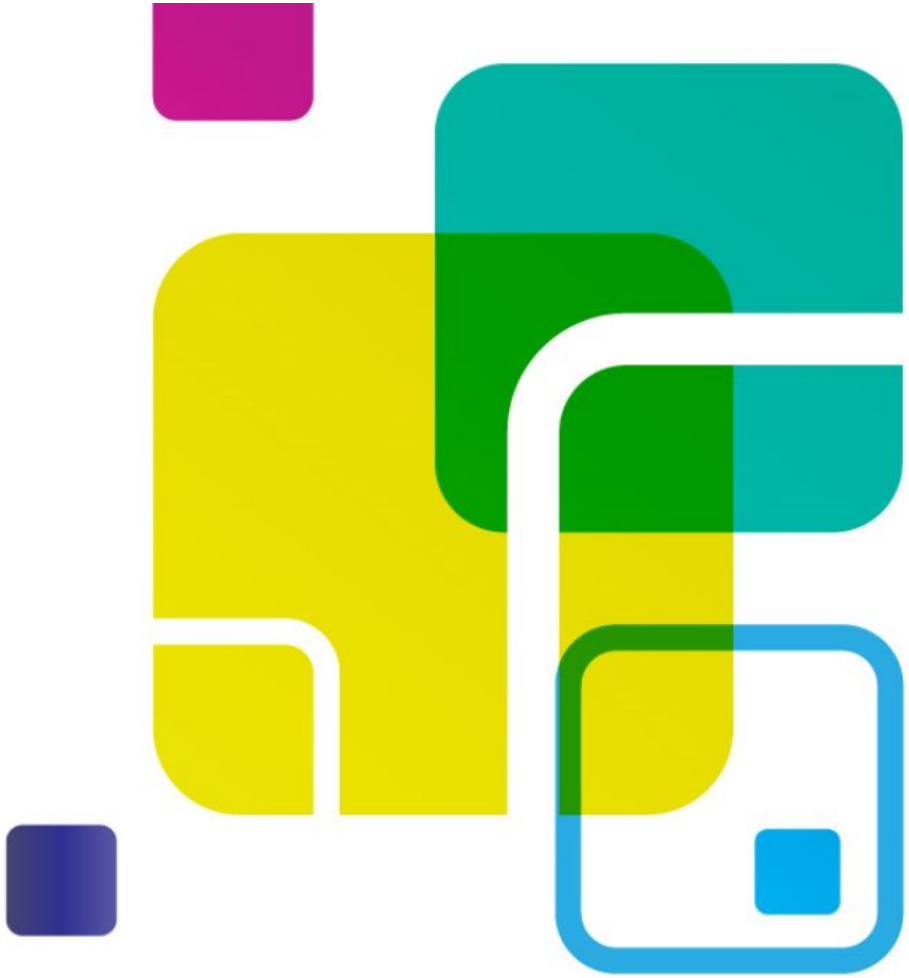
ID R&D



# Антиспуфинг в биометрических системах

# О Компании

## Технологическая миссия

- **Frictionless Biometrics**

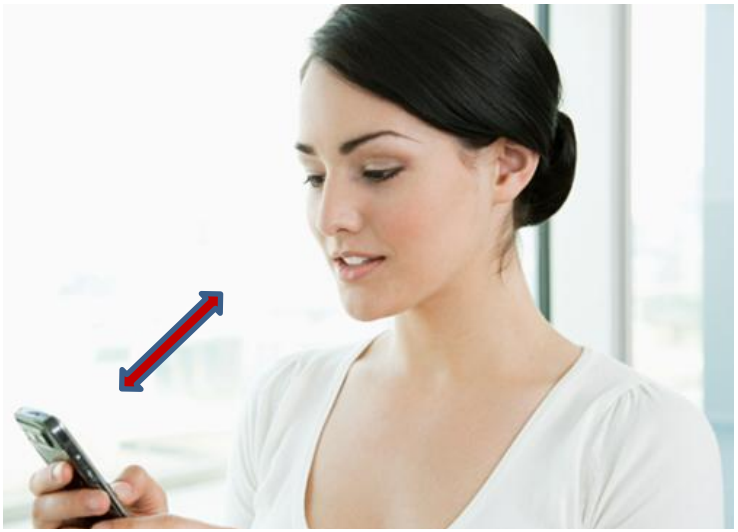
Разработка системы “узнавания” Пользователя без действий со стороны Пользователя

## Технологии

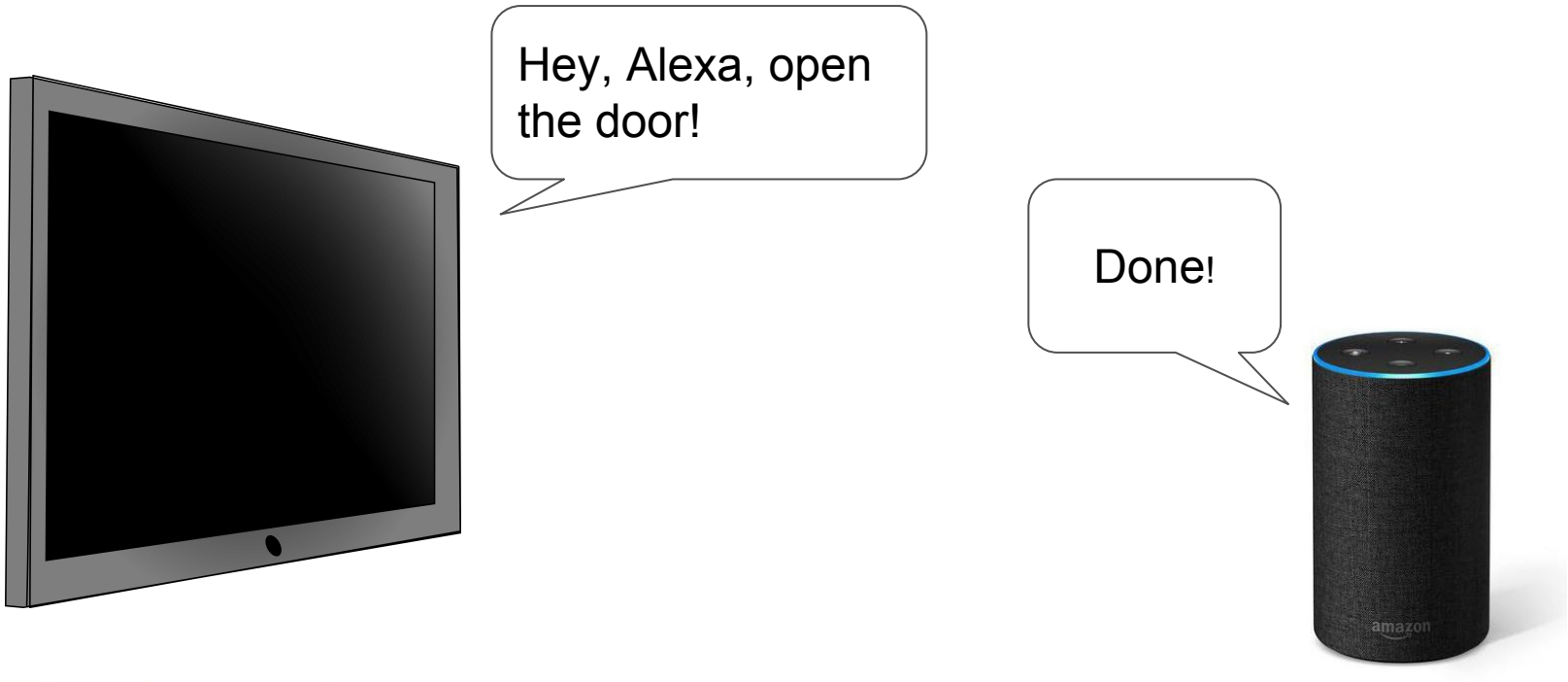
**Распознавание лиц/голоса/поведения**

Для

**Вирт Асс-ов/Чатботов/Мессенджеров**



# Спуфинг Виртуальных Ассистентов




TECH | AMAZON | CIRCUIT BREAKER

## Amazon's Alexa started ordering people dollhouses after hearing its name on TV

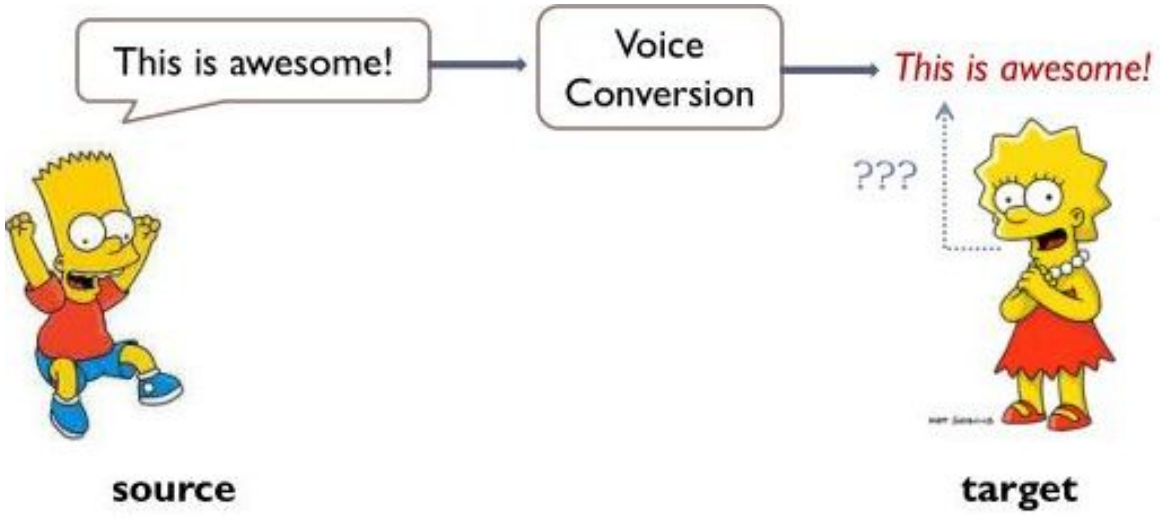
33

*Check your settings*

By Andrew Liptak | @AndrewLiptak | Jan 7, 2017, 5:52pm EST

# Способы Спуфинга

Запись с диктофона

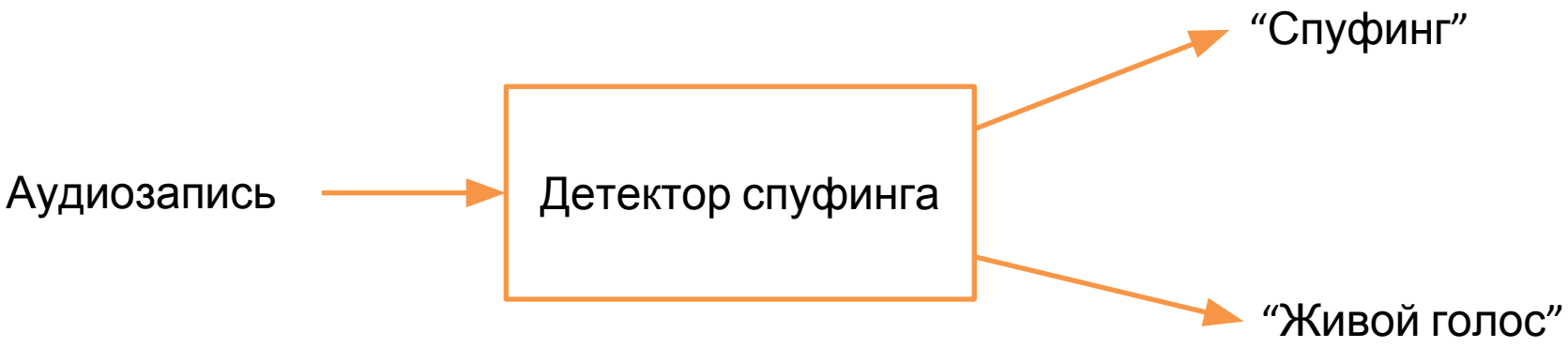



Конверсия голоса

# Постановка задачи

Детектор спуфинга должен определять, является ли голос **оригинальным** или **спуфингом**:

- Воспроизведенный с диктофона/радио/телевизора
- Сгенерированный синтезатором речи
- Сконвертированный из другого голоса



