УДК 004.352.242 https://doi.org/10.26160/2541-8637-2026-17-75-78

ПОВЫШЕНИЕ КАЧЕСТВА РАСПОЗНАВАНИЯ МЕТРОЛОГИЧЕСКОЙ ДОКУМЕНТАЦИИ ДЛЯ УЗЛОВ УЧЕТА НЕФТИ И ГАЗА ПУТЕМ ДООБУЧЕНИЯ ВИЗУАЛЬНО-ЯЗЫКОВОЙ МОДЕЛИ

Галиакберов И.В., Хасанова А.М.
Уфимский государственный нефтяной технический университет, Уфа

Ключевые слова: визуально-языковые модели, дообучение, метрологическая документация, нефтегазовая отрасль, распознавание документов, оптическое распознавание символов.
Аннотация. Рассматривается задача улучшения распознавания метрологической документации с помощью дообучения визуально-языковой модели LightOnOCR-2-1B. Применен метод LoRA, позволяющий адаптировать модель в условиях ограниченных ресурсов. Экспериментально подтверждено снижение CER и WER. Полученные результаты создают основу для автоматизированного анализа формул и таблиц в нефтегазовой отрасли.

IMPROVING THE QUALITY OF METROLOGICAL DOCUMENTATION RECOGNITION FOR OIL AND GAS METERING STATION THROUGH FINETUNNING OF A VISUAL-LANGUAGE MODEL

Galiakberov I.V., Khasanova A.M.
Ufa State Petroleum Technological University, Ufa

Keywords: visual-language models, fine-tuning, metrological documentation, oil and gas industry, document recognition, and optical character recognition.
Abstract. The task of improving the recognition of metrological documentation is considered using the LightOnOCR-2-1B visual-language model. The LoRA method is applied to adapt the model in a resource-constrained environment. The reduction of CER and WER is experimentally confirmed. The obtained results create a basis for automated analysis of formulas and tables in the oil and gas industry.

Введение

Исторически для целей оптического распознавания символов (Optical Character Recognition, OCR) использовалось шаблонное сопоставление. Процесс был поэтапным:

– предобработка изображения, что включало в себя повышение контраста, удаление шумов, выравнивание, преобразование в черно-белый формат;

– сегментация, где изображение разбивалось на отдельные символы;

– сопоставление с шаблонами, что представляло из себя сравнение изображения символа с библиотекой шаблонов (эталонных шрифтов).

Такие системы имели ряд недостатков: чувствительность к качеству изображения, шрифту, наклону; плохая работа с рукописным текстом; сложность разделения «слипшихся» символов.

С развитием вычислительных мощностей и искусственного интеллекта, начали появляться системы интеллектуального распознавания символов (Intelligent Character Recognition, ICR). Для их разработки применялось глубокое обучение. Широкое применение нашли сверточные (Convolutional Neural Networks, CNN) и рекуррентные нейронные сети (Recurrent Neural Network, RNN), что дало существенное повышение точности и производительности распознавания [1].

Передовой разработкой глубокого обучения являются «Трансформеры», представленные в работе «Attention is All You Need» группой ученных из Google в 2017 году [2]. Данная архитектура, основанная на механизме внимания, легла в основу современных языковых моделей, таких как GPT (Generative Pre-trained Transformer), BERT (Bidirectional Encoder Representations from Transformers), T5 и многих других [3].

Также широкое применение трансформеры нашли в задачах компьютерного зрения, объединившись с языковыми моделями, в результате чего появились визуально-языковые модели (Visual Language Models, VLM). Они используют мультимодальные данные, которые предварительно проходят предобработку в токенизаторе – модуле, преобразующем сырую информацию в идентификационные входы для модели. Такой принцип работы позволяет системе понимать контекст, с которым традиционные OCR не справляются, что делает их наиболее эффективными для любого типа документации, в том числе метрологической [1].

В открытом ресурсе «Hugging Face» [4] имеется большое количество свободно распространяемых моделей, которые хорошо справляются с распознаванием документов. Однако для раскрытия их полного потенциала, в условиях специфики конкретной предметной области, требуется их дообучение (Finetuning). Универсальные модели, обученные на общих наборах данных (книги, новости, веб-страницы), могут демонстрировать высокую точность на типовых текстах, но их эффективность существенно снижается при работе с узкоспециализированной документацией, где имеется множество профессиональной терминологии, специфические формулы, таблицы и аббревиатуры.

Основная часть

В данной работе будет рассматриваться задача улучшения распознавания метрологической документации визуально-языковой моделью для узлов учета нефти и газа. Данный этап является ключевым для последующих алгоритмов, например экспертных систем или верификации расчетов. Актуальность этой задачи продиктована высокой значимостью коммерческого учета углеводородов, где ошибки в показаниях или паспортных данных приборов могут приводить к серьезным финансовым и репутационным потерям.

