

Decision Worksheet по предобработке CV

Ответьте на восемь чисел ниже до выбора модели. Одна страница; один проект; одна печать.

1. Вход камеры — что приходит по проводу

- ☐ Разрешение и кодек (например, 1080p H.265 или 720p WebRTC VP8).
- ☐ Частота кадров на источнике (обычно 15, 24, 30 или 60 fps).
- ☐ Цветовое пространство и диапазон — BT.709 studio, BT.2020 studio или sRGB full.

2. Класс события и бюджет латентности

- ☐ Назовите класс события одним предложением («человек падает»).
- ☐ Бюджет латентности end-to-end в мс (камера → исполнитель).
- ☐ Топология под бюджет: только edge, только cloud или гибрид.

3. Выбор декодера

- ☐ PyAV — продакшн CPU-дефолт; читает цветовые метаданные.
- ☐ Decord — random-access ML-ридер с NVDEC.
- ☐ TorchCodec — PyTorch-нативный, тензоры на GPU.
- ☐ NVIDIA DALI — только для high-volume тренировки.

4. Каденция выборки кадров

- ☐ 30 fps — субсекундные события, быстрые номера и жесты.
- ☐ 6–15 fps — большинство action recognition, surveillance, pose.
- ☐ 1–2 fps — расписание / presence, дефолты VLM.
- ☐ Stride окна по карточке модели (например, SlowFast 16 / 2).

5. Пространственное преобразование

- ☐ Letterbox resize — обязателен для детекторов объектов.
- ☐ Centre crop после resize-shortest-side — для классификаторов.
- ☐ Интерполяция: bilinear (YOLO), bicubic (CLIP / ViT).

6. Константы нормализации

- ☐ ImageNet — ResNet, EfficientNet, ConvNeXt, большинство TIMM.
- ☐ CLIP — CLIP, SigLIP, LLaVA, Qwen-VL, Pixtral, InternVL.
- ☐ Raw 0–255 uint8 — Ultralytics YOLO, MediaPipe, многие ONNX.
- ☐ Загружены через ``AutoImageProcessor.from_pretrained(model_id)``.

7. Режим батчизации

- ☐ Dynamic batching на Triton — продакшн-дефолт.
- ☐ Continuous batching на vLLM — для VLM-нагрузок.
- ☐ Padded batching — для клипов переменной длины.
- ☐ Целевая утилизация GPU выше 70%.

8. Политика аугментации (тренировка)

- ☐ RandAugment + horizontal flip + colour jitter.
- ☐ Mixup или CutMix для image, VideoMix для клипов.
- ☐ Temporal dropout 20% для action recognition.
- ☐ Ablation — каждая аугментация даёт минимум 0,3 пп.

Резюме из восьми чисел

Разрешение / кодек источника · fps источника · цветовое пространство · класс события · бюджет латентности · каденция выборки · входная форма модели · набор констант нормализации.

Если хотя бы одного из этих восьми чисел нет в ТЗ — пайплайн предобработки не спроектировать корректно. Заполните все восемь до выбора модели.