

**Городенко Родион Дмитриевич**

бакалавр, Национальный исследовательский университет МЭИ, Москва.

Электронный адрес: domonion@list.ru

**Rodion D. Gorodenko**

bachelor, National Research University "Moscow Power Engineering Institute", Moscow.

E-mail address: domonion@list.ru

**Петров Сергей Андреевич**

кандидат технических наук, доцент кафедры безопасности и информационных технологий, Национальный исследовательский университет МЭИ, Москва. SPIN-код: 4284-6705, AuthorID: 1266446

Электронный адрес: PetrovSA@mpei.ru

**Sergei A. Petrov**

Ph.D. of Technical Sciences, Associate Professor at the Department of security and information technology, National Research University "Moscow Power Engineering Institute", Moscow. SPIN-code: 4284-6705, AuthorID: 1266446

E-mail address: PetrovSA@mpei.ru

**Баланев Кирилл Сергеевич**

аспирант, ассистент кафедры безопасности и информационных технологий, Национальный исследовательский университет МЭИ, Москва. SPIN-код: 8192-1861, AuthorID: 1204517

Электронный адрес: balanев.kirill@yandex.ru

**Kirill S. Balanев**

Postgraduate, Assistant at the Department of security and information technology, National Research University "Moscow Power Engineering Institute", Moscow. SPIN-code: 8192-1861, AuthorID: 1204517

E-mail address: balanев.kirill@yandex.ru

---

**МЕТОД ОЦЕНКИ КЛАССИФИКАЦИОННЫХ МОДЕЛЕЙ  
ДЛЯ АНАЛИЗА ВИДЕОПОТОКА**

---

**Аннотация.** В последние годы актуальность технологий компьютерного зрения значительно возросла, особенно в задачах анализа видеопотока, таких как обнаружение усталости, защита от подделок лиц и распознавание жестов. В известных подходах к оценке моделей, применяемых в этих задачах, широко используется метрика ROC AUC. Однако ее применение, основанное на покадровом анализе, имеет определенные ограничения. Они связаны с неустойчивостью результатов из-за нестабильных выходов модели между кадрами и отсутствием учета временных зависимостей в данных, что снижает точность оценки в реальных условиях использования стриминговых видео. Предлагаемый метод решает эти проблемы путем построения ROC-кривых для видео в целом, а не для отдельных кадров. Реализация этого процесса происходит посредством анализа выходов модели на каждом кадре, что позволяет более надежно различать положительные и отрицательные примеры на уровне всего видеофрагмента. Такой подход позволяет реалистично оценивать соотношение ложных срабатываний и полноты, а также улучшает чувствительность моделей за счет подбора более релевантных порогов классификации. Таким образом, новый метод предлагает более релевантную оценку моделей для потоковой обработки видео, обеспечивая более точное распознавание событий и их контекстуальную интерпретацию, что особенно важно в критически значимых приложениях.

**Ключевые слова:** компьютерное зрение, потоковая обработка видео, ROC AUC, оценка моделей, динамический анализ, контекстуальная интерпретация, защита от подделок, восприятие временных зависимостей, классификация видео.

**Для цитирования:** Городенко Р.Д., Петров С.А., Баланев К.С. Метод оценки классификационных моделей для анализа видеопотока // Вестник Российской нового университета. Серия: Сложные системы: модели, анализ, управление. 2025. № 2. С. 123 – 129. DOI: 10.18137/RNUV9187.25.02.P.123

---

## METHOD FOR EVALUATING CLASSIFICATION MODELS FOR VIDEO STREAM ANALYSIS

---

**Abstract.** In recent years, the relevance of computer vision technologies has significantly increased, especially in tasks related to video stream analysis, such as fatigue detection, face spoofing prevention, and gesture recognition. In well-known approaches for evaluating models used in these tasks, the ROC AUC metric is widely used. However, its application, based on frame-by-frame analysis, has certain limitations. These limitations are related to the instability of results due to inconsistent model outputs between frames and the lack of consideration for temporal dependencies in the data, which reduces the accuracy of the evaluation in real-world streaming video usage. The proposed methodology addresses these issues by constructing ROC curves for the video as a whole, rather than for individual frames. This process is implemented through the analysis of model outputs on each frame, allowing for more reliable differentiation between positive and negative examples at the level of the entire video segment. This approach enables a more realistic assessment of the trade-off between false positives and recall, and it also improves the sensitivity of models by selecting more relevant classification thresholds. Thus, the new methodology offers a more relevant evaluation of models for streaming video processing, ensuring more accurate event recognition and their contextual interpretation, which is especially important in critically significant applications.

