GRAPHICS & MEDIA LAB
VIDEO GROUP

О проведенном крупном челлендже детекторов дипфейков

Ватолин Дмитрий Сергеевич

ISP RAS Research Center for Trusted Artificial Intelligence
CS MSU Graphics&Media Lab

Об авторе

- Заведущий Graphics&Media Lab ВМК МГУ и AI Video Lab ИИИ МГУ
- Создатель сайтов по алгоритмам
  - https://compression.ru/video
  - https://videoprocessing.ai
  - https://videoprocessing.github.io/
- Области интересов: современное сжатие видео, измерение качества видео, четырехмерное видео
- Руководил 40+ проектами с компаниями **Intel, Cisco, Samsung, Huawei, Broadcom** и др.
- Автор №1* на Habr.com в хабах **«AR и VR»**, **«Работа с видео»** и **«Видеотехника»,** автор №2* **«Искусственный интеллект»**
- Сомневается в разумности Homo Sapiens

* Без учета редакторов

Работа с компаниями

- **90% of our projects** are sponsored by companies
- We have experience of **long-term collaboration** with Intel, Samsung, Huawei and other
- All our research is aimed to be **extremely practical** for industry

and many others...

Наши ключевые результаты

- 20+ лет анализа видеокодеков (до 5 отчетов в год по 400 страниц и 20000 графиков)
  compression.ru/video/codec_comparison
- Топ-1 Метрики оценки качества 3D-видео (18 метрик, больше всех в мире)
  videoprocessing.ai/stereo_quality
- Топ-1 Коллекция бенчмарков обработки видео (18 бенчмарков)
  videoprocessing.ai/benchmarks
- Провели два крупнейших в мире конкурса по сжатию без потерь (50 и 200 тысяч евро) globalcompetition.compression.ru www.gdcc.tech
- Наиболее известный студент **Карен Симонян** (**автор VGG**, Business Insiders's Top 100 AI persons 2023)
- Наиболее успешный исследователь **Анастасия Анциферова** (премия «Лидер ИИ 2025»)
- 14 A* статей за 5 лет (ICLR, ICML, AAAI, NeurIPS)

GRAPHICS & MEDIA LAB
VIDEO GROUP

Челлендж дипфейков CVPR NTIRE 2026

Для справки:

CVPR — №1 по h5-index среди всех конференций и журналов в мире

NTIRE — “New Trends in Image Restoration and Enhancement” воркшоп в рамках CVPR, обьединяющий 44 челленджа в 2026

Детекция дипфейков

О проекте

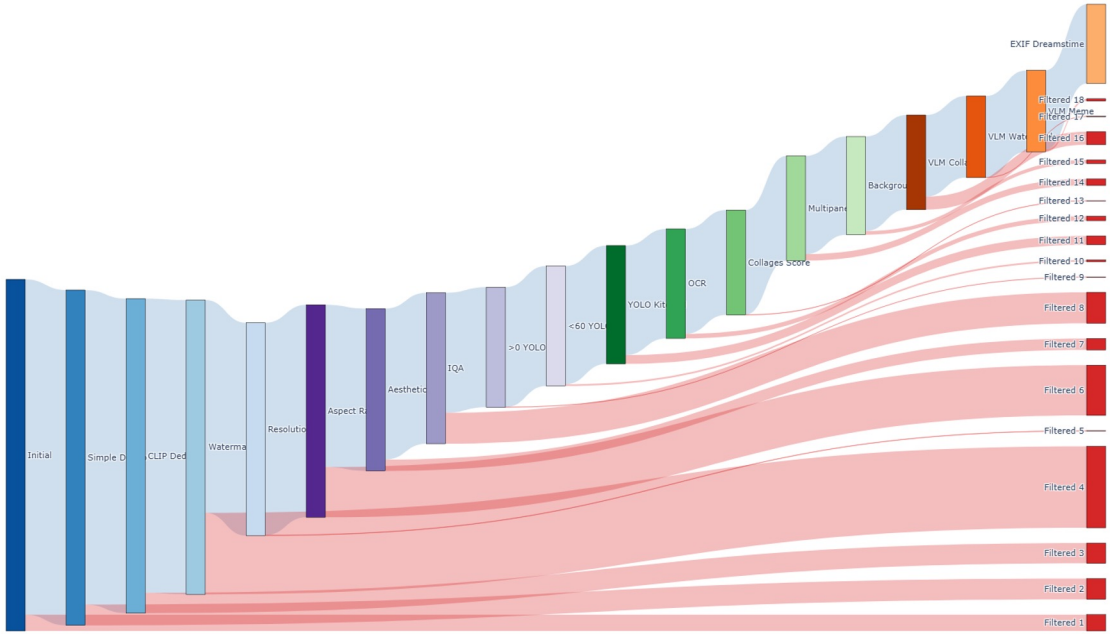
Мотивация: современные генераторы создают высоко-реалистичные изображения, при этом детекторы часто переобучаются под конкретные генераторы или выходят из строя при стандартной постобработке (сжатие, изменение размера, лёгкое редактирование)

Цель: создать челлендж, позволяющий наиболее полно и близко к реальным условиям оценить качество детекторов AI изображений

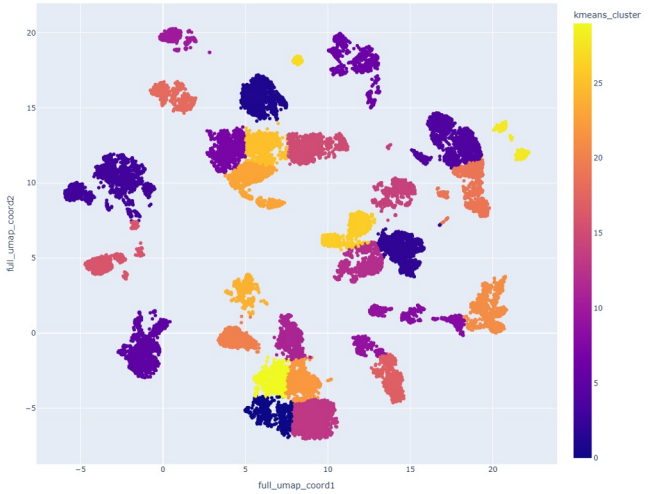
Примеры высоко-реалистичных AI изображений

Детекция дипфейков
Реальный датасет

- Взят достаточно большой датасет на 13 млн изображений
- Проведена фильтрация: убраны дубликаты, изображения с вотермарками, по разрешению, по качеству и другие
- Проведена кластеризации: по эстетике, качеству, эмбеддингу VLM тэгов, CLIP эмбеддингов, детектциям и



