Clasificación de piezas mediante Deep Learning para inspección industrial

Análisis de Imagen y Visión Artificial

Autor: Martín Solano Martinez

Introducción y objetivos

El objetivo principal es desarrollar un sistema de visión artificial capaz de clasificar pequeño material mecánico que circula por una cinta transportadora a alta velocidad.

Requisitos del sistema:

- Velocidad de cinta: 1 m/s (1000 mm/s)
- Distancia de trabajo: 500 mm - 700mm
- Resolución mínima: La tuerca más pequeña (6mm) debe representarse con al menos 165 píxeles

Dimensionamiento del sistema de captura

Para garantizar la viabilidad del software, el primer paso es dimensionar el hardware (cámara y óptica). El requisito crítico de resolución define la densidad de píxeles necesaria.

1. Requisito

135 Píxeles para un objeto de 6 mm.

2. Cálculo de densidad

$$\frac{135 \text{ px}}{6 \text{ mm}} = \mathbf{22{,}5 \text{ px/mm}}$$

Este es el número clave que garantiza la calidad de la imagen.

3. Cálculo del campo de visión (FOV)

Usando un sensor estándar de 5MP (2448x2048 px), calculamos el área máxima visible.

$$\text{FOV}_{\text{horizontal}} = \frac{2448 \text{ px}}{22{,}5 \text{ px/mm}} \approx 108{,}8 \text{ mm}$$

Conclusión clave: El sistema puede capturar un área de hasta 108,8 mm de ancho manteniendo la resolución mínima exigida para las piezas más pequeñas.

Validación rango dinámico de tamaños

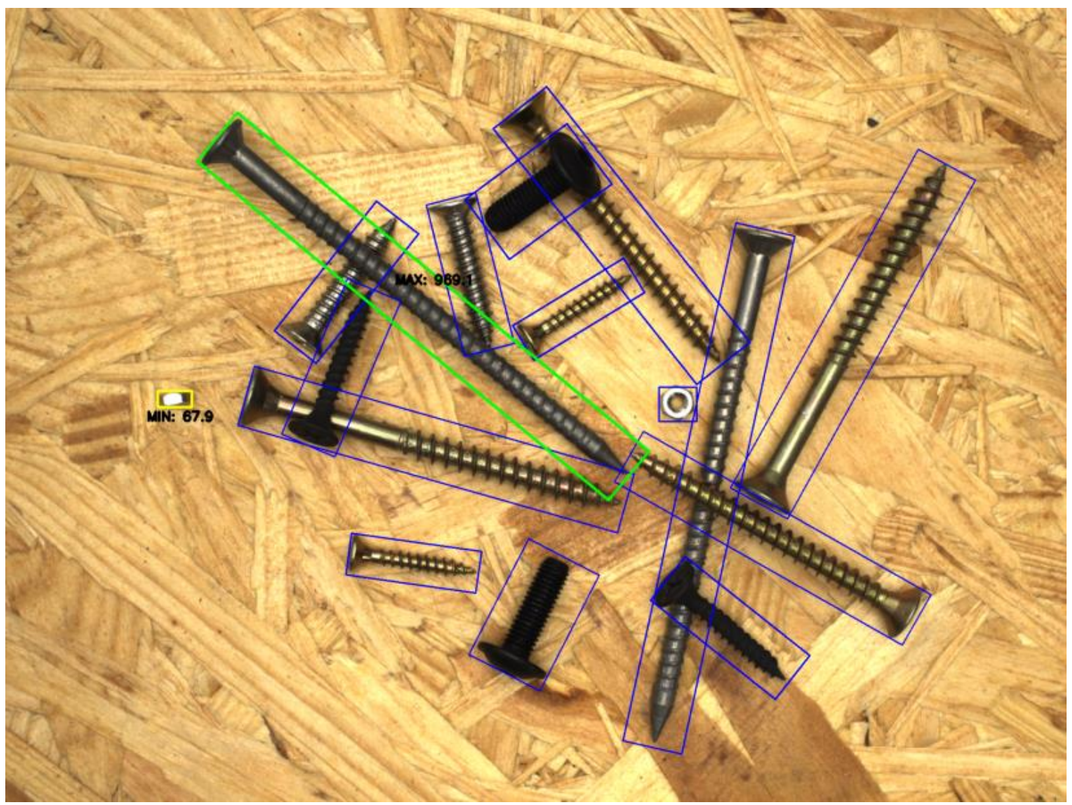
Se verificó que el FOV de 108,8 mm es suficiente para las piezas más grandes desarrollando un algoritmo para encontrar el 'peor caso' en todo el dataset: la imagen con la mayor disparidad de tamaños.

