	0.053
Суммарная важность **Total importance**	**0.419**	**0.303**

Дополнительный анализ важности признаков градиентного бустинга, представленный в табл. 3, показывает, что наибольший вклад в итоговое решение вносят параметры, связанные с вертикальной составляющей направления взгляда. Для обеих моделей оценки направления взгляда суммарная важность признаков (k_y, b_y, a_y) превышает суммарную важность признаков (k_x, b_x, a_x), причем три наиболее значимых признака в обоих случаях также относятся к вертикальной оси. Это согласуется с гипотезой о том, что генераторы хуже воспроизводят область глаз при крайних отклонениях взгляда вверх и вниз. Таким образом, параметры калибровки направления взгляда можно рассматривать как информативный признак подлинности видеотрека, пригодный как для самостоятельного использования, так и для объединения с существующими детекторами поддельного видео.

6. Обсуждение и ограничения. Предложенный подход показывает, что параметры калибровки, получаемые при переводе оценок направления взгляда в координаты на экране, несут устойчивую информацию, позволяющую отличать поддельные видеозаписи от подлинных на уровне целых видеотреков. Существенно, что эффект сохраняется при смене исходных наборов данных, на которых обучались как алгоритм оценки направления взгляда, так и детектор подделок, что указывает на относительную независимость признаков калибровки от конкретного источника обучения и на их переносимость между условиями съемки. Наблюдаемое преобладание вклада параметров, связанных с вертикальным отклонением взгляда, согласуется с предположением о том, что генераторы чаще допускают систематические артефакты при моделировании глаз и век в крайних положениях “вверх–вниз”, и эти артефакты проявляются не столько в отдельных кадрах, сколько в накопленных статистиках, отраженных в калибровочных преобразованиях.

С практической точки зрения метод особенно уместен в сценариях, где взаимодействие пользователя с экраном является естественной частью процедуры, а видеопоток поступает с потребительских камер без специализированного оборудования. В качестве примера можно рассматривать системы видеоконференц-

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

связи при подключении к рабочим встречам или удаленным сессиям, где требуется быстрая проверка подлинности участника без навязчивых действий: краткая калибровка с переводом взгляда на экранные метки может выполняться как этап входа или как периодическая проверка в ходе разговора. В таких условиях преимущества дает то, что решение опирается на динамическое поведение и согласованность реакции зрительной системы с предъявляемым стимулом, а не только на статические визуальные признаки лица, которые чаще становятся объектом имитации.

Вместе с тем у подхода есть ограничения, связанные как с постановкой задачи, так и с допущениями модели. Метод опирается на наличие калибровочной процедуры и достаточной длины видеотрека, поэтому он неприменим к коротким фрагментам и к записям, где пользователь не взаимодействует с экраном или не выполняет инструкцию. На качество также влияют геометрия сцены и условия съемки: расстояние до камеры, положение устройства относительно экрана, оптические искажения, а также вариативность освещения могут менять стабильность оценок направления взгляда и, как следствие, параметры калибровки. Дополнительным источником неопределенности является физиологическое и поведенческое разнообразие людей.

Ограничением текущей работы является то, что эксперименты выполнены в рамках активного кооперативного сценария на данных GazeT и на подделках, сгенерированных всего лишь двумя генераторами поддельных изображений. Поэтому полученные результаты следует интерпретировать не как доказательство универсальности предлагаемого признака для любых типов подделок, а как свидетельство его информативности и переносимости в исследованной постановке. Расширение экспериментов на дополнительные наборы данных, более широкий спектр методов генерации и сравнение с другими активными детекторами является важным направлением дальнейшей работы. Дополнительного исследования требует сравнение с более широким кругом современных пассивных и активных методов, а также оценка прироста от включения предлагаемых признаков в уже опубликованные пассивные SotA-подходы. Эти вопросы выходят за рамки настоящей работы и также рассматриваются нами как направление дальнейших исследований.