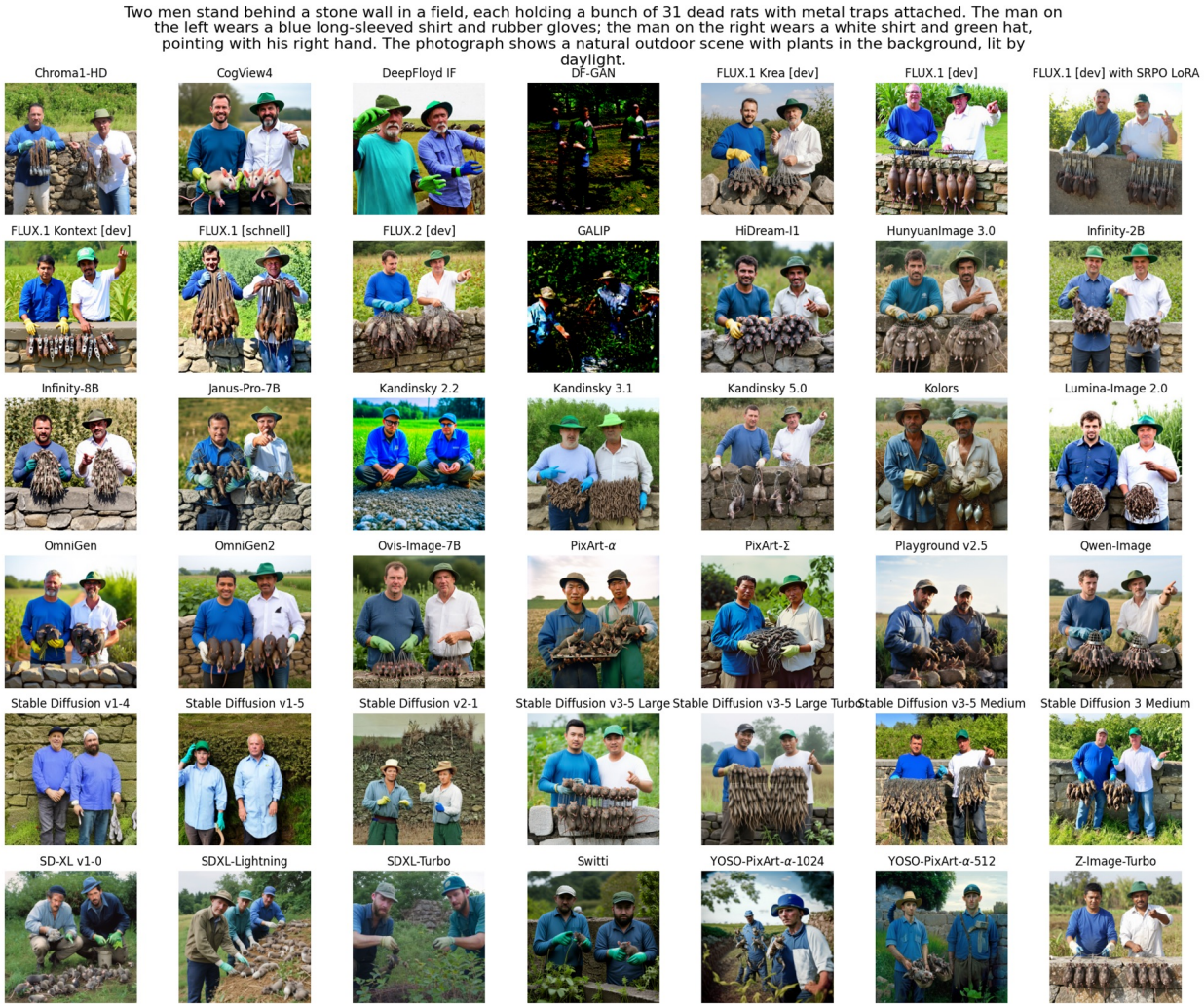
Пайплайн фильтрации (13M -> 2.9M)



