MMG
MMG
MMG

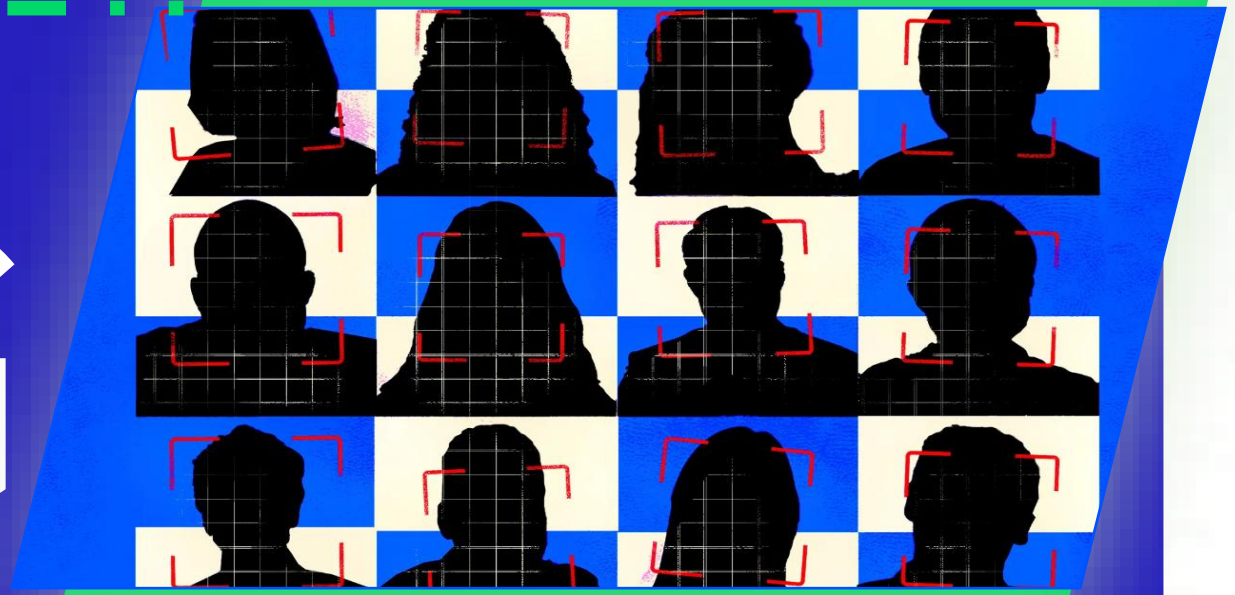
MMG

Команда Финансового Университета
при Правительстве РФ

FINANCIAL UNIVERSITY
UNDER THE GOVERNMENT OF THE RUSSIAN FEDERATION

×

ИТиАБД



Kryptonite
ML Challenge

“Ложь успевает обойти полмира,
пока правда надевает штаны.”

КРИПТОНИТ

XXX XXX XXX XXX XXX XX

О задаче

Задача обучить **энкодер**, работающий для изображений.
Он должен обладать следующими характеристиками:

Cosine similarity :

→ **1** для фото одного человека
→ **0** для фото разных людей
→ **0** для фото и дипфейка

