УДК 004.932.

Сравнительный анализ моделей компьютерного зрения для обработки изображений в реальном времени

Сосновский М.А.

Саратовский национальный
исследовательский государственный
университет имени Н.Г. Чернышевского
г. Саратов, Российская Федерация

1. Описание задачи

Целью работы было провести сравнительный анализ моделей компьютерного зрения на наборе изображений скриншотов игры в условиях работы в реальном времени.

2. Сбор и предобработка данных

Для достижения цели работы были собраны и предобработаны изображения всех 112 персонажей игры Genshin Impact, доступных на момент версии 6.4.

Производилась съёмка видео продолжительностью 5–8 минут, на котором персонаж произвольно меняет позы, а также вручную меняются ракурсы съёмки и приближение камеры. Из каждого такого видео выбирались наиболее уникальные положения персонажа в кадре. Таким образом извлечено 5 965 кадров, в среднем около 50–60 на каждого персонажа.

Затем каждый кадр обрезался до квадрата и создавалась его копия, на которой цветной фон заменялся нейтральным серым с помощью пользовательской Python-библиотеки или вручную в стандартном приложении «Фотографии» операционной системы Windows. В итоге, было собрано 14 000 изображений, в среднем около 125 на каждого персонажа.

3. Сравнительный анализ моделей

Сначала были отобраны 7 архитектур свёрточных нейронных сетей (VGGNet, Inception, ResNet, MobileNet, EfficientNet, ResNeXt, ConvNeXt) [1, 2] для первичного анализа необходимых вычислительных затрат для достижения

сравнительно хорошего качества на небольшой выборке из всего набора данных. Затем с тремя наиболее подходящими архитектурами (ResNet, MobileNet, EfficientNet) по соотношению длительности обучения и качества классификации [3, 4] проводились дальнейшие эксперименты на всём наборе данных. С этого момента все модели инициализировались с весами, предобученными на наборе данных ImageNet-1K. В итоге, по результатам экспериментов был сделан вывод, что среди рассмотренных модель с архитектурой EfficientNet-B0 наилучшим образом подходит для классификации персонажей на изображениях скриншотов игры в условиях работы в реальном времени.

4. Обучение и тестирование лучшей модели

Обучение лучшей модели проходило на 3, 9, 21, 69, 106 и 112 классах. Каждое состояние модели тестировалось на «сырых» скриншотах игры (по 10 на каждого персонажа). После подбора параметров предобработки изображений и аугментаций, скорости обучения и количества эпох до и после разморозки итоговая модель показала Accuracy = 1,0 на 1120 тестовых примерах. Также, проводились тесты на более сложных примерах с неоднородным фоном, нестандартным освещением и прочими дополнительными условиями. На таких изображениях уверенность предсказания истинного класса стала меньше, а в особо сложных случаях стали возникать ошибки.

5. Итоги, известные проблемы и будущее развитие

В итоге, был собран свой набор данных, состоящий из 14 000 изображений и 112 классов, и на нём дообучена наиболее эффективная для данной задачи модель с легковесной архитектурой EfficientNet-B0. Достигнута стопроцентная точность предсказаний на тестовых примерах и сравнительно высокая точность на более сложных примерах.

Проблемы и их будущее решение:

- На изображениях с неоднородным фоном или нестандартным освещением может значительно повышаться риск ошибочной классификации. Планиру-

ется добавить в набор данных более разнообразные изображения в контексте фона и освещения, а также улучшить качество аугментаций.

- Нет обработки out-of-distribution. Планируется собрать достаточно разнообразный набор изображений, не содержащих персонажей, и обучить дополнительную модель на них и всех уже имеющихся в наборе данных изображениях, которая будет различать персонажей и неперсонажей.

Список использованной литературы

1. Борзов С.М. *Нейросетевые технологии в задачах обнаружения и классификации объектов* / С.М. Борзов, Е.С. Нежевенко // Автометрия. – 2023. – Т. 59, № 3. – С. 52–71.
2. Younesi A. *A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends* / A. Younesi, M. Ansari, M. Fazli et al // IEEE Access. – 2024. – Vol. 12. – P. 41180–41218.
3. Losinski J. *Comparison of CNN architectures for INTERCO* / J. Losinski, I. Czarnowski // Procedia Computer Science. – 2025. – Vol. 270. – P. 861–870.
4. Zapalski M.K. *Comparison of chosen image classification methods on Android* / M.K. Zapalski, P.K. Zabczynski, P. Powroznik // Journal of Computer Sciences Institute. – 2025. – Vol. 36. – P. 342–349.