



Pontificia Universidad
JAVERIANA
Cali

**USO DE MODELOS DE APRENDIZAJE PROFUNDO PARA DETECTAR
DEFECTOS DE SOLDADURA EN PIEZAS METÁLICAS**

Programa de Maestría en Ingeniería

Presentado por:

JULIÁN ZULUAGA BOLÍVAR

Dirigido por:

JULIÁN GIL GONZÁLEZ Ph.D.

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Junio de 2025

Uso de modelos de aprendizaje profundo para detectar defectos de soldadura en piezas metálicas

J. Zuluaga

Facultad de ingeniería, Pontificia Universidad Javeriana de Cali, Colombia

julianzu@javerianacali.edu.co

1. Abstract

La inspección automatizada de defectos en soldaduras es un factor crítico para asegurar la calidad y seguridad en la industria. Este estudio propone el uso de modelos de aprendizaje profundo YOLO (You Only Look Once), entrenados mediante transferencia de aprendizaje, para la detección de defectos en soldaduras a partir de imágenes fotográficas. Empleando un conjunto de datos público, clasificado inicialmente en "soldadura mala", "soldadura buena" y "defecto", se realizaron tres fases experimentales. En la fase inicial, se compararon YOLOv8m y YOLO11m, revelando un buen desempeño para soldaduras aceptables, pero una precisión limitada ($\approx 48\%$) en la identificación específica de defectos además de presentarse confusión con el fondo (defecto no identificado). La segunda fase se centró en la optimización de hiperparámetros de YOLO11m, logrando mejoras moderadas, pero persistiendo la baja precisión ($\approx 49\%$) identificando la clase defecto. Para abordar esta limitación, la tercera fase fusionó las categorías "defecto" y "soldadura mala" en una única clase denominada "soldadura no conforme". Esta estrategia simplificó la clasificación y mejoró significativamente el rendimiento global del modelo, alcanzando una precisión de aproximadamente 75% en la detección de soldaduras no conformes. Los resultados demuestran el potencial de esta

metodología para optimizar los procesos de control de calidad en la industria, reduciendo la dependencia de inspecciones manuales y mejorando la eficiencia.

2. Introducción

La detección de defectos en soldadura es un aspecto fundamental en la industria manufacturera ya que las imperfecciones en las uniones soldadas pueden comprometer la integridad estructural de los productos y generar fallas que, dependiendo de la aplicación, podrían ser catastróficas [1]. Sectores como el aeroespacial, la industria automotriz y la construcción dependen de soldaduras de alta calidad para garantizar la seguridad y durabilidad de sus componentes. Tradicionalmente, la inspección de soldaduras se ha llevado a cabo mediante técnicas de ensayos no destructivos (NDT, por sus siglas en inglés), tales como la radiografía, ultrasonido y técnicas visuales [2]. Sin embargo, estos métodos suelen ser costosos, requieren personal capacitado y pueden estar sujetos a interpretaciones subjetivas.

Una alternativa accesible consiste en el uso de imágenes RGB capturadas con cámaras convencionales, que permiten realizar inspección visual sin necesidad de equipos NDT especializados. No obstante, factores como

iluminación variable y ángulos de captura pueden dificultar la detección manual de defectos, motivando el uso de modelos automatizados basados en aprendizaje profundo [3]. Estas problemáticas pueden ser abordadas mediante aplicaciones basadas en inteligencia artificial (IA), que mejoran la precisión y robustez en la identificación de defectos mediante técnicas de procesamiento de imágenes y aprendizaje profundo [4].

En los últimos años, el uso de técnicas basadas en IA, particularmente el aprendizaje profundo (*deep learning*), ha emergido como una solución prometedora para la automatización de la detección de defectos en soldadura. Modelos de redes neuronales convolucionales (CNNs) han demostrado un rendimiento sobresaliente en tareas de clasificación y detección de defectos en imágenes de soldaduras, superando las limitaciones de los enfoques tradicionales [5]. Entre las arquitecturas más utilizadas, YOLO se destaca por su capacidad de realizar detección en tiempo real con alta precisión, lo que la hace adecuada para sistemas de inspección industrial automatizados [6]. Estudios recientes han explorado diversas arquitecturas de segmentación y clasificación de defectos [7, 8], a partir del uso de modelos preentrenados con transferencia de aprendizaje [5]. Sin embargo, a pesar de estos avances, existen desafíos significativos relacionados con la disponibilidad y calidad de los conjuntos de datos, la interpretabilidad de los modelos y la generalización a distintos tipos de soldadura y materiales [9, 10].

