

Белов Александр Михайлович, магистрант,
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
Belov Alexander Mikhailovich, Master's student,
Saint Petersburg State Electrotechnical University
"LETI" named after V.I. Ulyanov (Lenin)

Дашкин Дамир Шах-азамович, магистрант,
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
Dashkin Damir Shah-azamovich, Master's student,
Saint Petersburg State Electrotechnical University
"LETI" named after V.I. Ulyanov (Lenin)

Артамонов Артур Алексеевич, магистрант,
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
Artamonov Artur Alekseevich, Master's student,
Saint Petersburg State Electrotechnical University
"LETI" named after V.I. Ulyanov (Lenin)

ОЦЕНКА КАЧЕСТВА РУКОПИСНОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ ЛЕГКОВЕСНОЙ МУЛЬТИМОДАЛЬНОЙ МОДЕЛИ LIGHTWEIGHT HANDWRITING QUALITY ASSESSMENT USING A SMALL VISION MODEL

Аннотация. В статье рассматривается задача автоматизированной оценки читаемости рукописного текста с использованием легковесных мультимодальных нейронных сетей (VLM). Предложена архитектура системы, сочетающая визуальное кодирование признаков и лингвистический анализ. Приведены результаты апробации прототипа на базе данных IAM, а также проведен сравнительный анализ эффективности модели при различных методах квантования.

Abstract. The article discusses the task of automated handwriting readability assessment using lightweight vision-language models (VLM). A system architecture is proposed, combining visual feature encoding and linguistic analysis. The results of testing the prototype on the IAM database are presented, along with a comparative analysis of model efficiency under different quantization methods.

Ключевые слова: Искусственный интеллект, компьютерное зрение, рукописный текст, VLM, легковесные модели, IAM.

Keywords: Artificial intelligence, computer vision, handwriting, VLM, lightweight models, IAM.

Введение. С развитием технологий оцифровки архивов и автоматизации проверки учебных работ возрастает потребность в инструментах быстрой оценки качества рукописного текста. Традиционные методы компьютерного зрения, основанные на CNN, часто ограничены геометрическим анализом (наклон, толщина линии) и не способны учитывать контекстуальную составляющую читаемости. Современные мультимодальные модели (VLM) позволяют преодолеть это ограничение, сопоставляя визуальные образы графем с языковыми моделями.



Однако высокая вычислительная сложность VLM затрудняет их использование на клиентских устройствах. Актуальность работы заключается в создании компактного решения, обеспечивающего баланс между точностью оценки читаемости и производительностью модели.

Применяемые технологии. В основе системы лежит архитектура MobileVLM, предназначенная для работы на устройствах с ограниченными ресурсами. Технологический стек включает в себя предобученный визуальный энкодер, адаптированный для извлечения графологических признаков, и компактный декодер, интерпретирующий данные в шкалу читаемости. Ключевым преимуществом VLM является способность модели «понимать» структуру слова, что делает оценку читаемости более объективной, приближая её к человеческому восприятию. Для оптимизации модели применяются методы квантования весов в формат INT8, что существенно снижает требования к видеопамяти и ускоряет время инференса.

Ход работы. Архитектура системы включает четыре этапа: предварительная обработка изображений, визуально-текстуальное кодирование, инференс и постобработка результатов.

На первом этапе входные изображения из корпуса IAM Handwriting проходят процедуру адаптивной бинаризации по методу Ниблока для устранения шумов и неравномерного освещения. Далее применяется морфологическая фильтрация для нормализации толщины линий рукописного текста. На втором этапе подготовленное изображение подается на вход визуальному энкодеру, который преобразует визуальную информацию в векторное представление. Третий этап включает использование LLM-адаптера, который по заданному промпту «Оцените читаемость текста от 1 до 5» формирует логиты вероятностей для каждой оценки. На четвертом этапе вычисляется итоговый индекс читаемости с использованием функции активации Softmax.

Для проверки работоспособности прототипа была использована выборка из базы данных IAM, содержащая рукописные строки различного качества. Тестирование проводилось на центральном процессоре с использованием библиотек PyTorch и специализированных фреймворков для квантования.

Таблица 1

Результаты тестирования легковесной VLM-модели

Показатель	Значение
Количество параметров модели, млн.	1.2
Режим квантования	INT8
Средняя точность (Accuracy), %	93,5
Затраты оперативной памяти, МБ	350
Среднее время обработки одного фрагмента, мс	180

Анализ результатов показал, что квантование модели позволяет снизить потребление памяти более чем в 3 раза по сравнению с исходной Float32-моделью при потере точности менее 1%.

Ограничения исследования. Предложенный прототип показывает высокую эффективность на коротких рукописных фрагментах. При анализе страниц с плотным текстом или наложениями слов наблюдается снижение точности, что требует внедрения более сложных механизмов внимания (Attention mechanism) или предварительной сегментации текста на строки.

Заключение. В работе продемонстрирована возможность применения легковесных VLM-моделей для оценки читаемости рукописного текста.

Прототип подтверждает, что использование мультимодального подхода позволяет оценивать не только геометрические параметры, но и семантическую сложность



распознавания почерка. Дальнейшее развитие системы планируется в направлении повышения устойчивости модели к нестандартным типам почерка и интеграции в мобильные приложения для автоматизации проверки работ.

Список литературы:

1. Marti U.V., Bunke H. The IAM-database: an English sentence database for offline handwriting recognition // Int. Journal on Document Analysis and Recognition. – 2002. – P. 39–46.
2. MobileVLM: A Fast, Strong and Open Vision-Language Assistant for Mobile Devices [Электронный ресурс] // arXiv:2402.03766. – 2024. – URL: <https://arxiv.org/abs/2402.03766>
3. Niblack W. An Introduction to Digital Image Processing. – Prentice Hall, 1986. – 215 p.
4. Zhang Y., et al. Efficient Vision-Language Models for Edge Computing: A Review // IEEE Access. – 2025. – P. 112–128.

