Deep 

или Не верь глазам своим

Неделя безопасного Рунета – 2019

CSF-19, 14 февраля 2019 г.

Что такое Deep Fakes?

Если научно, то это «использование искусственного интеллекта для синтезирования человеческого изображения», на основе нейросетей и машинного обучения.

Очень официально, но… мало что понятно ☺

А точнее, это технология, которая позволяет создать видеоизображение с реалистичными действиями и словами **реально существующего** человека.

Происхождение термина: deep learning (глубокое машинное обучение) + fakes (фальшивки).

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Как это было раньше?

В доцифровую эпоху - дублер (двойник), обилие грима и микширование голоса при «озвучке» - максимально близко к оригиналу.

Начало «цифрового века» – учимся синтезировать голос.

Photoshop и аналоги – легче изменять статичные изображения: сканам и фотографиям верить уже нельзя…

Но видео доверяли - считалось, что идеально «подделать» видео невозможно (дублера все равно можно отличить)…

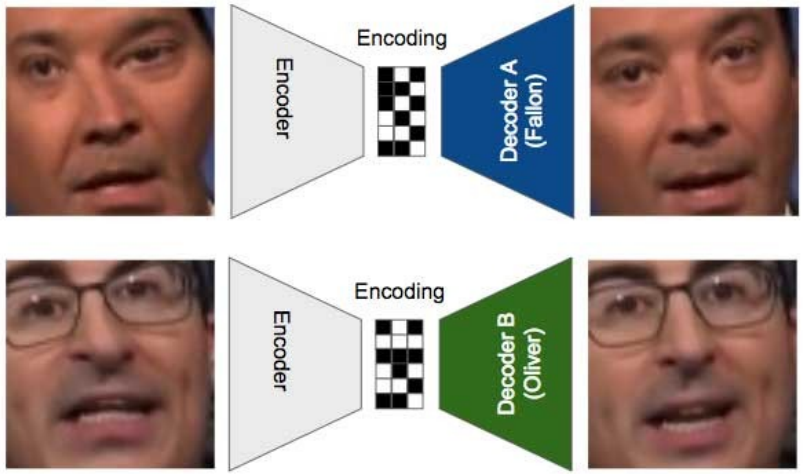
ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Как это выглядит сейчас?

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

А теперь, как это делается.

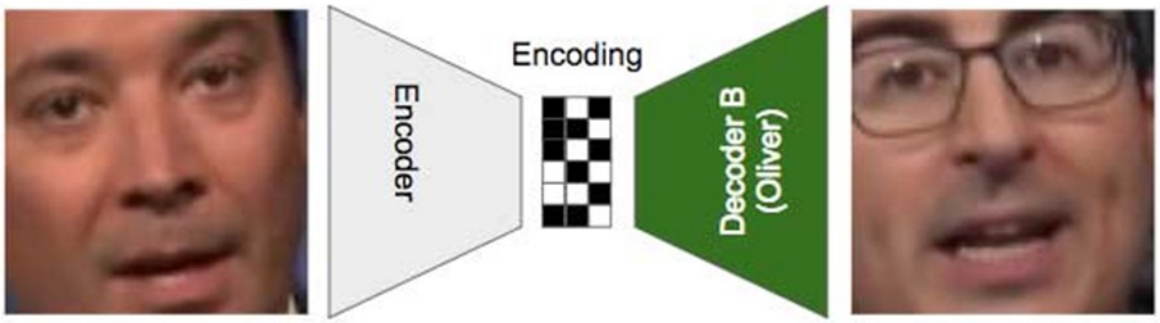
В основе лежит **автокодировщик** (autoencoder) – обучаемая нейросеть. Она учится «сжимать» входящее видео и восстанавливать оригинал из сжатого видео. В ходе этого она «понимает» движения, мимику и т.п. человека.



ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

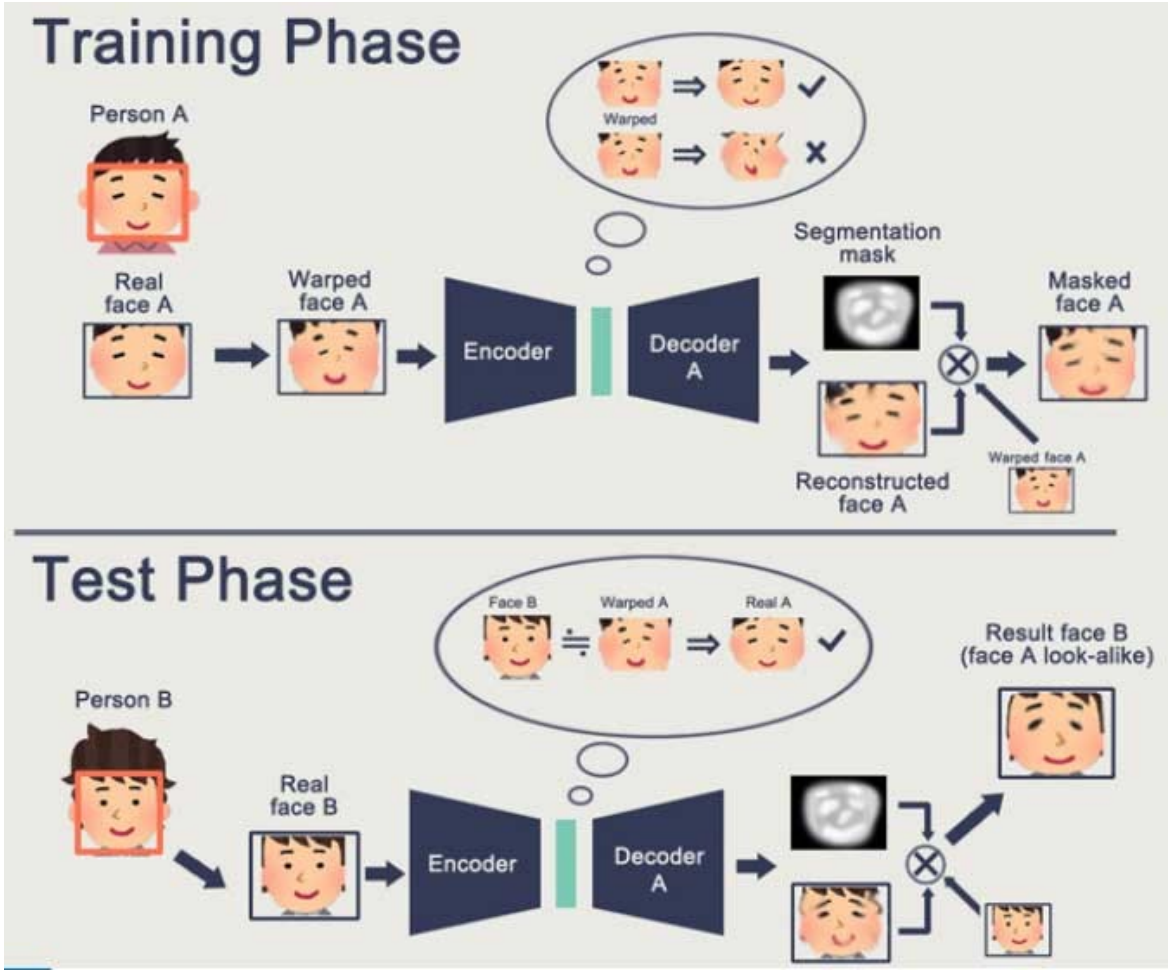
А теперь, как это делается.

Суть метода – мимика и движения от Персоны 1 (модели), внешний вид – от Персоны 2 (имитируемой).