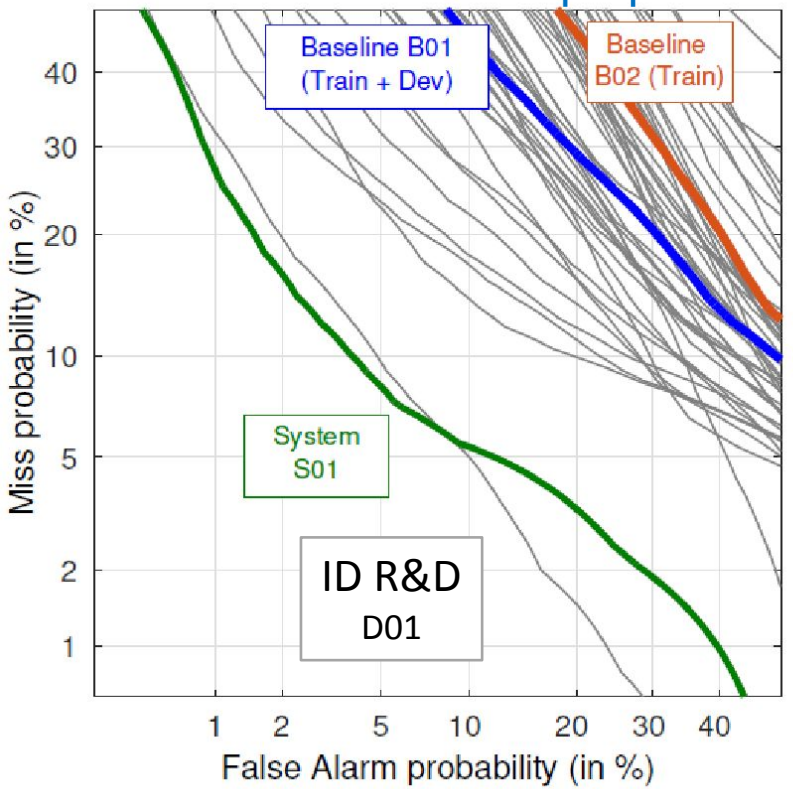
# Конкурс Алгоритмов Антиспуфинга

**ASVspoof 2017: Automatic Speaker Verification Spoofing and Countermeasures Challenge** был проведен с целью изучить возможности технологии детектирования ситуации воспроизведения аудиозаписи с диктофона/колонки/любого устройства

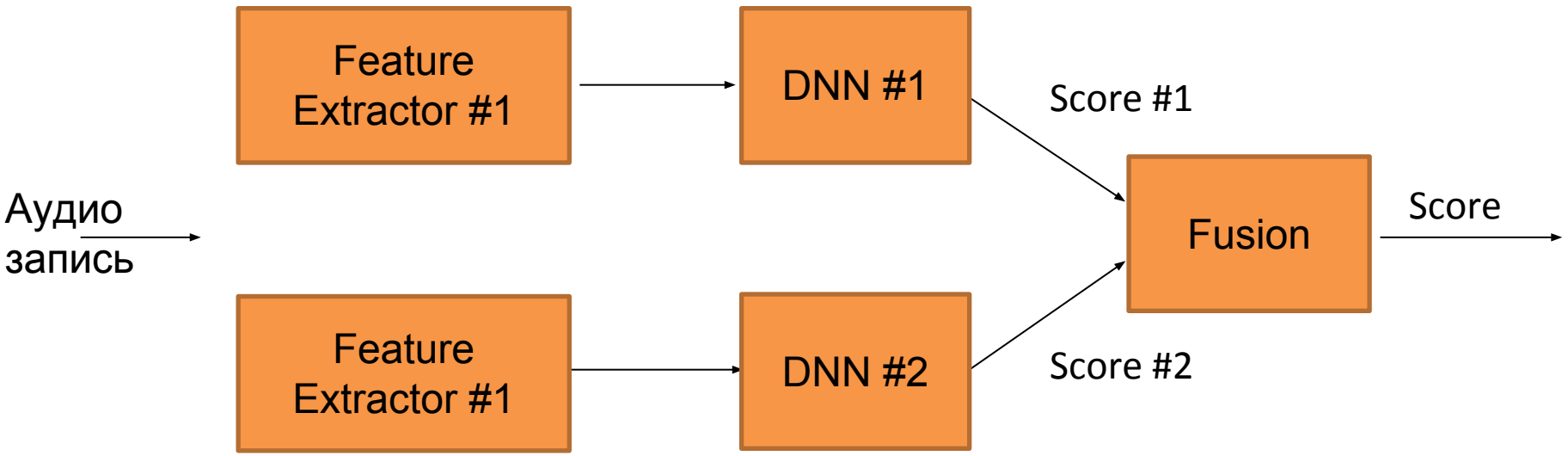
## Условияконкурса

- Короткие записи 1-3 секунд речи, необходимо определить класс записи: GENUINE (живой голос) или REPLAY (перезапись)
- 15 устройств воспроизведения
- 16 устройств записи
- 14000 тестовых записей

## Финальный DET график



# Схема речевого антиспуфинга



# Поиск Оптимальных Признаков

- Mel-frequency Cepstral Coefficients (MFCCs)
- Inverted Mel-frequency Cepstral Coefficients (IMFCCs)
- Spectral Centroid Magnitude Coefficients (SCMCs)
- Constant Q Cepstral Coefficients (CQCCs)
- Mean Hilbert Envelope Coefficients (MHECs)
- Relative-Phase Shift (RPS) Features
- Modified Group Delay Function
- Discrete Cosine Transform (DCT)
- Cosine Phase (CosPhase) Features
- Fast Fourier Transform (FFT)