Análisis del peor caso:

- Ratio máximo: 14.27

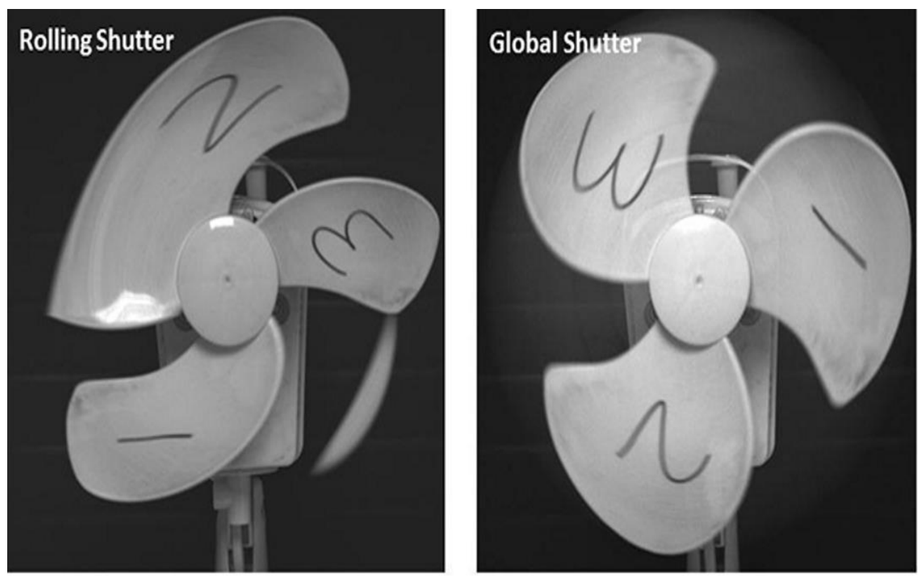
(en la imagen screws_005.png)

- Estimación de longitud máxima real:
  6 mm (pieza mín.) x 14.27 = 85,62 mm



Elección del obturador global

A una velocidad de 1 m/s, un sensor convencional (Rolling Shutter) produciría distorsiones geométricas, falseando las mediciones. La tecnología Global Shutter expone todos los píxeles simultáneamente, garantizando la fidelidad geométrica.



Conclusión: La selección de un sensor con obturador global es un requisito indispensable para la integridad de los datos en esta aplicación de alta velocidad.

Configuración final de hardware

Sensor seleccionado

- **Sony IMX547**
  - Especificaciones: 5MP, formato 1/1.8", ancho del sensor 6,71 mm.

Cálculo de la lente

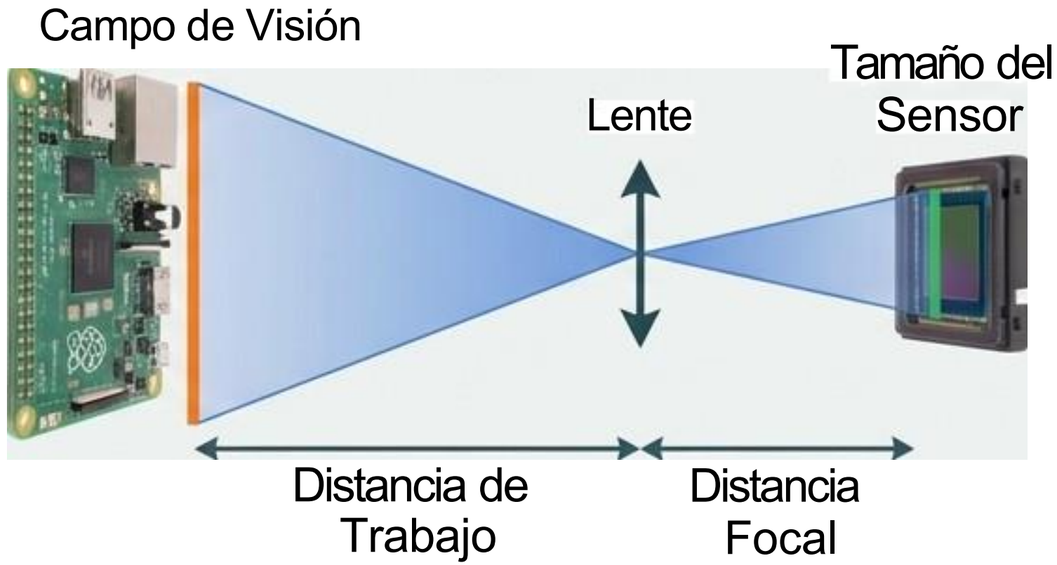
Usando la fórmula del modelo Pinhole:

$$f = (h \times WD)/FOV$$

Para una distancia de trabajo (WD) de 600 mm, se obtiene una distancia focal teórica de 37,0 mm.

