

Гибридный метод Kling-Grok: от изображений к реалистичным видеонаративам

Пайдершев Алексей Константинович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Цель исследования — показать возможности аудиовизуальной нейросетевой платформы Kling для поэтапной генерации реалистичных видеороликов путём комбинации визуальных референсов и детализированных текстовых промптов, а также мультимодальной платформы Grok для создания фотореалистичных изображений. Такой подход гарантирует единство композиционных, стилистических и эмоционально-смысловых элементов визуальных решений в когерентной нарративной структуре. Kling выбрана за высокую эффективность синтеза реалистичного видео через интеграцию текста и изображений, а Grok — за способность генерировать реалистичные изображения на основе текстовых промптов. Гибридный метод позволяет точно воспроизводить черты объектов и целенаправленно корректировать параметры композиции, освещения и стиля в соответствии с исследовательскими и творческими задачами.

Ключевые слова: изображение, видео, промт, нейросеть, Kling, Grok.

The Kling-Grok hybrid method: from images to realistic video narratives

Paydershev Alexey Konstantinovich

Sholom-Aleichem Priamurskiy State University

Student

Abstract

The purpose of the study is to demonstrate the capabilities of the Kling audiovisual neural network platform for the step-by-step generation of realistic videos by combining visual references and detailed text prompts, as well as the Grok multimodal platform for creating photorealistic images. This approach ensures the unity of compositional, stylistic, and emotional-semantic elements of visual solutions within a coherent narrative structure. Kling was chosen for its high efficiency in synthesizing realistic videos through the integration of text and images, while Grok was selected for its ability to generate realistic images based on text prompts. The hybrid method allows you to accurately reproduce the features of objects and purposefully adjust the composition, lighting, and style parameters according to your research and creative objectives.

Keywords: image, video, prompt, neural network, Kling, Grok.

1 Введение

1.1 Актуальность

Актуальность исследования обусловлена бурным развитием мультимодальных нейросетей, которые кардинально меняют методы создания визуального контента в рекламе, PR и образовании. В эпоху ускорения цифровых коммуникаций растёт спрос на фотореалистичные изображения и видео, адаптированные под бренд и обладающие сильным эмоциональным влиянием на аудиторию. Этот спрос значительно превышает возможности традиционной фото- и видеосъёмки, требующей больших затрат времени и ресурсов.

1.2 Обзор исследований

Е.В. Ильинская, Е.Н. Голышева, А.А. Медведев, Н.С. Масалитин описали применение генеративно-сопоставительных нейросетей для генерации изображений [1]. Способы генерации изображений при помощи искусственных нейронных сетей применили И. С. Шустов, Ю. С. Белов [2]. Е.В. Кагарлык, Ю. Ю. Балахонов создали картинки и описали обработку в новых технологиях нейронной сети [3]. Генеративно-сопоставительные сети для генерации изображений по тексту использовали А. О. Левин, Ю. С. Белов [4]. И. В. Филиппский проанализировал художественные особенности нейросетей и сгенерировал изображения и видео в контексте музейной практики [5].

1.3 Цель исследования

Цель настоящего исследования заключается в демонстрации функциональных возможностей синтеза мультимодальных нейросетевых платформ Kling и Grok, применяемых для поэтапной генерации фотореалистичных изображений и видеороликов на основе совмещения визуальных референсов и детализированных текстовых промтов.

2 Материалы и методы

В качестве основного инструментария исследования использовались мультимодальные нейросетевые платформы Kling и Grok, реализующие технологии генеративного синтеза изображений и видеороликов посредством интеграции визуальных референсов (исходных фотографий) и текстовых промтов. Такой подход обеспечивает высокий уровень точности в управлении параметрами композиции, стилистики и визуальных характеристик итогового контента.

