

CV Pre-Processing Decision Worksheet

Answer the eight numbers below before model selection. One page; one project; one print.

1. Camera input — what comes off the wire

- ☐ Resolution and codec (e.g. 1080p H.265 or 720p WebRTC VP8).
- ☐ Frame rate at source (typical: 15, 24, 30, or 60 fps).
- ☐ Colour space and range — BT.709 studio, BT.2020 studio, or sRGB full.

2. Event class and latency budget

- ☐ Name the event class in one sentence (e.g. 'person collapses').
- ☐ Latency budget end-to-end in ms (camera to actuator).
- ☐ Topology that fits the budget: edge only, cloud only, or hybrid.

3. Decoder choice

- ☐ PyAV — production CPU default; reads colour metadata.
- ☐ Decord — random-access ML loader with NVDEC.
- ☐ TorchCodec — PyTorch-native, tensors on GPU.
- ☐ NVIDIA DALI — high-volume training only.

4. Frame sampling cadence

- ☐ 30 fps — sub-second events, fast plates and gestures.
- ☐ 6 to 15 fps — most action recognition, surveillance, pose.
- ☐ 1 to 2 fps — schedule / presence, VLM defaults.
- ☐ Window stride matches model card (e.g. SlowFast 16 / 2).

5. Spatial transformation

- ☐ Letterbox resize — required for object detectors.
- ☐ Centre crop after resize-shortest-side — for classifiers.
- ☐ Interpolation: bilinear (YOLO), bicubic (CLIP / ViT).

6. Normalisation constants

- ☐ ImageNet — ResNet, EfficientNet, ConvNeXt, most TIMM.
- ☐ CLIP — CLIP, SigLIP, LLaVA, Qwen-VL, Pixtral, InternVL.
- ☐ Raw 0 to 255 uint8 — Ultralytics YOLO, MediaPipe, many ONNX.
- ☐ Loaded from ``AutoImageProcessor.from_pretrained(model_id)``.

7. Batching mode

- ☐ Dynamic batching on Triton — production default.
- ☐ Continuous batching on vLLM — for VLM workloads.
- ☐ Padded batching — for variable-length clips.
- ☐ GPU utilisation target greater than 70%.

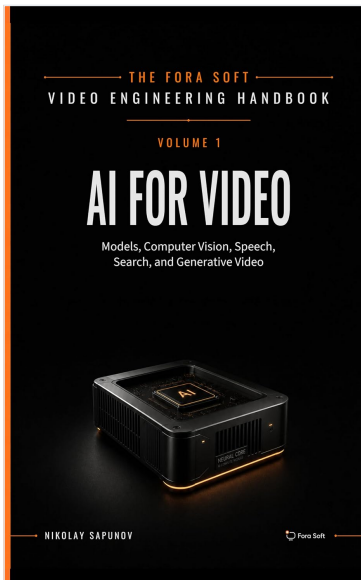
8. Augmentation policy (training)

- ☐ RandAugment + horizontal flip + colour jitter.
- ☐ Mixup or CutMix for image work, VideoMix for clips.
- ☐ Temporal dropout 20% on action recognition.
- ☐ Ablation tested — each augmentation earns at least 0.3 points.

The eight-number summary

Source resolution / codec · source fps · colour space · event class · latency budget · sampling cadence · model input shape · normalisation set.

If any of these eight is missing from the project spec, the pre-processing pipeline cannot be designed correctly. Fill all eight before picking a model.



GO DEEPER · VOLUME 6 OF 7

AI for Video: A Complete Guide to Intelligence in Video Products

From Models and Inference to Computer Vision, Speech, Search, and Generative Video

By Nikolay Sapunov, CEO at Fora Soft · 1,444 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH6TKPZ1

This worksheet fixes the eight pre-processing numbers before model selection, from colour space to normalisation. Volume 6 covers computer vision in video products in depth: detection, tracking, segmentation, pose and OCR, model artefacts, deployment topologies and the ML minimum you need to make these decisions.

WHAT YOU WILL FIND INSIDE

- Detection, tracking and segmentation
- Edge, cloud and hybrid topologies
- Model artefacts and input contracts
- ML minimum for engineering decisions

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Starting a computer vision feature on live or recorded video?

We will fix your pre-processing spec, pick the decoder and model, and measure accuracy and latency on your footage. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)