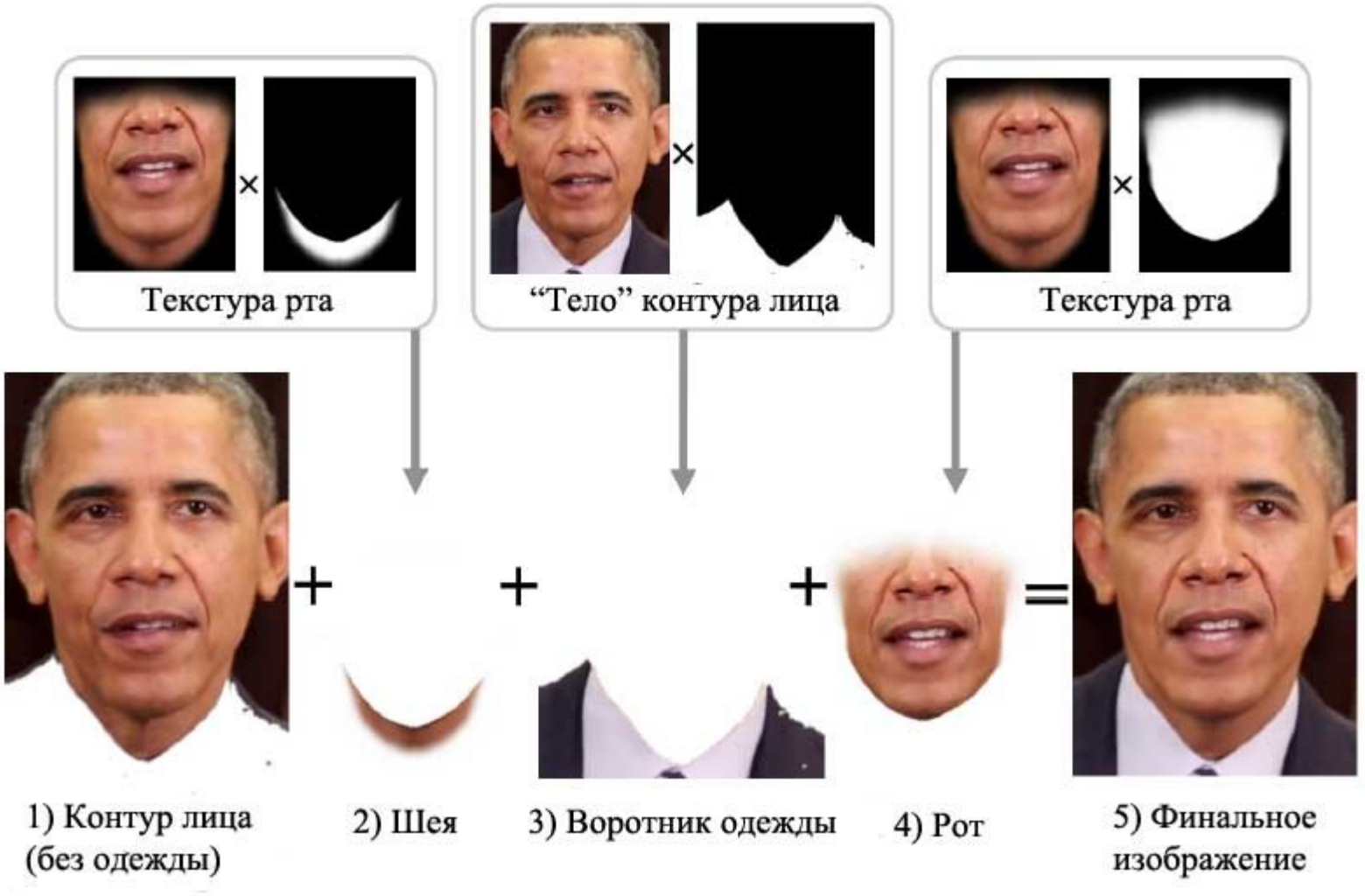
ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Нейросеть обучается и тестируется



ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Вот так «конструируется» видео

