

# Нейронные поля яркости – сделать большое из малого

Михаил Житомирский

Последние несколько лет в лексикон создателей медиа-контента вошел термин Neural Radiance Fields (NeRF) – нейронные поля яркости. По сути, это еще одно воплощение алгоритмов искусственного интеллекта. Определение такого рода сети выглядит следующим образом: «*Neural Radiance Field – это полностью связанная нейронная сеть, способная генерировать новые виды сложных 3D-сцен на основе частично набора двумерных изображений*».

Проще говоря, располагая некоторым количеством фотографий достаточного качества, можно с помощью NeRF создать – смоделировать – полноценные трехмерные сцены с такой степенью реалистичности, как будто они были сняты камерой. Конечно же, кинематографисты не могли пройти мимо этой технологии, особенно те, что любят экспериментировать с новыми методами и способами создания изображения.

Если вдуматься, это вполне закономерно. Ведь в основе кинематографа лежит именно фотография. Из нее вышел иллюзион движущихся картинок, и кинематографисты до сих пор пользуются фотографией, но уже по-новому, применяя ее в сфере визуальных эффектов. Достаточно вспомнить массив фотокамер, применявшийся в фильме «Матрица» для создания эффекта Bullet Time (время полета пули). Этот эпизод вызывает ощущение, что время в кадре в какой-то момент практически полностью остановилось, но камера продолжает двигаться.

Чтобы добиться этого эффекта, применили технику фотосъемки, известную как «временной срез». Суть ее в том, что вокруг объекта устанавливают большое количество фотокамер, одновременно снимающих этот объект. Затем формируется следующая последовательность, первый кадр с первой камеры, первый кадр со второй камеры, первый кадр с третьей камеры и т. д. В результате получается как бы круговой облет неподвижного объекта. Правда, в «Матрице» этот метод был немного модернизирован так, чтобы в кадре были еще медленно движущиеся персонажи фильма, а скорость итоговой последовательности составила 12 тыс. кадр/с. Так что при воспроизведении со стандартной скоростью 24 кадр/с эффект «остановки времени» получился более чем впечатляющим. И это лишь один из примеров успешного применения фотографии в кинематографе.

На новом витке технологического прогресса фотография встретила с технологией NeRF. Она, как уже отмечалось, позволяет создавать трехмерные модели объектов и пространств на базе фотоснимков или иных двумерных изображений. Пока основной сферой применения нового метода является создание превизов, контента для визуализации в пространствах, построенных из светодиодных экранов, для съемки с применением технологий виртуальной реальности.

Надо сказать, что еще в начале августа 2020 года была опубликована работа под названием «NeRF: представление сцен в виде нейронных полей яркости для синтеза изо-

## NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

Ben Mildenhall<sup>1\*</sup> Pratul P. Srinivasan<sup>1\*</sup> Matthew Tancik<sup>1\*</sup>  
Jonathan T. Barron<sup>2</sup> Ravi Ramamoorthi<sup>3</sup> Ren Ng<sup>1</sup>

<sup>1</sup>UC Berkeley <sup>2</sup>Google Research <sup>3</sup>UC San Diego

**Abstract.** We present a method that achieves state-of-the-art results for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views. Our algorithm represents a scene using a fully-connected (non-convolutional) deep network, whose input is a single continuous 5D coordinate (spatial location  $(x, y, z)$  and viewing direction  $(\theta, \phi)$ ) and whose output is the volume density and view-dependent emitted radiance at that spatial location. We synthesize views by querying 5D coordinates along camera rays and use classic volume rendering techniques to project the output colors and densities into an image. Because volume rendering is naturally differentiable, the only input required to optimize our representation is a set of images with known camera poses. We describe how to effectively optimize neural radiance fields to render photorealistic novel views of scenes with complicated geometry and appearance, and demonstrate results that outperform prior work on neural rendering and view synthesis. View synthesis results are best viewed as videos, so we urge readers to view our supplementary video for convincing comparisons.

**Keywords:** scene representation, view synthesis, image-based rendering, volume rendering, 3D deep learning

### 1 Introduction

In this work, we address the long-standing problem of view synthesis in a new way by directly optimizing parameters of a continuous 5D scene representation to minimize the error of rendering a set of captured images.