Для решения этой проблемы предлагается использовать легковесную визуально-языковую модель LightOnOCR-2-1B-base [5], которая относится к классу энкодер-декодер моделей, оптимизированных для распознавания текста на изображения. Выбор данной модели обусловлен ее архитектурными особенностями и оптимальным соотношением производительности и вычислительной сложности.

Базовые веса модели LightOnOCR-2-1B-base содержат общее понимание текста и изображений. Для адаптации под нашу задачу планируется провести этап пилотного дообучения на специализированном наборе данных. Этот набор будет включать аннотированные изображения реальных документов, характерных для нефтегазовой метрологии: МИ 3275-2010, МИ 3532-2015.

Для формирования обучающей выборки было отобрано 46 страниц упомянутых методик измерений. Аннотация каждого документа будет представлять собой структурированный текст в формате Markdown, с формулами LaTeX и таблицами HTML, где каждой странице ставится в соответствие его значение, извлеченное из изображения (скан DPI 200). Процесс дообучения будет направлен на минимизацию ошибки модели при генерации именно такого структурированного вывода.

Стендовым оборудованием для дообучения служит персональный компьютер со следующими характеристиками:

– операционная система Linux - Ubuntu 24.04;

– центральный процессор Intel Core i5-10400F;

– оперативная память 16 ГБ с частотой 3200 МГц и 32 ГБ файл подкачки;

– графический процессор NVIDIA RTX 3050 с CUDA 12.9, который позволяет ускорить работу моделей за счет параллельных вычислений;

– среда исполнения Python 3.12 с фреймворком для машинного обучения PyTorch.

Тренировочные аргументы для класса Trainer [6] приведены на рисунке 1. Дообучение проводилось с использованием метода LoRA (Low-Rank Adaptation). Этот подход позволяет адаптировать большие предобученные модели, замораживая исходные веса и добавляя небольшое количество обучаемых параметров.

```
output_dir = "./LightOnOCR-2-lora-ft-3"
training_args = TrainingArguments(
    output_dir=output_dir,
    num_train_epochs=100,
    per_device_train_batch_size=1,
    per_device_eval_batch_size=1,
    gradient_accumulation_steps=4,
    learning_rate=2e-4,
    weight_decay=0.0,
    logging_steps=10,
    eval_strategy="steps",
    eval_steps=50,
    save_strategy="steps",
    save_steps=500,
    save_total_limit=2,
    load_best_model_at_end=True,
    metric_for_best_model="eval_loss",
    bf16=torch.cuda.is_available(),
    fp16=False,
    remove_unused_columns=False,
    dataloader_pin_memory=False,
    gradient_checkpointing=True,
    optim="adamw_torch_fused",
    warmup_steps=10,
    lr_scheduler_type="linear",
    dataloader_num_workers=0,
    report_to="none",
    logging_dir=f"{output_dir}/logs",
)
```

Рис. 1. Тренировочные аргументы

Полное качество распознавания текста оценивается с помощью метрик CER (символьная ошибка) и WER (словесная ошибка). Полученные результаты дообучения приведены в таблице 1.

Хотя абсолютное улучшение может показаться незначительным, в контексте обработки метрологической документации даже такое снижение ошибки критически важно. Ошибки в распознавании одного символа могут привести к

неверной интерпретации числовых значений или искажению индексов в формулах, что в дальнейшем повлияет на корректность расчётов.

Табл. 1. Результаты дообучения

Метрика	Базовая модель	Дообученная модель	Прирост точности
CER, %	3.30	2.54	0,76
WER, %	6.01	5.49	0,52

Заключение

Полученная модель способна более точно распознавать сложноструктурированные элементы документов – формулы (в формате LaTeX) и таблицы (в формате HTML), что является важным шагом к созданию полноценной системы автоматизированного анализа нормативно-технической документации. В перспективе планируется расширить датасет за счёт других ГОСТов и методик.

Список литературы

1. Захаров Р.К., Кулагин В.П. Анализ подходов к использованию визуальных языковых моделей в задачах оптического распознавания символов // Информатизация образования и науки. – 2025. – №2(66). – С. 71-77.
2. Ashish V. Attention is all you need // 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
3. Жусип М.Н., Жаксыбаев Д.О. Сравнение чат-ботов с использованием трансформеров и нейросетей: исследование применения архитектур GPT и BERT // Вестник науки. – 2024. – №9(78). – С. 287-290.
4. Hugging Face. [Электронный ресурс] – Режим доступа: https://huggingface.co/
5. LightOnOCR-1B-1025. LightOn AI. [Электронный ресурс]. – Режим доступа: https://huggingface.co/lightonai/LightOnOCR-1B-1025.
6. Trainer. Hugging Face. [Электронный ресурс]. – Режим доступа: https://huggingface.co/docs/transformers/main_classes/trainer.

Сведения об авторах:
Галиакберов Ильназ Венерович – студент;
Хасанова Айгуль Марсовна – доцент.