En este estudio se propone un enfoque basado en aprendizaje profundo para la detección automatizada de soldaduras defectuosas mediante

el uso de modelos YOLO preentrenados (YOLOv8m y YOLO11m), entrenados sobre un conjunto de datos anotado en formato YOLO que incluye imágenes RGB clasificadas en tres categorías: soldaduras buenas, malas y defectos [2], las soldaduras buenas enmarcan un cordón de soldadura bien aplicado, las soldaduras malas enmarcan un cordón de soldadura que no cumple con las características necesarias para garantizar una buena unión entre las piezas, y por último la clase defecto enmarca aspectos puntuales de una soldadura defectuosa como lo son poros o grietas. Una ventaja clave del sistema es el uso de imágenes fotográficas capturadas con dispositivos convencionales, como teléfonos móviles, lo que facilita su implementación en entornos industriales sin necesidad de equipos especializados, a diferencia de métodos como la radiografía o el ultrasonido [11]. El estudio explora diferentes estrategias de ajuste de hiperparámetros y redefinición de clases para optimizar el rendimiento del modelo, alcanzando una precisión cercana al 75,4% [10].

Los hallazgos obtenidos evidencian el potencial de este enfoque para mejorar los procesos de control de calidad en soldadura, reduciendo la dependencia de inspecciones visuales manuales. La siguiente sección describe la metodología empleada para alcanzar estos resultados.

3. Métodos

Redes neuronales convolucionales

Las CNNs son un tipo de arquitectura de aprendizaje profundo diseñada para el procesamiento de imágenes. Estas redes se han utilizado ampliamente en la visión por

computadora debido a su capacidad para aprender automáticamente representaciones jerárquicas de características a partir de imágenes de entrada [11]. El funcionamiento de una CNN se basa en varias capas fundamentales:

- *Capas Convolucionales:* aplican filtros (*kernels*) sobre la imagen de entrada para extraer características relevantes, como bordes, texturas y patrones más complejos a medida que se avanza hacia las capas más profundas.
- *Capas de Pooling:* reducen la dimensionalidad de las características extraídas, mejorando la eficiencia computacional y permitiendo que la red generalice mejor.
- *Capas Completamente Conectadas:* interpretan las características aprendidas y generan una salida final, ya sea una clasificación o una detección de objetos.
- *Funciones de Activación:* introducen no linealidad en la red para que pueda aprender representaciones más complejas.

Para la detección de defectos en soldadura, se utilizaron CNNs entrenadas con imágenes previamente etiquetadas, lo cual permite al modelo aprender a distinguir entre soldaduras aceptables y no conformes. En este contexto, la arquitectura YOLO constituye el núcleo del sistema propuesto por su eficiencia y precisión en tareas de detección en tiempo real. YOLO integra una CNN profunda como extractor de características, capaz de procesar la imagen completa y extraer patrones jerárquicos desde bordes y texturas hasta formas más complejas a lo largo de sus múltiples capas convolucionales [6]. Estas características se utilizan luego en una capa final para predecir directamente, las coordenadas de los objetos, la categoría asignada y la puntuación de confianza. En

la figura 1 se ilustra la arquitectura de una CNN YOLO.

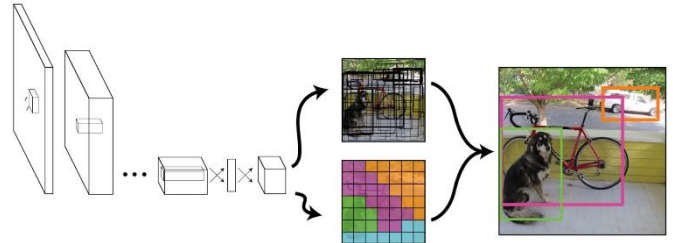


Figura 1. Arquitectura de una CNN de YOLO. [12]

A nivel operativo, YOLO divide la imagen en una cuadrícula, y cada celda predice múltiples cajas delimitadoras (*bounding boxes*) y sus etiquetas asociadas. Esta estrategia convierte el problema de detección en una tarea de regresión directa, lo que permite un procesamiento altamente eficiente. En resumen, el modelo YOLO detecta objetos en una imagen en un solo paso (*one shot*), dividiendo la imagen en celdas y prediciendo múltiples objetos simultáneamente. Es rápido y eficiente, ideal para aplicaciones en tiempo real.

En contraste, los modelos de dos etapas, como R-CNN (*Regions with CNN features*) primero generan propuestas de regiones (posibles ubicaciones de objetos) y luego clasifican cada una usando una CNN, lo que incrementa significativamente el tiempo de inferencia [13]. Gracias a su diseño unificado, YOLO resulta ideal para aplicaciones industriales donde se requiere detectar múltiples objetos en tiempo real, como en la inspección visual automatizada de soldaduras [14, 15].

Diseño del experimento

Para el desarrollo de la investigación se utilizó un conjunto de datos publicado en el repositorio Kaggle, denominado “Welding Defect – Object Detection” [16]. El conjunto de datos se encuentra en formato YOLO y está dividido en 2 versiones, cada una con imágenes de soldadura de piezas metálicas separadas en tres grupos: entrenamiento, validación y prueba. Cada carpeta contiene las imágenes y sus etiquetas correspondientes. La versión 1 del conjunto de datos cuenta con 827 imágenes de entrenamiento, 176 de validación y 73 de prueba, mientras que la versión 2 tiene 1.597, 279 y 124 imágenes respectivamente.