# Поиск Оптимальной Архитектуры

| ID | DATA | FEAT | MODEL | TYPE | Test 1 | Test 2 | Test 3 | Test 4 | Test 5 | Test 6 | Test 7 | Test 8 | Test 9 | Test 10 | Test 11 | Test 12 | Test 13 | Test 14 | Test 15 | Test 1C | Test 2C | WEIGHTED | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | val Human<br>val Reply | val Human<br>val TV | conf Human<br>conf Spoof | ASV_2017 Humans DEV<br>ASV_2017 Reply DEV | Mix Humans<br>Studio | Mix Humans<br>TURK & Toloka | Mix Humans<br>New Conf | ASV_2015 Humans DEV<br>ASV_2015 TTS DEV | ASV_2015 Humans DEV<br>ASV_2015 VC DEV | Mix Humans<br>Google TTS | val2 Human<br>val2 Reply | Human CV Exp 11 + Exp 12<br>Reply CV Exp 7 + Exp 12 | Human RU Books Exp 9 + Exp 22<br>Reply RU Books Exp 21 + Exp 23 | Human MTurk_4 EN Books<br>Reply MTurk_4 EN Books | Studio v2 Libri Speech Humans<br>Studio v2 Libri Speech Reply | val Human Close<br>val Reply | val Human Close<br>val TV | REPLY | VC & TTS | ASV | TOTAL SPOOF |
| a19 | DS_v7 +Studio | FFTZ (256x256) | SqueezeNet (9Mb) | EER | 12.50% | 16.16% | 6.75% | **3.80%** | 12.13% | 3.13% | 0.43% | **0.06%** | **0.09%** | 0.53% | 22.22% | 27.13% | 22.35% | 77.62% | - | 2.40% | 6.62% | 11.33% | 0.23% | 1.52% | **5.78%** |
| | | | | FR | 77.62% | 26.08% | 58.11% | 1.64% | 1.15% | 0.46% | 0.85% | 0.04% | 0.03% | 1.15% | 3.09% | 20.81% | 11.85% | 16.46% | - | 4.80% | 7.98% | 10.83% | 0.42% | 0.57% | 5.83% |
| | | | | FA | 0.26% | 1.44% | 0.00% | 2.11% | 13.47% | 2.96% | 0.00% | 0.02% | 0.05% | 0.00% | 17.59% | 10.39% | 10.39% | 9.05% | - | 0.32% | 2.00% | 6.21% | 0.02% | 0.72% | 5.11% |
| a24 | DS_v8 +Studio+ | DCT (128x128) | SqueezeNet (9Mb) | EER | 8.35% | 7.12% | 13.31% | **4.97%** | 7.55% | 2.97% | 1.30% | **0.45%** | **0.98%** | 0.73% | 8.93% | 4.67% | 8.66% | 6.62% | - | 4.23% | 3.68% | 6.08% | 0.72% | 2.13% | **3.40%** |
| | | | | FR | 7.58% | 11.23% | 56.04% | 0.23% | 0.82% | 0.35% | 0.62% | 0.61% | 0.43% | 0.82% | 1.24% | 1.52% | 4.09% | 1.44% | - | 1.07% | 1.79% | 4.46% | 0.82% | 0.42% | 2.54% |
| | | | | FA | 2.66% | 1.21% | 0.00% | 5.85% | 13.04% | 7.30% | 1.05% | 0.03% | 0.09% | 0.09% | 8.03% | 2.89% | 4.47% | 3.25% | - | 3.28% | 1.68% | 4.80% | 0.07% | 1.99% | 2.44% |
| a25 | DS_v9 +Studio+ | DCT (256x256) | SqueezeNet (9Mb) | EER | 12.86% | 18.64% | 9.22% | **2.75%** | 11.47% | 3.13% | 2.13% | **0.03%** | **0.00%** | 0.67% | 0.36% | 1.63% | 8.69% | 2.01% | - | 5.68% | 9.77% | 5.37% | 0.23% | 0.93% | **2.80%** |
| | | | | FR | 13.73% | 20.32% | 77.12% | 0.41% | 3.22% | 1.90% | 2.42% | 0.00% | 0.00% | 3.22% | 0.31% | 0.65% | 4.46% | 0.31% | - | 3.55% | 9.88% | 3.59% | 1.08% | 0.14% | 2.35% |
| | | | | FA | 1.59% | 2.88% | 1.80% | 3.63% | 8.20% | 4.50% | 0.30% | 0.03% | 0.00% | 0.02% | 0.31% | 0.98% | 4.12% | 2.37% | - | 1.96% | 3.99% | 2.92% | 0.02% | 1.22% | 1.47% |
| a26 | DS_v9 +Studio+ | DCT (256x256) | SqueezeNet (9Mb) | EER | 14.91% | 17.84% | 11.02% | **3.68%** | 11.20% | 3.84% | 1.20% | **0.06%** | **0.20%** | 0.47% | 8.93% | 2.45% | 8.81% | 2.89% | - | 7.56% | 8.30% | 6.35% | 0.24% | 1.31% | **3.30%** |
| | | | | FR | 14.40% | 21.30% | 13.51% | 0.59% | 0.89% | 0.36% | 0.67% | 0.36% | 0.1% | 0.89% | 1.54% | 0.40% | 3.21% | 0.12% | - | 3.91% | 6.31% | 2.88% | 0.39% | 0.29% | 1.84% |
| | | | | FA | 3.38% | 1.82% | 1.80% | 3.39% | 13.62% | 4.87% | 0.57% | 0.00% | 0.00% | 0.07% | 3.09% | 2.47% | 5.33% | 3.98% | - | 4.16% | 2.52% | 4.16% | 0.02% | 1.13% | 2.09% |
| a27 | DS_v9 +Studio+ | FFTZ (256x256) | SqueezeNet (9Mb) | EER | 8.91% | 21.08% | 10.36% | **2.92%** | 5.80% | 4.29% | 2.60% | **0.97%** | **0.09%** | 0.85% | 6.74% | 2.30% | 8.68% | 2.26% | - | 4.09% | 14.19% | 5.84% | 0.64% | 1.33% | **3.24%** |
| | | | | FR | 13.0% | 19.26% | 16.69% | 0.06% | 0.88% | 0.36% | 0.67% | 0.09% | 0.07% | 0.89% | 1.62% | 0.70% | 3.81% | 0.25% | - | 3.47% | 5.78% | 3.19% | 0.35% | 0.07% | 1.77% |
| | | | | FA | 1.23% | 3.31% | 1.13% | 5.79% | 9.07% | 3.70% | 3.10% | 1.14% | 0.52% | 0.35% | 11.42% | 1.86% | 4.80% | 2.86% | - | 1.91% | 7.35% | 5.13% | 0.43% | 2.32% | 2.79% |
| a29 | DS_v9 +Studio+ | FFTZ (256x256) | SqueezeNet (9Mb) | EER | 11.84% | 24.03% | 12.71% | **3.04%** | 6.93% | 4.90% | 2.50% | **0.37%** | **0.08%** | 0.33% | 8.93% | 2.40% | 8.10% | 2.71% | - | 4.79% | 14.79% | 6.42% | 0.26% | 1.16% | **3.34%** |
| | | | | FR | 14.80% | 21.81% | 19.37% | 0.18% | 0.82% | 0.35% | 0.62% | 0.13% | 0.08% | 0.82% | 1.24% | 0.58% | 4.1% | 0.20% | - | 3.22% | 5.38% | 3.27% | 0.34% | 0.13% | 1.81% |
| | | | | FA | 1.38% | 5.53% | 2.25% | 6.43% | 7.78% | 8.46% | 3.02% | 0.14% | 0.02% | 0.00% | 10.19% | 2.00% | 3.88% | 2.75% | - | 1.70% | 7.67% | 5.10% | 0.05% | 2.20% | 2.58% |
| a30 | DS_v9 +Studio+ | DCT (256x256) | SqueezeNet (9Mb) | EER | 12.60% | 29.47% | 11.5% | **3.97%** | 6.13% | 1.60% | 0.63% | **0.00%** | **0.02%** | 0.25% | 8.93% | 1.96% | 8.22% | 2.85% | - | 6.11% | 16.70% | 6.24% | 0.09% | 1.33% | **3.16%** |
| | | | | FR | 17.98% | 26.61% | 21.85% | 0.12% | 0.49% | 0.20% | 0.37% | 0.00% | 0.00% | 0.49% | 1.24% | 0.39% | 3.50% | 0.04% | - | 5.62% | 9.35% | 3.82% | 0.16% | 0.04% | 2.04% |
| | | | | FA | 1.79% | 5.61% | 2.48% | 4.15% | 9.24% | 2.99% | 0.17% | 0.00% | 0.02% | 0.00% | 3.40% | 1.67% | 4.65% | 3.63% | - | 2.21% | 7.77% | 3.85% | 0.01% | 1.39% | 1.93% |
| a31 | DS_v9 +Studio+ | FFTZ (256x256) | ResNet50 (280Mb) | EER | 7.23% | 17.89% | 8.55% | **2.92%** | 3.18% | 3.47% | 1.97% | **0.35%** | **0.29%** | 0.32% | 4.01% | 1.43% | 10.19% | 2.22% | - | 3.10% | 10.60% | 4.70% | 0.32% | 1.19% | **2.51%** |
| | | | | FR | 15.11% | 22.37% | 21.62% | 0.06% | 0.56% | 0.23% | 0.42% | 0.17% | 0.12% | 0.56% | 0.00% | 0.50% | 7.74% | 0.71% | - | 4.23% | 7.04% | 3.94% | 0.25% | 0.11% | 2.11% |
| | | | | FA | 0.87% | 3.26% | 1.58% | 6.54% | 4.44% | 7.48% | 2.43% | 0.09% | 0.03% | 0.00% | 12.96% | 0.87% | 3.38% | 1.63% | - | 0.82% | 4.32% | 4.25% | 0.04% | 2.32% | 2.14% |
| a32 | DS_v10 +Studio+ | DCT (256x256) | MobileNetV2 (27.4Mb) | EER | 11.32% | 23.28% | 11.02% | **3.68%** | 6.47% | 2.46% | 0.33% | **0.02%** | **0.12%** | 0.12% | 4.37% | 0.97% | 7.62% | 1.69% | - | 3.34% | 11.55% | 4.86% | 0.09% | 1.27% | **2.48%** |
| | | | | FR | 22.49% | 33.28% | 39.87% | 0.23% | 1.29% | 0.52% | 0.97% | 0.12% | 0.06% | 1.29% | 1.85% | 0.40% | 5.40% | 0.40% | - | 9.97% | 16.60% | 7.05% | 0.50% | 0.15% | 3.77% |
| | | | | FA | 0.00% | 1.82% | 0.00% | 5.39% | 5.1% | 2.77% | 0.00% | 0.00% | 0.00% | 0.00% | 5.3% | 0.52% | 2.99% | 1.15% | - | 0.00% | 2.52% | 1.65% | 0.00% | 1.13% | 0.85% |
| k1 | DS_v10 +DSP_files +TTS,VC | FFT (512x200) | LCNN_9 (#13) | EER | 7.50% | 6.90% | 4.60% | **20.20%** | 15.50% | | | | | | 2.70% | | | | - | | | | | | |
| | | | | FR | | | | | | | | | | | | | | | - | | | | | | |
| | | | | FA | | | | | | | | | | | | | | | - | | | | | | |
| k2 | DS_v10 +TTS,VC | FFT (512x200) | LCNN_9 (#15) | EER | 9.70% | 7.60% | 3.80% | **20.90%** | 17.90% | | | | | | 4.40% | | | | - | | | | | | |
| | | | | FR | | | | | | | | | | | | | | | - | | | | | | |
| | | | | FA | | | | | | | | | | | | | | | - | | | | | | |
| p1 | DS_v10 -Studio +LibriSpeech strict balance | DCT (256x400) | ResNet34 (85.4Mb) | EER | 6.56% | 8.16% | 2.52% | **10.30%** | 12.73% | 6.81% | 2.15% | **0.49%** | **1.64%** | 0.46% | 9.10% | 5.76% | 8.68% | | - | | | | | | |
| | | | | FR | | | | | | | | | | | | | | | - | | | | | | |
| | | | | FA | | | | | | | | | | | | | | | - | | | | | | |
| p2 | DS_v10 -Studio +LibriSpeech strict balance | ROW (1x48000) | wavenet (1.6Mb) | EER | 7.85% | 8.92% | 3.36% | **11.18%** | 12.71% | 6.81% | 1.66% | **0.15%** | **0.22%** | 1.40% | 0.00% | 4.30% | 7.18% | 3.04% | - | 2.79% | 3.82% | 5.66% | 0.59% | 3.85% | **3.13%** |
| | | | | FR | | | | | | | | | | | | | | | - | | | | | | |
| | | | | FA | | | | | | | | | | | | | | | - | | | | | | |
| p3 | DS_v10 -Studio +LibriSpeech soft balance | ROW (1x48000) | wavenet (1.6Mb) | EER | 8.24% | 12.32% | 4.09% | **4.92%** | 7.46% | 3.87% | 0.93% | **0.31%** | **0.41%** | 0.86% | 0.00% | 3.74% | 7.64% | 1.31% | - | 1.31% | 4.19% | 3.69% | 0.53% | 1.88% | **2.06%** |
| | | | | FR | | | | | | | | | | | | | | | - | | | | | | |
| | | | | FA | | | | | | | | | | | | | | | - | | | | | | |
| p2 A | DS_v10 -Studio +LibriSpeech strict balance | ROW (1x48000) | wavenet (1.6Mb) | EER | 6.81% | 8.11% | 3.13% | **11.58%** | 12.80% | 7.10% | 1.57% | **0.77%** | **0.11%** | 1.53% | 0.00% | 4.32% | 7.71% | 3.04% | - | 2.58% | 3.46% | 5.52% | 0.60% | 3.95% | **3.06%** |
| | | | | FR | 19.42% | 28.73% | 23.87% | 0.76% | 5.80% | 2.35% | 4.35% | 0.42% | 0.29% | 5.80% | 0.00% | 6.92% | 5.70% | 1.89% | - | 6.06% | 10.08% | 6.16% | 2.17% | 0.49% | 4.17% |
| | | | | FA | 0.00% | 0.38% | 0.00% | 15.44% | 5.56% | 4.68% | 0.02% | 0.03% | 0.00% | 0.11% | 0.00% | 0.32% | 2.54% | 1.39% | - | 0.00% | 0.53% | 2.77% | 0.05% | 5.16% | 1.41% |
| p3 A | DS_v10 -Studio +LibriSpeech soft balance | ROW (1x48000) | wavenet (1.6Mb) | EER | 8.25% | 12.28% | 4.28% | **5.03%** | 7.47% | 3.73% | 0.93% | **0.31%** | **0.34%** | 0.87% | 0.00% | 3.75% | 7.65% | 3.14% | - | 1.32% | 4.20% | 3.77% | 0.51% | 1.89% | **2.14%** |
| | | | | FR | 12.86% | 19.03% | 18.24% | 0.23% | 1.04% | 0.42% | 0.78% | 0.33% | 0.23% | 1.04% | 0.00% | 3.08% | 2.98% | 0.61% | - | 1.70% | 2.84% | 2.90% | 0.54% | 0.27% | 1.72% |
| | | | | FA | 0.26% | 1.2% | 0.00% | 14.15% | 9.47% | 5.37% | 0.12% | 0.03% | 0.03% | 0.16% | 0.00% | 1.08% | 4.39% | 2.57% | - | 0.32% | 1.68% | 3.36% | 0.07% | 4.73% | 1.81% |
| f9 | DS_v7 / DS_v8 | FFT + DCT | a19 + a24 | EER | 7.58% | 8.57% | 12.83% | 1.82% | 6.72% | 1.83% | 0.37% | 0.03% | 0.06% | 0.20% | 13.30% | 15.02% | 13.44% | 9.85% | - | 1.64% | 3.16% | 7.27% | 0.10% | 0.63% | **3.68%** |
| | | | | EER* | 7.58% | 8.57% | 12.83% | 1.82% | 6.72% | 1.83% | 0.37% | 0.03% | 0.06% | 0.20% | 13.30% | 7.70% | 7.70% | 7.70% | - | 1.64% | 3.16% | 4.61% | 0.10% | 0.63% | 2.35% |
| | | | | FR | 13.53% | 20.02% | 56.98% | 0.29% | 0.42% | 0.17% | 0.32% | 0.02% | 0.02% | 0.42% | 1.85% | 10.80% | 6.96% | 3.96% | - | 1.33% | 2.21% | 7.75% | 0.15% | 0.11% | 3.95% |
| | | | | FA | 0.10% | 0.83% | 0.00% | 1.52% | 12.22% | 2.35% | 0.00% | 0.00% | 0.02% | 0.00% | 6.48% | 5.66% | 6.72% | 6.43% | - | 0.13% | 1.16% | 3.88% | 0.01% | 0.51% | 1.94% |
| f14 | DS_v10 +Studio+ | FFTZ + DCT (256x256) | a31 + a32 | EER | 6.92% | 17.14% | 8.55% | **2.11%** | 4.40% | 2.34% | 0.33% | **0.05%** | **0.03%** | 0.13% | 1.83% | 0.57% | 8.02% | 1.06% | - | 1.83% | 8.52% | 3.60% | 0.07% | 0.73% | **1.83%** |
| | | | | FR | 10.65% | 15.69% | 14.87% | 0.00% | 0.11% | 0.05% | 0.08% | 0.00% | 0.00% | 0.11% | 0.00% | 0.07% | 2.42% | 0.02% | - | 1.01% | 1.68% | 1.85% | 0.04% | 0.00% | 0.94% |
| | | | | FA | 0.72% | 5.09% | 1.58% | 9.12% | 7.64% | 8.58% | 2.48% | 0.00% | 0.04% | 0.00% | 14.5% | 1.25% | 4.98% | 2.71% | - | 0.88% | 5.87% | 5.40% | 0.02% | 3.08% | 2.72% |
| f15 | DS_v9 / DS_v10 | FFTZ + DCT | a25 + a31 + a32 | EER | 8.76% | 17.37% | 7.89% | **1.58%** | 5.30% | 2.16% | 0.30% | **0.00%** | **0.00%** | 0.12% | 0.18% | 0.48% | 7.61% | 0.76% | - | 2.07% | 5.88% | 3.10% | 0.04% | 0.53% | **1.57%** |
| | | | | FR | 16.83% | 24.64% | 34.10% | 0.00% | 0.31% | 0.11% | 0.38% | 0.00% | 0.00% | 0.31% | 0.00% | 0.22% | 4.98% | 0.09% | - | 3.72% | 6.33% | 3.67% | 0.17% | 0.00% | 1.92% |
| | | | | FA | 0.26% | 1.52% | 0.90% | 5.28% | 4.73% | 5.26% | 0.02% | 0.00% | 0.00% | 0.00% | 0.31% | 0.29% | 3.09% | 0.95% | - | 0.32% | 2.10% | 1.75% | 0.00% | 1.09% | 0.87% |
| f16 | DS_v10 +Studio+ | FFTZ + DCT | p2 + a31 + a32 | EER | 6.56% | 13.72% | 4.94% | **1.34%** | 1.95% | 1.67% | 0.23% | **0.03%** | **0.03%** | 0.13% | 0.18% | 0.53% | 6.27% | 0.47% | - | 0.89% | 5.67% | 2.19% | 0.06% | 0.47% | **1.13%** |
| | | | | FR | 16.9% | 25.02% | 27.25% | 0.00% | 0.24% | 0.10% | 0.18% | 0.04% | 0.03% | 0.24% | 0.00% | 0.57% | 4.1% | 0.15% | - | 3.16% | 5.25% | 3.76% | 0.10% | 0.02% | 1.93% |
| | | | | FA | 0.00% | 1.37% | 0.00% | 4.39% | 3.02% | 3.18% | 0.00% | 0.00% | 0.00% | 0.00% | 0.31% | 0.1% | 2.43% | 0.37% | - | 0.00% | 1.89% | 1.43% | 0.00% | 1.46% | 0.71% |
| f17 | DS_v10 +Studio+ | RAW + DCT + FFT | p2 + a32 + a27 | EER | 6.15% | 16.2% | 4.94% | **1.46%** | 2.33% | 1.57% | 0.30% | **0.06%** | **0.02%** | 0.27% | 0.18% | 0.70% | 5.85% | 0.33% | - | 0.75% | 6.40% | 2.25% | 0.12% | 0.53% | **1.18%** |
| | | | | FR | 15.93% | 23.58% | 26.80% | 0.00% | 0.29% | 0.12% | 0.22% | 0.02% | 0.01% | 0.29% | 0.31% | 0.54% | 3.00% | 0.02% | - | 2.84% | 4.73% | 3.53% | 0.11% | 0.01% | 1.82% |
| | | | | FA | 0.00% | 1.87% | 0.00% | 3.63% | 4.78% | 2.97% | 0.02% | 0.00% | 0.00% | 0.00% | 0.31% | 0.19% | 2.78% | 0.60% | - | 0.00% | 2.3% | 1.60% | 0.00% | 1.27% | 0.80% |
| f18 | DS_v10 +Studio+ | RAW + DCT + FFT | p2 + a32 + a27 (13, 11, 5) | EER | 6.35% | 14.24% | 3.62% | **1.70%** | 2.87% | 2.10% | 0.13% | **0.03%** | **0.00%** | 0.13% | 0.18% | 0.73% | 5.53% | 0.39% | - | 0.52% | 4.72% | 2.04% | 0.05% | 0.58% | **1.05%** |
| | | | | FR | 17.78% | 26.3% | 29.09% | 0.00% | 0.35% | 0.15% | 0.23% | 0.03% | 0.02% | 0.35% | 0.31% | 0.80% | 3.06% | 0.05% | - | 3.22% | 5.38% | 3.87% | 0.12% | 0.02% | 1.99% |
| | | | | FA | 0.00% | 0.46% | 0.00% | 3.48% | 4.07% | 2.53% | 0.00% | 0.00% | 0.00% | 0.00% | 0.31% | 0.19% | 2.42% | 0.44% | - | 0.00% | 0.46% | 1.76% | 0.00% | 1.23% | 0.66% |