7. Заключение. Показано, что параметры калибровки модели оценки направления взгляда могут служить информативными признаками для распознавания поддельных видеозаписей с подменой лица. Проведенные эксперименты на записях, полученных на основе тестовой части набора GazeT и дополненных синтезированными подделками, подтвердили, что такие признаки позволяют устойчиво различать подлинные и поддельные треки даже при использовании моделей оценки направления взгляда и детекторов поддельных видео, обученных на различных выборках. Установлено также, что объединение калибровочных признаков с предсказаниями обычных детекторов поддельных видео дает дополнительный прирост качества, что указывает на взаимодополняющий характер этих источников информации. Анализ значимости признаков показал повышенную роль параметров, связанных с вертикальной составляющей взгляда, что косвенно подтверждает гипотезу о недостаточно правдоподобном воспроизведении области глаз при больших отклонениях вверх и вниз в современных методах синтеза лица. Полученные результаты позволяют рассматривать предложенный подход как перспективное направление развития кооперативных методов проверки подлинности видеопотока.

Список литературы

1. *Krafka K., Khosla A., Kellnhofer P., et al.* Eye tracking for everyone // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA, Las Vegas, June 27–30, 2016. IEEE, 2016. pp. 2176–2184. doi 10.1109/CVPR.2016.239.
2. *Красильников М.Д., Никитин М.Ю.* GazeT: улучшение определения трехмерного вектора направления взгляда оператора // Информационные процессы. 2024. **24**, № 4. 421–429. http://www.jip.ru/2024/421-429-2024.pdf. Дата обращения: 14 мая 2026.
3. *Zhang X., Sugano Y., Fritz M., Bulling A.* MPIIGaze: real-world dataset and deep appearance-based gaze estimation // IEEE Trans. Pattern Anal. Mach. Intell. 2019. **41**, N 1. 162–175. doi 10.1109/TPAMI.2017.2778103.
4. *Funes Mora K.A., Monay F., Odobez J.-M.* EYEDIAP: a database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras // Proceedings of the Symposium on Eye Tracking Research and Applications. Safety Harbor Florida: ACM, 2014. pp. 255–258. doi 10.1145/2578153.2578190.
5. *Zhang X., Park S., Beeler T., et al.* ETH-XGaze: a large scale dataset for gaze estimation under extreme head pose and gaze variation // Computer Vision — ECCV 2020. Lecture Notes in Computer Science. Vol. 12350. Cham: Springer International Publishing, 2020. pp. 365–381. doi 10.1007/978-3-030-58558-7_22.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