Датасет

10.000людей
98.000фото

Метрика

EER

Содержание

MMG
MMG
MMG

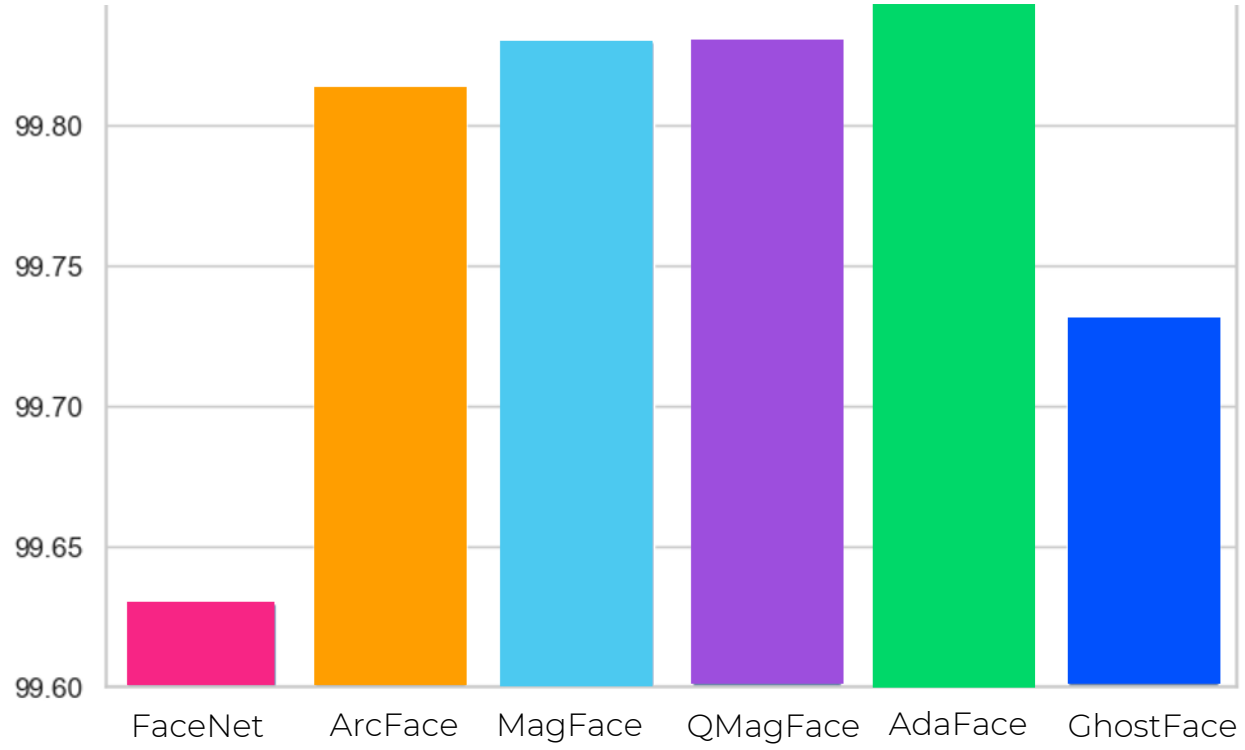
MMG
MMG
MMG

Модели

точность, скорость, дипфейки

FaceNet	2015
ArcFace	2019
MagFace	2021
QMagFace	2021
AdaFace	2022
GhostFace	2023

Бенчмарк LFW

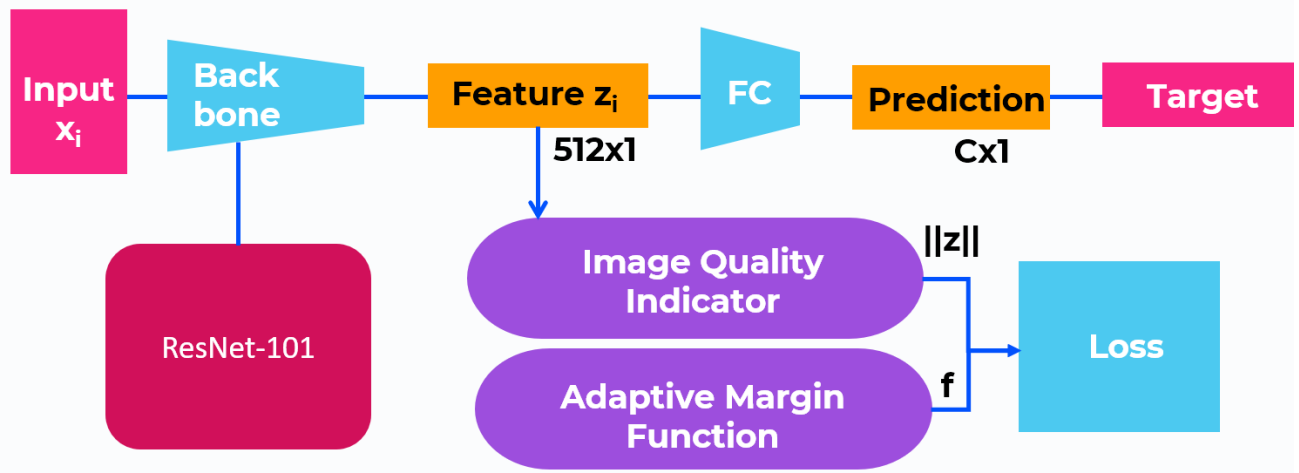


Разнообразие лосс-функций

- Contrastive Loss
- Triplet Loss
- Circle Loss
- Center Loss
- Hard Negative Mining
- Knowledge Distillation

MMG MMG MMG

AdaFace



$$\mathcal{L}_{\text{AdaFace}}(x_i) = -\log \frac{\exp\left(s\left[\cos\left(\theta_{y_i} + g_{\text{angle}}\left(\|\hat{z}_i\|\right)\right) - g_{\text{add}}\left(\|\hat{z}_i\|\right)\right]\right)}{\exp\left(s\left[\cos\left(\theta_{y_i} + g_{\text{angle}}\left(\|\hat{z}_i\|\right)\right) - g_{\text{add}}\left(\|\hat{z}_i\|\right)\right]\right) + \sum_{j \neq y_i} \exp\left(s\cos\theta_j\right)}$$

Где:

$$g_{\text{angle}}\left(\|\hat{z}_i\|\right) = -m \cdot \|\hat{z}_i\|, \quad g_{\text{add}}\left(\|\hat{z}_i\|\right) = m \cdot \|\hat{z}_i\| + m, \quad \|\hat{z}_i\| = \text{clip}\left(\frac{\|z_i\| - \mu_z}{\sigma_z / h}, -1, 1\right).$$

Loss

cos θ — косинусное сходство

y ∈ {0, 1} — метка

s — масштабный коэффициент

$\mathbf{m_b}$ — базовый margin

α — кф. динамического масштабирования

1. Динамический **margin**:

$$m_d = m_b + \alpha \cdot (1 - \cos\theta)$$

2. Определение **логита**:

$$\text{logit} = s \cdot \left(\cos\theta - m_d \cdot y\right) = s \cdot \left(\cos\theta - \left(m_b + \alpha \cdot (1 - \cos\theta)\right)y\right)$$

3. Функция потерь **Adaptive Margin Loss** :

$$\mathcal{L} = -\left[y \cdot \log\left(\sigma(\text{logit})\right) + (1 - y) \cdot \log\left(1 - \sigma(\text{logit})\right)\right]$$

MMG MMG

MMG

Aligner 1

1. MTCNN
2. RetinaFace
3. **CVLFace DFA**

Сбор и подготовка данных 2

Датасет: **CelebA**

10.000 персон

200.000 фото

Генерация **дипфейков**:

Original Arc2Face InstantID Ghost Roop

Пары 3

Real + **Real** = 1 (50%)

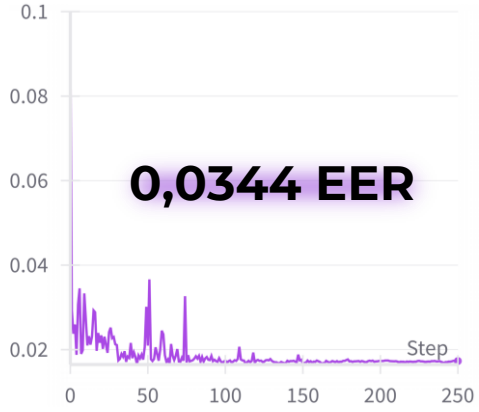
Real + **Fake** = 0 (25%)

Real + **Other** = 0 (25%)

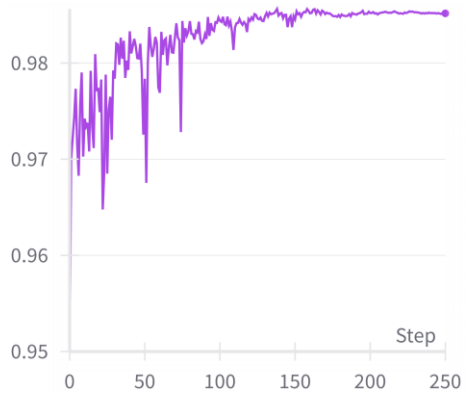
Пре трейн!

Обучение AdaFace 4

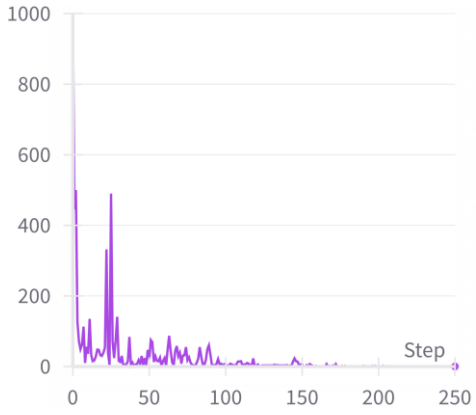
EER



val_accuracy



error_real_deepfake



MMG
MMG
MMG

xxx xxx xxx xxx

MMG
MMG
MMG

Датасет

Обнаруженная проблема:

~10% реальных фото неправильно размечены

000010

0.jpg

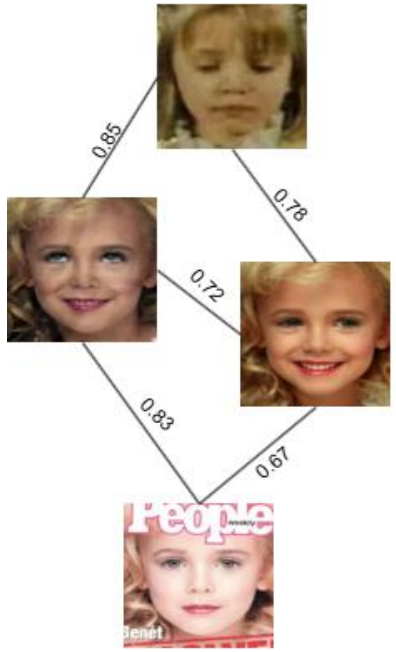
1.jpg

000204

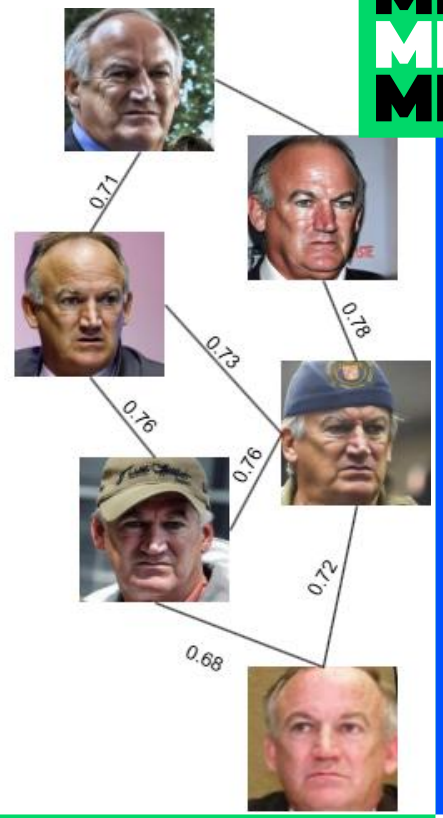
1.jpg

5.jpg

Папка: **000513**



Кластеризация лиц



Попытки очистить датасет:

1. Использование модели для построения матриц сходств
2. Графовая кластеризация по cosine similarity
3. Удаление выбивающихся изображений

Результаты:

1. Проблема некорректной разметки частично **решена**

2. Качество модели без привязки к лидерборду **выросло**

XXX
XXX
XXX
XXX
XXX
XXX
XXX

Обучение

Гиперпараметры

Epochs: **2**

Batch_size: **32**

Learning rate: **5e-6**

Checkpoint_inter: **100**

Base_margin: **0.35**

Dynamic_scale: **0.1**

Weight_decay: **1e-2**

Оптимизатор

AdamW:

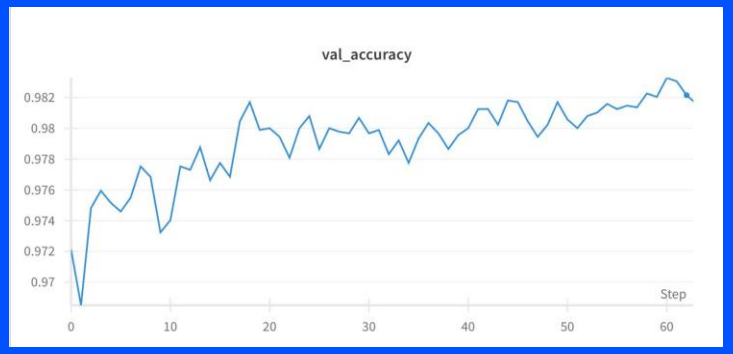
LR: **5e-6**
Weight decay: **1e-2**

Scheduler

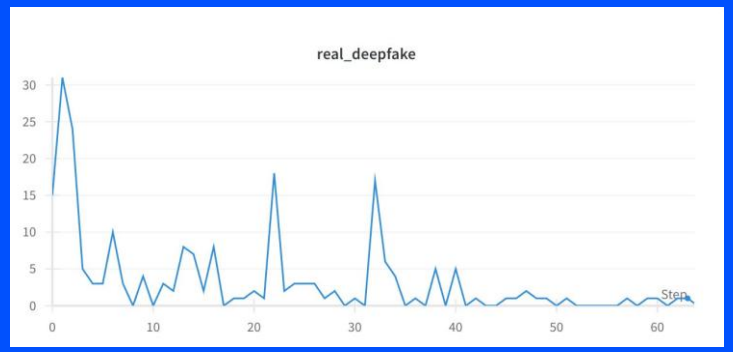
ReduceLROnPlateau:

Mode: **min**
Factor: **0.5**
Patience: **10**
Threshold: **1e-4**

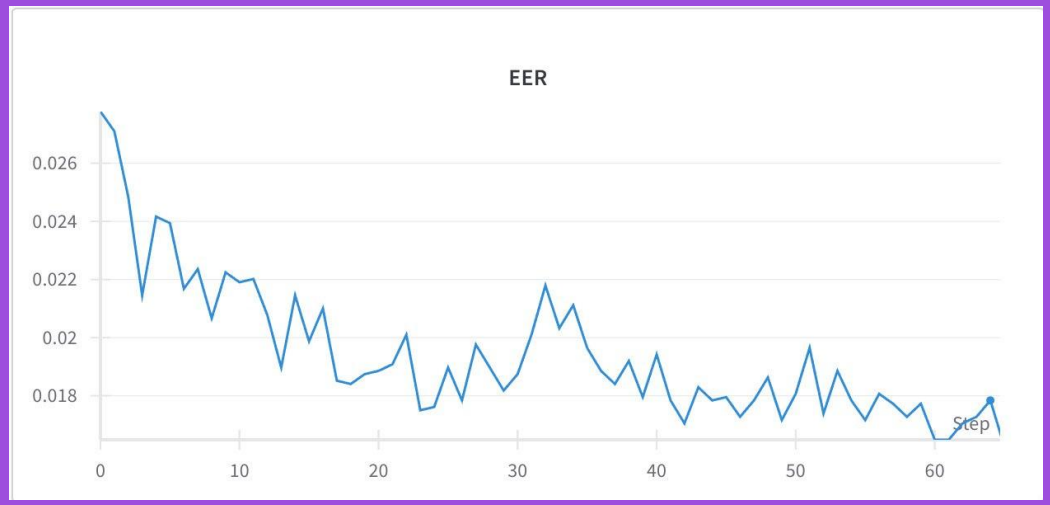
Validation threshold: **0.2**



val_accuracy



error_real_deepfake



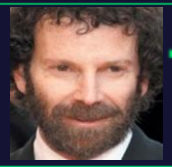
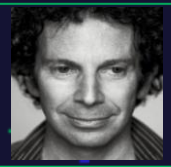
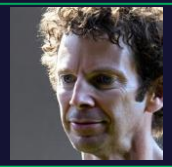
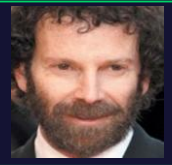
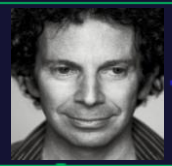
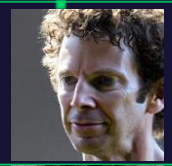
EER: 0.0264