Кластеризация датасета по части признаков

Детекция дипфейков

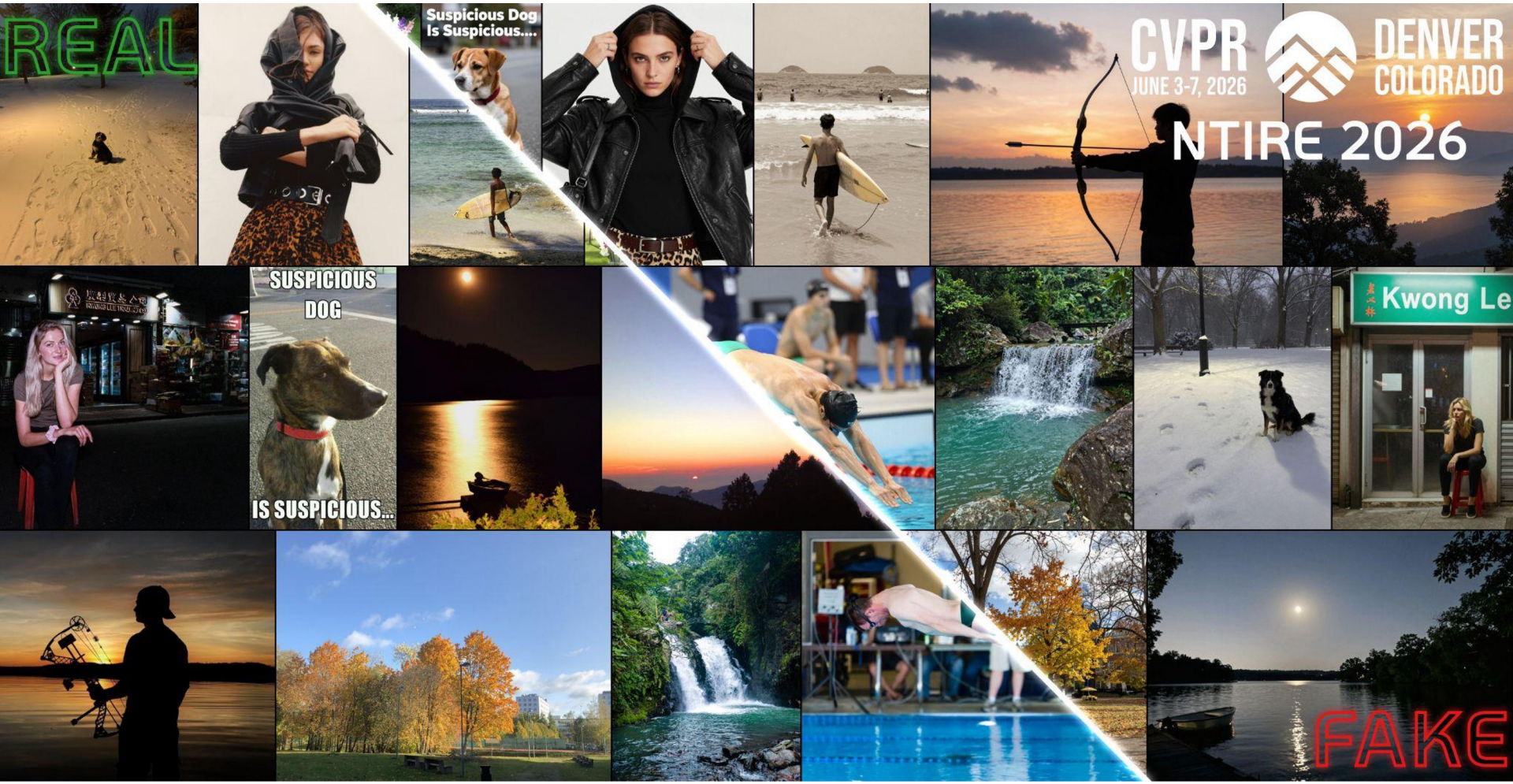
Фейковый датасет



- По полученным реальным изображениям с помощью VLM получены промпты
- Прогнано > 40 генераторов
- Сгенерировано > 9 млн фейковых изображений
- К реальным и фейковым изображениям добавлены искажения