3 Результаты

В процессе первого этапа генерации в платформу Grok необходимо загрузить текстовый промт для создания первого референсного изображения. На снимке изображена сибирская хаски в зимнем лесу средним планом, стиль кинематографичный, высокое разрешение. (рис. 1)

Промт: Realistic photo of a Siberian husky in the winter forest, middle plan, thick fluffy woolen fur, bright blue eyes, high pine forest around, tree trunks in the

snow, soft diffused winter light, light snowfall, cold bluish-white shades, cinematic style, high detail of wool and snow, traces on the snow, depth of field with blurred background, 4K, ultra-high resolution, 16:9. (Реалистичная фотография сибирского хаски в зимнем лесу, средний план, густой пушистый шерстяной мех, яркие голубые глаза, высокий сосновый лес вокруг, стволы деревьев в снегу, мягкий рассеянный зимний свет, легкий снегопад, холодные голубовато-белые оттенки, кинематографичный стиль, высокая детализация шерсти и снега, следы на снегу. снег, глубина резкости с размытым фоном, 4K, сверхвысокое разрешение, 16:9.)



Рисунок 1. Средний план

С помощью второго промта получилось создать высококачественное изображение этой же хаски общим планом. Стиль так же кинематографичный, разрешение высокое.

Промт: The same husky from afar from above, as if shooting from a quadcopter. 4K, 16:9 (Тот же хаски издалека, сверху, как будто снимали с квадрокоптера. 4K, 16:9)



Рисунок 2. Общий план

Второй этап – создание видеоролика с помощью платформы Kling. Первым делом на этом этапе в платформу следует загрузить два референсных изображения, которые ранее сгенерировала нейросеть Grok. Далее нейросеть DeepSeek, которая встроена в платформу для создания описаний видео, сгенерировала промт. Помимо основного промта в платформу можно загрузить описание звуковых эффектов, которыми должен сопровождаться видеоролик.

С помощью промта Kling сгенерировал высококачественное видео, в котором все действия главного героя и оператора естественны. Единственная ошибка, которую допустила нейросеть – собака идёт задом наперёд.

Ссылка на первое видео: <https://disk.yandex.ru/i/0s5LDBa8SExuBg>.

Промт: Follow the initial frame (close-up of a husky in a winter forest) and the final frame (general view of the same husky in a winter forest). Smooth camera zoom-out from a close-up of the husky's face to a general view of it standing/walking in the snow among snow-covered pine trees, then the camera rises and flies over the treetops of the forest, showing a snow-covered panorama with the husky in the center below. Light snowfall, the wind ruffling the husky's fur and branches, and realistic winter day lighting with soft shadows. Keep the consistency of the husky (fur color, posture, and facial expression), and the natural movement of the husky turning its head or taking a step forward. Camera: slow zoom out, then smooth drone fly-up and overhead pan, cinematic motion, high detail of the snow and forest, 24 fps, ultra-realistic, 4K. (Проследите за начальным кадром (крупный план хаски в зимнем лесу) и заключительным кадром (общий вид той же хаски в зимнем лесу). Плавное уменьшение масштаба изображения с крупного плана лица хаски до общего вида, когда он стоит / прогуливается по снегу среди заснеженных сосен, затем камера поднимается и пролетает над верхушками деревьев в лесу,

показывая заснеженную панораму с хаски в центре внизу. Легкий снегопад, ветер, взъерошивающий шерсть хаски и ветки, реалистичное освещение зимнего дня с мягкими тенями. Сохраняйте характер хаски (цвет шерсти, позу и выражение лица), а также естественные движения хаски, когда она поворачивает голову или делает шаг вперед. Камера: медленное уменьшение масштаба, затем плавный взлет дрона и панорамирование над головой, кинематографичное движение, высокая детализация снега и леса, 24 кадра в секунду, ультрареалистичная съемка в формате 4K.)

Для исправления ошибки, допущенной нейросетью, первоначальный промпт необходимо отредактировать. Появился акцент на том, что собака должна идти вперёд. Данная попытка так же увенчалась успехом – собака всё так же идёт задом наперёд.

Ссылка на второе видео: https://disk.yandex.ru/i/SqLSbyLZ_tutGQ.