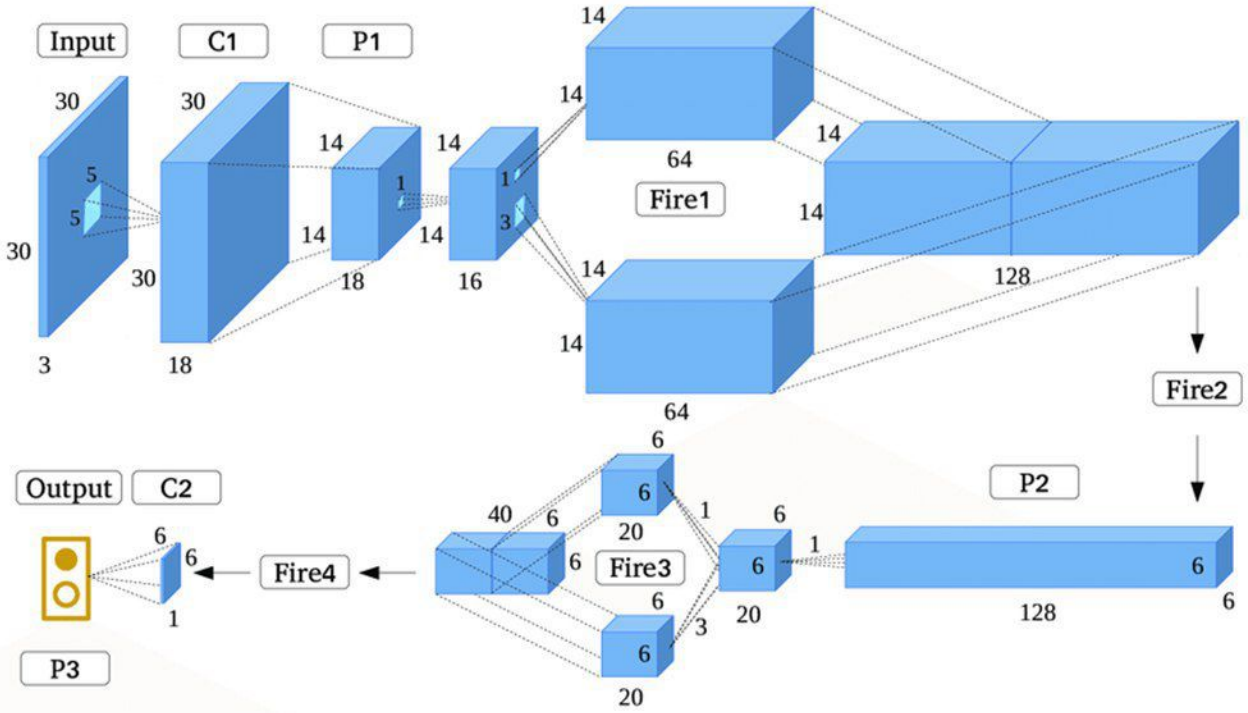
ID R&D

# Архитектуры DNN

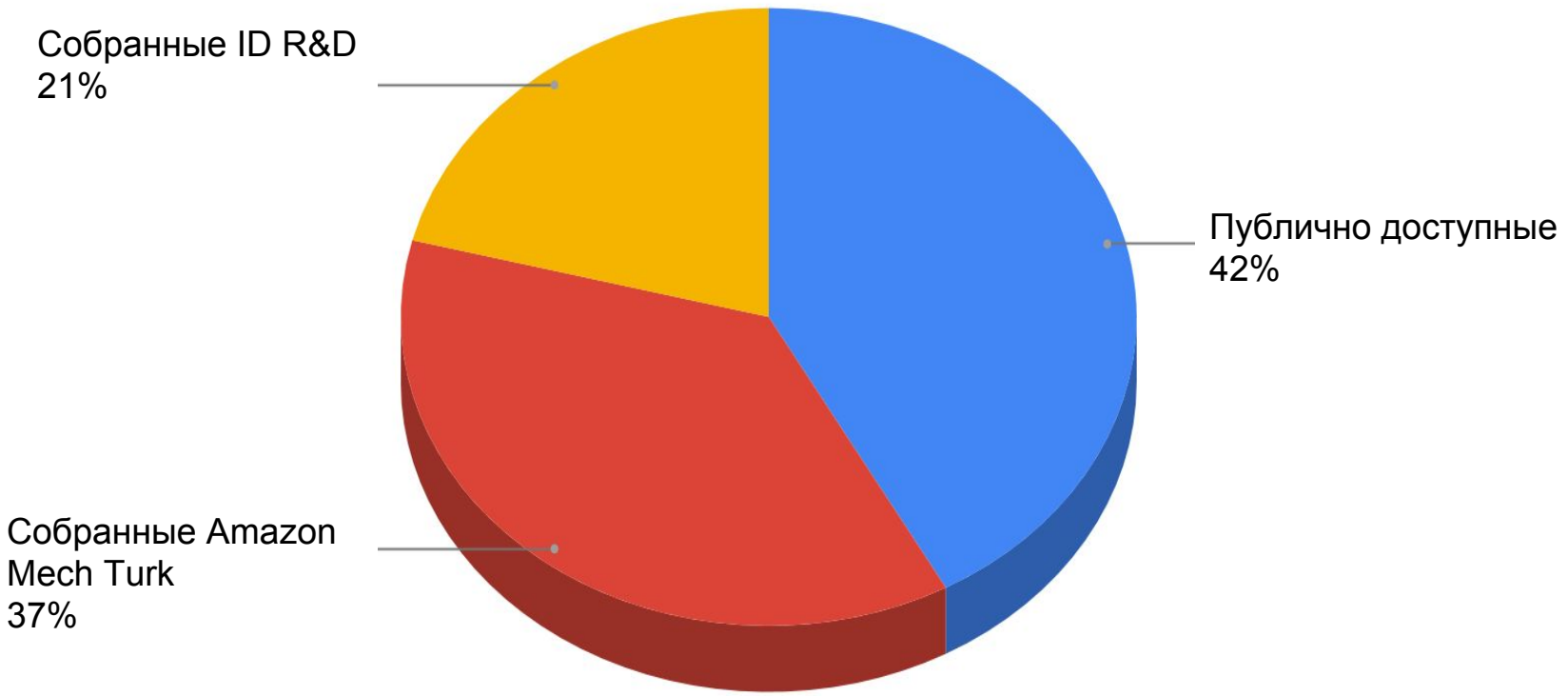
| Архитектура | Точность | Примечание |
|---|---|---|
| LCNN | Средняя | Широко используется в литературе |
| ShuffleNet | Средняя | |
| Time DNN | Средняя | |
| Temporal CNN | Высокая | Большие требования RAM |
| ResNet-50 | Высокая | Большой размер сети (260MB) |
| SqueezeNet | Высокая | Малый размер сети 3MB |
| WaveNet | Высокая | Малый размер сети 3MB |

# SqueezeNet

SqueezeNet архитектура достигает уровня точности AlexNet на ImageNet с количеством параметром в 50 раз меньше



# Обучающие Базы Данных

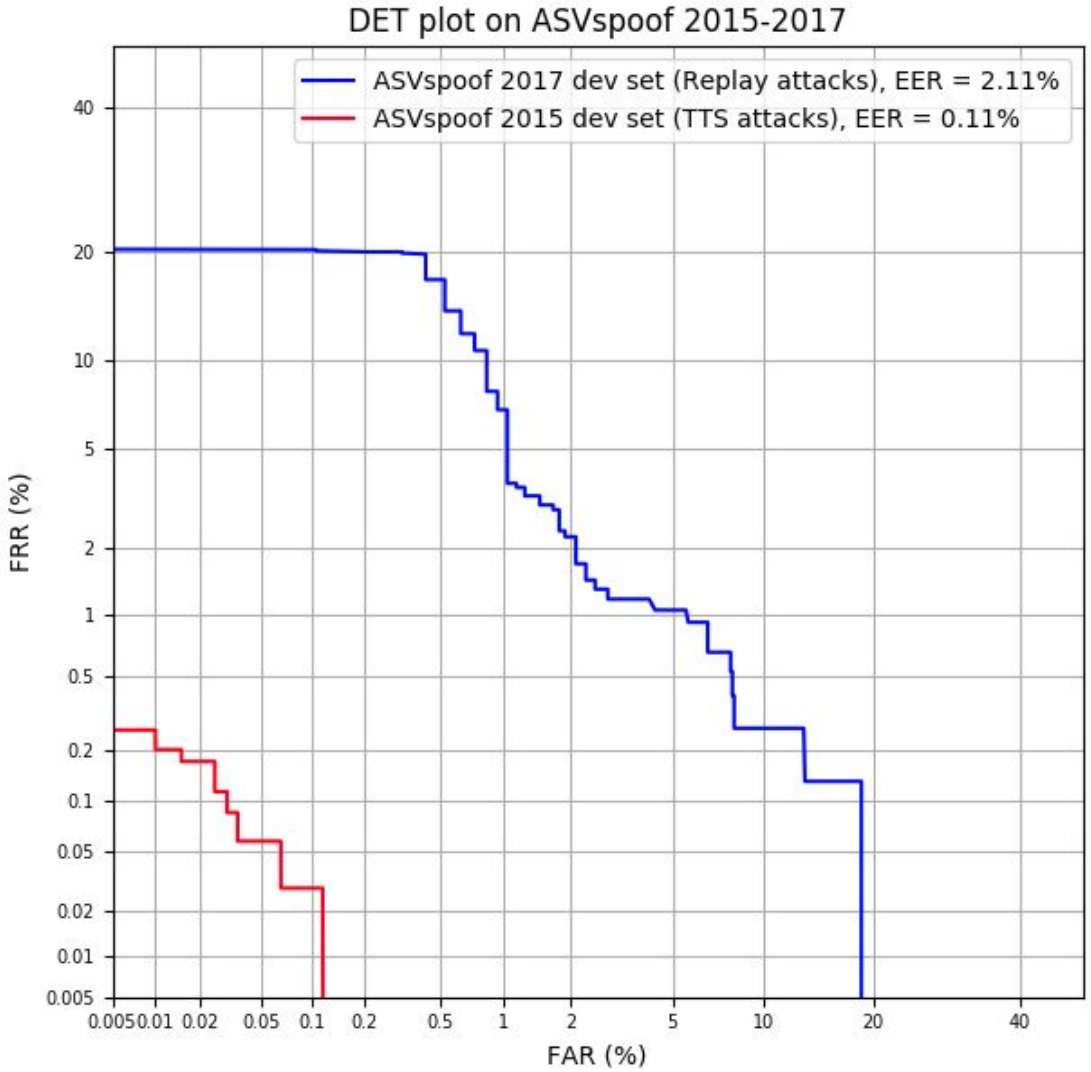


# Обучающие Базы Данных

Не смотря на то, что выбор архитектуры нейронной сети является одним из основных параметров, влияющих на качество всей системы в целом, обучающая база данных сильно влияет на устойчивость к типам диктофонов, синтезаторам речи и т.п.