Las etiquetas están almacenadas en archivos de texto que contienen un *ID* de clasificación (0, 1 o 2) junto con las coordenadas normalizadas de la caja delimitadora para cada detección presente en la imagen. Los *IDs* corresponden a: 0 –soldadura mala, 1 – soldadura buena, 2 – defecto. El nombre de cada archivo de texto de etiquetas coincide con el nombre de la imagen a la que pertenece.

Los conjuntos de datos de entrenamiento presentan las siguientes proporciones de etiquetas por clase. La versión 1 contiene 731 de soldadura mala, 1.109 de soldadura buena y 1.279 de defectos, mientras que la versión 2 tiene 1.076, 1.890 y 1.559 respectivamente. Cabe resaltar que una sola imagen puede contener múltiples defectos o regiones de interés etiquetadas.

Para evaluar el desempeño del sistema propuesto, se realizaron experimentos utilizando ambas versiones del conjunto de datos y dos modelos base: YOLOv8m.pt y YOLO11m.pt, los cuales son modelos preentrenados para detección de objetos.

Ambos fueron reentrenados mediante transferencia de aprendizaje. Este enfoque ha demostrado ser efectivo en aplicaciones industriales al permitir la adaptación de modelos preentrenados a tareas específicas con conjuntos de datos limitados [5].

En este estudio se plantearon tres fases experimentales secuenciales, cada una con un objetivo concreto. Todos los experimentos se ejecutaron en un entorno con GPU (NVIDIA Tesla T4) utilizando la plataforma Google Colab, lo que permitió reducir significativamente los tiempos de entrenamiento.

Fase 1: Resultados del entrenamiento inicial

En la primera fase experimental se entrenaron los modelos YOLOv8m y YOLO11m utilizando tanto la versión 1 como la versión 2 del conjunto de datos dando lugar a cuatro experimentos:

- Modelo YOLOv8m y conjunto de datos 1
- Modelo YOLOv8m y conjunto de datos 2
- Modelo YOLO11m y conjunto de datos 1
- Modelo YOLO11m y conjunto de datos 2

Todos los experimentos de esta fase se llevaron a cabo bajo los siguientes hiperparámetros iniciales: 150 épocas, tamaño de entrada de imagen de 640×640 píxeles, *batch size* de 16, tasa de aprendizaje inicial (lr_0) de 0,005, *momentum* de 0,937 y *weight decay* de 0,0005. Se aplicó un esquema de calentamiento (*warmup*) de 5 épocas, se incorporaron técnicas de aumento de datos (*data augmentation*) como *mosaic*=0,8 y *mixup*=0,3, y se ajustaron los coeficientes de la función de pérdida (*loss*) (*box*=0,15, *cls*=0,35,

iou=0,3) para equilibrar la contribución de las distintas tareas (localización vs clasificación).

Fase 2: Mejora del modelo con ajuste de hiperparámetros

El segundo experimento consistió en refinar el modelo con el fin de mejorar la detección de la clase defecto mediante un proceso de ajuste de hiperparámetros a partir de validación cruzada. En esta fase 2, el enfoque se centró exclusivamente en el modelo YOLO11m.pt entrenado sobre la versión 2 del conjunto de datos ya que esta combinación mostró resultados superiores. Se redujo el número de épocas de entrenamiento de 150 a 100 para mejorar la generalización y evitar el sobreajuste. Asimismo, se aumentó la tasa de aprendizaje inicial de 0,005 a 0,01 lo que facilita una convergencia más rápida. Adicionalmente, se redujo el período de *warmup* de 5 a 3 épocas acelerando la adaptación inicial del modelo.

También se ajustaron las ponderaciones de la función de pérdida y se refinaron las técnicas de *data augmentation* para mejorar la detección de defectos. En la función de pérdida, se asignaron nuevos pesos a sus componentes (box=0,05, cls=0,5, IoU=0,2), mejorando la precisión de la predicción de cajas delimitadoras (*box loss*), mientras que las pérdidas de clasificación (cls) e IoU (*Intersection over Union*) se calibraron para mantener un equilibrio adecuado en el aprendizaje mejorando la diferenciación entre clases. Por otra parte, se ajustaron las transformaciones de aumento de datos estableciendo mosaic=1,0 y mixup=0,5, favoreciendo la generalización del modelo ante variaciones de fondo y objetos [15].

En conclusión, los ajustes fueron dirigidos a mejorar la convergencia del modelo, evitar sobreajuste, incrementar la variabilidad en los datos de entrenamiento y optimizar la precisión en la clasificación y detección de clases. Estos cambios permitieron que el modelo YOLO capturara mejor las características relevantes de los defectos en soldadura, resultando en una mejora de rendimiento.