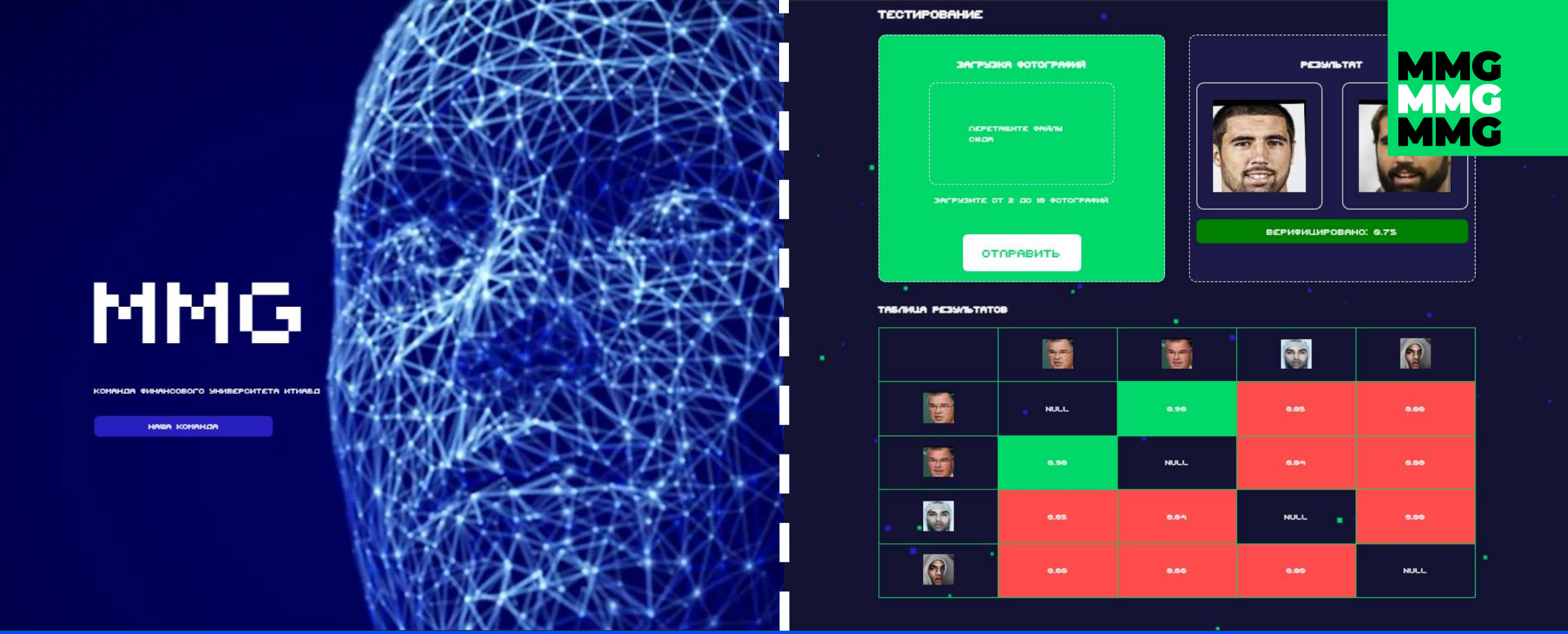
Результаты

Скорость:
RTX 4090 — ~**900** изображений в секунду

Стабильность:
Быстрый **энкодер**, стабильно работает даже на среднем «железе»