Selección comercial y verificación

- Lente comercial: 35 mm (la opción de 50 mm excedía la distancia de trabajo máxima).
- Distancia de trabajo real:
  WD= (35 mm x 108,8 mm)/ 6,71 mm =567,4 mm.



Presupuesto

- Cámara Basler (Sensor IMX547): = 650€
- Óptica computar MPW3 35 mm : 240€

Total: 890 €

Arquitectura eficiente para inferencia en tiempo real

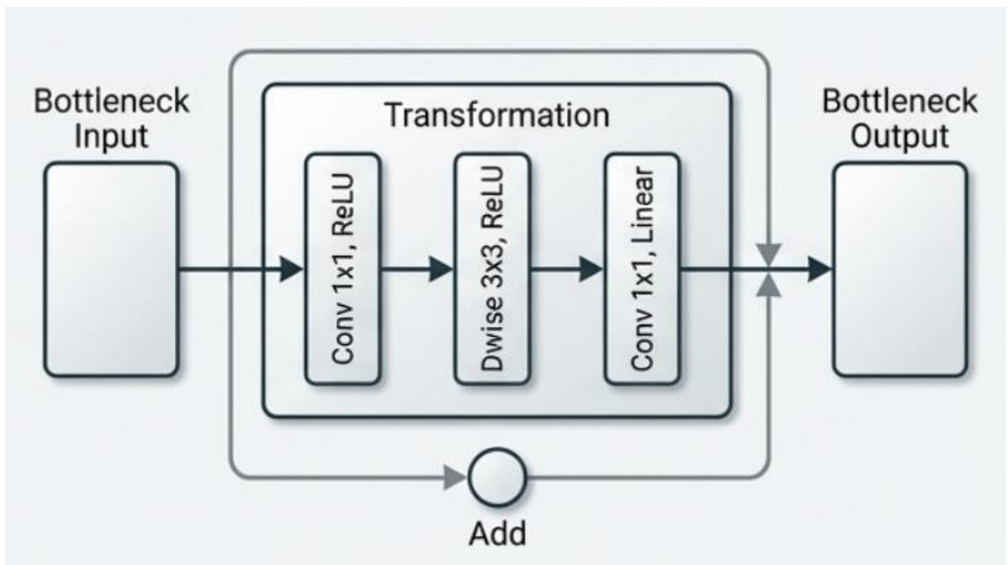
Para garantizar la clasificación en tiempo real, se seleccionó la arquitectura MobileNetV2. Su diseño es fundamental para lograr una latencia mínima.

Convoluciones Separables en Profundidad

- Esta técnica factoriza la convolución estándar en dos operaciones más ligeras: Depthwise (3x3) y Pointwise (1x1).

- Resultado: Reduce las operaciones de punto flotante (FLOPs) en un factor de 8 a 9 veces.

- Parámetros: = 3.5 millones (un 86% menos que ResNet50).