Fase 3: Resultados con clases fusionadas

Luego de una inspección del conjunto de datos (ver figura 2) se pudo evidenciar un traslape considerable entre las clases soldadura mala y defecto, lo cual puede generar un impacto negativo en el rendimiento del modelo. En este sentido, en la tercera fase de los experimentos se decidió fusionar las clases soldadura mala y defecto en una sola llamada soldadura no conforme. Concretamente, en la versión 2 del conjunto de datos se eliminaron las etiquetas de la clase defecto (tipo 2) y se unificaron con la clase soldadura mala (tipo 0). En la práctica, si una imagen contenía únicamente etiquetas tipo 2 (defectos aislados), se conservó solo la anotación con la caja delimitadora de mayor área y esa región se reclasificó como tipo 0 (es decir, como soldadura no conforme). De este modo, el modelo ya no distingue entre un defecto puntual y una soldadura globalmente mala: ambas situaciones pasan a etiquetarse como soldadura no conforme (clase 0), frente a la soldadura buena (clase 1). Utilizando este conjunto de datos modificado, se reentrenó el modelo YOLO11m con la misma configuración de hiperparámetros empleada en la fase anterior. En la figura 2 se ilustran ejemplos de imágenes con etiquetas superpuestas.

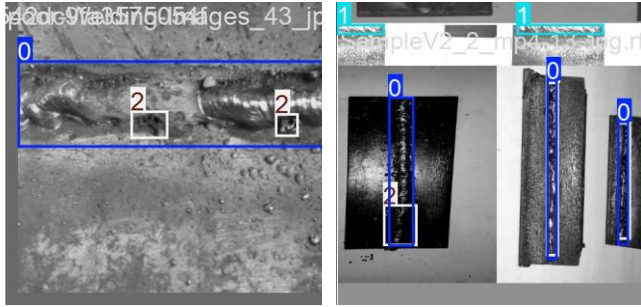


Figura 2. Imágenes con etiquetas superpuestas en el conjunto de datos de entrenamiento.

4. Resultados y Discusión

A continuación, se describen los resultados obtenidos en cada una de las fases experimentales descritas en la sección anterior.

Fase 1: Resultados del entrenamiento inicial

En esta primera fase se entrenaron los modelos YOLOv8m y YOLO11m, cada uno con ambas versiones del conjunto de datos (DS1 y DS2). En la tabla 1 se comparan los resultados de las diagonales de las matrices de confusión normalizadas obtenidas para cada combinación.

Tabla 1. Métricas de precisión por clase, calculadas a partir de las matrices de confusión normalizadas.

Métrica / Modelo	YOLO8 DS1	YOLO8 DS2	YOLO11 DS1	YOLO11 DS2
Soldadura mala	0.66	0.71	0.55	0.73
Soldadura buena	0.72	0.86	0.71	0.86
Defecto	0.44	0.53	0.41	0.49

Al comparar los resultados de ambos modelos y conjuntos de datos, YOLO11m.pt mostró métricas superiores a YOLOv8m.pt en todos los casos, y la versión 2 del conjunto de datos arrojó resultados significativamente mejores que la versión 1. Este último hallazgo se atribuye al mayor volumen de imágenes y a una mejor distribución de clases en la versión 2, lo cual se alinea con estudios que resaltan la importancia de conjuntos de datos balanceados y representativos para el éxito de los modelos de aprendizaje profundo [8].

Los resultados fueron evaluados principalmente mediante la matriz de confusión de la figura 3, correspondiente al mejor modelo (YOLO11m) entrenado con la versión 2 del conjunto de datos.

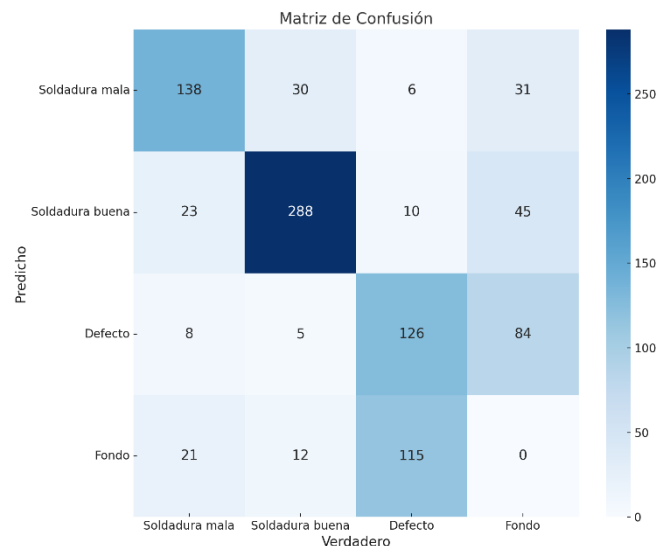


Figura 3. Matriz de confusión del modelo YOLO11m entrenado con la versión 2 del conjunto de datos (fase 1 de entrenamiento).