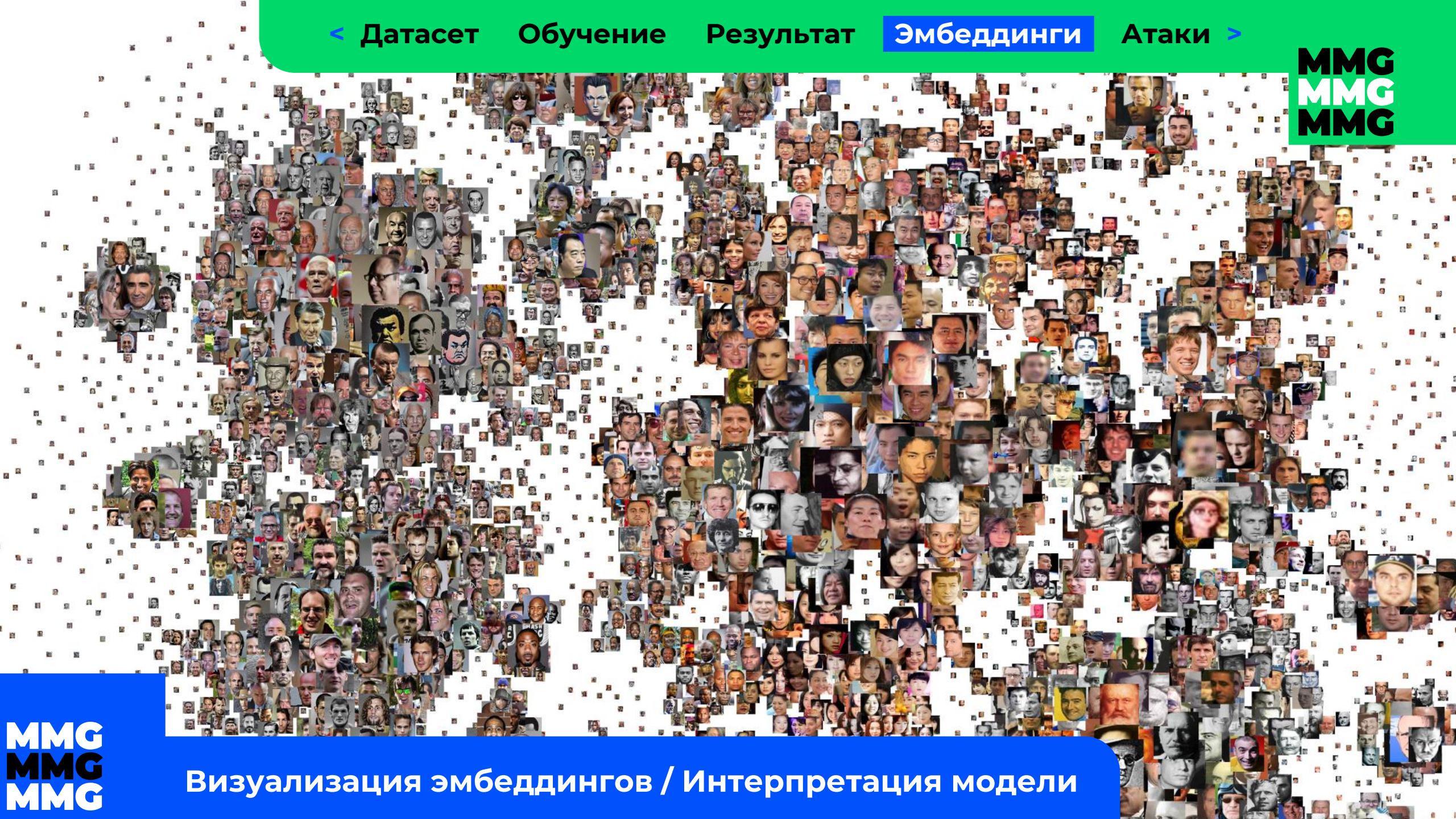
ТАБЛИЦА РЕЗУЛЬТАТОВ

						
	NULL	0.62	0.66	0.00	0.00	0.00
	0.62	NULL	0.60	0.00	0.00	0.00
	0.66	0.60	NULL	0.00	0.00	0.00
	0.00	0.00	0.00	NULL	0.96	0.95
	0.00	0.00	0.00	0.96	NULL	0.97
	0.00	0.00	0.00	0.95	0.97	NULL