**Keywords:** computer vision, video stream processing, ROC AUC, model evaluation, dynamic analysis, contextual interpretation, spoofing prevention, perception of temporal dependencies, video classification.

**For citation:** Gorodenko R.D., Petrov S.A., Balanov K.S. (2025) Method for evaluating classification models for video stream analysis. *Vestnik of Russian New University. Series: Complex Systems: Models, analysis, management*. No. 2. Pp. 123 – 129. DOI: 10.18137/RNUV9187.25.02.P.123 (In Russian).

### *Анализ предметной области*

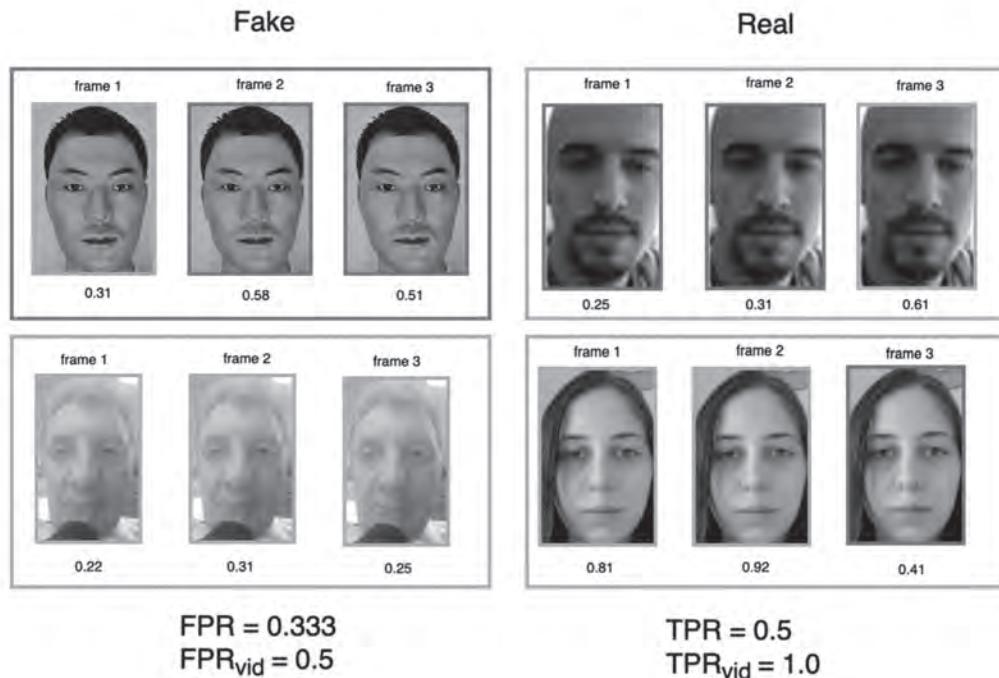
Современные технологии компьютерного зрения находят широкое применение в задачах классификации и обработки видеопотока [1], что вызывает возрастающий интерес исследователей к методам оценивания их эффективности. Задачи, в которых используемые методы могут быть крайне полезны, включают в себя такие направления, как обнаружение усталости (drowsiness detection) [2], защита от подделок лиц (antispoofing detection) [3] и обнаружение жестов (gesture detection) [4]. Кроме того, методы применимы в областях, связанных с подсчетом людей в толпе, распознаванием лиц и отслеживанием объектов. Однако, несмотря на широкий спектр приложений, большинство современных методов оценивания моделей компьютерного зрения основываются на анализе качества на уровне отдельных кадров.

Основная проблема использования покадрового подхода заключается в его ограниченности, особенно при внедрении моделей для потоковой обработки кадров. Такие методы часто не отражают полную картину, так как не учитывают контекст и динамику изменений в видеопотоке. Одной из причин этого является их неустойчивость к резким скачкам значения уверенности модели – показателей вероятности принадлежности ка-

дра к определенному классу – между соседними кадрами [5]. Вместо этого акцент должен быть сделан на оценивании модели на уровне видеопотока или его отдельных фрагментов, что позволит более точно оценить, как модель функционирует в реальных условиях использования.

Для задач, в которых скорость реакции модели и количество ложных срабатываний являются приоритетной метрикой, как, например, в задаче обнаружения подделок лиц [6], модель должна оперативно реагировать на реальные угрозы, чтобы это не привело к значительным последствиям. Поэтому существует необходимость разработки и применения методов оценивания, фокусирующихся на анализе модели в процессе обработки видеоконтента, что позволит улучшить точность и надежность моделей в приложениях реального времени.