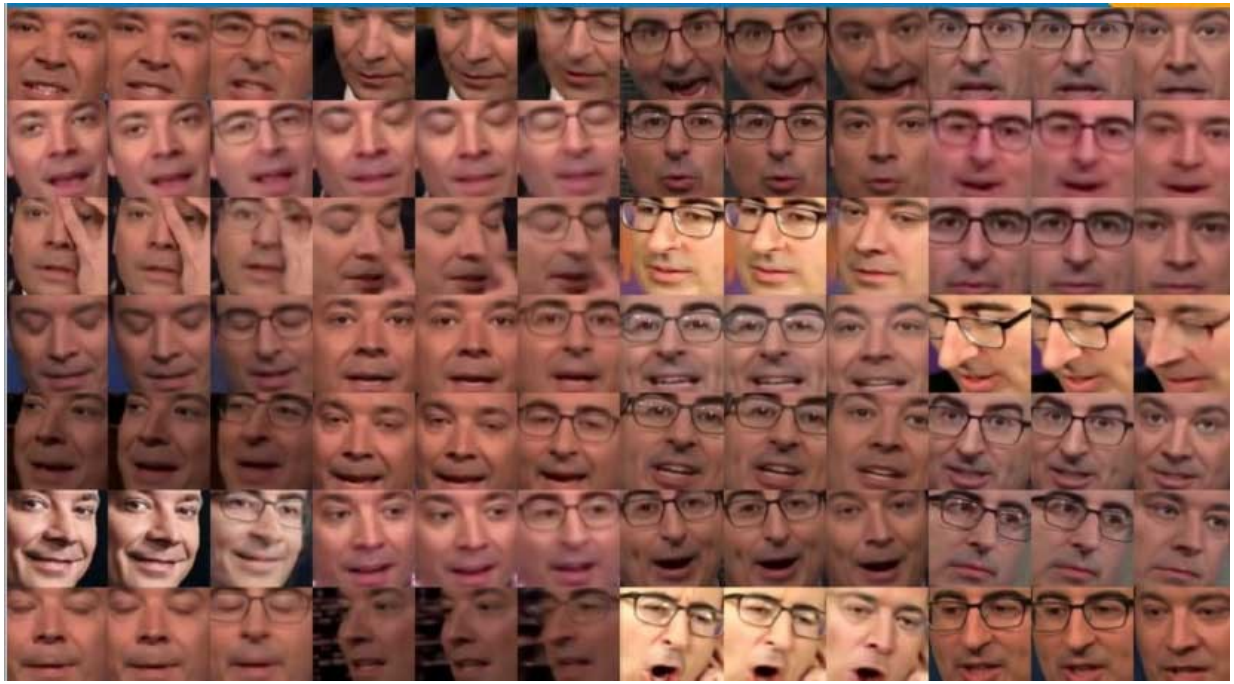


ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Все так просто? Нет.

Для достоверного воспроизведения, нейросети нужны 300 – 2000 фотоизображений воспроизводимой персоны (МИНИМУМ) в разных ракурсах, с разной мимикой и т.п. Работает принцип, аналогичный возрастной реконструкции: «чем больше, тем лучше». Наилучший вариант – несколько часов видео.

Качество вводимого материала имеет значение.



ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Иначе будет так…

Нейросети оказалось «недостаточно информации», чтобы качественно воспроизвести лицо человека в движении.

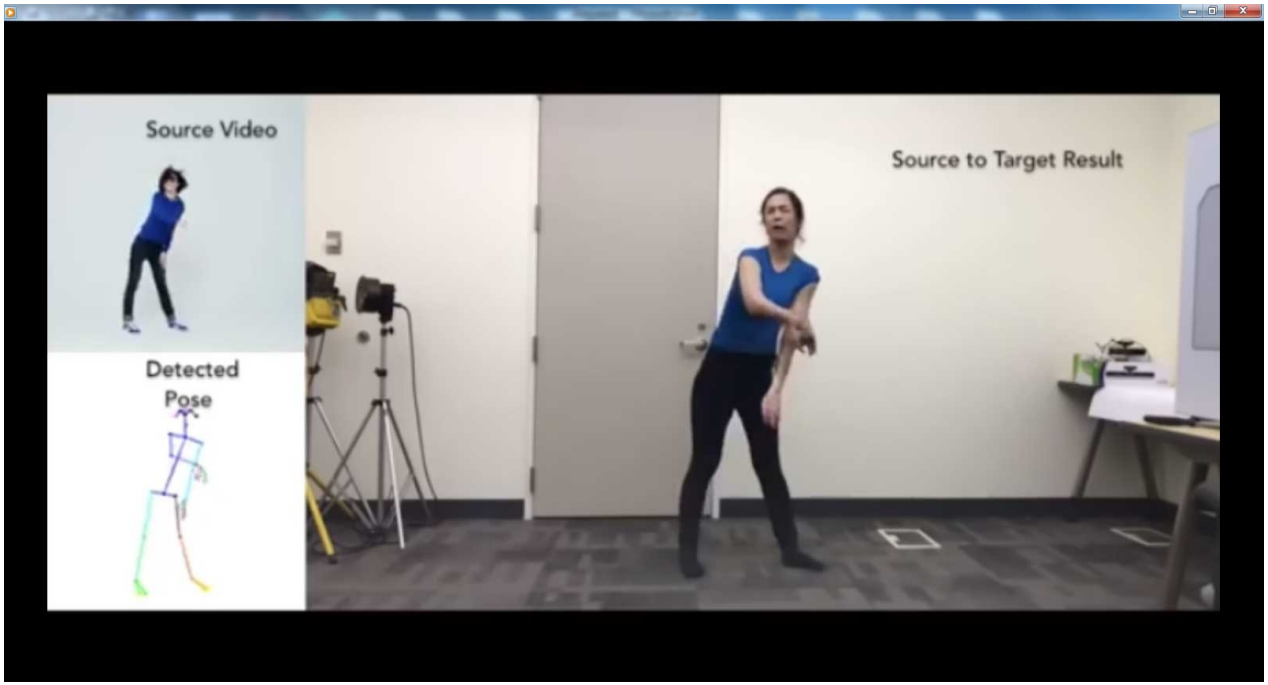
В результате места, «не понятые» искусственным интеллектом, оказываются «размытыми».

