

0.1. Шакин В.Ю., Сухов Ф.В. Разработка модуля идентификации для системы роботизированной маркировки образцов керна

Ручная маркировка образцов керна занимает длительное время, из-за чего существенно затягивается процесс лабораторных исследований, поэтому внедрение автоматизированной системы маркировки очень актуально для крупных нефтегазовых компаний. Основной целью представленной работы является разработка модуля идентификации текста на образце керна для системы роботизированной маркировки, над созданием которой работает новосибирская компания АО «Геологика». Модуль идентификации текста необходим для проверки качества маркировки, произведённой системой. Он представляет собой клиент-серверное приложение, написанное на языке Python. В разработке используется следующий стек технологий:

- Python 3.9 – язык программирования;
- Roboflow – приложение для разметки данных и создания выборок, использующихся для обучения нейронной сети [1];
- YOLOv5 – современный алгоритм глубокого обучения, который широко используется для обнаружения объектов [2];
- Tesseract-OCR – библиотека для распознавания текста, написанная на языке C++ [3];
- Docker – платформа для разработки, доставки и запуска контейнерных приложений [4].

Процесс идентификации текста был разделён на два этапа: на начальном этапе фото (скан) образца обрабатывается с помощью нейронной сети, которая определяет участки, на которых содержится текст и вырезает их с основного фото, на основном этапе происходит распознавание текста на каждом из кусочков с помощью Tesseract-OCR. Применение комбинации нейронной сети и Tesseract-OCR позволило существенно повысить точность распознавания текста на образце керна.

Работа клиент-серверного приложения происходит следующим образом: оператор вводит номер фото (скана) образца, после чего происходит отправка данных на сервер в формате «json». Сервер при помощи нейронной сети и Tesseract-OCR определяет текст на образце керна и отправляет оператору json-файл с результатами работы: номером запроса, текстом. С помощью Docker был создан контейнер, содержащий все необходимые зависимости для корректной работы приложения на ПК стороннего пользователя. На данный момент происходит интеграция модуля идентификации в разрабатываемую систему.

Научный руководитель — к.т.н. Ланграф С. В.

Список литературы

- [1] Roboflow documentation. [Электронный ресурс]. URL: <https://github.com/opencv-ai/roboflow> (дата обращения 01.07.2022).
- [2] YOLOv5 documentation. [Электронный ресурс]. URL: <https://docs.ultralytics.com> (дата обращения 05.07.2022).
- [3] Tesseract-OCR documentation [Электронный ресурс]. URL: <https://github.com/tesseract-ocr/tesseract> (дата обращения 06.07.2022).
- [4] Docker, official site [Электронный ресурс]. URL: <https://www.docker.com> (дата обращения 08.07.2022).