В качестве примера на Рисунке 1 приведена визуализация того, как вычисляются метрики полноты (TPR) и доли ложных срабатываний (FPR) при решении задачи распознавания подделок лиц.



**Рисунок 1.** Пример подсчета метрик при решении задачи распознавания подделок лиц  
для двух случаев: покадровый и на уровне видео  
Источник: рисунок выполнен авторами.

Метрики подсчитаны на искусственном наборе данных из четырех видеофрагментов, каждый из которых состоит из трех кадров, на которых была запущена модель. Красная рамка кадра означает некорректное распознавание, зеленая – корректное. Аналогичным образом нанесены рамки для каждого видеофрагмента. Под каждым кадром приведена уверенность в положительном классе (живость лица), выдаваемая моделью. Порог по уве-

ренности выбран 0,5; это означает, что если уверенность модели больше этого числа, то кадру присваивается положительная метка, то есть произошла активация модели на кадре (и, соответственно, на всем видеофрагменте). Метрики рассчитаны двумя способами: покадровые и на уровне видеофрагмента (обозначено через суффикс *-vid*). Как можно заметить, значение  $FPR_{vid}$  хуже  $FPR$ , при этом  $TPR_{vid}$  лучше  $TPR$ . Причина в том, что метрики с суффиксом *-vid* более чувствительны к активациям, происходящим хотя бы на одном кадре всего видеофрагмента. Чтобы снизить  $FPR_{vid}$  можно повысить порог до 0,6. В этом случае  $FPR_{vid} = 0$  и  $TPR_{vid} = 1$ , в то время как в случае покадрового подсчета  $FPR = 0$  и  $TPR = 0,5$ , что не отражает реальной картины при потоковой обработке видео.

### ROC-кривая для оценки классификационных моделей (см. Рисунок 2)

ROC AUC (Receiver Operating Characteristic – Area Under the Curve) – это метрика, часто используемая для оценки качества бинарных классификаторов [7]. Она позволяет оценить вероятность того, что классификатор правильно ранжирует случайно выбранную пару объектов из разных классов. ROC-кривая строится по параметрам истинно положительных (True Positive Rate – TPR) и ложноположительных (False Positive Rate – FPR) показателей для различных порогов бинаризации.

Основные этапы построения ROC-кривой включают следующие шаги.

1. Классификатор присваивает каждому объекту вероятность принадлежности к положительному классу (скор).
2. Для каждого возможного порога скоры сравниваются с этим порогом, и объекты классифицируются как положительные или отрицательные.
3. На основе полученных классификаций рассчитываются значения TPR и FPR.
4. Построение ROC-кривой происходит путем нанесения TPR против FPR для всех порогов, а AUC измеряет площадь под этой кривой.

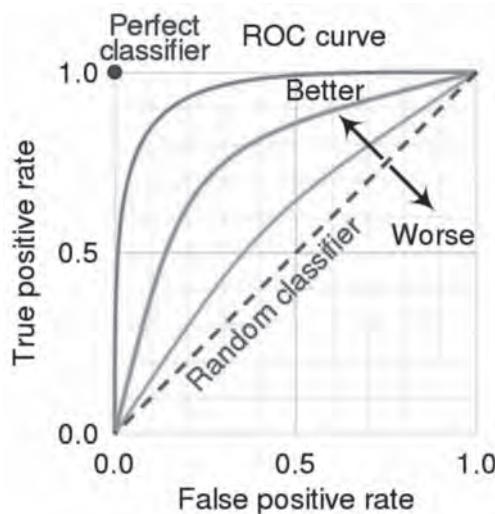


Рисунок 2. ROC-кривая

Источник: Wikimedia Foundation. URL: [https://upload.wikimedia.org/wikipedia/commons/thumb/1/13/Roc\\_curve.svg/250px-Roc\\_curve.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/1/13/Roc_curve.svg/250px-Roc_curve.svg.png) (дата обращения: 15.03.2025).

К преимуществам ROC AUC можно отнести следующие:

- позволяет оценить качество классификации независимо от распределения классов;
- предоставляет информацию обо всех возможных порогах, а не только об одном конкретном, что делает оценку более универсальной;
- оценивая TPR и FPR, дает комплексное представление о способности модели различать между положительными и отрицательными классами.