Детекция дипфейков

Примеры данных



Детекция дипфейков

Примеры данных

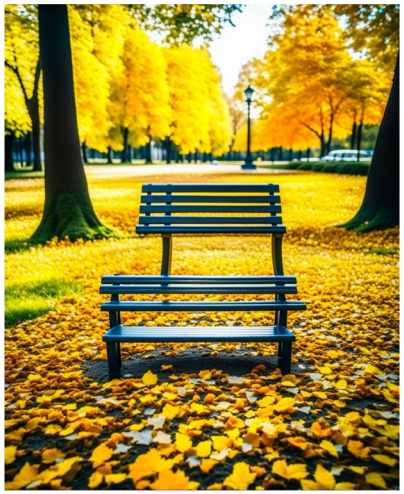
GRAPHICS & MEDIA LAB
VIDEO GROUP

Real

AI

Детекция дипфейков

Развитие генераторов



Скамейка

Несколько самолетов

Kandinsky 2
2023

Kandinsky 3
2024

Kandinsky 5
2025

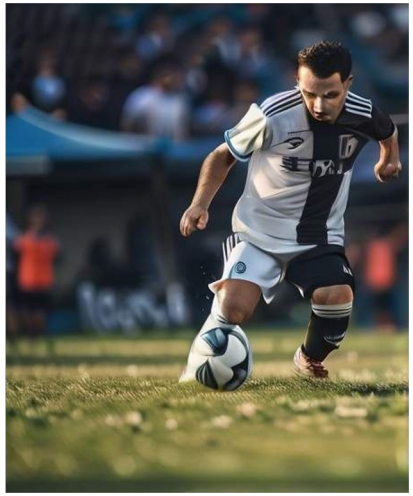
Детекция дипфейков

Фейл-кейсы

Ovis Image
2025

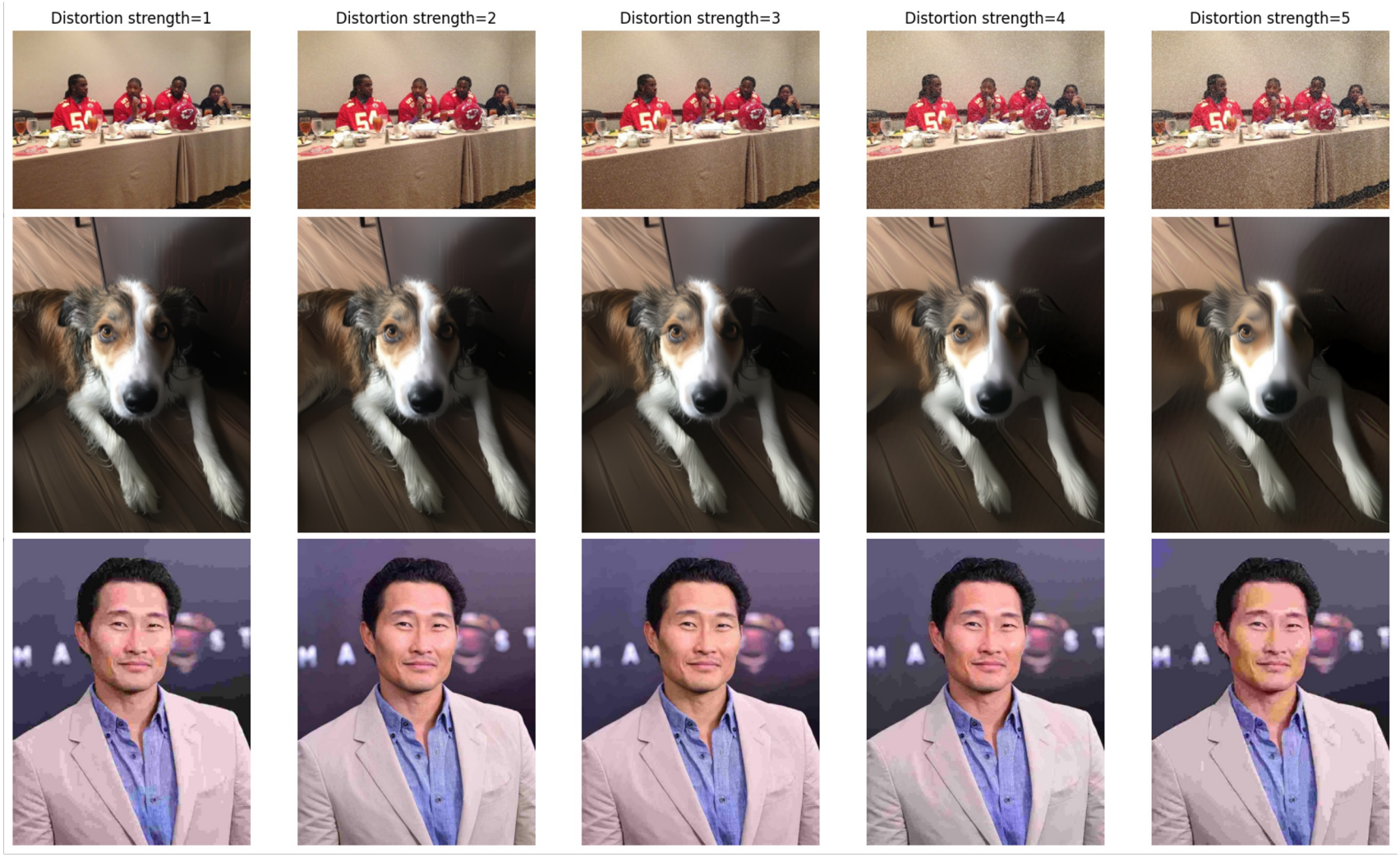
Kandinsky 3
2024

Stable
Diffusion
3.5 Large
2025



Playground
2.5
2024

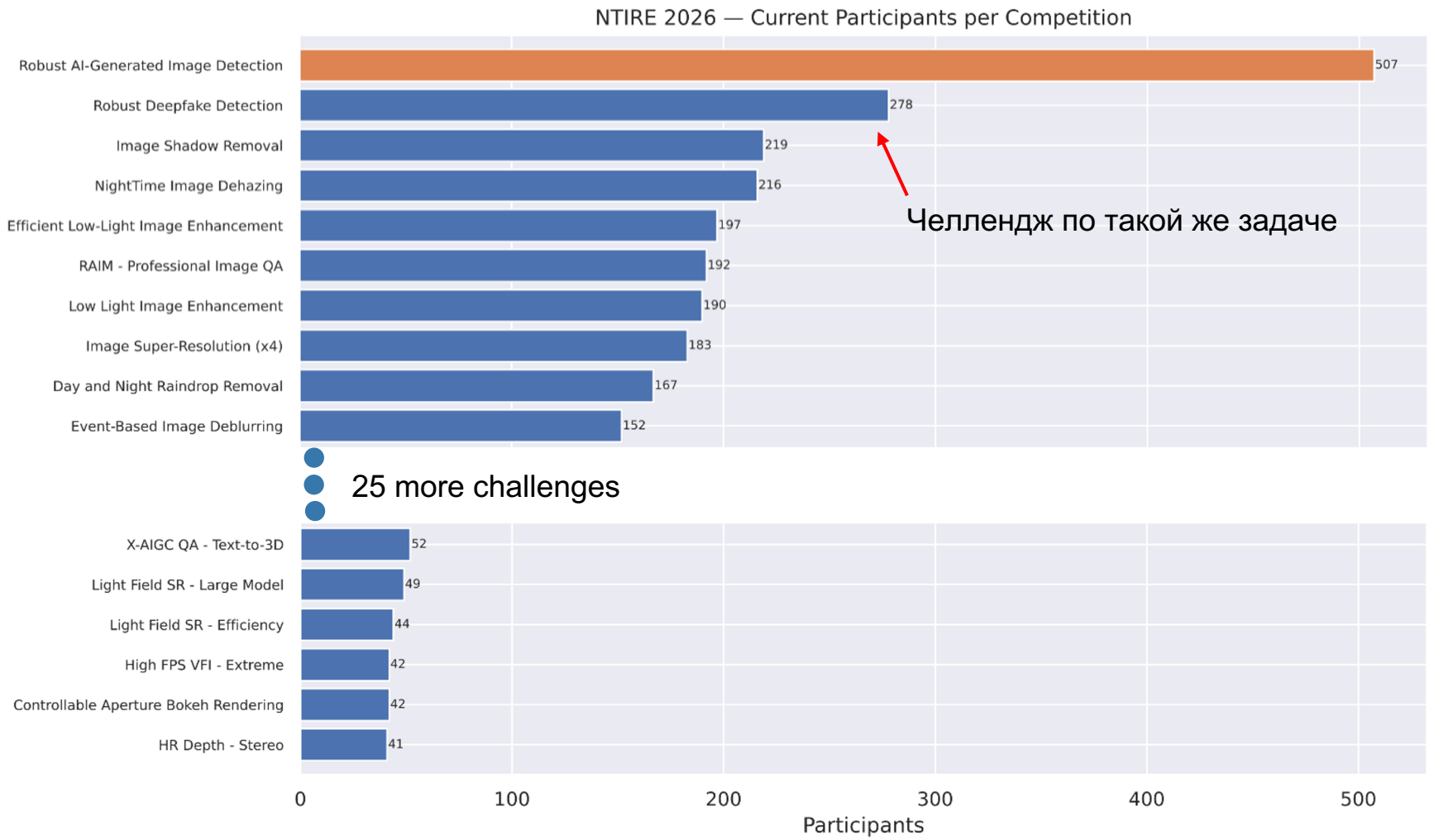
Детекция дипфейков
Примеры искажений



Сила искажения

Челлендж дипфейков

Количество участников

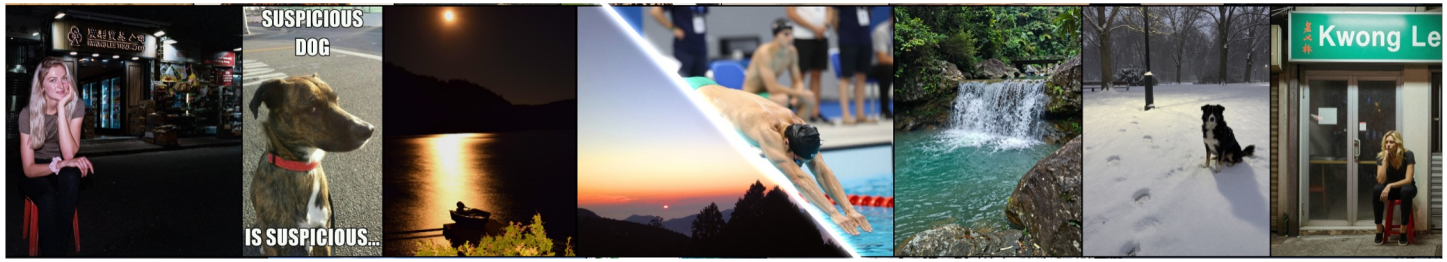


Челлендж дипфейков

Сравнение с соседним челленджом

GRAPHICS & MEDIA LAB
VIDEO GROUP

	Наш челелендж	Другой
Размер трейна	277к	1к
Размер теста	5000	2000
Домен картинок	General, лица, анимация, документы, текст, т.д.	Лица
Кол-во генераторов	35	~10?
Кол-во искажений	35	~10?



