

Обзор Interplay Mode

Google Interplay mode предназначен для внедрения интерактивных элементов в обучающие видео.

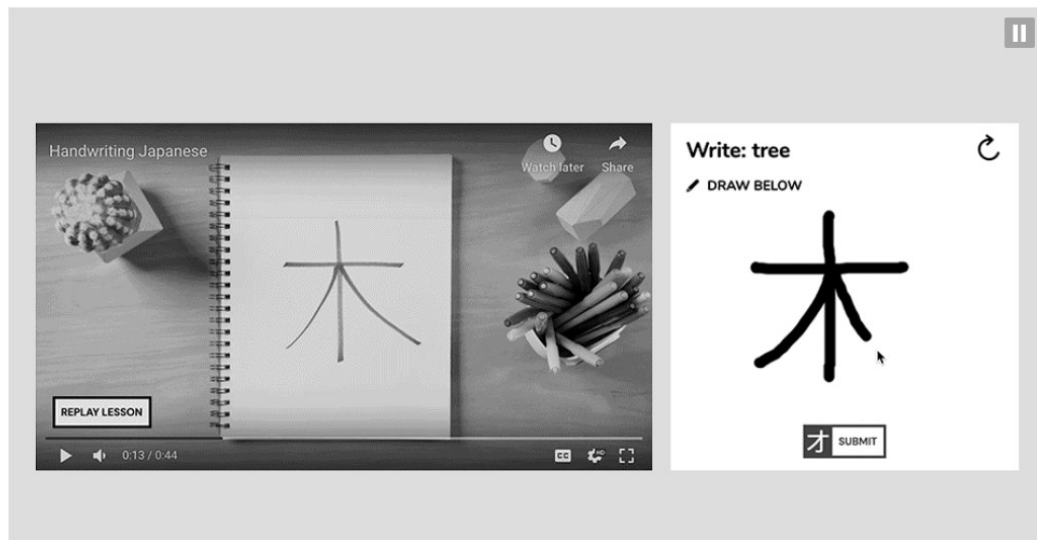


Рисунок 1: Пример использования ИИ в обучающем видео

В приведённом выше примере в видео для обучения иероглифам японского языка используется ИИ который распознает ручной ввод пользователя. В определённый момент видео останавливается и просит пользователя повторить иероглиф который нарисовал инструктор.

В случае если пользователь ввёл правильный иероглиф урок продолжается. В противном случае пользователя просят ввести данные заново.

Для распознавания текста используется Google Handwriting Recognition для распознавания текста и Teachable Machine для обучения ИИ.

Принцип работы распознавания текста

Для распознавания текста в Google используется принцип Кривых Безье. Рассмотрим на примере:

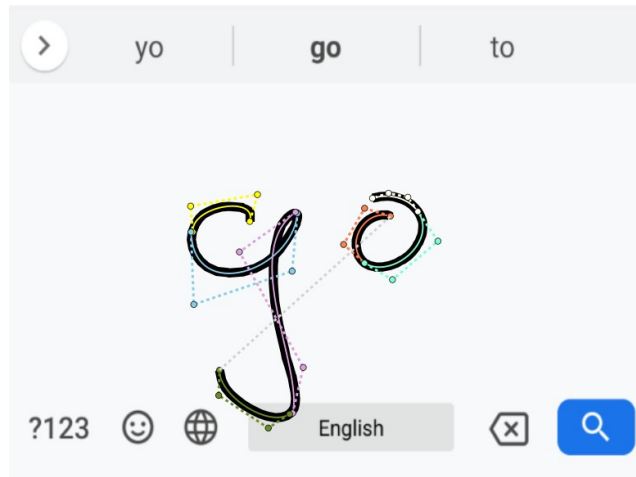


Рисунок 2: Пример входных данных

Введённое пользователем слово разбивается на Touch points или «Точки прикосновения», которые затем преобразуются в кривые Безье. При этом кривые хранят следующие данные:

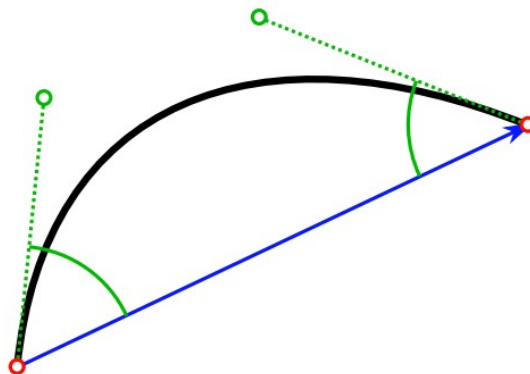


Рисунок 3: Пример кривой Безье

- Вектор между конечными точками (красные точки). (Голубая линия на рисунке 3)
- Расстояния между контрольными точками (зеленые точки) и конечными точками. Это зеленые пунктирные линии на рисунке.
- Углы между каждой контрольной точкой и конечной точкой. Это зеленые дуги на рисунке.
- Три временных коэффициента (не показаны на рисунке)
- Булевый индикатор, отображающий состояния pen-up или pen-down