Тем не менее существуют проблемы, связанные с использованием ROC AUC, особенно при применении к покадровой оценке моделей в стриминговой обработке видео. В условиях видеопотока, где результаты классификации могут быть нестабильными между последовательными кадрами, это часто приводит к резким скачкам уверенности модели, что негативно сказывается на устойчивости и интерпретируемости AUC. Кроме того, ROC AUC не учитывает временные зависимости и последовательные корреляции между кадрами, хотя в потоковой обработке важно реагировать на события в контексте всей последовательности изображений, а не смотреть на отдельные кадры. Также в покадровых моделях, которые используются в задачах безопасности как обнаружение поддельного лица, значимость различных ошибок может существенно различаться в зависимости от контекста, но ROC AUC не может должным образом учесть это различие, так как рассматривает все ошибки как равнозначные.

### Предлагаемое решение

Чтобы наилучшим образом применить ROC AUC при анализе видео, предлагается метод построения ROC-кривой, адаптированный для видеопотоков. Предлагается строить кривую не по выходам классификатора на отдельных кадрах, а по агрегированным значениям для отдельных видео, состоящих из кадров. В данном случае по оси  $Oy$  откладываются «положительные» отсечки (число видео с позитивной меткой), а по оси  $Ox$  – «отрицательные» (число видео с негативной меткой). Видео упорядочиваются в соответствии с максимальным значением уверенности в положительном классе среди всех кадров одного видео согласно формуле

$$score_v = \max_{f \in v} score_f,$$

где  $score$  – вероятность принадлежности к положительному классу;  $v$  – видео;  $f$  – кадр. На основе полученных  $score_v$  и имеющихся меток видео  $label_v$  строится ROC-кривая. В каждой ее точке можно определить показатели TPR и FPR для видео с учетом подобранного порога.

Для видео с негативной меткой акцент делается на максимальном значении уверенности модели среди всех кадров видео – это опорная точка, от которой зависит, будет ли пример (видео) рассмотрен как ложное срабатывание. Если в качестве порога выбрать значение больше этого значения, то видео уже не будет рассматриваться как ложное срабатывание. Для видео с положительной меткой достаточно активации хотя бы на одном кадре, то есть активация на видео также определяется максимальным значением уверенности модели. И если это значение выше заданного порога, то видео считается корректным срабатыванием.

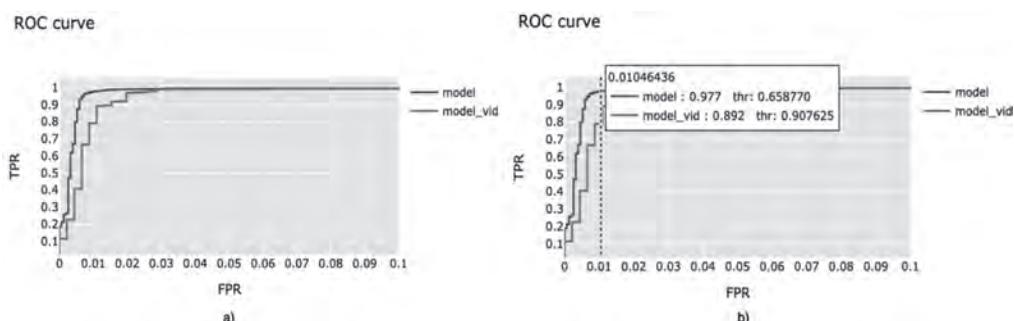
К преимуществам предложенного метода относится более корректная оценка модели в целевом случае использования в анализе видеопотока. Таким образом, получается реалистичное представление о работе модели в реальных условиях, где срабатывание модели определяется только частью кадров. Анализ на уровне видео обеспечивает более устойчивую оценку моделей, так как снижает влияние резких скачков значений уверенности на

отдельных кадрах, которые могут искажать результат при классическом подходе. Метод также обеспечивает более релевантные пороговые значения, поскольку кривая строится на основе максимального значения уверенности среди всех кадров видео, что позволяет корректно оценивать метрики точности для видео.

Однако у метода есть и недостатки. Он требует большего количества уникальных видео для получения устойчивых (статистически значимых) ROC-кривых, что может стать проблемой при отсутствии доступа к большим наборам видеоданных. Более того, анализ длинных видео может потребовать значительных вычислительных ресурсов, так как необходимо обрабатывать и получать выходы модели на каждом кадре (с учетом задержки модели). Если видеоданные неоднородны или содержат значительные вариации в длительности, это может потребовать дополнительного времени обработки, что необходимо учитывать в условиях ограниченных сроков.