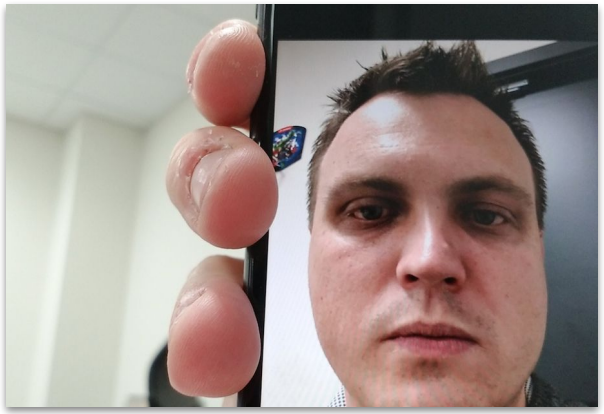
| Класс | Файлов |
|---|---|
| **“Живой” голос** | 198 221 |
| **Спуфинг:** Записи воспроизведения | 199 229 |
| **Спуфинг:** Синтез речи | 111 066 |
| **Спуфинг:** Конверсия речи | 157 295 |
| ВСЕГО | 665 811 |

# Точность Детектирования



# Конкурс по лицевому антиспуфингу

# Атаки на системы распознавания лиц



Replay-attack

Printed photo attack

3D-mask attack

# 2D vs 3D

- IDC: в 2018 году будет поставлено около 1.4 млрд смартфонов.
- Sigmaintell: около 100 млн выпущенных в 2018 году устройств будут иметь 3D-камеры.
- Различные источники: количество устройств с 3D камерами вырастет в 4-5 раз к 2022 году.

# Простые движения (liveness)

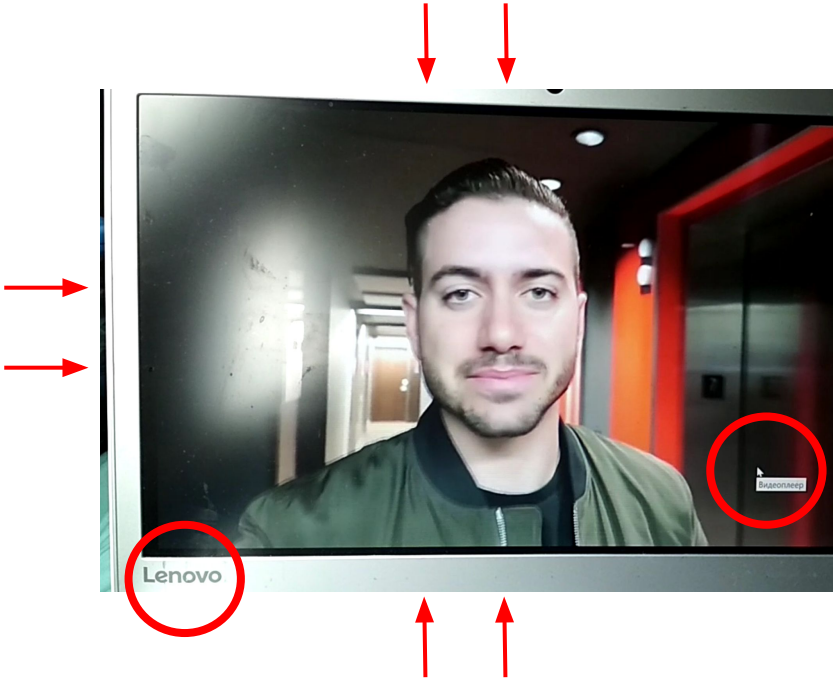
- Моргание
  - каждые 2-4 секунды
  - средняя длительность ~250 мс
- Движения зрачков
- Движения губ
- Движения головы
- Специальные жесты или мимика

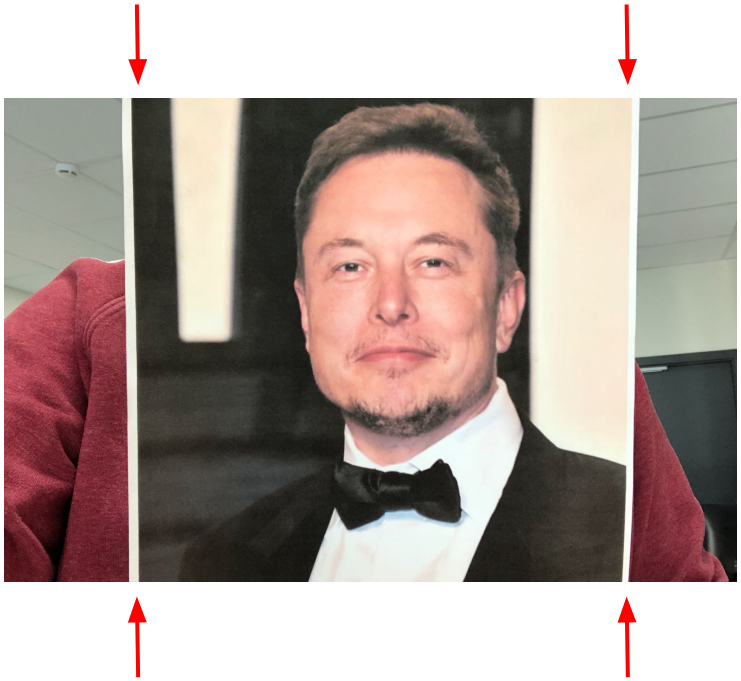
# Детектирование объектов

Что общего у этих двух атак?