Примеры нашего датасета

Примеры датасета другого челленджа

Разница размеров датасетов у нас и конкурентов на 2 порядка

GRAPHICS & MEDIA LAB VIDEO GROUP

3100

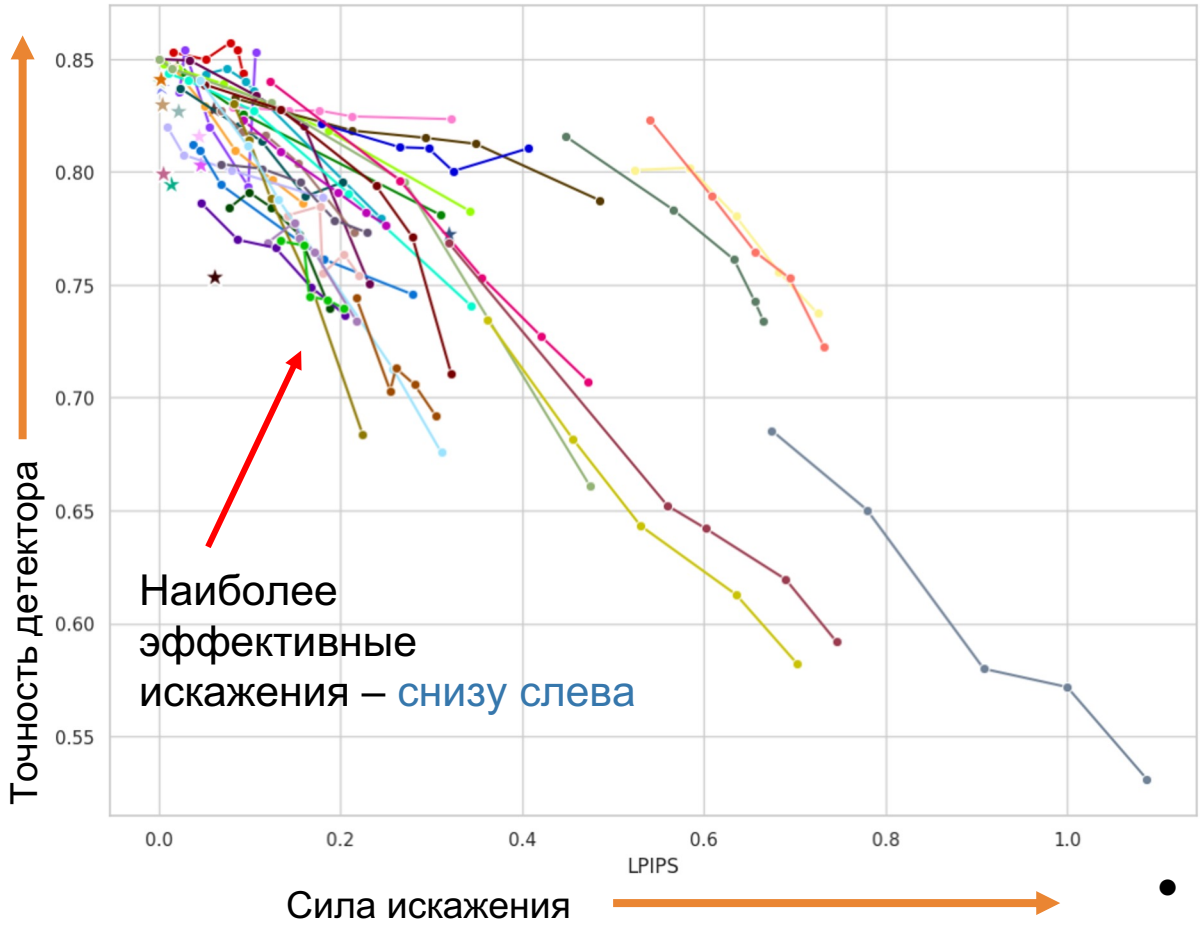
Split	Samples	Labels Provided	Degradations	Usage Policy
Training	1,000	Yes	Simple	Primary training data.
Validation	100	No	Additional	Daily testing; **Do not train on these.**
Public Test	1,000	No	Different	Final leaderboard ranking.
Private Test	1,000	No	Different	Manual verification & fairness check.

Split	Images	Labels Provided	Generator models	# Transformations	Intended usage
Toy dataset	1,000	No	10	0	Get familiar with data and submission format
Train	~277,000	Yes	20	0*	Primary training data.
Validation 1st	10,000	No	25	5 (simple)	Automatic evaluation during Development Phase.
Hard Validation	2,500	No	25	5 (simple)	Automatic evaluation during Development Phase. Closer to test compared to Val 1st.
Public test	2,500	No	30	7 (simple + complex)	Automatic evaluation during Testing Phase.
Private test	2,500	No	35	9 (simple + complex)	Manual verification & fairness check.