Se observan las tres clases originales: soldadura mala (0), soldadura buena (1) y defecto (2). Los valores en la diagonal indican el número de ejemplos correctamente clasificados por clase; fuera de la diagonal aparecen los números de confusiones entre clases.

En las imágenes de prueba se identificaron los principales errores cometidos por el modelo en esta primera fase. Las limitaciones más relevantes incluyen confusión entre clases (defectos clasificados como fondo) y omisión de defectos pequeños, lo cual refleja la dificultad del modelo para aprender a distinguir correctamente entre las clases. En la figura 4 se ilustran ejemplos de detecciones erróneas del modelo en esta fase, donde se observan detecciones superpuestas de los tres tipos de etiquetas.

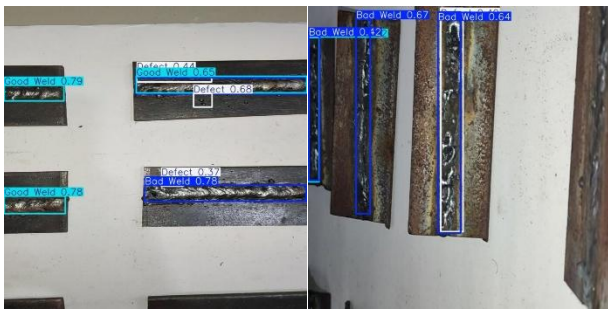


Figura 4. Ejemplos de detecciones con el modelo entrenado en la fase 1. Se observan etiquetas de los 3 tipos superpuestas.

En general, el modelo logró una detección robusta de la clase soldadura buena, con 288 predicciones correctas de un total de 335, lo que representa una exactitud cercana al 86%. En la clase soldadura mala, se detectaron correctamente 138 instancias sobre 190 en total para una exactitud del 73%, aunque se presentaron errores cruzados significativos con soldaduras buenas (31 instancias fueron mal clasificadas) y 21 casos se confundieron con el fondo.

En cuanto a la clase defecto, se alcanzaron 126 aciertos sobre 257 instancias reales, lo que

equivale a una tasa de detección de aproximadamente 49%. No obstante, este desempeño estuvo limitado por una alta tasa de omisión: 122 defectos fueron completamente pasados por alto, al ser clasificados como fondo.

En conjunto, estos resultados muestran que el modelo se desempeñó bien en la clasificación de soldaduras buenas, tuvo un rendimiento aceptable en soldaduras malas, pero aún enfrentó importantes desafíos en la identificación precisa de los defectos. La alta confusión entre defectos y fondo, junto con errores cruzados entre las clases activas, sugiere que el modelo necesitaba ajustes adicionales para mejorar su capacidad de generalización y discriminación. Por tanto, se procedió a realizar una segunda fase experimental, esta vez utilizando una configuración de entrenamiento ajustada.

Fase 2: Mejora del modelo con ajuste de hiperparámetros

Para esta segunda fase se utilizó el modelo YOLO11m con la versión 2 del conjunto de datos, ya que esta combinación arrojó mejores métricas durante la fase 1, además se ajustaron los hiperparámetros de entrenamiento con el fin de mejorar la detección de la clase defecto. Este entrenamiento ajustado mostró una mejora en la detección global del modelo. Sin embargo, persistieron errores en la clasificación específica de la clase defecto. La figura 5 muestra la matriz de confusión normalizada correspondiente a esta segunda fase experimental.

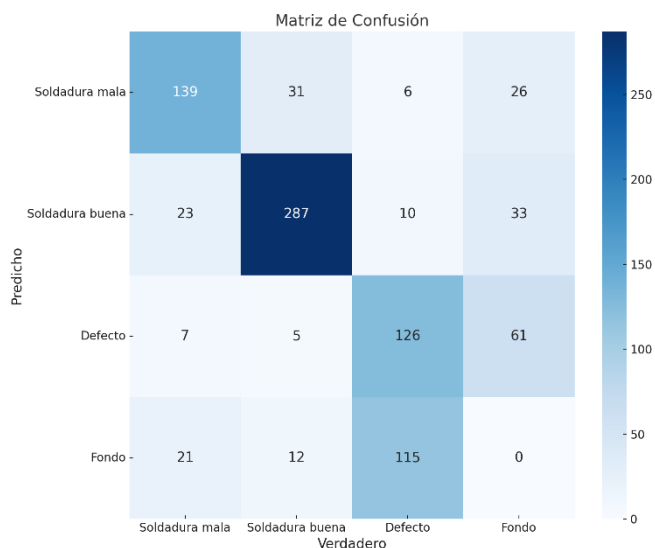


Figura 5. Matriz de confusión del modelo YOLO11m entrenado con la versión 2 del conjunto de datos, segunda fase de entrenamiento (tras ajuste de hiperparámetros).