We represent a static scene as a continuous 5D function that outputs the radiance emitted in each direction  $(\theta, \phi)$  at each point  $(x, y, z)$  in space, and a density at each point which acts like a differential opacity controlling how much radiance is accumulated by a ray passing through  $(x, y, z)$ . Our method optimizes a deep fully-connected neural network without any convolutional layers (often referred to as a multilayer perceptron or MLP) to represent this function by regressing from a single 5D coordinate  $(x, y, z, \theta, \phi)$  to a single volume density and view-dependent RGB color. To render this *neural radiance field* (NeRF)

\* Authors contributed equally to this work.

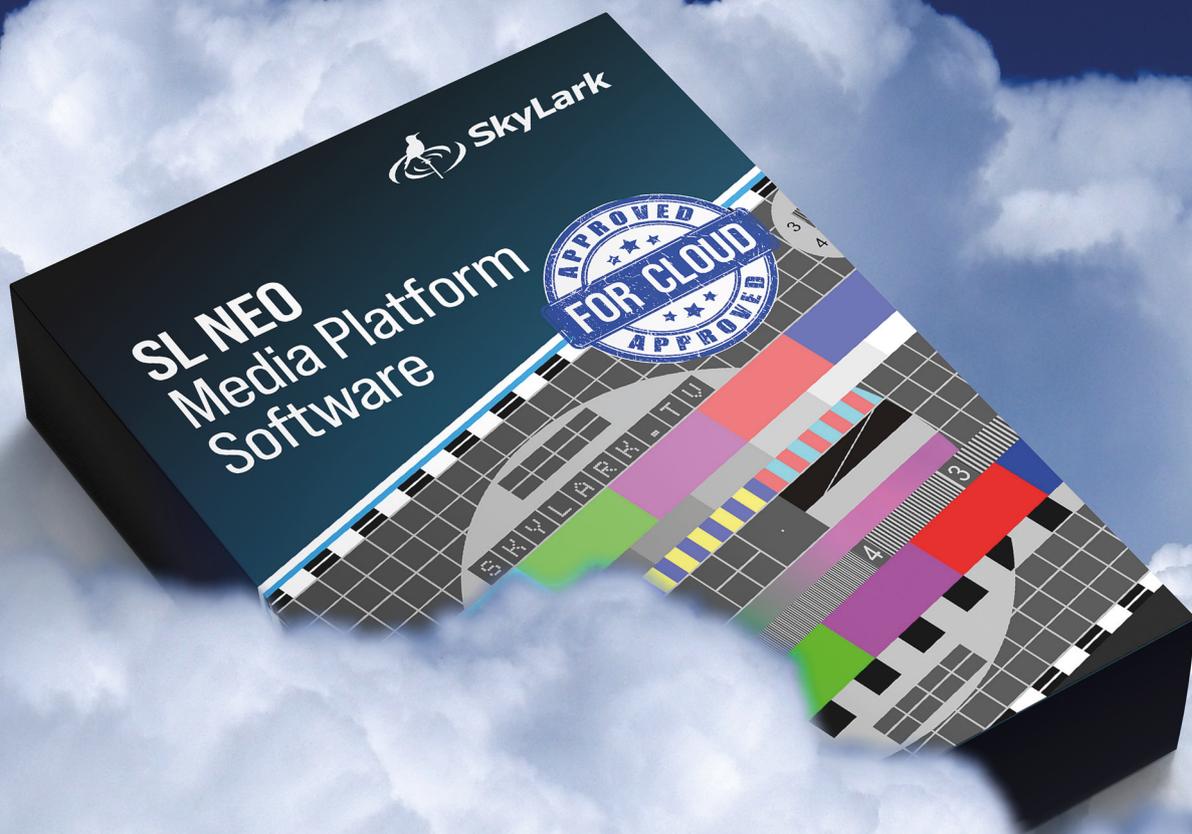
*Титульный лист работы «NeRF: представление сцен в виде нейронных полей яркости для синтеза изображений»*

бражений» (NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis). Авторами были Бен Майлденхолл, Пратул П. Сринивасан, Мэтью Танцик и Рен Нг (все из Калифорнийского университета в Беркли), Джонатан Т. Баррон из Google Research и Рави Рамамурти из Калифорнийского университета в Сан-Диего.