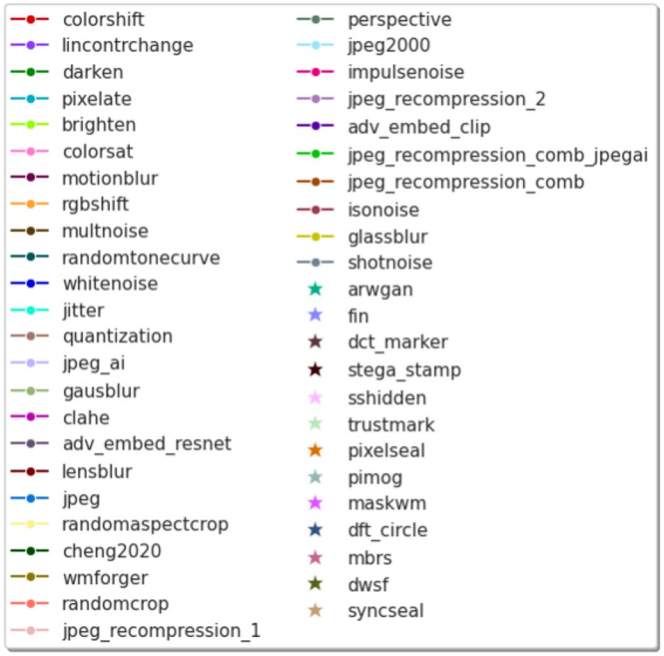
295500

Челлендж дипфейков

Влияние искажений на детекцию



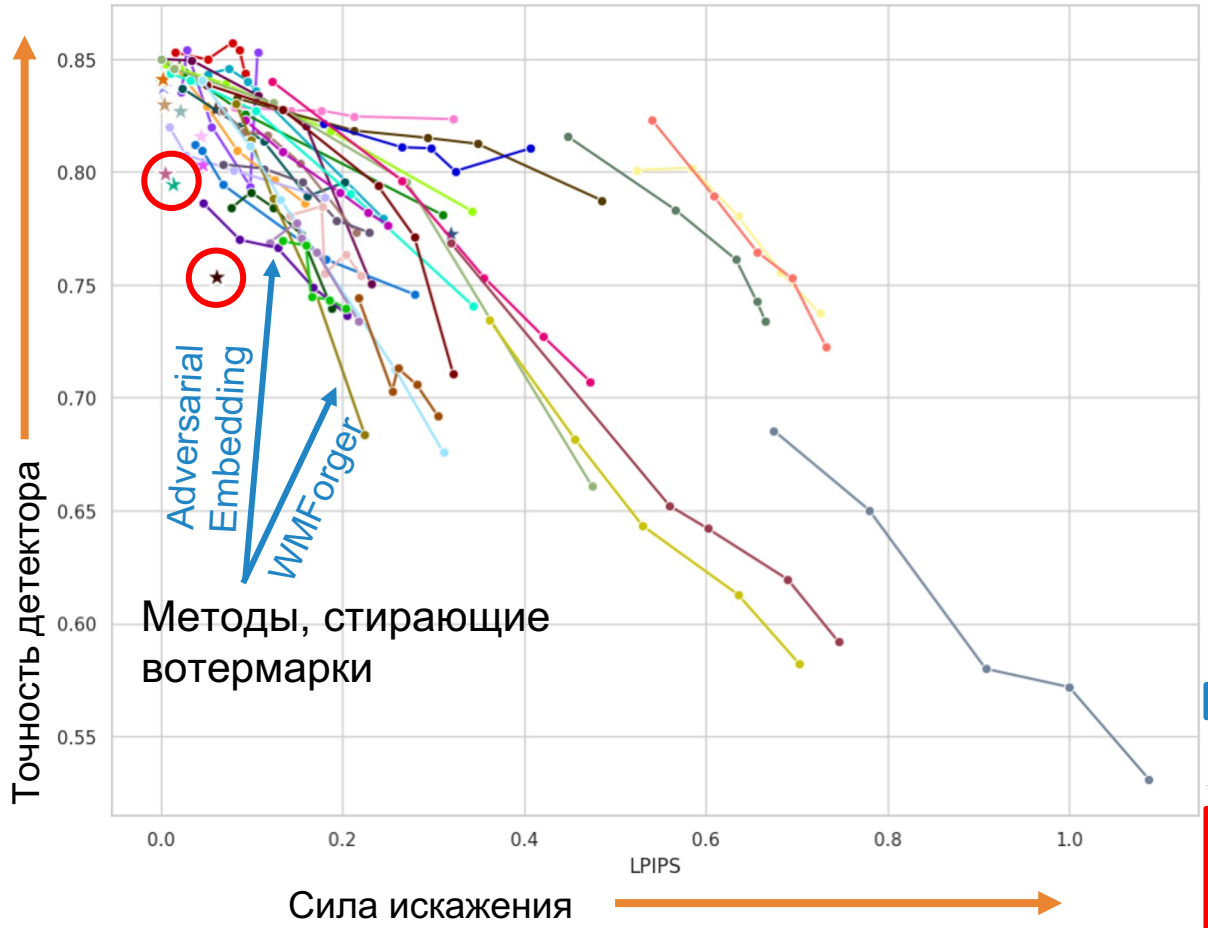
Протестировали >40 разных искажений:



- С ростом силы искажений точность детекции ИИ-картинок сильно падает
- Наиболее эффективный тип искажений – стиратели вотермарок генеративного контента

Челлендж дипфейков

Влияние искажений на детекцию

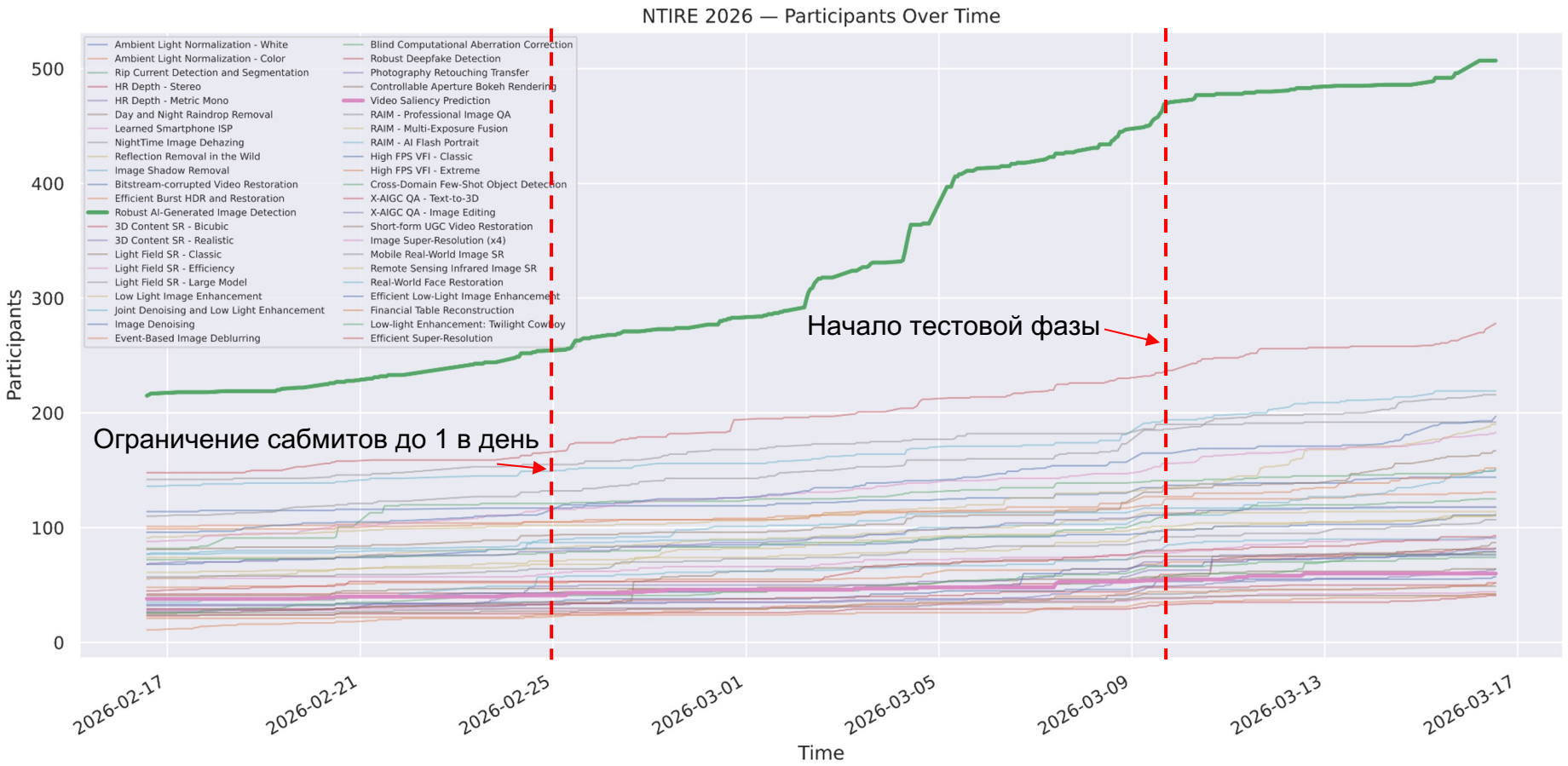


Протестировали >40 разных искажений:

★ – Методы незаметного маркирования изображений

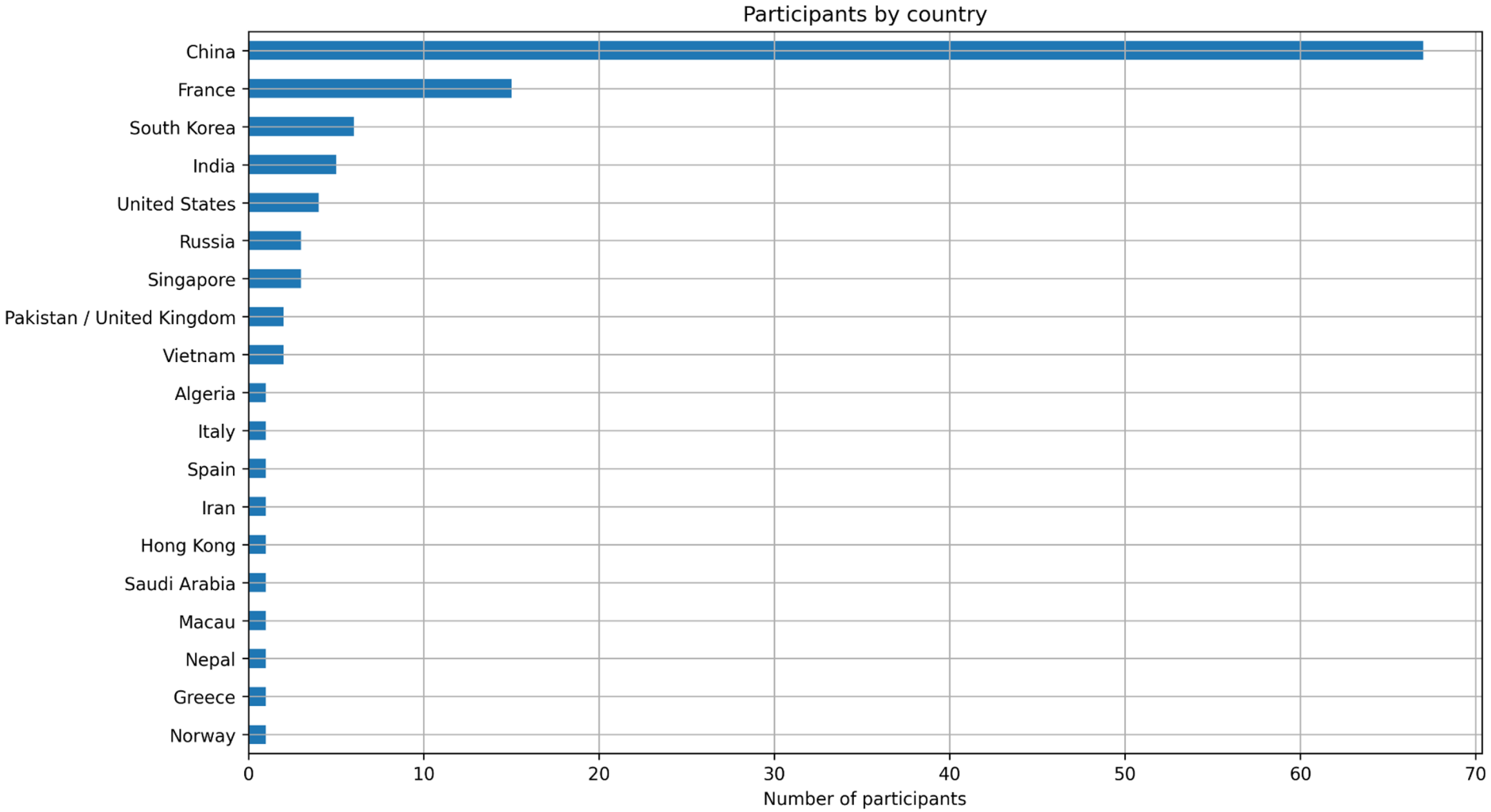
Челлендж дипфейков

Участники



Челлендж дипфейков

Участники по странам



Челлендж дипфейков

Участвующие компании и институты

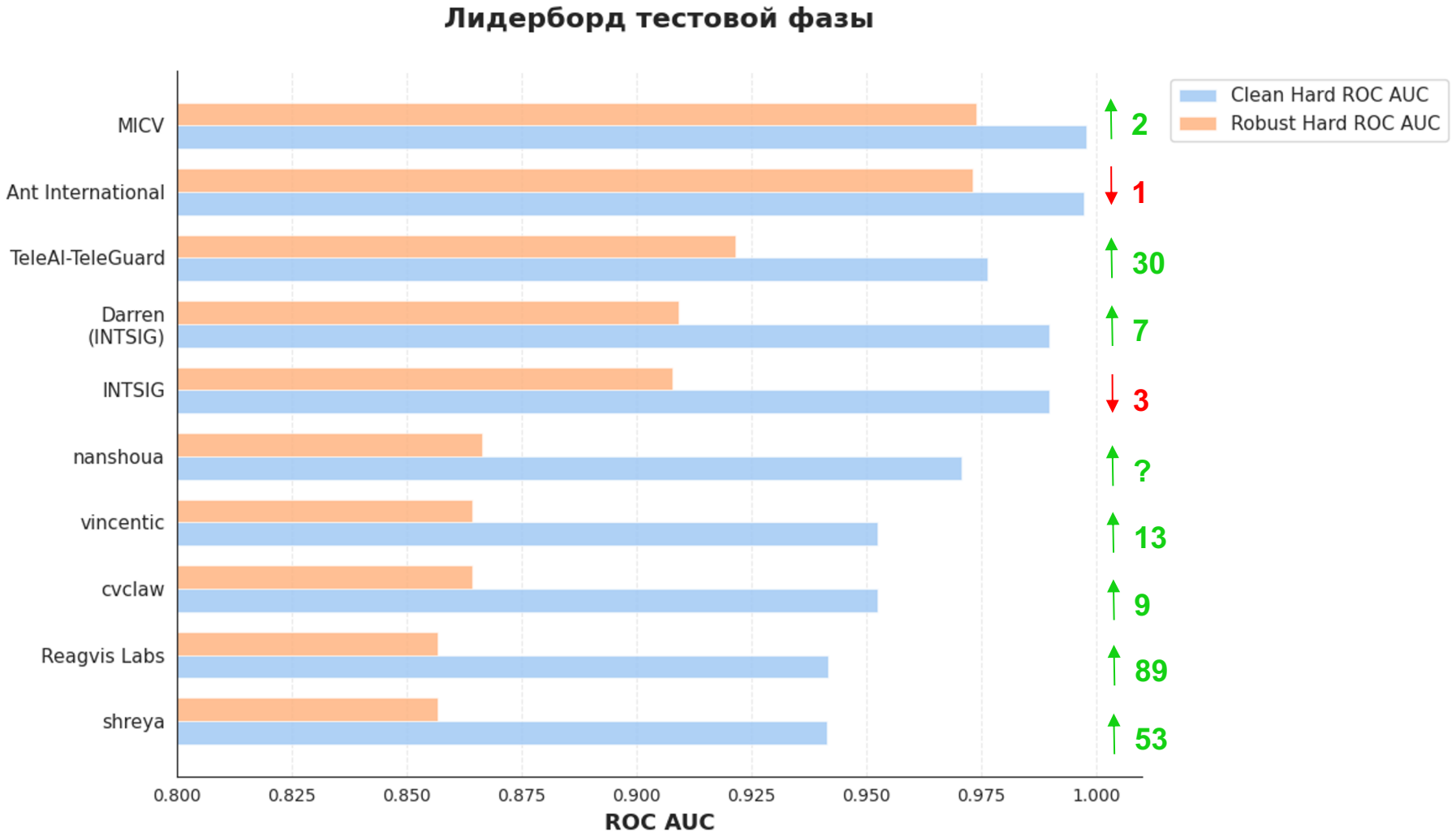
Отдельные компании:

- ByteDance (150 тысяч сотрудников, 155$ млрд выручки)
- Tencent (105 тысяч сотрудников, 92$ млрд выручки)
- ZTE (75 тысяч сотрудников, 15-17$ млрд выручки)
- Xiaomi (44 тысячи, 53$ млрд выручки)
- Ant Group (16 тысяч, ~17$ млрд выручки)

Отдельные университеты:

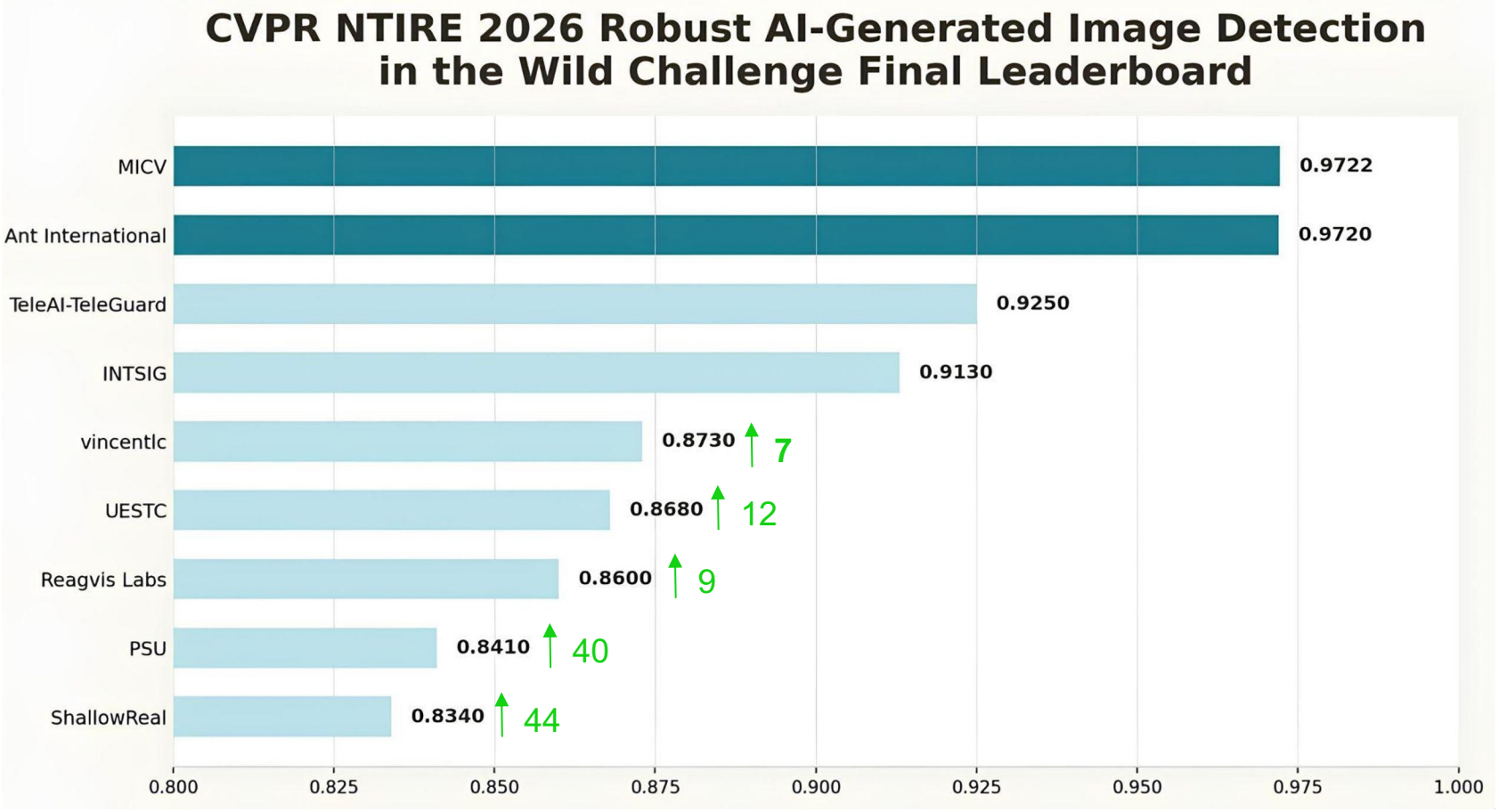
- Shanghai University
- Beijing University of Technology
- Xiamen University
- Nanjing University of Science
- Harbin Institute of Technology
- EURECOM (France)
- City University of Hong Kong
- Seoul National University
- National University of Singapore
- Norwegian University of Science

Челлендж дипфейков
Лидерборд тестовой фазы



Челлендж дипфейков

Лидерборд на скрытой выборке



Челлендж дипфейков

Финальный лидерборд; участники

1,2.  MICV, Ant International — **Две команды Ant Group, наиболее известные в мире продукты AliPay и AliExpress**

3.  **TeleAI-TeleGuard — TeleAI, подразделение China Telecom**

4.  **INTSIG — INTSIG Information Co, разработчики популярных программ CamScanner и CamCard**

5.  **vincentlc — Cong Luo,** индивидуальный участник

6.  UESTC — **команда University of Electronic Science and Technology of China (Ченду)**

7.  Reagvis Labs — индийский стартап, специализирующийся на детекте дипфейков

8.  PSU (ОАЭ) — **Prince Sultan University**

9.  **shallowReal — South China University of Technology, Zhejiang University (ZJU, Гуанчжоу)**

Передача на Россия 24 о проекте

Что есть еще и что дальше?

Реализовано:

- 35 состязательных атак на изображения и видео (и бенчмарк)
- 30+ защит (и бенчмарк)
- В процессе публикации самый большой в мире датасет для детекторов дипфейков

В планах:

- челлендж на редактирование картинок
- датасет для детекторов дипфейков в видео и челлендж
- стирание признаков генерации
- решение вопросов практического применения (скорость, новые генераторы и т.д.)

Contacts

Dmitriy Vatolin
e-mail: dmitriy@graphics.cs.msu.ru

- videoprocessing.ai/about
- compression.ru/video
- compression.ru/vqmt
- videocompletion.org
- videomatting.com
- subjectify.us
- evt.guru