

Программный продукт «Helmetnet»

Инструкция по установке



Инструкция по установке

Программный продукт «Helmetnet» поставляется в виде готового Docker-образа, содержащего инференс-сервис, и комплекта файлов для развёртывания сервера инференса (Triton Inference Server) с нейросетевой моделью. Установка выполняется в инфраструктуре заказчика администратором, обладающим базовыми навыками работы с ОС Linux и Docker.

Требования к среде установки: сервер под управлением ОС Linux (дистрибутив не регламентируется) с установленными Docker и Docker Compose; для ускорения вычислений может использоваться GPU NVIDIA с установленным NVIDIA Container Toolkit (опционально, поддерживается работа на CPU); не менее 8 ГБ оперативной памяти и 20 ГБ свободного дискового пространства.

Порядок установки:

1. Загрузите поставляемый Docker-образ на сервер и импортируйте его командой `docker load`, либо соберите образ из комплекта поставки командой `docker build`.
2. Создайте файл конфигурации `.env` на основе входящего в комплект поставки примера (`.env.example`) и укажите параметры: порт сервиса (`APP_PORT`, по умолчанию 5656), устройство вычислений (`DEVICE=cuda` или `cpu`), адрес сервера инференса (`TRITON_URL`), пути к файлам лицензии.
3. Разместите файлы нейросетевой модели в каталоге репозитория моделей Triton (`triton_repo`) в соответствии с комплектом поставки.
4. Запустите программный продукт командой `docker-compose -f docker-compose.yml -f docker-compose-triton.yml up -d`. Будут запущены два контейнера: инференс-сервис «Helmetnet» и сервер инференса Triton Inference Server.
5. Активируйте лицензию: откройте в веб-браузере страницу `http://<адрес_сервера>:5656/license/ui` и введите лицензионный ключ из комплекта поставки, либо выполните активацию через API-метод `POST /license/activate`.
6. Проверьте работоспособность: запрос `GET http://<адрес_сервера>:5656/health` должен вернуть ответ `{"status": "ok"}` (HTTP 200), запрос `GET /status` — служебную информацию о версии сервиса и статусе подключения к серверу инференса.

После успешной установки и активации лицензии программный продукт «Helmetnet» готов к работе согласно Инструкции по эксплуатации.