Для проверки валидности нового метода был взят набор данных для предсказания подделок лиц “Live ness Detection – Zalo AI Challenge 2022”<sup>1</sup>. На этом наборе данных был обучен классификатор, после чего были построены ROC-кривые для тестовой части данных двумя способами: классическим методом по кадрам (model) и по предложенному методу (model\_vid). Результаты представлены на Рисунке 3. Можно заметить, что ROC AUC для model\_vid и model различаются. Кроме того, график для model\_vid менее гладкий, чем для model, так как он построен на меньшем количестве примеров: видеофрагментов значительно меньше, чем кадров. На Рисунке 3, б приведены значения TPR с соответствующими порогами в фиксированной точке FPR = 1 %. Можно заметить, что для достижения FPR = 1 % на уровне видео необходимо повысить пороговое значение с 0,66 до 0,91.

Стоит учитывать, что при использовании метода, помимо полноты обнаружения позитивных примеров, необходимо принимать во внимание метрику задержки. Задержка определяется как время между кадром с фактическим появлением позитивного примера и кадром с активацией модели при выбранном пороге. Эта метрика важна, поскольку позволяет оценить быстроту отклика модели, что особенно критично для приложений реального времени, где своевременное обнаружение существенно влияет на общую эффективность системы.



**Рисунок 3.** ROC-кривые для стандартного покадрового подсчета TPR/FPR (model) и на уровне видео (model\_vid)

Источник: рисунок выполнен авторами.

<sup>1</sup> *Live ness Detection – Zalo AI Challenge 2022* // Kaggle.com. URL: <https://www.kaggle.com/datasets/hlly34/liveness-detection-zalo-2022> (дата обращения: 15.03.2025).

Также для приближения замеров к реальным условиям при разбиении видеофрагментов на кадры стоит учитывать задержку самой модели. Вместо анализа всех кадров из видеофрагмента отбирается каждый  $n$ -й кадр, где  $n$  вычисляется по формуле  $n = \text{fps} \times \text{latency}$ . Например, если работа модели ( $\text{latency}$ ) занимает 0,1 сек. и частота кадров видеопотока ( $\text{fps}$ ) 30 кадров/сек., следует анализировать каждый 3-й кадр.

### Заключение

Внедрение нового метода построения ROC-кривой, адаптированной для анализа видеопотока, представляет собой усовершенствование в оценивании моделей компьютерного зрения, предназначенных для работы в условиях стриминговых данных. Подход позволяет обеспечить более точное и контекстуально обоснованное оценивание, минимизируя влияние нестабильных выходов модели и улучшая релевантность порогов классификации. Несмотря на необходимость в большем количестве видеоданных и возможные временные затраты на обработку, главным преимуществом является более корректная оценка модели, что имеет особенное значение в критически важных приложениях реального времени. Таким образом, метод можно считать полезным инструментом оценивания моделей в арсенале исследователей и разработчиков, работающих в области компьютерного зрения, в частности, в обработке видео.

### Литература / References

1. Patil P., Pawar V., Pawar Y., Pisel S. (2021) Video Content Classification using Deep Learning. *arXiv*. DOI: <https://arxiv.org/abs/2111.13813>
2. Sengar S.S., Kumar A., Singh O. (2024) VigilEye – Artificial Intelligence-based Real-time Driver Drowsiness Detection. *arXiv*. DOI: <https://arxiv.org/pdf/2406.15646>
3. Hao Fang, Ajian Liu, Jun Wan, Escalera S., Chenxu Zhao, Xu Zhang, Stan Z. Li, Zhen Lei (2023) Surveillance Face Anti-spoofing. *arXiv*. DOI: <https://arxiv.org/abs/2301.00975>
4. Kapitanov A., Kvachiani K., Nagaev A., Kraynov R., Makhliarchuk A. (2022) HaGRID – Hand Gesture Recognition Image Dataset. *arXiv*. DOI: <https://arxiv.org/abs/2206.08219>
5. Harrington A., DuTell V., Tewari A., Hamilton M., Stent S., Rosenholtz R., Freeman W.T. (2023) Exploring perceptual straightness in learned visual representations. In: The Eleventh International Conference on Learning Representations – ICLR 2023. URL: <https://openreview.net/forum?id=4cOfD2qL6T> (accessed 15.03.2025).
6. Zhuo Wang, Zeheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, Zhongyuan Wang (2022) Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing. *arXiv*. DOI: <https://arxiv.org/abs/2203.05340>
7. Gneiting T., Vogel P. (2018) Receiver Operating Characteristic (ROC) Curves. *arXiv*. DOI: <https://arxiv.org/abs/1809.04808>

Поступила в редакцию: 24.04.2025

Received: 24.04.2025

Поступила после рецензирования: 21.05.2025

Revised: 21.05.2025

Принята к публикации: 07.06.2025

Accepted: 07.06.2025