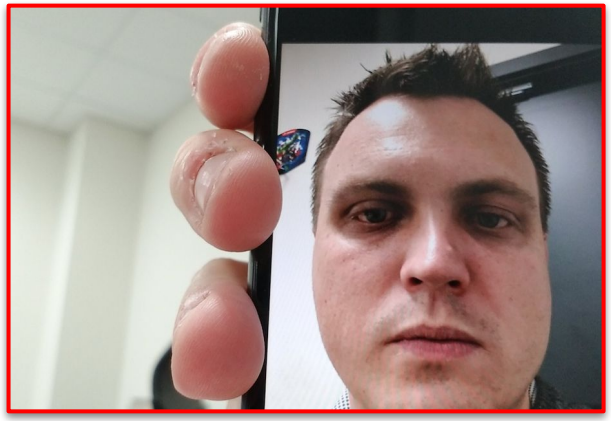

# Детектирование объектов

Хак для детектирования непрофессиональных атак



# Анализ текстуры

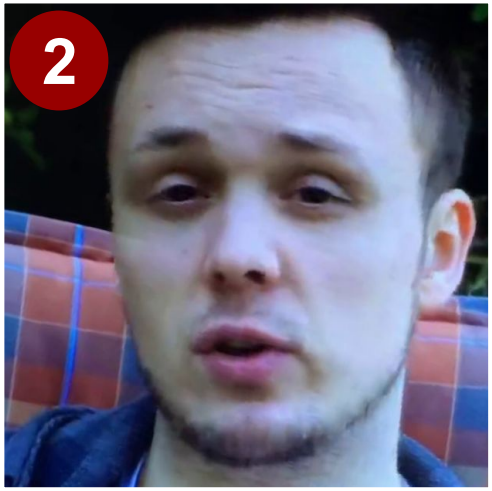


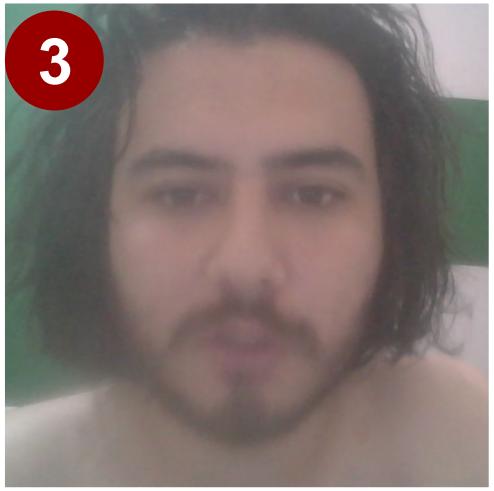
**Replay-attack (texture-based)**

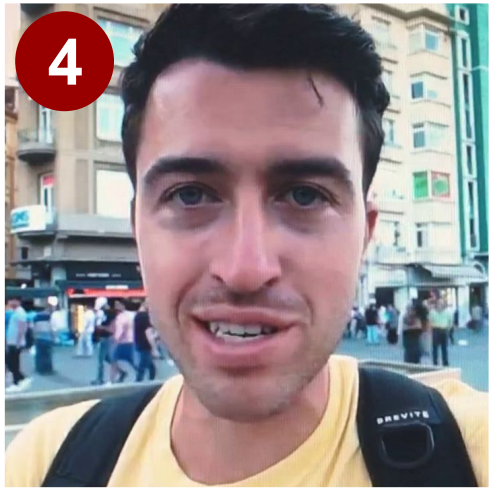
Printed photo attack

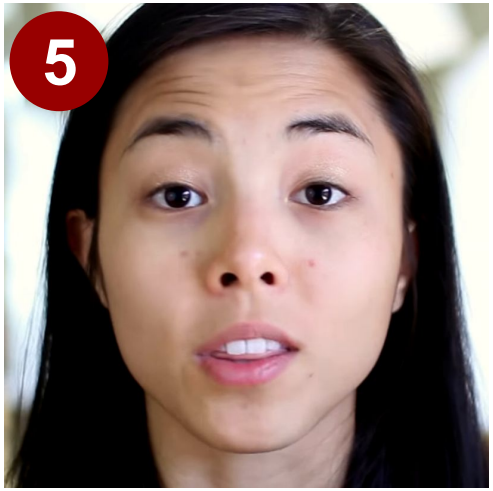
3D-mask attack

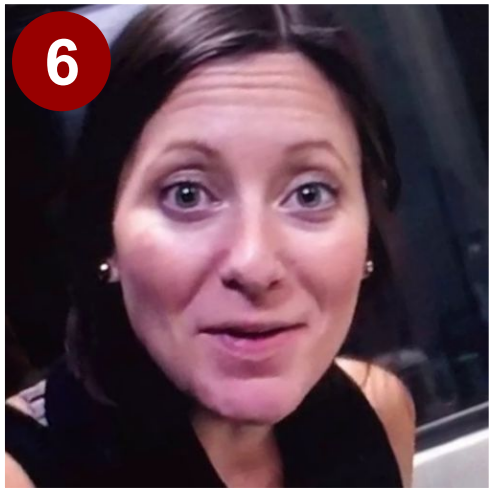
# Анализ текстуры

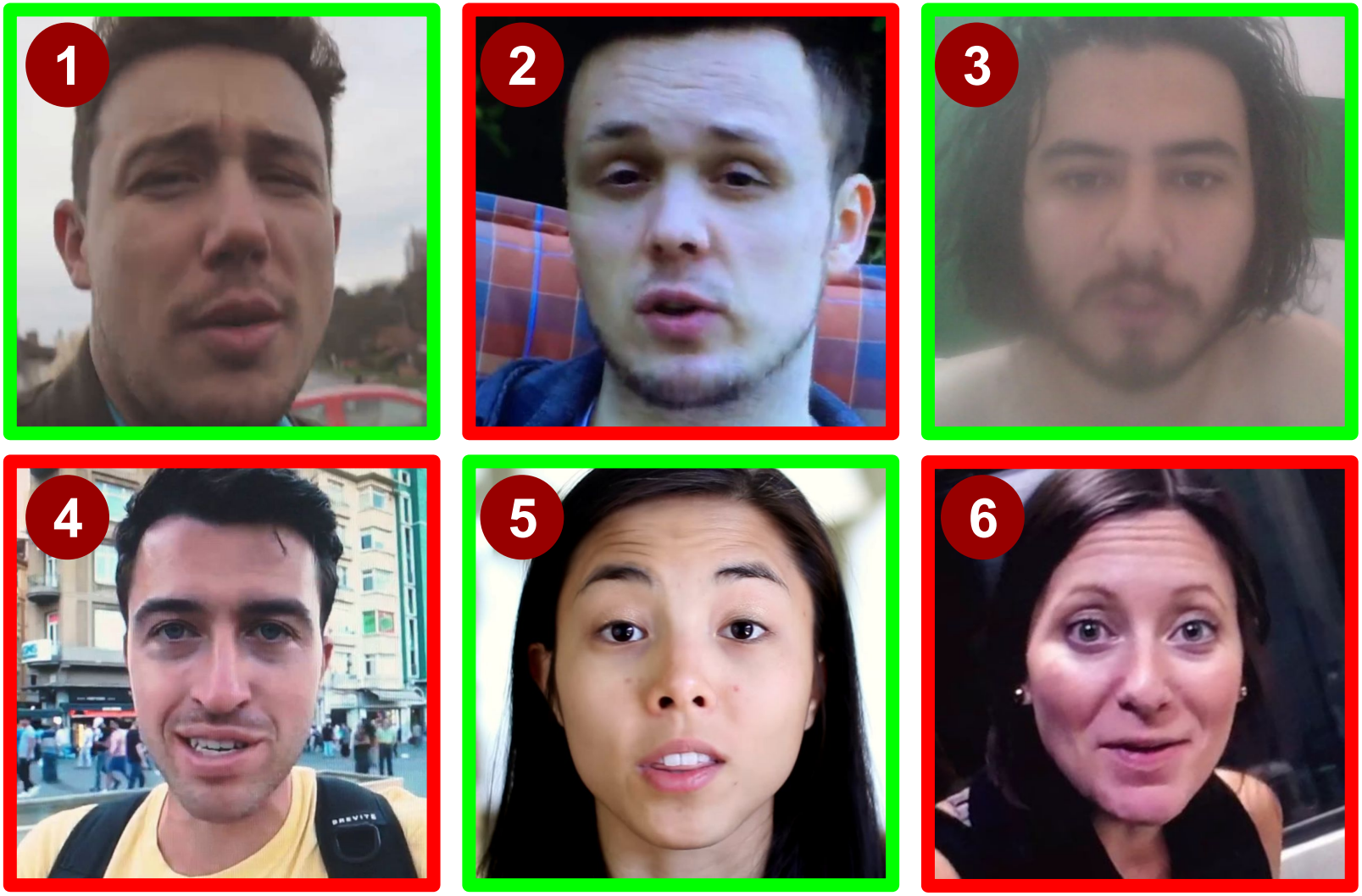



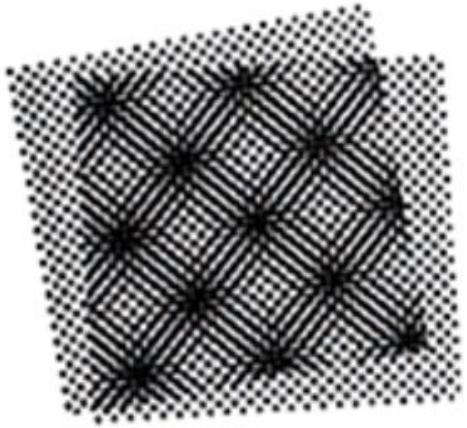



ID R&D

# Анализ текстуры

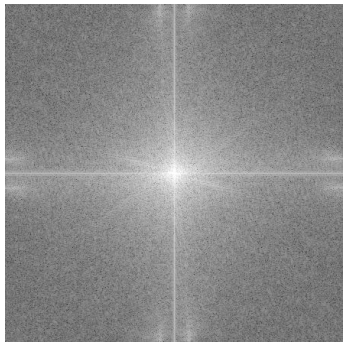


# Муаровый узор

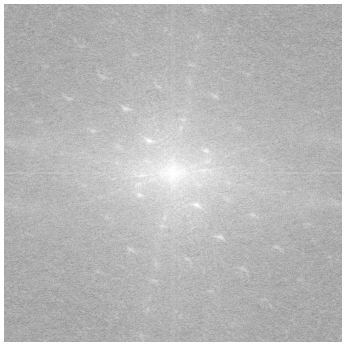


Муаровые узоры - нежелательные артефакты на изображениях, появляющиеся за счет наложения цифровых решеток оптических матриц. Довольно часто возникают при фотосъемке цифровых экранов.

Оригинал



Атака



ID R&D

# Базы данных

| База | Количество человек | Количество экранов | Количество устройств | Количество записей |
|---|---|---|---|---|
| CASIA-FASD | 50 | 1 | 1 | 600 |
| Idiap Replay-Attack | 50 | 3 | 1 | 1300 |
| Replay-Mobile | 40 | 2 | 2 | 1190 |
| MSU-MFSD | 35 | 2 | 2 | 280 |
| OULU NPU | 55 | 6 | 6 | 4950 |
| Spoof in the Wild | 165 | 2 | 4 | 4478 |
| *MSU USSA* | *1000* | *1* | *2* | *9000 im.* |

ID R&D

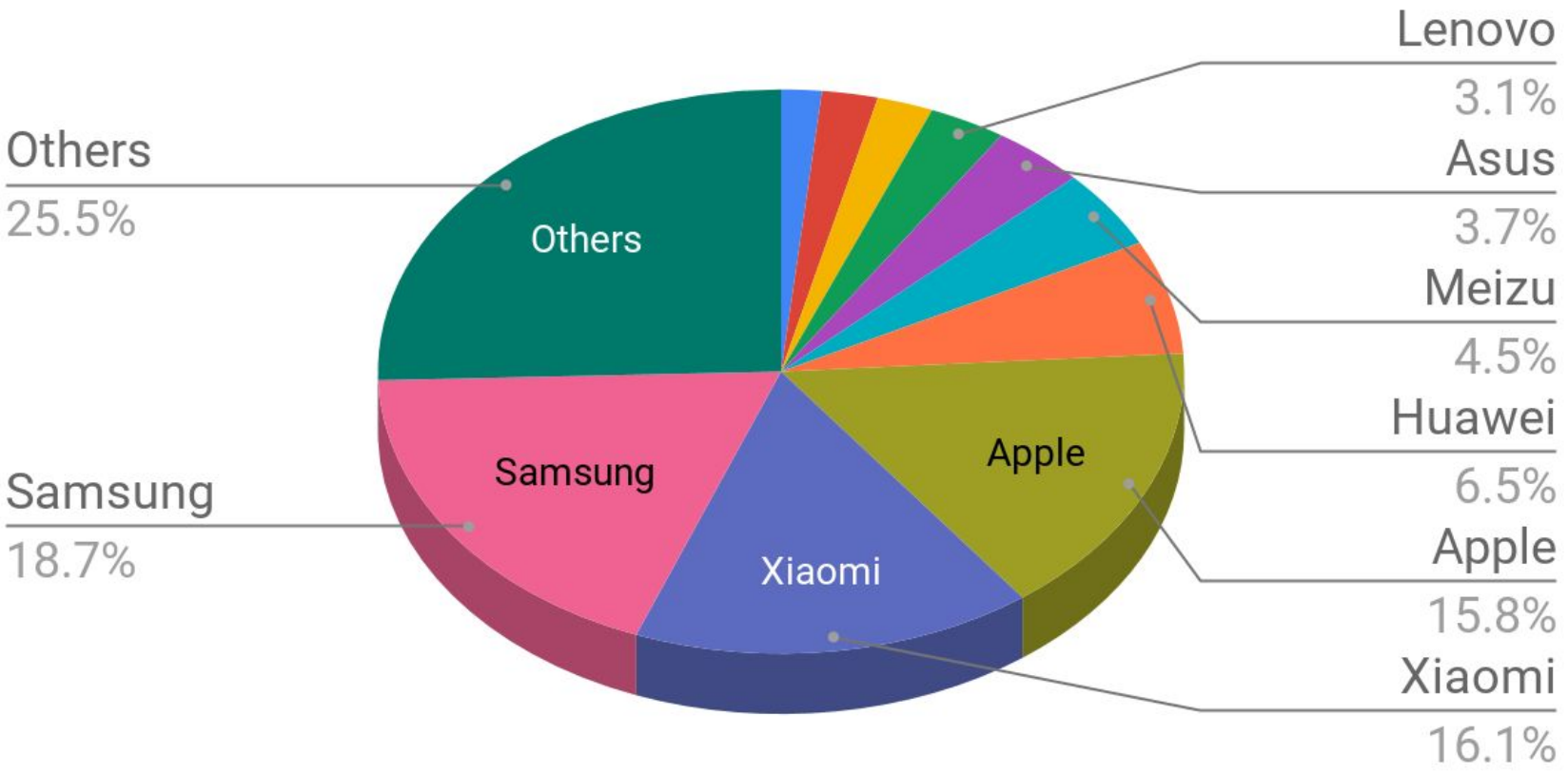
# Базы данных

| База | Количество человек | Количество экранов | Количество устройств | Количество записей |
|---|---|---|---|---|
| CASIA-FASD | 50 | 1 | 1 | 600 |
| Idiap Replay-Attack | 50 | 3 | 1 | 1300 |
| Replay-Mobile | 40 | 2 | 2 | 1190 |
| MSU-MFSD | 35 | 2 | 2 | 280 |
| OULU NPU | 55 | 6 | 6 | 4950 |
| Spoof in the Wild | 165 | 2 | 4 | 4478 |
| *MSU USSA* | *1000* | *1* | *2* | *9000 im.* |
| **IDRND-FASD** | 1092 (778+**314**) | **161** | **480** | **8005** (1212+6793) |

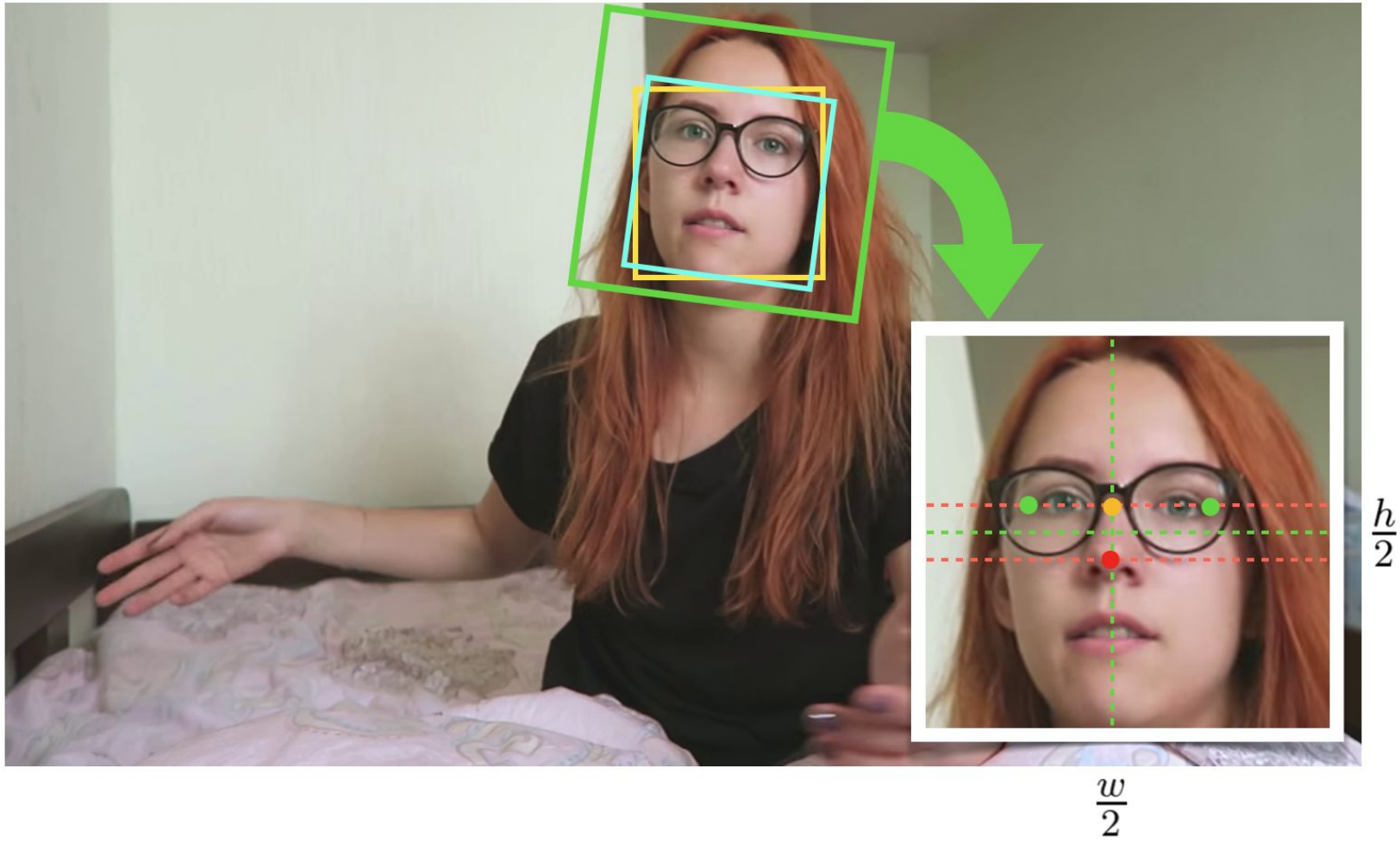
# IDRND_FASD

- Мужчины (654), Женщины (549)
- Общая длительность оригиналов ~ 5 ч.
- Общая длительность атак ~ 11 ч.
- HD Quality ~ 44% , SHD Quality ~ 46%
- Около 5% атак собрано в лаборатории

