

Название статьи (авторы статьи): “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis” (Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng)

Автор рецензии: Александра Сендерович

В статье предлагается метод, способный по нескольким фотографиям объекта с разных ракурсов закодировать 3D-модель этого объекта как нейронную сеть с полносвязными слоями и генерировать изображения объекта с новых ракурсов. Описана дифференцируемая процедура рендеринга изображения, включающая в себя предложенное авторами иерархическое сэмплирование, и позиционное кодирование. Иерархическое сэмплирование точек позволяет более эффективно восстанавливать изображение, избегая точек, попадающих в пустые или закрытые участки. Позиционное кодирование улучшает способности модели по воспроизведению мелких деталей.

Сильные стороны:

1. Авторы статьи протестировали свой метод на нескольких наборах данных и привели сравнение своего метода с предыдущими методами, использующимися для этой же цели. Почти во всех случаях метрики предложенного метода оказались лучше метрик предыдущих методов, что позволяет сделать вывод о полезности предложенного метода. В пользу этого вывода говорят и приведённые авторами сравнения картинок, сгенерированных различными методами: NeRF генерирует более детальные картинки, чем остальные алгоритмы.
2. Чтобы было проще иллюстрировать результаты работы алгоритма, авторы собрали свой небольшой набор данных.
3. Кроме этого, интересно, что авторы провели ablation study, позволяющий оценить важность предложенных авторами модификаций (иерархический семплинг и positional encoding), а также влияние числа входных картинок на качество результата. Такое сравнение говорит о том, что авторы хорошо изучили возможные варианты запуска модели и предложили наиболее оптимальный.
4. Сам по себе предложенный метод довольно простой: используются MLP и базовые алгоритмы рендеринга.

Слабые стороны:

1. По моему мнению, главный недостаток, бросающийся в глаза и не позволяющий в полной мере понять значимость и применимость предложенного алгоритма -- это отсутствие детального сравнения по времени и памяти с другими методами. В статье есть небольшой абзац, в котором говорится о компромиссе между памятью и временем обучения других методов восстановления изображений. Однако в нём не для всех методов приведены точные цифры, наглядно показывающие упомянутый компромисс. Почему-то сравнение по занимаемой моделью памяти проведено только с методом LLFF. Кроме этого, интересно также сравнить время применения разных методов (а не только время обучения).
2. Кроме отсутствия сравнения работы разных алгоритмов по памяти и времени интересно было бы посмотреть на сравнение работы рассмотренных методов при разном числе входных изображений.

В самой статье довольно сложно разбираться, не имея предварительного опыта чтения статей на ту же тему: используется много специализированных терминов, которые никак не поясняются в тексте. При этом статья очень хорошо иллюстрирована, что помогает разобраться в идее, предложенной авторами (но не помогает читателю поместить работу авторов в более глобальный контекст). Кроме этого, для иллюстрации работы алгоритма авторы записали вспомогательное видео.

Мелкие детали реализации и некоторые длинные выкладки в самой статье опущены и приведены в аппендиксах, что только улучшает её читаемость. Статья достаточно подробна, чтобы её можно было воспроизвести, кроме того, авторы предоставили гитхаб-репозиторий с кодом и инструкциями по его запуску.

Оценка: 10, уверенность: 3 (так как я не профессионал в области компьютерной графики).