

Streszczenie popularnonaukowe

Neuronowe pola radiacyjne (zwane z ang. NeRFami) to rodzaj algorytmu głębokiego uczenia, który jest wykorzystywany do generowania realistycznie wyglądających obrazów obiektów trójwymiarowych. NeRFy są w stanie uwzględnić skomplikowane parametry oświetlenia i cieniowania obrazu, dzięki czemu są stosowane w wielu obszarach, takich jak grafika komputerowa, rzeczywistość wirtualna i obrazowanie medyczne. Jednym z ich głównych ograniczeń jest złożoność obliczeniowa, co może utrudniać ich wykorzystanie w aplikacjach na dużą skalę lub nawet, w czasie rzeczywistym. Przykładowo, niedawno opracowany model do kodowania filmów, potrzebuje 56 dni obliczeń na karcie graficznej do zakodowania zaledwie 10-sekundową sekwencję.

Aby rozwiązać problem długiego trenowania NeRFów, badacze zaproponowali zastąpienie nadmiernie sparametryzowanych modeli sieci neuronowych prostszymi podejściami, inspirowanymi tradycyjną grafiką komputerową. Prowadzi to do szybszego działania, ale ma dwie istotne wady: zużycie pamięci wymagane do trenowania tych metod rośnie sześciennie, a charakterystyka danych wizualnych jakimi są filmy prowadzi do powstawania artefaktów renderingu i obniża jakość końcowego wyniku.

W tym projekcie, naszym celem jest zwiększenie efektywności treningu NeRFów, co ma doprowadzić do ich szerszej stosowalności w różnych rzeczywistych aplikacjach. Nasza wstępna praca nad kontrolowanymi NeRFami (CoNeRF), opublikowana na konferencji CVPR'22, pokazuje, że pierwsze kroki w kierunku rozwiązania powyższego celu są wykonalne, jednak rozszerzenie tego podejścia na bardziej elastyczną konfigurację bez adnotacji pozostaje otwartym wyzwaniem.

Aby sprostać temu wyzwaniu, zauważmy, że filmy zawierają wiele nadmiarowych informacji, jako że wiele kolejnych klatek jest podobnych do siebie. Postawiliśmy więc hipotezę, że możemy wykorzystać tę nadmiarowość do zakodowania filmów jako pól radiacyjnych w sposób wydajny obliczeniowo. Powinno to zredukować wymagania obliczeniowe postawione sieciom neuronowym, a im samym skupić się na szczegółach wizualnych, których brakuje w podejściach literaturowych. Dodatkowo, biorąc pod uwagę, że w przypadku filmów kręconych pod różnymi kątami widzenia, dane treningowe składają się z pojedynczych klatek przechwyconych jednocześnie przez wiele kamer. Jedną z alternatyw byłoby wytrenowanie oddzielnej sieci NeRF dla każdej klatki. Takie podejście pozwala na istotne zrównoleglenie całego procesu.

Finalnym rezultatem naszego projektu będzie wydajna obliczeniowo metoda tworzenia realistycznych i kontrolowanych wyjść wideo w trybie free-viewpoint, wykorzystująca neuronowe pola radiacyjne.

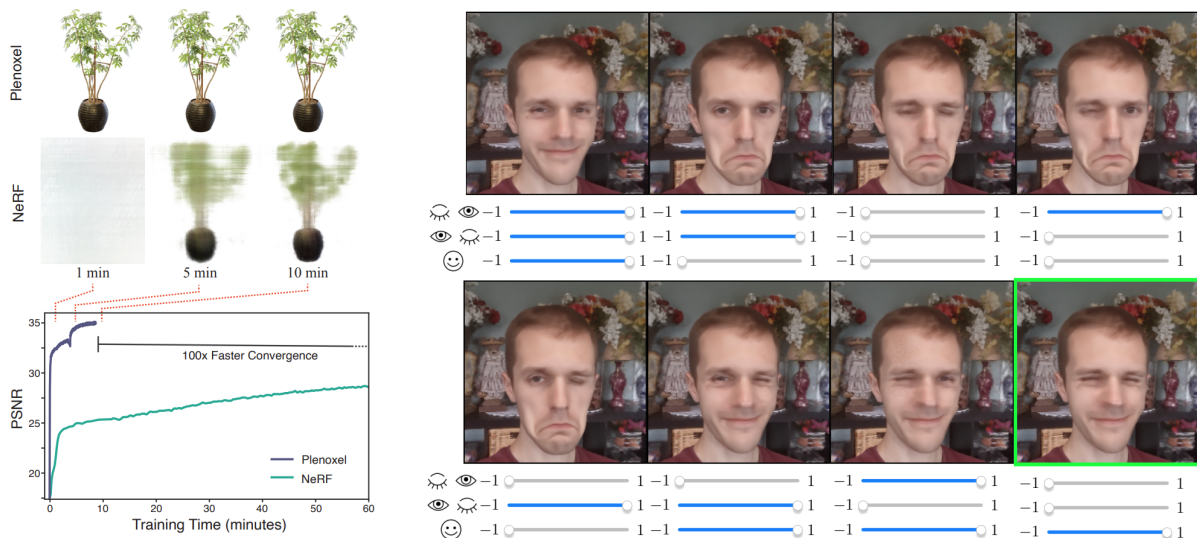


Figure 1: NeRFy pozwalają użytkownikom na syntezę nowych, niewidzianych widoków 3D sceny lub obiektu, np. ludzkiej twarzy, na podstawie zestawu zdjęć lub krótkiego filmu. NeRFy znalazły zastosowanie w wielu rzeczywistych aplikacjach, które wymagają renderowania nowych scen i widoków, takich jak animacja twarzy lub ludzkich awatarów, generowanie postaci w grach komputerowych czy estymacja pozycji kamer w robotyce. Metody te opierają się jednak na dużych modelach sieci neuronowych z milionami parametrów, które wymagają czasochłonnego treningu. Choć udało się znacząco zredukować czas treningu NeRFów dla scen statycznych (dzięki metodom takim jak przedstawiony na rysunku po lewej Plenoxels), metody operujące na filmach wideo nakręconych ze zmienną zawartością obrazu (rysunek po prawej) nadal wymagają dużej ilości zasobów do ich treningu. W naszym projekcie zajmujemy się tym ograniczeniem i proponujemy szybsze, bardziej wydajne metody treningu NeRFów dla filmów.