

Метод создания анимируемых аватаров с использованием нейронных полей излучения и двухмерного нейронного рендеринга

Ю.Х. Козлова

Самарский национальный исследовательский университет им. академика С.П. Королева

Самара, Россия

jpganeeva99@gmail.com

Аннотация — В работе представлен подход к созданию анимируемых аватаров. Метод комбинирует в себе использование двух технологий — нейронных полей излучения, обусловленных априорной информацией, полученной из параметрической модели FLAME, для неявного представления модели головы человека в трехмерном пространстве и 2D нейронного рендеринга для синтеза новых ракурсов на основе обученного представления. В отличие от существующих подходов, которые требуют наличие 3D-сканов или последовательности большого числа изображений человека с разных ракурсов, предложенный подход использует синтетический набор данных, полученный на основе параметрической модели FLAME. Разработанный подход позволяет синтезировать вид человека с новых ракурсов с возможностью изменять выражение лица.

Ключевые слова — нейронные поля излучения, 2D нейронный рендеринг, аватар, неявное представление, синтез новых видов, FLAME

I. ВВЕДЕНИЕ

В связи с развитием и внедрением в повседневную жизнь технологий виртуальной и дополненной реальности методы для создания цифровых аватаров людей становятся всё более востребованными [1]. Под цифровым аватаром человека понимается трехмерное представление человека с анимируемой мимикой, доступное для просмотра с разных точек обзора.

Существующие методы для создания цифровых аватаров можно подразделить на 3 класса: параметрические [2,3], непараметрические [4,5] и комбинированные [6-9] (объединяют преимущества параметрических и непараметрических подходов). В рамках работы для создания анимируемого аватара используется смешанный подход. Параметрическая составляющая позволяет производить экстраполяцию в тех областях, которые не доступны во входных данных, однако она не учитывает наличие волос, аксессуаров, морщин и т.п. Комбинирование с непараметрическим подходом позволяет дополнить эти детали, создавая более полные и реалистичные представления.

II. ОПИСАНИЕ ПРЕДЛОЖЕННОГО МЕТОДА

Предложенный метод создания анимируемого аватара состоит из трех последовательных этапов оптимизации параметров модели. Здесь и далее будем именовать их α , β и γ .

A. Этап α

На рис. 1 представлена схема этапа α . На текущем этапе производится обучение на синтетическом наборе данных, сформированном на основе открытого набора FFHQ [10]. Параметры для инициализации набора данных получены путем оптимизации модели FLAME

на основе метода, представленного в [11]. В результате предварительной обработки каждому изображению ставится в соответствие набор векторов, описывающих форму (shape), выражение лица (expression), позу (pose; челюсть, шея, глаза), текстуру (albedo) и освещение (illumination), а также маска сегментации, внутренние (f_x , f_y , s_x , s_y) и внешние параметры камеры (матрица поворота R , вектор смещения t). Затем для каждого набора фиксируются параметры формы и текстуры, а к остальным параметрам добавляется шум с равномерным распределением в заранее определенном диапазоне.

Для формирования представления анимируемого аватара в трехмерном пространстве используется технология нейронных полей излучения [12]. В отличие от базовой идеи было предложено обусловить модель векторами формы, выражения лица, текстуры и освещения, а также отказаться от вектора направления, чтобы избежать смещения, обусловленного набором данных. Одним из главных ограничений базовой идеи нейронных полей изучения является скорость отрисовки новых видов, поэтому предлагается выполнять синтез новых видов с использованием двухмерной нейронной сети [13]. Информация об ориентации головы в пространстве учитывается при нейронном рендеринге на этапе семплирования лучей.

В качестве функции потерь на всех этапах используется взвешенная сумма компонент: mean square error и perceptual loss [14], на вход которых поступают истинное и предсказанное изображения.