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Дело только в лице?

Не только – хотя лицо и есть самый «важный» элемент. В принципе, ИИ все равно, и он может «копировать» движения всего тела целиком.

Благодаря этому… любой сможет танцевать рэп или «Лебединое озеро»



ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Более «продвинутый» вариант

ИИ выстраивает 3D-модель человека. Возможны некоторые корректировки. Машина «учится» дополнять недостающие элементы…

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Киноиндустрия

Цифровое воссоздание актеров

«Звездные войны» – молодая принцесса Лейя обошлась в 200 млн. долларов…

Но сейчас эта технология стала доступна не только «хакеру с банкой пива», но и обычному юзеру соцсети

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Новости

Между прочим, это «робот» ☺ То есть синтезированный ведущий на основе реального, в которого можно «вложить» любой текст. Он будет работать 24 часа в сутки и не требовать высоких гонораров ☺

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Унижение чести и достоинства

Уже сейчас самой «перспективной» технологией «гражданского» применения Deep Fakes видят ее применение в порнографии – для так называемой non-consentual pornography:

- Порнографические ролики со знаменитостями, медийными личностями (создателям фильма «Юлия» дублеры бы не понадобились);
- «Порнография из мести».

Опасность в том, что многие выкладывают в соцсети сотни своих фото и видео – это уже достаточно для обучения ИИ и создания deep fake video.

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Сцены сексуальной эксплуатации несовершеннолетних, груминг

Несовершеннолетние среднего и старшего школьного возраста оставляют достаточно «фотовидеоследов» в Интернете для создания фейкового видео с их участием.

Возможность «продать» и распространить видео CSAM без реальной сексуальной эксплуатации несовершеннолетнего в ее «классическом» понимании (18-я глава УК РФ);

Возможность вымогательства, в том числе реальных сексуальных действий, под угрозой распространения (родители, старшие взрослые, сверстники и учителя поверят!)

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Экстремизм, насильственное изменение политического режима и политической системы

Роберт Хайнлайн, «Если это будет продолжаться», 1940 – подполье «синтезирует» Первого Пророка и вклинивается в трансляцию – «поддельный» вождь призывает к революции – дезориентация населения и правоохранителей - взятие подпольем власти.

В современных условиях – фальшивые видео с дискредитацией политических противников, призывами к смене власти (1916 и 1941 – листовки), провокация массовых беспорядков с возможностью захвата власти

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Пропаганда терроризма, внешняя политика

«Вечно живые» вожди террористических организаций;

Дискредитация противников, «фальшивые признания» без физического контроля за живым субъектом;

«Фальшивые теракты» с целью распыления сил;

Поддельные высказывания политиков с целью нагнетания напряженности в международных отношениях, вплоть до провокации военных действий (Бисмарк, 1870)

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Вопросы безопасности:

КАК это распознавать программно-технически?

КАК строить информационно-просветительскую работу в плане факт-чекинга и доверия информации?

Достаточно ли «блокировок по категориям»?

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА

Спасибо за внимание!

Урван Парфентьев
urvan@nedopusti.ru

ЦЕНТР БЕЗОПАСНОГО ИНТЕРНЕТА