Они представили метод, позволяющий достичь высоких результатов при синтезировании новых ракурсов сложных сцен путем оптимизации исходной функции непрерывной объемной сцены с использованием разреженного набора входных представлений. Предложенный алгоритм позволяет синтезировать сцену, используя полностью связанную глубокую сеть, входными данными для которой является единое непрерывное 5-координатное пространственное положение, а на выходе формируются данные объемной плотности и зависимый от ракурса просмотра излучаемый свет в данной пространственной точке. Пять координат – это пространственные  $(x, y, z)$  и координаты вектора направления просмотра  $(\theta, \phi)$ .

arXiv:2003.08934v2 [cs.CV] 3 Aug 2020

skylark.ru  
скайларк.рф



ТВОРИТЕ И СОЗДАВАЙТЕ  
МЫ ПОЗАБОТИМСЯ ОБО ВСЁМ ОСТАЛЬНОМ



Авторы работы синтезировали новые ракурсы, запрашивая 5D-координаты вдоль лучей камеры, и использовали классические методы объемной визуализации (рендеринга) для вывода цветов и плотностей в изображение. Поскольку объемный рендеринг естественно дифференцируем, единственными входными данными, необходимыми для оптимизации синтеза изображений, является набор изображений, для которых точно известны положения камер.

А сама работа посвящена тому, как эффективно оптимизировать нейронные поля яркости для визуализации фотореалистичных новых ракурсов сцен со сложными геометрией и внешним видом. Результаты уже тогда – более трех лет назад – оказались лучше того, что делалось ранее в сфере нейронного рендеринга и синтеза изображений.

Чтобы завершить это краткое научное отступление, можно привести пример того, как используется метод оптимизации непрерывного 5D-представления нейронного поля яркости (объемная плотность и зависящий от угла наблюдения цвет в любой точке непрерывного пространства) для сцены на базе набора входных изображений.

Для сбора исходных изображений сцены использовались методы пространственного рендеринга в направлении оптических осей камер для визуализации сцены под любым ракурсом. В данном примере визуализация проводилась на базе 100 исходных ракурсов синтетической сцены с ударной установкой. Исходные ракурсы были взяты случайным образом вдоль поверхности воображаемой полусферы вокруг установки, а результатом стали два новых ракурса, синтезированные с помощью NeRF.

По большому счету, данный метод можно рассматривать как эволюцию фотограмметрии – еще одного метода, предполагающего использование определенного числа фотографий для извлечения 3D-информации об объектах и пространствах.

Однако есть и одно существенное различие между этими двумя методами. При фотограмметрии текстура фотографии применяется напрямую к синтезируемой модели. А значит, если на фотографии есть, в частности, тени, то они будут и на модели, что ограничивает возможности использования модели в условиях освещения, отличных от тех, что были при съемке исходной фотографии.

NeRF, в отличие от этого, представляет яркость в каждой точке трехмерного пространства. То есть вместо того, чтобы фиксировать, как свет взаимодействует с объектом или пространством в момент съемки фотографии, NeRF

фокусируется на том, как свет МОЖЕТ взаимодействовать с объектом практически в любой ситуации. Это позволяет менять освещение на стадии последующей обработки в дополнение к визуализации сцены под новыми ракурсами.

Суть в том, что, как уже отмечалось, метод NeRF опирается на объемную фиксацию данных о пространстве или объекте, где цвет зависит от ракурса. Поэтому при визуализации синтетических изображений они получаются фотореалистичными, а при изменении ракурса соответственно меняются отражения и отраженный свет, и при этом изображение будет по-прежнему восприниматься как реально снятое.

Надо признать, что метод NeRF настолько нов, что пока не получил широкого распространения. Тем не менее прогресс в сфере искусственного интеллекта (AI) и машинного обучения поможет изменить ситуацию. В частности, активные усилия для этого прилагают компании Nvidia и Luma. Причем Luma AI продвинулась настолько, что выпустила одноименное приложение для смартфона, обеспечивающее и упрощающее фиксацию и обработку NeRF, а совсем недавно компания анонсировала и программный модуль для Unreal Engine, позволяющий выполнять рендеринг NeRF в режиме реального времени.