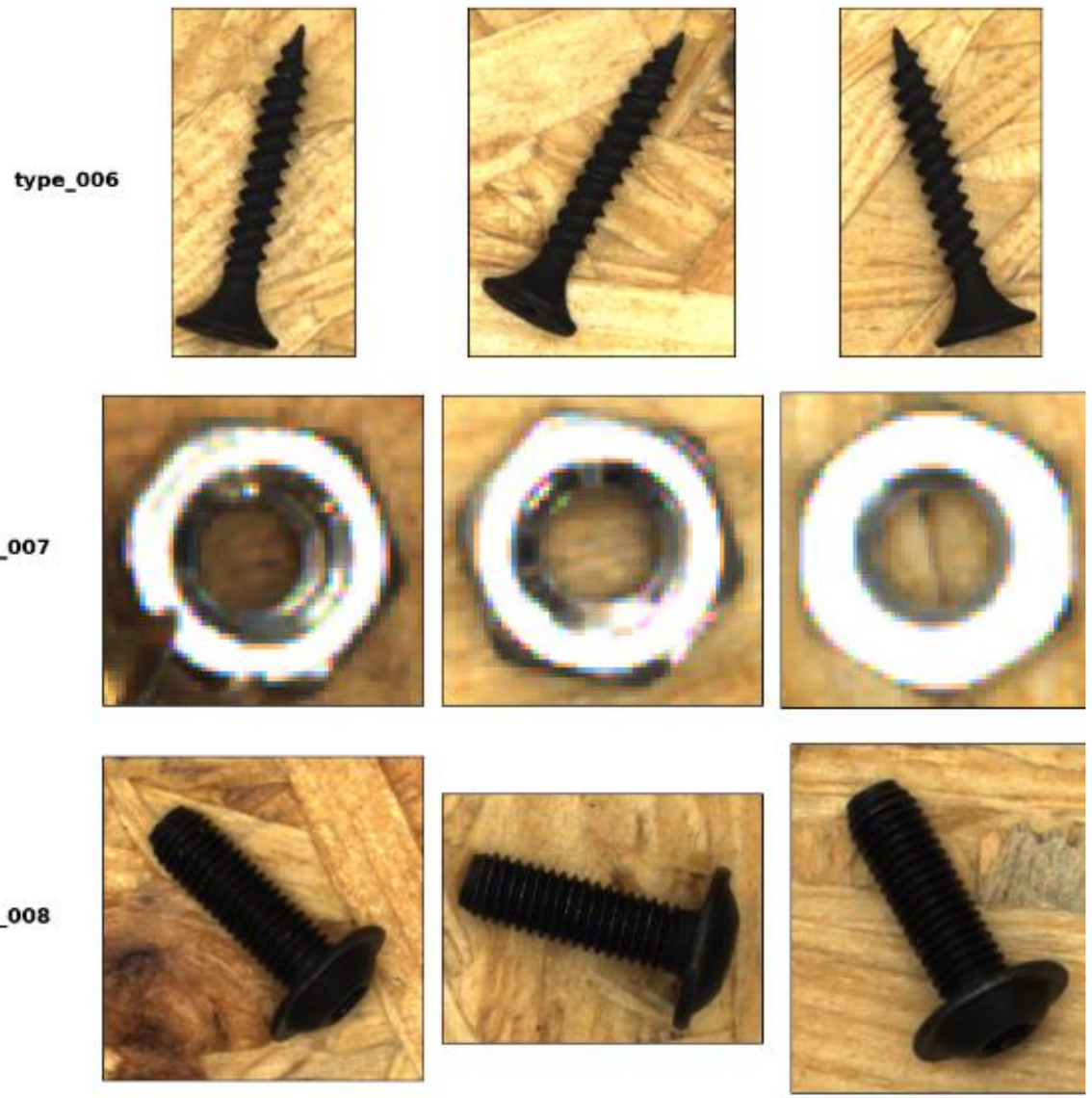
Preparación del dataset

Proceso de Extracción

1. Se leen las anotaciones de cajas rotadas (Oriented Bounding Box).
2. Se calculan los vértices de cada pieza.
3. Se genera el rectángulo envolvente recto mínimo (Axis-Aligned Bounding Box).

¿Por qué no rotar la imagen?

- Evita artefactos de interpolación: Preserva la calidad original de los píxeles.
- Conserva la variabilidad natural: Obliga al modelo a aprender características invariantes a la rotación, lo que lo hace más robusto.



Estrategia de entrenamiento: Transfer Learning

Infraestructura computacional

- Plataforma: Google Colab
- Hardware: GPU NVIDIA Tesla T4 para acelerar el entrenamiento.

Arquitectura del modelo final

1. Base (Congelada): MobileNetV2 con pesos de ImageNet. Sus capas no se re-entrenan.
2. Nueva cabeza de clasificación (Entrenable): Se añade una nueva sección final con:
   - GlobalAveragePooling2D
   - Capas dense con dropout (para evitar sobreajuste).
   - Capa de salida Softmax para las 13 clases

block_16_project (Conv2D)	(None, 7, 7, 320)	307,200	block_16_depthwi…
block_16_project_BN (BatchNormalizatio…	(None, 7, 7, 320)	1,280	block_16_project…
Conv_1 (Conv2D)	(None, 7, 7, 1280)	409,600	block_16_project…
Conv_1_bn (BatchNormalizatio…	(None, 7, 7, 1280)	5,120	Conv_1[0][0]
out_relu (ReLU)	(None, 7, 7, 1280)	0	Conv_1_bn[0][0]
global_average_poo… (GlobalAveragePool…	(None, 1280)	0	out_relu[0][0]
dense (Dense)	(None, 512)	655,872	global_average_p…
dropout (Dropout)	(None, 512)	0	dense[0][0]
dense_1 (Dense)	(None, 13)	6,669	dropout[0][0]

