

YOLO26n · FP32

yolo26n.onnx · identidad del grafo f21c7ca23a529ada... · medido el 3 de octubre de 2026

[Descargar este reporte en PDF](#) — es exactamente lo que recibe un cliente.

El veredicto

Corre en las 3 configuraciones medidas, con resultados correctos. La más rápida: Intel N100 · ONNX Runtime, con 57.6 ms por inferencia, unas 17 por segundo.

RASPBERRY PI 5 · ONNX RUNTIME Corre 126.4 ms por inferencia, p50 · p95 134.4 ms · 8 por segundo	INTEL N100 · ONNX RUNTIME Corre 57.6 ms por inferencia, p50 · p95 58.9 ms · 17 por segundo	INTEL N100 · OPENVINO Corre 64.0 ms por inferencia, p50 · p95 65.6 ms · 16 por segundo
---	--	--

Rendimiento

Inferencia del modelo, en milisegundos. «De punta a punta» suma preparar la entrada y procesar la salida.

CONFIGURACIÓN	P50	P95	P99	DE PUNTA A PUNTA P50	POR SEGUNDO
Raspberry Pi 5 · 8 GB · ONNX Runtime 1.30.0	126.4	134.4	138.3	138.1	8
Intel N100 · 8 GB · ONNX Runtime 1.30.0	57.6	58.9	59.3	67.2	17
Intel N100 · 8 GB · OpenVINO 2026.4.0	64.0	65.6	71.3	72.4	16

Resultados correctos

Las salidas se comparan con una referencia calculada con ONNX Runtime en un equipo de referencia, con las mismas entradas.

Raspberry Pi 5 · ONNX Runtime ✓ encuentra el 100 % de los objetos de la referencia, con una precisión del 100 %, en 12 imágenes

Intel N100 · ONNX Runtime ✓ encuentra el 100 % de los objetos de la referencia, con una precisión del 100 %, en 12 imágenes

Intel N100 · OpenVINO ✓ encuentra el 100 % de los objetos de la referencia, con una precisión del 100 %, en 12 imágenes

Lo que se esperaba y lo que pasó

Antes de medir, el protocolo deja escrito qué espera en cada dimensión. Después se contrasta con lo medido.

DIMENSIÓN	RASPBERRY PI 5 · ONNX RUNTIME	INTEL N100 · ONNX RUNTIME	INTEL N100 · OPENVINO
Modelo ONNX válido	✓ pasa onnx.checker OK	✓ pasa onnx.checker OK	—
Carga en el runtime	✓ pasa carga 57.68 ms	✓ pasa carga 116.609 ms	✓ pasa carga 359.373 ms
Operadores soportados	✓ pasa sesión creada sin operadores rechazados	✓ pasa sesión creada sin operadores rechazados	—
Cabe en memoria	✓ pasa pico RSS = 3.0% de la RAM	✓ pasa pico RSS = 3.3% de la RAM	✓ pasa pico RSS = 3.5% de la RAM
Resultados correctos	✓ pasa recall 100% · precision 100% vs referencia	✓ pasa recall 100% · precision 100% vs referencia	✓ pasa recall 100% · precision 100% vs referencia
p95 ≤ 200 ms	◆ cumple p95 134.3902 ms vs ≤ 200 ms	◆ cumple p95 58.8727 ms vs ≤ 200 ms	◆ cumple p95 65.5797 ms vs ≤ 200 ms

Estabilidad térmica	◆ pasa Tmax 65.3 °C	—	—
Soporte de la CPU	—	—	✓ pasa todo se ejecuta en CPU
Más rápido que ONNX Runtime	—	—	◆ no p50 64.045 ms vs 57.5655 ms en C08 (−11.3%)

✓ se confirmó lo que se esperaba · ◆ era una incógnita y quedó medida · · no se evalúa en esta medición

14 predicciones confirmadas, ninguna equivocada.

No se evaluó en esta medición: exportación.

Memoria, carga y temperatura

	RASPBERRY PI 5 · ONNX RUNTIME	INTEL N100 · ONNX RUNTIME	INTEL N100 · OPENVINO
Carga del modelo	58 ms	117 ms	359 ms
Memoria pico	239 MB · 3.0 % de la RAM	255 MB · 3.3 % de la RAM	265 MB · 3.5 % de la RAM
Temperatura	57 → 65 °C	53 → 62 °C	54 → 64 °C
Frecuencia de la CPU	2400–2400 MHz	2816–2900 MHz	2821–3150 MHz
Reducción térmica	no	no, por frecuencia (sin lectura directa)	no, por frecuencia (sin lectura directa)

El modelo

Archivo	yolo26n.onnx · 9.5 MB · FP32
Origen	Ultralytics yolo26n.pt (oficial; nunca una conversión de terceros)
Licencia	AGPL-3.0
Entradas	images FLOAT [1, 3, 640, 640]
Salidas	output0 FLOAT [1, 300, 6]
Grafo	ONNX opset 17 · 384 nodos · 23 tipos de operador
Identidad del archivo	843640bc7ae797dfbf3d9828ced4bfce393a78949b6ed282a0db874f64ef2680
Identidad del grafo	f21c7ca23a529adaf73a497aba4b366aa0d94f9f5bdf02c55aebbbe1460bd90f

Condiciones de la medición

	RASPBERRY PI 5 · ONNX RUNTIME	INTEL N100 · ONNX RUNTIME	INTEL N100 · OPENVINO
Sistema	Debian GNU/Linux 13 (trixie)	Ubuntu 24.04.5 LTS	Ubuntu 24.04.5 LTS
Python	3.13.5	3.12.3	3.12.3
Precisión efectiva	FP32, la del modelo	FP32, la del modelo	f32, fijada y comprobada
Hilos	los del runtime, por defecto	los del runtime, por defecto	los del runtime, por defecto
Medidas	100 tras 20 de calentamiento	100 tras 20 de calentamiento	100 tras 20 de calentamiento
Antes de medir	equipo recién reiniciado	equipo recién reiniciado	equipo recién reiniciado
Entorno congelado	ENV-T1-RPI5-8GB- LAB01-v01	ENV-T2-N100-8GB- LAB01-v02	ENV-T2-N100-8GB- LAB01-v02
Observación	OBS-C07- 20261003T215041Z- b5c74f	OBS-C08- 20261003T224756Z- 1c9208	OBS-C09- 20261003T224811Z- 9534d9

Lo que este reporte no cubre

- El consumo eléctrico: no se midió.

- Gráficas integradas y aceleradores: se midió sólo la CPU.
- Más de una entrada a la vez: todo es con batch 1, midiendo latencia.
- Otros runtimes y otros equipos distintos de los que aparecen aquí.

Es un reporte de muestra. El modelo es el YOLO26n oficial de Ultralytics, que es público, y por eso este reporte se puede publicar. Los reportes de clientes son confidenciales, y el modelo se borra al entregarlos.

Generado el 5 de octubre de 2026.