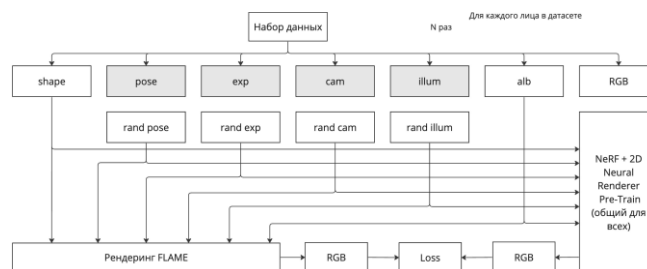


Рис. 1. Схематичное представление этапа α

B. Этап β

На рис. 2 представлена схема этапа β . На текущем этапе производится обучение на наборе данных FFHQ, но в отличие от этапа α , для вычисления функции потерь используются не синтетические изображения, а изображения из набора данных. На этом этапе обучение начинается из весов, полученных на этапе α . Данный этап является опциональным, однако его наличие позволяет повысить детализацию текстуры.

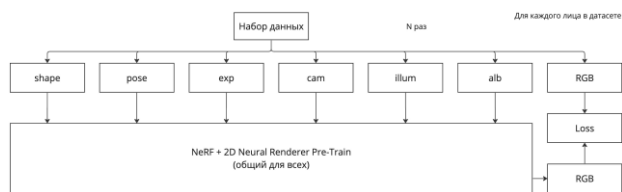


Рис. 2. Схематическое представление этапа β

С. Этап γ

На рис. 3 представлена схема этапа γ. На текущем этапе производится настройка весов модели под конкретного человека. На вход поступает видеопоследовательность, по кадрам которой производится предобработка (как описано на этапе α) и оптимизация. В результате веса модели, векторы формы и текстуры фиксируются и могут использоваться далее для синтеза новых видов.

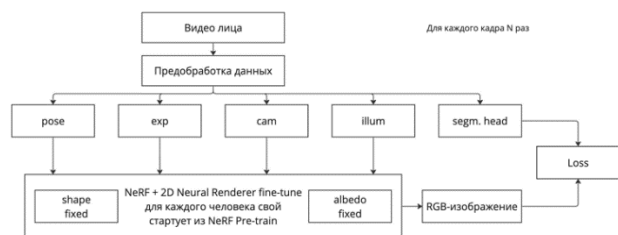


Рис. 3. Схематическое представление этапа γ

D. Синтез новых видов

На основе модели, полученной на этапе γ, можно выполнять синтез новых видов с помощью задания внешних параметров камеры (матрица поворота и вектор смещения). На рис. 4 представлена схема данного этапа.

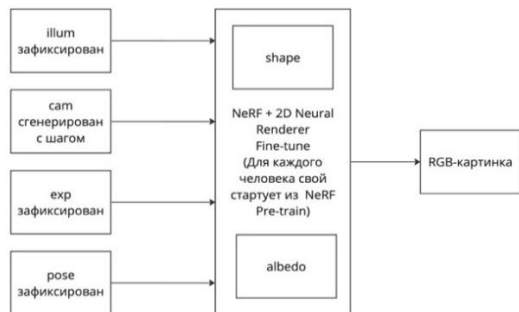


Рис. 4. Схематическое представление этапа синтеза новых видов

Кроме синтеза новых видов представленный метод делает возможным анимацию аватара, что подразумевает изменение вектора выражения лица на каждом кадре. Эти векторы могут быть получены путем предобработки новой видеопоследовательности (см. этап α). Важно отметить, что при этом веса модели и все векторы, за исключением вектора выражения лица, фиксируются. Перенос выражения лица может быть выполнен на основе видеопоследовательности с участием другого человека.