6. *Kellnhofer P., Recasens A., Stent S., et al.* Gaze360: physically unconstrained gaze estimation in the wild // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Korea (South), Seoul, October 27 – November 02, 2019. IEEE, 2020. pp. 6911–6920. doi 10.1109/ICCV.2019.00701.
7. *Jabber N.H., Hashim I.A.* Robust eye features extraction based on eye angles for efficient gaze classification system // 2018 Third Scientific Conference of Electrical Engineering (SCEE), Iraq, Baghdad, December 19–20, 2018. IEEE, 2019. pp. 13–18. doi 10.1109/SCEE.2018.8684107.
8. *Vicente F., Huang Z., Xiong X., et al.* Driver gaze tracking and eyes off the road detection system // IEEE Trans. Intell. Transport. Syst. 2015. **16**, N 4. 2014–2027. doi 10.1109/TITS.2015.2396031.
9. *George A., Routray A.* Real-time eye gaze direction classification using convolutional neural network // 2016 International Conference on Signal Processing and Communications (SPCOM), India, Bangalore, June 12–15, 2016. IEEE, 2016. pp. 1–5. doi 10.1109/SPCOM.2016.7746701.
10. *Park S., Spurr A., Hilliges O.* Deep pictorial gaze estimation // Computer Vision — ECCV 2018. Lecture Notes in Computer Science. Vol. 11217. Cham: Springer International Publishing, 2018. pp. 741–757. doi 10.1007/978-3-030-01261-8_44.
11. *Zhou X., Jiang J., Liu Q., et al.* Learning a 3D gaze estimator with adaptive weighted strategy // IEEE Access. 2020. **8**. 82142–82152. doi 10.1109/ACCESS.2020.2990685.
12. *Chen Z., Shi B.E.* Appearance-based gaze estimation using dilated-convolutions // Computer Vision — ACCV 2018. Lecture Notes in Computer Science. Vol. 11366. Cham: Springer International Publishing, 2019. pp. 309–324. doi 10.1007/978-3-030-20876-9_20.
13. *Cheng Y., Huang S., Wang F., et al.* A Coarse-to-fine adaptive network for appearance-based gaze estimation // AAAI-20 Technical Tracks 7. 2020. **34**, N 07. 10623–10630. doi 10.1609/aaai.v34i07.6636.
14. *He J., Pham K., Valliappan N., et al.* On-device few-shot personalization for real-time gaze estimation // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Korea (South), Seoul, October 27–28, 2019. IEEE, 2020. pp. 1149–1158. doi 10.1109/ICCVW.2019.00146.
15. *Rössler A., Cozzolino D., Verdoliva L., et al.* FaceForensics++: learning to detect manipulated facial images // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Korea (South), Seoul, October 27 – November 02, 2019. IEEE, 2020. pp. 1–11. doi 10.1109/ICCV.2019.00009.
16. *Li Y., Yang X., Sun P., et al.* Celeb-DF: a large-scale challenging dataset for deepfake forensics // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Seattle, June 13–19, 2020. IEEE, 2020. pp. 3204–3213. doi 10.1109/CVPR42600.2020.00327.
17. *Dolhansky B., Bitton J., Pflaum B., et al.* The deepfake detection challenge (DFDC) dataset // doi 10.48550/ARXIV.2006.07397.
18. *Pu J., Mangaokar N., Kelly L., et al.* Deepfake videos in the wild: analysis and detection // Proceedings of the Web Conference, Slovenia, Ljubljana, April 19–23, 2021. ACM, 2021. pp. 981–992. doi 10.1145/3442381.3449978.
19. *He Y., Gan B., Chen S., et al.* ForgeryNet: a versatile benchmark for comprehensive forgery analysis // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021. IEEE, 2021. pp. 4358–4367. doi 10.1109/CVPR46437.2021.00434.
20. *Afchar D., Nozick V., Yamagishi J., Echizen I.* MesoNet: a compact facial video forgery detection network // 2018 IEEE International Workshop on Information Forensics and Security (WIFS), China, Hong Kong, December 11–13, 2018. IEEE, 2019. pp. 1–7. doi 10.1109/WIFS.2018.8630761.
21. *Nguyen H.H., Yamagishi J., Echizen I.* Capsule-forensics: using capsule networks to detect forged images and videos // ICASSP 2019 — 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, May 12–17, 2019. IEEE, 2019. pp. 2307–2311. doi 10.1109/ICASSP.2019.8682602.
22. *Qian Y., Yin G., Sheng L., et al.* Thinking in frequency: face forgery detection by mining frequency-aware clues // Computer Vision — ECCV 2020. Lecture Notes in Computer Science. Vol. 12357. Cham: Springer International Publishing, 2020. pp. 86–103. doi 10.1007/978-3-030-58610-2_6.
23. *Dang H., Liu F., Stehouwer J., et al.* On the detection of digital face manipulation // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Seattle, June 13–19, 2020. IEEE, 2020. pp. 5780–5789. doi 10.1109/CVPR42600.2020.00582.
24. *Luo Y., Zhang Y., Yan J., Liu W.* Generalizing face forgery detection with high-frequency features // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021. IEEE, 2021. pp. 16312–16321. doi 10.1109/CVPR46437.2021.01605.
25. *Liu H., Li X., Zhou W., et al.* Spatial-phase shallow learning: rethinking face forgery detection in frequency domain // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021. IEEE, 2021. pp. 772–781. doi 10.1109/CVPR46437.2021.00083.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