Как считают в Luma AI, одним из вариантов эффективного применения NeRF-технологии является создание превизов. Как минимум, в плане существенной экономии финансов, ведь вместо полноценной съемочной группы на предполагаемую локацию можно отправить всего несколько человек со смартфонами, а на основе снятого ими материала синтезировать изображение, превосходящее то, что позволяет сделать фотограмметрия. Ну и, разумеется, создание изображений для светодиодных экранов с использованием возможностей Unreal Engine.

Но этим возможности NeRF-визуализации не исчерпываются. Все чаще кинематографисты используют приложение Luma AI для сканирования предметов реквизита с последующей интеграцией полученной трехмерной модели в реальную сцену. Конечно, не всегда удастся получить полную реалистичность, часто приходится дорабатывать модель уже на компьютере. Но получить представление о том, как будет выглядеть сцена с тем или иным реквизитом, используя его модели, вполне можно.

Оптимистично настроенные профессионалы считают, что NeRF в сочетании с другими методами синтеза изображений на базе AI, такими как Midjourney и Stable Diffusion,



Синтезирование новых ракурсов на базе набора исходных двумерных изображений

# ГЕНЕРАТОРЫ ОПОРНЫХ СИНХРОСИГНАЛОВ

## Генераторы автономные:



### PSGP-2059 – Генератор опорных видеосигналов и сигналов 1PPS, 10 МГц, PTP, NTP, LTC, WC

- автономный и ведомый режимы работы;
- стабильность в автономном режиме –  $1 \times 10^{-10}$
- ведение от GPS/GLONASS, PTP
- формирует видеосигналы синхронизации: «чёрное поле», Tri-Level и импульсные синхросигналы 1PPS, 10 МГц, LTC, WC; поддержка ST 2059
- формирует сигналы синхронизации времени NTP, PTP ST 1588
- встроенный приемник GPS/GLONASS
- два порта Ethernet – PTP и Control, порт RS-232 для навигационной информации
- в ведомом режиме ошибка положения импульса 1PPS не превышает 100 нс
- в автономном режиме уход импульса 1PPS не превышает 1 мкс за 3 ч

### Модель PSGP-2059RR:

- работает с выносным приемником GPS/GLONASS PGL-259
- компенсация задержки импульса 1PPS – в зависимости от длины кабеля от приемника до генератора

### PSG-2070 – Генератор синхросигналов 3G/HD/SD и испытательных сигналов

- автономный и ведомый режимы работы;
- стабильность в автономном режиме –  $1 \times 10^{-10}$
- ведение от опорных видеосигналов и от GPS/GLONASS
- формирует видеосигналы синхронизации: «чёрное поле», Tri-Level и импульсные синхросигналы 1PPS, 10 МГц, WC, LTC, аудио
- испытательные сигналы: аналоговые (PAL/SECAM), цифровые HD/SD-SDI, аудио аналоговые и цифровые AES/EBU
- измерение расхождения во времени видео- и аудиосигналов в аналоговых, цифровых и смешанных комплексах
- NTP-сервер



### PGL-259 – приемник GPS/GLONASS

- фантомное питание
- изолированная шина питания
- длина кабеля от генератора до приемника – до 300



### PNTIP-5021 – Сервер точного времени

- стабильность в автономном режиме –  $1 \times 10^{-10}$
- выполнение функций сервера NTP/STRATUM 1) в сетях IP
- формирование 1PPS, 10 МГц, LTC
- измерение временного интервала между внутренним 1PPS и внешним TIME CAPTURE сигналами
- приемник GPS/GLONASS

## Генераторы модульные:

### Модули PROFNEXT



#### PN-SGP-321 – Генератор сигналов 1PPS, 10 МГц, PTP, NTP

- автономный и ведомый режимы
- стабильность в автономном режиме –  $1 \times 10^{-10}$
- ведение от GPS/GLONASS, PTP
- выносной приемник GPS/GLONASS PGL-259, длина кабеля до генератора – до 300 м
- формирует импульсы 1PPS, 10 МГц (форма прямоугольная или синусоидальная)
- формирует сигналы синхронизации времени NTP, PTP ST 1588
- два порта Ethernet – PTP (слот SFP) и Control.