Дальше полученные данные проходят через нейросеть QRNN-based recognizer, которая преобразует результат в матрицу в которой каждая строка матрицы — буква алфавита, а каждая колонка связана с кривой полученной ранее. В самой же ячейке содержится вероятность совпадения с буквой.

Затем полученная матрица пропускается через CTC Decoder, который преобразует полученную матрицу в строку которая может быть использована в дальнейшем.

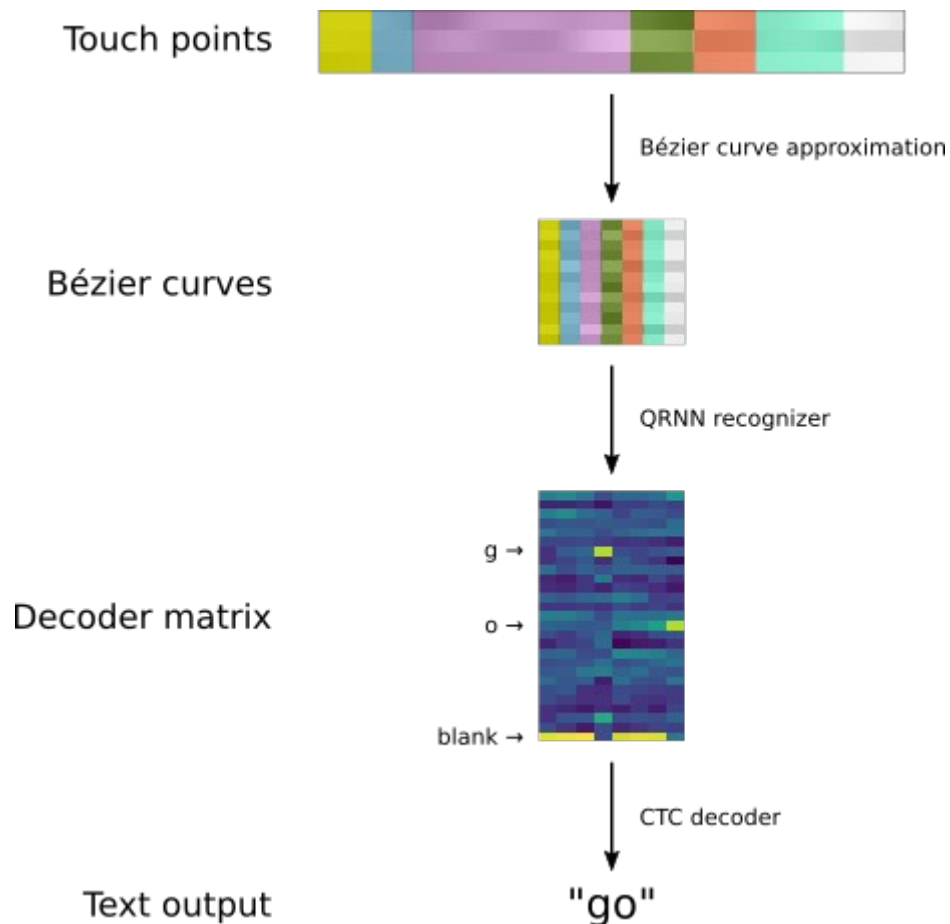
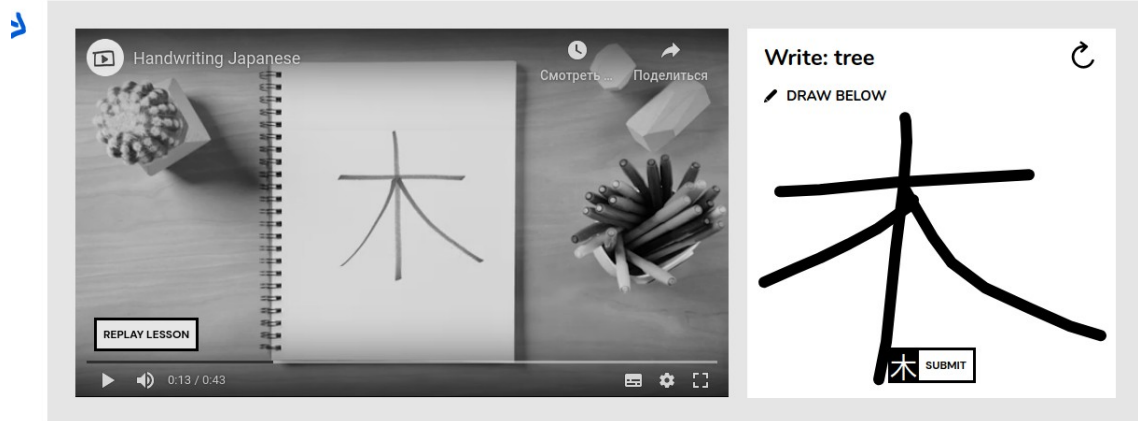