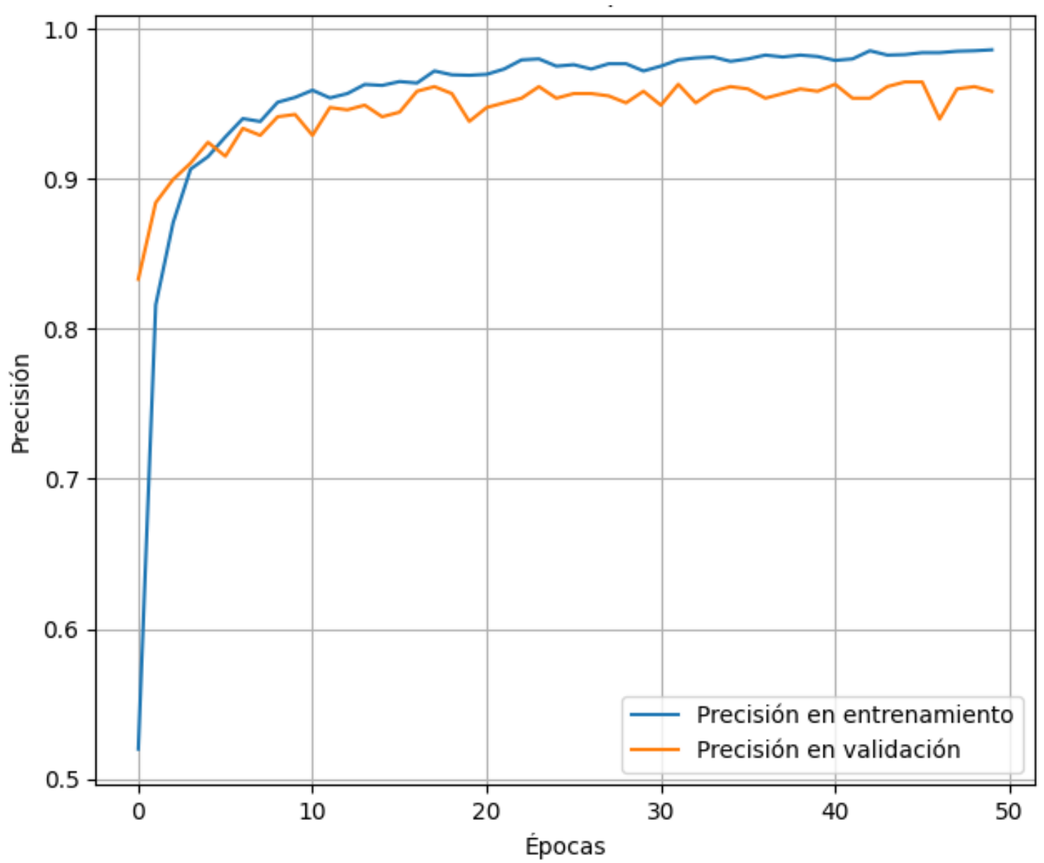
Total params: 2,920,525 (11.14 MB)