III. ЗАКЛЮЧЕНИЕ

В данной статье предложен трехэтапный подход к созданию анимируемых аватаров, основанный на использовании нейронных полей излучения и 2D

нейронного рендеринга. Основным преимуществом представленного подхода является использование синтетического набора данных, сформированного на базе параметрической модели FLAME, что позволяет избежать необходимости в наличии 3D-сканов или изображений людей с большого числа ракурсов в обучающей выборке. Этот подход упрощает создание аватаров, снижая зависимость от сложных процессов 3D-сканирования и увеличивая доступность технологии.

ЛИТЕРАТУРА

- [1] Korban, M. A Survey on Applications of Digital Human Avatars toward Virtual Co-presence / M. Korban, X. Li. // arXiv:2201.04168. – 2022. DOI: 10.48550/arXiv.2201.04168.
- [2] Gerig, T. Morphable Face Models - An Open Framework / T. Gerig, A. Morel-Forster, C. Blumer et al. // 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). – 2018. – P. 75-82. DOI: 10.1109/FG.2018.00021.
- [3] Li, T. Learning a model of facial shape and expression from 4D scans / T. Li, T. Bolkart, M.J. Black et al. // ACM Transactions on Graphics. – 2017. – Vol. 36(6). – P. 1-17. DOI: 10.1145/3130800.3130813.
- [4] Dou, P. End-to-end 3D face reconstruction with deep neural networks / P. Dou, S.K. Shah, I.A. Kakadiaris // Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. – 2017. – P. 1503-1512. DOI: 10.1109/CVPR.2017.164.
- [5] Jackson, A.S. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression / A.S. Jackson, A. Bulat, V. Argyriou, G. Tzimiropoulos // 2017 IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 1031-1039. DOI: 10.1109/ICCV.2017.117.
- [6] Feng, Y. Learning an animatable detailed 3D face model from in-the-wild images / Y. Feng, H. Feng, M.J. Black, T. Bolkart // ACM Transactions on Graphics. – 2021. – T. 40(4). – P. 88:1-88:13. DOI: 10.1145/3450626.3459936.
- [7] Grassal, P.-W. Neural Head Avatars from Monocular RGB Videos / P.-W. Grassal, M. Prinzel, T. Leistner et al. // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2022. – P. 18632-18643. DOI: 10.1109/CVPR52688.2022.01810.
- [8] Hong, Y. HeadNeRF: A Realtime NeRF-based Parametric Head Model / Y. Hong, B. Peng, H. Xiao et al. // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2022. – P. 20342-20352. DOI: 10.1109/CVPR52688.2022.01973.
- [9] Zhuang, Y. MoFaNeRF: Morphable Facial Neural Radiance Field / Y. Zhuang, H. Zhu, X. Sun, X. Cao // Computer Vision – ECCV 2022: Lecture Notes in Computer Science. – 2022. – Vol. 13663. – P. 268-285. DOI: 10.1007/978-3-031-20062-5_16.
- [10] Karras, T. A Style-Based Generator Architecture for Generative Adversarial Networks / T. Karras, S. Laine, T. Aila // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 4396-4405. DOI: 10.1109/CVPR.2019.00453.
- [11] HavenFeng/photometric_optimization: Photometric optimization code for creating the FLAME texture space and other applications [Electronic resource]. – Access mode: https://github.com/HavenFeng/photometric_optimization (03.12.2023).
- [12] Mildenhall, B. NeRF: representing scenes as neural radiance fields for view synthesis / B. Mildenhall, P.P. Srinivasan, M. Tancik et al. // Communications of the ACM. – 2021. – T. 65(1). – P. 99-106. DOI: 10.1145/3503250.
- [13] StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis / J. Gu, L. Liu, P. Wang, C. Theobalt. – 2021. DOI: 10.48550/arXiv.2110.08985.
- [14] Johnson, J. Perceptual Losses for Real-Time Style Transfer and Super-Resolution / J. Johnson, A. Alahi, L. Fei-Fei // Computer Vision – ECCV 2016: Lecture Notes in Computer Science. – 2016. – Vol. 9906. – P. 694-711. DOI: 10.1007/978-3-319-46475-6_43.