Рисунок 4: Общий алгоритм обработки рукописного ввода

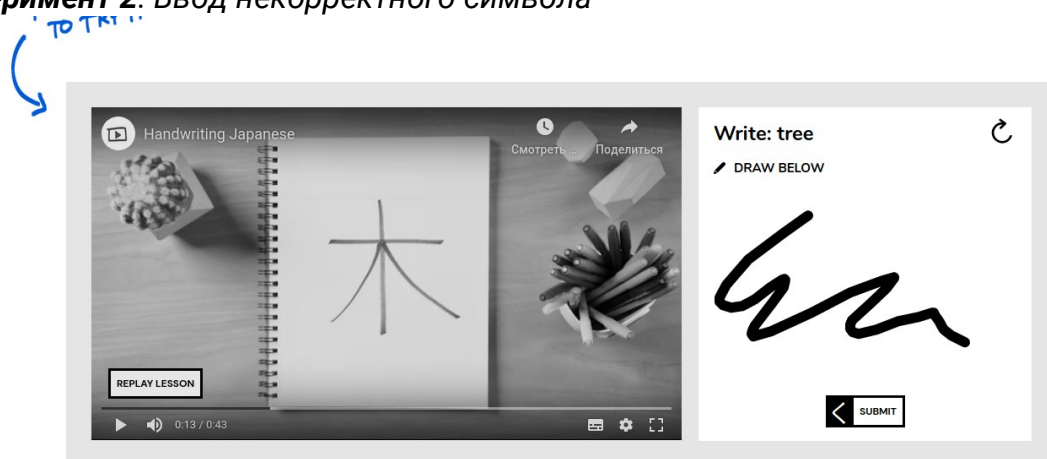
Проведённые эксперименты

Эксперимент 1: Ввод корректного символа



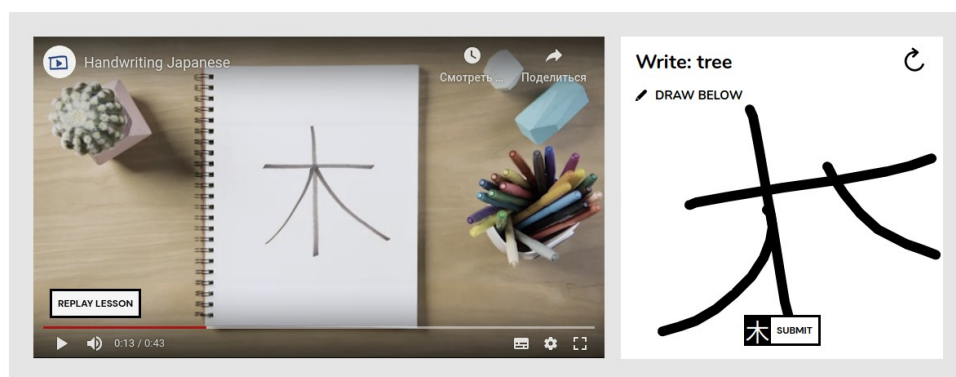
Результат: Система распознала введённый символ

Эксперимент 2: Ввод некорректного символа



Результат: Система распознала введённый символ, но отклонила его так как он не совпадает с необходимым

Эксперимент 3: Ввод корректного символа с искажением



Результат: Символ был распознан и принят несмотря на искажение

Вывод

В ходе работы был рассмотрен принцип работы распознавания текста в Interplay Mode приложении, которое позволяет создавать обучающие видео.

Изученный пример обладает продвинутым ИИ который может распознавать текст, однако у него тоже есть «ложные срабатывания», которые показывают что система все еще нуждается в доработке.