26. *Yan Z., Zhang Y., Fan Y., Wu B.* UCF: Uncovering common features for generalizable deepfake detection. doi 10.48550/ARXIV.2304.13949.
27. *Cao J., Ma C., Yao T., et al.* End-to-end reconstruction-classification learning for face forgery detection // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, New Orleans, June 18–24, 2022. IEEE, 2022. pp. 4103–4112. doi 10.1109/CVPR52688.2022.00408.
28. *Gerstner C.R., Farid H.* Detecting real-time deep-fake videos using active illumination // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), USA, New Orleans, June 19–20, 2022. IEEE, 2022. pp. 53–60. doi 10.1109/CVPRW56347.2022.00015.
29. *Guo H., Wang X., Lyu S.* Detection of real-time deepfakes in video conferencing with active probing and corneal reflection // ICASSP 2023 — 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Greece, Rhodes Island, June 04–10, 2023. IEEE, 2023. pp. 1–5. doi 10.1109/ICASSP49357.2023.10094720.
30. *Mittal G., Hegde C., Memon N.* Gotcha: real-time video deepfake detection via challenge-response // 2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P), Austria, Vienna, July 08–12, 2024. IEEE, 2024. pp. 1–20. doi 10.1109/EuroSP60621.2024.00009.
31. *Chen R., Chen X., Ni B., Ge Y.* SimSwap: an efficient framework for high fidelity face swapping // Proceedings of the 28th ACM International Conference on Multimedia, USA, Seattle, October 12–16, 2020. ACM, 2020. pp. 2003–2011. doi 10.48550/arXiv.2106.06340.
32. *Estanislao K.* Deep-live-cam. https://github.com/hacksider/Deep-Live-Cam. Cited May 16, 2026.

Получена	*Принята*	*Опубликована*
25 марта 2026 г.	*10 мая 2026 г.*	*28 мая 2026 г.*

Информация об авторах

Максим Денисович Красильников — аспирант; 1) Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, Ленинские горы, 1, стр. 52, 119991, Москва, Российская Федерация; 2) ООО "Технологии Видеоанализа", ул. Ефремова, 10, к. 2, 119048, Москва, Российская Федерация.

Никитин Михаил Юрьевич — ведущий исследователь; ООО "Технологии Видеоанализа", ул. Ефремова, 10, к. 2, 119048, Москва, Российская Федерация.

Конушин Антон Сергеевич — к.ф.-м.н., доцент; Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, Ленинские горы, 1, стр. 52, 119991, Москва, Российская Федерация.

References

1. K. Krafka, A. Khosla, P. Kellnhofer, et al., "Eye Tracking for Everyone," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA, Las Vegas, June 27–30, 2016* (IEEE, 2016), pp. 2176–2184. doi 10.1109/CVPR.2016.239.
2. M. D. Krasil'nikov and M. Yu. Nikitin, "GazeT: Improved Estimation of the Three-Dimensional Vector of the Operator's Gaze Direction," Information processes **24** (4), 421–429 (2024). http://www.jip.ru/2024/421-429-2024.pdf. Cited May 14, 2026.
3. X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation," IEEE Trans. Pattern Anal. Mach. Intell. **41** (1), 162–175 (2019). doi 10.1109/TPAMI.2017.2778103.
4. K. A. Funes Mora, F. Monay, and J.-M. Odobez, "EYEDIAP: a Database for the Development and Evaluation of Gaze Estimation Algorithms from RGB and RGB-D Cameras," in *Proceedings of the Symposium on Eye Tracking Research and Applications* (ACM, Safety Harbor Florida, 2014), pp. 255–258. doi 10.1145/2578153.2578190.
5. X. Zhang, S. Park, T. Beeler, et al., "ETH-XGaze: A Large Scale Dataset for Gaze Estimation Under Extreme Head Pose and Gaze Variation," in *Computer Vision –– ECCV 2020.* Lecture Notes in Computer Science. Vol 12350. (Springer International Publishing, Cham, 2020), pp. 365–381. doi 10.1007/978-3-030-58558-7_22.
6. P. Kellnhofer, A. Recasens, S. Stent, et al., "Gaze360: Physically Unconstrained Gaze Estimation in the Wild," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV), Korea (South), Seoul, October 27 – November 02, 2019* (IEEE, 2020), pp. 6911–6920. doi 10.1109/ICCV.2019.00701.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

