ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2022, том 508, с. 109–110

ПЕРЕДОВЫЕ ИССЛЕДОВАНИЯ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАШИННОГО ОБУЧЕНИЯ

УДК 004.8

ТЕХНОЛОГИИ КОМПЬЮТЕРНОГО ЗРЕНИЯ В ЗАДАЧАХ СИНТЕЗА ВЫСОКОКАЧЕСТВЕННОГО МУЛЬТИМЕДИЙНОГО КОНТЕНТА

© 2022 г. А. В. Кузнецов[1,*], Д. В. Димитров[1], А. Ю. Грошев[1], П. П. Парамонов[1], А. А. Мальцева[1]

Представлено академиком РАН А.Л. Семеновым
Поступило 28.10.2022 г.
После доработки 28.10.2022 г.
Принято к публикации 01.11.2022 г.

Развитие технологий глубокого обучения неизбежно порождает новые задачи и их решения в таких направлениях, как компьютерное зрение, VR/AR технологии, видеоаналитика, мультимодальное обучение и др. С ростом доступности высокопроизводительных вычислительных устройств многие современные методы и средства обработки цифровых данных становятся широко применимыми в том числе в рамках частных прикладных исследований. Данную тенденцию можно легко проследить по росту количества open-source решений, которые без труда запускаются на таких известных ресурсах, как, например, Google Colab. В рамках данного материала мы поделимся полученными результатами в части разработки и исследования прорывных технологий синтеза высококачественного мультимедийного контента, которые имеют широкое применение в таких задачах, как перенос лица.

Ключевые слова: перенос лица, GHOST, one shot, синтез фото, синтез видео

DOI: 10.31857/S2686954322070141

Существование технологии автоматического переноса лица человека на фото или видеоконтент всегда было и будет объектом спора научных и общественных групп по той причине, что алгоритмы создания deep fake контента часто звучат в СМИ в негативном контексте и описываются как средства создания компромата, манипуляции общественным мнением, дезориентации общества в части интерпретации каких-либо событий. Все это в первую очередь несет репутационные риски для физических, юридических лиц, так и государственных структур. Более того, даже сам термин “fake”, входящий в общеупотребимое словосочетании “deep fake”, обозначает подделку, что безусловно вызывает естественное негативное восприятие этого слова. Более того, в условиях распространенной в различных источниках информации о парадигме “цифровой гигиены” или чистоте данных, общество все меньше начинает доверять контенту, демонстрируемому в сети Интернет, что приводит к очевидному снижению степени доверия к мультимедийным данным, в особенности новостного характера.

Несмотря на потенциальный вред, который может нести технология синтеза мультимедийного контента, аппарат таких методов, как перенос лиц на фото или видео безусловно позволяет приносить пользу в различных задачах: съемка фильмов и видеороликов с участием актеров, которые по тем или иным причинам не могут физически присутствовать на съемках, создание нового образовательного, развлекательного и рекламного контента для привлечения пользователей, повышение качества мультимедийного контента и т.д. Стоит перенести фокус внимания с термина “fake” на технологические преимущества, которые дают методы переноса лиц, как становятся понятны очевидные плюсы для общества.

Наша команда давно занимается технологией переноса лица, и одним из главных достижений является алгоритм GHOST (Generative High-fidelity One Shot Transfer) [1]. Он позволяет выполнять перенос лица всего лишь с одного изображения-источника на целевые изображение или видео. Стоит отметить, что превалирующее большинство существующих методов решают задачу переноса лица на видео посредством использования набора кадров, на которых обучается специальная модель извлечения признаков. В основе решения лежит базовая архитектура FaceShifter [2] (перенос лица с изображения на изображение), которая была значительно улучшена в ходе исследований за счет таких нововведений, как функция потерь для области глаз, алгоритм сглаживания маски лица, алгоритм замены лица на видео,

[1] *Sber AI, Москва, Россия*
*****E-mail: AVladimirKuznetsov@sberbank.ru*

а также новый метод стабилизации ключевых точек лица для уменьшения его дрожания на соседних кадрах и этап повышения разрешения. Все это позволило обойти существующие SoTA решения по известным метрикам на 1–2%, а также снизить вычислительную сложность технологии переноса за счет One Shot подхода.

Как мы уже сказали ранее, модель, способная генерировать качественный синтезированный фото и видео контент, при неправильном умысле может нести риски особенно в эпоху информационных противостояний. Поэтому, чувствуя бремя этой большой ответственности, мы разработали модель обнаружения синтезированного моделью GHOST фото и видео контента, которая в отличие от существующих в открытом доступе детекторов deepfake с высокой точностью определяет контент, сгенерированный алгоритмом GHOST. Не ставя перед собой задачу поиска научной новизны, а рассматривая детекцию как чисто инженерную задачу, мы взяли за основу модель [3], которая выиграла на соревновании Kaggle в 2020 г. Архитектура состоит из сверточной сети извлечения векторов признаков изображений EfficientNet-B7 и добавленного слоя классификатора. Данная модель предобучалась на данных, предоставленных организаторами соревнования DeepFakeDetectionChallenge (DFDC). Датасет DFDC – это огромный набор оригинальных видео и полученных на их основе дипфейков с помощью различных доступных на тот момент алгоритмов генерации. В сжатом виде этот датасет занимает почти 500 Гб. Несмотря на то что автором лучшего решения была проделана огромная кропотливая работа по предобработке данных и различным аугментациям, его модель была ориентирована только на дипфейки, представленные в рамках DFDC, а для других методов, включая наш метод GHOST, она была нечувствительна. В ходе экспериментов на основе синтезированной моделью GHOST выборке мы получили обновленные веса модели детекции, которые позволяют значительно повысить качество (F1_Score) обнаружения контента, синтезированного моделью GHOST, с 0.14 до 0.98, сохранив при этом исходные значения качества обнаружения других способов создания deepfake фото и видео.

В заключение хочется отметить, что важно устанавливать нормативные рамки разработки любой технологии, чтобы минимизировать возможные риски использования ее злоумышленниками. Мы показали важность развития аппарата методов синтеза мультимедийного контента для решения полезных обществу задач и планируем дальше повышать качество модели GHOST путем учета оценки положения головы в трехмерном пространстве для улучшения переноса лица в экстремальных углах поворота.

СПИСОК ЛИТЕРАТУРЫ

1. *Groshev A. et al.* GHOST – A New Face Swap Approach for Image and Video Domains // IEEE Access. 2022. Т. 10. С. 83452–83462.
2. *Li L. et al.* Faceshifter: Towards high fidelity and occlusion aware face swapping // arXiv preprint arXiv:1912.13457. 2019.
3. *Das S. et al.* Towards solving the deepfake problem: An analysis on improving deepfake detection using dynamic face augmentation // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. С. 3776–3785.