### Модули PROFLEX



#### PFSG-7317 – Генератор синхросигналов ТВ высокой и стандартной четкости

- автономный и ведомый режимы
- стабильность в автономном режиме –  $1 \times 10^{-6}$
- ведение от опорных видеосигналов
- формирует видеосигналы синхронизации «чёрное поле» и Tri-Level.

## Общее для всех моделей:

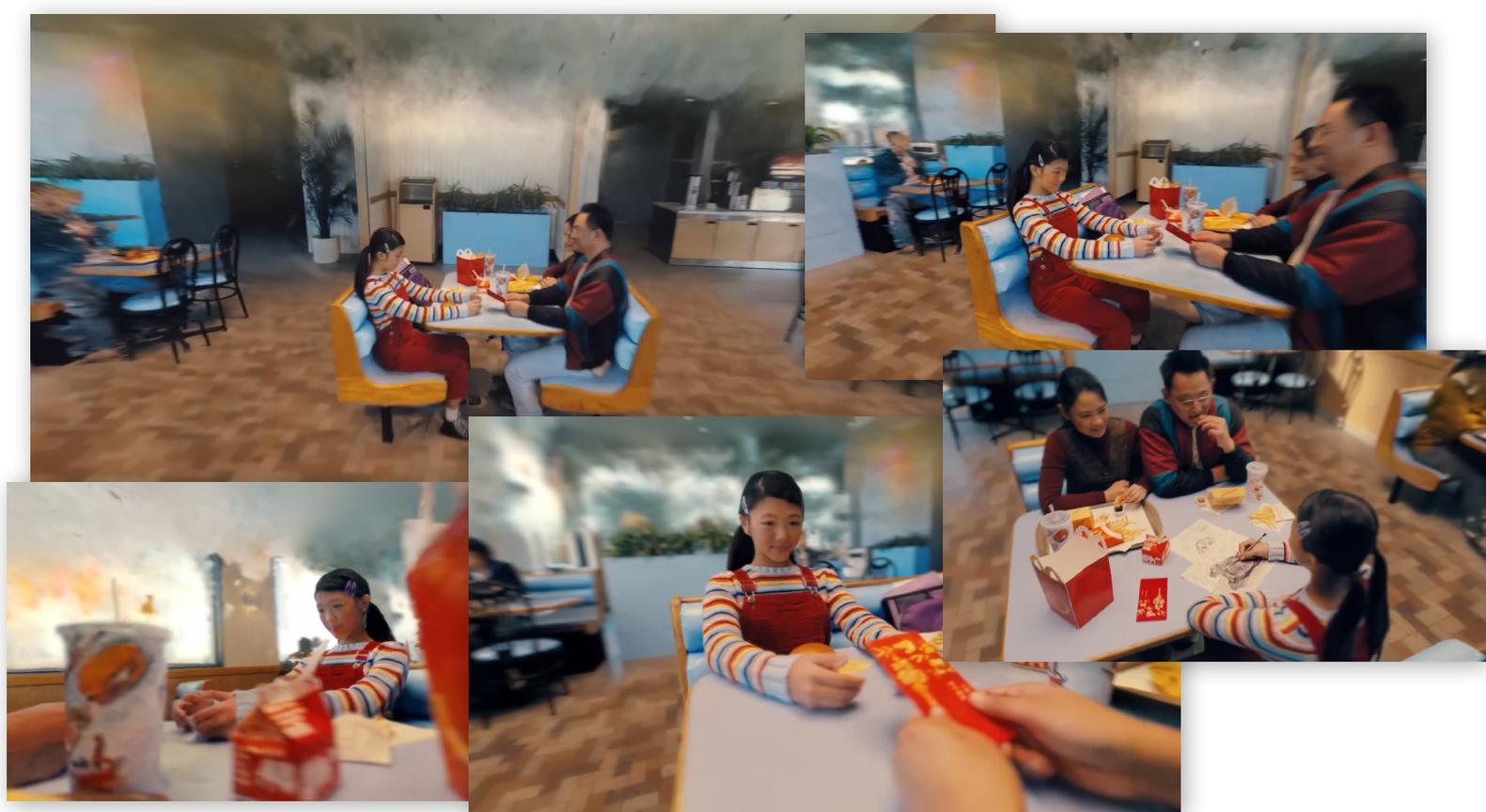
- ♦ Управление генераторами, серверами точного времени – web-интерфейс, SNMP
- ♦ Горячие резерв и замена блоков питания (кроме PNTIP-5021)
- ♦ Генераторы, сервер точного времени и выносной приемник комплектуются магнитной антенной с кабелем длиной 10 м
- ♦ Могут комплектоваться наружной антенной с кабелем длиной до 80 м без усилителя и до 140 м с усилителем