7. N. H. Jabber and I. A. Hashim, "Robust Eye Features Extraction Based on Eye Angles for Efficient Gaze Classification System," in *2018 Third Scientific Conference of Electrical Engineering (SCEE), Iraq, Baghdad, December 19–20, 2018* (IEEE, 2019), pp. 13–18. doi 10.1109/SCEE.2018.8684107.
8. F. Vicente, Z. Huang, X. Xiong, et al., "Driver Gaze Tracking and Eyes Off the Road Detection System," IEEE Trans. Intell. Transport. Syst. **16** (4), 2014–2027 (2015). doi 10.1109/TITS.2015.2396031.
9. A. George and A. Routray, "Real-Time Eye Gaze Direction Classification Using Convolutional Neural Network," in *2016 International Conference on Signal Processing and Communications (SPCOM), India, Bangalore, June 12–15, 2016* (IEEE, 2016), pp. 1–5. doi 10.1109/SPCOM.2016.7746701.
10. S. Park, A. Spurr, and O. Hilliges, "Deep Pictorial Gaze Estimation," in *Computer Vision — ECCV 2018. Lecture Notes in Computer Science.* Vol 11217. (Springer International Publishing, Cham, 2018), pp. 741–757. doi 10.1007/978-3-030-01261-8_44.
11. X. Zhou, J. Jiang, Q. Liu, et al., "Learning a 3D Gaze Estimator with Adaptive Weighted Strategy," IEEE Access **8**, 82142–82152 (2020). doi 10.1109/ACCESS.2020.2990685.
12. Z. Chen and B. E. Shi, "Appearance-Based Gaze Estimation Using Dilated-Convolutions," in *Computer Vision — ACCV 2018. Lecture Notes in Computer Science.* Vol. 11366. (Springer International Publishing, Cham, 2019), pp. 309–324. doi 10.1007/978-3-030-20876-9_20.
13. Y. Cheng, S. Huang, F. Wang, et al., "A Coarse-to-Fine Adaptive Network for Appearance-Based Gaze Estimation," AAAI-20 Technical Tracks 7. **34** (07), 10623–10630 (2020). doi 10.1609/aaai.v34i07.6636.
14. J. He, K. Pham, N. Valliappan, et al., "On-Device Few-Shot Personalization for Real-Time Gaze Estimation," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Korea (South), Seoul, October 27–28, 2019* (IEEE, 2020), pp. 1149–1158. doi 10.1109/ICCVW.2019.00146.
15. A. Rössler, D. Cozzolino, L. Verdoliva, et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV), Korea (South), Seoul, October 27 – November 02, 2019* (IEEE, 2020), pp. 1–11. doi 10.1109/ICCV.2019.00009.
16. Y. Li, X. Yang, P. Sun, et al., "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Seattle, June 13–19, 2020* (IEEE, 2020), pp. 3204–3213. doi 10.1109/CVPR42600.2020.00327.
17. B. Dolhansky, J. Bitton, B. Pflaum, et al., "The DeepFake Detection Challenge (DFDC) Dataset," doi 10.48550/ARXIV.2006.07397.
18. J. Pu, N. Mangaokar, L. Kelly, et al., "Deepfake Videos in the Wild: Analysis and Detection," in *Proceedings of the Web Conference, Slovenia, Ljubljana, April 19–23, 2021* (ACM, 2021), pp. 981–992. doi 10.1145/3442381.3449978.
19. Y. He, B. Gan, S. Chen, et al., "ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021* (IEEE, 2021), pp. 4358–4367. doi 10.1109/CVPR46437.2021.00434.
20. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS), China, Hong Kong, December 11–13, 2018* (IEEE, 2019), pp. 1–7. doi 10.1109/WIFS.2018.8630761.
21. H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos," in *ICASSP 2019 — 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, May 12–17, 2019* (IEEE, 2019), pp. 2307–2311. doi 10.1109/ICASSP.2019.8682602.
22. Y. Qian, G. Yin, L. Sheng, et al., "Thinking in Frequency: Face Forgery Detection by Mining Frequency-Aware Clues," in *Computer Vision — ECCV 2020*, Lecture Notes in Computer Science. Vol. 12357. (Springer International Publishing, Cham, 2020), pp. 86–103. doi 10.1007/978-3-030-58610-2_6.
23. H. Dang, F. Liu, J. Stehouwer, et al., "On the Detection of Digital Face Manipulation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Seattle, June 13–19, 2020* (IEEE, 2020), pp. 5780–5789. doi 10.1109/CVPR42600.2020.00582.
24. Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing Face Forgery Detection with High-frequency Features," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021* (IEEE, 2021), pp. 16312–16321. doi 10.1109/CVPR46437.2021.01605.
25. H. Liu, X. Li, W. Zhou, et al., "Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, Nashville, June 20–25, 2021* (IEEE, 2021), pp. 772–781. doi 10.1109/CVPR46437.2021.00083.
26. Z. Yan, Y. Zhang, Y. Fan, and B. Wu, UCF: Uncovering Common Features for Generalizable Deepfake Detection. doi 10.48550/ARXIV.2304.13949.

