

PaddleOCR Deployment Checklist

PP-OCRv5 — map your video pipeline to the right model and avoid the six production failure modes.

1. Pick the right model

- ☐ CPU only or edge box? → PP-OCRv5_mobile_det + PP-OCRv5_mobile_rec (~8 MB, ~80 ms / 1080p frame)
- ☐ GPU (T4 / A10 / L4)? → PP-OCRv5_server_det + PP-OCRv5_server_rec (~100 MB, ~30-60 ms / frame)
- ☐ Multilingual feed? → Use the 106-language multilingual checkpoint, not the per-language one
- ☐ Export once to ONNX → serve through ONNX Runtime, OpenVINO, or Triton alongside other video AI models

2. Fix the six failure modes before going live

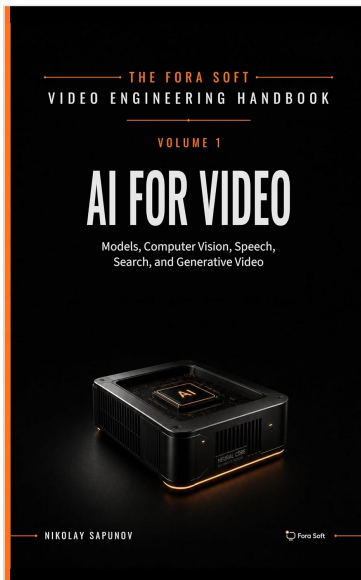
- ☐ Failure 1: Sampling every frame. Fix: run a motion detector or tracker first; OCR every 10th frame in each tracked region.
- ☐ Failure 2: Server model on CPU. Fix: match model to hardware. Mobile for CPU, server for GPU. Never cross the streams.
- ☐ Failure 3: Disabled direction classifier. Fix: keep use_textline_orientation=True. Costs 2 ms; saves 3 percentage points of accuracy.
- ☐ Failure 4: Small text at native size. Fix: 2x bicubic upscale before detection on burned-in subtitles. +8 ms, +5-10 points F-score.
- ☐ Failure 5: Per-frame reads without aggregation. Fix: confidence-weighted majority vote across all frames of the same tracked object.
- ☐ Failure 6: Hard-coded language. Fix: use the multilingual model with a per-track language-ID head; re-detect on scene change.

3. Pre-flight checks before going live

- ☐ Per-frame latency stays under your real-time budget (33 ms at 30 FPS, 16 ms at 60 FPS).
- ☐ GPU utilisation stays under 80% at peak load (room for traffic spikes and other AI stages).
- ☐ Aggregated string accuracy on hold-out test set is $\geq 95\%$ character-level on real footage from the target camera.
- ☐ Language-ID accuracy is $\geq 98\%$ on the multilingual hold-out set; misclassifications are logged for review.
- ☐ ONNX export is reproducible from the PaddlePaddle checkpoint; CI rebuilds it on each model refresh.
- ☐ PP-OCRv5 license note: Apache 2.0 commercial-use is fine; no per-page or per-frame fee.

4. Cost sanity check (2026 numbers)

- ☐ Self-hosted PP-OCRv5 on T4 GPU at ~\$0.35/hour: ~\$0.0006 per minute of 30-FPS video at 10 regions/frame.
- ☐ Google Document AI: \$1.50 per 1,000 pages (right for batch document workloads, wrong for continuous video).
- ☐ AWS Textract: \$1.50 per 1,000 pages (same reasoning as Document AI).
- ☐ Frontier VLM (Gemini / Claude Opus 4 / GPT-4o): ~\$3-15 per 1,000 frames (right when context matters, slower).



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 checklist maps a video pipeline to the right PP-OCRv5 model, fixes six failure modes and checks the cost. Volume 6 covers the computer vision behind it in depth: detection, tracking and OCR in video, model artifacts and deployment topologies, plus search over a video archive.

WHAT YOU WILL FIND INSIDE

- Text detection and OCR on video frames
- Tracking objects across frames
- Model formats and deployment topologies
- Search and RAG over a video archive

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

Need reliable text reading from live or archived video?

We will pick the OCR model for your hardware, add tracking and frame sampling, and hit your accuracy and latency targets. 250+ video products shipped since 2005.

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