

Метод хранения данных в CVAT'е и создание backup'ов задач и проектов

В CVAT'е существуют при создании задачи данные могут быть взяты из следующих источников:

- ☒ через браузер
- ☒ через **share** директорию
- ☐ **Remote Sources** (url к данным - не тестировался)
- ☐ **Cloud Storage** (диск в облаке - не тестировался)

Введение

CVAT при создании задачи разбивает данные на *чанки* (Chunk). По умолчанию, сам разметчик определяет размер чанка, если этот параметр не указан явно (**Chunk size**). Во время разметки подгружается нужный *чанк* (который соответствует открытому *frame'у*) и после этого начинается работа с этим *чанком*.

Если нужно будет более детальное описание работы с **чанками** следует об этом сообщить.

На метод хранения данных влияет параметр **Use cache**.

Используя **share** директорию

Если **Use cache** не выбран

- создание задачи займёт какое-то время для конвертирования изображений/ работы с видео
- при создании задачи все данные сохраняются в *volume data* в директорию **data/original** и **data/compressed** и таким образом задача не привязан к исходным данным (их можно удалять из **share** директории)
- в **data/raw** ничего не будет храниться - из-за этого для задачи нельзя выполнить команду **Backup Task / Backup/Project**
- при работе с задачей *чанки* подгружаются из **data** директории (Важно! *id* директории не связан с *id* директории)

Если **Use cache** выбран

- создание задачи проходит быстро - данные не обрабатываются, а только строится их список. Как правило, это занимает несколько секунд
- данные в задаче привязаны к исходным данным (их нельзя удалять из **share** директории)
- в директории **data/raw** хранится список файлов и их разбиение на *чанки*, но сами *чанки* не создаются
- *чанки* создаются в момент обращения к *frame'у* - в некоторых случаях этот процесс проходит быстро и незаметно для пользователя, в некоторых - требует времени и загрузки *frame* придется ждать

- созданный *чанк* хранится в *volume data*. К сожалению, точный путь директории не удалось определить. НО! Если все *чанки* создаются только один раз и если после этого удалить исходные данные из директории **share**, доступ к данным таска не будет утрачен, т.к. *чанки* уже существует. Однако, при этом будет утрачена опция создания *backup*'а + через какое-то время разметчик выдает ошибки об отсутствии данных. Дальнейшее поведение таска (после удаления исходных данных) считаю непредсказуемым
- возможно выполнение команд **Backup Task / Backup/Project**, при условии что данные не удалены/перемещены из директории **share**

Загрузка через браузер

В этом случае вне зависимости от параметра **Use cache** данные будут сохранены в **data/raw**. И *backup* возможен

Примечания

1. В директории **data/tasks/*/export_cache/** хранятся закешированные дампы (разметок, бекапов, прочего). Разметчик сам чистит этот кэш раз в какой-то период
2. Разметчик периодически сам очищает ненужную информацию (данные в **data**, записи в **postgres_db**)