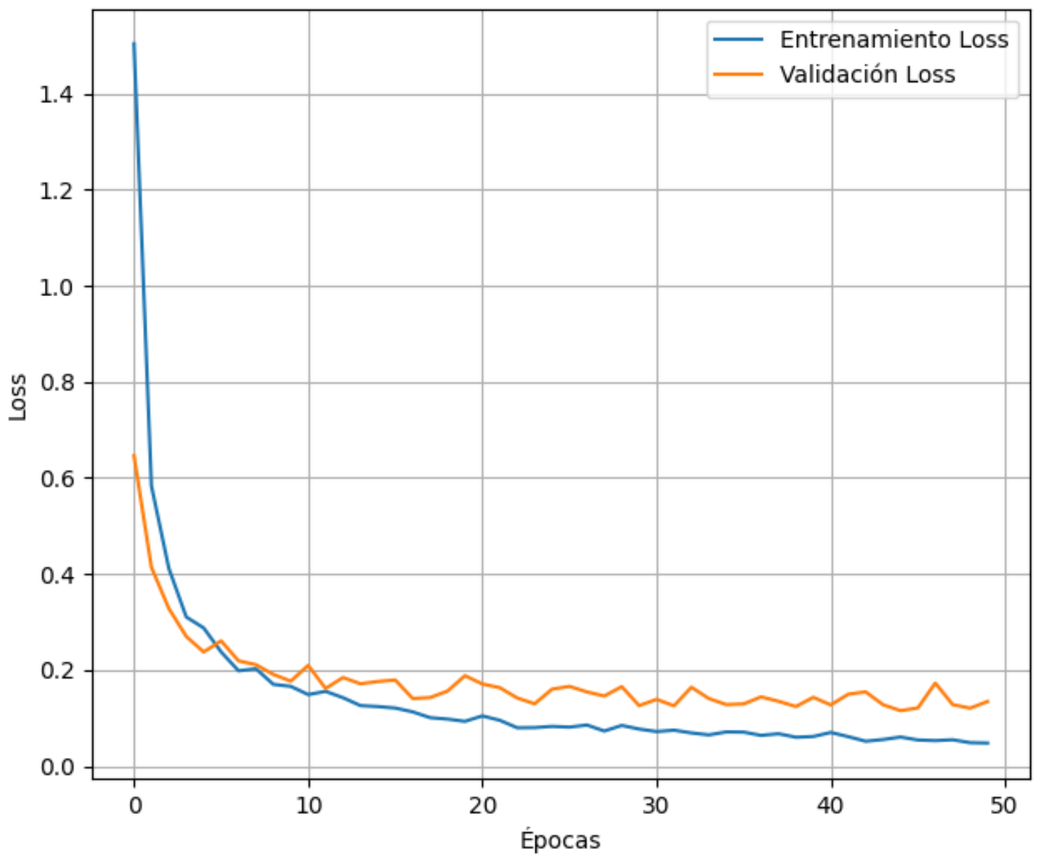
Trainable params: 662,541 (2.53 MB)

Non-trainable params: 2,257,984 (8.61 MB)

Resultados del entrenamiento

El modelo fue entrenado durante 50 épocas. Las curvas de aprendizaje demuestran la eficacia del Transfer Learning, alcanzando un rendimiento excelente.



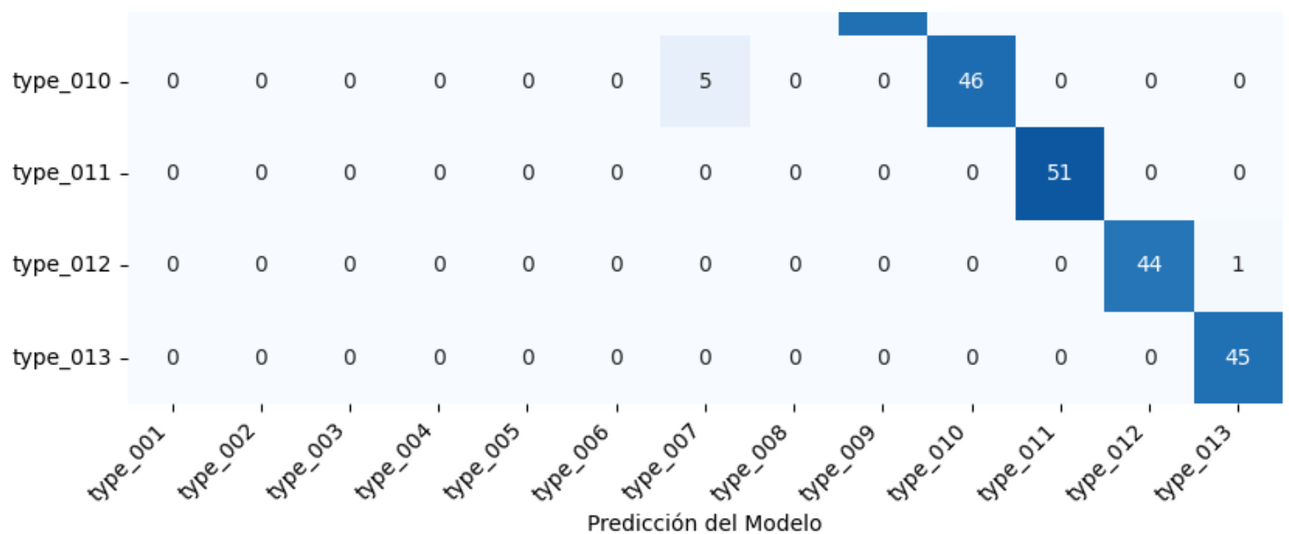


Análisis de resultados

La matriz de confusión revela un comportamiento desigual pero controlado. Analizar el rendimiento por clase es clave para la validación industrial.