Сайт для инференса

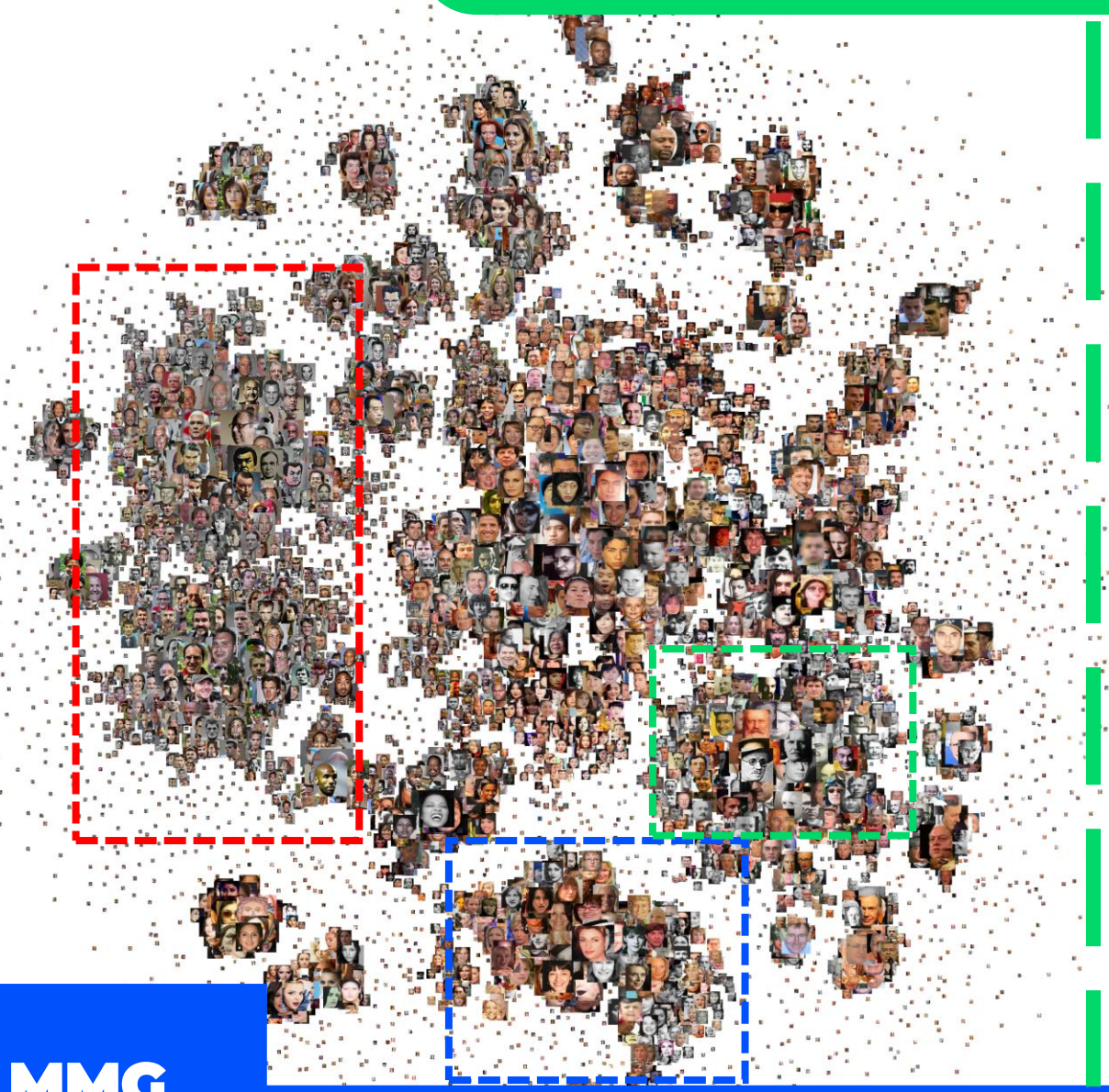
MMG
MMG
MMG

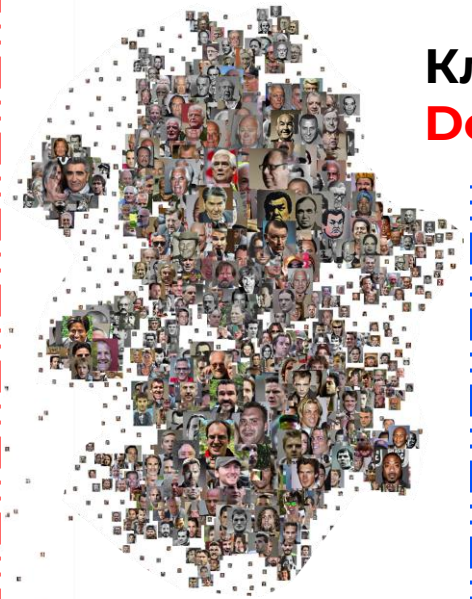


MMG
MMG
MMG

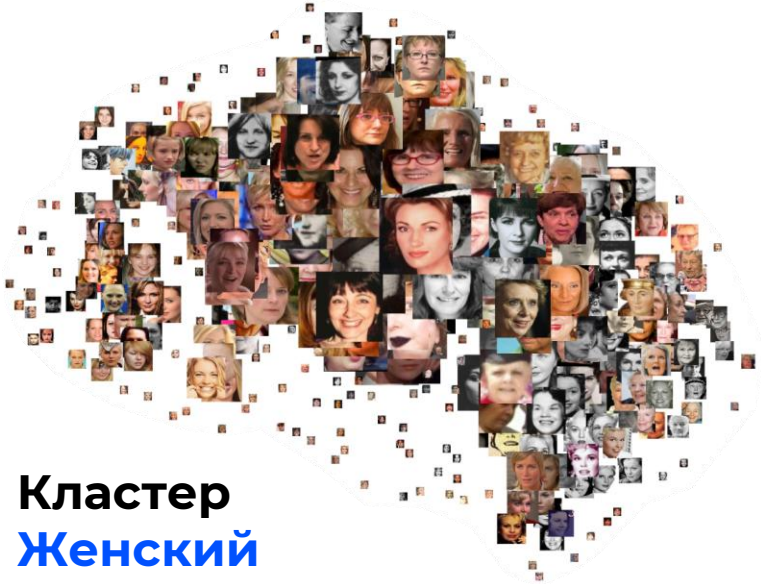
Визуализация эмбеддингов / Интерпретация модели

MMG
MMG
MMG





Кластер
DeepFake



Кластер
Женский

Кластер
Мужской

MMG
MMG
MMG

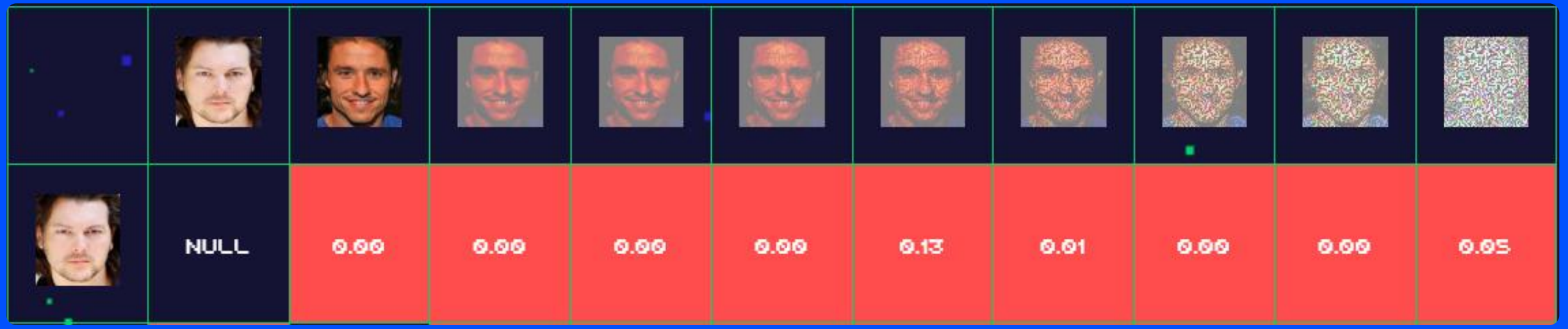
Визуализация эмбеддингов / Интерпретация модели

Атаки

ШУМ

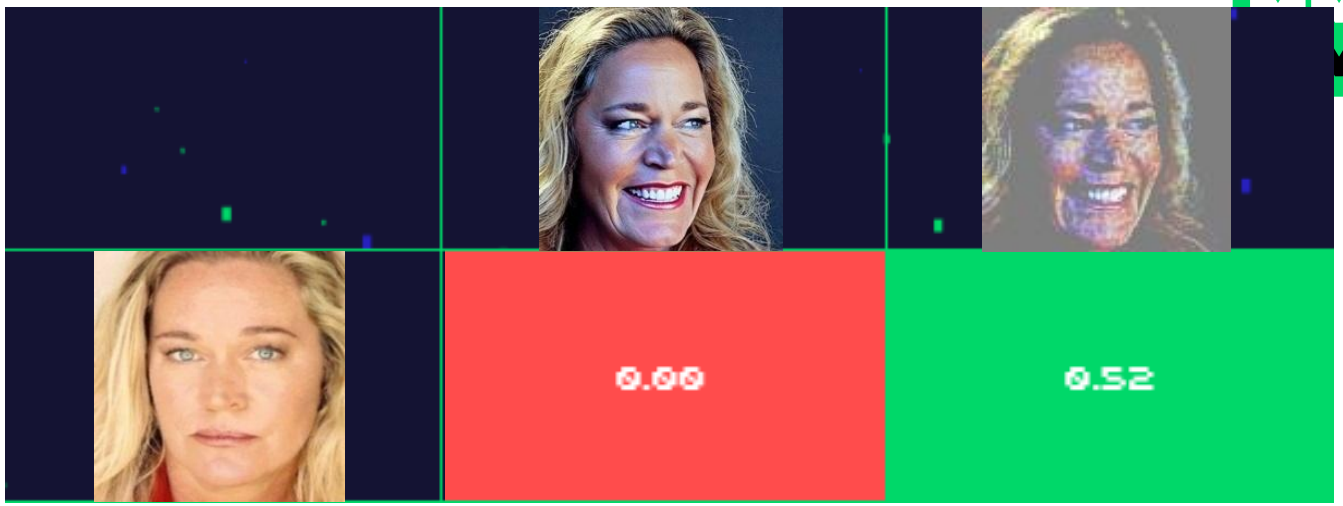
Вывод:

Шум может увеличить схожесть с другой фотографией, но это не оказывает критического влияния.

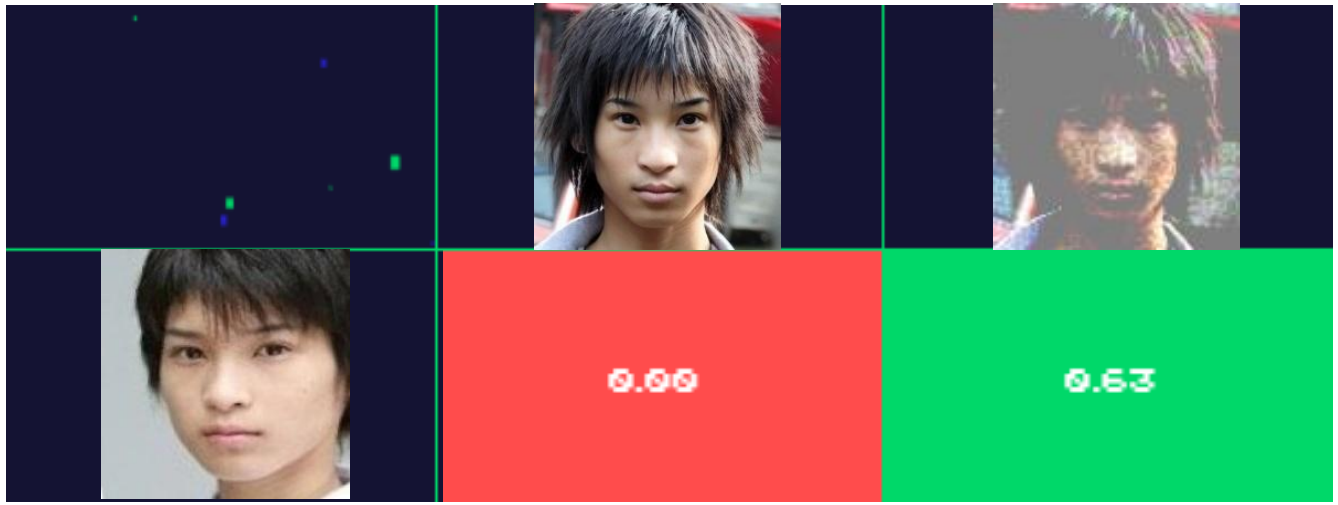


FGSM

Epsilon = 0.1



Real DeepFake FGSM



Real DeepFake FGSM

Тупики

Попробовали, но не использовали в финальной модели

Выводы:

1. **Глубокий ресерч – важно**
2. **Важность предобучения**
3. **Сложность ≠ качество**
4. **Важность предобучения**
5. **Агрессивная аугментация**

1

Triplet loss

Не дало прироста качества

2

Обучение без претрейна

Чекпоинт **Кима Минчула** не дал прироста по метрике

3

Заморозка модели

Полная заморозка — недообучение

Дополнительные ветки с **attention** - усложнение без улучшения

4

Гиперпараметры и оптимизаторы

Перепробовали множество конфигураций – не все давали улучшения

5

Аугментация Автолейблинг

Ухудшило качество модели

Масштаб

MMG
MMG
MMG

1. Использование более продвинутых и современных DeepFake и FaceSwap моделей

2. Увеличение количества обучающих данных

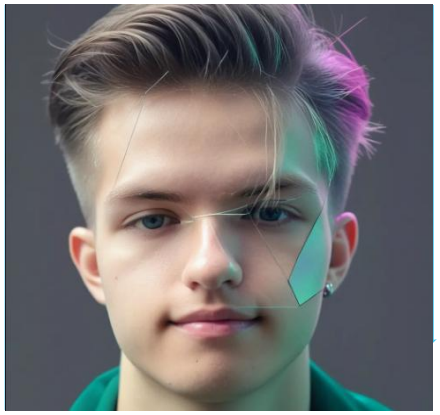
3. Борьба с FGSM-подобными атаками

MMG
MMG
MMG

Команда Финансового Университета
при Правительстве РФ

MMG

Артем Таратин



Денис Маликов

Даниил Аль-Натор

Илья Обухов

Команда **MMG** благодарит **Криптонит** за возможность участия в хакатоне!

КРИПТОНИТ

XXX XXX XXX XXX XXX XX

MMG
MMG
MMG

MMG
MMG
MMG