Промт: Observe the initial frame (close-up of a husky in a winter forest) and the final frame (general view of the same husky in a winter forest). Smooth camera zoom-out from a close-up of the husky's face to a general view of it walking through the snow among snow-covered pine trees, then the camera rises and flies over the treetops of the forest, revealing a snow-covered panorama with the husky in the center below. Light snowfall, the wind ruffling the husky's fur and branches, and realistic winter day lighting with soft shadows. Keep the consistency of the husky (fur color, posture, and facial expression), and the natural movement of the husky turning its head or taking a step forward. Camera: slow zoom out, then smooth drone fly-up and overhead pan, cinematic motion, high detail of the snow and forest, 24 fps, ultra-realistic, 4K. (Обратите внимание на начальный кадр (крупный план хаски в зимнем лесу) и заключительный кадр (общий вид той же хаски в зимнем лесу). Плавное уменьшение масштаба изображения с крупного плана лица хаски до общего вида, когда он идет вперёд по снегу среди заснеженных сосен, затем камера поднимается и пролетает над верхушками деревьев в лесу, открывая заснеженную панораму с хаски в центре. Легкий снегопад, ветер, взъерошивающий шерсть хаски и ветки, реалистичное освещение зимнего дня с мягкими тенями. Сохраняйте характер хаски (цвет шерсти, позу и выражение лица), а также естественные движения хаски, когда она поворачивает голову или делает шаг вперед. Камера: медленное уменьшение масштаба, затем плавный взлет дрона и панорамирование над головой, кинематографичное движение, высокая детализация снега и леса, 24 кадра в секунду, ультрареалистичная съемка в формате 4K.)

4 Вывод

В результате процесса генерации аудиовизуального контента на платформах Kling и Grok, основанного на текстовых описаниях и референсных снимков, получилось два видеоролика. Grok оправдал ожидания и создал качественные изображения даже несмотря на то, что текстовые описания оказались не подробными. Kling на основе текстовых описаний и референсных изображений сгенерировал два качественных видеоролика, которые не до конца совпадают с ожидаемым результатом. Это говорит о том, что платформу

можно успешно использовать для создания простых видео, а также оставляет поле для дальнейшего исследования нейросети и экспериментов с ней.

Библиографический список

1. Ильинская Е.В., Голышева Е.Н., Медведев А.А., Масалитин Н.С. Применение генеративно-сопоставительных нейросетей для генерации изображений // Научный результат. Информационные технологии. 2024. №1. С. 73-78. URL: <https://cyberleninka.ru/article/n/primenenie-generativno-sostyazatelnyh-neyrosetey-dlya-generatsii-izobrazheniy>.
2. Шустов И.С., Белов Ю.С. Способы генерации изображений при помощи искусственных нейронных сетей // Наукоемкие технологии в приборостроении и развитии инновационной деятельности в вузе: материалы Всероссийской научно-технической конференции, 15–17 ноября 2016 г. Т. 2. Калуга: Издательство МГТУ им. Н. Э. Баумана, 2016. URL: https://conference.kf.bmstu.ru/uploads/userfiles/december_2016_2.pdf#page=198.
3. Кагарлык Е. В., Балахонов Ю. Ю. Создание картинки и обработка её в новых технологиях нейронной сети // Современные информационные технологии в образовании, науке и промышленности. Искусственный интеллект в создании картин: Сборник трудов XVIII Международной конференции и XVI Международного конкурса научных и научно-методических работ, Москва, 12–16 февраля 2021 года. М.: Общество с ограниченной ответственностью "Издательство "Экон-Информ", 2021. С. 198-201. URL: <https://elibrary.ru/item.asp?id=45709355>.
4. Левин А. О., Белов Ю. С. Использование генеративно-сопоставительных сетей для генерации изображений по тексту // Научное обозрение - технические науки. 2023. №2. С. 10. URL: https://s.science-engineering.ru/pdf/2023/2023_2.pdf#page=10.
5. Филиппский И. В. Художественные особенности нейросетей: генерация изображений и видео в контексте музейной практики // Управление культурой. 2024. Т. 3. №. 4. С. 44-51. URL: <https://cyberleninka.ru/article/n/hudozhestvennye-osobennosti-neyrosetey-generatsiya-izobrazheniy-i-video-v-kontekste-muzeynoy-praktiki>.