

- Konfigurasi Test

- Models
- Epochs
- Batch
- Tflite

Konfigurasi Test

Models

Model	Jenis	Ukuran (approx)	Kecepatan	Akurasi	Cocok Untuk
yolov8n-cls.pt	Nano	◆ 6 MB	⚡ Cepat sekali	😐 Cukup baik	Edge device / Mobile
yolov8s-cls.pt	Small	◆ 12 MB	⚡ Cepat	👍 Lebih akurat dari n	Laptop / PC biasa
yolov8m-cls.pt	Medium	◆ 25 MB	⚙️ Sedang	✅ Bagus	Training dengan dataset sedang
yolov8l-cls.pt	Large	◆ 50 MB	🐢 Lebih lambat	🌟 Sangat akurat	GPU kuat
yolov8x-cls.pt	X-Large	◆ 95 MB	🐢🐢 Lambat	🌟🌟 Paling akurat	Training di Colab / GPU RTX
yolov9c-cls.pt	Compact (YOLOv9)	◆ 20 MB	⚙️ Cepat	✅ Akurasi tinggi	Edge device modern
yolov9e-cls.pt	Enhanced (YOLOv9)	◆ 60 MB	🐢	🌟 Akurasi tinggi banget	Server / GPU kuat

Epochs

Epochs	Durasi Training	Overfitting Risk	Hasil Umum	Rekomendasi Penggunaan	Keterangan Visual
10	⚡ Sangat cepat	▼ Sangat rendah	✗ Belum stabil	Cek awal model / uji pipeline	⌚ Tes awal, belum cukup untuk akurasi tinggi
25	⚙️ Cepat	▼ Rendah	😐 Masih dasar	Training awal atau dataset kecil	🔁 Cocok buat “sanity check” model

Epochs	Durasi Training	Overfitting Risk	Hasil Umum	Rekomendasi Penggunaan	Keterangan Visual
50	⚙️ Sedang	⚠️ Sedikit	😊 Mulai stabil	Dataset kecil-menengah	Akurasi mulai terlihat meningkat
75	⌚ Cukup lama	⚠️ Sedang	👍 Stabil	Dataset menengah	Titik awal perbaikan loss
100	🕒 Lama	⚠️ Sedang-tinggi	✅ Bagus	Dataset sedang-besar	Biasanya titik optimal akurasi
150	🕒 Lebih lama	⚠️ Tinggi	🌟 Sangat baik	Dataset besar, GPU kuat	Model mulai "fokus" dan spesifik
200	⌚ Panjang sekali	⚠️⚠️ Overfitting rawan	⭐⭐ Akurasi tinggi	Eksperimen akhir / model final	Butuh early stopping
300+	🕒 Sangat lama	🔥 Sangat tinggi	⚖️ Kadang turun	Penelitian / fine-tuning ekstrem	Waktu besar, hasil bisa stagnan

Batch

Batch Size	Kecepatan Training	Kebutuhan Memori	Akurasi	Stabilitas Gradien	Cocok Untuk	Keterangan Visual
4	🐢 Sangat lambat	♦ Sangat kecil	😐 Kurang stabil	⚠️ Cenderung fluktuatif	CPU / GPU lemah	Cocok untuk debugging awal
8	🐢 Lambat	♦ Kecil	😊 Cukup stabil	⚙️ Masih fluktuatif	Laptop / Colab gratis	Aman untuk dataset kecil
16	⚙️ Sedang	♦ Menengah	👍 Stabil	✅ Seimbang	GPU menengah (RTX 2060-3060)	Kombinasi efisiensi dan stabilitas
32	⚡ Cepat	♦ Cukup besar	🌟 Stabil & cepat	✅✅ Sangat baik	GPU kuat (RTX 3070+)	Umumnya paling optimal
64	⚡⚡ Cepat	♦ Besar	⭐⭐ Tinggi	✅ Stabil	GPU besar / Server	Untuk dataset besar & model besar
128	⚡⚡⚡ Sangat cepat	▲ Sangat besar	⭐⭐⭐ Sedikit turun	⚠️ Terlalu rata (loss lambat)	TPU / Server AI	Eksperimen efisiensi tinggi

Batch Size	Kecepatan Training	Kebutuhan Memori	Akurasi	Stabilitas Gradien	Cocok Untuk	Keterangan Visual
256+	⚡ Maksimal	▲ ▲ Ekstrem	⚖ Kadang menurun	⚠ Sulit konvergen	TPU / Multi-GPU cluster	💻 Untuk riset / benchmark

Tflite

Mode Export	Jenis Kuantisasi	Ukuran Model	Kecepatan Inference	Akurasi	Kompatibilitas Perangkat	Cocok Untuk	Keterangan Visual
Default (FP32)	Tanpa kuantisasi (float 32-bit)	Besar (~25–60 MB)	🐢 Normal	⭐ 100% (asli)	Semua perangkat dengan CPU/GPU modern	Laptop, Android kelas menengah ke atas	📦 Model utuh tanpa kompresi
FP16 (Half Precision)	Float 16-bit	Sedang (~12–25 MB)	⚡ Cepat	⭐ 98–99%	GPU / Android modern / Jetson	Mobile AI / Jetson Nano / Raspberry Pi 4	⚖ Seimbang antara akurasi dan ukuran
INT8 (Integer 8-bit)	Kuantisasi penuh 8-bit	Kecil (~6–10 MB)	⚡ Cepat	👍 93–96%	MCU / Edge Device / ESP32 AI Kit	IoT & perangkat low-power	💾 Sangat efisien & ringan
Dynamic Range	Kuantisasi dinamis (mixed float–int)	Sedang (~10–20 MB)	⚡ Cepat di CPU	✓ 96–98%	Laptop / CPU biasa	Desktop CPU / Colab / Edge ringan	🔄 Balance otomatis oleh TensorFlow
Float32 → INT8 hybrid	Kuantisasi pasca-training	Kecil (~8–12 MB)	⚙️ Cepat	👍 95–97%	Android & Edge modern	Deployment umum	🔧 Optimasi pasca-training
Optimized (Size)	Prioritaskan ukuran model kecil	Sangat kecil (~5–8 MB)	⚡ Cepat	😐 Turun 2–4%	IoT & Microcontroller	Deployment hemat memori	📈 Ukuran super efisien

Mode Export	Jenis Kuantisasi	Ukuran Model	Kecepatan Inference	Akurasi	Kompatibilitas Perangkat	Cocok Untuk	Keterangan Visual
Optimized (Latency)	Prioritaskan kecepatan inferensi	♦ Sedang (~10–15 MB)	⚡⚡⚡ Sangat cepat	⭐ Stabil	Android / Jetson / Edge AI	Aplikasi real-time	 Fokus pada real-time response