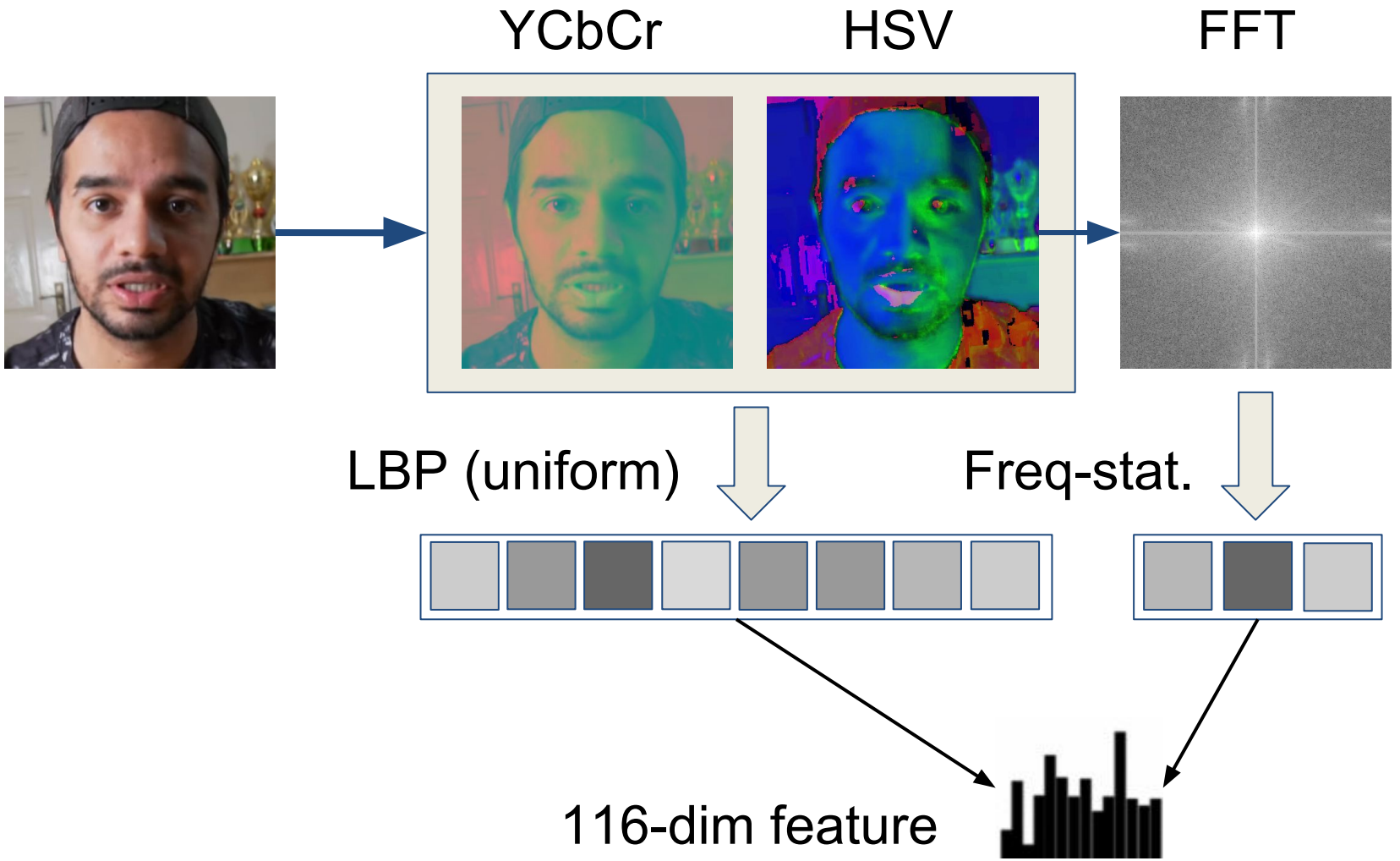
# Mobile brands (66)



# Эксперименты: предобработка



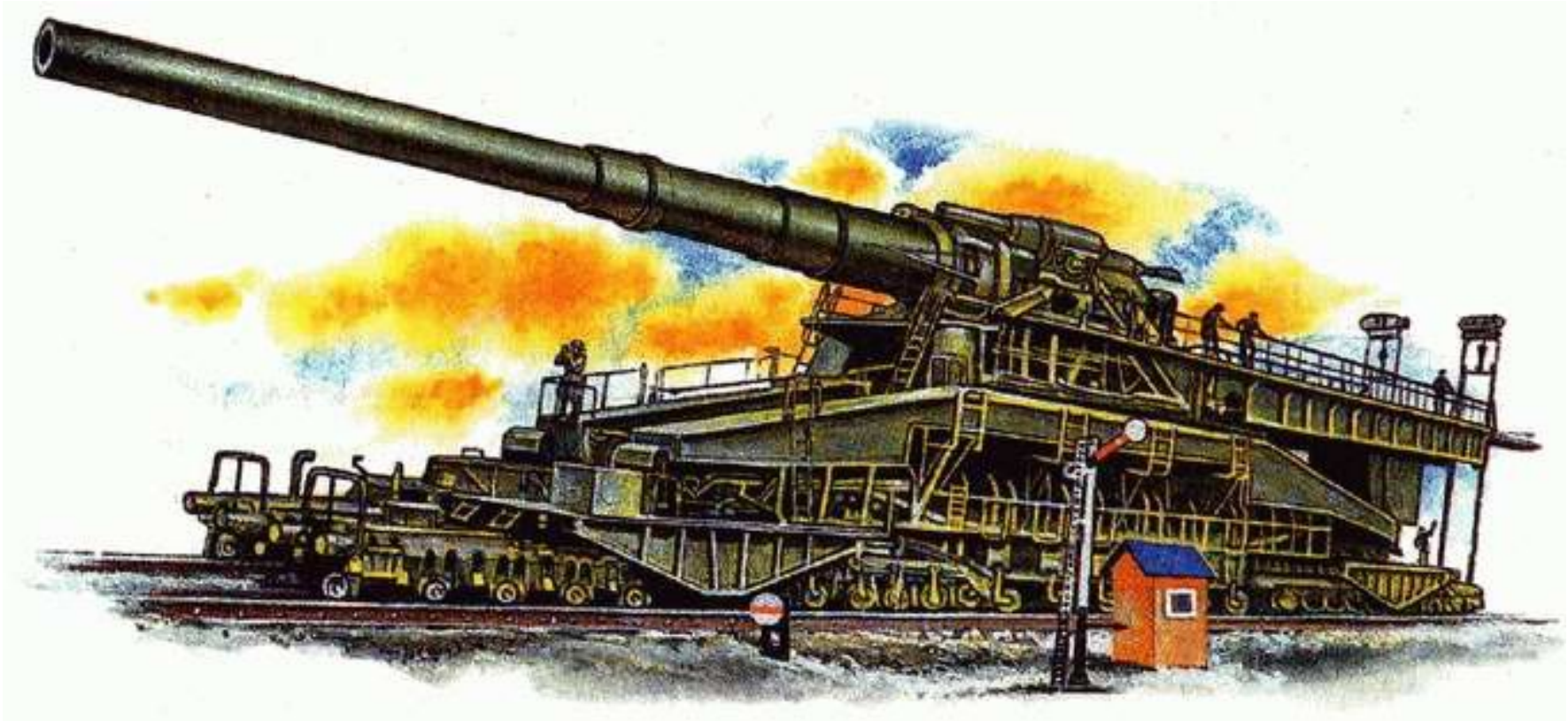
# Эксперименты (Baseline)



# Наши эксперименты (Baseline)


| | IDRND | Oulu Pr.I |
|---|---|---|
| EER | 11,4 % | 7,6 % |
| AUC | 0,974 | |

# Решения участников



ID R&D

# Решения участников

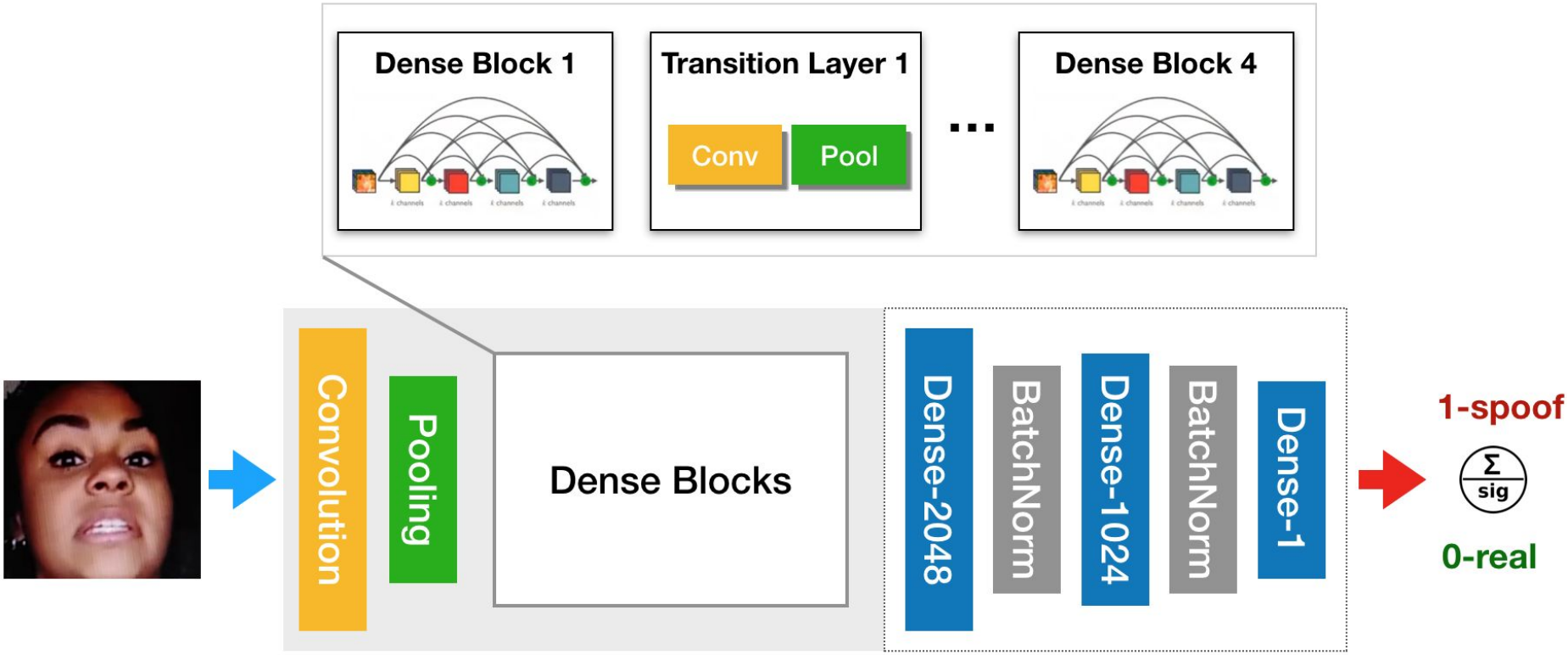
| Метод | Предобработка | Стек | Фьюжн | EER | ROC AUC |
|---|---|---|---|---|---|
| DenseNet-201 | TTA-4 (+flips) | 5 | Rank average | 0.0417 | 0.99934 |
| SE-Inception | TTA-16 (+flip) | 5 | Weighted sum | **0.0285** | 0.999336 |
| SE-ResNext 50 | Basic augmentation (+flips) | 2 | Mean | 0.0538 | 0.998732 |