A partir de la figura 5 se aprecia que el modelo alcanzó un desempeño destacado en la detección de soldaduras buenas, con 287 aciertos sobre un total de 335 instancias reales, lo que representa una exactitud del 85.7% (Igual que en la fase anterior). Esta clase fue la más confiable para el modelo, aunque se observaron errores de clasificación hacia soldadura mala (31), defecto (5) y fondo (12).

En la clase soldadura mala, el modelo logró 139 aciertos sobre 190 casos reales, para una exactitud aproximada del 73.2% que representa una mejora leve respecto a la fase 1. Los errores se distribuyeron principalmente hacia soldadura buena (23 instancias) y fondo (21), lo cual indica cierta ambigüedad en la diferenciación entre tipos de soldadura.

En cuanto a la clase defecto, se identificaron correctamente 126 de 257 instancias reales, lo que representa un desempeño del 49.0%. Esta clase sigue siendo la más difícil de detectar: 115 defectos fueron clasificados erróneamente como fondo, mientras que otros errores menores se dieron hacia soldadura mala (6) y buena (10). La elevada confusión con el fondo pone en evidencia la dificultad del modelo para detectar anomalías sutiles o pequeñas en el cordón de soldadura. En las imágenes de ejemplo de la figura 6 se observan a la izquierda las etiquetas de validación y a la derecha las predicciones. Observamos que de la clase defecto hay 10 etiquetas de las cuales detectó únicamente 3, mientras que detectó erróneamente como soldadura mala una imagen que solo contenía defectos.

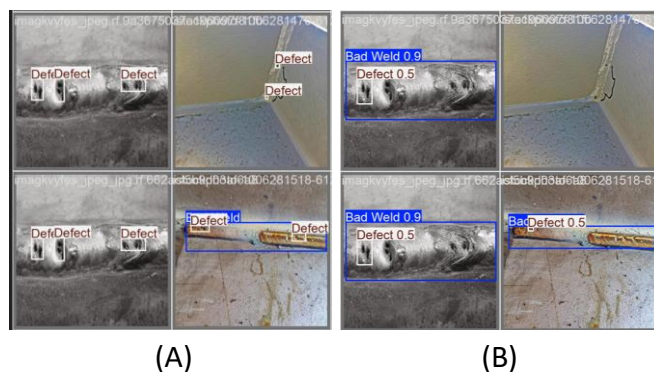


Figura 6. Ejemplos de detecciones con el modelo entrenado en la fase 2. La figura (A) muestra las etiquetas y la (B) muestra las predicciones.

Una observación relevante es que esta confusión entre defecto y soldadura mala se debe en parte a la forma en que fueron etiquetadas las imágenes de entrenamiento: en muchos casos, las mismas regiones que contienen un defecto también están

etiquetadas simultáneamente como soldadura mala, generando una superposición que dificulta el aprendizaje discriminativo del modelo. Este tipo de inconsistencia en el etiquetado ha sido identificado como una fuente importante de error en sistemas de visión artificial [17]. De hecho, la figura 2 ilustra un ejemplo de etiquetas superpuestas para las clases defecto y soldadura mala en una misma imagen, evidenciando el problema.

Se planteó que fusionar las clases defecto y soldadura mala en una sola categoría de soldadura no conforme podría mejorar el desempeño del modelo. Esta simplificación refleja la realidad del proceso de inspección, donde cualquier irregularidad implica que la soldadura requiere intervención. Además, permite reducir el ruido causado por etiquetas superpuestas. Para validar esta hipótesis, se llevó a cabo una tercera fase experimental.

Fase 3: Resultados con clases fusionadas

El efecto positivo de la fusión de clases se evidenció de inmediato en los resultados. La figura 7 muestra la matriz de confusión resultante de este entrenamiento final, con las categorías soldadura buena, soldadura no conforme (resultado de fusionar soldadura mala y defecto) y fondo.

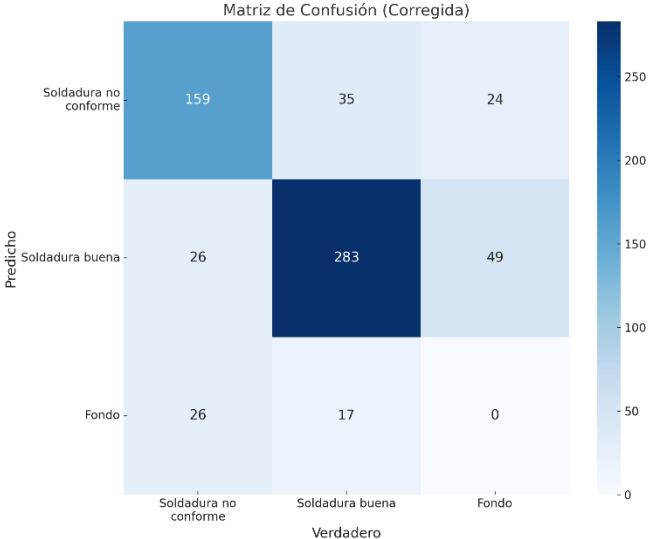


Figura 7. Matriz de confusión del modelo YOLO11m entrenado en la fase 3 con clases fusionadas.