Conflicto geométrico crítico:

La clase “type_010” es confundida con “type_007 en un 10% de los casos debido a su alta similitud visual



REPORTE DE CLASIFICACIÓN

	precision	recall	f1-score	support
type_001	0.94	1.00	0.97	47
type_002	1.00	0.98	0.99	48
type_003	1.00	0.96	0.98	49
type_004	0.92	1.00	0.96	48
type_005	0.96	0.93	0.95	59
type_006	1.00	0.95	0.98	63
type_007	0.92	1.00	0.96	55
type_008	1.00	1.00	1.00	55
type_009	1.00	1.00	1.00	45
type_010	1.00	0.90	0.95	51
type_011	1.00	1.00	1.00	51
type_012	0.98	0.98	0.98	45
type_013	0.98	1.00	0.99	45
accuracy			0.98	661
macro avg	0.98	0.98	0.98	661
weighted avg	0.98	0.98	0.98	661

Auditoría visual de errores

1. Similitud visual

Conflictos entre “type_010” y “type_007” (tuercas) o “type_004” y “type_005” (tornillos). La pérdida de resolución a 224x224 px difumina detalles finos.

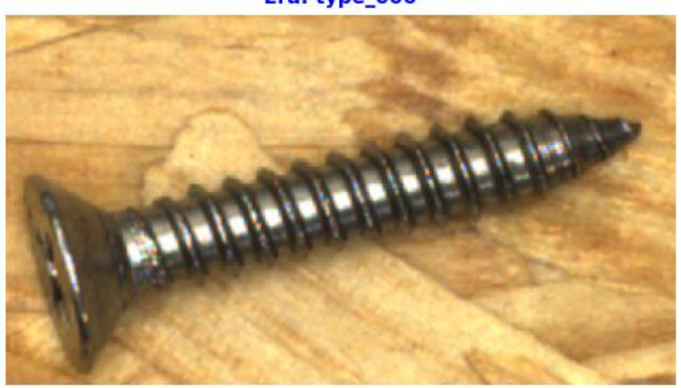
INPUT (error)
Archivo: 36_36005.png
Se predijo: type_007
Era: type_010

Ejemplo típico de la predicción
Archivo: 10_10001.png
Clase: type_007
(set de entrenamiento)

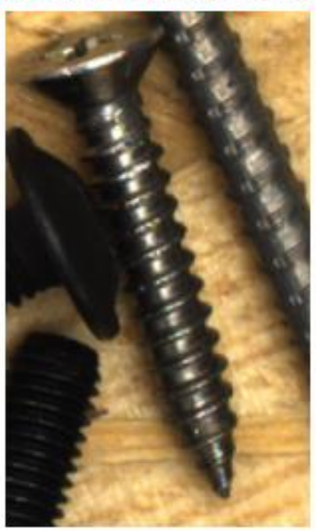
2. Errores en el etiquetado original

En varios casos, el modelo clasificó correctamente la pieza, pero la etiqueta del dataset era errónea. Esto demuestra una capacidad de generalización superior al propio etiquetado manual.

INPUT (error)
Archivo: 45_45006.png
Se predijo: type_005
Era: type_006



Ejemplo típico de la predicción
Archivo: 10_10011.png
Clase: type_005
(set de entrenamiento)



¿Cumple el sistema con la velocidad de la línea de producción?

Cálculo de la ventana de tiempo (T_window)

Datos de entrada:

- Velocidad de la cinta (v): 1000 mm/s
- Campo de visión longitudinal (FOV): 108,8 mm

Cálculo:

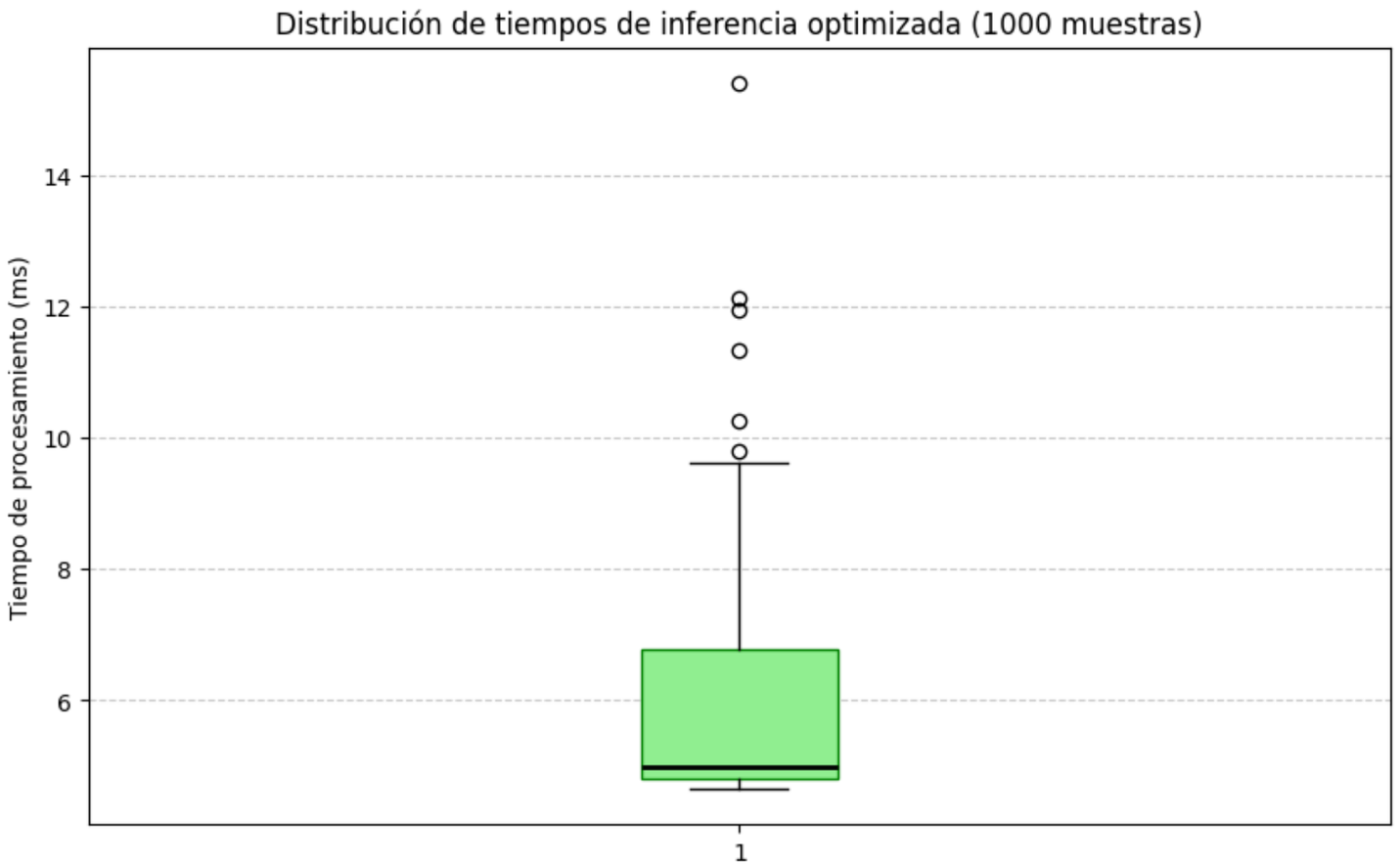
$T\_window = FOV / v = 108{,}8\ mm / 1000\ mm/s = 0{,}1088\ s = 108{,}8\ ms$

El requisito: tiempo máximo de procesamiento

108 milisegundos

Prueba de estrés

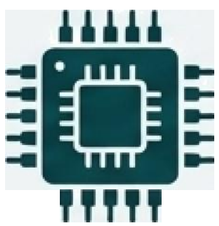
Se realizó una prueba de estrés con 1000 inferencias consecutivas en un entorno de producción simulado, utilizando optimización de grafos estáticos para ejecutar el modelo directamente sobre el motor de C++ y CUDA, maximizando así el rendimiento hardware.



Conclusiones y líneas futuras

El sistema valida el uso de redes neuronales ligeras para inspección de alta velocidad, operando a un equivalente de 173 FPS. La latencia de 5.77 ms está tan por debajo del límite de 108 ms que el sistema está “sobredimensionado", abriendo nuevas vías estratégicas.

Líneas futuras de mejora



1. Reducción de costes (Edge Computing)

Portar el modelo a hardware de bajo coste (ej. Raspberry Pi, NVIDIA Jetson Nano). El enorme margen de tiempo permite esta migración, reduciendo drásticamente el coste por estación de inspección.

2. Optimización de la precisión y producción

- Aumentar velocidad: La planta podría incrementar la velocidad de la cinta por encima de 1 m/s.
- Inspección multi-vista: Usar el tiempo sobrante (>100 ms) para procesar múltiples capturas de la misma pieza desde distintos ángulos, elevando la precisión operativa.