ПРОФИТТ

www.profit.ru

E-mail: info@profit.ru

Тел./факс: (812) 297-7032, 297-7120/22/23, 297-5193



Кадры из рекламного ролика «Лунный Новый год 2023»

со временем станет способна формировать изображения такого качества, которое вполне подходит для включения в финальный релиз кинофильма.

Что касается Nvidia, то эта компания тоже не теряет времени даром. Она выпустила Instant NeRF – серию приложений и кодов, позволяющих разработчикам быстро трансформировать статичные изображения в объемные модели и пространства. При использовании графических процессоров Nvidia процедура занимает считанные секунды. Ну а поскольку кинематографисты – люди творческие, они порой находят разным технологиям неожиданное применение. К примеру, уже есть случаи, когда творческая группа, снимающая сериал, применяет NeRF для сбора данных о той или иной локации, чтобы, если спустя какое-то время придется вернуться к ней, иметь возможность воссоздать ее в том виде, в котором она была в момент съемки. Ведь за два-три года многое может измениться, например, могут исчезнуть какие-то здания либо, наоборот, будут возведены новые объекты. Как правило, такое синтезированное изображение подается на светодиодные фоны.

Одним из первых коммерческих проектов, на котором применялась технология NeRF, стал 30-секундный рекламный ролик «Лунный Новый год 2023» (Lunar New Year 2023) компании Макдональдс, режиссером которого был Хосе М. Нортон, а оператором-постановщиком – Синг Хоуи Ям. Для визуализации детских воспоминаний главной героини ролика съемочная группа применила приложение Luma AI. Основу сцены составляло детское фото героини с ее родителями в ресторане Макдональдс. Персонажи были статичны, а камера облетала их, поднималась над ними, заглядывала в коробку Happy Meal.

Съемку делали камерой Red Monstro 8K VV с объективами Panavision Primo 70 и дискретными объективами VV, камеру ставили на стедикам. Это было нужно для полу-

чения максимально стабильного изображения, из которого потом извлекали нужные статичные кадры для использования в качестве исходных для NeRF.

Конечно, реальный рабочий процесс создания ролика был существенно сложнее, чем может показаться после прочтения его предельно краткого описания. На самом деле потрудиться пришлось и актерам, и операторам, и специалистам по обработке снятого материала. Но в итоге все получилось.

Как отмечают те, кто уже применял NeRF, технология хороша еще и тем, что позволяет совершать камерой движения, невозможные в реальности. Но есть и подводные камни – если виртуальная камера окажется в точке, в которой реальная камера не сняла ничего, то синтезируемое для этой точки изображение будет состоять полностью из интерполированных данных, и какими они будут, пока отдается на откуп «воображению» искусственного интеллекта. Так что, как и при реальной съемке, ключом к успеху остается тщательное планирование.

Тема NeRF богата, и здесь есть о чем говорить. Как и у любой новой технологии, у нейронных полей яркости есть и многочисленные достоинства, и определенные сложности, которые, несомненно, будут преодолены. В целом же NeRF – это еще одна форма машинного обучения в быстро эволюционирующем мире создания изображений и манипуляций с ними с использованием искусственного интеллекта. Нейронные поля яркости будут применяться в сочетании с другими технологиями и методами синтеза изображений. Вполне вероятно, что появятся целые направления создания контента без использования традиционной съемки, а только с помощью его генерирования с применением средств AI. Однако не думаю, что в обозримом будущем такая профессия, как кинооператор, уступит место искусственному интеллекту. ►