Los números representan el conteo de predicciones en cada categoría. Por ejemplo, el modelo predijo correctamente 159 regiones como soldadura no conforme lo cual equivale a un 75,4% y 283 como soldadura buena representando un 84,5%. Las celdas fuera de la diagonal principal reflejan las confusiones: casos de soldaduras no conformes clasificadas como buenas (35 instancias) y viceversa (26 instancias). Se observa un marcado incremento en las detecciones correctas de soldaduras no conformes en comparación con las fases previas.

Los resultados cuantitativos del modelo final resultan satisfactorios. El modelo muestra un equilibrio sólido entre precisión y *recall* al detectar soldaduras no conformes (incluyendo defectos) versus soldaduras buenas.

Los siguientes ejemplos ilustran el buen desempeño del modelo YOLO11m entrenado en la fase 3. La fusión de clases mejoró

significativamente la capacidad del modelo para distinguir entre soldaduras buenas y no conformes, como se evidencia en las detecciones precisas mostradas en la figura 8.



Figura 8. Ejemplos de detecciones con el modelo entrenado en la fase 3 (clases fusionadas).

Comparación de resultados entre fases experimentales

Para evaluar el desempeño del modelo en las distintas fases experimentales, se realizó un análisis comparativo basado en la precisión de detección por clase y la confusión entre categorías. En la tabla 2, se presentan los resultados obtenidos en cada fase, lo que permite visualizar la evolución del rendimiento del sistema tras los ajustes en los hiperparámetros y la redefinición de clases. Esta comparación ayuda a identificar mejoras clave en la detección de soldaduras no conformes y el impacto de la estrategia adoptada en la fase final

Tabla 2. Métricas de precisión por clase, calculadas a partir de las matrices de confusión de cada fase experimental.

Métrica	Fase 1: Entrenamiento inicial	Fase 2: Ajuste de hiperparámetros	Fase 3: Clases fusionadas
Precisión Soldadura Buena	85,9%	85,7%	84,5%
Precisión Soldadura No Conforme	72,6%	73,2%	75,4%
Precisión Defecto	49%	49%	Fusionada
Confusión con el fondo	Alta (115 defectos clasificados como fondo)	Moderada (115 defectos clasificados como fondo)	Baja (Optimización del modelo)

La fusión de las clases defecto y soldadura mala en la fase 3 permitió reducir los errores de clasificación, mejorando la precisión global en la detección de soldaduras no conformes y aumentando el rendimiento general del modelo.

En síntesis, los principales hallazgos de las tres fases experimentales son:

- El modelo YOLO11m entrenado con la versión 2 del conjunto de datos ofreció el mejor rendimiento general en comparación con YOLOv8m y con la versión 1, corroborando la relevancia de emplear arquitecturas robustas y datos más abundantes y balanceados.
- La detección de soldaduras buenas resultó robusta desde las primeras fases, con altas tasas de acierto y pocos falsos positivos.

- La detección de defectos en la soldadura fue inicialmente limitada debido a problemas en la calidad del etiquetado y a la superposición de etiquetas entre las clases defecto y soldadura mala, lo que generó confusiones significativas en las fases 1 y 2.
- La fusión de las clases defecto y soldadura mala en una sola categoría de soldadura no conforme mejoró notablemente la precisión del sistema para identificar soldaduras defectuosas.

5. Conclusiones

Los resultados obtenidos en este estudio demuestran el potencial del aprendizaje profundo, y particularmente de los modelos basados en YOLO, para la detección automatizada de defectos en soldaduras a partir de imágenes fotográficas.

El modelo YOLO11m mostró un desempeño superior en comparación con YOLOv8m, y la utilización de la versión 2 del conjunto de datos más extensa y equilibrada permitió alcanzar una mayor precisión, destacando la importancia de contar con datos suficientemente representativos. Sin embargo, se evidenció que la detección de defectos específicos estaba limitada por la superposición de etiquetas entre defectos y soldaduras malas, lo cual dificultaba que el modelo distinguiera con claridad entre ambas categorías.

Para superar esta limitación, se fusionaron dichas clases en una sola categoría que representa cualquier condición de soldadura no aceptable. Esta simplificación mejoró significativamente el desempeño del modelo en la fase final.