Данил Ахметов

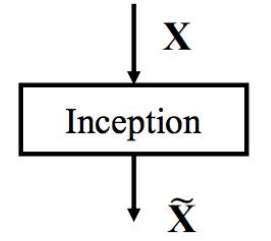
Алексей Алексеев

Илья Тетерников

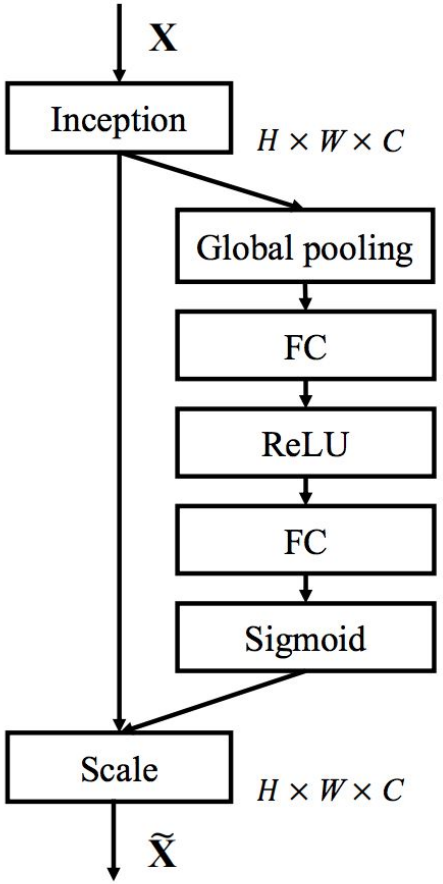
# Архитектуры из топ-3



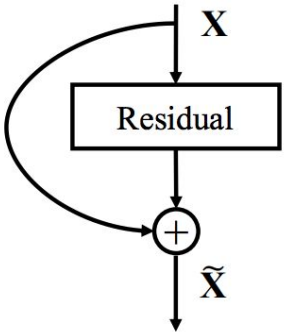
# Архитектуры из топ-3



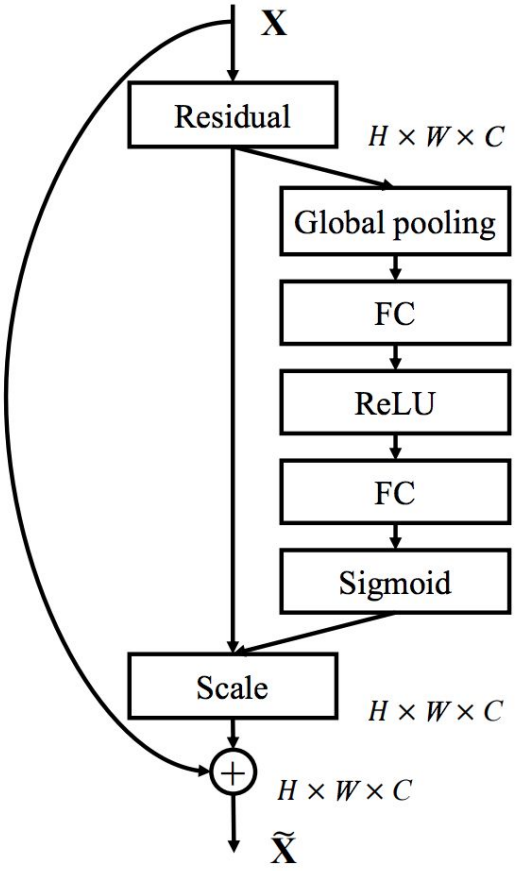

**Inception Module**




**SE-Inception Module**

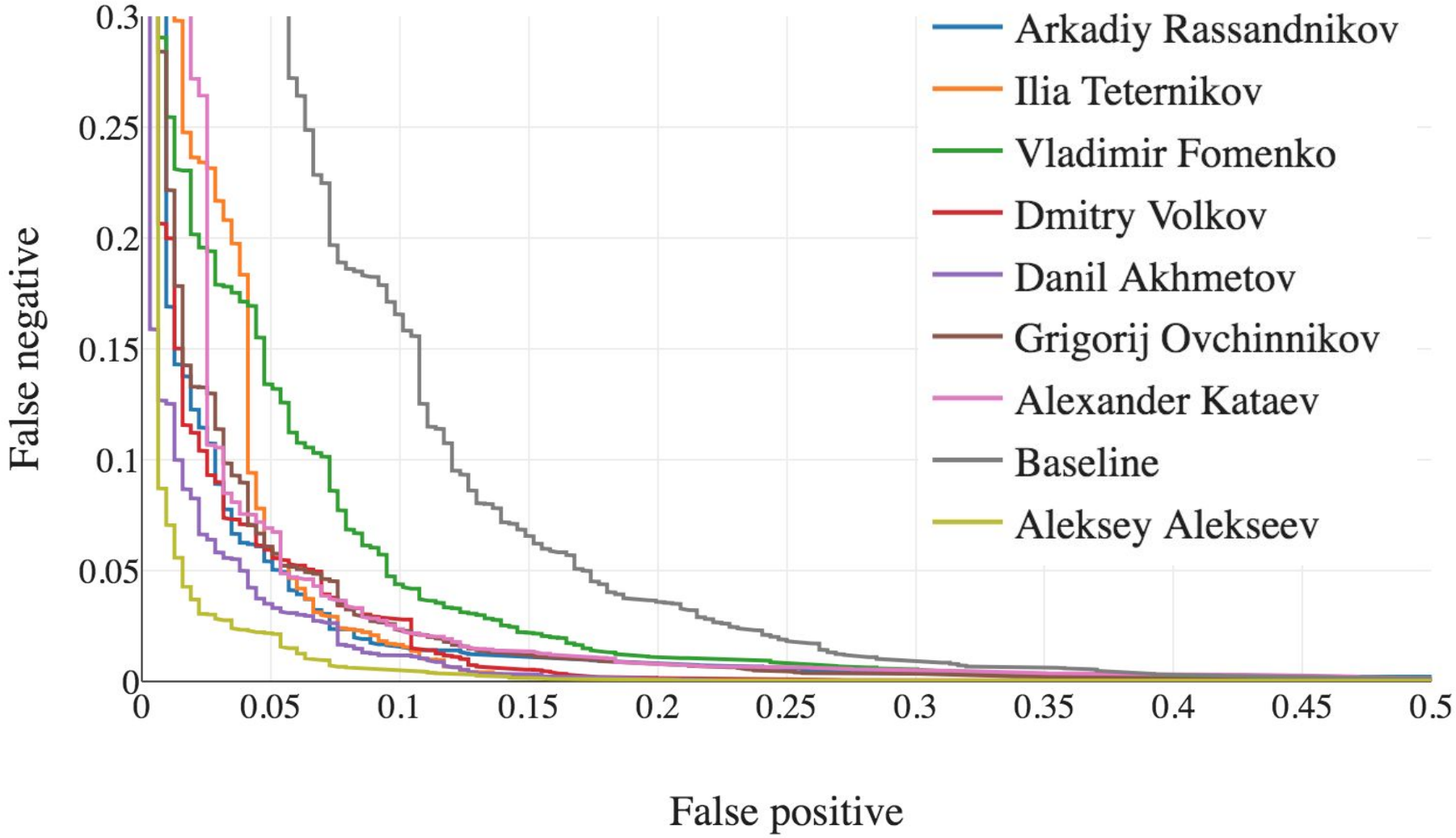



**ResNet Module**




**SE-ResNet Module**

# Решения участников (топ-20)

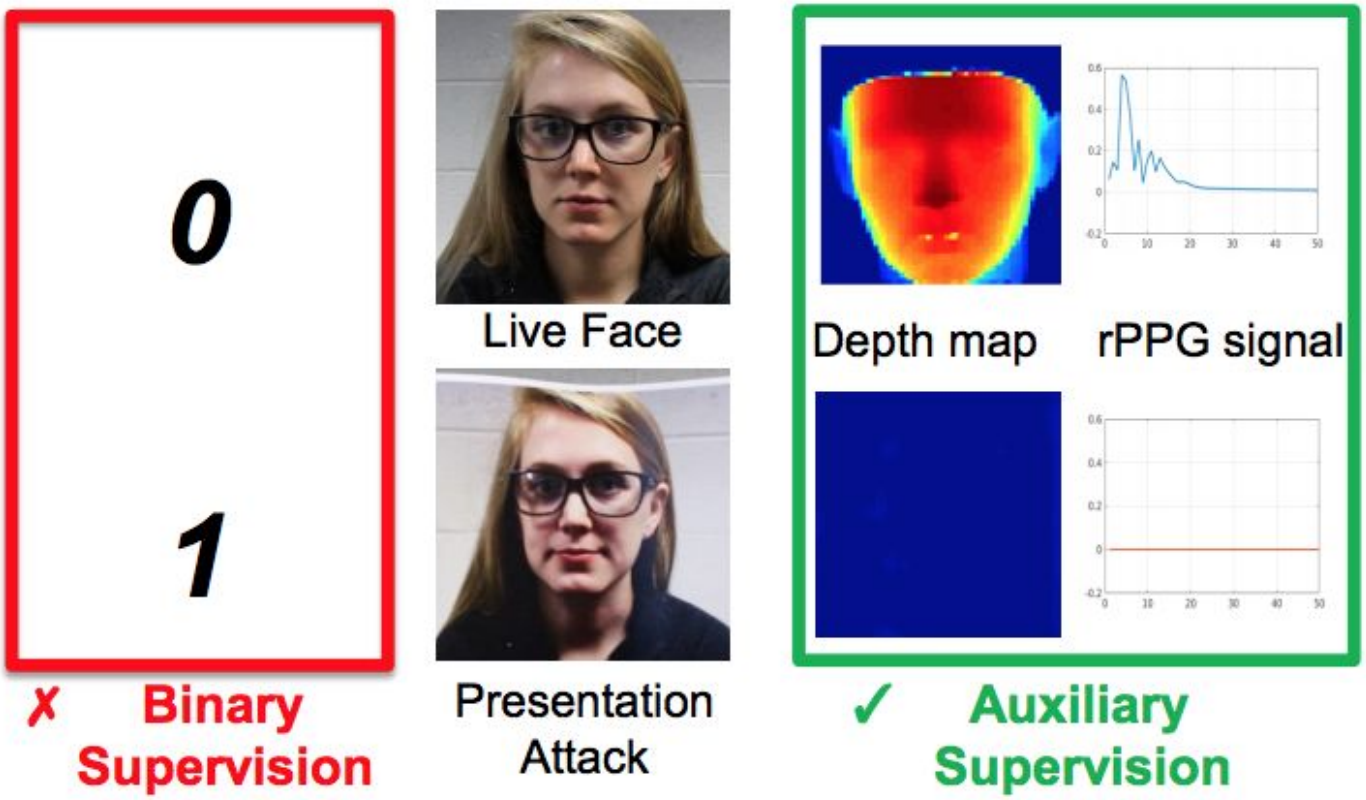


# Cross-dataset testing

| | IDRND_FASD | OULU_NPU (Protocol I) | |
|---|---|---|---|
| EER | 5,1 % | 12,6 % | Replay-attack |
| EER | 5,1 % | 29,9 % | Printed photo |

# Future steps

Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision.
Yaojie Liu,  Amin Jourabloo, Xiaoming Liu (MSU)



# Выводы

- Анализ 2D-атак будет актуален в ближайшие 5 лет для систем лицевого антиспуфинга.
- Детектирование атак на отдельных кадрах лучше всего подходит для frictionless-биометрии.
- Cross-dataset еще не побежден.
- Конкурсы - это здорово (со стороны организатора - тоже).
- Следующий конкурс - антиспуфинг по голосу.