

Чеклист развёртывания on-device CV

Depth Anything V2 + SmolVLM2 — выбрать правильные варианты и обойти ловушки лицензий и конвейера.

1. Выберите правильный вариант Depth Anything V2

- ☐ Коммерческий продукт? → только Depth-Anything-V2-Small (Apache 2.0, 24,8М параметров, 50+ FPS на потребительском GPU).
- ☐ Только research / внутренняя оценка? → Base / Large / Giant доступны, но под CC-BY-NC-4.0 (без коммерческого использования).
- ☐ Помещения (офисы, телемедицина, e-learning)? → metric-вариант Hypersim, max-глубина 20 м.
- ☐ Улица (surveillance, дрон, автопром)? → metric-вариант VKITTI, max-глубина 80 м.
- ☐ Сохранённый клип нужен с консистентной глубиной? → Video Depth Anything (CVPR 2025) вместо пер-кадрового V2.
- ☐ Live-стрим с low latency на 2K? → FlashDepth (2025) — backbone V2 + рекуррентное temporal-выравнивание.

2. Выберите правильный размер SmolVLM2

- ☐ Вкладка браузера / WebGPU? → SmolVLM2-256M, <1 GB GPU RAM, ~80 decode-токенов/с на M4 Max.
- ☐ Смартфон (iPhone 13+, Pixel 8+)? → SmolVLM2-500M, анализ видео на устройстве, есть демо-приложение Hugging Face.
- ☐ Jetson Orin / ноутбук / потребительский GPU? → SmolVLM2-2.2B, ~5,2 GB GPU RAM, лучшее понимание видео в классе.
- ☐ Лицензия? → все три размера под Apache 2.0. Без NC-оговорок. Коммерческое использование разрешено.

3. Стройте конвейер как воронку

- ☐ Стадия 1 — motion gate (background subtraction / крошечный YOLO-N) отбрасывает ~99% кадров на NPU камеры.
- ☐ Стадия 2 — depth + детекция параллельно; объединить в 3D-позицию; гейт по геометрической зоне.
- ☐ Стадия 3 — SmolVLM2 со structured yes/no-промптом, только на редких кадрах, которые окупают стоимость.
- ☐ Сопровождайте каждый вызов SmolVLM вторым проходом на другом сэмплированном кадре; действуйте только при совпадении.
- ☐ Для high-stakes-решений роутите survivors к frontier-VLM на сервере; on-device VLM — это дешёвый фильтр.

4. Шесть режимов отказа для аудита перед релизом

- ☐ Мерцание в depth-треке → ванильный V2 на видео дрожит; переключитесь на Video Depth Anything или FlashDepth.
- ☐ Неверный metric-вариант → Hypersim в помещении, VKITTI на улице; роутьте на runtime, если камера видит оба.
- ☐ Лицензионная ловушка → прототип на Large, релиз на Small. Аудит каждого чекпоинта в бинарнике.
- ☐ Галлюцинации VLM → ограничьте выход structured-промптами; требуйте совпадения на разных кадрах.
- ☐ Спрашивать у VLM геометрию → никогда не просите малый VLM об абсолютных расстояниях; числа — у depth-модели.
- ☐ Порог памяти → нужно 4 GB RAM устройства + AI-ускоритель; iPhone 13 / Pixel 8 / Jetson Orin Nano проходят.