El modelo resultante es capaz de detectar con alta fiabilidad las soldaduras que no cumplen con los

estándares de calidad, reduciendo a su vez las falsas alarmas en soldaduras buenas. En conclusión, la integración de técnicas de aprendizaje profundo para la inspección de soldaduras es viable y efectiva. La mejora lograda con la fusión de clases resalta la importancia de una correcta definición de categorías en problemas de visión artificial industriales.

El sistema propuesto, basado en YOLO11m entrenado con clases fusionadas, proporciona resultados precisos y robustos que lo acercan a un nivel adecuado para su implementación en entornos industriales reales. Esto abre la posibilidad de integrar modelos de detección de soldaduras no conformes en líneas de producción, habilitando la inspección automatizada en tiempo real. Una aplicación práctica de este sistema podría traducirse en un control de calidad más eficiente, consistente y económico, disminuyendo la dependencia de inspecciones visuales manuales y de métodos tradicionales costosos.

En futuros trabajos se podría extender la evaluación a diferentes tipos de soldadura y contextos industriales. Asimismo, sería importante explorar la integración de este modelo en sistemas de visión artificial en planta, con miras a consolidar su uso como herramienta de apoyo en la industria manufacturera.

6. Referencias

- [1] V. C. Kantumchu, et al, Automation in the Welding Industry: Incorporating Artificial Intelligence, Machine Learning and Other Technologies, Wiley, 2024, <https://doi.org/10.1002/97811394172948>
- [2] S. B. Block, et al, LoHi-WELD: A Novel Industrial Dataset for Weld Defect Detection and

Classification, a Deep Learning Study, and Future Perspectives, IEEE Access 12 (2024) 77442-77457, <https://doi.org/10.1109/ACCESS.2024.3407019>

[3] K. Nawaz, et al, Skin cancer detection using dermoscopic images with convolutional neural network. Sci Rep 15 (2025) 7252, <https://doi.org/10.1038/s41598-025-91446-6>

[4] S. Deepan, et al, Deep Learning Approaches for Signal and Image Processing: State-of-the-Art and Future Directions, IC3I 7 (2024) 468-472, <https://doi.org/10.1109/IC3I61595.2024.10829368>

[5] S. Kumaresan, et al, Transfer Learning With CNN for Classification of Weld Defect, IEEE Access 9 (2021) 95097-95108, <https://doi.org/10.1109/ACCESS.2021.3093487>

[6] J. Terven, et al, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS, Machine Learning and Knowledge Extraction 5 (2023) 1680-1716, <https://doi.org/10.3390/make5040083>

[7] Y. Zhao, et al, Grammar Induction from Visual, Speech, and Text, Artificial Intelligence 341 (2025) 104306, <https://doi.org/10.1016/j.artint.2025.104306>

[8] D. Ortiz-Perez, et al, Optimizing IoT Video Data: Dimensionality Reduction for Efficient Deep Learning on Edge Computing, Future Internet 17 (2025) 53, <https://doi.org/10.3390/fi17020053>

[9] L. Yang, et al, An Automatic Deep Segmentation Network for Pixel-Level Welding Defect Detection, IEEE Transactions on Instrumentation and Measurement 71 (2021) 5003510, <https://doi.org/10.1109/TIM.2021.3127645>

[10] K. Young-Min, et al, Bi-Directional Convolutional Recurrent Reconstructive Network for Welding Defect Detection. IEEE Access 9 (2021) 135316-135325, <https://doi.org/10.1109/ACCESS.2021.3116799>

[11] K. Kyeen, et al, Multi-Branch Deep Fusion Network-Based Automatic Detection of Weld Defects Using Non-Destructive Ultrasonic Test, IEEE Access 11 (2023) 114489-114496, <https://doi.org/10.1109/ACCESS.2023.3324717>

[12] J. Redmon, et al, You Only Look Once: Unified, Real-Time Object Detection, In Proceedings of the IEEE conference on computer vision and pattern recognition CVPR, (2016) 779-788, <https://doi.org/10.48550/arXiv.1506.02640>

[13] R. Girshick, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, IEEE Conference on Computer Vision and Pattern Recognition CVPR, (2014) 580-587, <https://doi.org/10.1109/CVPR.2014.81>

[14] A. Krizhevsky, et al, ImageNet Classification with Deep Convolutional Neural Networks, Communications of the ACM 60 (2017) 84-90, <https://doi.org/10.1145/3065386>

[15] A. Zhang, et al, Dive into Deep Learning, https://d2l.ai/chapter_convolutional-neural-networks/index.htm, 2023

[16] S. Adhi, Welding Defect - Object Detection, <https://www.kaggle.com/datasets/sukmaadhijaya/welding-defect-object-detection/data>, 2024

[17] Z. Gao, et al, Computer Vision -- ECCV 2022: Learning from Multiple Annotator Noisy Labels via Sample-Wise Label Fusion, 24, Springer Nature Switzerland, 2022, https://doi.org/10.1007/978-3-031-20053-3_24