2026, **27** (2), 274–289. doi 10.26089/NumMet.v27r218

27. J. Cao, C. Ma, T. Yao, et al., "End-to-End Reconstruction-Classification Learning for Face Forgery Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA, New Orleans, June 18–24, 2022* (IEEE, 2022), pp. 4103–4112. doi 10.1109/CVPR52688.2022.00408.
28. C. R. Gerstner and H. Farid, "Detecting Real-Time Deep-Fake Videos Using Active Illumination," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), USA, New Orleans, June 19–20, 2022* (IEEE, 2022), pp. 53–60. doi 10.1109/CVPRW56347.2022.00015.
29. H. Guo, X. Wang, and S. Lyu, "Detection of Real-Time Deepfakes in Video Conferencing with Active Probing and Corneal Reflection," in *ICASSP 2023 — 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Greece, Rhodes Island, June 04–10, 2023* (IEEE, 2023), pp. 1–5. doi 10.1109/ICASSP49357.2023.10094720.
30. G. Mittal, C. Hegde, and N. Memon, "Gotcha: Real-Time Video Deepfake Detection via Challenge-Response," in *2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P), Austria, Vienna, July 08–12, 2024* (IEEE, 2024), pp. 1–20. doi 10.1109/EuroSP60621.2024.00009.
31. R. Chen, X. Chen, B. Ni, and Y. Ge, "SimSwap: An Efficient Framework For High Fidelity Face Swapping," in *Proceedings of the 28th ACM International Conference on Multimedia, USA, Seattle, October 12–16, 2020* (ACM, 2020), pp. 2003–2011. doi 10.48550/arXiv.2106.06340.
32. K. Estanislao, "Deep-Live-Cam." https://github.com/hacksider/Deep-Live-Cam. Cited May 16, 2026.

Received *March 25, 2026*
Accepted *May 10, 2026*
Published *May 28, 2026*

Information about the authors

Maksim D. Krasilnikov — PhD student; 1) Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Leninskie Gory, 1, building 52, 119991, Moscow, Russia; 2) LLC Tevian, Efremova ulitsa, 10, building 2, 119048, Moscow, Russia.

Mikhail Yu. Nikitin — Lead researcher; LLC Tevian, Efremova ulitsa, 10, building 2, 119048, Moscow, Russia.

Anton S. Konushin — Ph.D., Associate Professor; Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Leninskie Gory, 1, building 52, 119991, Moscow, Russia.